

Formation initiale des bibliothécaires d'État

**La gestion des collections de la  
Bibliothèque numérique de l'Institut  
National d'Histoire de l'Art (INHA) :  
fichiers sources et fichiers dérivés**

**Elodie DESSERLE**

Sous la direction de Sandra Rony

Chef du service informatique documentaire, Département de la Bibliothèque  
et de la Documentation - Institut National d'Histoire de l'Art

## **Remerciements**

Je tiens à exprimer toute ma gratitude à Mr Jean-François Vincent, responsable de la bibliothèque numérique Medic@ de la BIUM, ainsi qu'à Mr Olivier Rouchon, chargé de la plate-forme d'archivage du CINES qui ont bien voulu répondre à mes nombreuses questions.

Je souhaite également remercier Mr Dominique Maniez pour sa grande disponibilité et Mr Damien Fuvelle pour son soutien permanent tout au long de l'année.

Enfin, mes remerciements vont à l'ensemble du personnel de l'INHA et tout particulièrement à Mme Anne Weber et Mr Alexandre Bezsonoff qui pilotent la bibliothèque numérique, à Mr Pascal Presle du Service informatique et à Mme Sandra Rony, responsable du Service de l'informatique documentaire.

**Résumé :**

La bibliothèque numérique de l'INHA ne dispose pas de système de gestion adapté. Ce rapport a pour objectif d'aider à la mise en place d'un système efficace. Pour ce faire, il aborde la question de la gestion des collections numérisées au sein des bibliothèques, il analyse diverses stratégies de gestion puis formule des recommandations.

**Descripteurs :**

Institut national d'histoire de l'art. Bibliothèque

Bibliothèques virtuelles

Numérisation

Gestion électronique de documents

Archivage électronique

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée.
--

**Abstract :**

The digital library of the INHA does not have adapted management system. This report tries to help in the implementation of an effective system. For that purpose, it tells about the question of the management of digitized collections within libraries, it analyzes various strategies of management, then it formulates recommendations.

**Keywords :**

Virtual library

Hybrid library

Digital library

Digital preservation

Document imaging systems

Electronic filing systems

# Sommaire

<b>INTRODUCTION .....</b>	<b>9</b>
<b>PARTIE 1 : L'INSTITUT NATIONAL D'HISTOIRE DE L'ART .....</b>	<b>11</b>
1. L'ÉTABLISSEMENT .....	11
1.1. <i>La naissance d'un projet</i> .....	11
1.2. <i>Statut actuel et missions</i> .....	12
1.3. <i>L'état actuel du projet</i> .....	13
1.3.1. Le Département des études et de la recherche .....	13
1.3.2. La Bibliothèque .....	13
1.3.3. La rénovation du Quadrilatère Richelieu .....	13
2. LE DÉPARTEMENT DE LA BIBLIOTHÈQUE ET DE LA DOCUMENTATION .....	14
2.1. <i>Historique des collections</i> .....	14
2.1.1. Le fonds Doucet .....	14
2.1.2. Le fonds de la BENSBA .....	15
2.1.3. Le fonds de la BCMN .....	16
2.2. <i>Fonctionnement</i> .....	17
2.2.1. Principes fondateurs .....	17
2.2.2. Moyens humains .....	17
2.2.3. Public .....	18
<b>PARTIE 2 : LA BIBLIOTHÈQUE NUMÉRIQUE DE L'INHA .....</b>	<b>19</b>
1. QU'EST-CE QU'UNE BIBLIOTHÈQUE NUMÉRIQUE ? .....	19
1.1. <i>Définition</i> .....	19
1.2. <i>Typologie</i> .....	20
1.2.1. Bibliothèque virtuelle .....	20
1.2.2. Bibliothèque numérique ou électronique .....	20
1.2.3. Bibliothèque hybride .....	21
2. LA POLITIQUE DOCUMENTAIRE DE LA BIBLIOTHÈQUE NUMÉRIQUE DE L'INHA .....	21
2.1. <i>Objectifs et publics visés</i> .....	22
2.2. <i>Organisation de la collection</i> .....	23
3. LA CHAÎNE DE NUMÉRISATION .....	24

3.1.	<i>Choix techniques et contraintes juridiques</i> .....	24
3.1.1.	Numérisation en mode image.....	24
3.1.2.	La résolution et le format .....	25
3.1.2.1.	La résolution .....	25
3.1.2.2.	Le format.....	26
3.1.3.	Une prestation extérieure .....	27
3.2.	<i>Les étapes de la chaîne</i> .....	27
3.2.1.	Sélection et préparation.....	27
3.2.2.	Numérisation et contrôle de la livraison.....	28
3.2.3.	Traitement pour la mise en ligne .....	28

**PARTIE 3 : LA GESTION DES COLLECTIONS NUMÉRISÉES AU SEIN DES BIBLIOTHÈQUES .....30**

1.	CONTEXTE ET PROBLÉMATIQUE .....	30
1.1.	<i>Les enjeux</i> .....	30
1.2.	<i>Problèmes posés par la préservation de l'information numérique</i> .....	31
1.2.1.	Facteurs techniques.....	31
1.2.2.	Facteurs organisationnels et administratifs .....	32
1.3.	<i>Travaux et réflexion en cours</i> .....	32
2.	STRATÉGIES DE GESTION .....	35
2.1.	<i>Un système de gestion de documents numériques</i> .....	36
2.1.1.	Définition.....	36
2.1.2.	Principales fonctions.....	36
2.2.	<i>Un modèle fonctionnel : la norme OAIS</i> .....	37
2.2.1.	Les acteurs de l'archivage .....	37
2.2.2.	Les fonctions de l'archivage .....	38
2.3.	<i>Stratégies de conservation</i> .....	39
2.4.	<i>Les métadonnées</i> .....	41
2.4.1.	Définition.....	41
2.4.2.	Fonctions .....	42

**PARTIE 4 : EXPÉRIENCES DE GESTION DES DOCUMENTS POUR UNE BIBLIOTHÈQUE NUMÉRIQUE .....43**

1.	LA GESTION ET L'ARCHIVAGE DE LA BIBLIOTHÈQUE NUMÉRIQUE DE L'INHA.....	43
1.1.	<i>Les fichiers sources</i> .....	43
1.1.1.	Sur cédéroms .....	43
1.1.2.	Sur serveur .....	44
1.2.	<i>Les fichiers dérivés</i> .....	44
1.2.1.	Le format.....	44
1.2.2.	Les services proposés par Picturelan.....	45
2.	EXEMPLES D'AUTRES EXPÉRIENCES .....	45
2.1.	<i>La Bibliothèque numérique Médic@</i> .....	46
2.1.1.	Choix opérés pour la gestion des documents numériques.....	46
2.1.2.	Analyse.....	47
2.2.	<i>Le CINES (Centre Informatique National de l'Enseignement Supérieur)</i> .....	49
2.2.1.	La gestion des documents numériques au CINES .....	49
2.2.2.	Fonctionnement de la plate-forme d'archivage .....	49
2.2.3.	Analyse.....	51

**PARTIE 5 : RECOMMANDATIONS ET PROPOSITIONS POUR AMÉLIORER LA GESTION DES DOCUMENTS DE LA BIBLIOTHÈQUE NUMÉRIQUE DE L'INHA .....53**

1.	ANALYSE DES STRATÉGIES DE CONSERVATION .....	53
1.1.	<i>Pratiques non applicables à l'INHA</i> .....	53
1.2.	<i>Une stratégie envisageable : la migration</i> .....	55
2.	L'ORGANISATION D'UN SYSTÈME DE GESTION .....	56
2.1.	<i>Service des « entrées »</i> .....	56
2.1.1.	Le versement.....	56
2.1.2.	Après le versement .....	57
2.2.	<i>Le service « stockage »</i> .....	57
2.2.1.	L'archivage .....	57
2.2.2.	Surveillance des fichiers.....	58
2.3.	<i>Service de gestion des données et communication</i> .....	58
2.3.1.	La recherche de fichiers .....	58
2.3.2.	L'accès et la récupération de fichiers.....	59

<b>CONCLUSION .....</b>	<b>60</b>
<b>BIBLIOGRAPHIE .....</b>	<b>62</b>
1. GÉNÉRALITÉS SUR L'INHA .....	62
2. GÉNÉRALITÉS SUR LES BIBLIOTHÈQUES NUMÉRIQUES .....	62
3. STRATÉGIES DE GESTION DES BIBLIOTHÈQUES NUMÉRIQUES .....	64
4. PÉRENNISATION DES DOCUMENTS ÉLECTRONIQUES .....	66
5. RECOMMANDATIONS, NORMES, STANDARDS .....	68
<b>TABLE DES ANNEXES .....</b>	<b>70</b>

## ***Introduction***

Aujourd'hui, de nombreuses initiatives internationales, nationales, régionales et locales ont été lancées, grâce à des financements des secteurs public et privé, pour assurer la numérisation de notre patrimoine culturel. Bien que très diversifiés de par leurs objectifs, le type de ressources utilisées (ouvrages, textes de références, documents audio, vidéos, ....) et les utilisateurs visés, ces projets se caractérisent par un souci commun : optimiser la valeur de l'investissement que représente la numérisation en souhaitant que le contenu numérique produit soit aussi utile, portable et pérenne que possible.

Si dans un premier temps, les bibliothèques numériques ont créé, recueilli et donné accès à de l'information sans nécessairement s'engager à la conserver à long terme, la situation est en train d'évoluer. De grandes institutions ont pris conscience des risques encourus par ce nouveau patrimoine numérique et commencent à mettre en place des stratégies pour garantir son archivage et sa préservation.

En lançant son projet de bibliothèque numérique en 2002, l'Institut National d'Histoire de l'Art s'est inscrit dans cette dynamique de création de contenus culturels numériques. Partant du constat d'une faiblesse concernant la gestion, le contrôle et la préservation à long terme des documents produits dans le cadre de sa bibliothèque numérique, l'INHA a souhaité entreprendre une réflexion afin de garantir la fiabilité et l'exploitabilité de ces nouvelles ressources. Actuellement, les premiers problèmes (tels que des fichiers devenus illisibles sur des cédéroms gravés en 2002) ont déjà surgi et les difficultés sont exacerbées par le fait que l'INHA ne dispose pas, pour l'instant, de véritable système cohérent de gestion pour ses documents numériques.

Ce présent rapport se propose de fournir des éléments pour nourrir cette réflexion sur la gestion du cycle de vie des documents numériques et ceci, de leur création à leur archivage. Il a pour principal objectif d'assurer dans le futur système une approche cohérente de la création, de la gestion et de l'édition des ressources numériques au

travers de l'utilisation efficace de standards, de règles et de recommandations qui constituent de bonnes pratiques.

Après une présentation de l'établissement et de la bibliothèque numérique en particulier, ce rapport aborde la question de la gestion des collections numérisées au sein des bibliothèques. Il analyse ensuite divers exemples de stratégies de gestion adoptées par quelques bibliothèques numériques (dont celle de l'INHA). Enfin, le rapport soumet des orientations stratégiques et formule des recommandations en proposant un modèle de gestion de documents numériques qui pourra servir de base de négociation et de travail lors de la réalisation du futur système de banque de données multimédias.

## **Partie 1 : L'Institut National d'Histoire de l'Art**

### **1. L'établissement**

#### **1.1. La naissance d'un projet**

Dès 1973, Jacques Thuillier, conseiller du président Georges Pompidou, constate un manque en France en ce qui concerne l'enseignement et la recherche en histoire de l'art ; il émet, alors, l'idée de la création d'un institut français spécialisé dans ce domaine. Dix ans plus tard, en 1983, André Chastel<sup>1</sup>, mandaté par le gouvernement, rédige un rapport. Selon lui, la France accuse un sérieux retard en matière d'histoire de l'art et doit se doter d'un institut de recherche et d'enseignement comparable aux grands instituts étrangers, tels le *Courtauld Institute* et *Warburg Institute* de Londres, le *Zentralinstitut für Kunstgeschichte* de Munich ou encore le *Getty Research Institute for the History of Art* de Los Angeles.

D'autres rapports suivront : celui de Françoise Benhamou<sup>2</sup> et Pierre Encrevé<sup>3</sup> en 1992, ceux de Michel Laclotte<sup>4</sup> en 1995 et 1996. Ils précisent les contours du futur établissement. Ses principales missions seront la recherche, la documentation et la formation. Une grande bibliothèque devra être constituée, à partir de fonds déjà existants. L'établissement prendra place au sein du quadrilatère Richelieu-Vivienne, à Paris, dans les locaux laissés vacants, en 1998, par le déménagement du département des imprimés de la Bibliothèque nationale de France (BnF) vers le site de Tolbiac.

En 1999, l'Association de préfiguration de l'Institut national d'histoire de l'art est formée. Elle est présidée par un professeur de l'Université Paris I, Alain Schnapp, qui deviendra, par la suite, le premier directeur de l'établissement.

---

<sup>1</sup> Historien d'art disparu en 1990, qui fut membre de l'Institut et professeur au Collège de France.

<sup>2</sup> Professeur des universités à Rouen et à Paris I, économiste, chercheur au MATISSE-CNRS (économie de la culture).

<sup>3</sup> Linguiste et historien d'art.

<sup>4</sup> Historien d'art spécialiste de la peinture française et italienne des XIVe et XVe siècles, conservateur et directeur de musées

## 1.2. Statut actuel et missions

L'Institut national d'histoire de l'art est un établissement public à caractère scientifique, culturel et professionnel créé par décret le 12 juillet 2001<sup>5</sup>. Il est placé sous la tutelle conjointe des ministères de l'Education nationale (Direction de l'enseignement supérieur), de la Recherche et de la Culture.

L'objectif premier de l'INHA est de regrouper sur un même site des activités de recherche, de documentation, d'enseignement et de formation en histoire de l'art et archéologie. En rassemblant différentes ressources et institutions auparavant dispersées et en mettant à leur disposition les outils qui leur sont nécessaires, l'établissement a pour mission d'assurer une meilleure diffusion des connaissances en histoire de l'art.

Le soutien à la recherche s'opère à un niveau national et international par la mise en œuvre de programmes scientifiques, l'organisation de colloques, de journées d'études, d'expositions, la publication de thèses et l'accueil de chercheurs, de boursiers et de professeurs de toute la France et de l'étranger.

Ces activités de recherche ne peuvent s'envisager sans la présence d'un vaste ensemble de ressources documentaires. C'est pourquoi l'établissement est en train de constituer une grande bibliothèque et procède à la mise en place de programmes documentaires, auxquels sont associés les chercheurs (par exemple pour la constitution d'inventaires, de bibliographies ou de banques de données).

En ce qui concerne l'enseignement et la formation, l'INHA héberge des institutions d'enseignement de niveau recherche (troisièmes cycles d'universités, laboratoires du CNRS, sociétés savantes, Ecole pratique des hautes études, Ecole des hautes études en sciences sociales, Ecole nationale du patrimoine, ...)

L'établissement est piloté par un conseil d'administration<sup>6</sup> assisté par un conseil scientifique « amené à se prononcer sur toute question pouvant avoir une incidence en matière de recherche »<sup>7</sup>. L'INHA est structuré en deux départements : le Département des études et de la recherche et le Département de la bibliothèque et de la documentation.

---

<sup>5</sup> Décret n°2001-621 du 12-07-2001 portant création de l'Institut national d'histoire de l'art (J.O. n°162 du 14-07-2001)

<sup>6</sup> Annexe 1 : L'organigramme de l'INHA

<sup>7</sup> Termes du décret du 12 juillet 2001

## 1.3. L'état actuel du projet

### 1.3.1. Le Département des études et de la recherche

L'installation du Département des études et de la recherche au 2 rue Vivienne a débuté en 2004, le « Carré Colbert » a été officiellement inauguré en février 2005 après restauration et réhabilitation complètes des locaux. Les premiers programmes de recherche commencent à donner des résultats tangibles et le déploiement de capacités d'accueil et d'animation scientifique ont permis d'organiser un ensemble de séminaires, conférences, journées d'étude, rencontres et colloques qui inscrit d'ores et déjà ce campus de formation et de recherche dans le paysage français et international de l'histoire de l'art.

### 1.3.2. La Bibliothèque

La Bibliothèque, quant à elle, résultera de la fusion de trois grandes bibliothèques d'art :

- la Bibliothèque d'art et d'archéologie Jacques Doucet qui a rejoint l'INHA au 1er janvier 2003 : 807 580 documents,
- la Bibliothèque centrale des musées nationaux (BCMN) : 252 160 documents,
- la Bibliothèque de l'Ecole nationale supérieure des Beaux-Arts (BENSBA) : 128 000 imprimés.

Par ailleurs, il faut ajouter la Bibliothèque de l'Ecole nationale des Chartes (149 000 documents) qui s'installera sur le site et deviendra « bibliothèque associée ». L'ensemble représentera à l'ouverture un total d'environ 1 340 000 documents.

Les collections et les services seront déployés dans l'historique salle Labrousse de la Bibliothèque nationale de France et dans les quatre étages de magasins attenants. 445 places seront offertes aux lecteurs, une bonne partie des collections (environ 265 000 documents) sera disponible en libre accès.

### 1.3.3. La rénovation du Quadrilatère Richelieu

Le projet de rénovation du Quadrilatère Richelieu a été lancé en 2004. Il concerne l'INHA mais aussi les départements spécialisés de la BnF (Arts du spectacle, Cartes et plans, Estampes et photographie, Manuscrits, Musique, Monnaies, médailles et antiques) toujours localisés sur le site historique de la Bibliothèque Nationale. Le tout représente un ensemble complexe de 50 000 m<sup>2</sup> de bâtiments anciens en partie classés ou inscrits à l'inventaire des monuments historiques. Le chantier vise trois objectifs principaux : le

renforcement de la sécurité par des réaménagements et une mise aux normes techniques (accès, circulation, installations électriques ...), la modernisation et l'extension des services offerts au public présent et distant ainsi que l'amélioration des conditions de conservation des collections. Il est confié à l'Etablissement public de maîtrise d'ouvrage des travaux culturels (EMOC). Suite au choix de l'équipe de maîtrise d'œuvre architecturale, en 2007, les études techniques complémentaires se dérouleront de mai 2007 à mars 2009, elles aboutiront à une esquisse architecturale d'ensemble et à un schéma directeur technique. Le site sera maintenu en fonctionnement pendant les travaux qui s'échelonneront en plusieurs phases du printemps 2009 au printemps 2015. L'ouverture de la Salle Labrouste est prévue pour 2012.

## **2. Le département de la bibliothèque et de la documentation**

Comme le stipule une convention, signée par les universités Paris IV et Paris I, la Chancellerie des universités de Paris, et l'Institut national d'histoire de l'art, « la gestion des collections composant la Bibliothèque d'art et d'archéologie Jacques Doucet est confiée à l'Institut national d'histoire de l'art pour participer à la constitution d'un grand ensemble documentaire dans le domaine de l'histoire de l'art ». Ainsi l'Institut national d'histoire de l'art a accueilli, au 1er janvier 2003, les personnels et collections de la Bibliothèque d'art et d'archéologie, première des trois bibliothèques d'art concernées à rejoindre l'Institut. La volonté de l'Institut est d'inscrire sa politique documentaire dans le respect des conceptions fondatrices de Jacques Doucet.

### **2.1. Historique des collections**

#### **2.1.1. Le fonds Doucet**

Amateur et collectionneur, le couturier Jacques Doucet (1853-1929) avait rassemblé une riche collection d'œuvres d'art. Constatant l'état de pénurie où se trouvait alors la recherche en histoire de l'art, il décida de constituer une bibliothèque, en réunissant « les documents, imprimés, estampes, photographies nécessaires à l'étude de l'art de tous les temps et de tous les pays et de mettre cet ensemble à la disposition des travailleurs ». Il s'entoura de spécialistes (universitaires, conservateurs, galeristes, ...)

pour constituer ses collections d'art et pour dresser un recensement des textes, des recueils de gravures et des reproductions nécessaires à la connaissance de l'évolution artistique. Il se préoccupa d'acquérir systématiquement les sources elles-mêmes (manuscrits, textes imprimés fondamentaux), mais également des documents graphiques, recueils d'ornements et estampes, qu'il envisageait comme un moyen de diffuser les techniques de travail et comme un trésor d'images reflétant le goût de chaque époque. Afin d'illustrer les tendances de l'art contemporain, il créa un Cabinet d'estampes modernes (10 000 pièces) où les plus grands artistes sont représentés : Goya, Manet, Matisse, Toulouse-Lautrec...

