The i2010 digital libraries initiative: Europe's cultural and scientific information at the click of a mouse

Horst Forster

APE Conference
Berlin, 23 January 2007
The digital libraries initiative

- Part of i2010, the Commission strategy for the information society
- Letter of 6 Heads of state and government of 28 April 2005
- Aim of the initiative: build on the potential of ICT to improve access for all
- Two main blocks: cultural content, scientific information

- Ladies and gentlemen. In September 2005, the Commissioner for Information Society and media, Mrs. Reding, launched the digital libraries initiative. The initiative is part of i2010, the broader Commission strategy for the information society. It follows up a letter of 6 Heads of state and government, together representing two thirds of the population of the European Union. In this letter the Heads of state and government asked the European Commission to take steps to make Europe's cultural and scientific heritage accessible online and to work towards a European digital library. The Commission has reacted quickly in order to make concrete and rapid progress.

- The aim of the digital libraries initiative is simple but ambitious: to build on the potential of information and communication technologies to improve the access to information for all. This is also the guiding principle for the two main blocks of the initiative. The first block deals with cultural content, the second deals with scientific and scholarly content.
Access to cultural content

- Wealth of content in Europe’s cultural institutions: books, newspapers, films, archival records
- Using the Internet to make it more accessible
- At present only a small percentage digitised
- Member States have embraced Commission approach

Access to cultural content

- Before I come to the issue of scientific information - which is at the heart of this conference - let me say a few words about our actions to make cultural content more accessible.

- Europe’s cultural institutions - libraries, archives, museums - contain a wealth of materials representing the richness of Europe’s history, acquired over the centuries. These materials cover a range of forms - books, newspapers, films, photographs and maps. To take the example of books: there are roughly 2.5 billion books in the libraries of the EU25.

- The Internet is increasingly the information source privileged above all others. Can it be used to make Europe’s cultural heritage more visible and accessible? We think it can. But a lot has to be done. Accessing the material through the Internet presupposes the existence of a digital copy; but at present only a very small part of the relevant
collections – less than 1% – has been digitised. A major effort by the Member States is necessary to digitise these collections and we see that an increasing number of Member States is actually willing to make such an effort.

- Member States, through their Ministers of culture have embraced the Commission approach to the European digital library and to the broader issues of digitisation and online accessibility. If all works out well, by 2010 European citizens will have online access to a large, diverse and ever-increasing share of Europe’s cultural heritage.

### The European digital library

- Common access point to Europe’s cultural heritage
- Build on existing TEL framework
- Targets:
  - By 2006: Europe wide collaboration between libraries
  - By 2008: 2 million objects searchable and accessible
  - By 2010: 6 million objects searchable and accessible, content of all types and from different institutions

### Building the European digital library

- The availability of digital content does not automatically mean you can find it through one single access point. And that is what a user would like. We are therefore working with cultural organisations towards a real European digital library. It will build upon the organisational
framework of TEL (the European library) set up by CENL, the conference of European national librarians.

- We have set some concrete targets. By 2006, there should be EU wide co-operation between national libraries with the aim to build the European digital library. This target was met, since the collaboration is now in place. By 2008, there will be at least 2 million digital objects ready to be searched and accessed through the European Digital Library. By 2010 this number must have risen to at least 6 million books, newspapers, journals, films, photographs and archival documents.

- This does not only concern objects from national libraries but from all kinds of cultural organisations at all geographic levels. After all, Europe's heritage is not confined to a few privileged large organisations but has been collected over the centuries in many regional and local archives, libraries and museums.

- Recent experiences have shown that there is great and growing interest for this type of material. When the British Library first made its Gutenberg Bible available on the web in November 2000, the pages received one million 'hits' in the first six months. In 2006, the French 'Institut national de l'audiovisuel' made 10 000 hours from its audiovisual archives accessible online. In the first few days, 6 million hits overloaded the service.
Towards a 20th century black hole?

- Just public domain material?
- Risk: absence of 20th century on the Internet
- Simple licensing systems for out of print
- Solutions for orphan works
- Attract private content holders to make their content accessible through the European digital library

Avoiding a 20th century black hole

- Digitisation and online accessibility is relatively straightforward for material that is in the public domain – that is: material not covered by intellectual property rights. But we cannot afford to limit the European digital library to public domain material. For many young people what you cannot find on the Internet does not exist. We should avoid a situation in which all information from roughly before 1900 is available and in which the 20th century is absent.

