



# **Journée d'étude sur la documentation informatisée au sein du ministère de la culture et de la communication**

**19 novembre 1999**

## Sommaire

<b>Introduction</b> .....	3
<b>La documentation structurée</b> .....	4
<b>La situation du plan de numérisation et des programmes associés</b>	
La numérisation à la Direction des musées de France .....	8
La situation du plan de numérisation et des programmes associés dans les bibliothèques .....	11
La numérisation aux Archives de France .....	14
Documentation structurée au sein de la Direction de l'architecture et du patrimoine (DAPA).....	17
La documentation informatisée au Ministère de la culture et de la communication.....	20
<b>Tables rondes</b>	
Évolution du plan de numérisation des fonds d'État - bilan pratique 1996-1999 - nouvelles orientations pour 2000-2001 .....	25
L'organisation de la production des connaissances scientifiques et de la diffusion documentaire au ministère dans le contexte du plan d'action du Gouvernement pour la société de l'information .....	29
<b>Conclusions</b> .....	34

## Introduction

Lancé en 1996, le plan de numérisation des fonds d'État géré par la Direction de l'administration générale (mission de la recherche et de la technologie (MRT), département de l'organisation et des systèmes d'information (DOSI) dans le cadre du comité scientifique pour la documentation informatisée et le multimédia du conseil ministériel de la recherche a fait la preuve de son utilité en accélérant la modernisation des systèmes d'information du ministère, en favorisant la formation des personnels aux nouvelles techniques et en incitant à mettre en œuvre de nouveaux modes de production et de diffusion des connaissances sur le patrimoine culturel.

Ce plan s'inscrit dans le programme d'action gouvernemental pour la société de l'information.

Le travail collectif effectué autour des projets a permis la constitution d'une communauté de spécialistes de la numérisation à l'intérieur du ministère ; les compétences acquises dans le cadre du plan sont devenues des atouts pour les services au moment où les collaborations avec les collectivités territoriales sur ces sujets se multiplient.

Après la mise en place de la convention de prix pour la numérisation des phototypes et des microformes, une nouvelle étape du plan vient d'être franchie avec la signature de marchés d'aide à l'indexation des fonds et à la numérisation des collections *in situ*.

La procédure d'appel à projets de numérisation permet la sélection des meilleurs programmes de numérisation soumis au comité par les directions centrales et régionales, tout en préservant les priorités scientifiques, techniques et éditoriales.

Les projets d'étude et de recherche financés par le ministère (MRT/DOSI, autres directions) mais également par la Commission Européenne ont facilité la définition de l'environnement informatique nécessaire pour assurer l'exploitation et la diffusion des fonds numérisés.

Avant la réunion des grands conseils scientifiques et techniques de fin d'année où seront discutés les orientations pour l'an 2000, il convient de s'assurer de l'adéquation des priorités techniques avec les besoins et les moyens des utilisateurs.

Cette journée d'étude interne aux services du ministère a pour but de dresser un bilan du plan de numérisation depuis son lancement, de déterminer ses réussites et ses faiblesses, de réfléchir à ses articulations avec les autres programmes, de réfléchir aux outils informatiques à mettre en œuvre et d'alimenter la réflexion sur les études et recherches nécessaires à la modernisation du service public culturel.

# La documentation structurée

par **Antoine Rizk**

Responsable du DESS ingénierie documentaire et multimédia  
de l'Université d'Evry Val d'Essonne  
Société Euroclid, partenaire du programme Médiaculture/dossier électronique

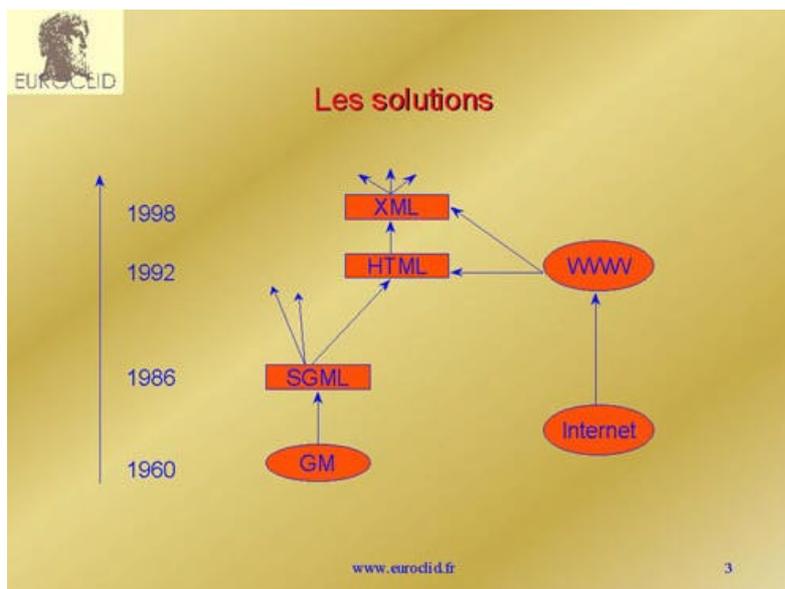
En introduction, quelques données statistiques provenant des Etats-Unis, soulignent l'importance prise par la production, le traitement et la diffusion d'informations dans la société d'aujourd'hui.

- La production de l'information par les entreprises et les administrations représente 20 % du PNB,
- 90 % de la documentation n'est pas structurée,
- Le temps consacré à la production ou à la recherche de l'information est important,
- Le contexte technologique évolue rapidement.

## 1. XML (eXtensible Markup Language)

Des solutions techniques en vue d'organiser l'information ont été recherchées dès les années 60 avec le langage GML (Generalized Markup Language), balisage généralisé, qui a donné naissance à SGML (Standard Generalized Markup Language), norme existant depuis 1986, dont le domaine d'application est resté limité à la documentation technique et à l'informatique éditoriale.

Avec l'arrivée du Web et l'émergence du langage HTML (HyperText Markup Language), qui a rapidement été rendu compatible avec SGML, un nouvel essor a été donné à SGML. Le besoin d'un méta-langage adapté au web est à l'origine de la création de XML, sous-ensemble de la norme SGML. XML est ainsi une double norme, en tant que recommandation du Consortium W3C et en tant que sous-ensemble d'une norme ISO.



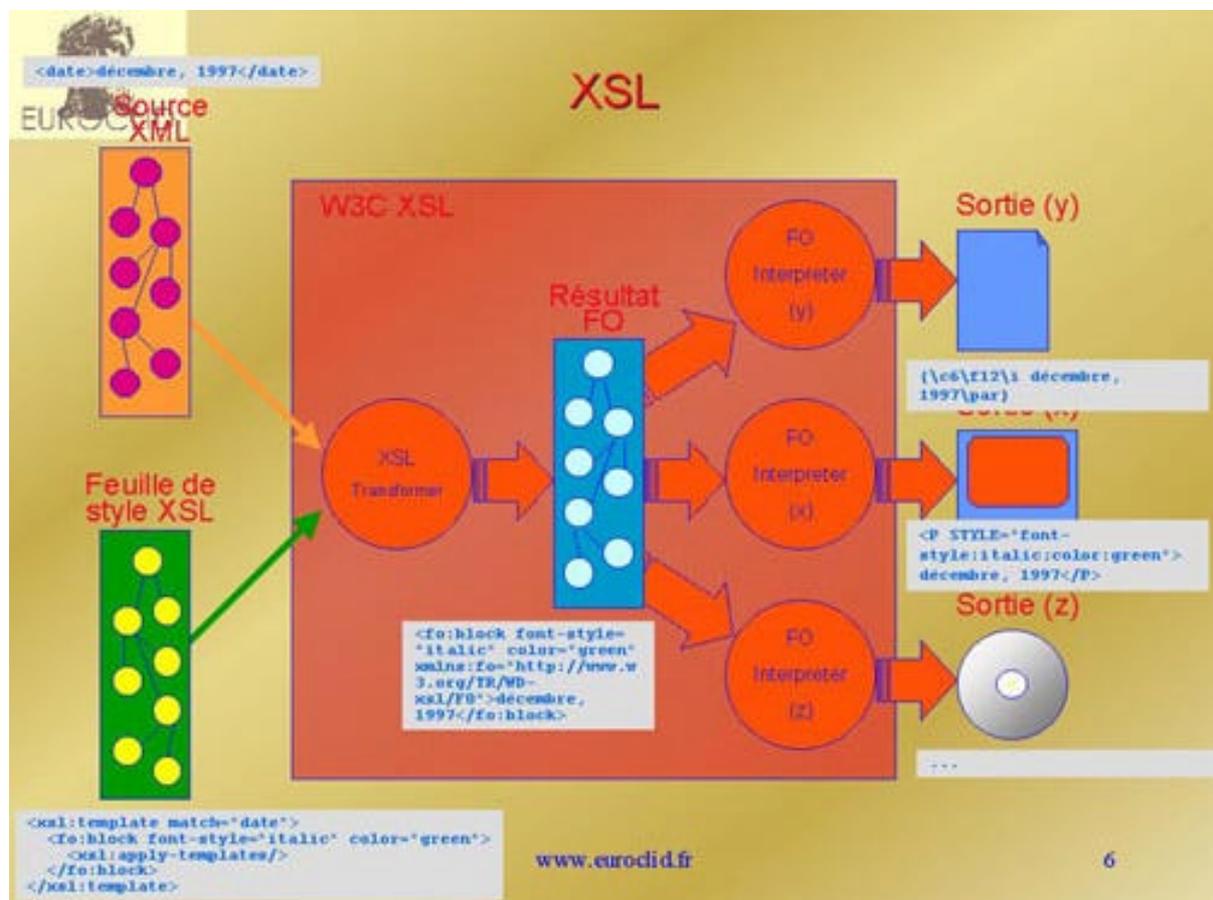
### 1.1 Les avantages d'XML sont :

- L'amélioration de la productivité.
- L'indépendance de l'aspect physique du document vis à vis de sa structure logique (le document devient manipulable selon le profil de l'utilisateur et selon la plate forme d'exploitation).
- La pérennité de l'information, XML rend possible un retraitement des données sur les plates formes du futur.

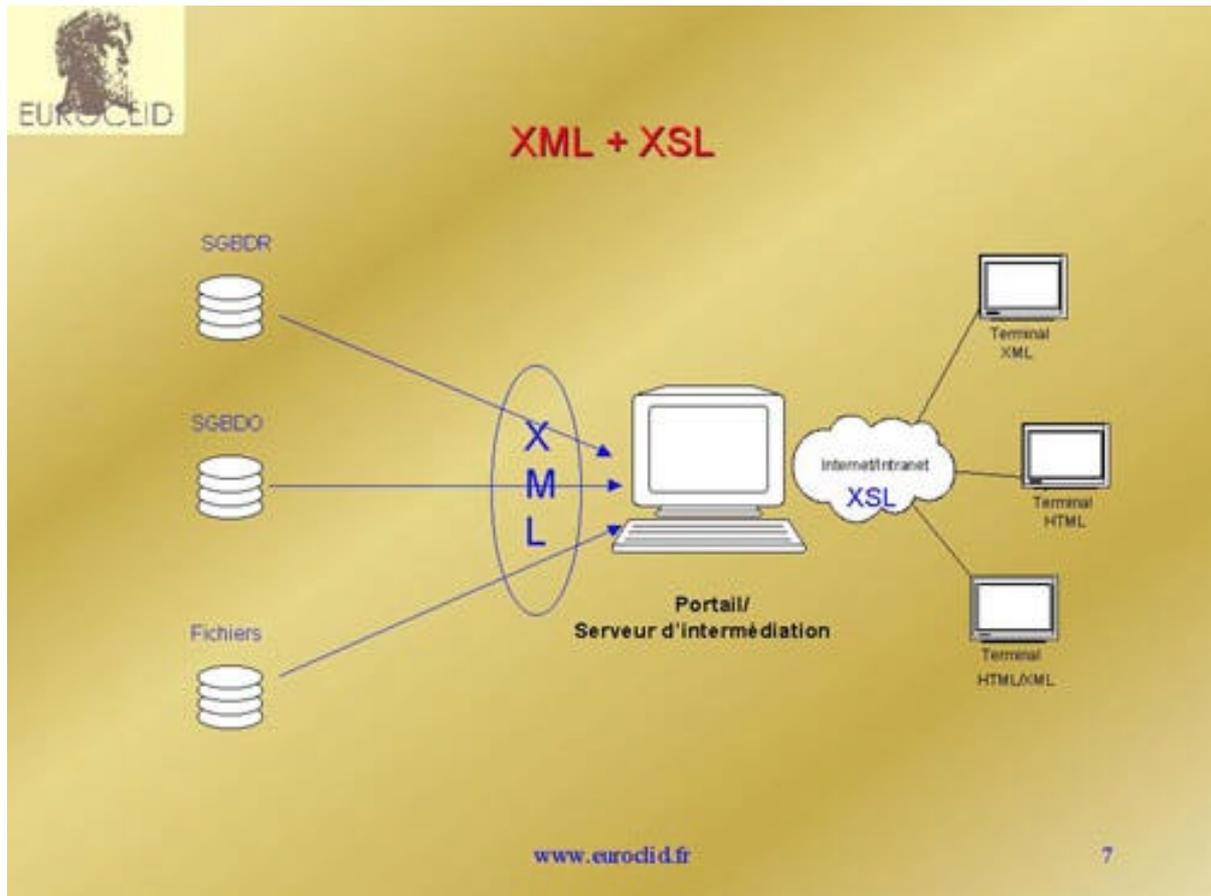
47 % des entreprises interrogées pensent utiliser XML dans les années à venir. XML sera la norme du commerce électronique. 47 % des entreprises comptent passer à Linux, qui cohabitera avec d'autres systèmes d'exploitation.