Doucet consacra à sa bibliothèque un million de francs or par an et la collection s'accrut à un rythme extraordinairement rapide. Le souci de la diffusion complétait chez Doucet la passion du collectionneur. Il n'eut de cesse de multiplier les accès possibles à la recherche : ainsi, il fonda le Répertoire d'art et d'archéologie, outil de recherches bibliographiques, et ouvrit à tout le public des amateurs éclairés sa bibliothèque, installée dans son hôtel de la rue Spontini depuis 1909. Il fit établir des fichiers, copier des documents d'archives et des catalogues d'autographes, il lança des programmes de recherche. Il constitua une photothèque consacrée à l'art en achetant de nombreux clichés et en finançant des campagnes de prises de vue dans le monde entier. Il chercha aussi à rassembler de façon systématique des spécimens d'écriture de tous les acteurs du monde de l'art, collectionneurs, critiques, historiens...

En 1918, Jacques Doucet fit la donation de ses prestigieuses collections à l'Université de Paris : la Bibliothèque d'Art et d'Archéologie (BAA) prit le statut de bibliothèque universitaire. Elle fut par la suite rattachée à deux universités, Paris IV et Paris I. Logée initialement dans l'hôtel particulier de Jacques Doucet, rue Spontini, puis implantée dans les années trente rue Michelet, la bibliothèque a déménagé en 1993, pour s'installer dans le bâtiment Richelieu appartenant à la Bibliothèque nationale de France.

Elle relève de l'INHA depuis janvier 2003.

### 2.1.2. Le fonds de la BENSBA

La BENSBA est l'héritière de la bibliothèque de l'Académie royale de peinture et de sculpture créée par Louis XIV en 1648. En 1793, le gouvernement de la Convention décide de fusionner les écoles des anciennes académies de peinture, de sculpture et d'architecture en un seul établissement : l'École nationale supérieure des beaux-arts. La bibliothèque hérite des archives et des inventaires de l'Académie, d'un certain nombre

de morceaux de réception gravés, des ouvrages, des dessins, des moulages, des œuvres primées au concours annuel du Grand Prix de Rome. Les fonds sont constitués dans une visée pédagogique puisque la bibliothèque doit être une source de documentation et de modèles pour les élèves architectes, peintres et sculpteurs. La collection s'enrichit au fil des années grâce aux donations, pour devenir par son importance et sa qualité une des plus belles de France. L'école possède des ouvrages (relatifs à l'histoire de l'architecture, de la peinture, de la sculpture, du dessin et des arts décoratifs occidentaux), des périodiques (une collection close de 900 titres), des estampes, des dessins de maître (dont un large panorama de l'étude du nu du XVIIIe siècle à la première moitié du XXe siècle avec des œuvres de Le Brun, Mignard, Van Loo ou Boucher), des dessins d'architecture, des photographies (datant principalement de la période 1850-1914), des manuscrits, des peintures et sculptures (prix de Rome et concours divers depuis le début du XVIIIe siècle jusqu'à 1968, ainsi qu'un certain nombre d'œuvres, portraits et morceaux de réception dont ceux de Fragonard, David, Ingres, Houdon...).

En 1994, la bibliothèque est divisée en deux services documentaires :

- une médiathèque proposant aux étudiants une documentation sur l'art contemporain et un fonds général d'histoire de l'art en libre accès,
- le service des collections consacré à l'histoire de l'art et dans lequel a été regroupé la quasi-totalité du fonds de la bibliothèque d'origine.

Actuellement, le service des collections de l'ENSBA possède plus de 300 000 œuvres. Seule une partie des ouvrages qu'il conserve rejoindra la bibliothèque de l'INHA au moment de l'ouverture de la salle de lecture définitive.

### 2.1.3. Le fonds de la BCMN

La BCMN est rattachée à la Direction des musées de France du Ministère de la culture. Située au Louvre, dans le Pavillon des arts et dans le Pavillon de Flore, elle a pour mission de constituer, conserver et mettre à disposition les ouvrages et les périodiques nécessaires aux travaux menés dans les musées nationaux sur l'histoire de l'art occidental, de l'Antiquité au XIXe siècle.

Les premières acquisitions régulières d'ouvrages s'effectuent à partir de la Monarchie de Juillet. Mais, c'est au cours des années 1920 que la bibliothèque prend toute son ampleur en endossant le rôle de service central d'acquisition pour les bibliothèques d'un certain nombre de musées nationaux.

Actuellement, la moitié de ses acquisitions se fait par les échanges qu'elle réalise avec cinq cents institutions dans quarante-huit pays, grâce aux publications mises à sa disposition par la Réunion des musées nationaux.

Les fonds se composent de 160 000 ouvrages, d'une importante collection de catalogues de ventes publiques, de 1800 périodiques (dont 1000 en cours) ainsi que de collections spécialisées (lettres d'artistes, dossiers et correspondances de conservateurs, d'archéologues, de collectionneurs et marchands, dessins, inventaires). La bibliothèque couvre les domaines de l'art occidental (de l'antiquité classique au XIXe siècle), de l'antiquité égyptienne, de l'antiquité du Proche et du Moyen Orient et de l'art islamique.

## 2.2. Fonctionnement

### 2.2.1. Principes fondateurs

La conception de la bibliothèque a été pensée en adoptant un certain nombre de principes fondateurs, exposés dans les différents rapports ayant accompagné la naissance de l'INHA. Ces grands principes sont les suivants :

- la bibliothèque est destinée à combler le retard de la France en matière de sources documentaires pour la recherche en histoire de l'art,
- la bibliothèque formera, avec la recherche, l'un des deux socles de l'INHA, pensés en complémentarité et enrichissement réciproques,
- la bibliothèque offrira à un public de chercheurs, de professionnels, d'amateurs une part importante de ses collections d'ouvrages et de périodiques en libre accès,
- la bibliothèque reflétera l'actualité de la recherche au niveau international, par un recours important aux sources et réseaux informatiques de tous types,
- la bibliothèque fera de la qualité de ses collections et de ses services aux utilisateurs ses exigences premières ; elle offrira à son public un ensemble de services documentaires approfondis,
- une bibliothèque numérique offrira accès à distance à ses principaux fonds (manuscrits, livres anciens, estampes, photographies).

### 2.2.2. Moyens humains

L'équipe se compose de 42 personnes titulaires, aidées par une cinquantaine de moniteurs étudiants employés à temps partiel, par quelques vacataires et chargés d'étude. L'intégration de la BAA au sein de l'INHA n'a pas été suivie de l'augmentation

en moyens humains espérée et, actuellement, le département de la bibliothèque doit fonctionner avec un effectif presque identique à celui de l'ancienne Bibliothèque d'art et d'archéologie. La direction de l'INHA a estimé que 101 personnes seraient nécessaires pour l'ouverture de la bibliothèque en Salle Labrouste. Cependant, même lorsque le personnel de l'actuelle BCMN aura rejoint l'Institut (comme le prévoit le projet), la bibliothèque ne comptera que 68 agents.

Le personnel est réparti entre cinq services : Développement des collections et politique documentaire, Catalogue, Patrimoine, Service aux publics et magasins, Informatique documentaire.

Le parti pris a été de ne pas fermer la bibliothèque et il faut aujourd'hui assurer la gestion quotidienne tout en menant les grands chantiers préparatoires pour l'ouverture en Salle Labrouste : intégration des autres bibliothèques, développement des collections imprimées et électroniques, reliure, recotation dans la classification de la Bibliothèque du Congrès, sélection pour le libre-accès, achèvement du catalogue commun, numérisation... L'enjeu principal est de continuer à assurer la qualité et le développement des services sur place et à distance tout en préparant l'implantation en Salle Labrouste.

### 2.2.3. Public

La bibliothèque est installée dans la Salle Ovale (anciennement Salle des périodiques de la BnF) qu'elle partage avec le Département de Recherches bibliographiques des collections spécialisées de la Bibliothèque Nationale. Elle accueille les étudiants en histoire de l'art et archéologie à partir de la maîtrise, les enseignants, les chercheurs et tous les professionnels et amateurs d'art. La fréquentation est en constante progression (plus de 10% par an) et les files d'attente pour obtenir une place de lecture sont quotidiennes. Le vieillissement des infrastructures du Quadrilatère pèse lourdement sur le fonctionnement : magasins aux conditions de conservation et de stockage insuffisantes, nécessité de délocaliser l'équivalent des accroissements annuels au Centre Technique du Livre de l'Enseignement Supérieur (CTLes) par manque de place, circuits de circulation vieillissants subissant de fréquentes pannes.

## **Partie 2 : La bibliothèque numérique de l'INHA**

Comme nous l'avons vu, la création d'une bibliothèque numérique offrant un accès à distance aux principaux fonds de l'INHA figurait, dès le début, parmi les principes fondateurs de la constitution de l'ensemble documentaire. Après avoir défini le concept, nous nous intéresserons plus précisément aux caractéristiques propres de la bibliothèque numérique de l'INHA.

### **1. Qu'est-ce qu'une bibliothèque numérique ?**

#### **1.1. Définition**

Le concept de bibliothèque numérique est défini par Charles Oppenheim et Daniel Smithson<sup>8</sup>, dans leur article «What is the hybrid library ?» paru dans le *Journal of Information science* en 1999, comme « un système d'information dans lequel toutes les ressources d'information sont disponibles sous une forme traitable par ordinateur et dans lequel toutes les fonctions d'acquisition, de stockage, de conservation, de recherche, d'accès et de visualisation utilisent les techniques numériques ». Le concept est utilisé par les spécialistes des sciences de l'information pour désigner tout système, qu'il soit ou non appuyé sur une bibliothèque traditionnelle, qui gère des ressources électroniques. Une ressource électronique est décrite, par Catherine Lupovici<sup>9</sup>, comme «un ensemble complexe de contenus d'informations et de packaging de ces informations dans un format de document accompagné d'un programme informatique ou application offrant des fonctionnalités de manipulation telles que la recherche, l'affichage, la navigation ». La bibliothèque numérique se distingue de la bibliothèque traditionnelle par le fait qu'elle traite uniquement des informations sous format numérique. Cette information peut avoir été créée initialement sous forme numérique, on parlera alors de document numérique natif. Mais elle peut aussi être le résultat d'une numérisation, à l'aide d'un scanner, d'un document existant au préalable sous format papier.

---

<sup>8</sup> Professeurs des Sciences de l'information, Loughborough University, Royaume-Uni

<sup>9</sup> Conservateur au Département de la bibliothèque numérique, Direction des services et des réseaux, de la Bibliothèque nationale de France jusqu'en juin 2007

En langage courant, les deux expressions « bibliothèque virtuelle » et « bibliothèque numérique » sont souvent employées indifféremment pour désigner une sorte de bibliothèque « immatérielle », fondée sur des technologies numériques, accessible au travers des réseaux, et consultable via un ordinateur. Cependant, ces dénominations ne recouvrent pas totalement les mêmes notions et il convient d'établir une brève typologie en rappelant rapidement ce que sont fondamentalement les bibliothèques « virtuelles », les bibliothèques « numériques » et les bibliothèques « hybrides ».

## 1.2. Typologie

### 1.2.1. Bibliothèque virtuelle

L'expression « bibliothèque virtuelle » semble trop souvent employée de façon abusive. En effet, comme Jean-Claude Le Moal<sup>10</sup> l'a précisé, en préambule de la Journée d'étude « Expériences et projets de bibliothèques virtuelles du Sud », organisée par l'ABDS<sup>11</sup> le 18 décembre 2001 : « le terme virtuel, qui suppose une non-existence physique, doit être réservé aux bibliothèques de liens qui pointent vers des sites divers à travers le monde ».

### 1.2.2. Bibliothèque numérique ou électronique

Toujours selon Jean-Claude Le Moal, une bibliothèque « numérique » (ou « bibliothèque électronique », les deux étant synonymes) « recouvre une réalité concrète et correspond à la mise en mémoire des documents pour en assurer une diffusion électronique. La bibliothèque numérique s'inscrit ainsi dans la continuité des bibliothèques traditionnelles, dont la fonction fondamentale est bien de rendre accessibles leurs ressources ».

Au sein des bibliothèques numériques, la bibliothèque numérique patrimoniale numérise des collections rares et/ou précieuses, en vue de les rendre accessibles à un large public. Cette politique de numérisation correspond la plupart du temps à une approche par fonds, dans une perspective d'accès à l'héritage culturel. C'est cette approche qu'a

---

<sup>10</sup> Chargé de mission à la Direction de l'information scientifique et de la communication (DISC) de l'Institut national de recherche en informatique et en automatique (INRIA)

<sup>11</sup> Association des professionnels de l'information et de la documentation

privilegié, notamment, la Bibliothèque nationale de France, pour la numérisation des fonds offerts sur le site Internet Gallica<sup>12</sup>.

### 1.2.3. Bibliothèque hybride

On parlera de «bibliothèque hybride » pour désigner une bibliothèque qui ajoute à ses ressources classiques, des ressources électroniques sur supports, ainsi que des ressources accessibles sur le Web, et qui gère le tout dans une même politique documentaire.

La bibliothèque hybride est donc une alternative aux bibliothèques traditionnelles (qui ne contiennent que des ressources papier) et aux bibliothèques numériques (qui ne contiennent que des ressources électroniques). Dans le processus de transformation des bibliothèques, on peut la voir soit comme l'étape nécessaire (vers un passage au « tout numérique »), soit comme l'étape finale (en considérant que le numérique ne pourra pas se substituer intégralement au papier).

Le propos de ce rapport se limitera à la notion de bibliothèque numérique patrimoniale puisque celle de l'INHA est effectivement construite à partir des fonds précieux ou remarquables de sa bibliothèque traditionnelle.

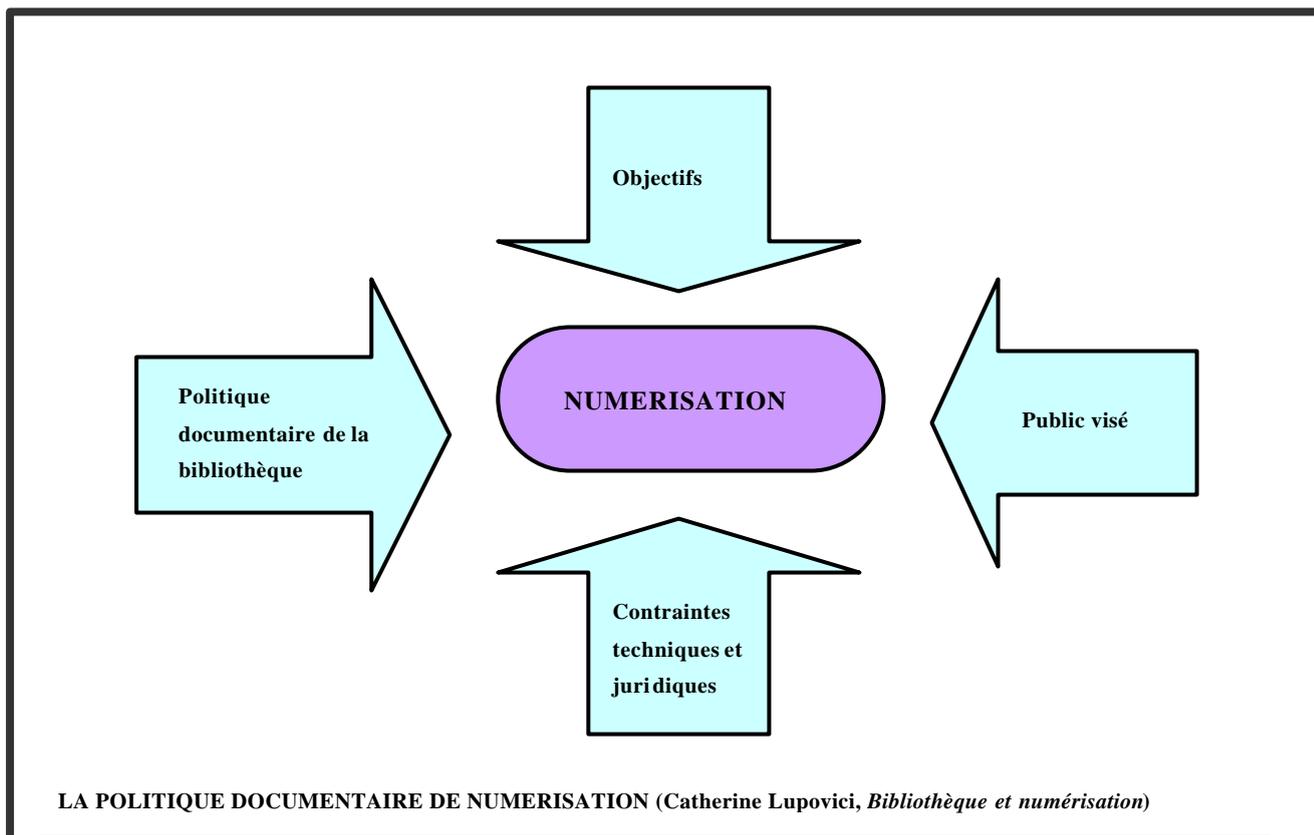
## 2. La politique documentaire de la bibliothèque numérique de l'INHA

Le préalable à tout projet de création de bibliothèque numérique doit être la définition d'une politique documentaire de numérisation qui s'intègre dans celle de l'établissement et pose les questions suivantes : que va-t-on numériser ? pourquoi ? pour qui ?

Ces interrogations permettent de déterminer les objectifs de la numérisation (intérêt scientifique, intérêt patrimonial et de valorisation, conservation préventive...) et les publics visés. Les contraintes techniques et juridiques ainsi que les impératifs financiers entrent également en compte dans la politique documentaire puisqu'ils influent sur la sélection, les possibilités de numérisation et les modes de diffusion.

---

<sup>12</sup> <http://gallica.bnf.fr/>



## 2.1. Objectifs et publics visés

La Bibliothèque de l'Institut national d'histoire de l'art constitue progressivement une bibliothèque numérique en histoire de l'art en sélectionnant, sur des critères scientifiques, parmi les documents qu'elle conserve, des ensembles cohérents et complets. L'objectif premier est de proposer, sur Internet, une bibliothèque de référence idéale pour un chercheur en histoire de l'art. De plus, ce programme s'inscrit dans la problématique de conservation puisque la numérisation peut servir de substitution de communication pour les documents anciens, fragiles, ou précieux qui s'avèrent être très consultés, tant par le public que par le personnel de l'INHA.

Le public visé est essentiellement un public de chercheurs (français et étrangers) menant des travaux en histoire de l'art et archéologie. Mais, l'INHA ambitionne également de toucher un plus vaste public : le public des amateurs intéressés, le grand public, et le public des scolaires (par exemple pour des besoins de textes, de présentations ou d'images pour un exposé).

## 2.2. Organisation de la collection

Les fonds concernés par la numérisation proviennent des trois établissements qui formeront à terme la future Bibliothèque de l'INHA (INHA - collection Jacques Doucet, BCMN et BENSBA). De 2002 à 2006, près de 240 000 vues ont été numérisées<sup>13</sup>.

Actuellement, la bibliothèque numérique est constituée pour sa plus grande part d'un fonds d'imprimés dénommé les « Classiques de l'histoire de l'art » qui sont des textes du XVI<sup>ème</sup> au XIX<sup>ème</sup> siècle, en français ou publiés en France en d'autres langues. Les choix reposent sur une étude des bibliographies, l'avis des spécialistes et les statistiques de communication des ouvrages. Les principales thématiques<sup>14</sup> sont l'architecture, les livres de fêtes, les guides de Paris anciens (fondamentaux pour la description des anciennes collections, les monuments détruits ou transformés, les tableaux déplacés etc.), l'archéologie, les arts figuratifs, les collections, l'art des jardins et l'esthétique.

A ceci s'ajoute la collection des catalogues du Musée du Louvre antérieurs à 1920 de la Bibliothèque centrale des musées nationaux qui comprend les catalogues généraux et ceux des départements spécialisés.

Dans le domaine iconographique, la bibliothèque propose des photographies, des plaques stéréoscopiques, des gravures, des plans architecturaux, des estampes provenant de la collection des fonds patrimoniaux Jacques Doucet (Goya, Manet, Toulouse-Lautrec, Pillement, ...) et des recueils de dessins (notamment ceux de la galerie du Palais royal datant du XVIII<sup>ème</sup> siècle c'est-à-dire antérieurs à la dispersion de cette fameuse collection de peintures).

