



MINISTÈRE DE
L'ÉDUCATION NATIONALE

MINISTÈRE DE
L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE



CRDP de l'académie de Versailles

2 rue Pierre-Bourdan
78160 Marly-le-Roi
Tél. 01 78 64 52 00

Formats de documents, avoir de bonnes pratiques

4 septembre 2013

Mission TICE Médiapôles

<http://www.tice.ac-versailles.fr/-Médiapoles-.html>



Document réalisé par :



ÉNUMA Logiciels Libres

Louis-Maurice De Sousa
François Lermigeaux



Ce document est mis à disposition selon les termes de la
Licence Creative Commons Paternité - Partage à l'identique.

Document réalisé sous [Debian GNU/Linux](#) avec [L^AT_EX](#)

Ce document a pour objet de définir un certain nombre de règles pour l'échange de documents au sein de la mission Tice académique. Ces préconisations ne concernent que les formats d'échange, chacun étant libre d'utiliser pour son propre compte le logiciel et le format de son choix.

1 Introduction

Le développement d'Internet à la fin du 20^{ème} siècle a considérablement multiplié les possibilités de partage et d'échange de documents. Ces possibilités ont pourtant été grandement diminuées par des pratiques issues de l'utilisation de l'ordinateur avant l'arrivée de l'Internet. Chaque logiciel utilise généralement un format de document qui lui est propre. Il est devenu courant d'échanger des documents dans ce format propre au logiciel sans se préoccuper de savoir si le destinataire du document a la possibilité d'en prendre connaissance.

À ces premières préoccupations s'ajoute la problématique de l'archivage à moyen terme des documents produits : il faut s'assurer qu'ils restent lisibles dans le temps, ce qui implique de choisir un format d'archivage et d'avoir une homogénéité certaine de ses fichiers, de manière à pouvoir les faire évoluer facilement dans le temps.

Cette problématique sera tout particulièrement forte dans les usages qui imposent la constitution d'un fond d'archives, ce qui est le cas des établissements scolaires.

Il convient donc toujours de se poser la question de savoir ce qu'on fera du document dans, 5, 10 ou 15 ans.



Le lecteur de ce document trouvera beaucoup de prolongements sur le wiki de l'université Paris 5 :

<http://wiki.univ-paris5.fr/wiki/Format> (dernière mise à jour en 2009)

2 Bureautique

L'un des types de documents le plus communément échangé est le document bureautique, texte, feuille de calcul ou diaporama. Comme les autres logiciels, les logiciels de bureautique utilisent souvent, chacun, un format de document spécifique qui n'est lisible que par le logiciel lui-même. Conscients des problèmes que cela pose dans l'entreprise, comme dans l'administration, différents acteurs ont proposé de définir une norme. En 2006, l'organisme de normalisation ISO a donc défini comme une norme¹ le format OpenDocument², qui est un format ouvert³. Tous les éditeurs peuvent donc, sans payer de droits particuliers, implémenter ce format de document dont les spécifications sont publiées. Tous les logiciels de bureautique libres utilisent donc ce format de documents et très récemment, MS Office® aussi. Tout autre format de document est donc à exclure même si plusieurs logiciels le lisent, comme les formats `doc` ou `docx`. Cette lecture n'est possible que par un travail de rétro-ingénierie, puisque leurs spécifications ne sont pas publiées. Il n'y a donc aucune garantie d'inter-opérabilité complète du document, ni de recouvrir l'ensemble de ses fonctionnalités.

A été ajouté au type de documents bien connu de traitement de texte ou de tableur, celui de « tableau à feuilles », plus connu sous l'anglicisme de « paperboard », utilisé par les logiciels de TNI (Tableau Numérique Interactif) ou de TBI (Tableau Blanc Interactif). Ces logiciels utilisent tous un format de documents qui leur est propre empêchant l'utilisation d'un document créé par un logiciel particulier d'une marque de TNI, avec le logiciel d'une autre marque de TNI. Cela malgré des fonctionnalités très proches de celles d'un logiciel de présentation assisté par ordinateur, ou de ce que permet le format vectoriel `svg`. Fort heureusement, à l'initiative du BECTA, un nouveau format libre de documents a été développé, à destination de ces logiciels, le format « `iwb` ».