Les langages d'avenir sont : Java, langage de programmation multi-plateforme et XML, langage de structuration des données.

Après l'arrivée en février 1998 d'XML, la norme **XSL** (eXtensible Style Language) est maintenant adoptée.



A un fragment d'information XML est associée une feuille de style XLS. Les deux sources, XML ou XSL sont indépendantes ; elles passent dans un transformateur XSL (formatting object) et ensuite par un interpréteur pour un affichage dans des formats divers (HTML, Word, Audio etc.).



Exemple d'association XML XSL mise en œuvre dans une architecture typique : le portail cherche l'information dans des bases de données hétérogènes, extrait un fragment de l'ensemble sur le portail et effectue une présentation selon la demande de l'utilisateur. L'arrivée des données sur le portail se fait en XML. Les requêtes du portail vers ces systèmes hétérogènes se feront dans le langage de requête natif de ces systèmes (SQL...) ou par des requêtes Z 39-50 (la question est encore en suspens) ou en interrogeant l'information HTML à la manière des moteurs de recherche tels "Altavista", mais avec une structure plus riche.

Un système de fichiers XML présente beaucoup d'avantages, comme en témoigne l'application sur le patrimoine historique de la ville de La Rochelle. Si l'interrogation s'effectue avec un formulaire structuré, comme dans les bases de données traditionnelles, la différence réside dans le fait que l'information n'est pas dans un système de bases de données mais dans des fichiers au format XML dont le nombre et le volume sont illimités.

## 1.2 Les limites d'XML

XML traite la structure syntaxique des documents mais ne permet de spécifier ni la sémantique d'une balise, ni le typage de son contenu, ni des relations normalisées entre les balises.

## 2. Autres normes utilisées

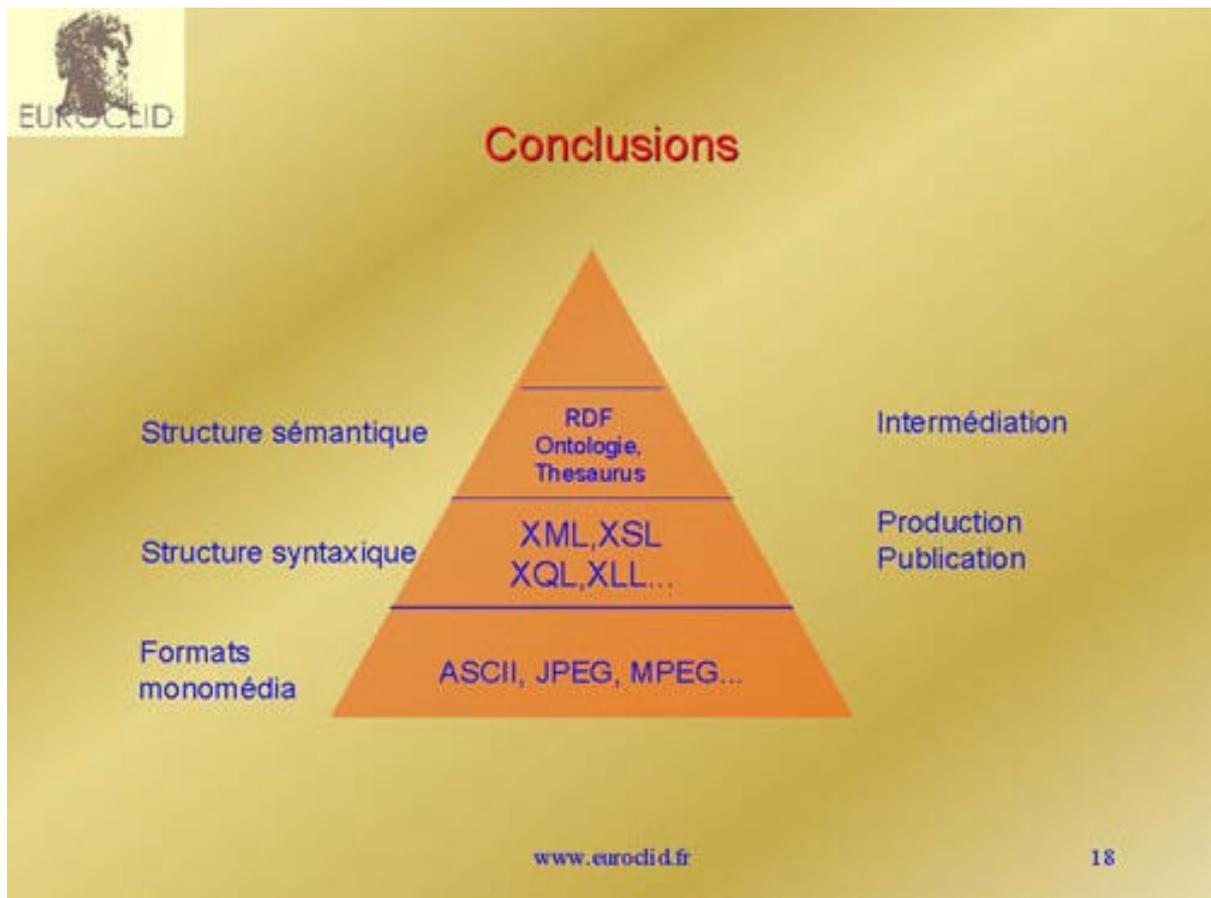
**XLL** pour la création de liens hypertextes (liens simples et étendus, multiples vers des destinations d'information différentes). Il est possible de représenter des liens hypertextes sans toucher à la source du lien ni à sa destination.

**RDF** (recommandation depuis 1999) spécifie les métadonnées,

**XML Data**, pour le typage des contenus des éléments,

**Topics Maps**, norme ISO basée sur SGML et HyTime, se fondant sur les mécanismes d'XLL. Concernant les liens, il est possible de typer les relations et de mettre les *Topics* en relation entre eux pour construire des strates d'information et bâtir des domaines de connaissances.

L'objectif est de pouvoir organiser l'information sans tomber dans les travers actuels d'HTML. Une situation comparable risque de se produire lorsque des centaines de milliers de DTD seront utilisées. Il est impératif d'organiser des strates d'information dès aujourd'hui, au moyen de la définition d'une ontologie et de l'organisation de systèmes de navigation de type *Topics Maps*. L'ontologie pourra être associée à un thésaurus, permettant la création de schémas automatiques RDF. Un langage commun pour définir les données verra le jour progressivement.



Aujourd'hui on sait stocker l'information mono média, l'indexer, l'organiser, bâtir une structure syntaxique (XML, XSL). Demain le langage XQL (équivalent de SQL pour les données en XML natif) permettra d'interroger des bases XML de manière riche portant sur la structure de l'information. Il sera ainsi possible de mettre en place rapidement des outils de production et de publication.

L'enjeu de la recherche (d'ici à dans trois ans) est l'organisation des données XML dans des systèmes d'intermédiation (thésaurus, ontologie, RDF). Le document n'est plus à penser comme un fragment d'information isolé, mais en tant qu'entité en relation avec d'autres documents, de manière à concevoir un réseau de connaissances autour d'XML, Internet et RDF.

# La numérisation à la Direction des musées de France

par **Philippe Avenier**,  
Direction des musées de France  
Bureau de l'informatique et de la recherche

Le développement du *World Wide Web* dans le domaine des musées a vu progresser de façon concomitante trois types de produits qui contiennent des images numériques :

- les bases de données de gestion et de documentation des collections : elles sont constituées d'abord localement, dans les établissements, pour assurer notamment l'informatisation des inventaires. Leur contenu documentaire (texte et image) est ensuite utilisé pour enrichir les bases nationales.
- les sites Web, destinés principalement à fournir des informations sur les musées : présentation générale des collections, renseignements pratiques, activités, manifestations...
- les expositions virtuelles ainsi que d'autres produits (présentations pédagogiques, jeux, etc.), en ligne ou sur cédérom.

Les usages que l'on peut faire de ces produits correspondent à trois grandes fonctions :

- **la recherche documentaire**
- **l'information et la communication**
- **la présentation des collections**

Ces produits, s'ils ont chacun leur logique propre, sont aussi complémentaires et contribuent à une meilleure diffusion de la connaissance sur les collections des musées.

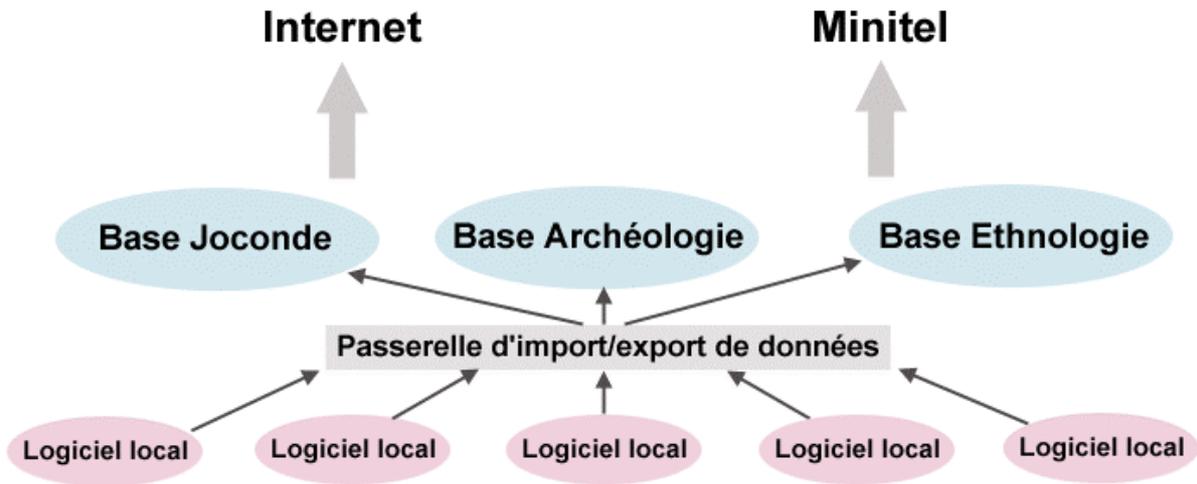
Les bases de données jouent, de ce point de vue, un rôle fondamental dans la mesure où elles rassemblent l'essentiel des informations sur les fonds : elles constituent le socle sur lequel s'appuient les autres réalisations.

## 1. L'enrichissement des bases de données nationales.

Administrées par la direction des musées de France, les bases de données nationales (beaux-arts, archéologie et ethnologie) sont enrichies en permanence par les contributions des équipes de conservation et de documentation de tous les musées.

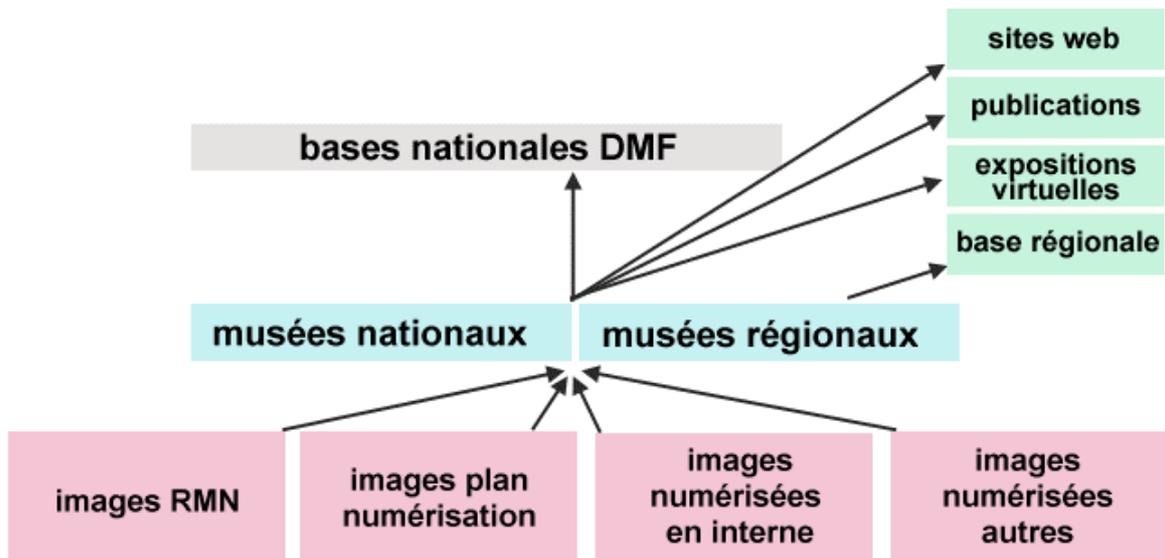
Cet enrichissement (schéma 1) s'effectue quel que soit le logiciel utilisé par les établissements. Des passerelles d'export, mises au point en concertation avec les sociétés qui ont développé des logiciels de gestion des collections, permettent d'automatiser en grande partie le reversement des informations lorsqu'elles ont été élaborées dans le respect des normes documentaires et techniques édictées par le Ministère de la culture et de la communication.

