

ELECTRONIC PUBLISHING IN SCIENCE - WHERE ARE WE NOW?

F A Mastroddi

Paper presented at a joint ICSU Press/UNESCO Expert Conference on Electronic Publishing in Science, Paris, 19-23 February 1996

This paper was first published by UNESCO as Proceedings, Conference of Experts on Electronic Publishing in Science, Paris, 1996, and is reprinted with kind permission.

Mr Franco Mastroddi is personal assistant to the Director General, Directorate General Telecommunications, information market and exploitation of research, Commission of the European Communities, DG XIII, Batiment Jean Monnet, Plateau du Kirchberg, Luxembourg

Introduction

"The information content sector will be very important for the future information society... It is necessary to develop the information services sector extensively" European Council of Ministers Resolution, 7th November 1995.

The progress of information and communication technologies over the past years is undeniable. Today, over 25% of offices in the European Union and 15% of homes are equipped with computers. Data networks span the globe. Scientists, engineers and businesses now have tools at their disposal to circulate data very quickly to worldwide audiences. This potential has far-reaching implications, not only for science publishing but for research, innovation and the economy in general. The 1995 G-7 Summit on the Information Society made clear links between the emerging information highways and economic competitiveness, job creation and quality of life.

There are three main driving forces behind this progress. The first is computing power. This has, according to Microsoft, made the equivalent of a million-fold improvement in the last 20 years, due to the improving price/performance ratio¹. The second force, data communications and networking, has become commonplace in the same time-frame. The third factor, often underestimated, is information content. Scholarly literature alone doubles every ten to fifteen years. A vast proportion of this literature is now being generated in or converted to electronic form every day, stimulating the progress and acceptance of technology-mediated communications.

What kinds of changes are being wrought in the publishing and information industries? Some of the main trends can be characterised as follows:

- from conduit to content. Less emphasis is being placed on technology and more on information products and services.
- from scribe to screen. Authors are generating more and more material - not just text, but increasingly multimedia, electronically.

- from local to global. Today, electronic information suppliers and customers can take a more global perspective, due to the distance-independent nature of networks like the Internet. But the necessary infrastructure is not installed in every country or region. Over half of the world's population has never placed a telephone call.
- from supply-driven to demand-driven services. Electronic information products and services (EIPS) will need to centre their systems design more on the actual and emerging needs of users.

This paper looks in more depth at these trends, covering the rapid growth of electronic information supply, the new kinds of user demands that are emerging, new models to improve the value-added information chain between author, publisher and user, and finally activities funded by the European Union (EU) relevant to electronic publishing in science.

The supply side: context and trends

The electronic publishing market

"Publishers are largely unaware of the potential of electronic publishing."

A 1994 EU study by Consulting Trust² estimates that the overall potential market for electronic publishing in Western Europe could be as high as 12,000 million ecu's (Mecus) by the year 2000. This would represent about 11% of the overall print publishing market, so the demise of the printed word is not yet forecast!

The main segments of the market are corporate publishing and communications, finance, entertainment, directories, reference material, legal, science, technology and medicine (STM), education, travel, hobby and specialist interests. The STM share is forecast at between 20% and 30%, representing some 625m ecus.

In contrast, the study revealed that many publishers in Europe are largely unaware of the dramatic changes in the information industry, of new opportunities and threats and the strategic implications of new media. There is significant uncertainty about technologies, markets and

economics and a lack of vision of the future perspectives. The overall electronic publishing market in Western Europe lags some three to five years behind the United States, where there exists a large, and homogeneous market without the linguistic and cultural differences which characterise the EU. Exceptionally, in the STM market, some of the larger European publishers in the English language are well established in the North American market.

Worldwide, STM is strongly represented in the database market.

STM databases

"Bibliographic and textual online databases still predominate in STM."

Both the number and variety of electronic information services in the STM area have seen continuous expansion in the last 20 years, but the core supply is still represented by classic online bibliographic databases. In total, approximately one thousand of the world's 8000-odd online databases are listed as being in the STM area, although the figure is probably nearer to 700-800, allowing for different versions of the same databases and sub-files of a database series.³

The majority of STM databases are either online (54%) or on CD-ROM (23%). Others are available usually on tape, diskette or other means. STM only represents a small portion of CD-ROM products, between 10-15 percent according to different estimates by the European Information Industry Association and other bodies.

