



Recommandations techniques pour les programmes de création de contenus culturels numériques

Version 1.0: Révisée le 7 mai 2004

Ce document a été élaboré dans le cadre du projet Minerva par l'UKOLN, Université de Bath, en collaboration avec *MLA : le Conseil pour les Musées, Bibliothèques et Archives*.

L'édition française a été réalisée pour le compte de la Mission de la recherche et de la technologie du Ministère de la culture et de la communication (partenaire français du projet Minerva) par Muriel Foulonneau (Relais Culture Europe), **Sarah Faraud (Relais Culture Europe)** et Alexandra Bonnamy (traductrice).



Le Groupe des Représentants Nationaux sur la numérisation et le projet européen MINERVA

MINERVA : Réseau Ministériel pour la Valorisation des Activités de Numérisation (*Ministerial Network for Valorising Digitisation Activities*)

<<http://www.minervaeurope.org/>>

Le groupe des représentants nationaux (GRN) des Etats-membres de l'Union Européenne sur la numérisation du patrimoine culturel et scientifique¹ a été créé en avril 2001 à Lund (Suède). Les "principes de Lund " adoptés par les Etats-membres définissent le cadre de la coopération européenne sur la mise en œuvre des politiques de numérisation, en identifiant les bonnes pratiques, en favorisant la création et l'identification de centres de compétences, en coopérant pour faciliter l'accès aux ressources numérisées et en développant des grilles d'évaluation comparative. Le projet MINERVA est financé par la Commission européenne (direction générale sur la société de l'information) dans le cadre du 5e Programme cadre pour la Recherche & Développement, depuis le 1er mars 2002 et agit comme secrétariat opérationnel des groupes de travail issus du plan d'action de Lund.

D'une durée de 3 ans, le projet MINERVA, coordonné par le ministère italien de la culture, a pour objectif de développer des actions concrètes dans les domaines suivants :

- évaluation de la **qualité des sites Internet culturels** pour faciliter l'accès aux ressources,
- identification des **bonnes pratiques** dans la conduite des projets de numérisation,
- **évaluation comparative** des politiques, programmes et projets de numérisation, en adoptant une grille commune,
- **inventaires des fonds et projets de numérisation** afin de permettre un recensement coordonné et comparable des activités de numérisation et étude des **aspects multilingues** de l'accès aux ressources culturelles numérisées,
- **interopérabilité** des systèmes documentaires d'accès aux ressources et gestion des **droits de propriété intellectuelle** afin de favoriser la constitution d'un "environnement européen de l'information", cadre de développement des contenus sur les réseaux (eEurope).

Ces thématiques ont donné naissance à des groupes de travail constitués d'experts de différents Etats-membres, en favorisant une approche transversale (musées, bibliothèques, archives, patrimoine bâti, archéologie) et l'utilisation de technologies ouvertes.

Au sein de MINERVA, le Ministère de la Culture et de la Communication (mission de la recherche et de la technologie et département des affaires internationales) en collaboration avec l'association Relais Culture Europe est chargé de coordonner les travaux sur le recensement des activités de numérisation, en se basant sur l'expérience de l'annuaire français des fonds numérisés².

Un guide réalisé avec les projets européens EMII-DCF, ERPANET et PULMAN

Le présent guide a été réalisé en collaboration avec d'autres projets européens relatifs à la numérisation des données culturelles. Les projets ERPANET (réseau européen sur l'accès et la préservation des ressources électroniques), PULMAN (Bibliothèques publiques gérant des réseaux avancés) et EMII-DCF (Institut européen d'information des Musées - Plate-forme de contenus distribués) ont participé aux discussions sur le guide et ont contribué aux études sur les standards et technologies utilisés pour construire des services culturels en ligne.

Ce rapport, rédigé par Pete Johnston de l'UKOLN (Royaume-Uni), a bénéficié des contributions du groupe de travail de MINERVA sur l'interopérabilité des systèmes d'information culturels animé par le Royaume-Uni.

Projet ERPANET (réseau européen sur l'accès et la préservation des ressources électroniques)
<<http://www.erpanet.org/>>

¹ Christophe Dessaux (Mission de la recherche et de la technologie du ministère de la culture et de la communication) et Daniel Malbert (Département des affaires internationales du ministère de la culture et de la communication) y représentent la France à la date de l'édition du présent document.

² http://www.culture.gouv.fr/culture/mrtnumerisation/fr/f_02.htm

Projet EMII-DCF (Institut européen d'information des Musées- Plate-forme de contenus distribués)
<<http://www.emii-dcf.org/>>

Projet PULMAN (Bibliothèques publiques gérant des réseaux avancés)
<<http://www.pulmanweb.org/>>

La version française du guide

La version originale de ce document mentionne essentiellement des références en anglais, c'est pourquoi des références en français ont été ajoutées dans certaines sections. Les sites ont été consultés entre le 20 avril 2004 et le 10 juin 2004. Les lecteurs sont invités à consulter la version anglaise de ce document sur le site du projet Minerva.

Version originale du présent document :

<<http://www.minervaeurope.org/publications/technicalguidelines.htm> >

Sous la direction de Pete Johnston, UKOLN avec les contributions des membres du groupe de travail 4 de Minerva :

- Eelco Bruinsma, Consultant, Hollande
- Rob Davies, MDR Partners / Projet PULMAN, Royaume Uni
- David Dawson, MLA, Royaume Uni
- Bert Degenhard Drenth, Adlib Information Systems / Projet EMII-DCF, Hollande
- Giuliana De Francesco, Ministero per i beni e le attività culturali, Italie
- Muriel Foulonneau, Relais Culture Europe / Projet EMII-DCF, France
- Gordon McKenna, mda / Projet EMII-DCF, Royaume Uni
- Paul Miller, UKOLN, Royaume Uni
- Maureen Potter, Projet ERPANET, Hollande
- Jos Taekema, Digital Erfgoed Hollande, Hollande
- Chris Turner, MLA, Royaume Uni

Crédits

Ce document est principalement basé sur quatre sources.

- Les *Recommandations et Standards Techniques de NOF-digitise* (Version 5, février 2003), réalisées pour le compte du fonds britannique *New Opportunities Fund* (NOF), par l'UKOLN, Université de Bath, en collaboration avec l'agence britannique *Resource: le Conseil pour les Musées, Archives et Bibliothèques*.
- Des informations complémentaires fournies aux projets *NOF-digitise* comme support aux *Standards et Recommandations* du Service de Conseil Technique de *NOF-digitise*, développées pour le NOF par l'UKOLN et le Service des Données sur les Arts et les Humanités³ britannique (AHDS), sous la forme du manuel du programme, d'instructions et de FAQs.
- Le *Rapport Cadre* (septembre 2003), publié par le projet de l'Institut Européen d'Information des Musées *Plate-forme pour les contenus distribués* (EMII-DCF)⁴, en particulier le *Modèle de numérisation des données* décrit dans le chapitre 16.
- Le *Guide des Bonnes Pratiques* (Version 1.2, novembre 2003), réalisé par le projet Minerva (Groupe de Travail n°6).

³ AHDS - Arts and Humanities Data Service

⁴ EMII-DCF European Museums' Information Institute – Distributed Content Framework

Les standards et recommandations techniques de *NOF-digitise*
<<http://www.peoplesnetwork.gov.uk/content/technical.asp>>

Le manuel du service de conseil technique du programme *NOF-digitise*
<<http://www.ukoln.ac.uk/nof/support/manual/>>

La FAQ du service de conseil technique du programme *NOF-digitise*
<<http://www.ukoln.ac.uk/nof/support/help/faqs/>>

Le *Rapport cadre* EMII-DCF
<<http://www.emii-dcf.org/dokument/frame.pdf>>

Le *Guide des bonnes pratiques* du Groupe de Travail n°6 de Minerva , et la version française datant de juin 2004
< <http://www.minervaeurope.org/publications/goodhand.htm> >

Il s'inspire également des documents suivants :

Le *Cadre d'orientation pour construire des collections numériques de bonne qualité* de l'Institut américain des Services aux Bibliothèques et aux Musées (IMLS)
<<http://www.ims.gov/pubs/forumframework.htm>>

Le *Guide des bonnes pratiques de la représentation et la gestion numérique des ressources patrimoniales* du NINCH (initiative nationale américaine pour un patrimoine en réseau)
<<http://www.ninch.org/programs/practice/>>

Les *Recommandations pour numériser des ressources* pour les *Ressources Culturelles* du Groupe des bibliothèques de recherche (RLG) américain
<<http://www.rlg.ac.uk/culturalres/prospective.html>>

Recommandations pour les descriptions de l'Initiative pour les ressources culturelles du Groupe des bibliothèques de recherche (RLG) américain
<<http://www.rlg.ac.uk/culturalres/descguide.html>>

Les standards et recommandations pour les projets de numérisation du Patrimoine canadien
<http://www.pch.gc.ca/progs/pcce-ccop/pubs/ccop-pccegide_e.pdf>

Travailler avec les ressources nationales numériques distribuées (DNER): Standards et recommandations pour réaliser une ressource nationale
<http://www.jisc.ac.uk/index.cfm?name=projman_standards>

L'Ensemble des standards de l'architecture de l'environnement de l'information du JISC (comité commun des systèmes d'information britannique)
<<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/standards/>>

Les recommandations des Bibliothèques publiques gérant des réseaux avancés (PULMAN)
<<http://www.pulmanweb.org/DGMs/DGMs.htm>>

La majeure partie des références en français est tirée des documents suivants :

Les éléments techniques et juridiques du site de la numérisation du ministère de la culture et de la communication
<<http://www.numerique.culture.fr>>

Le guide de la numérisation des collections de la sous -direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche
<<http://www.sup.adc.education.fr/bib/>>

Le guide de la création et gestion de contenu numérique du Musée Virtuel du Canada
<http://www.chin.gc.ca/Francais/Contenu_Numerique/index.html>

Le didacticiel de l'Université de Cornell en français : De la théorie à la pratique : didacticiel de l'imagerie numérique
<<http://www.library.cornell.edu/preservation/tutorialfrench/contents.html>>

Le cadre commun d'interopérabilité des systèmes d'information publics Version 2.1 (septembre 2003) publié par l'ADAE (Agence pour le développement de l'administration électronique)

<http://www.adae.pm.gouv.fr/article.php3?id_article=219>

Guide de conservation des informations et des documents numériques de l'ADAE (Agence pour le développement de l'administration électronique)

<http://www.adae.gouv.fr/article.php3?id_article=7>

L'UKOLN est un centre d'expertise sur la gestion de l'information numérique qui fournit des conseils et des services dans les secteurs des bibliothèques, des centres d'informations, de l'éducation et du patrimoine. Il est financé par l'agence britannique *MLA : le conseil pour les musées, les archives et les bibliothèques*, le *comité commun pour les systèmes d'information (JISC)* des *Conseils de financement de l'éducation supérieure* et par des financements de projets du JISC et de l'Union européenne. L'UKOLN reçoit également un soutien de l'Université de Bath où il est basé.

Sommaire

1.

1.Introduction.....	8
1.1.L'objectif de ce rapport.....	8
1.2.Le rôle des standards techniques.....	9
1.3.Les avantages des standards.....	9
1.4.L'approche du cycle de numérisation.....	10
1.5.Les niveaux d'exigence.....	11
2.La préparation à la numérisation.....	12
2.1.Le matériel.....	12
2.2.Les logiciels.....	12
2.3.L'environnement.....	12
3.La manipulation des documents originaux.....	14
3.1.Déplacement et manipulation appropriés des documents sources.....	14
3.2.Formation du personnel.....	14
4.Le processus de numérisation.....	15
5.Stockage et gestion des enregistrements de référence.....	17
5.1.Les formats de fichiers.....	17
5.1.1.Numérisation et stockage de texte.....	18
5.1.1.1.Le codage de caractères.....	18
5.1.1.2.Les Formats de documents.....	18
5.1.2.Numérisation et stockage d'une image fixe.....	19
5.1.2.1.Les images matricielles.....	19
5.1.2.2.Images graphiques non vectorielles.....	21
5.1.2.3.Images vectorielles.....	21
5.1.3.Numérisation et stockage de vidéo.....	21
5.1.4.Numérisation et stockage du son.....	22
5.2.Le choix des supports de stockage.....	22
5.3.Les stratégies de préservation.....	23
6.Création/saisie de métadonnées.....	24
6.1.La portée des métadonnées.....	24
6.2.Les standards appropriés.....	24
6.2.1.Les métadonnées descriptives.....	25
6.2.2.Les métadonnées administratives.....	26
6.2.3.Les métadonnées de conservation.....	27
6.2.4.Les métadonnées de structure.....	28
6.2.5.La description au niveau des collections.....	28
6.2.6.Les standards terminologiques.....	29
7.La publication.....	31
7.1.Les traitements pour la publication.....	31
7.1.1.La publication de texte.....	31
7.1.1.1.Le codage de caractères.....	31
7.1.1.2.Les formats de documents.....	31
7.1.2.La diffusion d'images fixes.....	32
7.1.2.1.Les images photographiques.....	32
7.1.2.2.Les images graphiques non-vectorielles.....	32
7.1.2.3.Les images graphiques vectorielles.....	32
7.1.3.La publication de vidéo.....	32

7.1.3.1.Le téléchargement.....	32
7.1.3.2.La diffusion en flux continu.....	32
7.1.4.La publication de matériel audio.....	33
7.1.4.1.Le téléchargement.....	33
7.1.4.2.La diffusion en flux continu	33
7.1.5.L'identification.....	33
7.2.Questions relatives à la 3D et à la réalité virtuelle.....	33
7.3.Les systèmes d'informations géographiques.....	34
7.4.Les sites Internet.....	35
7.4.1.L'accessibilité.....	35
7.4.2.La sécurité.....	36
7.4.3.L'authenticité.....	37
7.4.4.L'authentification des utilisateurs.....	37
7.4.5.Les indicateurs de performance.....	38
8.La diffusion des ressources.....	39
8.1.La collecte de métadonnées.....	39
8.2.La recherche en architecture distribuée.....	40
8.3.Les alertes.....	41
8.4.Les services WEB.....	41
8.5.Cadre de description des données (RDF) et ontologies Web.....	42
9.La réutilisation pour d'autres objectifs.....	44
9.1.La création de ressources éducatives.....	44
10.Droits de propriété intellectuelle	45
10.1.Identifier les droits d'auteur.....	45
10.2.Préserver les droits d'auteur.....	46
10.2.1.Le commerce électronique.....	46
10.2.2.Les licences Creative Commons.....	47
10.2.3.Les tatouages et empreintes numériques.....	47
11.Résumé.....	48
11.1.Maintenance.....	48

1. Introduction

A travers l'Europe, des initiatives internationales, nationales, régionales et locales voient le jour grâce à des financements des secteurs public et privé pour permettre l'accès à de vastes ensembles de ressources patrimoniales par différents canaux numériques. Les motivations de ces projets peuvent varier considérablement ; ils peuvent concerner différents types de ressources, répondre aux besoins de groupes spécifiques d'utilisateurs ou viser des objectifs sociaux et économiques particuliers.

