French government activity in the conservation of data and electronic documents

Serge NOVARETTI

Mission interministérielle de soutien technique pour le développement des technologies de l'information et de la communication dans l'administration
66 rue de Bellechasse - 75007 PARIS
FRANCE
mtic@mtic.pm.gouv.fr

Abstract

French government has launch in 2000 a public debate about conservation of data and electronic documents.
Due to the widespread use of Internet and extranet technologies, especially with electronic document exchange, we had have to adapt the politic of document conservation to this new challenge: "avoid the lack of memory in public administration".

A guide-book was produced. This guide-book is an important step to introduce a reflection about the conservation and to give some guidelines.
Improving XML is the other principal result of the guide-book

Making of a guide-book

French government has published a "request for comment" on its Internet site during three months.

It was issued from a document produced by the European community in 1996. This document described organisation, formats and support for data and document conservation, but it was out of date.

About fifty answers were made by different experts, so we were able to produce a synthesis and a guide-book to the conservation of data and electronic documents for e-services, intranet and internet sites (formats and media).

Experts of conservation and the «Archives Nationales» collaborated to produce the guide-book. A public meeting was organised by the French Prime Minister services to present the guide-book.

A lawyer, specialised in digital document legal issues, was involved in this meeting.. His presentation emphasised the semantic value has to be developed in regards to the present court knowledge.

On another hand, due to the need of specific drivers to visualise a document, the need to archive together the documents and the relevant software was pointed out

Despite their virtual nature, digital documents are threatened by the lack of long-term stability of their media.

The French standard NF Z42-013 and law on the validity of digital documents as formal proof require that documents be written on non-rewritable media, guaranteed only over ten years – a very brief period of time from the archivist's point of view (but further than magnetic tape).

Content of the guide-book

Conservation refers to an organisational, functional and material system which *permits data and documents to be accessed and restored over time.*

The process of conservation differs from computer backups in that the transferred files are selected, and the duration of conservation of the data may range from the useful working life within a department to statutory

Proceedings of the 27th VLDB Conference,
Roma, Italy, 2001

duration such as 10 years for contractual documents, 30 or 40 years for credit dossiers, or unlimited duration for documents of historical value.

It will be noted that conservation can be handled either internally or externally, and may be contracted out.

Some documents to be conserved are probative in nature: they serve as evidence of a transaction. This quality demands the adoption of an appropriate system for guaranteeing data and documents vis-à-vis third parties.

The guide-book is intended chiefly for managers in charge of teleprocedures and Intranet and Internet sites. In the modern environment, these managers have to take account of the technical and organisational aspects of the conservation of data and digital documents.

The guide-book describes the formats, metadata, media and organisation to determine conditions for the conservation of digital documents produced and received, and proposes recommendations for the practical implementation of such conservation.

The expected results after implementing these measures are as follows:

- Permanent access to and use of the data contained in medium to long-term documents, independently of any technological changes (software, hardware, specifications and standards);

- Lower service operating costs due to a reduction in circulation, redundancy of copies and a reduction in the storage of documents on paper media;

- Where applicable, the ability to produce legal proof of a deed by means of a digital document which can be used within the scope of legal proceedings.

The guide-book offers the following elements:

- A model for the processing of digital documents to be conserved;

- The formats, metadata and types of media to use;

- Reference action plans capable of implementation;

- Recommendations for the tools to be applied;

- Orders of magnitude for budgets;

- A framework for the formulation of a set of technical conditions of tender.

Matters relating to the digitisation of paper documents or microfilms are not covered in this guide-book, since digitisation is carried out upstream of the conservation process.

In order to satisfy requirements for the conservation of data and digital documents, information systems must incorporate the following processes and elements

Data conservation comprises three distinct processes: the integration process, the perpetuation process and the access and communication process.

- *The integration process chiefly concerns identification of the digital documents to be conserved, the recommended formats and the metadata to be associated with the documents.*

- *The perpetuation process concerns storage and the recommended media.*

- *The access and communication process concerns the management of access rights.*

The identification and evaluation of documents to be conserved are two indispensable stages whose proper implementations undoubtedly has a significant influence on the success of the project in terms of the conservation component.