Attention : les formats bureautiques n'intègrent pas les fontes de caractères. Un affichage correct suppose donc l'installation de la même fonte que celle utilisée dans le document. L'usage de fontes libres de droits est donc à privilégier.

	Modifiable	Diffusable	Archivable	Toléré	À exclure
Traitement de texte	.odt, .tex	.pdf	.odt, .tex, .pdf	.doc	.docx ⁴
Tableur	.ods, .csv	.pdf	.ods, .pdf	.xls	.xlsx ³
Diaporama	.odp, .tex	.pdf	.odp, .tex, .pdf	.ppt	.pptx ³
Texte	.txt, .rtf				
Tableau à feuilles	.svg, .odf, .iwb	.pdf			

1. http://www.iso.org/iso/fr/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43485

2. <http://fr.wikipedia.org/wiki/OpenDocument>

3. Un autre format bureautique, poussé par Microsoft®, a été normalisé en 2008, ooXML. Mais il n'est pas complètement techniquement défini, et donc, aucun logiciel, autre que ceux de Microsoft® n'est en mesure de l'utiliser. Ce format ne correspond pas exactement au `docx` et l'un comme l'autre évoluent régulièrement au gré des besoins de Microsoft® <http://en.wikipedia.org/wiki/DOCX>.

4. Et tout format propre à une application particulière, iWorks®, Lotus®...

3 Image

Comme pour la bureautique, la plupart des logiciels de traitement d'image utilisent un format de document qui leur est propre. Il y a une différence toutefois, le « web » a imposé une standardisation.

En effet, pour qu'une image soit vue dans un navigateur, elle doit être dans un format particulier. Ces formats sont `gif`, `png` et `jpeg` pour les images matricielles, et `svg` pour les images vectorielles. Ce sont donc des formats d'échanges privilégiés, assez largement utilisés.

Hormis le `svg`, ces formats ne gardent pas trace des informations de construction (modifications apportées, calques...). Il faut donc toujours garder l'image dans le format du logiciel utilisé et ce format ne permet pas d'échange. Toutefois, l'utilisation d'un logiciel libre n'impose à l'utilisateur, que l'installation du logiciel, et pas l'acquisition d'une licence particulière.

On distingue deux types d'images, les images matricielles, où les images sont enregistrées sous la forme d'une série de points, et les images vectorielles, où les images sont enregistrées sous la forme de formules qui permettent le calcul de l'image.

Avec l'apparition des appareils photos numériques, d'autres problèmes sont apparus. En effet, certaines gammes d'appareils permettent d'enregistrer l'image dans un format brut non compressé, `raw`. Malheureusement, bien qu'il existe un format standard et ouvert, `dng`, beaucoup d'appareils enregistrent l'image dans un format qui est propre à la marque, la rendant inexploitable sur certains systèmes, ou logiciels.

	Modifiable	Diffusable	Archivable	Toléré	À exclure
Matricielle	.bmp, .jpg, .png, .dng	.jpg, .png	.jpg, .png, .dng		.psd, .xcf, .tiff, .raw ⁵
Vectorielle	.svg	.svg	.svg		.ai ⁴

4 Son

Dans de domaine de l'audio, il est fréquent de confondre le conteneur et le codec. `Wav` et `ogg` sont des conteneurs. Lorsque le conteneur `ogg` contient le codec `vorbis`, le terme `ogg` sert souvent à définir le codec, alors que le conteneur `ogg` peut inclure de la vidéo (codec « Theora ») ou des sous-titres.

La plupart des conteneurs et des codecs privés sont payants. Ils ne le sont toutefois pas, ou très rarement, ou à son insu, pour l'utilisateur final. Ainsi, le très répandu codec `mp3` dispose d'un décodeur gratuit pour l'utilisateur d'ordinateur, mais est payant ⁶ pour le fabricant de baladeur ou de téléphone mobile, s'il souhaite permettre la lecture de ce format. Cette stratégie a permis la très grande diffusion de beaucoup de codecs, créant des relations de dépendance problématiques.

Nous devons donc nous tourner progressivement vers des formats libres et gratuits comme `flac` ou `ogg`.

