



# InterPARES 2 Project

International Research on Permanent Authentic Records in Electronic Systems

Projet InterPARES 2

Recherche internationale sur les documents d'archives authentiques  
permanents dans les systèmes électroniques

**Titre: Rapport final de l'Étude générale 11 :  
Sélection de formats de fichiers  
numériques pour préservation à long  
terme**

**Statut:** Final (public)

**Version:** 1.1

**Date de Soumission:** Décembre 2006

**Date de Parution:** Mars 2007

**Auteur:** Projet InterePARES 2

**Écrivain:** Evelyn Peters McLellan,  
Insurance Corporation of British Columbia

**Traducteur:** Société Gamma Inc.

**Étude Unité:** Domaine 3

**URL:** [http://www.interpares.org/display\\_file.cfm?doc=ip2\\_file\\_formats\(complete\).pdf](http://www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf) [Anglais]

[http://www.interpares.org/display\\_file.cfm?doc=ip2\\_file\\_formats\\_fichiers\\_numériques.pdf](http://www.interpares.org/display_file.cfm?doc=ip2_file_formats_fichiers_numériques.pdf) [Français]

## Table des matières

<b>Introduction</b> .....	1
<b>1. Terminologie</b> .....	1
1.1 Qu'est-ce qu'un format de fichier? .....	1
1.2 Formats de fichiers « ouverts » .....	3
1.3 Formats de fichiers « standard » .....	4
1.4 Formats de fichiers « stables » .....	6
1.5 Normalisation des termes .....	6
<b>2. Critères de sélection</b> .....	7
2.1 Utilisation largement répandue .....	7
2.2 Origine non exclusive .....	8
2.3 Large diffusion des spécifications .....	9
2.4 Indépendance des plates-formes (interopérabilité) .....	10
2.5 Compression .....	12
2.6 Analyse des critères .....	13
<b>3. Implications pour les politiques</b> .....	15
<b>4. Recommandations pour l'élaboration et la mise en œuvre des politiques</b> .....	19
Annexe A : Liste des dépôts examinés .....	21
Annexe B : URL des documents examinés, par dépôt .....	22
Bibliographie .....	24

# Sélection de formats de fichiers numériques pour préservation à long terme

## Introduction

Ces dernières années, c'est devenu pratique courante pour les dépôts de documents numériques — y compris les archives — d'accepter certains formats de fichiers numériques pour préservation à long terme tout en en rejetant d'autres. La question de savoir si les archives devraient limiter le nombre des formats de fichiers à préserver et les critères de sélection de ces formats soulèvent d'importantes questions théoriques et stratégiques dont doivent tenir compte tant les chercheurs dans le domaine de la préservation des documents numériques que les archivistes chargés de la gestion des dépôts numériques. Nous tentons ici de répondre à ces questions, puisque ce rapport est une analyse des questions et des tendances relatives à la sélection de formats de fichiers pour la préservation des documents numériques qui, compte tenu de la portée de la recherche d'InterPARES 2, est focalisée sur les secteurs du gouvernement électronique, des sciences et des arts. C'est un examen essentiellement qualitatif de la documentation disponible sur les sites internet de vingt dépôts et de quatre groupes de collaboration multi-institutionnels ayant des politiques et des procédures ou des lignes directrices d'inclusion bien établies à l'intention des organismes qui leur transfèrent des documents.<sup>1</sup> Invariablement, ces groupes et ces dépôts précisent les formats qu'ils acceptent ou donnent des indications des formats qui se prêtent le mieux à la préservation à long terme des documents. La liste des dépôts et des groupes de recherche étudiés figure en annexe.

## 1. Terminologie

L'examen de la documentation a été conçu pour recueillir des renseignements sur les critères appliqués dans diverses institutions afin de choisir les formats des fichiers numériques qu'elles entendent préserver à long terme. L'inexistence d'une terminologie uniforme pour décrire les caractéristiques des différents formats a nui dans une certaine mesure à la compréhension et à l'évaluation de ces critères.

### 1.1 Qu'est-ce qu'un format de fichier?

La plupart des institutions que nous avons étudiées ne définissent pas le terme *format de fichier* qu'on emploie toutefois typiquement pour désigner la structure d'organisation des données dans un dossier. InterPARES 2 définit ce terme en partie comme « l'organisation des données dans les fichiers,

---

<sup>1</sup> La recherche nécessaire à la production de ce document a été réalisée avec l'aide de Tracey Krause et Yvonne Loiselle, étudiantes en études supérieures à la School of Library, Archival and Information Studies de l'Université de la Colombie-Britannique.

habituellement conçue pour faciliter le stockage, l'extraction, le traitement, la présentation et/ou la transmission des données avec des logiciels ». <sup>2</sup> Pour sa part, le *Data Dictionary for Preservation Metadata* PREMIS définit ce terme en tant que « structure spécifique préétablie pour l'organisation d'un fichier numérique ou d'un train de bits ». <sup>3</sup> Cette structure préétablie comprend le codage des données, autrement dit la façon d'interpréter les bits pour produire du texte, des images et des sons. Certains types de codages sont spécifiquement associés à des formats de fichiers; par exemple, le codage MP3 sert à encoder le format de fichier audio numérique MP3. Pourtant, bien des formats peuvent avoir différents codages, puisque même un fichier plein texte peut être encodé en formats ASCII, EBCDIC ou Unicode, qui ont tous de nombreuses variantes. Le codage peut poser des problèmes pour les formats de fichier audio et vidéo parce que le codage optimal pour le stockage et la transmission exige souvent une compression (la suppression de bits des fichiers numériques pour en réduire la taille) qui risque de saper les efforts déployés pour leur préservation. <sup>4</sup> Une étude réalisée par l'Arts and Humanities Data Service (AHDS) a souligné qu'il existe une centaine de codages avec compression pour le format largement utilisé WAVE, par exemple, ce qui peut signifier que les fichiers WAVE peuvent être ouverts par certains logiciels, mais pas par d'autres. <sup>5</sup> Même TIFF, un format souvent employé pour la préservation à long terme d'images, peut être encodé de différentes façons, certaines faisant appel à une compression avec perte (une technique entraînant la perte définitive de certaines données).

Le problème du codage est d'autant plus complexe que les formats TIFF, WAVE, AVI et d'autres formats courants de fichiers d'images ainsi que d'autres fichiers audiovisuels ne sont pas des « formats » de fichiers en eux-mêmes, mais plutôt des « formats enveloppeurs » de fichiers, aussi appelés « formats conteneurs », conçus pour combiner de multiples trains de bits en un seul fichier. En ce qui concerne le format enveloppeur WAVE, l'AHDS a souligné dans son étude que « c'est seulement lorsqu'on ouvre le fichier que les spécifications détaillées — la plus importante étant la méthode de compression — sont révélées. Si le codec du logiciel [c.-à-d. le programme de compression] est spécifié et disponible, on a accès au fichier; sinon, on se retrouve avec un problème tel qu'il faudra probablement avoir recours à un logiciel d'enquête pour le diagnostic ». <sup>6</sup> Et l'AHDS recommande qu'on utilise les enveloppeurs AVI,

---

<sup>2</sup> « File format », The InterPARES 2 Project Dictionary, site internet d'InterPARES 2, à [http://www.interpares.org/ip2/display\\_file.cfm?doc=ip2\\_dictionary.pdf](http://www.interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf).

<sup>3</sup> Preservation Metadata: Implementation Strategies (PREMIS), *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*. États-Unis : On-Line Computer Library Center and Research Libraries Group, mai 2005, p. 237. <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>.

<sup>4</sup> Voir Caroline R. Arms et Carl Fleischhauer, compilateurs, *Sustainability of Digital Formats: Planning for Library of Congress Collections*. Washington (D.C.) : Bibliothèque du Congrès, mise à jour le 6 mars 2006. <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>. Pour une introduction pratique à la compression, voir *File Formats and Compression*. Royaume-Uni : Technical Advisory Service for Images, mars 2005. <http://www.tasi.ac.uk/advice/creating/format.html#ff3>.

<sup>5</sup> Andrew Wilson et al., *Moving Images and Sound Archiving Study*. Arts and Humanities Data Service, Royaume-Uni : Final Draft, juin 2006, p. 29.

<http://roda.ianit.pt/?q=en/system/files/Moving+Images+and+Sound+Archiving+Study1.doc>. Définition du terme *codec* insérée par l'auteur.

<sup>6</sup> *Ibid.*

QuickTime et WMV pour des formats de fichiers d'archives seulement si « les trains de bits audio et vidéo dans les enveloppeurs ne sont pas comprimés, ou si la compression ne cause aucune perte de données ».<sup>7</sup>

Plusieurs des institutions étudiées recommandent ou acceptent le format de fichier XML. Nous devons toutefois nous demander si c'est bien un format de fichier en soi, puisque XML est un métalangage permettant à l'utilisateur de créer des ensembles spécifiques de balises ou de langages marqueurs pour spécifier la structure de certains types de documents. Ainsi, GML (Langage balisé en géographie) et XHTML (Langage hypertexte extensible) sont des sous-types de XML, et aucun des deux ne peut être considéré comme un format de fichier au sens strict du terme, puisque ce sont plutôt des fichiers-texte avec des balises définies en XML. Cela dit, XML est plus générique qu'un simple langage de balisage contenant au moins un ensemble précis de balises conçues pour structurer des données dans un dossier. Par conséquent, il n'est pas en lui-même un format de fichier, mais rend simplement possible la structuration et la définition de formats de fichiers et de formats d'enveloppeurs. (Nous reviendrons à un examen plus étoffé des implications de l'acceptation des formats XML.)

Comme la définition d'un *format de fichier* n'est donc pas toujours simple, les institutions qui établissent des lignes directrices sur les formats de fichiers pour leur préservation ou qui précisent quels formats sont susceptibles d'être acceptés dans un dépôt de fichiers numériques doivent clairement stipuler si les « formats » qu'elles acceptent sont des formats de fichiers, des formats enveloppeurs ou des fichiers balisés. Dans le cas des formats enveloppeurs, il faut absolument préciser les codages des trains de bits constituants; dans celui des fichiers XML, le format du fichier, le codage et le schéma XML ou la définition du type de document doivent être spécifiés. Il faudrait aussi préciser le numéro de la version de chaque format existant en plus d'une version, étant donné que des versions différentes peuvent avoir des exigences de codage et d'autres caractéristiques différentes.

