

## **Les systèmes d'information archivistiques en France**

(Claire Sibille, Direction des Archives de France)

### **Cet exposé abordera :**

- le schéma directeur des systèmes d'information du ministère de la Culture et de la Communication (MCC)
- le schéma directeur du système d'information des Archives de France (avec quelques exemples d'application en cours)
- les approches techniques (les outils de diffusion utilisés par la direction des Archives de France)

### **1. Le schéma directeur des systèmes d'information du MCC**

#### **Faciliter l'accès du public au patrimoine culturel**

- Le ministère de la Culture et de la Communication élabore et met en œuvre la politique culturelle de l'État
- Sa tâche principale est d'offrir une information complète et homogène sur le patrimoine culturel
- Un moyen pour cela est l'utilisation de l'Internet et des nouvelles technologies

#### **Problématique**

- S'adapter à de nouvelles modalités de diffusion et de valorisation du patrimoine culturel (jusqu'ici, juxtaposition de ressources numériques sans cohérence globale d'affichage et de contenu)
- Simplifier et clarifier pour le public l'accès aux ressources électroniques du MCC (bases de données, sites Web, œuvres numériques...)

#### **Besoins prioritaires**

- Prendre en compte les différentes catégories d'utilisateurs (chercheurs, étudiants, grand public, etc.)
- Offrir au public un accès unique aux données sur le patrimoine culturel (descriptions documentaires et images numériques), avec différents modes et niveaux de recherche et de consultation, via le portail culture.fr

#### **Diversité des applications en cours, quelques exemples**

- Bases de données comprenant la description d'institutions patrimoniales :
  - Museofile (Direction des Musées de France)

- 
- Catalogue des fonds culturels numérisés (Mission Recherche et Technologie)
  - Bases lexicographiques :
    - Thesaurus (Direction de l'Architecture et du Patrimoine)

### **Museofile**

<http://museofile.culture.fr/>

- Contient des informations sur environ 1300 musées en France :
  - chaque fiche comprend : le nom du musée, son sigle, sa localisation, ses collections et les services proposés ainsi que des liens vers le site web (s'il existe)
- Constitue le point d'accès central au web des musées de France
- Est destinée à fusionner dans un référentiel unique des institutions culturelles du MCC, en cours de développement

### **Catalogue des fonds culturels numérisés**

[http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f\\_02.htm](http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_02.htm)

- Développé sur une plate-forme open source (base de données XML ; 1000 notices et 500 images)
- Fournit des renseignements à la fois sur 800 collections culturelles (textes, images, audiovisuel...) numérisées dans le cadre du Plan annuel de numérisation et sur les institutions chargées de la conservation et/ou de la numérisation des collections physiques correspondantes
- En cours de transformation vers un portail européen Michael (actuellement France, Royaume-Uni et Italie)

### **Thesaurus**

<http://www.culture.gouv.fr/documentation/thesarch/pres.htm>

- Contient 6380 termes pour la description d'œuvres architecturales et mobilières
- Utilisé pour l'indexation dans les bases de données nationales Mérimée et Palissy
- Comprend des définitions et des notes d'usage, des équivalents anglais et italiens
- OAI en cours d'étude

### **Intégration des bases de données archivistiques**

- Deux facteurs importants faciliteront l'intégration des bases de données archivistiques au sein du futur portail culture.fr :
  - l'utilisation du XML et des DTD EAD et EAC
  - l'utilisation d'outils open source permettant la publication de documents XML et la mise en œuvre de l'OAI-PMH

### **culture.fr**

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

- Pour le moment, ne contient que du contenu éditorial

- Site web référençant plusieurs bases de données
- Il n'y a pas de recherche transversale
- Toutefois, une technologie de flux (ZOPE/CPS) permet la saisie locale de données dans un environnement distribué

