

# Electronic books and journals : challenges and opportunities

A. Endres

Technical University Munich, Germany *E-mail: endres@informatik.tu-muenchen.de*

## Abstract

Electronic books and journals are both desirable and feasible. From a user's point of view they solve some shortcomings of conventional media, but more important are the new opportunities they open. To exploit those, a number of technical challenges have to be addressed. Besides, a series of business and organisational issues have to be taken care of. The German project MeDoc had to cope with a subset of these questions and found pragmatic answers. Libraries and librarians are well advised to embrace electronic media as an important new opportunity.

## 1. Introduction

The vision of an electronic digital library has been around for while. Two of its key ingredients are electronic books and electronic journals. Others are electronic archives of images (graphs, pictures), drawings (maps, designs), audio (music, voice) and video. Software libraries are another, somewhat unique category. Because of the simultaneous growth in storage and transmission capacity the idea of a single centralized library did not get its chance. The more adequate solution is a distributed implementation, interconnecting multiple (medium size) libraries into a integrated information system, also called a virtual library.

In the following, electronic books and electronic journals will be discussed in detail. I will first show why they are useful and desirable. In the next chapters I will describe some of the technical and non-technical problems that have to be solved when trying to come up with acceptable solutions. Drawing on the experience of a major German undertaking that I am associated with I like to give an illustration of one particular solution. Finally I will try to convey my ideas as to what implications the new developments have for libraries and librarians.

## 2. Electronic books — why are they desirable ?

Hard- and softcover books with several hundreds of printed pages between the two covers play a key role in our society. They are the carrier of our literature and of most of our cultural and scientific heritage. For many people this has resulted in an emotional binding to books, unparalleled by other goods of daily life. This emotional binding can entice us to ignore or to forget most of the shortcomings that we have to struggle with when using a book. By mentioning some of these drawbacks, I will show where electronic books have a chance to come in, where their opportunities are.

Clearly, not all books contain works of art or works of literature. There are such earthly things as phone books, time tables (for trains and buses), income tax

declaration guides and airplane maintenance manuals. For them no tears will be shed if they move away from their paper incarnation. In fact, most of them have already. In the case of phone books, only the French were successful so far because they decided to provide every household with a appropriate display device, called Minitel.

If we now look at those books that we cherish so much, that we like to have around us in our office or at home. For some of them the medium paper is just perfect. In the case of Sophokles, Shakespeare or Goethe plays, we know that their content does not age. It could have been carved in stone or iron as well, not considering the size and weight problems that this might have caused. Others, like my favorite encyclopaedia I should have replaced every 3-5~years. Its 12~volumes, however, take up so much space in my bookshelf that I was forced to place other books behind each other or had to store them away in the basement.

In the above we have seen several arguments already where electronic books seem to provide a better solution:

- the possibility to keep the information current through frequent updates, -
- the reduction in physical storage space through the highly compressed and miniaturized storage medium.

There are other advantages that an electronic book has over its paper version:

- it is possible to add high quality color illustrations or fancy layouts without affecting the cost per piece,
- as soon as a copy has been located it can be transmitted with the speed of light and at almost no cost,
- if one knows a reliable source, one never will run out of copies. Distributing an electronic book does not exhaust its stock, and
- provided the appropriate software tools are available, one can search the entire content automatically for certain topics.

These advantages come to bear even in the case that an existing book is simply digitized without any restructuring or additions. The true value of an electronic book will materialize, however, if the new medium is genuinely exploited. This means that use is made of the fact that electronic storage can easily be interconnected and that most storage devices have processing power as well. The first property allows for links to be established between any entity of information and any other entity. It is the concept of hypertext that is referred to. The second property gives rise to the addition of dynamic capabilities. A math book can allow for formulas to be executed, a physics book can provide rotating bodies or a chemistry book can simulate experiments or processes. It is the door to computer-based visualization and animation of complex processes and systems that opens up.