- Obviously digitisation and online accessibility has to be achieved in full respect of legislation on intellectual property rights. But we are looking together with publishers and libraries to simple licensing systems for out of print works and solutions for orphan works – where it is impossible to locate the rightholders. We also expect the European Digital Library to attract publishers and other private content holders to make their content accessible and searchable online through the common access point.
Access to Scientific Information.

- I now come to the second block of the digital libraries initiative that deals with scientific information. And when I talk about information I mean journal articles, research results and relevant research data. The Commission is about to publish a Communication – a policy statement – that will deal with the essential topic of access to and preservation of scientific information. We have prepared this Communication jointly with our colleagues in the directorate general for Research. It looks at the changes brought about by the digital environment in the ways in which scientific information can be accessed now and stored for future use.
- Let me first of all say a few words about the role of the European Community in this area: Research and dissemination of research information are broadly international endeavours. They are European, or rather global, but certainly not just national. The Commission is
itself a major financing body, with a budget of some 50 billion EUR in the 7th Framework programme. Furthermore access to scientific information is part of our policies for the information society and for improving the conditions in the European research area.

- The issues related to scientific information are different from the ones in the area of cultural content, where digitisation is a key theme. In the area of scientific information, a lot of old issues of journals have been digitised and can be searched in databases. Practically all 'new' scientific information is produced in digital format. 90% of scientific journals (STM journals in English) are available online, mostly along with a printed version. A study done for the British library indicates that by 2020, 90% of all journals will only be available in digital format.

A polarised debate

- So does this mean that all is well? That depends on whom you talk to. Over the last few years we have had extensive discussions and consultations with stakeholders. These contacts show that positions have become rather polarised.

- Roughly speaking you have on the one hand part of the scientific community, some funding bodies and major libraries that question the simple transfer of old practices to the new environment. They advocate immediate and free access to research results on the Internet. The Internet environment has offered new possibilities, so why not use them?

- On the other hand you have the scientific publishers that underline the merits of the present system. These are commercial publishers, but also learned societies, publishing on behalf of or for a specific research community. Scientific publishing is an area of activity where Europe is doing really well and has outpaced other continents.
Europe's scientific publishers employ 36,000 people directly. Publishers realise things are changing, but simply do not see a sustainable alternative way of publishing and even less so a clear roadmap towards such a new model.

- What we observe is that the different stakeholders are sitting in their trenches and shoot at each other. We would like to help building consensus about the way forward. Such a consensus should take into account the strengths of the present system, but also the need to better exploit the opportunities of the information society.

A heated debate...

Those who advocate open access:
- Beneficial to the research system
- Triple payment argument
- There is an access problem

Publishers:
- Publishing has a cost
- The present system is efficient
- No undue government intervention that puts the peer review system at risk

Arguments in the open access discussions

- Let's have a look at the arguments used by both sides. Those that support open access are using roughly the following arguments
  - Free access to all scientific information through the Internet will benefit the scientific system and make it more efficient.
The public purse pays three times: for producing the results and writing the article; for the peer-review by professors; and for buying the results through libraries. Therefore the public sector should have a bigger say in the system and not just rely on rules dictated by the publishers.

Prices have gone up so much over the last years that there is an access problem.

- The publishers put forward a number of counter-arguments:
  - Publishing has a cost and someone has to bear it. The 'no-one pays' model does not exist.
  - Publishers have made the system more efficient with the new ICT tools and access to information has greatly improved.
  - The third key argument is related to the peer review system. At present results are consolidated in articles which are then submitted to colleagues for a thorough check. Once this stage has been passed, you can be reasonably sure of the value of the results. Changing the present model could put at risk this well working peer review system.

- Abstracting from these arguments, there seems to be an agreement on the two basic goals to be achieved: a rapid flow of information and guaranteeing quality. In the Internet environment, these goals do not always go hand in hand. A rapid dissemination of information that is not properly checked can lead to the trap of 'Internet superficiality'. I said before that today many children assume that you can find everything about the world on the Internet. What is more, they also assume that all you find there is true. Obviously this is not the case. To translate that to the scientific world: how can you assure that researchers and students get their information instantly but only read what is worthwhile reading?
• The upcoming policy statement of the Commission will deal with the issue of access to scientific information. It will announce a number of actions and some experiments. The aim is to accompany in a realistic way processes that are under way and to facilitate an organised debate at European level between politicians, and between stakeholders.

