

#19

NUMÉRISER ET METTRE EN LIGNE

sous la direction de
thierry claerr et isabelle westeel

**NUMÉRISER ET
METTRE EN LIGNE**

LA BOÎTE À OUTILS

Les manuels de cette collection visent à fournir aux professionnels des ouvrages pratiques pour conduire des projets bibliothéconomiques d'actualité concernant aussi bien des bibliothèques publiques que des bibliothèques universitaires ou de recherche.

Écrit à plusieurs mains, chaque volume est coordonné par un ou plusieurs professionnels des bibliothèques.

Catherine Jackson

Directrice de collection

Déjà parus :

+++++

BAO #16

Gérer les périodiques

sous la direction de

Géraldine Barron (2008)

BAO # 17

Favoriser la réussite des étudiants

sous la direction de

Carine El Bekri-Dinoird (2009)

BAO # 18

Mettre en œuvre

un plan de classement

sous la direction de

Bertrand Calenge (2009)

À paraître :

+++++

BAO #20

Mettre en œuvre un service

de questions-réponses en ligne

sous la direction de

Claire NGuyen (2010)

+++++

PRESSES DE L'enssib

École nationale supérieure des sciences

de l'information et des bibliothèques

17-21 boulevard du 11 novembre 1918

69623 Villeurbanne CEDEX

Tél. 04 72 44 43 43 Fax 04 72 44 43 44

<<http://www.enssib.fr/presses>>

#19

NUMÉRISER ET METTRE EN LIGNE

sous la direction de
thierry claerr et isabelle westeel

Ont contribué à cet ouvrage :

Stéphane Alcandre

Thierry Claerr

Bernard Fages

François Houste

Patricia Le Galèze

Catherine Mocellin

Sophie Sepetjan

Isabelle Westeel

Nous remercions pour leurs relectures et leurs conseils avisés

Emmanuelle Bermès, Laurent Duploux, Marie-Elise Fréon, Nicolas Planckaert,

Jean-Pierre Sakoun et Arnaud Wael.

Numériser et mettre en ligne [Texte imprimé] / Sous la direction de Thierry Claerr et Isabelle Westeel. - Villeurbanne : Presses de l'enssib, cop. 2010. - 1 vol. (200 p.) ; 21 cm. - (La Boîte à outils ; 19). ISBN 978-2-910227-80-7. (br.) : 22 €

Dewey : 025.002 / 070.579 7

Rameau : Documentation de bibliothèque - - Numérisation

Bibliothèques virtuelles - - Guides, manuels, etc.

SOMMAIRE

MODE D'EMPLOI	p. 7
<i>par Thierry Claerr et Isabelle Westeel</i>	

PARTIE I : MAÎTRISER LES ASPECTS TECHNIQUES DE LA NUMÉRISATION

<i>par Catherine Mocellin</i>	p. 19
-------------------------------------	-------

ENCADRÉ : Choisir un numériseur <i>par Bernard Fages</i>	p. 40
--	-------

PARTIE II : RESPECTER LE DROIT DE LA PROPRIÉTÉ LITTÉRAIRE ET ARTISTIQUE

<i>par Sophie Sepetjan</i>	p. 45
----------------------------------	-------

PARTIE III : MONTER UN MARCHÉ DE NUMÉRISATION

<i>par Stéphane Alcandre</i>	p. 65
------------------------------------	-------

PARTIE IV : RÉDIGER UN CAHIER DES CHARGES DE NUMÉRISATION

<i>par Catherine Mocellin</i>	p. 89
-------------------------------------	-------

PARTIE V : INDEXER, STRUCTURER, ÉCHANGER : MÉTADONNÉES ET INTEROPÉRABILITÉ

<i>par Isabelle Westeel</i>	p. 99
-----------------------------------	-------

PARTIE VI : METTRE EN LIGNE DES DOCUMENTS NUMÉRISÉS, ÉVALUER, RÉFÉRENCER

<i>par François Houste</i> <i>et Isabelle Westeel</i>	p. 125
--	--------

PARTIE VII : CONSERVER LES DOCUMENTS NUMÉRISÉS

<i>par Patricia Le Galèze</i> <i>et Isabelle Westeel</i>	p. 157
---	--------

ENCADRÉ : Plateformes d'archivage d'établissements publics	p. 179
ENCADRÉ : Plateformes d'archivage privées	p. 180
ENCADRÉ : Plateformes d'archivage open source	p. 182

MÉMENTO

<i>par Thierry Claerr</i> <i>et Isabelle Westeel</i>	p. 183
---	--------

INDEX	p. 187
SIGLES ET ABRÉVIATIONS ...	p. 189
BIBLIOGRAPHIE	p. 192
Liste des illustrations .	p. 197
Liste des auteurs	p. 199

MODE D'EMPLOI

par Thierry Claerr et Isabelle Westeel

La numérisation avait déjà fait l'objet dans cette même collection, d'un ouvrage intitulé *Conduire un projet de numérisation*¹ publié en 2002, mais, compte tenu des évolutions importantes dans ce domaine, il nous a semblé important non seulement d'actualiser le propos mais surtout de lui apporter de nouvelles dimensions que ce soit en matière de métadonnées, de mise en ligne et de conservation des fichiers numérisés.

Même si cet ouvrage s'est voulu le plus large possible, en parlant autant des ouvrages patrimoniaux que des ouvrages contemporains, des documents sous droits que livres de droit, il n'abordera cependant pas la numérisation des collections vidéo, audio ou cinéma (cf. guide de numérisation sur le site du ministère de la Culture² et les formations assurées par l'INA³). Le souci permanent des auteurs a été de s'adresser à des établissements de toute taille, selon des logiques d'institution.

LES ENJEUX

+++++

La numérisation des données culturelles est un véritable enjeu pour la diffusion des connaissances et pour la démocratisation de la culture. Elle est aujourd'hui entrée dans les pratiques courantes des professionnels de la conservation et de la diffusion du patrimoine culturel. Sujet d'actualité, le numérique a un impact considérable sur le rôle des bibliothèques et sur l'évolution de leurs missions.

Au temps des pionniers, marqué par *The Gutenberg Project*⁴ en 1971, a succédé en France l'époque des expérimentations réussies avec dès le milieu des années 1990 : *Gallica*, la bibliothèque numérique de la BnF mise en ligne en 1997, la bibliothèque électronique de Lisieux ou encore les premières bibliothèques numériques patrimoniales comme *Medic@*, bibliothèque

1. *Conduire un projet de numérisation*, sous la direction de Charlette Buresi et Laure Cédelle-Joubert. Villeurbanne : Presses de l'enssib ; Paris : Éditions Tec&Doc, 2002 (La Boîte à outils #13).

2. [En ligne] < http://www.culture.gouv.fr/culture/mrt/numerisation/fr/f_04.htm > (consulté le 29 janvier 2010).

3. [En ligne] < <http://www.ina-sup.com/> > (consulté le 29 janvier 2010).

4. [En ligne] < <http://www.gutenberg.org> > (consulté le 9 février 2010).

numérique de la Bibliothèque interuniversitaire de médecine de Paris, ou le *CNUM*, bibliothèque numérique du Conservatoire national des arts et métiers. Nous sommes à présent entrés dans une phase de généralisation puisque plusieurs centaines d'institutions culturelles réparties sur tout le territoire sont impliquées aujourd'hui dans un plan de numérisation de leurs collections. Le rôle de producteur et de diffuseur des ressources numériques n'est plus dévolu aux seuls grands établissements engagés dans des programmes d'envergure. Depuis plusieurs années, de nombreuses institutions culturelles ont développé une offre numérique avec le soutien de leur collectivité et/ou de l'État.

L'initiative d'envergure lancée en décembre 2004 par Google a suscité des réactions qui ont permis de construire une offre européenne, *Europeana*, et de faire évoluer *Gallica*. Acteurs majeurs de ces politiques de diffusion et de valorisation par le numérique, les bibliothèques sont pleinement conscientes de leurs responsabilités et tiennent une place de choix au sein des acteurs : les institutions culturelles qui conservent les fonds, les utilisateurs (internauts et utilisateurs de livres numériques), les laboratoires de recherche, les éditeurs, les prestataires techniques et les industriels (numérisation et logiciels de gestion électronique de documents).

Souvent mis en avant, les enjeux des projets de numérisation sont multiples et doivent s'inscrire dans une politique générale d'établissement. Ils concernent :

- **la préservation et la conservation des collections :**

- > préserver des documents originaux fragiles ou fragilisés (parfois de grandes dimensions) en proposant un document de substitution sous une forme numérique, aisément manipulable et en favorisant ainsi des perspectives de conservation partagée ;
- > faire connaître des documents non communicables en raison de leur état matériel ou des difficultés de manutention.

- **la diffusion et la valorisation d'un corpus à des fins :**

- > culturelles : « tourisme culturel », valorisation du territoire, création de produits multimédia, de dossiers documentaires et d'expositions virtuelles ;

> scientifiques et pédagogiques :

diffusion de contenus comme matériau pour la recherche, en articulation avec les programmes de recherche et les nouveaux dispositifs en place dans l'Université : enseignement à distance (EAD), espaces numériques de travail (ENT), universités numériques régionales (UNR) et campus numériques ;

diffusion de la production scientifique résultant de la recherche (littérature grise, revues) : outil essentiel de visibilité et de coopération scientifique au niveau national et international, (par exemple le portail Persée ou encore les Thèses numériques-STAR).

• **l'amélioration des services rendus aux usagers :**

> en terme d'accès : accès à distance, accès simultanés ;

> en terme de consultation : lisibilité, reproduction et comparaison de documents ;

> en terme de recherche et d'appropriation : exploitation de corpus physiquement éloignés ; exploitation enrichie grâce à la combinaison d'un travail scientifique d'indexation et de structuration des documents ; développement d'outils logiciels pour optimiser la recherche et apporter une plus-value au document numérique lui-même (annotations...) ; mise en place d'espaces personnels ou collaboratifs de travail entre chercheurs (notion de parcours de lecture).

Les enjeux se trouvent désormais dans une offre de qualité adaptée à un contenu, à un service et à un public. Il s'agit de développer des contenus répondant aux attentes du public et de faire émerger de nouveaux types d'exploitation des documents numérisés. Ils doivent notamment tenir compte des nouveaux usages culturels et scientifiques liés à l'Internet et aux possibilités de coopération entre institutions.

LES PUBLICS ET LES USAGES

+++++

Il s'agit de savoir pourquoi et pour qui on numérise. Réfléchir aux publics-cibles et aux usages permet de travailler sur la question des contenus et des services.

TABEAU PUBLICS VISÉS ET SERVICES LIÉS

	Objectifs	Services et fonctionnalités	Contraintes et qualités requises
tout public	Communication autour des collections ; valorisation de l'institution ou de la collectivité ; réduction du fossé intergénérationnel vis-à-vis du numérique	Extension ludo-éducative des thématiques ; encyclopédie contributive ; espaces récréatifs ; parcours intergénérationnels...	Actualisation et renouvellement régulier du site, traitement multimédia et ludique des contenus ; opérations médiatiques régulières ; multimédia... Prise en compte de l'appropriation différente du numérique entre les générations.
population de la collectivité	Implication des habitants dans le développement des NTIC ; constitution d'une base d'informations dynamique, territorialisation	Encyclopédie contributive ; informations locales	Mise en place de dispositifs d'accompagnement des habitants ; simplicité et interactivité de l'interface d'accès ; multimédia...
enseignants et scolaires	Instaurer des liens avec l'ensemble des sites éducatifs et culturels ; institutionnaliser des liens avec l'Éducation nationale...	Ressources pédagogiques, approche critique diversifiée par thème...	Clarté et simplicité de l'interface ; affichage différencié par niveaux ; participation des enseignants à l'alimentation de la base, scénarisation pédagogique
érudits, amateurs et curieux	Susciter la curiosité vis-à-vis des collections, répondre aux attentes et associer éventuellement à l'indexation ou à l'éditorialisation des contenus	Ressources critiques, espaces de travail, services collaboratifs (annotations, indexation...)	Approche critique et érudite ; scientificité des données ; capacité à associer ce public
spécialistes, universitaires et étudiants	Fédérer le réseau des spécialistes, assurer l'identification et la reconnaissance institutionnelle ; mondialiser les collections ; crédibiliser... (nécessaire traçabilité et fiabilité des ressources)	Numérisation de haute qualité, indexation fine, ressources critiques, espaces de travail...	Scientificité et exhaustivité des données ; stabilité et mise à jour ; capacité à impliquer les chercheurs ; multilinguisme...
professionnels et administration de l'établissement	Gérer des collections, mettre au point des instruments de repérage et d'identification, disposer d'une photothèque...	Inventaire et traitement des collections ; gestion et consultation des documents...	Exhaustivité des données ; indexation ; enrichissement régulier et mise à jour...

Tableau établi à partir d'une étude menée en 2001 pour le Pôle Diderot de Langres.

LES PRINCIPAUX DÉFIS DE LA NUMÉRISATION ET DE LA MISE EN LIGNE

Numériser des documents et les mettre en ligne relèvent de défis multiples :

- **un défi technologique** : le projet doit être judicieusement évalué qu'il s'agisse de prestations, de choix techniques relatifs à l'achat de matériel ou au mode de numérisation ;
- **un défi économique** : une opération de numérisation implique l'engagement de dépenses qui dépasseront largement le simple budget lié à l'opération de numérisation. Le mode de financement doit être défini pour décider de la viabilité ou de la faisabilité de l'opération. Le projet peut également être construit en plusieurs étapes : court, moyen et long terme ;
- **un défi juridique** qui doit autant tenir compte de la législation en matière de propriété intellectuelle et artistique que du code des marchés publics. Il est indispensable de s'assurer d'emblée des droits attachés aux documents à numériser (documents libres de droits ou a contrario identification des ayants droit et prise en compte des bases de négociations avec ces derniers) ;
- **un défi professionnel et organisationnel** : il s'agit de savoir gérer le *back-office* (gestion de projet, place du bibliothécaire dans le projet...) et le *front-office* (valorisation éditoriale, formation des usagers, communication) ;
- **un défi scientifique** : trop souvent négligé, ce défi devrait constituer la première étape d'un projet (choix du corpus, complémentarité avec d'autres fonds numérisés...). Il nécessite de connaître les collections, de sélectionner celles qui seront prioritaires à la numérisation, de savoir élaborer un projet à géométrie variable à court et à long terme. Le chef de projet doit également inscrire le projet dans un paysage numérique foisonnant et, pour certaines thématiques, concurrentiel.

ÉTAT DES LIEUX

Le catalogue des collections numérisées *Patrimoine numérique*⁵ dresse un panorama de la numérisation du patrimoine culturel en France : il recense plus de 500 institutions culturelles impliquées dans plus de 1 400 opérations de numérisation. Ces chiffres attestent une activité certaine. Les projets en ligne et les véritables bibliothèques numériques (on entend par là des collections organisées) sont au nombre d'une cinquantaine : *Gallica*, la bibliothèque numérique de la BnF bien sûr, mais aussi les bibliothèques numériques de Bourg-en-Bresse, de Dole, de Grenoble, de Lille, de Lyon, du Mans, de Montpellier, d'Orléans, de Roubaix, de Rouen, de Saint-Étienne, de Toulouse, de Troyes, de Valenciennes et de Versailles pour les bibliothèques municipales ; le Conservatoire numérique des Arts et Métiers (*CNUM*), *Medic@* pour la Bibliothèque interuniversitaire de médecine (BIUM) de Paris, *Internum/e-corpus*, les *Bibliothèques virtuelles humanistes*, ainsi que les bibliothèques numériques de la bibliothèque interuniversitaire de Cujas, des Services communs de la documentation (SCD) de Poitiers et de Strasbourg notamment pour l'enseignement supérieur.

Depuis 2009, *Patrimoine numérique* est interconnecté avec le portail *Numes* < <http://www.numes.fr> >, inventaire en ligne des corpus numérisés et projets de numérisation des établissements et organismes d'enseignement et de recherche, qui a été développé sous la responsabilité de l'Agence bibliographique de l'enseignement supérieur (Abes) avec le soutien du Très grand équipement (TGE) Adonis.

5. Créé en 2001 par le ministère de la Culture et de la Communication, *Patrimoine numérique*, catalogue national des collections numérisées, a vocation à recenser et décrire les collections patrimoniales numérisées ou en cours de numérisation en France. Il a connu une évolution majeure en 2006 grâce à son intégration dans le portail européen MICHAEL (Multilingual Inventory of the Cultural Heritage in Europe – Inventaire multilingue du patrimoine culturel en Europe). [En ligne] < www.michael-culture.org > et < www.numerique.culture.fr > (consultés le 9 février 2010).

Ces deux outils ont vocation à assurer une couverture nationale de la numérisation, pour acquérir une meilleure connaissance de projets parfois confidentiels⁶, pour favoriser les partenariats entre institutions, pour permettre des échanges d'expérience sur les aspects techniques des projets et accroître la visibilité de ces projets auprès des utilisateurs.

Dans les petits et moyens établissements, construire et surtout maintenir dans la durée des projets de valorisation du patrimoine par le numérique est un travail compliqué. Aux difficultés liées au respect du calendrier budgétaire, s'ajoute la carence de l'offre logicielle dans le domaine de la mise en ligne des contenus numérisés. En France, quelques établissements complètent la mise en place de bases de données traditionnelles par des sites particuliers qui permettent d'éditorialiser les contenus : *Jean Carbonnier : 1908-2003* (Bibliothèque Interuniversitaire Cujas⁷), *La bibliothèque virtuelle sur les premiers socialismes* (BU Droit-Lettres Poitiers⁸), *La fleur au fusil* (Bibliothèque municipale de Roubaix), les expositions virtuelles de la région Rhône-Alpes présentées sur le portail *Lectura : La révolte et les rêves de Stendhal*, (Bibliothèque municipale de Grenoble), *Jules Massenet, morceaux choisis* (Médiathèque de Saint-Étienne), *Roger Vaillant, 1907-1965* (Médiathèque de Bourg-en-Bresse).

LES POLITIQUES DE NUMÉRISATION DES BIBLIOTHÈQUES EN FRANCE

+++++

En France, les récentes innovations industrielles, techniques et organisationnelles, stimulées par de grands projets (numérisation de grandes séries d'archives comme les registres paroissiaux et d'état civil ou les cadastres, programme de numérisation de masse de la Bibliothèque nationale de France) ont mis la numérisation à la portée de nombreuses institutions soucieuses de rendre accessibles et de valoriser leurs collections par les

6. Pour signaler les projets de numérisation dans *Patrimoine numérique*, rubrique « inscrire vos collections » dans l'espace professionnel du site.

7. [En ligne] < <http://expocujas.univ-paris1.fr/Carbonnier> > (consulté le 9 février 2010).

8. [En ligne] < <http://premierssocialismes.edel.univ-poitiers.fr/> > et < <http://lafleuraufusil.bn-r.fr/> > (consulté le 9 février 2010).

médias numériques. Si les initiatives locales se multiplient et ont trouvé leur place dans le paysage numérique, les réalisations restent encore relativement dispersées et souvent limitées à des corpus peu nombreux.

Depuis les années 1990, l'État a progressivement formalisé sa politique en matière de numérisation du patrimoine culturel à partir des objectifs stratégiques suivants :

- augmenter l'offre culturelle numérique pour répondre à l'attente des publics ;
- favoriser l'accès à des fonds et à des corpus documentaires, pour certains fragiles ou fragilisés (presse, audiovisuel analogique...) ;
- développer les ressources numériques pour des usages innovants ainsi que les partenariats et la constitution de réseaux de contenus (autour de thèmes ou de typologie de documents).

Depuis 1996, le Plan national de numérisation du ministère de la Culture et de la Communication favorise, sous forme d'appels à projets, les initiatives des acteurs culturels pour la numérisation et la mise en réseau de leurs fonds patrimoniaux. Ce plan a joué un rôle déterminant dans le paysage de la numérisation non seulement par les moyens engagés, mais aussi par la diffusion des recommandations du ministère sur les normes et les standards documentaires et par l'enrichissement du portail *Collections* du ministère de la Culture et de la Communication.

Les établissements publics nationaux, en particulier la BnF (avec sa bibliothèque numérique *Gallica*) ou l'Institut national de l'audiovisuel (INA), sont aussi des acteurs et des partenaires essentiels pour la politique nationale de numérisation.

Les collectivités territoriales participent largement à l'activité de numérisation et certaines en ont même été, il y a plus de dix ans, les promoteurs. Les politiques régionales de numérisation s'inscrivent parfois dans des contrats de plan État-Région (notamment Aquitaine, Lorraine et Franche-Comté), ainsi que dans le cadre du Plan d'action pour le patrimoine écrit.

La diffusion en ligne des ressources numérisées reste cependant encore trop souvent limitée, malgré la mise en place de portails régionaux de ressources numériques (ex. Banque numérique du savoir en Aquitaine) qui permettent de coordonner les projets et d'améliorer la visibilité et l'exploitation des documents numérisés.

La multiplication des initiatives rend d'autant plus urgente la nécessité d'une concertation accrue en matière de canalisation des énergies, de coordination des efforts (éviter les doublons et mutualiser les moyens) et de bonnes pratiques (recommandations techniques et conservation des documents numériques). Le rapport *Schéma numérique des bibliothèques* a été rendu en mars 2010 par Bruno Racine, président de la BnF au ministre de la Culture et de la Communication.

FINANCEMENTS⁹

+++++

En matière de financements, les institutions peuvent utiliser les dispositifs européens, le mécénat et les partenariats public/privé. Les programmes de numérisation sont également soutenus par des politiques particulières.

Voir tableau p. 16.

9. Établie en février 2010, cette présentation des financements pourra être amenée à évoluer dans le cadre de l'Emprunt national, la numérisation du patrimoine culturel faisant partie des actions à soutenir à hauteur de 750 millions d'euros.

DISPOSITIFS SPÉCIFIQUES DE FINANCEMENTS DE LA NUMÉRISATION

TABLEAU

pour les institutions relevant du ministère de l'enseignement supérieur et de la recherche

La politique contractuelle est le principal outil de financement de la numérisation :

- > opérations de numérisation des bibliothèques universitaires via les contrats des établissements ;
- > programmes de constitution de corpus scientifiques via les contrats des organismes de recherche.

Le financement au titre des très grandes infrastructures de recherche, par les appels à projets : Agence nationale de la Recherche (ANR) et Très grand équipement (TGE) Adonis.

Les programmes de numérisation concertée mis en place par le département coopération de la BnF (Sciences juridiques avec la Bibliothèque interuniversitaire Cujas, Histoire de l'art avec la bibliothèque de l'Institut national d'histoire de l'Art...).

Aides :

- > les contrats : passage au nouveau modèle d'allocation des moyens depuis janvier 2009 ;
- > les conventions pôles associés de la Bibliothèque nationale de France ;
- > la Bibliothèque Scientifique Numérique ;
- > l'appel aux opérateurs (Persée, Centre informatique national de l'enseignement supérieur-CINES, Agence bibliographique de l'enseignement supérieur-ABES).

pour les institutions relevant du ministère de la culture et de la communication

> Le plan national de numérisation du ministère

Fonctionnement par appel à projets.

Objectif : numériser les contenus culturels pour les diffuser gratuitement.

Financement : jusqu'à 50 %.

Depuis 2005, la valorisation et la mise en ligne des fonds numérisés sont devenues prioritaires.

Deux objectifs clairement affichés :

- > augmenter significativement l'offre de ressources numériques culturelles constituée d'ensembles cohérents et de taille critique ;
- > promouvoir une consultation libre et ouverte des ressources numériques culturelles pour tous les internautes.

Projets d'une durée d'un an maximum ; définition de six programmes thématiques ; priorité aux partenariats régionaux.

Pour les institutions relevant des collectivités territoriales

- > les financements locaux : municipalité, conseil général, région... ;
- > les contrats de projet État-Région (CPER) dans certaines régions ;
- > certains fonds régionaux d'acquisition des bibliothèques-FRAB (ex. Basse-Normandie et Midi-Pyrénées) ;
- > les crédits déconcentrés en région auprès de la direction régionale des affaires culturelles ;
- > les programmes de numérisation concertée de la Bibliothèque nationale de France.

Financement spécifique aux bibliothèques des collectivités territoriales

Le concours particulier « bibliothèques » au sein de la dotation générale de décentralisation (DGD) peut soutenir la numérisation (production et diffusion) selon deux axes :

- > aménagement d'un atelier et acquisition de matériels de numérisation ;
- > actions de numérisation, signalement et diffusion des collections.

Contact : conseiller livre et lecture à la direction régionale des affaires culturelles (DRAC).

PERSPECTIVES POUR UNE NUMÉRISATION DURABLE

+++++

Remis le 12 janvier 2010 par Marc Tessier au ministre de la Culture et de la Communication, le *rapport sur la numérisation du patrimoine écrit*¹⁰ a notamment rappelé l'importance de la diffusion des données numériques et le rôle d'*Europeana*¹¹, portail européen lancé en novembre 2008, pour faire émerger une approche européenne concertée sur la numérisation.

La numérisation et l'accessibilité en ligne des contenus culturels sont essentielles à la valorisation du patrimoine, au dynamisme de la création de contenus et à l'émergence de nouveaux services en ligne. Elles contribuent à la démocratisation de l'accès à la culture, au développement de la société de l'information et de l'économie de la connaissance et nécessitent :

- le développement de la coopération numérique et une meilleure articulation des politiques de numérisation à tous les niveaux ;
- la conservation à long terme des documents numériques ;
- l'amélioration de l'accès aux documents en ligne et la prise en compte des nouveaux usages culturels et scientifiques liés à l'Internet, ce qui implique pour les chefs de projets une expertise, une veille technologique et des formations appropriées¹².

10. Ministère de la Culture et de la Communication. [En ligne] < <http://www.culture.gouv.fr/mcc/Actualites/A-la-une/Mission-sur-la-numerisation-du-patrimoine-ecrit> > (consulté le 9 février 2010).

11. [En ligne] < www.europeana.eu > (consulté le 9 février 2010).

12. Une étude commandée en 2009 par la Bibliothèque Archives nationales du Québec (BANQ) et par la Société des musées québécois a montré que l'accès à l'expertise et à la formation constitue une priorité pour près de la moitié des professionnels ayant répondu à l'enquête. [En ligne] < <http://collections.banq.qc.ca/ark:/52327/bs1870360> > (consulté le 9 février 2010).

PARTIE I

MAÎTRISER LES ASPECTS TECHNIQUES DE LA NUMÉRISATION

MAÎTRISER LES ASPECTS TECHNIQUES DE LA NUMÉRISATION

par Catherine
Mocellin

LA NUMÉRISATION DES IMAGES FIXES (IMPRIMÉS, ICONOGRAPHIE)

+++++

La numérisation est la conversion d'une information présente sur un support physique en codage numérique en mode binaire (en 0 et 1) sur un ou plusieurs bits¹, lisible par un programme (logiciel, périphérique, etc.). Cette conversion aboutit à une image (numérisation en mode image, ou fac-similé) ou à une information textuelle : conversion en mode texte du mode image, automatiquement par un logiciel de reconnaissance optique des caractères (OCR) ou par saisie manuelle.

NUMÉRISATION EN MODE IMAGE

caractéristiques de l'image numérique

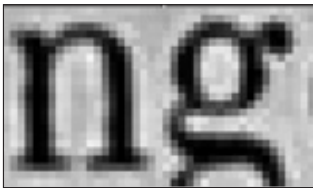
Dans le numériseur ou l'appareil photo, un convertisseur analogique-numérique comportant des capteurs de lumière opère la transformation de l'image analogique en sa version numérique. Selon les appareils, ces capteurs sont organisés en matrice ou de manière linéaire (voir encadré sur les numériseurs p. 40).

La lumière réfléchie par l'objet est exprimée par 0 (pas de lumière, pas de courant) ou 1 ; la luminosité de chaque pixel est donc codée par une valeur numérique. L'image résultante est de type matricielle ou *bitmap*², c'est-à-dire une juxtaposition de pixels³ en rangées et en colonnes dont la valeur est générée par les capteurs :

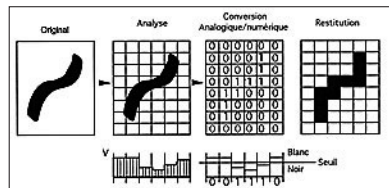
1. Un bit est la plus petite unité de traitement de l'information par un ordinateur. 8 bits font un octet. 1024 octets font un kilo-octet (Ko), 1024 Ko font un méga-octet (Mo) et ainsi de suite.

2. « Carte de bits ». Les images vectorielles, traduisant l'information par équations, ne sont pas utiles pour les projets des bibliothèques ou des services d'archives.

3. Pixel : *picture element*, plus petite unité d'image numérique.



Grille de pixels en niveaux de gris

Conversion du signal analogique en signal numérique⁴

La profondeur d'acquisition⁵ est le nombre de bits codant chaque pixel ; elle influence donc le poids du fichier. Voici les trois principaux modes d'encodage des couleurs : un codage sur 1 bit restitue une image **bitonale** (noire et blanche) avec 2 valeurs possibles pour chaque pixel, 0 et 1. Un codage de chaque pixel sur 8 bits restitue une image à 256 valeurs (2^8 niveaux de gris). Un codage sur 24 bits (3 octets) restitue une image à 16,7 millions (2^{24}) de couleurs (couleurs vraies⁶). Chaque octet contient la valeur d'une couche de couleur (rouge, vert, bleu) comprise entre 0 et 255. Ce codage à 24 bits multiplie par 3 le poids informatique du fichier par rapport à un codage en niveaux de gris. On obtient en général une image bitonale par la **binarisation**, ou réduction des valeurs de pixels en niveaux de gris ou en couleurs à 0 ou 1, à partir d'une valeur de référence fixée dans les gris. Les pixels plus sombres que la valeur du seuil seront mis en noir, les pixels plus clairs en blanc, ce qui implique un éclairage suffisant et un bon réglage du point blanc. Le seuil de binarisation doit être choisi avec précaution : s'il tend trop vers le noir, on risque la perte d'information (trop de pixels gris sont convertis en blanc) ; s'il tend trop vers le blanc, il y a risque d'illisibilité (trop de pixels gris deviennent noirs). La binarisation est pertinente sur les imprimés de bonne qualité, elle facilite la lecture et permet de diminuer le poids des fichiers à stocker. Cependant elle est complexe à définir surtout si le papier et/ou l'impression sont hétérogènes ; les documents trop tachés ou à trop faible

4. Extrait, avec l'autorisation des auteurs, de Gilles Kagan, Thierry Buquet, *Numérisation d'images : de l'acquisition à la publication*. Cours proposé par le service photographique de l'IRHT, Orléans, IRHT, 2002 (Ædilis, Publications pédagogiques, 1) < <http://www.irht.cnrs.fr/cours/acq/numerique.htm> > (consulté le 19 février 2010).

5. Profondeur d'acquisition ou dynamique de l'image.

6. On rappellera pour mémoire le codage en couleurs indexées, selon une palette de 256 valeurs de couleur, donc sur 8 bits. La représentation des couleurs est très pauvre mais cette méthode est économe en espace de stockage.

contraste ne pourront pas être binarisés sauf à devenir illisibles. La **profondeur d'acquisition** est donc fonction du support et de son contenu. Le niveau de gris est très adapté pour la presse, du fait de la finesse des caractères, de la fréquente hétérogénéité du papier et des illustrations :

TABLEAU **PROFONDEUR D'ACQUISITION RECOMMANDÉE
SELON LE SUPPORT**

profondeur d'acquisition	support
Couleurs	Tout ou majoritairement en couleur
Niveaux de gris	Documents en demi-teintes (lavis, fusain, photographies, dessins à détails très fins...) Imprimés contenant un grand nombre de photographies : presse par exemple Documents avec de fortes rousseurs, très tachés sur la zone imprimée ; faiblement contrastés ; à l'impression irrégulière (report d'encre, caractères abîmés...)
Noir et blanc	Imprimés courants ; schémas, tableaux ; dessins au trait

La définition et la résolution définissent l'image numérique quantitative-ment. La définition est le nombre total de pixels de l'image : nombre en abscisses multiplié par le nombre en ordonnées, par exemple $2\,048 \times 3\,072 = 6,29$ millions de pixels. La résolution est le nombre de pixels par unité de mesure, en général le pouce⁷. L'unité est le dpi (*dot per inch*) ou ppi (*pixel per inch*). « 300 dpi » signifie que chaque pouce en abscisses et en ordonnées comprend 300 pixels juxtaposés.

Après la capture, il est possible d'accroître la résolution par ajout artificiel de pixels de valeurs similaires aux pixels proches (rééchantillonnage ou interpolation). Cette création d'information artificielle est à éviter dans un projet de numérisation qui vise à la préservation à long terme de docu-

7. Inch, un pouce = 2,54 cm.

ments, on s'assurera donc que le prestataire ne l'a pas mis en œuvre à la production. On peut aussi diminuer le nombre de pixels par la compression (voir *infra*).

La résolution exprimant une fréquence d'échantillonnage, plus elle est haute, plus la finesse de détails sur l'image restituée sera importante et donc plus le poids du fichier non compressé sera lourd. Ce choix est donc à faire soigneusement, en préservant l'avenir, et selon les caractéristiques de l'original et les modalités de restitution : plus le document est petit, plus la résolution doit être élevée pour capturer les détails et permettre des agrandissements. Pour les documents de très grand format (> A1) on pourra prendre plusieurs vues de détail à une résolution standard (300 dpi) et une vue globale à une résolution plus faible (150 dpi). 300 dpi permet la substitution sur écran et l'impression à l'échelle originale, puisque les écrans ont en général une résolution entre 72 et 96 dpi et les imprimantes standard une résolution d'impression minimale de 150 lpi (lignes par pouce). Une image inférieure à 150 dpi ne pourra donc pas être imprimée avec qualité mais sera lisible sur écran.

Le poids de l'image est déterminé par la résolution et la profondeur d'acquisition et se distingue du poids du fichier définitif, lié au format et à une éventuelle compression. Ce poids informatique, en octets, est calculé avec la formule :

définition * profondeur d'acquisition/8.

On divisera par 1 024 pour avoir le poids en kilo-octet. Pour un ouvrage de 100 pages A5 (15*21 cm) en noir et blanc en 300 dpi non compressé, on calculera le poids en Ko en exprimant la définition d'une image en cm (division par 2,54), puis en divisant par 8 puis par 1 024 et en multipliant le tout par 100. Le poids final est de 53 640 Ko soit environ 52,38 Mo⁸.

8. Détail du calcul : $100 * (15/2.54 * 300 * 21/2.54 * 300 * 1/8 / 1024) = 53\,640\text{ Ko}$, ~ 52,38 Mo

La résolution sera donc choisie selon la taille de l'original, l'utilisation prévue, les capacités du système d'archivage et ses évolutions. Pour une conservation à long terme, les résolutions suivantes par type de document, issues des règles de la BnF⁹, sont proposées :

TABLEAU **RÉSOLUTIONS PRÉCONISÉES SELON LE TYPE ET LE FORMAT DES ORIGINAUX**

support	cas général	exception	remarques
<i>Originaux opaques</i> > = A6 (imprimés, manuscrits, calques...)	300 ou 400 dpi	600 dpi	600 dpi si le document présente des variations importantes de format, d'informations (très petits caractères par exemple)
<i>Originaux opaques</i> < A6 (cartes postales, médailles et monnaies...)	600 dpi		
<i>Reproductions transparentes :</i> (ektas, diapositives, microformes...)	300 ou 400 dpi	600 dpi	Pour une échelle 1 de l'original. Mêmes recommandations que pour les opaques
<i>Documents transparents :</i> Originaux (plaques de verre...) Reproductions dont l'échelle de réduction est inconnue	Test préalable pour définir la résolution adaptée Règle : 1 000 points au cm		Cette recommandation s'appuie sur un agrandissement de 8 d'un négatif photo. Sa limite sera la capacité des numériseurs et les capacités d'archivage

9. Détail des choix techniques de la BnF : Bibliothèque nationale de France.

[En ligne] < <http://www.bnf.fr/fr/professionnels/numeriser.html> > (consulté le 19 février 2010).

Formats et compression

Un format de fichier est une manière de coder les données. Il peut être propriétaire ou non, éventuellement composé de sous-ensembles. Selon leurs caractéristiques, certains formats sont plus adaptés à la numérisation que d'autres. Le format natif généré par l'appareil, propre au constructeur, est dénommé par le terme générique RAW (« brut »)¹⁰. C'est en quelque sorte le négatif de l'image : ce fichier contient les informations relatives aux conditions de prise de vue et aux réglages. Il permet d'évaluer la qualité d'une photo avant traitement mais il est consultable à l'aide d'un convertisseur spécifique et ce n'est pas un format d'archivage. Il n'est donc pas très utile de conserver les fichiers RAW. Adobe® a défini un format RAW ouvert et le plus universel possible, le DNG, dont l'utilisation n'est cependant pas encore généralisée.

Les caractéristiques des formats et de leurs sous-ensembles étant variables, le poids du fichier dans un format donné sera différent, *a fortiori* si l'image est compressée.

La compression consiste à réduire le nombre de pixels de l'image, avec ou sans perte d'informations. Le quotient ou le taux¹¹ de compression doit être choisi pour fournir le meilleur équilibre entre la qualité d'image et l'économie en poids finales. Le mode avec perte est irréversible : à la réouverture de l'image, les données compressées ne sont pas restituées. C'est le cas de la compression JPEG. Plus le taux est fort, plus l'image sera « pixelisée » et présentera des artefacts, mais moins elle sera lourde. Un taux de compression de 75-80 % donne généralement le meilleur ratio qualité/taille mais même à une qualité de 100 % la compression JPEG détériore l'image¹². Si l'image est destinée à être réutilisée et archivée, il est préférable de choisir la compression sans perte consistant généralement à factoriser les pixels identiques et à mémoriser leur emplacement dans l'image pour les restituer à leur place à la décompression. Parmi les algo-

10. Exemples de fichiers RAW : CR2, NEF, CRW...

11. Quotient de compression = taille initiale de l'image/taille finale ; plus la compression est forte, plus le quotient est élevé. Taux = taille finale/taille initiale *100 (s'exprime en %). À un taux de 10 %, l'image finale est 10 fois moins importante que l'image initiale.

12. Voir le Mémento établi par la BnF début 2009 : Bibliothèque nationale de France. [En ligne] < http://www.bnf.fr/fr/professionnels/num_metadonnees/s.num_metadonnees_formats.html > (consulté le 19 février 2010).

rithmes sans perte, celui du groupe 4 de l'UIT (Union Internationale des Télécommunications) défini d'abord pour les fax est très intéressant, mais valable uniquement pour les images bitonales. La compression **LZW**¹³ procède par constitution de dictionnaires de données récurrentes auxquels un code est affecté. Lors de la compression, c'est ce code qui est attribué pour les données déjà connues dans un dictionnaire. **JPEG 2000** est un nouvel algorithme de compression avec ou sans perte plus puissant que le JPEG, auquel est associé un format. Sans entrer dans les détails on précisera que la méthode permet, à partir de l'image initiale, la création d'images de résolutions et de tailles différentes (« tuiles »). C'est pourquoi il est aussi considéré comme un format de diffusion, à ceci près qu'il faut un plug-in pour le consulter en ligne. Notamment grâce à sa compression sans perte, JPEG 2000 est un format prometteur dont l'utilisation est à l'étude à la BnF, et déjà actée par de grandes bibliothèques étrangères (BN de Norvège, des Pays-Bas, LoC...) pour l'archivage à long terme et la diffusion¹⁴. Il permet de diminuer fortement les espaces de stockage nécessaires à l'archivage. Le format PNG a été créé en 1995 afin de remplacer le GIF et de repousser les limites du JPEG. Il supporte jusqu'à 16 bits par couche de couleur et sa compression est non destructive. Indépendant des plateformes et des systèmes d'exploitation, ce format ouvert possède de nombreuses fonctionnalités mais il est encore peu répandu. Enfin le TIFF est un format très répandu et préconisé pour l'archivage à long terme.

Quel que soit le format, les fichiers non compressés sont très lourds : voici quelques grandeurs arrondies pour un original A4, en TIFF v. 6 mono-page non compressé et pour un ouvrage de 200 pages en 300 dpi :

Le format e-pub (.epub), libre et ouvert, développé par l'IDPF¹⁷ depuis septem-

13. Du nom de ses inventeurs Lempel, ZiV puis Welch (fin des années 1970 - début des années 1980).

14. National Digital Newspaper Program par ex. : Bibliothèque du Congrès. [En ligne]

< <http://chroniclingamerica.loc.gov/> > (consulté le 19 février 2010).

TABLEAU **RAPPORTS DE POIDS SELON LA PROFONDEUR D'ACQUISITION ET LA RÉOLUTION**

TIFF v. 6 non compressé	300 dpi		600 dpi		ouvrage de 200 p. à 300 dpi		
	Mo	Ko	Mo	Ko	Mo	Ko	Go
Noir et blanc	1.0	1 054	4.1	4 260	200	210 800	0,195
Niveaux de gris	8.2	8 432	33.1	33 996	1 640	1 686 400	1,601
Couleur	24.7	25 296	87.5	89 660	4 940	5 059 200	4,824

Remarque : doubler la résolution revient à quadrupler le volume¹⁵.

Ci-dessous les principaux formats actuels recommandés et leurs caractéristiques¹⁶.

