



Date : 23/11/2006

**Pushing the boundaries of traditional heritage policy:
Maintaining long-term access to multimedia content by
introducing emulation and contextualization instead of
accepting inevitable loss**

Grégory Miura

Library curator Electronic Documents Section – Multimedia
Service

gregory.miura@bnf.fr

Département de l'Audiovisuel – Bibliothèque nationale de
France

Quai François Mauriac 75706 Paris cedex 13 France

Meeting:	91. Information Technology with Audiovisual and Multimedia and National Libraries
-----------------	--

Simultaneous Interpretation:	No
-------------------------------------	----

WORLD LIBRARY AND INFORMATION CONGRESS: 72ND IFLA GENERAL CONFERENCE AND COUNCIL
20-24 August 2006, Seoul, Korea
<http://www.ifla.org/IV/ifla72/index.htm>

Abstract

This paper will address the direction chosen by the Audiovisual Department of the Bibliothèque nationale de France (BnF) in order to maintain access to a comprehensive collection of multimedia and electronic documents. It will describe several experimental technical solutions for preservation, emphasizing the consequences of a requirement for future users to be able to use these documents for research purposes. These result in the use of emulation solutions as well as new practices of collection development policy with regard to the specific structure of information within a multimedia context

Multimedia, software and other computer documents are posing a new challenge for cultural institutions, which is shared but sometimes underestimated by publishers, researchers and the general public. Even the most up-to-date medium or mode of expression is heavily constrained by technological and economic cycles that reduce its cultural impact and future visibility.

The French tradition of legal deposit legislation integrated multimedia and software documents into the patrimony in a two-phase process. Firstly in 1975, the *Phonothèque nationale* (the future *Audiovisual department* of the *Bibliothèque nationale de France*) was charged with collecting mixed media publications alongside moving images. During the eighties the first examples of software, educational and gaming documents began to appear through legal deposit. Then in 1992, new legislation established that single carrier multimedia documents, software and databases must also be covered by the obligation of legal deposit, since commercial operators of every kind were disseminating samples of such documents.

As the latest of a long tradition of legal deposit, multimedia and software are no exception to the established practices of exhaustiveness and neutrality regarding content. With the coming of web archiving, French heritage collections are now able to display the whole history of media and modes of

expression in a complementary panorama, including the *Institut national de l'Audiovisuel* for radio and television broadcasting collections and the *Centre national de la Cinématographie* for cinema. It is important to underline this point, as off-line multimedia and software are most often considered to be the forgotten link in the chronology of media: after printed books, sound and moving images materials and before on-line and internet resources. By now the total collection represents an amount of 30,000 documents with an annual growth of 4,000 items.

Homogeneity is definitely not a characteristic that we should attribute to such a collection. In trying to keep track of these documents, an institution like the *Bibliothèque nationale de France* is finding that different types of diversity are interlaced in combination with each other. An attempted panoramic view of the whole collection began by isolating a publishing diversity with regard to intellectual content, public use and production pattern. This brings us beyond the traditional distinction between fiction and non-fiction. Multimedia and software documents are introducing new objects and new uses (training, game, tool, communication...). These kinds of use are unusual in a library, but they are also genuine contemporary fields of research interest, and a national library should be the depository of all material dedicated to scientific and professional studies. A listing of the various types covered might include:

- Multimedia with documentary content (from general public use to scientific and artistic documents)
- Video games
- Software
- Children's
- Educational, professional and self-training
- Databases
- Electronic journals
- Corporate, institutional and association publications
- Digitized documents