**Digital preservation**

• Major theme throughout the digital libraries initiative
• Why preserving digital information?
  – Mind the gap
• Economic potential
  – Investment in new technologies
• Commission action on digital preservation
  – Co-funding projects
  – Raising the political profile of the issue

**Digital preservation**

• The policy statement will also draw the attention to another topic which needs to be urgently addressed. That is: digital preservation or, if you wish, the preservation of digital material. If you look at it from the preservation point of view, the 'old technologies' were amazingly efficient. Take the example of clay-tablets. They easily lasted for thousands of years. Paper is not so bad either, lasting a few hundreds of years. In comparison: the average life-time of a web-page is 44 days. This is about the average lifespan of a housefly.
The issue of digital preservation is a major theme throughout the digital libraries initiative. It is relevant for cultural material that has been digitised, it is relevant for born digital scientific information and in fact it is crucial for any organisation that produces and stores information.

The problem is that digital information is much less stable than paper. Hardware and software change very rapidly, which can make the information unreadable. And storage devices (CD-ROMs) have a limited lifetime. So in many cases constant migration of the information is necessary to keep it alive over time.

And migration does not simply mean making a copy, but unfortunately you also have to check the result and sometimes adjust it. If you lose a few accents in the migration of a French text to a new software that may be annoying. But in other cases little changes can be crucial. An example: navy engineers working with diagrams of the propulsion, electrical and other systems of an aircraft carrier discovered that there were subtle changes when you opened them on newer versions of the software. The changes were little, for example a dotted line instead of a dashed line. But this kind of changes could be critical for the operations of a ship that was powered by two nuclear reactions. Not exactly the type of operation where you would like to take any risk.

When you talk about digital preservation, obviously a first question is whether we need to keep all the information produced by researchers: all the research data, all the draft articles, all the relevant websites. Not all would be my answer, but perhaps more than we now think. You need to keep the data of experiments to repeat them or check their validity. You need information from the past to study phenomena over time. And you need to keep information to consider results in
their context. For example: could this company have known, given the state of the art that their product was dangerous?

- If we do not actively pursue digital preservation now, we risk having a gap in our intellectual record. No one seems to feel responsible for keeping and storing digital research data.

- Legal deposit laws are being updated in most Member States to take into account digital material and not just paper books and journals. But still the risk of gaps remains. Within 50 years we may have lost relevant information or need to do very costly recovery work that could have been prevented.

- This is the risk of not-preserving, but digital preservation also presents considerable business opportunities. There is an untapped industrial potential, with service providers emerging and new preservation technologies being developed. This is an area where Europe has a chance to take a lead, if the necessary investments are made.

- Indeed, research into this area is crucial, for example to find ways to handle and preserve large volumes of data and/or dynamic content. Within the 7th Framework programme for R&D, digital preservation is one of the priority areas where my directorate will co-fund projects.

- Next to co-funding relevant projects, the Commission is trying to get the issue of digital preservation more firmly on the political agenda. A recent Commission Recommendation (August 2006) asked the Member States to address digital preservation in a structured way: through action plans with a concrete definition of responsibilities and target outputs, and by updating the relevant legislation (for example the legal deposit legislation and the copyright legislation). The upcoming policy statement on scientific information will stress the need to take up the challenges also in this area.
Conclusion

• Work ongoing towards a real European digital library
• Scientific information: different appreciations on the way forward
  – Experiments needed...and dialogue!
• We are only at the beginning of addressing the challenges of digital preservation
  – Think in term of costs, but also benefits

Conclusion

• I come to my conclusion. Our actions on scientific information are part of the broader digital libraries initiative that aims to improve access to information in the Internet age. Within the initiative, we are working towards a real European digital library of cultural material of all types. And we are addressing the legal and other conditions that are relevant for a better access to cultural material.

• As for the dissemination of and access to scientific information, there are different appreciations on the best way forward. Indeed there is a lot of tension between stakeholders which could stand in the way of sensible solutions to improve access. In the coming years we have to experiment with new models that improve access to the information, while building on the strengths of the existing system. And we have to get a dialogue going between stakeholders to ensure that energy is devoted to improving access and not to fighting each other.
• As for the digital preservation issue, we are only at the beginning of addressing it. The main problem here is to make people at all levels aware of the organisational, technical and legal issues at stake. We have to take up the digital preservation challenge, not defining it only in terms of costs and problems, but also in terms of benefits and economic opportunities.