TABLEAU **CARACTÉRISTIQUES DES PRINCIPAUX FORMATS DE FICHIER D'IMAGES FIXES**

p. 28

15. Exemple : « passer de 300 dpi à 600 dpi ».

16. Voir le Mémento sur les formats de la BnF, *op. cit.*, et les pages de la Bibliothèque du Congrès. [En ligne] < <http://www.digitalpreservation.gov/formats> > (consulté le 19 février 2010).

nom/producteur	documentation	caractéristiques
TIFF - <i>Tagged Image File Format</i> /Aldus Corp. (rachat par Adobe®) et Microsoft®	ISO 12639 : 2004. Spécification (v. 6, 1992) : < http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf >.	Extension : .tiff, .tif Monopage (un fichier, une image) ou multipage (un fichier pour une séquence d'images) Archivage d'images lourdes indépendamment des plateformes et des périphériques. Largement supporté par les applications de gestion d'images
JFIF – <i>JPEG File Interchange Format</i> /C-Cube Microsystems	JFIF : < http://www.jpeg.org/public/jfif.pdf > JPEG : ISO : IS 10918-1 - Recommandation T.81	Extension : .jpeg, .jpg, .jpe, .jfif, .jfi Format d'image pour les fichiers compressés avec l'algorithme JPEG (compression avec perte pour échanger des images sur les réseaux avec une qualité acceptable)
JPEG 2000/ <i>Joint Photographic Experts Group</i> , ISO, UIT	ISO – Technologies de l'information – SC 29 – UIT < http://www.jpeg.org/jpeg2000/index.html >	Extension : .jp2 Ensemble de méthodes de compression avec ou sans perte et de formats de fichier
PNG - <i>Portable Networks Graphics</i> /W3C	ISO/IEC 15948 : 2004. Spécification v. 1.2 : < http://www.libpng.org/pub/png/spec/1.2/PNG-Contents.html >	Extension : .png Conçu pour supprimer les limitations du GIF et du JPEG. Ouvert, flexible, robuste. Indépendant des plateformes. Contient de nombreuses informations
PDF – <i>Portable Document Format</i> /Adobe®	ISO 32000-1 : 2008 (v 1.7). Suppléments à la norme : < http://www.adobe.com/devnet/pdf/pdf_reference.html >	Extension : .pdf Créé d'abord pour l'impression. Indépendant du système d'exploitation. Peut englober plusieurs contenus dans d'autres formats et comporter des formulaires
PDF/A-Sous-comité SC 2, ISO/TC 171	Partie 1 : ISO 19005-1 : 2005	Extension : .pdf Créé pour l'archivage à long terme, basé sur la v. 1.4 de PDF

compression	gestion des couleurs	utilisation préconisée
LZW UIT Groupe 4 (réversible) Chaque format compressé selon un algorithme forme un sous-ensemble du TIFF v.6.	Étendue (jusqu'à 24 bits) Supporte plusieurs espaces de couleur et plusieurs profils ICC	Très utilisé pour l'archivage à long terme (souplesse d'utilisation, bonne gestion des couleurs, documentation). Peu adapté pour la diffusion (lourdeur des fichiers). Tous les navigateurs ne le supportent pas encore sous sa forme native. Compression UIT Groupe 4 seulement pour les images bitonales
JPEG. Taux à choisir ; 100 % = pas de compression ; 75-80 % fournit un bon ratio	Étendue (jusqu'à 24 bits). Utilisé de préférence pour les photos.	Peu adapté pour l'archivage. Conçu pour la diffusion et l'échange d'images sur Internet, entre de multiples plateformes et applications
JPEG 2000	Très étendue (jusqu'à 32 bits). Supporte plusieurs espaces de couleur et plusieurs profils ICC	Utilisation accrue à partir de 2005. Utile pour la diffusion car génère des images de résolution variables (« tuiles »), ce qui évite la gestion d'un jeu de fichiers de diffusion
Oui, sans perte. Algorithme PNG non breveté	Très étendue (jusqu'à 48 bits)	Format du Web Tous les outils graphiques ne gèrent pas toutes ses fonctionnalités. Gère la transparence
Oui, plusieurs algorithmes	Étendue Supporte plusieurs espaces de couleur et plusieurs profils ICC	Adapté pour la diffusion et l'échange de fichiers sur Internet, textuels surtout ; peu adapté pour les images lourdes. Fonctions de recherche plein texte utiles pour les projets incluant l'OCR.
Idem	Les espaces de couleurs doivent être spécifiés de manière indépendante	Recommandé par les archives. Contenus audio et vidéos interdits ; liste obligatoire de métadonnées ; les fontes doivent être encapsulées. La norme comprend aussi la façon de développer un outil de visualisation de fichier PDF/A

bre 2007 est utilisé pour la publication sur liseuses (e-books) car il permet d'ajuster le contenu au type de support électronique. Aujourd'hui il est adapté surtout aux contenus textuels simples, pas pour les documents à présentation complexe, et se compose d'un ensemble de fichiers zippés.

gestion des couleurs

La gestion des couleurs est l'un des points délicats dans un projet de numérisation. Dans la chaîne numérique coexistent deux principaux espaces de couleurs (manières de coder celles-ci) : RVB (rouge, vert, bleu) à l'acquisition et à la restitution sur écran, CMJN (cyan, magenta, jaune, noir) à l'impression. Adobe® 98 est un espace RVB très courant ; sRGB, plus étendu, est également utilisé. Les couleurs n'étant pas traduites de manière identique dans chaque espace, des fichiers normalisés – profils ICC¹⁸ – permettront de convertir les références d'un espace à l'autre tout au long de la chaîne pour garantir une impression des couleurs fidèle à l'original. Un profil fait le lien d'une part entre les couleurs de modes différents (RVB vers CMJN) et d'autre part entre les coordonnées des couleurs de chaque appareil travaillant dans un même mode (numériseur, écran). En début de projet et pour chaque lot homogène de documents les appareils doivent être calibrés à l'aide d'outils normalisés : il s'agit de définir leur état de référence (pour les couleurs à partir du point blanc, la résolution, etc.) et de constituer les référentiels utilisés lors de la production courante. Cette étape est à suivre avec attention : un réel dialogue fourni doit être instauré avec le prestataire ou l'opérateur si la numérisation est faite en interne, pour veiller à ce que les réglages permettent d'atteindre le niveau de qualité exigé. Pour une bonne restitution des couleurs au long du projet, on veillera à :

- avoir un système de numérisation performant et adapté au projet avec

17. International Digital Publishing Forum. [En ligne] < <http://www.idpf.org> > (consulté le 19 février 2010).

18. International Color Consortium. [En ligne] < <http://www.color.org/index.xalter> > (consulté le 22 janvier 2010). Organisme regroupant les fabricants d'appareils pour établir les profils normalisés.

une bonne gestion de la lumière à la capture (lumière blanche calibrée avec illumination suffisante, pas de lumière parasite) ;

- étalonner le numériseur : définir un état colorimétrique de référence à partir du point blanc et d'outils (mires normalisées...), utiliser des profils ICC et vérifier régulièrement l'absence de dérives ;
- ajouter à chaque document numérique l'image des mires numérisées le jour du traitement du document ;
- calibrer les écrans de contrôle à l'aide d'un spectrophotomètre ou d'un colorimètre et vérifier régulièrement l'absence de dérives ;
- contrôler les images avec les originaux sous les yeux.

CONVERSION EN MODE TEXTE

Ce procédé permet la recherche plein texte sur des imprimés numérisés, donc de valoriser le contenu et de proposer de nouveaux services aux utilisateurs :

apports de l'ocr et difficultés posées par la conversion

Intérêts

- Recherche plein texte, manipulations multiples
- Amélioration du niveau de services : navigabilité, facilité de consultation, annotations...
- Accessibilité pour les personnes handicapées

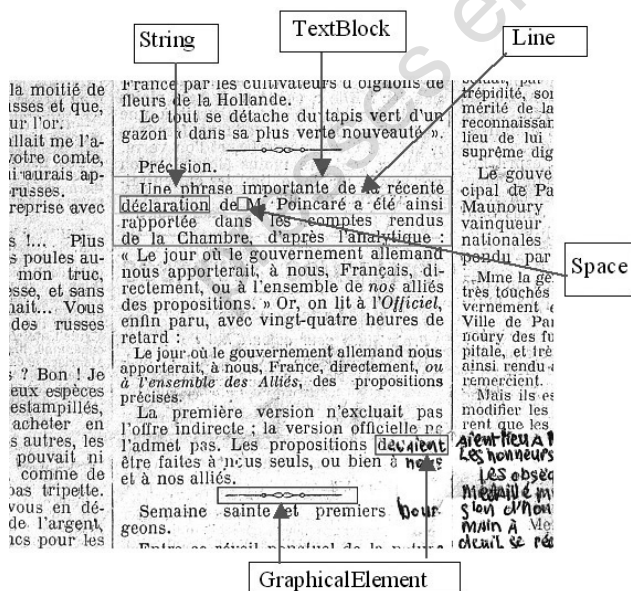
Difficultés

- Lourdeur (balisage conforme à un schéma, relectures...)
- Coûts élevés
- Chaîne et outils de contrôle, voire d'extraction d'images, à développer

La conversion en mode texte est faite par saisie manuelle et/ou par reconnaissance optique des caractères (OCR), méthode automatique. La technologie OCR est en évolution permanente, mais la méthode reste fondamentalement la même : **segmentation** de l'image en plusieurs « boîtes » de texte caractérisées (de la zone imprimée jusqu'à l'espace et la césure entre deux termes), enregistrement de leurs coordonnées en pixels dans un repère orthonormé dont l'origine est le point supérieur gauche de la page, puis **reconnaissance des caractères** de chaque boîte par comparaison avec des ensembles de formes et à l'aide d'outils linguistiques, conversion enfin dans un codage donné (ISO-8859-1, UTF-8...). Ce schéma montre la segmentation et la caractérisation des boîtes dans un texte.

SCHEMA

OCR - SEGMENTATION D'UNE ZONE IMPRIMÉE ET CARACTÉRISATION DES BOÎTES



Le résultat est un fichier texte dans un format simple (.txt) ou xml (format ALTO par exemple) dont le contenu peut être inclus dans un fichier PDF multicouche pour l'affichage du texte en transparence avec le mode image et autoriser la recherche. L'indexation du texte permet d'offrir à l'utilisateur un moteur de recherche qui ne se limite pas aux fonctionnalités du PDF.

La **segmentation** et la **reconnaissance** sont qualifiées par un taux de qualité en pourcentage (100 % signifiant aucune erreur) : ce taux peut être déterminé – OCR corrigé, pour atteindre un niveau de qualité fixe, par exemple 98,5 % – ou pas (OCR brut). Même en combinant plusieurs logiciels d'OCR sur un même texte, atteindre une haute qualité (entre 98 et 100 %) implique une reprise manuelle et devient donc plus coûteux que l'OCR brut qui peut fournir de bons résultats si le support et l'impression sont de qualité. Pour la segmentation, l'unité d'évaluation est la zone couverte par la boîte ; pour la reconnaissance, l'unité est le mot ou le caractère : 99,985 % de reconnaissance au caractère (soit 15 erreurs tous les 100 000 caractères) signifie que sur un échantillon de 1 000 caractères espaces compris, la première erreur provoquera le rejet du document. Ce taux proche de la saisie manuelle risque de retarder la production du fait de reprises très importantes. On choisira donc un taux légèrement plus bas, ou une qualité au mot, selon chaque situation. Pour l'OCR brut, on vérifiera que le taux annoncé par le titulaire est bien celui constaté par le commanditaire.

On prendra en compte en particulier les caractéristiques du support et le contenu. La presse présentant un très grand nombre de caractères par page, un niveau trop proche de 100 % risque d'élever les coûts, alors que l'on pourra choisir un taux très élevé pour des périodiques ou des monographies, et surtout pour des dictionnaires ou des index. En général pour des documents de ce type (aux formats et nombre de caractères par pages standards), l'OCR haute qualité est en moyenne entre 30 et 35 % plus cher que le brut. Les pratiques des bibliothèques françaises sont variables : à la BnF en 2009 le taux de reconnaissance est de 98,5 % au fascicule pour la presse (taux fixe), et pour les imprimés c'est un OCR brut sauf sur une sélection d'entre eux (99,9 % au document). Les autres bibliothèques ayant conduit des projets d'OCR ont en général choisi de l'OCR brut.

La qualité de l'OCR dépend de la qualité de l'original et de sa version numérisée. Même si la recherche dans ce domaine avance vite, on signalera au Cahier des clauses techniques particulières (CCTP) les caractéristiques suivantes, qui n'empêchent pas la production mais limitent la qualité finale :

- mauvaise qualité de l'original et de l'image numérique : cas de courbures et d'inclinaisons de lignes, de caractères déformés : la segmentation est difficile, donc la reconnaissance mauvaise ;
- impression de l'original hétérogène, trop grasse, page tachée (rousseurs...), surtout si l'image est binarisée : les mots seront mal reconnus ;
- zones ou supports difficilement reconnaissables : zones illustrées ; manuscrits ;
- polices de caractères trop resserrées, trop irrégulières, caractères trop gras, trop grands : risques de confusion entre les caractères. Les alphabets non latins ou non ISO sont mal reconnus (grec, fraktur,...) ;
- structure du texte : s'il est en colonne, il faut définir un ordre de segmentation ; signaler les éléments non textuels imbriqués (graphiques, illustrations...).

La résolution doit également être suffisante et la binarisation correcte pour une bonne reconnaissance. Le fonds à traiter doit être le plus homogène possible afin d'éviter des ajustements fréquents des outils, toujours coûteux.

Un format courant pour l'OCR est le ALTO, défini par la Bibliothèque du Congrès et la BnF¹⁹. Reposant sur un schéma XML, il est libre, ouvert et permet l'archivage et la réutilisation à long terme des données dans la mesure où il conserve notamment pour chaque zone de texte, les coordonnées, le taux de confiance de reconnaissance, des informations de présentation.

19. Bibliothèque nationale de France. [En ligne] < http://bibnum.bnf.fr/ns/alto_prod.xsd. > (consulté le 19 février 2010) et la Charte de conversion OCR (*op. cit.*).

Il est donc très répandu mais la gestion des fichiers ALTO nécessite une chaîne de contrôle et de conservation spécifique. Une solution plus légère mais moins pérenne consiste à demander au prestataire une livraison du texte incluse dans les fichiers PDF ou PNG de diffusion. Enfin, selon les moyens disponibles, la conversion des tables des matières et des index en mode texte est une solution intermédiaire satisfaisante pour offrir des accès fins au contenu maintenu en mode image.

SUIVI DE LA QUALITÉ

Le contrôle qualité passe par la définition, par prestation, de **critères** de qualité, d'un ou des **niveaux de qualité acceptables (NQA)** et de méthodes de contrôle, explicités dans le CCTP. Chaque critère mentionné doit être contrôlé. Les erreurs peuvent être classées selon leur gravité avec un NQA différent (par exemple 1 % pour les erreurs qualifiées comme majeures, 1,5 % pour les erreurs mineures). Les contrôles concerneront la qualité des images, de la conversion, des métadonnées, de l'écriture et de l'organisation des fichiers sur les supports de livraison et tout autre point demandé. Le contrôle visuel implique des écrans calibrés et de grande dimension, du personnel suffisant et compétent dont les activités seront diversifiées pour limiter la monotonie du contrôle, source de lassitude et de diminution de l'acuité visuelle. Plusieurs méthodes de suivi de la qualité existent et peuvent être combinées, selon la situation de chaque établissement, en voici quelques-unes ci-dessous :

**MÉTHODES DE CONTRÔLE QUALITÉ POSSIBLES
ET LEURS PRINCIPALES CARACTÉRISTIQUES**

méthode	Avantages	Inconvénients	Remarques
Contrôle exhaustif	Exhaustivité	Coûteux et long	Faible volumétrie, ou en phase de test, ou automatisé
Contrôle par échantillonnage	Peut reposer sur une norme ²⁰ (dans ce cas, l'appliquer rigoureusement)	Suivi précis et régulier du respect de la norme : calcul de la taille de l'échantillon selon les NQA, la taille des lots à contrôler, et les changements de niveau de contrôle (renforcé, normal, réduit) selon le nombre de lots admis ou rejetés	Projets à volumétrie de moyenne à forte et de supports identiques ; contrôle renforcé en phase de test
Mise à disposition d'outils de contrôle calibrés	Investissements très limités ; s'assurer de la maintenance et du calibrage par la société	Maîtrise limitée sur les outils Augmentation des coûts	CCTP rigoureux sur les besoins et les critères techniques ; limiter la demande
Audits de la chaîne du prestataire (contrôle du respect des engagements qualité et des méthodes de production, en amont des livraisons)	> Anticipe sur les reprises à faire et donc réduit les coûts > Englobe l'ensemble des paramètres de production et pas uniquement le produit fini > C'est un dialogue avec le prestataire pour l'amélioration continue	Lourd à mettre en place et à mener. Pour être crédible et efficace doit s'appuyer sur des méthodes standardisées et des documents contractuels (plan assurance qualité incluant des procédures et des indicateurs qualité relevés régulièrement)	Pour tous projets. Ne se substitue pas au contrôle du prestataire sur sa production. Implique de la régularité et la rédaction de rapports d'audits validés par les deux parties ; la mise en œuvre des recommandations d'un audit est vérifiée à l'audit suivant

20. Norme NF-ISO 2859-1 : règles d'échantillonnage pour les contrôles par attributs.

Sous-traiter le contrôle qualité est envisageable si le commanditaire a les moyens de suivre, voire d'arbitrer deux marchés articulés entre eux ou deux lots d'un même marché.

TABLEAU **AVANTAGES ET LIMITES D'UN CONTRÔLE QUALITÉ SOUS-TRAITÉ**

Avantages	Inconvénients
Fiabilité du contrôle et du respect de la norme d'échantillonnage (compétence de la société possédant les ressources techniques et logicielles)	<ul style="list-style-type: none"> • Fiabilité réelle seulement si l'institution a les moyens de « contrôler le contrôleur » et si des règles écrites et précises de production et de contrôle sont appliquées par toutes les parties
Possibilité de traiter des flux plus importants (car les échantillons à contrôler sur les livraisons du contrôleur sont de petite taille)	<ul style="list-style-type: none"> • Flux importants de documents à suivre entre deux prestataires et selon les résultats de contrôle, d'où un suivi de la production très lourd
	<ul style="list-style-type: none"> • Lourdeur de gestion due à la complexité du montage administratif et technique, des plannings, de l'arbitrage nécessaire en cas de désaccord sur les résultats de contrôle • Délais de réalisation et de paiement très fortement allongés • Si les deux marchés sont réalisés en même temps, risques pour un prestataire en cas de difficultés pour l'autre

Au lancement du projet, **la phase de test** vise à valider avec le prestataire les calibrages et les ajustements des outils de production permettant d'obtenir, sur un ensemble de documents représentatifs du fonds à traiter (jeu de tests) la qualité requise au cahier des charges. Cette phase déterminante implique un dialogue rapproché avec le prestataire pour détermi-

ner les procédures et les réglages fins (image exhaustive, nette et redressée, remise en polarité positive pour les originaux négatifs, profondeur de codage et seuil de binarisation adaptés, taux qualité pour l'OCR...). Si plusieurs prestations sont requises, des contrôles unitaires par prestation précéderont les contrôles globaux portant sur l'ensemble. La validation des tests par le commanditaire signifie le lancement de la **phase de production courante**. Les moyens de production étant calés sur les choix et référentiels convenus en phase de test, demander leur modification en production courante est très délicat puisque tout changement sur une chaîne de production a un coût. La réussite de la phase de test conditionne donc largement celle de la production courante.

numérisation et microfilms

Pour certaines collections, en particulier la presse ancienne, en fonction des moyens de l'établissement, de la politique de conservation et de communication des originaux, on choisira de numériser la collection telle qu'elle se présente, de la consolider avant ou après traitement, de faire appel à une collection d'une autre institution en meilleur état, ou à des microformes sous réserve de la qualité des masters et de leur disponibilité pour la numérisation.

Recommandations

Il est préférable de choisir de traiter des masters de très bonne qualité non communiqués aux lecteurs plutôt que des copies. Il est donc nécessaire d'effectuer une visualisation préalable des collections de films par sondage. Lors de cette étape, on prendra particulièrement garde aux points suivants :

- homogénéité de la densité sur une bobine ;
- longueur et solidité du film (on vérifiera la présence ou non de collages) ;
- qualité de l'émulsion ;
- type de reproduction : en noir et blanc ou en modelé continu ;
- état matériel du film : absence de poussières, de rayures ;
- caractéristiques du document reproduit : une reliure très serrée se traduit sur le film par une zone d'ombre pouvant masquer les premiers caractères de chaque ligne et induit des difficultés supplémentaires pour l'OCR. Si cette

zone est large et sur de nombreuses pages, il faudra envisager de traiter l'original papier dérelié, s'il est disponible ; dans le cas contraire, on veillera à signaler au lecteur que la numérisation est conforme à l'original ;

- propriété des films : elle doit appartenir à l'institution, on veillera donc à identifier les modalités de reproduction avant de lancer le programme de numérisation. On notera que les matrices des films reproduits par l'Association pour la conservation et la reproduction photographique de la presse (ACRPP²¹) sont en général la propriété de l'Association, il est donc nécessaire d'obtenir son autorisation avant d'entreprendre leur numérisation.

Les numériseurs de microformes ne permettent pas à ce jour (2009) de numériser tous les formats originaux si la résolution dépasse 300 dpi. Pour traiter des documents à une résolution supérieure à cette dernière, il faudra par exemple :

- se limiter dans les formats des originaux (le A3 représente une limite sûre) ;
- prévoir deux résolutions : 400 dpi jusqu'au A3 et 300 ou moins au-delà.

21. ACRPP : [En ligne] < <http://www.acrpp.fr/> >

CHOISIR UN NUMÉRISEUR

ENCADRÉ

par Bernard Fages

Un grand nombre de numériseurs de documents sont disponibles sur le marché français mais tous ne sont pas destinés aux mêmes applications. On peut distinguer trois grandes typologies d'équipement :

- les numériseurs destinés à être mis en libre-service dans une salle de lecture, l'image numérique est généralement destinée à l'usage propre du lecteur qui peut obtenir soit un fichier numérique via un port USB ou une copie papier directe via une imprimante directement connectée au numériseur ou en réseau ;
- les numériseurs de production, dont les fichiers sont destinés à l'archivage à des fins de conservation et de diffusion. Ils peuvent être répartis en deux groupes : ceux jusqu'au format A2 et ceux d'un format supérieur ;
- les numériseurs automatiques, équipés d'un robot tourne pages, sont destinés à la numérisation de masse. Ils nécessitent des corpus relativement homogènes en terme de dimensions d'ouvrages et de qualité des papiers. Certains d'entre eux peuvent aussi convenir pour des documents dont la reliure ne permet pas une ouverture supérieure à 90°. Le feuilletage automatique des pages peut être débrayé dans le cas d'une trop grande hétérogénéité des feuillets.

Enfin, un boîtier photographique ou dos

numérique montés sur un statif de reproduction²² et un système d'éclairage font un excellent numériseur. Aujourd'hui, les plus avancés de ces systèmes ne permettent d'enregistrer guère plus qu'un format A2 en 400 dpi. Contrairement aux typologies de numériseurs citées plus haut, ce type de configuration exige d'être servi par un photographe.

CHOIX D'UN NUMÉRISEUR

Il est difficile de donner une règle générale pour choisir un numériseur. Au moins trois critères sont à prendre en compte.

Le porte livres

Il doit être adapté aux dimensions (hauteur, largeur et épaisseur) et à la typologie (reliés, reliés serrés, en feuilles, etc.) des documents qui seront enregistrés. Attention : les dimensions admissibles sur un même équipement varient suivant l'utilisation d'un porte livres à 180° ou d'un porte livres à ouverture partielle. La plupart des porte livres peuvent être utilisés avec ou sans vitre. L'utilisation de la vitre nécessite de vérifier si la pression exercée par les plateaux est réglable et si leur vitesse de montée est ajustable.

La résolution sur l'original

La résolution²³ dépend directement du nombre de capteurs sur la barrette d'analyse (cas des numériseurs à balayage) ou du nombre de capteurs sur la matrice et

22. Partie métallique servant de support à un appareil optique.

23. Exprimée en dpi ou ppp : point par pouce.

...

de la distance entre le document original et les capteurs. Pour un même capteur, la qualité dépend aussi de la quantité de lumière insolant les capteurs. C'est ainsi qu'un balayage rapide se traduira par un rendu des couleurs moins correct par rapport à un balayage plus lent. Le logiciel de traitement de l'image du numériseur intervient aussi largement dans la qualité de l'image reproduite.

Le type d'éclairage

Pour avoir un rendu colorimétrique fidèle au document, il faut que le numériseur soit doté d'un système d'éclairage qui lui soit propre. Sans un éclairage adapté, le résultat de numérisation dépendrait du type d'éclairage (lumière du jour, incandescent, fluorescent, LED²⁴...) de la pièce où se trouve le numériseur.

On l'a bien compris, le choix d'un numériseur résulte d'un compromis. Il conviendra de privilégier les critères qui apparaissent les plus importants pour la réussite du projet.

Le travail de numérisation

L'aspect « numérisation » ne représente qu'une partie du travail à effectuer par le photographe. Il est en effet nécessaire de rattacher les fichiers numériques à la notice qui décrit intellectuellement le document et, éventuellement, de donner un certain nombre d'indications sur l'aspect physique du document original : feuillet manquant, erreur de pagination ou de foliotation, pages maculées,

déchirées, etc. et aussi création d'une table des matières en correspondance avec le numéro du fichier. Ces saisies sont à la charge du photographe qui, document en main, doit renseigner ces champs annexes du fichier image, pour faciliter la lecture de l'utilisateur.

Les caractéristiques physiques des documents influent sur la difficulté à les numériser. La principale contrainte est due aux reliures précieuses et fragiles, qui imposent le recours à un porte livre à ouverture partielle (100 à 140 °). C'est également le cas des papiers brillants qui induisent des reflets sur les images numériques ou, à l'inverse, la nécessité de faire briller les « ors » pour certains documents enluminés.

Ces difficultés de numérisation induisent fortement sur la productivité des numériseurs. En pratique, la numérisation automatique d'un document relié, homogène en terme de qualité de papier et de taille de pages permettra d'enregistrer en une heure ce que d'autres feront en une journée avec des documents fragiles, hétérogènes (type de papier et taille de page différents, dépliants, etc.) ou demandant une manipulation méticuleuse.

Avec des documents patrimoniaux, en automatique on ne peut guère dépasser 3 000 pages par jour, et en manuel, suivant les difficultés rencontrées, la productivité peut varier de 200 jusqu'à 1 000 pages par jour, ce qui semble un maximum.

24. LED : *Light-Emitting Diode* – DEL (diode électroluminescente) : petit composant électronique capable d'émettre de la lumière lorsqu'il est soumis à un courant électrique.

**VUE, NON EXHAUSTIVE, DES MARQUES
TABLEAU ET DES NUMÉRISEURS DISPONIBLES EN FRANCE**

marques de numériseurs	distributeur en France	observations	format
Assy-I2S	I2S	Numériseur/tourne page automatique (ouverture à 180° du document)	
Cruse	Art Petrus (Monaco)	Numériseur	jusqu'au A0
I2S	Spigraph	Numériseur	du format A2 au format double A0
ImageWare	Data Technology, Agedoc	Numériseur	du format A2 au format A1
Kirtas	Kirtas	Numériseur/tourne page automatique (ouverture partielle du document)	
Konica Minolta	Agedoc	Numériseur	A3
Metis	L'Equipement Intégral	Numériseur	du format 50 x 70 cm au format double A0
Microbox	Mach 33	Numériseur	A2
Ormec Gestion	Ormec Gestion	Numériseur	A2
Proserv	Agedoc	Numériseur	du format A2 au format A0
Quidenius EB	Consulting/4Capture	Numériseur/tourne page automatique (ouverture partielle du document)	
Solar	Mach 33	Numériseur, éclairage en option	A2
Treventus	Data Technology	Numériseur/tourne page automatique (ouverture 60° du document)	
Zeutschel	EB Consulting/4Capture, Mach 33 et Ormec Gestion	Numériseur	du format A2 au format A0

NUMÉRISATION INTERNE/EXTERNE**TABLEAU***par Thierry Claerr et Isabelle Westeel*

	Avantages	Inconvénients
Numérisation interne	<ul style="list-style-type: none"> • souplesse • possibilité d'articuler avec la numérisation à la demande et les demandes de reproduction • facilité d'adaptation, possibilité de réaliser un travail « à la carte » de très bonne qualité • sécurité des collections convient à des documents fragiles et à des ensembles hétérogènes • développement de compétences en interne • maîtrise d'une chaîne de production (préparation, numérisation, traitement et contrôle-qualité) 	<ul style="list-style-type: none"> • investissement dans un matériel lourd et complexe qui doit être renouvelé régulièrement ; coût de maintenance du matériel élevé • nécessité de mettre en place les outils nécessaires à la chaîne de production (logiciels, écrans calibrés, ordinateurs...) • production souvent lente vu les lots hétérogènes, nécessité de réglages fréquents de la machine • poids sur les ressources humaines et l'organisation interne de l'établissement • nécessité de compétences en numérique, veille, formations
Numérisation <i>in situ</i> sous-traitée à un prestataire	<ul style="list-style-type: none"> • meilleure solution pour les documents fragiles (pour éviter les déplacements) et précieux (en raison de leur valeur d'assurance) • sécurité des collections • possibilité d'organiser des démonstrations pour le personnel et/ou des visites pour les élus, les partenaires et le public 	<ul style="list-style-type: none"> • coût élevé • prise en compte des contraintes d'installation (local adapté) et d'horaires de l'établissement • coordination parfois délicate des tests, de la numérisation et de la post-production avec le prestataire • vigilance sur le poids qui peut peser sur l'organisation interne de l'établissement
Numérisation externe	<ul style="list-style-type: none"> • rapidité d'exécution • convient aux ensembles homogènes • maîtrise du coût de numérisation établi par le marché (prix à l'image ou prix forfaitaire pour un ensemble) • coût de l'obsolescence technologique pris en charge par le prestataire • risque limité • prestations diverses (en fonction des besoins exprimés dans le marché) 	<ul style="list-style-type: none"> • contrôle de la sortie des documents hors des dépôts (vigilance sur les conditions de stockage et les risques liés aux transports) • coût du transport et de l'assurance • compétences techniques assurées principalement par le prestataire • coordination parfois délicate des tests, de la numérisation et de la post-production avec le prestataire

presses ensib

PARTIE II

**RESPECTER
LE DROIT DE
LA PROPRIÉTÉ
LITTÉRAIRE
ET ARTISTIQUE**

RESPECTER LE DROIT DE LA PROPRIÉTÉ LITTÉRAIRE ET ARTISTIQUE

*par Sophie
Sepetjan*

Rares sont les projets de constitution d'une bibliothèque numérique pouvant ignorer le droit de la propriété littéraire et artistique. Les bibliothèques, à l'exception de celles constituées exclusivement de fonds tombés dans le domaine public, conservent en effet de nombreux documents protégés par le droit d'auteur et ce, quels que soient la nature de leurs collections, leurs modes d'entrée ou leur ancienneté.

Avant d'engager un projet de numérisation et pour en cerner les contours, la bibliothèque devra répondre aux questions suivantes : les documents concernés sont-ils des œuvres protégées ? Quelle est la durée de leur protection ? Certains usages sont-ils possibles sans autorisation ? Le cas échéant, auprès de qui faut-il demander ces autorisations et sous quelle forme les obtenir ?

Cette contribution se propose d'apporter des éléments de réponse à ces différentes questions pour orienter les choix de la bibliothèque. Toutefois, le recours aux textes de lois, voire aux services d'experts juridiques, s'avère parfois nécessaire pour écarter ou réduire les risques encourus.

Car reproduire et diffuser une œuvre protégée, sans avoir requis au préalable les autorisations nécessaires et acquitté les droits correspondants, c'est risquer de commettre un acte de contrefaçon et d'être poursuivis devant les tribunaux civils, pour indemnisation du dommage subi, et/ou devant le juge pénal, la contrefaçon étant un délit.

DES DROITS D'AUTEUR ET DES DROITS VOISINS

+++++

Le droit d'auteur distingue traditionnellement les droits patrimoniaux, prérogatives permettant à l'auteur de tirer un profit pécuniaire de l'ex-

ploitation de son œuvre et le droit moral, dont l'objectif essentiel est de permettre à l'auteur de défendre son œuvre contre les atteintes qui pourraient lui être portées.

LES DROITS PATRIMONIAUX

Les droits patrimoniaux sont temporaires. Ils rassemblent le droit de reproduction et le droit de représentation. Ils appartiennent en premier lieu exclusivement à l'auteur. Pour le droit de reproduction, l'élément déterminant est la fixation de l'œuvre sur un support, quel qu'il soit. Ainsi, la numérisation, procédé de reproduction permettant de fixer une œuvre sur un support numérique, ou plus virtuellement à lui donner forme dans un fichier électronique, met en œuvre le droit de reproduction. Pour le droit de représentation, l'élément essentiel est la transmission de l'œuvre au public, quels que soient le mode et le procédé utilisés.

Si le plus souvent l'exploitation de l'œuvre va mettre en jeu simultanément ces deux droits, il faut garder à l'esprit qu'ils sont parfaitement distincts l'un de l'autre. Ceci a pour conséquence qu'une opération de numérisation et de mise en ligne suppose d'obtenir une double autorisation du titulaire des droits. Ainsi, comme nous le verrons plus loin, le contrat par lequel l'auteur cédera son droit de reproduction à des fins de numérisation et de représentation à des fins de mise en ligne devra explicitement viser ces deux droits.

Les droits d'auteur sont des droits de nature incorporelle, distincts de la propriété matérielle d'un bien. Cette règle essentielle explique pourquoi le fait de conserver une œuvre dans ses collections ne confère pas à la bibliothèque le droit de la reproduire et d'en diffuser la reproduction.

Avant 1910, la loi prévoyait que les achats de collections entraînaient la cession des droits d'auteur à l'acquéreur. Ainsi les collections achetées avant 1910 sont réputées avoir été transférées à l'acquéreur avec les droits

d'auteur²⁵. Cependant, la portée de ce texte est limitée en pratique. En effet, le principe d'interprétation restrictive des cessions impose que chacun des droits et des usages envisagés fasse l'objet d'une cession précise et explicite dans le contrat. Bien entendu, les clauses de cession du début du xx^e siècle ne pouvaient prévoir la reproduction sur support numérique ni la diffusion sur Internet. Des avenants aux contrats de l'époque sont donc à prévoir afin d'élargir les droits cédés à la numérisation et à la mise en ligne.

LE DROIT MORAL

Le droit moral est inaliénable, imprescriptible, perpétuel.

Inaliénable, car l'auteur ne peut céder son droit moral par contrat ni renoncer par avance à le faire valoir. Imprescriptible, car l'auteur et ses ayants droit ne sont limités par aucun délai pour exercer la défense de ce droit en justice. Perpétuel car, à la différence des droits patrimoniaux qui ne durent qu'un temps, le droit moral ne s'éteint jamais.

Le droit moral est un droit absolu, que l'auteur ou ses ayants droit peuvent défendre sans aucune limite temporelle.

Le droit moral est constitué de quatre prérogatives :

- le droit de divulgation ;
- le droit à la paternité ;
- le droit au respect de l'œuvre ;
- le droit de retrait et de repentir.

La période de protection est appelée « monopole d'exploitation » de l'auteur. Une fois les droits patrimoniaux expirés, l'œuvre « tombe » dans le domaine public. Son exploitation est alors possible sans autorisation ni rémunération des ayants droit de l'auteur. Mais le droit moral, lui, ne connaît pas de limite temporelle : il doit toujours être respecté, y compris au-delà du monopole d'exploitation.

25. « Les ventes pleines et entières effectuées avant la loi du 19 avril 1910, sans aucune réserve, de la propriété d'une œuvre littéraire transmettent à l'acquéreur la pleine et absolue propriété de l'œuvre ainsi que l'ensemble des droits patrimoniaux de l'œuvre » (Cass. 1^{re} civ., 25 mai 2005).

Le droit de divulgation

Seul l'auteur peut décider de divulguer ou non son œuvre. Le dépôt, le don, le legs de collections inédites par un auteur à une bibliothèque ne constituent pas une divulgation de leur contenu. Le droit moral étant perpétuel, la publication ou la diffusion d'inédits, sous quelque forme que ce soit (publication, exposition, mise en ligne...) ne peut se faire sans la volonté clairement exprimée de l'auteur, et ce, quelle que soit la date de cette divulgation. Avant de procéder à la numérisation d'un inédit, il convient de rechercher la trace d'un souhait qui aurait été exprimé par l'auteur (dans ses publications et papiers, dans un testament ou toute autre expression claire de sa volonté). À défaut de trouver trace d'un souhait clairement exprimé du vivant de l'auteur, il conviendra de s'assurer de l'existence d'ayants droit et, le cas échéant, de recueillir leur accord pour la divulgation.

Précisons que l'exercice du droit de divulgation par les ayants droit n'est pas absolu. En cas de refus de ceux-ci, il est possible de demander au tribunal de contrôler le caractère abusif ou non de ce refus.

Le droit à la paternité

L'auteur jouit du droit au respect de son nom et de sa qualité²⁶. Il peut choisir à l'inverse de publier son œuvre anonymement ou sous un nom d'emprunt son œuvre. Reproduire une œuvre et/ou la diffuser sans mentionner le nom de son auteur est une violation du droit moral.

Le droit au respect de l'œuvre

L'auteur peut s'opposer à toute modification, suppression, ajout ou altération quelconque de son œuvre. Le respect de l'intégrité de l'œuvre commande que la reproduction n'opère aucun changement dans la structure, l'apparence, le contenu de l'œuvre. Ce point est délicat en matière de numérisation. Il faut veiller à ne pas dénaturer l'œuvre, notamment par un changement de format ou de qualité. Ainsi, l'océrisation d'un texte peut, dans certains cas, conduire à une reproduction tronquée de l'œuvre, susceptible de porter atteinte à son intégrité et de constituer une atteinte au droit moral de l'auteur.

26. Article L121-1 al.1 du CPI. Disponible sur le site Legifrance.
[En ligne] < <http://www.legifrance.gouv.fr> > (consulté le 22 janvier 2010).

Le droit de retrait et de repentir

Maître de sa divulgation, l'auteur est également libre d'éprouver des regrets et de demander le retrait définitif (droit de retrait) ou temporaire de son œuvre, le temps d'y apporter des modifications (droit de repentir). Parce qu'il est susceptible de créer un lourd préjudice pour l'éditeur ou le producteur, cessionnaires du droit d'exploitation de l'œuvre, l'exercice de ce droit suppose le versement par l'auteur d'une indemnisation. Cette prérogative est rarement mise en œuvre. Le cas échéant, le retrait d'un ouvrage du commerce par son auteur aurait certainement des répercussions sur sa mise en ligne pourtant préalablement autorisée.

LES DROITS VOISINS

Des droits voisins du droit d'auteur bénéficient aux catégories suivantes : artistes-interprètes, producteurs de phonogrammes et de vidéogrammes, entreprises de communication audiovisuelle. Seules les deux premières catégories seront évoquées ici.

On l'aura compris, les droits voisins ne concernent que les œuvres audiovisuelles et sonores. Pour celles-ci, les droits voisins s'ajoutent au droit d'auteur, de telle sorte que la reproduction et la diffusion d'œuvres sonores et audiovisuelles supposent l'obtention d'autorisations tant des auteurs que des titulaires des droits voisins. Les autorisations seront délivrées selon les cas par les sociétés de gestion collective et/ou les producteurs.

L'artiste interprète bénéficie d'un droit moral limité au droit au respect de son nom et de son interprétation. Il dispose en revanche d'un droit exclusif d'autoriser l'utilisation de son interprétation, en autorisant ou non la fixation, la reproduction et la communication de l'interprétation. L'artiste interprète contrôle la destination de son interprétation. En d'autres termes, l'autorisation donnée par contrat pour l'enregistrement de son interprétation en vue d'une production phonographique et de sa distribution ne vaudra pas pour les autres utilisations.

LE DROIT SUI GENERIS DES BASES DE DONNÉES ²⁷

Les bases de données, qu'elles soient électroniques ou sur support traditionnel, bénéficient d'une protection spécifique, lorsque la présentation du contenu atteste « d'un investissement financier, matériel ou humain substantiel ». Cette protection bénéficie au producteur de la base de données, personne physique ou morale.

Elle peut se doubler d'une protection du contenu de la base au titre du droit d'auteur et/ou des droits voisins.

COMMENT RECONNAÎTRE UNE ŒUVRE PROTÉGÉE ?

+++++

C'EST UNE ŒUVRE DE L'ESPRIT

Au sens du droit d'auteur, l'œuvre est une création intellectuelle, une œuvre de l'esprit. Pour être protégée, l'œuvre doit avoir été matérialisée, réalisée sous une forme quelconque. Ainsi une simple idée – de livre, de scénario, de tableau – ne peut être protégée en tant que telle, seule sa matérialisation, écrite ou orale, musicale ou parlée, graphique, plastique etc. pourra l'être. L'article L.112-2 du Code de la propriété intellectuelle dresse d'ailleurs une liste non exhaustive d'œuvres de l'esprit, à titre purement indicatif. Le législateur s'est bien gardé de clore la liste des œuvres tant l'évolution des techniques, de la société et la richesse de l'imagination humaine sont de nature à renouveler en permanence la nature et le genre des créations.

Ainsi en pratique, peu de documents conservés dans les collections des bibliothèques échappent à la qualification d'œuvres.

L'ŒUVRE EST PROTÉGÉABLE QUELS QUE SOIENT SON GENRE, SA FORME, SON MÉRITE ET SA DESTINATION

Ceci est d'autant plus vrai que la qualification d'œuvre est indépendante d'un certain nombre de critères. Elle peut être protégée indépendamment de son genre (la catégorie à laquelle elle appartient, littéraire, musicale,

27. Articles L.112-3 et L.341-1 et suivants du CPI.

audiovisuelle ou autre), de sa forme d'expression (écrite, orale, musicale...), de son mérite (la qualité et l'esthétique n'entrent pas en considération) et de sa destination (peu importe l'usage qui en sera fait).

L'ŒUVRE NE REQUIERT AUCUNE FORMALITÉ DE DÉPÔT

À la différence des marques ou des brevets, qui supposent l'accomplissement de formalités préalables et d'un dépôt pour bénéficier d'une protection juridique, l'œuvre, elle, n'est soumise à aucune formalité pour bénéficier de celle du droit d'auteur²⁸. C'est une distinction fondamentale entre propriété industrielle et propriété intellectuelle.

L'ŒUVRE DOIT ÊTRE ORIGINALE

La seule condition requise pour qu'une œuvre bénéficie de la protection est son originalité. Une œuvre originale est une œuvre qui porte l'empreinte de la personnalité de son auteur : elle est une création intellectuelle qui lui est propre. Pour mieux comprendre cette notion d'originalité, qui est plus proche de la notion de liberté créative que de la nouveauté, il faut imaginer que l'œuvre est le résultat de choix techniques, artistiques et de ce qui fait l'unicité de son auteur, sa personnalité. Ainsi, la jurisprudence ne reconnaît pas de droit d'auteur au copiste qui reproduit à l'identique, sans la moindre différence visible, l'œuvre d'un tiers.

Dans le domaine du livre, le dépôt légal, créé en 1537 par François I^{er}, impose aux éditeurs le dépôt aujourd'hui en deux exemplaires à la BnF de toutes les publications produites ou diffusées en France en vue de leur conservation et de leur consultation. Le dépôt légal n'a aucune incidence sur le régime du droit d'auteur. Il ne crée pas de protection et n'est pas une condition à la protection. En la matière, le dépôt légal peut tout au plus servir de preuve pour établir la date de création d'une œuvre.

28. La convention de Berne, traité international fondamental en matière de propriété littéraire et artistique, dispose dans son article 5.2 que « la jouissance et l'exercice des droits [d'auteur] ne sont subordonnés à aucune formalité ». Dans le même esprit, l'article L.III-1 du Code de la propriété intellectuelle énonce que « l'auteur d'une œuvre de l'esprit jouit sur cette œuvre, du seul fait de sa création, d'un droit de propriété incorporelle exclusif et opposable à tous ».

Où s'arrête la reproduction servile ? Où commence la création ? Une telle distinction trouve son application dans le domaine des arts graphiques et plastiques, avec les copies de tableaux par exemple, mais également dans le domaine littéraire. Ainsi, il est d'usage de ne reconnaître aucune originalité à un texte de pure transcription, d'un manuscrit par exemple, dès lors qu'elle n'a laissé à son auteur aucune liberté de création intellectuelle.

Les déclinaisons de cette notion sont nombreuses. Les traductions sont des œuvres dérivées de l'œuvre première. Elles font partie des œuvres susceptibles de protection, car il est peu probable que deux traducteurs confrontés au même exercice livrent deux traductions rigoureusement identiques du même texte.