## 1.2 Formats de fichiers « ouverts »

De nombreuses institutions étudiées emploient des termes tels que format de fichier *ouvert* pour décrire un format de fichier ayant une ou plusieurs caractéristiques, la principale étant que ses spécifications sont publiées et largement diffusées, en plus du fait que le format a été créé par un logiciel non exclusif (ou à source générale). Toutefois, cela peut aussi vouloir dire que le format n'est pas protégé par un brevet, que son utilisation n'implique ni le versement de redevances, ni la possibilité qu'on doive en verser plus tard, ou encore qu'il est largement utilisé. Dans bien des cas, la signification du terme n'est pas claire. Par exemple, les Archives nationales d'Australie acceptent les formats de fichiers « à la fois ouverts et étayés par une solide documentation »,<sup>8</sup> ce qui laisse entendre qu'un format de fichier dont les spécifications sont

<sup>7</sup> *Ibid.*, p. 48.

<sup>8</sup> Simon Davis, *Recordkeeping Issues Forum: Digital Preservation Strategy*. Australie : National Archives of Australia, 19 novembre 2002, p. 4. [http://www.naa.gov.au/recordkeeping/rkpubs/fora/02nov/digital\\_preservation.pdf](http://www.naa.gov.au/recordkeeping/rkpubs/fora/02nov/digital_preservation.pdf).

largement connues n'est pas synonyme de format de fichier ouvert, tandis que la Research Libraries Group and Digital Library Federation parle de « formats ouverts et non exclusifs »,<sup>9</sup> ce qui revient à dire que les caractéristiques d'être ouvert et non exclusif d'un format sont distinctes. Au Royaume-Uni, le Technical Advisory Service for Images a écrit qu'un format choisi pour la préservation de fichiers numériques devrait « être un format de fichier standard ouvert; on ne devrait pas se servir de formats de fichiers exclusifs, puisqu'il n'est pas certain qu'on pourra ouvrir les fichiers dans l'avenir »,<sup>10</sup> sans préciser que « l'ouverture » s'entendrait de l'existence d'une spécification publiée et largement diffusée. Le Groupe de travail de l'Union européenne et des États-Unis sur les collections audio de la parole a déclaré qu'il « existe ce qu'on pourrait appeler un "spectre de l'ouverture", avec des degrés de protection appliqués par les concepteurs de l'industrie informatique... On trouve une documentation publique exhaustive pour certains formats exclusifs, p. ex. TIFF ». <sup>11</sup> Les Data Archive du Royaume-Uni font une distinction entre les formats exclusifs, les formats « diffusés » (c.-à-d. les formats exclusifs dont les spécifications sont largement diffusées) et les formats « ouverts », autrement dit « créés par un groupe coopératif, puis largement diffusés pour que tout le monde puisse s'en servir, sans restriction ». <sup>12</sup> Il vaut la peine de préciser que les formats « ouverts » définis comme étant non exclusifs et comme ayant des spécifications largement diffusées ne sont pas nécessairement synonymes de formats produits par des logiciels à source générale. Cette dernière expression s'entend de logiciels dont les codes sont largement diffusés et peuvent être modifiés. Les logiciels à source générale ne produisent pas toujours des formats non exclusifs : par exemple, certains logiciels à source générale peuvent être utilisés pour créer le format exclusif PDF. <sup>13</sup>

### 1.3 Formats de fichiers « standard »

D'autres institutions — moins nombreuses — qualifient de *standard* les formats de fichiers acceptés ou recommandés. Le terme a une large acception; on s'en sert pour indiquer que le format a plusieurs caractéristiques désirables. L'Institut des services d'information scientifique des Pays-Bas, par exemple, définit les « formats de fichiers d'images standard » comme des formats largement adoptés, dont les spécifications sont publiées, d'une grande

<sup>9</sup> Franziska Frey, *Guides to Quality in Visual Resource Imaging: 5. File Formats for Digital Masters*. États-Unis : Research Libraries Group and Digital Library Federation, 2000. Section 2.3, *Pros and Cons of Various Formats*. <http://www.rlg.org/visguides/visguide5.html>.

<sup>10</sup> Technical Advisory Service for Images, *Advice Paper: Choosing a File Format*. United Kingdom: Technical Advisory Services, mai 2006. <http://www.tasi.ac.uk/advice/creating/format.html>.

<sup>11</sup> Steve Renals et Jerry Goldman et al., *EU-US Working Group on Spoken-Word Audio Collections, Final Report*. Union européenne et États-Unis : EU-US Working Group on Spoken-Word Audio Collections, 18 juin 2003, p. 33. <http://www.dcs.shef.ac.uk/spandh/projects/swag/swagReport.pdf>.

<sup>12</sup> UK Data Archive, *UK Data Archive Preservation Policy*. Colchester: University of Essex, version 2.0, septembre 2005, p. 19.

<http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf>.

<sup>13</sup> PDF (Format de document portable) est protégé par plusieurs brevets détenus par la société qui l'a créé, Adobe Systems Inc. Toutefois, ses spécifications sont publiques, et Adobe Systems autorise les producteurs de logiciels qui s'y conforment à s'en servir sans lui verser de redevances. En outre, l'Organisation internationale de normalisation a récemment approuvé PDF/Archival (PDF/A), une version élargie de PDF 1.4, pour utilisation comme format de préservation à long terme (ISO 19005-1:2005, publiée le 1<sup>er</sup> octobre 2005).

interopérabilité, n'exigeant aucune compression de données et capables de soutenir les métadonnées qu'on veut préserver.<sup>14</sup> Bibliothèque et Archives Canada parle pour sa part de formats « *de facto* standard » qui, bien qu'exclusifs, sont « reconnus comme des formats et des types de fichiers qui sont devenus la norme dans l'industrie en raison de leur utilisation et de leur soutien très répandus, et non parce qu'ils ont été officiellement approuvés par une organisation d'établissement de normes ». <sup>15</sup> Le Bureau des documents d'archives publics de l'État de Victoria, en Australie, recommande qu'on opte pour un format devenu « la norme dans l'industrie » faute de format non exclusif dont les spécifications seraient publiées:

[Traduction]

La raison pour laquelle on utilise un format qui est « devenu la norme dans l'industrie » est économique. Si ce produit détient la plus grande part du marché, il est peu probable qu'il soit remplacé facilement ou rapidement. Par conséquent, les documents d'archives conservés dans ce format seront accessibles pour une période raisonnable. En outre, si le format devenu la norme dans l'industrie devait être remplacé, les produits de remplacement seront presque inévitablement capables de lire le format exclusif de celui qu'ils remplaceront.<sup>16</sup>

D'autres organisations emploient le mot *standard* pour indiquer qu'un format a été accepté ou que des organisations nationales ou internationales de normalisation comme l'ANSI ou l'ISO en encouragent l'utilisation,<sup>17</sup> un critère qui peut être substitué à d'autres ou complété par eux. Les responsables du projet DAVID, par exemple, ont déclaré que les « formats de fichiers standard doivent leur statut à des initiatives (officielles) de normalisation », en ajoutant que ce statut peut aussi être dérivé de « leur utilisation générale ». <sup>18</sup> Les documents produits dans le cadre de ce projet ne contiennent pas de renseignements détaillés sur ce qui constitue une initiative officielle de normalisation d'un format de fichier. Le Musée des beaux-arts/Archives du cinéma du Pacifique de l'Université de la Californie à Berkeley parle d'utilisation des « normes nationales et internationales dans toute la mesure du possible », en précisant que la durabilité du contenu numérique est accrue lorsqu'il est conservé « dans des formats standard soutenus par une documentation diffusée dans une vaste

<sup>14</sup> René van Horik, *Image Formats: Practical Experiences*. Pays-Bas : Institut des services d'information scientifique des Pays-Bas, 2004. Présentation Erpanet, Vienne, mai 2004, p. 22.

[http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna\\_Horik.pdf](http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna_Horik.pdf).

<sup>15</sup> David L. Brown et Mike Swan. *Lignes directrices relatives aux types de fichiers informatiques, aux formats de transfert et aux normes d'information*. Ottawa : Bibliothèque et Archives Canada (BAC), version 1.1, 28 juin 2004, section 1.3, *Concept*. <http://www.collectionscanada.ca/information-management/002/007002-3017-f.html>.

<sup>16</sup> Bureau des documents d'archives publics de l'État de Victoria, *Advice 13: Long-Term Preservation Formats*. North Melbourne: Public Records Office Victoria, 2 septembre 2004, section 6.3, *Industry standard formats*. [http://www.prov.vic.gov.au/vers/standard/advice\\_13/](http://www.prov.vic.gov.au/vers/standard/advice_13/).

<sup>17</sup> American National Standards Institute et Organisation internationale de normalisation, respectivement.

<sup>18</sup> Projet DAVID (Archivering in Vlaamse Instellingen en Diensten/archivage numérique dans les institutions et les administrations flamandes), *Digital Archiving, Guideline and Advice 4: Standards for Fileformats*. Anvers, 2003, p. 1. <http://www.expertisecentrumdavid.be/davidproject/teksten/guideline4.pdf>.



communauté, sans tomber sous le contrôle d'une compagnie quelconque ». Cette organisation qualifie aussi de tels formats de « neutres ». <sup>19</sup>

#### 1.4 Formats de fichiers « stables »

Bibliothèque et Archives Canada, les Archives numériques du Florida Center for Library Automation (FCLA), les Archives nationales du Royaume-Uni et les Archives de données du Royaume-Uni parlent de *stabilité* des formats. Un format de fichiers stable semble être à la fois rétrocompatible (autrement dit compatible avec ses versions antérieures) et bien soutenu par l'industrie des logiciels. Bibliothèque et Archives Canada a écrit qu'AVI est devenu une norme *de facto*, mais que « Microsoft a annoncé qu'il suspendrait sous peu son soutien du format. À court terme, les fichiers en AVI devraient être convertis en un format plus stable, [car] les chances de ce format de continuer à être soutenu sont précaires ». <sup>20</sup> Pour leur part, les Archives nationales du Royaume-Uni ont écrit qu'il « est souhaitable que les spécifications du format soient stables et non susceptibles de changer constamment (ou nettement) avec le temps. Les nouvelles versions du format devraient aussi être rétrocompatibles ». <sup>21</sup> Les Archives numériques du FCLA et les Archives de données du Royaume-Uni recommandent l'utilisation de formats stables, sans toutefois définir ce qu'elles entendent par là. <sup>22</sup>

#### 1.5 Normalisation des termes

Les termes employés pour décrire les caractéristiques des formats de fichiers numériques qui optimisent les possibilités de préservation à long terme ne sont donc pas définis de façon uniforme dans toutes les institutions, voire pas définis du tout. La normalisation des définitions de termes tels qu'*ouvert*, *standard* et *stable* serait une étape utile, qui aiderait les institutions chargées de préserver les fichiers à communiquer efficacement avec leurs donateurs, leurs institutions mères et les autres dépôts de fichiers numériques. L'emploi des termes *ouvert* et *standard* semble particulièrement épineux, mais en poursuivant la lecture de ce rapport, le lecteur verra qu'il existe énormément de variantes dans la terminologie employée pour décrire de nombreuses autres caractéristiques des formats de fichiers propices à leur préservation.

