DTIC FILE COPY

②

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

Technical Evaluation Report

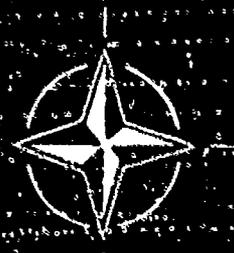# Electronic Transfer of Information and its Impact on Aerospace and Defence Research and Development

Abstracts

Paper Reprinted from

AGARD Conference Proceedings No.466

1990

DTIC
ELECT
AUG 23 199
B

NORTH ATLANTIC TREATY ORGANIZATION

90 08 20 150

PREFACE

This Technical Evaluation Report was prepared at the request
of the Technical Information Panel of AGARD by

> Mr. Carlo Vernimb
> Information Services Consultants
> Terrestlaan 20
> B-1900 Overijse
> Belgium

It has been published in the Proceedings of the Symposium,
"Electronic Transfer of Information and its Impact on
Aerospace and Defence Research and Development". AGARD
Conference Proceedings, CP-460, in March 1990.

This reprint has been prepared by AGARD to give a wider
distribution to the report.

CONTENTS

TECHNICAL EVALUATION REPORT

by

Carlo Vernimb
Information Services Consultants
Terrestlaan 20
B-1900 Overijse
Belgium

SUMMARY

The report summarises 16 papers (including discussions) presented at the AGARD Technical Information Panel (TIP) Specialists' Meeting on "Electronic Transfer of Information and its Impact on Aerospace and Defence Research and Development" which was held from 17th—19th October 1989 in Brussels, Belgium; and comments on the state of the art of the technologies presented, discusses their possible introduction into scientific and technical organisations, and provides recommendations on the use of technologies with emphasis on the aerospace and defence community. The following topics were addressed: Technologies for electronic transfer of information, electronic storage and delivery, electronic publishing and communication, applications to the aerospace and defence R & D community. (SDW) kr

1. INTRODUCTION

The AGARD Technical Information Panel (TIP) Specialists' Meeting on "Electronic Transfer of Information and its Impact on Aerospace and Defence Research and Development" was held (with 182 participants) from 17th—19th October 1989 in Brussels, Belgium. Its theme was defined as follows:

"The technology for electronic transfer of information is developing rapidly; electronic publishing, electronic storage, processing and delivery of information including text and images are all operational or feasible. Innovations in tele-communication networks play an important role in the electronic transmission of information. Applications of advanced information technologies impact the production, transfer and use of scientific and technical information as well as the communication and working conditions in the international scientific community."

The meeting was broken down into 4 Sessions and a Forum Discussion:

Session I	— Technologies for Electronic Transfer of Information
Session II	— Electronic Storage and Delivery: On-going Experiments and Operational Systems
Session III	— Electronic Publishing and Communication: On-going Experiments and Operational Systems
Session IV	— Applications to the Aerospace and Defence R & D Community
Forum Discussion	

The Evaluation Report consists of two main parts, a descriptive and a reflective part. The descriptive part contains summaries of the 16 papers presented, the discussions on the papers, the forum discussion and some points made by participants in response to a questionnaire. The descriptive part refrains as much as possible from comments by the rapporteur (the author of the Evaluation Report). Of course, summarizing means selecting items which the rapporteur believes are important. As a consequence, *indirect* comments cannot completely be avoided.

The reflective part contains the rapporteur's comments on the state of the art of the technologies presented, discusses their possible introduction into scientific and technical organisations, and provides recommendations on the use of technologies with emphasis on the aerospace and defence community. In contrast to the descriptive part, the reflective part expresses the personal opinion of the rapporteur.

2. SUMMARIES OF PAPERS AND DISCUSSIONS

Paper No. 1.

F.Mastroddi described the present state of electronic publishing in Europe and reports on several trends and actions in this field. The figures presented were derived from various reports and from the findings of the European Information Market Observatory, a panel of 500 observers (users and suppliers) set up by the Commission of the European Communities (CEC). The actions described are those which were or are being encouraged by the CEC. Information Technology: Partly owing to national (e.g. the Alvey Programme in the UK), intergovernmental (e.g. EUREKA) and European Community actions (e.g. ESPRIT, the European Strategic Programme for Research and development in Information Technology) European IT companies have significantly increased their own domestic market share from 33% in 1983 to 50% today. Further support (e.g. through ESPRIT II and EUREKA) for IT actions will be devoted to IT components, semiconductors and information processing systems.

Telecommunications

The European PTTs had revenues from public telecomms services of 53 billion ECU in 1987, and they invest 20 billion ECU annually. However, the US figures are twice as large and grow faster; e.g. the telecomms equipment market in Europe is

growing by less than 5% per year, compared with 8% in the USA. It is assumed that this discrepancy is partly due to the fragmentation of the European market which is intended to be overcome by the single market from 1992 on. In order to speed up the establishment of a single telecomms market the CEC pleads in favour of a liberalised regulatory environment ("Green Paper"), e.g. by means of Directives such as the one for the liberalisation of the supply of telecommunications equipment. Another measure is the stimulation of the construction, implementantion and use of advanced digital data networks and especially the coordination of the development of an ISDN infrastructure; the CEC recommends that the PTTs should plan their networks so as to provide 5 million ISDN subscriber lines by end 1992. Thirdly, the CEC encourages the development of IBC (Integrated Broadband Communications) with transmission rates of 0.5 to 2 Mbits/s as compared with 64 Kbits/s for ISDN. The implementation of IBC is planned to start in 1995. With its STAR Programme the CEC intends to ensure that no Member State will fall behind in the implementation and use of these new telecommunications.

Database development

A distinction is being made of

— online ASCII (bibliographic, numeric, factual, full text)
— online non-ASCII (videotex, teletext)
— local magnetic media (floppy disk, hard disk)
— local optical media (WORM, CD-ROM, videodisk, CD-I) and
— online optical media (image servers, remote jukeboxes).

There is already a substantial online market in Western Europe with a turnover of about 2 billion ECU, and it grows by approximately 20% per year. More than 80% of the turnover is generated by financial information services, notably Reuters. Of the worldwide 3240 online databases in 1987 only 27% originated in Europe, the bulk coming from the USA. In Europe 70% of the databases are funded by the public, in the USA only 25%.

As an example for the application of Digital Optical Disks (DOD) the TRANSDOC technology for storage of documents in facsimile form is mentioned. It is used e.g. by the European Patent Office and by the CNRS in Nancy (France) for many millions of pages. DOD systems are already linked to the telephone network for fax transmission and recently also to an ISDN network in France. For database publishing CD-ROM seems to have even better prospects than DOD. There are more than 300 commercial CD-ROM titles and 170,000 CD-ROM drives worldwide. The figures for Europe are 50 titles and 25,000 drives. In addition to full text, bibliographic/referral and numeric databases more and more mixed databases (text and graphics etc.) are put on CD-ROM. The role of in-house corporate publishing is likely to increase.

European Community initiatives in the information market

The IMPACT (Information Market Policy Actions) Programme of Directorate General XIII set up a European Market Observatory, overcomes technical, administrative and legal barriers to the flow of information, improves the synergy between public and private sectors, promotes the use of European information services, prepares actions in favour of libraries, and encourages pilot/ demonstration projects in the following areas: Patent information, image banks, intelligent interfaces, tourism information, information on standards. An IMPACT II Programme is in preparation.

Discussion

The discussion revealed interest in the figures presented and their re-use (Yanez). Interested people and organisations are invited to request from DG XIII Technical Reports (4 so far) issued by the European Information Market Observatory which was set up by the CEC and consists of a network of about 500 observers.

A question on broadband communications (Molholm) triggered the reply (Dunning) that there is a need for transmission rates of 2 Mbits/s which would require "backbone" transmission rates of 140 or even 540 Mbits/s. Each of the various potential user groups cannot afford to pay the rather high network costs alone. Discussions are under way between the CEC and the European Space Agency (ESA) on even higher transmission rates (Gigabits/s) which are also planned by the ARPA network in the USA for the end of the century. In the UK (Holmes) requirements for transmission rates of 2 Mbits/s are met by optical fibers which are already accessible at user premises e.g. in the London area.