Le cas des notices bibliographiques est également intéressant. Il convient de distinguer celles conçues à partir d'un cadre ne laissant aucune marge de manœuvre à leur auteur, par exemple lorsque celui-ci doit remplir des champs prédéfinis, et celles pour lesquelles l'auteur a pu faire œuvre de création intellectuelle, comme dans le cas des notices d'autorité.

QUELLE EST LA DURÉE DE PROTECTION ?

+++++

DURÉE DU DROIT D'AUTEUR

cas général

Le droit moral est perpétuel. En théorie, il ne s'éteint jamais. En pratique, il perd de sa vigueur au fur et à mesure que les générations se succèdent et se trouvent plus ou moins disposées à agir pour en faire assurer le respect.

Les droits patrimoniaux ont une durée de vie limitée. La règle générale est que ces droits perdurent pendant la durée de vie de l'auteur et 70 ans après sa mort.

Pour les œuvres de collaboration, conçues par une multiplicité d'auteur, la date à retenir pour le calcul de la durée de protection est la date de décès du dernier des co-auteurs survivants.

cas particuliers : les œuvres anonymes, pseudonymes et collectives

Par définition, l'auteur d'une œuvre anonyme ou pseudonyme est inconnu. Faute de pouvoir calculer le point de départ de la durée de protection à compter de la date de décès de l'auteur, ce calcul devra être fait à compter de la date de publication.

Dans l'hypothèse où la date de publication serait elle-même inconnue, il convient d'essayer d'en faire l'estimation, grâce à des éléments d'appréciation contenus dans le document lui-même : références historiques, contenu d'une image, etc. ou éléments extérieurs au document : articles de presse, critiques littéraires etc.

La date certaine ou estimée de publication permettra de définir une date approximative de décès de l'auteur²⁹. Il conviendra dans ce cas de procéder au calcul en ménageant une importante marge de sécurité.

Les œuvres collectives sont des œuvres créées à l'initiative et sous la direction d'une personne physique ou morale – un éditeur –, dans lesquelles les contributions de chacun se fondent dans l'œuvre commune sans qu'il soit possible d'attribuer à chacun la part qui lui revient. Ce cas vise quasi exclusivement les dictionnaires, et encore, seulement lorsque les articles ne sont pas signés ou paraphés par leurs auteurs. Dans ce cas, qui reste très exceptionnel, la protection est de 70 ans à compter de la publication.

Les œuvres orphelines sont des œuvres protégées dont les titulaires de droits ne peuvent être identifiés, ou ne peuvent être retrouvés malgré des recherches avérées et sérieuses. Les autorités communautaires et nationales travaillent à la mise en place d'un cadre légal qui permettra l'utilisation des œuvres orphelines tout en assurant le respect du droit et la rémunération des éventuels titulaires de droits, s'ils venaient à se faire connaître. Pour l'instant, les œuvres orphelines constituent dans les bibliothèques des stocks d'œuvres protégées³⁰ dont l'utilisation est bloquée faute d'autorisation expresse³¹. La prise de risque accompagnée d'une mention « droits réservés » est à proscrire.

Les répercussions de la guerre sur les durées de protection

Les auteurs « morts pour la France »

Pour les auteurs « morts pour la France » - cette mention figurant sur leur acte de décès - une période de protection supplémentaire de 30 ans s'applique³².

Les prorogations de guerre

Destinées à compenser le manque à gagner subi par les auteurs pendant les guerres, les prorogations de guerre ont eu pour effet de prolonger les durées de protection d'une durée équivalente à celles des périodes de conflit.

En France cependant, bien que les prorogations de guerre figurent toujours dans le Code de la propriété intellectuelle, les jurisprudences rendues par la Cour de cassation permettent de retenir une durée de la protection de 70 ans après la mort de l'auteur, incluant les prorogations de guerre. La seule exception concerne les cas où au 1^{er} juillet 1995, date d'entrée en vigueur de la directive européenne sur l'harmonisation des durées de protection, une période de protection plus longue avait commencé à courir, les droits acquis étant dans ce dernier cas respectés.

Ainsi « Le Petit Prince » de Saint-Exupéry, qui bénéficiait d'une durée de protection de 50 ans, plus 8 ans de prorogation pour la première guerre mondiale, plus 30 ans du fait qu'Antoine de Saint-Exupéry est « mort pour la France » en juillet 1944, est protégé pendant 88 ans au total, soit jusqu'en 2032. Un tel cas ne devrait plus concerner que de très rares auteurs en France.

29. Ainsi, pour écarter le risque encouru, on pourra considérer que l'auteur a écrit à l'âge de 18 ans et qu'il est mort à l'âge de 80 ans. Ainsi pour un ouvrage paru en 1921, une extrapolation conduira à considérer que l'auteur a pu naître en 1903, mourir en 1983, laissant ainsi cette œuvre protégée jusqu'au 1^{er} janvier 2054...

30. La British Library a récemment estimé que 40 % de son fonds serait constitué d'œuvres orphelines. Une telle estimation ne peut être que très approximative, faute d'une analyse détaillée des œuvres, de leur statut, de leurs ayants droit et des adresses de ces derniers. Malgré cette imprécision, l'importance du chiffre avancé donne une idée de l'ampleur du phénomène et de l'urgence à y remédier par un système légal.

31. Le Code de la propriété intellectuelle (art. L.121-3 et L.211-2) prévoit toutefois une procédure de saisine du juge en cas de vacance, de déshérence ou d'absence d'ayant droit connu.

32. Voir notamment : ministère de la Défense. [En ligne]

< <http://www.memoiredeshommes.sga.defense.gouv.fr> > (consulté le 22 janvier 2010).

durée des droits voisins

Elle est de 50 ans à compter du 1^{er} janvier de l'année civile à partir :

- de l'interprétation de l'œuvre (pour les artistes interprètes) ;
- de la première fixation du phonogramme ou du vidéogramme (pour les producteurs de phonogrammes et de vidéogrammes).

Toutefois, le point de départ du délai de protection peut être la date de première communication au public si celle-ci intervient pendant la durée du monopole.

Compte tenu des durées différentes des droits voisins et du droit d'auteur ainsi que des points de départ des délais de protection, certains documents peuvent être libres de droits voisins mais encore protégés au titre du droit d'auteur (c'est le cas par exemple de la version du *Requiem* de Maurice Duruflé, compositeur mort en 1986 jouée pour la première fois en 1947) ou libres de droit d'auteur mais protégés au titre des droits voisins (ainsi la version chorégraphiée en 1961 par Maurice Béjart du *Boléro* de Maurice Ravel décédé en 1937).

FOCUS SUR QUELQUES CATÉGORIES DE DOCUMENTS

LES MANUSCRITS ET CORRESPONDANCES

Ces documents présentent plusieurs particularités :

- ils sont souvent inédits. Leur utilisation suppose donc l'autorisation des ayants droit de l'auteur au titre du droit moral de divulgation de l'auteur.
- la communication de certains manuscrits, en particulier les correspondances, journaux intimes ou brouillons, doit être faite dans le respect de la vie privée de l'auteur, du destinataire des lettres ou de toute personne citée ou mise en cause dans le manuscrit³³ Au cas par cas, lorsque le contenu du document est susceptible de mettre en cause la vie privée d'une ou

33. Art. 9 du Code civil : « chacun a droit au respect de sa vie privée ». Ce droit absolu peut être défendu par la personne de son vivant et par ses héritiers.

plusieurs personnes (révélation de faits intimes, de situations compromettantes, etc.), l'autorisation des personnes concernées pourra être requise.

- le droit au respect de la vie privée ne concerne que les personnes vivantes et la faculté d'agir s'éteint avec le décès de la personne concernée, seule titulaire de ce droit (Civ. 1^{re}, 14 décembre 1999, Bull. n° 345). Pour les personnes décédées, seuls sont exigés le respect de la vie privée familiale et pour les personnes notoirement connues, le « respect dû à la vérité ».
- des réserves de communication peuvent être posées par les auteurs ou leurs ayants droit, ainsi que par les propriétaires d'archives privées, au moment de la remise de leurs fonds en don ou en dépôt auprès de l'institution. Ces réserves doivent être scrupuleusement respectées par la bibliothèque³⁴.

LA PRESSE PÉRIODIQUE

Les journaux, revues, périodiques sont la plupart du temps des œuvres de collaboration. Sauf exception, le régime des œuvres collectives ne leur est pas applicable. Pour déterminer si un périodique est encore protégé, il convient de retenir la date de décès du collaborateur décédé le dernier.

Les droits d'auteurs appartiennent au journaliste. Il existe une cession automatique des droits patrimoniaux du journaliste à son employeur pour la première diffusion de son article. Tous les autres usages, notamment par des tiers, doivent être expressément autorisés.

En outre, si le périodique concerné est toujours commercialisé, il convient de s'assurer de l'autorisation de la société qui en assure l'exploitation, notamment au regard du droit d'auteur (protection du titre), du droit de la concurrence et/ou du droit des marques.

Si le titre a disparu, mais que le document est encore protégé, l'utilisation de cette œuvre considérée comme orpheline n'est pas permise.

34. Art. L213-6 du Code du patrimoine : « Les services publics d'archives qui reçoivent des archives privées à titre de don, de legs, de cession ou de dépôt sont tenus de respecter les stipulations du donateur, de l'auteur du legs, du cédant ou du déposant quant à la conservation et à la communication de ces archives ».

LES DOCUMENTS ICONOGRAPHIQUES : ESTAMPES, PHOTOGRAPHIES, CARTES POSTALES ETC.

Les documents iconographiques présentent la particularité de superposer plusieurs droits. Ainsi, imaginons une photographie qui représenterait le peintre Pierre Soulages, posant devant l'un de ses tableaux, lui-même installé devant la pyramide du Louvre de l'architecte Pei. Une telle photographie mettrait en jeu le droit de l'auteur de la photographie, le droit d'auteur de l'artiste sur son tableau, son droit sur son image et le droit d'auteur de Pei comme architecte-auteur de la pyramide.

Le droit à l'image des personnes : toute personne dispose d'un droit absolu sur son image et sur l'utilisation qui en est faite. Elle peut ainsi s'opposer à sa reproduction et à sa diffusion sans son autorisation expresse, quel que soit le support utilisé. En outre, l'autorisation donnée par une personne à l'utilisation de son image pour une exploitation déterminée ne vaut que pour cette seule exploitation. Toutefois, comme le droit au respect de la vie privée auquel il se rattache, le droit à l'image des personnes cesse à leur décès : *« le droit d'agir pour le respect de la vie privée s'éteint au décès de la personne concernée, seule titulaire de ce droit »* (Cour de cassation, arrêt du 15 février 2005).

Le droit à l'image des biens : la jurisprudence de la Cour de Cassation est désormais constante. Le propriétaire d'un bien ne dispose pas d'un droit exclusif sur l'image de celui-ci et il ne peut s'opposer à l'utilisation de l'image de son bien que lorsque cet usage lui cause « un trouble anormal ».

LES ŒUVRES AUDIOVISUELLES

Les documents audiovisuels, comme les œuvres sonores, combinent droits d'auteur (auteurs de la composition musicale, des paroles, du livret de l'œuvre lyrique...) et droits voisins (interprètes, musiciens, chanteurs, producteurs...). Pour le calcul des droits d'auteur, il faut retenir la date de décès du dernier auteur décédé. Sont présumés auteurs, sauf preuve contraire, l'auteur du scénario, l'auteur de l'adaptation, l'auteur du texte parlé, l'auteur des compositions musicales avec ou sans paroles spécialement réalisées pour l'œuvre, le réalisateur (article L. 113-7 al 2 du CPI).

La présomption est simple et peut être renversée par la preuve de la non-participation de ces intervenants à la création de l'œuvre. En outre, la liste des coauteurs n'est pas limitative et tout autre intervenant peut rapporter la preuve de sa participation à la création intellectuelle de l'œuvre audiovisuelle.

Il existe par ailleurs une présomption de cession des droits d'exploitation au producteur, ce qui facilite la gestion des droits. Attention cependant, cette présomption de cession ne vaut pas pour « l'auteur de la composition musicale avec ou sans paroles ». En pratique, les demandes d'autorisation devront donc être formulées en parallèle au producteur et à l'auteur de la musique, via la SACEM/SDRM.

QUELS SONT LES USAGES PERMIS SANS AUTORISATION ?

+++++

Les évolutions technologiques importantes intervenues ces dernières années ont conduit le législateur à adapter la propriété littéraire et artistique aux réalités de notre société. Une avancée importante a été réalisée par l'adoption de la directive européenne du 22 mai 2001 sur « l'harmonisation de certains droits d'auteur et droits voisins dans la société de l'information » et par sa transposition en droit français par la loi connue sous l'acronyme Dadvsi (loi du 1^{er} août 2006 « droits d'auteur et droits voisins dans la société de l'information »).

Ces textes ont introduit dans notre droit de nouvelles exceptions au droit d'auteur.

Pour mémoire, les exceptions au droit d'auteur sont des cas limitativement énumérés par le Code de la propriété intellectuelle, dans lesquels certains usages sont permis sans autorisation préalable, ce qui constitue en soi une exception à la règle absolue de l'autorisation préalable imposée par le droit d'auteur. Parmi les exceptions au droit d'auteur, on peut citer la représentation dans le cercle de famille, le droit de courte citation, la copie privée, l'exception pédagogique ou l'exception handicap. Dans certains cas, l'exception n'est pas totale puisque les titulaires de droits perçoivent quand même une rémunération par le biais d'un système de licence légale : une partie du prix payé par l'utilisateur vient ainsi compenser l'autorisation de

reproduction donnée. Ainsi l'utilisateur paie, par exemple en achetant un support vierge de reproduction (copie privée) ou en payant le prix de sa photocopie (droit de reprographie). Un système analogue est envisagé en France pour permettre l'utilisation des œuvres orphelines.

Dans les situations qui nous occupent, numérisation et mise en ligne, les exceptions au droit d'auteur sont rares, voire inexistantes.

Seule la numérisation, réalisée par une bibliothèque à des fins de conservation, entre à coup sûr dans le cadre de l'exception bibliothèque³⁵. La mise en ligne dans les emprises des bibliothèques accessibles au public, des musées et des services d'archives n'est possible qu'à la condition qu'elle respecte le test en trois étapes³⁶ ce qui reste encore à démontrer...

Une certitude est qu'aucune exception ne peut permettre de couvrir la mise en ligne sur Internet. Ce type d'utilisation suppose l'autorisation préalable et, le cas échéant, la rémunération des titulaires de droits.

À QUI S'ADRESSER POUR OBTENIR LES AUTORISATIONS ?

S'ADRESSER À L'AUTEUR

« La qualité d'auteur appartient sauf preuve contraire, à celui sous le nom de qui l'œuvre est divulguée » (art. L 113-1 CPI). La qualité d'auteur revient donc à celui ou celle qui se fait le premier connaître en tant que tel. Il ou elle bénéficie d'une présomption que seule une preuve contraire juridiquement valable peut écarter.

35. Un ajout au 8° de l'article L.122-5 du Code de la propriété intellectuelle dispose : « Les actes de reproduction spécifiques effectués à des fins de conservation ou destinés à préserver les conditions de sa consultation sur place, par des bibliothèques accessibles au public, des musées ou par des services d'archives, qui ne recherchent aucun avantage commercial ou économique direct à condition de ne pas porter atteinte à l'exploitation normale de l'œuvre ni causer un préjudice injustifié aux intérêts légitimes de l'auteur ».

36. Les exceptions au droit d'auteur ne peuvent être mises en œuvre que si elles respectent le test en trois étapes : elles doivent viser un cas spécifique, ne pas porter une atteinte injustifiée aux intérêts légitimes des auteurs et ne pas porter préjudice à l'exploitation normale de l'œuvre.

En tant que titulaire originel des droits, c'est donc l'auteur qui délivre les autorisations de reproduction et de représentation de son œuvre, en fixe les contours, les limites et les conditions financières.

Il peut, par contrat, céder l'exploitation de ses droits à un exploitant, éditeur ou producteur par exemple. Il peut également confier la gestion de ses droits à un tiers à qui il demande d'en assurer la gestion en délivrant les autorisations à sa place et de percevoir pour son compte les rémunérations correspondantes.

Sur une œuvre de collaboration, les coauteurs sont cotitulaires des droits. Cela suppose de recueillir les autorisations individuellement auprès de chaque coauteur ou auprès du producteur lorsque ce dernier est cessionnaire de l'ensemble des droits (voir *infra* pour les œuvres audiovisuelles).

S'ADRESSER AUX AYANTS DROIT DE L'AUTEUR

Le Code de la propriété intellectuelle prévoit que l'auteur peut céder ses droits patrimoniaux, notamment par testament dans le cadre de sa succession. Si rien de particulier n'a été prévu par l'auteur, les règles normales de dévolution du Code civil s'appliquent. C'est le cas également pour le droit au respect du nom et le droit au respect de l'œuvre.

Seul le droit de divulgation obéit à des règles particulières. Sauf volonté contraire exprimée par l'auteur, la loi confie l'exercice de ce droit à l'exécuteur testamentaire, puis, à défaut d'exécuteur testamentaire ou en cas de refus ou de décès de celui-ci, aux descendants, à défaut de descendants au conjoint, et à défaut de conjoint aux autres héritiers.

L'auteur peut choisir de placer son œuvre sous un régime de réutilisation prédéfini par une licence, telles que les licences *creative commons*. Il peut même décider de renoncer à l'exercice de ses droits patrimoniaux, en plaçant volontairement ses œuvres sous licence « CC0 » (CC-zéro). Cette licence créée très récemment prévoit l'abdication totale par l'auteur de ses droits, conduisant ainsi à placer volontairement une œuvre protégeable sous un régime de domaine public. Attention cependant : ce régime est révocable (l'auteur pourra toujours renoncer à l'application de la licence et revenir au régime commun du droit d'auteur) et le droit moral ne cesse de devoir être respecté...

S'ADRESSER AUX SOCIÉTÉS DE GESTION COLLECTIVE

Ces sociétés ont pour double mission de délivrer pour le compte de leurs adhérents les autorisations d'utilisation de leurs œuvres et d'en fixer les conditions d'utilisation ainsi que de percevoir et répartir les redevances payées par les utilisateurs. Les sociétés de gestion tiennent à la disposition des utilisateurs le répertoire des auteurs dont elles gèrent les droits. Ces sociétés sont constituées sous forme de sociétés civiles et sont contrôlées par le ministère de la Culture et par la Commission permanente de contrôle des sociétés de perception et de répartition des droits³⁷.

Les auteurs et titulaires de droits voisins peuvent confier la gestion de leurs droits à une société soit en leur confiant un simple mandat, soit en leur cédant tout ou partie de ses droits d'exploitation sur leurs œuvres. Le caractère exclusif de ces cessions explique pourquoi l'auteur, une fois ce mandat de gestion confié à une société de gestion, ne peut plus gérer ses droits lui-même. Ainsi, si un auteur est adhérent de l'ADAGP ou de la SACEM, il ne pourra plus délivrer d'autorisation directement à l'utilisateur qui sera contraint de passer par la société de gestion pour l'obtenir.

QUELLE FORME DOIT REVÊTIR L'AUTORISATION ?

+++++

LES ÉLÉMENTS DU CONTRAT

Si le Code de la propriété intellectuelle n'exige l'existence d'un écrit que pour certains contrats (d'édition, de représentation et d'adaptation audiovisuelle notamment), il est fermement conseillé d'obtenir des autorisations écrites de la part des titulaires de droits, ne serait-ce que pour des raisons de preuve. Selon le principe d'interprétation restrictive des cessions, la liste et la portée des droits cédés par l'auteur s'apprécient de façon restrictive. En d'autres termes, ce qui n'a pas fait l'objet d'une cession précise et explicite dans le contrat est présumé ne pas avoir été cédé.

37. Voir la liste du ministère de la Culture et de la Communication des 22 sociétés de perception et de répartition des droits d'auteur (SPRD). [En ligne] : < <http://www.droitsdauteur.culture.gouv.fr/index-pla.htm> > (consulté le 6 mars 2010).

L'article L.131-3 impose sous peine de nullité de la cession, que quatre éléments figurent dans le contrat : l'étendue des droits cédés, c'est-à-dire leur nature du droit cédé et leur champ (droit de reproduction et/ou de représentation et modes d'exploitation prévus), la destination des droits cédés (la finalité poursuivie, par exemple diffusion sur Internet dans le cadre d'une bibliothèque numérique), le lieu et la durée de la cession (territoire national ou monde entier, pour combien de temps)

Dans le cadre des projets de bibliothèque numérique d'accès gratuit, il est recommandé de prévoir une cession des droits sans exclusivité. En effet, ce type d'usage ne justifie pas que l'auteur se dessaisisse totalement de ses droits au profit de l'institution. Une telle exclusivité serait d'ailleurs probablement hors de portée financière de celle-ci. Dans ces projets, les contrats de cession des droits s'apparentent plus à des autorisations de reproduction et de diffusion, à des concessions, qu'à de véritables cessions des droits, comme lorsque l'auteur d'un ouvrage confie la commercialisation de son œuvre à un éditeur.

LE PRIX

Il est en théorie fixé librement entre les parties. Dans de nombreux cas, notamment lorsque l'autorisation est délivrée par une société de gestion collective, les utilisations sont facturées selon des barèmes existants.

Le prix de la cession doit figurer dans le contrat.

La rémunération de l'auteur doit être proportionnelle aux produits de l'exploitation. Dans certains cas limitativement énumérés par le CPI³⁸ la cession peut être forfaitaire.

Une cession gracieuse des droits ne peut être envisagée qu'en cas d'exploitation strictement non commerciale de l'œuvre.

Pour aller plus loin...

Bien que fondamental, le respect de la propriété littéraire et artistique ne sera pas toujours suffisant. En effet, une fois tombée dans le domaine public, l'œuvre ne sera pas nécessairement libre de réutilisation. Certaines conditions liées au respect du droit moral, de la vie privée, du droit à l'image mais également de la domanialité publique ou du droit de la concurrence pourront en limiter l'utilisation.

SCHEMA RÉCAPITULATIF DU DROIT D'AUTEUR**DÉCÈS DE L'AUTEUR**

ou date de publication
pour les œuvres collectives,
anonymes, pseudonymes

+ 30 ANS EN
CAS DE
« MORT POUR
LA FRANCE »

70 ANS**monopole d'exploitation de l'auteur****domaine public***oui*

Auteur / Ayants droit
Éditeur (y compris pour
les œuvres épuisées)
Producteur
Société de gestion

Le titulaire
des droits est
identifiable et
peut être
contacté :

non

L'œuvre est orpheline :
pas d'utilisation
possible à ce jour
(régime de licence
légale en projet)

CAS DES INÉDITS
POSTHUMES :
25 ANS DE
PROTECTION
À COMPTER DE
LA PUBLICATION

PARTIE III

MONTER UN MARCHÉ PUBLIC DE NUMÉRIISATION

MONTER UN MARCHÉ PUBLIC DE NUMÉRIISATION

par Stéphane Alcantre

LES GRANDS PRINCIPES DU CODE DES MARCHÉS PUBLICS

+++++

DÉFINITION ET CHAMP D'APPLICATION

Le Code des marchés publics, en vigueur depuis le 1^{er} septembre 2006, résulte du décret n° 2006-975 du 1^{er} août 2006. Ce texte a été modifié³⁸ plusieurs fois depuis cette date mais les modifications apportées ont essentiellement consisté à y introduire des assouplissements, des éclaircissements rédactionnels et des adaptations techniques.

L'article 1^{er} du code définit la notion de marché public et donne ainsi la dimension matérielle du champ d'application du code. Ainsi, les marchés publics sont définis comme étant les contrats conclus à titre onéreux par

38. Référence des textes du Code des marchés : Décret n° 2009-1702 du 30 décembre 2009 modifiant les seuils applicables aux marchés passés en application du Code des marchés publics et de l'ordonnance n° 2005-649 du 6 juin 2005 relative aux marchés passés par certaines personnes publiques ou privées non soumises au Code des marchés publics et aux contrats de partenariat.

Décret n° 2009-1456 du 27 novembre 2009 relatif aux procédures de recours applicables aux contrats de la commande publique.

Décret n° 2009-1086 du 2 septembre 2009 tendant à assurer l'effet utile des directives 89/665/CEE et 92/13/CEE et modifiant certaines dispositions applicables aux marchés publics.

Décret n° 2008-1356 du 19 décembre 2008 relatif au relèvement de certains seuils du Code des marchés publics. Par un arrêt du 10 février 2010, Maître Perez c/ ministre de l'Économie, le Conseil d'État a annulé partiellement ce décret en ce qu'il a relevé le seuil de 4 000 € HT à 20 000 € HT. Cette annulation entre en vigueur à compter du 1^{er} mai 2010. À compter de cette date, le seuil de 20 000 € HT est donc remplacé par 4 000 € HT.

Décret n° 2008-1355 du 19 décembre 2008 de mise en œuvre du plan de relance économique dans les marchés publics.

Décret n° 2008-1334 du 17 décembre 2008 modifiant diverses dispositions régissant les marchés soumis au Code des marchés publics et aux décrets pris pour l'application de l'ordonnance n° 2005-649 du 6 juin 2005 relative aux marchés passés par certaines personnes publiques ou privées non soumises au Code des marchés publics.

Décret n° 2008-585 du 19 juin 2008 adaptant à Mayotte certaines dispositions du Code des marchés publics.

Décret n° 2008-407 du 28 avril 2008 modifiant l'article 98 du Code des marchés publics.

Décret n° 2008-206 du 29 février 2008 modifiant le Code des marchés publics.

Décret n° 2007-1850 du 26 décembre 2007 modifiant les seuils applicables aux marchés passés en application du Code des marchés publics et de l'ordonnance n° 2005-649 du 6 juin 2005 relative aux marchés passés par certaines personnes publiques ou privées non soumises au Code des marchés publics et aux contrats de partenariat.

les pouvoirs adjudicateurs³⁹ définis à l'article 2 et des opérateurs économiques publics ou privés, pour répondre à leurs besoins en matière de travaux, de fournitures ou de services.

Le caractère onéreux du marché s'apprécie de manière extensive. Ainsi, il n'est pas nécessaire qu'un prix soit payé par la personne publique à son cocontractant pour que le marché soit considéré comme conclu à titre gratuit. En effet, d'après une jurisprudence du Conseil d'État désormais bien établie et illustrée par un exemple bien connu, un contrat par lequel une collectivité locale confie à une entreprise le soin d'implanter du mobilier urbain sur son domaine en l'exonérant de redevance d'occupation et en l'autorisant à exploiter pour son propre compte les emplacements publicitaires ménagés sur ledit mobilier, est bien un marché public car il est conclu à titre onéreux. Il en irait de même si une bibliothèque publique venait à confier à un prestataire la numérisation de ses collections sans lui en payer le prix mais en l'autorisant à effectuer une exploitation commerciale des documents numérisés sans qu'il ait à verser une redevance au titre de cette exploitation.

L'article 1^{er} définit également les notions de travaux, de prestations de service et de fournitures. La numérisation de collections relève à l'évidence de la catégorie des prestations de service. Toutefois, un marché dont l'objet concerne l'achat de matériel de numérisation relève de la catégorie des fournitures. La distinction entre ces deux catégories ne constitue toutefois pas un enjeu juridique majeur dans la mesure où le Code des marchés leur applique le même traitement, contrairement à ce qu'il prévoit en matière de travaux.

L'article 2 du code donne, quant à lui, la dimension organique du champ d'application de son dispositif : il définit les personnes auxquelles le code s'applique. Il s'agit de l'État et de ses établissements autres que ceux ayant un caractère industriel et commercial, des collectivités territoriales et des établissements publics locaux.

39. La notion de pouvoir adjudicateur est issue de la réglementation communautaire et vise à recouvrir toutes les formes de structures, différentes selon les pays de l'Union, qui sont soumises en code. Afin de faciliter la compréhension, nous utiliserons la notion de personne publique qui est moins large que celle de pouvoir adjudicateur.

PRINCIPES FONDAMENTAUX

L'article 1^{er} du code les énumère. Les marchés publics respectent les principes de liberté d'accès à la commande publique, d'égalité de traitement des candidats et de transparence des procédures. Et le code ajoute que ces principes permettent d'assurer l'efficacité de la commande publique et la bonne utilisation des deniers publics.

Quel que soit son montant, un marché public est soumis au respect de ces principes fondamentaux. Ils irriguent l'ensemble des 113 articles de la première partie du code dont les dispositions visent à permettre de les respecter au mieux. Le code est donc avant tout un outil pour l'acheteur. Il lui permet de garantir le respect des principes fondamentaux et de se prémunir contre une infraction qui pourrait avoir pour lui des conséquences pénales.

En effet, l'article 432-14 du Code pénal dispose : *« Est puni de deux ans d'emprisonnement et de 30 000 euros d'amende le fait par une personne dépositaire de l'autorité publique ou chargée d'une mission de service public ou investie d'un mandat électif public ou exerçant les fonctions de représentant, administrateur ou agent de l'État, des collectivités territoriales, des établissements publics, des sociétés d'économie mixte d'intérêt national chargées d'une mission de service public et des sociétés d'économie mixte locales ou par toute personne agissant pour le compte de l'une de celles susmentionnées de procurer ou de tenter de procurer à autrui un avantage injustifié par un acte contraire aux dispositions législatives ou réglementaires ayant pour objet de garantir la liberté d'accès et l'égalité des candidats dans les marchés publics et les délégations de service public. »*

Le délit prévu par l'article 432-14 du Code pénal est donc constitué par la simple infraction aux principes fondamentaux posés par le Code des marchés publics. Le non-respect de l'une des procédures qu'il prévoit, dès lors que celles-ci visent justement à garantir le respect desdits principes, suffit à constituer l'infraction pénale. Contrairement à une idée reçue, il n'est point besoin d'en avoir tiré un avantage matériel ou financier particulier pour être passible d'une condamnation.

LES SEUILS D'APPLICATION DU CODE DES MARCHÉS PUBLICS

LES MODALITÉS D'APPRÉCIATION DES SEUILS

Le Code des marchés distingue deux catégories de procédures : les procédures qu'il formalise de manière expresse et celles qu'il ne formalise pas tout en les encadrant. Les marchés conclus en application d'une procédure formalisée seront dits marchés formalisés, les autres seront dits marchés à procédure adaptée (MAPA).

Le code fixe en son article 26 les seuils financiers au-delà desquels le recours à une procédure formalisée sera obligatoire pour passer un marché. Pour les prestations de services et les fournitures, ce seuil est de 125 000 € HT pour l'État et ses établissements publics et de 193 000 € HT⁴⁰ pour les collectivités territoriales et les établissements publics locaux.

Le code fixe les modalités financières d'appréciation du seuil (articles 27, 72, 76 et 77). Ce dernier s'entend sur la durée totale du marché, qu'il soit pluriannuel ou non. Ainsi, s'il s'agit d'un marché d'une année reconductible une ou plusieurs fois, c'est le montant global du marché sur sa durée maximale qui devra être pris en compte. De même, lorsqu'un marché comporte une tranche ferme et une ou plusieurs tranches conditionnelles, c'est le montant toutes tranches comprises qui servira de référence. S'il s'agit d'un marché à bons de commande comportant un montant maximum, c'est ce montant maximum qui doit être utilisé. Si le marché à bons de commande ne comporte pas de montant maximum, il sera présumé dépasser le seuil. Enfin, si la procédure de passation comporte plusieurs lots, c'est le montant cumulé des lots qui sera pris en compte.

Le code fixe également les modalités matérielles d'appréciation du seuil (article 27). Il doit être procédé à une estimation de la valeur totale des fournitures ou des services qui peuvent être considérés comme homogènes soit en raison de leurs caractéristiques propres, soit parce qu'ils constituent une unité fonctionnelle. Ainsi, tous les besoins de même nature de la personne publique, qu'ils fassent l'objet de commandes par un ou plusieurs

40. Ces seuils sont susceptibles d'évoluer tous les deux ans en fonction des accords conclus par l'Union européenne dans le cadre de l'Organisation mondiale du commerce. La dernière modification en date est entrée en vigueur le 1^{er} janvier 2010.

services doivent être cumulés. Si la personne publique entend lancer plusieurs marchés de numérisation au cours de la même année, leurs montants devront être additionnés et c'est ce montant global qui devra être pris en compte pour l'appréciation du seuil. Qu'il s'agisse de numériser des imprimés (monographies ou presse), des documents audiovisuels ou photographiques, les prestations sont de même nature et doivent être appréciées dans leur globalité. De même, si une personne publique se lance dans un vaste programme de numérisation de ses collections, sur plusieurs années, le programme en question constituera une unité fonctionnelle au sens de l'article 27 du code. Rappelons enfin que le code précise que la personne publique ne peut pas se soustraire à l'application des dispositions qu'il prévoit en scindant ses achats ou en utilisant des modalités de calcul de la valeur estimée des marchés autres que celles issues de l'article 27.

Pour la passation des marchés formalisés, outre le respect de la procédure fixée par le code et décrite ci-après, la consultation doit faire l'objet d'une mesure de publicité préalable au *Bulletin officiel des annonces des marchés publics* (BOAMP), au *Journal officiel de l'Union européenne* (JOUE) et sur le profil d'acheteur⁴¹ de la personne publique.

LES MARCHÉS À PROCÉDURE ADAPTÉE (MAPA)

Lorsque les seuils précités ne sont pas franchis, l'article 28 du code prévoit que les marchés concernés sont passés selon une procédure adaptée dont les modalités sont librement fixées par la personne publique en fonction de la nature et des caractéristiques du besoin à satisfaire, du nombre ou de la localisation des opérateurs économiques susceptibles d'y répondre ainsi que des circonstances de l'achat. Ces modalités peuvent s'inspirer des procédures formalisées prévues par le code.

41. Le profil d'acheteur est un site, généralement appelée « plateforme », accessible en ligne, par l'intermédiaire du réseau Internet, offrant toutes les fonctionnalités nécessaires à la dématérialisation des procédures : au minimum, information des candidats et réception des candidatures et des offres. Le site Internet d'une personne publique ne peut être qualifié de profil d'acheteur que s'il offre l'accès à ces fonctionnalités. Il constitue donc généralement une rubrique à l'intérieur d'un site sur la toile.

Afin de mettre en œuvre ces mesures, qui laissent en réalité une grande latitude aux personnes publiques, la plupart d'entre elles se sont dotées d'une procédure interne dont le degré de précision et d'exigence formelle va croissant en fonction du montant du marché (voir *infra* pour les précautions à prendre). Que ces règles générales internes existent ou non au sein de la personne publique concernée, cette dernière peut toujours décider de mettre en œuvre une procédure spécifique propre à un marché. En tout état de cause, l'édition de ces mesures devra respecter les règles applicables à chaque personne publique (décision du représentant légal, délibération du conseil d'administration, délibération du conseil délibérant de la collectivité territoriale...).

Le tableau ci-après récapitule les seuils prévus par le code pour les fournitures et les services :

TABEAU	LES SEUILS PRÉVUS PAR LE CODE DES MARCHÉS PUBLICS POUR LES MARCHÉS FOURNITURES ET SERVICES p. 72
---------------	---

montant du marché	publicité	mise en concurrence
En dessous de 20 000 euros HT jusqu'au 30 avril 2010, en dessous de 4 000 euros HT à compter du 1 ^{er} mai 2010	Pas de modalités formelles de publicité préalables	Pas de modalités formelles de mise en concurrence préalables. La personne publique peut décider qu'il sera passé sans publicité ni mise en concurrence préalable. Il peut être recommandé de mettre en œuvre une procédure formalisée de comparaison par exemple avec la production de plusieurs devis. Négociation possible.
Entre 20 000 euros HT et 89 999 euros HT	Publicité préalable obligatoire adaptée. À considérer en proportion avec le montant du marché. La personne publique décide en fonction de différents critères mais les modalités doivent « susciter la concurrence nécessaire »	Procédure préalable obligatoire adaptée. Mise en concurrence obligatoire selon les modalités du « marché passé selon une procédure adaptée » (MAPA) article 28 du Code des marchés. La personne publique doit définir elle-même les règles de mise en concurrence proportionnées à l'objet et au montant du marché. Négociation possible.
Entre 90 000 euros HT et les seuils communautaires (soit 124 999 euros HT pour les marchés Fournitures et services de l'État et 192 999 euros HT pour les marchés Fournitures et services des collectivités territoriales)	Publicité obligatoire. Les modalités sont strictement définies par le Code des marchés (article 40) : publication au <i>Bulletin officiel des annonces des marchés publics</i> (BOAMP) ou dans un <i>Journal d'annonces légales</i> (JAL) ainsi que sur le profil d'acheteur de la personne publique. L'annonce dans la presse spécialisée est recommandée.	Procédure adaptée. Cadre du MAPA. Négociation possible.
Au-dessus des seuils communautaires (supérieur à 125 000 euros HT ou 193 000 euros HT)	Publicité obligatoire. Les modalités sont strictement définies par le Code des marchés : publication au <i>Bulletin officiel des annonces des marchés publics</i> (BOAMP) et au <i>Journal officiel de l'Union européenne</i> (JOUE). L'annonce dans la presse spécialisée est recommandée.	Procédure formalisée. Mise en concurrence obligatoire. Marchés passés sur appels d'offres mais possibilité sous certaines conditions de recourir à un marché négocié (art. 35), à une procédure de dialogue compétitif (article 36). D'autres types de procédures sont prévus par le Code des marchés.

LES PROCÉDURES DE PASSATION

+++++

L'appel d'offres est la procédure formalisée de droit commun prévue par le code pour la passation des marchés. Contrairement aux autres procédures, son utilisation est inconditionnelle.

L'APPEL D'OFFRES OUVERT

L'appel d'offres ouvert est traité par les articles 57 à 60 du Code des marchés. Il est la procédure formalisée la plus couramment utilisée. Un avis d'appel public à la concurrence est publié au BOAMP⁴² et au JOUE. Les entreprises disposent alors d'un délai minimum de 52 jours pour présenter leur candidature et leur offre. Il s'agit là du délai minimum de droit commun. Il peut être réduit dans les conditions fixées par l'article 57⁴². Toutefois, les prestations de numérisation sont souvent complexes et la rédaction des offres par les entreprises peut s'avérer fastidieuse. Il convient donc de n'utiliser les facultés de réduction des délais offertes par le code qu'avec parcimonie. Les candidats doivent pouvoir disposer du temps suffisant pour élaborer une offre complète et pertinente tant sur le plan technique que sur le plan financier.

Le point de départ de ce délai est le moment de l'envoi de l'avis au BOAMP et non pas sa date de publication effective⁴³. En pratique, elle a lieu en général 48 heures après l'envoi.

L'avis est constitué d'une annonce rédigée selon un modèle fixé par le règlement communautaire n° 1564/2005 repris par un arrêté ministériel et par le BOAMP. Le modèle comporte des rubriques dont beaucoup sont obligatoires et qu'il suffit de compléter. Dès réception, le BOAMP procède lui-même à l'envoi de l'avis pour publication au JOUE.

Outre l'annonce publiée, le dossier de consultation des entreprises (DCE)

42. Le délai peut notamment être réduit de sept jours lorsque l'avis d'appel public à la concurrence est envoyé par voie électronique. Il peut encore être réduit de cinq jours supplémentaires lorsque l'ensemble des pièces du dossier de consultation sont accessibles librement, directement et intégralement par voie électronique à une adresse Internet précisée dans l'avis. Le délai de remise des offres peut ainsi être porté à 40 jours minimum.

43. Le BOAMP dispose d'un délai de six jours au maximum pour procéder à la publication.

est constitué d'un règlement particulier de la consultation (RPC). Il peut reprendre les informations essentielles figurant dans l'annonce tout en les complétant, en précisant les modalités pratiques de remise des offres (lieu et heure précis de remise des offres par exemple, documents administratifs à fournir, conditions éventuelles d'analyse des candidatures...). Il convient d'apporter le plus grand soin à sa rédaction. Le règlement complète et précise l'annonce, il ne doit pas s'y substituer ou la vider de sa substance (cela vaut également pour les MAPA). Le dossier comporte ensuite l'ensemble des pièces qui deviendront celles du marché ainsi que toutes les documentations techniques qui pourraient être utiles aux entreprises pour constituer leur offre. Il est par exemple possible, lorsque le marché portera sur la réalisation d'une prestation de conversion en mode texte, de fournir un échantillon de documents déjà numérisés représentatifs afin de permettre aux entreprises de les traiter et de mesurer ainsi l'ampleur de la prestation.

Les candidatures et les offres seront remises par les soumissionnaires aux heures et dates indiquées dans le DCE. Le code précise qu'il s'agit bien d'une date limite. Le cachet de la Poste, qui est représentatif d'une date d'envoi d'un pli, n'aura donc aucune valeur. Il est donc fortement conseillé d'indiquer dans le RPC la date et l'heure limites de remise des offres ainsi que le lieu de cette remise, car à défaut de pouvoir effectuer un envoi postal, les entreprises feront directement déposer leur pli dans les locaux de la personne publique.

Les plis sont ensuite ouverts. Il est fortement conseillé de dresser un procès-verbal de cette ouverture. Il est ensuite procédé à l'**analyse des candidatures**. Celle-ci ne peut être faite que sur la base de critères ou de niveaux de capacité minimum à condition qu'ils soient énoncés dans le RPC. Il est possible de demander un certain nombre de renseignements aux entreprises, sur la base desquels l'analyse des candidatures sera effectuée. Toutefois, la nature des renseignements qui peuvent être demandés au stade de la candidature est limitativement fixée par un arrêté ministériel d'application du Code des marchés en date du 28 août 2006. Il précise que les renseignements demandés doivent être nécessaires à l'appréciation des capacités des candidats. Par ailleurs, afin de ne pas pénaliser les entreprises nouvellement créées, l'article 52 précise que l'absence de référé-

rence d'un candidat ne suffit pas à l'écarter et ne dispense pas la personne publique de procéder à une véritable analyse de ses capacités techniques, financières et professionnelles. Une fois l'analyse effectuée, la liste des candidats retenus est arrêtée. Pour les collectivités territoriales, c'est la commission d'appel d'offres qui arrête cette liste et dresse le procès-verbal de la réunion qu'elle a tenue à cet effet. Les candidats non retenus en sont immédiatement informés et les motifs de rejet de leur candidature leur sont communiqués.

En matière d'appel d'offres ouvert et lorsque le marché ne nécessite pas de capacités particulières, l'analyse des candidatures peut se réduire à une simple vérification de points élémentaires comme le chiffre d'affaires, l'ampleur du personnel et les références. Il y est alors procédé en même temps que l'ouverture des plis (nombre de collectivités territoriales y font procéder en séance de la commission d'appel d'offres qui procède ainsi également à l'ouverture des plis). Si la personne publique souhaite apporter un soin particulier à l'analyse des candidatures la procédure de l'appel d'offres restreint [cf. plus loin] peut s'avérer beaucoup plus adaptée à la situation.

