

Organising the Digitisation of Collections at a National or International Level: Models for Research Library Development

by DANIEL RENOULT

Libraries today, amongst other research institutions, are solicited more and more to take part in programmes of digitisation of their collections. Apart from the way these programmes affect the public image of the institution and the prestige that is to be gained, participating in these programmes leads us to consider not only technical decisions but above all long term strategic choices.

I have been asked, as the person in charge of the digitisation programme of a major European research library, to present today some of the elements which characterise these strategic choices and to launch a discussion on this complex question which has already been raised in this very place during the „Digitising Journals“ conference in March. First of all, I take this opportunity to thank the *Liber* conference organisers for entrusting me with such an honour.

I propose to approach this question from three successive angles:

- the economic, legal and scientific context,
- economic development models,
- the global collection concept and sharing responsibility for content.

I shall be illustrating my remarks by referring to some projects and I have deliberately avoided limiting these to Europe. I shall present the choices recently confirmed by the Bibliothèque nationale de France not as a model but as an example amongst others, and I hope that we will have time to discuss together the various problems.

1. THE ECONOMIC, LEGAL AND SCIENTIFIC CONTEXT

While not insisting on the point, which is well known to most of you, we must remember that the development of digitisation is seen today in an exceptional context.

The first aspect of this context is the growth of what is known as the new economy whose driving forces are the multimedia industry (film, television), the telecommunications and the computer industries of which an important factor is of course the Internet. In an atmosphere of technological and financial euphoria we hear every week of new mergers and the emergence of new ventures: new forms of economy and communication materialise in a highly competitive context where the sums of venture capital are measured in millions or billions of dollars. To take a recent example, the objective of the new Vivendi-Universal group which associates Havas, Canal+ and Vivendi, Universal Pictures (film studios) and Universal Music (music industry), is to generate an annual cash-flow of 65 billion dollars. This recent example of alliance illustrates the vertical regrouping of these new industrial sectors. These companies which are developing today and tomorrow's means of communication and exchange are looking for multimedia contents, which in library language we call *collections*: catalogues of discs, of films from the film industry, of images from photographic agencies etc. A common characteristic of the strategies of these groups is indeed to propose a package of value-added services which are thus highly profitable: it is not just a question of providing „channels“ or machines but above all consumer products in fields such as recreation (sports, games), information and culture (encyclopædia for example) which will be distributed through multi-access Internet portals, i.e. available just as easily through new-generation cellphones as through television, personal computers or e-books. In this field, the collections of libraries and museums appear immensely rich to these companies: books often out of print, in the public domain and often difficult to obtain second-hand, images (prints and photographs) and even in some cases sound recordings, but also reference data (catalogues, bibliographies, databases) allowing efficient identification of these contents which, to the eyes of the multimedia industry constitute raw materials which are freely accessible or at most cheaply available¹.

The second element, which is closely linked to the first, concerns the evolving legal situation and in particular the question of authors' rights. While this is a subject of debate at a national level arbitration is today international and as far as we are concerned European. Even more than for books, the question of periodicals audio-visual material is decisive as can be seen with the debates on journalists' author's rights, or with the court cases brought over the diffusion of MP3 format musical recordings over the

Internet. Whatever the results may be, we can be sure that they will tend towards affording better protection and equitable remuneration of owners. Let us not forget either, that, over and above the question of authors' rights, other questions such as telecommunications operation and therefore network tariffs also depend on the way international and EU law evolves.

The development of digital publication forms the third element of the context. Let us first note that more and more mainstream publishers are embarking on the digitisation of their own collections. In addition to the ventures of such already highly specialised companies as, in France, *Bibliopolis* who have just signed a contract with the publisher Gallimard², the major groups are investing in the digitisation of their catalogues: in Paris, Havas admits to having already invested 70 million francs in digitising while at Amsterdam Reed Elsevier, which represents 27,000 employees, has already announced an investment of £80m per year for the last two years in making publications available on line.

What is less known and often underestimated, because less visible in the media, is the development by research laboratories of specific digitisation programmes which are in general independent of the initiatives of both publishers and libraries. During a recent conference on digital library content organised in Paris by the BnF and the New York Public Library³, it became clear that scientists are becoming more and more important in perfecting content search engines, including the human and social sciences. The digitisation of manuscripts in image and text modes under the guidance of researchers is making important contributions to scientific knowledge. We know for example that this type of research has enabled the recovery of fragments from the pre-Socratic philosophers, which had previously been thought to be lost, and our knowledge of the medieval corpus has been widened. Closer to the present, erudite work on the texts of Rabelais, Voltaire or William Blake, making use of digital resources, has thrown new light on works of literature and enabled the transformation of the notion of scholastic text editions. The use of computing techniques for research at any rate enables us to witness the fact that there is no comparison to be made between a hasty fac-simile reproduction of a document by a publisher out for short term profit and a serious publication or republication of a text. We may note in passing the meeting of the research policies and the development policies of libraries could throw much light on choices to be made in digitisation: this is one of the first points that I would make on the subject of a global strategy.