Over forty percent of the databases are bibliographic in format. A substantial number (29%) contain full text. Only a small percentage of the databases carry images (4%), and it will be interesting to see whether this figure changes significantly over the next years.

Not surprisingly, the vast majority of STM databases (88%) are in the English language. Other languages covered are French (in 4% of databases), German (3%), Spanish (2%) and other languages (3%). This factor is probably not seen as a major hindrance to the increasingly international academic and research communities in STM, but could become a factor if electronic publishers wish to

target broader international markets or localise their products. Two-thirds of the STM databases surveyed originate in the US or Canada. Just under one-third originate from within the EU member countries, whilst a small percentage comes from Australasia, East European or other countries.

Internet-based services

"Over 40 million people have access to the Internet. The question now is: what will fill the information highways?"

The Internet provides a parallel evolutionary path for electronic publishing in science. Although the Internet is based on its own set of protocols, it is possible to cross between classic ASCII-based online hosts and Internet services, such as bulletin boards, telnet services or file transfer. Bridges with Web sites are also possible through gateways or intelligent agents. The European Commission Host Organisation (ECHO) is setting up gateway protocols between Web sites and its databases. Intelligent agents which can handle different environments for the users are being announced by General Magic.

The Internet is growing fast, numerically. The EU's Information Market Observatory noted in mid-1994 that there were some 32 million users on the Internet. One year later, the market research firm IDC estimated the number of users worldwide at over 40 million, growing to 128 million by 1997. In comparison, customers for classic online hosts are counted in terms of hundreds of thousands.

Science publishing is a major activity on the Internet, and academic and scholarly publishers in Europe were amongst the first to recognise the potential importance of the Internet as a publishing medium. Projects like TULIP and Red Sage illustrate this point. Today, there are several thousand science sites on the Internet. The top science subjects are engineering (880 sites), computer sciences (727), medicine (612), biology (509), earth sciences (473) and physics (469).

Out of all the services springing up on the Internet, probably the most attractive ones in terms of publishing are on the WorldWideWeb

(WWW). From fifty servers in 1992, registered WWW sites total approximately seventy thousand by end 1995. There are in excess of 800 new web sites appearing worldwide every day according to an announcement in late 1995 by the W3 Consortium. The sites carry articles, news, references to documents, publicity, announcements, job seeking, calls for papers, information for authors, and user feedback forms. Many Web sites are based on small computers and are for promotional purposes, rather than providing a full search service. Instead, they point the user to other types of access modalities, such as e-mail, file transfer protocol (FTP) addresses or telnet access. DIALOG for example, refers Web users to telnet connections. A number of search engines are available on the Web. Their capabilities range from simple browsing based on one-word searches to allowing Boolean operators, string searching on titles and hypertext-marked words, word proximity and weighting for search terms. However, the indexing is not as extensive as with classic databases. Also, a Web search often does not give detailed information about a site's contents, but a title or sometimes just a journal issue number. In such cases, the results of a search often need to be verified by downloading the full text, with all the potential waste of time and energy.

One Web-based catalogue⁴ lists 1500 electronic journals. Amongst the academic journals, around fifty are actually classified as peer reviewed although many more are to be found. Only a modest number of journals have shown up in the non-reviewed category, leading one to believe that this label is not regarded as essential by the publisher who fills in the directory entry. Over eighty science and academic publishers are listed on the WWW, from the commercial sector, higher education, institutes and associations. In addition to the journals side, STM text books are also becoming available on the networks.

At this point in time, interest from the supply side in STM looks firm and growing, although many of the activities are promotional and not for delivering content. This may change, as in other areas, where classic servers (e.g. MAID) are starting to offer full search capacity on the Web.

The demand side: new user needs

"The degree of acceptance of new information services by users is one of the determining factors for an effective development of the information society"
European Council of Ministers Resolution, 7th November 1995.