Paper No. 2. 'The Present and Future in North America'

B.Unruh compared certain characteristics (economics, ease of distribution, market, and timing) of printing ("Gutenberg technology") with those of several new technologies: online, optical disks, knowledge software, videotex and gateways. The Information industry today is more driven by technology than by market forces. It is therefore important to look at the potential a technology has in meeting user requirements. In doing so, the young CD-ROM and gateway technologies seem to score better than the well established online technology. "Online systems and the products they offer, because of their high acceptance levels in certain communities and because of their failure to provide affordable services and easy-to-access systems, have been the impetus for the technologies that today are under development or being incorporated into the dissimination process". These technologies mainly aim at improving access to information: retrieval packages, knowledge software, faster telecommunications etc. They ought to address the end-user and not so much the skilled documentalist working in a library. Worth noticing is a trend towards clustering of information: Subject areas, such as medical or business information, and the corresponding databases (which should cover the subject area rather completely) are pre-selected so that marketing and user training can be focussed on better identifyable target groups of users thus allowing a better market penetration of the information products. Another interesting development is expected after the recent decision of the (often quoted) U.S. Federal Court Judge Harold Greene that the mighty AT&T is allowed, from 24 August 1989 on, to enter electronic publishing.

Discussion

On a question (Lawrence) whether AT&T will address the general consumer market place or the professional user community Unruh replied that if AT&T will be successful then for two reasons: choice of software for easy access to information and choice or production of databases which are of interest for the user.

On a question (Bullock) whether software for searching by end-users would be made available also to the scientific and technical community or whether ST laboratories would have to develop their own software the reply was that more and more software will be written "for you" than "by you", and the advice was given to keep close contact with software producers to make sure they will meet the user's requirements.

On a question on videotex (Chambaud) Unruh explained that in the USA videotex and gateways are used as near-synonyms, that Judge Greene was in favour of videotex after having seen a Minitel demonstration in France, however, that no videotex experiment in the USA has succeeded so far.

Dunning referred to the domestic (or consumer) market versus the professional market and wanted to know if there would be a future for service providers in addition to giants such as AT&T. Unruh felt that the potential unreached market is in the professional area because today's products have no appeal to domestic users, even if more artificial intelligence would be used. There is, of course, a big potential consumer market, but there is no product for it. A chance is seen for services which involve transactions (ticket ordering, home shopping etc.).

Paper No. 3. 'The Present State and Trends in Japan'

U.Wattenberg gave a full picture of electronic transfer of information in Japan. The huge number (6000) of Kanji characters caused many difficulties for the information industry: a Japanese-Japanese dictionary is needed for inputting characters via keyboard; the characters require high resolution screens etc. Having solved these problems Japan has turned such disadvantages into advantages: It is leading e.g. in the production of high resolution screens and fax machines. Together with China and Korea, Japan is establishing a 2-byte standard character set for up to 65000 characters. The Japan Information Center of Science and Technology (JICST) is one of the largest hosts and document supply centres in the world. The JICST file on science and technology e.g. grows by 700 000 documents per year, and the number of documents copied per year is 700 000 too. 100 000 documents per year are translated into English; machine translation systems are applied mainly for translations from English into Japanese. JICST hosts one of the STN nodes (Columbus, Karlsruhe, Tokyo) through which the JICST databases are worldwide accessible, those in English really, and those in Japanese "in principle". Of high importance are also the activities of the Japan Patent Information Organization (JAPIO) which cooperates with the US Patent and Trade Mark Office and the European Patent Office e.g. in establishing a worldwide image file of all patents since 1920. Instead of the traditional 16 mm microfilm and microfiche the JAPIO now distributes to its subscribers 12 cm CD-ROM including full text images and index data. There is a sophisticated retrieval system for trade marks. And patents can be applied now either online, or on floppy disk, or in printed form. Japanese newspaper companies now hold their journals in electronic form. There is a database with 500 000 photos on CD-ROM. Each day 500 new photos enter the database. Nippon Telegraph and Telephone Corporation (NTT) has not been split up and is, therefore, very powerful, in spite of some competition from other organizations. ISDN (64 Kbit x 2 channel) started last year; this year a 64 Kbit x 23 system has been added, enabling 1.5 Mbit picture transmission. Experiences with videotex ("Captain") were disappointing. The paper contains a wealth of detailed information including contact addresses.

Discussion

Japan seems to be very successful in many respects including database generation; but what about retrieval? (Lawrence).

Japan is excellent in hardware and must catch up in software. JICST provides multi-database searching; the trend is towards "one-button" searches, in particular for SDI.

Is Group 4 facsimile transmission widespread? Is ISDN generally available? What is done for transforming text images into character codes? (Dunning).

Group 4 facsimile transmission (1 second per page) requires special lines (the normal telephone network not being sufficient); these exist between premises of big companies and are used for Group 4 ... larger and use ISDN is needed which starts just now in Japan. There are optical character readers for Japanese texts, but re-typing is still cheaper.

Questions on the information flow from and to Japan (Chambaud, Yanez, Vanautryve) yielded the following reply:

Japan is with 10% the third largest producer of scientific and technical information (after the US and the UK), however, there are only a few users of information in Japanese outside Japan. The situation is better for the subset of information translated into English by Japanese translators. Japan puts much effort in making their information available to the West. The West should put more effort than now into helping Japan in selecting information for translation and into using Japanese information. A comment from the floor (Lawrence) confirmed that there is virtually no use of Japanese databases in the USA.

Paper No. 4. 'Innovations in Telecommunications Networks'

C.Holmes of British Telecom (BT) reported on innovations in telecommunication networks. With regard to telecomms BT is a user, a vendor of networks and components, a systems integrator, a conformance tester, and a standard maker. Of main interest to the STI community are, of course, the services provided by BT. In this respect BT decided (1) to run its telephone services properly, (2) to go into new services such as value added networks, mobile communications ("cellular radio") for even the transmission of images, and (3) to go international. With deregulation in the UK, competition in the value added services marketplace, and the advent of the single European market in 1992 BT has reacted: The UK market alone is not viable for certain services; a pan-European market is aimed at. The Dialcom Public Services are available already in 23 countries and with 400,000 users; Tymnet will be added. Applications of the X.400 protocol are being developed: Electronic Data Interchange

(EDIfact for international fund transfers), Office Document Architecture (ODA), formatted messaging activities and user interface architectures. X.25 PSPDN commenced operation 8 years ago and has links to 100 other networks in 80 countries. BT offers an OSI infrastructure Local Area Network (LAN) and commenced an ISDN trial in Spring 1985. "The agreement of the European standard for ISDN later this year will provide a sound basis for investment for both PTT and multinational users alike". More and more telecomms developments are user driven.

Discussion:

Which mechanisms are used and should be used to put pressure on PTTs to react to requirements of users of scientific and technical information? (Lawrence).

Such requirements ought to be focussed.

How will optical fibers be brought to the workplace? (Molholm).

At present fibers are brought "to the door" not to the workplace. For buildings this would mean that the vertical infrastructure would be made with fibers and the horizontal distribution with copper wires. In new buildings plastic tubes should be installed in walls through which bundles of unshielded optical fibers could be blown when needed. Very soon fibers should go on to the desk.

Paper No. 5. 'Megatrends'

S.Chambaud had the courage to predict the future of electronic information transfer in form of 5 "megatrends":

1. Internal information originated in an organization will be merged with external information available from commercial services. This process will be facilitated by technologies such as desk top publishing (internal information), ISDN (external information), CD-ROM (storage of both types of information), and hypertext (user-friendly access to information) and could result in, what could be called, an "information machine" accessible even for the famous end-user.

2. More value will be added to information. There is progress from straightforward ("one-dimensional") text to multidimensional information including graphics, audio and moving images.

3. There will be vertical integration of documentary functions, such as bibliographic searches, localization, ordering and delivery of full textndocuments (e.g. to the local laser printer); and one might add ordering and delivery of automatic (raw) translations.

4. Commercial actors will integrate vertically and horizontally. These actors are the publishers, the database producers, the service providers, the intermediaries or information brokers, and the users. There are many examples of publishers and new actors in the information services market buying database producers and service providers (vertical integration); and there are horizontal acquisitions aiming at scale of economy and reduced competition. These developments will continue.