2. CONSTRUCTION OF ECONOMIC MODELS

In this rich, fluid and competitive context, it is hardly necessary to point out that the financial weight of libraries is small, as is the influence that they can bring to bear on legal developments. The choice of an economic model of development for collections and services thus becomes an even more important strategic factor. Indeed, even for a project limited to a specific thematic corpus, the investment required for digitising a collection is an important element of the budget of the average library, and the necessary resources especially in terms of personnel for the functioning and maintenance of digital libraries are considerable. While accepting the risk involved in a certain simplification, I will restrict myself to citing three main models illustrated with some examples.

The first development model is a classical economic model founded on market analysis and profit requirements which enables funding of growth and distribution of dividends.

We are concerned here with digitisation projects specifically associated with libraries and for this we may turn to a recent and very interesting example with the FATHOM project⁴ which brings together institutions of higher education (*Columbia University, London School of Economics and Political Science*), research libraries (British Library, New York Public Library), museums (Smithsonian Institution's National Museum of Natural History) and a publisher (Cambridge University Press).

The object of this project is to make available over the Internet a subscriber service portal allowing access to university courses and documents held both by libraries and by museums. The FATHOM project is clearly aimed at a clientele concerned with education in English speaking countries, a market whose economic potential is judged to be very promising. According to some sources⁵, the American market for online education will already have reached a value of 6 billion dollars in 2002 and 9 billion in 2003, and could reach 30% to 35% of the American higher education market which is today valued at about 750 billion dollars. One of the strengths of the FATHOM project is in associating university courses, a classical educational product, with the provision of digitised documents on line. More than a simple remote university, the ambition is to build an *interactive service for the distribution of knowledge on an international scale and on the basis of the market economy*. This idea is not limited only to the FATHOM project: it is behind a number of competing projects from universities on the Internet. Without going into the details of the FATHOM project, which is not our aim today, we may simply note that the juxtaposition of digitised collections, however prestigious they may be (Magna Carta, photographs of New York, the notebooks of an anthropologist), may be an indispensable initial stage but

does not in itself constitute an interactive educational service. In-depth editorial work is necessary to advance from simple collections of documents to an interactive application of this breadth.

The second development model is one founded less on profit than on the principal of self-financing and is most often less directly oriented to a market economy than towards objectives of general scientific interest. This type of model has tended to be favoured in North America. Fairly typical of this economic strategy are the projects developed with the support of The Andrew W. Mellon Foundation, such as the *Journal Storage* programme (JSTOR) or those of the *Research Library Group* such as *Cultural Material Initiative*.

As with many projects for digitising scientific journals and associating libraries, the objective of JSTOR⁶ which was presented during the *Digitising Journals* conference⁷ at Copenhagen in March, is to encourage co-ordinated digitisation of scientific periodicals, and at the same time improve their access and conservation in libraries. However, even if the initial investments have been financed in part by the Mellon Foundation, the principle is one of a self-financing project. Access to JSTOR is therefore payable through annual licensing whose cost is calculated from the size of the subscribing institution which, in turn, is free to recoup the costs from its users if it should choose to do so. JSTOR today provides access in text mode to 117 university collections of English language periodicals.

The *Research Library Group*, aiming to share access to the digitised collections of its members, has recently launched the *Cultural Material Initiative*⁸. The idea is to create an Internet portal giving access to multimedia documents from a single search engine. For example, it is possible for a researcher working on the sculptor Rodin to find both articles about « The Burghers of Calais » and also archive documents or images of the sculpture from the participating museums. Payment for access represents one of the sources of funding for the development of the digitised collections.

The third and final development model for digitised collections is that of local, national or international public funding. The services are in this case available free of charge on the Internet either to a restricted community of researchers authenticated by means of passwords, or unrestrictedly to all users. This economic model which mirrors the principle of public financing for research, is the most frequently encountered model in Europe. It concerns above all public institutions, scientific research establishments, libraries, museums and archives. There are numerous examples. Since we are today at Copenhagen we may cite the *Danmarks Elektroniske Forskningsbibliotek* programme⁹ which is funded at a national level with a budget of 200 million crowns for the period 1998-2002. It is also thanks to public funds that the Bibliothèque nationale de France has assembled digitised collections and

gone on to make public domain documents available on the Internet through the *Gallica* service. Without taking into account the cost of computer infrastructure, the global investment for the digitisation project of the BnF since 1992 amounts to 70 million francs (\$10 million) to which is added an annual budget for digitisation of collections of 3.5 million francs. Apart from the BnF, the Ministry of Culture will be allocating some 12 million francs to digitisation projects in archives, museums and libraries. The Ministry for Education has in turn made a 14 millions francs grant available to the Maison des Sciences de l'Homme for furthering digitisation in the research sector.