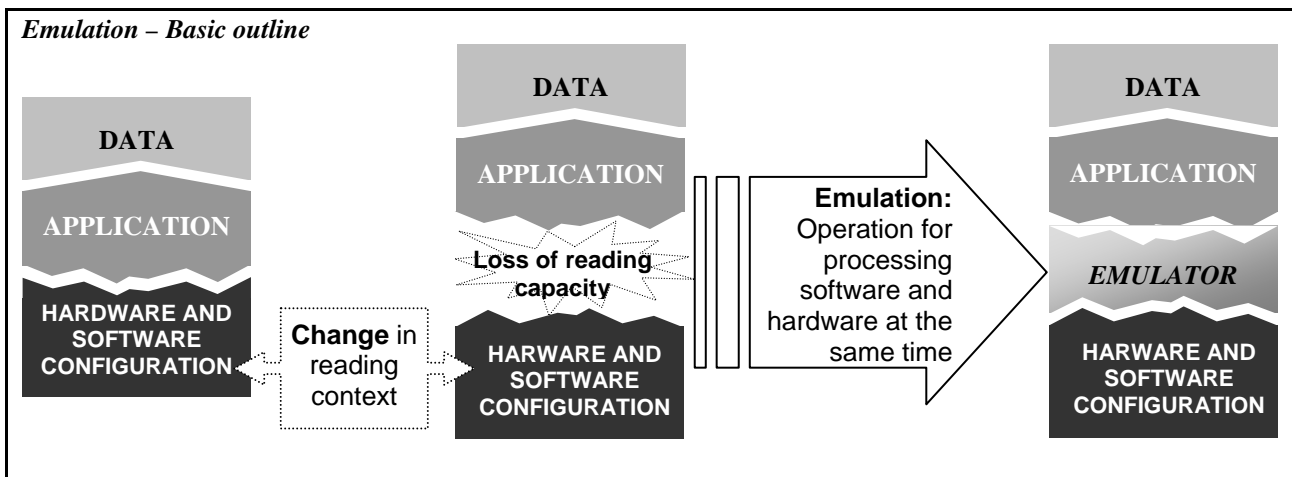
It is important to note, in analysing of the number of documents received each year by legal deposit, that on-line documents have not put an end to the publishing of off-line material. With an experience of more than 10 years now, we may say that there is long term evolution as well as anecdotal change during the period. The other type of diversity is that almost every different platform and carrier and generation of computing and electronic history is represented in the collection:

- Computer audio cassette (TO7, MO5, Amstrad CPC, etc.)
- Floppy discs 5" ¼ and 3" ½
- CD-I, CD-ROM et DVD-ROM, CD-R et DVD-R
- Cartridge (Megadrive, Game Boy, Master System, etc.)
- Proprietary format of previous carrier (Dreamcast, Playstation2, Xbox etc.)
- And now multimedia card, HDVD, UMD, Blu-Ray, USB devices...

Facing this heterogeneity, the preservation of access is a crucial point because it is one of the three pillars of legal deposit responsibility. Today's way of doing this is quite pragmatic but we may ask how long it will continue working, since technological change is a permanent trend of the hardware industry, just as it is in multimedia and software publishing. As time goes by, downward compatibility and expertise in the operation of older platforms are dramatically decreasing. The whole collection is fully accessible to researchers in a specific area of the audiovisual reading room on the Tolbiac site of the BnF. Each document is installed by a technical staff member on demand on its original platform and configuration or a compatible one. Requesting a document may be done in advance or on the day itself. PC and Mac hardware are pre-installed in the reading room; other specific hardware may be installed on demand for any document that can't be played on a standard workstation.

Concern about the inherent frailty of this kind of media is an old source of questions inside and outside the world of libraries, but until the past year we have not managed to solve the problem with an applicable solution. Issued in 2005, our *Electronic documents work plan for preservation* establishes emulation as one of the main solutions to integrate into general preservation policy. It provides a preservation plan for multimedia and software collections, made up from state of the art solutions.

Emulation implies that you accept that at a certain point in technological evolution the original context of consultation must be abandoned. We move slightly or sometimes radically from an issue of preservation to a new way of accessing multimedia and software material. Despite common opinion, emulation, and more broadly how to preserve the content of documents running on computers and similar platforms, is an old debate¹. It seems to be the only effective way for long term access, in combination with the more traditional technique of migration. We now know that just transferring data and programs from an endangered carrier to a current and safer one cannot in itself help with the application contained in most of these documents. Emulation is based on a software development (an emulator) that allows the user to have access to any kind of application on an up-to-date platform while the software thinks it is in its original environment. Emulator programs either reproduce the original instruction set, processor speed, graphic and sound resolution, or translate today's user inputs into old ones:



As emulation is not a mainstream strategy, we have chosen a rather pragmatic, experimental approach to integrate it. Part of the solution of this problem is to work closely with the most advanced people in emulation development: the community of collectors of old computing and console platforms. We try to have contacts with three of the main associations in France in this field: MO5.com in Paris, ACONIT in Grenoble and Silicium in Toulouse. We use their knowledge and federative influence as an entry point to networks of specialists in several platforms and emulations techniques. Based on an emulator test bed study, we have chosen some types of collections to explore new way of accessing these documents

Our work plan covered several complementary areas allowing us to move on from trial and error to a first consulting process. We also took into account circumstances that allowed us to connect theory to concrete implementations. We worked to emulate two different platforms.

- In 2004, we receive as a gift from the *Centre d'études et des systèmes des technologies avancées* a collection of educational and children's multimedia and video games for the Thomson TO7 and MO5 (1982-1986). This collection had been assembled to disseminate computing knowledge as part of the mission of one of the very first multidisciplinary think tanks on information technology.
- One year later we started the integration of documents from the pre legal deposit period by defining simple criteria for an acquisition policy for old software and multimedia documents. A designated provider was charged with exploring the whole catalogue for a specific platform, looking for original documents which were complete and in good condition (including carrier, box and documentation). This first acquisition campaign assembled 300 documents in Amiga floppy disc proprietary format that needed emulation development as this Commodore format was only in use from 1985 to the beginning of the nineties.