The understanding of user needs is not a new problem. Studies on both sides of the Atlantic as far back as the mid-1970s have tried to quantify the elusive requirements of the online user. There are still major stumbling blocks to determining user needs. Firstly, there is a lack of objective measurement criteria, especially for judging information content. User acceptance criteria such as relevance, comprehensiveness, timeliness or cost-benefit are important of course, but need to be put into a more objective context. Usability matrix in this respect is a research topic in EU programmes. Another problem is measuring the performance of information delivery mechanisms. There are some established metrics here, but the technological goal posts keep shifting. New technologies, techniques and services keep emerging. There are few stable benchmarks against which user acceptance and performance criteria can be measured. One example is in the networks area, where the availability of more bandwidth can instantly change the way an information service is delivered. An additional factor of complexity is the current concern for information and document exchange over different hardware and software platforms.

The quality of today's information services

"At least one in five online searches meets quality problems."⁵

The quality of electronic information products and services is a source of potential concern. Clifford Stoll⁶ claims that new technologies cannot be cheaper, faster and more effective all at the same time. The makers of PCs may disagree! The Internet is quoted as an example of this limitation. Users may appreciate getting cheap access to information in some cases, but they can suffer in terms of extended search times, ineffectiveness of the retrieval software and occasionally ending up with goods of

dubious value which have just been dumped on the Net.

What is the extent of the problem? A user survey carried out by the European Commission's Information Market Observatory (IMO) revealed that the average respondent meets major quality problems in one in five online searches. In these cases, the quality problems concerned irrelevant or unwanted data, the slowness of the search process, formatting problems which made the results of the search unusable, or a mixture of the above.

Not surprisingly, the study concluded that improved quality is a prerequisite for user satisfaction and a major factor in expanding markets. Suppliers should work with users to develop User Requirements Specifications and measurement methods. It was felt that standardised approaches to quality management do not in themselves satisfy users. This is mainly because relevant standards like the ISO 9000 series are about reporting on quality, not implementing it. The study claims that the Internet needs the help of the professional information community for classifying and indexing information, but is this cost-effective?

There is a lack of legislation in this area. Responsibilities are not clear, especially when it comes to commercial or legal liability. It is often up to the individual organisation to set quality standards. INIST (Institut de l'information scientifique et technique) for example, has a quality assurance plan for its STM services and databases like PASCAL. The study also points out that the information industry associations like EUSIDIC (European Association of Information Services), ADBS (Association des documentalistes et bibliothécaires spécialisés), NFAIS (National Federation of Abstracting and Indexing Services) and CIQM (Centre for Information Quality Management) are starting to address this issue. In view of the present situation, however, it is still a case of 'buyer beware'.

A new problem area: non-text-based information retrieval

"Multimedia information availability does not necessarily imply information accessibility."⁷

Documents, still images and video, as well as sound files can nowadays be located and downloaded relatively easily. But there are several problems.

Firstly, the user and the information source usually need to have the same data exchange protocols.

Secondly, there is a lack of bandwidth and high-capacity switching nodes. The current situation with WWW can be painfully slow, as the communications lines or gateway switches have become saturated. This neutralises the attractive browsing capabilities of Web protocols. Users are often encouraged to disengage the graphics mode. Future broad band networks offer a more exciting prospect, and several broad band facilities already exist in the research communities. The EU programme ACTS (Advanced Communications Technologies and Services) is currently experimenting with this approach, for example with groups of museums and galleries. The Telematics Applications programme is experimenting along similar lines with research networks for scientists and engineers.

The lack of searchability of graphics and images is another issue. The problem is in searching the *contents* of the multimedia object, for example to locate a certain detail, shape or pattern, to be able to interpret molecular structures in a visual way, or to extract data from medical images and could avoid the need for the user to make the search on accompanying text fields. Current work in the area of geographical information systems, for example in the EU's IMPACT programme (Information Market Policy Actions), offers some interesting leads in this area.

