

D1.3 :Tenir compte des enjeux de l'interopérabilité

Le format des fichiers

Le **format d'un fichier** décrit la façon dont l'information est codée dans le fichier.

Il ne faut pas confondre le **format de fichier** et l'**extension** : le format de fichier décrit la façon dont l'information est codée ; l'extension est une partie du nom du fichier.

Quand vous voulez enregistrer un document au format Microsoft Word, le logiciel propose trois formats possibles (Microsoft Word 97/2000/XP, Microsoft Word 95 ou Microsoft Word 6,0). Ce sont trois formats de fichiers différents mais l'extension du fichier sera la même (doc).

L'extension d'un fichier permet au système de lui associer une application.

- Quand on tente d'ouvrir un fichier (en double-cliquant dessus par exemple), c'est l'application associée par défaut à cette extension qui est lancée et ouvre le fichier ;
- L'extension d'un fichier n'est pas obligatoire. Si celle-ci est omise, il faudra certainement indiquer avec quelle application le fichier doit s'ouvrir.

Les extensions de fichiers ne sont pas toujours visibles. Vous pouvez paramétrer votre gestionnaire de fichiers pour qu'elles le soient.

Pour changer le format d'un fichier :

- il ne faut surtout pas le renommer en changeant son extension car cela ne modifiera pas la façon dont l'information a été codée à l'intérieur du fichier ;
- il faut l'ouvrir avec l'application appropriée et l'enregistrer sous un autre nom en changeant le format dans la liste déroulante « type de fichier ».

On distingue deux catégories de fichiers :

- Les **fichiers textes** : les informations sont stockées sous forme de caractères lisibles par tout éditeur de texte.
- Les **fichiers binaires** : les informations sont stockées en binaire et ne peuvent être lues qu'avec le logiciel adéquat.

Attention, les fichiers textes ne concernent pas forcément les documents issus d'un traitement de texte.

Exemples de fichiers textes :

Un classeur au format ODF est constitué d'une archive de fichiers XML (Extensible Markup Language), qui sont des fichiers textes dont l'information est enrichie à l'aide de balises textuelles.

Un fichier au format CSV est un fichier texte qui contient les données d'un tableau séparées par des virgules (séparateur de colonnes) ou des marques de fin de paragraphe (séparateur de lignes). Ce format de fichier peut être exploité comme feuille de calcul dans un tableur.

Une page web au format HTML est un fichier texte utilisant des balises textuelles pour mettre en forme le texte, insérer une image ou définir un lien hypertexte.

Exemples de fichiers de données binaires :

Un document au format Microsoft Word 97/2000/XP.

Une photo au format d'image JPEG.

Un fichier exécutable d'application.

L'interopérabilité

Pour qu'un logiciel puisse exploiter les fichiers produits par un autre logiciel, il doit connaître les spécifications du format des fichiers. Si les spécifications sont publiées par un organisme de normalisation il s'agit d'une **norme** ; si elles sont publiées par un autre type d'organisme, il s'agit d'un **standard**.

Un format est dit **ouvert** si ses spécifications sont publiées et accessibles de tous. Dans le cas contraire, on parle de format **fermé** ou **propriétaire**.

Qu'est ce que l'interopérabilité ?

La **compatibilité** est la possibilité pour deux systèmes de types différents de fonctionner ensemble, notamment en utilisant les mêmes fichiers de données.

Lorsqu'un acteur devient dominant dans un domaine, les autres acteurs font en sorte d'être compatibles avec lui.

- Avantage : l'ensemble des systèmes peuvent à peu près fonctionner ensemble.
- Inconvénient : l'acteur dominant contrôle d'une certaine manière cette possibilité.

L'**interopérabilité** est la possibilité pour différents systèmes de fonctionner ensemble sans dépendre d'un acteur particulier. Elle repose sur la présence d'un standard ouvert.

L'**interopérabilité** est la capacité que possède un produit ou un système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre.

Explications et définitions adaptées du [groupe de travail Interop de l'AFUL](#) (consulté le 16 août 2011).

Exemple :

Le format de fichier Microsoft Word 97/2000/XP est un format fermé. Quand on tente d'enregistrer un document réalisé avec Writer d'Open Office dans ce format, le message suivant apparaît : « Il est possible qu'une partie du formatage et du contenu de ce document ne puisse pas être enregistrée dans le format de fichier Microsoft Word 97/2000/XP. Souhaitez-vous continuer ? »

En effet, OpenOffice ne peut garantir de traiter parfaitement tous les éléments du format Microsoft Word 97/2000/XP car son référentiel n'est pas public.

Les formats de documents

Un **document texte** est un fichier créé par un traitement de texte. C'est un fichier qui contient non seulement des caractères, mais aussi une structure, des mises en forme, des objets, etc.

Voici une liste non exhaustive de formats de fichiers pour les documents texte :

- le format **texte** (txt) est un format de fichiers basique : il est composé d'une suite de caractères simples (l'espace et le retour à la ligne étant considérés comme des caractères) ;
- le format **Microsoft Word** (doc) est un format propriétaire très courant ;
- le format **Rich Text Format** (rtf) est un format ouvert développé par Microsoft dans un souci d'interopérabilité ;
- le format **Texte de l'Open Document Format** (odt) est un format ouvert de traitement de texte ; il est utilisé par l'application Writer d'Open Office ;
- le format **Texte de l'Office Open XML** (docx) est un format de traitement de texte créé par Microsoft pour répondre à une demande d'interopérabilité ; il est utilisé par Microsoft Word depuis 2007.

Un **classeur** est un fichier manipulé par un tableur. Il regroupe plusieurs feuilles de calcul.