<sup>19</sup> Richard Rinehart et Guenter Waibel, *Strategies for Digital Media Asset Management*. Californie : UC Berkeley Art Museum/Pacific Film Archive, 26 avril 2001. Document de politique interne non publié.

<sup>20</sup> Brown et Swan, *Lignes directrices*, op. cit., section 3.3.2.1. *Entrelacement audio-vidéo (AVI)*.

<sup>21</sup> Adrian Brown, *Digital Preservation Guidance Note 1: Selecting File Formats for Long-Term Preservation*. Surrey (R.-U.) : National Archives of the United Kingdom, 19 juin 2003, p. 6.

[http://www.nationalarchives.gov.uk/preservation/advice/pdf/selecting\\_file\\_formats.pdf](http://www.nationalarchives.gov.uk/preservation/advice/pdf/selecting_file_formats.pdf).

<sup>22</sup> « En règle générale, servez-vous de formats indépendants des plates-formes et des fournisseurs, non exclusifs, stables, ouverts et bien soutenus. » Florida Center for Library Automation (FCLA), *Recommended Data Formats for Preservation Purposes in the FCLA Digital Archive*. Gainesville (Floride) : Florida Center for Library Automation, juin 2005, p. 2. <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>; « Les Archives de données du Royaume-Uni préfèrent conserver les fichiers dans des formats basés sur des normes stables, ouvertes et largement diffusées... », Archives de données du Royaume-Uni, *Preservation Policy*, op. cit., p. 23. <http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf>.



## 2. Critères de sélection

L'examen de la documentation a révélé plusieurs caractéristiques des formats de fichiers numériques généralement considérées comme nécessaires pour la préservation à long terme des fichiers numériques. Les principales sont qu'ils soient largement adoptés, non exclusifs, soutenus par une solide documentation (autrement dit assortis de spécifications largement diffusées), indépendants des plates-formes (ou interopérables), et soit non comprimés, soit comprimés avec une technique ne causant aucune perte de données. Ces critères sont analysés de façon détaillée dans les pages qui suivent.

### 2.1 Utilisation largement répandue

Dix-huit des vingt-quatre institutions<sup>23</sup> étudiées considèrent l'adoption largement répandue des formats de fichiers comme un critère de leur sélection pour la préservation à long terme de données numériques. L'On-line Computer Library Center, par exemple, a déclaré que les dépôts de fichiers numériques appelés à évaluer les facteurs de risque se doivent de tenir compte du fait qu'un format est « largement accepté ou n'occupe au contraire qu'un petit créneau ».<sup>24</sup> Le Massachusetts Institute of Technology (MIT) accepte les fichiers en Microsoft Word et PowerPoint ainsi qu'en Lotus 1-2-3 et en WordPerfect (sans toutefois en garantir la préservation intégrale) parce que ces formats « sont si largement utilisés que des outils de migration conçus par des tiers vont vraisemblablement être créés pour faciliter la migration entre eux et d'autres formats ».<sup>25</sup> Les Archives nationales du Royaume-Uni sont favorables aux formats largement répandus parce que leur utilisation généralisée tend à perpétuer leur soutien par l'industrie des logiciels:

Les lois de l'offre et de la demande dictent que les formats bien établis dont l'utilisation est largement répandue tendent à être plus généralement soutenus — et pour plus longtemps — par les fournisseurs de logiciels que ceux cantonnés dans une niche. Les formats les plus largement utilisés, qui sont soutenus par une aussi vaste gamme de logiciels que possible, doivent donc être préférés dans toute la mesure du possible.<sup>26</sup>

La Bibliothèque du Congrès estime quant à elle que « si un format est largement adopté, il est moins susceptible de devenir rapidement désuet, et la création d'outils de migration et d'émulation par l'industrie est plus vraisemblable dans son cas, sans que les institutions d'archives ne doivent investir spécifiquement en ce sens ».<sup>27</sup> (L'utilisation largement répandue en tant

<sup>23</sup> Par souci de simplicité, nous incluons dans les « institutions » les quatre groupes de collaboration pour la recherche figurant à l'Annexe A.

<sup>24</sup> Andreas Stanesco, « Assessing the Durability of Formats in a Digital Preservation Environment: the INFORM Methodology », *D-Lib Magazine* 10(11), novembre 2004.  
<http://www.dlib.org/dlib/november04/stanesco/11stanesco.html>.

<sup>25</sup> Massachusetts Institute of Technology (MIT), *General DSpace FAQ*. Cambridge (Massachusetts) : MIT Libraries, non daté. <http://libraries.mit.edu/DSpace-mit/about/faq.html>.

<sup>26</sup> Brown, *Digital Preservation Guidance Note 1*, op. cit., p. 6.

<sup>27</sup> Arms et Fleischhauer, *Sustainability of Digital Formats*, op cit.

qu'élément clé de l'élaboration de normes devenues la règle dans l'industrie ou un standard *de facto* a été décrite plus haut.)

Déterminer ce qui constitue une utilisation largement répandue est une démarche subjective. Les auteurs d'un rapport produit pour la Bibliothèque de l'Université Cornell ont déclaré qu'un format devrait être « largement adopté par de vastes communautés et par de gros groupes pour accroître les chances qu'ils établissent des voies de migration bien définies », <sup>28</sup> tandis que l'Institut des services d'information scientifique des Pays-Bas a déclaré qu'il devrait avoir été « utilisé par une vaste communauté durant une longue période ». <sup>29</sup> Pour sa part, Bibliothèque et Archives Canada recommande le format MPEG-2 plutôt que MPEG-4 pour la conservation des images animées en raison notamment de l'« acceptation dans le marché et [de la] pénétration » du premier, alors que la norme MPEG-4 « n'a toujours pas été adoptée par un certain nombre de créateurs et de fabricants de logiciels ». <sup>30</sup> La Bibliothèque du Congrès a reconnu que le degré d'utilisation est difficile à quantifier en écrivant que:

Les preuves de l'adoption largement répandue d'un format numérique comprennent le groupage d'outils employés avec des ordinateurs personnels, le soutien local dans les navigateurs Web ou les outils de création de contenu de pointe sur le marché, notamment ceux qui sont conçus pour les professionnels, ainsi que l'existence de nombreux produits rivaux pour la création, la manipulation ou le rendu des objets numériques dans ce format. <sup>31</sup>

Fait à signaler, la Bibliothèque du Congrès ajoute qu'un « format qui a été étudié par d'autres institutions d'archives et accepté comme format d'archives de choix ou soutenu a aussi fait la preuve de son adoption ». <sup>32</sup> L'adoption d'un format par des organismes d'archives est cruciale pour le Département d'architecture de l'Art Institute de Chicago, qui recommande TIFF et PDF parce qu'ils sont « largement utilisés pour fins d'archives ». <sup>33</sup>

## 2.2 Origine non exclusive

Dix-sept des institutions étudiées considèrent les formats non exclusifs comme préférables aux exclusifs. Certaines d'entre elles les qualifient de formats *ouverts*, tandis que d'autres, comme l'On-line Computer Library Center et la Bibliothèque du Congrès, déclarent expressément que l'utilisation d'un format devrait être possible sans qu'on doive verser des redevances, des frais

---

<sup>28</sup> Anne R. Kenney et al., *Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy: Final Project Report*. Ithaca (New York) : Cornell University Library, mai 2001, p. 8. <http://www.library.cornell.edu/ims/IMLS-CULfinalreport2.pdf>.

<sup>29</sup> Horik, *Image Formats*, op. cit., p. 22.

<sup>30</sup> Brown et Swan, *Lignes directrices*, op. cit., sections 3.3.1.1, *Groupe d'experts pour le codage d'images animées (MPEG-2)* et 3.3.2.2, *MPEG-4*.

<sup>31</sup> Arms et Fleischhauer, *Sustainability of Digital Formats*, op. cit.