One approach to this problem, indicated in an information engineering study by GMD, Germany, is to combine the ability of Internet protocols to locate and deliver multimedia objects to the user, with the power of Boolean retrieval systems into a single engine. Of course, the objects would need to be tagged with 'meta-information'. Examples suggested of useful combinations are:

- classification of images through embedded text. In this case, searchable text objects are linked to parts of the image;

- classification of images through textual description, using automatically generated algorithms. This could be developed for example to improve communication between chemists by means of graphical structure diagrams. The diagrams could be expressed in textual form, to help users retrieve the required chemical structures.

New supplier-user links: a review of the information value chain

Managing change

"A survey of 259 companies conducted by the American Management Association finds that 84% of them are going through a business transformation. Information technology is cited as the main area of change."⁸

ICT has led to introduction of new working practices in many different sectors of the economy, ranging from financial services to shipyards. Publishing can be considered as one of the traditional sectors, where successful practices built up over hundreds of years are not discarded lightly. However, there are pressures to change. Many companies and organisations outside publishing are realising that their second business is the information business. Those which master new technologies, such as telecoms operators and IT companies, are now looking to master the digital information content. On the other hand, traditional media companies nowadays need to be at the cutting edge of online technology, by providing marketing on the Web, interactive CD products and so on. The picture is further complicated by the 'do-it-yourself'; publishing possibilities. It is technically quick and easy for an author to provide Internet access to his work. However, for an established publishing organisation, there are always logistic and organisational questions, and before the go/no go decision is made, cost-benefits and the impact on present activities need to be explored.

Apart from the type of pressures outlined above, there are also opportunities. Many science publishers already implement electronic publishing in part, for example by accepting diskettes or e-mail from authors. Technically, it

is not a great step towards full electronic publishing on the various online networks or through CD products.

The changes also put more pressure on the user. Electronic publishing at present tends to make the middle links of the information chain more efficient, by displacing effort towards the two ends, namely the user and the author. In STM, these are often the same people, but the needs are different. Users have to learn esoteric retrieval systems, deal with temperamental networks or CD-ROM drivers, spend hours navigating and retrieving texts, reformatting them and printing them out. Authors need to learn new authoring systems, mark-up techniques and increasingly, multimedia object linking and embedding. This hidden effort is usually not costed, but at some point in time may need to be accounted for.

A new information chain

"A recent EU study has re-examined the business model used by publishers, and concludes that a new kind of information chain is evolving."

The publishing process is being disrupted. It is possible to bypass certain steps (e.g. typesetting and proofreading) and to make endless iterations on other steps (e.g. peer review and updating or redrafting of material). A piece of electronic information can be cycled and recycled so easily that it can no longer be considered as stable in the same way as with printed publications. Information technology has made the information chain more flexible and less linear. The traditional distinctions in the old value chain are being blurred. Conventional players like typesetters, printers and distributors (including libraries, book stores and documentation centres) are having to re-evaluate their functions, roles and skills. New players, often from the information and communications technologies sector, are entering the exclusive domain of publishers and information providers. They can be from the supply industries like telecommunications organisations or hardware/software manufacturers, or from other sectors of the multimedia market, such as television, cinema and even video games manufacturers.

The information engineering study entitled IE 2001 examines the new pressures and opportunities provided by information technologies for publishing. It comments on the most influential technologies, such as the increasing bandwidth becoming available, and on the ways in which publishers can react. Its main conclusions are that:

- the emergence of more targeted products, e.g. in the STM area, has become easier. It is also becoming a necessity. The notion of geographic marketplaces is hardly relevant on today's data networks. It is being overtaken by the concept of 'affinity groups', namely sets of users with like interests but geographically dispersed.
- the long term balance of power is shifting, as the role of producers, vendors and distributors is continually changing and no-one can claim to have an exclusive role in electronic publishing. This evolution could easily mean that publishers who do not adapt appropriately will be left behind.
- print-on-paper is still a strong force. Digital printing is a transitional product paving the way for paperless products and services
- publishers should be 'digitally prepared'.

It comments that STM publishing has enormous potential for promoting this evolution, particularly with the support of academic networks for gathering and distributing electronic texts.