Voici une liste non exhaustive de formats de fichiers pour les classeurs :

- le format **Comma-Separated Values** (csv) est un format texte ouvert qui se présente sous forme de lignes de valeurs (sans formules) ; c'est un format d'échange souvent utilisé pour exporter des listes de nombres (liste de notes, etc.) ;
- le format **Microsoft Excel** (xls) est un format propriétaire ;
- le format **Classeur de l'Open Document Format** (ods) est un format de classeur ouvert ; il est utilisé par l'application Calc d'Open Office ;
- le format **Classeur de l'Office Open XML** (xlsx) est un format de classeur créé par Microsoft pour répondre à une demande d'interopérabilité. Il est utilisé par Microsoft Excel depuis 2007.

Pour une présentation en face à face, on accompagne souvent son exposé d'un **diaporama** (suite de diapositives) créé par un logiciel de **présentation**.

Voici une liste non exhaustive de formats de fichiers pour les présentations :

- le format **Microsoft PowerPoint** (ppt) est un format propriétaire ;
- le format **Présentation de l'Open Document Format** (odp) est un format ouvert de présentation ; il est utilisé par l'application Impress d'Open Office ;
- le format **Présentation de l'Office Open XML** (xlsx) est un format de présentation créé par Microsoft pour répondre à une demande d'interopérabilité ; il est utilisé par Microsoft PowerPoint depuis 2007.

Pour une présentation en ligne, on peut créer un site constitué d'une ou plusieurs **pages web**.

Voici un exemple de format de fichier pour les pages web :

- le format de document **HyperText Markup Language** (html) est un format ouvert pour la conception de pages web.

Les formats d'images

Pour représenter une image, on peut la décrire à l'aide de fonctions mathématiques (représentation vectorielle) ou par l'ensemble des points qui la composent (représentation matricielle).

Une **image vectorielle** est définie par un ensemble de données mathématiques : coordonnées, fonctions, attributs, etc.

Un image vectorielle peut être agrandie ou rétrécie sans dégradation car l'image sera recalculée précisément en fonction de la taille souhaitée. En général, le fichier correspondant est peu volumineux.

Quelques formats d'images vectorielles

- Le format **Scalable Vector Graphics** (svg) est un format ouvert d'image vectorielle ; il est surtout utilisé en cartographie et sur les téléphones portables.
- Le format **Dessin de l'Open Document Format** (odg) est un format ouvert de dessin vectoriel ; il est utilisé par l'application Draw d'Open Office.

Un dessin réalisé avec la barre d'outils dessin des logiciels de bureautique ou un diagramme créé dans un logiciel de présentation sont des images vectorielles.

Une **image matricielle** est définie par une grille de points ou pixels auxquels est associée une couleur.

Une image matricielle se dégrade si on l'agrandit : la pixellisation devient visible. En fonction de la taille de l'image et du nombre de couleurs utilisées, le fichier correspondant peut devenir volumineux. Pour transiter sur Internet, on utilisera des formats matriciels compressés.

Quelques formats d'images matricielles

- Le format **BitMaP** (bmp) est un format d'image matricielle développé par Microsoft et IBM. C'est un format simple, ouvert et facilement utilisable par les logiciels mais il est peu utilisé sur le web à cause de la taille volumineuse de ses fichiers, la couleur de chaque pixel étant codée sans effort de compression.
- Le format **Joint Photographic Experts Group** (jpeg ou jpg) est un format d'image ouvert compressé pour des images de type photographie (16 millions de couleurs). Le principe de compression JPEG est de coder la couleur de rectangles dont les pixels ont la même couleur (compression sans perte) ou des couleurs très proches (compression avec perte, taux de qualité paramétrable).
- Le format **Graphics Interchange Format** (gif) est un format d'image ouvert compressé pour des images de type dessin. Il permet de définir une couleur de fond transparente, de proposer un mode d'affichage entrelacé (aperçu de l'image qui se précise en cours de téléchargement) et de créer des images animées. Les couleurs d'une image sont définies sur une palette de 256 couleurs choisies parmi 16 millions de couleurs possibles : c'est le principe de base de la compression GIF, dite en couleurs indexées.
- Le format **Portable Network Graphics** (png) est un format d'image ouvert compressé. Il effectue une compression en couleurs indexées, mais la palette n'est pas limitée à 256 couleurs.

En pratique, on règle le taux de compression de façon à conserver une bonne qualité visuelle de l'image.

En général, la compression JPEG est préconisée pour les photographies, tandis que la compression GIF ou PNG se prête mieux aux dessins.

Les autres formats

Le format **Portable Document Format** (pdf) est un format ouvert développé par la société Adobe Systems qui préserve la mise en forme d'un document :

- il est généralement utilisé pour la diffusion de documents formatés qui n'ont pas vocation à être modifiés ;
- il s'applique à tous types de documents imprimables : document de traitement de texte, feuille de calcul, présentation, etc.

Pour lire un fichier au format PDF, vous devez utiliser un lecteur spécifique : Adobe Reader est un lecteur gratuit diffusé par la société Adobe Systems.

La numérisation du son et de la vidéo permet de stocker des musiques et des films sous forme de **fichiers**.

Étant donné la taille importante des fichiers audio et vidéo, il est fortement conseillé d'utiliser des formats audio-visuels compressés.

- Le format **Windows Media Audio** (wma) est un format propriétaire de compression audio développé par Microsoft.

Le **Moving Pictures Expert Group** a développé des méthodes de compression pour les fichiers audio et vidéo.

Parmi les plus connus, on distingue :

- le format MPEG-1/2 Audio Layer 3 ou **MP3** (mp3) est un format compressé ouvert pour les fichiers audio ; il s'agit d'une compression avec des pertes non perceptibles par l'oreille humaine ;
- le format **MPEG-4** (mpg, mp4, etc.) est un format compressé ouvert pour les fichiers vidéos.