Les divers ministères, établissements ou agences qui soutiennent des programmes de numérisation partagent généralement un souci commun : optimiser la valeur de leurs investissements en souhaitant que le contenu numérique produit soit aussi utile, portable et pérenne que possible. Ces critères se rattachent à la notion d'interopérabilité des ressources (et des mécanismes au travers desquels les ressources sont accessibles).

Le principe de l'interopérabilité est d'assurer une approche cohérente de la création, de la gestion et de l'édition des ressources numériques au travers de l'utilisation efficace de standards, de règles et de recommandations qui constituent des bonnes pratiques.

L'adoption d'ensembles communs de standards et recommandations techniques constitue souvent le premier pas vers un accord et une uniformisation au sein d'un programme. Ce document vise à fournir des recommandations pour l'usage de standards – principalement techniques. Il est avant tout destiné aux décideurs et chefs de projet qui mettent en œuvre les programmes de création de contenus culturels numériques.

1.1. L'objectif de ce rapport

Ce rapport ne vise pas à imposer un ensemble unique de prescriptions auxquelles tous les projets devraient se conformer. Il serait illusoire de rédiger un document qui rassemblerait toutes les exigences spécifiques aux contextes de nombreux programmes et il est normal que différentes approches des recommandations soient retenues. Il s'agit plutôt d'identifier les domaines pour lesquels il existe une similarité d'approche et de fournir un ensemble commun de prescriptions autour duquel d'autres prescriptions, spécifiques à chaque contexte, peuvent être élaborées. En ce sens, son cadre et son importance sont similaires à ceux du *Modèle de numérisation des données* d'EMII-DCF⁵ et certaines recommandations du présent document sont directement issues de ce dernier.

L'utilisation de ces recommandations ne peut pas garantir que les ressources seront "interopérables". Les prescriptions destinées à assurer l'utilité, la portabilité et la pérennité des ressources numériques varient d'un programme à l'autre et la forme sous laquelle les standards sont déployés par chaque projet reflète ces différences. Alors que les recommandations fournies dans ce document visent à être applicables de manière générale, chaque programme opère dans un contexte dans lequel les projets doivent se plier à des contraintes et des standards déterminés par différents acteurs (institutionnels, liés à des programmes, sectoriels, régionaux, nationaux, internationaux). Par exemple, les programmes financés par le secteur public doivent utiliser les standards imposés par les gouvernements nationaux.

Au cours de la vie d'un programme, l'environnement technologique peut changer car les standards évoluent. Les gestionnaires des programmes devraient se tenir informés de tous les développements en cours de standards utiles dans leur contexte d'exploitation.

Il est important qu'ils assurent un accompagnement des projets sous la forme de services conseil qui peuvent fournir des indications sur l'interprétation et la mise en œuvre des standards et des recommandations. Il garantit également que ces recommandations et ces standards sont mis à jour afin de refléter leurs évolutions significatives.

⁵ in *Rapport cadre* d'EMII-DCF

1.2. Le rôle des standards techniques

Le *Rapport cadre* d'EMII-DCF rappelle la définition d'un standard utilisée par l'Institution Britannique des Standards (BSI)⁶ :

“ Un standard est une spécification publiée qui établit un langage commun, il contient une spécification technique ou d'autres critères précis et est conçu pour être utilisé de manière invariable, comme une règle, une recommandation ou une définition. Les standards sont appliqués à de nombreuses ressources, des produits, des méthodes et des services. Ils contribuent à faciliter la vie et à augmenter la fiabilité et l'efficacité de nombreux biens et services que nous utilisons. ”

L'utilisation appropriée des standards de numérisation peut permettre d'atteindre le niveau d'invariabilité qui rend l'interopérabilité possible. Un haut niveau d'invariabilité entre les ressources numériques de nombreux fournisseurs signifie que l'outil ou le service opérant au travers de ces ressources ne doit travailler qu'avec un nombre limité de formats, d'interfaces et de protocoles clairement définis. En revanche, un nombre de plus en plus important de formats et de protocoles différents rendrait de tels développements complexes, coûteux et non fiables, si ce n'est impossibles. La procédure par laquelle les standards eux-mêmes sont développés signifie qu'ils incluent des bonnes pratiques basées sur les expériences passées et qu'ils imposent une plus grande rigueur dans les pratiques en cours.

Les standards sont souvent définis comme soit

- *de jure* – formellement reconnus par un organisme chargé de mettre en place et de diffuser des standards, en général développés au travers d'un accord commun entre un certain nombre d'acteurs intéressés. L'ensemble de protocoles TCP/IP, maintenus par l'*Internet Engineering Task Force* (IETF) en est un exemple.
- *de facto* – ne sont pas formellement reconnus par un organisme de standardisation, cependant largement utilisés et reconnus comme standards par ses utilisateurs. Il peut s'agir par exemple d'un format de fichier utilisé par un logiciel qui occupe une part prédominante ou importante du marché dans un domaine particulier tel que le format *Adobe Portable Document Format* (PDF).

Un critère important pour définir les standards est leur "ouverture". Elle peut se référer à un certain nombre de caractéristiques d'un standard et le *Rapport Cadre* EMI-DCF⁷ souligne trois aspects d'un intérêt majeur pour les utilisateurs de standards:

- l'accès ouvert (au standard et aux documents produits au cours de son développement);
- l'utilisation libre (implémenter le standard induit peu ou pas de coûts liés aux droits de propriété intellectuelle, au travers de licences d'utilisation par exemple); et
- le support orienté vers les besoins des utilisateurs plutôt que vers l'intérêt des fournisseurs de standards.

En fonction du scénario décrit ci-dessus, les spécifications des formats, interfaces et protocoles utilisés par les fournisseurs de ressources sont librement disponibles. De multiples développeurs peuvent donc développer des outils et services similaires en évitant la dépendance vis-à-vis d'un seul outil ou d'une seule plate-forme.

En général, les procédures formelles liées au développement de standards *de jure* sont perçues comme la garantie que ces standards sont réellement "ouverts". Dans les présentes recommandations, la préférence est accordée à des standards ouverts, mais dans certains cas, l'utilisation de standards industriels ou de facto peut être envisagée.

1.3. Les avantages des standards

Il est important de prendre en considération les aspects suivants :

⁶ *British Standards Institution* (BSI)

⁷ *Framework Report*

- **L’interopérabilité.** Les contenus devraient être accessibles similairement par les utilisateurs, quel que soit le projet et quel que soit le programme de financement. Il devrait être possible de localiser et interagir avec les contenus de manière invariable, de les utiliser facilement, sans outil spécialisé et de les gérer efficacement.
- **L’accessibilité.** Ces ressources devraient être aussi accessibles que possible et qu’elles soient publiées dans des formats normalisés ouverts et non-propriétaires. Si ces ressources sont utiles à un large ensemble d’utilisateurs, il est nécessaire d’envisager l’accès pour de multiples communautés linguistiques et d’assurer l’accessibilité à des citoyens affectés de divers handicaps.
- **La conservation.** L’avenir à long terme des ressources devrait être garanti de sorte que le retour sur investissement soit maximisé et les données culturelles maintenues dans leur continuité historique et la diversité de leurs supports.
- **La sécurité.** A l’âge des réseaux, il devrait être possible d’établir l’identité des contenus et des projets (et le cas échéant, des utilisateurs) ; de protéger les droits de propriété intellectuelle et la confidentialité ; enfin de déterminer l’intégrité et l’authenticité des ressources.

Faute de traiter efficacement ces différentes questions, de sérieuses difficultés peuvent se poser, entraînant le gaspillage de ressources par différents acteurs:

- **Les utilisateurs – le citoyen, l’élève, l’enfant.** Ils engageront vainement du temps et des efforts s’ils ne trouvent pas immédiatement ou n’utilisent pas ce qui est le plus adapté à leurs besoins parce que ce n’est pas décrit correctement ou parce que ce n’est pas disponible dans un format particulier ou parce que cela nécessite des outils spécialisés pour être exploité ou parce qu’enfin ce n’est pas numérisé sous une forme facilement utilisable.
- **Les fournisseurs et les gestionnaires d’informations.** Leur investissement peut être redondant et gaspillé si leurs ressources ne révèlent pas leur valeur à l’usage, si leurs produits n’atteignent qu’une partie de leur public, enfin s’ils investissent dans des pratiques non standard ou dépassées.
- **Les financeurs.** Ils doivent payer pour l’effort redondant et fragmenté, pour la répétition superflue des procédures d’apprentissage, pour les équipes projet qui opèrent moins efficacement qu’elles ne devraient et déploient des techniques moins optimales, pour les contenus qui ne correspondent pas aux besoins des utilisateurs ou aux nécessités du marché.
- **Les créateurs et auteurs.** Ce qu’ils lèguent à l’avenir peut être perdu.

1.4. L’approche du cycle de numérisation

La structure du présent rapport reflète une approche basée sur le “cycle de vie” du processus de numérisation et correspond à la structure du *Guide des Bonnes Pratiques* développé par le groupe de travail n° 6 du projet Minerva.

Il est divisé en sections, chacune reflétant une étape de ce cycle. En pratique, il existe des relations et des dépendances entre les activités développées au cours de ces différentes étapes et il est possible que certaines de ces étapes ne soient pas strictement séquentielles.

a. La préparation à la numérisation,

1. La manipulation des originaux,
2. Le processus de numérisation,
3. Le stockage et la conservation des enregistrements de référence,
4. La saisie des métadonnées,
5. La publication,
6. La diffusion,
7. La réutilisation pour d’autres objectifs,
8. Les droits de propriété intellectuelle.

1.5. Les niveaux d'exigence

La prise en compte des standards et des recommandations varient en fonction des programmes suivant un éventail allant de l'adoption de bonnes pratiques à l'obligation de compatibilité avec les standards comme condition d'attribution des financements. Les standards et recommandations adoptés par les programmes correspondent généralement à différents niveaux d'exigence et il est possible de distinguer entre :

- **Les prescriptions** : les standards largement admis et utilisés. Les projets **doivent** mettre en œuvre les standards qui sont identifiés comme des exigences.
- **Les recommandations** qui représentent les bonnes pratiques mais il est possible qu'il y ait des raisons de ne pas les appliquer en totalité, par exemple parce que ces standards sont encore en développement. Les équipes projets **devraient** rester attentives à l'évolution de ces standards, et de leurs applications potentielles.

La distinction entre les prescriptions et les recommandations est en général effectuée dans le contexte d'un programme particulier.

Dans le contexte d'un programme spécifique, les auteurs de recommandations pour l'usage de standards techniques devront faire la distinction entre exigences et conseils.

Dans les documents de standardisation, les mots “ **doit**, **devrait** et **peut** ”, revêtent des significations précises sur les niveaux d'exigence:

- **doit** : ce verbe indique une prescription technique à laquelle tous les projets doivent se plier.
- **devrait** : ce verbe est employé lorsqu'il existe des raisons valables de ne pas considérer ce conseil comme une exigence absolue mais toutes les implications que cela entraîne doivent alors être clairement comprises. Le verbe '**devrait**' est utilisé avec des standards techniques qui seront probablement largement utilisés dans le futur, au cours du cycle de vie du projet mais qui sont encore dans une phase de diffusion.
- **peut** : ce verbe indique que le sujet mérite attention mais qu'il ne s'agit pas d'une contrainte. Le terme '**peut**' a été utilisé pour se référer à des standards qui sont encore en développement.

Ce vocabulaire est basé sur la terminologie utilisée par la documentation de l'*Internet Engineering Task Force* (IETF).

Ces mots-clés sont utilisés dans le reste de ce document. Pour un programme spécifique, les auteurs **devraient** adapter les niveaux d'exigence spécifiés à ceux de leur propre contexte.

IETF RFC 2119 Mots-clés pour l'usage dans les RFCs pour indiquer les niveaux d'exigence de l' <i>Internet Engineering Task Force</i> < http://www.ietf.org/ >

2. La préparation à la numérisation

Les équipes projet **doivent** acquérir une bonne connaissance des collections destinées à être numérisées et de l'utilisation des ressources numériques créées. Lors du choix du matériel et des logiciels de numérisation, les équipes projet **doivent** prendre en compte les caractéristiques des documents originaux telles que leur format, leur taille, leur état et être conscientes de l'importance d'enregistrer précisément leurs attributs tels que la couleur.

Recommandations:

TASI: Conseils: créer des images numériques du Service britannique de conseil technique pour les images

<<http://www.tasi.ac.uk/advice/creating/creating.html>>

Le processus de numérisation, article écrit pour le programme *New Opportunities Fund* par l'UKOLN (université de Bath, Royaume-Uni)

<http://www.ukoln.ac.uk/nof/support/help/papers/digitisation_process/>

En français

Le guide de la numérisation des collections de la sous-direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche

<<http://www.sup.adc.education.fr/bib/Acti/Num/analyse.htm#preparation>>

Le guide de la création et gestion de contenu numérique du Musée Virtuel du Canada

<http://www.chin.gc.ca/Francais/Contenu_Numerique/index.html>

2.1. Le matériel

Ce document ne donne pas de conseils précis quant au choix de l'équipement de numérisation. Les équipes projet **doivent** démontrer qu'elles savent quels sont les équipements disponibles sur le marché, les facteurs déterminant s'il est possible ou non de les utiliser avec différents objets physiques ainsi que la manière de les connecter à du matériel externe, tel qu'un ordinateur personnel.

Les équipes projet **doivent** s'assurer que l'équipement choisi produit des objets numériques d'une qualité adaptée aux utilisations prévues, à un coût acceptable.

Avant d'acquérir un équipement de numérisation ou de sous-traiter à des prestataires, l'équipe projet **devrait** se faire conseiller de manière appropriée et **devrait** en estimer le coût exact, en se basant sur les conditions spécifiques du projet.

2.2. Les logiciels

Ce document ne donne pas de conseils précis quant au choix des logiciels de numérisation. Les équipes projet **doivent** démontrer qu'elles connaissent l'usage des logiciels pour la saisie et le traitement des images, ainsi que les pré-requis et compatibilités matériels et logiciels.

L'équipe projet **doit** s'assurer que le logiciel dispose des fonctionnalités nécessaires à l'utilisation des objets numériques créés, à un coût acceptable et qu'il est utilisable par les personnels concernés.