¹ KAPLAN, Elizabeth. « A Response to "Preserving Software: Why and How" ». *Iterations, an interdisciplinary Journal of Software History* [On-line]. 1 (September 13, 2002). Available at < <http://www.cbi.umn.edu/iterations/kaplan.html> > (last access : 06.10.13)

Building on the success of these experiences, we next tried, as a matter of advocacy, to legitimate our first conclusions within the institution with a further step. We decided to experiment with using emulation to solve an old problem caused by the dispersal of cartographic documents between two departments of the library: Maps and Charts on the Richelieu site and Audiovisual in the new building, 5km away at Tolbiac. The separation of these closely related documents had been the result of setting up workflows based on media and carrier and not on disciplinary content. The final goal of this project is to make available to users a new remote drive workstation for born-digital cartographic and geographic documents and applications. The basic principle is to give access within the Maps and Charts reading room to more than one thousand cartographic documents, of any type and platform generation from a digitized map to portable geographic information system. Each document will be installed on a virtual machine (a specific emulated platform, eg Windows 95). This virtual machine is defined by two files: one describing hardware and peripheral components and the second, the virtual hard disk. After testing network quality, it is now possible to display the original document on a remote workstation.

There are several advantages in this new approach to access. No more need to maintain old platforms and hardware in their original configuration. No more need for multi boot situations. We also have the opportunity to build a library of past and existing operating systems and related software which will be directly available for the creation of the right virtual machine. Such a collection of environments could be reusable for other collections. There will no longer be a need to manipulate original documents in the reading room. This will be a significant gain for the security of collections. Remote access will be available from now on with the possibility of a multimedia approach for any a discipline in several places in the library

We have to keep clear in mind that emulation is not the ideal solution: indeed there probably is no ideal solution. The main limitation is that emulation implies a loss of original context. Differences between original access and emulated access may only grow as time goes by in the actual state-of-the-art of emulation. For example, processing or access speed may greatly differ. It may also be difficult to reproduce the experience of specific peripheral devices. We consider it to be a further priority to organize contextual materials unifying two visions: the original document and the non-published and self-produced documentation linked to it - even beyond technical answers. This direction is illustrated with several initiatives.

In parallel with collecting documents by legal deposit, gifts or acquisitions, we are committed to building a hardware collection to help researchers understand how a document was played originally. And in addition to that, for staff members, it's a way to become familiar with how old technologies and platforms were used. A complementary approach resides in keeping track of expert skills on old platforms through close links with collectors' associations. As the web archiving project is gaining cruising speed, we have focused, as domain experts, on resources linked to multimedia and software publications in several areas of the web such as publisher official sites, user communities, emulation databases and unofficial expert resources. The Internet appears to be the first place of disseminated knowledge on these documents and a common platform for developers, publishers, researchers and users to share information on. It's another distinct feature of the medium that so much interest has been revealed in developing web material complementary to the legal deposit itself.

Technical and even collections analyses are not the only directions to think about for this emergent patrimony. A user study focused on the whole problem of access has been completed by students in librarianship in order to have a wider range of more suitable services and a better understanding of new critical and analytical methodologies in information science, computing or 3D graphics research. The results point to a set of future enhancements that would require further funding in order to develop new patterns of use. Future developments to come, like a library of backup copies, or video recordings of consultations made by both researchers and users, would benefit the whole community. Multimedia and software implies management of non documentary uses (game, tool, and communication), randomly organized content or using context strongly linked to content.

All of the above points to a need for us to offer more than simply direct access to original documents, even through emulation. We may now look to the future on the basis of this work in progress. It introduces

a lot of innovations both in content and access, because this is critical due to the frailty of these media and the complexity of relationships between data and application. The whole procedure needs to be strengthened by establishing this experimental process as a formal workflow starting as soon as a document enters the library. Beyond its strict use for published multimedia and software documents we are convinced that emulation is a key technique for dealing with all multifaceted entities: digitized documents, web archiving and digital library. It is time also to join forces within the library world: this year the *Bibliothèque nationale de France* has joined the *Koninklijke Bibliothek* in a process to review their well-established project on modular emulation. Off-line born-digital documents should be a cause of humility for the librarian, because it only take a few years for them to be lost and can take ages to manage developing the right tool to regain access. But at the same time, they challenge our way of thinking, traditionally based on the printed culture, in order to let us imagine new, more effective means of access for researchers and for the general public alike.