The present core process of electronic publishing is threefold: input, processing and output. This general process still holds, but there are now more distinct stages and the process is no longer linear. Once a document is captured, it can be processed, formatted, delivered and retrieved along several different, parallel paths. It can be changed and recycled several times en route.

The new stages of the chain are portrayed as follows:

Content creation and manipulation: the desktop publishing upswing of past years has created a relatively mature market in text and document processing. So much information now exists in electronic form and can be syndicated or sold as a commodity. In this case, the repackaging of

information becomes a priority issue. There is also an increasing number of tools available to deal with graphics and most recently, video and animation, although there are not many at desktop prices. There are probably some 70 authoring packages on the market. However, the full multimedia authoring package which is cross-platform and which can be customised to the individual needs of vertical area - such as STM- is still around the corner.

Document representation, interchange and delivery: the continued progress of norms like SGML (standard generalised mark-up language) and its hypertext equivalent HTML, their derivatives and image encoding norms such as JPEG (Joint Photographic Experts Group) and the MPEG (Motion Picture Experts Group) series is promising. In practice, the most successful software viewers and browsers can be re-configured to cope with different graphics formats. There is still a problem of video formats, due mainly to low processor speeds.

Database storage and retrieval: relational database management systems still dominate. There is continued research interest in object-related systems, where the notion of a record or a document could be replaced by the concept of the 'information object'¹⁰. The object in this case would be a self-standing, discrete piece of information enriched with meta-information on its context, status, related objects and so on. This approach goes a step further than the logical tags given to HTML documents, which appear as hot-spots on Web Servers. On the retrieval side, there is currently great interest in intuitive graphical user interfaces, such as can be found in some game packages, and, on a further horizon, techniques such as information visualisation.

Networks and services: there are several evolutionary paths for networkers. Service providers like CompuServe, Prodigy and America Online as well the plethora of specialised services, need to support an increasing number of network protocols, as well as their proprietary ones. Packet-switched networks are widely available. ISDN (Integrated services digital network) has been introduced to many European countries, but is not yet in widespread use. Many global commercial networks are springing up, such as the

Microsoft network, Infonet and AT&T Interchange. These are commonly used in conjunction with the normal telephone Internet access, unless the user has a connection to a research network. One major advance in this area is the continual increase in modem speeds and bandwidth, as traffic increases on the networks. However Web access and multimedia applications in Europe at least rely much at present on the research networks. Pilot applications on national research networks such as SuperJANET and at European level in the ACTS programme are currently demonstrating the technical possibilities of the impending broadband networks which promise to provide full multimedia services.

Work flow and transaction management: this element in the model helps to facilitate secure contacts between closed user groups, for cooperative group working, or between suppliers and users, for payment systems. It is a stage previously under-represented in the electronic information chain. It is currently the subject of research worldwide, not only in the EU's Information Engineering research projects, but also in the US Digital Library projects and in Japanese projects. Amongst other factors, this link in the chain should aim to provide a subsystem for royalty payments or copyright clearance.

Content identification and usage: best practice models are needed to help guide user behaviour when performing information access. Apart from the concerns cited above, new problem areas are: identifying multiple sources, searching meta-information (for example for information visualisation), intelligent agents, interleaved parallel access to databases in a distributed system, iterative refinement of queries, combining information retrieval with other work, such as populating models with new data, feeding multimedia presentations and designs, and incorporating external results in technical reports.

In the case of STM publishing, the current status with regard to this chain is illustrated in the study by the following characteristics:

- few STM authors currently produce multimedia content or have the necessary skills, although many scientists nowadays

- have easy access to the emerging research networks
- nearly all publishers have started to implement digital storage and SGML tagging, some with the help of authors
 - most interactive media STM publications are produced only on CD-ROM, or experimentally online.

The study predicts that in five years time, new elements will come into play. For example, digital printing will facilitate on-demand STM publishing. Paper-based publications 'of record' may disappear and be replaced by validated databases. Multimedia content databases will appear online as much as off-line. Some of the above links can be joined directly through the Internet, skipping the intermediaries. This could lead to a situation where fast scientific communications might obviate the need for some STM journals.