Although the potential offered by the new medium is impressive, it will take quite some time before it will be fully exploited. As far as books are concerned, the migration to electronic forms will start out with those types where the content is either rather volatile (like legal texts or technical manuals) or where the content is more an

accumulation of pieces (like in an encyclopaedia) then a consecutive story (like in a novel). Adding executable examples or even multimedia extensions (audio and video) to a book is very time consuming on the part of the author. Only few single authors will either spend the time or develop the skill to become masters in that respect. This is a field, however, for teams of authors. The pressure to move in this directions will be highest in those sciences where dynamic processes play a key role, i. e. physics, chemistry, biology and medicine. For others, it will be quite hard to justify the additional effort.

In my field of speciality, computer science, I am aware of a few dozen electronic books. They are in different stages as far as their useability is concerned. Some are simple printable texts, most of them have some search facility associated with them, some are true hypertext documents, very few have dynamic capabilities.

### **3. Electronic journals — why are they here?**

While electronic books are frequently discussed as a controversial issue, this is not the case for journals. There is wide agreement that electronic journals are the way of the future, at least for the majority of the scientific journals. Why is the situation so different?

Journals play quite a different role than books. They are the carriers of highly time dependent, volatile or specialized information. Of course, not all journals are the same. There is one category that is referred to as trade journals. They contain product related information for people in industry and other professional positions. Their essence are the most recent announcements from vendors, product evaluations and method descriptions. They also carry a heavy load of advertisement. A completely different category are the newsletters issued by professional societies for their members. Again, the information is factual and time dependent. The category of which people are most concerned about are the scholarly journals. They are either produced by professional societies or by commercial publishing houses. For them, timeliness may be less important, but usually they are very specialized, i. e. they are covering material that is of interest to only a few hundred people working in that field.

Today's journals have the following disadvantages, if I take the perspective of either a reader or an author:

- to stay abreast of what is going on in my field of speciality I have to follow what is published in about 20 different journals,
- of the journals for which I have a subscription only a fraction of the papers is relevant for my work,
- international journals usually arrive with a 2-3 months delay unless I pay for airfreight,
- if a field is very active, papers are queuing up for certain journals. This may result in the creation of additional journals with further limitation of scope, and
- the more specialized a journal becomes the higher the price.

Electronic journals solve some of the above problems in the following way:

- postal delays no longer exist,
- it is easy to search the content of a higher number of journals and to obtain individual papers quickly,
- journals supporting a small community of readers are economically feasible, and
- individual papers can be made available without waiting for the deadline of the respective issue.

In addition to that, all the advantages mentioned for electronic books also apply.

In computer science, there are several hundred electronic journals in existence today. There will be more every week. Because of real (or assumed) problems in accounting, some of the publishers of paper-based journals only offer the electronic version to subscribers of the paper version. This reminds me of the development in the software industry, where software for a while was bundled with hardware. It is a temporary solution only.

#### **4. Technical problems and challenges**

The field of digital libraries has become a very active area of research and development. As in most other areas that relate to new electronic media, contributions from different disciplines are required to come up with optimal solutions. In the following, some problems and challenges are discussed that require technical solutions, i. e. solutions to be provided mainly by computer scientists.

##### **4.1. *Digitization and conversion***

For many fields of study it is necessary to recapture the material that may exist in analog (i. e. paper) form only and transform it into electronic and digital form. This is particularly true for all historical documents. After the document has been scanned it may be necessary to add some descriptive information to it so that it can be retrieved automatically. In some cases it may be possible or advisable to derive the descriptive information through the help of character recognition (OCR) software. Unfortunately, the quality of these routines is not good enough to convert large bodies of text without manual intervention. The difference in storage consumption may, however, make the conversion from pixel image to coded representation desirable. For most contemporary publications, the author has already used electronic text processing means to produce the source material, even if the book or article was only published in paper form. In this case it is advisable to start from the authors electronic source and convert that into whatever format is required for electronic storage and dissemination.

##### **4.2. *Storage media and systems***

Digital libraries require large amounts of storage. Fortunately, disk type storage in the range of gigabytes is feasible for individual workstations. What is required in addition at some central server station are storage devices that can provide backup capability for multiple users or handle input from multiple CD-ROM readers simultaneously (devices referred to as juke boxes). To administer the large text or multimedia objects

traditional database management systems (DBMS) may not always suffice. A new category of fulltext or multimedia DBMS is often the more appropriate solution.