L'**analyse des offres** est ensuite entamée sur la base des **critères** énoncés dans le DCE (annonce et RPC). Si plusieurs critères sont utilisés, ils doivent être pondérés. Si des sous-critères sont utilisés à l'intérieur des critères, dès lors qu'ils ne se déduisent pas directement des critères eux-mêmes, ils doivent également figurer dans le DCE et être également pondérés. Si un seul critère est utilisé, ce sera obligatoirement le prix. Toute négociation avec les entreprises est strictement interdite.

Un **rapport d'analyse des offres** est dressé et le marché est attribué à l'entreprise ayant présenté l'offre économiquement la plus avantageuse (le mot économie n'est ici pas entendu par le Code des marchés dans sa seule dimension financière). Pour les collectivités territoriales, c'est la commission d'appel d'offres qui procède à cette attribution sur la base du rapport d'analyse.

Les entreprises, dont les offres n'ont pas été retenues, en sont immédiatement informées et les motifs de rejet de leurs offres communiqués. La personne publique doit enfin attendre un délai minimal de onze jours avant de signer le marché. Ce délai doit permettre aux entreprises évincées qui

le souhaitent d'exercer un référé précontractuel devant le tribunal administratif si elles estiment que les obligations de publicité et de mise en concurrence prévues par le Code des marchés n'ont pas été respectées par la personne publique. Une fois signé, le marché est notifié à son titulaire. Son exécution ne peut commencer qu'une fois cette notification effectuée.

L'APPEL D'OFFRES RESTREINT

Il est prévu par les articles 60 à 64 du code. Toutes les règles applicables à l'appel d'offres ouvert relatives à la publication au BOAMP et au JOUE, au contenu du dossier de consultation et aux modalités d'appréciation des candidatures et des offres sont identiques.

Il organise la sélection des soumissionnaires en deux étapes. La première concerne les candidatures. L'annonce publiée fixe leurs conditions d'examen. À ce stade, il ne comporte aucun dossier de consultation et se contente de donner les grandes lignes du marché dont la procédure de passation est lancée. Il peut prévoir qu'un nombre minimum et/ou un nombre maximum de candidats sera retenu. Si un nombre minimum est prévu, il ne peut être inférieur à cinq. Les candidats disposent d'un délai minimum de 37 jours pour déposer leur candidature⁴⁴

Les candidatures sont ensuite analysées et retenues selon les mêmes modalités que pour l'appel d'offres ouvert. Après information des candidats évincés, une lettre de consultation est adressée simultanément aux candidats retenus. Cette lettre comporte l'ensemble des pièces du dossier de consultation (ou les moyens de les obtenir, adresse physique ou électronique par exemple), la date, l'heure limite et le lieu de remise des offres, la mention indiquant que les documents de l'offre doivent être rédigés en français et la liste des documents à fournir avec l'offre. Le délai minimum de remise des offres est de 40 jours à compter de l'envoi de la lettre de consultation. Il peut être réduit dans les conditions fixées par l'article 57⁴⁵. Toutefois, comme pour l'appel d'offres ouvert, il convient de laisser aux entreprises

44. Le délai peut être ramené à 30 jours si l'avis a été adressé par voie électronique.

45. Le délai peut notamment être réduit de cinq jours lorsque l'ensemble des pièces du dossier sont librement, directement et intégralement accessibles par voie électronique à une adresse Internet précisée dans la lettre de consultation.

un délai suffisant pour établir leurs offres. L'ouverture des plis, l'analyse des offres, l'attribution du marché et l'information des soumissionnaires non retenus sont effectuées dans les mêmes conditions que pour l'appel d'offres ouvert. Là encore, toute négociation avec les entreprises est strictement interdite.

LES AUTRES PROCÉDURES FORMALISÉES

Les autres procédures formalisées nécessitent qu'un certain nombre de conditions soient remplies pour pouvoir être utilisées.

Comme son nom l'indique, la procédure du marché négocié autorise la personne publique à négocier avec les soumissionnaires. Il existe deux catégories de marchés négociés : ceux négociés après publicité et mise en concurrence et ceux négociés sans publicité et sans mise en concurrence.

> Les marchés négociés après publicité et mise en concurrence (article 35-I). Parmi les cinq cas possibles de recours au marché négocié, seul le premier est possible pour un marché de numérisation. Il s'agit de la situation où un appel d'offres a été lancé et les offres se sont révélées irrégulières⁴⁶ ou inacceptables⁴⁷. Il est alors possible de relancer une nouvelle procédure à condition que le dossier de consultation ne soit pas modifié. Les modalités de publicité⁴⁸ sont identiques à celles de l'appel d'offres et le déroulement de la procédure s'apparente à celles de l'appel d'offres restreint. Il existe toutefois trois différences notables :

- le nombre minimum de candidats, s'il est prévu, ne peut être inférieur à trois ;
- le délai de remise des offres est librement fixé par la personne publique (en prenant soin qu'il soit suffisant) ;
- les offres font l'objet d'une négociation. La négociation doit être menée

46. Une offre qui apporte une réponse au besoin mais est incomplète ou ne respecte pas les exigences formulées dans le dossier de consultation.

47. Une offre qui méconnaît la réglementation en vigueur ou dont les crédits budgétaires qui ont été alloués sur la base d'une estimation ne permettent pas de la financer.

48. Il est possible de se dispenser d'une mesure de publicité lorsque la procédure nouvelle de passation du marché est conduite avec les seules entreprises ayant présenté une offre à l'occasion du premier appel d'offres déclaré infructueux.

dans le respect de l'égalité de traitement des soumissionnaires et les informations qui leur sont communiquées le cas échéant, ne doivent pas en avantager certains.

> Les marchés négociés sans publicité et sans mise en concurrence (article 35-II) peuvent être utilisés dans dix cas de figure différents. Ces cas doivent être interprétés de manière restrictive et les conditions prévues par le code strictement respectées.

Le **dialogue compétitif** est la procédure par laquelle la personne publique conduit un dialogue avec les soumissionnaires. Elle ne sera citée ici que pour mémoire. Ses conditions restrictives d'utilisation, sa longueur et sa complexité rendent son usage peu fréquent. En tout état de cause il ne saurait être utilisé sans une analyse préalable approfondie, avec l'assistance de spécialistes tant sur le plan juridique que sur le plan technique.

LES MAPA

Tous les marchés qui n'atteignent pas le seuil des procédures formalisées prévues par le code, sont des MAPA.

Dès lors que les principes fondamentaux énoncés à l'article 1^{er} du code (liberté d'accès à la commande publique, égalité de traitement des candidats, transparence des procédures, efficacité de la commande publique et bonne utilisation des deniers publics) sont applicables à l'ensemble des marchés quel que soit leur montant, les procédures imaginées par les personnes publiques pour encadrer la passation des MAPA devront tendre à en garantir le respect. Il est donc recommandé de s'inspirer des règles prévues par le code pour la passation des marchés formalisés⁴⁹ Il conviendra ainsi de ne pas hésiter à utiliser largement des mesures de publicité préalables : dans un journal d'annonces légales, dans un journal professionnel à diffusion large, sur un site Internet spécialisé dans l'achat public, sur le site Internet de la personne publique...

49. Sachant qu'il est toujours possible, si la personne publique le souhaite, de passer un marché selon une procédure formalisée même s'il n'atteint pas le seuil prévu par le code. L'ensemble des règles de la procédure devra alors être respecté sans le moindre aménagement.

Si aucune publicité n'est effectuée dans un journal d'annonces légales, il est conseillé de ne pas se limiter à publier une annonce sur le seul site Internet de la personne publique, en complétant par une publicité sur un site Internet spécialisé dans la publication d'annonces relatives aux marchés publics. Nombre d'opérateurs proposent à ce jour ce service. Plus le montant du marché sera élevé, plus les mesures de publicité devront être efficaces.

Il est également indispensable, pour garantir l'égalité de traitement des candidats et la transparence de la procédure que le déroulement de cette dernière soit décrit dans le dossier de consultation qui doit lui-même comprendre une description claire des modalités d'attribution du marché et en particulier des critères d'analyse des offres et de leur poids respectif (notion de pondération).

De même, le fait que le montant du marché soit faible, ne dispense pas de bien définir le besoin. En matière de numérisation, cette bonne définition ne saurait se passer d'indications aussi précises que possible sur les documents à numériser, sur la nature et les caractéristiques des fichiers contenant les données numérisées, sur le niveau de qualité attendu pour la numérisation en mode texte, sur les modalités de mise en ligne des fichiers le cas échéant... À défaut de précisions suffisantes, les principes fondamentaux de la commande publique pourraient être considérés comme non respectés.

Il est également conseillé de dresser des procès-verbaux lors de la réception des plis et de leur ouverture et de rédiger un **rapport d'analyse des offres** qui relate le contenu des offres, avec leurs défauts et leurs qualités au regard du besoin tel qu'il a été défini dans le cahier des charges et qui justifie que le choix du titulaire du marché s'est bien porté sur l'offre économiquement la plus avantageuse.

Quelle que soit la procédure mise en place pour leur passation, les MAPA peuvent faire l'objet d'une négociation. La possibilité de négocier doit être annoncée dans les documents de consultation. Elle peut porter sur tous les éléments de l'offre et notamment sur le prix (article 28). Elle doit toutefois respecter les principes fondamentaux du Code des marchés. Ainsi, elle ne doit pas conduire à une modification substantielle du cahier des charges initial, car cela remettrait en question la publicité et la mise en concurrence qui ont été effectuées. De même, elle ne doit pas porter atteinte au principe d'égalité de traitement des candidats qui doivent tous bénéficier

du même niveau d'information durant toute la négociation. Si un candidat n'est pas retenu pour la négociation, cette décision doit être motivée par des éléments objectifs. Par ailleurs, la négociation ne doit pas conduire à permettre à une entreprise de refaire intégralement son offre, ce serait lui donner un avantage discriminatoire vis-à-vis des autres soumissionnaires. Une offre qui est trop éloignée du besoin, qui ne satisfait largement pas au cahier des charges ou dont le prix est trop éloigné des autres offres ou de l'estimation n'a pas à faire l'objet d'une négociation car celle-ci conduirait inévitablement à refaire l'offre initiale.

Pour encadrer cette négociation, il est conseillé de procéder à une analyse des offres complète et de rédiger un rapport d'analyse avant négociation. Ce rapport comportera une liste de questions qui seront posées par écrit aux entreprises. Un délai raisonnable sera laissé à celles-ci afin de répondre, puis les réponses seront à leur tour analysées et le rapport complété. L'ensemble du processus se déroulera en ayant toujours à l'esprit le cahier des charges et les critères d'analyse qu'il comportait.

LES GROUPEMENTS DE COMMANDES

Le code prévoit la possibilité de regrouper plusieurs personnes morales pour organiser ensemble la passation d'un marché. Ces personnes peuvent être l'État, des établissements publics de l'État, des collectivités territoriales, des établissements publics locaux et des personnes privées. Toutes les combinaisons sont possibles à condition que l'ensemble des membres du groupement appliquent les règles du Code des marchés. Une convention doit être signée entre tous les membres. Elle désigne parmi eux un coordonnateur qui sera chargé de mettre en œuvre la procédure de passation. Chaque membre s'engage dans la convention à passer un marché à hauteur de ses besoins préalablement définis. Lorsque le groupement comporte au moins une collectivité territoriale ou un établissement public local, une commission d'appel d'offres est constituée. Chaque membre du groupement y est représenté et elle est présidée par le représentant du coordonnateur. Lorsque les collectivités territoriales ou les établissements publics locaux sont majoritaires, le marché sera attribué par la commission d'appel d'offres. Lorsqu'ils sont minoritaires, le marché est attribué après avis de la commission. Lorsque le groupement ne comporte ni collecti-

vité territoriale, ni établissement public local, c'est la convention constitutive du groupement qui définit les modalités d'attribution du marché.

Le groupement peut revêtir trois formes différentes :

- chaque membre signera un marché avec le titulaire désigné collectivement et suivra l'exécution de son propre marché. Les partenaires du groupement se séparent donc au moment de l'attribution du marché ;
- le coordonnateur signe et notifie le marché, mais chaque membre en suit l'exécution pour la partie qui le concerne ;
- le coordonnateur, signe, notifie et suit l'exécution du marché pour le compte du groupement. La collaboration se poursuit ainsi sur toute la durée du marché.

La mise en place du groupement peut être lourde. Toutefois, dès lors qu'il est en place, il peut permettre de rassembler dans un même marché les collections de plusieurs institutions que certaines n'auraient pas possédées seules, ainsi que les moyens techniques, administratifs ou humains nécessaires au lancement de l'opération de numérisation. Par ailleurs, le groupement peut développer des synergies entre les institutions. Enfin, lorsque l'opération nécessite des investissements de la part du titulaire du marché, le groupement peut permettre d'atteindre un volume financier suffisant pour que ceux-ci puissent être amortis. Il est en cela facteur d'efficacité de l'utilisation des deniers publics.

LE CONTENU DU MARCHÉ

+++++

LA FORME DU MARCHÉ

L'allotissement

L'article 10 du Code des marchés publics prévoit que les personnes publiques passent les marchés par lots séparés. Il s'agit d'un principe dont l'objectif est de susciter la plus large concurrence possible. Lorsque le besoin est important, l'allotissement permettra de constituer des lots dont l'ampleur sera moindre, ce qui permettra à des entreprises de taille plus réduite de soumissionner. Chaque lot constituera un marché indépendant des autres

lots même si certaines pièces du marché peuvent être identiques à tous les lots. En matière de numérisation, on peut envisager des lots en fonction de la nature des collections à traiter et de la chaîne de production qui en découle. Il est également possible de prévoir, par exemple, un lot pour la numérisation et un lot pour la mise en ligne des documents. Une entreprise peut présenter une offre pour plusieurs lots et se trouver attributaire d'un ou plusieurs lots. Elle ne peut toutefois pas proposer des offres variables en fonction du nombre de lots dont elle serait susceptible d'être attributaire. La personne publique analysera, quant à elle, chaque offre pour chaque lot de manière indépendante.

Le principe posé par l'article 10 n'est applicable que lorsqu'il est possible d'identifier des prestations distinctes les unes des autres. Il ne s'agit pas d'opérer des découpages artificiels pour créer des lots qui, en réalité, ne peuvent s'exécuter de manière séparée.

L'article 10 prévoit également des exceptions. L'**allotissement** est ainsi à écarter si la personne publique estime que la dévolution en lots séparés est de nature à restreindre la concurrence (le volume financier est trop faible pour susciter un intérêt pour les opérateurs économiques), ou risque de rendre techniquement difficile l'exécution des prestations (par exemple lorsque la coordination de plusieurs entreprises s'avère difficile), ou risque de rendre financièrement coûteuse l'opération (par exemple l'exécution de la prestation nécessite un volume financier suffisant pour permettre l'amortissement d'un investissement initial du titulaire). Les personnes publiques peuvent donc ne pas allouer le marché, à condition que ce choix soit motivé.

Formes de prix

Le marché peut revêtir deux formes de prix différentes : un prix forfaitaire ou des prix unitaires. Le marché conclu au forfait comporte un ensemble de prestations qui feront l'objet d'un paiement global du prix sur lequel le titulaire s'est engagé en remettant son offre. Cette forme de marché nécessite que les prestations soient clairement et précisément définies et en particulier leurs quantités, puisque même si la quantité finale traitée est différente, seul le montant forfaitaire sera réglé. En matière de numérisation, ce type de marché ne pourra être utilisé que lorsque le fonds à numériser est parfaitement identifié et qu'il nécessite un traitement relativement homogène (par

exemple le nombre de pages des ouvrages imprimés est connu et stable). En cas de doute, il est possible d'introduire dans le marché une variable d'ajustement en prévoyant par exemple que le prix forfaitaire s'applique à une quantité définie qui peut varier de plus ou moins 5 à 10 %. Dès lors que ce taux ne sera pas dépassé, le prix du marché ne sera pas modifié.

Lorsque les quantités à traiter ne peuvent être évaluées avec précision (par exemple le nombre d'ouvrages à traiter est connu mais pas leur nombre de pages), le marché sera conclu à prix unitaires. Les prix sont définis dans un bordereau, le bordereau de prix unitaires (BPU), qui est constitué par la personne publique en fonction de la nature des prestations qu'il recouvre (par exemple un prix à la page numérisée, qui peut être lui-même décomposé en plusieurs prix en fonction du traitement demandé). Les prix unitaires sont renseignés par les soumissionnaires. Le marché sera donc dit à bons de commande (article 77 du Code des marchés) et ceux-ci seront émis par la personne publique en fonction de ses besoins et de la capacité de traitement du titulaire. Le marché comportera un montant ou une quantité minimum et un montant ou une quantité maximum⁵⁰. La personne publique sera engagée juridiquement à hauteur de ce montant ou de cette quantité minimum. S'il n'est pas atteint au cours du marché, le titulaire aura droit au paiement de la marge bénéficiaire qu'il aurait dû réaliser s'il avait été atteint. Il convient donc d'apporter un grand soin à la définition de ce montant ou de cette quantité car c'est lui qui permettra aux entreprises qui remettront une offre de connaître le volume financier minimum du marché qu'elles peuvent escompter.

Les tranches

L'article 72 du Code des marchés prévoit que les marchés peuvent comporter une tranche ferme et une ou plusieurs tranches conditionnelles. Le contenu de chacune des tranches doit être clairement défini dans le marché. Les prestations d'une tranche ne doivent pas dépendre de la suivante. En revanche, il est possible que le recours à une tranche conditionnelle dépende de la bonne exécution de la tranche ferme. Chacune des tranches est affermée de manière

50. Il peut également être conclu sans montant minimum mais avec un maximum, sans maximum mais avec un minimum et enfin sans minimum ni maximum.

expresse et écrite par la personne publique qui notifie sa décision au titulaire. Le marché fixe les modalités de cet affermissement et notamment la période au cours de laquelle il y sera procédé (à partir de quand et jusqu'à quand il peut avoir lieu). La présence de tranches permet d'échelonner dans le temps l'exécution des prestations. Elle est très appréciée par nombre de personnes publiques qui y trouvent le moyen confortable de contractualiser leurs incertitudes budgétaires. Elles ne seront liées que pour le montant prévu pour la tranche ferme. Le titulaire n'aura aucun droit à ce que la ou les tranches conditionnelles soient affermies. Afin d'atténuer cet inconvénient, l'article 72 précise que le marché peut prévoir le versement d'une indemnité de dédit au titulaire en cas de non-affermissment d'une tranche.

LES PIÈCES ET LES CLAUSES DU MARCHÉ

Elles sont définies par les articles 11 à 13 du Code des marchés publics. Tout marché dont le montant atteint la somme de 4 000 € HT doit être conclu sous forme écrite.

Le marché comporte un acte d'engagement (AE). Il n'est pas obligatoire pour les MAPA mais il est recommandé de rédiger un acte simplifié s'en inspirant car ce document comporte des éléments essentiels. Il est la pièce signée par le candidat au moment de la remise de son offre qui sera également signée par le représentant de la personne publique une fois le marché attribué.

Le marché comportera ensuite un cahier des charges constitué de documents généraux et particuliers. Parmi les premiers figureront les cahiers des clauses administratives générales (CCAG, voir *supra*) qui fixent les dispositions générales applicables à une catégorie de marchés. Ces documents sont approuvés par des arrêtés du ministre chargé de l'économie. Parmi les seconds figurera le cahier des clauses administratives particulières (CCAP) qui fixe les dispositions administratives propres à chaque marché et le cahier des clauses techniques particulières (CCTP) qui fixe les dispositions techniques nécessaires à l'exécution des prestations de chaque marché. Ces règles ne sont pas obligatoires pour les MAPA, il est donc possible de regrouper dans un même document les clauses administratives et techniques. Toutefois, la description des prestations de numérisation nécessite d'importants développements relatifs aux documents, aux fichiers informatiques et à la qualité de la numérisation attendue. Il est donc déconseillé

de les mêler aux clauses administratives dont ils pourraient venir atténuer la clarté et la compréhension. Les pièces constitutives du marché doivent obligatoirement comprendre les clauses suivantes :

TABEAU CLAUSES DES PIÈCES CONSTITUTIVES D'UN MARCHÉ PUBLIC

Identification des parties contractantes	AE
Justification de la qualité de la personne signataire pour le compte de la personne publique et le cas échéant la délibération autorisant la signature du marché	AE
Définition de l'objet du marché	AE CCAP CCTP
Référence aux articles et alinéas du code des marchés en application duquel le marché est passé	AE CCAP
Énumération des pièces du marché et ordre de priorité	CCAP
Prix ou modalités de sa détermination	AE
Durée d'exécution du marché ou dates prévisionnelles de début d'exécution et d'achèvement	AE CCAP
Conditions de réception, de livraison ou d'admission des prestations	CCAP
Conditions de règlement et délais de paiement	CCAP
Conditions de résiliation	CCAP
Date de notification du marché	AE
Comptable assignataire des paiements	AE CCAP
Éléments propres aux marchés à tranches conditionnelles le cas échéant	AE CCAP CCTP

Ces règles ne sont pas obligatoires pour les MAPA. Toutefois, ces clauses constituent le minimum nécessaire que doit contenir un contrat. On imagine mal comment un MAPA pourrait être valablement conclu si l'une d'elle venait à manquer.

LES CAHIERS DE CLAUSES GÉNÉRALES APPLICABLES À UNE CATÉGORIE DE MARCHÉS : LES CCAG-TIC ET FCS

Les CCAG sont applicables dès lors que le marché y fait référence dans ses documents particuliers. Cela suffit à rendre contractuel l'ensemble de leurs clauses. Il est possible d'y déroger à condition de mentionner dans les documents particuliers du marché quelles sont les clauses qui y dérogent. Ces dérogations doivent être récapitulées à la fin du CCAP. Il est possible de ne pas faire référence à un CCAG. Toutefois, en particulier pour les MAPA, il est fortement conseillé de les utiliser. En effet, les CCAG contiennent toutes les clauses administratives essentielles qui sont de nature à protéger la personne publique sur le plan juridique⁵¹.

Deux CCAG sont susceptibles de s'appliquer en matière de marchés de numérisation. Si le marché comporte des prestations de numérisation en mode image et/ou en mode texte, c'est le CCAG applicable aux marchés de fournitures courantes et de services⁵² (CCAG-FCS) qui sera utilisé. Si le marché porte sur la mise en ligne de collections numérisées, c'est le CCAG applicable aux marchés publics de techniques de l'information et de la communication⁵³ (CCAG-TIC), édicté par arrêté ministériel du 16 septembre 2009, qui sera utilisé. Il n'est pas possible de faire référence à deux CCAG à la fois.

Un aspect particulièrement important de chacun des CCAG sera ici abordé : la vérification et l'admission des prestations.

51. Ils comportent notamment d'importants articles relatifs à la protection de la main d'œuvre, au respect des règles relatives à la protection de l'environnement, à la protection des données à caractère personnel, à la protection du secret et de la confidentialité, à la réparation des dommages, à l'assurance, à la garantie, au prix et aux modes de règlement, aux délais, aux pénalités, aux opérations de vérification et d'admission des prestations, à la cession des droits de propriété intellectuelle le cas échéant, à la résiliation et aux différends et litiges. Grâce au CCAG, un MAPA dont le texte se réduirait à des dispositions techniques, pourrait néanmoins s'exécuter dans des conditions satisfaisantes.

52. Arrêté ministériel du 19 janvier 2009.

53. Arrêté ministériel du 16 septembre 2009.

Pour le **CCAG-FCS**⁵⁴ : le titulaire avise la personne publique de l'achèvement de la prestation. Celle-ci dispose d'un délai de 15 jours pour opérer les vérifications quantitatives et qualitatives et prendre une décision. À défaut de décision expresse, le silence de la personne publique vaut admission des prestations, passé le délai de 15 jours. La personne publique prononce l'admission lorsque la prestation est réalisée conformément aux prescriptions du cahier des charges. Lorsqu'elle estime que des prestations ne peuvent être admises que moyennant certaines mises au point, elle peut décider d'ajourner l'admission des prestations par une décision motivée.

Cette décision invite le titulaire à présenter à nouveau les prestations mises au point, dans un délai de 15 jours. Le titulaire doit faire connaître son acceptation dans un délai de 10 jours à compter de la notification de la décision d'ajournement. En cas de refus du titulaire ou de silence gardé par lui durant ce délai, le pouvoir adjudicateur a le choix d'admettre les prestations avec réfaction ou de les rejeter dans un délai de 15 jours. À défaut de décision expresse dans ce délai, la prestation est réputée rejetée. Lorsque la personne publique estime que des prestations, sans être entièrement conformes aux stipulations du marché, peuvent néanmoins être admises en l'état, il peut les admettre avec réfaction de prix proportionnelle à l'importance des imperfections constatées. Cette décision doit être motivée. Elle ne peut être notifiée au titulaire qu'après qu'il a été mis à même de présenter ses observations. La réfaction vaut renonciation aux prestations concernées de la part de la personne publique. Enfin, lorsque la personne publique estime que les prestations ne peuvent être admises en l'état, il en prononce le rejet partiel ou total. La décision de rejet doit être motivée. Elle ne peut être prise qu'après que le titulaire a été mis à même de présenter ses observations. Il est tenu d'exécuter à nouveau la prestation prévue par le marché.

Pour le **CCAG-TIC**⁵⁵ : l'installation et la mise en ordre sont réalisées par le titulaire. Il dispose d'un mois à compter de la date contractuelle de livrai-

54. Elles figurent aux articles 22 à 26 du CCAG-FCS.

55. Elles figurent aux articles 23 à 28 du CCAG-TIC.

son pour effectuer la mise en ordre de marche (MOM). Il avise ensuite la personne publique de la date à partir de laquelle les prestations pourront être présentées en vue d'être vérifiées. La personne publique procède alors à des vérifications quantitatives et qualitatives. Les opérations de vérifications qualitatives comprennent deux étapes :

- la première est la **vérification d'aptitude (VA)** qui intervient après la mise en ordre de marche. Elle a pour objet de constater que les prestations, livrées ou exécutées, présentent les caractéristiques techniques qui les rendent aptes à remplir les fonctions précisées dans les documents particuliers du marché. Le délai imparti à la personne publique pour y procéder et notifier sa décision est d'un mois à compter de la mise en ordre de marche ;
- si la décision de VA est positive, la **vérification de service régulier (VSR)** débute. Elle a pour objet de constater que les prestations fournies sont capables d'assurer un service régulier dans les conditions normales d'exploitation prévues dans les documents particuliers du marché. Elle s'observe pendant un mois, à partir du jour de la décision positive de VA. Le service est réputé régulier si la durée cumulée, sur le mois, des indisponibilités imputables à chaque élément de matériel ne dépasse pas 2 % de la durée d'utilisation effective.

À chacune des deux étapes, la personne publique peut prononcer la réception, l'ajournement, la refaction ou le rejet des prestations dans des conditions équivalentes à celles du CCAG-FCS. Elle dispose d'un délai de 7 jours pour le faire. À défaut de décision expresse, la prestation sera réputée réceptionnée.

Il est possible de déroger dans le marché à ces règles en allongeant par exemple les délais de vérification. Il convient d'apporter la plus grande attention au traitement de cette phase d'admission/réception dans le CCAP. La conséquence d'un mauvais traitement de cet aspect pourrait bien être que la personne publique valide, par son silence, des prestations qui ne devraient pas l'être.

Lorsque le marché comporte à la fois des prestations de numérisation et de mise en ligne, il peut faire référence au CCAG-FCS pour la numérisation et les prestations informatiques de mise en ligne seront traitées par un article du CCAP qui recopiera *in extenso* les articles du CCAG-TIC.

PARTIE IV

RÉDIGER UN CAHIER DES CHARGES DE NUMÉRISATION

RÉDIGER UN CAHIER DES CHARGES DE NUMÉRISATION

par Catherine
Mocellin

Le cahier des charges doit permettre de répondre au(x) besoin(s) et objectifs de numérisation précis et quantifiés, délimités lors d'une phase d'instruction préalable à la rédaction, faisant apparaître les éléments clés du projet.

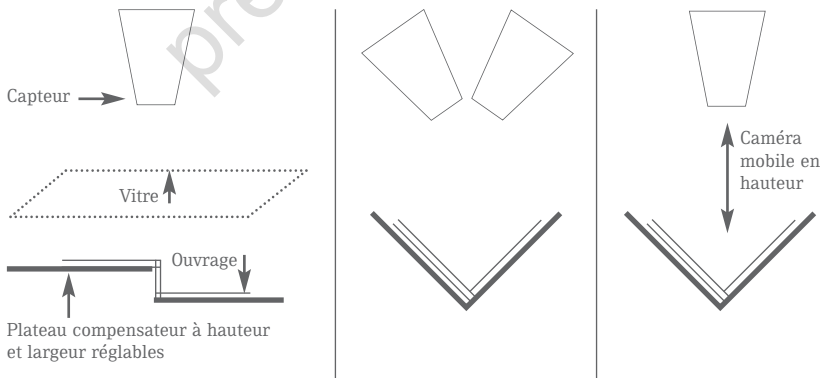
Il formalise le cadre contractuel d'une offre concurrentielle (prix, organisation, ressources, compétences...) permettant l'évaluation technique et financière des candidats, et dans lequel le projet sera réalisé. Ces documents engageant le commanditaire doivent donc être très soigneusement rédigés.

Quel que soit le type de marché, l'attention de ses rédacteurs devra être focalisée sur tous les points où se rencontrent d'une part les prestations du titulaire et d'autre part les tâches de la personne publique. Ils ne devront également jamais perdre de vue que toute prestation mise à la charge du titulaire doit avoir pour corollaire sa vérification, sa validation et son paiement par la personne publique. La répartition des tâches devra donc être clairement identifiée sur toute la chaîne de production.

L'instruction préalable vise à identifier les collections à numériser, les tâches de préparation indispensables, les modalités de réalisation du projet selon les moyens humains, techniques, financiers et les objectifs de mise en ligne. Ces éléments conditionnent le type et l'étendue des prestations à demander. Il est essentiel d'abord d'**expertiser les originaux à traiter**, en articulant la politique de numérisation et les choix techniques aux politiques documentaires, de conservation et de communication des originaux : quel est le fonds à numériser ? Envoie-t-on des documents très dégradés que l'on ne communiquera plus ensuite, ou des ouvrages en bon état ? Une restauration est-elle nécessaire, réalisable ? Quand est-elle faite ? Certains documents sont-ils trop fragiles pour être numérisés sans dégradation importante ? Dans certains cas, il ne faudra pas hésiter à renoncer à la numérisation.

On veillera à numériser **un fonds cohérent** sur le plan documentaire et, sur le plan des supports, le plus homogène possible (ou réparti en sous-ensembles pouvant constituer des lots dans un marché). Une expertise fine des collections détaillera le **volume à traiter**, le(s) **support(s)** et leurs **caractéristiques matérielles** (type, valeur particulière, formats, état de conservation, contenus, qualité d'impression, variation de densité, de contraste, logos sur les microformes...). On sera vigilant sur les difficultés posées par les originaux selon les prestations souhaitées : pour des documents reliés, le degré d'ouverture s'il est inférieur à 120°, l'étroitesse de la marge de fond de cahier et la fragilité des reliures nécessitent parfois de dérelier les ouvrages pour permettre une prise de vue exhaustive. La reliure après numérisation ne sera, en général, pas possible du fait de la fragilité des fonds de cahier. Voici trois manières de numériser un ouvrage sur un porte livre (parmi d'autres) :

SCHEMA TROIS DISPOSITIONS POSSIBLES POUR LA PRISE DE VUE DES ORIGINAUX (PARMI D'AUTRES)



Pour une prestation d'OCR, on signalera les caractéristiques détaillées dans la Partie I. Proposer un jeu d'images à convertir et demander aux candidats de fournir dans le délai de réponse un test de leur OCR sur ce jeu facilite la sélection lors de l'analyse. Pour la saisie de tables des matières et d'index, on mentionnera, par exemple, l'existence éventuelle de tables multi-volumes (tables décennales, annuelles...) et la manière de les traiter pour les lier à chaque volume qu'elles référencent. Cette expertise conduit à déterminer les paramètres techniques par prestation selon l'utilisation des documents numérisés et le système de gestion (archivage, mise en ligne), les contraintes sur les originaux et les précautions de manipulations.

Une **fiche d'état physique par ouvrage** doit en décrire les principales caractéristiques matérielles, selon un niveau de détail suffisant pour informer les opérateurs, mais pas trop fin pour ne pas conduire le prestataire à refuser abusivement des ouvrages qui lui sembleraient trop fragiles – chaque ouvrage refusé devant être remplacé par le commanditaire pour tenir les volumes contractuels. La fiche essentielle au transfert de responsabilité sur les originaux est la base de toute réclamation auprès du prestataire en cas de dégradation. Enfin, on fixera une valeur d'assurance, globale et/ou particulière à certains originaux, qu'il est préférable de formuler sous forme de fourchette de deux valeurs.

L'expertise permet donc de définir l'étendue du fonds à traiter, les opérations préalables sur les originaux, les règles de numérisation, la gestion des responsabilités dont, en particulier, le seuil d'acceptabilité des dégradations sur les originaux (à partir de quelle gravité de dégradation l'établissement procédera en général à une réclamation auprès du prestataire) – sachant que toute utilisation **normale** de documents fragiles présente un risque et que la réclamation sera faite au cas par cas.

Les autres points clés à instruire avant la rédaction du cahier des charges sont présentés ci-dessous.

thème	principaux points	remarques
Originaux à traiter <i>CCTP, CCAP</i>	Supports et état matériel, contraintes de manipulation, répartition en nombre de pages ou en pourcentage Numérisation dans les locaux du prestataire ou sur site	Prévoir si besoin une formation des opérateurs à la manipulation Allotissement possible pour des prestations très différentes (envisager alors le suivi d'un prestataire par lot)
Paramètres techniques : <i>CCTP</i>	Toutes les règles nécessaires à la numérisation et/ou la conversion en mode texte, selon les originaux	Prendre en compte l'utilisation future des documents numérisés ; préciser les règles pour les fichiers d'archivage et ceux de diffusion
Transfert de responsabilité : <i>CCAP, CCTP</i>	Valeur d'assurance générale et/ou particulière à certains documents Mentionner la fiche d'état physique par original sans la décrire Modalités pratiques des transferts, documents de suivi Pénalités pour dégradation, perte ou retard	Formuler la valeur d'assurance sous forme de fourchette de valeurs Déterminer le niveau de détail pertinent dans la fiche d'état
Accès aux documents numérisés : <i>CCTP</i>	Types d'accès par Internet et par le catalogue, métadonnées d'identification des documents numérisés Tâches préalables (correction des notices, dépouillement des originaux...)	Connaître les droits d'auteur sur les originaux pour distinguer accès libre et restreint. Adapter les accès aux supports et au type de contenu (accès au fascicule pour la presse par ex.)
Personnel : <i>CCAP, CCTP</i>	Organisation des équipes et charge de travail + sous-traitance éventuelle : logistique des enlèvements, création du fichier de récolement, catalogage...	Connaître les contraintes présentes et à venir sur les ressources humaines
Planning : <i>CCAP, CCTP</i>	Délais globaux (exécution, tranches éventuelles, garantie, phases successives) et particuliers (par bon de commande, délai de validation), selon les capacités de suivi et de contrôle du commanditaire	La forme du marché, les tranches éventuelles et les modalités d'affermissement sont liées notamment au planning, au montant financier et aux volumes. Un marché à prix global et forfaitaire est adapté pour des volumes précis et fixes ; la gestion est facilitée.
Montant financier : <i>CCAP</i>	Forme du marché selon le code des marchés publics montant minimum/maximum pour l'ensemble du marché et par tranche éventuelle (ferme, conditionnelle)	Un marché à bons de commande offre une souplesse dans les volumes et les montants financiers, mais est plus lourd à gérer. Un découpage en tranches est utile en cas d'incertitude sur les budgets et pour pouvoir résilier le marché au terme d'une tranche
Moyens techniques et modalités de contrôle : <i>CCAP, CCTP</i>	Chaîne de gestion de la production (outils de suivi des originaux, des documents numériques), fichiers d'accompagnement Modalités de livraison, modalités de contrôle du commanditaire et niveaux de qualité exigés	Identifier les évolutions à porter sur les chaînes pour assurer un suivi fiable et une mise en ligne rapide des documents validés. Ajuster les méthodes de contrôle aux capacités de l'institution

Si le budget le permet mais que les moyens humains ou techniques manquent, on pourra demander des ressources complémentaires au prestataire (caisses et scellés, personnels, livraison d'indicateurs de suivi du projet,...) ou décider de **sous-traiter tout ou partie du contrôle qualité** (voir Partie I). On signalera également les points incertains à la date de la publication mais réglés lors de la phase de test ou pour cette date.

Outre la satisfaction du besoin, le cahier des charges a vocation à permettre aux soumissionnaires de constituer une offre pertinente. Il s'agit de permettre aux candidats de comprendre finement le(s) besoin(s) pour ajuster leur offre (prix, organisation, moyens, planning...). La rédaction doit être la plus précise et complète possible et préserver l'avenir, sachant que tous les documents sont contractuels – en particulier les volumes à traiter ou les montants financiers mentionnés. On **s'engagera sur les quantités et/ou les montants**, mais, dans un marché à bons de commande, pour conserver une marge d'ajustement on pourra préciser ces chiffres sous forme de fourchette. On évitera également de mentionner à la fois les volumes précis et un montant financier fixe car cela fausse la mise en concurrence. Si l'on fournit des volumes à traiter, on les exprimera en pages (unité de comptage la plus simple et la seule univoque, même pour le mode texte).

Pour préserver l'avenir et tenir compte de l'hétérogénéité des fonds, la description de ceux-ci doit **avoir un niveau de détail pertinent** pour ne pas conduire le candidat à gonfler ses prix face à de l'imprécision ou à ce qu'il pourrait estimer être d'une trop grande complexité. Être complet et précis mais générique garantit en cours de projet une souplesse dans les sélections. Toutefois, la responsabilité du bibliothécaire chargé des sélections est essentielle pour la tenue des délais et des volumes : elles doivent être conformes aux descriptions matérielles et aux règles de traitement données au CCTP. Organiser pour les candidats, dans le délai de réponse, une **visite** de collections représentatives du projet leur permet de mieux comprendre les contraintes des ouvrages et donc d'ajuster leur offre.

Le plan du CCTP suit la chaîne de production, après une introduction du contexte, des objectifs du projet et une liste résumée des prestations⁵⁶ En annexe du CCTP, on indiquera les données techniques fines (schémas XML, métadonnées, schémas explicatifs, etc.). Les tableaux p. 97-98 détaillent les différentes parties et leurs contenus essentiels.

Principaux champs d'un CCTP d'un marché de numérisation		
Chapitres	Principaux points	Remarques
Contexte et objectifs	<ul style="list-style-type: none"> • Présenter clairement le besoin, l'existant, les objectifs généraux du projet • Proposer aux sociétés une visite d'un échantillon des collections 	Une offre de qualité révèle une compréhension fine du besoin
Résumé des prestations	<ul style="list-style-type: none"> • Présentation en liste exhaustive organisée selon la chaîne de production 	Il s'agit de permettre d'embrasser rapidement les caractéristiques du projet
Répartition, typologie des documents	<ul style="list-style-type: none"> • Volumes précis par support, exprimés en % de l'ensemble, en pages ou demande d'atteindre le maximum de pages selon le montant financier 	Le montage financier et volumétrique favorise la baisse des prix par le jeu de la concurrence
Mise à disposition des originaux, échanges	<ul style="list-style-type: none"> • Transferts : logistique aller-retour (dont véhicule aux normes de conservation) ; organisation des locaux, horaires, accès si prestation sur site ; • Identification des originaux éventuellement refusés 	Renvoyer au CCAP pour toutes les précisions administratives sur les assurances, le calcul des pénalités, les délais
Description des prestations	<ul style="list-style-type: none"> • Dépouillement, identification des documents originaux et numérisés ; • Numérisation et/ou conversion en mode texte (y compris pour les tables des matières et index) : caractéristiques matérielles détaillées des originaux, règles techniques précises pour chaque prestation ; • Livrables de suivi de production (rapports de production, statistiques...) ; • Autres prestations (mise à disposition de moyens complémentaires) 	Être précis et complet sur les caractéristiques des originaux en restant générique (ne pas entrer dans les cas particuliers) ; au besoin préciser que la liste des caractéristiques n'est pas exhaustive pour conserver une liberté d'action dans les sélections et la négociation
Contrôles de conformité et suivi de production	<ul style="list-style-type: none"> • Suivi du projet (tenue des délais, des volumes) ; • Modalités de contrôle et de changement de phase (phase de test, de production courante...) ; • Normes appliquées ; • Critères ; niveaux de qualité chiffrés par prestation ; • Réfections : modalités, délais de contrôle du commanditaire, de relivraison 	Un plan assurance qualité contractuel peut être utile. Préciser que la liste des critères de qualité n'est pas exhaustive et offre une marge de négociation si la qualité est médiocre. Le niveau de qualité impacte les coûts.
Modalités de livraison	<ul style="list-style-type: none"> • Structure des livraisons selon les outils de contrôle, archivage, mise en ligne ; • Arborescence et règles de nommage des fichiers et des répertoires ; • Supports et modalités de livraison selon les volumes prévus 	Insérer des croquis, détailler les protocoles techniques. Séparer la livraison des réfections des nouveaux documents numérisés.
Annexes	<ul style="list-style-type: none"> • Spécifications techniques détaillées : fichier d'identification, fichiers image, fichiers texte (y compris tables et index) et formalisme ; • Règles de pagination ; • Autres référentiels utiles fournis par le commanditaire 	Schémas XML et leur URL s'ils sont en ligne, modèles de documents tabulés. Privilégier les tableaux et les croquis. S'appuyer sur des normes et des standards.