What the user sees

"The importance of a good user metaphor cannot be underestimated." M. McAdams, content designer of the online Washington Post.

The changes in the information chain, will necessarily lead to a different organisation and presentation of information products and services. Care should be taken at this stage to target any restructuring towards the needs of the user. The new or casual user should be given a clear picture in advance of the contents, capabilities and limits of any service, the electronic equivalent of window-shopping.

One of the main research topics in recent years, amongst EU programmes like IT (Information Technology), IMPACT, Information and Language Engineering, has been in computer-mediated access to information. Today, menus, guides, directories and retrieval software are all commonly available. However, they can be slow, cumbersome and linear, guiding users narrowly along a certain route at the expense of retaining an overview. There are several new approaches to this key problem.

One approach is through better interfaces. The current trend towards Graphical User

Interfaces (GUI) working in client/server mode is very attractive, as illustrated by products such as SciFinder for Chemical Information. However, they currently rely on establishing a one-to-one protocol with an information source, and their usefulness is limited in multiple source searches. This particular problem is being tackled in the STM area, under the MIME (Multipurpose Internet Mail Extension) initiative for linking chemical information sources under a common protocol.

Interfaces will gradually become more user-responsive, through the integration of natural language features, speech recognition and localised adaptations. One popular topic at present is that of so-called 'smart software'. This is not usually portrayed as artificial intelligence, but rather in the form of software agents or knowbots: self-contained programmes which help out the user to make choices when using unfamiliar products or services. Software wizards found in today's desktop and games products are a precursor of this technique. Software agents are being developed, for example at the MIT Media Lab, to help users on the information highways. They aim to detect the user's individual search patterns, help guide users through the Web by 'looking' some steps ahead and reporting back, by providing personalized memory banks for the user and even by swapping information with other agents, e.g. to put users in touch with each other. Although such developments sound futuristic, the process has already started, e.g. electronic agents in the Sony Magic Link device and responsive cartoon characters in interactive games. Bill Gates has announced recently that Microsoft will be launching smart products in the near future.

Another approach to this challenge is non-technical, namely the creation and realisation of new user metaphors, to guide users through retrieval systems and information content. A metaphor should help adjust a person's expectations and assumptions about how a service works and what it can (or cannot) offer. The notions of 'global village' or 'electronic malls' are examples of such metaphors. With information services, there is too often a perception gap between users and the information system. Too little is known about

the way people envisage information systems and navigate around different databases and services.

What are the ground rules for user metaphors? Firstly, the electronic version should provide more added-value than the print version, both in terms of content and search facilities. Secondly, the structure of the product or service needs to be clear and flexible, so that users are not confronted with a rigid set of approach paths to information items. For example, the first screen that a user sees could provide different angles of attack to the ensuing information. A balance is needed between simplicity and sophistication. After following a series of hypertext links on the Web, the user more often than not strays from the area of interest or ends up in a blind alley. Thirdly, it is vital for individual works to be put in the proper context of established knowledge, especially in academic research. Logical search paths should be kept consistent, particularly if hypertext linkages are extended to any significant word in the document. Projects like the UK's Electronic Libraries Programme's Open Journal and the Human Genome project are trying out fresh approaches to this problem, including the possibility of user-generated links between previously unconnected hypertext objects based on cognitive methods.

This is not yet the case today, however. Users lack sufficient means to locate information quickly or even to find out if the information exists on a network, and can easily wander aimlessly through online services before giving up.

Appendix

EU-funded activities

"It is essential for users to be involved in all phases of research projects so that they can express their needs..."¹¹

Today, there are several EU-funded research and market stimulation initiatives which cover different aspects of electronic publishing. The most relevant of these are the new research activities in information engineering and libraries and the new INFO 2000 market stimulation initiative.

Projects must involve users at every stage. A typical project life cycle starts with drawing up of user requirements, and concludes with demonstration and validation by users in the field. The main resources are spent not on technology development but on obtaining user feedback. The importance of this approach, as compared with the 'technology-push' approach of previous years, cannot be underestimated.