<sup>32</sup> *Ibid.*

<sup>33</sup> Art Institute of Chicago, Department of Architecture. *Collecting, Archiving and Exhibiting Digital Design Data*. Chicago : Kristine Fallon Associates Inc., 2004, p. 35. [http://www.artic.edu/aic/collections/dept\\_architecture/dddreport/0C.pdf](http://www.artic.edu/aic/collections/dept_architecture/dddreport/0C.pdf).

d'utilisation sous licence ou des droits exigibles en vertu d'un brevet.<sup>34</sup> Cela dit, bien des institutions font des exceptions pour certains formats en se fondant sur des facteurs comme leur utilisation largement répandue, la diffusion de leurs spécifications ou la possibilité de convertir les fichiers visés dans des formats non exclusifs. Le Bureau des documents d'archives publics de l'État de Victoria recommande l'utilisation de formats non exclusifs dans toute la mesure du possible, tout en se servant de PDF comme format de préservation dans le cadre de sa Stratégie des documents d'archives électroniques de l'État de Victoria (*Victorian Electronic Records Strategy*, VERS), en partie parce que ses spécifications sont publiques.<sup>35</sup> L'Ohio Electronic Records Committee recommande les formats d'images numériques non exclusifs en précisant que si une institution accepte un format exclusif, celui-ci devrait « faire le pont avec un format de fichier d'images numériques non exclusif ».<sup>36</sup> De même, d'après la Digital Preservation Coalition de Grande-Bretagne, lorsque seuls des formats exclusifs sont disponibles, « l'existence de formats d'exportation soutenus de façon qu'il soit possible de transférer des données hors de [...] ces environnements exclusifs est un facteur d'importance cruciale ».<sup>37</sup> Comme nous l'avons déjà dit, le MIT accepte Microsoft et d'autres formats exclusifs en raison de leur utilisation très répandue, mais avec un bémol, en avouant qu'il ne pourra peut-être pas assurer la préservation intégrale du contenu des fichiers. La Bibliothèque du Congrès, qui n'emploie pas l'épithète *non exclusif*, préférant parler de l'existence ou de l'absence de brevets, fait valoir que:

L'impact des brevets n'est peut-être pas suffisamment important en lui-même pour justifier qu'on le considère comme un facteur indépendant... L'adoption [largement répandue] d'un format peut être un bon indice qu'il ne sapera pas la capacité des institutions d'archives de soutenir l'accès au contenu grâce à la migration, la génération dynamique de copies de service ou d'autres techniques.<sup>38</sup>

### 2.3 Large diffusion des spécifications

Pour dix-sept des institutions étudiées, la large diffusion des spécifications semble être un facteur important de la sélection des formats de fichiers pour la préservation à long terme des données. Certaines institutions préfèrent parler de « documentation » plutôt que de « spécifications », ou de formats de fichiers « étayés par une solide documentation ». La Bibliothèque de l'Université Cornell, par exemple, préconise la sélection de formats qui ont fait l'objet « d'un développement non exclusif approfondi, avec la documentation

---

<sup>34</sup> Stanescu, « Assessing the Durability of Formats », *op. cit.*; Arms et Fleischhauer, *Sustainability of Digital Formats*, *op. cit.*

<sup>35</sup> Public Records Office of Victoria, *Advice 13*, *op. cit.*, section 6.2.2, *Published Formats*, et section 7.2 *PDF (Portable Document Format)*.

<sup>36</sup> Ohio Electronic Records Committee, *Revised Digital Imaging Guidelines: Guidelines for State of Ohio Executive Agencies and Local Governments*. Ohio : Ohio Electronic Records Committee, 26 juin 2003. <http://www.ohiojunction.net/erc/imagingrevision/revisedimaging2003.html>.

<sup>37</sup> Neil Beagrie et Maggie Jones, *Digital Preservation Coalition Handbook*. Royaume-Uni : Digital Preservation Coalition, mise à jour en août 2006, section 5.2, *File Format and Standards*. <http://www.dpconline.org/graphics/handbook/>.

<sup>38</sup> Arms et Fleischhauer, *Sustainability of Digital Formats*, *op. cit.*

correspondante ». <sup>39</sup> Que la documentation et les spécifications soient synonymes n'est toutefois pas évident. Par exemple, l'On-line Computer Library Center considère comme un facteur de risque pour les formats le fait que la source ou les spécifications puissent être inspectées indépendamment ou pas, ou encore que le format soit complexe ou suffisamment étayé par des documents ou pas. <sup>40</sup> Bref, il semblerait que la documentation et les spécifications soient des réalités distinctes.

La plupart des institutions qui recommandent des formats dont les spécifications ont été publiées ou sont largement diffusées font le rapport entre cette exigence et celle voulant que le format ne soit pas exclusif. Néanmoins, bon nombre d'institutions sont disposées à recommander ou accepter des formats exclusifs, pourvu que leurs spécifications soient publiées. Les institutions qui acceptent PDF, comme la California Digital Library et le Bureau des documents d'archives publics de l'État de Victoria, sont des exemples de cette tendance. La Bibliothèque du Congrès va plus loin; en ce qui concerne les formats exclusifs dont les spécifications ne sont pas publiées, elle estime que « dans l'avenir, le dépôt d'une documentation complète en fiducie dans une organisation d'archives de confiance assurerait un certain degré de divulgation pour soutenir la préservation de renseignements dans des formats exclusifs à l'égard desquels la documentation n'a pas été rendue publique ». <sup>41</sup> Par conséquent, l'inexistence de spécifications publiées ou largement diffusées n'est pas nécessairement un obstacle insurmontable à la préservation à long terme des fichiers numériques dans un format donné.

## 2.4 Indépendance des plates-formes (interopérabilité)

Ce critère a été mentionné sous une forme ou une autre par treize des institutions étudiées. L'Institut des services d'information scientifique des Pays-Bas, par exemple, a déclaré que « le format doit être soutenu par une vaste gamme de systèmes », <sup>42</sup> tandis que la Bibliothèque d'État et universitaire, Arhus, et la Bibliothèque royale de Copenhague ont recommandé des formats indépendants du matériel, des systèmes d'exploitation et d'autres logiciels. <sup>43</sup> La Bibliothèque de l'Université Cornell parle d'uniformité entre les plates-formes et de « dépendances minimales du matériel et des logiciels », <sup>44</sup> tandis que les Archives nationales du Royaume-Uni préfèrent le terme « interopérabilité », qu'elles expliquent de la façon suivante:

La capacité d'échanger des documents d'archives électroniques avec d'autres utilisateurs et systèmes de TI est fréquemment un facteur important. Les formats soutenus par une vaste gamme de logiciels ou

<sup>39</sup> Anne R. Kenney et al., *Preserving Cornell's Digital Image Collections*, op. cit., p. 8.

<http://www.library.cornell.edu/imls/MLS-CULfinalreport2.pdf>.

<sup>40</sup> Stanesco, « Assessing the Durability of Formats », op. cit.

<sup>41</sup> Arms et Fleischhauer, *Sustainability of Digital Formats*, op. cit.

<sup>42</sup> Horik, *Image Formats*, op. cit., p. 22.

<sup>43</sup> Lars R. Clausen, principal auteur. *Handling File Formats*. Danemark : Bibliothèque d'État et universitaire, Arhus, et Bibliothèque royale, Copenhague, mai 2004, p. 12.

<http://netarchive.dk/publikationer/FileFormats-2004.pdf>.

<sup>44</sup> Kenney et al., *Preserving Cornell's Digital Image Collections*, op. cit., p. 8.

indépendants des plates-formes sont donc très désirables dans bien des situations. Cette caractéristique tend aussi à soutenir la viabilité à long terme des données, en facilitant leur migration d'un environnement technique à un autre.<sup>45</sup>

La Bibliothèque du Congrès considère les « dépendances de l'extérieur », autrement dit « le degré auquel un format donné dépend d'un matériel, d'un système d'exploitation ou de logiciels donnés pour le rendu ou l'utilisation » comme un des facteurs du soutien des formats de fichiers.<sup>46</sup> Digital Image Archive of Medieval Music (DIAMM) du Royaume-Uni a simplement déclaré qu'elle se sert de TIFF parce que c'est le format actuellement considéré comme « le plus largement compatible et facilement transférable pour les fichiers d'images ».<sup>47</sup>

Ces dernières années, on a beaucoup vanté XML en tant que mécanisme de facilitation de l'échange de données et de l'interopérabilité. De nombreuses institutions étudiées ont déclaré que les formats balisés en XML sont acceptables ou recommandés pour la préservation à long terme de données numériques, particulièrement lorsqu'ils font appel à des grammaires bien établies et normalisées comme GML et XHTML, dont on se sert pour faciliter l'interopérabilité respectivement des données géospatiales et des pages Web. Néanmoins, XML n'est pas une panacée pour résoudre les problèmes d'échange de données. Les documents en XML doivent être conformes aux définitions de type de document (DTD) déclarant leurs attributs, leurs éléments et leur syntaxe. Une évaluation des formats de documents ouverts réalisée pour la Commission européenne a révélé que le fait que les fabricants de logiciels, voire les créateurs individuels de documents d'archives créent des DTD peut générer des formats XML exclusifs dépendant d'un logiciel donné : « Les documents qui obéissent à [sic] différents formats basés sur XML ou fondés sur des DTD ne sont pas nécessairement compatibles. La conversion entre les deux formats peut même se révéler extrêmement difficile, sinon impossible. » Les auteurs du rapport ont ajouté que « le potentiel de création de formats de documents "exclusifs" en XML inquiète une grande partie de la communauté qui préconise des sources ouvertes ».<sup>48</sup> La plupart des institutions étudiées exigent que les formats basés sur XML soient accompagnés de leur DTD lorsqu'ils sont stockés dans un dépôt numérique. La maintenance de multiples DTD et le potentiel de création de formats exclusifs basés sur XML risquent toutefois de poser des problèmes pour les dépôts numériques, de sorte que toute politique sur les formats à sélectionner doit tenir compte de ces facteurs.

<sup>45</sup> Brown, *Digital Preservation Guidance Note 1, op. cit.*, p. 7.

<sup>46</sup> Arms et Fleischhauer, *Sustainability of Digital Formats, op. cit.*

<sup>47</sup> *Archiving*, site Internet de Digital Image Archive of Medieval Music (DIAMM), Royaume-Uni : Universités d'Oxford et de Londres, 15 juin 2006. <http://www.diamm.ac.uk/content/description/archiving.html>.

<sup>48</sup> Valoris, *Comparative Assessment of Open Documents Formats: Market Overview*. Commission européenne : 30 décembre 2003, p. 17. <http://europa.eu.int/idabc/servlets/Doc?id=17982>.