Principaux champs d'un CCAP d'un marché de numérisation		
Chapitres	Principaux points	Remarques
Objet, forme, pièces constitutives	<ul style="list-style-type: none"> • Rappeler l'objet du marché • Forme (pour un marché à bons de commandes, définir l'unité commandée : pages, documents numériques) • Découpage en tranches, et modalités d'affermissement • N° du lot si le marché est alloté • Montants financiers minimaux/maximaux ou volumétrie à produire • Pièces constitutives 	
Durée et conditions d'exécution	<ul style="list-style-type: none"> • Organisation des relations (interlocuteurs...) • Éventuellement, visite préalable à la remise des offres et définition des modalités ; • Durée du marché et délai d'exécution ; • Délais par phases : tranche ferme et conditionnelle(s) ; • Phase de test, de production courante, de réfection, de garantie 	Avoir un délai de contrôle suffisant facilite la gestion du projet par le commanditaire (chaque délai indiqué au CCAP engage contractuellement le commanditaire et doit être respecté)
Exécution des prestations, contrôles	<ul style="list-style-type: none"> • Mise à disposition des documents à traiter • Délai de réalisation d'un bon de commande et de livraison • Contrôles par le prestataire • Visite des ateliers du prestataire par le commanditaire et modalités • Délais de vérification et d'admission par le commanditaire 	Si le contrôle par audit est choisi, le CCAP précisera les modalités et son caractère contractuel
Secret professionnel, droits de propriété	<ul style="list-style-type: none"> • Secret professionnel • Clause de renoncement du titulaire à la propriété sur les fichiers numériques 	
Garanties et assurances	<ul style="list-style-type: none"> • Délai de garantie • Assurances dommages aux biens • Délai de sauvegarde des documents numérisés à la fin du marché 	Demander au prestataire de sauvegarder la production durant 6 mois est utile mais a un coût pour l'entreprise
Prix, paiement	<ul style="list-style-type: none"> • Contenu et variation des prix – acomptes, avances éventuels • Conditions et délais de paiement 	
Pénalités, règlement des différends	<ul style="list-style-type: none"> • Pénalités pour perte de documents, retard, dégradation • Modalités d'application et calcul • Règlement des différends, modalités de résiliation du marché 	
Modification en cours d'exécution, dérogations	<ul style="list-style-type: none"> • Modalités de modifications et règles de dérogation au CCAG 	

On précisera dans le RPC et le BPU (voir Partie III) les spécificités du marché et ce qui est attendu des candidats. Dans le RPC, les critères de notation – en général une valeur égale pour la note technique et la note financière (50 %) – peuvent être pondérés pour insister sur les points du besoin sur lesquels les candidats sont particulièrement attendus, par exemple : 20 % de la note technique sur la cohérence de l'organisation, des équipes, de la gestion des flux, 20 % sur les outils de production et de contrôle (numériseurs, logiciels pour l'OCR et les contrôles, base de suivi de production), 10 % sur le respect des délais (justesse et réalisme du planning). On donnera également au RPC **les points obligatoires dans les offres et les contraintes de délais** (proposition d'un plan d'assurance qualité – PAQ –, d'un planning de réalisation...) : une bonne gestion de projet par les deux parties conditionne la réussite du marché.

Il faut fournir suffisamment de critères d'analyse des offres, qui permettront de départager les candidats et seront repris dans le rapport d'analyse. On sera très attentif au contenu et à la formulation de ce dernier, car il peut être demandé par un candidat non retenu pour faire valoir tous les points de contestation possibles. Ce rapport devra donc permettre au commanditaire de justifier son choix en transparence, sur des points objectifs à partir des documents contractuels.

Le BPU est nécessaire pour un marché à bons de commande. Établi par le commanditaire par type de prestation selon les caractéristiques des originaux (format, degré d'ouverture, traitement automatique ou manuel pour les microformes...), et/ou des répartitions fixées au CCTP et au CCAP, il est rempli par le candidat avec ses prix. Pour la conversion, il sera organisé selon le niveau de qualité (brut et/ou corrigé).

56. Voir modèles de cahiers des charges de numérisation en ligne : ministère de la Culture et de la Communication. Archives de France. Écrire un cahier des charges de numérisation du patrimoine : guide technique. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/static/1308> > (consulté le 19 février 2010).

Un marché à bons de commande permet d'ajuster les volumes par commande aux capacités de la bibliothèque et aux contraintes apparaissant durant le marché. Le **suivi de la production** est central pour respecter les engagements contractuels. La production mensuelle doit être lissée à l'aide d'un planning annuel ajusté avec le prestataire. Le CCAP précise le délai de réalisation d'un bon de commande et stipule éventuellement que ce délai est donné sur chaque bon. Les pages rejetées au contrôle doivent être relivrées dans les délais d'exécution, même en cas de renvoi des originaux ; selon la complexité du projet, ce délai doit donc être suffisamment long, mais pas trop pour éviter que de nombreux bons de commande soient en cours de réalisation en même temps. Commander un nombre de pages sans détailler les cotes des originaux sur le bon facilite le suivi administratif du marché dans la mesure où le bon peut être soldé une fois que le volume en pages commandé est validé, quels que soient les originaux concernés.

Enfin, si aucun délai de validation n'est donné au CCAP, ceux du **CCAG-FCS** font foi : 2 semaines. Dans tous les cas, il faut prononcer la validation des livraisons dans les délais contractuels, sous peine de ne plus pouvoir rejeter les non-conformités, alors validées d'office. Les **applications** et les **documents de suivi** sont donc essentiels pour la réussite du marché : c'est toute la chaîne de production qui doit être organisée pour respecter chaque échéance.

PARTIE V

**INDEXER,
STRUCTURER,
ÉCHANGER :
MÉTADONNÉES ET
INTEROPÉRABILITÉ**

INDEXER, STRUCTURER, ÉCHANGER : MÉTADONNÉES ET INTEROPÉRABILITÉ

par Isabelle
Westeel

DÉFINITIONS ET TYPOLOGIE

+++++

Il est de la responsabilité des chefs de projet chargés de monter des applications documentaires de maintenir les données et les métadonnées stables et pérennes. La description, l'indexation et la structuration des documents, expertises des « métiers de la documentation » doivent dépasser le temps du projet et pouvoir s'appuyer sur une réflexion approfondie. Les métadonnées sont les garants de fonctionnalités de recherche et de navigation satisfaisantes dans une bibliothèque numérique. Les outils de diffusion étant amenés à évoluer, il convient donc de rendre les données et leur description stables dans le temps et de les faire reposer sur des normes et des standards.

DÉFINITIONS

Une métadonnée (du grec *meta* : au-delà de) est une donnée à propos d'un ensemble signifiant de données pouvant correspondre à un document papier ou électronique. Les textes de normalisation proposent deux définitions qui permettent de cerner leur nature fonctionnelle :

- « donnée qui définit et décrit une autre donnée » (ISO/IEC 11179-3, 2003) ;
- « de l'information structurée qui décrit, explique, localise la ressource et en facilite la recherche, l'usage et la gestion » (NISO, *Understanding Metadata*, 2004).

À l'origine destinées principalement à rendre les documents lisibles par les machines, les métadonnées servent à exploiter le document et à garantir son accès. Certaines métadonnées doivent être directement compréhensibles. Dans les documents bureautiques, les métadonnées correspondent souvent aux « propriétés » du document qui sont accessibles par le menu

Fichier. Cette grille de description doit être complétée pour pouvoir servir à la recherche d'une information précise. Elle comporte des zones relatives à la description sémantique du document. De la même façon, les logiciels de retouche d'images permettent de saisir des métadonnées techniques ou sémantiques. La mise en relation des différentes grilles de métadonnées impose une « normalisation » quant à l'intitulé des zones et à la saisie de l'information. Dans les deux cas cités ci-dessus, les métadonnées sont « internes » c'est-à-dire qu'elles sont enregistrées à l'intérieur même du fichier informatique auquel elles s'appliquent (texte, image...). Ces métadonnées internes sont souvent limitées en nombre et délicates à mettre à jour. Elles peuvent en outre être altérées ou disparaître lors de la conversion du fichier dans un autre format. Les métadonnées peuvent aussi être stockées à l'extérieur du fichier (métadonnées externes). Ce système permet d'établir des métadonnées plus riches. Lors du processus de gestion des informations, il convient de veiller à ne pas perdre le lien entre les deux fichiers, ce lien pouvant être signifié par un nom de fichier identique mais des extensions différentes (par exemple fichier du document en .jpg et fichier de métadonnées en .xml). Certaines métadonnées doivent être saisies, c'est le cas en général pour les métadonnées sémantiques. De très nombreuses métadonnées peuvent être produites de manière automatique par des scripts ou des conversions de fichiers.

Recommandations

Lors de la phase de numérisation, on pourra demander au prestataire ou à l'atelier de l'institution d'établir des métadonnées internes essentielles : identifiant(s), titre du document, nom de l'établissement de conservation, date de numérisation, type de scanner, résolution de numérisation.

TPOLOGIE

Les métadonnées permettent ainsi d'identifier chaque constituant du document numérique et de le relier à l'ensemble. Elles sont en quelque sorte la « reliure » du document dans le monde numérique. Trois types de métadonnées peuvent caractériser un document numérique :

- des **métadonnées descriptives** regroupant les informations de contenu et d'identification du document (auteur, titre, mots-clés, identifiant...). Cette description est établie dans un format normalisé (Dublin Core, EAD, MarcXML...) qui permettra les échanges ;
- des **métadonnées techniques ou de structure**, c'est-à-dire les informations relevant de la version du document (date, format de fichier, taille, compression...). Elles permettent de connaître tous les fichiers composant un document ainsi que les relations (physique et logique) entre les fichiers (par exemple ordre d'affichage des fichiers...). Elles incluent aussi des informations sur l'environnement matériel et logiciel en vue d'une migration ;
- des **métadonnées administratives** composées :
 - > des métadonnées de gestion des droits : droits d'accès (droits d'auteur principalement) et d'usage (droits d'impression, de modification...),
 - > des métadonnées techniques de préservation (comment conserver et restituer le document ? Comment conserver l'historique des modifications des fichiers ?) (voir Partie VII).

granularité

Toute gestion de collection numérique impose de tenir compte de la granularité (physique ou logique). La granularité est définie par les différents niveaux d'information à prendre en compte : collection, ressource, article, paragraphe de l'article...

LE PRÉALABLE XML

La nécessité de structurer l'information est apparue dans les années 1970 avec les réflexions ayant abouti à l'émergence des langages de balises. Un document est composé d'une structure logique (contenu organisé) et d'une

structure physique (mise en forme de l'information). Ce constat général a amené la naissance du langage SGML (*Standard Generalized Markup Language*) et de ses principaux descendants HTML et XML. Le langage XML (*Extensible Markup Language* = Langage de balisage extensible) est la syntaxe des principales grammaires de structuration. Il s'agit d'un cadre général qui se préoccupe du contenu du document et non pas de sa mise en forme (contrairement au HTML qui gère des balises de mise en forme). Un document XML peut être représenté sous la forme d'un arbre (ou hiérarchie) composé d'éléments ou balises. Les éléments sont donc les objets fondamentaux d'un document XML. Celui-ci est toujours composé d'un élément racine (ou élément supérieur) et d'un ensemble d'éléments contenant eux-mêmes d'autres éléments ou du texte. Les attributs permettent de qualifier les éléments. Les documents XML sont des documents « texte ». Le langage est indépendant de toute application logicielle. On peut noter que le XML intègre la gestion de l'Unicode, norme informatique permettant le codage de tous les caractères quel que soit le système d'écriture.

Voici un exemple très simple de document XML décrivant un livre.

L'élément racine <livre> contient un <auteur> et un <titre>

```
<?xml version="1.0" encoding="UTF-8"?>
<livre>
  <auteur>
    <nom>Dumas</nom>
    <prenom>Alexandre</prenom>
  </auteur>
  <titre>Les trois mousquetaires</titre>
</livre>
```

La recommandation XML permet ainsi la création libre d'éléments mais donne quelques règles concernant l'organisation du document lui-même. Si un document XML suit l'ensemble des règles de la recommandation, on dit qu'il est « bien formé ». Les navigateurs pourront le lire sans problème.

La recommandation XML ne peut cependant suffire à exprimer la structure logique d'un document. Elle est complétée par des grammaires (DTD,

Document Type Definition et schémas XML) adaptées à certains types de document. Ces grammaires permettent de fixer la structuration pour un certain type de document et de rendre un document XML valide. Les métiers de l'information et de la documentation utilisent plusieurs DTD détaillées plus loin. Une DTD liste ainsi les éléments permis dans le document, le contenu possible de ces éléments (texte ou autres éléments), les attributs associés aux éléments et enfin les valeurs permises par les attributs. Les schémas XML, programme du W3C, permettent un contrôle plus strict et une structuration plus forte de l'information.

La mise en forme des documents

Puisque XML ne se préoccupe pas de la mise en forme des documents, celle-ci est assurée par l'utilisation de feuilles de style, réglées par trois langages définis par le W3C : le CSS (*Cascading Style Sheets*), XSLT (*eXtensible Stylesheet Language Transformations*) et XSL-FO (*eXtensible Stylesheet Language – Formatting Objects*). Ces différents langages permettent donc de mettre en forme les documents XML et sont utilisés par les informaticiens au moment de la construction du site Web.

comment produire du XML ?

On peut produire du XML de différentes manières : par la saisie dans un éditeur XML, par la conversion automatique de données à partir d'un tableur de type Microsoft® Excel®... Certains logiciels documentaires permettent la saisie de documents XML de façon transparente à partir d'un masque de saisie. Il convient alors de s'assurer que ces logiciels respectent bien les standards de description.

LES PRINCIPALES GRAMMAIRES XML

+++++

LE SCHÉMA METS

METS (*Metadata Encoding and Transmission Standard* = Standard d'encodage et d'échange de métadonnées ; < <http://www.loc.gov/standards/mets> > est un schéma XML défini pour permettre la création d'un jeu de métadonnées. Il permet de créer un document qui encapsule toutes les métadon-

nées d'un objet numérique. À un objet numérique correspond un fichier METS.

Ce schéma permet de regrouper dans un seul fichier :

- toutes les métadonnées d'un document ;
- des pointeurs vers des métadonnées qui décrivent le document et sont stockées ailleurs ;
- les fichiers qui composent le document.

Il permet de préserver l'intégrité des documents complexes grâce à une carte de structure qui permet de reconstruire la navigation dans le document. Il comprend sept sections :

- **Header** (informations sur le document METS) ;
- **Descriptive metadata Section** (métadonnées descriptives externes ou encapsulées : Dublin Core, EAD...) ;
- **Administrative metadata Section** (métadonnées techniques, droit, source, provenance) ;
- **File Section** (liste des fichiers composant l'objet) ;
- **Structural map** (carte de structure, peut être logique ou physique, répétable) ;
- **Structural links** (liens entre différents éléments de la carte de structure) ;
- **Behavior** (association d'exécutables). METS permet de préparer des données conformément à la norme OAIS.

La mise en place de METS dans un établissement demande que le XML soit bien maîtrisé et que l'intérêt du METS pour les projets soit bien compris. Utilisé par la BnF dans le cadre du système d'archivage SPAR, METS est également utilisé pour Persée < <http://www.persee.fr> > et Cairn < <http://www.cairn.info> > ainsi que dans le cadre des éditions électroniques de l'École des chartes. Le schéma est supporté par les logiciels Fedora, Greenstone et DSpace.

TABEAU MÉTADONNÉES TECHNIQUES ET DESCRIPTIVES

Le tableau suivant dresse un panorama des principales métadonnées techniques et sémantiques utilisées dans les bibliothèques numériques.

Métadonnées	commentaires
<p>EAD <i>Encoded Archival Description</i> = Description archivistique encodée. Maintenue par la Bibliothèque du Congrès en partenariat avec la Société des archivistes américains. Maintenance assurée en France par le groupe AFNOR CG46/GE3 (EAD). Version actuelle DTD EAD 2002 Site officiel : < http://www.locweb.loc.gov/ead/ > Site français de référence : Archives de France < http://www.archivesdefrance.culture.gouv.fr/gerer/classement/normes-outils/ead/ > (liens vers la traduction française des balises et vers le référentiel des applications françaises).</p> <p>Complété par le schéma EAC-CPF <i>Encoded Archival Context</i> = Contexte archivistique encodé – Collectivités, personnes, familles. Site dédié au projet hébergé par la Staatsbibliothek de Berlin (liste des projets et applications, dictionnaire des balises, schéma en cours de normalisation) < http://eac.staatsbibliothek-berlin.de/ ></p>	<p>DTD destinée à l'origine à l'encodage d'instruments de recherche archivistique. Elle est utilisée également pour décrire les collections de manuscrits et a été choisie pour la conversion rétrospective du CGM (Catalogue général des manuscrits des bibliothèques publiques de France). Les notices EAD sont visibles sur le site du CCFr (Catalogue collectif de France) et sur Calames (Catalogue en ligne des archives et des manuscrits de l'enseignement supérieur)</p> <p>Publication prévue en 2010 d'un « Guide sur les bonnes pratiques de l'EAD en bibliothèque ».</p> <p>Données d'autorité sur les personnes, familles et collectivités. Il suit les travaux internationaux sur le sujet</p>
<p>TEI <i>Text Encoding Initiative</i> = Initiative pour l'encodage ou le balisage des textes. Dernière version : Recommandations TEI P5 < http://www.tei-c.org/index.xml > L'utilisation de la TEI Lite n'est pas recommandée car elle ne suit pas toujours les changements de la TEI.</p>	<p>Ensemble de recommandations (<i>Guidelines</i>) pour établir un standard d'encodage qui permettra d'échanger des textes de toutes natures. La TEI est composée de modules (ou <i>tags sets</i>). Chaque module correspond à un type de texte (prose, vers, drame, parole, dictionnaires, terminologie). Il est possible de combiner ces modules et ainsi de construire des DTD adaptées au texte à encoder.</p>

	<p>Exemple : Base Epistemon des Bibliothèques Virtuelles Humanistes (BVH) du Centre d'Etudes Supérieures de la Renaissance (CESR) à Tours < http://www.bvh.univ-tours.fr/ > ; textes et images encodés selon les recommandations de la TEI. Sur ce site, mise en ligne du Manuel d'encodage XML-TEI. Renaissance et Temps modernes. (imprimés-manuscrits), version 2, 1^{er} juillet 2009.</p>
<p>TEF Thèses électroniques de France < http://www.abes.fr/abes/documents/tef/ ></p>	<p>Profil METS qui associe à une thèse électronique des métadonnées descriptives (auteur, titre, sujet...), des métadonnées administratives liées au diplôme (doctorant, jury, établissement de soutenance, discipline), des métadonnées de gestion (circuit administratif, archivage, droits) et des métadonnées de conservation. TEF utilise l'enveloppe METS, qui lie entre elles différents types de métadonnées exprimées dans des vocabulaires standards (Dublin Core, MADS cf. plus bas). Le modèle de données TEF peut être exprimé en RDF.</p>
<p>MARXML < http://www.loc.gov/standards/marcxml/ ></p> <p>MODS <i>Metadata Object Description Schema</i> < http://www.loc.gov/standards/mods/ ></p>	<p>Schéma XML s'appuyant sur Marc21 et permettant une conversion réversible sans perte d'une notice Marc21 d'ISO 2709 à XML. Élaboré en 2002, MODS est un schéma XML qui établit un jeu de métadonnées bibliographiques permettant la conversion de notices en format Marc21 mais aussi la création originale de notices. Il utilise des éléments explicites (<i>title, name, publisher...</i>) plutôt que les étiquettes numériques du Marc. Il permet de décrire des ressources très diverses et est utilisé par exemple dans la base de données DAPHNE (Données en Archéologie, Préhistoire et Histoire sur le Net ; http://www.daphne.cnrs.fr)</p>

...

<p>Complété par MADS <i>Metadata Authority Description Schema</i> < http://www.loc.gov/standards/mads/ ></p> <p>Remarque : Les formats bibliographiques sont en cours d'évolution Évolution de l'ISO 2709 par MarcXchange < http://www.loc.gov/standards/iso25577/ > Travaux autour de RDA (<i>Resource Description and Access</i>) : futur code international de catalogage, destiné à remplacer les AACR avec prise en compte des FRBR (<i>Functional Requirements for Bibliographic Records</i>), modèle conceptuel pour les notices bibliographiques élaboré par l'IFLA.</p>	<p>Schéma XML créé en complément de MODS et en lien avec le format Marc21 Autorités permettant de créer des notices d'autorités (personnes, événements, sujets...).</p> <p>Mapping entre RDA et MODS disponible en mars 2009.</p>
<p>ALTO <i>Analysed layout and text object</i> Site officiel : < http://www.loc.gov/standards/alto/ > Site français de référence : pages professionnelles de la BnF Schéma ALTO tel qu'utilisé à la BnF : < http://bibnum.bnf.fr/ns/alto_prod.xsd ></p>	<p>Format issu du projet européen METAe permettant d'utiliser le texte converti à partir d'un format image par l'OCR et de rendre compte de la mise en page physique d'un document. Ce schéma XML conserve les coordonnées du contenu (texte, illustrations, graphiques) dans l'image et permet la superposition de l'image et du texte (fichier PDF multicouches).</p>
<p>LOM <i>Learning Object Metadata</i>, conçu, en 2002, par le comité « IEEE-LTSC-LOM, Institute of Electrical and Electronics Engineers, Inc. - Learning Technology Standards Committee - Learning Objects Metadata Working Group » Le LOMFR est le profil d'application français du LOM. Il correspond à la norme française NF Z76-040 publiée par l'AFNOR en décembre 2006. < http://www.lom-fr.fr/ ></p>	<p>Modèle de description ; métadonnées associées à des ressources d'enseignement et d'apprentissage (numériques et non numériques).</p>

TABEAU MÉTADONNÉES POUR LES INFORMATIONS GÉOGRAPHIQUES

métadonnées	commentaires
Norme ISO 19115 (2003) . Version 19115-2 en 2009 Déclinée en 2007 pour donner la norme ISO 19139 , compatible XML.	Norme pour la description des informations géographiques (imprimé ou numérique).
La directive européenne INSPIRE (<i>Infrastructure for spatial information in Europe</i>) établit une infrastructure d'information géographique pour l'ensemble de l'Union.	INSPIRE s'intéresse aux métadonnées numériques (création et mises à jour) et concerne 34 domaines thématiques. On peut retenir l'annexe 1 qui intéresse les données nécessaires au repérage sur le territoire, telles que systèmes de coordonnées, unités administratives, réseaux de transport, hydrographie, parcellaire cadastral, adresses et noms de lieux ainsi que les données relatives aux sites protégés. INSPIRE est basée sur ISO 19115 (métadonnées) et ISO 19119 (services). Elle est compatible avec le Dublin Core.
Les Geoformats : formats issus du monde géographique et mis en place pour diffuser l'information géographique sous forme de services Web	
GML (<i>Geography Markup Language</i>) est un langage XML qui permet d'encoder et d'échanger des données géographiques. Correspond à la norme ISO 19136 : 2007.	Il fournit un modèle de données aux systèmes d'information géographique. Format directement issu du monde géographique et plutôt réservé à interopérer des applications cartographiques, il est donc riche et complexe.
KML (<i>Keyhole Markup Language</i>) est un langage XML qui a été développé pour permettre la gestion de l'affichage des données géospatiales et géographiques sur Google Earth, Google Maps, Google Mobile et WorldWind.	Ce format n'est pas seulement destiné aux applications géographiques et il est donc plus pauvre que GML. Depuis avril 2008, KML (à partir de la version 2.2) est une « norme » ouverte, reconnue et supportée par l'OGC (<i>Open Geospatial Consortium</i>) et maintenant largement supportée par les systèmes d'information géographiques traditionnels.

TABLEAU LES MÉTADONNÉES IMAGES

Métadonnées	Commentaires
<p>EXif (<i>EXchangeable Image File</i>) : format développé en 1995 par la JEIDA (Japan Electronic Industry Development Association) (version 2.2 de 2002). Ce n'est pas un standard mais il est très bien supporté par les logiciels.</p> <p>< http://www.exif.org ></p> <p>Métadonnées internes</p>	<p>Métadonnées techniques de l'image c'est-à-dire relatives à la prise de vue (marque et modèle de l'appareil, hauteur et largeur de l'image, réglages, date et heure, ouverture, résolution, taux de compression, données géographiques (données GPS ou données de géolocalisation...). Métadonnées produites automatiquement par l'intermédiaire du numériseur ou de l'appareil photo. Ces métadonnées sont supportées par les formats TIFF et JPEG.</p>
<p>MIX (<i>Metadata for Images in XML</i>) : déclinaison sous forme de schéma XML de la norme NISO Z39.87 (Data Dictionary – Technical Metadata for Digital Still Images – Dictionnaire de données – Métadonnées techniques pour des images numériques fixes).</p> <p>Dictionnaire des données : < http://www.niso.org/standards/index.html ></p> <p>Documentation MIX : < http://www.loc.gov/standards/mix/ ></p> <p>Métadonnées externes</p>	<p>Métadonnées techniques : 200 éléments répartis en 5 familles :</p> <ul style="list-style-type: none"> • information de base sur l'objet numérique (identifiant, taille, format, compression, fixité) • informations de base sur l'image (dimensions, couleur...) • métadonnées de capture de l'image (taille de la source, date de capture, informations sur le scanner, informations sur la caméra numérique, coordonnées géographiques) • métadonnées d'évaluation de l'image (échantillonnage de capture, échantillonnage colorimétrique...) • historique des modifications <p>MIX est utilisé par la BnF dans le cadre de SPAR</p>

...

...

<p>IPTC-NAA IIM (<i>Information Interchange Model</i>) ou IIM : format d'échange défini en 1991 par l'IPTC (<i>International Press and Telecommunications Council</i>) et la NAA (<i>Newspaper Association of America</i>).</p> <p>< http://www.iptc.org/cms/site/ ></p> <p>La version actuelle du modèle IPTC/IIM est la version 4.1 qui date de 1999. En mars 2005, est annoncé le schéma XMP IPTC Core, version 1.0 qui s'inspire du format XMP d'Adobe® et qui est appelé à remplacer l'IPTC/IIM. Ce schéma intègre l'Unicode et ne devrait donc plus poser de problèmes pour le codage des caractères. En juillet 2009 est publiée la spécification IPTC Photo Metadata Standard incluant l'IPTC Core 1.1 et l'IPTC Extension 1.1.</p>	<p>Format d'indexation permettant de noter les caractéristiques sémantiques de l'image (identifiant, nom de l'auteur, date de création, titre de l'image, description du contenu de l'image, copyright, mots-clés). C'est ce modèle qui a servi de base à la société Adobe® pour inscrire des métadonnées dans les en-têtes des images gérées par le logiciel Adobe® Photoshop®. Ces données peuvent être intégrées dans l'en-tête de l'image manuellement par un opérateur ou par l'intermédiaire d'un script. Ces métadonnées peuvent être insérées dans les images JPEG.</p> <p>On trouve sur Internet des outils gratuits permettant de travailler sur les métadonnées IPTC/IIM ou XMP : XnView, Exiftool...</p>
<p>Métadonnées internes.</p> <p>XMP (<i>eXtensible Metadata Platform</i>) : format d'échange créé par la société Adobe® en 2001. Ce format est fondé sur le XML et s'appuie sur RDF et le Dublin Core. Il converge actuellement vers l'IPTC. Il ne concerne pas seulement les images numériques.</p> <p>Métadonnées internes ou externes.</p>	<p>Format pouvant être généré à partir de la suite logicielle de la société Adobe®. Ce fichier peut être créé directement lors de l'acquisition de l'image par le numériseur. Il ne s'agit pas d'une norme mais de nombreux outils supportent ce format. Ce format est adopté par un nombre croissant d'outils et de logiciels. Dans la dernière version de Adobe® Photoshop® CS4, il est possible d'introduire des métadonnées XMP pour chaque calque et donc de suivre l'historique des changements subis par le document. La version CS4 de Adobe® Photoshop® permet également d'associer des métadonnées aux formats audio et vidéo.</p>

LE SCHÉMA GÉNÉRAL DUBLIN CORE

Le schéma Dublin Core est un ensemble de métadonnées devant permettre de décrire de manière univoque n'importe quelle ressource en ligne. Créé en 1995 à Dublin (Ohio), siège de l'OCLC, ce schéma minimal devait permettre à tout créateur de ressource d'établir une description de base : le « cœur » de l'information. Finalement peu utilisé par les moteurs de recherche, le vocabulaire doit finalement sa force à l'utilisation qu'en ont faite les professionnels de la documentation. Normalisé en 2003 par l'ISO (15836), il est devenu la base de protocoles d'interrogation permettant de rendre des bases interoperables.

Dans sa version 1.1, le Dublin Core comporte quinze éléments, tous optionnels et tous répétables, regroupés à l'origine en trois groupes : contenu de la ressource (*coverage, description, relation, source, subject, title, type*), questions de propriété intellectuelle liées à la ressource (*contributor, creator, publisher, rights*), version de la ressource (*date, format, identifier, language*).

Le Dublin Core simple (ou non qualifié) est donc composé de ces quinze éléments. Des éléments de raffinements peuvent compléter certains éléments, formant ainsi ce que l'on appelle le **Dublin Core qualifié**. Ils permettent de remplacer les éléments de base par des éléments plus précis. Ainsi, au lieu de *title* (dc : title), on pourra utiliser alternative (dcterms : alternative) pour désigner un titre abrégé ou un titre traduit. De même, pour tenir compte du cycle de vie d'un document numérique, on pourra préciser la date : dcterms : created, dcterms : modified...

Le Dublin Core ne donne pas d'indications sur la manière de remplir les différents éléments mais conseille pour certains d'utiliser des normes ou des listes d'autorités : les **schémas d'encodage**. Cela permet d'utiliser un élément Dublin Core en prenant une valeur dans un vocabulaire contrôlé ou en la formatant selon une convention particulière. Ainsi, dc : subject peut être renseigné en s'appuyant sur la classification décimale Dewey ou les vedettes matière Rameau. Dans la façon de saisir les informations, on peut conseiller d'appliquer les règles bien connues des bibliothécaires permettant de normaliser la saisie : formes d'autorités pour les éléments < creator > ou < contributor > par exemple.

TABEAU Liste des quinze éléments dublin core

<contributor> (contributeur)	Contributeur (auteur secondaire) du contenu de la ressource : personne physique ou collectivité. [On peut conseiller de saisir cet élément de façon normalisée].
<coverage> (couverture)	Couverture spatio-temporelle ou juridictionnelle du contenu de la ressource. Le contenu de cet élément est, le plus souvent, un nom de lieu et/ou des dates en utilisant un vocabulaire contrôlé : par exemple « France » ; « 19 ^e siècle »...
<creator> (createur)	Créateur du contenu de la ressource : personne physique ou collectivité. [On peut conseiller de saisir cet élément de façon normalisée].
<date> (date)	Date de la création du document. Il est fortement conseillé d'utiliser la norme ISO 8601 qui code la date sous la forme AAAA-MM-JJ. Il est possible de « raffiner » cet élément pour indiquer une date de modification, de sélection de la ressource...
<description> (description)	Description du contenu de la ressource. Le contenu de cet élément peut être un résumé, une table des matières ou encore une description en texte libre.
<format> (format)	Manifestation physique ou numérique de la ressource. Le contenu de cet élément est, le plus souvent, le support ou/et le type d'application informatique nécessaire à sa lecture, par exemple image/jpeg ou text/pdf. On peut également donner des indications concernant la taille du document (par exemple nombre d'octets). L'utilisation de la liste MIME (liste des « types de média définis sur Internet ») est recommandée.
<identifier> (identifiant)	Référence unique et non ambiguë de la ressource. Il est recommandé d'utiliser un système d'identification, comme URI (<i>Uniform Resource Identifier</i>) ou URL (<i>Uniform Resource Locator</i>).
<language> (langue)	Langue du contenu de la ressource. L'utilisation de la norme ISO 639 est recommandée.
<publisher> (éditeur)	Éditeur (au sens commercial) de la ressource : personne physique ou collectivité.
<relation> (relation)	Ressource(s) liée(s) à la ressource décrite (autres versions, formats...). L'utilisation d'un système d'identification (comme URI, URL...) est recommandée.
<rights> (droits)	Informations sur les droits concernant la ressource décrite (gestion des droits, droits d'auteur...).
<source> (source)	Référence d'origine de la ressource décrite. L'utilisation d'un système d'identification est recommandée.

...

<subject> (sujet)	Sujet du document. L'utilisation d'un système de mots-clefs ou d'indices de classification est recommandée, en particulier les classifications reconnues ou les vocabulaires contrôlés.
<title> (titre)	Titre principal de la ressource, le plus souvent donné par son créateur. [Même s'il n'en est pas question dans la norme, on peut recommander de rendre cet élément obligatoire pour la bonne gestion des systèmes et des échanges de données.]
<type> (type)	Information sur la nature ou le genre de la ressource (texte, image...). L'utilisation d'un terme choisi dans une liste contrôlée, par exemple celle proposée sur le site Dublin Core, est recommandée.

Il est possible de placer ces métadonnées Dublin Core dans l'en-tête de la page HTML, à l'intérieur de l'élément head, sous la forme de balises META. L'attribut *name* contient le nom de l'élément préfixé par « dc » et l'attribut *content* contient la valeur de l'élément.

recommandations pour une bonne gestion des métadonnées

- > Travailler en cohérence sur la granularité,
- > Travailler en cohérence sur les métadonnées sans trop se préoccuper des outils de diffusion dans un premier temps,
- > Travailler sur des formats riches, le Dublin Core n'étant qu'un format d'échanges,
- > Élaborer des « guides pratiques » d'application des formats dans les établissements,
- > Réfléchir en début de projet sur les métadonnées de préservation, de gestion des droits et de provenance,
- > Faire de la veille et consulter les grands établissements pour valider les choix.

INTEROPÉRABILITÉ : LES MÉTADONNÉES COMME FORMAT DE BASE DES PROTOCOLES

Pour dialoguer ensemble, deux ordinateurs ont besoin d'un langage commun. C'est à cela que servent les protocoles, ils définissent le format suivant lequel un ordinateur client et un serveur doivent échanger leurs données pour pouvoir se comprendre.

Les protocoles SRU/SRW

Les protocoles SRU/SRW sont les successeurs de Z39.50 dans le cadre du Web. Les deux protocoles (SRU : *Search and Retrieve via URL* ; SRW : *Search and Retrieve via Web Services*) ont été fondus en 2007. Ce sont des Web services normalisés qui fonctionnent avec les moyens ordinaires du Web : le protocole HTTP lance la requête, la réponse revient en XML. Comme en Z39.50, il est nécessaire de connaître la structure des bases cibles pour lancer la requête.

une requête SRU vers la bibliothèque du congrès

Une requête est donc une simple URL qui contient différents paramètres de recherche :

< <http://z3950.loc.gov:7090/voyager?version=1.1&operation=searchRetrieve&query=title=bear&maximumRecords=10&recordSchema=dc> >

searchRetrieve est l'opération effectuée : en l'occurrence une recherche.
query = title = bear cherche le mot « bear » dans le titre.

maximumRecords = 10 indique que les résultats seront retournés par tranches de 10.

recordSchema = dc précise le format XML des résultats.

SRU < <http://www.loc.gov/standards/sru/> > permet donc de construire une interface de recherche purement Web. Le navigateur présente les résultats. C'est selon cette logique qu'est construite The European Library (TEL), mise en place par la Conférence européenne des directeurs des bibliothèques nationales (CENL), pour les bibliothèques nationales < <http://www.the-europeanlibrary.org> >.

OAI-PMH : open Archives Initiative Protocol for Metadata Harvesting

L'OAI < <http://www.openarchives.org> > est née en 1999 à Santa Fé d'une rencontre autour du Web et du développement des systèmes de publication électronique. Cette rencontre a initié la création d'un protocole d'échange de métadonnées basé sur le Dublin Core, dont l'objectif est de faciliter le signalement des documents et la création de collections intéro-

pérables. Le protocole lui-même s'intitule *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH), et il est actuellement dans sa version 2 (juin 2002), considérée comme définitive. Ce protocole asynchrone utilise des standards du Web (le protocole HTTP et le langage XML) et son fonctionnement repose sur deux acteurs :

- les *Data providers* (fournisseurs de données) qui « exposent » les métadonnées dans des entrepôts ;
 - les *Service providers* (fournisseurs de services) qui mettent en place des opérations de moissonnage ou de collecte des métadonnées exposées dans les entrepôts de façon automatique et incrémentale.
- Une bibliothèque peut choisir d'être « entrepôt » ou « moissonneur » ou peut proposer les deux services.

Pour en savoir plus

Sur le site de la Bibliothèque nationale de France : Stutzmann Dominique. Présentation de l'OAI-PMH : *historique, objectifs, règles de fonctionnement*, 18 novembre 2009. [En ligne] < http://www.bnf.fr/fr/professionnels/anx_poles_autres/a.po_formation_supports.html > et le guide d'utilisation de l'OAI. [En ligne] < <http://bibnum.bnf.fr> > (consultés le 22 janvier 2010).

Dans les entrepôts, les métadonnées sont exposées obligatoirement au minimum en Dublin Core non qualifié, mais d'autres formats de métadonnées plus complexes peuvent être utilisés (par exemple l'EAD) à condition d'être exprimés en XML suivant un schéma accessible en ligne. La base de données de production peut gérer des formats plus riches. D'un point de vue technique, le travail principal pour le fournisseur de données consiste à exposer une base de données accessible via un serveur web et une application capable de répondre aux six requêtes OAI-PMH (*verbs*) générées par un moissonneur et de renvoyer des documents XML valides.

TABEAU Les six requêtes OAI-PMH

verbes	Actions
GetRecord	Demande d'un enregistrement précis, souvent à partir de son identifiant. La requête précise le format de données à utiliser (au minimum le Dublin Core).
Identify	Demande d'informations à propos de l'entrepôt OAI.
ListIdentifiers	Demande de la liste des items contenus dans l'entrepôt (liste complète ou établie selon des critères) avec informations de gestion.
ListMetadataFormats	Demande de la liste des formats de métadonnées disponibles pour un entrepôt
ListRecords	Demande de la liste des items contenus dans l'entrepôt (liste complète ou établie selon des critères) avec informations de gestion et métadonnées elles-mêmes. Cette action permet de récupérer des données de l'entrepôt OAI.
ListSets	Demande des ensembles disponibles dans un entrepôt.

Le protocole OAI est simple à mettre en place pour un service informatique. La structure de la base de données n'est pas modifiée. Les fournisseurs de données stockent simplement les métadonnées et les documents numérisés sont consultables sur le site d'origine. La mise en place de ce protocole est particulièrement adaptée aux projets de bibliothèques numériques partagées. La Bibliothèque nationale de France encourage les partenariats de ce type et moissonne les métadonnées de plusieurs établissements

< http://www.bnf.fr/fr/professionnels/anx_poles_autres/a.po_formation_supports.html >. La liste des partenaires OAI de *Gallica* est disponible par la Recherche avancée de la bibliothèque numérique de la BnF. Dans la liste des résultats, les documents des partenaires sont identifiés par une imagerie ressemblant au Panthéon avec derrière un globe terrestre. Comme le portail

Collections du ministère de la Culture, la bibliothèque numérique européenne *Europeana*, point d'accès commun aux collections patrimoniales européennes (musées, bibliothèques, archives, collections audiovisuelles), fonctionne par le protocole OAI.

Le protocole OAI doit maintenant faire partie intégrante des projets. La visibilité des données qu'il procure doit être complétée par la mise en place de Web services ou API (voir Partie VI).

L'ORI-OAI

L'ORI-OAI (outil de référencement et d'indexation – réseau de portails OAI ; < <http://www.ori-oai.org/> >) permet de gérer et de rendre accessible l'ensemble des documents numériques d'un établissement d'enseignement supérieur et de recherche. Les métadonnées principalement basées sur le Dublin Core et sur LOMFR sont créées dans un système de *workflow*. L'intérêt de l'outil réside dans les possibilités très souples de moissonnage utilisant le protocole OAI-PMH.

LE WEB SÉMANTIQUE ET L'INITIATIVE LINKED DATA

Esquissé dès 1998 par l'inventeur du Web, Tim Berners-Lee, le Web sémantique est un concept étudié depuis 2001 par le W3C et qui se veut l'avenir du Web. Il consiste à créer un Web de données (*a Web of data*) en complément du Web des documents. Le Web sémantique ne changera rien à la manière dont les machines interprètent les documents, mais il leur permettra d'accéder à des contenus qui sont aujourd'hui cachés dans les bases de données (le Web invisible ou *deep Web*).

LES STANDARDS DU WEB SÉMANTIQUE

Le Web sémantique se sert essentiellement des métadonnées sémantiques ou de contenu et vise à créer et à établir des liens entre elles. Reposant donc sur un modèle XML plutôt que HTML, le Web sémantique demande de travailler à plusieurs niveaux, d'une part sur l'exposition optimale des métadonnées sur le Web et d'autre part sur l'enrichissement des métadon-

nées par des méthodes communes et donc rigoureuses d'indexation, reposant sur des normes et des recommandations.

Le standard indispensable à la construction du Web sémantique est l'URI (*Uniform Resource Identifier* = identifiant uniforme de ressource) qui est une syntaxe d'identifiant générique utilisée sur le Web. L'URI permet d'identifier d'une façon certaine et sans équivoque une ressource et doit être unique. Les URL (*Uniform Resource Locator*) sont des sortes d'URI.

Le modèle de données servant de base au Web sémantique est le RDF (*Resource Description Framework*), cadre de description de ressources. Il s'agit d'une infrastructure permettant de décrire formellement une « ressource » (selon le vocabulaire utilisé par le modèle RDF) d'une manière simple, autrement dit un « cadre » théorique et formel pour définir les métadonnées. RDF permet donc de décrire des ressources avec différents jeux de métadonnées (par exemple Dublin Core, XMP...) et dans des implémentations différentes : le plus souvent la syntaxe XML. RDF permet donc de faire entrer de multiples vocabulaires dans un « cadre » commun, rendant possible l'interopérabilité et les échanges de données. Ce cadre commun permet également de passer au-dessus des applications métier.

Ainsi, chaque ressource est décrite par un ensemble de trois balises : le **triplet**. On attribue à la ressource une valeur pour une certaine propriété. Cette information peut être manipulée par un programme informatique. Le triplet est composé du « sujet », du « prédicat » et de l'« objet », autrement dit un sujet, un verbe et un complément. Le sujet est signifié par l'URI de la ressource.

Pour résumer, le **triplet** est constitué de la façon suivante :

<subject> = sujet = ressource décrite

<predicate> = prédicat = propriété ou attribut associée à la ressource

<object> = objet = valeur prise par la propriété

Bien que pouvant être exprimé dans une syntaxe XML, RDF est assez différent de XML. Alors qu'un document XML peut être traduit sous la forme d'un arbre, **RDF** est basé sur le concept de **graphes**. RDF décrit une ressource, relie les données à une catégorie et par conséquent lie les ressources entre elles par l'intermédiaire de ces catégories.

Prenons l'exemple d'un document XML simple :

```
<livre> <auteur>Alexandre Dumas</auteur> </livre>
```

En RDF, ce document est traduit de façon explicite par : Ce livre/a pour auteur/Alexandre Dumas.

vocabulaires RDF

Le modèle RDF est complété par des schémas RDF, vocabulaires définissant des propriétés pouvant être utilisées dans les descriptions. Un schéma RDF est un vocabulaire (donc RDF) écrit avec la norme RDF Schema (RDFS).

grammaire et vocabulaire

Le terme de grammaire est utilisé pour XML. Il est obligatoire de respecter une grammaire dans son ensemble si on veut que le document XML soit valide. Cette notion de validation n'existe pas en RDF. On parle de vocabulaires, des réservoirs de propriétés et de classes dans lesquels on peut puiser à volonté, qu'on peut utiliser en tout ou en partie, mais avec une certaine liberté.

TABLEAU **Exemples de vocabulaires RDF**

FOAF <i>Friend of a friend</i> < http://www.foaf-project.org/ >	Description de personnes et des relations qu'elles entretiennent entre elles.
SKOS <i>Simple Knowledge Organisation System</i> = Système simple d'organisation des connaissances < http://www.w3.org/2004/02/skos/ >	Représentation formelle de systèmes d'organisation des connaissances comme les thesauri ou les vocabulaires contrôlés

ontologies et owl

OWL (*Ontology Web Language*), standard W3C depuis février 2004, est un autre formalisme pour décrire des vocabulaires RDF. OWL vise à décrire, représenter et modéliser des vocabulaires complexes ou ontologies. Une ontologie permet d'exprimer les concepts et connaissances relatifs à un domaine d'application, d'écrire les liens entre les concepts et le sens (ou la sémantique) de ces liens. Un document OWL s'écrit comme un document RDF, donc sous la forme d'un triplet.