## Schéma d'alimentation des bases documentaires nationales



## 2. La production des images numériques

Les images numériques intégrées dans les bases nationales peuvent provenir de diverses origines.



Elles peuvent avoir été produites par l'Agence photographique de la Réunion des musées nationaux, par la société choisie pour le plan de numérisation du ministère (fonds d'État), par les musées eux-mêmes, lorsqu'ils ont fait le choix de numériser en interne une partie de leurs fonds, ou par d'autres acteurs extérieurs aux musées.

### **3. Questions posées**

La constitution de la documentation numérique soulève deux principaux types de questions :

#### ***des questions d'organisation interne :***

- qu'il s'agisse de l'informatisation de la documentation sur les oeuvres ou de la numérisation des clichés, il importe que la chaîne de traitement de l'information soit mise au point avec rigueur et cohérence. Tous les aspects pratiques (gestion de la photothèque, accessibilité aux oeuvres...) doivent être considérés.

#### ***des questions de politique générale :***

- la variété des situations de chacun des établissements et la diversité des logiques qui sont à l'œuvre dans la constitution des produits utilisant des images numériques impliquent que la complémentarité entre les initiatives soit pensée et organisée,
- le choix du type de numérisation doit se faire en fonction des objectifs que l'on se fixe,
- la prise en compte des problèmes de droits liés à la diffusion des images doit être effective dès le départ de chaque projet.

# La situation du plan de numérisation et des programmes associés dans les bibliothèques

par **Gaëlle Bequet**

Conservateur

Direction du livre et de la lecture

Département des politiques documentaires et patrimoniales

## 1. Les chiffres disponibles aujourd'hui

La direction du livre et de la lecture a publié en 1999 une enquête sur l'équipement informatique des bibliothèques réalisée en 1998 dont l'échantillon était de 2 188 bibliothèques de communes de plus de 2 000 habitants. Elle est disponible sur le site du ministère de la culture et de la communication à l'adresse suivante : <http://www.culture.fr/culture/dll/lmeca.pdf>

Cette enquête a montré que vingt-cinq bibliothèques menaient des opérations de numérisation. Parmi celles-ci, dix-huit numérisaient elles-mêmes leurs documents :

- Ballancourt (Essonnes),
- Baud (Morbihan),
- Grasse (Alpes Maritimes),
- La Ferté sous Jouarre (Seine et Marne),
- Lille (Nord),
- Limoges (Hte Vienne),
- Lisieux (Calvados),
- Lorient (Morbihan),
- Lyon (Rhône),
- Melun (Seine et Marne),
- Noisy le Sec (Seine Saint Denis),
- Orange (Vaucluse),
- Roanne (Loire),
- Saint-Omer (Pas de Calais),
- Senlis (Oise),
- Taverny (Val d'Oise),
- Troyes (Aube),
- Valenciennes (Nord).

Sur ces dix-huit bibliothèques, neuf bibliothèques numérisaient des fonds patrimoniaux, quatre des fonds courants (couvertures d'imprimés, extraits vidéos et sonores), cinq les deux types de fonds.

Sept bibliothèques ont recours à des prestataires extérieurs pour la numérisation : Avignon, Belfort, Dijon, Nancy, Nîmes, Saint-Quentin (Aisne), Toulouse.

Enfin, les bibliothèques de Limoges, Lisieux, Lorient, Lyon, Senlis, Troyes ont aussi recours à des prestataires extérieurs lorsqu'elles ne numérisent pas elles-mêmes.

Plusieurs de ces bibliothèques ont un site internet où sont consultables les documents numérisés (voir le portail culturel à l'adresse suivante : <http://www.culture.fr/culture/autserv/biblio.htm#pub>).

La proportion de bibliothèques ayant numérisé des documents est encore modeste puisqu'elle est de 1%. Néanmoins, il faut comprendre que les 2 188 bibliothèques de l'échantillon ne seront pas toutes concernées par ce type d'opération.

## 2. Les enquêtes futures

L'enquête sur l'équipement informatique des bibliothèques a lieu tous les trois ans. Ce n'est pas un rythme adapté au développement rapide des pratiques de numérisation. La Direction du livre et de la lecture collecte aussi chaque année les rapports annuels des bibliothèques. Le questionnaire envoyé est déjà très fourni, il sera néanmoins complété par une question sur la numérisation. Celle-ci permettra par la suite d'envoyer de façon ciblée un questionnaire plus détaillé.

Celui-ci, déjà rédigé, portera notamment sur les points suivants :

- le contenu des collections numérisées, leur volumétrie et leur traitement bibliographique
- les modes de numérisation (directe ou indirecte, type de fichiers, supports de conservation)
- les moyens dont dispose la bibliothèque (personnel, matériel, budget)
- les partenariats éventuels avec des projets nationaux, européens ou avec un éditeur.

Elle a été envoyée le lundi 29 novembre 1999 à une centaine de bibliothèques pour commencer à rassembler des données pour l'année 2000 afin de préparer les prochains appels à projets de la direction de l'administration générale. Les résultats seront connus en février.

La direction du livre et de la lecture souhaite que l'exploitation de cette enquête puisse se faire sous la forme d'une base de données comparable à celle mise en oeuvre par la bibliothèque nationale du Canada (<http://www.nlc-bnc.ca/initiatives/erella.htm>) où sont en effet recensés les projets de numérisation de plusieurs institutions canadiennes.

## 3. Le plan de numérisation

Trois bibliothèques ont répondu jusqu'à présent aux appels à projets de la direction de l'administration générale :

- la bibliothèque municipale de Troyes pour des manuscrits enluminés ;
- la bibliothèque municipale d'Amiens pour des documents similaires ;
- la bibliothèque municipale de Saint-Etienne pour des cours et dessins de l'École nationale des Mines.

Pour le dernier appel à projets qui s'est achevé le 17 novembre, cinq bibliothèques ont postulé :

- la bibliothèque municipale de Pau pour des gravures, dessins et aquarelles illustrant l'histoire du Béarn ;
- la bibliothèque municipale de Valenciennes pour des incunables et post-incunables ;
- la bibliothèque municipale de Metz pour des cartes réunies vers 1750 par le bibliothécaire de Saint Arnoul ;
- les bibliothèques municipales de Châlons en Champagne et de Dôle pour des manuscrits enluminés.

## 4. Le programme des pôles associés de la Bibliothèque nationale de France

Il faut rappeler ici que la Bibliothèque nationale de France, par le biais de son programme des pôles associés, coopère avec des institutions en région. Cette coopération porte principalement sur un partage des acquisitions documentaires et sur la gestion du dépôt légal imprimeur. Elle se diversifie aujourd'hui pour toucher au domaine de la numérisation avec un projet consacré aux publications des sociétés savantes du XIX<sup>e</sup> siècle (des origines à 1914). L'opération concerne actuellement les archives départementales des régions Aquitaine et Lorraine.

Elle porte sur la numérisation à partir de microfilms :

- de 560 000 pages en mode image ;
- de 7 000 pages de tables des matières en mode texte.

La bibliothèque nationale de France a acquis une expérience inégalée dans le domaine de la mise à disposition des documents numériques grâce à son serveur Gallica (<http://gallica.bnf.fr>). Elle joue un rôle pilote sur le plan national, notamment dans la formation des bibliothécaires dans le domaine de la gestion électronique de documents :

- par l'intervention de ses équipes au cours de stages professionnels ;
- par l'organisation de journées d'information comme celle qui a eu lieu en octobre 1999 consacrée aux questions juridiques posées par la numérisation des documents ;
- par la participation de ses agents à des instances de réflexion et de normalisation.

## 5. Les aides spécifiques destinées aux bibliothèques

La Direction du livre et de la lecture, en accord avec la Direction générale des collectivités territoriales du Ministère de l'intérieur, a modifié en mars 1999 la circulaire d'application du concours particulier des bibliothèques de la Dotation générale de décentralisation.

Deux possibilités sont offertes aux établissements :

- l'aménagement d'un atelier de numérisation dans la bibliothèque dont le financement par l'Etat peut atteindre 50 %. Cette installation est subordonnée à la présence d'un personnel formé affecté à temps plein ou temps partiel à cet équipement,
- la sous-traitance des opérations de numérisation pour lesquelles le financement est compris dans une fourchette de 20 à 40 % du montant hors taxes des devis. Sont prises ici en compte les opérations de conversion des catalogues ou inventaires signalant les documents numérisés, la production éventuelle de supports de substitution.

Il faut rappeler aussi que les conseillers pour le livre et la lecture des directions régionales des affaires culturelles peuvent aider sur d'autres crédits d'intervention des opérations de numérisation. C'est le cas par exemple en Franche-Comté pour les bibliothèques de Besançon et de Dôle.

Dans le cadre de la mise en œuvre de ce programme, il est prévu des formations pour les conseillers pour le livre et la lecture. La base de données qui sera constituée à partir des résultats de l'enquête sur la numérisation et qui sera disponible en ligne leur permettra à tout moment de s'informer sur les projets en cours.

# La numérisation aux Archives de France

par Joël Poivre

Conservateur à la Direction des Archives de France,  
Service Technique

Les Archives de France ont participé au plan de numérisation dès son lancement, d'abord modestement, puis avec de plus en plus de projets. Plusieurs services d'archives avaient d'ailleurs engagé réflexions et travaux dans ce domaine, dès le début des années 90. L'exposé s'attachera principalement à tracer un bilan de ces réalisations et d'en tirer quelques perspectives pour l'avenir.

## 1. Bilan chiffré

Depuis le lancement du plan, en 1996, vingt-trois projets ont bénéficié d'une aide, pour un montant total d'environ 2 MF. Ces projets provenaient d'une trentaine de services, principalement des Archives départementales, mais aussi des Archives nationales et d'un service d'archives municipales.

## 2. Typologie des fonds numérisés

### 2.1 Les documents d'état civil et assimilés

Il faut entendre par là, aussi bien les documents d'état civil moderne (postérieurs à la Révolution) que les registres paroissiaux de l'Ancien Régime, source très consultée par les généalogistes qui constituent une partie importante du public des services d'archives.

Parmi les exemples les plus frappants, on peut citer les projets menés depuis deux ans par les services de la région Champagne-Ardenne, sur la numérisation des tables décennales et, en 1999, à l'occasion du premier appel à projets, l'opération lancée en Saône-et-Loire sur des registres paroissiaux, dont les plus anciens remontent au XIV<sup>e</sup> siècle.

Globalement, on peut estimer que ce type de projets représente entre 30 et 40% des projets de numérisation de fonds d'archives.

### 2.2 Les documents à contenu iconographique riche : cadastre, antiquités et objets d'art, etc.

Il s'agit des documents les plus spectaculaires : plans anciens, cadastre napoléonien, photographies des antiquités et objets d'art (lorsque ce service est rattaché aux Archives), photographies diverses etc.

Par souci de clarté, il convient de distinguer les opérations relevant des Archives nationales, de celles relevant des Archives départementales ou municipales.

Au niveau national, on peut citer deux cas particulièrement intéressants : la numérisation de la photothèque de la France d'Outre-mer, conservée au Centre des archives d'outre-mer à Aix-en-Provence, qui a bénéficié du plan de numérisation, et la base ARCHIM (<http://www.culture.fr/documentation/caran/sommaire3.htm>) créée au Centre historique des Archives nationales, base qui a peu bénéficié du plan de numérisation. Constituée au départ, par les reproductions demandés par les lecteurs, elle ne cesse de s'accroître de collections iconographiques thématiques.

Dans les départements, la numérisation a surtout porté sur des plans anciens ou modernes, tels les plans d'intendance (série C des Archives départementales), pour l'Ancien Régime (par exemple dans

l'Essonne ou l'Oise) ou le cadastre napoléonien, pour lequel plusieurs projets viennent d'être présentés à l'occasion du dernier appel à projets de 1999.

Les archivistes se sont aussi intéressés aux Antiquités et objets d'art. Ainsi le département des Yvelines a bénéficié d'un financement du Conseil général pour numériser ses plans d'Ancien Régime. Un site web a été réalisé, il est hébergé par le Ministère de la Culture à l'adresse suivante : <http://www.culture.fr/ady/>.

### **2.3 Les autres types de documents**

Dans cette catégorie, on trouve des projets de numérisation d'archives de type varié, allant des minutes notariales anciennes (projet des Archives du Var) aux documents relatifs aux deux guerres mondiales en Picardie, en passant par les délibérations d'un conseil municipal (ce dernier cas étant hors plan de numérisation, ne s'agissant pas de fonds appartenant à l'État).