2.3. L'environnement

Il est important d'établir un environnement approprié au processus de numérisation. Il en assure l'efficacité pour la création de ressources numériques facilement utilisables et garantit que le moins possible de dommages seront occasionnés aux documents physiques qui sont les sources.

La numérisation peut, soit être effectuée au sein même du service sur des équipements existants ou acquis à cet effet, soit être sous-traitée. L'équipe projet **doit** comprendre les facteurs qui déterminent ce choix, non seulement en termes de coûts mais également de spécifications lors de la manipulation des documents physiques et de la création d'objets numériques.

3. La manipulation des documents originaux

3.1. Déplacement et manipulation appropriés des documents sources

Les questions de conservation portent à la fois sur l'objet physique numérisé et sur sa représentation numérique, une fois celle-ci créée. Les responsables du projet **doivent** mesurer les risques que présente l'exposition des originaux à un processus de numérisation, notamment lorsque les objets sont uniques, précieux ou fragiles. Ils **doivent** discuter du processus avec les personnes responsables des originaux.

Recommandations:

Recommandations pour les images numériques de la conférence commune sur la conservation du Bureau national de la conservation (NPO, Irlande et Royaume-Uni) et du Groupe des bibliothèques de recherche (RLG, Etats-Unis) 28 - 30 Septembre 1998
<<http://www.rlg.org/preserv/joint/>>

3.2. Formation du personnel

Les équipes projet **doivent** s'assurer que l'ensemble du personnel reçoit une formation à l'utilisation des matériels et logiciels ainsi qu'à une manipulation correcte des originaux lors du processus de numérisation. Cela garantit l'efficacité du processus et réduit les risques de dégradation des originaux.

Recommandations:

TASI: Conseil: Créer des images numériques du Service britannique de conseil technique pour les images
<<http://www.tasi.ac.uk/advice/creating/creating.html>>

4. Le processus de numérisation

La numérisation est la conversion de fonds analogiques (ou physiques) en fonds numériques en vue d'une utilisation par des logiciels. Les décisions prises au moment de la numérisation ont un impact sur la facilité avec laquelle les ressources créées pourront être gérées et accessibles, mais également sur leur viabilité.

Il est difficile de spécifier de manière exhaustive les standards requis pour la capture initiale des données car ils évoluent avec le temps et différents types de ressources peuvent nécessiter des standards différents. Cependant, les équipes projet **doivent** démontrer qu'elles ont pris en considération les implications des trois étapes suivantes :

- la sélection des fonds à numériser,
- la préparation physique des fonds pour la numérisation,
- le processus technique de numérisation.

L'Initiative pour la numérisation des images du JISC (JIDI), le Service des Données sur les Arts et les Humanités britannique (AHDS)⁸ et le Service britannique de conseil technique pour les images (TASI)⁹ fournissent des recommandations plus développées sur ces questions.

De nombreuses recommandations sur la numérisation sont également disponibles dans diverses publications, notamment le texte récent d'Anne R. Kenney et d'Oya Y. Rieger, *De la théorie à la pratique : imagerie numérique pour les bibliothèques et les archives*¹⁰ (Groupe des bibliothèques de recherche¹¹, 2000).

Il est également important de signaler les actes de conférence du RLG/NPO, rassemblés dans les *Recommandations pour les images numériques*¹² (Bureau National de Conservation¹³, 1998). De plus, la Fédération américaine des bibliothèques électroniques¹⁴, le Conseil des bibliothèques et ressources d'information¹⁵ et le Groupe des bibliothèques de recherche¹⁶ ont récemment publié plusieurs guides utiles sur la qualité des images¹⁷.

⁸ *Arts and Humanities Data Service*

⁹ *Technical Advisory Service for Images*

¹⁰ *Moving Theory into Practice: digital imaging for libraries and archives*

¹¹ *Research Libraries Group*

¹² *Guidelines for Digital Imaging*, Recommandations pour les images numériques de la conférence commune sur la conservation du Bureau National de la Conservation (NPO, Irlande et Royaume-Uni) et du Groupe des Bibliothèques de Recherche (RLG, Etats-Unis) 28 - 30 Septembre 1998

¹³ *National Preservation Office*

¹⁴ *Digital Library Federation*

¹⁵ *Council on Library and Information Resources*

¹⁶ *Research Libraries Group*

¹⁷ *Guides to Quality in Visual Resource Imaging*

Recommandations:

JIDI, l'Initiative pour la numérisation des images du JISC (comité commun des systèmes d'informations britannique)

<<http://www.ilrt.bris.ac.uk/jidi/>>

AHDS: Guides des bonnes pratiques de la création et de l'usage des ressources numériques du Service britannique des Données sur les Arts et les Humanités

<<http://ahds.ac.uk/guides.htm>>

TASI (Service britannique de conseil technique pour les images)

<<http://www.tasi.ac.uk/>>

Une étude de faisabilité pour l'Initiative de numérisation des images du JISC (JIDI)

<http://heds.herts.ac.uk/resources/papers/jidi_fs.html>

Recommandations pour les images numériques de la conférence commune sur la conservation du Bureau national de la conservation (NPO, Irlande et Royaume-Uni) et du Groupe des bibliothèques de recherche (RLG, Etats-Unis) 28 - 30 Septembre 1998

<<http://www.rlg.org/preserv/joint/>>

Guide de la qualité des images des ressources visuelles du Groupe des bibliothèques de recherche (RLG)

<<http://www.rlg.ac.uk/visguides/>>

En français

Le guide de la numérisation des collections de la sous-direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche

<<http://www.sup.adc.education.fr/bib/Acti/Num/notech.htm>>

Le guide de la création et gestion de contenu numérique du Musée Virtuel du Canada

<http://www.chin.gc.ca/Francais/Contenu_Numerique/index.html>

5. Stockage et gestion des enregistrements de référence¹⁸

La conservation **doit** être considérée comme faisant partie intégrante du processus de création numérique. Elle dépend de la documentation de l'ensemble des procédures techniques qui ont conduit à la création d'un objet. Beaucoup d'informations importantes peuvent – dans la plupart des cas – être enregistrées lors de la création.

Les équipes projet **doivent** être conscientes de l'intérêt que représente la création d'un "enregistrement de référence" en haute définition et intégralement documenté, à partir duquel toutes les autres versions (par exemple des versions comprimées pour l'accès Internet) peuvent être dérivées. Cela facilite la migration périodique des données et le développement de nouveaux produits et de nouvelles ressources.

Il est important de comprendre que la conservation ne consiste pas seulement à choisir des formats de fichiers ou des types de supports appropriés. Elle devrait être considérée comme une responsabilité de gestion fondamentale pour ceux qui possèdent et gèrent des contenus numériques, assurant leur utilisation et réutilisation à long terme. Elle dépend d'un grand nombre de facteurs extérieurs au processus de numérisation, tels que la stabilité institutionnelle, la continuité du financement et la détention des droits de propriété intellectuelle.

Des stratégies techniques peuvent cependant être adoptées au cours du processus de numérisation pour faciliter la conservation. Par exemple, de nombreux projets de numérisation ont commencé à élaborer des stratégies basées sur la création d' "enregistrements de référence" contenant des métadonnées. Un bref panorama technique de la stratégie des "enregistrements de référence" est décrite dans l'article produit par HEDS¹⁹ pour le programme britannique *NOF-digitise* sur le processus de numérisation.

Recommandations:

Recommandations pour les images numériques de la conférence commune sur la conservation du Bureau national de la conservation (NPO, Irlande et Royaume-Uni) et du Groupe des bibliothèques de recherche (RLG, Etats-Unis) 28 - 30 Septembre 1998

[<http://www.rlg.org/preserv/joint/>](http://www.rlg.org/preserv/joint/)

Guide de gestion de la conservation des ressources numériques de la Coalition pour la conservation numérique

[<http://www.dpconline.org/graphics/handbook/>](http://www.dpconline.org/graphics/handbook/)

Le processus de numérisation, article écrit pour le programme *New Opportunities Fund* par l'UKOLN (université de Bath, Royaume-Uni)

[<http://www.ukoln.ac.uk/nof/support/help/papers/digitisation_process/>](http://www.ukoln.ac.uk/nof/support/help/papers/digitisation_process/)

En français

Le guide de la création et gestion de contenu numérique du Musée Virtuel du Canada

[<http://www.chin.gc.ca/Francais/Contenu_Numerique/Numerisez_Collections/normes.html#guidelines>](http://www.chin.gc.ca/Francais/Contenu_Numerique/Numerisez_Collections/normes.html#guidelines)

5.1. Les formats de fichiers

Des formats ouverts **devraient** être utilisés lors de la création de ressources numériques afin d'en maximiser l'accès. (N.B. : les formats de fichiers pour la publication d'enregistrements numériques sont décrits dans la section .)

L'utilisation de formats de fichiers ouverts favorise l'interopérabilité, garantissant la possible réutilisation des ressources ainsi que leur création et modification par un grand nombre d'applications. Cela évite également la dépendance vis-à-vis d'un fournisseur particulier.

¹⁸ Egalement appelés "masters" ou "documents maîtres", ce sont les originaux numériques.

¹⁹ Service de numérisation de l'enseignement supérieur britannique <http://heds.herts.ac.uk/>

Il se peut qu'aucun standard ouvert approprié n'existe ou que les standards soient trop récents pour que les outils conformes soient tous disponibles. Dans certains cas, l'utilisation de formats propriétaires **peut** donc être acceptable. Lorsque des formats propriétaires sont utilisés, l'équipe projet **devrait** envisager une stratégie de migration permettant une transition ultérieure vers des standards ouverts.

Si les standards ouverts ne sont pas utilisés, les équipes projet **devraient** justifier le recours à des formats propriétaires dans leurs propositions, en portant une attention particulière aux questions d'accessibilité.

5.1.1. Numérisation et stockage de texte

5.1.1.1. Le codage de caractères

Un codage de caractères est un algorithme permettant de représenter des caractères sous une forme numérique en définissant une équivalence entre des séquences de code de caractères (les entiers correspondant à des caractères dans un répertoire) et des séquences de valeurs de 8 bits (octets). Pour pouvoir interpréter les bits qui composent un objet numérique, une application doit connaître le mode de codage des caractères utilisé.

Le codage de caractères utilisé par les documents textuels **devrait** être explicitement signalé. Pour les documents XML, le codage des caractères **devrait** être enregistré dans la déclaration d'encodage de la déclaration XML.

Pour les documents XHTML, la déclaration XML **peut** être omise, mais le codage **doit** être enregistré dans la valeur de l'attribut `http-equiv` d'un élément `meta`.

Pour les questions relatives au codage pour la publication de documents, se reporter à la section 5.1.1

Standards:

Le Consortium Unicode. Le standard Unicode, Version 4.0.0, défini par: *The Unicode Standard, Version 4.0* (Boston, MA, Addison-Wesley, 2003. ISBN 0-321-18578-1)

<<http://www.unicode.org/versions/Unicode4.0.0/>>

Extensible Markup Language (XML) 1.0, site du consortium *World Wide Web*

<<http://www.w3.org/TR/REC-xml/>>

XHTML 1.0 The Extensible HyperText Markup Language, site du consortium *World Wide Web*

<<http://www.w3.org/TR/xhtml1/>>

Recommandations:

Un tutoriel sur les questions de codage de caractères de Jukka Korpela de l'Université de Tampere

<<http://www.cs.tut.fi/~jkorpela/chars.html>>

En français

Site Babel de l'*Internet Society*

<<http://alis.isoc.org/codage/>>

5.1.1.2. Les Formats de documents

Le contenu textuel **devrait** être créé et géré dans un format structuré, adapté à la génération de documents HTML ou XHTML destinés à être publiés.

Dans la plupart des cas, stocker du contenu textuel en SGML ou en XML, en se conformant à une définition de type de document (DTD) ou à un schéma XML sera l'option la plus appropriée. Les équipes projet **peuvent** choisir de stocker ce contenu soit dans des fichiers plats, soit à l'intérieur d'une base de données. Tous les documents **devraient** être validés en fonction de la DTD ou du schéma XML correspondant.

Les équipes projets **devraient** connaître et comprendre l'intérêt de formats standardisés pour l'encodage de textes, tels que la *Text Encoding Initiative* (TEI), et **devraient** stocker le contenu textuel sous de tels formats, lorsque cela paraît approprié. Les équipes projet **peuvent** stocker le contenu textuel en HTML 4 ou XHTML 1.0 (ou des versions ultérieures). Elles **peuvent** stocker le

contenu textuel sous des formats SGML ou XML, conformes à d'autres DTD ou schémas, mais elles **doivent** définir des équivalences avec un schéma reconnu.

Dans certains cas, les équipes projets **peuvent** choisir de stocker les fonds textuels en utilisant le format *Portable Document Format* (PDF) d'*Adobe*. Le PDF est un format de fichier propriétaire détenu par *Adobe* qui préserve les polices, le format, les couleurs et les graphiques du document source. Les fichiers PDF sont compacts et peuvent être visualisés et imprimés grâce au visualiseur gratuit *Acrobat Reader*. Cependant, comme pour toute solution propriétaire, leur adoption présente des dangers et les équipes projet **devraient** être conscientes des coûts potentiels de cette approche. Elles **devraient** étudier une stratégie de migration permettant une transition ultérieure vers des standards ouverts.

Standards:

ISO 8879:1986. Information Processing -- Text and Office Systems -- Standard Generalized Markup Language (SGML)

Extensible Markup Language (XML) 1.0, site du consortium *World Wide Web*
<<http://www.w3.org/TR/REC-xml/>>

Text Encoding Initiative (TEI) du consortium international TEI
<<http://www.tei-c.org/>>

HTML 4.01 HyperText Markup Language, site du consortium *World Wide Web*
<<http://www.w3.org/TR/html401/>>

XHTML 1.0 The Extensible HyperText Markup Language, site du consortium *World Wide Web*
<<http://www.w3.org/TR/xhtml1/>>

Autres références:

Portable Document Format (PDF) d'Adobe
<<http://www.adobe.com/products/acrobat/adobepdf.html>>

Recommandations:

Guide des bonnes pratiques de l'AHDS (Service britannique des Données sur les Arts et les Humanités) : *Créer et documenter des textes électroniques*
<<http://ota.ahds.ac.uk/documents/creating/>>

5.1.2. Numérisation et stockage d'une image fixe

Les images numériques fixes entrent dans deux catégories principales : les images matricielles (ou "pixelisées") et les images vectorielles ("orientées objet"). Les images matricielles prennent la forme d'une grille ou matrice, où chaque "élément d'image" (pixel) a un emplacement unique dans la matrice et une valeur de couleur indépendante pouvant être modifiée séparément. Les fichiers vectoriels fournissent un ensemble d'instructions mathématiques utilisées par un programme de dessin pour construire une image.