## 2.5 Compression

Il est difficile de quantifier les attitudes sur la compression, parce que les mandats et les politiques d'acquisition des institutions étudiées varient énormément. Ainsi, Digital Image Archive of Medieval Music est une institution d'archives essentiellement constituées d'images photographiques qui accepte seulement des images en TIFF sans compression. Bibliothèque et Archives Canada se situe à l'autre pôle, puisqu'elle acquiert des fichiers de texte et d'images ainsi que des fichiers audio et vidéo d'une grande variété, en formats aussi bien non comprimés que comprimés. Toutefois, il est possible de généraliser en concluant à l'existence d'une préférence marquée pour des fichiers non comprimés ou peu comprimés, si possible : quatre institutions acceptent ou recommandent seulement des fichiers non comprimés, et dix préfèrent ou recommandent des fichiers non comprimés tout en acceptant des fichiers avec compression sans perte. Cinq seulement des institutions acceptent explicitement des fichiers comprimés avec perte de données. Enfin, trois institutions, l'On-line Computer Library Center, le Groupe de travail UE-US sur les collections audio de la parole et la Bibliothèque d'État et universitaire, Arhus/Bibliothèque royale de Copenhague estiment que la compression est dangereuse ou cause des problèmes, sans toutefois la rejeter expressément.<sup>49</sup>

La plupart des institutions qui acceptent ou recommandent des fichiers comprimés stipulent que seulement certains types de fichiers peuvent être comprimés ou que certains types de compression sont acceptables et d'autres pas. La National Archives and Records Administration des États-Unis (NARA), par exemple, préfère les fichiers en TIFF non comprimés, mais accepte ceux qui sont comprimés sans perte de données, tout comme les fichiers JPEG comprimés avec perte, à certaines conditions.<sup>50</sup> La Bibliothèque de l'Université Cornell et la Research Libraries Group/Digital Library Federation acceptent des fichiers avec compression « sans perte visuelle » (une compression qui supprime certaines données sans réduire la qualité de l'image).<sup>51</sup> Certaines institutions, comme l'Ohio Electronic Records Committee et le Bureau des documents d'archives publics de l'État de Victoria, précisent quelles normes ou techniques de compression on doit employer.<sup>52</sup> La NARA pose des exigences générales et recommande des techniques de compression capables d'y répondre, en stipulant que l'organisme qui transfère les données doit « préciser la méthode de compression des fichiers utilisée (le cas échéant) et le degré de compression (p. ex. moyen, élevé) sélectionné pour la ou les images ». <sup>53</sup> La Bibliothèque du Congrès recommande que les dépôts de documents numériques « acceptent le contenu comprimé avec des algorithmes publiquement connus et largement

<sup>49</sup> Stanescu, « Assessing the Durability of Formats », *op. cit.*; EU-US Working Group on Spoken-Word Audio Collections, *Final Report*, p. 23 et 34; Clausen, *Handling File Formats*, *op. cit.*, p. 16.

<sup>50</sup> U.S. National Archives and Records Administration (NARA), *Expanding Acceptable Transfer Requirements: Transfer Instructions for Permanent Electronic Records: Digital Photographic Records*. Maryland, National Archives and Records Administration : non daté.

<http://www.archives.gov/records-mgmt/initiatives/erm-guidance.html>.

<sup>51</sup> Kenney et al., *Preserving Cornell's Digital Image Collections*, *op. cit.*, p. 37; Frey, *File Formats for Digital Masters*, *op. cit.*, section 3.4, *Compression*.

<sup>52</sup> Ohio Electronic Records Committee, *Guidelines for State of Ohio Executive Agencies*, *op. cit.*; Public Records Office of Victoria, *Advice 13*, *op. cit.*, sections 7.3 TIFF (Tagged Image File Format) et 7.4 JPEG.

<sup>53</sup> NARA, *Expanding Acceptable Transfer Requirements*, *op. cit.*

adoptés soit sans perte de données, soit avec des pertes dues à la compression, mais restant acceptable pour le créateur, l'éditeur ou le principal utilisateur en tant que version originale ». <sup>54</sup>

## 2.6 Analyse des critères

L'utilisation largement répandue des formats et le fait qu'ils ne sont pas exclusifs, que leurs spécifications sont publiées, qu'ils sont interopérables et sans compression (ou avec compression sans perte de données) semblent être les plus importants facteurs de sélection de formats de fichiers numériques pour préservation à long terme, si variables que soient les termes employés pour les décrire. Bref, toutes les politiques ou stratégies de préservation qui font une distinction entre les formats de fichiers acceptables et inacceptables devraient en tenir compte. Pourtant, de nombreuses institutions reconnaissent tacitement qu'il n'existe pas toujours de formats réunissant toutes ces caractéristiques, de sorte qu'il faut souvent trouver des formats quand même acceptables. Nous avons vu par exemple que le caractère exclusif d'un format peut être considéré comme mitigé par l'existence de spécifications largement diffusées ou par son degré d'utilisation et de soutien dans l'industrie des logiciels. Dans bien des cas, il n'existe tout simplement aucun format satisfaisant aux critères désirés. Bibliothèque et Archives Canada admet avoir tenté « d'équilibrer les exigences de qualité, de stabilité, d'éventuelle longévité et d'acceptation de l'industrie » en précisant n'avoir sélectionné des formats non exclusifs ou standard *de facto* que lorsque c'était possible. <sup>55</sup> Le Florida Center for Library Automation recommande « comme règle générale » <sup>56</sup> des formats indépendants des plates-formes et non exclusifs. Dans leurs conseils à l'intention de ceux qui déposent des données, les Archives de données du Royaume-Uni stipulent qu'on « devrait utiliser des formats ouverts dans toute la mesure du possible. Ensuite, il faudrait envisager de se servir de formats disponibles comme RTF Microsoft et PDF Adobe. On ne devrait songer à des formats exclusifs qu'en dernier recours ». <sup>57</sup>

Il est plus difficile encore de trouver des formats idéaux dans les domaines scientifique et artistique. La Bibliothèque du Congrès a souligné que l'interopérabilité est particulièrement importante lorsqu'il s'agit de préserver des données scientifiques, puisque « les ensembles de données scientifiques établis à partir des données de capteurs peuvent n'avoir aucune utilité s'il n'existe pas de logiciels spécialisés pour les analyser et les visualiser, des logiciels qui risquent d'être eux-mêmes difficiles à soutenir même avec le code source disponible ». <sup>58</sup> La NASA et le Goddard Space Center sont conscients de ce problème, et c'est pourquoi ils ont conçu leur format de données commun (*Common Data Format*), avec un langage de balisage en XML appelé Common Data Format Markup Language (CDFML) pour échanger et préserver des

<sup>54</sup> Arms et Fleischhauer, *Sustainability Factors*, *op. cit.*

<sup>55</sup> Brown et Swan, *Lignes directrices*, *op. cit.*, section 1.3, *Concept*.

<sup>56</sup> FCLA, *Recommended Data Formats*, *op. cit.*, p. 2.

<sup>57</sup> UK Data Archive, *Preservation Policy*, *op. cit.*, p. 20. Cet organisme définit les « normes disponibles » comme des formats exclusifs dont les spécifications sont largement diffusées, et les normes « exclusives » comme des « formats appartenant à une compagnie qui ne sont pas généralement disponibles ». *Ibid.*

<sup>58</sup> Arms et Fleischhauer, *Sustainability of Digital Formats*, *op. cit.*



données astronomiques et autres recueillies par télécaptures. De nombreuses autres initiatives font appel à des formats normalisés pour la conservation des données scientifiques, notamment le Système de données planétaires (PDS)<sup>59</sup> et le Format d'archives normalisé pour l'Europe (SAFE); à cela s'ajoute la création de plusieurs langages de balisage en XML, comme le Langage balisé en géographie (GML), le Langage d'échanges scientifiques extensible (XSIL), l'Unité de données formatées en XML (XFDU), le Langage de balisage chimique (CML), le Langage de balisage des matériaux (MatML) et l'Initiative de documentation des données (DDI), pour les sciences sociales et les sciences du comportement. Beaucoup de ces initiatives n'en sont toutefois encore qu'à leurs premières étapes, et la communauté scientifique n'y a toujours pas largement souscrit.<sup>60</sup> En outre, elles n'ont pas nécessairement été acceptées non plus comme des formats standard de transfert aux dépôts de données numériques, essentiellement parce que bien peu des dépôts établis préservent des données scientifiques.<sup>61</sup> Les difficultés sont analogues pour les œuvres d'art numérisées. Le format le plus largement accepté pour la conservation des images numériques est TIFF, et la Research Libraries Group and Digital Library Federation recommande ce format exclusif parce que, « de tous les formats actuellement disponibles, TIFF est celui qu'on peut considérer comme le plus "archivistique" ». <sup>62</sup> Dans le cas des fichiers audionumériques, où le codage avec compression pose un problème critique, les auteurs de l'étude de l'AHDS ont conclu que:

Avec le meilleur scénario, nous aurions des fichiers originaux bruts ou des trains de données sonores non « édités » à préserver, mais il est peu probable qu'on se propose de préserver de tels objets dans la plupart des cas... Toutes choses égales par ailleurs, le meilleur objet à préserver fait appel à un codage conservant un maximum de profondeur de bits avec la fréquence d'échantillonnage la plus élevée et un minimum de compression.<sup>63</sup>

Richard Rinehart, du Musée d'art et des archives cinématographiques du Pacifique de l'Université de la Californie à Berkeley, a déclaré que son institution sélectionne « un petit ensemble de formats par défaut, en ne déviant de ce choix que si nécessaire. » Il a précisé que cette stratégie fonctionne assez bien pour la documentation sur les œuvres d'art, quoique pas pour les œuvres elles-mêmes:

<sup>59</sup> Voir le rapport final de l'étude de cas 8 d'InterPARES 2 : William Underwood, *Mars Global Surveyor Data Records in the Planetary Data System: A Case Study*, septembre 2005 (ébauche).

<sup>60</sup> L'Initiative de documentation des données (*Data Documentation Initiative*) est l'un des plus avancés de ces projets; elle a publié la première version publique de sa DTD en 2001, et cette DTD a été utilisée par la California Digital Library ainsi que par d'autres grands dépôts. Voir le site internet de la DDI, à <http://www.icpsr.umich.edu/DDI/>. Pour une analyse de l'élaboration des formats et des autres normes applicables aux données dans les sciences en général, voir John Rumble, Jr. et al., *Developing and Using Standards for Data and Information in Science and Technology*. Tennessee : Information International Associates, 2006. <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/025.pdf>.

<sup>61</sup> Les données géospatiales font parfois exception, parce que le SIG s'applique largement dans le secteur du gouvernement électronique. La NARA et BAC, par exemple, acceptent les données géospatiales en précisant quels sont les formats acceptables pour leur transfert.

<sup>62</sup> Frey, *File Formats for Digital Masters*, op. cit., section 2.3 *Pros and Cons of Various Formats*.