The advantages of free access for the user are obvious: it is in principle possible to provide the widest possible distribution of information, in particular within schools and universities. The disadvantages are not totally negligible: the notion of the cost of these services is ignored whilst the development of these very services requires greater and greater public funding. And while the results of these various projects are encouraging, their life span is only relative. The initial investment is often linked to the launch of major projects: this is typically the case of the BnF. Financing the running costs in the longer term is linked to the fluctuations of public finance and the constraints of an annual budget. Furthermore, growth perspectives are limited by the European budgeting policy, which aims at reducing taxation and public expenditure.

It is important that we make the point that these economic models are not necessarily mutually exclusive.

For the research institutions that are the most used, the question becomes one of the coherence of these economic models within a generalised strategy. Already today we are seeing the participation of great research libraries in both FATHOM and JSTOR. The Bibliothèque nationale de France is looking in a similar fashion at diversifying its strategy for the development of digitised collections in opening, alongside the *Gallica* service other charging services, either through subscription or by charging for each usage. This would both avoid letting the economic future of the programmes rely entirely on State funding, including private sector sponsorship, and also enable services paid for by the user to be offered such as high definition image banks which could be of as much interest to the multimedia industry as to researchers.

This brings us to the question of the roles of different libraries and their various types of users.

3. A „GLOBAL” DIGITISED COLLECTION.

In some ways the objectives of the multimedia industry resemble those of libraries. It is true that co-operation between research libraries, like the international strategy of companies, depends to a greater or lesser degree on the notion of a *global virtual collection*. Whatever may be the size of a library, none can be self-sufficient. Strategies for co-operation are based both on documentary and economic aspects, and strongly favour the definition and adoption of identical or at least inter-operable technical standards. The encouragement of a basis for exchange in scientific, economic and technical terms is, alongside common objectives of distribution of knowledge and culture, the principal means whereby public policies of funding of digitisation can be justified. These policies are undertaken by the state, the regions, but also by the European Community, which also plays an important role, as do professional associations. At world level, the major industrialised countries (*Bibliotheca Universalis*¹⁰) and Unesco (*Memory of the World*¹¹) also encourage co-operation.

However, over and above the very generalised concept of the global virtual collection, we must today question the real coherence of these various initiatives. A universal library in which each project would find its place as though in a great jigsaw will not just happen by adding together all the local, national and international initiatives. As with *universal bibliographic referencing*, the *global virtual collection* is an idealist concept whose dynamic function is unquestionable but whose realisation remains at some distance. Thus each country, each cultural sphere, tends to promote its own language and tradition just as it preserves its own heritage. Furthermore from one project to another, co-operation is organised according to different areas of interest. Some programmes will work towards a common technical approach irrespective of contents. As concerns the content of collections variety of policies is the rule: in one case the priority is given to heritage¹², in another to thematic corpora, in yet another to certain types of documents such as scientific journals or documents whose only common denominator is their poor state of conservation. Finally, and despite the numerous initiatives to promote co-operation between libraries which the numbers of conferences and symposia prove take place, the use of public funds is inevitable for the digitisation of identical collections. Thus several research institutions across the world have unwittingly funded the digitisation of the same titles of major scientific journals such as the *Philosophical Transactions of the Royal Society*. This sort of mistake entirely justifies such initiatives as DIEPER¹³ which we will be discussing later, and makes clear the need to look at more detailed sharing of tasks in digitisation of collections as well as the gaps in the

mechanisms of control and co-operation, given that these mechanisms cannot be limited to the national level.

How can we best share out the tasks and responsibilities in this domain? When confronted with this question there is a strong temptation to respond first with technical and administrative details. We could imagine for example a generalised European authority to survey the initiatives and improve co-ordination. This has indeed already been suggested and supported by a number of colleagues. However such an idea encounters a number of problems of methodology and could in any case only deal with the regulation of public funding.

Elsewhere, a number of scientific circles have no intention of being content with only a European and technical policy for digitisation. This is the case for example with mathematicians and physicists who favour a fully international approach. Incidentally, whether it be in the physical sciences or in the social sciences, a detailed analysis of the digitisation programmes undertaken by research institutions shows that scientific co-operation knows *a priori* no political or geographic frontiers. The relevance of co-operation within a field of research rather than guided by other documentary criteria is widely subscribed to in scientific circles. This method points to the employment of organisations for co-operation specialised by subject and reminds us of the fundamental role of the end user in determining our strategies.