En général, le processus de numérisation génère une image matricielle, les images vectorielles étant plus souvent le produit d'un logiciel de dessin.

5.1.2.1. Les images matricielles

Lors de la création et du stockage d'images matricielles, deux facteurs doivent être pris en considération : le format de fichier et les paramètres de qualité.

Les images matricielles **devraient** en principe être stockées sous une forme non comprimée générée par le processus de numérisation, sans aucun retraitement. Les images matricielles **doivent** être créées en utilisant l'un des formats suivants : *Tagged Image File Format* (TIFF), *Portable Network Graphics* (PNG), *Graphical Interchange Format* (GIF) ou *JPEG Still Picture Interchange File Format* (JPEG/SPIFF).

Deux paramètres fondamentaux doivent être pris en compte :

- **La résolution spatiale** : la fréquence à laquelle des échantillons de l'original sont capturés par le dispositif de numérisation, exprimée sous la forme d'un nombre d'échantillons par pouce (spi)²⁰ ou plus communément sous la forme de pixels par pouce, ppp²¹ dans l'image numérique qui en résulte.
- **La résolution des couleurs (profondeur de bits)** : le nombre de couleurs (ou de niveaux de luminosité) disponibles pour représenter différentes couleurs (ou tons de gris) dans l'original, exprimé en nombre de bits. Par exemple, une résolution de couleurs de 8 bits signifie que 256 couleurs différentes sont disponibles.

De manière générale, les images photographiques **devraient** être créées au format TIFF.

La sélection des paramètres de qualité nécessaires à la numérisation d'une ressource est déterminée par la taille de l'original, la quantité de détails présents dans l'original et les utilisations prévues de l'image numérique. Numériser une diapositive de 35mm exige une résolution plus élevée que dans le cas d'une lithographie de 6x4 car elle est plus petite et plus détaillée. Si l'une des utilisations de l'image d'une aquarelle requiert de pouvoir analyser d'infimes détails de coups de pinceaux, la résolution nécessaire est plus élevée que pour le seul affichage de l'image à l'écran.

Les images **devraient** être créées à la résolution adaptée et à la profondeur de bits la plus élevée, à un coût acceptable et en demeurant pratiques et maniables au vu des utilisations envisagées. Chaque équipe projet **doit** identifier le niveau minimal de qualité et de densité d'informations dont elle a besoin.

A titre d'exemple, une résolution de 600 points par pouce (ppp) et une profondeur de bits de 24 bits couleur ou de 8 bits à échelle de gris **devraient** être envisagées pour les impressions photographiques. Une résolution de 2400 ppp **devrait** être appliquée pour des diapositives de 35 mm afin de capturer la plus grande densité d'informations. (Source : EMII DCF)

Dans certains cas, par exemple lors de l'utilisation d'appareils photo numériques bon marché, il **peut** être indiqué de stocker les images sous un format JPEG/SPIFF, comme alternative au format TIFF. Les images seront alors plus petites mais de plus basse qualité. De telles images peuvent être adaptées à la présentation de photographies d'événements, etc., sur un site Internet mais l'utilisation de tels appareils photos n'est pas recommandée pour la numérisation de contenu à grande échelle. (Source : *NOF-digitise*)

Standards:

Tagged Image File Format (TIFF) sur le site d'Adobe

< <http://www.itu.int/itudoc/itu-t/com16/tiff-fx/docs/tiff6.pdf> >

Joint Photographic Expert Group (JPEG), site du consortium *World Wide Web*

<<http://www.w3.org/Graphics/JPEG/>>

JPEG Still Picture Interchange File Format (SPIFF), site du Groupe JPEG

<<http://www.jpeg.org/public/spiff.pdf>>

Recommandations:

TASI: Conseil: *Créer des images numériques* du Service britannique de conseil technique pour les images

<<http://www.tasi.ac.uk/advice/creating/creating.html>>

En français

De la théorie à la pratique : didacticiel de l'imagerie numérique, didacticiel de l'Université de Cornell

<<http://www.library.cornell.edu/preservation/tutorial-french/presentation/presentation-02.html>>

Le guide de la numérisation des collections de la sous-direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche

<<http://www.sup.adc.education.fr/bib/Acti/Num/notech.htm>>

²⁰ *samples per inch - spi*

²¹ *pixels per inch - ppi*

5.1.2.2. Images graphiques non vectorielles

Les images créées par ordinateur, telles que les logos, les icônes et les dessins au trait **devraient** en principe être créées en tant qu'images PNG ou GIF à une résolution de 72 ppp. (N.B. : les images issues de la numérisation de dessins au trait devraient être gérées comme décrit dans la section précédente).

Standards:

Portable Network Graphics (PNG), site du consortium *World Wide Web*
<<http://www.w3.org/TR/PNG>>

5.1.2.3. Images vectorielles

Les images vectorielles sont constituées de multiples objets géométriques (lignes, ellipses, polygones, entre autres formes) construits grâce à une séquence de commandes ou de formulations mathématiques pour tracer des lignes et des formes. Les graphiques vectoriels **devraient** être créés et stockés en utilisant un format ouvert tel que *Scalable Vector Graphics (SVG)*, un langage XML permettant de décrire de tels graphiques. Les dessins SVG peuvent être interactifs et dynamiques, ils sont modulables en fonction des différents affichages possibles à l'écran et en fonction des différentes résolutions d'imprimante.

L'utilisation du format propriétaire *Flash* de *Macromedia* peut également être approprié ; cependant les équipes projet **devraient** envisager une stratégie de migration vers des formats plus ouverts lorsque ceux-ci seront plus largement déployés. Toutefois, l'utilisation du texte dans le format Flash devra être évitée afin de permettre le développement de versions multilingues.

Standards:

Scalable Vector Graphics (SVG), site du consortium *World Wide Web*
<<http://www.w3.org/TR/SVG/>>

Autres références:

Macromedia Flash
<<http://www.macromedia.com/>>

5.1.3. Numérisation et stockage de vidéo

La vidéo **devrait** en principe être stockée sous une forme non comprimée obtenue à partir du matériel de numérisation, sans traitement ultérieur. La vidéo **devrait** être créée à la plus haute résolution possible, à une intensité de couleur et à une fréquence de trame²² abordables et pratiques compte tenu des utilisations prévues. Chaque équipe projet **doit** identifier le niveau minimal de qualité requis.

La vidéo **devrait** être stockée en utilisant le format non compressé RAW AVI, sans aucun codec²³, à une taille de trame de 720x576 pixels, une fréquence de trame de 25 images par seconde, avec 24 bits couleur. Le codage de couleurs PAL devrait être utilisé.

La vidéo **peut** être créée et stockée en utilisant le format MPEG approprié (MPEG-1, MPEG-2 ou MPEG-4) ou les formats propriétaires *WMF* de *Microsoft*, *ASF* ou *Quicktime*.

²² Nombre de fois par seconde où l'image est redessinée

²³ Un codec est un algorithme permettant de réduire significativement les flux de données en COmpressant / DEcompressant les données vidéos.

Standards:

Moving Pictures Experts Group (MPEG), Groupe d'experts sur les images animées
<<http://www.cselt.it/mpeg/>>

5.1.4. Numérisation et stockage du son

Le son **devrait** en principe être stocké sous une forme non comprimée obtenue à partir des équipements de numérisation, sans traitement ultérieur tel que la réduction de bruit. Les enregistrements audio **devraient** être créés et stockés sous un format non comprimé tel que WAV de *Microsoft* ou AIFF d'*Apple*. Un son stéréo de 24 bits à un taux d'échantillonnage de 48/ 96 KHz **devrait** être utilisé pour des copies de référence. Ce taux est recommandé par la Société d'ingénierie du son(AES) et par l'Association internationale des archives sonores et audiovisuelles (IASA).

Les enregistrements audio **peuvent** être créés et stockés sous des formats comprimés tels que MP3, WMA, *RealAudio*, ou *Sun AU*.

5.2. Le choix des supports de stockage

Les supports de stockage numérique ont des spécifications logicielles et matérielles différentes en matière d'accès. Leurs caractéristiques de stockage et de gestion diffèrent également. Deux types de difficultés peuvent entraver l'accès continu aux supports numériques :

- La détérioration physique ou l'endommagement du support lui-même,
- Le changement technologique conduisant à l'obsolescence de l'infrastructure logicielle et matérielle nécessaire pour accéder au support.

Les ressources générées au cours du projet de numérisation sont, en général, stockées sur les disques durs d'un ou plusieurs serveurs de fichiers, ainsi que sur des supports de stockage portables. Au moment où ce document est écrit, les types de supports de stockage portables les plus communément utilisés sont la bande magnétique et le support optique (CD-R et DVD).

Le support portable choisi **devrait** être de bonne qualité, de marques et de fournisseurs réputés. Les nouveaux devraient faire l'objet de test pour en identifier les défauts. Les supports **devraient** être manipulés, utilisés et stockés selon les instructions de leur fournisseur.

Les équipes projet **devraient** réaliser des copies de toutes leurs ressources numériques, – des enregistrements de métadonnées et des objets numérisés – sur deux types différents de supports de stockage. Au moins une copie **devrait** être conservée dans un endroit distinct de leur localisation principale afin de garantir leur sauvegarde au cas où le site principal serait endommagé. Tous les transferts vers des supports portables **devraient** être consignés. (Source : Guide des bonnes pratiques Minerva, *Digital Preservation Coalition*)

Les supports **devraient** être “ renouvelés²⁴ ” (c'est-à-dire que les données devraient être copiées sur un nouveau support, de même nature que le premier) de manière régulière tout au long de la vie du support. L'activité de renouvellement **devrait** être consignée. (Source : Minerva GPH, *Digital Preservation Coalition*)

²⁴ ou “rafraichis”

Gestion de la conservation des ressources numériques de la Digital Preservation Coalition
<<http://www.dpconline.org/graphics/handbook/>>

TASI: Conseil: *Utiliser les CD-R et les DVD-R pour la conservation numérique* du Service britannique de conseil technique pour les images
<<http://www.tasi.ac.uk/advice/delivering/cdr-dvdr.html>>

En français

Le guide de la numérisation des collections de la sous-direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche
<<http://www.sup.adc.education.fr/bib/Acti/Num/notech.htm#supports>>

De la théorie à la pratique : didacticiel de l'imagerie numérique, didacticiel de l'Université de Cornell
<<http://www.library.cornell.edu/preservation/tutorial-french/preservation/preservation-01.html>>

5.3. Les stratégies de préservation

Il existe trois approches techniques principales à la conservation numérique : la conservation de technologie, l'émulation de technologie et la migration de données. Les deux premières approches sont centrées sur la technologie utilisée pour accéder à l'objet, soit en maintenant le matériel et le logiciel d'origine, soit en utilisant la technologie actuelle pour reproduire l'environnement original. Les stratégies de migration sont orientées vers la maintenance des objets numériques sous une forme accessible par la technologie actuelle. Dans cette optique, le travail sur les "archives persistantes"²⁵ basé sur l'articulation des caractéristiques essentielles des objets à préserver, sera très utile.

Dans ce scénario, les objets sont régulièrement transférés d'un environnement technique à un autre, plus récent, tout en conservant le contenu, le contexte, la convivialité et la fonctionnalité de l'original dans la mesure où cela est possible. Certaines migrations peuvent nécessiter la copie de l'objet d'un support ou matériel vers un nouveau support ou matériel et/ou la transformation de l'objet d'un format à un autre. Certaines migrations peuvent ne nécessiter qu'une transformation de format relativement simple, une migration vers un environnement très différent peut engager un processus complexe avec un effort de conception considérable.

Les équipes projet **devraient** être conscientes des exigences d'une stratégie de conservation basée sur la migration et **devraient** développer des politiques et des recommandations favorisant sa mise en oeuvre.

Dans une stratégie de conservation basée sur la migration des données, le transfert des métadonnées peut être une phase critique (Voir). Les métadonnées sont nécessaires à la gestion des objets et au processus de migration. La migration conduit inévitablement, au moins à long terme, à certaines modifications ou pertes dans les fonctionnalités d'origine. Lorsque celles-ci contribuaient à l'interprétation de l'objet, les utilisateurs font appel aux métadonnées relatives à la migration, à celles relatives à l'objet d'origine et à ses transformations pour appréhender la fonctionnalité fournie dans l'environnement technologique d'origine.

En français

Les principes techniques et organisationnels de la préservation des documents numériques, article de Catherine Lupovici, journée d'étude 2001 sur la préservation des documents numériques
<http://www-sv.cict.fr/adbu/actes_et_je/je2001/CathLUPO_140901.html>

De la théorie à la pratique : didacticiel de l'imagerie numérique, didacticiel de l'Université de Cornell
<<http://www.library.cornell.edu/preservation/tutorial-french/preservation/preservation-03.html>>

²⁵ Notion développée par un groupe de travail du Forum Grid http://www.gridforum.org/6_DATA/persist.htm

6. Création/saisie de métadonnées

Les métadonnées peuvent être définies littéralement comme “des données sur les données”, mais le terme se rapporte généralement à des données structurées à propos de ressources qui peuvent être utilisées pour effectuer un grand nombre d’opérations. Une ressource peut correspondre à tout document identifié, qu’il soit numérique ou analogique. Parmi les opérations qu’il est possible de réaliser grâce aux métadonnées, la diffusion, la localisation, la gestion (y compris la gestion des droits) et la conservation à long terme d’une ressource. Pour une seule ressource, différents types de métadonnées peuvent être nécessaires afin d’assurer ces différentes fonctions.

6.1. La portée des métadonnées

Il peut être nécessaire de fournir des métadonnées décrivant différentes classes de ressources, notamment :

- les objets physiques numérisés ;
- les objets numériques créés au cours du processus de numérisation et stockés en tant qu’ “enregistrements de référence” ;
- les objets numériques dérivés de ces “enregistrements de référence” pour une diffusion en réseau ;
- de nouvelles ressources créées en utilisant ces objets numériques ;
- des collections de l’un des types d’objets mentionnés ci-dessus.

6.2. Les standards appropriés

Une métadonnée est parfois classée par rapport aux fonctions qu’elle vise à remplir. Dans la pratique, les schémas de métadonnées ont souvent de multiples fonctions et chevauchent les catégories mentionnées ci-dessous.

Les communautés de conservateurs, responsables de la gestion de différents types de ressources ont développé leurs propres standards de métadonnées afin d’effectuer des opérations sur ces ressources. La communauté muséale a créé les standards SPECTRUM et CDWA pour la gestion des objets muséographiques ; la communauté archivistique a développé les standards ISAD(G), ISAAR(CPF) et EAD pour l’administration et la localisation des enregistrements archivistiques ; enfin la communauté des bibliothèques utilise la famille de standards MARC pour représenter et échanger des métadonnées bibliographiques.