#### **4.3. *Formats for storage and representation***

One of the most difficult problems seems to be the selection of the appropriate formats for storage and presentation. Because an electronic document is amenable to entirely different modes of usage no single format is optimal. Presentation on the screen requires other formats than printing on paper, and neither of them may be optimal for either long-time and compressed storage or for content-based searching. It appears that the most successful approach uses a storage-efficient format like SGML that preserves all information that is needed for on-the-fly conversion to screen or printable format. If the problems are complex for pure text corpora, they reach another dimension if formulas, graphics, images, audio and video are involved. There are dozens of defacto, de jure and proposed standards that are applicable. They both alleviate and complicate the situation.

#### **4.4. *Classification, indexing and retrieval***

The problem of classification is well known to every librarian. If it was a bottleneck already for a traditional library; it will be much more so for a digital library, particularly if it will contain more than just text material. While classification requires subject-related knowledge, the indexing task is more of a syntactical nature. Here the processing power of modern computers are brought to bear as some of the widely used search engines demonstrate. Again there are large gaps between text-based indexing and searching and indexing of non-text objects. For whatever material is properly described and classified, automatic retrieval is the outstanding advantage of electronic libraries, be it in a single-location digital archive or in a network-based world wide distributed virtual library. In the latter case, the individual sources need to be described and classified.

#### **4.5. *Delivery, presentation and usage***

For documents stored as fulltext delivery is easy, provided the necessary communication bandwidth exists. Here some countries are making big strides. It may take some time, however, before audio and video delivery of good quality is achievable outside metropolitan areas. The presentation of delivered objects has to fit into the users working environment. Here the various browsers for the World Wide Web (Netscape, Explorer) are establishing defacto standards. For some material it is not enough that it can be presented or viewed. It must be processable by the local system, i.e. compatible with the available environment and platform. This was the problem addressed by the team that provided the new interpretive approach called Java.

#### **4.6. *Privacy and security***

One of the greatest challenges for electronic commerce is the need to assure that transactions are done reliably and exactly between the partners that want to be

involved. This is the problem of authentication, i. e. proving ones identity without seeing each other. If the nature of the transaction or the details of it should not be observable by others this requires security measures of a certain type. When access to the capability or resources of a network should be controlled this requires other measures. Although academic networks are usually somewhat liberal in this respect, technical and organisational means are available to improve the situation. In some cases, the advent of electronic digital libraries motivates people to take steps in that direction.

## **5. Business, organisational and legal issues**

Technical problems are certainly only one side of the coin. Business, organisation and legal issues may easily become a stumbling block, delaying the application of new technologies. In our case, I believe that the pressure of the maturing technology is such that they are being addressed in due course.

### **5.7. *Cost of conversion and operation***

It would be a mistake to ignore that the preparation of material for use in electronic books and journals will be possible without significant additional effort. This is particularly true if attempts are made to exploit the capabilities of the new medium. Also the costs for the operation of a network of servers is not negligible. In an academic environment not all costs may be visible, making people to believe that they do not exist.

### **5.2. *Distribution and prices***

Whenever costs have to be recovered, pricing becomes a critical issue. We are well accustomed to associate prices with physical goods we want to acquire. Therefore producers of electronic books encounter the least difficulties if they offer their product as CD-ROM. They fit nicely in the bookshelves of bookstores or of supermarkets, and can be ordered, stored and accounted for in the same way as other goods. Things are different for online, i. e. network-based offerings. In this case, the distribution and pricing models that seem to work resemble those of a utility, be it electricity, phone, water or sewage. There is usually some fixed base fee which gives access to the service and some variable part which reflects the actual resource consumption or value obtained during a certain time period. The base fee would correspond to the annual subscription price of a journal, the variable part to the copy fees of a delivery service. Budgeting and accounting for the latter part may not be easy. It is therefore often avoided or circumvented.

### **5.3. *User rights and vendor obligations***

There are people who request that copyright law should be revised for electronic media. It is certainly true that some of the terms used (like copy, publication) have meanings that come from traditional analog media. These terms need to be clarified. For the term 'copy' a similar operational definition has to be found as was

done for software. Here some recent revisions of copyright law allow that a backup copy can be made without permission of the copyright owner.