## **3. Bilan qualitatif**

### **3.1 Avancement des travaux et problèmes techniques**

On constate, dans la pratique, une certaine lenteur des procédures d'exécution du plan de numérisation. C'est ainsi que s'achèvent seulement ces jours-ci des opérations sélectionnées en 1998.

De plus, certains problèmes techniques se sont posés et restent non résolus comme la numérisation des phototypes négatifs en couleurs (ex. : Atlas Trudaine, mappes sardes)

### **3.2 Les opérations de partenariat avec les autres institutions patrimoniales**

Il apparaît important de souligner que certaines opérations de numérisation ont été menées en partenariat avec d'autres institutions culturelles, comme par exemple pour les fonds de l'abbaye de Clairvaux (associant Bibliothèque municipale et Archives départementales) et pour l'opération nommée "heurs et malheurs des populations picardes" qui associe plusieurs services patrimoniaux (inventaire, archives et conservation régionale des monuments historiques).

### **3.3 L'effet d'entraînement**

En dehors du plan national, ces dernières années, de nombreux services d'archives, notamment dans les collectivités territoriales, ont engagé des opérations de numérisation. C'est ainsi qu'en 1998, vingt-sept services d'archives départementales et vingt et un services d'archives communales ont numérisé ou fait numériser des documents. On peut voir là un effet d'entraînement du plan de numérisation. La diffusion des documents numériques a été assurée soit par le biais de cédéroms, soit par la mise en ligne sur un site internet.

## **4. Perspectives**

### **4.1 Améliorer les procédures**

Il apparaît nécessaire d'alléger les procédures et de résoudre certains problèmes techniques. De ce point de vue, l'élargissement du plan, en l'an 2000, par le biais de subventions à des programmes de numérisation (crédits d'intervention sur le titre VI du budget) sera un atout majeur.

## **4.2 L'élargissement à de nouveaux types de documents**

### **4.2.1 Le cadastre napoléonien**

La Direction des Archives de France entend favoriser la numérisation du cadastre napoléonien, série homogène et de grande valeur, présente dans tous les départements. Plusieurs projets de ce type ont été présentés lors du dernier appel.

### **4.2.2 Les documents sonores**

Les documents sonores sont également un gisement important dans les Archives, leur numérisation en favorisera la diffusion. Un projet de cette nature a été présenté par le Centre des Archives Contemporaines de Fontainebleau.

### **4.2.3 Des séries administratives**

Enfin, la numérisation doit permettre de mettre à la disposition du public certaines séries homogènes et très consultées de documents. C'est l'objectif du projet de numérisation des originaux des décrets de naturalisation que vient de proposer le Centre historique des Archives nationales.

## **4.3 La conservation**

Enfin, il faudra s'attacher à réfléchir aux moyens de conserver à long terme les documents numérisés qui deviennent à leur tour des documents d'archives. La question de la pérennité des supports et des matériels de lecture est complexe, elle mérite des programmes de recherche et d'étude.

# Documentation structurée au sein de la Direction de l'architecture et du patrimoine (DAPA)

par **Christophe Dessaux et Luc-Florent Lièvre**  
Sous-Direction des ressources humaines et des affaires générales

Le Ministère de la Culture et de la Communication souhaite développer les dossiers documentaires sous forme électronique structurés dans le standard SGML/XML.

Les éléments constitutifs de cette documentation proviennent de sources variées (saisie directe lors de la constitution du dossier ou récupération de données externes : images, textes, éléments de bases de données existantes), il sont plus ou moins structurés (on y trouve aussi bien des textes librement rédigés, que des extractions de bases de données, structurées en champs typés pouvant être associés à des vocabulaires contrôlés ou à des thésaurus). Les dossiers contiennent des relations (relations internes au document, relations avec d'autres documents électroniques ou des ressources externes), ces relations sont de différents types (hiérarchiques, sémantiques, associatives...).

L'objectif est de parvenir à une méthode générique de modélisation des dossiers permettant d'aboutir à un "*schéma de documentation*", puis d'appliquer cette méthode aux dossiers produits par la Direction de l'Architecture et du patrimoine.

## I - Contexte du projet au sein de la Direction de l'Architecture et du Patrimoine

La DAPA comprend différents services intervenant sur le patrimoine (Archéologie, Inventaire Général, Monuments Historiques, Qualité des espaces et de l'architecture et Ethnologie).

Des projets de passage au "dossier électronique" existent au sein de la DAPA depuis 1995, avec deux expérimentations fondées sur l'utilisation du standard SGML.

Pour l'Archéologie, le travail s'est engagé sur les Bilans Scientifiques Régionaux (BSR), avec la mise au point d'une DTD spécifique.

Pour l'Inventaire Général, le projet avait pour objectif la constitution directement en numérique et l'archivage au format SGML de l'ensemble de la documentation produite sur les œuvres patrimoniales.

La documentation produite par l'Inventaire comprend des dossiers documentaires "papier", issus des travaux de recherche et d'enquêtes des équipes régionales, indexés par des notices informatiques dans les bases de données documentaires (base Mérimée pour les dossiers d'œuvres architecturales et base Palissy pour les dossiers d'objets mobiliers). Sur chaque œuvre, les dossiers sont organisés en différents chapitres : identification (désignation et localisation), historique, description (...), ces dossiers comprenant également des illustrations (sous forme de plans, photos, relevés...), indexées par ailleurs dans la base nationale d'images Mémoire, et des références bibliographiques, pouvant être versées dans la base nationale Archidoc.

Les expérimentations réalisées dans le cadre de deux conventions de recherche passées entre l'INRIA et la Mission de la Recherche et de la Technologie du Ministère ont déjà permis de valider la faisabilité du dossier électronique : mise au point d'une première version de la DTD des dossiers d'Inventaire (DTD CI V2.x) et test d'une chaîne de production dans deux régions pilotes (Bretagne et Poitou-Charentes).

Un CD-ROM présentant les dossiers réalisés à partir de l'inventaire mené sur la ville de La Rochelle a été largement diffusé fin 1998 (<http://www.culture.fr/culture/inventai/itiinv/larochelle/index.html>).

Le bilan de trois années d'expérimentation en Poitou-Charentes a été présenté dans le cadre de la journée d'étude organisée en 1998

([http://www.culture.gouv.fr/culture/mrt/numerisation/fr/seminaire\\_du\\_161198/index.htm](http://www.culture.gouv.fr/culture/mrt/numerisation/fr/seminaire_du_161198/index.htm)).

En Bretagne, la production de dossiers au format électronique est d'ores et déjà opérationnelle, dans le cadre de l'opération d'Inventaire Préliminaire d'Ille et Vilaine (inventaire sur 6 ans), grâce au développement d'un outil de production - RENABL - au sein du service régional, qui compte la présence d'un informaticien, Pierrick Brihaye. Un CD-ROM sur le premier canton ainsi inventorié - Dol de Bretagne - a été remis au Conseil Général d'Ille et Vilaine fin octobre 1999. L'opération menée en Bretagne avait également été présentée lors de la journée consacrée à la documentation numérique en novembre 1998.

Le déploiement de dossiers électroniques dans l'ensemble des régions et son extension au sein des services de la DAPA peuvent désormais être envisagés.

Cela suppose la mise au point préalable, par le service informatique du ministère (Département de l'Organisation et des Systèmes d'Information - DOSI- de la DAG), des outils logiciels permettant la production, l'exploitation et la diffusion des dossiers électroniques. Le développement de ces outils va faire l'objet d'un dossier déposé au prochain Comité Ministériel des Systèmes d'Information, et sera intégré au schéma directeur informatique de la DAPA en cours d'élaboration.

Mais cela demande également de poursuivre le travail de recherche sur la structuration de la modélisation, et cela selon deux axes :

- un premier travail d'approfondissement du travail sur les DTD mises au point dans le cadre des opérations de recherches déjà réalisées : réflexion sur la " granularité des DTD " et mise au point d'un schéma de documentation,
- un second travail sur la transversalité des DTD : il s'agira d'étudier la faisabilité de structures de documentation génériques ou interopérables au sein des différents services travaillant sur le patrimoine, avec pour objectif la facilitation des échanges de données et la mise au point d'interrogation conjointe de plusieurs sources d'information.

## **II - Travail sur la structuration de la documentation (DTD)**

Ce travail comprend deux objectifs :

### **1) travail sur la granularité des DTD et mise au point d'un schéma de documentation**

La DTD mise au point pour l'Inventaire correspond à la structure de dossiers finalisés, c'est-à-dire une fois que tous les éléments constitutifs de la documentation ont été rassemblés.

Elle ne prend pas suffisamment en compte les différentes étapes de la constitution du dossier, qui amènent les chercheurs à ne travailler que sur des sous-éléments de cette DTD, en plusieurs étapes. Elle ne prend pas non plus suffisamment en compte le caractère réutilisable de certains de ces éléments : certains éléments, comme l'illustration ou la bibliographie, peuvent être partagés entre plusieurs dossiers, être stockés dans une base de données partagée, être importés ou exportés depuis cette base ...

Il importe donc d'approfondir le travail déjà réalisé par un travail sur la "modularité" ou sur la "granularité" des DTD.

Ce travail consistera à expliciter les différents "objets" étudiés (œuvres architecturales et mobilières, ensembles...), et à formaliser les relations qu'ils entretiennent entre eux dans le cadre des travaux de recherche opérés par l'Inventaire. Les différents éléments de documentation liés à chacun de ces objets devront être formalisés, débouchant sur la définition des différents dossiers documentaires (dossiers d'œuvre, dossiers de synthèse...), qui seront à modéliser dans le cadre d'un "*schéma de documentation*" prenant en compte l'aspect "modulaire" de la documentation.

Le "*schéma de documentation*" doit permettre de prendre en compte les trois aspects suivants :

1-1) modélisation de la structure de la documentation, débouchant sur l'établissement d'une DTD ;

1-2) modélisation des relations mises en œuvre dans la documentation : relations inter et intra documents (à spécifier et à typer dans la DTD) mais aussi relations avec des ressources externes (gestion des import/export de données, liens dynamiques vers des ressources externes). Ce dernier aspect devra permettre en particulier de préciser le degré de "modularité" à prévoir au sein de la DTD. Certains éléments réutilisables (entre plusieurs dossiers) ou récupérés depuis des ressources externes (que ce soit dynamiquement ou non) devront être identifiés et modélisés ("sous" DTD) ;

1-3) modélisation de la nature des éléments de la documentation, débouchant sur la spécification des contrôles à appliquer aux différents éléments et aux liens : typage des éléments et des liens, association à des listes d'autorités, cohérence...

Cette modélisation devra prendre en compte les différentes étapes de la constitution du dossier (différentes étapes de saisie et d'intégration des données externes, de contrôles, de publication...), et donc prévoir les différents états des dossiers ("provisoires", "à valider", "publiables"...).

Compte tenu de l'avancement de la réflexion sur la documentation de l'Inventaire (existence d'une DTD), il est proposé que cette partie de la réflexion soit poursuivie dans le cadre du groupe de travail sur la DTD Inventaire mis en place à la sous-direction.

## **2) travail sur la DTD transversale**

L'intégration à la démarche de modélisation des préoccupations des autres services patrimoniaux (Archéologie, Monuments Historiques, Services de l'architecture) permettra de valider la faisabilité d'une structure documentaire générique.

Il est déjà apparu en effet que de nombreux volets documentaires sont également utilisés par d'autres services patrimoniaux, que ce soit dans des bases de données (AGRIPPA, pour les Monuments Historiques, ou PATRIARCHE, pour les services de l'Archéologie) ou dans de la documentation papier. Ils pourraient donc faire l'objet d'une description transversale au sein de la DAPA. Il s'agit principalement des volets liés :

- à l'identification de l'œuvre patrimoniale (désignation et localisation des œuvres),
- aux sources documentaires (bibliographie, archives...),
- aux illustrations.

La réflexion permettra notamment d'établir :

- la pertinence de la mise en commun de certains éléments documentaires, compte tenu des usages différents et des modes d'alimentation,
- la description des éléments communs ou à rendre interopérables : au cas par cas, il devra être précisé si la structuration de l'information peut être commune aux différents services, ou si elle doit être rendue compatible (tables de correspondance ou d'appariement, travail sur les vocabulaires), ou si les éléments manipulés par chaque service peuvent être simplement reliés entre eux par des relations (à définir et à typer),
- la description des éléments transversaux ("sous" DTD bibliographie...) partagés par tous,
- l'étude du mode d'échange et de mise à jour des données (dynamique ou non).

Cette réflexion devra étendre son spectre à la structuration de dossiers documentaires qui ne sont pas forcément liés directement aux dossiers d'Inventaire et dont le traitement en DTD ferait l'objet d'un développement ultérieur : dossiers de secteurs sauvegardé, création à partir de Patriarche et en liaison avec les Bilans Scientifiques Régionaux des dossiers de site archéologique liés à un SIG, des dossiers scientifiques d'œuvre ayant fait l'objet d'une étude par le Laboratoire de Recherche de Monuments Historiques ...