L’équipe projet **devrait** connaître les prescriptions de standards de métadonnées spécifiques à sa communauté ou à son domaine.

Les équipes projet **devraient** s’assurer que le(s) schéma(s) de métadonnées adopté(s) est (sont) documenté(s) de manière complète. Cette documentation **devrait** inclure des directives de catalogage détaillées listant les éléments de métadonnées à utiliser et définissant la manière dont ces éléments doivent être employés pour décrire les types de ressources créées et gérées par le projet. Ces directives sont nécessaires même lorsqu’un schéma standard de métadonnées est utilisé, afin d’expliquer comment ce schéma doit être mis en oeuvre dans le contexte spécifique du projet.

Standards:

SPECTRUM, le standard de documentation des musées britanniques, 2^e édition

Institut de Recherche du Getty, *Categories for the Description of Works of Art* (CDWA)

<http://www.getty.edu/research/conducting_research/standards/cdwa/>

International Standard for Archival Description (General) (ISAD(G)). 2^e édition du Comité international des archives

<http://www.ica.org/biblio/isad_g_2e.pdf>

International Standard Archival Authority Record for Corporate Bodies, Persons and Families du Comité international des archives

<http://www.ica.org/biblio/isaar_eng.pdf>

Encoded Archival Description (EAD) maintenu par la Bibliothèque du Congrès

<<http://www.loc.gov/ead/>>

Machine Readable Cataloguing (MARC): MARC 21 maintenu par la bibliothèque du Congrès

<<http://www.loc.gov/marc/>>

Recommandations:

Guide des meilleures pratiques pour les objets numériques des archives en ligne de Californie (OAC BPG DO), Version 1.0

<<http://www.cdlib.org/inside/projects/oac/bpgdo/>>

En français

Normes relatives aux métadonnées de CHIN (Réseau d'information sur le patrimoine canadien)

<http://www.chin.gc.ca/Francais/Normes/metadonnees_multimedia.html>

Métadonnées, normes et standards d'EduNet

<<http://www.educnet.education.fr/dossier/metadata/>>

De la théorie à la pratique : didacticiel de l'imagerie numérique, didacticiel de l'Université de Cornell

<<http://www.library.cornell.edu/preservation/tutorial-french/metadata/metadata-01.html>>

Dossier de la bibliothèque départementale du Val d'Oise sur UNIMARC et ISO 2709

<<http://www.cg95.fr/biblio/bdvo/coinbib/iso2709/>>

Site EAD de la Direction des archives de France

<<http://www.archivesdefrance.culture.gouv.fr/fr/archivistique/DAFlangage.html>>

Systèmes descriptifs de l'Inventaire

<<http://www.culture.gouv.fr/culture/inventai/extranet/sysdesc.htm>>

Site BiblioML

<<http://www.biblioml.org/>>

Méthode d'inventaire informatique des objets beaux-arts et arts décoratifs de la Direction des Musées de France

<<http://www.culture.gouv.fr/documentation/joconde/methode.htm>>

6.2.1. Les métadonnées descriptives

Les métadonnées descriptives sont utilisées pour localiser et interpréter l'objet numérique.

Les équipes projet **devraient** connaître les prescriptions en matière de métadonnées descriptives pour les objets numériques.

Afin que leurs ressources soient localisables par de nombreuses applications et de nombreux services, les équipes projet **doivent** saisir et stocker suffisamment de métadonnées descriptives pour générer un enregistrement de métadonnées pour chaque élément en utilisant l'ensemble d'éléments de métadonnées du "Dublin Core" (DCMES)²⁶ sous sa forme simple, "non-qualifiée". Le DCMES est un schéma de métadonnées descriptives très simple, développé dans le cadre d'une initiative multidisciplinaire et conçu pour localiser des ressources de nombreux domaines. Il définit quinze éléments simples permettant de localiser des ressources multidisciplinaires : Titre, Créateur, Sujet, Description, Editeur, Collaborateur, Date, Type, Format, Identifiant, Source, Langue, Relation, Couverture et Droits.

Cela ne signifie pas que seules les métadonnées en Dublin Core (DC) simple devraient être enregistrées pour chaque élément, mais plutôt que la capacité à fournir des métadonnées en DC simple est une prescription minimale pour la localisation de ressources. Dans la pratique, ces métadonnées en DC simple seront probablement un sous-ensemble d'un ensemble plus important de métadonnées portant sur un même élément.

Pour localiser une ressource au sein du secteur patrimonial, les équipes projet **devraient** également envisager de fournir une description de métadonnées pour chaque élément selon le schéma DC Culture.

Les équipes projet **devraient** être conscientes de tous les autres cas qui exigeraient des métadonnées descriptives et **peuvent** avoir besoin de saisir et de stocker des métadonnées descriptives supplémentaires.

Standards:

Dublin Core Metadata Element Set, Version 1.1 du Comité Dublin Core

<<http://dublincore.org/documents/dces/>>

Le *DC.Culture* du projet Minerva

<<http://www.minervaeurope.org/DC.Culture.htm>>

Recommandations:

Utiliser le Dublin Core sur le site du Dublin Core

<<http://dublincore.org/documents/usageguide/>>

En français

Le guide d'utilisation du Dublin Core de l'Université de Laval

<<http://www.bibl.ulaval.ca/DublinCore/usageguide-20000716fr.htm>>

Dublin Core et ressources culturelles, article de Muriel Foulonneau dans *Culture & Recherche*

<<http://www.culture.gouv.fr/culture/editions/r-cr/crf97.pdf>>

En français

Le guide de la numérisation des collections de la sous-direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche

<<http://www.sup.adc.education.fr/bib/Acti/Num/notech.htm#1.6.1.>>

6.2.2. Les métadonnées administratives

Les métadonnées administratives permettent de gérer un objet numérique et de fournir des informations sur sa création et sur toute contrainte liée à son utilisation. Elles peuvent inclure :

- Les métadonnées techniques, décrivant les caractéristiques techniques d'une ressource numérique ;
- Les métadonnées sur le source, décrivant l'objet à partir duquel la ressource numérique a été produite ;
- Les métadonnées de provenance numérique, décrivant l'historique des opérations effectuées sur un objet numérique depuis sa création/saisie ;

²⁶ *Dublin Core Metadata Element Set*

- Les métadonnées de gestion des droits, décrivant les droits d'auteur, les restrictions d'utilisation et les accords de licence auxquels est soumise pour l'utilisation de la ressource.

Les métadonnées techniques comprennent des informations qui ne peuvent être saisies efficacement que dans le cadre du processus de numérisation : par exemple, des informations concernant la nature de l'objet source, l'équipement de numérisation utilisé et ses paramètres (formats, types de compression, etc.) et portant sur les agents responsables du processus de numérisation. Il est parfois possible de générer certaines de ces métadonnées à partir du logiciel de numérisation utilisé.

Il n'existe cependant pas de standard unique pour ce type de métadonnées. Pour les images, un comité de l'Organisation nationale des standards sur l'information (NISO)²⁷, aux États-Unis, a produit une version de travail d'un dictionnaire de métadonnées techniques pour images numériques fixes.

Un ensemble plus compact de seize éléments de métadonnées a été publié en 1998 par un groupe de travail sur la conservation des métadonnées, constitué par le Groupe des bibliothèques de recherche (RLG)²⁸.

Les équipes projet **devraient** connaître les prescriptions concernant les métadonnées administratives pour les objets numériques.

Les équipes projet **doivent** saisir et stocker suffisamment de métadonnées administratives pour la gestion de leurs ressources numériques.

Standards:

Dictionnaire de métadonnées techniques pour images numériques fixes du NISO, organisme de standardisation américain

<<http://www.niso.org/pdfs/DataDict.pdf>>

Groupe de travail du RLG (Groupe des bibliothèques de recherche) sur les questions de conservation des métadonnées

<<http://www.rlg.org/preserv/presmeta.html>>

En français

Normes relatives aux métadonnées de CHIN (Réseau d'information sur le patrimoine canadien)

<http://www.chin.gc.ca/Francais/Normes/metadonnees_droits.html>

6.2.3. Les métadonnées de conservation

Les métadonnées de conservation contiennent les informations nécessaires à l'usage d'une ressource à long terme. Un ensemble de 16 métadonnées de base pour accompagner la conservation a été publié en 1998 par un groupe de travail du Groupe des bibliothèques de recherche (RLG).

Le modèle de référence pour un système d'archivage ouvert (OAIS) fournit un cadre conceptuel pour le développement et la comparaison des archives numériques. Il propose à la fois un volet fonctionnel soulignant les opérations nécessaires au développement d'une archive et un volet informatif décrivant les métadonnées devant accompagner la création d'une archive.

Un groupe de travail OCL/RLG travaillant sur les métadonnées de conservation²⁹ a utilisé le cadre du modèle OAIS. Il a pu ainsi développer deux aspects du modèle OAIS portant directement sur les métadonnées de conservation : Information sur le contenu et Conservation, description, information.

²⁷ *National Information Standards Organization*

²⁸ *Research Libraries Group*

²⁹ Se reporter à http://www.rlg.org/en/pdfs/pm_framework.pdf

Standards

Reference Model for an Open Archival Information System (OAIS) du Comité consultative pour les systèmes de données spatiaux

<<http://www.ccsds.org/documents/650x0b1.pdf>>

Structure de métadonnées pour la préservation des objets numériques du Centre coopératif des bibliothèques OCLC

<http://www.oclc.org/research/projects/pmwg/pm_framework.pdf>

En français

Guide de conservation des informations et des documents numériques de l'ADAE (Agence pour le développement de l'administration électronique)

<http://www.adae.pm.gouv.fr/article.php?id_article=7>

L'information technique nécessaire à la préservation à long terme des documents numériques, Article de Julien Masanès de la Bibliothèque Nationale de France

<http://bibnum.bnf.fr/conservation/infopreservation_fr.pdf>

Normes ouvertes en technologies de l'information, sur le site du Québec sur les Autoroutes de l'information

<<http://www.autoroute.gouv.qc.ca/publica/normes/norme35.htm>>

6.2.4. Les métadonnées de structure

Les métadonnées de structure décrivent les relations logiques ou physiques entre les différentes parties d'un objet composé. Par exemple, un livre physique est constitué d'une séquence de pages. Le processus de numérisation peut générer un certain nombre de ressources numériques séparées, éventuellement une image par page, mais le fait que ces ressources forment une séquence et que cette séquence constitue un objet composé est essentiel à leur utilisation et à leur interprétation.

Le *Standard d'Encodage et de Transmission de Métadonnées (METS)*³⁰ est un format d'encodage pour les métadonnées descriptives, administratives et de structure. Il est conçu pour traiter à la fois la gestion et la diffusion d'objets numériques ainsi que leur échange entre différents systèmes d'informations.

Les spécifications de l'*IMS Content Packaging* définissent une manière de décrire la structure des ressources éducatives composées et de les organiser.

Les équipes projet **devraient** connaître les spécifications des métadonnées de structure pour des ressources numériques, le rôle de METS pour le "conditionnement" de métadonnées et d'objets numériques et celui des spécifications de l'*IMS Content Packaging* pour l'échange de ressources éducatives réutilisables.

Standards:

Metadata Encoding and Transmission Standard (METS) de la Bibliothèque du Congrès

<<http://www.loc.gov/standards/mets/>>

IMS Content Packaging de l'organisation IMS de standardisation des systèmes éducatifs

<<http://www.imsproject.org/content/packaging/>>

6.2.5. La description au niveau des collections

Une ressource numérique est créée non pas de manière isolée mais comme partie d'une collection. Elle devrait être considérée dans le contexte de cette collection et de son développement. En effet,

³⁰ *Metadata Encoding and Transmission Standard*

les collections sont perçues comme des composants autour desquels de nombreux types de services numériques peuvent être élaborés.

Les collections **devraient** être décrites de telle sorte qu'un utilisateur puisse identifier les caractéristiques importantes de la collection et que les collections s'intègrent à l'ensemble plus large des collections numériques existantes et aux services numériques les utilisant.

Les équipes projet **devraient** démontrer leur connaissance des initiatives qui visent à améliorer la diffusion et la localisation de collections, telles que les inventaires de contenus numérisés et d'activités de numérisation, couvrant des programmes, des communautés, des secteurs, des domaines, nationaux ou internationaux. Les équipes projets **devraient** mettre leurs métadonnées à disposition de ces services lorsque cela est pertinent.

Les équipes projet **devraient** fournir des descriptions au niveau collection en utilisant un schéma de métadonnées approprié. Les équipes projet **devraient** connaître le schéma de description des collections du programme britannique de recherche pour les bibliothèques (RSLP)³¹, le schéma de description de niveau collection défini par Minerva, ainsi que le tout nouveau profil d'application de description des collections du Dublin Core.

Standards:

Description des collections du RSLP (Research Support Libraries Programme)

<<http://www.ukoln.ac.uk/metadata/rslp/>>

Minerva: Rapport D3.2: Inventaires, localisation de contenus numérisés & questions multilingues: *Etude de faisabilité pour une plate-forme commune*

<http://www.minervaeurope.org/intranet/reports/D3_2.pdf>

Profil d'application de description des collections du Dublin Core

<<http://dublincore.org/groups/collections/>>

Recommandations:

Rapport D3.1: Inventaires, localisation de contenus numérisés & questions multilingues: *Analyse du contenu existant*

<http://www.minervaeurope.org/intranet/reports/D3_1.pdf>

Collection description focus de l'UKOLN (université de Bath, Royaume-Uni)

<<http://www.ukoln.ac.uk/cd-focus/>>

En français

Normes relatives aux métadonnées de CHIN (Réseau d'information sur le patrimoine canadien)

<http://www.chin.gc.ca/Francais/Normes/metadonnees_description.html>

6.2.6. Les standards terminologiques

La transmission efficace des informations contenues dans les enregistrements de métadonnées exige davantage qu'une conception commune du schéma de métadonnées adopté et des éléments de métadonnées qui le constituent. Elle dépend également d'une compréhension partagée des termes utilisés comme valeurs de ces éléments de métadonnées par l'adoption soit de terminologies communes, soit de terminologies différentes pour lesquelles les relations entre les termes sont clairement définies.

Lorsque cela est possible, les équipes projet **devraient** utiliser des sources terminologiques multilingues reconnues pour fournir des valeurs aux éléments de métadonnées. S'il n'existe aucune terminologie standard, elles **peuvent** avoir recours à des terminologies locales. Lorsque des terminologies locales sont mises en oeuvre, les informations concernant la terminologie, les termes qui la constituent et leur signification **doivent** être publiés.

L'utilisation d'une terminologie, qu'elle soit standard ou spécifique au projet, **doit** être indiquée sans aucune ambiguïté dans les enregistrements de métadonnées.