Le modèle CIDOC/CRM (*Conceptual Reference Model* = modèle conceptuel de référence ; < <http://cidoc.ics.forth.gr/> >) est une ontologie éditée par le Groupe de normalisation documentaire du Comité international pour la documentation du Conseil international des musées (ICOM-CIDOC), permettant de décrire les objets patrimoniaux. Passé standard ISO 21127 : 2006, la version actuelle 4.2.2 date d'août 2007. La BnF est un acteur majeur dans les travaux menés autour du modèle CIDOC/CRM.

Le *Linked data* (données liées)

Le *Linked Data* est une initiative visant à inciter les détenteurs de données à les passer en RDF et à les rendre accessibles librement sur le Web, ce qui nécessite d'utiliser des URI en HTTP et de les relier avec d'autres données existantes. Les vocabulaires RDF sont un des moyens qui existent pour relier les données.

EXEMPLES D'APPLICATIONS ACTUELLES DU WEB SÉMANTIQUE

Pour que cette exposition de métadonnées puisse améliorer les performances de recherche des machines et ainsi aider les utilisateurs du Web, on compte sur la puissance du protocole d'échange du Web, le HTTP, et sur l'efficacité naissante du XHTML. Le W3C a stabilisé au début de l'année 2008 un langage de requêtes qui reconnaît et exploite les données RDF en permettant de les interroger et de les extraire : il s'agit du langage SPARQL (à prononcer *sparkle*), langage SQL pour le Web sémantique.

Les applications concrètes du Web sémantique commencent à apparaître. Quelques applications peuvent être listées même si la puissance actuelle des moteurs de recherche empêche encore d'entrevoir à l'heure actuelle l'application qui ferait du Web sémantique le tournant décisif du Web.

L'exemple incontournable d'une grande base de données en RDF est DBpedia < <http://dbpedia.org/> > et ses millions de données extraites de Wikipedia. Le but de ce projet encore expérimental et mené par des chercheurs des universités allemandes de Leipzig et de Berlin est d'extraire les informations présentes dans les « infobox » de Wikipedia (infobox = « table préformatée de données dynamiques qui est présente dans des articles sur un même sujet ») et de les transformer de façon automatique selon le modèle RDF. L'intérêt réside dans la masse de ressources RDF ainsi générées associée aux aspects collaboratifs et multilingues de l'encyclopédie. Le langage SPARQL est utilisé pour lancer les requêtes. DBpedia Mobile est une application pour naviguer dans le *Web of data* sur un téléphone portable. L'outil permet un point d'entrée géospatial dans le Web des données. L'idée est de géolocaliser automatiquement l'utilisateur, puis de lui proposer une carte sur laquelle il peut naviguer pour obtenir des informations sur les lieux remarquables, des photos, des services... et également publier ses propres photos ou autres documents. L'idée est de se promener sur les autres sites du *Web of data*. Si les données d'un site sont sur le *Web of data*, il n'est plus nécessaire de développer des interfaces compliquées pour favoriser la mobilité : les données seront accessibles à partir d'applications développées par d'autres.

LIBRIS, < <http://libris.kb.se/> > catalogue collectif des bibliothèques suédoises (170 bibliothèques connectées), développé en Méthode Agile, est le premier catalogue de bibliothèque intégralement disponible dans le *Linked Data*. Ce catalogue gère 6 millions de notices bibliographiques, 20 millions de notices d'exemplaires et 200 000 notices autorités (auteurs, titres, sujets). Ce catalogue utilise plusieurs ontologies : FOAF, SKOS et Bibliontology. Le modèle RDF utilise également les possibilités des FRBR permettant de relier les enregistrements ou notices relevant de la même œuvre. Les concepteurs de ce catalogue nouvelle génération ont en projet de le lier à DBpedia.

Les liens vers les données externes sont réalisés par différentes techniques : des API, OpenSearch, Z39.50, SRU et OpenURL-Coins.

Les technologies du web sémantique appliquées aux bibliothèques numériques

Le Thought Lab d'*Europeana* < <http://www.europeana.eu/portal/thought-lab.html> > et The MultimediaN E-culture < <http://e-culture.multimedien.nl/> > expérimentent le Web sémantique et proposent des interfaces à facettes dans des images à partir des données en RDF. Il est possible de rebondir de concept en concept.

Depuis quelques mois, les moteurs de recherche généralistes commencent à tirer parti de la richesse des données et proposent des affichages plus complets dans leurs résultats de recherche.

- Yahoo ! propose aux webmasters de créer une application sur SearchMonkey pour afficher de l'information riche sur les extraits affichés dans les résultats de recherche.

Ainsi Calames < <http://www.calames.abes.fr/> > utilise cette possibilité donnée par le Web sémantique (cf. le blog de Calames < <http://calames.wordpress.com/> >).

- Depuis quelques mois, Google commence à utiliser les métadonnées du Web sémantique : RdFa ou microformats pour afficher des textes de présentation (*Snippets*) enrichis dans ses résultats de recherche.

Le Web sémantique intéresse directement les métiers de la documentation par l'intérêt commun porté aux métadonnées et à l'information structurée. La mise en œuvre du Web sémantique est d'ailleurs identifiée comme un grand chantier de la BnF pour les années à venir. Le rapport Tessier⁵⁷ sur la « numérisation du patrimoine écrit » (janvier 2010) mentionne le Web sémantique comme l'un des enjeux majeurs du développement des bibliothèques numériques.

57. Ministère de la Culture et de la Communication. [En ligne] < <http://www.culture.gouv.fr/mcc/Actualites/A-la-une/Mission-sur-la-numerisation-du-patrimoine-ecrit> > (consulté le 22 janvier 2010).

presses ensib

PARTIE VI

**METTRE EN
LIGNE DES
DOCUMENTS
NUMÉRISÉS,
ÉVALUER,
RÉFÉRENCER**

METTRE EN LIGNE DES DOCUMENTS NUMÉRISÉS, ÉVALUER, RÉFÉRENCER

par François
Houste
et Isabelle
Westeel

METTRE EN LIGNE

Mettre en ligne des documents numérisés consiste à mettre en place sur le Web un système de recherche documentaire et une interface proposant des fonctionnalités et des services permettant la consultation de documents et leur réappropriation par le public. Ce type de projet entre dans les missions de valorisation des bibliothèques en plaçant les collections au sein d'un réseau.

CONTEXTE ET TYPOLOGIE DES SYSTÈMES

Les bibliothèques ont des besoins très variés en terme de construction d'un système d'information souvent assimilable à une gestion électronique de documents (GED). La typologie des documents entraîne des usages très divers : on ne consulte pas une collection iconographique ou photographique comme on feuillette un ouvrage ou comme on recherche un numéro précis dans une collection de presse.

Dans les bibliothèques, les logiciels permettant la mise en ligne de contenus relèvent du marché de l'informatique documentaire qui souffre de l'absence d'une véritable demande et de la crise économique qui a amené certains fournisseurs étrangers à se retirer du marché français. Quelques prestataires traditionnels des bibliothèques présentant une offre logicielle essentiellement tournée vers les systèmes intégrés de gestion de bibliothèque (SIGB) proposent également des modules correspondant aux besoins de mise en ligne d'une bibliothèque numérique ou des offres globales et intégrées de navigation dans les données et les documents au travers d'un portail. Cette offre traditionnelle reste peu satisfaisante au regard des besoins et de la demande du public habitué aux moteurs de recherche du Web et à leur offre de services performants et intuitifs. Dans « l'état des lieux 2009 des logiciels métier pour bibliothèque » publié dans le magazine *Archimag* de juin 2009, Marc Maisonneuve de la société Tosca

Consultants cite 32 gestionnaires de bibliothèques numériques⁵⁸, essentiellement issus des logiciels de gestion. On y trouve des prestataires traditionnels comme les sociétés Archimed, Cadic, Ever Team (solution Flora) ou encore Decalog, Ex libris (Digitool) ou GB Concept (Alexandrie). On peut ajouter à cette liste eDip, Cindoc ou JLBnet GEIDE.

Il existe également une offre commerciale spécifique correspondant à la catégorie des « photothèques » et équipant par exemple les agences de presse qui ont de gros besoins de gestion de ressources numériques, par exemple Orphea Studio de la Société Algoba, Ajaris Pro d'Orkis, Armadillo, Phraseanet, e-photo d'Einden Studio ou encore Bagherra de Propixo. Dans le monde universitaire, plusieurs systèmes de publication adaptés aux « archives ouvertes » sont utilisés comme Eprints ou DSpace. Devant la faiblesse de l'offre, certains prestataires de numérisation ont développé des offres « clés en main » et des solutions « packagées » couplant la numérisation des documents et leur mise en ligne, y compris bien souvent l'hébergement. Ces solutions attractives en raison de leur coût raisonnable séduisent certains établissements. Il convient dans ce cas de s'assurer de la bonne administration des données mises en ligne, des conditions financières liées à l'importation de données nouvelles dans le système de publication, des modalités d'hébergement ainsi que de l'évolutivité et de la visibilité du site.

Plusieurs solutions libres peuvent être adaptées aux besoins des établissements, leur stabilité dépendant essentiellement de la force de la communauté entourant le logiciel (développeurs et utilisateurs) et de l'environnement informatique de l'établissement. Dans ce domaine, les solutions les plus stables sont Greenstone, logiciel néo-zélandais et Alfresco ou Nuxeo, ces deux derniers relevant plus de la gestion des contenus d'entreprise. Le monde du libre propose également plusieurs outils intéressants pour la mise en place d'un système. MySQL est une base de données open source très efficace. Lucene (< <http://lucene.apache.org/> >) est un moteur

58. Étude également publiée sur le site de Tosca Consultants. On peut également consulter la page « Solutions logicielles pour bibliothèques numériques » de Bibliopedia. [En ligne] < http://www.bibliopedia.fr/index.php/Solutions_logicielles_pour_biblioth%C3%A8ques_num%C3%A9riques > (consulté le 22 janvier 2010).

de recherche open source de la fondation Apache qui permet d'indexer et de rechercher du texte sur des volumes importants de données. Il est utilisé par *Gallica* et *Europeana*.

Très peu de solutions « clés en main » existent réellement sur le marché actuel et la mise en ligne constitue un frein majeur à la valorisation des ressources dans la plupart des établissements. Dans la plupart des cas, la mise en ligne de documents numérisés ne peut se contenter de systèmes de gestion de contenus (SGC ou CMS *Content Management System*) de type SPIP, Joomla ou encore Typo 3, étant donné la richesse des informations contenues dans les documents et la performance attendue des moteurs de recherche.

Quelle que soit la solution choisie, la véritable difficulté pour les bibliothèques reste d'intégrer cette nouvelle brique dans le système d'information existant : SIGB, portail et site Web, en conservant une cohérence globale du point de vue des usagers (fluidité de navigation entre les différentes applications, lien entre une référence du catalogue et un document de la bibliothèque numérique). Du point de vue des professionnels, il va s'agir d'administrer et d'exploiter une application supplémentaire, pas forcément conforme aux règles du système d'information en place (prise en compte d'une gestion des utilisateurs, d'une gestion des droits sur les documents, choix de formats de données normalisées et établissement d'une politique d'archivage).

LES FONCTIONNALITÉS ET LES SERVICES ATTENDUS D'UNE BIBLIOTHEQUE NUMERIQUE

Pour faire l'acquisition d'une solution logicielle de mise en ligne, les responsables de projet seront le plus souvent amenés à rédiger un cahier des charges listant les fonctionnalités attendues du système sachant qu'une bibliothèque numérique est constituée de trois éléments principaux : une base de données, un moteur de recherche et un système de visualisation des données et des documents. On peut penser que la solution peut venir de ce type d'assemblage à condition que les différentes briques soient ouvertes et que le projet soit très sérieusement documenté. De nombreuses bibliothèques américaines fonctionnent de manière très pragmatique sur ces questions : à chaque type de documents son système de navigation et son type d'interface⁵⁹.

Accompagnement de la prestation

Dans le choix de la mise en ligne on veillera à ce que les conditions de stabilité du système soient réunies. L'offre du prestataire devra détailler les points suivants :

- installation des logiciels et fourniture de la documentation détaillant les outils utilisés et les procédures d'installation et d'exploitation,
- paramétrabilité du système en fonction des souhaits de l'établissement,
- scalabilité du système permettant de faire monter le logiciel en puissance en fonction des volumétries à mettre en ligne,
- formation du personnel (généraliste et spécifique), transfert de compétences et assistance au démarrage,
- maintenance et suivi des évolutions majeures et mineures,
- existence d'un club utilisateurs ou étendue de la communauté dans le monde du libre,
- moyens mis en œuvre en matière de suivi (centre d'appels, hotline, horaires, délais d'intervention, effectifs...).

Il est donc nécessaire de mettre en place un groupe de travail qui élaborera un cahier des charges suffisamment ouvert pour permettre aux prestataires de répondre et de proposer des solutions. L'écriture du cahier des charges impose de définir avec soin les fonctionnalités et les services que l'établissement veut offrir à travers le système d'information. Étant donné l'état du marché, il est conseillé de trouver une juste mesure entre un cahier des charges très (trop) précis et contraint qui empêcherait les sociétés de répondre et un cadre trop large qui ne permettrait pas la réalisation du marché et donc du contrat passé entre la personne publique et le prestataire. En ce qui concerne la procédure et dans le cas de projets importants, il peut être opportun d'utiliser la procédure du dialogue compétitif évoqué dans la Partie III « autres procédures

59. Bibliothèques américaines fonctionnant sur différents types de documents : New York Public Library < www.nypl.org > et MLibrary-University of Michigan-Library Information Technology (LIT) < <http://www.lib.umich.edu/lit/> > (consultés le 5 mars 2010).

formalisées (voir p. 77) et éventuellement de se faire accompagner d'une assistance à maîtrise d'ouvrage (AMO).

La « Méthode Agile » qui consiste à utiliser une procédure de construction commune (client/prestataire) d'une solution logicielle avec développements itératifs et tests progressifs peut être envisagée pour ce type de prestations, à condition d'être appuyée par le service informatique et le service des marchés.

mettre en œuvre la méthode agile pour des projets innovants

Cette méthode a fait l'objet d'une expérience concluante pour la construction du catalogue collectif des bibliothèques suédoises, Libris. [En ligne] < <http://libris.kb.se/> > (consulté le 22 janvier 2010).

Lindström Henrik, Malmsten, Martin. User-Centred Design and Agile Development : Rebuilding the Swedish National Union Catalogue, *Code{4}Lib Journal*, 2008, Issue 5, p. 12-15. [En ligne] < <http://journal.code4lib.org/articles/561> > Voir l'analyse sur Figoblog : < <http://www.figoblog.org/node/1946> > (consultés le 22 janvier 2010).

La liste des fonctionnalités et des services fera l'objet, au sein de réunions du groupe projet, d'une étude approfondie. Il se déroule au minimum 18 mois entre l'écriture du cahier des charges et le lancement du projet. Il conviendra donc d'être particulièrement vigilant sur les fonctionnalités liées au Web, de faire une veille technologique importante afin de prévoir les mécanismes auxquels sont habitués les internautes.

Un système de GED traite de façon classique de la production/acquisition, de la gestion, de la recherche et de la consultation ainsi que du stockage (vers archivage) des documents. Les fonctionnalités classiques d'un tel système seront donc reprises dans le cahier des charges. Il conviendra également d'évaluer en premier lieu les supports (texte, image, son, vidéo) et les formats qui devront être acceptés par le système. On pourra ensuite insister sur les fonctionnalités et les services suivants :

Fonctionnalités	
Recherche	<p>Recherche libre, guidée, avancée, par index</p> <p>Recherche paramétrée selon les types de documents (ex. la presse demande des modes de recherche spécifiques (date, période, mot du texte)</p> <p>Recherche par mots dans le document</p> <p>Recherche à facettes (les filtres sont proposés <i>a posteriori</i> et non <i>a priori</i>) Ce type de recherche permet de raffiner pas à pas.</p> <p>Recherche sémantique</p> <p>Limitations de recherche (périodes, types de documents...), filtres (tri par date, par pertinence, affichage)</p> <p>Remarque : un moteur d'indexation est efficace si les métadonnées sont bien établies (« la bonne donnée au bon endroit ») et riches (gestion des thesauri)</p>
Affichage	<p>Affichage de la liste des résultats (classement des réponses, affichage des informations)</p> <p>Affichage des résultats (affichage plein-écran, notice + image, image : format, zoom avant-arrière, <i>watermarking</i>...)</p>
Navigation	<p>Gestion de l'hypertexte, rebonds et filtres</p> <p>Passage d'une image à l'autre, fonctionnalité d'un tourne-page</p> <p>Gestion de la granularité : comment remonter dans une collection ou dans un dossier/navigation dans les corpus et pas seulement dans les documents</p> <p>Gestion de la table des matières en mode texte</p> <p><i>Clustering</i> (données proches)</p>
Visualisation	<p>Consultation, lecture, annotations</p> <p>Rotations, miroirs, planches-contact, mosaïques, comparaison d'images</p>
Téléchargement et impression	<p>« origine » (mention incrustée dans l'image « bibliothèque de... cote... »)</p> <p>Téléchargement (complet ou partiel, en PDF...), commande : formulaire pré-rempli</p>
Guides d'utilisation	Aides, FAQ

...

Services	
Gestion de l'historique de recherche et des paniers	Intra-session (anonyme) ou inter-session (personnalisé) ; Gestion de la fonctionnalité : « Envoyer à un ami »
Annonce des nouveautés	Fils RSS ou Diffusion sélective de l'information
Web 2.0	Syndication de contenus (fils RSS, API...), services liés aux réseaux sociaux et aux communautés d'utilisateurs, participation à l'enrichissement du contenu (tagging...)
Interopérabilité, connexion et réappropriation	Gestion des identifiants ; Gestion des protocoles SRU/SRW et OAI-PMH Gestion des systèmes de recommandations (ouvrages liés, mots-clés liés...) Articulation avec le catalogue de la bibliothèque par le champ 856 Accessibilité par recherche fédérée Compatibilité avec les lecteurs e-books du marché ; Compatibilité avec les terminaux mobiles (smartphones, PDA, téléphones, liseuses...) ; Adaptabilité avec les logiciels de reconnaissance vocale Adaptabilité avec les logiciels de sauvegarde de bibliographie : Zotero, EndNote et RefWorks Connexions avec les moteurs de recherche (cf. firefox...) ; Gestions d'identité
Produits éditoriaux	Cédéroms, expositions, parcours chronologique, géographique (ex. Bibliothèque numérique de Roubaix)

Identifiants et URL

Les URL stables (voire pérennes) permettent d'assurer un lien permanent vers la ressource même si son emplacement physique ou le logiciel d'accès changent. Elles sont les moyens de connecter différents services informatiques (catalogues, blogs, sites Web...) avec la bibliothèque numérique. Pour établir ces URL on peut choisir plusieurs solutions résumées ci-dessous :

- conception d'un **système de nommage signifiant** s'appuyant sur l'arborescence hiérarchique de la collection : exemple retenu à la bibliothèque nationale d'Australie (NLA) ou reposant sur une carte de structure.

Ressource : Persistent Identifier scheme for digital collections at the National Library of Australia. [En ligne]

< <http://www.nla.gov.au/initiatives/nlapi.html> > (consulté le 22 janvier 2010).

- choix d'un **système de nommage opaque** basé sur des combinaisons alphanumériques non signifiantes. Les identifiants opaques sont attribués au niveau de l'objet exclusivement, et ils ne gèrent pas l'arborescence de la collection ni sa structure comme peuvent le faire les identifiants signifiants.

Le système d'identifiants pérennes créé et maintenu par la California Digital Library et choisi par la BnF, *Archival Resource Key* (ARK) combine nommage opaque pour les objets primaires, et nommage signifiant pour les composantes et les variantes.

LE WEB 2.0

+++++

Le Web 2.0 est un ensemble de techniques et de fonctionnalités visant à rendre l'internaute acteur, plus que spectateur, des sites Web qu'il fréquente.

LES PRINCIPAUX SERVICES DU WEB 2.0

On retrouve sous le terme de Web 2.0 un catalogue de fonctionnalités et de sites assez éparés qui gardent comme point commun le fait de pousser l'internaute à agir et à participer.

- les plateformes d'hébergement vidéo (YouTube, DailyMotion...) sur lesquelles l'internaute est invité à verser ses propres films ;
- les blogs, sites Web composés de billets personnels, où l'internaute peut à la fois jouer le rôle d'auteur (le blogueur lui-même) et de commentateur. Extension de ce moyen s'apparentant à un journal intime, l'outil de micro-

blogging Twitter, créé en 2006 et très populaire en 2009, permet de poster des messages brefs par Internet, par messagerie instantanée ou par SMS ;

- les wikis, systèmes de gestion de contenus sur le Web, sur les pages desquels les internautes autorisés ont la possibilité d'intervenir. Le wiki le plus utilisé est l'encyclopédie en ligne Wikipédia ;

- les réseaux sociaux (MySpace, Facebook...) basés avant tout sur la création de liens et de contacts entre les différents utilisateurs ;

- l'utilisation de la folksonomie et du *tagging* (qualification de l'information) collaboratif. Mis en place originellement par Flickr, puis relayé par d'autres sites de partage de contenus tels que delicious ou Technorati, le *tagging* associe un mot-clé librement choisi à un contenu afin de qualifier celui-ci. Cette indexation libre permet d'ajouter de nouveaux points d'accès et peut être proposée sous différentes interfaces de classement sur les sites. La technique du nuage de mots (ou de *tags*) figure de façon graphique les mots-clés les plus utilisés en présentant ceux-ci avec une taille plus grande ou parfois à l'aide de couleurs différentes ;

- les flux RSS qui permettent à un internaute d'obtenir sur sa machine, à l'aide d'un agrégateur, l'ensemble des mises à jour des sites auxquels il s'est abonné sans avoir à se rendre sur les sites eux-mêmes. [Techniquement, un fichier RSS n'est rien d'autre qu'un fichier de type XML, structuré de manière fixe, et présentant les contenus exportés d'un site Web].

QUELQUES EXEMPLES LIÉS AUX BIBLIOTHÈQUES NUMÉRIQUES

Ces services directement issus du mouvement du Web 2.0 et très demandés par les internautes ont trouvé des applications concrètes dans certaines bibliothèques numériques. On peut classer ces services en trois catégories⁶⁰ : la syndication de contenus, les services liés aux réseaux sociaux et aux communautés d'utilisateurs et enfin la participation à l'enrichissement du contenu.

60. Selon la typologie de Dominique Stutzmann dans *Gallica 2 : les pratiques collaboratives dans l'environnement d'une bibliothèque numérique et les développements Web 2.0 de la BnF, Rencontres 2008 des professionnels de l'IST « Information numérique : nouveaux produits – nouveaux usages »* Nancy – 16-18 juin 2008. [En ligne] < <http://rpist.inist.fr/sites/rpist/upload/supports2008/stutzmann.pps> > (consulté le 22 janvier 2010).

TABLEAU **QUELQUES EXEMPLES D'APPLICATIONS WEB 2.0
DANS LES BIBLIOTHÈQUES NUMÉRIQUES**

syndication de contenus	
Flux RSS, agrégateurs	<i>Gallica</i> , liste des derniers documents numérisés ou recherche précise
API, <i>mashup</i> , <i>widget</i>	Voir plus loin
communautés	
Réseaux sociaux	Second life (Stanford Libraries, Bayerische Staatsbibliothek) Facebook, Delicious : (Bibliothèque municipale de Toulouse)
Recommandations de type Amazon (les personnes qui ont consulté ce document ont aussi consulté)	Bibliothèque numérique de la Médiathèque de Dole
Intégration dans un espace personnel	<i>Gallica</i> , projet MonCarnet 2.0 du CN2SV (Centre national pour la numérisation de sources visuelles)
participation à l'enrichissement du contenu	
Annotations collaboratives	Archives départementales des Yvelines, Archives départementales de la Mayenne, Archives municipales de Rennes
Blogs	New York Public Library Digital Labs : blog sur les différents outils numériques en usage à la New York Public Library Blog <i>Gallica</i> http://blog.bnf.fr/gallica/
Wikis	Accès aux ressources de <i>Gallica</i> par les articles de la Wikipedia (par exemple article Les Misérables)
Folksonomie	Flickr Commons mis en œuvre par exemple en France par la Bibliothèque municipale de Toulouse ; Exemple de <i>tagging</i> aux États-Unis : Indianapolis Museum of Art Online Collection < http://www.imamuseum.org/connect/tags >
Corrections d'OCR	Wikisource

L'IMPACT DU WEB 2.0

Si les technologies et les systèmes de navigation liés au Web 2.0 sont des opportunités formidables pour la diffusion de base de données, il faut toutefois sérieusement étudier l'impact et les possibilités réelles de celles-ci. La tentation de l'utilisation de technologies avancées pour le simple plaisir de leur utilisation doit à tout prix être évitée. Les interfaces graphiques développées doivent répondre à des normes d'accessibilité clairement définies et servir avant tout les demandes de l'utilisateur. L'impact du Web 2.0 et des médias participatifs sur la vie d'un site Web ou d'une base de données est à relativiser. Permettre aux visiteurs extérieurs de qualifier une information ne veut pas dire que ceux-ci vont profiter réellement de cette opportunité. D'une façon générale et en particulier pour les métiers de la documentation on obtient de meilleurs résultats en matière de collaboration en ligne en ciblant une communauté d'intérêt proche (généalogistes, historiens amateurs...) et en permettant à celle-ci de s'exprimer, qu'en essayant de cibler un public très large mais peu impliqué sur un sujet.

METTRE EN ŒUVRE DES APPLICATIONS COMPOSITES ET DISSÉMINER LES DONNÉES

Le mouvement du Web 2.0 a également provoqué l'essor de la mise en œuvre d'applications composites utilisant des programmes mis à disposition dans d'autres systèmes et par d'autres établissements : les API, les *mashups* et les *widgets*, ces moyens participant également à la syndication de contenus.

- Une *Application Programming Interface* (API) ou interface de programmation est constituée de fonctions et de procédures qu'un programme informatique met à la disposition d'autres programmes. Elle permet l'interopérabilité entre plusieurs systèmes informatiques. Les fonctionnalités d'un système sont ainsi rendues disponibles à d'autres. De plus en plus, les services mis en ligne fournissent des API permettant à d'autres logiciels de récupérer leurs fonctionnalités : on peut citer par exemple l'API Google Maps dynamique permettant d'intégrer des éléments de cartographie sur un site ou encore l'API de Géoportail pour intégrer les données de l'IGN et des outils cartographiques sur un site Web. L'utilisation de ces programmes extérieurs permet de proposer des interfaces correspondant au goût des

internautes. L'arrivée de Google Maps et de Google Earth a par exemple montré l'engouement du grand public pour la cartographie en ligne.

un exemple d'utilisation d'API

Sur le site *New York then and now*, un développeur australien a récupéré sur Flickr les photographies de New York mises à disposition par la New York Public Library et a utilisé l'API de Google Maps pour les placer sur une carte géographique [En ligne] < <http://www.paulhagon.com/thenand-now/nypl/> >.

La mise en relation des photographies anciennes avec les vues actuelles de la ville est rendue possible par le géocodage.

Il est bien entendu possible d'utiliser ce système d'API pour mettre la bibliothèque numérique en relation avec d'autres sites de bibliothèques ou de librairies.

- Un *mashup* : un site Web composé de la combinaison sur une page d'au moins deux sources de données forme un *mashup* ou application composée. Le *mashup* combine les API pour extraire les données et fonctionnalités d'autres sites. Le contenu du site est composé de façon dynamique pour répondre aux demandes des internautes.

un exemple de mashup : le site worldcat identities

Le site WorlCat Identities < <http://orlabs.oclc.org/Identities/> > est un *mashup* de données sur les auteurs. Ce site fait appel à plusieurs API : WorlCat, dial-a-book (librairie en ligne), AddThis (gestion de bookmarks), Google books, Amazon, weRead (réseau social de lecture).

- Un *widget* (contraction des mots Windows® et gadget) est un composant graphique que l'on peut placer dans une application, sur un bureau ou sur un site Web. Son contenu est composé de façon dynamique. Un *widget* permet d'afficher par exemple des données météo, des actualités, un traducteur sur une interface. Il permet aussi de construire une interface de

recherche à installer sur un site Web ou sur son environnement de travail. Un *widget Gallica* est disponible sur l'univers Netvibes.

Il est donc possible d'utiliser des services offerts à l'extérieur pour construire une interface ou des compléments d'interfaces. Il est alors nécessaire que le système d'information de l'établissement soit suffisamment ouvert pour les accepter. *A contrario*, la bibliothèque peut programmer et proposer des API qui seront utilisées par d'autres établissements.

FAIRE ÉVOLUER LES INTERFACES

L'évolution rapide du Web et l'essor de la culture numérique obligent les bibliothèques à penser autrement la valorisation de leurs collections pour s'adapter aux pratiques culturelles.

FEUILLETER/PARCOURIR

Le feuilletage sur écran de pages d'un ouvrage numérisé a fait l'objet d'un développement mené par la British Library dès 1997. L'application Turning the pages™ dans sa version 2.0 permet de zoomer, de lire des commentaires et de poser des notes. Il existe à présent plusieurs outils (libres ou payants), souvent en technologie Adobe® Flash®, permettant de mettre en ligne un livre virtuel. En novembre 2009, la BnF a mis en place sur son site une fonctionnalité permettant le feuilletage de tous les livres de *Gallica* en utilisant un nouveau visualiseur en technologie Adobe® Flex®. Ce visualiseur, complémentaire de la navigation plus traditionnelle, permet de faire défiler les pages verticalement ou horizontalement, deux par deux en vis-à-vis ou page par page. Enfin les laboratoires de recherche publics et privés réfléchissent à de nouvelles possibilités d'interfaces (par exemple, la technologie Pivot™ de Microsoft® < <http://www.getpivot.com/> >).

scénariser

Certains établissements prennent le parti de valoriser leurs collections à l'aide de parcours ou de scénarios. La construction de ces sites utilise des technologies assez classiques du Web que de nombreuses agences multimédia maîtrisent. L'important dans ces projets est de réfléchir aux objectifs, aux particularités des collections et aux publics-cibles.

On peut citer en exemple les dossiers pédagogiques de la BnF ([En ligne] < <http://classes.bnf.fr> >), l'*ABC décoiffé du dragon ailé* de la Bibliothèque municipale de Rouen

([En ligne] < <http://bibliotheque.rouen.fr/creaplus/abcd/> > ou encore *Le Mystère de la Cordelière* des Archives départementales de l'Aube ([En ligne] < http://www.archives-aube.com/atelier_pedagogique/travailler_en_classe/mystere_cordeliere.php >)

PENSER SA BIBLIOTHÈQUE NUMÉRIQUE SUR MOBILE

Avec les évolutions technologiques, les usages d'Internet évoluent et de plus en plus d'accès au Web se font aujourd'hui via des terminaux mobiles (31 % des possesseurs de téléphone mobile en France surfent sur Internet selon le Baromètre de suivi des services mobiles – Ipsos AFMM). Parmi ces terminaux, on notera particulièrement la part de marché croissante des Smartphones - téléphone mobile couplé aux fonctionnalités d'un PDA (16 % des ventes en novembre 2009 selon une étude GfK), parmi lesquels l'iPhone fait figure de produit phare (50 % des ventes de Smartphones en novembre 2009). Cette évolution des terminaux oblige désormais à penser sa présence sur téléphone mobile de deux façons distinctes : un site mobile qui peut proposer des informations de base et une application destinée aux Smartphones qui se doit d'apporter une valeur ajoutée forte aux visiteurs via des fonctionnalités graphiques avancées (voir l'application iPhone du Louvre lancée fin 2009) ou l'utilisation de la géolocalisation. Par exemple, depuis avril 2009, une nouvelle application, baptisée *Places* disponible sur la téléphonie mobile permet d'accéder en un seul clic à une information complète sur une localité. Il s'agit de la première tentative de recherche transversale appliquée à Google Earth et disponible sur téléphonie mobile. Le patrimoine pourrait emboîter ce pas majeur dans l'évolution du eTourisme. Le marché de l'Internet mobile est encore en 2010 en pleine

mutation, une phase d'étude reste donc indispensable avant le lancement d'un projet pour cette nouvelle audience.

LES ÉCRANS TACTILES (LE MULTITOUCH)

Il est également nécessaire de réfléchir à la valorisation de l'objet numérique au travers de nouvelles interfaces qui ne reproduisent pas les interactions traditionnelles. Les applications multitouch peuvent constituer des alternatives permettant de mettre en place des interfaces multimédia innovantes et d'élaborer des visites interactives multi-utilisateurs complémentaires des navigations sur le Web. Une application multitouch⁶¹ s'inscrit dans la famille des interfaces tangibles (*Tangible User Interface*, TUI). Le clavier et le clic de la souris sont remplacés par le doigt ou par le déplacement d'un objet. Lorsqu'on évoque le multitouch, on pense à Apple qui a permis de démocratiser cette technologie auprès du grand public, mais d'autres modèles apparaissent permettant l'amplitude du geste sur de grandes surfaces et surtout la coopération entre plusieurs utilisateurs autour d'une seule machine : Microsoft® avec la table Surface® ou Jazzmutant⁶² avec la technologie Lemur. Le tactile devient partie intégrante des interfaces utilisateurs (cf. Microsoft® Windows® 7). Le multitouch peut prendre alors toute sa dimension sur une autre échelle que celle du portable. Les applications à concevoir sont multiples : interface murale avec capteurs analogiques ou avec pointeurs type télécommandes infrarouges ou télécommande wiimote, table multitouch avec ou sans objets. Ce type d'application a déjà donné lieu à plusieurs expérimentations dans le cadre du patrimoine. En 2007, l'Australian Museum⁶³ à Sydney a installé la table interactive Surface®-Like (LightWell) permettant aux visiteurs de découvrir et de s'informer sur la faune australienne. En France,

61. Voir les articles de l'encyclopédie Wikipédia. [En ligne] < <http://en.wikipedia.org/wiki/Multi-touch> > ; < http://en.wikipedia.org/wiki/Tangible_User_Interface > et < http://en.wikipedia.org/wiki/Surface_computing > (consultés le 22 janvier 2010).

62. Jazzmutant. [En ligne] < http://www.jazzmutant.com/lemur_overview.php > (consulté le 22 janvier 2010).

63. [En ligne] < <http://www.nowhereelse.fr/une-table-interactive-surface-like-sinville-au-museo-vido-9618/> > (consulté le 22 janvier 2010).

le Centre expérimental Erasme⁶⁴ (Département du Rhône) a développé des applications de type pédagogique. Dans le cadre des bibliothèques numériques, on peut penser à des interfaces intuitives à inventer avec par exemple échange, manipulation et classement de photographies.

RENDRE ACCESSIBLE

+++++

Rendre un site Web accessible, c'est le mettre à disposition de tous les individus, quel que soit leur matériel ou logiciel, leur infrastructure réseau, leur langue maternelle, leur culture, leur localisation géographique, ou leurs aptitudes physiques ou mentales. Cela permet, en particulier aux personnes handicapées, d'avoir accès à l'information, à l'éducation et à la culture.

LA WEB ACCESSIBILITY INITIATIVE (WAI)

En 1996, le W3C, consortium chargé de réguler le Web, a créé la *Web Accessibility Initiative* (WAI) dans le but d'améliorer l'accessibilité. En 1999, les *Web Content Accessibility Guidelines* (WCAG) version 1.0 ont été créés en direction des créateurs de contenu qui définissent 3 niveaux de priorité correspondant à 3 niveaux d'accessibilité :

- A (*Web developers must satisfy these requirements*) ;
- AA (*Web developers should satisfy these requirements*) ;
- AAA (*Web developers may satisfy these requirements*).

La version 2.0 des WCAG a été publiée le 11 décembre 2008. La recommandation est structurée autour de quatre principes : les contenus doivent être « perceptibles », « utilisables », « compréhensibles » et « robustes ». Les niveaux de recommandation A, AA et AAA sont conservés.

64. [En ligne] < <http://www.erasme.org/Tabletouch-et-Papas-Andina> > (consulté le 22 janvier 2010).

LE RÉFÉRENTIEL GÉNÉRAL D'ACCESSIBILITÉ POUR LES ADMINISTRATIONS

En 2002, le Parlement européen a adopté une résolution qui reconnaît l'importance de l'accessibilité du Web dans les institutions européennes et les États membres. Elle indique que le niveau AA est requis pour les sites publics. En France, en février 2004 l'ADAE (Agence pour le Développement de l'Administration Electronique) publie le premier « Référentiel accessibilité des services Internet de l'administration ». Ce référentiel est issu d'un travail avec le service AccessiWeb de l'association BrailleNet, association majeure œuvrant pour améliorer l'accessibilité numérique, notamment à destination des personnes handicapées visuelles.

La loi du 11 février 2005 pour « l'égalité des droits et des chances, la participation et la citoyenneté des personnes handicapées » impose aux services de communication en ligne du secteur public le respect de recommandations internationales pour l'accessibilité. L'article 47 stipule que « les services de communication publique en ligne des services de l'État, des collectivités territoriales et des établissements publics qui en dépendent doivent être accessibles aux personnes handicapées... ». Le décret n° 2009-546 du 14 mai 2009 pris en application de l'article 47 impose une mise en œuvre de l'accessibilité dans un délai de deux ans (à partir de la publication du décret) pour les services de communication publique en ligne de l'État et des établissements publics qui en dépendent, et de trois ans pour les services de communication publique en ligne des collectivités territoriales et des établissements qui en dépendent.

Dans un souci de normalisation, l'État publie des référentiels. En mai 2007, était parue une première version du Référentiel général d'accessibilité pour les administrations (RGAA). Le 23 octobre 2009 paraît la version 2.2 du RGAA (version 2.2.1 légèrement modifié publié le 16 novembre 2009).

< <http://references.modernisation.gouv.fr/rgaa-accessibilite> >

« Le niveau attendu pour les sites concernés par le RGAA est le niveau AA du WAI. Pour être conforme au RGAA, il est donc nécessaire de valider l'ensemble des tests ayant un niveau WCAG déduit A et AA. Les critères de succès associés au niveau AAA peuvent être pris en compte dans certains contextes, lorsque cela est possible et pertinent. »

< <http://references.modernisation.gouv.fr/5-modalites-dapplication-du-rgaa> >

LE LABEL ACCESSIWEB

Parallèlement, le groupe AccessiWeb (Association BrailleNet) a fait des propositions pour adapter le premier référentiel de l'administration à l'évolution du Web (standards européens et transition vers le WCAG 2.0). Le référentiel AccessiWeb v.1.1 a été publié le 9 juin 2008, la version 3.1 le 2 mars 2009. Le but de ce guide est de permettre le développement de sites Web accessibles. Il a débouché sur la création d'un label de qualité pouvant donner lieu à une certification par une tierce partie. Le 17 décembre 2009 est paru officiellement le référentiel AccessiWeb 2.0. en correspondance avec les WCAG 2.0. Le référentiel AccessiWeb 2.0 établit 131 critères d'accessibilité répartis en trois niveaux de priorité/conformité (Bronze, Argent et Or). Il peut être utilisé dans le cadre d'une procédure de certification de conformité d'un site ou d'un service WAI Web aussi bien aux WCAG 2.0 qu'au RGAA 2.2. Il est donc recommandé aux établissements de faire labéliser leur site dans ce cadre. Les nouvelles règles d'application pour le label AccessiWeb seront proposées par BrailleNet dans le courant du 1^{er} semestre 2010. (Association BrailleNet < <http://www.accessiweb.org> >).

TABLEAU LES NIVEAUX DU LABEL ACCESSIWEB

Niveau Bronze = Niveau A de WAI	Site accessible de façon satisfaisante. Possibilité pour les personnes handicapées de le consulter avec des logiciels adaptés (synthèse vocale, agrandissement...) 81 critères respectés.
Niveau Argent = Niveau AA de WAI	Site accessible de façon très satisfaisante. Des règles sont respectées pour permettre la consultation. Niveau Bronze + 20 critères respectés
Niveau Or = Niveau AAA de WAI	Toutes les recommandations sont respectées. Niveau Argent + 30 critères respectifs.

APPLICATIONS PRATIQUES

Pour qu'un contenu soit réellement accessible, on peut considérer de manière synthétique :

- que le contenu et sa mise en forme doivent être clairement disjoints ;
- que tout composant non-textuel doit disposer d'une alternative textuelle.

Si les règles à respecter sont donc établies, les moyens techniques pour les appliquer ne sont pas toujours familiers. Il convient de se pencher sur la norme pour en tirer les préceptes et règles qui s'appliquent au contenu que l'on doit produire.

Pour en savoir plus

On lira avec profit l'article de Dominique Stutzmann, « La prise en compte du handicap dans la bibliothèque numérique Gallica » in Andissac Marie-Noëlle et Poitevin Marie-José (dir.). *Handicap et bibliothèque*. Paris, Association des bibliothécaires français, 2^e édition revue et augmentée, 2009 (Médiathèmes), p. 110-117.

ÉVALUER LA BIBLIOTHÈQUE NUMÉRIQUE

Les outils statistiques permettent de connaître l'audience d'un site, ainsi que la provenance et le comportement des visiteurs sur celui-ci. Si les données disponibles sont en général assez larges, il est important d'avoir une vision claire de ses besoins d'analyse avant de se lancer dans la sélection d'un outil en particulier.

Le choix d'une plateforme statistique doit donc avant tout être fonction des données qu'il est important de suivre sur un site.

Par exemple :

- **audience générale du site** : visiteurs, pages affichées ;
- **veille des contenus les plus populaires** ;
- **origine des visiteurs**, localisation géographique ;
- **parcours de l'utilisateur sur le site** ;
- type de connexion et de matériel informatique utilisé.

TABLEAU **DONNÉES DISPONIBLES
SUR LES PLATEFORMES STATISTIQUES**

Les données liées au contenu	Les données liées au visiteur	Les données liées à la navigation du site	Les données liées à l'origine des visites
Nombre de pages affichées	Nombre de visites	Liens utilisés	Part du trafic par levier (moteur, accès direct, liens extérieurs...) ;
Pages visitées au cours de la période	Nombre de visiteurs uniques	Chemin parcouru sur le site	Sites référents ;
Pages les plus populaires	Origine géographique	Nombre de pages vues par session	Répartition par moteur de recherche
Taux de rebond par page	Langue	Durée de la visite	Mots clés d'accès au site
Pages d'entrée et de sortie	Configuration matérielle et logicielle	Accès à un objectif de trafic	

LES DIFFÉRENTS TYPES DE PLATEFORMES STATISTIQUES

Les différents types d'outils statistiques sont virtuellement capables de donner accès à toutes ces informations. Mais certains présentent ces données de façon plus naturelle ou de manière plus rapide. Les plateformes statistiques se distinguent principalement par l'origine de leurs informations, ou par l'angle pris dans le traitement de celles-ci :

- **les plateformes basées sur l'analyse des *logs*** (journaux d'activité) serveur sont souvent les plus précises quant aux fichiers les plus populaires. Elles permettent de *tracker* l'accès à n'importe quel type de fichier hébergé sur le serveur du site. Elles peinent toutefois, pour des raisons techniques, à identifier les visiteurs et le comportement de ceux-ci. Ces plateformes doivent être hébergées elles-mêmes sur le serveur du site ;
- **les plateformes User-Centric** basent leur analyse sur l'identification des visiteurs. Le plus souvent construits sur la base d'une écriture et d'une lecture de *cookies*, les outils User-Centrics permettent une mesure fiable des chiffres de visiteurs uniques d'un site et peuvent même présenter des lacunes dans l'analyse du comportement de ceux-ci une fois sur le site ;
- **les plateformes Site-Centric** basent leur analyse de l'audience sur le comportement de l'internaute et s'assurent de suivre celui-ci. Souvent implémentés à l'aide de codes JavaScript, les outils Site-Centrics permettent de mesurer de manière fiable les pages vues et l'inventaire d'un site ;
- enfin, **les plateformes Media-Centric** se concentrent principalement sur le point d'entrée d'un visiteur (le média utilisé pour accéder à un site) et se préoccupent peu du comportement du visiteur par la suite. Elles sont principalement utilisées à des fins publicitaires.