Enfin, la Bibliothèque présente des documents manuscrits et des fonds d'archives, des papiers, des journaux et des lettres d'artistes. Un important fonds Delacroix a été constitué en rassemblant des volumes du Journal de Delacroix, des manuscrits et des correspondances issus à la fois des collections Jacques Doucet et des collections de la Bibliothèque centrale des musées nationaux.

Pour l'accroissement de cette collection numérique, la prochaine campagne de numérisation 2007-2010 portera sur :

- un ensemble de livres imprimés du XIX<sup>e</sup> siècle et du XX<sup>e</sup> siècle, dont un ensemble de catalogues de musées et d'expositions,

---

<sup>13</sup> Annexe 2 : Nombre de vues numérisées

<sup>14</sup> Annexe 3 : Liste des rubriques et fonds de la bibliothèque numérique

- un ensemble de livres imprimés du XVIe au début du XXe siècle et un ensemble de photographies, en recueils et à la pièce,
- un ensemble de registres et un fichier décrivant 300 000 photographies entrées par don en mars 2007.

### **3. La chaîne de numérisation**

#### **3.1. Choix techniques et contraintes juridiques**

##### **3.1.1. Numérisation en mode image**

Le but principal poursuivi par l'INHA en créant sa collection numérique est de préserver les originaux tout en facilitant leur accès au public. En effet, la numérisation rend possible une diffusion très large avec de bonnes conditions de consultation sans mettre en péril les originaux, elle semble ainsi pouvoir apporter une réponse à l'éternel dilemme des bibliothèques entre conservation et communication. Cependant, il faut prendre en compte un certain nombre de contraintes techniques et juridiques. En effet, l'accès au plus grand nombre n'est possible que dans le cas où la bibliothèque dispose du droit de donner cet accès ou si le document est libre de droits<sup>15</sup>. D'autre part, les chercheurs ont généralement besoin d'une numérisation très fidèle à l'original, ce qui suppose une excellente qualité d'image. Mettre en ligne de telles images pose des problèmes techniques (images très lourdes donc temps de chargement très longs) et des problèmes juridiques (risque de piratage).

Afin de préserver l'intégrité de l'original et de pouvoir proposer aux usagers un véritable fac-similé, le choix a été celui d'une numérisation en mode image. Cependant, ce mode de numérisation ne permet pas la manipulation du texte et offre comme seul point d'accès la notice bibliographique du catalogue. Pour pallier cet inconvénient, une solution intermédiaire a été adoptée en combinant une numérisation en mode image avec des instruments d'accès pour permettre de restituer le feuilletage des documents. Ces données sont regroupées en 3 types de signets :

- signets « pages » : il s'agit d'un relevé strict de la pagination utile lorsque l'on cherche à consulter une page précise (ils sont nécessaires car les numéros des pages

---

<sup>15</sup> Annexe 4 : Rappel sur le droit d'auteur appliqué à la numérisation

proposés par le navigateur pdf (Acrobat Reader) ne correspondent qu'à des numéros séquentiels des images et non aux numéros réels des pages de l'original) ;

- signets «chapitres, tables et index» : il s'agit d'un relevé de la structure de l'ouvrage qui fait apparaître les principales parties, sous-parties, tables et index ;

- signets «planches» : il s'agit d'un relevé des planches et des illustrations.

Par ailleurs, une notice est associée à chaque objet numérique. Elle indique les caractéristiques de l'original, entre autres : la localisation (bibliothèque d'origine), la cote, le type de document, le titre, l'auteur, la provenance...

### 3.1.2. La résolution et le format

Les décisions prises au moment de la numérisation d'une collection sont déterminantes pour sa gestion future, son stockage et sa conservation. Lors d'une numérisation en mode image, deux facteurs entrent en considération : la résolution (ou paramètres de qualité) et le format.

#### 3.1.2.1. *La résolution*

Le processus de numérisation génère des images «matricielles» ou «pixelisées» c'est-à-dire qu'elles prennent la forme d'une grille (ou matrice) où chaque élément de l'image (ou pixel) a un emplacement unique et une valeur de couleurs indépendante. Il s'agit de deux paramètres fondamentaux pour la qualité des images.

- *La résolution spatiale*

Elle correspond à la fréquence à laquelle des échantillons de l'original sont capturés par le dispositif de numérisation. Elle s'exprime sous la forme d'un nombre de points par pouce (dpi<sup>16</sup>) (pour mémoire : la résolution d'un écran est de 72 à 96 dpi, celle d'une impression professionnelle de 300 dpi, celle d'une impression bureautique de 150 à 200 dpi). Plus il y a de points, plus on a de détails (c'est-à-dire d'informations par cm) et plus les fichiers sont volumineux.

La résolution adoptée par l'INHA diffère selon le type de documents :

- 8 points/mm (200 dpi) pour les catalogues du Louvre,
- 12 points/mm (300 dpi) pour les pages de textes,
- 16 points/mm (400 dpi) pour les pages d'illustrations,
- 24 points/mm (600 dpi) pour les photographies,
- 96 points/mm (2 400 dpi) pour les microformes ou les diapositives.

---

<sup>16</sup> Dot per inch

- *La résolution des couleurs (profondeur de bits)*

Le nombre de couleurs (ou niveaux de luminosité) disponibles pour représenter différentes couleurs (ou tons de gris) dans l'original est exprimé en nombre de bits. Une résolution de couleurs de 8 bits signifie que 256 couleurs différentes sont disponibles et une résolution de couleurs de 24 bits représente 16 777 216 couleurs codables (l'œil humain distingue jusqu'à 350 000 couleurs).

Selon les caractéristiques des documents originaux, la numérisation pour l'INHA s'effectue, en niveaux de gris (échantillonnés sur 8 bits par pixel) ou en couleurs (RVB échantillonnés sur 24 bits par pixel).

### 3.1.2.2. *Le format*

- *Format des fichiers sources*

Les fichiers sources sont les images en haute définition générées par le processus de numérisation qui servent au stockage. Elles sont non comprimées et n'ont subi aucun retraitement. Selon les documents et les exigences de chaque institution possédant les originaux, ces images sont créées au format TIFF (Tagged Image File Format), PNG (Portable Network Graphics) ou JPEG (Joint Photographic Expert Group).

- *Format des fichiers dérivés*

Pour les rendre plus maniables en vue de leur mise en ligne, les fichiers sources sont dégradés en fichiers dérivés. Les images sont allégées grâce à une opération de compression qui rend les fichiers moins dévoreurs d'espaces de stockage et abaisse, par conséquent, les temps de téléchargement. Toutefois, la compression utilisée autorise un zoom à l'affichage jusqu'à 200% sans dégradation, ce qui permet de bien voir les textes si nécessaire.

Pour les images isolées (estampes, photographies, ...) le format utilisé est le JPEG (qualité 7 sur l'échelle de Photoshop). Pour les images en recueils (livres, manuscrits, recueils d'ouvrages, ...) le format utilisé est le PDF (Portable Document Format). Les fichiers PDF sont réalisés à l'aide d'Adobe Acrobat, logiciel de publication et de visualisation qui, dans ses nouvelles versions, propose un accès plus rapide puisque le téléchargement se fait en arrière fond c'est-à-dire que l'on peut consulter les premières pages sans attendre le chargement de l'ensemble du fichier.

### 3.1.3. Une prestation extérieure

La bibliothèque a fait le choix d'avoir recours à des prestataires extérieurs car une numérisation en interne aurait représenté un investissement de départ trop lourd à supporter. La particularité et la diversité des fonds à traiter (documents fragiles appelant une numérisation de qualité en couleurs, estampes, photographies, plaques de verre...) aurait nécessité l'achat de matériels spécifiques et coûteux. En outre, de par son effectif réduit, la bibliothèque ne pouvait pas former puis mobiliser du personnel pour assurer la numérisation.

Cependant, le recours à une prestation externalisée est plus contraignant en termes de préparation, de suivi et de contrôle. Chaque campagne de numérisation donne lieu à un appel d'offres avec rédaction d'un CCTP (cahier des clauses techniques particulières) et d'un CCAP (cahier des clauses administratives particulières) qui fixent les dispositions techniques et administratives propres à l'exécution de chaque marché.

## 3.2. Les étapes de la chaîne

### 3.2.1. Sélection et préparation

La première phase consiste en la sélection des documents à numériser. Cette sélection est réalisée en fonction de la politique documentaire mentionnée plus haut. Les conservateurs du Service du patrimoine de la bibliothèque ainsi que les chercheurs du Département des études et de la recherche sont sollicités pour établir des listes de documents qu'il leur semble pertinent de numériser. Cette liste est, ensuite, soumise aux équipes de Gallica de la BnF qui signalent à l'INHA les documents qu'ils auraient éventuellement déjà inscrits dans leurs prochains programmes de numérisation. Cette vérification a pour but d'éviter les pertes d'argent et de temps que représenteraient les doublons de numérisation. La liste est, enfin, validée après un contrôle en magasins portant sur la présence et l'état physique des pièces à numériser.

Chaque document fait l'objet d'une fiche de travail<sup>17</sup> et d'un constat d'état<sup>18</sup> destinés à faciliter le travail de l'opérateur qui va le numériser et à lui permettre de faire connaître les éventuels problèmes auxquels il a été confronté.

---

<sup>17</sup> Annexe 5 : Une fiche de travail

<sup>18</sup> Annexe 6 : Un constat d'état

### 3.2.2. Numérisation et contrôle de la livraison

La numérisation peut se dérouler selon deux modalités. S'il s'agit de documents particulièrement fragiles, précieux ou difficiles à numériser, la numérisation est réalisée dans un atelier installé sur place à la bibliothèque par le prestataire désigné<sup>19</sup>. S'il s'agit de documents ne nécessitant pas de précautions spécifiques et pouvant être aisément expédiés et transportés, la numérisation est traitée dans les ateliers propres du prestataire.

Les fichiers numérisés sont alors remis à la bibliothèque qui effectue un premier contrôle et signale les éventuelles reprises (oubli de pages, ordre des pages, cadrage, orientation de l'image, validation des éventuels montages, netteté, distorsions géométriques...). L'intégrité et le nommage des fichiers sont aussi vérifiés à partir du bordereau de retour.

### 3.2.3. Traitement pour la mise en ligne

Les fichiers issus de la production sont livrés sur cédéroms ou sur disque dur par le prestataire. Ils sont chargés sur un serveur de stockage de la bibliothèque. Les images sont ensuite compressées avant d'être mises en ligne pour limiter les espaces de stockage nécessaires et pour offrir aux usagers une rapidité de téléchargement satisfaisante ; le poids maximal des fichiers est limité à 2 Go (gigaoctets)<sup>20</sup>. Pour les ouvrages ou les pièces nécessitant un regroupement (recto et verso d'une lettre, vues de détails d'une planche grand format...) les fichiers PDF (Portable Document Format) sont fabriqués et les signets rédigés (pagination, table des matières, table des illustrations)<sup>21</sup>. Chaque objet numérique est alors protégé par copyright apposé au bas de chaque image et fait l'objet d'une notice bibliographique.

Dans l'attente du futur système de banque de données multimédias, l'INHA a choisi de faire appel à un prestataire, la Société Atchik Image, qui propose le service Picturelan pour l'hébergement et la fourniture d'accès aux documents numériques. Picturelan permet de stocker les fichiers, propose une interface pour la diffusion sur le Web et un module d'administration<sup>22</sup> à distance pour le chargement des données et leur organisation. La mise en ligne se fait par téléchargement classique http pour les téléchargements à l'unité ou chargement ftp pour les téléchargements par lots.

---

<sup>19</sup> Annexe 7 : L'atelier de numérisation en interne

<sup>20</sup> Un Gigaoctet correspond à 1024 Mégaoctets soit 1 073 741 824 Octets

<sup>21</sup> Annexe 8 : Un pdf avec signets

Actuellement, environ 130 000 vues ont été chargées sur Picturelan<sup>23</sup>. Bien évidemment, ce chiffre évolue constamment au rythme des nouveaux chargements de documents (un à deux par mois).

La liste des documents en ligne est, ensuite, transmise au Service du Catalogue qui peut procéder au catalogage dans le SUDOC.

---

<sup>22</sup> Annexe 9 : Le module d'administration de Picturelan

<sup>23</sup> Annexe 10 : Nombre de vues chargées

## **Partie 3 : La gestion des collections numérisées au sein des bibliothèques**

Au cours des dernières années, les documents numériques ont pris une place croissante au sein des bibliothèques. Les établissements qui acquièrent ou produisent de l'information sous forme numérique sont confrontés à une série de questions concernant le cycle de vie de ces nouvelles ressources et plus particulièrement leur gestion et leur préservation. Conscient de la fragilité de ce nouveau patrimoine, l'Unesco<sup>24</sup>, dans sa publication *La mémoire de la société de l'information*, parue en 2003 pour le Sommet mondial sur la société de l'information, présente ainsi le problème de la conservation du numérique : « tant que l'information a emprunté des supports physiques pour circuler, elle faisait trace. Même sans s'en préoccuper il reste toujours quelque chose de la trace, quelque chose qui se transforme en archive. Mais l'informatique a ce défaut congénital : si l'on ne sauvegarde pas on efface ».

### **1. Contexte et problématique**

#### **1.1. Les enjeux**

Partant du constat que la plupart des fichiers informatiques créés il y a plus de vingt ans sont aujourd'hui illisibles (format de fichier inconnu, support physique détérioré, logiciel ou matériel de lecture disparu...), on ne peut que prendre conscience des enjeux liés à la gestion et à la préservation des collections numériques. Il s'agit de :

- conserver le document et l'information qu'il contient dans son aspect physique comme dans son aspect intellectuel,
- sur le très long terme (30 ans et au-delà),
- de manière à pouvoir le rendre accessible et compréhensible.

La présence des ressources numériques introduit, donc, une nouvelle dimension dans la gestion informatisée des bibliothèques puisqu'il faut assurer la gestion des ressources

---

<sup>24</sup> United Nations Educational, Scientific and Cultural Organization

numériques elles-mêmes en conjonction avec leur signalement dans le catalogue informatisé.

## **1.2. Problèmes posés par la préservation de l'information numérique**

Un document numérique enregistré il y a dix ans ou parfois moins peut déjà être dans une situation vulnérable au regard de sa préservation. Cette vulnérabilité peut provenir de causes multiples.

### **1.2.1. Facteurs techniques**

Tout d'abord, un certain nombre de facteurs techniques entrent en jeu puisque les documents numériques sont soumis à l'obsolescence des logiciels, des systèmes et des technologies de stockage. La conservation du support ne suffit pas à assurer la lecture voire la compréhension. La pérennisation et l'intelligibilité des contenus numériques est, donc, devenu un véritable défi face à l'accélération des évolutions de la technologie (à titre d'exemple : plus de six versions de Word sous Windows se sont succédées de 1995 à nos jours). De plus, la compatibilité ascendante entre les différentes versions ne peut être considérée comme une solution viable à long terme car elle reste souvent limitée.

Pour que les objets numériques restent utilisables à moyen et à très long termes, il faut non seulement que leur intégrité initiale soit conservée mais également qu'ils restent compréhensibles pour les utilisateurs une fois que les programmes et les plates-formes techniques de consultation contemporaines de leur création auront disparu. Ce sont ces plates-formes, en effet, qui permettent la transformation d'un train de bits en un contenu signifiant. Les documents numériques sont, donc, caractérisés par une grande dépendance entre données créées et environnements de création. La conservation des connaissances relatives au contexte de leur création est primordiale pour l'intelligibilité des contenus, car pour les données numériques les programmes de codage et d'interprétation du code ont une durée de vie inférieure à dix ans.

Cependant, cette vulnérabilité n'est pas uniquement technique mais résulte, également, de l'insuffisance de l'organisation généralement en place dans les organismes et de l'incapacité à penser l'information et le document numérique en terme de flux et de cycle de vie depuis leur intégration dans un système d'information jusqu'à leur archivage.

### 1.2.2. Facteurs organisationnels et administratifs

Les facteurs organisationnels et financiers représentent, effectivement, un autre aspect du problème qu'il ne faut pas négliger. L'engagement institutionnel concernant la préservation à long terme est encore, bien souvent, insuffisant. L'absence de politiques et de procédures de préservation ainsi que la raréfaction des ressources humaines et financières mettent en danger ce nouveau patrimoine numérique. Il s'avère parfois difficile d'inscrire la pérennisation de l'information électronique dans les priorités d'un établissement. Cependant, elle constitue une activité en soi qui nécessite un budget, la mise en place d'une organisation du travail ainsi qu'un partage des responsabilités.

S'ajoute, encore, l'insuffisance du réseau de normes, une législation et une réglementation peu appropriées à ce nouveau contexte ainsi qu'une évolution constante de la nature des lois régissant les droits d'auteur et l'usage s'appliquant aux collections numériques.

Fort heureusement, dès le début des années 90, des regroupements d'institutions de recherche scientifique, d'archives ou de bibliothèques, ont commencé à explorer les manières de prendre en compte les spécificités des documents numériques dans les politiques d'archivage et de conservation.

### 1.3. Travaux et réflexion en cours

Les initiatives sont nombreuses mais nous retiendrons simplement quelques programmes désignés comme pilote en raison de leur positionnement institutionnel et scientifique, en particulier parce qu'ils mobilisent des expertises internationales et que la diffusion large de leurs travaux permet une mise en commun de leurs résultats.

- **Pandora (Preserving and Accessing Networked Documentary Resources of Australia)**<sup>25</sup>

Il s'agit d'un des premiers programmes d'envergure soutenus par les pouvoirs publics australiens qui a pour objectif la mise en œuvre effective d'un système d'archivage des publications électroniques et des sites australiens du Web. A l'issue d'une première étape de travaux menés dès 1996 par la Bibliothèque nationale d'Australie, un ensemble de recommandations ont donné lieu à une publication qui définit des principes et méthodes de sélection ainsi que les modes d'organisation et de gestion de l'archive. Une

---

<sup>25</sup> <http://pandora.nla.gov.au/index.html>

seconde phase de travaux « PADI » (Preserving Access to Digital Information) a permis de concevoir et de mettre en œuvre un modèle national d'archivage partagé des données électroniques et du Web australien.

- **NEDLIB (Networked European Deposit Library)<sup>26</sup>**

Ce programme, soutenu par la Commission Européenne de 1998 à 2001, regroupe huit grandes bibliothèques européennes : la Bibliothèque Royale des Pays-Bas, la Bibliothèque nationale de Norvège, la Bibliothèque nationale d'Allemagne, la Bibliothèque nationale du Portugal, la Bibliothèque nationale de Suisse, la Bibliothèque de Florence, la Bibliothèque universitaire d'Helsinki et la Bibliothèque nationale de France. Des sociétés d'édition telles qu'Elsevier Science, Kluwer Academic et Springer Verlag sont également impliquées. Le programme vise la constitution d'un modèle de spécifications fonctionnelles et techniques relatif aux documents électroniques publiés sur support ou diffusés sur le Web. Il prend en considération l'ensemble des fonctions, depuis la collecte jusqu'à la conservation incluant la mise à plat des techniques de migration et d'émulation. Parmi les questions clefs qu'il aborde, figurent les procédures de maintenance des archives et les liens entre les besoins de métadonnées et les stratégies de conservation.

- **CEDARS (CURL Exemplars in Digital Archives)<sup>27</sup>**

Ce programme est développé par un consortium de bibliothèques de recherche universitaires britanniques (Oxford, Leeds et Cambridge) et américaines (Université du Michigan). Démarré en avril 1998, il se propose d'explorer les questions relatives à l'acquisition des publications électroniques, à leur conservation, à leur description et enfin à leur accès. Son objectif est de mettre à la disposition de la communauté des bibliothèques un guide décrivant, tant d'un point de vue méthodologique qu'opérationnel, les bonnes pratiques permettant d'intégrer dans les collections préexistantes des corpus de textes numériques. Depuis 1999, le programme a mis en place un nouveau volet appelé CAMILEON<sup>28</sup> (Creative Archiving at Michigan and Leeds Emulating the Old on the New) visant à approfondir les problématiques des techniques de l'émulation (cf. §2.3 définition des stratégies de conservation). Il explore les conditions de recours à l'émulation comme stratégie de conservation capable de

<sup>26</sup> <http://www.kb.nl.coop/nedlib>

<sup>27</sup> <http://www.leeds.ac.uk/cedars>

remédier à l'obsolescence des systèmes de production de données numériques et de maintenir les possibilités d'accès à ces contenus dans leur configuration de création initiale. Il a pour objectif le développement et l'expérimentation d'outils d'émulation ainsi que l'élaboration d'un ensemble de recommandations concernant tant la migration que l'émulation. Il vise, également, à créer un cadre stratégique concernant les politiques de gestion des collections numériques, et à promouvoir les méthodes appropriées à la préservation à long terme de différents types de ressources numériques (y compris la création des métadonnées appropriées).