# La documentation informatisée au ministère de la culture et de la communication

par Jean-Louis Pascon<sup>1</sup>  
Département de l'organisation et des systèmes d'information

Au sein du ministère de la culture, la documentation informatisée sur le patrimoine repose sur deux éléments complémentaires : d'une part le dossier électronique qui est l'outil de production et de consultation de la documentation et d'autre part le plan de numérisation qui est le moyen d'illustrer ces dossiers à partir de l'immense fonds iconographique dont dispose le ministère.

## 1. La documentation électronique

### 1.1. Les recherches en 1999

#### *La plate-forme technique de saisie des dossiers électroniques*

Depuis trois ans, un programme de recherche et développement, destiné à explorer les voies permettant de renouveler les méthodes et les outils employés par les directions pour leur production documentaire, a été mis en place.

La première phase de ce programme de recherche a été centrée sur la réalisation d'un prototype de diffusion qui a permis de présenter les potentialités des outils numériques exploitant les normes de structuration de l'information, comme SGML, d'aborder les questions de cartographie électronique, ainsi que les possibilités d'interrogation via Internet.

Des serveurs tests ont été expérimentés pour l'archéologie (bilans scientifiques régionaux) et pour l'Inventaire (inventaires de la ville de la Rochelle, du canton de Dôle en Bretagne). Ils ont démontré tout l'intérêt de ce type de systèmes pour faciliter l'accès en ligne aux dossiers patrimoniaux complets. La deuxième phase, qui est sur le point d'être terminée, vise à définir la totalité d'une chaîne de production d'informations.

Un prototype a été mis au point et sera présenté à la communauté scientifique du ministère au cours du premier trimestre 2000. Un cahier des charges pour le développement d'une application destinée à être déployée dans les services et les DRAC résultera de cette phase d'étude, le DOSI en coordonnera la réalisation en raison des objectifs directement opérationnels du dispositif prévu.

### 1.2. L'étude de la structuration des documents

Durant l'année 1999, toutes les questions relatives à la documentation électronique des différents domaines patrimoniaux ont été examinées. Des "Définitions de Types de Documents" (DTD), qui sont des schémas de structuration d'ensemble de données, ont été élaborées pour les archives, l'archéologie, le patrimoine bâti et les dossiers d'artistes contemporains.

#### *La DTD EAD*

Les services d'archives ont des volumes considérables de données à gérer, sur des supports diversifiés (documents papiers, photographies, cartes anciennes et modernes, plans, etc.). Ils doivent faire face à des demandes d'usage très variées.

---

<sup>1</sup> Jean-Louis Pascon est responsable avec Alain Maulny du comité scientifique "informatique documentaire et multimédia" du conseil ministériel de la recherche

Les archives disposent d'une longue tradition en matière de méthode de dépouillement des fonds et d'un système de description à plusieurs niveaux (de l'inventaire général des fonds aux inventaires pièce à pièce).

C'est pourquoi, il a semblé important d'examiner l'utilisation de la DTD EAD (Encoded archival description) mise au point par les Archives nationales américaines afin de fournir un cadre à un plan de développement de la documentation électronique dans les services d'archives. L'étude réalisée en 1999 a fourni la définition des typologies des instruments de recherche, les liaisons entre ces instruments, les principaux outils informatiques nécessaires aux archivistes, ainsi que les apports de ce type d'outils lors de l'usage des instruments ainsi produits.

### ***L'archéologie***

Une DTD permettant la saisie et le stockage des bilans scientifiques régionaux (BSR) a été élaborée. Le but est d'accélérer la production et la diffusion de ces bilans à la communauté scientifique en permettant aux services régionaux de l'archéologie de participer à un réseau national de diffusion en ligne et en leur offrant des outils numériques cohérents facilitant la communication rapide des résultats.

Quelques documents ont été produits par cette méthode et installés sur un serveur afin de réaliser les premiers tests. Des saisies complémentaires devront étoffer le contenu de ce système d'informations en 2000, afin de pouvoir faire des démonstrations auprès des services intéressés.

Les documents finaux de synthèse (DFS) devraient, à leur tour, faire l'objet d'une étude pour examiner, en relation avec la programme "Patriarche", les améliorations qui pourraient être apportées au dispositif actuel de communication des rapports de fouilles archéologiques.

### ***Les dossiers d'artistes***

En collaboration entre la Direction des musées de France et la Délégation aux arts plastiques, une recherche a été entreprise en vue de définir les éléments documentaires principaux qui constituent un dossier d'artiste, dans le but de rationaliser leur production et de faciliter les échanges.

Une première définition de la structure de ces dossiers a été réalisée. Au début de l'année prochaine, un ensemble significatif de dossiers va être saisi par des documentalistes intéressés par l'expérimentation afin de pouvoir estimer la validité de cette structure documentaire et ses possibilités d'exploitation par les musées et centres d'art.

### ***L'inventaire général***

L'inventaire général a poursuivi ses travaux sur la modélisation de sa documentation. Avec le recul permis par des opérations menées en vraie grandeur, comme en Bretagne, il a été estimé que la structure générale, telle qu'elle a été définie, notamment dans le cadre du programme sur la ville de La Rochelle, est satisfaisante.

Cependant, il est apparu que pour une pratique quotidienne il fallait simplifier certains aspects de la DTD afin de pouvoir manipuler des parties de documents (par exemple pour partager des images numériques entre des dossiers correspondant à plusieurs monuments).

### ***Les notices bibliographiques***

Afin de créer des ressources documentaires partageables par tous les services, il a été décidé de lancer une première étude sur les notices bibliographiques qui sont en effet très largement utilisées au sein du ministère.

Le but de cette étude n'est pas de redéfinir les notices bibliographiques au format UNIMARC préconisées par la Direction du livre et de la lecture mais de les transformer au format XML, directement exploitable sur Internet. A terme, ce format de présentation pourrait devenir une norme d'échange de notices bibliographiques.

### **1.3. Les projets de recherche pour 2000**

Les recherches génériques concernant la structuration des dossiers documentaires intéressant les directions devraient être achevées en 2000. Dans les années à venir, ce sont les services qui vont devoir désormais gérer ces travaux documentaires.

Les outils numériques permettant la création et la mise à jour de documents fondés sur ces structures, vont être développés par le Département de l'organisation et des systèmes d'information (DOSI). Ces outils, reprendront des éléments des chaînes actuelles de traitement des données documentaires afin d'assurer une transition souple entre la situation présente et la future documentation électronique.

En 2000, un programme de recherche sur la structuration des documents et la mise au point de DTD sera mené pour assurer l'harmonisation des concepts communs aux divers services du ministère et garantir une plus grande fluidité des échanges d'informations entre les documents.

Le plan de numérisation des fonds d'Etat et les recherches sur la documentation électronique ont fait émerger de nouveaux problèmes ; le principal d'entre eux concerne le traitement de masse.

En effet, le développement du plan de numérisation se heurte aux problèmes d'indexation des documents. Si la numérisation physique d'un millier de documents ne prend qu'une trentaine de minutes que dire du temps nécessaire à leur indexation individuelle lorsque seule existe une indexation globale de l'ensemble ?

La seule solution pour avancer rapidement dans le processus de numérisation des fonds patrimoniaux passe par l'utilisation de systèmes automatisés d'aide à l'indexation.

Trois voies de recherche sont proposées par le comité.

#### ***– L'aide à l'analyse du texte***

Dans le passé, on a souvent opposé indexation plein texte et indexation par mots clés ou par thesaurus. Il semble préférable, comme le confirment certaines recherches, de conjuguer ces deux techniques. L'étude proposée en 2000 vise à tester un système d'analyse sémantique de documents qui suggérerait des éléments d'indexation au documentaliste. Si une telle approche était viable, le temps de saisie de nouvelles notices pourrait être fortement réduit.

#### ***– La reconnaissance de textes manuscrits***

Des prototypes de logiciels permettant l'analyse de documents manuscrits sont disponibles depuis quelques mois. Il semble donc intéressant d'examiner rapidement les possibilités de ces nouveaux outils pour accélérer l'indexation de documents, tels les registres d'état-civil.

#### ***– L'indexation directe d'images numériques est une voie nouvelle***

Des travaux menés ces dernières années par l'INRIA ou le CEA ont conduit à des logiciels d'indexation d'images utilisant des algorithmes qui s'adaptent aux documents traités. Ils permettent des indexations rapides et des consultations par typologie d'images. Il est envisagé en 2000, d'explorer la pertinence de ces outils pour certains domaines d'indexation intéressant le ministère.

#### ***La consultation des données***

Enfin, il convient de continuer à explorer les aspects liés à la consultation des données. Le projet "Aquarelle" a permis de montrer comment pouvaient être effectuées, sur un même sujet, des interrogations simultanées de bases de données distinctes mais portant sur des thématiques semblables. Par contre, l'interrogation simultanée de bases de données traitant de sujets différents pose des problèmes beaucoup plus complexes. Il est donc proposé de tester de nouvelles méthodes de consultation qui nécessitent de définir des modèles (ou des ontologies) du domaine traité.

## **2. Le plan de numérisation**

### **2.1. Les appels à projets**

Deux appels à projets ont été lancés en 1999 auprès des services centraux et déconcentrés du ministère par le comité scientifique "informatique documentaire et multimédia". Le premier portait sur la numérisation de phototypes ou de microformes. Le deuxième appel à projets a eu pour priorité d'une part l'aide à la documentation pour réaliser des reprises d'inventaire et la numérisation en direct de plans, cartes, registres ou albums.

Au total, 74 dossiers de candidature ont été reçus, 59 dossiers ont été retenus dans leur totalité ou en partie pour un financement total de cinq millions de francs.

En 2000, en plus des huit millions de francs (titre V, hors BCRD) destinés à la numérisation des fonds d'Etat et qui continueront à être attribués aux services par l'intermédiaire d'appels à projets, le ministère va disposer de quatre millions de francs (titre VI) qui vont permettre de subventionner la numérisation de fonds patrimoniaux n'appartenant pas à l'Etat.

Le comité propose de distribuer ces crédits par la voie d'appel à projets.

Quatre critères d'éligibilité pourraient être retenus : le caractère remarquable du fonds, l'existence d'un véritable projet de valorisation et de partenaires éditoriaux, la garantie concernant la titularité des droits sur les documents et leur exploitation sous forme électronique, la présence d'une équipe capable de mener à bien le projet dans le délai d'une année.

A l'issue du processus de sélection, une convention devrait fixer les engagements de chacune des parties.

Enfin, il semble, au regard des résultats des précédents appels à projets, qu'il soit devenu nécessaire d'afficher certaines priorités sectorielles et d'inciter à la numérisation de fonds au travers d'une politique concertée. Pour ne citer qu'un exemple, une part importante du cadastre napoléonien a été numérisée, il conviendrait maintenant de terminer la couverture complète de la France.

Le comité se réunira dans le courant du mois de janvier 2000 pour discuter de telles priorités.

Les représentants de la Bibliothèque nationale de France seront invités à se joindre aux discussions et des complémentarités seront recherchées avec le programme de numérisation de cet établissement lorsqu'il associe des services de l'Etat en région.

### **2.2. La convention de prix**

La convention de prix avec la société Jouve qui a remporté le marché national s'est mise progressivement en place en 1999. Le démarrage de cette convention a été plus lent que prévu en raison de difficultés techniques. Un type particulier de phototypes (les négatifs en couleur) a posé des problèmes d'étalonnage de la chaîne de production.

Bien que cette convention soit accompagnée de plusieurs manuels et de logiciels pour élaborer les documents de suivi et que des formations aient été organisées pour les agents chargés de mettre en œuvre les programmes de numérisation, il existe encore des dysfonctionnements lors des premiers envois de phototypes au prestataire.

En 2000, il conviendrait de mettre en place des formations in situ pour aider les services à démarrer les campagnes de numérisation en fonction de leurs besoins spécifiques. Des réunions ont eu lieu avec le prestataire afin d'améliorer cette partie de la procédure et un avenant à la convention est en cours de négociation pour intégrer ces évolutions.

A la date du 15 décembre 1999, près de 35 000 images ont été numérisées en 6 mois, représentant environ 200 cédéroms. Il semble donc possible de parvenir à un total de 100 000 images numérisées en année pleine et de graver 1 200 cédéroms (en tenant compte des doubles de sécurité).

### **2.3. Les nouveaux marchés**

Plusieurs marchés sont en cours de mise au point et devraient être opérationnels au début de l'année. Cinq de ces marchés sont destinés à fournir des prestations de numérisation " in situ ". Les cartes et plans, les registres (comme les registres d'Etat Civil), les objets, les plaques de verre et certains types de monuments sont concernés par ces marchés.

Un marché particulier sera passé en vue de mettre en place une assistance documentaire. Celle-ci n'a pas pour objectif de fournir un travail documentaire se substituant à celui des services, il s'agit d'une aide ponctuelle motivée par l'importance scientifique du fonds ou par une opération d'édition.