5. Systems will become more intelligent. This depends on research: with new computer systems imitating functions of the brain one can expect natural language understanding, automatic translations, intelligent character reading (ICR), speach synthesis etc.

From these 5 megatrends one can derive 2 hypertrends: There will be more integrated, intelligent and user-friendly information services, and there will be a viable information industry.

Discussion

Is there a tendency also to integrate the production of documents into the chain of documentary functions? (Beauvais).

Yes; the increased use of the Standard Generalized Markup Language (SGML) by authors and editors can serve as an example.

If one uses machine translation as an example for intelligent systems, then there was not much development over the last 20 years. How soon will intelligent machine translations be available? (Wattenberg).

We will have to wait for the new generation of computers. A comment from Yanez pointed to the fact that one already now speaks of a language industry and that results from more mature automatic translation systems such as Eurotra can be expected soon.

Paper No. 6. 'Optical Disc Systems'

A.de Ridder gave an animating description of three types of optical disc systems: ROM, WORM and Reversible all of which are complementary and fit in a range of applications.

ROM (read only memory)
CD — audio, 12 cm, 74 min hifi
CD — video, 12 cm, 5 min video (clip) + 20 min audio, hifi TV
LV — Laser Vision, 33 cm, 154,000 still frames, 90 min video + audio
CD — ROM, 12 cm, 654 Mbyte, MSDOS etc., text + bit map pictures

CD — ROM-XA, 12 cm, 654 Mbyte, text, audio, still + moving pictures
CD — I, 12 cm, real time oper. system, text + audio + still + moving pictures

WORM (write once read many)
5¼" WORM, 800 Mbyte
12" WORM, 2 Gigabyte

Reversible (under development), erasing and writing in one step,
could, when available, replace hard disc in PC.

The CD-ROM drive is more expensive than the CD-audio drive because it uses a more reliable (and expensive) motor. After error detection and correction the error rate is 10^{-16} compared with 10^{-2} for consumer audio. 275,000 pages (in character mode) can be stored on a CD-ROM. Program, index and information can be stored on the same disc.

Discussion

Is multiple access to CD-ROM possible? (Brendreth).

It is indeed. Many PCs can be linked to many CD-ROM drives, e.g. in a UNIX environment, allowing multiple users to access multiple databases.

The response time for CD-ROM is shorter than for online.

How does copyright effect the use of CD-ROMs? (Stolk).

One can arrange that a program copied from a CD-ROM can be used only when disk drive and CD-ROM are accessible.

Paper No. 7. 'Document Delivery via ISDN or Satellite Networks'

A.Dunning began his presentation on document delivery via ISDN or satellite networks with some fundamental considerations on the characteristics of documents, document delivery systems, archives and document receiving terminals. Taking into account the various forms in which a document can be stored today, a document is defined as "a structured data set of which there is a lasting independent record for the purpose of transferring information or knowledge between human beings". A document delivery system comprises three components: a document store or archive, a means of delivery, i.e. a telecomms system, and a receiving station. Archives range from group 3 telecopiers to powerful PCs with at least 1 Mbyte RAM and a 20 MHz Intel 80386 or Motorola 68020, or better a 25—30 MHz 80386 or 68030 and 8 Mbyte RAM. A scanner and a printer are needed with selectable resolution (200, 300 and 400 pixels per inch) and one or more communications modules. Adequate software for the handling of requests and deliveries is not yet available. The receiving station would have to have characteristics which match the characteristics of the archive and a memory buffer. For electronic document delivery there are the telex, public switched telephone, circuit switched data, packet switched data, integrated services digital, and the satellite networks. Two main ISDN user/network access capabilities will be offered: basic access with 144 Kbits/sec and primary access with 1544 Kbits/sec in North America or with 2048 Kbits/sec in Europe. ISDN is closer than many users expect; and information and communication managers should now plan to use it. ISDN standards are well underway. The advantages of satellites for telecomms services are: some are already in orbit and operational, they have a wide geographical coverage, they can be used in broadcast and multicast mode of use; they suggest distance independent tariffs; and they permit high data rates (> 2 Mbit/sec). The APOLLO (Article Procurement On Line with Local Ordering) system is a "multi-hub unidirectional communication system" (several transmit and many receive only stations). It is designed to use the Satellite Multiservice System transponders of EUTELSAT 1 satellites with a user data transmission rate of 1.536 Mbits/sec. On the assumption that a 300 pixels per inch A4 page, compressed according to the modified READ algorithm (group 4 facsimile), contains about 1 million bits, approximately 90 pages per minute could be transmitted over the satellite channel. The microterminals will cost between $1 000 and $10 000. The price per page for the enduser will be in the order of $1.

Discussion

The Canada Institute for Scientific and Technical Information provides a rather large document delivery service (about 1/10 of the British Library's document supply service); with the costs for journal subscriptions going up Brendreth is absolutely sure that people will pay $1 or even $2—3 per page for fast document delivery; they already now pay a surcharge of $10 per document (about $1 per page) for delivery by telefax.

How can APOLLO services be integrated with existing facilities? (Lhullier).

Through an X.25 port connected to an APOLLO data station controller.

APOLLO offers the opportunity for multiple distribution of documents. Has the CEC considered the copyright problems which could arise? (Lawrence).

The CEC has recently issued a report with proposals on copyright. The unit dealing with these questions is aware also of the implications caused by new technologies.

Are there people in the audience who could report on experience with APOLLO? (Searle).

The Fachinformationszentrum Karlsruhe together with the German Patent Office and the Technische Informations-Bibliothek in Hanover was interested in APOLLO but is now looking ahead to use the forthcoming ISDN for document delivery. (Tittbach). The British Library Document Supply Centre was and possibly still is keen to pursue APOLLO (Dunning).

Are there similar plans in the United States? (Tittlbach).

In Japan there are plans to use TV satellite channels over night for document delivery. (Wattenberg).

The Wall Street Journal is using satellites for printing on both sides of the USA, and the Financial Times uses satellites for printing in Germany. There is a new generation of satellites with much more power which will permit the use of small receive stations perhaps even for data transmission to portable computers. (Dunning).

Paper No. 8. 'The Weapons Laboratory Technical Library — Automating with "STILAS"

B.Newton described STILAS, the Scientific and Technical Information Library Automated System operated by the United States Air Force Weapons Laboratory Technical Library. STILAS provides, in addition to the traditional library functions of circulation, serials control, acquisitions, and inventory control, the features of an integrated library system with gateway reference access to up to four remote databases simultaneously. STILAS runs on the UNIX 5.3 operating system implemented on a UNISYS 5000/95 supermini computer to which 16 workstations are linked. The retrieval software (in C) is based on the BRS/Search system. A comprehensive training package is available. The system permits endusers to search across a spectrum of remote databases while simultaneously searching local library files. For searching in up to four systems simultaneously the same command language is used; each search statement must be formulated only once; STILAS will translate it into the appropriate forms for each database being searched. There are 700 personal computers in the Weapons Laboratory outside the Library. Their users will be able to connect to remote databases through the STILAS host via a Local Area Network (LAN).

Discussion

A question on extra staffing required and another one on backfiling (Andrews) prompted the reply that all additional work had to be done by the Library's own permanent staff of 12 (and some co-op students) and that backfiling is aimed at covering the last five years.

The Redstone Scientific Information Center is going to use STILAS, which is felt to be a very good system, for their rather large collection of 3 million items; it is anticipated to extra employ a system administrator and an automation technician for this purpose; in the end STILAS will make savings (Bullock).

Staff resources becoming available due to avoiding duplicate cataloguing could be used for serving customers (7000 in the Air Force Weapons Laboratory) better than now (Newton).

A question on collective cataloguing (Yanez) prompted the reply that it is of advantage if specialised centers spread over the country would catalogue, abstract and index the material they know best and then exchange their products with the other centers.

Are there already end-users of STILAS, and what is done for training them? (Searle).

Many of the potential end-users of Weapons Laboratory are computer literate; to alert them to use STILAS orientation programmes with videocassettes and training packages are being prepared.

Should STILAS not be recommended for general use by NATO organisations? (Tittlbach).

STILAS could fulfill the required functions, however, people are needed to run such systems; and there is no commitment from the NATO headquarters to make the managerial staff, the technical staff and the information specialists available. (Molholm).