### **Aquitaine patrimoines**

<http://demo.ajlsm.com/ap/web/>

- Préfigure ce que culture.fr deviendra
- Un fournisseur de services culturels (données moissonnées via le protocole OAI)
- Le portail regroupe diverses informations culturelles provenant de bibliothèques, archives, musées, etc.
- Différents modes de recherche : recherche simple (plein texte), avancée, en cliquant sur une carte géographique

## **2. Le schéma directeur du système d'information des Archives de France**

### **Faciliter l'accès du public aux archives**

- Les guides de recherche ou de bases de données d'intérêt national accessibles par Internet visent :
  - à faire connaître les documents conservés dans les services publics d'archives (archives nationales, départementales, municipales, régionales)
  - dans certains cas à permettre l'accès aux documents eux-mêmes, lorsqu'ils ont été numérisés

### **Problématique**

- Faire appliquer des normes en matière de description des fonds et de numérisation de documents
- Définir des priorités pour la numérisation des fonds d'archives (marchés de numérisation des centres des Archives nationales, plan national de numérisation)
- Mettre en place des portails nationaux de diffusion d'instruments de recherche et d'archives numérisées

### **Une politique fondée sur la définition de thèmes prioritaires**

- Définir des thèmes prioritaires, correspondant à des types de documents, dont la DAF encouragera la diffusion en mettant en place des portails nationaux, eux-mêmes exploitables dans le cadre de portails plus vastes
- Interface unique d'interrogation des instruments de recherche de différentes institutions (moissonnage de données via le protocole OAI-PMH), puis renvoi sur les sites d'origine pour l'affichage des documents

### **Nomina**

<http://www.france-genealogie.fr/>

<http://nomina.france-genealogie.fr/nomina/>

- Par exemple, sur le thème de l'histoire des familles et de la généalogie, le portail Nomina permet, à l'aide du protocole OAI-PMH, la consultation de plusieurs bases de données généalogiques réparties (Archives départementales de la Mayenne, Centre des archives d'outre-mer...) et l'accès à des documents numérisés
  - Environ 13 000 000 noms de personnes et de familles interrogeables

#### **Thèmes prioritaires envisagés**

- Les instruments de recherche
- Archives Canada-France
- Les grandes sources de données fondamentales
- Les images du patrimoine
- Les documents fondateurs de la mémoire commune
- Les fonds menacés

#### **Les instruments de recherche**

- La numérisation des instruments de recherche disponibles dans les services d'archives est vivement encouragée, même lorsque les documents eux-mêmes n'ont pas été numérisés
- Ces outils permettent en effet une première orientation des chercheurs, à partir des états généraux des fonds, vers des instruments de recherche détaillés et dernière étape, la consultation des documents eux-mêmes, lorsque ceux-ci ont fait l'objet d'une numérisation

#### **Base d'Orientation et de Recherche dans les Archives**

<http://daf.archivesdefrance.culture.gouv.fr/sdx/ap/>

- La direction des Archives de France a mis en œuvre une Base de données et d'Orientation dans les Archives (BORA), destinée à terme à devenir un portail d'information national pour la description, conforme à ISAD(G) de toutes les ressources archivistiques des services publics français
- BORA comprend actuellement deux modules, archives privées et archives photographiques
- BORA pourrait être étendue à d'autres types de documents (registres cadastraux, d'état civil, archives notariales)
- La base de données est progressivement alimentée, après validation, par des fichiers transmis par les institutions de conservation de ces ressources (8 000 notices descriptives de fonds conservés dans trois centres nationaux et cinquante services territoriaux)
- L'application BORA a été développée sur la plate-forme logicielle SDX, outil pour la consultation de documents XML utilisant le moteur de recherche Lucene
- Les notices sont structurées en EAD 1.0