- **INTERPARES (International Research on Permanent Authentic Records in Electronic Systems)<sup>29</sup>**

Il s'agit de l'un des plus importants programmes de recherche et de développement consacré à la conservation à long terme des données numériques puisque plus d'une vingtaine de pays y participent (par le concours de chercheurs, d'institutions d'archives, de producteurs de contenus et d'industriels). Ce programme tente de proposer des actions concourant à la conservation des données numériques par la mise en place de standards, de principes d'organisation et de recommandations à l'attention des gouvernements, des acteurs industriels et des institutions culturelles et patrimoniales.

- **PREMIS (Preservation Metadata for Digital Materials)<sup>30</sup>**

Depuis 2000, les deux associations professionnelles de bibliothèques, d'archives et de musées OCLC (Online Computer Library Center) et RLG (Research Libraries Group) ont axé une partie de leurs recherches sur la préservation et l'accès à long terme aux ressources numériques. Ces travaux ont abouti à la rédaction de divers documents de travail (*Preservation metadata for longterm retention, Attributes of digital archive for research repositories*) ainsi qu'à la mise en place du programme PREMIS (Preservation Metadata for Digital Materials) centré sur les métadonnées de préservation.

- **Groupe PIN (Pérennisation des Informations Numériques)<sup>31</sup>**

En France, depuis 2000, le groupe PIN de l'association Aristote, constitue « un lieu d'échange de savoirs, de savoir faire et d'actions dans le domaine de la pérennisation des informations représentées sous forme numérique ». Son activité est dédiée à l'ensemble des problèmes techniques, organisationnels, financiers, normatifs, méthodologiques et

<sup>28</sup> <http://www.si.umich.edu/CAMILEON/>

<sup>29</sup> <http://www.interpares.org/>

<sup>30</sup> <http://www.oclc.org/research/projects/pmwg/>

juridiques posés par la préservation à long terme de l'information sous forme numérique. Ce groupe expert rassemble de nombreux organismes : le CNES, la BnF, les Archives de France, le CEA, ... et se caractérise, par conséquent, par une grande diversité de compétences et d'expériences professionnelles. Le groupe PIN dispose d'un forum de discussion par messagerie et tient des réunions trimestrielles à Paris. Les thèmes traités ont vocation à couvrir l'ensemble des facettes de la problématique. Le programme d'actions vise à l'élaboration de manuels pratiques et à l'évaluation des formats de données pour la pérennisation ; il comprend, également, des actions de formation.

Enfin, il convient de citer les recommandations pour la conservation pérenne de l'information numérique émises par l'Unesco (*Charte de l'Unesco sur la conservation du patrimoine numérique*<sup>32</sup>) et celles publiées conjointement par la DAF et la DGME<sup>33</sup> (*Standard d'échange de données pour l'archivage*<sup>34</sup>) qui ont pour but de faciliter l'interopérabilité entre deux systèmes d'information (service d'archives, producteurs, utilisateurs...).

L'étude de ces divers travaux et réflexions permet de dégager des stratégies de gestion et de conservation.

## 2. Stratégies de gestion

Tout d'abord, il faut rappeler que la gestion et la préservation des documents numériques nécessitent l'appropriation par les bibliothécaires de nouveaux domaines de compétences qui portent sur :

- les supports et la préservation des fichiers,
- les formats de représentation de l'information,
- les techniques de mise à disposition de l'information numérique.

Nous examinerons les caractéristiques fonctionnelles d'un système de gestion d'objets numériques puis le modèle OAIS qui apparaît aujourd'hui comme une référence incontestable internationale et multidisciplinaire pour l'archivage pérenne.

<sup>31</sup> <http://vds.cnes.fr/pin/>

<sup>32</sup> [http://portal.unesco.org/ci/fr/ev.php-URL\\_ID=13366&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/fr/ev.php-URL_ID=13366&URL_DO=DO_TOPIC&URL_SECTION=201.html)

<sup>33</sup> DAF : Direction des Archives de France, et DGME : Direction Générale de la Modernisation de l'Etat

<sup>34</sup> [https://www.ateliers.modernisation.gouv.fr/ministeres/projets\\_adele/a103\\_archivage\\_elect/public/standard\\_d\\_echange\\_d\\_folder\\_contents](https://www.ateliers.modernisation.gouv.fr/ministeres/projets_adele/a103_archivage_elect/public/standard_d_echange_d_folder_contents)

## 2.1. Un système de gestion de documents numériques

### 2.1.1. Définition

Selon la définition donnée par le CREPUQ<sup>35</sup> (dans *La gestion des documents numériques des établissements universitaires du Québec*), la gestion des documents numériques recouvre l'« ensemble des procédures, opérations et techniques visant à contrôler, tout au long du cycle de vie des documents dans l'organisation, la création, l'utilisation, et à en assurer l'authenticité et l'intégrité ».

Un système de gestion des collections numériques est une sorte de magasin numérique qui permet de stocker, conserver et communiquer ces nouvelles ressources. Il consiste, au premier abord, en la gestion informatique des fonctions de magasinage et de communication. Mais, plus précisément, un système de gestion de ressources numériques doit permettre la gestion du processus de création des documents (gestion du circuit de production), la gestion du contenu du document lui-même (gérer les différentes instances de chaque document) et enfin la gestion de la signification de son contenu (gestion sémantique).

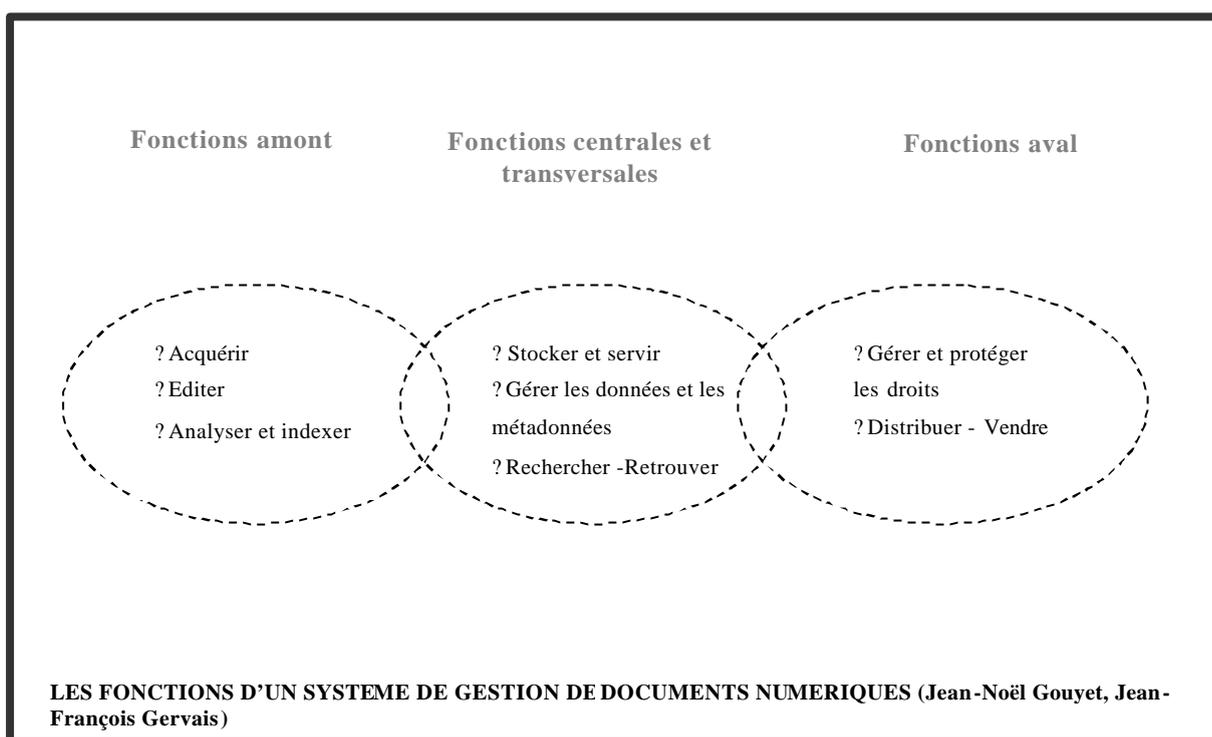
### 2.1.2. Principales fonctions

Les principales fonctions d'un système de gestion suivent le cycle de vie du document numérique. Elles peuvent être regroupées en trois grands domaines :

- les fonctions amont qui gèrent la production et l'acquisition du document et de ses métadonnées,
- les fonctions centrales et transversales qui stockent les objets numériques et les métadonnées, gèrent leurs différentes instances et permettent de retrouver les données,
- les fonctions aval qui assurent la gestion des droits liés à chaque donnée et s'occupent des éventuels services associés (par exemple la vente d'images haute-définition pour les éditeurs ou les commandes de reprints).

---

<sup>35</sup> La Conférence des recteurs et des principaux des universités du Québec (CREPUQ) est un organisme privé qui regroupe, sur une base volontaire, tous les établissements universitaires québécois



## 2.2. Un modèle fonctionnel : la norme OAIS

La plupart des efforts de préservation numérique étudiés prônent un déroulement des activités reposant sur l'OAIS<sup>36</sup> (Reference Model for an Open Archival Information System) comme exemple-repère. Le modèle de référence OAIS est né d'un travail mené par le CCSDS<sup>37</sup> (Consultative Committee for Space Data Systems). Il spécifie de manière très générale l'architecture logique et les fonctionnalités d'un système d'archivage. OAIS est un modèle abstrait définissant une terminologie et des concepts. Il identifie les acteurs, décrit les fonctions et les flux d'information.

### 2.2.1. Les acteurs de l'archivage

Le modèle OAIS définit 4 types d'acteurs :

- l'« archive » représente l'opérateur du système d'archivage,
- le « management » est le décideur politique qui établit un plan stratégique global et soutient le dispositif, politiquement, financièrement, et sur le très long terme,

<sup>36</sup> Traduction en français : [http://vds.cnes.fr/pin/documents/projet\\_norme\\_oais\\_version\\_francaise.pdf](http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf)

<sup>37</sup> Organisme international de normalisation des agences spatiales

- les « producteurs » sont les organismes qui fournissent les objets à archiver,
- les « utilisateurs » sont les organismes ou personnes qui ont accès aux objets archivés.

Les objets numériques connaissent trois phases : de SIP (submission information package) avant leur archivage, ils deviennent AIP (archival information package) une fois archivés pour, enfin, se transformer en DIP (dissemination information package) lorsqu'ils sont mis à la disposition des utilisateurs.

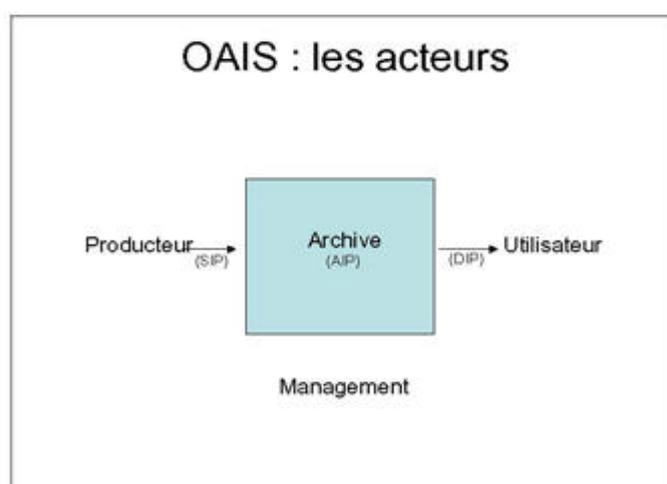


Figure 1 - Les acteurs selon le modèle OAIS

### 2.2.2. Les fonctions de l'archivage

OAIS présente un système d'archivage selon six grands domaines fonctionnels.

- L'entité « entrées » reçoit, contrôle et valide les objets à archiver. Puis elle les transmet à l'entité « stockage » tout en fournissant les informations nécessaires à leur description et à leur gestion dans le temps à l'entité « gestion des données ».

- L'entité « stockage » assure la conservation physique des objets archivés et les tient à la disposition de l'entité « accès », conformément aux règles établies par l'entité « administration ». Elle prend, également, en charge la réalisation des copies multiples et le renouvellement des supports anciens.

- L'entité « gestion des données » prend en charge la tenue à jour de toutes les informations internes nécessaires au système d'archivage. Elle fournit les informations descriptives, techniques et archivistiques des objets archivés nécessaires aux autres entités.

- L'entité « administration » remplit la fonction de coordination générale du système en établissant les règles internes, en veillant à la qualité globale du service rendu et à son amélioration, et en rendant compte au management.

- L'entité « planification de la pérennisation » veille sur l'environnement extérieur pour préparer et planifier les évolutions nécessaires, notamment technologiques. Mais elle doit aussi surveiller et anticiper les changements qui peuvent s'opérer dans la « communauté d'utilisateurs » en vue de garantir que le service d'accès reste conforme aux attentes nouvelles des utilisateurs.

- L'entité « accès » permet aux utilisateurs de rechercher dans le « catalogue » des objets archivés. Elle contrôle l'accès aux documents et les transmet aux utilisateurs.

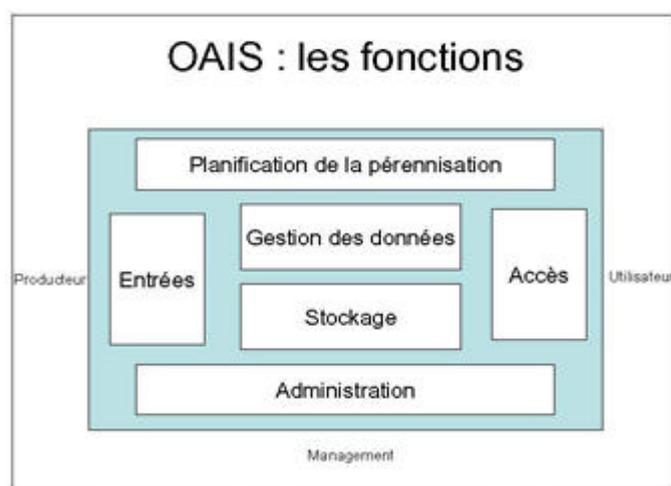


Figure 2 - Les 6 fonctions du modèle OAIS

### 2.3. Stratégies de conservation

L'archivage des données numériques ne peut se limiter à une sauvegarde ou à un backup système mais doit faire l'objet de la mise en place d'une véritable stratégie. Diverses méthodes de conservation d'objets numériques s'offrent actuellement à nous.

- **Copie de la chaîne de bits**

La copie de la chaîne de bits, ou appelée plus souvent « copie de sauvegarde », produit un double exact d'un objet numérique. Cette copie de sauvegarde doit être stockée en un lieu éloigné, afin que l'original et la copie ne soient pas vulnérables à un même sinistre.

- **Régénération**

Elle consiste en la copie de données numériques d'un support de stockage à long terme vers un autre support du même type, sans modification aucune de la chaîne de bits (par exemple d'un ancien disque compact réinscriptible vers un disque compact réinscriptible neuf). La « régénération modifiée » est la copie sur un autre support différent mais suffisamment semblable pour que la chaîne de bits n'ait à subir aucune modification concernant l'application et le système d'exploitation qui utilisent les données (par exemple d'un disque Zip de 100 Mo vers un disque Zip de 750 Mo).

- **Supports durables ou persistants**

Les supports durables (par exemple les CD à couche réfléchissante en or) peuvent réduire le besoin de régénération. Associés à une manipulation soignée, à un contrôle de la température et de l'humidité, et à un entreposage approprié, ils peuvent diminuer les pertes consécutives à la détérioration du support.

- **Conservation technologique**

Il s'agit de préserver l'environnement technique de fonctionnement du système, y compris le système d'exploitation, les logiciels d'application originaux, les lecteurs de support, etc. La conservation technologique correspond davantage à une stratégie de récupération en cas de désastre, à utiliser pour des objets numériques qui n'ont pas bénéficié d'une stratégie appropriée de conservation. Elle permet de faire face à l'obsolescence des supports, en supposant que ceux-ci ne se sont pas dégradés au point d'être illisibles.

- **Archéologie numérique**

Son objectif est de récupérer des chaînes de bits lisibles à partir d'un support lourdement abimé ou d'un environnement matériel et logiciel endommagé ou périmé. Stratégie de récupération d'urgence, elle fait intervenir des techniques et méthodes spécialisées de récupération de chaînes de bits à partir de supports devenus illisibles, en raison de dommages physiques ou d'une panne de matériel.

- **Sauvegarde analogique**

Cette stratégie repose sur la conversion d'objets numériques en une forme analogique. L'expression la plus simple de ce principe est l'impression sur papier mais elle devient rapidement impraticable à large échelle. La sauvegarde peut se faire sur supports analogiques durables (par exemple des microfilms aux halogénures d'argent). Elle permet de conserver le contenu et de le protéger de l'obsolescence.

- **L'émulation**

Ce procédé consiste à utiliser un programme qui simule l'environnement informatique nécessaire à l'exploitation des données sur une plate-forme courante alors que leur environnement technique est obsolète. L'émulateur se charge d'imiter la technologie disparue afin de procéder à la lecture d'un fichier dans son format original.

- **La migration**

La migration est l'opération au cours de laquelle les bits d'un fichier ou d'un programme sont modifiés afin de pouvoir être lus par un nouveau système d'exploitation ou une nouvelle version d'une application. Il s'agit d'un retraitement des données anciennes visant à les rendre compatibles avec un nouvel environnement. La migration peut conduire à changer le format ou le codage des données. Données et informations sont, alors, séparées de tout outil logiciel ayant servi à leur constitution et sont converties dans un langage différent.

## **2.4. Les métadonnées**

En matière de gestion et de conservation de documents numériques, le rôle des métadonnées est essentiel.

### **2.4.1. Définition**

Le terme « métadonnée » existe depuis longtemps mais il est employé dans les métiers de la conservation de l'information depuis seulement une dizaine d'années. Littéralement, une métadonnée est une donnée sur une donnée. Plus précisément, c'est un ensemble structuré d'informations décrivant ou caractérisant une ressource quelconque. Les métadonnées peuvent être internes ou externes à l'objet numérique.

Les métadonnées nécessaires peuvent être regroupées en trois sous-groupes à savoir, les métadonnées d'identification du document (auteur, organisme, titre, signataire, langue, date, etc.), celles relatives à la description du document (logiciel, version, système, méthode de compression, encryptage, etc.) et celles concernant l'administration du document (contrôle des droits d'accès, codification de sécurité, règle de conservation, création, identifiant unique, signature électronique, etc.). Un quatrième sous-groupe, dépendant de l'architecture du système choisi, pourrait englober les métadonnées relatives à leur administration (création horodatée de l'entrée, journalisation des opérations et modifications, y compris les noms des modificateurs de l'entrée).

## 2.4.2. Fonctions

Parmi les opérations qu'il est possible de réaliser grâce aux métadonnées, figurent la diffusion, la localisation, la gestion (y compris la gestion des droits) et la conservation à long terme d'une ressource. Pour une seule ressource, différents types de métadonnées peuvent être nécessaires afin d'assurer ces différentes fonctions.

Les métadonnées permettent la gestion, la manipulation, le partage, l'exploitation et la préservation des documents numériques. Elles en définissent les droits d'utilisation. Elles favorisent, également, la recherche et le repérage des documents ; et, est-il utile de rappeler que dans l'immensité du Web, un document introuvable est un document inexistant.

Il existe plusieurs standards de métadonnées (EAD, TEI, MARC-XML, BiblioML, MODS<sup>38</sup>, ...) mais, comme le signale la BnF, le Dublin Core est « l'un des formats de métadonnées les plus répandus sur Internet ». Ce schéma de métadonnées générique permet de décrire des ressources numériques ou physiques et d'établir des relations avec d'autres ressources. Il comprend officiellement 15 éléments<sup>39</sup> de description formels (titre, créateur, éditeur, ...), intellectuels (sujet, description, langue,...) et relatifs à la propriété intellectuelle.