5. Et tout format propre à une application ou un appareil particulier

6. <http://openmoko-fr.org/blog/index.php?post/2008/11/16/Openmoko-contraint-de-retirer-le-support-du-MP2/MP3>

	<i>Modifiable</i>	<i>Diffusable</i>	<i>Archivable</i>	<i>Toléré</i>	<i>À exclure</i>
Compression destructive	.ogg	.ogg	.ogg	.mp3	.wma, .aac
Compression non destructive	.flac	.flac			

5 Vidéo

Dans le domaine de la vidéo, la dissociation entre le conteneur et le codec est plus claire bien que souvent ignorée. `avi` ou `mp4` ne sont que des conteneurs pouvant accueillir beaucoup de codecs différents. On ne peut donc être certain de pouvoir lire un `avi` ou un `mp4`.

Là encore, nous devons nous tourner progressivement vers des formats libres, ouverts et gratuits comme le conteneur `mkv` ou le codec « Theora ».

Un codec particulier crée une situation originale, le H.264. Très efficace, largement utilisé par défaut sur les appareils du commerce, promu par Apple®, Microsoft®, ... , ce codec n'en est pas moins soumis à une licence d'utilisation. Si, comme dans le cas de `flash`® ou `mp3`®, son utilisation en lecture est gratuite, la diffusion de contenus au format H.264 est soumise au paiement de droits de diffusion au consortium qui gère H.264. Consortium dont font bien sûr partie, Apple®, Microsoft®, ... Ainsi, une société éditrice de vidéos qui acquiert un appareil professionnel ou semi-professionnel, dans le commerce, pour réaliser des vidéos à usage commercial, et qui encode en H.264, devra s'acquitter d'une licence pour pouvoir vendre ces vidéos. Contrat signé à l'insu de l'acheteur, et dont les clauses peuvent bien sûr évoluer en fonction des besoins du consortium. Pour ces différentes raisons le format H.264 est à exclure. En 2013, un nouveau codec H.265 a été développé avec les mêmes restrictions que H.264.

Pourtant, une solution existe. Le projet VideoLAN, éditeur du célèbre lecteur vidéo VLC, a conçu des bibliothèques libres, d'encodage et de décodage de H.264. Ce sont les bibliothèques X.264. Il est donc possible d'encoder ou de décoder H.264 avec ces bibliothèques. Elles sont donc à privilégier, même si, au final, le résultat pour le lecteur est le même.

En 2010 Google® a lancé un nouveau conteneur libre, `webm`. Ce conteneur comprend une vidéo au codec libre VP8 et du son en OGG. En 2013, ce conteneur contiendra le codec vidéo libre VP9, qui, comme H265, permet d'économiser de la bande passante.

	<i>Modifiable</i>	<i>Diffusable</i>	<i>Archivable</i>	<i>Toléré</i>	<i>À exclure</i>
Conteneurs	.mkv, .ogg, .webm	.mkv, .ogg, .webm	.mkv, .webm	.avi, .mov .mp4	
Codecs	Theora, Xvid, X.264, VP8, VP9	Theora, Xvid, VP8, VP9	.dv		H.264, H.265

6 Pour conclure

Dans le milieu éducatif, la normalisation des formats d'échange de fichier est une nécessité éducative et civique : les formats ouverts sont les seuls à utiliser pour échanger des documents avec des élèves. C'est l'occasion d'aborder avec eux la problématique des données, de leur format et de leur conservation, de travailler autour de plusieurs items du B2i et c'est surtout, ne pas les obliger à utiliser des logiciels privés.

Le Journal Officiel n° 143 du 22 juin 2004 publie la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (référence NOR : ECOX0200175L).

On y trouve, Titre I^{er} (De la liberté de communication en ligne), Chapitre I^{er} (La communication au public en ligne), Article 4 :

« On entend par standard ouvert tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérable et dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en œuvre. »

Sources :

Version papier : la loi, pages 11168 à 11182 ; l'article 4, page 11169 (colonne de droite).

Version en ligne : <http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=ECOX0200175L>

Cela signifie donc :

- des protocoles et des formats de données **indépendants** d'un logiciel particulier, d'un système d'exploitation ou d'une société
- des spécifications techniques **documentées, publiées, non payantes, sans brevet** dessus, **sans « royalties »** dessus.