#### **Les Archives nationales**

- Par ailleurs, les centres des Archives nationales mettent en ligne des états généraux de fonds :
  - états généraux partiels des fonds au centre historique des archives nationales (état sommaire des fonds privés, application ETANOT)  
<http://www.archivesnationales.culture.gouv.fr/chan/>
  - état général des fonds et inventaires pour le centre des archives du monde du travail  
<http://www.archivesnationales.culture.gouv.fr/camt/>
  - application IREL pour le centre des archives d'outre-mer  
<http://www.archivesnationales.culture.gouv.fr/caom/fr/index.html>

- état général des fonds sous la forme d'une base de données pour le Centre des archives contemporaines  
<http://www.archivesnationales.culture.gouv.fr/cac/fr/index.html>

- Le nouveau centre des Archives nationales de Pierrefitte comprendra une salle des inventaires virtuelle
- Dans une première étape, seront mis en ligne les guides et instruments de recherche décrivant les documents conservés dans les trois centres généralistes des Archives nationales (Paris, Fontainebleau et Pierrefitte), en attendant de pouvoir consulter les documents eux-mêmes

### **Application IREL (Instruments de recherche en ligne)**

[http://www.archivesnationales.culture.gouv.fr/caom/fr/index\\_irel.html](http://www.archivesnationales.culture.gouv.fr/caom/fr/index_irel.html)

- IREL est une application en ligne des Archives nationales à Aix-en-Provence (CAOM)
- L'accent a été mis dans un premier temps sur l'état des fonds
  - le nombre d'inventaires détaillés publiés est encore limité : bases de données XML comme l'état civil des Français d'Algérie, liens vers d'autres sites web comme Canada-France, instruments de recherche en EAD)
- Les résultats s'affichent de façon hiérarchisée, en fonction du cadre de classement
- Le choix d'une seule fenêtre pour le résultat et pour le contexte des unités documentaires pertinentes a été retenu
- La table des matières a été supprimée dans la mesure où, dans la partie gauche de l'écran, le contexte de la recherche permet de remonter dans les niveaux de description supérieurs
- Il est aussi possible de naviguer dans l'arborescence de l'état général sans passer par le formulaire de recherche
- La date de la dernière mise à jour est visible sur l'écran d'accueil
- Quand on descend dans les niveaux de description, on quitte le fragment de la réponse
- L'historique des recherches a été abandonné volontairement, c'est le contexte de la recherche qui a été considéré comme important
- Un lien permet de remonter à la liste des résultats, on peut aussi relancer très facilement la recherche
- Il n'y a par contre pas de possibilité de stocker les résultats dans un "caddie" ou dans un "panier"
- Il est possible de naviguer vers le résultat précédent ou suivant

### **Les archives territoriales**

- Les archives territoriales, de leur côté, sont de plus en plus nombreuses à constituer des outils d'orientation :
  - Archives départementales de l'Aube, de l'Aude, du Puy-de-Dôme, de la Haute-Vienne, de la Savoie, du Val-de-Marne, etc.

### **Archives départementales du Puy-de-Dôme**

<http://www.archivesdepartementales.puydedome.com/>

- En amont, l'application utilise les exports au format XML conformes à la DTD EAD 1.0 générés par le logiciel d'archives Arkhéia

- L'internaute accède à deux types d'instruments de recherche : état général des fonds (1091 fonds décrits à ce jour) et inventaires détaillés (56 en ligne)
- La publication des instruments de recherche s'effectuera progressivement au fur et à mesure de l'état d'avancement des classements et des reprises d'inventaires anciens sous Arkhéia
- Les internautes effectuant des recherches sur l'histoire des familles peuvent accéder directement à l'inventaire des archives de l'état civil et à l'état des fonds des notaires
- L'internaute dispose de plusieurs possibilités de recherche (sur l'état des fonds, les inventaires ou les deux à la fois) :
  - une recherche arborescente dans l'état des fonds présenté selon le cadre de classement des archives (possibilité d'afficher la liste de tous les inventaires en ligne et d'y accéder directement)
  - une recherche guidée limitée à une interrogation sur le champ d'indexation matières de l'état des fonds
  - une recherche libre en plein texte
- Les résultats s'affichent en suivant le plan de classement, le nombre d'occurrences étant reporté derrière chaque symbole d'affichage
- Au niveau de l'inventaire, l'affichage des occurrences (mots surlignés) se fait en tenant compte du contexte