One project, for example (BASELINE) will produce guidelines for system developers in EU projects and elsewhere on new methods of user validation and usability engineering. Another project (INUSE) will set up a network of usability support centres which will not only serve EU projects but become available to a wider audience for the purposes of user-centred design. This is not a new idea, and will build on existing facilities, training the trainers. It is hoped that the network will become widespread and easily accessible to both the industrial and academic communities.

Information engineering research

The aim of Information Engineering is to permit easier and more selective access and better usability of electronic information in all its forms. The work covers the principal links in the electronic information chain (production, dissemination, retrieval, etc.) and focuses on four priorities: meeting user requirements; improving integration into user-friendly systems of the tools and methods used at different stages in the information chain; improving the value and usability of information; and managing information in the form of images, sound and other non-textual forms of representation.

During 1995, a Call for Proposals attracted 128 proposals, including 13 STM proposals. The STM proposals drew some general comments from evaluators:

- general awareness of mark-up languages like SGML was high but there was a lack of knowledge about their implementation problems e.g. well adapted document type definitions, clashes with HTML structures;
- little reference was made to international standards for multimedia (MHEG (Multimedia/Hypermedia Experts Group) or HyTime);

- most STM proposals were Internet-oriented, but did not make the distinctions between the various protocol layers,
- most proposers fully recognised the importance of intellectual property rights (IPR) and billing mechanisms, but did not show awareness of existing models, or of existing or promising technologies such as digital cash and encryption techniques,
- information security issues were generally underestimated,
- key issues relating to information formats, such as naming, indexing and retrieval of documents or multimedia objects were not properly addressed. Many proposers relied on very classical methods exploiting Boolean logic or string-searching techniques, without recognising that non-textual retrieval is a new problem that needs attention,
- a methodology for evaluating usability and for incorporating user-centred design in the project which can be improved in an iterative way was generally lacking.

Project examples

Electronic multimedia materials online. In this project, medical publishers and educational course producers will pilot methods for improving multimedia publishing in the educational area, with emphasis on design, reuse of material, flexible delivery and business models to protect IPR. It will operate on the whole information chain of STM, through collaborative working between authors, publishers and other intermediaries. As an example, material will be provided by publishers to intermediaries, who will then repackage and customise it for end users (usually students) and even for onward sale to third parties. This part of the project strikes at the heart of concerns about electronic publishing! Presentation formats for disabled students will also be tested. Demonstration sites are foreseen in Italy, Belgium, Germany and the UK. The project will encourage the transfer of best practices by developing usability guidelines within a Handbook for Multimedia design.

Multimedia publishing information engine. The aim of this feasibility project is to demonstrate advanced facilities to authors and publishers for the production of multimedia technical titles. It will concern books, reference works, and didactics and will aim to integrate various approaches such as hypertext, interactive multimedia and knowledge-based technologies, in order to get a more organised information space for storage and retrieval. The consortium includes publishers from several countries, technology companies as well as academic organisations who will provide direct access to users.

There are also projects like NESSTAR, which concerns statistical archive access and visualisation, which involves environmental information from many small publishers, and GEOMED, a geographical information system project which involves STM data.

Telematics for libraries

Libraries are inevitably becoming part of the electronic information chain. In the US, the \$50 million National Digital Library Project and other projects aim at massive digitisation and access to public libraries over the next five years. These projects include several STM publishers. In Japan, NACSIS (National Centre for Scientific and Technical Information) has a similar programme. Institutes like INIST in France and British Library in the UK are also closely involved in publisher-library ventures.

The EU Telematics for Libraries programme is designed to help increase the ready availability of library resources and to facilitate their interconnection. Inevitably, there are many strong links with academic publishing. It is not just a question of which kind of document supply model to use, or whether to set up an online public access catalogue or document store on the Internet. The larger issue at stake is how the interfaces between libraries and these actors are being redefined. Currently, non-electronic document delivery services are usually based on a 'fair-dealing' principle, where no specific payment is made to the publisher. In future, this delicate balance could well be upset if electronic services live up to their potential. There is still debate about the

extra costs and who will pay - certainly the user, but maybe not the end-user of library services.