³¹ *Research Support Libraries Programme*

Les enregistrements de métadonnées de niveau collection **devraient** utiliser les terminologies recommandées avec le schéma Minerva de description de niveau collection.

Standards:

Minerva: Rapport D3.2: Inventaires, localisation de contenus numérisés & questions multilingues:
Etude de faisabilité pour une plate-forme commune

<http://www.minervaeurope.org/intranet/reports/D3_2.pdf>

En français

Normes de vocabulaire et schémas de classification du CHIN (Réseau d'Information sur le patrimoine canadien)

<http://www.chin.gc.ca/Francais/Normes/vocabulaire_intro.html>

Le langage d'indexation de la Bibliothèque nationale de France Rameau

<<http://rameau.bnf.fr/>>

Lexiques et thésaurus de l'Inventaire

<<http://www.culture.gouv.fr/culture/inventai/extranet/lexthe.htm>>

Thésaurus de l'UNESCO

<<http://www.unesco.org/general/fre/whatsnew/thesaurus.html>>

Accéder aux ressources patrimoniales multilingues, compte-rendu du séminaire Minerva du 3 octobre 2003

<http://www.culture.gouv.fr/culture/mrt/numerisation/fr/eeurope/minerva_031003.htm>

7. La publication

Le mode principal d'accès de l'utilisateur final aux ressources est présumé être l'Internet. La préparation pour la publication nécessite le traitement de l' "enregistrement de référence" afin de produire des objets numériques adaptés au contexte de l'Internet. Il s'agit généralement d'en réduire la qualité afin de générer des fichiers de taille convenable pour le transfert à travers les réseaux.

Le son et la vidéo peuvent être disponibles, soit en téléchargement, soit en "flux continu"³², c'est-à-dire qu'au lieu de transférer l'intégralité du fichier avant que la lecture puisse commencer, un petit espace de mémoire tampon est créé sur l'ordinateur de l'utilisateur. Les données sont alors transférées dans cette mémoire tampon. Dès qu'elle est pleine, le fichier passe en mode lecture, alors que les données restantes continuent à être transférées.

Les équipes projets **doivent** prendre en considération le fait qu'il existe des variations au sein :

- des types de matériels et de logiciels client employés par les utilisateurs,
- des largeurs de bandes avec lesquelles les utilisateurs travaillent.

Pour atteindre le plus grand nombre de personnes, les équipes projet **devraient** mettre à disposition leurs ressources dans des formats ou tailles différents ou à des résolutions/taux de bits différents. L'équipe projet **devrait** réviser régulièrement les critères sur lesquels elle décide des formats et des paramètres de publication.

Note. : les recommandations suivantes, portant sur les formats de diffusion, devraient être lues conjointement avec les spécifications sur les formats de fichiers pour le stockage des ressources.

7.1. Les traitements pour la publication

7.1.1. La publication de texte

7.1.1.1. Le codage de caractères

Le codage de caractères utilisé dans les documents textuels **devrait** être transmis dans l'en-tête HTTP et enregistré au sein des documents, en fonction des besoins. Il est nécessaire de souligner que certains protocoles basés sur le XML peuvent rendre obligatoire l'utilisation d'un codage de caractères spécifique. Par exemple le protocole OAI pour la collecte de métadonnées requiert l'utilisation du codage de caractères UTF-8.

Recommandations:

Un tutoriel sur les questions de codage de caractères de Jukka Korpela de l'Université de Tampere
<<http://www.cs.tut.fi/~jkorpela/chars.html>>

En français

Site Babel de l'*Internet Society*
<<http://alis.isoc.org/codage/>>

7.1.1.2. Les formats de documents

Le contenu textuel **doit** être publié en XHTML 1.0 ou HTML 4 (ou versions ultérieures), même si l'utilisation des formats SGML ou XML pour des documents conformes à d'autres DTD ou schémas **peut** parfois être appropriée (source : *NOF-digitise*).

Dans certains cas, la publication sous des formats propriétaires tels que PDF, RTF ou *Microsoft Word* **peut** être appropriée s'ils s'ajoutent au XHTML/HTML, mais les équipes projet **doivent** s'assurer que les questions d'accessibilité ont été traitées (source : *NOF-digitise*).

³² "streaming"

Standards:

HTML 4.01 HyperText Markup Language, site du consortium *World Wide Web*

<<http://www.w3.org/TR/html401/>>

XHTML 1.0 The Extensible HyperText Markup Language, site du consortium *World Wide Web*

<<http://www.w3.org/TR/xhtml1/>>

Autres références:

Portable Document Format (PDF), *Adobe*

<<http://www.adobe.com/products/acrobat/adobepdf.html>>

7.1.2. La diffusion d'images fixes

7.1.2.1. Les images photographiques

Les images **doivent** être fournies sur Internet en JPEG/SPIFF.

Pour offrir un accès adapté au contexte d'utilisation, plusieurs tailles **devraient** être disponibles. Les questions relatives à la propriété intellectuelle peuvent être prise en considération dans la décision de la taille et de la qualité de l'image fournie.

Les icônes **devraient** être diffusées à une résolution de 72 ppp, en utilisant une profondeur de bits de 24 bits couleur ou de 8 bits en échelle de gris avec un maximum de 100-200 pixels pour la plus grande dimension (source : EMII-DCF).

Les images destinées à être présentées en plein écran **devraient** être diffusées à une résolution de 150 ppp, en utilisant une profondeur de bits de 24 bits couleur ou de 8 bits en échelle de gris avec un maximum de 600 pixels pour la plus grande dimension. Cette résolution demeure inférieure aux prescriptions de la reproduction imprimée en haute définition (source : EMII-DCF).

7.1.2.2. Les images graphiques non-vectorielles

Les images **devraient** être publiées sur Internet soit au format *Graphical Interchange Format* (GIF), soit au format *Portable Network Graphics* (PNG).

7.1.2.3. Les images graphiques vectorielles

Les images **devraient** être publiées sur Internet au format *Scalable Vector Graphics* (SVG).

7.1.3. La publication de vidéo

L'équipe projet **devrait** tenir compte du fait que l'accès des utilisateurs à la vidéo peut être limité par la largeur de bande utilisée et il peut être préférable de fournir une gamme de fichiers ou de flux de données de différentes qualités.

7.1.3.1. Le téléchargement

La vidéo destinée à être téléchargée **devrait** être publiée sur Internet au format MPEG-1 ou aux formats propriétaires *Audio Video Interleave* (AVI) de *Microsoft*, *Windows Media Video* (WMV) ou *Quicktime* d'*Apple*. (Source : *NOF-digitise*, EMII-DCF)

Groupe d'Expert sur les Images Animées (MPEG)

<<http://www.cselt.it/mpeg/>>

7.1.3.2. La diffusion en flux continu³³

La vidéo destinée à la diffusion en flux continu **devrait** être publiée sur Internet aux formats *Advanced Streaming Format* (ASF) de *Microsoft*, *Windows Media Video* (WMV) ou *Quicktime* d'*Apple*. (Source : *NOF-digitise*, EMII-DCF)

³³ ou "streaming"

7.1.4. La publication de matériel audio

L'équipe projet **devrait** tenir compte du fait que l'accès des utilisateurs au son peut être limité par la largeur de bande utilisée et qu'il peut être préférable de fournir des fichiers ou des flots de données de différentes qualités.

7.1.4.1. Le téléchargement

Le son **devrait** être disponible sur Internet sous une forme compressée, au format *MPEG Layer 3* (MP3) ou aux formats propriétaires *RealAudio* (RA) ou *Windows Media Audio* (WMA) de *Microsoft*. Un débit de 256 Kbps devrait être utilisé lorsqu'une qualité proche de celle des CD est nécessaire. Un débit binaire de 160 Kbps fournit une bonne qualité (source : EMII-DCF).

Le son **peut** être publié sous des formes non compressées grâce aux formats WAV/AIFF de *Microsoft* ou AU de *Sun*.

7.1.4.2. La diffusion en flux continu

Le son destiné à la diffusion en flux continu **devrait** être publié sur Internet au format *MPEG Layer 3* (MP3) ou aux formats propriétaires *RealAudio* (RA) ou *Windows Media Audio* (WMA) de *Microsoft*.

7.1.5. L'identification

Les ressources numérisées **devraient** être identifiées sans ambiguïté et adressables de manière unique directement à partir du navigateur Internet d'un utilisateur. Il est important, par exemple, que l'utilisateur final ait la possibilité de faire référence directement et en toute confiance à une ressource individuelle, plutôt que de devoir faire un lien sur le site Internet d'un projet entier. Les équipes projet **devraient** utiliser l'*Uniform Resource Identifier* (URI) à cet effet et **devraient** s'assurer que cet URI est raisonnablement pérenne (persistant). Les URI **ne devraient pas** inclure d'informations sur le format de fichier, la technologie du serveur, la structure organisationnelle du fournisseur de services ou toute autre information susceptible de changer au cours de la durée de vie de la ressource. (Source : *NOF-digitise*, *JISC Information Environment*)

Lorsque cela est pertinent, les équipes projet **peuvent** envisager l'utilisation d'identifiants *Digital Object Identifiers* (DOI) ou d'identifiants pérennes basés sur un autre schéma d'identifiants.

Les équipes projet **peuvent** également souhaiter garantir que les ensembles logiques se trouvant au sein des ressources qu'ils fournissent sont adressables de manière unique et pérenne.

Standards:

Uniform Resource Identifiers (URI), site du consortium *World Wide Web*

<<http://www.w3.org/Addressing/>>

Digital Object Identifier (DOI)

<<http://www.doi.org/>>

En français

Le guide de la numérisation des collections de la sous-direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche

<<http://www.sup.adc.education.fr/bib/Acti/Num/notech.htm#toto>>

7.2. Questions relatives à la 3D et à la réalité virtuelle

Les équipes projet qui utilisent des animations de réalité virtuelle (RV) tri-dimensionnelles "*fly throughs*" et des modèles 3D **doivent** prendre en compte les besoins des utilisateurs qui accèdent à leur site en utilisant des ordinateurs classiques et des connexions par modem.

Ces modèles sont généralement utilisés pour la reconstruction d'édifices et d'autres structures ou pour la simulation de zones entières d'un paysage. Les modèles ont traditionnellement été construits

et publiés avec de puissantes stations de travail et c'est toujours le cas pour les plus détaillés. Pour les équipes projet qui doivent diffuser le résultat de leurs travaux auprès d'un large public via l'Internet, des modèles aussi détaillés peuvent n'être d'aucune utilité. Cependant, il est possible d'incorporer des modèles moins complexes à des sites Internet, plus accessibles.

En générant ces modèles, les équipes projet **doivent** savoir que la plus grande partie de leurs utilisateurs, dans un avenir proche, continueront à accéder à l'Internet en utilisant un modem de 56k ou une connexion partagée, plutôt que toute autre technologie à haut débit. Les caractéristiques des ordinateurs utilisés par la plupart des visiteurs sont également susceptibles d'être de qualité significativement moindre que celles des machines sur lesquelles les équipes projet ont généré et testé ces modèles. Les équipes projet **doivent** par conséquent évaluer si leurs modèles sont utilisables dans de telles conditions et **doivent** les tester en utilisant des connexions par modem classiques et des systèmes informatiques similaires à ceux que possèdent les particuliers, les écoles ou les bibliothèques avec les systèmes d'exploitation et les navigateurs les plus courants.

Dans ce domaine, les standards sont en évolution constante, mais les équipes projet **devraient** produire des modèles de RV compatibles avec la spécification X3D.

La RV *QuickTime* d'Apple (QTVR) n'est pas un véritable format d'images en 3D, mais elle offre des fonctionnalités utiles. Les équipes projet qui n'ont pas besoin de l'intégralité des fonctionnalités de X3D **peuvent** envisager d'utiliser plutôt QTVR.

Standards:

Consortium Web3D

<<http://www.web3d.org/>>

X3D du Consortium Web3D

<<http://www.web3d.org/x3d.html>>

QuickTime VR d'Apple

<<http://www.apple.com/quicktime/qtvr/>>

Recommandations:

Guide des bonnes pratiques du service des données archéologiques RV de l'AHDS (Service britannique des Données sur les Arts et les Humanités)

<<http://ads.ahds.ac.uk/project/goodguides/g2gp.html>>

7.3. Les systèmes d'informations géographiques

Les contenus culturels ont souvent une composante géographique sur laquelle on peut baser un outil puissant de classement et de récupération des données. Les Systèmes d'information géographique (SIG) sont des applications spécialement conçues pour conserver, manipuler et récupérer une information géographique. Les SIG sont de plus en plus implémentés dans le domaine du patrimoine culturel.

Toutefois, il n'est pas nécessaire d'installer et de maintenir un SIG si l'on veut conserver des informations géolocalisées ou inclure un plan sur le site web d'un projet. En effet, une partie des données relatives à la localisation géographique peuvent être conservées dans des bases de données traditionnelles et des plans simples de localisation peuvent être créés de diverses façons. Les équipes projet **doivent** s'assurer qu'elles pourront supporter l'implémentation d'un SIG dans le futur, même si cela n'est pas prévu dans le projet. Le protocole de collecte des métadonnées OAI-PMH (utilisant un schéma DC. Culture) pourra alors permettre la présentation des données par un SIG externe (voir section 8.1)

Les équipes projet qui désirent employer des Systèmes d'information géographique (SIG) **doivent** obtenir les autorisations appropriées pour utiliser les données des cartes réalisées par des tiers et s'assurer que les licences sont également valables pour la diffusion sur l'Internet.

Les équipes projet **doivent** s'assurer que les ensembles de données combinés afin de fournir un service sont d'une échelle et d'une résolution similaires et peuvent être utilisés ensemble et de cette manière.

Les produits commerciaux de SIG sélectionnés **devraient** être conformes aux standards industriels émergents de l'*Open GIS Consortium*.

Lors de l'enregistrement de données spatiales, les équipes projet **doivent** utiliser un système de référence de coordonnées normalisé adapté.

Les équipes projet **doivent** utiliser des standards nationaux adaptés à l'enregistrement d'adresses physiques (rues).

Standards:

Consortium OpenGIS
<<http://www.opengis.org/>>

Recommandations:

Guide des bonnes pratiques du service des données archéologiques SIG du Service britannique des Données sur les Arts et les Humanités
<<http://ads.ahds.ac.uk/project/goodguides/gis/>>

7.4. Les sites Internet

Les ressources d'un projet **doivent** être accessibles par les navigateurs Internet, généralement en utilisant de l'HTML ou du XHTML ainsi que le protocole HTTP 1.1. Si d'autres protocoles sont mis en oeuvre (par exemple le Z39.50), des passerelles **doivent** être disponibles pour permettre d'accéder à ces ressources via un navigateur.