### **Archives Canada-France**

<http://www.archivescanadafrance.org>

<http://bd.archivescanadafrance.org/acf/>

<http://demo.ajlsm.com/sdx-221/acf/home.shtm>

- Dans le cadre d'un projet de coopération franco-canadien, la numérisation de documents conservés en France et au Canada sur la Nouvelle-France sont encouragés
- Le portail Archives Canada France permet la diffusion de ces documents
- Le portail donne accès à une exposition virtuelle et une base de données contenant déjà plus de 20 000 notices descriptives, structurées en XML/EAD, des documents tirés des fonds français et canadiens, ainsi que les images des pages des documents
- Les fiches descriptives ont été exportées en EAD depuis une base de données
- Les fonds sont décrits dans des pages HTML indépendantes des documents EAD
- Une nouvelle version de la base devrait permettre de contextualiser une unité numérisée lors de sa consultation, afin de comprendre dans quel fonds, série organique, etc. elle se situe, quelles sont les autres unités de ces ensembles, etc.
- Par ailleurs, dans une prochaine phase, un volet pédagogique et un volet généalogique seront ouverts :
  - pour le premier, des outils de travail collaboratifs devront être réalisés afin de permettre à des classes de part et d'autre de l'Atlantique, de monter des projets communs de valorisation

-----

- pour le second, devront être à la fois atteints par le portail, des bases de dépouillements nominatifs déjà effectués ainsi que des bases à faire réaliser sur la base de grilles normées
- ces ressources seront elles-mêmes moissonnées par Nomina

### **Les grandes sources de données fondamentales**

- Les différents publics sont intéressés par certains types de documents, que les services publics d'archives numérisent ou envisagent de numériser en priorité :
  - documents sur l'histoire des personnes (état civil, recensement de la population, registres matricules de recrutement militaire...), des territoires (cadastre, terriers), des biens (actes des notaires, hypothèques), des entreprises, documents officiels (délibérations des collectivités territoriales, arrêtés, décisions de justice)...
- Dans le domaine des archives, un objectif de la numérisation et de la diffusion pourrait donc être de parvenir, pour quelques documents de base (concernant l'histoire des personnes et des territoires par exemple), à une couverture totale du territoire national
- Cela contribuerait pour le public, à la "lisibilité" des ressources disponibles
- La mise en place de cette infrastructure bénéficierait de conditions pratiques de réalisation très favorables :
  - sélection très facile des documents à numériser, qui représentent des masses importantes (un million de volumes environ pour l'état civil, 500 000 plans cadastraux, etc.) ;
  - documents non soumis aux droits de la propriété littéraire et artistique ;
  - documents existant en exemplaires uniques ou en un petit nombre de copies bien localisées (dans ce cas, une des copies doit être privilégiée)
- Par exemple, les documents sur l'histoire des personnes, qui font déjà l'objet de très nombreuses opérations de numérisation dans l'ensemble du réseau des services publics d'archives, peuvent, s'ils ont été indexés nominativement, être diffusés par le portail Nomina

### **Les images du patrimoine**

- Très nombreuses dans bien des services publics d'archives, les photographies et les cartes postales font partie des fonds les plus consultés par le grand public et ont déjà fait l'objet de nombreuses opérations de numérisation
- Sont en particulier concernées les images du patrimoine (monuments, objets, paysages,...)
- La numérisation et la mise en ligne des images du patrimoine (monuments, objets, paysages, portraits...) sont à encourager, à condition que les droits de propriété littéraire et artistiques aient pu être négociés
- Le Centre des archives d'outre-mer a déjà monté une base images pour ses propres fonds images (Ulysse) <http://demo.ajlsm.com/sdx-221/ulysses/index>