Cette assistance va permettre d'indexer rapidement certains fonds mais son coût très élevé ne permettra pas de généraliser un tel soutien à tous les services.

Enfin un dernier marché est destiné à mettre en place un contrôle de la qualité des supports d'archivage, notamment pour veiller à leur conservation dans le temps.

Par ailleurs, une convention va être passée avec la RIEP (Régie Industrielle des établissements Pénitenciers) pour la numérisation d'archives sonores. Cette convention est établie en partenariat avec l'INA qui utilisera également les services de cette régie pour ses opérations de numérisation de fonds sonores.

La mise en place de ces marchés est longue et complexe, tant au niveau de la définition des prestations (quel découpage des tâches faut-il adopter pour des opérations de numérisation ?) que de la recherche d'entreprises compétentes en France et Europe. Ainsi trois lots des marchés de numérisation ont du être déclarés infructueux du fait de l'absence de réponse.

### **3. Conclusion**

D'autres sujets seront à traiter dans les années à venir, comme les SIG (les Systèmes d'Information Géographique) ou la préservation de l'information sous forme numérique, enjeu majeur.

L'année 2000 sera celle des premiers déploiements des nouveaux outils de saisie et de l'exploitation des premiers résultats du plan de numérisation.

## **Table ronde**

### **Évolution du plan de numérisation des fonds d'État bilan pratique 1996-1999 nouvelles orientations pour 2000-2001**

#### **Intervenants :**

Françoise Baligand, directeur du Musée de la Chartreuse de Douai,  
Xavier de La Selle, directeur des archives départementales de l'Aube  
Thierry Delcourt, directeur de la bibliothèque municipale de Troyes  
Frédéric Novicki, direction régionale des affaires culturelles de Picardie

#### **Modérateur :**

Alain Maulny, direction régionale des affaires culturelles de Poitou-Charentes

**Alain Maulny** invite les intervenants à décrire leur projet, les problèmes rencontrés lors de la mise en place des opérations de numérisation, les solutions mises en œuvre, ainsi que les aspects relevant de la diffusion et de la restitution au public.

**Françoise Baligand**, qui intervient comme responsable d'un projet de numérisation mené depuis une dizaine d'années par l'association des conservateurs des musées du Nord-Pas-de-Calais qui regroupe 34 musées, rappelle que le projet d'informatisation des collections a démarré en 1990 et s'est caractérisé par la recherche de compatibilité avec les systèmes nationaux et le choix d'un logiciel commun "Micromusée". Cette opération a été financée par les collectivités locales qui ont pris à leur charge l'équipement, le travail des personnels.

En 1993 a démarré le plan de numérisation, piloté par l'association qui finance notamment :

- le salaire d'un responsable qui assure la coordination des documentalistes et des photographes, suit les opérations de numérisation et fait le lien entre les conservateurs et les informaticiens ;
- les salaires de documentalistes (étudiants d'histoire de l'art formés à l'informatique) qui assurent la saisie des données ;
- la campagne photographique.

Les opérations sont destinées à la mise en œuvre du serveur internet *Musenor* (<http://www.musenor.org/>) sur lequel sont stockés textes et images.

*Musenor* comprend :

- une base de données (50 000 images et fiches à terme)
- des informations sur les activités des musées, leurs produits...
- des parcours thématiques (sur les thèmes des expositions temporaires, ou, comme c'est le cas pour un document sur la peinture nordique de 1450 à 1550 en cours de réalisation, véritables expositions virtuelles réunissant des œuvres qui ne peuvent être déplacées).

Le financement est assuré à hauteur de 50 % par la communauté européenne (FEDER), 25 % étant à la charge de l'État et 25 % à la charge des collectivités.

- 1<sup>re</sup> tranche 1997-1998 : 1,740 MF
- 2<sup>e</sup> tranche 1999-2000 : 2,603 MF
- 3<sup>e</sup> tranche fin année 2000 : 3,218 MF

Dans le cadre du contrat de plan 2000-2006, une demande de subvention de 1,2 MF a été faite (1/3 Etat, 1/3 région, 1/3 collectivités territoriales) pour poursuivre la numérisation et mettre l'accent sur la diffusion (bornes dans les musées...).

Les problèmes rencontrés portent sur le décalage entre la décision de subvention et son versement (6 mois à un an) mais sont surtout d'ordre technique, et achoppent sur les questions d'hébergement (le pôle universitaire européen de Lille (PUEL) s'étant révélé défaillant, le choix s'est porté sur un hébergeur privé, mais un retour est prévu vers le PUEL à nouveau), avec, en suspens, la question de l'accès du réseau Rénater hors du monde universitaire, celle de la création éventuelle d'un site miroir etc.

Autre problème rencontré, la signature de la convention avec la ville de Lille qui préfère réserver ses images à son propre serveur ; pour les collections d'art contemporain (Musée de Villeneuve d'Ascq et Musée Matisse du Cateau), il est prévu de créer des liens avec le réseau Videomuseum.

**Frédéric Novicki** indique que le projet de numérisation a démarré en 1997 en Picardie. Dans un premier temps, limité aux services de l'inventaire et de l'archéologie, il s'est élargi au service régional de l'archéologie, aux archives départementales de l'Oise et de la Somme, à la bibliothèque municipale d'Amiens. C'est au total six services qui sont concernés par le projet sur des thématiques communes : "Heurs et malheurs de la population picarde aux XIXe et XXe siècles", et "Patrimoine bâti en Picardie".

Le service de l'inventaire assure la coordination entre les partenaires régionaux, les directions centrales et la société Jouve.

Deux mille documents (NB et couleur) ont été numérisés à ce jour par la société Jouve, soit quarante-quatre cédéroms.

Les problèmes rencontrés sont les suivants :

- les délais de numérisation : 12 semaines pour les archives départementales de la Somme, 21 semaines pour les archives départementales de l'Oise, 9 semaines pour l'inventaire ;
- le manuel "Phototypes et microformes", peu compréhensible, devrait être réécrit, à la lumière de solutions mises au point récemment ;
- l'interface entre partenaires (régionaux, centrale, société Jouve) ;
- l'harmonisation des grilles descriptives entre services (inventaire / monuments historiques / archéologie) ;
- les conditions de diffusion.

La qualité de la numérisation est satisfaisante.

**Jean-Louis Pascon** rappelle qu'il est fondamental que les services se posent la question de la diffusion des images numérisées, des conditions d'exploitation, au moment de la conception du projet (édition sur papier, cédéroms de présentation des fonds, outils de recherche). Les réponses appartiennent aux services. Des essais d'impression en quadrichromie seront faits en 2000.

Il est également rappelé qu'il faut se conformer au protocole existant (seuil minimum d'images à traiter etc.).

Enfin, des délais de six à neuf mois ont été observés dans certaines DRAC pour la signature des marchés.

Les **services de la direction régionale des affaires culturelles de la région Poitou-Charentes** font également état de retards dans les livraisons des documents numérisés : aucun retour sur les dix mille documents envoyés depuis juin 99 à la société Jouve.

**Thierry Delcourt** et **Xavier de La Selle** présentent conjointement l'opération menée de façon concertée sur les fonds relatifs à l'abbaye de Clairvaux, qui a été engagée avec des objectifs communs :

- la mise à disposition de fonds rares et précieux provenant de l'abbaye<sup>2</sup>, auprès du public sur Internet
- la réalisation de produits dérivés, ainsi que des publications imprimées
- le projet pourrait s'élargir au patrimoine mobilier et immobilier de l'abbaye

À la bibliothèque municipale, le choix s'est porté sur des manuscrits des XII<sup>e</sup> et XIII<sup>e</sup> s. et, dans certains cas, des XIV<sup>e</sup> et XV<sup>e</sup> s., soit 6 000 vues (présentation en pleine page ainsi que de détails d'enluminures et de miniatures).

Aux Archives départementales, c'est le chartrier et les sceaux qui ont été photographiés et numérisés, soit 1 500 clichés (2/3 des documents concernent le cartulaire, 1/3 les sceaux).

Les spécifications techniques ont été définies en commun, notamment en ce qui concerne la production des clichés, ce souci a été poussé jusqu'à l'achat de lots identiques de pellicules. Il a porté également sur la constitution de classeurs mis à disposition de la société Jouve.

Aux archives départementales, cette opération a donné lieu à un travail documentaire important, plus lourd que la numérisation, car il n'existait ni inventaire ni photographies. La rétro-conversion des fichiers manuscrits datant de 1930 et comportant une analyse des chartes, a permis la création d'une base de données autorisant des recherches fines.

Pour la bibliothèque municipale (BM), les aspects positifs concernent :

- le traitement scientifique des documents, la découverte de certains manuscrits (ce travail est à l'origine de deux expositions)
- la mise en place de collaborations scientifiques palliant l'isolement de la bibliothèque et l'absence de personnels qualifiés, avec l'Institut de recherche et d'histoire des textes (IRHT), la prise en charge de la numérisation des manuscrits de la BnF dans le cadre de la numérisation des fonds de la BM, le montage d'un projet de collaboration avec le Musée national des arts et traditions populaires sur la littérature de colportage...
- l'obtention de crédits complémentaires attribués par la ville, conséquence du choix par le ministère de la culture des établissements comme sites pilotes.

Les aspects négatifs rencontrés par la bibliothèque municipale portent sur :

- le retard dans la numérisation (suite à des retards dans la passation des marchés)
- les difficultés de restitution des blancs pour les manuscrits sur vélin
- l'adressage, la limitation du nombre de caractères pour les fichiers de récolement ne permettant pas de retranscrire le RBCCN
- les niveaux de restitution prévus, aucun niveau n'autorisant la lecture à l'écran pour une diffusion sur internet, un niveau de résolution intermédiaire a fait l'objet d'une commande complémentaire sur des crédits de la collectivité locale.

Les archives départementales, qui achèveront la base de données d'ici à trois ou quatre mois, n'ont pas rencontré de problème de délai, ni d'adressage, et soulignent l'excellente qualité de la numérisation.

**Jean-Pierre Dalbéra** souligne les forces de ces projets de numérisation (la qualité des ressources, les compétences mises en œuvre à cette occasion, les banques d'images qui se créent associées à de nouvelles bases de données) et les faiblesses (peu d'édition électronique, peu d'opérations concertées de valorisation culturelle, peu de contacts avec les mondes de l'éducation et du tourisme susceptibles d'utiliser les nouvelles ressources multimédias). Il rappelle les coopérations possibles avec les grands établissements publics, les autres ministères, le monde associatif ainsi que la

<sup>2</sup> Les manuscrits provenant de l'abbaye de Clairvaux sont dispersés dans trois bibliothèques : bibliothèque municipale de Troyes, BnF, Bibliothèque inter-universitaire de Montpellier.

panoplie de soutiens existants au ministère de la culture et de la communication (aide à l'édition multimédia au CNC, dans les directions etc.) et au ministère de l'éducation nationale, de la recherche et de la technologie (labelisation des projets à vocation éducative, etc.).

## Table ronde

### L'organisation de la production des connaissances scientifiques et de la diffusion documentaire au ministère dans le contexte du plan d'action du Gouvernement pour la société de l'information

#### Participants :

Bruno Bachimont, directeur de la recherche à l'Institut national de l'audiovisuel,  
Christophe Dessaux et Luc Lièvre, direction de l'architecture et du patrimoine,  
Michel Fingerhut, directeur de la médiathèque de l'Ircam,  
Michel Florenzano, chercheur CNRS et directeur du MAP (Modélisation, simulation de l'architecture et du paysage), unité mixte de recherche Ministère de la culture / Ecole d'architecture de Marseille / CNRS)

#### Modérateur :

Jean-Pierre Dalbéra, chef de la mission de la recherche et de la technologie de la direction de l'administration générale

**Michel Florenzano** décrit l'un des programmes de recherche de l'unité mixte de recherche (<http://www.map.archi.fr/>) qu'il dirige qui porte sur le relevé d'architecture pour reconstruire ou construire des édifices architecturaux.

Il s'agit de formaliser la connaissance sur l'architecture en l'instrumentalisant par des processus informatisés et préenregistrés. Le relevé est une opération spécifique qui s'inscrit dans un cadre générique, le relevé renseigne le modèle.

Deux outils techniques sont utilisés (<http://www.map.archi.fr/table ronde99/>) :

- Un système de balayage laser 3D, dispositif dérivé de l'industrie, qui peut relever une grande quantité de points. Son domaine d'application : les objets de dimensions moyennes, sculptures, bas reliefs ...
- La photogrammétrie, qui à la différence du balayage laser, est un moyen semi-automatique, à post traitement.

Les étapes du traitement :

- modèles géométriques, base de départ au calcul de l'image 3D, obtenus par le relevé
- texturation
- positionnement des lumières
- étapes liées à l'animation

La problématique est opérationnelle, et laisse ouvertes les possibilités de coopération avec les opérateurs du patrimoine. Il s'agit d'attacher aux images produites une valeur sémantique plus forte que l'image elle-même, d'utiliser le modèle de l'objet pour naviguer dans l'information qui a présidé à sa restitution, voire à sa conception. L'image est une représentation parmi d'autres de l'objet dont on a le modèle.