Paper No. 9. 'The Automation Plan of DC1 (Search) of the European Patent Office'

R.Bare reported on the Automation Plan of the Directorate General 1 (Search) of the European Patent Office (EPO) according to which 300 million DM will be spent over 8 years. The 500 examiners of the EPO need to have access to more than 20 million documents with an annual increase of 700,000 new documents, most of them patent literature, but also journal articles etc. There is also a need for searching tools using keywords or full text techniques that complement the use of the International Patent Classification (IPC) for searching. As a consequence three lines of action will be pursued: processing of textual information, processing of images, and personal systems. The main system for textual information is EPOQUE, an internal host computer loaded with internal (FAMIly, INVEntory, Classification) and extensively used external databases (DERWENT, INPADOC). EPOQUE permits, with one and the same query language, crossfile searching (among different hosts), cluster searching, online help, thesaurus management, search strategy saves, and downloading to personal files. It will require a mainframe power of 30 Mips and a disk capacity of 60 Gbytes to serve up to 500 users simultaneously. With regard to the processing of images, in the framework of the BACON project 125 million of pages of patent documents are being captured (in form of images) and stored on optical disks. 65 million pages are being processed by EPO, the rest by the US Patent and Trademark Office and the Japanese Patent Office in the framework of a trilateral agreement. The examiners will maintain on their PC's personal files with material downloaded from various sources and used as a starting point in patent examination. The expected benefits of automation will be gain in productivity, an increased quality of searching, and a higher motivation of people.

Discussion

Relevant documents could be missed in searching, possibly due to inadequate indexing, thus causing problems with patent infringement litigation. Which level of indexing is applied in order to avoid such problems? (Searle).

All documents entering the EPO are re-classified by specialists thus reducing the danger of by-passing a document to a minimum. Nevertheless, as on average 1300 documents have to be consulted per working day, there is the possibility of

overlooking a relevant document. In any case it is ruled that "the EPO cannot be held responsible".

There are at least 100 database management systems with quite similar functions. Are we not reinventing the wheel with systems such as EPOQUE? (Dunning).

Each user has specific requirements. When EPO called for proposals, none of the 8 bidders had the required software on the shelf; it had to be developed (by a consortium consisting of SARIN, Telesystemes and EPS) for 20 million DM. In a comment Molholm agreed with Dunning that there is a general problem with a multitude of similar systems. He felt that the functions should be broken up into at least three different parts: the ability to connect and to develop protocols; to know with which systems to connect; and what to do with the information in the environment into which it is delivered. Part of the problem could be solved by "intelligent gateway systems".

Tittlbach pointed to the novelty of EPOQUE of integrating access to external, internal and personal information and felt that many organizations could learn from it. He asked Bare to report on the policy of EPO on the provision of patent information to users outside the EPO. EPO's policy in this respect has been formulated in December 1988 by EPO's Administrative Council: The EPO is a producer of information, and the task of dissimination is left to the National Patent Offices (NPO) of the Member States of the European Patent Organisation. In case the NPOs will not dissiminate the information it can be made available to European information providers under conditions which would result in fair competion with existing information providers.

Paper No. 10. 'Desk-Top Publishing — What you need to know'

M.Taylor gave a comprehensive view on desk top publishing (DTP). DTP is the way many people can use some of the 'traditional' publishing skills. DTP can cope with different types of input (text, images etc.), is easier to read, more authoritive, more effective in communicating, uses less paper, and facilitates multiple and multimedia output. When the documentation of a device (e.g. an aircraft) outweighs the device itself and when the documentation needs frequent updating then DTP can save much money. The minimum requirements for DTP are a PC, a screen to show input or output, a printer (preferably a laser printer) to print the output, a keyboard, a mouse and some DTP software. In addition to creating pages the software should support document retrieval, document management, revision control, revision tracking, distribution, and distribution control. The system should accept texts from other devices such as networks, should permit optical as well as intelligent character reading, should accept images such as line drawings and photographs, and should permit the creation of graphics. Publishing standards and architectures such as the Standard Generalised Markup Language (SGML) and the Office Document Architecture (ODA) play an increasing role in DTP.

Discussion

How to convince management that a DTP system should be acquired? (Correia).

The first point would be to improve the documents to be published, the second would be cost savings; for several applications also the time factor is of importance (there are cases where the documentation was out of date when the hardware was shipped).

There are scientists using DTP as authors and also as editors of their own products. Do we want such a development? Would we not loose something? And what are the prospects of colour prints created by DTP? (Dimond).

Certainly technology should not dictate whether authors become their own editors; it should depend on the author's organisation. Organisations can demand their authors to adhere to certain rules such as SGML. DTP offers flexibility. For example speed of publishing can considerably be increased with DTP; e.g. the changes a technician makes in a technical drawing can instantly be made seen on the other side of the Atlantic. And a large set of heavy manuals can be replaced by a CD-ROM which then easily can be mailed. With regard to colour, the technology is available, however, the costs of using colour in DTP are still very high.

There are several examples of misuse of DTP (red letters printed on yellow paper, strange page numbering systems etc.) (Hart). Training in using DTP is absolutely necessary, in particular with regard to basic layout, use of typography, and mouse control.

Paper No. 11. 'Electronic Publishing with SGML'

M.Krüger described electronic publishing with the Standard Generalized Markup Language (SGML). SGML is a language which permits the description of the structure of a document; i.e. of the presentation of elements of documents. The elements could be texts, images etc.; the presentation refers to the location of the elements, to typography etc. If authors could be induced to apply SGML the publishing process could considerably be facilitated and speeded up. This happened already in certain environments: The US Department of Defense (DoD) established a plan to acquire, process and use logistic technical information in digital form and decided to apply the plan. This initiative is known under the name CALS (Computer-aided Acquisition and Logistics Support). It includes the application of SGML. DoD's suppliers of hard and software such as weapon systems are bound to deliver the accompanying information (documentation) in CALS format. This led, after a time of stagnation, to a more general acceptance and application of SGML. Defence departments in other countries seem to follow the DoD example. Also commercial publishers (Elsevier and Springer with 250 journal titles each) are using SGML.

Discussion

If authors are not using SGML could SGML be fed into the process later? (Lawrence).

No, the structure (SGML) must come first, before authors put information on paper or other media.

In order to take on board SGML, does a Ministry (or other organisation) have to have a policy to begin with? (Andrews).
Yes, CALS provides a good example.

How could we apply SGML to AGARD publications such as conference reports with up to 40 authors from quite different types of organisations? (Hart).

On the advice (Krueger) to have the AGARD proceedings published by a commercial publisher, Hart replied that this was tried and then turned down by national delegates; in the trial the publisher typeset the material; as a consequence the authors had to proof-read their documents again.

What is the relationship between SGML and the forthcoming ODA (Office Document Architecture) standard? (Holmes).

ODA is not directly related to publishing, even if a manuscript prepared according to ODA can be published. SGML is a language with which quite different items or events such as documents, a breakfeast and a dancing performance can be described. It is not competing with ODA, it can be interfaced with ODA.

Paper No. 12. 'Electronic Mail Systems'

In his paper on Electronic Mail Systems B. MAHON defined the most basic level of E-mail as a service providing a capability to write text messages into a system for delivery to other users in the same system. To this editing capabilities can be added and the possibility of distribution to a multitude of receivers. One of the principal difficulties with any messaging system is ensuring that it is used. There are various techniques to overcome these difficulties, however, none can guarantee usage. Personal computers which are more and more used for word processing can be linked for the exchange of texts. This type of E-mail depends on the interoperability between PCs and networks. The new CCITT/ISO standard X.400 will facilitate the interaction between previously incompatible electronic messaging systems. In a future scenario an electronic mail server (computer and software) which is available through the telecommunications can provide the following functions: message sending and receiving, telex sending and receiving, fax sending for all and receiving for some specially equipped devices, and voice messaging for all. All this is technically possible today but not yet as easy to use as today's methods and, therefore, not yet generally acceptable. Augmentation with ergonomic and functional features is required in order to overcome practices such as sending faxes after they have been created in word processors, printed and then fed into the fax machine.

Discussion

Molholm decided for his environment not to use E-mail now and for a while for two reasons: the necessary PCs with modems are not available to all associates and the discipline to really use the system could not be imposed. For the time being telefax was chosen instead.