<sup>63</sup> Wilson et al., *Moving Images and Sound Archiving Study*, op. cit., p. 54.

Nous nous écartons principalement des formats standard dans nos collections numériques d'art et de cinéma dès le départ. Ces œuvres d'art numériques nous arrivent dans divers formats... Les formats de préservation d'une partie de ces œuvres d'art numériques sont fonction de la nature des œuvres elles-mêmes. Nous comptons adopter une stratégie médiatique variable qui nous permettra de transférer des œuvres d'art numérique dans de nouveaux médias au fur et à mesure qu'ils évolueront, mais chaque œuvre d'art est pour nous un cas d'espèce dans lequel nous devons nous conformer aux désirs de l'artiste, en respectant notre contrat avec lui.<sup>64</sup>

Le secteur du gouvernement électronique paraît plus prometteur en ce qui concerne la production de documents d'archives dans des formats qui se prêtent à leur préservation en raison de la taille des organisations et des éléments communs des programmes et des services qui tendent à s'imposer avec le temps. Tous les gouvernements ont des systèmes de traitement de texte et de bases de données, par exemple, et beaucoup d'entre eux sont en train d'adopter des plates-formes de service normalisées basées sur Internet ainsi que des bases de données basées sur le SIG. En outre, les organismes gouvernementaux ont besoin d'échanger des données aussi bien à l'interne qu'entre eux, et cette démarche sera vraisemblablement bien soutenue dans l'avenir grâce à des formats basés sur XML, comme le format de document ouvert conçu par l'Organisation pour l'avancement des normes sur l'information structurée (OASIS).<sup>65</sup> À l'heure actuelle, quand on se sert de systèmes exclusifs, la normalisation (c.-à-d. la conversion à des formats propices à la préservation) est plus susceptible de se faire dans le cas des courriels, des fichiers de traitement de texte, des bases de données et des documents typiquement produits par les gouvernements plutôt que dans celui des éléments comparables créés dans le contexte de démarches scientifiques ou artistiques très spécialisées.

### 3. Implications pour les politiques

Jusqu'à présent, nous n'avons pas abordé la question de savoir si un dépôt d'archives devrait limiter le nombre de formats de fichiers qu'il accepte pour préservation. Bien entendu, la plupart des institutions étudiées ont fait savoir que telle est leur politique soit explicitement, soit implicitement. L'adoption d'une telle politique est purement une question de choix basée sur des facteurs commerciaux pour les dépôts de documents numériques de tierces parties fonctionnant comme des entreprises et acceptant de préserver à contrat des documents d'archives numériques. Cette situation soulève toutefois des

<sup>64</sup> Courriel à Yvonne Loiselle, 11 août 2005.

<sup>65</sup> Voir Organization for the Advancement of Structured Information Standards, *OASIS Open Document Format for Office Applications (OpenDocument) Technical Committee FAQ*, 2006, à <http://www.oasis-open.org/committees/office/fag.php>. L'élaboration d'OpenDocument a été largement inspirée par les besoins du monde des affaires; toutefois, étant donné que les activités du secteur privé et des gouvernements génèrent souvent les mêmes types de documents, on s'attend à ce que ce format soit largement applicable aussi bien dans les administrations gouvernementales que dans les entreprises.

dilemmes pour d'autres types d'institutions. Les archives gouvernementales devraient-elles refuser l'accession de certains formats utilisés par leurs institutions mères? Ont-elles le pouvoir de le faire? L'éthique le leur permet-elle? Pour les organisations qui préservent des documents d'archives artistiques et scientifiques, une politique comme celle-là aboutira-t-elle à une situation dans laquelle des documents d'archives seront perdus ou des processus créatifs ou scientifiques indûment sapés par les préférences des dépôts de documents numériques?

Plusieurs grands dépôts comme le MIT, le Florida Center for Library Automation, les OCLC Digital Archive, la California Digital Library et le NARA ont tenté de surmonter ces difficultés en offrant différents degrés de préservation en fonction du format. C'est ainsi que le guide de la politique des archives numériques du FCLA précise que:

N'importe quel format de fichiers peut être déposé dans les FDA [Florida Digital Archives]. Toutefois, seuls les fichiers dans des formats soutenus peuvent bénéficier de services de préservation intégrale visant à en maintenir l'utilité. Les fichiers dans des formats non soutenus seront préservés dans leur version originale (soumise) seulement, avec une préservation au niveau des bits.<sup>66</sup>

La politique du MIT ressemble beaucoup à celle du FCLA:

Au MIT, pour le moment, nous reconnaissons que les formats dans lesquels les professeurs créent leur matériel de recherche sont impossibles à prédire ou contrôler... C'est pour cette raison que nous avons défini trois degrés de préservation pour un format donné : soutenu, connu, non soutenu. Les formats soutenus bénéficieront d'une préservation fonctionnelle grâce à des techniques de migration ou d'émulation des formats. Les formats connus sont ceux que nous ne pouvons pas promettre de préserver (p. ex. les formats exclusifs ou binaires), mais qui sont si répandus que nous croyons que des tiers créeront des outils de migration pour faciliter la migration entre formats. Enfin, les formats non soutenus sont ceux sur lesquels nous n'en savons pas assez pour assurer une préservation fonctionnelle quelconque. À chacun de ces trois degrés, nous entreprendrons une préservation au niveau des bits afin que les « archéologues numériques » de l'avenir aient la matière brute sur laquelle ils pourront travailler si le contenu en vaut la peine.<sup>67</sup>

Pour sa part, la California Digital Library « garantit seulement la préservation du train de bits original » pour les « formats de fichiers nouveaux ou

---

<sup>66</sup> Florida Center for Library Automation, *FCLA Digital Archive (FDA) Policy Guide*. Gainesville (Floride) : Florida Center for Library Automation, version 1.1 décembre 2004, p. 1.

[http://www.fcla.edu/digitalArchive/pdfs/DigitalArchivePolicyGuide1\\_1.pdf](http://www.fcla.edu/digitalArchive/pdfs/DigitalArchivePolicyGuide1_1.pdf). Le *Guide* précise que, pour les objets logiques incluant des fichiers « en formats tant soutenus que non soutenus, il n'existe aucune garantie que l'objet logique restera utilisable comme on l'entendait. »

<sup>67</sup> MIT, *General DSpace FAQ*, *op. cit.*

inconnus ». <sup>68</sup> Le NARA offre trois degrés de préservation, essentiellement en fonction du format, à savoir Préservation et accès de base, Préservation accrue et Accès et préservation et accès optimaux. <sup>69</sup> Cela dit, accepter des formats de fichiers sans s'engager à assurer leur protection intégrale pose des problèmes pour diverses raisons, puisque cela peut mener à un sentiment de sécurité injustifié chez les donateurs des documents d'archives, qui peuvent se conforter d'avoir fait don de leurs documents d'archives à des dépôts de documents numériques respectés et bien établis, en partant du principe que ces documents vont être accessibles d'une façon ou d'une autre dans l'avenir, même s'ils ne sont pas dans un des formats qu'on préfère actuellement. Il peut en résulter des problèmes stratégiques et éthiques pour les archives gouvernementales, dont le mandat consiste non seulement à acquérir, mais aussi à préserver les documents d'archives de leurs institutions mères ainsi qu'à les rendre accessibles. C'est aussi un engagement d'assumer une charge de travail indéterminée peut-être sans soutien dans l'avenir, sans aucune garantie de succès. Il n'en reste pas moins que c'est une tentative pour répondre aux besoins de diverses populations de déposants, et qu'une telle politique pourrait être nécessaire dans les institutions ayant de vastes mandats d'acquisition.

Quant à savoir si les institutions d'archives devraient limiter le nombre de formats de fichiers qu'elles acceptent, comme on l'a déjà dit, nous estimons qu'elles doivent le faire, mais avec des réserves dépendant de leur mandat et de leurs relations avec leurs institutions mères ainsi qu'avec les organisations de donateurs. Dans la plupart des cas, la question est incontournable sauf si l'organisation d'archives est disposée soit à accepter tous les formats en s'engageant à les préserver tous (il est peu probable qu'elle réussisse à le faire), soit, à l'instar des institutions dont nous avons parlé, à accepter des documents d'archives pour lesquels il n'existe pas de plans de préservation entièrement établis, en postulant qu'on inventera des outils de migration et de rendu dans l'avenir. Le volume de documentation disponible sur le thème des formats de fichiers désirables confirme que les gestionnaires de dépôts de documents numériques sont arrivés à la conclusion qu'il est souvent nécessaire de limiter le nombre des formats de fichiers pour accession. Cela signifie-t-il nécessairement que certains documents d'archives numériques sont destinés à être éphémères dès leur création? La réponse à cette question dépend en partie de la sensibilité du dépôt intéressé aux besoins de ses institutions mères et de ses donateurs. Le fait que l'utilisation largement répandue d'un format donné est considérée comme un facteur important par tant des institutions que nous avons étudiées laisse entendre que les besoins des archives et ceux d'au moins certains des créateurs coïncident. L'interopérabilité et la rétrocompatibilité sont aussi des caractéristiques propices à la préservation aussi susceptibles de répondre aux besoins des créateurs de documents d'archives que de satisfaire aux exigences des institutions d'archives elles-mêmes. Ce sera vraisemblablement le cas

<sup>68</sup> CDL Digital Library Services Advisory Group, *CDL Guidelines for Digital Objects*. California Digital Library : version 2.0, ébauche, novembre 2005, p. 5. [http://www.cdlib.org/inside/diglib/guidelines/cdl\\_gdo\\_v2\\_draft.pdf](http://www.cdlib.org/inside/diglib/guidelines/cdl_gdo_v2_draft.pdf).

<sup>69</sup> Electronic Records Archives Program Management Office, *Introduction to Preservation and Access Levels Concepts*. Maryland : National Archives and Records Administration, 5 décembre 2003, p. 7 et 8. <http://www.archives.gov/era/pdf/preservation-and-access-levels.pdf>.

particulièrement dans les secteurs du gouvernement électronique et des sciences, mais moins dans celui des arts, où le processus constant de création individuelle est moins tributaire de l'échange de données entre des institutions.