---

<sup>38</sup> EAD : Encoded Archival Description, TEI : Text Encoding Initiative, BiblioML : Bibliographic Markup Language, MODS : Metadata Object Description Schema

<sup>39</sup> Annexe 11 : Les 15 éléments du Dublin Core

## ***Partie 4 : Expériences de gestion des documents pour une bibliothèque numérique***

### **1. La gestion et l'archivage de la bibliothèque numérique de l'INHA**

#### **1.1. Les fichiers sources**

##### **1.1.1. Sur cédéroms**

Les images haute définition sont livrées par les prestataires sur cédéroms en deux exemplaires. La bibliothèque numérique de l'INHA représente, actuellement, environ 3000 cédéroms. Cependant, il est difficile, à l'heure actuelle, de donner un ordre de grandeur de la vitesse de dégradation des cédéroms. La durée de vie d'un disque optique (c'est-à-dire la période pendant laquelle le disque peut être lu) dépend de nombreux facteurs : qualité de fabrication, condition du disque avant enregistrement, qualité de l'enregistrement, conditions environnementales, manipulations et entretien. Des cédéroms gravés en 2002 ont déjà présenté des problèmes et un certain nombre de fichiers ont été difficilement récupérables.

L'INHA tente, pourtant, de respecter certaines règles lors de la gravure et du stockage pour permettre une conservation des données sur, au minimum, une dizaine d'années. Ainsi, il est demandé aux prestataires de ne rien écrire sur le disque, sinon un numéro d'ordre sur l'anneau central qui demeure le seul endroit sans risque où des informations peuvent être inscrites. Ensuite, les cédéroms sont conservés en respectant certaines normes sur la qualité de l'environnement et les conditions de stockage, cela afin de limiter leur détérioration. Ils sont conditionnés dans des boîtes opaques qui les préservent de la poussière et de la lumière. Le premier exemplaire est rangé dans le bureau de la numérisation (situé Carré Colbert) sur des étagères. Le deuxième exemplaire est archivé par la bibliothèque détentrice de l'original qui est numérisé sur le cédérom. Pour la collection Doucet, ce deuxième exemplaire est rangé dans les magasins

de la bibliothèque (sur le site Richelieu). Ce stockage en des lieux distants permet de se prémunir des risques liés aux catastrophes naturelles (incendie, inondation...).

### 1.1.2. Sur serveur

Les fichiers numériques sont, également, chargés sur un serveur dédié au stockage des données de la numérisation. L'ensemble des documents de la bibliothèque numérique occupe, actuellement, un espace de 1,9 To (téraoctets)<sup>40</sup>. Le service informatique s'est équipé, depuis la fin de l'année 2006, d'un serveur pouvant stocker de très grands volumes (8 To extensibles jusqu'à 64 To).

La copie des fichiers numériques est accompagnée du calcul d'une clef numérique. Cette clef numérique, générée par le logiciel MD5 (MD signifiant Message Digest), est infalsifiable et unique pour chaque fichier. Elle servira de témoin de l'intégrité des données pour les contrôles ultérieurs. Les fichiers sont classés dans des dossiers par bibliothèque d'origine, par campagne de numérisation, et par CD. Chaque dossier CD comprend un fichier MD5 renfermant les clefs des fichiers contenus dans le CD.

## 1.2. Les fichiers dérivés

### 1.2.1. Le format

Les fichiers sources sont compressés pour leur mise en ligne. Les formats utilisés pour créer ces fichiers dérivés sont le JPEG (Joint Photographic Expert Group) et le PDF (Portable Document Format).

Deux grandes catégories de formats existent : les formats propriétaires et les formats ouverts<sup>41</sup>. Les formats propriétaires, tels que Word, Excel, PowerPoint, PDF... sont des formats créés par des concepteurs de logiciels. Ils ont l'avantage d'être très répandus, connus d'un grand nombre d'utilisateurs et faciles à utiliser, ils ont par contre le désavantage d'être contrôlés par des entreprises obéissant surtout aux lois du marché. A l'inverse, les formats dits ouverts émanent d'un organisme international neutre, visant une durabilité plus grande des documents. Leur utilisation évite la dépendance vis-à-vis d'un fournisseur particulier et favorise l'interopérabilité, garantissant la possible réutilisation des ressources ainsi que leur création et modification par un grand nombre d'applications.

---

<sup>40</sup> Un téraoctet vaut 1 024 Go (gigaoctets), soit 1 099 milliards d'octets

PDF appartient à la Société Adobe. Il s'agit donc d'un format propriétaire mais, en raison de sa très grande popularité, il est reconnu, de facto, comme une norme. De plus, ce format est public c'est-à-dire que sa structure et ses spécifications sont connues car elles ont été publiées.

Le format JPEG est une norme établie par un organisme officiel de normalisation, il ne dépend pas d'un logiciel en particulier et son utilisation est libre de droits.

Face au grand nombre de formats en circulation sur le Web, le recours à ces deux formats a été privilégié car ils représentent une garantie en termes d'accessibilité et d'interopérabilité puisqu'ils sont considérés comme des normes stables.

### 1.2.2. Les services proposés par Picturelan

Puisque la réalisation du système de banque de données multimédias de l'INHA prendra un certain temps, la bibliothèque a décidé de faire appel à un prestataire extérieur pour mettre à disposition les fonds numérisés de la bibliothèque. La société Atchik Image et son logiciel Picturelan ont été retenus pour assurer l'hébergement et la fourniture d'accès aux images. Picturelan stocke les fichiers destinés à la mise en ligne sur un serveur (sauvegardé mensuellement sur bandes LTO 3) et propose une interface publique ainsi qu'un module d'administration à distance.

En mai 2007, l'espace de stockage dévolu à la Bibliothèque numérique de l'INHA est devenu insuffisant, étant donné l'ampleur et le rythme des mises en ligne. Les 100 Go prévus dans le cadre du marché initial ont été largement saturés (près du double). Pour pouvoir continuer à enrichir sa Bibliothèque numérique, l'INHA a demandé que la capacité de stockage dédiée s'élève à 500 Go.

## 2. Exemples d'autres expériences

Afin d'alimenter et d'élargir cette réflexion, il paraît intéressant d'analyser les situations d'autres établissements gérant des fonds numériques.

---

<sup>41</sup> Annexe 12 : Les formats

## 2.1. La Bibliothèque numérique Medic@

Depuis l'automne 2000, la Bibliothèque Interuniversitaire de Médecine de Paris (BIUM) a mis en place la bibliothèque numérique Medic@<sup>42</sup> afin de valoriser ses fonds qui figurent parmi les plus riches collections de documents patrimoniaux en histoire de la médecine au monde. Pour mener à bien ce projet de numérisation patrimoniale et d'édition scientifique, la BIUM dispose d'un atelier de numérisation en interne équipé d'un Digibook 2000 qui numérise en noir et blanc. La bibliothèque fait également appel à des prestataires extérieurs lors de campagnes de numérisation portant sur des documents spécifiques (par exemple à l'ACRPP<sup>43</sup> pour la numérisation des périodiques particulièrement fragiles du XIXe). Début 2007, Medic@ propose plus de 3 600 textes intégraux numérisés en mode image organisés en rubriques thématiques.

### 2.1.1. Choix opérés pour la gestion des documents numériques

Diverses bases de données FileMaker Pro ont été créées pour gérer la collection numérique. Une première base décrit les rubriques de Medic@, une deuxième chaque document contenu dans les diverses rubriques et, enfin, une dernière base sert à la description des vues composant chaque document. Les champs de ces bases de données ont été mis en conformité, en 2005, avec les éléments du Dublin Core.

- **Fichiers sources**

La gravure des images se fait au format JPEG avec un faible taux de compression, exclusivement sur des cédéroms MPO (fabricant de cédéroms certifiés Iso 9001/v2000) et les cédéroms sont ensuite stockés dans une armoire en magasin. Depuis sa création, Medic@ a déjà généré plus de 3 000 cédéroms.

Les premières années d'existence de Medic@, le seul support de stockage était le cédérom. Mais, après quelques années, la question de l'archivage des données de Medic@ s'est imposée aux membres de l'équipe puisque huit cédéroms gravés en 2003 sont devenus illisibles (la bibliothèque a perdu son travail de numérisation pour un ensemble de 141 thèses). Pour remédier à ce problème et minimiser les risques de perte de données, la BIUM a décidé de doubler l'archivage sur cédéroms d'un archivage sur serveur. Les cédéroms ont été transférés sur un serveur de l'Université Paris 5 (5,45 To à disposition), lui-même sauvegardé sur bandes.

---

<sup>42</sup> <http://www.bium.univ-paris5.fr/histmed/medica>

Actuellement, les documents numérisés sont, donc, archivés sur support optique contrôlé (cédéroms) et sur le serveur de l'université. La BIUM étudie la possibilité d'un archivage pérenne externalisé de ses données, éventuellement par le Centre Informatique National de l'Enseignement Supérieur (CINES) qui met actuellement en place un service générique d'archivage numérique qu'il proposera ensuite à l'ensemble de la communauté Enseignement supérieur et Recherche.

- **Fichiers dérivés**

Les images mises en ligne sont des images de résolution moyenne, dont la taille varie selon le format de l'ouvrage original (40 pix./cm, soit environ 100 dpi, ce qui correspond sur un écran CRT 17" en 1024x768 à une image légèrement plus grande que l'original). Une copie des images dégradées mises en ligne sur le site est sauvegardée sur un serveur appartenant à la bibliothèque.

Une adresse permanente est attribuée à chaque document (ex. : <http://www.bium.univ-paris5.fr/histmed/medica/cote?01894>) et, également, à chaque page (ex. : <http://www.bium.univ-paris5.fr/histmed/medica/page?01894&p=86>), ce qui permet de faire des citations sur le Web ou dans une publication imprimée.

## 2.1.2. Analyse

- **Faiblesses**

L'emploi de diverses bases FileMaker Pro rend la gestion de la collection numérique assez complexe puisque ce logiciel ne permet pas de créer de bases de données relationnelles. Par conséquent, il faut penser à toutes les répercussions que peut avoir une modification dans une base et les reporter manuellement. Ce mode de fonctionnement ne prémunit pas des risques d'oublis ou d'erreurs dans les mises à jour et complique la gestion de la relation entre l'unité numérique (le fichier informatique) et l'unité du document physique.

- **Points forts**

La BIUM propose des services commerciaux associés à Medic@ : ventes de cédéroms des ouvrages numérisés (les images sont fournies telles qu'elles sont accessibles en ligne, pour un usage strictement privé), ventes d'images en haute définition et vente de reprints (commande minimum de 30 exemplaires).

<sup>43</sup> Association pour la conservation et la reproduction photographique de la presse

Récemment, la bibliothèque a mis en place un réservoir OAI-PMH (Open Archives Initiative's Protocol for Metadata Harvesting), qui permet aux professionnels qui le souhaitent de récupérer les notices des documents chargés. Ce réservoir (accessible à l'adresse : <http://web2.bium.univ-paris5.fr/oai/oai2.php>) permet de moissonner toutes les notices en deux formats : Dublin Core non qualifié (Dublin Core «simple»), et Dublin Core qualifié. Les sets mis en place permettent de moissonner le réservoir par type de document, par série ou dossier documentaire et par siècle. Ce travail a été réalisé dans le cadre d'une étroite collaboration avec la bibliothèque numérique de la Bibliothèque nationale de France (la BIUM a le statut de Pôle associé de la BnF). En effet, Medic@ et Gallica ont mis en place une politique concertée de numérisation, un partage de compétences, un échange de fichiers via les réservoirs OAI-PMH ainsi qu'une passerelle d'interrogation commune. Ce réservoir OAI-PMH a, également, permis à la BIUM de développer ses partenariats à l'étranger puisque l'University of Michigan a d'ores et déjà commencé le moissonnage.

Pour une bibliothèque numérique, disposer d'un tel réservoir représente un double avantage : il améliore la visibilité des ressources sur le Web et il permet de mettre en place des collections numériques partagées ; il crée, ainsi, plus d'échanges avec plus de types d'acteurs différents. Dans un paysage numérique caractérisé par des volumes d'informations disponibles abondants et sans cesse croissants, les bibliothèques doivent trouver des solutions techniques qui leur donnent la possibilité de décloisonner les corpus et de faciliter les partenariats et la numérisation partagée. En répondant à cette attente, le protocole OAI-PMH est en train de devenir un standard pour l'interopérabilité et l'échange de métadonnées en ligne. La mise en place du réservoir a entraîné un dynamisme certain puisque l'équipe de Medic@ envisage, par la suite, de nouvelles collaborations (avec la National Library of Medicine (USA) et le Wellcome Institute de Londres) et souhaite participer au projet de Bibliothèque Numérique Européenne lancé en 2005.

## 2.2. Le CINES (Centre Informatique National de l'Enseignement Supérieur)

### 2.2.1. La gestion des documents numériques au CINES

Depuis 2004, le MENESR (ministère de l'Éducation nationale de l'enseignement supérieur et de la recherche) a chargé le CINES de la conservation à long terme des documents numériques du patrimoine scientifique Université-Recherche (pour les fonds jugés d'intérêt national devant être conservés indéfiniment). Dans un premier temps, il lui a demandé de travailler sur 2 projets pilotes : l'archivage pérenne des thèses électroniques et celui des revues en sciences humaines et sociales du portail Persée<sup>44</sup>.

Afin de remplir sa mission d'archivage numérique pérenne, le CINES a décidé de se doter d'une plate-forme : PAC (plate-forme d'archivage du CINES). PAC se compose d'une plate-forme matérielle (serveur de versement et serveur de stockage sauvegardé sur bandes Robothèque Storagetek), d'une partie logicielle développée en langage de programmation Java et d'un outil open source de validation Jhove (JSTOR/Harvard Object Validation Environment) pour détecter les formats des archives transférées.

Les référentiels privilégiés pour la conception de la plate-forme ont été le modèle OAIS (Open Archival Information System), le standard d'échange DAF-DGME et les travaux du groupe PIN (Pérennisation des Informations Numériques).

PAC dispose d'un serveur de versement où le producteur des documents pourra transmettre ses archives, d'un serveur de stockage où sont conservés les documents et d'un serveur d'accès où le producteur et les services demandeurs autorisés à consulter les archives peuvent rechercher et obtenir un document.

### 2.2.2. Fonctionnement de la plate-forme d'archivage

- **Le transfert**

Avant tout versement, le responsable du transfert contacte le CINES afin d'élaborer le protocole de transfert (identification des données à pérenniser, volumétrie, formats choisis, conformité avec les normes, précisions sur la communicabilité et modalités de versement).

Le transfert est effectué par réseau (via SFTP) ou sur support amovible, suivant le protocole de transfert préalablement établi. Lors de la réception de l'objet à archiver, le

---

<sup>44</sup> <http://www.persee.fr/>

service PAC envoie par courriel un accusé de réception de transfert. Après avoir repéré les données transférées, l'application contrôle leur validité technique (conformité des formats, structure de versement...) en s'assurant qu'elles respectent les conditions définies dans le protocole de transfert.

Si le contrôle s'avère positif, le versement est transféré sur le serveur de stockage où il est archivé. Le CINES envoie un certificat d'archivage comportant l'identifiant unique et pérenne de l'archive transférée, ses métadonnées ainsi que la liste des contrôles effectués par le CINES.

Si le contrôle est négatif, le service PAC enverra un avis de rejet, et le versement refusé est déplacé dans un répertoire spécifique où le service versant peut venir le récupérer afin de modifier les éléments non conformes.

- **Le stockage**

Les données contrôlées sont stockées sur serveur. Ce stockage s'accompagne d'une veille technologique réalisée par le CINES. Cette veille porte sur les évolutions d'environnement de stockage, des technologies et des formats. Le CINES exerce également une surveillance sur les archives déposées puisqu'il assure le contrôle des supports, les migrations de support à support, les restaurations et le contrôle des métadonnées. Le service versant est averti des modifications lorsqu'elles touchent à la mise à jour des métadonnées ou au format, grâce à un avis de modification précisant les références des archives et la nature des modifications.

Chaque année, le CINES envoie aux services producteurs une liste des documents archivés sur la plate-forme ainsi que les résultats des tests de contrôle des supports.

- **La communication**

La communication des archives s'opère via le serveur d'accès. PAC offre la possibilité de consulter le sous-catalogue d'un projet particulier et d'effectuer une recherche grâce à l'outil de recherche du serveur d'accès. La demande de communication d'archives peut concerner les données ou les métadonnées. Pour faire une demande de communication d'archives auprès du CINES, le service demandeur choisit le mode de restitution (consultation en ligne, support amovible ou transmission réseau) puis envoie une demande de communication. Le CINES envoie en retour un accusé de réception de la demande.

Si l'archive est librement communicable (critère défini préalablement dans le protocole de transfert), le CINES communique l'archive au service demandeur sans en référer au service producteur (le demandeur et le producteur peuvent être différents). Le service

demandeur retourne ensuite au CINES un accusé de réception de communication si l'archive est parvenue avec succès au service demandeur, ou un avis d'anomalie de communication dans le cas contraire.

Si l'archive n'est pas communicable, PAC envoie une demande d'autorisation de communication au service producteur qui peut agréer ou refuser la demande.

- **Eliminations de documents**

L'élimination d'objets conservés au CINES est un processus qui ne se réalise qu'avec l'accord explicite du service producteur. Elle peut être proposée par le CINES ou demandée par le service producteur. Une demande d'élimination d'archives doit comprendre également les métadonnées des documents concernés. Une fois la destruction effectuée, le CINES envoie au service producteur une notification d'élimination d'archives qui rappelle l'identification des archives concernées par cette élimination.

### 2.2.3. Analyse

- **Faiblesses**

La version actuelle de la plate-forme dispose d'un espace de stockage avec une capacité limitée à 100 Go ce qui apparaît comme largement insuffisant au regard des deux premiers projets dont le CINES a la charge. La prochaine version de PAC, prévue pour décembre 2007, devrait offrir 10 To. Mais l'archivage des thèses électroniques (environ 10000 thèses par an) nécessite 100 Go par an et celui des revues du portail Persée (environ 4000 revues par an) exige 10 To par an. Une montée rapide en puissance des capacités de stockage semble primordiale ; et l'étendue du service PAC à l'ensemble des établissements de l'Enseignement Supérieur (comme le prévoit la BIUM) paraît encore bien lointaine.

- **Points forts**

PAC a été créé en conformité avec le standard d'échanges de données pour l'archivage DAF-DGME et se conforme à la législation relative au droit intellectuel, au droit des auteurs, au droit de reproduction et au droit de représentation des œuvres de l'esprit. Par conséquent, la plate-forme intègre et respecte les délais prévus par le Code du patrimoine en matière de communicabilité des archives.

En formalisant les échanges entre PAC et le service producteur, le CINES offre un exemple concret d'application du modèle OAIS. Cette expérience est très intéressante car la principale difficulté du modèle est qu'il ne donne pas de méthodologie pour la

mise en œuvre d'un tel système. En effet, O AIS est un modèle conceptuel et ne précise pas, par conséquent, les détails pour sa mise en œuvre. Le CINES propose, ainsi, une réalisation concrète en conformité avec O AIS et permet ainsi de passer de l'abstraction à l'application.

## ***Partie 5 : Recommandations et propositions pour améliorer la gestion des documents de la bibliothèque numérique de l'INHA***

Un système de gestion adapté au cycle de vie du document numérique doit remplir trois exigences principales :

- maîtriser les outils de création et de gestion,
- faciliter les échanges et la diffusion,
- garantir l'accessibilité et la conservation.

Pérennité, intégrité, sécurité et traçabilité sont les quatre principaux sujets pouvant être définis en termes d'enjeux et d'objectifs opérationnels.

Les actions de conservation des documents numériques ne peuvent pas être des actions de conservation curative mais sont impérativement des actions qui doivent être gérées de manière préventive. La pérennisation d'objets numériques implique, donc, la mise en place d'un système de gestion spécifique. Comme pour la conservation des documents classiques, celle des objets numériques ne peut pas être le fruit du hasard, et les règles normalisées, définies pour un magasin dans lequel sont rangés et classés les documents papier, vont donc devoir être transposées aux règles de gestion et d'archivage pérenne des objets électroniques (en intégrant les spécificités du numérique). Ces règles doivent évidemment permettre non seulement la conservation mais aussi l'accès et l'utilisation.

### **1. Analyse des stratégies de conservation**

Comme nous l'avons vu plus haut, de nombreuses stratégies de conservation d'objets numériques existent mais aucune ne convient à tous les types de données, établissements ou situations.

#### **1.1. Pratiques non applicables à l'INHA**

La copie de la chaîne de bits représente une stratégie minimale de maintien des données. Cependant, elle ne résout que le problème de la perte de données due à une panne d'un appareil ou d'un support, à un mauvais fonctionnement, à une dégradation, à une destruction malveillante ou à un désastre naturel. La régénération, quant à elle, permet

essentiellement de répondre aux éventuels problèmes de dégradation et d'obsolescence du support de stockage. Composantes essentielles de tout programme de conservation d'objets numériques, ces deux premières stratégies ne peuvent toutefois pas être considérées comme un programme complet de conservation à long terme.