C'est la question de la documentation structurée au sein de la direction de l'architecture et du patrimoine que **Christophe Dessaux** s'attache à traiter, en rappelant que le projet sur le dossier électronique qui vise à dématérialiser l'ensemble de la documentation produite sur les œuvres, a fait l'objet, depuis 1995, de programmes de recherche menés dans le cadre d'une convention entre le ministère de la culture et l'INRIA. Ces programmes ont abouti à la création d'une définition type de documents numériques adaptés aux besoins des services patrimoniaux. Des tests de la chaîne de production ont été menés :

- en Poitou-Charentes, où un cédérom sur La Rochelle a été diffusé en 1998,

- en Bretagne, où la production de dossiers électroniques est maintenant opérationnelle ; un cédérom sur Dôle a été réalisé en 1999.

Les programmes futurs portent sur :

- les outils de production qui devraient être réalisés en l'an 2000 avec l'aide du Département de l'organisation et des systèmes d'information. Le schéma directeur de la direction de l'architecture et du patrimoine va intégrer ces nouveaux besoins.
- L'adaptation des structures organisant les documents numériques aux conditions réelle de travail des services (modularité des outils, interopérabilité, partage de vocabulaires et de thésaurus etc.).

**Bernard Randoïn** (sous-direction de l'archéologie) évoque la création de documents structurés pour les ***Bilans scientifiques régionaux*** (BSR), qui sont des résumés de la totalité des opérations archéologiques publiés sous forme papier par les services régionaux de l'archéologie. Les responsables du projet, soutenu par la mission de la recherche et de la technologie, sont confrontés au choix d'un éditeur SGML adapté aux utilisateurs.

**Jean-Pierre Dalbéra** rappelle par ailleurs qu'il est urgent ouvrir le chantier visant à structurer un dossier électronique destiné aux rapports de fouilles archéologiques (DFS) compatible avec le programme "Patriarche".

Il passe ensuite la parole à **Martin Sévigny** qui travaille actuellement, sous contrat avec la mission de la recherche et de la technologie, à l'élaboration de documents numériques structurés (DTD) sur les dossiers d'artistes (arts plastiques, musées etc.) et sur les références bibliographiques (format UNIMARC) : <http://www.culture.fr/BiblioML> . La livraison de ses travaux aura lieu au cours du premier semestre 2000.

Une question dans la salle porte sur la structuration des écrits administratifs, pour leur diffusion courante, leur consultation, et la constitution d'archives.

**Luc-Florent Lièvre** fait l'état des lieux de la cartographie à la DAPA qui passe de la production cartographique à l'information géographique, c'est à dire l'appréhension directe des réalités du territoire, en citant deux exemples :

- la nouvelle application de la carte archéologique "Patriarche" repose sur une base de données "ORACLE" et un système d'information géographique (SIG), l'unité d'enregistrement étant l'information archéologique localisée.
- Le projet d'atlas d'architecture et du patrimoine, dont le cœur du dispositif est un SIG, qui traite des édifices archéologiques et architecturaux.

Une nouvelle DTD " localisation ", transversale aux différents services, est en cours de définition.

Cette montée en puissance de l'information géographique induit des coûts importants, des besoins de formation, et pose la question de la disponibilité de fonds de cartes (cadastre numérisé, fonds de l'Institut géographique national IGN) et de la diffusion des données (la confidentialité de certaines données ayant des implications sur la définition du projet).

Le Ministère de la culture sera en mesure de devenir un producteur de données géographiques, s'il parvient à s'équiper et à maîtriser ces outils.

**Michel Fingerhut** rappelle le rôle de l'Institut de recherche et de coordination acoustique/ musique (Ircam) en tant que laboratoire de recherche associé au CNRS, producteur d'œuvres musicales (bandes analogiques, supports numériques) qui ne sont pas toujours édités et en tant que producteur d'écrits publiés dans des périodiques, des livres, des actes de conférence... L'Ircam disposait, depuis de nombreuses années, d'une bibliothèque spécialisée dans la musique contemporaine.

En 1995, il a été décidé de transformer cette bibliothèque en médiathèque (<http://mediatheque.ircam.fr/>), tout en gardant le modèle et les supports de bibliothèque (salle de lecture ouverte à tous publics, documents papier, prêt) et en y ajoutant les archives des fonds de l'Ircam, des films sur la musique, des disques compacts, des cédéroms, des bases de données, des

références de sites internet... L'objectif a été d'intégrer l'ensemble de ces supports dans un même système et de rendre accessible catalogue et documents par le même système de consultation.

En outre, la disponibilité informatique de tous ces fonds permet de réaliser des documents hybrides à but pédagogique (analyses musicologiques, présentation de partition et de musique simultanée, etc.). Six mois ont été nécessaires pour la définition du projet, six mois pour sa mise en œuvre.

### ***Formats et supports***

Le choix a été fait de numériser ces différents types de médias à des fins de conservation et de diffusion. La numérisation n'offrant pas de garantie de conservation vu les changements dans les formats et les standards, il peut être nécessaire de pouvoir transcoder les données (en changer le format avec son évolution), d'où le choix de formats structurés afin de faciliter les conversions ultérieures éventuelles.

Le système s'appuie sur les standards du web, ce qui assure une compatibilité relativement répandue. Les textes sont en HTML, plusieurs formats sont utilisés pour l'audiovisuel, selon qu'il s'agit de conservation (échantillons 44,1 KHz gravés sur disque compact) ou de diffusion (MPEG 1 Layer 2). D'autres formats sont à l'étude (Real audio). VRML est utilisé pour la réalisation d'une interface en réalité virtuelle de la Médiathèque et de son catalogue. Les métadonnées sont au standard UNIMARC pour tous les documents, comprenant un lien vers le document numérisé le cas échéant, et des extensions particulières à la musique, définies en collaboration avec d'autres bibliothèques musicales.

### ***Diffusion, confidentialité***

Outre l'accès direct aux documents "papier" disponibles sur les étagères, le système informatique offre plusieurs modes d'accès à leur catalogue et aux documents numérisés: formulaires, listes (par plan de classement ou catégories de documents), plan au sol, modèle 3D...

L'accès aux documents multimédia (son, vidéo) est fourni via des interfaces visant à reproduire des objets familiers (panneau de lecteur de CD ou de magnétoscope). Les cédéroms sont lancés via un lien hypertextuel, et retournent à la page Web consultée après leur exécution.

Les possibilités de consultation informatique des documents ne sont pas les mêmes selon que l'utilisateur est dans la médiathèque ou consulte les fonds via les réseaux, les fiches biographiques, par exemple, comprennent une photographie et certains renseignements textuels visibles en consultation interne mais pas en externe.

Enfin, le navigateur fourni sur site est simplifié, permet un contrôle d'accès vers des sites présélectionnés et préserve la plate-forme sous-jacente de tout accès (lecture, écriture) par l'utilisateur.

### ***Développements futurs***

La recherche portera sur l'ajout d'outils de recherche intelligents dans les corpus de musique (partition, son) et l'intégration du protocole Z39.50 pour la consultation simultanée de catalogues distants.

**Bruno Bachimont** rappelle que la première mission de l'INA est de conserver et de diffuser.

### ***La conservation***

La dégradation et l'obsolescence rapide des supports obligent à procéder à des transferts sur des supports numériques, l'hétérogénéité des supports et des formats posant de nombreux problèmes pour la maintenance du parc de matériels de lecture.

En raison de l'évolution rapide des formats, des critères de recherche, des exigences des utilisateurs et des futurs transferts à prévoir, la numérisation doit être capable d'anticiper ces besoins en prenant en compte toutes les données indispensables (numérisation riche).

## ***L'indexation***

Le document audiovisuel présente plusieurs spécificités :

- c'est un document temporel
- c'est un document non alphabétique, il n'est pas possible de repérer facilement une unité.

Dans la pratique, le document audiovisuel est visionné par un spécialiste, il fait l'objet d'une transcription en langue naturelle, à partir de laquelle est effectuée l'indexation. Le texte est nécessaire pour accéder rapidement à l'information, car l'accès à l'image est très long. C'est le texte qui prescrit sa signification à l'image.

L'indexation automatique donne des résultats encore incertains :

- La segmentation en plans est satisfaisante.
- La segmentation en scènes ne fonctionne pas ou mal : c'est encore un enjeu de recherche.
- La reconnaissance des locuteurs n'est pas fiable.
- La transcription automatique fonctionne à 60 %.

L'alignement automatique d'un texte sur la parole fonctionne mais est très peu robuste (cela requiert une bande son de bonne qualité, et une qualité de discours – diction, débit, etc. – très homogène).

L'analyse automatique ne peut pas en général être effectuée sur des flux compressés, il est nécessaire de décompresser les fichiers. En effet, il est difficile de travailler directement sur le flux MPEG : on pourrait, dans une première approche vouloir par exemple utiliser le vecteur mouvement présent dans l'encodage MPEG pour retrouver le mouvement constaté des objets sur l'image. Mais le calcul du vecteur mouvement par les encodeurs MPEG n'est pas prescrit par la norme, et plusieurs méthodes sont possibles. Ces méthodes ne calculent pas nécessairement des mouvements constatés entre des objets présents sur l'image, mais seulement qu'un bloc de pixel sur une image est très similaire à un bloc de même taille sur une image voisine, mais positionné à un autre endroit sur cette image. Le vecteur mouvement de MPEG correspond alors à la distance entre les positions des deux blocs : c'est le déplacement d'un bloc sur l'image. Rien n'assure que le bloc corresponde à un objet pertinent de l'image et il n'y a pas de recouvrement nécessaire entre le vecteur de mouvement et l'attribution d'un mouvement à un objet.

Par conséquent, on voit que les technologies conçues pour la compression et la transmission se prêtent mal à l'indexation et l'interprétation du contenu. C'est pourquoi, jusqu'à présent, seule une personne peut associer du sens à un flux, les outils de l'indexation automatique constituant seulement une aide au documentaliste.

## ***Le génie documentaire***

Dans le cadre de l'évolution de la documentation en XML, la spécificité de l'audiovisuel intervient, puisqu'il s'agit de deux éléments en parallèle, la vidéo et le son. En particulier, il faut pouvoir décrire l'intrication du son et de l'image : par exemple, qu'un segment sonore commence avant la fin d'un plan, et se termine après la rupture de plan : les segments sonores et visuels se chevauchent. Mais SGML et XML n'autorisent pas le chevauchement de deux unités de description et leur extension à l'audiovisuel n'est donc pas immédiate. A présent, dans le cadre de MPEG-7, les travaux s'orientent sur des schémas XML et RDF.

En résumé, plusieurs disciplines concourent au traitement de la documentation audiovisuelle :

- Le traitement du signal
- Le génie documentaire (XML, SGML, SMIL)
- Le monde de la représentation des connaissances

Les données complexes requièrent des langages riches tels que RDF qui a été proposé par le monde de l'intelligence artificielle. Est fait appel à la terminologie, et aux ontologies. Enfin les usagers doivent pouvoir jouer leur rôle en matière de définition de profils d'indexation.

Les normes représentent un enjeu majeur. Actuellement on normalise à priori, en créant un cadre de travail. Il importe que ce cadre soit suffisamment souple et large pour que les différents usages

puissent s'exprimer. Dans MPEG-7, le but de la normalisation est de proposer un langage commun pour l'indexation et la description des contenus. Ce langage commun est une syntaxe formelle informatique et s'exploite à travers des outils informatiques : la description devient un enjeu technique et industriel. Il importe que le point de vue de la description et des contenus soit pris en compte dans cette entreprise. Or, ce sont les ingénieurs du signal, auteurs de MPEG1, 2 et 4, qui participent pour le moment à MPEG-7, la communauté des usagers étant largement absente de ces travaux. L'enjeu est par conséquent que les praticiens des contenus, les communautés des spécialistes travaillant sur les contenus, interviennent dans MPEG-7, et se fassent entendre. Sinon, l'industrie des contenus, dont MPEG-7 n'est que la prémisse technologique, se constituera sans eux.

# **Conclusions de la journée d'étude sur la numérisation au ministère de la culture et de la communication**

**par Jean-Pierre Dalbéra**  
Chef de la mission de la recherche et de la technologie

## **1. Le plan de numérisation des fonds d'Etat**

Le principal objectif de la journée d'étude était d'établir un bilan du plan de numérisation et des programmes associés afin d'en améliorer le fonctionnement et d'établir les priorités des actions à mener en 2000.

### **a) un plan mobilisateur et incitatif**

Le plan de numérisation des fonds iconographiques et sonores appartenant à l'Etat a été lancé en 1996, avec le soutien initial de la DATAR et après une enquête sur les besoins des services et établissements. Depuis cette date, il s'est poursuivi chaque année et a bénéficié d'un financement croissant (5 MF en 1999 et 8 MF en 2000 sur le chapitre 56.91).