Recognising this issue, the Commission has initiated several actions which stimulate cooperation between the different parties on the relatively neutral territory of research projects and concertation platforms. This measure allows all parties to gain experience, gauge the extent of potential problems and test out different technological mechanisms, without the commitment of a commercial contract, and in anticipation of legislation. In some cases, this approach by the Commission has facilitated cooperation agreements between publishers and libraries which otherwise would have been difficult or impossible to reach.

Current library projects

Current library projects which look specifically at the electronic publishing and document delivery chain are EURILIA (technical documents in aerospace), DALI (multimedia scientific reports in oceanography), FASTDOC (document ordering and delivery). All of these projects, except EURILIA, include publishers in the supply of the materials and in assessing the methods of their use. FASTDOC combines online searching with ordering and delivery in the area of chemical journals. It integrates three different ordering mechanisms and aims at a 5-minute turnaround time, user friendly, reasonably priced and economically viable. The preliminary results, a prototype system, look promising. Also, the project ELSA, which includes STM titles, extends the TULIP experiment by remote delivery of full text from publisher to user in SGML format. These projects will mostly aim to finish during the course of 1996.

Another important initiative is ECUP, a concerted action to create awareness in libraries of copyright issues and to establish codes of good conduct for libraries in dealing with copyright issues related to various media and related library use. It will help valorise results and findings coming through current projects (DECOMATE and COPINET, also relevant to STM), which both deal with copyright issues.

This approach is taken a step further in the new projects to be launched early 1996.

New library projects

STM publishers in the LIBERATION project will prepare a range of multimedia materials for libraries, both on CD ROM and for local area network or Internet use. Various billing schemes will be tested in the participating libraries in networked environments. The effects of the availability of multimedia books on the Internet on sales of the printed products will be explored. User preferences regarding work location (library versus users workplace) will also be investigated. Due to the strong involvement of publishers, new alliances in this area will hopefully be encouraged.

Another project, ELITE, includes STM publishers and document delivery centres in an experiment with the interconnection of distributed library services. ELITE will combine the technologies in the areas of electronic document management, access and delivery using WWW services.

Future initiatives: STM market stimulation through INFO 2000

Consultations were held in mid-1995 between the European Commission and STM publishers, users, information disseminators and information hosts, on the INFO 2000 market stimulation programme. The important issue of closer interaction between suppliers and users was raised, given the shifting roles of parties in the marketplace. Another issue was the potential impact of 'do-it-yourself' publishing through the Internet. The extent and ramifications of this possible evolution are one of the unknowns of the future electronic publishing marketplace, and it was felt important to monitor this trend closely.

Other general issues raised were:

- economics: pricing and transaction systems,
- investments from the public sector side, information quality and security, and privacy needs.
- new markets: should STM providers attempt to enter broader markets? What kind of awareness and dissemination is needed at EU level to help promote an STM information culture amongst users.

- promotion of new skills amongst authors, users or intermediaries.

With the European Commission acting as facilitator, it is proposed that these consultations will continue in two specific sectors, pharmaceuticals and construction engineering, where the need for concrete actions will be explored.

References

1. Creating the global information society. Microsoft White Paper, October 1995
2. New opportunities for publishers in the information services market. Consulting Trust 1994, CEC, DG XIII-E
3. Gale Online database. Datastar, 1995
4. WorldWide Web Virtual Library, 1995
5. The quality of electronic information products and services. Information market Observatory, CEC, DG XIII-E. September 1995

6. Silicon Snake Oil: second thoughts on the information highway. Doubleday. 1995
7. Study on non-text-based information retrieval - state of the art. ELPUB 106, GMD. CEC, DG XIII-E. 1995
8. Electronic-Engineering-Times (ISSN: 0192-1541) ISS: n851. June 5, 1995. p128(1).
9. Information Engineering 2001 - identification of influential technologies, impact assessment and recommendations for action. Meta-Generics. CEC DG XIII-E. 1995
10. Document Interchange Standards. CEC DG XIII-E. November 1995
11. Council Decision of 23.11.94 on a specific programme of research and technology development, including demonstration in the field of telematics applications of common interest (1994-1998)