Fax has generally been chosen as it is convenient; even texts produced with a PC can directly be fed into a fax machine. Could E-mail successfully be applied in conferencing systems? (Wattenberg).

There is only an artificial separation between E-mail and computer conferencing; and there are many examples for successful applications of these systems, often between different sites of the same organization. A problem with E-mail is that many potential users are not prepared to touch a keyboard. (Molholm).

Paper No. 13. 'Computer Conferencing'

In his paper J.Black described computer conferencing as an effective means of group communication without the limitations posed by real-time interactions (e.g. audio and video teleconferences) or physical location (e.g. face-to-face 'traditional' meetings). Computer conferencing permits "sharing the collective memory of a discussion between two, two hundred or two thousand participants who may be scattered around the globe... Participants can join a computer conference at their own convenience (in terms of time of day and physical location), read what has already transpired in the meeting, add their 'voice' to the discussion or contribute new thoughts and th{ } exit from the 'conference' to return at another convenient time." A wide range of possible and actual applications is listed which could have and partially has been listed already by B. Mahon in his presentation on electronic mail services, thus underlining the similarity of E-mail and computer conferencing. However, there are also differences between both these techniques: computer conferencing normally requires a moderator who must stimulate and guide the discussion. The paper was not presented by John Black but by his collaborator, Ellen M. PEARSON. She pointed to certain characteristics of computer conferencing such as chronological storage (who said what and when?), logical relationships of text (linking blocks of contributions by subject, in order to follow the train of thoughts). That participants from various language groups may find written text easier to understand than spoken words is seen as one of the many advantages of computer conferencing. Pearson gave a number of examples of applications of computer conferencing and concluded with J. Black's words that "it is a 'natural' for use by AGARD."

Discussion

If one starts with E-mail, is then additional software required for switching to computer conferencing? (Wattenberg).

Additional software is needed in particular for the chronological storage and retrieval of conference contributions and for the logical relationship of texts and their retrieval by means of free text searching.

Do you have particular examples of the use of computer conferencing in project management scenarios? (Hall).
The answer was "No".

In one of the examples computer conferencing was used as a teaching tool. Can it really replace face-to-face communication? (Andrews).

It is not a complete substitute for human interaction; it is an additional tool which saves some of the teacher's time.

British Telecom applies an electronic publishing package (which includes searching capabilities) on top of their electronic mail system and thus achieve computer conferencing. (Holmes).

You mentioned multilingual computer conferences. Do you apply automatic translation facilities for this purpose? (Chevalier).

"I only referred to various language and cultural groups which may find written text easier to understand than spoken words and that computer conferencing gives them the time for thinking and formulating their reply." There are systems for automatic translation (French to English translation was demonstrated last fall in Greece) which could be integrated into computer conferencing in a not too distant future. (Molholm).

A bit out of context the question was discussed if a signature on a document was legal when the faxed document was used as a proof.

Paper No. 14. 'The DOD "CALS" Initiative'

K.Molholm reported on the Computer-aided Acquisition and Logistics Support (CALS) initiative of the US Department of Defense (DoD). CALS "is directed toward improving the design, development, and support of weapons systems throu. ne use of current and emerging computer technology," and aiming at the use of "electronic transfer of information to the maxi m amount possible". The CALS initiative is a "strategy designed to create a system that can create, transform, store, reproduce, change, distribute, and use information as it evolves through the design, manufacture, maintenance, and logistics support of Defense weapons systems and equipment". In this initiative also the weapons manufacturing industries as suppliers of DoD are involved. DoD will not contract with suppliers who are unwilling to do business (e.g. submit engineering drawings, blueprints) in a paperless mode. The basic objectives of CALS are: reduced acquisition and support costs; improved quality and timeliness of technical information; and improved responsiveness of industry. As an example for cost savings it was mentioned that the US Air Force will save $135 million annualy just in the cost of updating manuals. Use will be made of CALS by the various military communities responsible for standardization, system design and development, maintenance, and logistics. Since many CALS standards have or will become US Federal Information Processing Standards (FIPS) they will also be used by other US Government agencies. Even other NATO member nations might adopt CALS. CALS is accelerating the acceptance and work on several emerging text and graphical standards, including the CCITT Group 4 raster image specifications, the Standard Generalized Markup Language (SGML) for texts and the Initial Graphics Exchange Standard (IGES) for drawings.

This paper was discussed at the Forum Discussion.

Paper No. 15. 'Transfert Electronique de l'Information: Applications dans la Société, "Aérospatiale"'

O.Lavroff described the applications of electronic information transfer in Aérospatiale, a French company with 33,000 employees, a turnover of 28 billion French Francs, and involved in the production of Concorde, Airbus and Ariane. A number of electronic information projects is being studied or implemented: a bibliographic database in the field of technics for internal (knowhow of company) and external (patents etc.) information which should obviate intermediaries; a database on macroeconomics; a database on internal and external standards; a database on multilingual terminology, stored on CD-ROM, as an aid for translation; technical documentation, stored on optical media, for use by the clients, in particular a maintenance information planning system; computer assisted aircraft trouble shooting; an aircraft documentation retrieval system (more than 200,000 pages text and images); an advanced project for European information exchange on technical documentation; and a system for automated online order processing, e.g. for ordering spare parts. With 70% of the production being exported, automatic translation of technical documentation is of utmost importance.

This paper was discussed at the Forum Discussion.

Paper No. 16. 'Information Technology Applications: a British Aerospace Military Aircraft Ltd. View'

K.Hall reported on information technology applications in the company British Aerospace Military Aircraft Ltd. with 131,000 employees and an annual turnover of several billion Pounds Sterling. The company is guided by three main principles:
  (1) a capability to design, develop, make and market complete vehicles and systems must be maintained;
  (2) a technical competitive edge must be sustained; and
  (3) there must be continuing effort to reduce unit costs.

In order to achieve these goals information technology is extensively applied. "A company's success in the future will be linked closely to its success in the exploitation of information technology and in particular data." As a key system in the company's overall business architecture PROMIS is used which stands for PROject Management Integrated System. It permits active management control and feedback capability as well as the ability to produce credible executive management reports. A noteworthy policy was applied in using PROMIS: "Users must be discouraged from becoming 'experts' in the system or its language, but encouraged to use the system as a tool for managing projects". PROMIS handles data, tables, drawings, barcharts, critical path networks, milestone achievements, budget spending reports, resource histograms, cost/achievement reports and graphs etc.; it is coded in the Metier Management Systems Artemis language and runs in an IBM mainframe environment; access by PC will be added. "A most useful feature of PROMIS is the ability to predict outturn of a project and impact on expected milestones and end dates. This feature assumes more significance in the modelling role ('what if?'), during which a copy of the master data is used and modified to analyse a particular set of circumstances (sudden reduction of resources, change

of work content etc.)". "The introduction of such a comprehensive system represents something of a cultural shock to an organisation,". However, "the benefits derived are numerous".

This paper was discussed at the Forum Discussion.

Forum Discussion.

Impact of CALS

What impact will CALS have on the data distribution of the Defense Technical Information Center (DTIC)? And will CALS impact on technical base reporting?

There will be no direct impact; only those parts which are standards, like SGML, will have an impact. On the other hand, more and more parts of CALS are likely to become standards which will then generally be applied. (Molholm).

CD-ROM

What problems do you foresee in distributing CD-ROMs? Speakers have discussed the technical issues of CD-ROM etc. as data stores; the Aerospatiale examples are in-house and new; the only other effective example quoted has been the DEC software distribution service; all other CD-ROM based services are almost unusable; what prospects are there that this situation will improve?

CD-ROM can replace paper in certain areas. DTIC has recently put 250,000 bibliographic citations of unclassified material on CD-ROM for evaluation of the retrieval system by users. With classified material the pressing would have to be done in-house; this is, of course, an economic question. (Molholm).

When one has to deal with classified and unclassified documents CD-ROM is not economic. (Bullock).

The first generation of CD-ROM products was rather poor. The second generation with software interfaces was considerably better; and we now face a third generation with much more interesting products than just bibliographic tools. Bibliographic tools with simple access found their place e.g. in universities, whereas the new CD-ROMs, e.g. with full texts, software and data related to a specific subject area, bring information to the fingertips of the user. (Lawrence).