Certain formats are to be preferred according to the type of document.

A table in the guide-book gives a summary of the various formats and will help to make a choice on the basis of four criteria concerning the recommendability, durability, openness and frequency of use of a given format.

Depending on the extent of their use in the world at large, formats are classified as being either widely used or little used.

The transmission of documents to be conserved may be effected by means of e-services or office automation applications.

Such applications may allow users to verify the transmission (review and approval function). In any event, it is for the manager to make a choice regarding the incorporation of this function.

In addition to gathering the documents to be transferred, there is also the task of compiling the associated metadata file in one of the formats recommended.

As mentioned in the electronic archiving standard NF Z 42-013, the storage medium must be an "optical medium in which the writing of the bits encoding the data is effected by irreversible transformation of one or more constituents of that medium".

It is obligatory for the medium to be of the optical type, i.e. it must use laser technology (not magnetic or magneto-optical) and must be non-rewritable (or of the WORM type - Write Once Read Many).

The deformation of the active surface of the medium must be definitive and irreversible.

The medium to use at present is the *Recordable Compact Disc or CD-R*, which offers different advantages: Compatibility of the medium with the vast majority of CD-ROM readers; Reliability of the medium (approximately ten);Low cost; Standardisation;

In the medium term, the recordable digital versatile disc or *DVD-R* will make it possible to achieve capacities of 5 - 20 GB. Indeed, it is possible to use this technology today if a sufficient budget is allowed.

Note that for the purpose of accessibility by internal users in their routine activities, document bases may be replicated on departmental machines.

Destruction procedures will need to be defined, but these will not be implemented in the short term. However, incomplete or defective media must be physically destroyed, taking account of the security level applicable to the content of the documents contained on the media (crushing of media).

The guide-book indicates storage costs (January 2001).

Two types of access must be provided for:

- Access to the digital document by the generating departments, which is effected in the form of a "replay".

- This access limit for government departments means that the documentary database must incorporate access rules. These rules must be adapted to any organisational changes.

- Access to the digital document by departments other than the generating department, and by users and researchers.

Documents deriving from (issued or received by) centralised or decentralised government departments constitute public archives and are covered by the provisions of Article 3 of the Law of 3 January 1979.

Consequently, the Direction des Archives de France has the authority to oversee application of the law, and in particular to authorise access to public documents by private individuals and researchers.

Such authorisation may be granted by means of derogation where the documents concerned are not immediately communicable.

Access to documents is becoming an increasingly sensitive issue as "search and display" operations become technically easier thanks to Internet standards.

As an example, data concerning the salaries or performance evaluations of personnel is naturally confidential, and access is restricted to authorised parties.

The responsible departments may intervene to authorise a user to access a document. At this stage, the authorisation will be received either bye e-mail or by fax. If directly concerned, he or she may receive a paper copy.

In order to satisfy document access needs, use may be made of the functions offered by an Electronic Document Management (EDM) application.

The management of conservation activities demands a high degree of technicality and a global approach on the part of the organisation concerned.

It must therefore be handled very rigorously and as an ongoing concern.

Conservation activities must be auditable and periodically audited so that, where applicable, it can be demonstrated to any requesting authority that documents are being conserved in conformity with the proper conditions.

In order to prove the reliability of operation of the conservation system, audit and self-audit measures must be applied regularly and systematically.

The project leader who will supervise the conduct of this activity is identified by the management, which commits itself to supporting his actions.

The guide-book gives a plan of a set of technical conditions of tender.

Conclusion

This guide-book is an important step to introduce a reflection about the conservation and to give some guidelines.

A public recommendation has been produced in order to give guidelines for the different Administration departments.

Improving XML is the other principal result of the guide-book. XML can be used for different purposes :

- XML is a format that should be easy to migrate.
- XML allows the separation of content from the presentation and separate storage
- The guide-book recommends defining an XML envelope for the documents
- XML is a good candidate for describing the metadata associated with the document- possibly as a part of its envelope.

The guide-book will be updated often in the next few years.