Les équipes projet **devraient** faire en sorte que les sites Internet de leurs projets soient les plus disponibles possible. Des périodes significatives d'indisponibilité **devraient** être rapportées au programme de financement.

Standards:

Hypertext Transfer Protocol, HTTP/1.1, site du consortium *World Wide Web*
<<http://www.w3.org/Protocols/HTTP/>>

7.4.1. L'accessibilité

Les ressources des projets **doivent** être accessibles par de nombreux navigateurs, systèmes matériels, programmes automatisés et utilisateurs finaux.

Les sites Internet **doivent** être accessibles à de nombreux navigateurs et matériels (par exemple des assistants numériques personnels – PDA – et des PC). Les sites Internet **doivent** être utilisables par des navigateurs compatibles avec les recommandations du W3C telles que le HTML/XHTML, les CSS (feuilles de style)³⁴ et le DOM (modèle objet de document)³⁵. Les équipes projet qui utilisent des formats de fichiers propriétaires et des technologies de greffons de navigateurs³⁶ **doivent** s'assurer que leur contenu reste toujours utilisable par les navigateurs qui ne disposent pas des greffons. Par conséquent, l'utilisation de technologies telles que *Javascript* et *Flash* de *Macromedia* pour la navigation du site doit être envisagée avec précaution.

³⁴ CSS - Cascading Style Sheets

³⁵ DOM - Document Object Model

³⁶ "plug-in"

L'apparence d'un site Internet **devrait** être contrôlée par l'utilisation de feuilles de style conformes à l'architecture et aux recommandations sur l'accessibilité du W3C. La dernière version des feuilles de style (CSS) recommandée par le W3C (actuellement CSS 2) devrait être utilisée, même si, n'étant pas encore implémentée dans tous les navigateurs, toutes les caractéristiques définies par la CSS 2 ne sont pas utilisables.

Les équipes projet **doivent** appliquer les recommandations WAI (Initiative pour l'Accessibilité au Web)³⁷ du W3C et par conséquent assurer un degré élevé d'accessibilité aux personnes souffrant de certaines incapacités. Les projets **doivent** être conformes au niveau A du WAI et les équipes projet **devraient** avoir pour objectif d'être conformes au niveau AA du WAI.

Standards:

Cascading Style Sheets (CSS), niveau 2, site du consortium World Wide Web

<<http://www.w3.org/TR/REC-CSS2/>>

Web Content Accessibility Guidelines (WCAG) 1.0, site du consortium World Wide Web

<<http://www.w3.org/TR/WCAG10/>>

Recommandations:

Web Accessibility Initiative (WAI), site du consortium World Wide Web

<<http://www.w3.org/WAI/>>

RNIB (Institut Royal britannique des Aveugles): Accessible Web Design

<<http://www.rnib.org.uk/digital/hints.htm>>

Service en ligne Watchfire Bobby

<<http://bobby.watchfire.com/>>

Site de Jakob Nielsen, useit.com " Usable Information Technology "

<<http://www.useit.com/>>

En français

L'accessibilité vue par Visual Friendly, article publié sur le site de l'ADAE (Agence pour le développement de l'administration électronique)

<http://www.adae.pm.gouv.fr/article.php3?id_article=146>

7.4.2. La sécurité

Les machines utilisées pour la publication des projets **doivent** être les plus sécurisées possible. Les conseils formulés dans les manuels de systèmes d'exploitation concernant la sécurité **doivent** être suivis. Tous les correctifs de sécurité connus **doivent** être appliqués.

Les machines **devraient** être configurées de manière à ne faire fonctionner que le nombre minimal de services réseau. Les machines **devraient** être placées derrière un pare-feu lorsque cela est possible, avec un accès à l'Internet uniquement sur les ports nécessaires à la publication du projet.

Les équipes projet **devraient** connaître les règlements fournis par l'ISO/IEC 17799:2000. La gestion et l'utilisation des informations personnelles **doivent** être conformes à la législation nationale en vigueur.

Lorsque des informations sensibles passent d'un client à un serveur via le réseau, les équipes projet **doivent** utiliser le protocole *Secure Sockets Layer (SSL)* pour crypter les données. Cela concerne notamment les noms d'utilisateurs et les mots de passe, les informations portant sur les cartes de crédit et toute autre donnée personnelle. Il est à noter que l'utilisation de SSL permet à l'utilisateur final d'avoir davantage confiance en l'authenticité du service.

³⁷ WAI - Web Accessibility Initiative

Standards:

Standard international ISO/IEC 17799:2000 *Règlement pour la gestion des informations de sécurité* du NIST (institut national américain des standards et technologies)

<<http://csrc.nist.gov/publications/secpubs/otherpubs/revision-faq.pdf>>

Secure Sockets Layer (SSL) 3.0

<<http://wp.netscape.com/eng/ssl3/>>

Recommandations:

Introduction au *Secure Socket Layer (SSL)* sur le site de *Netscape*

<<http://developer.netscape.com/docs/manuals/security/sslin/index.htm>>

En français

Les protocoles de sécurité sur Internet, article de l'ADAE (Agence pour le développement de l'administration électronique)

<http://www.adae.gouv.fr/article.php3?id_article=106>

SSL, un protocole pour le commerce électronique, article de l'ADAE (Agence pour le développement de l'administration électronique)

<http://www.adae.gouv.fr/article.php3?id_article=111>

Sécurité, garantir de bonnes pratiques, article de *OINet*

<<http://www.01net.com/article/225266.html>>

7.4.3. L'authenticité

Les noms de domaine spécifiques au projet **devraient** être enregistrés auprès du système de nom de domaine (DNS)³⁸. Le nom de domaine fait partie de la “marque” du projet et permettra aux utilisateurs finaux d'identifier l'authenticité du contenu publié. Les noms de domaine **devraient** par conséquent être clairement étiquetés soit avec le nom du projet, soit avec le nom de l'organisation en charge du projet.

Dans certaines situations, il peut être nécessaire de sécuriser la connexion réseau entre le client et le serveur en utilisant *Sockets Layer (SSL)* pour que les utilisateurs soient sûrs qu'ils échangent des informations avec le site Internet du projet souhaité.

Les musées qui possèdent des sites Internet **devraient** envisager l'enregistrement d'un “.museum” comme nom de domaine de premier niveau afin d'améliorer la diffusion de leurs services et d'indiquer que leurs sites sont associés à de véritables musées.

Recommandations:

Répertoire de ressources du *Domain Names System (DNS)*

<<http://www.dns.net/dnsrd/>>

Site *Dot Museum* de l'Association de gestion du domaine des musées (MuseDoma)

<<http://about.museum/>>

7.4.4. L'authentification des utilisateurs

Certaines équipes projet **peuvent** souhaiter limiter l'accès à certaines parties de leurs ressources (par exemple aux images ou cartes en très haute résolution) à des utilisateurs spécifiques. L'authentification des utilisateurs est une fonctionnalité importante pour s'assurer que seuls les utilisateurs légitimes accèdent aux ressources en ligne du projet.

³⁸ DNS - *Domain Name System*

Si les équipes projet décident de mettre en place l'authentification des utilisateurs pour certaines ressources, celle-ci **devrait** être basée sur une combinaison identifiant et mot de passe. Dans le cas de projets Internet, l'authentification basique http³⁹ **doit** être utilisée pour transférer la combinaison identifiant/mot de passe du navigateur au serveur.

Dans certains cas, l'authentification basée sur le numéro IP (qui compare l'adresse IP du client à une liste d'adresses IP connues) **peut** être une bonne alternative aux identifiants et mots de passe. Cependant, l'utilisation de cette méthode d'authentification est fortement déconseillée puisque de plus en plus de fournisseurs de services Internet utilisent des adresses IP dynamiques, ce qui rend très difficile la gestion d'une liste d'adresses IP validées. De plus, la prise en compte des utilisateurs mobiles et des utilisateurs se trouvant derrière un pare-feu compromettra davantage encore la gestion de l'authentification par le numéro IP.

Les équipes projet **peuvent** décider d'avoir recours à des services d'authentification par un " tiers de confiance " ou " tiers certificateur " pour gérer les identifiants et les mots de passe, si cela est pertinent.

Standards:

Hypertext Transfer Protocol, HTTP/1.1, site du consortium World Wide Web
<<http://www.w3.org/Protocols/HTTP/>>

7.4.5. Les indicateurs de performance

Il est possible d'utiliser des indicateurs de performance pour fournir des mesures objectives sur l'utilisation d'un service Internet et des informations sur l'impact du projet de numérisation. L'indicateur de performance le plus connu utilise des fichiers de consignation⁴⁰ de serveur Internet. L'analyse des fichiers de consignation du serveur peut fournir des informations intéressantes sur le développement d'un service et sur ses modèles d'utilisation, bien que les rapports doivent être interprétés avec précaution.

Les équipes projet **doivent** établir des statistiques sur l'usage des sites Internet et **devraient** les utiliser de manière appropriée pour analyser l'usage des ressources numérisées.

³⁹ *HTTP Basic Authentication*

⁴⁰ "log files"

8. La diffusion des ressources

Les collections créées dans le cadre d'un projet de numérisation font partie d'un corpus de ressources plus large. Afin de faciliter la localisation de ressources au sein de ce corpus, les équipes projet **doivent**, pour chaque collection, penser à mettre les métadonnées sur leurs ressources à disposition de manière à ce qu'elles puissent être utilisées par d'autres applications et services.

L'identification des métadonnées nécessaires et la manière dont elles devraient être exposées dépend de la nature des ressources créées et des applications et services avec lesquels ces métadonnées sont partagées.

Les équipes projet **devraient** exposer un ou plusieurs enregistrement(s) de métadonnées de niveau collection décrivant chacune de leurs collections. Les équipes projet **peuvent** également diffuser des enregistrements de métadonnées décrivant les ressources numériques individuelles au sein de leur(s) collection(s).

A la fois les enregistrements de métadonnées de niveau collection et ceux de niveau élément **devraient** inclure les termes et conditions d'utilisation de la ressource.

Afin de faciliter les échanges potentiels et l'interopérabilité entre les services, les équipes projet **devraient** être capables de fournir des descriptions de niveau élément sous la forme de simples enregistrements de métadonnées de type Dublin Core non qualifié et **peuvent** fournir des descriptions de niveau élément conformément au schéma DC Culture.

Lorsqu'il s'agit de "ressources éducatives" ou de ressources présentant un intérêt pour les communautés de l'éducation et de l'enseignement, les équipes projet **devraient** également penser à fournir des descriptions sous la forme de métadonnées d'objets éducatifs IEEE⁴¹ (format LOM).

Les équipes projet **devraient** connaître toutes les spécifications additionnelles qui pourraient être imposées dans leurs contextes spécifiques (ex. : les standards de métadonnées gouvernementaux).

Les équipes projet **devraient** se tenir informées de toutes les questions légales de nature à affecter leurs enregistrements de métadonnées.

Standards:

Dublin Core Metadata Element Set, Version 1.1 du Comité Dublin Core

<<http://dublincore.org/documents/dces/>>

DC.Culture du projet Minerva

<<http://www.minervaeurope.org/DC.Culture.htm>>

IEEE Learning Object Metadata, de l'association américaine de standardisation IEEE

<<http://ltsc.ieee.org/wg12/>>

8.1. La collecte de métadonnées

Les équipes projet **devraient** être conscientes que le Protocole de collecte de métadonnées de l'Initiative des archives ouvertes (OAI-PMH) est un moyen de rendre leurs métadonnées disponibles aux fournisseurs de services.

Les équipes projet **peuvent** envisager de mettre leurs métadonnées à disposition pour être collectées en installant des entrepôts de métadonnées conformes à l'OAI. Les équipes projet qui

⁴¹ *IEEE Learning Object Metadata*

mettent en place ces entrepôts **devraient** envisager d'inclure une formulation des droits dans leurs métadonnées afin de garantir leurs droits de propriété sur leurs métadonnées.

Standards:

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) de l'Initiative des archives ouvertes

<<http://www.openarchives.org/>>

Recommandations:

FAQ sur l'OAI du JISC (comité commun des systèmes d'information britannique)

<<http://www.ukoln.ac.uk/distributed-systems/dner/arch/faq/oai/>>

[Open Archives Initiative - Protocol for Metadata Harvesting: Practices of the cultural heritage actors by Muriel Foulonneau and David Dawson on OAForum Website

<http://www.oaforum.org/otherfiles/oaf_d48_cser3_foullonneau.pdf>]

En français

Le protocole OAI-PMH, une opportunité pour le patrimoine numérique, article de Muriel Foulonneau sur le site de la numérisation du ministère de la culture et de la communication

<http://www.numerique.culture.fr/fr/technique/documents/guide_oai.pdf>

Présentation de l'OAI par Martin Sévigny et Christine Bourgoïn sur le site de la société AJLSM

<<http://www.ajlsm.com/projets/pp/technique/oai.html>>

8.2. La recherche en architecture distribuée

Les équipes projet **peuvent** avoir besoin de connaître le Z39.50, un protocole réseau qui permet d'effectuer des recherches dans des bases de données hétérogènes (et généralement distantes) et de retrouver des données via une interface utilisateur. Le Z39.50 est la plupart du temps, utilisé pour rechercher des enregistrements bibliographiques, bien qu'il existe également certaines implémentations non-bibliographiques. Les équipes projet qui utilisent le Z39.50 **doivent** connaître le "profil de Bath" et sa pertinence pour l'interopérabilité multi-sectorielle.

Les équipes projet **peuvent** également avoir besoin de connaître le protocole de services Web de recherche/localisation (SRW/SRU)⁴², élaboré à partir de la sémantique du Z39.50 pour offrir une fonctionnalité similaire en utilisant des technologies de service Web.

⁴² Search/Retrieve Web Service

Standards:

Agence de maintenance du Z39.50 à la Bibliothèque du Congrès
<<http://www.loc.gov/z3950/agency/>>

Profil de Bath (profil Z39.50), sur le site de la bibliothèque nationale du Canada
<<http://www.nlc-bnc.ca/bath/tp-bath2-e.htm>>

SRW: *Search/Retrieve Web Service* sur le site de la Bibliothèque du Congrès
<<http://lcweb.loc.gov/z3950/agency/zing/srw/>>

Recommandations:

Z39.50 pour tous, article de Paul Miller dans la revue britannique *Adriadne*
<<http://www.ariadne.ac.uk/issue21/z3950/>>

En français

Support de cours de Dominique Lahary de l'Université Paris X sur la norme Z39.50
<<http://membres.lycos.fr/vacher/profess/cours/mediadix/z3950/>>

Ressources de la bibliothèque nationale du Canada sur le profil de Bath en français
<<http://www.nlc-bnc.ca/bath/ap-bath-f.htm>>

8.3. Les alertes

Les équipes projet **peuvent** avoir besoin de connaître la famille de spécifications Sommaire de Site RDF (ou Riche) (RSS)⁴³. RSS fournit un mécanisme d'échange de métadonnées descriptives, généralement sous la forme d'une liste d'éléments, chacun contenant une brève description textuelle ainsi qu'un lien vers la source de données d'origine.