LES LOGICIELS DE STATISTIQUES

Il existe une très grande variété de plateformes statistiques, qui répondent souvent à des besoins précis. Parmi les logiciels les plus utilisés en France, on peut citer quelques exemples :

- Webalizer < www.mrunix.net/webalizer/ > est une extension gratuite du serveur Web Apache permettant d'analyser rapidement et graphiquement les journaux d'activité du serveur ;
- AWStats < awstats.sourceforge.net > disponible gratuitement sous licence GNU, est un outil fiable, basé sur l'analyse des journaux d'activité serveur. Il fournit des statistiques détaillées, mais fonctionne sans cookie et ne permet donc pas de suivre l'internaute précisément ;
- Google Analytics < www.google.com/analytics >, est une plateforme de statistiques site-centrique disponible gratuitement sur le Web depuis 2004. Fournissant des données très complètes et hautement personnalisables, elle appartient toutefois à Google qui se réserve sous certaines conditions de droit d'utiliser les données présentées ;
- Les outils AT Internet et la plateforme gratuite Xiti® ([En ligne] < www.atinternet.com/Produits/Solutions-XiTi.aspx >), sont parmi les plus utilisés en France et permettent d'avoir une vue d'ensemble des visites d'un site, mais sont limités dans l'analyse du comportement Internaute ;
- Des plateformes professionnelles payantes, telles eStat ([En ligne] < www.estat.com >) (plateforme de l'Institut français Mediamétrie), Omniture® ([En ligne] < www.omniture.com/fr > (propriété d'Adobe®), Coremetrics ([En ligne] < www.coremetrics.fr >) ou Nedstat ([En ligne] < <http://www.nedstat.fr/> >)) fournissent des options d'analyse très avancées mais souvent pour un prix de licence assez élevé.

Même si la plupart des hébergeurs proposent la mise à disposition de statistiques ainsi que des outils d'analyse, il convient toutefois de ne pas oublier cet outil central dans son cahier des charges. Il est éventuellement possible de faire appel à plusieurs outils dont les données se complètent, à condition de garder en tête que les méthodes appliquées diffèrent d'une plateforme à une autre. Pour un même site, sur une même période, deux outils de statistiques ne fourniront jamais des données identiques !

Cette analyse des statistiques doit être complétée par des études d'usages menées à partir de questionnaires sur place ou en ligne, d'entretiens individuels et de focus groups (ou groupe de discussion). On consultera avec profit les études d'usages menées sur *Gallica* et *Europeana*. [En ligne]
 < http://www.bnf.fr/fr/la_bnf/pub_dist/s.enquetes_gallica.html >.

RÉFÉRENCER

Le terme « référencement naturel » regroupe l'ensemble des techniques visant à améliorer la visibilité d'un site dans les moteurs de recherche. Suivant les profils de site, on estime que jusqu'à 80 % des visites peut provenir de ces moteurs. Il convient donc de ne pas considérer le référencement naturel (ou SEO) comme une simple opération technique, mais comme une stratégie d'acquisition d'audience globale, qui couvre les étapes allant de la définition même de cette audience à la satisfaction de celle-ci par le biais des contenus disponibles sur un site.

Les règles du référencement naturel évoluent en permanence, avec la mise à disposition de nouveaux outils et au fil de l'évolution des algorithmes des moteurs. Toutefois, on peut préciser quelques principes permanents qui doivent rester au centre de n'importe quelle stratégie de référencement naturel. La visibilité d'un site dépend de trois éléments majeurs :

- Son accessibilité technique vis-à-vis des robots de recherche ;
- Son optimisation sémantique ;
- Sa visibilité extérieure.

LE CHOIX DES MOTS-CLÉS CIBLÉS

Avant même d'aborder l'optimisation concrète d'un site, il est important de définir le ciblage de celui-ci. En référencement naturel, ce ciblage se traduit par la définition de mots clés, des requêtes sur lesquelles le site doit se positionner dans les moteurs de recherche. Le choix des mots clés peut se faire en fonction de différents paramètres :

- L'importance de ce mot-clé dans la stratégie de communication, il apparaît logique de cibler le mot-clé Livres quand on référence le site d'une bibliothèque ;
- L'audience potentielle sur les moteurs de recherche, inutile de cibler un mot que personne ne cherche ;
- La concurrence présente sur ce mot-clé, il est très difficile de se positionner sur certains mots très génériques où sont présents de nombreux sites ;
- L'affinité du mot-clé avec le contenu proposé, il est naturel de cibler les sujets qu'on aborde sur ses pages.

Si ces critères découlent avant tout du bon sens, plusieurs outils sont disponibles pour déterminer chacun d'eux :

- **Google Adwords** : Keyword Tool accessible gratuitement, permet de connaître le volume de requêtes généré par un mot-clé sur le Web ou sur un marché local. [En ligne] < <https://adwords.google.com/select/KeywordToolExternal> >.
- **Google Trends**, fourni lui aussi gratuitement et librement, permet de connaître l'évolution des recherches concernant un mot clé dans le temps [En ligne] < www.google.com/trends >.
- **L'interface même de Google**, lors de la saisie d'un mot-clé, permet de connaître le nombre de résultats disponibles pour ce mot et donc sa concurrence potentielle.

Il est également important de garder en tête que des mots clés très précis peuvent être très pertinents à cibler. Selon la théorie de la *Long Tail* : une somme de mots ciblés peut générer un trafic plus important que les seuls mots clés « majeurs ».

L'ACCESSIBILITÉ TECHNIQUE

Pour qu'un site soit visible sur Internet, son architecture doit être compatible avec les attentes des moteurs de recherche :

- le site est en accès libre : il n'est pas protégé par un mot de passe et ne demande pas d'identification pour afficher son contenu ;
- le site ne doit pas être orphelin : il doit être accessible, de manière naturelle, depuis d'autres sites sur Internet ;
- toutes les pages du site sont accessibles via des liens HTML classiques.

Enfin, le site ne doit pas utiliser pas de technologie « bloquante » pour le référencement. La navigation du site et ses contenus textuels doivent être développés en HTML pur, ou dans un langage dynamique générant du HTML afin d'être lisibles par les moteurs. Parmi ces points bloquants, on retiendra particulièrement les éléments suivants :

- **Les URL dynamiques** : les pages Web développées à l'aide d'une base de données et de langages dynamiques tels PHP, ASP... ont besoin de paramètres qui servent à connaître les données à extraire de la base de données et à utiliser lors de la construction de la page en question. Ces paramètres sont fournis à la page par l'intermédiaire de l'URL, à la suite de caractères spéciaux (par ex. : < <http://www.monsite.com/article.php?id=5> >). Ces paramètres peuvent également être présentés sous la forme d'une suite de chiffres. Dans une URL, la séparation entre le cœur de l'URL (la page appelée) et les paramètres d'appel de la page se fait par le caractère « ? ».

Même si les moteurs supportent de mieux en mieux les URL de pages dynamiques, ce type d'adresse utilisant des caractères spéciaux est à éviter le plus possible. En effet, certains moteurs ne voient pas les différents attributs présents à la suite des caractères spéciaux et sont donc incapables de trouver le contenu issu de la base de données qui en découle. Ces moteurs seront donc incapables de référencer correctement ces sites. Une technique d'URL rewriting doit alors impérativement être mise en place.

- **Les scripts de redirection** : les moteurs de recherche référencent assez mal les sites utilisant une redirection pour leur page d'accueil. Ainsi, si à

la saisie de l'adresse < <http://www.monsite.com> >, le moteur se trouve redirigé automatiquement vers la page < <http://www.monsite.com/site/default.asp> >, celui pourra référencer ces deux pages comme étant différentes. Cette technique de référencement se traduit principalement par un mauvais classement de la page d'accueil physique du site (/site/default.asp), celle-ci ne faisant en général l'objet d'aucun lien depuis l'extérieur. Il convient donc, d'une manière générale, que la page d'accueil d'un site soit physiquement la page Index hébergée à la racine du nom de domaine du site.

- **La technologie Ajax** : certaines technologies utilisées sur Internet permettent de concevoir des interfaces utilisateurs plus intuitives et dynamiques. Parmi ces technologies, on compte notamment la norme Ajax (*Asynchronous JavaScript And XML*), technologie utilisant les langages JavaScript et XML et permettant d'afficher un nouveau contenu dans une page sans avoir à recharger entièrement celle-ci.

Cette technologie est à éviter. En effet, les moteurs de recherche actuels ne peuvent indexer que des pages distinctes. Tout contenu destiné à être trouvé par l'intermédiaire des moteurs de recherche se doit donc d'être présent sur les pages du site dès le chargement de celles-ci.

- **Le JavaScript d'une manière générale** : il s'agit d'un langage de script qui permet d'ajouter des éléments d'interactivité à une page Web, par exemple l'ouverture de nouvelles fenêtres de navigateur. Il se présente sous la forme de script inclus à l'intérieur du code HTML des pages.

Les robots des moteurs de recherche ne voient pas les programmes JavaScript et sont donc incapables de suivre les liens qui pourraient y être inclus. Dans la mesure du possible, on évitera donc d'utiliser du JavaScript pour la navigation d'un site ou pour la mise en place de liens sur les pages de celui-ci.

- **La technologie Adobe® Flash®** : elle permet la création d'animations Web dynamiques et légères. Ces animations se traduisent, techniquement, par différents fichiers .swf insérés à l'intérieur des différentes pages des sites. Ces animations, compilées lors de leur création afin de leur garantir une légèreté optimale, ne sont pas consultables par les robots des moteurs de recherche. Les textes et liens qui y sont inclus sont donc invisibles pour ceux-ci.

Il convient donc d'éviter au maximum l'utilisation des animations Adobe® Flash® pour tout contenu devant être indexé dans les moteurs de recherche. De la même façon, on évitera au maximum l'utilisation d'animations Adobe® Flash® pour la navigation des sites, afin de s'assurer que les liens de celle-ci sont bien suivis par les robots des moteurs de recherche.

Cela ne veut pas dire qu'un site doit faire l'impasse sur ces technologies : Ajax, Adobe® Flash® ou JavaScript. Elles peuvent s'avérer pertinentes dans le cadre de la présentation dynamique de certains contenus. Il faut toutefois éviter d'utiliser celles-ci de manière systématique, et surtout dans la présentation de contenus qui ont vocation à être trouvés via les moteurs de recherche.

Le protocole sitemaps

Le protocole Sitemaps, initié par Google en 2005, propose aux webmasters de construire un fichier XML listant tout ou partie des pages de leur site Web dans le but d'informer Google de leur existence.

Une fois inscrit au service Google Webmaster Tools ([En ligne] < www.google.com/webmasters/tools/), il est possible d'indiquer à Google l'emplacement de ce fichier et permettre une meilleure indexation de la totalité des pages disponibles sur un site.

Depuis novembre 2006, quatre moteurs (Yahoo!, Bing, Ask et Exalead) ont annoncé leur ralliement au protocole Sitemaps et proposent désormais des interfaces permettant une soumission facilitée de ces listes de fichiers.

L'OPTIMISATION SÉMANTIQUE

Les robots des moteurs de recherche travaillent avec une vision la plus « humaine » possible des contenus. Ils cherchent avant tout à déterminer si un texte est compatible avec la lecture et les attentes d'un internaute humain. Ils traquent donc les rédactions « abusives » (répétitions trop fréquentes d'un mot-clé) et cherchent avant tout la cohérence des contenus proposés. Les recommandations d'écriture vont donc vers une plus grande cohérence des mots-clés utilisés dans les pages : contenus visibles pour les internautes (textes réels, titres...) et textes invisibles.

Le texte visible

- Le texte réel de la page, tel qu'il est visible par les internautes. Ce texte doit s'architecturer autour du sujet principal de la page et utiliser si possible les mots clés identifiés comme prioritaires sur cette page ;
- Les en-têtes HTML (< H1 >/< H2 >) servent à définir le texte le plus important d'une page.
- Le titre HTML de la page (présent dans la barre de titre du navigateur) est utilisé en tant que lien par les moteurs de recherche. Il bénéficie d'un poids important dans le calcul de la pertinence d'une page.

Le texte invisible

- Les balises META Description et META Keywords. Très peu utilisées aujourd'hui par les moteurs de recherche, elles restent toutefois pertinentes pour les moteurs mineurs. Google utilise parfois le contenu de la balise Description en tant que description d'un résultat ;
- Les légendes d'images (attribut ALT). Cet attribut peut permettre de renforcer la présence de mots clés au sein des pages. Il permet aussi de qualifier des contenus pour les moteurs de recherche d'images ;
- Les légendes des liens (attributs TITLE). Certains moteurs utilisent ce contenu pour qualifier plus avant les liens sortants provenant de certaines pages.

LA VISIBILITÉ EXTÉRIEURE

Pour être considéré comme pertinent, un site Web doit bénéficier d'importants liens depuis l'extérieur, si possible depuis des sites jugés eux-mêmes pertinents et populaires. Une stratégie de visibilité extérieure cohérente devra donc chercher la présence sur :

- **de nombreux sites extérieurs.** La popularité d'un site est proportionnelle à sa visibilité sur le Web et au nombre de liens menant à ses pages ;
- **des sites sémantiquement proches.** Les moteurs de recherche accordent plus de valeur aux liens provenant de sites de secteurs proches qu'aux liens plus généralistes ;
- **des sites « de confiance ».** (administration, universités, grandes entreprises...). Les moteurs accordent un bonus de pertinence aux sites institutionnels et prennent cette crédibilité en compte pour calculer la pertinence des sites liés.

Pour atteindre ces objectifs de visibilité, différentes stratégies d'acquisition de liens peuvent être envisagées, comme la mise en réseau de sites géographiquement ou thématiquement proches, ou l'acquisition de liens « profonds » (vers les pages internes du site) depuis des sites spécialisés sur un sujet.

Les moteurs accordent également beaucoup d'importance aux textes des liens utilisés pour lier un site. Il convient donc, dès que possible, d'obtenir des liens utilisant des mots clés forts, et identifiés comme cible, depuis d'éventuels partenaires extérieurs.

LE RÉFÉRENCIEMENT ÉTHIQUE

Des astuces, utilisées par certains sites, permettent de cacher des mots-clés à certains endroits d'une page, d'utiliser des pages satellites, etc. Ces pratiques sont interdites par les moteurs de recherche et peuvent entraîner une désindexation (*blacklistage*) des moteurs, comme ce fut le cas pour BMW en février 2006⁶⁵ ou Hermès en septembre 2006⁶⁶.

Les règles à respecter, édictées par Google, sont simples et reviennent principalement à concevoir son site en fonction des besoins des internautes, et non pas de ceux des moteurs :

Recommandations

Évitez les « astuces » destinées à améliorer le classement de votre site par les moteurs de recherche. Pour savoir si votre site Web respecte nos consignes, posez-vous simplement la question suivante : « Suis-je en mesure d'expliquer à un site Web concurrent quelles sont les solutions que j'ai adoptées pour améliorer mon site ? » Vous pouvez également vous poser les questions suivantes : « Ces solutions sont-elles d'une aide quelconque pour les internautes ? » ou « Aurions-nous fait appel à ces techniques si les moteurs de recherche n'existaient pas ? »⁶⁷

65. [En ligne] < <http://actu.abondance.com/2006-06/google-bmw.php> > (consulté le 22 janvier 2010).

66. [En ligne] < <http://www.journaldunet.com/solutions/0611/061109-hermes-disparition/01.shtml> > (consulté le 22 janvier 2010).

67. [En ligne] < <http://www.google.com/support/webmasters/bin/answer.py?hl=fr&answer=35769#3> > (consulté le 22 janvier 2010).

METTRE EN PLACE UN SUIVI DE RÉFÉRENCEMENT

La question du suivi du référencement est ensuite une question d'organisation et de budget. Le référencement naturel est une expertise qui demande des connaissances de plus en plus diverses et un temps de mise en place assez élevé. Le recours à une agence spécialisée est donc souvent recommandé pour l'exécution d'une véritable campagne de référencement naturel suivie. En France, les contrats de référencement naturel oscillent entre 3 000 et 18 000 euros à l'année suivant le degré d'implication de l'entreprise approchée et la complexité du référencement naturel du site.

Il reste toutefois important de bien saisir ce que comprennent de tels tarifs et quelle est la nature précise de prestation achetée. Certaines agences n'hésitent en effet pas à définir comme une offre de référencement la création de pages ciblées, totalement détachées de votre site. Avant d'envisager l'appel à une agence externe, qui ne concernera que les projets suffisamment avancés, un établissement peut, dans un premier temps, faire effectuer un diagnostic « référencement » par une agence spécialisée. Cette analyse correspondra souvent au coût d'une ou deux journées de travail.

Enfin, il paraît nécessaire à présent de faire figurer cette question du référencement dans les cahiers des charges de mise en ligne sous forme de fonctionnalités et résultats obligatoires faisant l'objet ou non d'un lot séparé. De plus, lors de la création d'un site, il est de plus en plus important de s'assurer du respect de quelques règles élémentaires d'accessibilité (points bloquants techniques, rédaction des titres de page...) dès la phase de réalisation. Il est en effet plus simple, et moins onéreux, de contrôler ces points pendant le développement que d'effectuer des modifications techniques lourdes une fois le site en ligne.

presses ensib

PARTIE VII**CONSERVER
LES DOCUMENTS
NUMÉRISÉS**

CONSERVER LES DOCUMENTS NUMÉRISÉS

*par Patricia
Le Galèze
et Isabelle
Westeel*

CONTEXTE ET ENJEUX

La pérennité de l'information numérique est une question complexe. Il n'existe pas de solutions prêtes à l'emploi dans ce domaine. En effet les technologies de préservation sont à renouveler régulièrement, le matériel informatique étant en évolution perpétuelle. L'accroissement exponentiel de l'information numérique est une difficulté supplémentaire.

Dans un souci de préservation des originaux comme de maîtrise des coûts, la préservation des documents numérisés cherche à éviter de devoir numériser à nouveau les originaux. Il s'agit de pérenniser l'investissement réalisé. La préservation numérique doit donc être prise en compte dès le début d'un projet de numérisation. Ceci afin de protéger les données numériques, mais aussi de les garder accessibles le plus longtemps possible.

La préservation numérique se heurte cependant à trois risques majeurs :

- l'obsolescence des applications informatiques et des formats de fichiers, qui incitent à garder les outils ou à utiliser la migration des données ou l'émulation des systèmes ;
- la fragilité et la durée de vie relativement faible des supports ;
- la difficulté de retrouver des fichiers informatiques s'ils n'ont pas été bien décrits à l'origine et bien suivis. Pour cela, les services informatiques ont besoin d'un maximum d'informations, et donc de métadonnées. Même s'il s'agit d'informations coûteuses, elles se révèlent indispensables à la pérennité.

L'enjeu est double :

- la conservation du support ;
- l'accès aux données en fonction des formats, de la structuration du document et de l'encodage de l'information numérique.

Les problèmes rencontrés par les établissements dans leurs stratégies de préservation sont essentiellement les capacités insuffisantes de stockage, la perte de données, les données devenues illisibles et enfin l'expertise insuffisante.

Il convient donc de mettre en place une politique de conservation numérique à son niveau, comprenant notamment une politique claire de nommage des fichiers, de formats, de stockage et de migration :

- préparer des données et des métadonnées ;
- rassembler les éléments nécessaires à la construction de l'arborescence des fichiers ;
- fournir les informations techniques permettant de surveiller et de faire migrer les données ;
- s'assurer que les informations sur les droits d'accès sont présentes ;
- mesurer le facteur de risque lié au degré d'incomplétude des métadonnées ;
- respecter les formats de données garantissant la pérennité et permettant de couvrir la majorité des besoins.

Il sera également nécessaire d'associer autant que possible les services informatiques, y compris des informaticiens spécialisés et des prestataires qui se verront attribuer les rôles et les obligations suivants :

- la gestion de grands ensembles de fichiers stockés, dupliqués sur différents types de supports ;

- les technologies réseau à haut débit permettant de communiquer avec les usagers du service ;
- les technologies de stockage à grande capacité, robots de stockage, supports de stockage avec leurs caractéristiques et leur fiabilité ;
- les moyens de surveillance de l'état des supports ;
- la capacité à maintenir en fonctionnement opérationnel un système ouvert très largement et à le faire évoluer en fonction des progrès de la technologie et des montées en charge.

RECOMMANDATIONS TECHNIQUES ET CONSEILS POUR UNE NUMÉRISATION DURABLE

En matière de conservation numérique, le Référentiel général d'interopérabilité (RGI)⁶⁸ définit des recommandations réparties entre différents volets :

- volet organisationnel : politique d'archivage sécurisé dans le secteur public ;
- volet sémantique : schéma d'échange de données pour l'archivage (visant à faciliter les échanges entre services producteurs d'archives et services d'archives) ;
- volet technique : préconisations en matière de formats pour la conservation ainsi qu'en matière de supports.

Ces recommandations constituent les objectifs à atteindre pour favoriser l'interopérabilité. Le RGI n'émet que des préconisations en matière de normes, de standards et de bonnes pratiques. Aucune norme ni standard n'est interdite ou déconseillée.

68. RGI Version 1.0, 12 juin 2009, version en vigueur. [En ligne]

< <http://www.referencess.modernisation.gouv.fr/rgi-interoperabilite> > (consulté le 22 janvier 2010) et
 < http://referencess.modernisation.gouv.fr/sites/default/files/RGI_Version1_0.pdf > (consulté le 22 janvier 2010).

sites de référence

- Ministère de la Culture et de la Communication. *Conservation à long terme des documents numérisés* : guide technique publié en 2008 et réalisé en coordination avec le comité de pilotage numérisation : normes et standards, politique de conservation, production des documents numérisés, supports et infrastructure de stockage, questions juridiques, projets et expériences. [En ligne] < <http://www.culture.gouv.fr/culture/mrt/numerisation/fr/technique/documents/conservation.pdf> > (consulté le 22 janvier 2010).
- Site de l'association Aristote. *Pérennisation des Informations Numériques (PIN)*. [En ligne] < <http://www.aristote.asso.fr/PIN/> > (consulté le 22 janvier 2010). Créé en 2000, le groupe de travail et d'échanges français PIN (Pérennisation de l'Information Numérique) est un lieu d'échange de savoirs, de savoir-faire et d'actions dans le domaine de la pérennisation des informations représentées sous forme numérique.
- Site de la Direction générale de modernisation de l'État (DGME) : la DGME regroupe, depuis le 1^{er} janvier 2006, l'ensemble des services œuvrant pour la réforme et la modernisation de l'État, dont l'ancienne agence pour le développement de l'administration électronique (ADAE), devenue au sein de la DGME, le service pour le développement de l'administration électronique (SDAE). C'est avec ce service que la direction des Archives de France met en œuvre l'action 103 du programme ADELE, sur l'archivage électronique. [En ligne] < <http://www.modernisation.gouv.fr/> > (consulté le 22 janvier 2010).

FORMATS

Il est non seulement essentiel de choisir les bons formats de numérisation, tant pour la qualité de la numérisation que pour la préservation à long terme des documents numériques, mais il est aussi important de distinguer le format de conservation du format de diffusion pour des raisons notamment d'accès aux documents en terme de temps d'affichage, de téléchargement et de distinction entre la consultation et la préservation.

Les formats sont des systèmes de codage de l'information, qui permettent de passer d'informations intelligibles à des données binaires, et vice-versa. Ces opérations de codage et de décodage sont généralement assurées par des

logiciels (par ex. Adobe® Photoshop® ...). Certains formats sont secrets (cf. Microsoft® Word), d'autres sont rares. Si les logiciels évoluent ou disparaissent, le sens des données binaires stockées risque de devenir inaccessible. C'est le cas lorsque le processus de codage est secret et que seul le logiciel d'origine est capable de le réaliser. Il est donc indispensable d'utiliser des formats dont les spécifications sont publiques, et de préférence des formats reconnus comme normes internationales. Il est également conseillé d'utiliser des formats courants. Des migrations seront nécessaires si les documents sont à conserver.

Le format doit donc être :

- un format ouvert, c'est-à-dire un format dont la documentation est librement accessible (un format peut être propriétaire mais il devra être suffisamment documenté pour en permettre une large utilisation) ;
- un format largement utilisé ;
- un format normalisé si possible ;
- un format indépendant (notamment des autres formats et des plateformes).

TABLEAU**Liste des formats recommandés en terme de conservation par le RGI (version 1.0)**

p. 163

type de document	format	commentaires
Documents	XML	Format dont les spécifications sont normalisées par l'ISO
	PDF/A	ISO 19005-1 Archivage des documents bureautiques statiques non révisables
	PDF 1.7	Conservation des documents bureautiques dynamiques
Images	GIF	
	PNG	ISO 15948
	JPEG	ISO 10918
	TIFF	
	DNG	
Son	MP3	ISO 11172-3 Conservation de séquences sonores de qualité ordinaire
	WAV	Codage PCM Sans compression - Conservation de séquences sonores de haute qualité sonore
Vidéo	MPEG-2	ISO 13818 Conservation de séquences vidéo en basse définition
	MPEG-4	ISO 14496 Conservation de séquences vidéo en haute définition
Objet graphique en 2D	SVG 1.1	Recommandation W3C
Objet et univers 3D	X3D	Normalisé par l'ISO
Bases de données	XML CSV	

Utilisé par la Bibliothèque du Congrès et par la Bibliothèque royale des Pays-Bas, le format JPEG 2000 peut être considéré comme un bon compromis pour le master de conservation. La BnF étudie la possibilité d'utiliser ce format pour l'archivage même si ce format qui n'est pas supporté par tous les visualiseurs, connaît des difficultés dans le domaine de la diffusion.

sites de référence

- De même que le Centre informatique de l'enseignement supérieur (CINES) publie sur son site une liste de formats archivables sur sa plateforme PAC⁶⁹, la BnF a réalisé un mémento sur les formats d'images. [En ligne]
< http://www.bnf.fr/fr/professionnels/num_metadonnees/s.num_metadonnees_formats.html?first_Art=non > (consulté le 22 février 2010).
- La base PRONOM des Archives nationales du Royaume-Uni permet de mettre à disposition des informations sur les formats de fichiers, les produits logiciels et autres composants techniques nécessaires pour assurer la préservation à long terme des documents électroniques. [En ligne]
< <http://www.nationalarchives.gov.uk/PRONOM/> > (consulté le 22 janvier 2010).
- La Bibliothèque du Congrès propose des pages consacrées à l'étude et à l'évaluation des formats. [En ligne]
< <http://www.digitalpreservation.gov/formats/> > (consulté le 22 janvier 2010).

NORMES ET STANDARDS

Des normes, des standards et des référentiels ont vu le jour et commencent à trouver des applications concrètes au niveau national.

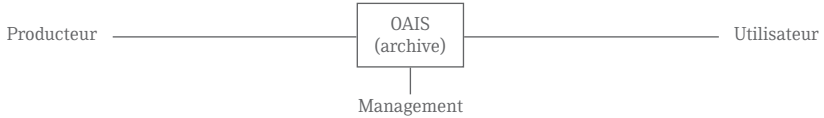
normes généralistes

OAIS (Open Archival Information System/système ouvert d'archivage d'information) : norme ISO 14721 : 2003⁷⁰ (Système de transfert des informations et données spatiales – Système ouvert d'archivage de l'information – Modèle de référence)

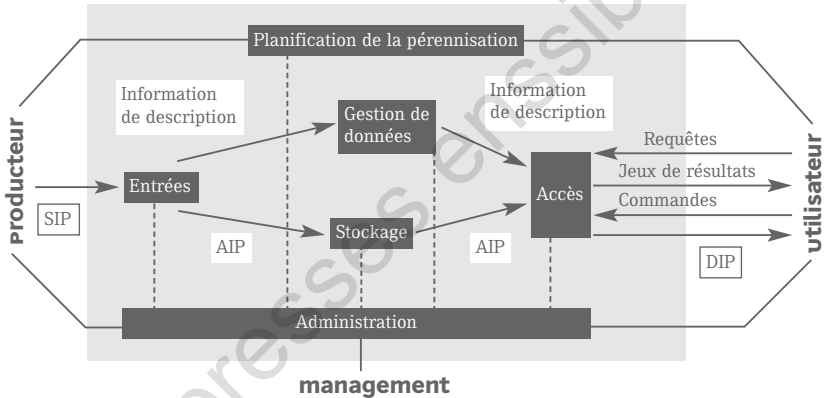
69. Cines. [En ligne] < <http://www.cines.fr/spip.php?article685> > (consulté le 22 janvier 2010).

70. [En ligne] < http://www.iso.org/iso/fr/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683 > (consulté le 22 janvier 2010). Norme en ligne < http://www.aristote.asso.fr/PIN/documents/OAIS_650_Obl.pdf > (consulté le 22 janvier 2010).

SCHEMA **Modèle d'environnement d'un OAIS** **(traduction française de l'OAIS p. 2-2)**



SCHEMA **Entités fonctionnelles OAIS** **(traduction française de l'OAIS p. 4-1)**



Norme généraliste mise au point par les principaux centres d'études spatiales du monde dont le Centre national d'études spatiales (CNES) en France, OAIS est un modèle de référence⁷¹ pour la définition des concepts indispensables à la compréhension de l'archivage numérique : l'organisation, l'architecture et le mode de fonctionnement des archives existantes et futures. Ce système a été adopté dès 1999 par tous les programmes de bibliothèques s'intéressant à la préservation et à l'accès à long terme et par les entrepôts institutionnels.

71. Traduction française, en cours de normalisation, consultable sur le site PIN. [En ligne] < http://www.aris-tote.asso.fr/PIN/documents/norme_oais_version_francaise.pdf > (consulté le 22 janvier 2010).

Ce modèle aborde non seulement toute une série de fonctions liées à la conservation d'informations archivées comprenant le versement des données, le stockage, la gestion des données, l'accès et la diffusion, mais aussi la migration d'informations numériques vers des supports et des formats nouveaux, les modèles de données utilisés pour représenter l'information, le rôle du logiciel dans la conservation de l'information, ainsi que l'échange d'informations numériques entre archives. Il identifie les interfaces internes et externes des fonctions d'archivage ainsi qu'un certain nombre de services de haut niveau relatifs à ces interfaces. Il détermine les responsabilités minimales auxquelles une archive doit souscrire pour être qualifiée d'OAIS.

Il définit :

- les objets d'information, les métadonnées nécessaires à leur préservation et l'organisation à mettre en place pour leur archivage, leur conservation et leur communication ;
- un vocabulaire et un ensemble de concepts permettant d'appréhender, de façon globale et complète, la question de l'archivage à long terme des données sous forme numérique quels que soient les secteurs professionnels concernés ;
- deux modèles complémentaires : un modèle d'information et un modèle fonctionnel détaillé.

Il propose également une classification des types de migration et des différents modes de coopération possibles entre archives. La norme ne préjuge pas d'une implémentation particulière et constitue un cadre de développement pour des normes complémentaires.

La norme NF Z 42-013 (mars 2009) : 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. La première version de la norme datait de 2001. Une nouvelle version, élaborée dans le cadre de la commission de normalisation 171

(« Archivage - Cycle de vie du document et sécurité des données »), a été adoptée le 4 février 2009 par le conseil d'administration de l'AFNOR. Cette norme fournit un ensemble de spécifications concernant les mesures à mettre en œuvre pour l'enregistrement, le stockage et la restitution de documents électroniques afin d'assurer la conservation et l'intégrité de ceux-ci. Cette nouvelle version reconnaît les supports réinscriptibles grâce aux garanties et protections offertes par les moyens technologiques reconnus tels la signature électronique et l'horodatage. Elle étend le périmètre d'application aux documents sonores et audiovisuels. Elle définit l'organisation des audits des systèmes d'archivage, ainsi que l'intervention des tiers dans le processus d'archivage (tiers archiveurs et autres prestataires de services).

Le **Records Management : norme ISO 15489** (*Information and documentation – Records management* octobre 2001, en français en 2002) qui a pour objectif la gestion des archives courantes et intermédiaires. Cette norme a pour « finalité de permettre à l'organisme de disposer à tout instant du document dont il a besoin pour conduire ses activités, répondre aux exigences légales et réglementaires, et se protéger en cas de contentieux⁷² ». La norme définit très précisément les processus à mettre en place depuis la création du document. Elle permet de prendre des décisions à propos des étapes de création, de capture, de contrôle, de conservation et d'archivage des documents⁷³.

Les **spécifications Moreq 2** : modèle européen pour la mise en œuvre du Records Management pour les documents électroniques et précisant les exigences de l'archivage électronique à des fins de preuve, disponible sur < <http://www.archivesdefrance.culture.gouv.fr/gerer/records-management-et-collecte/> >.

72. Groupe métier AAF – ADBS « Records management », Analyse de la norme ISO 15489 au regard des pratiques archivistiques françaises. Version 2 – 8 avril 2005, p. 4. [En ligne] < <http://www.adbs.fr/groupe-de-travail-aaf-adbs-records-management-29392.htm?RH=ACCUEIL> > (consulté le 4 mars 2010).

73. Voir le site des Archives de France. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/records-management-et-collecte/> > (consulté le 22 janvier 2010).

standards et normes relatifs à l'interface Producteur-Archive

PAIMAS (*Producer-Archive Interface Methodology Abstract Standard*) : norme ISO 20652. Cette norme concerne la méthodologie de conduite de l'interface Producteur-Archive et permet de disposer d'un guide standard.

PAIS (*Producer-Archive Interface Specification*) – XFDU (*XML Formatted Data Units*) spécifie des mécanismes formels (automatisation) permettant la définition, le transfert et la validation des documents à archiver et permet de mettre en œuvre PAIMAS.

Le protocole standard d'échange DGME/DAF⁷⁴ spécifie les processus et la mise en œuvre des transferts depuis le producteur vers l'utilisateur.

certification/audit

TRAC (*Trustworthy Repositories Audit and Certification*) définit des exigences de certification des archives en vue de la normalisation de ces exigences par l'ISO. Il fournit une liste de critères pour l'étude de la fiabilité des capacités d'une Archive à conserver des documents numériques de façon pérenne.

DRAMBORA (*Digital Repository Audit method Based on Risk Assessment*) est une méthodologie d'audit d'archives numériques basée sur l'évaluation des risques liés à l'archivage de documents numériques. Elle fournit des outils pour faciliter l'audit interne d'une archive en vue d'identifier ses forces et faiblesses.

Pour en savoir plus

Site de l'association Aristote. Pérennisation des Informations

Numériques (PIN). [En ligne]

< <http://www.aristote.asso.fr/PIN/> > (consulté le 22 janvier 2010).

74. Voir le site des Archives de France. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/standard/> > (consulté le 22 janvier 2010).

Métadonnées

Les métadonnées techniques permettent d'identifier et de caractériser les formats de représentation de l'information ou formats de données. (Pour une présentation complète des métadonnées, se reporter à la Partie V). Le dictionnaire PREMIS basé sur le format XML est, dans ce domaine, un outil très précis. Il vise à établir, à partir du modèle OAIS, une liste des métadonnées gérées dans le processus de conservation numérique⁷⁵. Il permet la spécification d'une stratégie pour le codage, le stockage et la gestion de métadonnées de préservation pour les systèmes d'archivage.

RECOMMANDATIONS PRATIQUES POUR LE STOCKAGE

CHOIX DU SUPPORT DE STOCKAGE

Les critères de choix du support de stockage se font en fonction :

- du volume à stocker ;
- du coût ;
- de la durée de conservation ;
- des caractères critiques des données à conserver, de la confidentialité de l'information et de la sécurité ;
- de l'accessibilité ;
- de la fréquence d'utilisation de l'information ;
- de la fiabilité par rapport aux mécanismes d'enregistrement, de la stabilité du support, de sa robustesse et de la large diffusion de la technologie ;
- de l'évolutivité (possibilités de migrations).

75. PREMIS : *Preservation metadata and the OAIS information model*. Groupe de travail OCLC/RLG sur les métadonnées de conservation, juin 2002. PREMIS Data Dictionary, mai 2005. Maintenance et suivi des travaux assurés par la Bibliothèque du Congrès. [En ligne] < <http://www.loc.gov/standards/premis> > (consulté le 22 janvier 2010).

Les supports sont variés (supports optiques, supports magnétiques) mais il n'y a pas aujourd'hui de support pérenne. Même si on constate une baisse continue du prix de ces supports, on se heurte à l'augmentation régulière des volumes.

Les supports sont fragiles et se détériorent. Des mauvaises conditions de stockage ou d'utilisation peuvent aboutir à une destruction rapide des supports, que ce soit en raison :

- d'un mauvais environnement (température, humidité, lumière, champ magnétique) ;
- d'erreurs de manipulation (rayures, produits toxiques, effacement accidentel).

Les supports ont aussi tendance à se dégrader naturellement par exemple par oxydation. En général, la durée de vie des supports ne dépasse guère 5 ou 10 ans.

Il faut donc :

- choisir des supports robustes, normalisés, produits par plusieurs fabricants,
- procéder à des renouvellements réguliers de supports,
- surveiller les conditions de stockage, l'état des supports ainsi que leur disponibilité sur le marché,
- créer au moins une copie de sauvegarde (inconvenient : on double au minimum les coûts de stockage), de préférence sur un support différent.

TABEAU SUPPORTS DE STOCKAGE

SUPPORTS DE STOCKAGE		
Supports magnétiques	Bandes magnétiques	LTO Support professionnel Qualité contrôlée Support assez utilisé Risque technologique élevé
	Disques magnétiques internes/externes	Disque dur Support professionnel Qualité contrôlée Support assez utilisé Risque technologique élevé
Supports optiques	CD = <i>compact disc</i> : disque compact non inscriptible i.e. pressé ou gravé en usine. CD-R (ou CD-WORM) = <i>compact disc recordable</i> : disque compact inscriptible. CD-RW = <i>compact disc rewritable</i> : disque compact réinscriptible.	Capacité : 700 Mo Support grand public Qualité inégale Risque technologique faible La fabrication doit répondre à la norme 42-011-2 La gravure doit répondre à la norme ISO 9660
	DVD, DVD-R, DVD-RW = <i>digital versatil disc, recordable, rewritable</i> : disque numérique polyvalent, inscriptible, réinscriptible	Capacité : 4,7 Go Support grand public Qualité très inégale Risque technologique faible Double couche ou double face capacité 17 Go

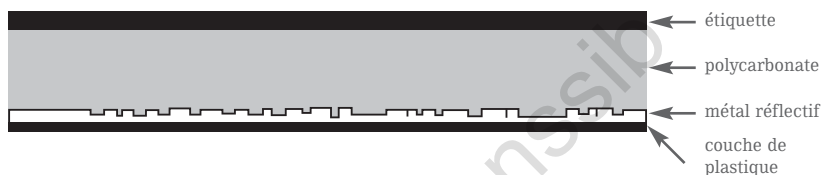
LES SUPPORTS OPTIQUES

Il est important que ce type de supports respecte :

- la norme NF Z 42-013 concernant les conditions et procédures de conservation (tests, copies régulières et transferts de supports) qui doivent être décrites en détail ;

- la norme d'enregistrement ISO 9660 (niveau 2). Cette norme garantit la lisibilité des noms sur tout type de plateforme. Les noms de répertoires et de fichiers (y compris le séparateur et l'extension) sont limités à 31 caractères qui doivent obligatoirement être des lettres capitales, des chiffres et des tirets bas (_). L'arborescence des fichiers est limitée à huit niveaux de répertoires.

SCHEMA *Composition d'un CD ou DVD*



Pour rappel, la fabrication d'un CD-R doit répondre à la norme 42-011-2 ⁷⁶. Un CD ou DVD, quel que soit son type, se compose en général d'une couche métallique réfléchissante (or, argent ou aluminium, suivant la qualité du support) et d'une couche de plastique à base de polycarbonate. Ces deux éléments sont immuables et servent à la fois à rendre le CD lisible, en réfléchissant la lumière du laser de lecture, et à solidifier le support en le rendant plus épais. La composition de la zone de données dépend du type de média utilisé : laque où les données sont pressées pour les CD-Rom, DVD-Rom et CD-Audio, couche de teinture réagissant à un type précis de laser pour les CD-R et DVD-R (inscriptibles). Les supports réinscriptibles (CD et DVD RW) utilisent un alliage à changement de phase, chauffé par un laser, pour stocker les données. Pour y effacer des informations, il suffit alors de redonner à l'alliage en question son état initial sur certaines sections du disque. Ces différences de structures suivant les différentes technologies de support impliquent bien entendu des méthodes de conservation, et des précautions d'usage adaptées.

76. Cette norme a été retirée par l'AFNOR mais constitue encore une référence.

Ces dernières années, plusieurs études ont été menées pour aboutir à des recommandations relatives à la conservation des supports optiques.

- Diffusion en 2005 par la Direction des Archives de France de Recommandations relatives à la gravure, à la conservation et l'évaluation des CD-R⁷⁷. Les principales recommandations sont les suivantes :

- > choisir des modèles dédiés à la conservation longue durée ;
- > remplir le disque à 90 % de sa capacité au maximum ;
- > éviter de graver à trop grande ou à trop basse vitesse ;
- > vérifier de temps en temps l'état des disques ;
- > respecter les conditions de stockage (température entre 16° et 23°, humidité entre 30 % et 50 %, luminosité) ;
- > graver au moins deux exemplaires : un master d'archivage et une copie de consultation ;
- > stocker les exemplaires dans deux endroits séparés et suffisamment distants.

- Constitution en 2005 d'un « Réseau national de recherche sur la conservation des informations enregistrées sur disques optiques numériques » sous la forme d'un Groupement d'Intérêt Scientifique (GIS-DON⁷⁸) donnant de nombreuses informations utiles sur les perspectives existantes dans ce domaine.

- Mise en place d'un observatoire permanent de la qualité des CD-R du marché avec le lancement d'une étude en 2006 sur les CD-R (analyse des couples CD/graveurs) dont les résultats sont disponibles sur le site des Archives de France⁷⁹ qui propose aussi un guide pour la réalisation de la migration de stocks de CD-R.

77. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/stockage> > (consulté le 22 janvier 2010).