Les archivistes peuvent choisir d'aller plus loin qu'une simple limitation des formats que leurs archives acceptent : ils peuvent activement encourager l'application de normes sur les formats employés pour la création des documents d'archives. À cet égard, il faut tenir compte de plusieurs facteurs liés notamment au fait que les activités qui génèrent des documents d'archives se déroulent ou non dans les secteurs du gouvernement électronique, des sciences ou des arts. Dans le cas des archives gouvernementales, préconiser tels ou tels formats pour la création de documents d'archives numériques peut être considéré comme une tentative d'encourager des pratiques exemplaires pour la création de documents d'archives en général. L'institution d'archives gouvernementale peut acquérir une maîtrise des formats de fichiers préservables dont ses institutions mères peuvent tirer profit lorsqu'elles envisagent l'élaboration de nouveaux programmes et services de gouvernement électronique, et peut-être devrait-elle le faire. Bibliothèque et Archives Canada est une institution d'archives gouvernementale qui encourage explicitement l'utilisation de certains formats pour la création de documents d'archives, et de formats qui se prêtent aussi au transfert de données et de documents d'archives entre les composantes de l'administration fédérale.<sup>70</sup> Les institutions d'archives gouvernementales ont toujours eu la responsabilité d'encourager une bonne « tenue des livres » dans leurs institutions mères et d'assurer la préservation des documents d'archives sur les activités gouvernementales. Recommander l'emploi de formats de fichiers particuliers pour la création des documents d'archives en fonction de leur potentiel de préservation à long terme peut être une extension de cette fonction traditionnelle.

On pourrait en dire largement autant pour les institutions qui préservent des documents d'archives scientifiques, puisque les besoins fondamentaux d'échange de données et de fiabilité des documents d'archives préservés pour étayer les résultats de la recherche ressemblent énormément aux besoins d'échange de données et d'imputabilité des gouvernements. Bien qu'il ne soit pas indispensable que les dépôts de documents numériques dictent les types de documents d'archives à créer, ils n'en peuvent pas moins indiquer clairement quelles caractéristiques des formats de fichiers sont les plus susceptibles de garantir la préservation des documents d'archives scientifiques. Les types de documents d'archives créés, la relation juridique entre les dépôts et les créateurs de ces documents et les mandats des premiers comme des seconds influent tous sur les interactions entre les institutions d'archives et les créateurs quant à la sélection des formats de création des documents d'archives.

Il est plus difficile d'en dire autant en ce qui concerne les arts. Dans bien des cas où il est question d'art numérique, le processus d'écrire un code et de concevoir des formats de fichiers peut faire partie d'un processus de création artistique dans lequel les institutions d'archives ne devraient pas s'ingérer. Néanmoins, elles peuvent toujours acquérir une maîtrise de la création et de la sélection des formats dont la communauté artistique peut bénéficier. Cette

---

<sup>70</sup> Brown et Swan, *Lignes directrices*, op. cit., section 1.1, *But et portée*.

relation peut se manifester entre le consultant (l'institution d'archives) et le client (le créateur), ou n'être tout simplement que le dialogue permanent entre ceux qui préservent les documents et leurs créateurs.<sup>71</sup> Il serait même possible que l'institution d'archives fasse le pont entre des groupes d'artistes, lorsqu'un groupe s'est fié à elle pour préserver une œuvre d'art et qu'un autre tire parti de cette expérience pour créer ses propres œuvres.

#### 4. Recommandations pour l'élaboration et la mise en œuvre des politiques

Déterminer les critères à employer pour la sélection des formats de fichiers en vue de leur préservation à long terme et, si nécessaire, recommander ces formats pour la création de documents d'archives sont des démarches qui devraient être considérées comme positives pour le développement de la capacité d'une institution d'archives de préserver des documents d'archives numériques. C'est pour cette raison que notre rapport se termine avec les recommandations suivantes pour l'élaboration et la mise en œuvre de politiques sur la sélection de formats de fichiers numériques pour préservation à long terme:

1. Clarifier la terminologie : déterminer ce qu'on entend par des termes comme *ouvert*, *standard*, *stable* et *étayé par une solide documentation*, et définir ces termes dans les politiques.
2. Faire une distinction entre les formats de fichiers eux-mêmes, les formats enveloppeurs (ou conteneurs) et les formats balisés, comme les fichiers balisés en XML, et veiller à ce que les versions et les autres caractéristiques telles que le codage soient comprises et entièrement spécifiées.
3. Dans le cas des fichiers en XML, exiger qu'ils soient bien formés, valides et accompagnés des DTD ou des schémas pertinents.
4. Dans toute la mesure du possible, sélectionner des formats largement utilisés, non exclusifs, indépendants des plates-formes et ayant des spécifications largement diffusées.
5. Préciser si les fichiers comprimés sont acceptables et, si oui, préciser le type de compression permise. Dans toute la mesure du possible, opter pour des techniques de compression sans perte de données et conformes aux normes internationales acceptées.
6. S'il est impossible de sélectionner des formats ayant les caractéristiques énumérées dans la recommandation 4, opter pour des

---

<sup>71</sup> C'est dans une certaine mesure l'approche du Variable Media Network, qui « apparie des artistes avec des consultants des musées et des médias » pour convertir leurs œuvres d'art analogiques et/ou numériques sous une forme compatible avec les nouveaux médias ou pour les rendre plus indépendantes des médias. Voir le site internet de ce projet, à <http://variablemedia.net/>.

formats préservés dans d'autres dépôts de documents numériques et collaborer avec ces dépôts pour concevoir des plans de préservation à leur égard.

7. Dans toute la mesure du possible, travailler de concert avec les créateurs de documents d'archives pour veiller à ce qu'ils utilisent des logiciels capables de créer ces documents dans des formats répondant aux critères énumérés dans la recommandation 4.



## Annexe A : Liste des dépôts examinés

La documentation accessible en direct sur les dépôts de documents numériques suivants a été examinée :

- Art Institute of Chicago, Department of Architecture
- Arts and Humanities Data Service, United Kingdom
- California Digital Library
- Cornell University Library
- Digital Image Archive of Medieval Music (DIAMM), Universities of Oxford and London
- Florida Center for Library Automation
- Library and Archives Canada
- Library of Congress
- Massachusetts Institute of Technology
- National Archives of Australia
- National Archives of the United Kingdom
- U.S. National Archives and Records Administration
- Netherlands Institute for Scientific Information Services
- Ohio Electronic Records Committee
- On-line Computer Library Center, United States
- Public Records Office of Victoria
- State and University Library, Arhus, and the Royal Library, Copenhagen
- Technical Advisory Service for Images, United Kingdom
- UC Berkeley Art Museum/Pacific Film Archive
- UK Data Archive

La documentation relative aux groupes de collaboration (ou non institutionnels) suivants a aussi été étudiée, parce qu'elle contient des recommandations spécifiques pour la sélection des formats de fichiers à préserver :

- DAVID Project (Digitale Archivering in Vlaamse Instellingen en Diensten, or Digital Archiving in Flemish Institutions and Administrations)
- Digital Preservation Coalition, United Kingdom
- EU-US Working Group on Spoken-Word Audio Collections
- Research Libraries Group and Digital Library Federation

## Annexe B : URL des documents examinés, par dépôt\*

Voici une liste des URL visités pour l'analyse quantitative des critères des formats de fichiers présentée à la section 2. Les références intégrales des documents examinés figurent dans la Bibliographie, à la fin du rapport.

Art Institute of Chicago, Department of Architecture

[http://www.artic.edu/aic/collections/dept\\_architecture/dddreport/0C.pdf](http://www.artic.edu/aic/collections/dept_architecture/dddreport/0C.pdf)

Arts and Humanities Data Service, United Kingdom

<http://ahds.ac.uk/depositing/deposit-formats.htm>

California Digital Library

<http://www.cdlib.org/inside/diglib/guidelines/bpgimages/reqs.html#reqformats>

[http://www.cdlib.org/inside/diglib/guidelines/cdl\\_gdo\\_v2\\_draft.pdf](http://www.cdlib.org/inside/diglib/guidelines/cdl_gdo_v2_draft.pdf)

Cornell University Library

<http://www.library.cornell.edu/imls/IMLS-CULfinalreport2.pdf>

Digital Image Archive of Medieval Music (DIAMM), Universities of Oxford and London

<http://www.diamm.ac.uk/content/description/archiving.html>

<http://www.diamm.ac.uk/content/description/quality.html>

Florida Center for Library Automation

<http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>

Library and Archives Canada

<http://www.collectionscanada.ca/information-management/002/007002-3017-e.html>

Library of Congress

<http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>

Massachusetts Institute of Technology

<http://libraries.mit.edu/dspace-mit/about/faq.html>

National Archives of Australia

[http://www.naa.gov.au/recordkeeping/rkpubs/fora/02nov/digital\\_preservation.pdf](http://www.naa.gov.au/recordkeeping/rkpubs/fora/02nov/digital_preservation.pdf)

National Archives of the United Kingdom

[http://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf)

---

\* Visite la plus récente de tous ces sites le 3 décembre 2006.

Netherlands Institute for Scientific Information Services

[http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna\\_Horik.pdf](http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna_Horik.pdf)

Ohio Electronic Records Committee

<http://ww.ohiojunction.net/erc/imagingrevision/revisedimaging2003.html>

On-line Computer Library Center, United States

<http://www.dlib.org/dlib/november04/stanescu/11stanescu.html>

<http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf>

Public Records Office of Victoria

[http://www.prov.vic.gov.au/vers/standard/advice\\_13/](http://www.prov.vic.gov.au/vers/standard/advice_13/)

State and University Library, Arhus, and the Royal Library, Copenhagen

<http://netarchive.dk/publikationer/FileFormats-2004.pdf>

Technical Advisory Service for Images, United Kingdom

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

<http://www.tasi.ac.uk/advice/creating/ffformat.html#ff3>

UK Data Archive

<http://www.data->

[archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf](http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf)

U.S. National Archives and Records Administration

<http://www.archives.gov/research/arc/digitizing-archival-materials.pdf>

<http://www.archives.gov/records-mgmt/initiatives/erm-guidance.html>

<http://www.archives.gov/about/regulations/part-1228/l.html>

DAVID Project (Digitale Archivering in Vlaamse Instellingen en Diensten, or Digital Archiving in Flemish Institutions and Administrations)

<http://www.expertisecentrumdavid.be/davidproject/teksten/guideline4.pdf>

Digital Preservation Coalition, United Kingdom

<http://ww.dpconline.org/graphics/handbook>

EU-US Working Group on Spoken-Word Audio Collections

<http://delos-noe.iei.pi.cnr.it/activities/internationalforum/Joint-WGs/spokenword/SpokenWord.pdf>

Research Libraries Group and Digital Library Federation

<http://www.rlg.org/legacy/visguides/>

## Bibliographie\*

Arms, Caroline R., and Carl Fleischhauer. *Sustainability of Digital Formats: Planning for Library of Congress Collections*. Washington, D.C.: Library of Congress, updated March 6, 2006.

