

The Impact of Electronic Publishing on the Academic Community

Session 4: Social and cultural issues

Electronic publishing trends and advances

Franco Mastroddi

Directorate General for Telecommunications, Information Market and Exploitation of Research, European Commission, Luxembourg, Franco.Mastroddi@lux.dg13.cec.be.

©Franco Mastroddi, 1997

[Copyright Information](#)

Publishing in transition

In 1973, two English communications technologists speculated about "one universal network which serves all purposes and carries speech, pictures, data and so forth" [1]. It took twenty years for this technological vision to take some shape, notably through the Internet and World Wide Web. Will it take another twenty years for effective worldwide electronic publishing?

One of the driving factors is the growing availability of multimedia content. Content capture and distribution through networks or optical disks is now technically possible for any company or organization. There are thousands of online databases, CD-ROMs and Web sites on all subjects available to anyone in the academic or commercial world.

However, there are several outstanding issues for publishing. How much longer do publishers have to take technology developments on board? Are new organizational models needed in the migration from print to electronic? Which real benefits will accrue, in terms of better communication of knowledge? How will key concerns about quality, authenticity, navigability and usability of digital information be met? This paper looks at some of these issues.

An emerging infrastructure

Information and communications technologies (ICTs) are central to the information society. In Europe, the telecommunications market, for example, is growing at 10% per annum. Personal computers are widely available. Key developments for multimedia are the growth of CD-ROM installations (5'7 million in the European Union) and networking connections. Internet connections are estimated at between 12'15 million in Europe. ISDN (Integrated Services Digital

Network), offering 64-kbit-per-second channels to the user, now numbers some 9 million users in Europe, but the majority (60%) are in Germany [2].

Liberalization of European telecommunications from 1998 onwards is expected to encourage a more varied range of interactive multimedia content services, such as interactive cable television. But widely available interactive broadband channels are still a long way off: at least not until 2005 [3]. Publishers will meanwhile need to rely on hybrid solutions.

Towards a new publishing model?

Global ICT developments are fast opening up new opportunities for publishing. Possibilities identified many years ago are now turning into reality, such as tele-authoring and on-demand publishing. A new value chain for publishing is emerging [4]: (i) the creation process is going 100% digital and multi-stage. The first stage is content creation and manipulation, including acquisition and design of documents and multimedia objects like animations, interactive graphics, knowledge representations and sounds. Secondly, the resulting digital elements go in the authoring process. The final product can then be customized, revised, repackaged or 're-purposed'.

(ii) Distribution will be more flexible, allowing distributed data delivery through parallel mechanisms: Internet, online, CD-ROM or print. This possibility also allows for on-demand publishing, just-in-time publishing and do-it-yourself publishing.

(iii) Information retrieval will become more interactive, as the reader will need to become more active in locating, identifying and using multimedia content. Help in the form of computer mediation is currently an important research topic.

(iv) Preservation and access. Publication was previously a matter of record, and typeset texts were practically cast in stone. Publications today can become dynamic, evolving through reader feedback and scientific advances. This causes problems for publishers (managing back copies), users (linking old to new versions) and especially for knowledge institutions like libraries, museums and archives. The job of migrating from paper to digital, and choosing the right digital format will become an impossible one unless techniques for preservation and access are properly integrated into the multimedia-content value-added chain.

(v) Embedding meta-information. 'Smart information' can be embedded into multimedia documents for origin, authenticity and transactions. This is crucial for correct handling. All actors in the chain capture, adapt and re-use copyrighted digital elements and in the current absence of a worldwide clearance system common for different multimedia objects, electronic publishing is caught in the dilemma of promoting fair use whilst discouraging abusive copying etc. Meta-information is also very pro-active, for example in the design and presentation of the document, for hypermedia navigational links and for linguistic information allowing semantic searches or easier translation.

For the author, the whole process is becoming non-linear, allowing collaborative tele-authoring, interactive peer review and circulation of intermediate work. For the publisher, digital asset

management will play a central role in the creation process, allowing individual digital elements (text, pictures, sound, etc.) to be managed separately, updated, integrated in different packages and published as the market demands.

The electronic publishing supply side

The so-called information explosion does not only refer to the printed word. Online and off-line databases are now numerous enough to be brought into the equation. Web services in particular are far too numerous to ignore. Electronic publishing is expected to take up a significant portion of the print publishing market by the year 2000. For science publishing, the electronic segment is expected to grow from 2% of the total science, technical and medical market today to around 15% of the market in 2002.

The Web explosion

The World Wide Web is probably the nearest implementation so far of the technology vision of a universal multimedia network. Science publishing has exploded on the Internet in the past 2'3 years. Today, there are approximately 24\000 science site entries listed on the Web, compared with some 4\000 in 1995. Zoology, engineering, biology and earth sciences are the most numerous, totalling some 10\000 sites between them.