La copie sur supports persistants n'a aucun effet sur un certain nombre de causes possibles de perte, dont la perte physique due à une catastrophe, l'obsolescence du support, l'obsolescence des algorithmes de codage ou de mise en forme. Un support durable pourrait même mettre le contenu en danger en donnant un faux sentiment de sécurité.

La conservation technologique élargit les possibilités d'accès pour des supports et des formats de fichier périmés mais conduit inévitablement à une impasse puisqu'aucune technologie obsolète ne peut être maintenue indéfiniment en état de fonctionnement. De plus, elle n'est pas à la portée d'un établissement seul car le maintien d'une technologie obsolète nécessite un investissement considérable en équipement et en personnel.

L'archéologie numérique est pratiquée prioritairement par des entreprises à but lucratif spécialisées dans la récupération de données car elle exige une grande variété de dispositifs (y compris de types obsolètes) ainsi que des installations spéciales.

La sauvegarde analogique est une technique coûteuse qui ne permet pas de conserver les spécificités de la forme numérique, ceci limite sa pertinence à certaines catégories de documents : les textes et les images fixes monochromatiques sont les mieux adaptés à ce genre de conversion.

L'émulation est une technique séduisante puisqu'elle ne nécessite pas de modification du document. Les pertes, de génération en génération, sont donc évitées. Cependant, cette méthode n'est pas encore véritablement applicable car l'intégrité et l'authenticité des données ne sont nullement garanties. Cette méthode paraît difficilement réalisable à long terme puisque chaque nouveau système impliquerait de réécrire tous les émulateurs pour les systèmes antérieurs, en nombre toujours croissant. Qui plus est, l'émulation de technologies anciennes requiert, pour son utilisation, un savoir-faire que ne possèdent pas nécessairement le personnel et les usagers souhaitant consulter les documents. Si nous voulons que les collections soient accessibles au plus grand nombre, nous ne pouvons exiger des utilisateurs qu'ils connaissent effectivement des logiciels obsolètes.

## 1.2. Une stratégie envisageable : la migration

Le procédé de migration, par lequel sont transférés les fichiers d'une technologie menacée d'obsolescence à une technologie plus récente, est celui qui, actuellement, est le plus en usage. Il est explicitement recommandé par de nombreuses institutions. La migration assure une longévité supérieure aux documents numériques. Mais, il faut bien garder à l'esprit qu'elle n'est pas sans faille.

Tout d'abord, la conversion entraîne parfois des pertes de données, dues à leur retraitement, ce qui menace l'intégrité des fichiers. Comme de telles pertes peuvent survenir à chaque migration, la qualité du fichier risque de s'amoinrir de génération en génération et la fiabilité du document peut être compromise.

Ensuite, les données sont séparées de leur mise en forme originelle et ne retrouveront plus, une fois remontées sur un autre outil logiciel, leur apparence et présentation d'origine. La conservation des données électroniques par la technique de la migration aboutit donc à un résultat très différent de ce qu'on connaît avec le papier et les autres supports analogues : pour ces derniers on s'attache à conserver en même temps l'information et la façon dont celle-ci a été, dès l'origine, mise en forme (et cette mise en forme est, souvent, elle-même source d'information). Quiconque consultera des documents ou des données ayant fait l'objet d'une migration sera donc tributaire des instruments de recherche et de la qualité des métadonnées, pour reconstituer l'aspect initial des documents et retrouver la perception que pouvaient en avoir les producteurs. De plus, de par la séparation des données de leur contexte technologique de création, la migration est dans une certaine mesure irréversible (d'autant plus lorsque le mode de codage utilisé est spécifique à un constructeur et à un système d'exploitation).

Enfin, la haute instabilité de la technologie fait en sorte qu'il est difficile de prévoir le moment où les migrations seront nécessaires (les formats et les supports ne changent pas tous au même rythme). Afin de limiter les risques, une planification stricte du processus de migration s'impose et une veille technologique permanente doit être mise en place. L'opération de migration exige, donc, de lourds investissements liés aux besoins technologiques, un énorme travail et doit mobiliser, parmi les effectifs en charge de l'archivage, des personnes averties de l'évolution des langages et des outils.

La migration apparaît, à plusieurs égards, risquée et coûteuse mais, en l'absence de procédés alternatifs viables, cette technique reste la plus probante pour assurer la pérennité des données et des informations.

## **2. L'organisation d'un système de gestion**

Un système d'archivage construit sur le modèle de référence de l'OAIS permet de mettre en place une gestion et une conservation des objets numériques correspondant aux objectifs de chaque institution, en fonction des moyens qu'elle affecte à cette mission. Le modèle permet de réaliser un système complet intégré complètement dans l'organisation de l'institution ou de passer un contrat avec un tiers archiveur sur des bases précises de service et de responsabilité.

Dans la vision proposée plus haut, un système de gestion doit prendre en charge les données à préserver, assurer la préservation de ces données quelle que soit la durée et restituer ces données à la demande. Pour schématiser, les caractéristiques attendues sont donc :

- la garantie de préservation des données sans altération (garantie d'intégrité),
- la capacité à s'adapter aux évolutions de la technologie sans impact sur la préservation des fichiers,
- la sécurité et la confidentialité.

L'analyse des divers modèles conduit à concevoir l'organisation d'un système de gestion basé sur trois services : un service des «entrées », un service de stockage et un service de gestion des données et communication.

### **2.1. Service des « entrées »**

La fonction «entrées » assure la réception des objets transmis par le producteur, contrôle leur validité, procède à l'ajout des métadonnées techniques et attribue à chaque objet un identifiant unique. Il transfère, ensuite, les objets numériques (données et métadonnées) au service de stockage et les métadonnées au service de gestion des données et communication.

#### **2.1.1. Le versement**

L'objectif est de fournir au service des «entrées » un ensemble complet, organisé et convenablement décrit d'objets numériques. Les compétences requises à cette étape portent sur la connaissance des informations à préserver, la vérification de leur intelligibilité et de leur complétude. La maîtrise des différents formats paraît indispensable pour assurer la qualité des données et des métadonnées introduites. Les critères de conformité des versements sont définis dans une phase préliminaire qui décrit

la structure formelle des documents, ils peuvent être repris dans une DTD (Document Type Definition). Une fois le système installé, l'introduction et la mise à jour des données et des métadonnées doivent être soigneusement planifiées et gérées.

### 2.1.2. Après le versement

Le service des «entrées» vérifie, ensuite, la conformité du versement. Ce contrôle porte sur la structure informatique, le format et les métadonnées. Un outil permettant d'appliquer les aspects d'identifications, de validation et de caractérisation d'un format a été développé par JSTOR et Harvard : JHOVE (JSTOR/Harvard Object Validation Environment). Il est distribué en logiciel libre et peut être intégré dans le déroulement des opérations d'entrées dans une archive OAIS en ce qui concerne la création du paquet d'information soumis (SIP) et la validation du versement.

Suite à ce contrôle et après validation, l'empreinte numérique de chaque fichier sera calculée. Le versement pourra alors être archivé sur le serveur.

## 2.2. Le service « stockage »

L'objectif principal du service « stockage » est de garder les documents en ordre, correctement nommés, dans un environnement sûr, dans le format le plus approprié pour la publication, la réutilisation et la conservation.

### 2.2.1. L'archivage

Le service « stockage » prend en charge les fichiers en vue de leur conservation à long terme. Tout d'abord, le service doit disposer de grandes capacités et de technologies avancées pour le stockage. En quatre ans d'existence, la bibliothèque numérique de l'INHA a déjà généré 1,9 To de fichiers sources auxquels il faut ajouter les 200 Go des fichiers dérivés. Le service doit pouvoir intégrer l'ensemble de ces fichiers mais aussi créer une arborescence virtuelle afin de proposer une vue détaillée permanente de tout ou partie de l'espace de stockage.

Lors de cette étape, chaque fichier se voit attribuer un « niveau » de préservation car les traitements pourront différer suivant le fichier déposé. On peut, par exemple, imaginer une duplication et un stockage en plusieurs sites séparés pour les fichiers sources, et un traitement plus léger pour les fichiers dérivés.

### 2.2.2. Surveillance des fichiers

Le service «stockage» doit garantir l'intégrité des fichiers et prendre en compte les évolutions de la technologie (techniques de stockage, systèmes, logiciels...).

Des moyens de surveillance de l'état des fichiers doivent être intégrés afin de garantir l'intégrité de l'information c'est-à-dire la lisibilité du document, la stabilité de son contenu informationnel ainsi que la traçabilité des opérations qu'il a subies. Les données doivent pouvoir toujours être interprétées quelle que soient les évolutions technologiques. La sécurisation contre la corruption des données s'appuie sur deux mécanismes principaux : les codes correcteurs d'erreurs qui pallient aux pertes et aux déformations du train de bits et l'empreinte numérique qui vise à contrôler l'intégrité d'une donnée.

Il existe des outils de surveillance dynamique des médias (StorSentry, AMASS Infinite File Life, Service Delivery Platform...) qui permettent la prévision des migrations en collectant les statistiques sur le comportement des lectures (erreurs, rétention, nombre de lectures, ...).

Il faut garder à l'esprit que plus les formats et les supports sont homogènes, plus le processus de migration s'en trouve facilité, ce qui se répercute, également, de façon positive sur les coûts de l'opération.

## **2.3. Service de gestion des données et communication**

Le serveur de communication gère la relation avec les utilisateurs. L'objectif principal d'un système de gestion d'informations numériques est de permettre à une communauté d'accéder aux informations et de les réutiliser.

### 2.3.1. La recherche de fichiers

Le service doit maîtriser les processus de recherche d'information et doit être capable de déterminer des critères de sélection des données adaptés aux utilisateurs, ces derniers étant définis comme la personne physique ou le programme qui active les fonctions du service de communication. Il doit proposer une interface d'interrogation avec plusieurs voies d'accès, évolutive en fonctions des besoins des usagers.

### 2.3.2. L'accès et la récupération de fichiers

Tous les utilisateurs n'ont pas nécessairement le même profil et le service doit gérer différents droits d'accès, contrôler ou limiter l'accès afin de savoir qui accède à quoi.

Après authentification, l'utilisateur peut récupérer les fichiers.

Le service doit fabriquer un index pour obtenir rapidement l'adresse mémoire où se trouve un document. Grâce à son identifiant, le document archivé est communiqué et/ou recopié sur un média approprié à la diffusion.

Pour conclure, le système de gestion entraînera une plus grande efficacité et réactivité grâce au temps épargné à retrouver les différentes versions de fichiers existants. Il permettra de suivre en temps réel la production.

Une telle organisation doit, enfin, pouvoir faire l'objet de contrôles et d'audits externes. Au travers de son organisation, de ses moyens, de ses équipes, des standards et des procédures appliquées, le système doit être en mesure d'apporter la démonstration de sa capacité à assurer sa mission c'est-à-dire à gérer et préserver à long terme les informations sous forme numérique dont il a la charge. Ceci ouvre un nouveau champ de réflexion sur la certification des archives numériques, qui pourra faire l'objet d'un prochain rapport !

## **Conclusion**

La gestion et la conservation des collections font, depuis longtemps, partie de la sphère traditionnelle des activités bibliothéconomiques. Ces activités prennent toutefois une nouvelle dimension avec l'introduction des documents numériques. La numérisation implique la prise en compte de nombreuses considérations qui ne peuvent pas être réglées définitivement. Comme nous l'avons vu, la rapidité avec laquelle se modifie la technologie appelle à la plus grande vigilance de la part des responsables de bibliothèques numériques, qui n'ont guère d'autre choix que d'adopter une approche proactive s'ils veulent que leurs collections demeurent accessibles et conservent leur intégrité malgré les changements technologiques.

Certes, la conservation à long terme des données numériques est une opération coûteuse. Mais, reconstruire des données perdues, détériorées ou détruites l'est encore plus. Un document électronique qui n'a pas été conservé ou qui a été mal conservé, de sorte que l'on ne puisse en garantir l'intégrité, est à jamais perdu.

Plus la collection numérisée de l'INHA prend de l'ampleur, plus l'absence de système adapté pour la gestion et la conservation des fichiers se fait sentir. Ce rapport a tenté de présenter les méthodes possibles à mettre en place, dans l'immédiat, pour favoriser l'émergence d'une gestion adéquate des documents numériques. Les choix opérés sont d'une importance considérable car ils détermineront les conditions futures de conservation de ce véritable patrimoine culturel. Mais, il faut veiller à ne pas tomber dans un écueil inverse qui consisterait en une focalisation excessive sur les questions techniques (choix de supports et formats). L'enjeu est de réussir à mettre en place une solution technique et organisationnelle viable.

La conservation des documents numériques sur le long terme reste, donc, un problème ouvert. Le manque de recul est certain. Il serait encore possible d'explorer bien d'autres voies, que ce rapport n'a pu qu'évoquer, notamment celle de la coopération entre institutions possédant des collections numérisées. En effet, qu'elle soit internationale, nationale ou locale, la collaboration paraît un moyen des plus efficaces pour faire face

aux défis de la numérisation et de la conservation des documents numériques. La coopération pourrait hautement favoriser la bonne marche des projets de numérisation et la conservation des documents numériques grâce à l'adoption de politiques et de standards communs, à la mutualisation des connaissances, au partenariat au sein de groupes de recherche alliant partage des expertises et des coûts. Ajoutons qu'un autre avantage indéniable serait l'augmentation de la visibilité des collections, ce qui pourrait se traduire par l'augmentation de leur utilisation.

Pour les institutions qui créent des informations numériques à caractère patrimonial, maîtriser le cycle de vie du document numérique représente, à bien des égards, un véritable défi, d'autant plus que les enjeux vont maintenant au-delà des seules questions techniques. En effet, les enjeux sont aujourd'hui, également, culturels, historiques, politiques, économiques et sociaux ; et, comme le souligne le Groupe PIN, on peut s'interroger sur ce « que deviendrait une société incapable d'assurer la protection de ses informations patrimoniales ».

## ***Bibliographie***

### **1. Généralités sur l'INHA**

**INSTITUT NATIONAL D'HISTOIRE DE L'ART.** *Rapport d'activité de l'Institut National d'Histoire de l'Art.* Paris : Institut national d'histoire de l'art, 2004. 63p.

**COMMENT, Bernard, CHAPON, François.** *Doucet de fonds en combles : trésors d'une bibliothèque d'art.* Paris : Institut national d'histoire de l'art, INHA : Herscher, 2004. 141 p. (La Galerie (Paris. 2004))

ISBN 2-7335-0361-8

**JOUGUELET, Suzanne.** La rénovation du Quadrilatère Richelieu est lancée ! *Chroniques de la Bnf*, printemps 2007, n°38, p.10.

**LE NORMAND ROMAIN, Antoinette, INSTITUT NATIONAL D'HISTOIRE DE L'ART.** *Site de [l'Institut National d'Histoire de l'Art]* [en ligne]. Paris : INHA, [2002]. [consulté le 16 juillet 2007]. Disponible sur Internet :  
< <http://www.inha.fr/> >

### **2. Généralités sur les bibliothèques numériques**

**ASSOCIATION DES BIBLIOTHECAIRES FRANÇAIS, BIBLIOTHEQUE NATIONALE DE FRANCE.** *Bibliothèques numériques : Où en sommes-nous ?* : [actes de la journée du 10 octobre 2005] [en ligne]. Paris : Association des Bibliothécaires Français, 2005. [consulté le 3 avril 2007]. Disponible sur Internet :

< [http://www.abf.asso.fr/article.php3?id\\_article=485](http://www.abf.asso.fr/article.php3?id_article=485) >

**BIBLIOTHEQUE NATIONALE DE FRANCE.** *Numérisation et données numériques* : [pages professionnelles] [en ligne]. Paris : Bibliothèque Nationale de

France, 2004. Mise à jour le 01-03-2007. [consulté le 3 avril 2007]. Disponible sur Internet :

< <http://www.bnf.fr/pages/zNavigat/frame/infopro.htm>>

**BURESI Charlette, CEDELLE-JOUBERT Laure.** *Conduire un projet de numérisation.* Villeurbanne : ENSSIB ; Paris : Éd. Tec et Doc : Lavoisier, 2002. 326 p. (Collection La boîte à outils ; 13)

ISBN 2-7430-0551-3

**JACQUESSON, Alain, RIVIER Alexis.** *Bibliothèques et documents numériques : concepts, composantes, techniques et enjeux.* Nouvelle éd. Paris : Ed. du Cercle de la librairie, 2005. 573 p. (Bibliothèques)

ISBN 2-7654-0915-3

**LAGOZE, Carl, KRAFFT, Dean B., PAYETTE, Sandy [et al.]**. Qu'est-ce qu'une bibliothèque numérique, au juste ? Au-delà des fonctions recherche et accès dans la National Science Digital Library [en ligne]. *AMETIST (Appropriation, Mutualisation, Expérimentations des technologies de l'IST)*, septembre 2006, n° 0, p.25-56 [en ligne]. Disponible sur Internet :

<<http://archivesic.ccsd.cnrs.fr/docs/00/12/34/23/PDF/lagoze70906.pdf>> [consulté le 3 avril 2007].

**LE MOAL, Jean-Claude.** Journée d'étude ADBS : Expériences et projets de bibliothèques virtuelles du Sud. *Documentaliste - Sciences de l'information*, avril 2002, Volume 39, n° 1-2, p.45-47 [en ligne]. [consulté le 1er avril 2007]. Disponible sur Internet :

<[http://www.adbs.fr/uploads/docsi/1209\\_fr.pdf](http://www.adbs.fr/uploads/docsi/1209_fr.pdf)>

**LUPOVICI, Catherine.** De la bibliothèque classique à la bibliothèque numérique : continuité et rupture. *Documentaliste - Sciences de l'information*, décembre 2000, Volume 37, n° 5-6, p.286-297 [en ligne]. [consulté le 1er avril 2007]. Disponible sur Internet :

< [http://www.adbs.fr/uploads/docsi/670\\_fr.pdf](http://www.adbs.fr/uploads/docsi/670_fr.pdf)>

**OPPENHEIM, Charles, SMITHSON, Daniel.** What is the hybrid library? *Journal of Information science*. Vol. 25, n°2, 1999, p.97-112.