Standards:

RDF Site Summary (RSS) 1.0, spécification publiée par le groupe RSS-DEV
<<http://purl.org/rss/1.0/spec>>

RSS 2.0 par le *Berkman Center for Internet & Society* à l'École de droit de Harvard
<<http://blogs.law.harvard.edu/tech/rss>>

Recommandations:

Contenu mutualisé: c'est plus que quelques formats de fichiers, article de Paul Miller dans la revue britannique *Adriadne*
<<http://www.ariadne.ac.uk/issue35/miller/>>

8.4 Les services WEB

Les équipes projet **devraient** connaître les spécifications des services web et notamment de la version 1.2 de SOAP⁴⁴ et du Langage de description des services web (WSDL)⁴⁵.

⁴³ *RDF (ou Rich) Site Summary*

⁴⁴ <http://ws.apache.org/soap/>

Pour les services en réseau non concernés par les protocoles spécifiques mentionnées précédemment, on **devrait** tout de même considérer l'utilisation de SOAP, bien que l'utilisation du style d'architecture REST combiné aux requêtes http 1.1 GET ou POST pour récupérer des documents XML **puisse** être appropriée.

Il se **peut** aussi que les équipes projet aient besoin de compétences dans la spécification Description, découverte et intégration universelle (UDDI)⁴⁶.

Standards:

SOAP Version 1.2 Part 1: Messaging Framework

<<http://www.w3.org/TR/soap12-part1/>>

Web Services Description Language (WSDL) 1.1

<<http://www.w3.org/TR/wsdl>>

Hypertext Transfer Protocol, HTTP/1.1

<<http://www.w3.org/Protocols/HTTP/>>

Guide:

SOAP Version 1.2 Part 0: Primer

<<http://www.w3.org/TR/soap12-part0/>>

En français :

SOAP version 1.2 Partie 0

<<http://www.w3.org/2002/07/soap-translation/soap12-part0.html> >

8.5 Cadre de description des données (RDF) et ontologies Web

Les équipes projet **pourront** tirer profit des capacités de partage et de réutilisation des données sur le Web générées par les spécifications du cadre de descriptions des données (RDF). RDF fournit un modèle simple de description des ressources. Cette partie sera développée ultérieurement afin de fournir des modèles d'interrogations pour les bases de données RDF.

Les équipes projet **peuvent** utiliser les ontologies basées sur le Langage web ontologique (OWL). OWL se sert des schémas RDF pour enrichir le vocabulaire descriptif portant sur les propriétés et les classes afin de faciliter la définition des concepts de base et de permettre leur traitement automatisé.

Les équipes projet **peuvent** explorer les possibilités d'interopérabilité sémantique offertes par des ontologies tel le Modèle conceptuel de référence (CRM) du CIDOC ou le modèle ontologique ABC Ontologie développé par le projet Harmony.

Le CRM propose un cadre sémantique commun et extensible adaptable à tout système d'information d'une institution patrimoniale et pouvant faire le lien entre différentes sources d'information.

⁴⁵ <http://www.w3.org/TR/wsdl>

⁴⁶ <http://www.uddi.org/>

Standards:

Cadre de description des données (RDF)

<<http://www.w3.org/RDF/>>

Langage ontologique Web (OWL)

<<http://www.w3.org/2001/sw/WebOnt/>>

CIDOC Modèle conceptuel de référence (CRM)

<<http://cidoc.ics.forth.gr/>>

Guide:

RDF

<<http://www.w3.org/TR/rdf-primer/>>

OWL présentation

<<http://www.w3.org/TR/owl-features/>>

Ontologie et Modèle ABC

<<http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/>>

9. La réutilisation pour d'autres objectifs

Il est possible que les utilisateurs veuillent re-conditionner les ressources qui ont été créées par des projets de numérisation et les représenter pour un autre usage. Afin de faciliter cette réutilisation, l'implémentation de standards est importante.

9.1. La création de ressources éducatives

Les équipes projet **devraient** envisager la réutilisation potentielle des ressources qu'ils créent et prendre en compte le fait que les utilisateurs finaux ou des tiers peuvent désirer extraire des éléments d'une ressource donnée et les re-conditionner avec des parties d'autres ressources, provenant de leurs propres collections et d'autres sources.

Ce cas est particulièrement courant dans le secteur éducatif. Dans la communauté éducative mondiale, un certain nombre d'initiatives sont en cours pour la création d'outils de gestion de ressources éducatives. Une partie de ces efforts se concentre sur la description de contenu tel que celui créé par des programmes de numérisation.

Les équipes projet qui développent des ressources éducatives **doivent** connaître le standard de métadonnées pour les objets éducatifs de l'IEEE (LOM)⁴⁷ et **devraient** envisager de fournir des descriptions LOM de leurs ressources éducatives.

L'équipe projet **devrait** se tenir informée du travail du consortium IMS de développement des spécifications pour l'interopérabilité dans des systèmes technologiques éducatifs. Les équipes projet qui conçoivent des ressources éducatives **devraient** envisager l'utilisation de l'*IMS Content Packaging* pour faciliter l'accès à ces ressources par des utilisateurs d'environnements éducatifs virtuels.

Standards:

IEEE Learning Object Metadata de l'association américaine de standardisation IEEE

<<http://ltsc.ieee.org/wg12/>>

IMS Global Learning Consortium, Inc. de l'organisation IMS de standardisation des systèmes éducatifs

<<http://www.imsproject.org/>>

IMS Content Packaging de l'organisation IMS de standardisation des systèmes éducatifs

<<http://www.imsproject.org/content/packaging/>>

En français

Normes relatives aux métadonnées de CHIN (Réseau d'information sur le patrimoine canadien)

<http://www.chin.gc.ca/Francais/Normes/metadonnees_education.html>

Séminaire *Normes et standards pour les activités numériques dans l'enseignement*, de Lyon et Saint-Clément-les-Places, 9 et 10 octobre 2003, EducNet

<<http://www.educnet.education.fr/tech/normes/seminaire-lyon-1003.htm>>

⁴⁷ *IEEE Learning Object Metadata*

10. Droits de propriété intellectuelle

10.1. Identifier les droits d'auteur

Les équipes projet **doivent** respecter les droits de propriété intellectuelle liés aux ressources avec lesquelles elles travaillent, y compris :

- les droits des propriétaires du matériel source qui est numérisé ;
- les droits des propriétaires des ressources numériques ;
- les droits ou autorisations concédés à un fournisseur de service pour rendre les ressources numériques disponibles ;
- les droits ou autorisations concédés aux utilisateurs des ressources numériques.

Les équipes projet **doivent** également respecter tous les droits issus des termes et conditions spécifiques à tout programme de numérisation au sein duquel elles opèrent.

Une attention particulière est recommandée dans les cas suivants :

- **Ressources publiées.** Il est peu probable que les éditeurs accordent l'autorisation de numériser des ressources soumises au droit d'auteur, à moins qu'ils n'en tirent des avantages. Il est possible que les fonds plus anciens ne soient plus soumis au droit d'auteur mais l'équipe projet est responsable de s'en assurer.
- **Productions internes.** Les droits relatifs à tout travail entrepris par le personnel d'un établissement dans le cadre de ses activités normales restent la propriété de l'établissement. Dans certains établissements universitaires, il est possible que ces droits n'aient pas été revendiqués et les auteurs peuvent les avoir assignés à des éditeurs externes. Les bénévoles restent détenteurs du droit d'auteur sur leur travail, même s'ils renoncent à leurs droits.
- **Institutions sous-traitant le travail.** Ce travail, par exemple de la photographie, aura normalement assuré des droits de reproduction, mais ils ne s'étendent pas nécessairement à la numérisation, sauf si cela est spécifiquement défini dans le contrat. Les équipes projet détiendront les droits sur les fonds numérisés uniquement cette autorisation leur a été explicitement accordée.
- **Dons, legs et prêts.** Les fonds peuvent être soumis à des conditions particulières susceptibles de limiter la possibilité de les numériser.

Les équipes projet **devraient** enregistrer l'existence des droits qu'ils détiennent et les autorisations qui leur ont été concédées.

Droits d'auteur et environnement en réseau, article du groupe de travail sur la politique des services en réseau

Créer des ressources numériques pour les arts visuels: Standards et bonnes pratiques du Service britannique des Données sur les Arts et les Humanités

<http://vads.ahds.ac.uk/guides/creating_guide/contents.html>

Instructions sur les droits d'auteur du JISC

<http://www.jisc.ac.uk/uploaded_documents/bp19.pdf>

Propriété Intellectuelle britannique

<<http://www.intellectual-property.gov.uk>>

Organisation mondiale de la propriété intellectuelle (OMPI)

<<http://www.wipo.org/>>

En français

Guide d'élaboration des contrats du ministère de la culture et de la communication

<http://www.culture.gouv.fr/culture/mrt/numerisation/fr/documents/guide_elaboration_contrats.pdf>

Le guide de la numérisation des collections de la sous-direction des bibliothèques et de la documentation du ministère de la jeunesse, de l'éducation et de la recherche

<<http://www.sup.adc.education.fr/bib/Acti/Num/notech.htm#1.6.1.>>

10.2. Préserver les droits d'auteur

Dans l'environnement réseau, toute transaction engageant la propriété intellectuelle est, par nature, une transaction de droits. L'expression de ces "conditions d'accès"⁴⁸ ou "règlement commercial" dépend des "métadonnées juridiques" – données qui identifient sans ambiguïté et de manière certaine la propriété intellectuelle, les droits spécifiques concédés (par exemple de lire, imprimer, copier, modifier), ainsi que les utilisateurs ou utilisateurs potentiels.

Les équipes projet **devraient** tenir à jour les informations sur les droits qu'ils détiennent et acquièrent de manière cohérente afin qu'elles puissent être communiquées sous un format standard.

Parmi les types d'informations nécessaires :

- L'identification de la ressource ;
- Le nom de la personne ou de l'organisation qui concède les droits ;
- Le ou les droit(s) précis qui sont concédé(s) (y compris, par exemple, si la modification est autorisée) – et toute exclusion spécifique ;
- La période pendant laquelle les droits sont concédés ;
- Le ou les groupe(s) d'utilisateurs qui sont autorisés à utiliser la ressource ;
- Toute obligation (y compris, mais pas seulement, les obligations financières) qui peuvent incomber aux utilisateurs de la ressource.

10.2.1. Le commerce électronique

Dans le cas de programmes de création de contenu du secteur public, il est courant de spécifier que le contenu créé doit être consultable gratuitement, au moins dans un but éducatif. Dans certains cas, les programmes encouragent ou requièrent que les projets génèrent des revenus à partir des ressources créées.

Les équipes projet **doivent** se conformer aux conditions définies dans le cadre du programme concernant l'accès aux ressources créées et leur utilisation.

Les équipes projet **doivent** s'assurer que les droits de propriété intellectuelle sont correctement protégés.

⁴⁸ 'Terms of Availability'

10.2.2. Les licences *Creative Commons*

[NDLR : Le projet des licences *Creative Commons* a défini des licences d'utilisation pour des auteurs qui souhaitent publier leurs œuvres en ligne en renonçant au paiement de droits d'auteur, éventuellement sous certaines conditions (par exemple des usages à but éducatif). Inspirées des licences publiques GNU GPL, les licences *Creative Commons* portent sur des œuvres de création, telles que des sites Internet, photographies, littérature, images... Les équipes projet **peuvent** publier leurs ressources sous licences *Creative Commons*.]

10.2.3. Les tatouages⁴⁹ et empreintes⁵⁰ numériques

Les équipes projet **devraient** envisager de tatouer et mettre des empreintes sur les fonds numériques qu'ils produisent.

Le tatouage numérique est l'incrustation d'une marque permanente au sein d'un fichier qui peut ensuite être utilisée pour prouver l'origine d'une image ou les droits d'auteurs afférents. Le tatouage numérique est en général intégré aux données de l'image de manière à ce qu'il soit virtuellement impossible de l'enlever. Les tatouages numériques peuvent être visibles, invisibles ou une combinaison des deux. Dans tous les cas, le tatouage numérique est introduit de manière à altérer le moins possible l'image originale. Les tatouages numériques invisibles **doivent** pouvoir résister au recadrage, à la rotation, à la compression et à la transformation d'une image.

De même qu'il est possible de tatouer numériquement les images avant de les publier, les empreintes des images peuvent être prises de manière dynamique au moment où elles sont diffusées, c'est-à-dire lorsque l'image est téléchargée d'un site Internet. Il est ensuite possible de coder d'autres informations telles que l'identifiant de l'utilisateur, la date, l'heure, l'adresse IP, etc. dans le tatouage numérique. Cela rend chaque téléchargement unique et localisable grâce à une base de données de transactions enregistrant qui télécharge des images. Des techniques similaires peuvent être utilisées pour les supports audio et vidéo.

Standards

[Creative Commons (projet américain)

<<http://creativecommons.org/>>]

Recommandations:

Dérober et chaparder, la bibliothèque virtuelle des développeurs Web, article de Linda Cole sur la Bibliothèque virtuelle des développeurs

<<http://www.wdvl.com/Authoring/Graphics/Theft/>>

En français

Travaux français menés par le CERSA (CNRS/Paris II) pour adapter les *Creative Commons*

<<http://creativecommons.org/projects/international/fr/>>

Mesures techniques de protection des oeuvres & DRMS – état des lieux janvier 2003 Etude de Philippe Chantepie

<<http://www.culture.gouv.fr/culture/cspla/Mptdrms.pdf>>

⁴⁹ *Watermarking*

⁵⁰ *Fingerprinting*

11. Résumé

Le but de ce document est de fournir un ensemble de recommandations plutôt que de refléter les différentes exigences des multiples projets et programmes. Les responsables de la mise en place de projets de numérisation devront adapter ces recommandations au contexte spécifique dans lequel ils se trouveront afin d'en sélectionner certains, d'en adapter d'autres ou même d'en ajouter. Toutefois, les auteurs espèrent que ce document pourra être une base utile dans des contextes différents.

11.1 Maintenance

Ces recommandations seront mises à jour et développées par le projet MINERVA. Tout commentaire ou suggestion concernant des changements ou des mises à jour devront être soumis au projet MINERVA.

Guide :

Site du projet MINERVA

<<http://www.minervaeurope.org/>>

<minerva2@beniculturali.it>