Conventional science databases

Both the number and variety of conventional databases have seen continuous expansion in the last 20 years, but the core supply is still represented by classic online bibliographic databases. In total, approximately 1\300 (1\000 in 1995) of the world's 9\000 or so online databases are listed as covering the science area, although the figure should be adjusted to allow for different versions of the same databases and sub-files of a database series [5].

CD-ROM is still alive and kicking. The CD-ROM format is showing a slight growth from approximately one fifth in 1995 to one third in 1997. About half of the database entries include online format, whilst the rest are either on diskette, magnetic tape or handheld. Multimedia is not common in this market place: bibliographic formats account for 53% of the entries and 36% carry full text. Only 8% of the databases are multimedia.

Who will manage the information explosion?

Science publishing produces some 1.5 million works (articles, papers, monographs, etc.) per year. The accumulating scientific and cultural heritage needs to be stored in future in a way which allows quick and easy access. Yet libraries, museums, galleries, archives and other organizations which preserve and give access to our heritage are currently fighting to keep up with the wired world.

The Clinton'Gore administration has announced its intention to help link every American library to the Internet. In Europe, there are 96\000 libraries, but only a handful are networked. Today, there are perhaps 250 libraries around Europe which have Web sites, compared with 800 in the

USA (including 475 in the academic area and 200 public libraries). The rest of the world accounts for some 130 sites (50 in Australasia, 60 in Canada and 20 in South America). The potential in Europe is growing, as illustrated by the UK and French national libraries, as well as library networks in Benelux or Scandinavia.

In the museums area, it is a similar scene. Too much material is still in its original space-consuming or fragile form, stored away from the public. European museums have recognized this problem, and 400 museums and art galleries have linked up through a Memorandum of Understanding to promote multi-media access.

Archives present a difficult problem. A 1994 report of the European Commission on archives asks the question: are we to witness a loss of memory in computerized archives? The Swedish national archive, for example, has over 4\000 magnetic tapes. New problems pertain to computer law, the rapid technology developments and the sheer effort needed, which could impede access to valuable public or private archives in digital form [6].

The demand side: new user models and interaction

There are many barriers to determining user behaviour in the field of electronic publishing. It is a young activity, fragmented across many different disciplines, income brackets, geographic areas and socio-economic profiles. Nevertheless, Web technology has obviously opened new perspectives in this area.

The Web interface allows users to fill in questionnaires quickly and transmit them for automated processing to a server. The GVU (Graphics Visualization & Usability) survey by Georgia Tech [7], for example, polled 23\000 responses on general demographic queries, usage patterns and geographic location. The last survey in April 1996 indicated that Web users are not yet representative of the working population. The typical user profile is a young male, very unwilling to pay fees yet spending on average more than 20|h per week online. This survey probably attracted many university students.

Where is the borderline with invasion of privacy?

Web software can go much further than putting up questionnaires. The client'server architecture allows the supplier to plant a profile in the user's computer, called a 'cookie', for use next time they call. The problem is that users could object to personal data being generated by an external body (sometimes without the user's specific permission). Another controversial feature is that the server can check on 'cookies' left by other servers: seen by some as a time saver and by others as an invasion of privacy. This may be why recent browsers allow users to disable cookie scripts.

Encryption software is also becoming easily available, for commercial and for moral purposes. However, legislation in Europe is not harmonized and such software has to be disabled in certain countries, for example.

Web navigation: the new search for longitude

Web hypermedia is a breakthrough in online navigation, but does it really work? In the 1700s, ships could calculate latitude for navigation, but could not calculate longitude properly. This had disastrous consequences, for example when Admiral Shovell's fleet crashed into the Scilly Isles' coastline in 1707, with the loss of 2\000 lives. Today, the problem is recurring with online networks. Users follow hypertext trails like a kind of paperchase, but have no sense of overall bearing: what is around the corner? What has been missed? A common result of browsing is a meandering voyage with dead ends, backtracking and the consequent loss of time, energy and goodwill.

Search tools exist and are always improving. Directories offer single-word searches according to categories. Search engines are available to make more powerful retrieval arguments with rudimentary Boolean logic. One popular search engine indexes 31 million pages on 476\000 servers, and four million articles from 14\000 Usenet news groups. It is accessed over 29 million times per weekday.

However, none of these above techniques can be fully workable without proper indexing, abstracting or, at least, fast browsing. Improved indexing and metadata tagging techniques are needed, whether for full text, images or other multimedia objects.

Other advances in information retrieval

Apart from better-quality indexing, there is great research interest in new kinds of information retrieval interfaces where user interaction can be improved.

(i) Broadcasting paradigms such as 'push technologies' can help to limit the information overload to the users main interest areas. Systems like BackWeb and PointCast are examples.

(ii) Personalized interfaces, such as the callable personal librarian, intelligent agents or search wizards to help in navigation, retrieval, sorting, filtering, abstracting and possibly translation.