On the other side, the vendor must assure that whatever he offers is clear of others peoples' rights and that he has made best efforts to transmit and present the author's work exactly and completely. I am concerned when I hear: since it is so much easier to steal other peoples work one should lower the legal barriers as well. Or put differently, why keep up the threat of prosecution if a high percentage of violations will not be detected. I have heard the same argument in favor of software piracy. I do not like it there either.

#### **5.4.      *Quality assurance***

In a network world, everybody can be his or her own publisher. There is no need to go through a reviewing process before putting something out for other peoples' consumption. So the quality level of electronic publications will, on average, be lower than that of paper publications. Statistically speaking this has to be so, since additional publications will see the light of the day that otherwise would not have made it. The only question is, which is the lesser evil, to starve or to have to make a choice.

For me, as editor of a journal that appears both on paper and electronically, there is no doubt that refereed journals will continue to be refereed. There will be new refereed electronic journals and new unrefereed publications. In fact, the refereeing process can be strengthened by involving more reviewers, and accelerated at the same time, thanks to electronic transmissions. For a scholar it becomes more critical to learn how to distinguish gold from ore, or chaff from wheat (as we say in German). This is what his/her education was supposed to achieve, anyhow. For students and public at large, scholars, teachers and librarians are called for to help and to offer their advice.

#### **5.5.      *Archiving***

A problem mentioned frequently in connection with electronic publications is that of archiving. What needs to be done that later generations (or historicans) are able to access important documents that were published electronically only? I do not believe that everything that is written on paper needs to or can be preserved indefinitely. The same applies to electronic documents. There certainly are legal documents (digital maps, contracts, patents) or pieces of art, literature and science that should be preserved. This, in my opinion, is a new responsibility of archives and libraries. I know that for Germany the German national library (Deutsche Bibliothek) has taken on this task. It involves selecting memorabilia, and porting them from one technology to the next, whenever that becomes necessary.

### **6. The MeDoc approach**

MeDoc (Multimedia Electronic Documents) is one of the several model projects funded by the German ministry of technology (BMBF) in the area of digital libraries. It is part of a government initiative to advance new information technologies in

science and engineering. MeDoc was started in September 1995 and involves 9 research partners and 24 pilot users, mainly from academic institutions. The project is lead by an alliance of three partners, the German Informatics society (GI), the FIZ Karlsruhe and the Springer-Verlag. Its main purpose is to build a prototyp of a digital library for computer science literature. In the following I will

highlight some aspects of this project by describing the approaches taken with respect to content, technical structure, business and legal considerations. The project is still going for at least one more year, meaning that some of the approaches may change.

### **6.1. *Library contents***

The MeDoc library will contain fulltext versions of some 30-40 journals, some 80-100 books and several 1000 technical reports. Some journals are back issues only, most of them include current issues as well. Journals will be presented in a variety of formats, including HTML, PostScript and PDF. Most books will be converted from the authors electronic source format (usually Word or LaTeX) to either HTML or PDF. Several will have multimedia supplements, i. e. animations or videos.

### **6.2. *Technical structure***

The MeDoc library will initially be distributed over six server locations. They are chosen based on networks capability or competence for technical support. The library content will be spread equally among the servers, providing for some mutual backup. Each server will be equipped with the same software to handle user access, accounting, and user statistics. A major new function will be an information broker, a software component that assists the user to find appropriate sources for information, inside and outside of MeDoc. It will provide access to library catalogues, reference data bases and other network based information resources.

### **6.3. *Business and legal considerations***

Most of the electronic journals and books offerd in MeDoc are titles owned by commercial publishers. The MeDoc project has aquired the rights of electronic use for a predefined group of project participants (research partners and pilot users). These titles will be offered in a fee structure which was specifically negotiated for the project. It normally consist of a base fee for electronic access (subscription) of the document and an additional page-based fee for printing. Users obtain the right for repeatable access, but for personal and non-commercial use only. Technical reports are normally free of charge, as they are typically offered by university or researche institutes.

### **6.4. *Long term considerations***

Although MeDoc can be considered as a large scale feasibility study, it is planned to continue the established service after the end of the project (August 1997). Details for the organisational setup are still to be negotiated. In this respect it is also important to find out whether the concepts developed can be applied to other fields like chemistry, mathematics or physics as well.