78. [En ligne] < http://www.lne.fr/fr/r_et_d/gis-don/conservation-donnees-numeriques-gis-don.asp > (consulté le 22 janvier 2010).

79. [En ligne] < <http://www.archivesdefrance.culture.gouv.fr/gerer/archives-electroniques/stockage> > (consulté le 22 janvier 2010).

Dans le cahier des charges, il conviendra de demander au prestataire de suivre ces différentes recommandations. On pourra distinguer le support de livraison et le support de conservation.

Le prestataire indiquera la marque et la capacité des disques utilisés ainsi que la nature du film métallique et du colorant organique. Il indiquera le nom du fabricant des supports et les numéros de série des disques utilisés. Il donnera également des indications précises sur le graveur et le logiciel de gravure utilisés.

Pour des questions de contrôle et de traçabilité, les pochettes des CD-R pourront comporter les indications suivantes :

- Nom de l'établissement ;
- Nom du CD-R selon les recommandations de la Mission de la recherche et de la technologie du ministère de la Culture et de la Communication⁸⁰ ;
- Titre et cote du document ;
- Date de production ;
- Logiciels utilisés pour la production ;
- Résolution et compression ;
- Nombre et poids total des fichiers ;
- Nom des fichiers extrêmes.

On pourra demander au prestataire la livraison d'un autre jeu d'images sur un support de disque dur externe connectables en USB 2.0. Le prestataire indiquera le type de disque utilisé et précisera la structuration du disque dur (*filesystem*) au moment du formatage : NTFS, FAT32, ext3, ext2...

Cette prestation est surtout demandée pour des questions de commodité de chargement de la base de données.

80. [En ligne] < http://www.culture.gouv.fr/mrt/numerisation/fr/gestion_fonds_images/gstim8.htm > (consulté le 22 janvier 2010).

LES PÉRIPHÉRIQUES DE STOCKAGE

Les types de périphériques de stockage sont différents suivant le support utilisé et suivant la capacité de stockage de la plateforme :

- pour les CD-R et DVD-R : stockage par *jukebox* ;
- pour les bandes magnétiques : lecteur de bandes ou « autoloader » pour les petits volumes, bibliothèques intégrant plusieurs drives pour les gros volumes ;
- pour les disques durs : baies de stockage.

Les types d'architecture de stockage

- DAS (*Direct Attached Storage*) : architecture la plus simple et la plus répandue, en attachement direct (un serveur avec un disque attaché) mais peu évolutive et peu adaptée à la volumétrie importante ;
- NAS (*Network Attached Storage*) : architecture en attachement réseau. Les données sont stockées sur des serveurs de fichiers reliés entre eux par le réseau local :
 - > avantages : simplicité, évolutivité, disponibilité des données, augmentation de la sécurité des données ;
 - > inconvénients : le réseau local partagé empêche la garantie des performances, administration complexe.
- SAN (*Storage Area Network*) : réseau informatique dédié au stockage, permettant la mutualisation des ressources de stockage :
 - > avantages : performances garanties et administration simplifiée ;
 - > inconvénients : coût très élevé, solution réservée aux grandes organisations, nécessite des compétences pointues.

Pour le réseau de stockage, les systèmes de type NAS et SAN ne sont prescrits que pour les plus grandes plateformes (gérant plusieurs dizaines de To).

LE STOCKAGE HIÉRARCHIQUE

Le stockage hiérarchique (HSM = Hierarchical Storage Management) permet d'optimiser l'utilisation d'un espace de stockage pour des volumes importants de données. Le principe consiste à utiliser des supports performants (disques durs) pour les données à haute criticité ou demandant un accès fréquents et des supports moins rapides (disques durs, bandes magnétiques...) pour les données demandées moins souvent ou pouvant être sollicitées par un accès plus lent. Ce fonctionnement est tout à fait transparent pour l'utilisateur. Des procédures automatiques permettent de dynamiser le système en migrant au besoin les fichiers les moins sollicités vers les supports adéquats, tout en en conservant la trace. Répondant à des questions d'économie, de gestion de performances et de bonne administration des capacités de stockage, cette possibilité est utilisée depuis quelques années dans les administrations et les collectivités. Elle peut répondre aux problématiques de stockage mais également d'archivage dans certains cas.

LA SAUVEGARDE DES DONNÉES ET L'ARCHIVAGE

Un système de stockage devra être accompagné d'outils de sauvegarde permettant de sécuriser les données. On pourra équiper les machines de disques en RAID (*Redundant Array of Independant Disks*)⁸¹. Les technologies RAID offrent différents niveaux de sécurisation de données. Il est conseillé d'utiliser des machines équipées de disques durs en RAID 1 ou 5. Le RAID 1 consiste à dupliquer en permanence l'information sur plusieurs disques selon un principe de miroir. La fiabilité du système est excellente mais le coût du stockage est doublé. Le RAID 5 est un système dans lequel les données sont réparties sur plusieurs disques selon le principe de l'agrégation avec parité répartie. Ce système moins coûteux offre une sécurité très élevée. Des rythmes de sauvegarde (quotidiens, hebdomadaires...) sont à mettre en place. De plus, il est parfois conseillé de stocker

81. Le système RAID permet de stocker les données sur plusieurs disques. Ils sont redondants, en totalité ou en partie, pour assurer une sauvegarde en cas de défaillance du système.

régulièrement à l'extérieur des sauvegardes sur bandes.

PRÉCONISATIONS AU NIVEAU SUPPORT ET PÉRIPHÉRIQUE DE STOCKAGE SELON LA VOLUMÉTRIE

- Utilisation de DVD-R et CD-R pour les très petits volumes, en raison de leur coût.
- Utilisation d'une plateforme automatisée à partir d'un certain volume (à partir de 2 To de données), en fonction des moyens humains.

exemples de solutions selon la volumétrie

Pour les petits volumes (50 Go) :

- utilisation de DVD-R ou CD-R pour le stockage (en 2 exemplaires) ;
- contrôle régulier des supports et une recopie à prévoir dans un délai de 5 à 7 ans ;
- faire tester un échantillon de CD-R régulièrement (par société privée ou par logiciel dédié).

Pour les volumes moyens (5 To) :

- copies multiples en diversifiant les technologies de stockage ;
- conserver sur site une copie disque et une copie bande et sur site distant une copie bande ;
- archivage de la totalité des fonds sur serveur en RAID 5 avec copie LTO (acquisition d'un auto-loader) en architecture de type DAS, NAS.

- le microforme⁸² comme archive informatique : microfilm argentique impressionné par un laser numérique, le microfilm numérique s'impose désormais comme un moyen sûr d'archiver à long terme des informations numériques. Ces microfilms sont réalisés par des traceurs micrographiques ou traceurs de fichiers numériques sur microfilm 35 mm et 16 mm.

82. La commission CGZ43 C (Archivage des données électroniques) est chargée de l'élaboration d'une norme française sur l'archivage des données électroniques sur microformes COM (Computer Output Microforms, sortie d'ordinateur sur microfiche) y associant éventuellement la technologie COLD (Computer Output to Laser Disk, sortie d'ordinateur sur disque laser).

SOLUTIONS DE TIERS-ARCHIVAGE

+++++

Des offres de tiers-archivage numérique (ou archivage électronique) sont proposées à la fois par des établissements publics et par des entreprises privées.

Du côté du privé, l'offre actuelle d'archivage numérique se structure principalement autour de la notion d'archivage sécurisé à valeur probante comprenant la dématérialisation, la sécurisation des échanges et la certification avec les tiers de confiance chargés de garantir l'authenticité des documents (signature électronique) et la sécurité des échanges.

Notions à prendre en compte pour choisir une solution pérenne de tiers-archivage :

- une gestion sur le long terme : format d'archivage, migration, respect des normes en terme d'archivage pérenne ;
- la taille des fichiers à archiver : possibilité d'archiver de grandes masses de données (ex. les masters) ;
- la place des métadonnées et l'indexation.

Les coûts sont actuellement de 7 000 euros le To par an. Ces coûts varient en fonction des services rendus, de la fréquence de l'accès aux données et en fonction de la mutualisation possible.

tiers stockage (court et moyen terme) / tiers-archivage (long terme)

Un système d'archivage n'est pas :

- un système de sauvegarde ou de secours ;
- un rangement définitif des données.

Il s'agit d'un ensemble de processus et de moyens concourant à la pérennisation des informations.

Un système de stockage est une plateforme ou un élément de base sur lequel s'appuie le système d'archivage. Le recours à l'externalisation du stockage permet de faire appel à des acteurs disposant de compétences et de faire des économies d'échelle, compte tenu de l'importance des coûts fixes de ce type de projet (personnel, matériel).

ENCADRÉ**PLATEFORMES D'ARCHIVAGE D'ÉTABLISSEMENTS PUBLICS****Nom : Plateforme d'Archivage du Cines (PAC)**

Établissement : Centre informatique national de l'enseignement supérieur (Cines)

Types de données archivées : Thèses électroniques déposées dans STAR (Abes), revues en SHS du portail Persée, documents sonores issus de la recherche dans le domaine de l'oral (Projet pilote CRDO dans le cadre du programme SHS du TGE-Adonis)

Volumétrie : 1 To archivé, capacité 20 To extensible à 40 To

Principes organisationnels :

- basée sur les standards du domaine : modèle OAIS, protocole standard d'échange de données pour l'archivage, métadonnées Dublin Core, identifiant pérenne (ARK), calcul d'empreinte numérique ;
- liste des formats de fichiers acceptés volontairement limitée : formats publiés, largement utilisés, normalisés si possible ;
- architecture basée sur du matériel SUN, le logiciel Arcsys et des logiciels libres (Java, MySQL, Jhove, ImageMagick).

Volonté de se positionner comme tiers archiveur à destination du monde de l'enseignement supérieur et de la recherche.

Lien : < <http://www.cines.fr/-D-I-S-T-.html> >

Nom : Système de Préservation et d'Archivage Réparti (SPAR)

Établissement : Bibliothèque nationale de France

Types de données archivées :

- production issue de la numérisation de conservation et de consultation (imprimés, images, audiovisuel, manuscrits...) ;
- production issue du dépôt légal automatique et négocié (support, Web, presse quotidienne régionale) ;
- production administrative et technique (documents internes).

Volumétrie : 400 To archivés, 1 Po attendu en 2010

Principes organisationnels :

- basée sur les standards du domaine : modèle OAIS, dictionnaire de métadonnées (METS, Dublin Core, PREMIS, ODRL), identifiant pérenne (ARK), calcul d'empreinte numérique ;
- infrastructure basée sur du matériel SUN, applicatif développé par Atos Origin basé sur Fedora-Commons

Volonté de se positionner comme tiers archiveur

Lien :

< <http://bibnum.bnf.fr/spar/index.html> >

En 2009, l'Association des développeurs et des utilisateurs de logiciels libres pour l'administration et les collectivités territoriales (ADULLACT) crée une coopérative API pour proposer de l'infogérance d'archivage électronique aux collectivités.

ENCADRÉ**PLATEFORMES D'ARCHIVAGE PRIVÉES****Nom : Rosetta**

Établissement : ExLibris, développé en collaboration avec la bibliothèque nationale de Nouvelle-Zélande (depuis 2008).

Principes organisationnels :

- progiciel à destination des bibliothèques ;
- se veut un système évolutif, prenant en compte les collections des bibliothèques de toute taille (documents numérisés, électroniques, audiovisuel, sites Web), tenant compte du cycle de vie du document, de l'évolutivité des formats ;
- système basé sur les standards du domaine tel que le modèle OAIS.

Lien : < <http://www.exlibrisgroup.com/category/ExLibrisRosettaOverview> >

Nom : E-coffre fort

Établissement : Office Français pour la Sécurité et l'Archivage des Documents (OFSAD)

Types de données archivées : documents papier et documents électroniques.

Volumétrie : capacité du coffre illimité et durée de conservation illimitée

Principes organisationnels :

offre s'adressant aux particuliers, aux associations et aux entreprises

- conversion automatique en PDF avec engagement à suivre les éventuelles modifications du format PDF ;
- engagement de restituer des copies numériques des documents d'origine papier toujours lisibles et exploitables plus de 30 ans après leur dépôt ;
- diversité des modes de dépôt (par imprimante virtuelle, par téléchargement par Drag & Drop (et copier-coller), par dépôt initial sur CD-Rom ;
- cryptage et empreinte numérique ;
- archivage de trois originaux sur deux sites distants ;
- certification ;
- référence à la norme NF Z 42-013

Archivage électronique.

Tarifs : offres distinctes pour particuliers (de 3 € HT à 12 € HT par mois) et professionnels (de 12 € HT à 28 € HT par mois). Labellisé par la FNTC en 2003

Lien : < <http://www.e-coffrefort.fr/index.htm> >

...

Nom : Le coffre fort électronique[®]**Établissement : CDC Arkhinéo****Types de données archivées :**

tout document

Principes organisationnels :

- enregistrement de quatre originaux sur plusieurs disques magnétiques répartis sur deux sites distants, *back up* automatique quotidien sur disques magnétiques sur un 3^e site distant ;
- constitution d'une enveloppe logique contenant le calcul d'empreinte, l'horodatage et le document reçu ;
- attribution automatique d'un identifiant unique d'archive et signature électronique de l'enveloppe d'une Public Key Infrastructure (PKI) interne.

Tarifs non communiqués, aucune migration de format semble prévue.

Lien : < <http://www.cdcarckhineo.com/index.php?page=accueil> >

Nom : Système d'Archivage Électronique**Établissement : Locarchives****Types de données archivées :**

tout document

Principes organisationnels :

- contrôle automatique des fichiers ;
- indexation des documents reçus ;
- conversion automatique en PDF-A,
- signature électronique (certification, empreinte, horodatage) ;
- certificat de réception ;
- stockage on line et off line (CD-R) sur deux sites distants ;
- utilisation de supports et formats normalisés : CD-R et PDF-A ;
- réalisation éventuelle de migration de supports en cas de non-lisibilité et possibilité d'assurer la conversion vers des formats pérennes.

Tarifs non communiqués, référence à la norme NF Z 42-013 Archivage électronique

Lien : < <http://www.locarchives.fr/solutions/archivage-documents-electroniques.html> >

Sur le site de la Fédération nationale des tiers de confiance (FNTC), liste des opérateurs et prestataires de services de confiance adhérents : < http://www.fntc.org/component/option,com_fntcmembre/task,college/Itemid,76/idCollege,3/ >

ENCADRÉ**PLATEFORMES D'ARCHIVAGE OPEN SOURCE**

Nom : Flexible Extensible Digital Object Repository Architecture (Fedora)

Développement par l'université de Cornell et l'université de Virginie

Lien : < <http://fedora-commons.org/> >

Nom : D-Space

Développement en 2002 par les bibliothèques du MIT et Hewlett-Packard.

Système tourné vers le monde académique.

Lien : < <http://dspace.org/> >

> **Plateformes** libres d'accès pouvant être adaptées aux spécificités de chaque service. Depuis 2009, ces deux systèmes sont unis dans Dura-Space.

Lien : < <http://duraspace.org/index.php> >

Nom : Lots of Copies Keep Stuff Safe (LOCKSS)

Mis en place par l'université de Stanford, ce système a pour objectif de fournir aux bibliothèques des outils et supports pour la préservation numérique.

Système basé sur le modèle OAIS, migration de formats, réparation des données.

Lien : < <http://www.lockss.org/lockss/Home> >

MÉMENTO

par Thierry Claerr et Isabelle Westeel

Le montage d'un projet de numérisation est délimité par l'état de l'art, les contraintes liées aux originaux à traiter, les objectifs de la numérisation, l'usage futur des fichiers et la situation de l'établissement (moyens budgétaires, humains, logistiques, positionnement par rapport à sa tutelle et au service informatique, niveau de compétences techniques). C'est un travail transversal avec plusieurs services, nécessitant une connaissance fine des aspects techniques de l'image numérique et des compétences en gestion de projet.

Ce mémento récapitule successivement les principales étapes d'un projet de numérisation et celles de la gestion de projet.

LES TROIS PRINCIPALES ÉTAPES

A. LA NUMÉRISATION

- Sélection intellectuelle du corpus (intérêt, homogénéité, complémentarité avec d'autres corpus, caractère « unique » des documents, nécessités de préservation des documents...)
- Étude des publics potentiels
- Analyse juridique
- Analyse du contexte institutionnel (moyens techniques, moyens humains disponibles, nécessité de formations)
- Recherche de partenariats extérieurs
- Analyse financière
- Formalisation des objectifs / Validation du projet (délibération, conseil scientifique ...)
- Planification du projet (ressources financières, matérielles et humaines)
- Sélection matérielle / Récolement / préparation du corpus
- Signalement / Catalogage des documents
- Rédaction du cahier des charges
- Numérisation / Suivi de la chaîne de production
- Suivi de la numérisation (contrôle qualité des images et des métadonnées)
- Retour des documents en magasin

- Stockage et conservation des supports informatiques (cédéroms, disques durs...)
- Migration des fichiers informatiques dans un système de stockage (sauvegarde / archivage)

B. LES MÉTADONNÉES

- Complétude du signalement / catalogage avant numérisation
- Choix des formats de métadonnées (privilégier des formats ouverts et normalisés)
- Élaboration d'un « guide pratique » d'application des formats
- Élaboration de métadonnées descriptives, administratives et de structure

C. LA BIBLIOTHÈQUE NUMÉRIQUE EN LIGNE

- Définition des besoins pour la mise en ligne
- Validation du projet
- Cadre de travail (mode projet : chef de projet, comité de suivi, planification...)
- Rédaction du cahier des charges GED (cahier de fonctionnalités et de services)
- Choix d'une solution logicielle
- Mise en place du flux documentaire (migration, alimentation du système...)
- Mise en place du flux éditorial par un comité de suivi : publication, valorisation, scénarisation...
- Interopérabilité (OAI, SRU/SRW...)
- Référencement du site sur les moteurs de recherche
- Dissémination sur le Web (APIs, *widgets*...), sur les terminaux mobiles
- Mise en production / Mise en ligne
- Communication
- Documentation validée du projet
- Évaluation des usages
- Évolution du système (possibilités de scalabilité, de flexibilité...)

Ces différentes étapes doivent correspondre aux logiques, aux missions et à la taille des établissements :

- pour la conservation : les fichiers correspondant à la numérisation de quelques manuscrits pourront être conservés correctement sur des CD. Par contre on pourra penser à un système de stockage/archivage plus élaboré si on se lance dans la numérisation des collections de presse.

- pour la mise en ligne : les objectifs peuvent également être gradués depuis la valorisation de quelques documents par l'intermédiaire d'un « produit multimédia » jusqu'à l'empilement quantitatif et qualitatif d'un corpus dans une bibliothèque numérique.

LA GESTION DE PROJET

+++++

Un projet est un ensemble d'activités qui sont prises en charge, dans un délai donné et dans les limites de ressources imparties, par des personnes qui y sont affectées dans le but d'atteindre des objectifs définis (norme AFNOR Z 67-100-1).

LE CADRE D'UN PROJET

- Nomination d'un chef de projet (qui peut être formalisée sous la forme d'une lettre de mission) ; le chef de projet supervise l'ensemble du programme du début à la fin
 - Mise en place d'un comité de pilotage (lieu décisionnel, validation des différentes étapes) et d'une équipe projet
 - Étude préalable, veille et *benchmarking* (comparaison avec des projets similaires mis en place dans d'autres établissements)
 - Mise en place d'un calendrier de projet (suivi des objectifs et des tâches). La technique du diagramme de Gantt permet de modéliser la planification des tâches (installation par un tableur, MS Project ou autres logiciels)
 - Définition des rôles et responsabilités de chacun au sein de l'équipe projet
 - Suivi du projet
 - Information de la direction et communication auprès des agents de l'établissement
 - Évaluation et suivi des risques
- > Périodes de l'année moins propices à la production
 - > Mauvaise coordination entre les différents plannings de réalisation
 - > Disponibilité des équipes de travail
 - > Départ d'un membre de l'équipe et son remplacement
 - > Avoir toujours un plan B et une solution de repli
- Rédaction de la synthèse à la fin du projet pour passage à une autre équipe, capitalisation des connaissances, montée en compétences des équipes et modélisation éventuelle.

LES CONDITIONS DE RÉUSSITE

- Validation politique du projet
- Inscription du projet dans le projet de l'établissement ou dans le projet de service (priorité et positionnement du projet par rapport aux autres missions)
- Inscription du projet dans le schéma directeur informatique
- Soutien du projet par la direction de l'établissement
- Collaboration avec le service informatique, le service des marchés, le service juridique
- Mise en place d'un plan de formation
- Liens avec la politique documentaire
- Suivi en mode projet
- Information institutionnelle sur l'avancée du projet
- Prise en compte de la « culture numérique » de l'établissement
- Gestion du changement : « les collaborateurs ne peuvent accepter un changement que s'ils comprennent ce qui ne changera pas »

EN GUISE DE CONCLUSION

+++++

Un projet de numérisation :

- s'inscrit dans un contexte scientifique, politique, administratif, juridique et financier précis dont il faut au préalable délimiter les contours,
- fait appel à des compétences variées tout en impliquant un nécessaire travail en équipe et en partenariat avec d'autres services et institutions.
- représente un investissement financier et humain de toute une équipe sur plusieurs mois, des budgets complémentaires en matériel pour l'exploitation et la maintenance, une coordination et un contrôle constants sur le long terme.

« Les bibliothèques numériques sont des organisations qui offrent des ressources, y compris en personnel, pour sélectionner, structurer, offrir un accès intellectuel, distribuer et conserver l'intégrité de documents sous une forme numérique »

IFLA. Digital libraries : definition, issues and challenges, mars 1998.

INDEX

- Accessibilité : 141-143
 Algorithme : 26, 28-29
 Allotissement : 81-82
 ALTO : 34-35, 108
 Analyse des offres : 75, 79, 97
 API : 118, 123, 136, 137
 Binarisation : 21, 38
 Bit : 20-21
 BOAMP : 70, 73
 BPU : 83, 97
 CCAG : 84, 86-88, 96, 98
 CCAP : 93, 95-98
 CCTP : 34, 36, 93, 95, 97
 CD : 171-175, 177, 184
 Colorimétrie : 21-22, 29-31
 Compression : 23
 Contrôle qualité : 35, 94
 Critères de sélection : 75, 79
 DADVSI : 59
 DAS : 175
 DCE : 73-75
 Définition : 23
 Dialogue compétitif : 72, 78
 Diagramme de Gantt : 185
 Disque dur externe : 171
 DVD : 171-175, 177
 Échantillonnage : 23, 37
 E-pub : 26
 Exception « bibliothèques » : 60
 Gallica : 7, 8, 14, 128
 GED : 8, 126, 130
 Granularité : 102
 JPEG : 25-26, 28-29, 110
 JPEG 2000 : 26, 164
 LED : 41
 Licence creative commons : 61
 Linked data : 121
 LTO : 171, 177
 LZW : 26, 29
 MAPA : 69-72, 78, 84, 86
 Marché négocié : 72, 77-78
 Mashup : 136, 138
 Métadonnées : 100-123, 169
 Méthode Agile : 122, 130
 METS : 104
 MICHAEL : 12
 Microfilm : 38, 91
 Microfilm numérique : 177
 Mire : 31
 Mode image/mode texte : 20, 33
 NAS : 175
 NQA : 35-36
 Numérisation en externe/
 en interne : 43
 Numériseur : 40, 91
 Numes : 12
 OAI-PMH : 115-118, 132
 OAIS : 105, 164-166
 Octet : 20
 OCR : 20, 33-35, 38, 92
 Œuvre collaborative
 ou de collaboration : 53, 61
 Œuvre collective : 54-55
 Œuvre orpheline : 54-55
 OWL : 121
 PAQ : 97
 Patrimoine numérique : 12-13
 Persée : 9, 105
 PDF : 28-29, 35, 108, 131, 163
 Pixel : 20-21
 Plan national de numérisation : 14
 PNG : 28-29, 35, 163

Poids d'une image numérique :	27
Point blanc :	30
Profil ICC :	29-31
Profondeur d'acquisition :	21-23
Protocole :	114-118
RAID :	176-177
RAW :	25
RDF :	119-120, 121
Reconnaissance optique	
de caractères :	voir OCR
Référentiel général d'interopérabilité-	
RGI :	142-144, 160, 162
Résolution :	22-24, 30, 40
RPC :	74-75, 97
RSS :	132-135
RVB :	23, 30
SAN :	175
SPAR :	105, 179
SRU/SRW :	115, 123, 132
Stockage hiérarchique :	176
Tiers-archivage :	178-181
TIFF :	28-29, 163
Tuile :	26, 29
XML :	34, 101, 103-104
WAI :	141-143
Web sémantique :	118
Widget :	136, 138

SIGLES ET ABRÉVIATIONS

ABES Agence bibliographique
de l'enseignement supérieur

ADAE Agence pour le développement
de l'administration électronique

ADAGP Société française de gestion
collective des droits d'auteur dans
les arts visuels (peinture, sculpture,
photographie, multimédia,...)

AE Acte d'engagement

AFNOR Association française
de normalisation

AJAX Asynchronous JavaScript
and XML

ALTO Analysed layout and text object

AMO Assistance à maîtrise d'ouvrage

API Application programming interface

ARK (Identifiant) Archival resource key

BnF Bibliothèque nationale de France

BOAMP Bulletin officiel des annonces
des marchés publics

BPU Bordereau de prix unitaires

CCAG Cahier des clauses
administratives générales

CCAG-FCS pour CCAG de fournitures
courantes et de services

CCAG-TIC pour CCAG de techniques de
l'information et de la communication

CCAP Cahier des clauses
administratives particulières

CCTP Cahier des clauses techniques
particulières

CD Compact disc

CIDOC/CRM Comité international
de la documentation/Conceptual
Reference Model

CGM Catalogue général des manuscrits
des bibliothèques publiques de France

CMS Content management system =
SGC

CNUM Conservatoire numérique
des arts et métiers

CPER Contrat de projet État-Région

CPI Code de la propriété intellectuelle

CSS Cascading Style Sheets

DADVSI [loi relative aux] Droits
d'auteur et droits voisins dans la société
de l'information

DAS Direct Attached Storage

DC Dublin core

DCE Dossier de consultation
des entreprises

DGD Dotation générale
de décentralisation

DOI Digital Object Identifier

DPI (dot per inch) ou **PPI** (pixel per
inch) Point par pouce ou point par pixel

DTD Document Type Definition

DVD Digital Versatile Disc

EAC-CPF Encoded Archival Context-
Collectivités, personnes, familles

EAD Encoded Archival Description

EXIF Exchangeable image file

FOAF Friend of a friend

FRBR Fonctionnal Requirements for
Bibliographic Records = spécifications
fonctionnelles des notices bibliographiques

GED Gestion électronique de documents

ICC International Color Consortium

IFLA International Federation
of Library Associations

IPTC International Press and
Telecommunications Council

JOUE Journal officiel
de l'Union européenne

JPEG Joint Photographic Experts Group

LOM Learning Object Metadata

MADS Metadata Authority
Description Schema

MAPA Marché à procédure adaptée

METS Metadata Encoding
and Transmission Standard

MODS Metadata Object
Description Schema

MRT Mission de la recherche
et de la technologie

NAS Network Attached Storage

NQA Niveaux de qualité acceptables

OAI-PMH Open Archive Initiative –
Protocol for Metadata Harvesting

OCR Optical character recognition =
reconnaissance optique de caractères

OWL Ontology Web Language

PAQ Plan assurance qualité

PDF Portable Document Format

PNG Portable Networks Graphics

RAID Redundant array of independent
disks/matrice redondante de disques
indépendants

RDA Resource Description and Access	TEF Thèses électroniques de France
RDF Resource Description Framework	TEI Text Encoding Initiative
RGAA Référentiel général d'accessibilité pour les administrations	TGE Très grand équipement
RPC Règlement particulier de la consultation	TIFF Tagged Image File Format
RGI Référentiel général d'interopérabilité	TUI Tangible User Interface
RVB Rouge, vert, bleu	UIT Union internationale des télécommunications
SACEM Société des auteurs, compositeurs et éditeurs de musique	URI Uniform Resource Identifier
SAN Storage Area Network	URL Uniform Resource Locator
SIG Système d'information géographique	VA Vérification d'aptitude
SIGB Système intégré de gestion de bibliothèque	VSR Vérification de service régulier
SGC Système de Gestion de Contenus = CMS	W3C World Wide Web Consortium
SGML Standard Generalized Markup Language	WAI Web Accessibility Initiative
SKOS Simple Knowledge Organisation System	WCAG Web Content Accessibility Guidelines
SPARQL Query Language for RDF	XML Extensible Markup Language
SRU/SRW Search and Retrieve via URL/Search and Retrieve via Web	XMP Extensible Metadata Platform
	XSLT Extensible Stylesheet Language Transformations
	XSL-FO Extensible Stylesheet Language – Formatting Objects

BIBLIOGRAPHIE

ÉTAT DES LIEUX, ÉTAT DE L'ART

+++++

Aubry Martine, Westeel Isabelle (dir.).

La numérisation des textes et des images : techniques et réalisations. Actes des journées d'étude organisées par l'Université de Lille III, 16 et 17 janvier 2003. Villeneuve-d'Asq, Presses de l'Université Charles de Gaulle-Lille III, 2003.

Bermès Emmanuelle, Fréon Marie-Élise, Martin Frédéric. « Tous les chemins mènent au numérique ». *Bulletin des bibliothèques de France*, t. 53, n° 6, 2008, p. 34-39.

Berners-Lee Tim. « Le Web va changer de dimension », *La Recherche*, novembre 2007, p. 34-38.

Chaumier Jacques. *Document et numérisation. Enjeux techniques, économiques, culturels et sociaux.* Paris, ADBS, 2006.

Claerr Thierry. « La conservation des collections numérisées dans les bibliothèques municipales : état des lieux et perspectives ». 4^e édition des Entretien de la BnF « Numérique et bibliothèques : le deuxième choc : 7 et 8 décembre 2006 ». [En ligne] < <http://www.bnf.fr/> > (consulté le 8 mars 2010). Rubrique Formations et journées professionnelles

« De la lettre au nombre. La numérisation en marche » dans *Arabesques*, n° 42, avril-juin 2006.

« Numérisation du patrimoine culturel » dans *Culture & Recherche*, n° 118-119, automne-hiver 2008-2009. [En ligne] < www.culture.fr/culture/editions/r-cr.htm > (consulté le 8 mars 2010).

Weber Anne. « La numérisation des sources de l'histoire de l'art : la bibliothèque numérique de l'Institut national d'histoire de l'art ». *Bulletin des bibliothèques de France*, t. 53, n° 6, 2008, p. 54-62.

Westeel Isabelle. « Le patrimoine passe au numérique ». *Bulletin des bibliothèques de France*, t. 54, n° 1, 2009, p. 28-35.

GÉSTION DE PROJET

+++++

Berdugo Alain. *Le maître d'ouvrage du système d'information : une fonction stratégique confirmée.* Paris, Hermès-Lavoisier, 2005.

Buresi Charlette, Cédelle-Joubert Laure. *Conduire un projet de numérisation.* Villeurbanne : Presses de l'enssib ; Paris : Éditions Tec & Doc, 2002 (La Boîte à outils ; 13).

Ferraille Jean-François, Machefert Sylvain, Scopsi Claire, Soual Laurent. *Mener un projet Open Source en bibliothèque, documentation et archives.* Paris, Éditions du Cercle de la Librairie, 2007 (Bibliothèques).

Kattnig Cécile. *Gestion et diffusion d'un fonds d'images.* Paris, Nathan université-ADBS, 2002 (Collection 128 : information et documentation).

ASPECTS JURIDIQUES

+++++

Chevry Emmanuelle (dir.). *Les droits de reproduction dans les bibliothèques, archives et musées. Actes de la journée d'étude à Troyes le 18 novembre 2004.* Reims, université de Reims Champagne-Ardenne, 2005.

Maurel Lionel. *Bibliothèques numériques : le défi du droit d'auteur.* Villeurbanne, Presses de l'enssib, 2008.

ASPECTS TECHNIQUES

+++++

Andrieu Olivier. *Réussir son référencement Web.* Paris, Eyrolles, 2008.

Bermès Emmanuelle. « Les moteurs de recherche. Petit précis de mécanique à l'usage des bibliothèques numériques ». *Bulletin des bibliothèques de France*, t. 52, n° 6, 2007, p. 5-10.

Bermès Emmanuelle. « Des identifiants pérennes pour les ressources numériques : l'expérience de la BnF ». *International Preservation News*, n° 40, décembre 2006, p. 16-21.
< <http://archive.ifla.org/VI/4/news/ipnn40.pdf> >

Griffey Jason. *Mobile Technology and Libraries.* New-York, Neal-Schuman, 2010.

Calderan Lisette, Hidoine Bernard, Millet Jacques. *Métadonnées : mutations et perspectives : séminaire INRIA, 29 septembre-3 octobre 2008, Dijon.* Paris, ADBS, 2008.

Evans Woody. *Building Library 3.0: Issues in Creating a Culture of Participation.* Oxford, Chandos, 2009.

Jacquesson Alain, Rivier Alexis. *Bibliothèques et documents numériques. Concepts, composantes, techniques et enjeux.* Paris, Éditions du Cercle de la Librairie, 2005 (Bibliothèques).

Lejeune Yannick. *XML.* Paris, Micro Application, 2004 (Guide Microapp).

Le Moal Jean-Claude, Hidoine Bernard. *Bibliothèques numériques : cours INRIA, 9-13 octobre 2000, La Bresse.* Paris, ADBS, 2000.

Le Moal Jean-Claude, Hidoine Bernard, Calderan Lisette. *La recherche d'information sur les réseaux : cours INRIA, 30 septembre - 4 octobre 2002, Le Bono (Morbihan).* Paris, ADBS, 2002.

Le Moal Jean-Claude, Hidoine Bernard, Calderan Lisette. *Publier sur Internet : séminaire INRIA, 27 septembre-1^{er} octobre 2004, Aix-les-Bains*. Paris, ADBS, 2004.

Ministère de la Culture et de la Communication. *Culture et handicap. Guide pratique de l'accessibilité*. 2007. [En ligne]
< <http://www.culture.gouv.fr/handicap/index.html> > (consulté le 8 mars 2010).

Nielsen Jakob. *Conception de sites web : l'art de la simplicité*. Paris, CampusPress, 2000.

Stutzmann Dominique. « La prise en compte du handicap dans la bibliothèque numérique *Gallica* » dans Andissac Marie-Noëlle et Poitevin Marie-José (dir.). *Handicap et bibliothèque*. Paris, Association des bibliothécaires français, 2^e édition revue et augmentée, 2009 (Médiathèmes), p. 110-117.

Thurrow Shari. *Être visible sur Internet*. Paris, Pearson Education, 2008 (Campuspress Reference).

Troncy Raphaël. « Nouveaux outils et documents audiovisuels : les innovations du web sémantique ». *Documentaliste – Sciences de l'information*, vol. 42, n° 6, décembre, 2005, p. 392-404.

CONSERVATION NUMÉRIQUE

+++++

Banat-Berger Françoise, Duploux Laurent, Huc Claude. *L'archivage numérique à long terme. Les débuts de la maturité ?* Paris, La Documentation française, 2009 (Manuels et guides pratiques/Archives nationales).

Calderan Lisette, Hidoine Bernard, Millet Jacques. *Pérenniser le document numérique : séminaire INRIA, 2-6 octobre 2006, Amboise*. Paris, ADBS, 2006.

Cathaly-Prétou Gérard, Fontaine Jean-Marc, Lafosse Jacques. *L'archivage sur CD-R : acquérir, graver, contrôler, conserver*. Paris, Eyrolles, 2006.

Leblond Corinne (dir.). *Archivage et stockage pérennes*. Paris, Londres, Hermès Science, 2009.

SITES DE RÉFÉRENCE

Agence bibliographique de
l'enseignement supérieur (Abes) :
*Numes, inventaire des corpus numérisés
dans l'enseignement et la recherche* :
< <http://www.numes.fr> >

Association Aristote. *Pérennisation
des Informations Numériques (PIN)* :
< <http://www.aristote.asso.fr/PIN/> >

Bibliothèque nationale de France :
actualités professionnelles.
Numérisation : < <http://www.bnf.fr> >
Charte technique de numérisation,
éléments sur la préservation
du document numérique :
< <http://bibnum.bnf.fr> >

Europeana :
< [http://www.europeana.eu/portal/
aboutus.html#technicalplans](http://www.europeana.eu/portal/aboutus.html#technicalplans) >

Groupement d'intérêt scientifique
sur la conservation des informations
enregistrées sur les disques optiques
numériques (GIS-DON) :
< [http://www.lne.fr/fr/r_et_d/gis-
don/conservation-donnees-
numeriques-gis-don.asp](http://www.lne.fr/fr/r_et_d/gis-don/conservation-donnees-numeriques-gis-don.asp) >

*Ministerial network for valorising
digitisation activities* (Minerva) –
Réseau ministériel pour la valorisation
des activités de numérisation :
< <http://www.minervaeurope.org/> >

Ministère de la Culture
et de la Communication/Mission
de la recherche et de la technologie.
Numérisation du patrimoine culturel :
< <http://www.culture.gouv.fr/mrt/numeration/index.html> > (actualités, fiches
techniques et juridiques et référentiel
conservation numérique).
Patrimoine numérique :
< <http://www.numerique.culture.fr> >

Ministère de la Culture et de la
Communication/Direction des archives
de France (DAF)
e_archives, numérisation :
< [http://www.archivesdefrance.culture.
gouv.fr](http://www.archivesdefrance.culture.gouv.fr) >

Portail international archivistique
francophone : < [http://www.piaf-
archives.org/](http://www.piaf-archives.org/) > Espace formation

Très grand équipement (TGE) Adonis.
« Guide des bonnes pratiques
numériques » (v1, décembre 2009)
< [http://www.tge-
adonis.fr/wiki/index.php/guides](http://www.tge-adonis.fr/wiki/index.php/guides) >

Université Cornell, Bibliothèque :
Didacticiel :
< [http://www.library.cornell.edu/
preservation/tutorial-
french/contents.html](http://www.library.cornell.edu/preservation/tutorial-french/contents.html) >

DES BLOGS ET DES WIKIS

+++++

Abondance.com, l'actualité
et l'information sur le référencement
et les moteurs de recherche :
< <http://www.abondance.com> >

Bibliopedia

Page « Bibliothèques numériques » :
< http://www.bibliopedia.fr/index.php/Biblioth%C3%A8ques_num%C3%A9riques >
Page « Archives en ligne » :
< http://www.bibliopedia.fr/index.php/Archives_en_ligne >

Delicious. Base sur les usages Web 2.0
en bibliothèques :
< http://delicious.com/webographie_cnft >

Figoblog : < <http://www.figoblog.org/> >

Lespetitescases.net. Blog spécialisé
sur le Web sémantique :
< <http://www.lespetitescases.net/> >

Wikipedia pour les aspects
techniques et informatiques :
< <http://www.wikipedia.fr> >

LISTE DES ILLUSTRATIONS

Publics visés et services liés	p. 10
Dispositifs spécifiques de financements de la numérisation	p. 16-17
Grille de pixels en niveaux de gris	p. 21
Conversion du signal analogique en signal numérique	p. 21
Profondeur d'acquisition recommandée selon le support	p. 22
Résolutions préconisées selon le type et le format des originaux	p. 24
Rapports de poids selon la profondeur d'acquisition et la résolution	p. 27
Caractéristiques des principaux formats de fichier d'images fixes	p. 28-29
Apports de l'OCR et difficultés posées par la conversion	p. 31
OCR - Segmentation d'une zone imprimée et caractérisation des boîtes	p. 32
Méthodes de contrôle qualité possibles et leurs principales caractéristiques ...	p. 36
Avantages et limites d'un contrôle qualité sous-traité	p. 37
Numérisation et microfilms	p. 38-39
Choisir un numériseur	p. 40-41
Vue, non exhaustive, des marques de numériseurs disponibles en France	p. 42
Numérisation interne/externe	p. 43
Le droit à l'image des personnes ..	p. 58
Le droit à l'image des biens	p. 58
Schéma récapitulatif du droit d'auteur .	p. 64
Les seuils prévus par le code des marchés publics pour les marchés fournitures et services	p. 72
Clauses des pièces constitutives d'un marché public	p. 85
Trois dispositions possibles pour la prise de vue des originaux (parmi d'autres) ...	p. 91
Principaux thèmes de l'instruction préalable d'un marché de numérisation	p. 93
Principaux champs d'un CCTP d'un marché de numérisation	p. 95
Principaux champs d'un CCAP d'un marché de numérisation	p. 96
Métadonnées techniques et descriptives	p. 106-108
Métadonnées pour les informations géographiques	p. 109
Les métadonnées images	p. 110-111
Liste des quinze éléments Dublin Core .	p. 113-114
Une requête SRU vers la bibliothèque du Congrès	p. 115
Les six requêtes OAI-PMH	p. 117
L'ORI-OAI	p. 118
Le triplet	p. 119
Vocabulaires RDF	p. 120
Exemples de vocabulaires RDF ..	p. 120
Fonctionnalités et services à prévoir dans un cahier des charges « Mise en ligne »	p. 131-132
Identifiants et URL	p. 133
Quelques exemples d'applications Web 2.0 dans les bibliothèques numériques	p. 135
Un exemple d'utilisation d'API	p. 137
Un exemple de <i>mashup</i>	p. 137
Le Label Accessiweb	p. 144
Données disponibles sur les plateformes statistiques	p. 145
Le protocole Sitemaps	p. 152

Liste des formats recommandés
en terme de conservation
par le RGI (version 1.0) p. 163
Modèle d’environnement
d’un OAIS p. 165
Entités fonctionnelles OAIS p. 165
Supports de stockage p. 171
Composition d’un CD ou DVD p. 172
Tiers stockage (court et moyen terme) /
Tiers-archivage (long terme) p. 178
Plateformes d’archivage d’établissements
publics p. 179
Plateformes d’archivage privées
..... p. 180-181
Plateformes d’archivage open source
..... p. 182

presses ensib

LISTE DES AUTEURS

Stéphane Alcandre

Bibliothèque nationale de France,
directeur adjoint du Département des
affaires juridiques et de la commande
publique, chef du Service des marchés

Thierry Claerr

ministère de la Culture
et de la Communication,
Service du livre et de la lecture

Bernard Fages

Bibliothèque nationale de France,
Direction des services et des réseaux,
Département de la conservation,
expert technique en reproduction

François Houste

LSF Interactive, directeur
« Projets spéciaux et Analytics »

Patricia Le Galèze

ministère de la Culture
et de la Communication,
Service du livre et de la lecture

Catherine Mocellin

Bibliothèque municipale d'Orléans,
directrice adjointe

Sophie Sepetjan

Bibliothèque nationale de France,
chef du Service juridique

Isabelle Westeel

Conseil régional Nord-Pas de Calais

Secrétariat d'édition :
Silvia Ceccani

Conception graphique
et mise en page :
atelier Perluette,
Charline Bourbon

Achevé d'imprimer
en mars 2010
par l'imprimerie Chirat

Dépôt légal :
1^{er} semestre 2010

presses ensib