En 2000, une dotation de 4 MF a été inscrite au chapitre 66.91 afin de subventionner des programmes de numérisation menés par des partenaires publics ou privés qui seraient complémentaires des projets des services du ministère.

Pour la numérisation de fonds d'Etat, la procédure mise en place a permis de soutenir de nombreuses initiatives, prises dans toutes les régions de France par des services centraux, à compétence nationale ou déconcentrés. Le rôle de coordination des DRAC est demeuré déterminant pour maintenir une cohérence dans les actions régionales, soutenir les services les moins bien pourvus et susciter des collaborations entre équipes.

En permettant à tous les services d'expérimenter les outils numériques, le plan de numérisation a incité au regroupement de moyens et à leur mutualisation. Ainsi, plusieurs DRAC (PACA, Pays de la Loire, Champagne-Ardenne, etc.) ont mis en place des pôles techniques sur l'image dont la vocation est de travailler pour tous les services et de rassembler les compétences.

Dans certaines régions, de véritables synergies se sont créées, en interne, mais également avec des partenaires extérieurs et des collectivités locales. Le financement du ministère de la culture et de la communication a permis de débloquer des crédits locaux qui ont accéléré les programmes de numérisation, menés notamment pour des archives départementales ou des bibliothèques municipales, dont les responsables ont montré un grand dynamisme.

Contribuant à la modernisation des méthodes de travail et de diffusion des services publics culturels, le plan de numérisation a incité au renouvellement d'équipements informatiques et à la formation de nombreux agents aux technologies de l'information et de la communication.

Les photographes ont été les premiers à s'approprier les techniques numériques de production et de traitement d'images. D'autres catégories de personnels (documentalistes, chercheurs, conservateurs, ingénieurs, etc.) ont suivi les stages de formation organisés par la direction de l'administration générale dont le nombre a augmenté chaque année.

L'Ecole nationale du patrimoine a également joué un rôle important de sensibilisation aux nouveaux outils de production, de traitement et de diffusion des connaissances.

Le plan de numérisation a conduit les services à s'interroger sur les moyens et systèmes d'information utilisés (bases de données, sites web, cédéroms, etc.) et à s'engager avec l'appui des directions centrales dans des recherches et des études pour les améliorer.

Cette dynamique collective doit être maintenue car elle est particulièrement bénéfique. Elle accélère la modernisation du ministère de la culture et de la communication en s'inscrivant pleinement dans le plan d'action du Gouvernement pour préparer l'entrée de la France dans la société de l'information. Elle offre au public de nouveaux services et des moyens de connaissance du patrimoine culturel.

Au moment où l'Etat va signer des contrats de plan avec les régions qui comportent d'importants projets de numérisation, tous les services du ministère doivent renforcer leurs capacités d'expertise et être capables d'apporter des réponses aux nombreuses questions scientifiques, techniques et culturelles qui sont posées à l'Etat.

Leur rôle est d'aider à la définition de cahiers des charges des projets de numérisation afin de maintenir une cohérence nationale et de garantir l'interopérabilité des systèmes d'information publics. Cet objectif sera atteint si les matériels, systèmes et logiciels retenus privilégient les normes internationales de codage de l'information et assurent la pérennité des informations numérisés.

## **b) les contraintes de la numérisation automatisée**

Le marché national passé avec la société Jouve après l'appel d'offres européen permet au ministère de bénéficier de prix réduits pour une numérisation de masse de très haute qualité.

Une organisation a été mise en place au niveau national pour mettre en œuvre le marché de sous-traitance, assurer le contrôle de la qualité des travaux et offrir l'accès aux documents numériques.

Cette méthode de traitement automatisé des phototypes et des microfilms comporte plusieurs contraintes au niveau des envois à effectuer au prestataire (conditionnement et quantité de documents), de l'établissement de bordereaux d'accompagnement des documents, de l'homogénéité des lots à numériser, des formats prévus dans le marché, de la procédure à respecter, de la vérification de la titularité des droits.

Certaines DRAC ont rencontré des difficultés pour tenir les délais imposés, pour remplir les bordereaux informatisés en raison de l'inadaptation de leur matériel informatique ou pour assurer l'indexation complète des ensembles de phototypes à numériser.

Durant la période de démarrage du processus, des retards dans la numérisation des premiers lots envoyés au prestataire ont pu être constatés mais au dernier trimestre de l'année le rythme de numérisation atteint 10 000 documents par mois.

Seule la numérisation automatisée de négatifs en couleur non étalonnés a posé des problèmes techniques qui n'ont pu être totalement résolus dans le contexte du marché. Dans ces cas particuliers, un traitement manuel doit être réalisé pour vérifier les résultats et corriger certains fichiers.

Soit une autre société devra effectuer les traitements complémentaires qui ne sont pas pris en compte dans le marché national, soit les traitements devront être faits par les services s'ils disposent de moyens. Des solutions techniques seront proposées par le comité de pilotage du plan de numérisation pour répondre aux besoins des services demandeurs.

Une réunion de travail sur ces sujets sera organisée en janvier par la MRT et le DOSI avec la société Jouve pour améliorer les procédures et la communication avec les services utilisateurs.

Enfin, une convention entre le ministère et la régie inter-établissements pénitentiaires (RIEP) du Ministère de la justice sera prochainement signée pour la numérisation des fonds sonores. L'atelier de

traitement mis en place par la RIEP est également utilisé par l'INA pour la numérisation d'émissions radiophoniques.

### c) la valorisation des fonds numérisés

Au sein des services, le manque de personnels disponibles et compétents pour assurer l'ensemble des tâches d'indexation, de préparation et de contrôle de la numérisation apparaît comme le principal facteur de ralentissement du dispositif national mis en œuvre. Des marchés complémentaires (numérisation in situ, aide à l'indexation, contrôle qualité, etc.) devraient apporter des solutions à ces problèmes. Toutefois, le coût de prestations de numérisation non automatisées ou d'indexation devient rapidement très élevé eu égard aux quantités de documents à traiter. En conséquence, ces types d'aides ne seront apportées en 2000 qu'à des projets majeurs, de haut niveau scientifique, fédérant plusieurs partenaires et comportant un volet de diffusion innovant.

Au cours de la journée d'étude, plusieurs interventions ont montré que la valorisation des fonds numérisés était insuffisamment prise en compte par les services alors qu'il s'agit d'une priorité du ministère. Des mesures incitatives seront étudiées début 2000 par le comité pour accélérer la diffusion en ligne des images dans les bases de données du ministère comme Mérimée, Mémoire ou des bases locales mais également sous des formes scénarisées : sites web, cédéroms thématiques, itinéraires du patrimoine en ligne, publications électroniques pédagogiques ou touristiques, etc.

Quelques opérations montrent néanmoins que des synergies importantes peuvent se constituer autour de projets scientifiques et culturels et être très productives en termes de développement et d'action ; c'est le cas par exemple à l'occasion de la numérisation des fonds de l'abbaye de Clairvaux, des archives historiques de Carcassonne ou des photographies historiques du musée Guimet.

Le catalogage des fonds numérisés doit être une des actions à mener prioritairement en 2000 afin d'informer les utilisateurs de leur existence et de faciliter leur accès. Actuellement, des fiches résumant la totalité des projets sélectionnés depuis 1996 par le comité de pilotage sont accessibles en ligne sur le site web consacré à la numérisation :

<http://www.culture.fr/culture/mrt/numerisation/>

Ces fiches décrivent sommairement les contenus des fonds à numériser et fournissent les coordonnées des services et des responsables ; quelques images illustrent certaines collections numérisées. Dans l'avenir, il serait souhaitable de disposer systématiquement d'exemples illustrant les fonds concernés, et surtout de créer un catalogue des fonds numérisés qui permettrait des interrogations par sujet, localisation, institution, etc. selon les normes des bibliothèques.

La direction du livre et de la lecture s'associera à ce catalogage en apportant des éléments d'information sur les fonds que les bibliothèques publiques numérisent avec leurs moyens propres et que la direction répertorie systématiquement.

## 2. Les programmes de recherche et développement associés

Depuis trois ans, un programme de recherche et développement, destiné à explorer les voies permettant de renouveler les méthodes et les outils employés par les directions pour leur production documentaire, a été mis en place.

La première phase de ce programme de recherche a été centrée sur la réalisation d'un prototype de diffusion qui a permis de présenter les potentialités des outils numériques exploitant les normes de structuration de l'information, comme SGML<sup>3</sup> et XML<sup>4</sup>, d'étudier les interfaces de cartographie électronique, ainsi que les possibilités d'interrogation via Internet.

<sup>3</sup> SGML : Standard generalized markup language

<sup>4</sup> XML : eXtensible markup language

Des " Définitions de Type de Documents " (DTD), qui sont des schémas de structuration d'ensemble de données, ont été élaborés pour les archives, l'archéologie, le patrimoine bâti et les dossiers d'artistes contemporains avec le soutien de la MRT.

Il est apparu que pour une pratique quotidienne il fallait simplifier certains aspects de ces schémas afin de pouvoir manipuler des parties de documents (pour partager par exemple des notices bibliographiques ou des images numériques entre des dossiers correspondant à plusieurs monuments).

Des serveurs tests ont été expérimentés pour l'archéologie (bilans scientifiques régionaux) et pour l'Inventaire (inventaires de la ville de la Rochelle<sup>5</sup> ou du canton de Dole en Bretagne). Ils ont démontré tout l'intérêt de ce type de système pour faciliter l'accès en ligne aux dossiers patrimoniaux complets.

La deuxième phase doit permettre de définir la totalité des chaînes de production d'informations afin de mettre en œuvre des dispositifs opérationnels adaptés à chaque secteur documentaire. Ces outils, reprendront des éléments des chaînes actuelles de traitement des données afin d'assurer une transition souple entre la situation présente et la future documentation électronique.

En 2000, un programme de recherche sur la structuration des documents et la mise au point des schémas des dossiers devra être mené pour assurer l'harmonisation des concepts communs aux divers services du ministère, garantir une plus grande fluidité des échanges d'information entre les documents et généraliser l'usage des interfaces graphiques.

Le plan de numérisation des fonds d'Etat et les recherches sur la documentation électronique ont mis en évidence les difficultés induites par le traitement de masse de documents hétérogènes auquel sont confrontés tous les services documentaires du ministère. Ils ont rappelé la nécessité de parvenir à des solutions numériques pérennes et normalisées, garantissant l'accès à distance et l'interopérabilité des systèmes d'information.

Une médiathèque comme celle de l'IRCAM<sup>6</sup> travaille dans cette voie pour résoudre les problèmes complexes qui se posent dans la conservation et la diffusion de dossiers comprenant des textes, des films, des partitions, des images fixes, des enregistrements musicaux, etc. Il en est de même à l'INA<sup>7</sup> où la direction de la recherche mène plusieurs programmes (en collaboration avec certaines chaînes de télévision et des établissements d'enseignement à distance) pour accélérer les méthodes d'indexation et de recherche de documents audiovisuels en tenant compte des besoins réels des utilisateurs et en offrant des outils de consultation performants.

Enfin, la documentation numérique intègre de plus en plus les images de synthèse notamment dans le domaine de l'architecture, de l'urbanisme et du patrimoine bâti mais également de certains musées (*par exemple au musée des arts asiatiques de Nice*). Des travaux sont menés notamment par le laboratoire « Modèles et simulations pour l'architecture, l'urbanisme et le paysage »<sup>8</sup> associant le CNRS et plusieurs écoles d'architecture pour améliorer les dispositifs de numérisation en trois dimensions, en faciliter l'usage et parvenir à mettre au point des systèmes documentaires de réalité augmentée.

Ces voies de recherche et développement trouvent dans la procédure d'aide PRIAMM<sup>9</sup>, mise en place par les ministères de la culture et de l'industrie de nouvelles possibilités de financement. Elles sont également au centre de plusieurs actions du programme cadre de recherche et développement : PCRD de l'Union européenne.

Au cours du deuxième semestre de l'année 2000, l'exercice de la présidence de l'Union européenne par la France conduira le ministère de la culture et de la communication à être particulièrement attentif à ces questions et à prendre des initiatives pour renforcer les coopérations entre les institutions culturelles, les laboratoires publics et les entreprises capables de faire progresser la maîtrise des outils numériques pour une meilleure diffusion des connaissances.

<sup>5</sup> <http://www.culture.fr/culture/inventai/itiinv/larochelle/index.html>

<sup>6</sup> IRCAM : Institut de recherche et coordination acoustique/musique (<http://mediatheque.ircam.fr/>)

<sup>7</sup> INA : Institut national de l'audiovisuel (<http://www.ina.fr/Recherche/>)

<sup>8</sup> <http://www.gamsau.archi.fr/UMR/SOURCES/umrmap.htm>

<sup>9</sup> PRIAMM : Programme pour la recherche et l'innovation dans l'audiovisuel et le multimédia ([http://www.cnc.fr/priamm/index\\_dyn2.htm](http://www.cnc.fr/priamm/index_dyn2.htm))