**PAPY, Fabrice.** *Les bibliothèques numériques*. Paris : Hermes science publ. : Lavoisier, 2005. 220 p. (Collection : IC2)  
ISBN 2-7462-1036-3

**RODES, Jean-Michel, PIEJUT, Geneviève, PLAS, Emmanuèle.** *La mémoire de la société de l'information*. Paris : Unesco, 2003. 103 p. [en ligne]. [consulté le 16 août 2007]. Disponible sur Internet :  
<<http://unesdoc.unesco.org/images/0013/001355/135529f.pdf>>

### **3. Stratégies de gestion des bibliothèques numériques**

**BIBLIOTHEQUE NATIONALE DE FRANCE.** *Numérique et bibliothèques : le deuxième choc* : 4<sup>ème</sup> édition des Entretiens de la BnF, 7-8 décembre 2006 [en ligne]. Paris : Bibliothèque nationale de France, 2006. [consulté le 3 avril 2007]. Disponible sur Internet :  
<[http://www.bnf.fr/pages/infopro/journeespro/jp\\_entretiens06.htm](http://www.bnf.fr/pages/infopro/journeespro/jp_entretiens06.htm)>

**CONFERENCE DES RECTEURS ET DES PRINCIPAUX DES UNIVERSITES DU QUEBEC (CREPUQ).** *La gestion des documents numériques des établissements universitaires du Québec état de situation et planification stratégique* [en ligne]. [Montréal] : CREPUQ, 2004. [consulté le 1<sup>er</sup> avril 2007]. Disponible sur Internet :  
<<http://www.crepuq.qc.ca/documents/arch/Rapport-GGDN.pdf>>

**FRANCE. Sous-direction des bibliothèques et de la documentation.** *A propos de la numérisation* : Notions et conseils techniques élémentaires [en ligne]. Paris : Sous-direction des bibliothèques et de la documentation, 1998. Mise à jour le 28-08-2002. [consulté le 3 avril 2007]. Disponible sur Internet :  
<http://www.sup.adc.education.fr/bib/Acti/Num/num.html>

**IMARK.** *Le kit de ressources pour la gestion de l'information* [en ligne]. Rome : Organisation des nations unies pour l'alimentation et l'agriculture, cop. 2006. [consulté le 5 avril 2007]. Disponible sur Internet :

<[http://www.imarkgroup.org/modulesintro\\_fr.asp](http://www.imarkgroup.org/modulesintro_fr.asp)>

**RESEAU CANADIEN D'INFORMATION SUR LE PATRIMOINE (RCIP).** *Création et gestion de contenu numérique* [en ligne]. Gatineau (Québec) : RCIP, 2002. Mise à jour le 05-03-2007. [consulté le 3 avril 2007]. Disponible sur Internet :

< [http://www.rcip.gc.ca/Francais/Contenu\\_Numerique/index.html](http://www.rcip.gc.ca/Francais/Contenu_Numerique/index.html)>

**RESEAU CANADIEN D'INFORMATION SUR LE PATRIMOINE (RCIP).** *Numérisez vos collections* : guide à l'intention des gestionnaires chargés de la planification et de la mise en œuvre de projets d'informatisation [en ligne]. Gatineau (Québec) : RCIP, 2002. Mise à jour le 27-04-2002. [consulté le 3 avril 2007]. Disponible sur Internet :

< [http://www.rcip.gc.ca/Francais/Contenu\\_Numerique/Guide\\_Gestionnaires/index.html](http://www.rcip.gc.ca/Francais/Contenu_Numerique/Guide_Gestionnaires/index.html)>

**FERCHAUD, Bernadette.** Journée d'étude ADBS : Expériences et projets de bibliothèques virtuelles du Sud. *Documentaliste - Sciences de l'information*, avril 2002, Volume 39, n° 1-2, p.45-47 [en ligne]. [consulté le 1er avril 2007]. Disponible sur Internet :

<[http://www.adbs.fr/uploads/docsi/1209\\_fr.pdf](http://www.adbs.fr/uploads/docsi/1209_fr.pdf)>

**GOUYET, Jean-Noël, GERVAIS, Jean-François.** *Gestion des médias numériques digital media asset management*. Paris : Dunod ; [Bry-sur-Marne] : INA, impr. 2006. 328 p. (Collection Audio-photo-vidéo, 1769-597X)  
ISBN 2-10-049738-3

**KENNEY, Anne R., RIEGER, Oya Y., ENTLICH, Richard.** *De la théorie à la pratique* : didacticiel d'imagerie numérique de l'Université de Cornell. [Ithaca]: Bibliothèque de l'Université Cornell, Département de Recherches, cop. 2000. Mise à jour le 21-03-2003. [consulté le 24 mars 2007]. Disponible sur Internet :

<<http://www.library.cornell.edu/preservation/tutorial-french/contents.html>>

**PHAN Françoise.** *De la numérisation à la production, au stockage et à la diffusion de documents numériques à l'Université de L'Artois* : Projet personnel de bibliothécaire. Villeurbanne : École nationale supérieure des sciences de l'information et des bibliothèques, 2002, 44 p.

**PRAX, Jean-Yves, LARCHER, Simon.** *La gestion électronique documentaire.* Paris : Dunod, 2004. 341 p.  
ISBN 2-10-007891-7

**ROCH, Marie-Ève.** *Dossier : La numérisation* [en ligne]. Montréal : Université de Montréal, Association internationale francophone des bibliothécaires documentalistes, 2006. Mise à jour le 26-02-2007. [consulté le 2 avril 2007]. Disponible sur Internet : <[http://bibliodoc.francophonie.org/article.php3?id\\_article=197](http://bibliodoc.francophonie.org/article.php3?id_article=197)>

**ROSSI, Christian.** *De la diffusion à la conservation des documents numériques* [en ligne]. [S.l.] : [s.n.], 2005. Mise à jour le 17-03-2005. [consulté le 4 avril 2007]. Disponible sur Internet : <[http://archivesic.ccsd.cnrs.fr/docs/00/06/25/89/PDF/sic\\_00001379.pdf](http://archivesic.ccsd.cnrs.fr/docs/00/06/25/89/PDF/sic_00001379.pdf)>

**SAVARD, Réjean.** *Le numérique : impact sur le cycle de vie du document* : actes du colloque organisé à l'Université de Montréal par l'EBSI et l'ENSSIB du 13 au 15 octobre 2004 [en ligne]. Montréal : Ecole de bibliothéconomie et des sciences de l'information, Université de Montréal, 2007. [consulté le 3 avril 2007]. Disponible sur Internet : <<http://babel.enssib.fr/sommaire.php?id=612>>

#### **4. Pérennisation des documents électroniques**

**ARCHIVES NATIONALES (FRANCE).** *Bulletin des Archives de France sur l'archivage à long terme des documents électroniques* [en ligne]. Paris : Direction des Archives de France, 2000- .[consulté le 3 avril 2007]. Disponible sur Internet : <<http://www.archivesdefrance.culture.gouv.fr/fr/publications/DAFbulelectronique.html>>

**ASSOCIATION ARISTOTE, CENTRE SPATIAL DE TOULOUSE AU CENTRE NATIONAL D'ETUDES SPATIALES (CNES).** *Serveur du groupe PIN : Pérennisation des Informations Numériques* [en ligne]. Toulouse : CNES, 2004. Mise à jour le 23-03-2007. [consulté le 2 avril 2007]. Disponible sur Internet :  
<<http://vds.cnes.fr/pin/>>

**BIBLIOTHEQUE NATIONALE D'AUSTRALIE.** *Directives pour la préservation du patrimoine numérique* [en ligne]. [S.l.] : UNESCO, 2003. [consulté le 3 avril 2007]. Disponible sur Internet :  
< <http://unesdoc.unesco.org/images/0013/001300/130071f.pdf>>

**BADOLATO, Anne-Marie.** *Gestion de la conservation de collections numériques* [en ligne]. [Ithaca, NY] : Bibliothèque de l'Université Cornell, cop. 2003. Mise à jour le 22-12-2006. [consulté le 2 avril 2007]. Disponible sur Internet :  
<<http://www.library.cornell.edu/iris/tutorial/dpm-french/index.html>>

**CALDERAN, Lisette, HIDOINE, Bernard, MILLET, Jacques, INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE.** *Pérenniser le document numérique : Séminaire INRIA, 2-6 octobre 2006, Amboise.* Paris : ADBS, 2006. 206 p.  
ISBN 2-84365-087-9

**LUPOVICI, Catherine.** Les stratégies de gestion et de conservation préventive des documents numériques. *Bulletin des bibliothèques de France*, 2000, Volume 45, n° 4, p.43-54 [en ligne]. [consulté le 1er avril 2007]. Disponible sur Internet :  
<<http://bbf.enssib.fr/sdx/BBF/frontoffice/2000/04/document.xsp?id=bbf-2000-04-0043-004/2000/04/fam-dossier/dossier&statutMaitre=non&statutFils=non>>

**MAHE, Annaïg.** Pérenniser le document numérique : Séminaire Inria 2006, *Bulletin des bibliothèques de France*, 2007, n° 2, p. 102-103 [en ligne]. Disponible sur Internet :  
<<http://bbf.enssib.fr>> [consulté le 3 avril 2007].

## 5. Recommandations, normes, standards

**ASSOCIATION FRANCAISE DE NORMALISATION.** *Archivage électronique : Spécifications relatives à la conception et à l'exploitation de systèmes informatiques en vue d'assurer la conservation et l'intégrité des documents stockés dans ces systèmes (Z 42-013)* [en ligne]. Paris : AFNOR, 2001. [consulté le 2 avril 2007]. Résumé, descripteurs et sommaire disponibles sur Internet :

<[http://www.boutique.afnor.org/NEL5DetailNormeEnLigne.aspx?&nivCtx=NELZNELZ1A10A101A107&ts=1402112&CLE\\_ART=FA118461](http://www.boutique.afnor.org/NEL5DetailNormeEnLigne.aspx?&nivCtx=NELZNELZ1A10A101A107&ts=1402112&CLE_ART=FA118461)>

**COMITE CONSULTATIF POUR LES SYSTEMES DE DONNEES SPATIALES (CCSDS).** *Modèle de référence pour un Système ouvert d'archivage d'information (OAIS)* [en ligne]. Washington : Comité Consultatif pour les Systèmes de Données Spatiales, 2005. [consulté le 3 avril 2007]. Disponible sur Internet :

< [http://vds.cnes.fr/pin/documents/projet\\_norme\\_oais\\_version\\_francaise.pdf](http://vds.cnes.fr/pin/documents/projet_norme_oais_version_francaise.pdf)>

**CHABIN, Marie-Anne.** *Specifications MoReq (Model requirements for the management of electronic records)* : modèle d'exigences pour l'organisation de l'archivage électronique [en ligne]. Paris : Archive 17, 2004. 132 p.[consulté le 3 avril 2007]. Disponible sur Internet :

< [http://www.archivistes.org/IMG/MoReq\\_francais.pdf](http://www.archivistes.org/IMG/MoReq_francais.pdf) >

**DHERENT, Catherine.** *Les archives électroniques : Manuel pratique* [en ligne]. Paris : Direction des Archives de France, 2002.[consulté le 3 avril 2007]. Disponible sur Internet :

<<http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/DAFmanuel%20version%207.html>>

**FOULONNEAU, Muriel, BONNAMY, Alexandra, UNIVERSITE DE BATH.** *Recommandations techniques pour les programmes de création de contenus culturels numériques* (édition française du projet Minerva) [en ligne]. [Paris] : Relais Culture Europe, 2004. [consulté le 12 juillet 2007]. Disponible sur Internet :

<[http://www.minervaeurope.org/structure/workinggroups/servprov/documents/techguid1\\_0-f.pdf](http://www.minervaeurope.org/structure/workinggroups/servprov/documents/techguid1_0-f.pdf)>

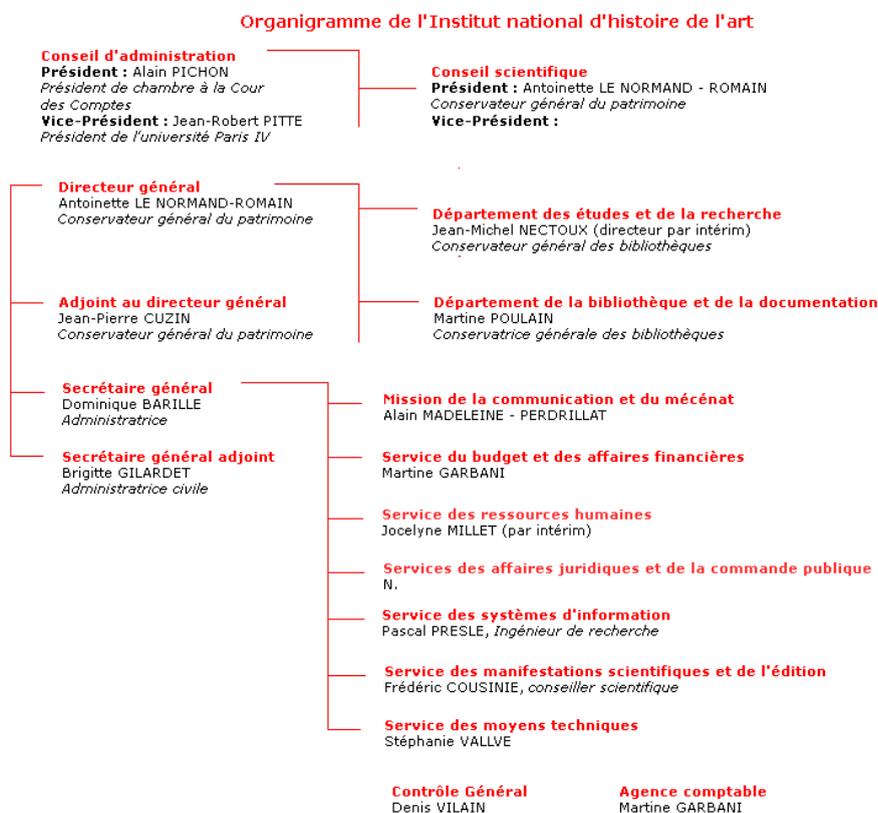
**HILLMANN, Diane, TEASDALE, Guy.** *Initiative de métadonnées du Dublin Core : Guide d'utilisation version française* [en ligne]. Québec : Bibliothèque de l'Université de Laval, 2001. [consulté le 4 avril 2007]. Disponible sur Internet :  
<<http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>>

**TAYEB, Martine, SEVIGNY, Martin, DELVALLEE, Anne.** *Numérisation du patrimoine culturel* [en ligne]. [Paris] : Comité scientifique pour la documentation informatisée & multimédia du Conseil ministériel de la recherche, Ministère de la culture et de la communication, cop. 1998.[consulté le 1<sup>er</sup> avril 2007]. Disponible sur Internet :  
<<http://www.culture.gouv.fr/culture/mrt/numerisation/>>

## ***Table des annexes***

<b>ANNEXE 1 : L'ORGANIGRAMME DE L'INHA .....</b>	<b>71</b>
<b>ANNEXE 2 : NOMBRE DE VUES NUMÉRISÉES .....</b>	<b>72</b>
<b>ANNEXE 3 : LISTE DES RUBRIQUES ET FONDS DE LA BIBLIOTHÈQUE NUMÉRIQUE .....</b>	<b>73</b>
<b>ANNEXE 4 : RAPPEL SUR LE DROIT D'AUTEUR APPLIQUÉ À LA NUMÉRISATION .....</b>	<b>75</b>
<b>ANNEXE 5 : UNE FICHE DE TRAVAIL .....</b>	<b>78</b>
<b>ANNEXE 6: UN CONSTAT D'ÉTAT .....</b>	<b>79</b>
<b>ANNEXE 7 : L'ATELIER DE NUMÉRISATION EN INTERNE .....</b>	<b>80</b>
<b>ANNEXE 8 : UN PDF AVEC ONGLETS .....</b>	<b>81</b>
<b>ANNEXE 9 : LE MODULE D'ADMINISTRATION DE PICTURELAN .....</b>	<b>82</b>
<b>ANNEXE 10 : NOMBRE DE VUES CHARGÉES .....</b>	<b>83</b>
<b>ANNEXE 11 : LES 15 ÉLÉMENTS DU DUBLIN CORE .....</b>	<b>86</b>
<b>ANNEXE 12 : LES FORMATS .....</b>	<b>88</b>

## Annexe 1 : L'organigramme de l'INHA



## Annexe 2 : Nombre de vues numérisées

Nombre d'images numérisées		
<b>par établissement</b>	ENSBA	12232
	BCMNI	81231
	INHA	145693
	<b>total</b>	<b>239156</b>
<b>par année</b>	2002	23825
	2003	22523
	2004/2005	168664
	2006	24144
	<b>total</b>	<b>239156</b>
<b>année</b>	<b>établissement</b>	<b>nb vues</b>
2002	ENSBA	12232
	BCMNI	11593
2003	INHA	18981
	BCMNI	3542
2004/2005	INHA	120170
	BCMNI	48494
2006	INHA	6542
	BCMNI	17602
	<b>total</b>	<b>239156</b>

## ***Annexe 3 : Liste des rubriques et fonds de la bibliothèque numérique***

- Archives, Manuscrits et autographes
  - Carnets de voyages et de notes de Maurice-Adolphe Linant de Bellefonds (1800-1883)
  - Catalogue des Tableaux du cabinet du Roi Placés dans l'Hôtel de la Sur-Intendance à Versailles fait en l'Année 1784 (...)
  - Collection de lettres autographes provenant du legs Robert Le Masle (1935-1971)
  - Collection d'autographes d'artistes du XIXème et de la première moitié du XXème siècle rassemblés par Albert S. Henraux (1881-1953)
  - Correspondance d'Alfred Bruyas (1821-1877)
  - Documentation, Notes, dessins d'Alexandre-Charles Sauvageot (1781-1860)
  - Ensemble de correspondances adressées à Léonce Bénédict (1856-1925)
  - Fonds Jean-Louis-Ernest Meissonier (1815-1891)
  - Inventaire des Tableaux du Cabinet du Roi, placés à la Sur-Intendance des Batimens de Sa Majesté à Versailles. Fait en l'Année 1784 (...)
  - Journal et correspondance d'Eugène Delacroix (1798-1863)
  - Trente cinq documents autographes, principalement d'artistes du XIXe siècle
  - Le voyage en Italie de François Debret (1777- 1850)
  - Vingt-six lettres autographes de peintres et de sculpteurs du XIXe siècle
- Dessins
  - Album de caricatures de Charles Garnier (1825-1898)
  - Dessins aquarellés de l'architecte Jean-Charles Geslin (1814-1885)
  - Dessins de la Galerie du palais royal (1786-1791)
- Estampes
  - Cabinet d'estampes Jacques Doucet. Estampes anciennes
  - Cabinet d'estampes Jacques Doucet. Estampes modernes
  - Recueils d'estampes représentant des portraits d'artistes (XVI-XIXe siècle)
  - Vues de villes, estampes du XVII au XIXe siècle
- Livres
  - Catalogues du Musée du Louvre publiés avant 1920 et déposés aux archives des musées nationaux
  - Classiques de l'Histoire de l'art. Livres imprimés en français ou publiés en France du XVIe- XIXe siècle

- Photographies
  - Fonds photographique Eugène Atget (1857-1927)
  - Fonds photographique Félix Bonfils (1831-1885)
  - Fonds photographique Collinet-Guérin (années 1910)
  - Fonds photographique de Hyacinthe -César Delmaet (1828-1862) et Louis - Emile Durandelle (1839-1917)
  - Fonds photographique de Jean Laurent (1816-1892)
  - Fonds photographique de G. Lékégian (18..-19..)
  - Fonds photographique de Séraphin Médéric Mieusement (1840-1905)
  - Fonds photographique des frères Neurdein (Etienne (1832-après 1915 ; Louis Antonin (1846-après 1915))
  - Fonds photographique de J. Pascal Sebah (1838-1910)
  
- Plans, dessins et relevés d'architecture
  - Albums de dessins d'architecture de Guillaume Abel Blouet (1795- 1853)
  - Albums de dessins d'architecture de Félix Duban (1798-1870)
  - Albums de dessins d'architecture de Lesueur (1794- 1883)
  - Album de Mandar (1757-1844)
  - Carnets Hittorff (1792-1867)
  - Dessins de Alexandre-Dominique Denuelle (1818-1880), Louis Hippolyte Lebas (1782-1867), Pierre-Joseph Garrez (1802-1852), et Charles Garnier(1825-1898)
  - Le Tombeau des Volumni, par Louis-Hippolyte Lebas (1782- 1867)

## ***Annexe 4 : Rappel sur le droit d'auteur appliqué à la numérisation***

### **Rappel sur le droit d'auteur appliqué à la numérisation**

#### **I. Principes généraux**

Dans le cadre de la définition de la propriété littéraire et artistique, l'auteur dispose de deux types de droits sur son œuvre :

##### **1. le droit moral**

Il concerne notamment :

- la paternité de l'œuvre : le nom de l'auteur doit être mentionné pour toute publication ou communication de l'œuvre (au minimum, le nom de l'auteur ; la forme usuelle est : nom, et si possible, prénom de l'auteur ; titre de l'œuvre ; si possible, date de création de l'œuvre ; lieu de conservation),

- le respect de l'œuvre : aucune modification portant atteinte à l'intégrité tant physique que morale de l'œuvre ne doit intervenir du fait de la numérisation et de la diffusion sous forme numérique,

- la divulgation de l'œuvre : l'auteur peut seul décider de faire connaître ou non son œuvre au public.

Le droit moral, étant perpétuel, inaliénable et imprescriptible, doit être respecté dans tous les cas, y compris pour une œuvre tombée dans le domaine public.

##### **2. les droits patrimoniaux**

La numérisation est concernée par le droit d'exploitation, qui comprend le droit de représentation et le droit de reproduction.

En vertu de ces droits, l'auteur dispose du droit exclusif d'autoriser ou d'interdire toute forme de reproduction et de représentation de son œuvre, ainsi que du droit de fixer les modalités financières de cette exploitation, et ce, que la finalité de l'exploitation soit commerciale ou non. Toute utilisation d'une œuvre sans autorisation de son auteur constitue une contrefaçon et est civilement et pénalement sanctionnée.

Les opérations de numérisation relèvent des deux types de droits :

- 1) par le droit de reproduction (transfert des données sur support numérique, impression/téléchargement)  
En effet, la reproduction consiste dans la fixation matérielle de l'œuvre par tous les procédés qui permettent de la communiquer au public de manière indirecte (Code de la propriété intellectuelle, art. L 122-3)
- 2) par le droit de représentation (diffusion sur le réseau internet - en interne et en externe)  
En effet, cette notion s'applique à la communication de l'œuvre au public par un procédé quelconque (CPI art L 122-2), dont la transmission sous forme numérique.