(iii) Natural language and gesture-based interfaces, relying on advanced audio-visual interaction between human and machine, speech and through gestures.

(iv) Information visualization, in order to retrieve, manage and manipulate large and complex data sets. The data sets can be represented in a highly visual way through three-dimensional 'virtual reality' interfaces;

Some of these are available today, but most are still in the research and development domain, whether in Massachusetts Institute of Technology Media Lab, in European universities or in corporate research. In all cases, these new methods need fast-track experimentation and feedback on their performance and benefits.

Is language diversity a problem?

"What is the point of having greater access to information if most of it is in a foreign language which the user cannot understand?" (Information Society Forum Report). Taking all subject

categories, around two-thirds of conventional online and CD-ROM databases are in the English language [5a]. In the science area, the figure for English is closer to 92%. Other languages covered in the science area are French (5%), German (4,5%), Spanish (3%) and other languages with 3%. This factor is probably not seen as a major hindrance to the international academic and research communities.

However, in other subject areas like commerce or culture, the language barriers are real. In 1995, the UK's Department of Trade and Industry reported that 30% of UK export companies were losing business due to the language barrier. The problems for continental Europeans, as well as for Asians and those in other parts of the world, would be bigger.

Today, over 50% of European Internet users are non-native English speakers. This percentage is expected to grow substantially over the next decade. This could explain why some consumer and commercial operations are beginning to integrate language technologies into their systems. For example, Telia has developed a 12-language multilingual browser for the Web. Speech-driven interfaces for online retrieval are being developed for European users by some telecommunications operators. The need to localize contents and software will be more pressing, and will provide opportunities for smaller content and service providers in European niche markets.

European research activities

Relevant activities to electronic publishing within the European Union's Fourth Framework Programme for Research and Technology Development 1994'1998 cover three main dimensions: information engineering, libraries and language engineering. Together they represent nearly 70 collaborative projects with several hundred partners [8].

Information engineering projects include Europe-MMM, a joint development of multimedia materials among authors, publishers and repackagers/educators for publishing multimedia educational material with rights handling, and Medform, a project for rapid multimedia publishing and analysis of the publisher organization model.

In the libraries area, LIBERATION develops electronic library services in academic environments, allowing new forms of co-operation between libraries and publishers by way of implementing relatively simple models for dealing with issues such as copyright and fair charging. A wide range of traditional and multimedia materials will be covered. Awareness of electronic copyright issues is high through ECUP (European Copyright User Platform).

In the language area, Parole concerns a large-scale, harmonized set of text databases (corpora) and lexica for all European Union languages for use in language learning and academic research. MAITS (Multilingual Application Interface for Telematic Services) develops a multilingual interface enabling users to have the language, character set and cultural environment of their choice for electronic communications. Sparkle involves multilingual information retrieval through parsing and knowledge extraction.

Future perspectives

The European Commission recently proposed an outline for the next framework programme (FP5) covering 1998'2002 [9]. Within this informal proposal, the information society plays a central role, and multimedia content is designated as a key action, covering: (i) interactive electronic publishing with new ways of creating and structuring publications and personalizing content delivery; this also encompasses the provision of cultural content, for example through electronic libraries and virtual museums; (ii) new language technologies to help to make information and communication systems more user friendly; (iii) advanced technologies for information access, filtering and analysis will help overcome information overload, make multimedia content easier to use and will also encompass geographic information and statistical systems.

Notes and references

1 Davies, D.W. and Barber D.L.A. (1973) Communication networks for computers. John Wiley and Sons, New York. The authors were with UK's National Physics Laboratory, working with the Advanced Research Projects Agency, on what later became the Internet.

2 The Information Society and the Citizen: A Status Report. European Commission, September 1996. At <http://www.ispo.cec.be>. Projected indicators for households in the five largest countries in Europe in 1998.

3 Strategic developments for the European publishing industry, towards the year 2000. (1996) Andersen Consulting, European Commission DG XIII/E

4 Information Engineering 2001 - Identification of influential technologies, impact assessment and recommendations for action. (1996) Meta-Generics. DG XIII-E4, European Commission, Luxembourg

5 Gale Research Database, Datastar host service, Switzerland. March 1997

5a Gale Research Datastar host service, Switzerland, 1996

6 Archives in the European Union. (1994) European Commission, Office for Official Publications of the European Communities, Luxembourg. ISBN 92-826-8233-1

7 Source: Georgia Tech, USA. 1996, http://www.cc.gatech.edu/gvu/user_surveys/

8 Telematics Applications Programme: Project Lists 1994'1998. European Commission. <http://www.echo.lu>

9 Towards the Fifth Framework Programme, Scientific and Technological Objectives. (1997) European Commission, Brussels, Luxembourg. ISBN 92-827-9259-5, <http://www.cordis.lu>

©Franco Mastroddi, 1997
[Copyright Information](#)