Economics of management systems

What happens to the workforce profile when a system like PROMIS is installed?

We would have had to increase our (non-productive) workforce in project management had we not had an automated system like PROMIS. We maintained the workforce and achieved a greater throughput. (Hall).

Multilingual dictionaries

Is Aerospatiale's multilingual dictionary available?

It exists as an inhouse database. We have received a subsidy from the Ministry for Industry in France to put the dictionary on CD-ROM and to make it available to private and public aerospace organisations. It will be available early next year; however, it needs to be further improved, not only with regard to the content and presentation, but also with regard to the retrieval software. This will be achieved by cooperation with aerospace organisations in other European countries, and in nine months or so a version will be available which can serve as a valuable tool for translators. (Lavroff).

Computer conferencing

Which procedures are necessary for starting a computer conference?

A moderator or chairman, an agenda or a decision on topics to be discussed, an agreement for people to come to the meeting, and material to be distributed before the meeting, are needed to start up any conference. An additional requirement of computer conferences is to introduce the people to the medium, "so that the medium does not get in the way of the message". Technicians are needed to deal with the mechanics; the participants of a computer conference must be allowed to concentrate on the intellectual part. Computer conferencing is a tool which cannot completely replace face-to-face conferences. In fact, for the same group of people the first conference should be a face-to-face conference which can then be followed by computer conferences. (Pearson).

Fiber optics

Could we have more information on fiber optics; in particular when will all of London have fiber optics, and what about using fiber optics for military telephone systems?

London has fiber optics in (or underneath) most of the major streets, and the major customers have access to it. The whole of Westminster up to St. John's Wood and down to the river, and from Fulham to Hackney is covered in fiber; about thirteen miles by five. The whole of the UK is covered with fiber for the trunk-network and for the junction network which links the telephone exchanges together. In buildings fibers are first installed vertically and then, when required, horizontally to the users' desks. In new buildings plastic tubes should be mounted behind the plaster work. Bundles of fibers can then be blown with compressed air from the basement through the tubes to the desks. Also in the military environment fibers are looked at with favour as, due to the absence of plastics, the risks of fire can be reduced. Another favourable aspect is the absence of induction. (Holmes).

Role of the information intermediary

Taking into account all the technical developments we have heard of these day, what will be the future role of librarians and information specialists? (Newton).

There will not be much change for the professions, however their functions will probably change. Demands are expressed for end-user searching, however, the end-user is not a good searcher. On the other hand, the information specialist (the intermediary) is not really the expert on that what he retrieved, in terms of content. Both are needed, retrieval skill and subject knowledge, therefore the intermediary and the end-user should closely work together, if need should be through telecommunications. An example is Easynet: when an enduser runs into trouble with his search he can enter SOS, and a real life person will help. (Molholm).

If a database is well structured permitting access via a menu then intermediaries are not required. If this is not the case then the intermediary has to stay in permanent contact with the user. As there are not many structured databases there is a good future for intermediaries. (Lavroff).

The role of the information intermediaries will change only in so far as personal contacts with the end-user will increasingly be replaced by contacts via telecommunications. (Van Leeuwen).

The main problem, at least in a library environment, is input, not output. More effort must be put into input, e.g. the use of an online thesaurus for indexing. (Heaston).

My concern is that the message is lost in the media. Computer experts create systems providing general access to 'something', but somebody has to decide how to categorize these 'things'. Librarians have done this already before the birth of Christ. We categorize the world as pre-history and history. And history started with the work of librarians. They hold and categorized information. Input now becomes more and more automated but not categorization, not placing information in a logical place for retrieval. Categorization work will have to grow, otherwise information will be lost. (Molholm).

Coming back to the role of the intermediary, many databases are not directly accessible by end-users; their use requires intermediaries who are familiar with the new information technologies. The number of information broker offices in France is increasing. (Masson).

As the number of databases is increasing and the access procedures are becoming more complicated the information brokers have to become more competent. If they are, they become 'rich', and the rich become richer, the poor become poorer. (Larue).

Technical manuals

Do you envisage a complete change in the provision of technical manuals as presently provided by the Armed Forces (such manuals have to be provided at very low levels such as to an airman mending an aircraft or a soldier mending a lorry, and these people may not necessarily have a read out capability) ?

The whole area of technical manuals was one of the driving forces of CALS. In the long run it is envisaged that the data would be delivered to hand-held screens, with drawings, video images, sound etc. (Molholm).

Supercomputer highways in the USA

Who can provide information on plans in the USA to connect 1000 laboratories with high speed communication lines?

There are in fact plans by the Executive Office of the President and a proposal in Congress for a telecomms 'highway' for linking supercomputer systems between universities. (Molholm).

The National Science Foundation is the primary co-ordinating body behind it. (Rice).

The Hubble Telescope when brought into orbit will deliver about 100 Gigabytes of data every day which will be processed by Canadian scientists. This is the driving force behind the establishment of a high-speed network in Canada which largely is following the ARPANET in the USA. (Brendreth).

Another example are the data from global change: environmental data collected from all over the world. (Molholm).

Standard terminology for the composition of source texts

Traditionally there has been an interest in indexing tools such as categorization schemes and thesauri; with large computer memories there is a tendency to rely on full text searching. This may work in a relatively small user community. However, large database systems searched by a variety of users with different backgrounds and nationalities may be less successful if only full text can be applied. Another problem is the treatment of images as well as text. Standards such as SGML might help to extract bibliographic data directly from source documents; categorization schemes become more and more sophisticated with the introduction of cross-references, and differences between categorization schemes and thesauri are becoming smaller and smaller; computer assisted translation is also becoming more popular; however, all these tools may fail if the source document is not written correctly (spelling checkers, style checkers, glossaries are only a partial solution), but there could be tools to check a document and to suggest a standard terminology. By applying glossary tools the advantages of a thesaurus as a searching tool are automatically built into the source text. Is such a development possible, feasible or not very likely? Would we agree also that indexing and searching should be done more and more by originators and end-users?

There is not one answer. Indexing will be required to identify a unity of concepts across a collection of documents because language changes over time; however, there will be tools (e.g automatic indexing aids) to help in these processes. The same goes for retrieval; there is and will be software for helping the end-user. (Lawrence).

We are experimenting with full text. Assistance is needed on two levels: for the novice user and for the experienced searcher. Some assistance is needed, but we must be flexible and look for new tools. e.g. the concepts used for automatic translation could be very useful for automated searches. Most of the tools will have to be automated in order to make the information systems practicable.(McCauley).

There are quite sophisticated systems for the storage and processing of information. Should we not put more effort in the preparation of texts from the start. Automatic translations are wasted if the source texts are poor. Should we not apply standard

structures and standard terminology in order to achieve better texts which then could be used for automatic capturing, indexing etc., instead of accepting any garbage as input for our marvelous systems? (Ampt).

Perhaps this topic could be discussed at the TIP meeting next year in Norway, in particular the question whether authors should be given guidelines for composing their papers. (Yanez).

If you really want to get into a 'standardized authorship' situation you should not start with authors but at school. Equally I do not believe that it would ever be practical, and I hope it will not. (Hall).

Experience shows that authors do not follow instructions. As long as authors cannot be automated there is a need for an information specialist for indexing. (Heaston).

Computer viruses

What are the real hazards of computer viruses which might enter our systems via networks?

I think we make too big an issue of viruses. (Molholm).

There are anti-virus tools. We would make the biggest mistake if we would not use networks for fear of viruses. (McCauley).

Use of data collected by satellites

Which use is made of data collected by satellites and transferred to the earth?

Very little! Satellites such as Landsat receive and re-distribute to earth stations data at rates of several 100 Mbytes/sec. These data are useful for agriculture, forestry, preservation of natural resources etc., however, only a very small fraction is used. It is recommended, therefore, that much more effort is put into processing these data. (Colens).

We are looking more at the handling of the data after getting them on the ground; however, it is a massive problem. We have concentrated so far more on gathering the data and making sure we loose no single bit and we get it correctly. We now pay more attention to processing and re-distribution. (Rice, NASA).

Several years ago it seemed that, due to the legal status of NASA, Landsat data could be distributed only to the public, not to the private sector. (Brendreth).