Pour toute opération de numérisation, il faut donc solliciter l'autorisation de l'auteur ou, après son décès, de ses héritiers (l'accord de tous est nécessaire en cas d'indivision).

La durée de la protection s'étend durant la vie de l'auteur et sur une période de 70 ans à compter du 1er janvier de l'année qui suit son décès (loi de 1997 transposant en droit français la directive européenne de 1993) ; on y ajoute (par prudence car la doctrine juridique à ce sujet n'est pas fixée) des prorogations de 14 ans et 262 jours pour les années correspondant aux deux guerres mondiales. Sur ces bases, la BnF propose dans un souci de simplification de considérer comme libres de droits les documents dont l'auteur est décédé avant 1919 (ou dont la date de parution est antérieure à 1870).

Voir les fiches techniques du ministère de la Culture

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/index.htm>

## **II. L'impact de la loi sur le droit d'auteur et les droits voisins dans la société de l'information sur l'exploitation des contenus numériques**

Voir Valette, Arnaud. *L'impact de la loi DADVSI sur l'exploitation des contenus numériques*. Babel - edit -, Droits d'auteur, loi DADVSI. ENSSIB - novembre 2006

[en ligne] <http://babel.enssib.fr/document.php?id=351>

## **III. Les procédures à suivre**

Le cadre dans lequel s'exerce aujourd'hui le droit d'auteur lié à la numérisation est un cadre jurisprudentiel, appuyé sur la réglementation du Code de la propriété intellectuelle.

Il n'existe pas de mode de gestion collective globale des droits liés à la numérisation (à la différence de ce qui existe pour la photocopie). Il est donc obligatoire, pour chaque établissement, de négocier les autorisations d'exploitation au cas par cas avec les auteurs ou leurs ayants droits. Ces autorisations peuvent être délivrées à titre onéreux ou gratuit.

La cession de droits est régie par le Code de la propriété intellectuelle. Elle doit être mentionnée par écrit et le domaine d'exploitation des droits cédés doit être précisé quant à :

- son étendue,
- sa destination,
- son lieu,
- sa durée (article L 131-3 du CPI),

et ce, que la cession s'accompagne ou non d'une rémunération.

Il est impératif de définir le plus complètement et le plus précisément possible les conditions d'exploitation des œuvres numérisées. Chaque usage ainsi que chaque droit nécessaire à cet usage doivent faire l'objet d'une cession explicitement mentionnée dans le contrat. La liste des œuvres concernées doit aussi être établie, afin de négocier précisément les droits attachés à chacune d'elles ; le Service juridique de la BnF, contacté par Réjane Quéméneur, a insisté sur l'importance de ce recensement préalable.

Remarque : dans le cas où les ayants droits n'auront pu être identifiés, on peut prévoir une provision forfaitaire constituant une « réserve de droits » qui pourront être acquittés au cas où un ayant droit se manifesterait ; la mention « droits réservés » est alors à apposer sur l'image.

Il est clair que ces procédures induisent de fortes contraintes de gestion. Pour cette raison, la Mission de la recherche et de la technologie du Ministère de la Culture conseille de privilégier les documents libres de droits (l'acquisition de droits n'est pas prise en compte pour la subvention accordée aux établissements qui participent au plan de numérisation).

Les établissements d'ailleurs entreprennent rarement de telles démarches. En particulier lorsque les volumétries sont importantes, ils se limitent aux documents libres de droits (BIUM ; BM de Lyon (estampes et affiches) ; Nordnum - SCD de Lille III, etc.)

Voir *Catalogue des fonds culturels numérisés mis en ligne par la MRT*  
[http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f\\_02.htm](http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_02.htm)

C'est le cas aussi de la BnF, qui souligne, dans le cadre de la numérisation massive d'imprimés dans le cadre de Gallica (76 000 textes), à la fois la difficulté financière et l'immensité des démarches à effectuer.

Il importe donc, avant de commencer la négociation des droits, de bien mesurer l'investissement nécessaire (en coût financier mais surtout en force de travail) et de l'évaluer au regard du bénéfice qui en sera retiré.

Françoise BERARD -

## Annexe 5 : Une fiche de travail

<b>COTE OUVRAGE:</b>		<b>8 I 295(3)</b>
<b>COTE NORMALISE:</b>		<b>0008I295_03</b>
		<b>Semaine prod</b>
		<b>Fiche N°84</b>
Type	II	Paramètres généraux en l'absence d'indications contraires
Titre court	Histoire de Paris et description de ses plus beaux monuments	
Ouverture	90	
Format	22 x 14,5 cm	
		300 dpi pages texte 400 dpi si illustration standard
Nb de fichiers estimés	455	Mono page cas général
Nb de pl. et depl.	20pl	
Nb de pages Numérisés	<b>OPERATEUR</b>	Double page si information continu sur deux pages en regard
Temps passé	<b>OPERATEUR</b>	
Date de Numérisation	<b>OPERATEUR</b>	JPEG qualité 12
Nb de pages blanches numérisées	<b>OPERATEUR</b>	Numériser les pages blanches sauf verso d'illustration sans annotation
Observations générales du BEE		
N°folio	Observations INHA	Observations Opérateur
ap le faux titre	planche	
ap p XII	planche	
ap p 36	planche	
ap p50	planche	

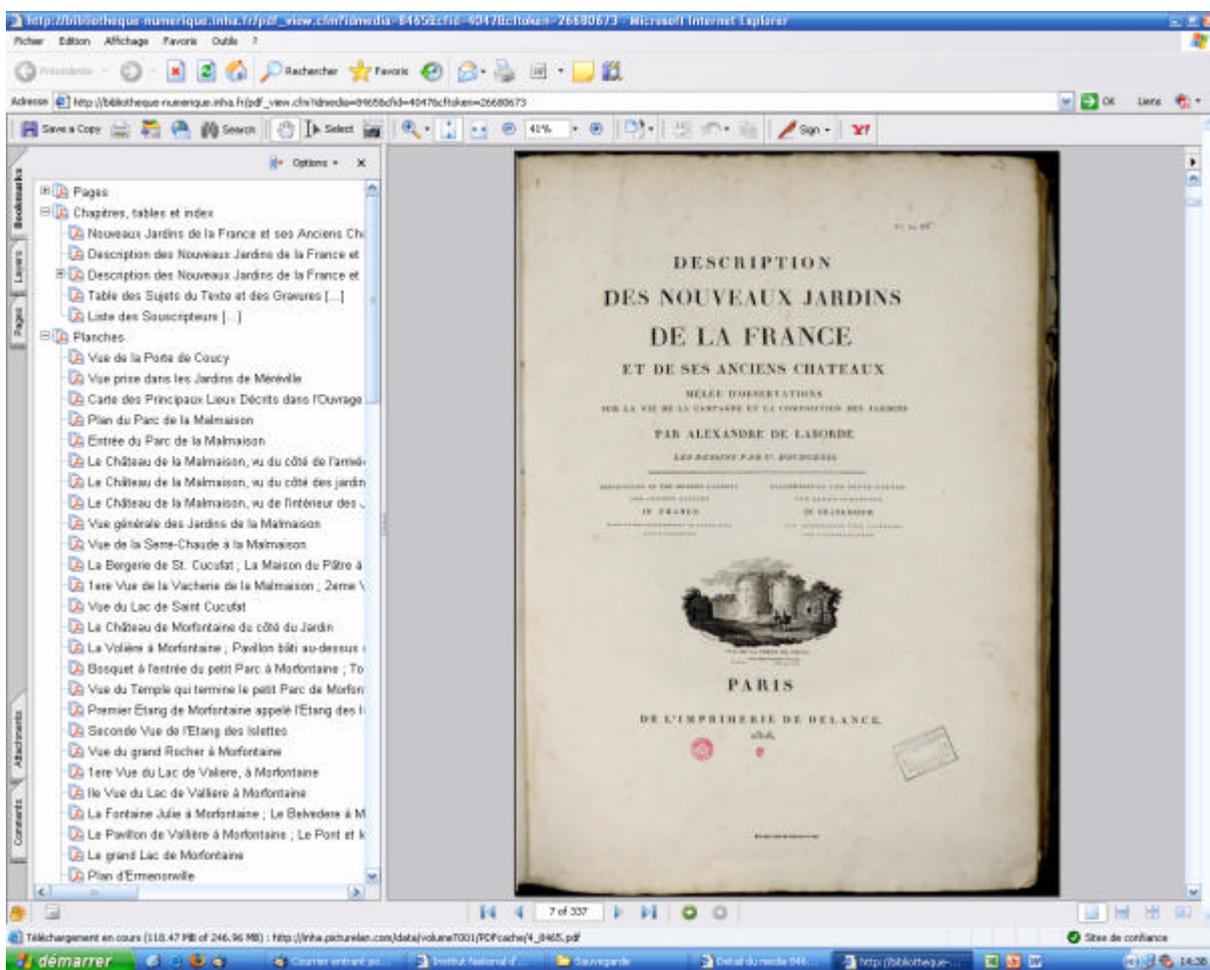
## Annexe 6: Un constat d'état

Cote	
Volumaison	
Titre, Auteur	
Adresse biblio, Date	
Collation	
Etat général du document	
Reliure	<input type="checkbox"/> Dos cassé <input type="checkbox"/> Articulation dos/plat fendu <input type="checkbox"/> Plat détaché  Notes :
Pages	<input type="checkbox"/> Pages arrachées : <input type="checkbox"/> Pages déchirées : <input type="checkbox"/> Pages tachées :  Notes :
Dépliants et doubles pages sur onglet	Nombre : Ouverture : Etat, Fragilité :  Notes :
Autres caractéristiques de l'exemplaire (illustrations couleur, cachet...):	

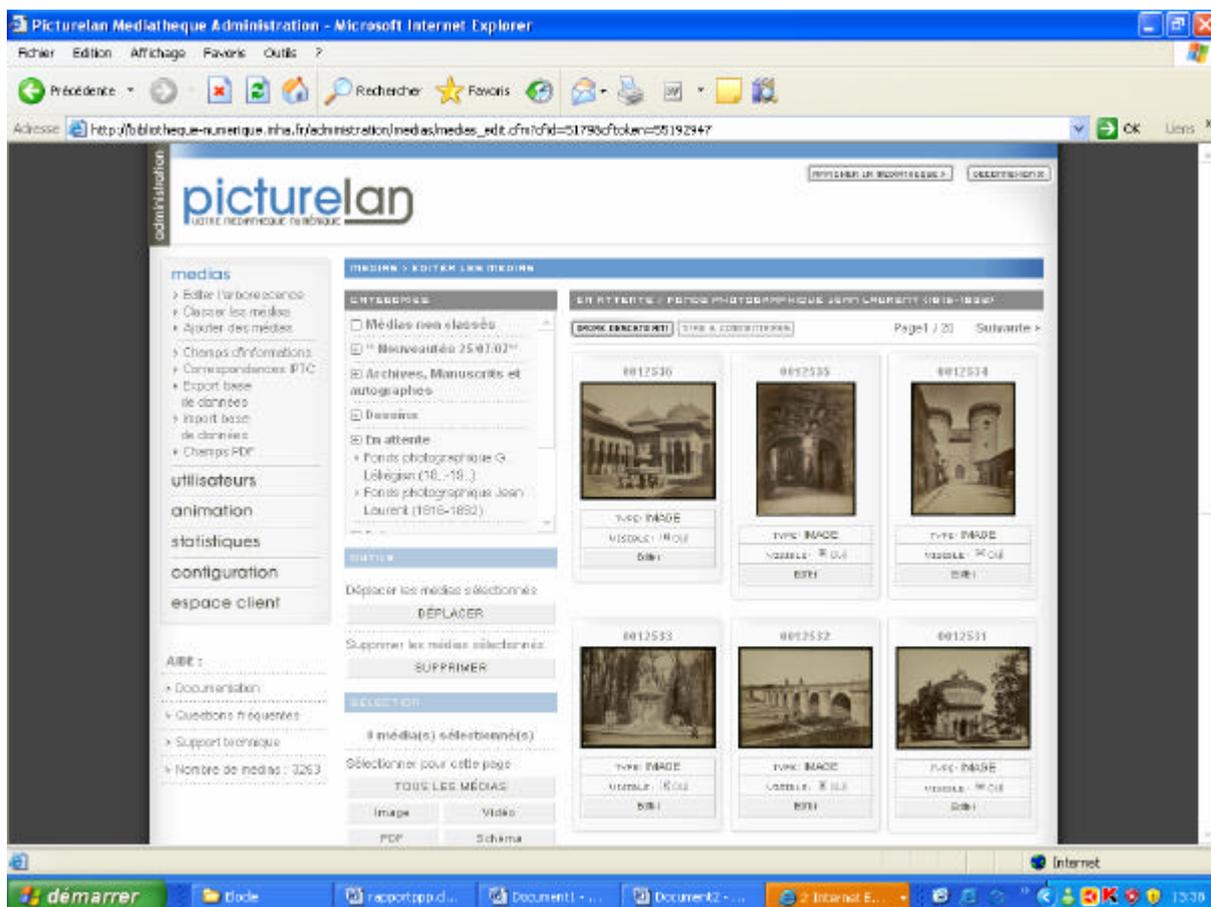
## ***Annexe 7 : L'atelier de numérisation en interne***



## Annexe 8 : Un Pdf avec onglets



## Annexe 9 : Le module d'administration de Picturelan



## Annexe 10 : Nombre de vues chargées

CATEGORIE	SOUS-CATEGORIE	NOMBRE DE DOCUMENTS CHARGES	DONT VOLUMES	NOMBRE DE VUES DES VOL.	NOMBRE DE VUES DES PIECES	TOTAL NOMBRE DE VUES
Archives, Manuscrits et autographes	Correspondance d'Alfred Bruyas (1821-1877)	6	6	1470	0	1470
Archives, Manuscrits et autographes	Inventaire des Tableaux du Cabinet du Roi, (...)	3	3	150	0	150
Archives, Manuscrits et autographes	Journal et correspondance d'Eugène Delacroix (1798-1863)	77	69	4336	8	4344
Archives, Manuscrits et autographes	35 documents autographes, principalement d'artistes du XIXe siècle	35	32	92	3	95
Archives, Manuscrits et autographes	26 lettres autographes de peintres et de sculpteurs du XIXe siècle	20	7	27	13	40
Archives, Manuscrits et autographes		3	3	112	0	112
<b>TOTAL Archives, Manuscrits et autographes</b>		<b>144</b>				<b>6211</b>
Dessins	Dessins aquarellés de l'architecte Jean-Charles Geslin (1814-1885)	2	2	399	0	399
Dessins	Dessins de la Galerie du palais royal (1786-1791)	6	6	308	0	308
<b>TOTAL Dessins</b>		<b>8</b>				<b>707</b>

Estampes	Cabinet d'estampes Jacques Doucet. Estampes anciennes	13	13	1256	0	1256
Estampes	Cabinet d'estampes Jacques Doucet. Estampes modernes	42	15	401	27	428
<b>TOTAL Estampes</b>		<b>55</b>				<b>1684</b>
Livres	Catalogues du Musée du Louvre publiés avant 1920 et déposés aux archives des musées nationaux	49	49	11669	0	11669
Livres	Classiques de l'Histoire de l'art. Archéologie	84	84	29431	0	29431
Livres	Classiques de l'Histoire de l'art.	22	22	6955	0	6955
Livres	Architecture Classiques de l'Histoire de l'art. Arts figuratifs	84	84	31975	0	31975
Livres	Classiques de l'Histoire de l'art. Collections	15	15	4213	0	4213
Livres	Classiques de l'Histoire de l'art. Esthétique	3	3	1245	0	1245
Livres	Classiques de l'Histoire de l'art. Guides de Paris	54	54	26276	0	26276
Livres	Classiques de l'Histoire de l'art. Jardins	23	23	2261	0	2261
Livres	Classiques de l'Histoire de l'art. Livres de Fêtes	33	33	4515	0	4515
<b>TOTAL Livres</b>		<b>367</b>				<b>118540</b>
Plans, dessins et relevés d'architecture	Carnets Hittorff (1792-1867)	1	1	60	0	60
Plans, dessins et relevés d'architecture	Dessins de Pierre-Léonard Laurécisque (1797-1860)	108	28	148	80	228
Plans, dessins et relevés d'architecture	Albums de dessins d'architecture de Guillaume Abel Blouet (1795-1853)	244	0	0	244	244
<b>TOTAL Plans, dessins et</b>		<b>353</b>				<b>532</b>

<b>relevés d'architecture</b>						
Photographies	Fonds photographique Félix Bonfils (1831-1885)	201	0	0	201	201
Photographies	Fonds photographique Collinet-Guérin (années 1910)	1864	0	0	1864	1864
<b>TOTAL Photographies</b>		<b>2065</b>				<b>2065</b>
<b>TOTAL</b>		<b>2992</b>	<b>552</b>	<b>127299</b>	<b>2440</b>	<b>129739</b>

## Annexe 11 : Les 15 éléments du Dublin Core

Les éléments du Dublin Core sont au nombre de 15 ; ils sont tous optionnels et répétables et peuvent apparaître dans n'importe quel ordre. On peut les classer selon trois rubriques : certains éléments concernent le contenu de la ressource, d'autres se rapportent à sa propriété intellectuelle. Enfin, le troisième type concerne la matérialisation de la ressource.

Contenu	Propriété intellectuelle	Matérialisation
Titre	Créateur	Date
Sujet	Editeur	Type
Description	Contributeur	Format
Source	Droits	Identifiant
Langage		
Relation		
Couverture		

- **le titre** : le nom donné à la ressource
- **le sujet** : le sujet du contenu de la ressource
- **la description** : une description du contenu de la ressource
- **la source** : une référence à une ressource à partir de laquelle la ressource actuelle a été dérivée
- **la langue** : la langue du contenu intellectuel de la ressource
- **la relation** : une référence à une autre ressource qui a un rapport avec cette ressource
- **la couverture** : la portée ou la couverture spatio-temporelle de la ressource
- **le créateur** : l'entité principalement responsable de la circulation du contenu de la ressource

- **l'éditeur** : l'entité responsable de la diffusion de la ressource
- **le contributeur** : une entité qui a contribué à la création du contenu de la ressource
- **les droits** : information sur les droits liés à la ressource
- **la date** : une date associée à un évènement dans le cycle de vie de la ressource (création, publication)
- **le type** : la nature ou le genre du contenu de la ressource
- **le format** : la matérialisation physique ou digitale de la ressource
- **l'identifiant** : une référence non ambiguë à la ressource dans un contexte donné

## Annexe 12 : Les formats

Tableau de [http://bibliodoc.francophonie.org/article.php3?id\\_article=209](http://bibliodoc.francophonie.org/article.php3?id_article=209)

*Numérisation : préservation du patrimoine numérique*

Format propriétaire	Élaboré par une compagnie privée et destiné à être utilisé avec les appareils ou les logiciels vendus par cette même entreprise et dont les modes de représentations ne sont pas publiées.	Formats issus de la suite bureautique Microsoft Office (antérieurs à 2007)
Format propriétaire public	Élaboré par une compagnie privée qui a en a publié les spécifications, permettant ainsi le développement de logiciels génériques.	<ul style="list-style-type: none"> <li>▶ RTF (Rich Text format), de Microsoft</li> <li>▶ PCD -Kodak Photo CD</li> </ul>
Norme de fait	Format propriétaire public reconnu de facto en raison de sa très grande popularité.	<ul style="list-style-type: none"> <li>▶ Gif (Graphic Interchange Format), de CompuServe</li> <li>▶ PDF (Portable Document Format), de Adobe</li> <li>▶ Tiff (Tagged Image File Format) de Adobe</li> </ul>
Norme officielle	Norme établie par un organisme officiel de normalisation au niveau national ou international qui ne dépend pas d'un logiciel en particulier pour être fonctionnelle et dont l'utilisation est libre de droits. L'Organisation internationale de normalisation (ISO) est l'organisme reconnu internationalement pour l'établissement de normes. Le World Wide Web Consortium (W3C), quant à lui, émet des recommandations visant une uniformisation des pratiques sur le Web.	<ul style="list-style-type: none"> <li>▶ Codes ASCII et UNICODE</li> <li>▶ Les langages balisés SGML, HTML, XML</li> <li>▶ Jpeg (Joint Photographic Expert Group)</li> </ul>