### **Les documents fondateurs de la mémoire commune**

- Au Centre historique des Archives nationales (Paris), la base Archim et la numérisation envisagée des archives exceptionnelles constituant le Musée de l'histoire de France entrent dans cette perspective  
<http://www.culture.gouv.fr/documentation/archim/accueil.html>

- Le site "L'histoire par l'image", mise en œuvre par la Réunion des musées nationaux, présente également de nombreux documents d'archives  
<http://www.histoire-image.org/>

### **Les fonds menacés**

- Même si la numérisation, en aboutissant à retirer les documents originaux de la consultation par le public, favorise toujours leur préservation, certains types de documents sont particulièrement sensibles et se dégradent naturellement même sans intervention extérieure : c'est le cas des enregistrements audiovisuels, de certains imprimés (par exemple la presse)...
- La numérisation de ces documents menacés devrait être encouragée, ainsi que leur mise en ligne dans la mesure du possible

### **Définir une structure de métadonnées**

- Pour chacun des types de documents concernés, une structure de métadonnées permettant leur interrogation commune devra être définie
- Cette structure doit être compatible à la fois avec les normes de descriptions des documents d'archives (ISAD(G), ISAAR(CPF), DTD EAD) et avec les standards techniques d'échange (OAI-PMH)
- S'inspirer pour cela de l'expérience du portail Nomina (structure commune définie pour les bases de données généalogiques)

## **3. Approches techniques : les outils de diffusion utilisés par la direction des Archives de France**

### **SDX (Système pour les Documents XML)**

- Développé avec le financement du MCC
- Logiciel open source permettant de construire des applications web où la recherche joue un rôle important
- Basée sur l'infrastructure Cocoon 2 de l'Apache foundation, SDX permet de construire des sites Web complexes adaptés à vos besoins
- SDX intègre Lucene (moteur de recherche entièrement développé en Java)

### **PLEADE**

- Logiciel libre permettant à chacun de publier des instruments de recherche archivistique en EAD dans un environnement Web (développé par les compagnies françaises AJLSM et Anaphore) <http://www.pleade.org/>
- repose sur d'autres technologies :
  - un serveur Web, comme Apache, utilisé pour gérer des requêtes HTTP depuis un navigateur Web et pour retourner l'information sous forme de pages HTML
  - un moteur de servlet Java, comme Tomcat d'Apache, avec une machine virtuelle Java
  - et SDX, la plate-forme de publication et de recherche de documents XML
- SDX supporte l'utilisation de thesaurus pour la recherche, le moissonnage ou les entrepôts OAI-PMH
- Ces fonctionnalités peuvent donc être ajoutées facilement à PLEADE

-----

### **Navimages**

- Suite logicielle permettant de préparer et de publier des séries d'images numériques sur le Web
- Comprend trois modules avec des fonctionnalités répondant aux trois objectifs généraux de Navimages : organisation et gestion d'images, interrogation des descriptions d'images, et visualisation des images et navigation dans un environnement distribué
- PLEADE et Navimages ont été conçus pour fonctionner ensemble

### **ZOPE/CPS**

- Zope est une application open source
- Zope est un serveur d'application Web qui peut être utilisé pour publier des " objets " sur le Web
- Ces objets peuvent être du plein texte, du contenu structuré, etc.
- Zope est particulièrement adapté pour la gestion de contenu et facilite la mise à jour de données

### **Conclusion**

- L'objectif principal du MCC est de mettre en œuvre un guichet unique donnant accès au patrimoine culturel français
- Dans ce contexte, le but de la DAF est d'avoir des bases de données interopérables et de les intégrer dans le portail culture.fr
- Ce travail sera facilité par la structuration des données (XML, EAD, EAC) et les outils libres