## **7. Implications for libraries and librarians**

As indicated before, electronic media become a very important complement of traditional media. It would be wrong to ignore this. As a consequence libraries and librarians are well advised to embrace electronic information as a part of their business. As far as journals are concerned, the directions taken by publishing houses (like Elsevier and Springer) and professional societies (like ACM and IEEE) are well known. In the case of books, more questions are open and more experimentation is going on. More cautious steps (electronic supplements) or temporary detours (distribution via CD-ROM) can occur.

Nevertheless, librarians should bring in their experience and help to shape the solutions in a way that they suit their customers. Electronic media will not be the end of scholarly or scientific publication. The contrary is true, they will accelerate the process even more. With increasing numbers of publications, the librarian's function will become more important. His or her role will change, however. There may be less time needed to physically locate, procure or distribute objects but increased demand for professional advice to users and for intellectual contributions to the never ending task of structuring, classifying and explaining the universe of human knowledge.

## **8. Concluding remark**

As can be seen from history, the appearance of a new medium or a new technology seldom completely replaces an existing medium or technology. It forces, however, the older technology to concentrate on its strengths and to abandon those applications where it was weak. As an example, automobile driving did not completely eliminate horse riding, nor did flying replace driving. Three examples that are closer to our field are photography versus painting, radio versus newspaper and movie versus life theatre. I **still** read a printed newspaper every morning. Maybe it allows me to determine myself the speed of receiving information. My capacity to draw or to paint, however; cannot substitute for the hundreds of pictures I am taking every year. And once in a while I enjoy life theatre or opera.

Books and journals on paper will survive, but they will serve different functions than today. They will be closely designed and styled for the purposes they will serve. This will be a narrower purpose than today. On the other hand, information stored in electronic form will have to serve multiple purposes. It will not only be multimedial, it will be tailorable, ubiquitous and interactive. We simply will have many more options.

## 9. References

The following references are thought of as a first introduction into the subject. They are structured by type of activity or source.

### A. Plans of professional societies

- American Chemical Society: Will Science Publishing Perish? The Paradox of Contemporary Science Journals. Special discussion paper of the ACS 1995
- American Mathematical Society. Electronic Publications: <http://www.acs.org/committee/publications/>
- Denning, P./ Rous, B.: The ACM Electronic Publication Plan. Comm. ACM 38.4 (1995) and [http://www.acm.org/pubs/epub/\\_plan.html](http://www.acm.org/pubs/epub/_plan.html)
- IEEE Computer Society Publications: <http://computer.org/publications/>
- Institute of Electrical Engineers, Publishing and Information Service <http://www.iee.org.uk/publish/>

### B. Plans of commercial publishers

- Elsevier Science, Amsterdam: <http://www.elsevier.nl/homepage/news/1996/science>
- Springer-Verlag, Heidelberg : <http://link.springer.de>

### C. Ongoing research projects

- Special Issue on Digital Libraries, Comm. ACM 38,4 (1995)
- European library projectsW <http://wwwz.echo.lu/libraries/en/>
- Special Issue on Digital Library Initiative. IEEE Computer May 1995
- MeDoc — The electronic computer science library. <http://medoc.informatik.tu-muenchen.de>

### D. Sources of E-Books and E-Journals

- Books and journals online : <http://www1.informatik.tu-muenchen.de/local/proj/Medocl/>
- Survey of electronic journals : <http://journals.ecs.soton.ac.uk/survey/>

### E. Other activities and positions

- Adam, N. R., Holowczak, R. et. al: Digital library task force. IEEE Computer, August 1996 and [http://cimic3.rutgers.edu/ieee\\_dltf.html](http://cimic3.rutgers.edu/ieee_dltf.html)
- Association of Research Libraries. Definition and purposes of digital library <http://sunsite.berkeley.edu/ARL/definition.html>
- The British Library Online Services : <http://portico.bl.uk/access/overview.html>
- Conferences on theory and practise of digital libraries, 1994 and 1995 Proceedings <http://www.csdl.tamu.edu/DL94> and [DL95](http://www.csdl.tamu.edu/DL95)
- Odlyzko, A. M.: Tragic loss or good riddance? The impending demise of traditional scholarly journals, J.UCS Vol. 0, No 0 Nov. 15,1994 (Pilot issue): <http://medoc.springer.de:8000/jucs>