Critically evaluated data.

Users in need of data, e.g. on laser damage assessment and laser hardening, need this data for immediate use. Data are spread over many sources, and even if their collection can be handled they are often contradictory. Therefore, a critical evaluation of these data is absolutely necessary. More should be done in this respect. (Cho Yen Ho).

The US Department of Defense has 23 Information Analysis Centers 13 of which are supported by the Defense Technical Information Center. They address the question of critical evaluation of data. (Molholm).

The more information is produced the greater is the need to prepare the information in such a way that it can be used directly and reliably. Computer assisted techniques providing access to information have to be improved to meet better and better the users' requirements. The more information is produced the more information analysis centres are needed. For some subject areas information analysis centres exist for many years, some, such as the Gmelin Institute for Anorganic Chemistry and the Beilstein Institute for Organic Chemistry, even since the last century. The Beilstein is now online available (e.g. through STN, Karlsruhe) covering information from 1830 on. Software is being developed for accessing not only bibliographic data and full texts but also tables and graphics. (Tittlbach).

Standards

There are various standards for the electronic data interchange which are partly conflicting. We are often lost not knowing which standard to apply; should we follow the standards of ISO or those of other organisations? (Dahev, Pakistan).

Everybody in principle agrees to apply standards; but often the better (standard) is the enemy of the good (standard). It takes a long time until a standard is adopted. In contrast to patents, standards are prepared mainly by volunteers. We should realize that there are economic benefits in standardisation, and we should therefore be prepared to invest money into standard making. (Ampt).

Closing remarks

Walter Blados, Chairman of the AGARD Technical Information Panel, thanked a large number of individuals for their contributions to the success of the meeting and announced that TIP is sponsoring lectures in June 1990 in the USA, UK and Belgium and that a TIP Specialists' Meeting "Bridging the Communication Gap with the help from Natural Language Processing" is scheduled to take place in Trondheim, Norway, from 5th–6th September 1990.

Questionaire

The questionaires completed by participants after the meeting revealed interest in technologies such as electronic mail, computer conferencing, video conferencing, hypertext, optical character reading (OCR), intelligent character reading (iCR), and barcode scanners; in the role of librarians in the next century: in CALS-like projects; in practical details of leading-edge systems; in interworking between systems; in relative capabilities of hard- and software; and in live demonstrations.

In general, there was more interest in practicalities than in statistics and considerations.

3. COMMENTS ON TECHNOLOGIES.

The most natural forms of communication are gestures, mimic expressions and the spoken word, all understood by kids. Civilization required words to be stored and transmitted over long distances. Writing and reading was the technique chosen, and the kids had to become literate, a process amplified by Gutenberg. Today the 'kids' have to be computer literate or even information specialists in order to find all the 'words' collected and stored somewhere. Is this inevitable? Mankind has developed techniques for the collecting, storing and distribution of information. Should mankind not be able to develop techniques for the easy access to information which would render computer literacy and special training superfluous? There are signs that we and our descendants still have to go a long way, but that this way will lead towards something like an information machine which answers (spoken or written) questions.

The information industry, on its way to become profitable, is proceeding from producing something technically feasible to producing something demanded by users. With profitableness comes competition and more respect for user demands. Therefore profitable information markets like the financial and business information market are served first with user-friendly information products. Medical doctors come next; technicians and scientists come further down the line. In any case there is a trend to break the information market down into fractions corresponding to well defined target groups of users who then are provided with all information they need for their profession. Much effort is put into mixed input (text, images, voice, video) and multimedia output (online, CD-ROM etc.), into the development of software for really easy access to information with a minimum of training, and into collecting and adding value to the relevant (internal and external) information. This development I interprete as a trend towards information machines.

As long as an information machine is not there, let alone on our desk, we, the information services community, must look for ways and means to serve our users best. We must look for relevant data and for technologies to evaluate, categorize, store, retrieve, distribute and make them consumerable for the end-user. Let us look at some of these technologies.

Data collection

Data include text, images, voice and video. With regard to text, traditionally we dealt with bibliographic data and abstracts. They refer to rather than contain information. Therefore the trend, supported by high capacity storage media, goes towards full texts. The texts can be captured in facsimile mode by means of scanners, as done e.g. for the BACON project for 125 million pages of patent documents scanned on behalf of the European, the US and the Japanese Patent Offices (see paper 9), for the ADONIS project which every week produces a CD-ROM with articles in the field of biomedicine contributed by a multitude of publishers, and for the Pressedatenbank of Gruner & Jahr with an input of 1000 press cuts per day. The technology is well developed and permits different degrees of resolution (e.g. 200, 300 or 400 pixels per inch); the scanning costs per page are in the order of 25 cents. Scanning covers images as well as texts, an advantage. However, if the text were available in coded form it would require much less storage space and could e.g. be extracted, edited, searched or translated by computer, transmitted at lower costs etc. Texts can be read into a computer by Optical Character Reading (OCR). If it is available already in scanned form OCR would require first printing on paper and then reading which is rather inconvenient. For this purpose Intelligent Character Reading (ICR) has been developed which in a first step identifies rows of scanned text and then the letters within the rows. The algorithms for identifying letters are more powerful than for OCR; as a consequence ICR can be applied to practically all type faces and even to handwriting. ICR is not yet a mature technology in spite of very remarkable results. As this technology seems to be important for dealing with (internal and external) full text documents it is recommended to study this topic further. As it is relevant to the theme of the next TIP Specialist's Meeting ("Bridging the Communication Gap with Help from Natural Language Processing") an invited paper could deal with it (proposed title: "Comparative Evaluation of ICR Systems"). Another approach could be to induce an organisation such as INK or International Resource Development Inc. to carry out a multi-client study on this topic and to subscribe to it.

Adding Value to Data

The critical evaluation of data by information analysis centres and the screening and condensation of information for inclusion in handbooks or more advanced devices is an eternal task. Even the envisaged information machine requires continuous updating with knowledge. Librarians and information centres, when dealing with electronic transfer of information, concentrated first on secondary (bibliographic) and now increasingly on primary (full texts, images etc.) information. More user-friendly, however, are evaluated data and handbook-type of information. Many librarians fear that some of their traditional functions will in future be fulfilled by electronic devices. I wonder if they should not take the initiative e.g. in cooperation with publishers, to urge for, to encourage or even to organise the adding of value to data: for the benefit of their users and their own. And TIP could well amplify such conceivable initiatives of its members by grouping and channelling them.

At the TIP meeting frequent reference was made to the Standard Generalised Markup Language (SGML) which adds value to data, not to their information content but to the facility with which they can be further processed. This reminds me of an automatic translation system (TITUS) which requires pre-editing (transformation of natural language into a language with a limited choice of syntax) before the translation algorithms can be applied. Are we on the right track? "Technology is the effort to avoid efforts" (Ortega y Gasset). With TITUS and SGML we reduce processing efforts but we increase text creation or preparation efforts. These thoughts might sound heretical in the light of the acceptence and adoption of SGML by (at least parts of) the information community. Perhaps it is only a question of finding the right balance between efforts.

Categorization (indexing) and searching (retrieval)

This was the only topic within the scope of the conference theme which was not adequately covered by presentations; it was, however, addressed at length at the final discussion. It appeared that those who spoke at the meeting were of the same opinion that for both, indexing and retrieval, information specialists are unavoidable ("as long as authors cannot be automated"). Knowing that one counter-example invalidates even the most convincing and familiar theory I report that I have recently seen a demonstration of an automatic indexing and retrieval system which can successfully compete with any specialist:

terms (except stopwords) are extracted from full texts (or abstracts) and reduced to stems; linear and an inverted files are established; the query is entered in natural language, and its terms are matched against the files; retrieved texts are relevance assessed; with terms occurring in the relevant documents tight sub-queries (retrieving a minimum of texts but at least one) are derived; the retrieved texts are assessed and the most successful (in terms of retrieving further relevant texts) sub-queries are selected and used for further retrieval and assessment; with now more relevant texts useful terms can more precisely be identified and used (automatically) for further generation of sub-queries etc. The natural language query can be vague; the target of relevant texts is solely defined by the user's relevance decisions. As a consequence, the end-user, who, of course, must be able to identify relevant information in a text displayed, achieves better results than the retrieval specialist who cannot be absolutely sure about the end-user's real information demands. And, the system retrieves relevant texts some of which the retrieval expert did not dream when formulating a (e.g. Boolean) query and therefore missed them. And, training is not required at all, just entering a natural language query and responding with yes or no to the texts displayed. And, indexing by information specialists is rendered superfluous.