<http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>.

Art Institute of Chicago, Department of Architecture. *Collecting, Archiving and Exhibiting Digital Design Data*. Chicago: Kristine Fallon Associates Inc., 2004.

[http://www.artic.edu/aic/collections/dept\\_architecture/dddreport/0C.pdf](http://www.artic.edu/aic/collections/dept_architecture/dddreport/0C.pdf).

Beagrie, Neil, and Maggie Jones. *Digital Preservation Coalition Handbook*. United Kingdom: Digital Preservation Coalition, updated August 2006.

<http://www.dpconline.org/graphics/handbook>.

Brown, Adrian, *Digital Preservation Guidance Note 1: Selecting File Formats for Long-Term Preservation*. Surrey, UK: National Archives of the United Kingdom, June 19, 2003.

[http://www.nationalarchives.gov.uk/documents/selecting\\_file\\_formats.pdf](http://www.nationalarchives.gov.uk/documents/selecting_file_formats.pdf).

Brown, David L., and Mike Swan. *Guidelines for Computer File Types, Interchange Formats and Information Standards*. Ottawa: Library and Archives Canada, version 1.1, June 28, 2004.

<http://www.collectionscanada.ca/information-management/002/007002-3017-e.html>.

CDL Digital Library Services Advisory Group. *CDL Guidelines for Digital Images*, chapter 3, *Requirements*, March 10, 2005.

<http://www.cdlib.org/inside/diglib/guidelines/bpgimages/regs.html#reqformats>.

\_\_\_\_\_. *CDL Guidelines for Digital Objects*. California Digital Library Advisory Group: version 2.0, draft, November 2005.

[http://www.cdlib.org/inside/diglib/guidelines/cdl\\_gdo\\_v2\\_draft.pdf](http://www.cdlib.org/inside/diglib/guidelines/cdl_gdo_v2_draft.pdf).

Clausen, Lars R., main author. *Handling File Formats*. Denmark: State and University Library, Aarhus, and the Royal Library, Copenhagen, May 2004.

<http://netarchive.dk/publikationer/FileFormats-2004.pdf>.

Data Documentation Initiative, Data Documentation Initiative Alliance, 2006.

<http://www.icpsr.umich.edu/DDI/>.

DAVID Project, *Digital Archiving, Guideline and Advice 4: Standards for Fileformats*. Antwerp, 2003.

<http://www.expertisecentrumdavid.be/davidproject/teksten/guideline4.pdf>.

---

\* Visite la plus récente de tous ces sites le 3 décembre 2006.

Davis, Simon. *Recordkeeping Issues Forum: Digital Preservation Strategy*. Australia: National Archives of Australia, November 19, 2002. [http://www.naa.gov.au/recordkeeping/rkpubs/fora/02nov/digital\\_preservation.pdf](http://www.naa.gov.au/recordkeeping/rkpubs/fora/02nov/digital_preservation.pdf).

Digital Image Archive of Medieval Music (DIAMM) Web site. United Kingdom: Universities of Oxford and London, June 15, 2006. <http://www.diamm.ac.uk/content/description/archiving.html>.

Florida Center for Library Automation. *FCLA Digital Archive (FDA) Policy Guide*. Gainesville, Florida: Florida Center for Library Automation, version 1.1 December 2004. [http://www.fcla.edu/digitalArchive/pdfs/DigitalArchivePolicyGuide1\\_1.pdf](http://www.fcla.edu/digitalArchive/pdfs/DigitalArchivePolicyGuide1_1.pdf).

\_\_\_\_\_. *Recommended Data Formats for Preservation Purposes in the FCLA Digital Archive*. Gainesville, Florida: Florida Center for Library Automation, June 2005. <http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>.

Frey, Franziska. *Guides to Quality in Visual Resource Imaging: 5. File Formats for Digital Masters*. United States: Research Libraries Group and Digital Library Federation, 2000. <http://www.rlg.org/legacy/visguides/>.

James, Hamish. *AHDS Deposit Formats*. Arts and Humanities Data Service, United Kingdom, August 22, 2006. <http://ahds.ac.uk/depositing/deposit-formats.htm>.

Kenney, Anne R., et al. *Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy: Final Project Report*. Ithaca, New York: Cornell University Library, May 2001. <http://www.library.cornell.edu/imls/IMLS-CULfinalreport2.pdf>.

Massachusetts Institute of Technology. *General DSpace FAQ*. Cambridge, Massachusetts: MIT Libraries, undated. <http://libraries.mit.edu/dspace-mit/about/faq.html>.

Ohio Electronic Records Committee, *Revised Digital Imaging Guidelines: Guidelines for State of Ohio Executive Agencies and Local Governments*. Ohio: Ohio Electronic Records Committee, June 26, 2003. [www.ohiojunction.net/erc/imagingrevision/revisedimaging2003.html](http://www.ohiojunction.net/erc/imagingrevision/revisedimaging2003.html).

On-Line Computer Library Center, *OCLC Digital Archive Preservation Policy and Supporting Documentation*, Dublin, Ohio, August 8, 2006. <http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf>.

Organization for the Advancement of Structured Information Standards. *OASIS Open Document Format for Office Applications (OpenDocument) Technical Committee FAQ*. 2006. <http://www.oasis-open.org/committees/office/faq.php>.

Preservation Metadata: Implementation Strategies (PREMIS). *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*. United States: On-Line Computer Library Center and Research Libraries Group, May 2005. <http://www.oclc.org/research/projects/pmwg/premis-final.pdf>.

Public Records Office of Victoria. *Advice 13: Long-Term Preservation Formats*. North Melbourne: Public Records Office Victoria, September 2, 2004. [http://www.prov.vic.gov.au/vers/standard/advice\\_13/](http://www.prov.vic.gov.au/vers/standard/advice_13/).

Puglia, Steve, Jeffrey Reed and Erin Rhodes, *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*. U.S. National Archives and Records Administration, June 2004. <http://www.archives.gov/research/arc/digitizing-archival-materials.pdf>.

Renals, Steve, and Jerry Goldman, et al. *EU-US Working Group on Spoken-Word Audio Collections, Final Report*. European Union and United States: EU-US Working Group on Spoken-Word Audio Collections, June 18, 2003. <http://www.dcs.shef.ac.uk/spandh/projects/swag/swagReport.pdf>.

Rinehart, Richard, and Guenter Waibel. *Strategies for Digital Media Asset Management*. California: UC Berkeley Art Museum/Pacific Film Archive, April 26, 2001. Unpublished internal policy document.

Rumble, Jr., John, et al. *Developing and Using Standards for Data and Information in Science and Technology*. Tennessee: Information International Associates, 2006. <http://www.ukoln.ac.uk/events/pv-2005/pv-2005-final-papers/025.pdf>.

Stanescu, Andreas. "Assessing the Durability of Formats in a Digital Preservation Environment: the INFORM Methodology," *D-Lib Magazine* 10(11), November 2004. <http://www.dlib.org/dlib/november04/stanescu/11stanescu.html>.

Technical Advisory Service for Images. *Advice Paper: Choosing a File Format*. United Kingdom: Technical Advisory Services, May 2006. <http://www.tasi.ac.uk/advice/creating/pdf/format.pdf>.

\_\_\_\_\_, *File Formats and Compression*. United Kingdom: Technical Advisory Service for Images, March 2005. <http://www.tasi.ac.uk/advice/creating/ffformat.html#ff3>.

UC Berkeley Art Museum/Pacific Film Archive, e-mail from Richard Rinehart to Yvonne Loiselle, August 11, 2005.

UK Data Archive. *UK Data Archive Preservation Policy*. Colchester: University of Essex, version 2.0, September 2005.

<http://www.data-archive.ac.uk/news/publications/UKDAPreservationPolicy0905.pdf>.

U.S. National Archives and Records Administration. *NARA Electronic Records Management (ERM) Guidance on the Web*. Maryland, National Archives and Records Administration, various dates. <http://www.archives.gov/records-mgmt/initiatives/erm-guidance.html>. Various sections starting with the title *Expanding Acceptable Transfer Requirements*....

\_\_\_\_\_, *Regulations, Sub-chapter B, Records Management, Part 1228 -- Disposition of Federal Records, Subpart L -- Transfer of Records to the National Archives of the United States*.

<http://www.archives.gov/about/regulations/part-1228/l.html>.

\_\_\_\_\_, *Electronic Records Archives Program Management Office. Introduction to Preservation and Access Levels Concepts*. Maryland: National Archives and Records Administration, December 5, 2003.

<http://www.archives.gov/era/pdf/preservation-and-access-levels.pdf>.

Valoris. *Comparative Assessment of Open Documents Formats: Market Overview*. European Commission, December 30, 2003.

<http://europa.eu.int/idabc/servlets/Doc?id=17982>.

van Horik, Rene. *Image Formats: Practical Experiences*. Netherlands: Netherlands Institute for Scientific Information Services, 2004. Erpanet presentation, Vienna, May 2004.

[http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna\\_Horik.pdf](http://www.erpanet.org/events/2004/vienna/presentations/erpaTrainingVienna_Horik.pdf).

Variable Media Network, Guggenheim Museum and Daniel Langlois Foundation for Art, Science, and Technology. Web site, undated. <http://variablemedia.net/>.

Wilson, Andrew, et al., *Moving Images and Sound Archiving Study*. Arts and Humanities Data Service, United Kingdom: Final Draft, June 2006.

<http://roda.iantht.pt/?q=en/system/files/Moving+Images+and+Sound+Archiving+Study1.doc>.