There is no single and systematic answer to these questions today. As an example, I would like to show how we at the Bibliothèque nationale de France are trying to find a solution given that our answers are far from exhaustive. We have in fact just completed our strategic plan for the coming three years and its preparation was for us the opportunity to reconsider most of the questions which have just been recited.

Our basic concept is one of a *library without walls*. We have thus chosen to place the emphasis on the Internet as the main means of distributing digitised documents and to participate as closely as possible in international projects while concentrating on scientific content rather than on technical solutions. This is why the BnF is giving priority for example to thematic corpora rather than pursuing the virtual encyclopædic library. The second road that we have chosen is to offer charged services alongside the free services which will be maintained or developed. Of course as a national library we shall be continuing our programmes to enable free access to the national collections and to a vast corpus of French language documents. But at the same time it seemed to us that the main thrust of our projects should be towards more diversified programmes and services, oriented towards research, and not exclusively funded by public resources but at least self-financing. We

are thus looking to commercialise an image bank containing manuscripts, prints and other documents from the BnF by the year 2002. We are also thus looking to participate in The Andrew W. Mellon Foundation's Dunhuang Archives project. Here we will be creating a subscriber-accessed site dealing with the Dunhuang archaeological site and which will enable a virtual reconstitution of the manuscript and painting collection, which today is dispersed between China, Russia, France and the United States.

CONCLUSIONS

To conclude this panorama of strategic questions that are raised when creating digitised collections, I would like to point to some subjects for reflection in the form of suggestions.

1. Define a business plan for each project

Co-operation between libraries in the field of digitisation is considerable and happily shows a tendency to grow: however, too few projects - you will be able to verify this by checking the information given on the web sites concerned - are clearly stating their business plan for the development of digitisation and the associated technical maintenance required. Many projects rely on prospects of growing public funding while we are in a sector where increases in costs can be considerable and I do not share the technocratic euphoria of those who predict each year drastic falls in costs. In some cases, even, the quest for public funds receives greater weight than the interests of the end user whose benefits may seem doubtful once the project has been completed.

2. Build alliances for digitisation policies

Whichever economic model is chosen, and whatever the size of the library, co-operation between libraries and other partners must be the rule. More synergy between research policy and library policy in public sector is suggested.

3. Keep the specific role of libraries

There is today a major risk of developing projects that go beyond the specific functions of libraries. Their job is not to publish texts: research laboratories and publishers are often better fitted for such work; the role of libraries is to contribute to enabling long term access to documents and information.

Reference services, long term access and conservation of digitised documents are irreplaceable missions. For national libraries, legal deposit of electronic material is also a core mission.

4. Produce fully interactive service

Only services with strong added value can succeed in the digitisation sector in the medium term. It would probably be a strategic error to offer simply a collection of digitised documents without any added editorial or retrieval content.

5. Offer client-oriented services

In the same way, we should beware of straying from our objectives, which may happen particularly because of the way the projects are financed, and we must be careful to keep our sights on the end user. What are the benefits that our users can expect to obtain in the short or medium term from the applications we offer? Ideally we should be able to point to a specific population of users for each application and determine our means of evaluation accordingly.

REFERENCES

- 1 Declarations made by digitisation project leaders from the Havas group are in this respect quite clear.
- 2 Gallimard have broken this agreement since the Copenhagen LIBER Conference.
See <<http://listes.cru.fr/arc/biblio-fr@cru.fr/2000-07/msg00069.html>>
- 3 Most of the papers will shortly be made available on the BnF web site:
<<http://www.bnf.fr>>
- 4 <<http://fathom.com/>>
- 5 IDC. 2000
- 6 <<http://www.jstor.org/>>
- 7 <<http://www.deflink.dk/english/def.ihtml?fil=digit>>
- 8 <<http://www.rlg.org/culturalres/>>

- 9 <<http://www.deflink.dk/english/def.ihtml>>
- 10 <<http://portico.bl.uk/gabriel/bibliotheca-universalis/index.htm>>
- 11 Abid, Abdelaziz. *Memory of the World: Preserving our Documentary Heritage*. <<http://www.unesco.org/webworld/memory/Abid.htm>>
- 12 This is often the case with national libraries, such as in Japan.
- 13 <<http://www.sub.uni-goettingen.de/gdz/dieper/>>

Daniel Renoult
Department of information systems
Bibliothèque nationale de France
11, quai François Mauriac
75706 Paris cedex 13, France
daniel.renoult@bnf.fr