This example shows that we must watch for new technologies even in areas the development of which we used to believe is rather stagnant. "New technologies for automatic (full text) indexing and retrieval" could be a topic for presentation and discussion at the next TIP Specialists' Meeting.

Storage

With CD-ROM and the other derivatives of the Compact Disk the technology for the storage of information has made a quantitative leap. 654 Mbytes per disk can meet many requirements. Many of the CD-ROM products we see today did not yet find their markets. Of the products which serve the information market probably those will be most successful which provide easy access to full texts and, even more, those which provide rather complete information to a well defined target group of users.

Organisations which handle classified information, which cannot entrust external organisations with pressing the information on CD-ROM and which cannot afford the uneconomic inhouse pressing are advised to consider WORM (Write Once Read Many disks) with 800 Mbyte (5 1/4") or 2 Gigabyte (12").

Distribution

Many speakers referred to the forthcoming high capacity communication channels, ISDN, fiber optics, satellites etc. which will permit the (mass) transfer of data and texts with acceptable delays. Two recommendations are repeated here: Information and communication managers should now plan to use ISDN, and architects should be urged to have plastic tubes mounted behind the plaster work in new buildings for taking up optical fibers.

An issue Dunning pointed out is the chicken-and-egg situation with regard to the transfer of data via satellites: there are many potential users of high capacity transmission channels but each of them alone would not need and pay for the whole channel and therefore will not approach the carriers (PTTs etc.). As a consequence, the carriers believe that there will not be enough traffic, and they proceed half-heartedly only. It is a question of identifying and collecting demands for information transfer with high bit rates. Should TIP not invite the AGARD community to identify high bit rate demands of their organisations, collect the demands if they exist, and try to join forces with other parties (banks, airlines, newspapers, research centres, universities etc.) in order to establish a sum of demands which would encourage the carriers to meet them?

Integrated Systems

There is a large number of software packages with some or all functions required by information managers and specialists. Many are very similar; therefore the question was raised if we are not quite often re-inventing the wheel. The reply was that user requirements differ from case to case and that therefore tailormade software is needed the development of which costs millions of dollars. STILAS (paper 8) and EPOQUE (paper 9) may serve as examples. I wonder if one could not think of more economic solutions. My suggestion would be to entrust an appropriate organisation with the collection of user requirements, with setting up user requirement specifications, with separating the requirements into many small modules, and with establishing protocols for linking the modules. The information industry should be involved in this process from the beginning, and the so established functional modules and protocols should be given the status of guidelines or better standards. I am sure the information industry would comply. Users could then select appropriate module and entrust system and software houses with the development of functions which are not yet adequately covered by existing modules. Who could sponsor the first step, the collection of user requirements and the setting up of user requirement specifications? Here Dunning's statement comes to my mind that standardisation and common procedures are close to the heart of the Commission of the European Communities (CEC). Perhaps TIP should ask the CEC to make a start with (co?) financing the first step.

Copyright

At the discussion of De Ridder's paper on optical disk systems the question "How does copyright effect the use of CD-ROM?" was raised but not adequately answered, in my opinion. The answer was: "One can arrange that a program copied from a CD-ROM can be used only when disk drive and CD-ROM are available". This is fine for programs but not for texts. I guess that the originator of the question (Stolk) wanted to draw the auditorium's attention to the fact that one can draw multiple paper copies from full texts stored on CD-ROMs and sell them, not to the liking of publishers. This dislike induced the ADONIS group of publishers to make their CD-ROMs, containing (biomedical) journal articles in facsimile mode, available to libraries only and this under the condition that they will get a fee for each paper copy drawn.

4. RECOMMONDATIONS

Recommendations Addressed to Information Managers and Specialists

1. Keep yourself informed about the development of the information services market, e.g. through the annual reports of the European Information Market Observatory which cover market statistics and major trends. Address: CEC, DG XIII, L-2920 Luxembourg.

2. Watch out for new technologies such as hypertext, CD-I, WORM, "Reversible" disk, ICR (Intelligent Character Reading), relational image databases, more advanced automatic translation systems, knowledge based (AI) systems and those which do not have a name yet.

3. Look which new technologies the European, US and Japanese Patent Offices apply; in some cases they are not only protagonists but also de facto standard makers.

4. Plan to make use of ISDN.

5. If you have the slightest influence on the planning of new buildings take care that tubes for taking up optical fibers are installed.

6. Study success stories such as STILAS (Paper 8), EPOQUE (Paper 9), CALS (Paper 14), PROMIS (Paper 16) etc. and consider to use (part of) these systems.

7. Consider the application of SGML (Paper 11).

8. Consider if you should not make use of Japanese information (Paper 3).

9. Consider if you should not make use of earth observation data provided by satellites. See the Section "Use of data collected by satellites" of the General Discussion!

10. If you need a computerized French-English dictionary ask Aerospatiale, B.P. 76, F-92152 Suresnes Cedex. It should be ready in about a year's time.

Recommendations addressed to AGARD's Technical Information Panel

11. At the next TIP Specialists' Meeting "Bridging the Communication Gap with the Help from Natural Language Processing", 5th—6th September 1990 in Trondheim, the following topics should be considered for presentation and/or discussion:

    — Comparative evaluation of ICR (Intelligent Character Reading) systems
    — New technologies for automatic full text indexing and retrieval systems
    — Comparative evaluation of automatic translation systems.

12. TIP should consider if it should not try out E-mail and/or computer conferencing between Panel members and, at a next step, between participants of the annual TIP Specialists' Meeting. A topic for discussion could be the Technical Evaluation Report. If such an exercise would be successful it could be recommended to other AGARD Panels. (Papers 12 and 13).

13. TIP should suggest to relevant information centres and libraries to consider if they should not encourage, group, channel or even organise the critical evaluation of data. See also the Section "Critically evaluated data" of the General Discussion and the Section "Adding value to data" of the Comments on Technologies!

14. TIP, together with interested AGARD organisation, should try to identify demands for data transfer with high bit rates in order to find out if a critical mass for high bit rate telecommunications could be reached in conjunction with non-AGARD-related organisations. The collection of demands above the level of AGARD could then probably best be done by NASA and ESA. See also the discussion of Paper 1 and the Section "Distribution" of the Comments on Technologies!

15. TIP should ask the Commission of the European Communities to encourage the establishment of user requirement specifications for all functions demanded by information managers and specialists (library functions; indexing, storage, retrieval, distribution of information; E-mail, computer conferencing etc.). These specifications could provide the basis for standard information technology modules and for standard protocols for linking the modules. See also the discussion on Paper 9 and the Section "Integrated systems" of the Comments on Technologies!

Recommendations to AGARD Related Organisations

16. Consider to adopt CALS (Computer-aided Acquisition and Logistics Support). See Paper 14!

17. For the storage of classified information consider the use of WORM disks. They can be loaded with information in-house. See also the Section "Storage" of the Comments on Technologies!

AGARD

NATO ⊕ OTAN

7 rue Ancelle · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone (1)47.38.57.00 · Telex 610 176

DISTRIBUTION OF UNCLASSIFIED
AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres.Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Sales Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD — VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Director Scientific Information Services
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Board
Ved Idraetsparken 4
2100 Copenhagen O

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92320 Châtillon

GERMANY
Fachinformationszentrum
Karlsruhe
D-7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reyjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3 Piazzale Adenauer
00144 Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
Kluyverweg 1
2629 HS Delft

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Portuguese National Coordinator to AGARD
Gabinete de Estudos e Programas
CLAFA
Base de Alfragide
Alfragide
2700 Amadora

SPAIN
INTA (AGARD Publications)
Pintor Rosales 34
28008 Madrid

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE Daire Başkanlığı (ARGE)
Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Research Center
M/S 180
Hampton, Virginia 23665

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD
STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE
DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

SALES AGENCIES

National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

ESA/Information Retrieval Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

The British Library
Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Branch
NASA Headquarters (NIT-40)
Washington D.C. 20546, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA