

SCI6005 Information numérique et informatique documentaire (H2023)

Anton Boudreau Ninkov, EBSI, UdeM

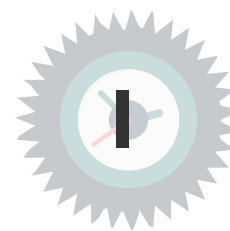
H2023 19 janvier 2023

*Cours 2 - Principes de base en gestion de l'information numérique
SCI6005*

Table des matières

I - Cours 2 - Principes de base en gestion de l'information numérique	3
1. + Au programme aujourd'hui	3
2. + Alignement pédagogique	3
3. Formats de fichiers	4
3.1. Définition de "formats de fichiers"	5
3.2. Utilité de connaître les formats de fichiers	6
3.3. Types de formats de fichiers	6
3.4. Jeux de caractères.....	12
3.5. Pistes pour l'identification du format d'un fichier.....	14
4. Nommage des fichiers et des dossiers.....	14
5. Organisation des fichiers.....	15
5.1. Organisation physique des fichiers	15
5.2. Organisation intellectuelle des fichiers.....	19
6. Conservation des fichiers	20
6.1. Conservation physique des fichiers	20
6.2. Conservation des fichiers du point de vue archivistique	21
7. Ressources en lien avec le cours	22
Abréviations	23
Bibliographie	24
Webographie	25
Index	26
Crédits des ressources	27

Cours 2 - Principes de base en gestion de l'information numérique



1. + Au programme aujourd'hui

- Principes de base en gestion de l'information numérique
 - Formats de fichiers
 - Nommage des fichiers
 - Organisation des fichiers
 - Conservation des fichiers

2. + Alignement pédagogique



Objectifs visés, matériel du cours et évaluation : Examen mi-session

Objectif général : Comprendre la place des technologies et de l'information numérique en contexte documentaire		
Objectifs spécifiques	Compétences à développer	Matériel associé
Décrire les principales technologies utilisées en milieu documentaires	Pouvoir expliquer ce qu'est un format de fichier (composantes syntaxique et sémantique)	Section <i>Formats de fichiers</i> > Définition de "formats de fichiers"
	Pouvoir donner des contextes où la connaissance des formats est utile	Section <i>Formats de fichiers</i> > Utilité de connaître les formats de fichiers
	Être en mesure de distinguer les caractéristiques des familles de formats (ouverts, fermés, etc.)	Section <i>Formats de fichiers</i> > Types de formats de fichiers
	Être en mesure de définir ce qu'est un jeu de caractères et à quoi il sert	Section <i>Formats de fichiers</i> > Jeux de caractères
	Pouvoir expliquer la différence entre les jeux de caractères ASCII, ASCII étendus, Unicode	Section <i>Formats de fichiers</i> > Jeux de caractères
	Pouvoir expliquer les principales règles de base pour nommer des fichiers et des dossiers	Section <i>Nommage des fichiers et des dossiers</i>

Pouvoir expliquer la différence entre un chemin absolu et un chemin relatif	Section <i>Organisation des fichiers</i> > <i>Organisation physique des fichiers</i>
Pouvoir construire des chemins absolus et relatifs à partir d'une arborescence donnée	Section <i>Organisation des fichiers</i> > <i>Organisation physique des fichiers</i>
Pouvoir décrire sommairement les considérations physiques et archivistiques qui prévalent pour la conservation des fichiers	Section <i>Conservation des fichiers</i>

Lien entre les objectifs, les compétences à développer et le matériel du cours 2



Objectifs visés, matière du cours et activités associées

Objectif général : Concevoir et implanter des systèmes et services d'information numérique		
Objectifs spécifiques	Matière	Activités
Exécuter des manipulations de base sur des documents, données et fichiers numériques	Nommage des fichiers et dossiers	TP Principes de base en gestion de l'information numérique Trousse sur les arborescences de fichiers et dossiers
	Organisation des fichiers	TP Principes de base en gestion de l'information numérique
	Conservation des fichiers	TP Principes de base en gestion de l'information numérique

Lien entre les objectifs, la matière du cours 2 et les activités associées

3. Formats de fichiers

Nous commencerons par examiner ce que l'on entend par un **format de fichier** et par décrire les **caractéristiques des principales familles** de formats de fichiers. Cette compréhension se révèle utile lorsque l'on gère de l'information numérique à plusieurs **égards** qui seront présentés. Cette section se terminera par la présentation d'une **méthode de résolution de problèmes** en lien avec les formats de fichiers.

3.1. Définition de "formats de fichiers"

L'OQLF^{p.23} définit un format ainsi :

Format



« [s]tructure définie de données contenues sur un support magnétique ou autre, établie selon des règles qui régissent le stockage, l'affichage, la manipulation, l'impression ou la transmission de ces données. »

(OQLF^{GDT p.25}, 2001, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8368613)

Ce que l'on doit retenir de cette définition est l'idée qu'un format est une **structure** de données qui répond à certaines **règles** propres à ce format en fonction des **actions** qu'il permet. PDF, par exemple, est un exemple d'un format de fichier, ainsi que JPEG et MP4. Ces trois formats représentent des types de fichier très différents, respectivement un fichier textuel, un fichier d'image et un fichier de son. Leur structure respecte ainsi des règles différentes qui permettent des opérations différentes.

Plus spécifiquement, ces règles doivent respecter une certaine **syntaxe** propre au format, c'est-à-dire une certaine **manière d'écrire les 0 et les 1**. C'est ce que l'on appelle la **composante syntaxique** d'un format. Pour pouvoir ouvrir un fichier sur votre ordinateur, il faut que les 0 et les 1 soient bien conformes à cette syntaxe. Il s'agit qu'un seul 0 soit changé pour que le fichier devienne illisible (corrompu).

En sus d'une certaine syntaxe, les règles d'un format décrivent une **manière d'en interpréter le contenu**, c'est-à-dire de faire du sens des 0 et des 1. On parle ici de la **composante sémantique** du format. Même si vous recevez par courriel un fichier dont la syntaxe est conforme au format, si vous n'avez aucun logiciel capable d'en interpréter le sens, vous ne serez pas capable de l'ouvrir.

Ainsi, pour que vous soyez en mesure d'**ouvrir un fichier** sur votre ordinateur, il faut qu'il en respecte bien la **syntaxe** ainsi que vous ayez ce qu'il faut pour en comprendre la **sémantique**. Ces deux composantes sont étroitement reliées.

Ce que n'est pas un format...



Un format n'est pas un logiciel

Il ne faut pas confondre le format d'un fichier avec le logiciel qui l'a créé. Il est ainsi plus ou moins juste de parler d'un fichier en format Word comme Word permet de sauvegarder un fichier en plusieurs formats (docx, rtf, txt, html).

Un format n'est pas une extension

Dans les systèmes d'exploitation qui affichent l'extension des fichiers (par exemple sous Windows), bien que cette extension soit normalement en lien avec le format, ce n'est pas l'extension qui crée le format. Si vous renommez un fichier en format HTML pour lui attribuer comme extension PDF, le fichier ne sera pas transformé en format PDF! L'*extension* est un moyen d'*associer un fichier à un logiciel* pour en accélérer l'ouverture. Changer l'extension d'un fichier en le renommant va simplement l'associer potentiellement à un autre logiciel qui sera fort probablement bien en peine de l'ouvrir comme il n'en connaîtra vraisemblablement pas la sémantique.

3.2. Utilité de connaître les formats de fichiers

On fait actuellement face à une problématique multiple : (1) la **prolifération** des formats, (2) l'**opacité** potentielle de l'information numérique si le format n'est pas reconnu, et (3) le **choix** du meilleur format pour une tâche donnée. Dans un contexte de gestion de l'information numérique il importe ainsi de s'attarder à connaître les différents formats de fichier pour **plusieurs raisons** :

- D'une part, il faut choisir des formats qui sont **compatibles** avec votre environnement informatique.
- D'autre part, cela vous aidera à **résoudre** certains **problèmes de lecture** de fichiers que vous pourriez rencontrer.
- De plus, vous serez ainsi en mesure de **mieux gérer** les **échanges** d'information et d'**optimiser** leur **diffusion** en choisissant les bons formats.
- Finalement, cela vous permettra de faire des **choix éclairés** selon les **besoins** que vous aurez en termes :
 - d'**accessibilité** pour utiliser des formats compris par une communauté de lecteurs la plus grande possible
 - de **pérennité** pour que vos fichiers demeurent accessibles dans le futur
 - de **réutilisation** pour pouvoir retravailler certains fichiers

3.3. Types de formats de fichiers

Formats ouverts versus formats fermés

On retrouve **différentes familles de formats**. Une première dichotomie à connaître est la distinction entre des **formats ouverts** et des **formats fermés** :

Formats ouverts	Formats fermés
<ul style="list-style-type: none"> • Spécifications (c'est-à-dire les règles définissant la syntaxe et la sémantique du format) publiques et accessibles • Favorisent l'interopérabilité entre les logiciels • Généralement produits ou préconisés par des organismes internationaux de normalisation • Par exemple, format ODT (traitement de texte de LibreOffice) 	<ul style="list-style-type: none"> • Spécifications secrètes et non publiées • Compatibilité limitée (avec un seul logiciel, ou bien avec ceux d'un même producteur donné ou, suite à une entente, de quelques producteurs) • Par exemple, format des fichiers produits par Adobe Photoshop (.psd)

Formats ouverts versus formats fermés

La **différence** entre ces deux familles de formats réside dans l'**accessibilité des spécifications** qui aura un impact sur l'interopérabilité de ces formats. Si vous cherchez à vous assurer qu'un plus grand nombre de personnes soit en mesure de consulter votre fichier, le format ouvert est en ce cas préférable.

Formats normalisés & formats propriétaires

On retrouve aussi deux autres familles de formats qui se distinguent cette fois-ci par leurs **producteurs** :

Formats normalisés	Formats propriétaires
<ul style="list-style-type: none"> • Formats définis et adoptés par une instance officielle de normalisation (deviennent des normes) • Offrent une plus grande garantie de pérennité et d'accessibilité qu'un format fermé • Par exemple <ul style="list-style-type: none"> ◦ Format PNG recommandé par le W3C (World Wide Web Consortium) ◦ Format HTML défini par le W3C 	<ul style="list-style-type: none"> • Formats dont les spécifications sont définies par un producteur commercial • Peuvent être ouverts si les spécifications sont disponibles publiquement (par ex. format PPTX) <ul style="list-style-type: none"> ◦ Un format propriétaire ouvert devenu très populaire et reconnu par un grand nombre de producteurs est appelé une norme de facto (par ex. format LaTeX) • Peuvent être fermés si les spécifications sont secrètes (par exemple format interne de DB/TextWorks (logiciel de bases de données))

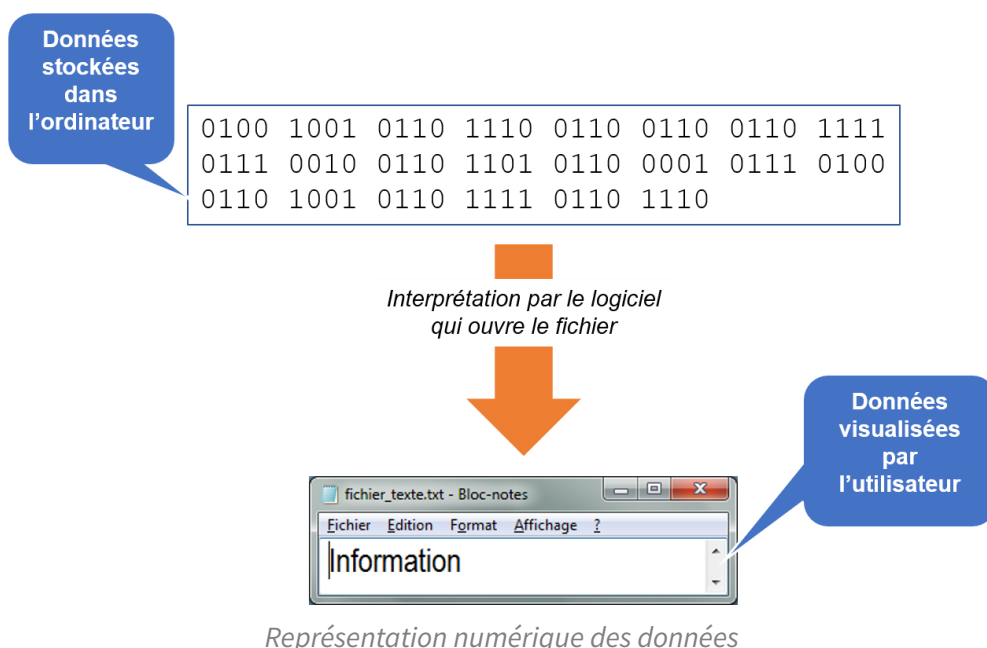
Formats normalisés & formats propriétaires

Si vous recherchez un format qui garantit une **meilleure conservation** dans le temps, les formats **normalisés** sont à prendre en considération. Le fait d'être endossés par une instance officielle de normalisation est en effet un garant sur le plan de la pérennité des formats.

Formats texte et formats binaires

Nous nous attarderons à une dernière dichotomie dans les familles de formats de fichiers, soit la distinction entre un **fichier en format texte** et un **fichier en format binaire**.

D'entrée de jeu, précisons que **tout fichier est enregistré sur le disque dur sous forme binaire**, c'est-à-dire dans des **suites de 0 et de 1**. Ces 0 et ces 1 sont agencés de manière à respecter la syntaxe du format et de manière à ce que leur interprétation par le logiciel utilisé pour l'ouvrir produise le résultat souhaité (une image, un son, un texte, etc.).



Heureusement pour nous, il y a donc quelque chose qui sert *d'interprète* afin de nous restituer les 0 et les 1 sous une forme qui nous est compréhensible. Pour certains types de fichiers, le **format est complexe** et demande un **logiciel spécialisé** pour en faire l'interprétation. C'est le cas, par exemple, d'une image. On parlera ici de **fichiers en format binaire**.

Dans d'autres cas, le fichier contient uniquement des **caractères alphanumériques** (c'est-à-dire des lettres, des chiffres et quelques symboles) sans aucune mise en forme, un contenu beaucoup plus simple que celui d'une image! L'interprétation de son format peut alors se faire par l'utilisation d'une simple table de correspondance entre des suites de 0 et de 1 et des caractères alphanumériques (ce que l'on appelle un **jeu de caractères**; cette notion sera abordée un peu plus loin). Ainsi, un éditeur de texte simple comme *Bloc-notes* (Windows) peut arriver à l'interpréter. C'est ce que l'on appelle un **fichier en format texte**.

Extrait du jeu de caractères ISO-8859-1

Caractère	Code hexadécimal	Code binaire
I	49	0100 1001
J	4A	0100 1010
[...]		
f	66	0110 0110
g	67	0110 0111
[...]		
n	6E	0110 1110
o	6F	0110 1111
p	70	0111 0000
[...]		
r	72	0111 0010
s	73	0111 0011

Jeu de caractères utilisé pour interpréter les données stockées dans l'ordinateur comme un format de fichier texte par Bloc-notes

Données stockées dans l'ordinateur

```

0100 1001 0110
1110 0110 0110
0110 1111 0111
0010 0110 1101
0110 0001 0111
0100 0110 1001
0110 1111 0110
1110
                    
```

Données visualisées par l'utilisateur

Jeu de caractères = table de correspondance entre un caractère et un groupe de bits

Par exemple, dans le jeu de caractères ISO-8859-1, le caractère « n » correspond à la série de bits **0110 1110**

Interprétation d'un fichier en format texte

La distinction entre un fichier en format binaire et un fichier en format texte repose ainsi sur la **complexité de leur sémantique** (complexe ou simple). Le tableau ci-dessous résume les différences entre ces deux types de formats :

Fichier en format texte	Fichier en format binaire
<ul style="list-style-type: none"> Fichiers dont le contenu est interprété comme une suite de caractères selon un jeu de caractères (sémantique = jeu de caractères) Aucun élément de formatage ou de présentation Peuvent être compris par un éditeur de texte tel Bloc-notes Exemple d'un fichier texte ouvert dans le Bloc-notes : https://cours.ebsi.umontreal.ca/sci6005/fichiers_demo/copie_txt.jpg 	<ul style="list-style-type: none"> Fichiers dont le contenu ne peut être interprété comme une suite de caractères selon un jeu de caractères en raison de leur sémantique plus complexe (par ex. mise en page, images) Besoin d'un logiciel compatible avec celui utilisé pour la création du fichier pour pouvoir afficher ou exécuter un fichier binaire Exemple d'un fichier binaire ouvert dans le Bloc-notes : https://cours.ebsi.umontreal.ca/sci6005/fichiers_demo/binaire_blocnotes.jpg Familles : fichiers d'images, fichiers sonores, fichiers vidéo, etc.

Fichier en format texte versus fichier en format binaire

Comme vous l'avez noté en ouvrant les deux exemples du tableau ci-dessus, une manière simple de savoir si un fichier est en format binaire ou en format texte est de l'ouvrir dans un éditeur de texte simple comme *Bloc-notes* : s'il est lisible, il s'agit d'un format texte, sinon, c'est un format binaire.

Il est à noter qu'il existe des formats de fichiers **basés sur un format texte**, c'est-à-dire interprétables à l'aide d'un jeu de caractères, mais avec des règles supplémentaires pour des opérations de mise en page ou de structuration. Il est ainsi possible de représenter des objets plus complexes. C'est le cas, par exemple, du format HTML que nous découvrirons un peu plus tard dans la session. Si vous l'ouvrez dans *Bloc-notes*, vous serez capable de le lire, mais il est possible que vous ne connaissiez pas certains codes utilisés comme, par exemple, `<h1>`. Il s'agit de balises HTML représentant certains types de contenu que le navigateur web saura interpréter pour lui associer une mise en page particulière. Dans l'exemple ci-dessous, la balise `<h1>` représente un titre de niveau 1 que le navigateur affichera dans une police de caractères plus grande et en gras.

```
1 <h1>Plan du cours SCI6005 - Information numérique et informatique documentaire
  (Hiver 2023)</h1>
2 <p><strong>Professeur : </strong>Anton Boudreau Ninkov</p>
```

Types de fichiers en format binaire

On retrouve une **multitude** de types de fichiers en format binaire. Nous en présenterons quelques-uns dans les tableaux ci-dessous à titre d'exemples.

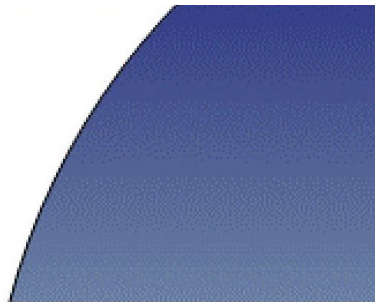
Formats	Caractéristiques
Graphic Interchange Format (extension <i>.gif</i>)	<ul style="list-style-type: none"> • Développé par CompuServe (années '80) • Maximum de 256 couleurs (8 bits) • Surtout pour graphiques et icônes • Gifs animés : série d'images gif affichées en séquence
Joint Photographic Expert Group (extensions <i>.jpeg, .jpg</i>)	<ul style="list-style-type: none"> • Norme de compression d'images selon le standard ISO-10918 • Palette de 16 millions de couleurs (24 pixels/bit) • Fichiers habituellement plus petits que les <i>.gif</i> donc transfert plus rapide et économie d'espace • Pour photographies, œuvres d'art et images texturées
Tagged File Format (extensions <i>.tif, .tiff</i>)	<ul style="list-style-type: none"> • Développé par Adobe • Format soutenu par la plupart des logiciels d'image • Pour archivage des images de qualité (très haute résolution) • Fichiers relativement volumineux
Portable Network Graphics (extension <i>.png</i>)	<ul style="list-style-type: none"> • Norme publique recommandée par le W3C • Spécialement conçu pour Internet en réponse à la « guerre des gifs » (https://www.gnu.org/philosophy/gif.fr.html) • 16 millions de couleurs • Pas supporté par les anciennes versions des navigateurs • Appelé à remplacer le <i>.gif</i> : taille plus réduite et palette plus grande

Principaux types de fichiers d'image (format binaire)

Le tableau ci-dessus illustre bien une des raisons principales pour retenir un format, soit l'**utilisation que l'on compte faire des fichiers**. On remarque en effet des **différences** notables pour les fichiers d'image pour ce qui est du **nombre de couleurs** possibles. De plus, certains formats comme JPEG utilisent un algorithme de **compression** pour obtenir des fichiers plus légers. Qui dit compression dit potentiellement **perte d'information**.

Dans l'exemple ci-dessous, trois images en formats différents ont été créées à partir de la même image de départ agrandie à 300%. Voici quelques observations par rapport à ces formats :

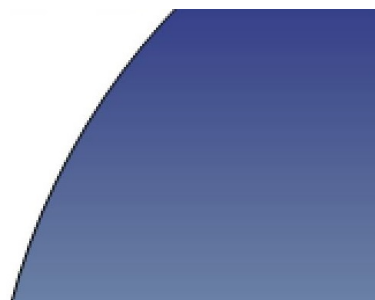
- On remarque en effet que le **format TIF** est **plus lourd** et que le **JPEG réduit la taille** du fichier grâce à la compression.
- Remarquez l'effet de **pixelisation** du fichier **GIF** du fait d'une **palette de couleurs moins grande**.
- Pour **JPEG**, on remarque une certaine **perte de précision** au pourtour du cercle.
- C'est le format **TIF** qui offre la **plus grande précision**.



Fichier GIF (79ko)



Fichier JPEG (21ko)



Fichier TIF (109ko)

C'est donc le type d'image et son utilisation qui influenceront vos choix. Dans un contexte archivistique de conservation à long terme d'images, la perte de précision n'est pas souhaitable rendant un format comme TIF plus intéressant. *Bibliothèque et Archives Canada*, par exemple, donne des lignes directrices pour aider au choix des formats : <https://www.bac-lac.gc.ca/fra/services/gestion-ressources-documentaires-gouvernement/lignes-directrices/Documents/formats-fichier-rdvc.pdf>.

Les tableaux qui suivent présentent des exemples d'autres types de fichiers en format binaire. Les caractéristiques présentées sont des pistes qui peuvent être utilisées pour faire un choix éclairé de formats.

Formats	Caractéristiques
Real Audio (extensions <i>.ra</i> , <i>.ram</i>)	<ul style="list-style-type: none"> Format propriétaire pour la diffusion en direct de séquences sonores et vidéo par Internet Permet la lecture en transit de fichiers audio (<i>audio streaming</i>)
Music Instrumental Digital Interface (extensions <i>.mid</i> , <i>.rmi</i>)	<ul style="list-style-type: none"> Fichier son, au format normalisé, contenant les instructions MIDI transmises à un synthétiseur qui le traduit en notes de musique Fichier de petite taille et de bonne qualité
MPEG1 Audio Layer 3 (extension <i>.mp3</i>)	<ul style="list-style-type: none"> Compression d'une séquence sonore dans un très petit fichier (1/12 de la taille originale) tout en conservant la qualité sonore initiale Télécharge et joue les fichiers ainsi que lecture en transit (<i>audio streaming</i>) On retrouve maintenant les fichiers .mp4
Wave Form Audio File Format (extension <i>.wav</i>)	<ul style="list-style-type: none"> Standard de son de Windows 3.11 créé par Microsoft Représentation numérique d'un signal analogique Fichiers volumineux (1 minute = 1 Meg)

Principaux types de fichiers de son (format binaire)

Formats	Caractéristiques
Real Audio (extensions <i>.ra</i> , <i>.ram</i>)	<ul style="list-style-type: none"> Format propriétaire pour la diffusion en direct de séquences sonores et vidéos par Internet Permet la lecture en transit de fichiers audio (<i>audio streaming</i>)
Quicktime (extensions <i>.qt</i> , <i>.mov</i>)	<ul style="list-style-type: none"> PC & Mac Norme ISO créée à l'origine par Apple Combine son, vidéo et animation Bonne qualité d'images Visualisation en différé
Audio Video Interleaved (extension <i>.avi</i>)	<ul style="list-style-type: none"> Standard de Microsoft Windows Utilisé pour le stockage des images vidéo et du son Données vidéo alternent avec celles audio, accélérant la vitesse de restitution
Moving Picture Experts Group (extensions <i>.mpg</i> , <i>.mpeg</i>)	<ul style="list-style-type: none"> Fichier vidéo comprimé, utilisé notamment pour la diffusion de films sur Internet Norme pour vidéo numérique et compression audio numérique développée par 70 compagnies, dont Sony, Philips et Apple Bonne qualité vidéo

Principaux types de fichiers vidéo (format binaire)

Autres types de fichiers en format binaire

- Fichiers de **traitement de texte** (extensions .doc, .docx, .wpd, .odt, ...)
- Fichiers **Portable Document Format** (extension .pdf)
 - Pour la diffusion de documents
 - Associé à Adobe Acrobat
 - Norme de facto qui conserve les éléments typographiques et la mise en page originale
- Fichiers **compressés** (extension .zip, .hqx, .qzip, ...)
 - Pour la **compression** et l'**archivage** de fichiers; à considérer aussi pour l'**échange** de fichiers
 - Utilisés par certains formats « **composites** » pour regrouper leurs composantes (par exemple, ePub, .docx)
 - Certains formats exploitent des formats existants pour représenter leurs composantes en les regroupant dans une archive compressée
 - Remplacer l'extension .epub par .zip permet de découvrir les différentes composantes d'un fichier ePub basé sur XML et XHTML
 - La compression **modifie** le format des fichiers
 - Traitement spécial applicable à tout format et complètement **réversible** (tout comme le cryptage ou le chiffrement)
- Etc.

3.4. Jeux de caractères

Dans la section précédente, nous avons vu que l'interprétation d'un fichier en format texte passe par l'utilisation d'un jeu de caractères. Bien que cela semble plutôt simple, l'**évolution des jeux de caractères** au fil du temps fait en sorte que certains **problèmes d'interprétation** peuvent survenir. Une connaissance de cette **évolution** permet de développer de **meilleurs réflexes de résolution de problèmes** lorsque surgit ce type de problèmes.

Décennies	Jeux de caractères	Caractéristiques
1960	ASCII (<i>American Standard Code for Information Interchange</i>) (ASCII pur)	<ul style="list-style-type: none"> • Créé en 1963 et normalisé en 1967 • Code sur 8 bits, mais en utilisant 7 donc possibilité de représenter 128 caractères seulement ($= 2^7$) • Initiative américaine, donc les caractères accentués non inclus
1980	ASCII étendus	<ul style="list-style-type: none"> • Développés au début des années 80 pour inclure les caractères accentués • ASCII pur « étendu » sur 8 bits donc 256 caractères, mieux, mais insuffisant pour toutes les langues • Plusieurs ASCII étendus existent ayant tous en commun les premiers 128 caractères de l'ASCII pur normalisé mais qui diffèrent pour les 128 suivants • Série ISO-8859 : 15 jeux normalisés dont le ISO-Latin-1 (ISO-8859-1)

Décennies	Jeux de caractères	Caractéristiques
		<ul style="list-style-type: none"> Autres exemples d'ASCII étendus, mais non normalisés : <ul style="list-style-type: none"> Jeu de caractères Windows : page de codes 1252 Jeu de caractères DOS : page de codes 437 Jeu de caractères sous Mac : Mac Os Roman
1990	Unicode	<ul style="list-style-type: none"> Standard informatique développé au début des années 1990 pour répondre aux problèmes nés de la multiplicité des ASCII étendus Code sur 16 bits donc potentiellement 65 536 caractères peuvent être représentés; suffisant pour contenir l'ensemble des caractères pour les principales langues du monde Plusieurs jeux différents : UTF-8, UTF-16BE, etc. UTF-8 : reconnu sur le Web Pour exemples, voir https://www.unicode.org/charts/

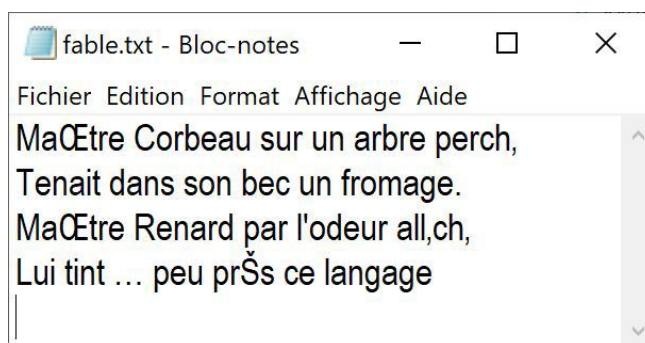
Bref historique des jeux de caractères

Ce qu'il faut retenir, d'une part, c'est l'**augmentation** au fil du temps du **nombre de caractères représentés** par les jeux de caractères afin de prendre en compte un nombre grandissant de langues et, d'autre part, le fait que pour les **ASCII étendus**, cela n'a pas fait l'objet d'un effort unique de normalisation ce qui a généré **plusieurs** ASCII étendus. Encore aujourd'hui, ces multiples ASCII étendus **coexistent**, et ce parfois sur un même ordinateur. Il est ainsi possible d'observer des **problèmes** dans la restitution des contenus si le mauvais jeu de caractères est utilisé.

C'est le cas de l'exemple ci-dessous. Le texte a été **écrit** à partir du **mode commande** de Windows et il est **affiché** dans le **mode graphique** de Windows sur le même poste. Remarquez que :

- Tous les caractères **non accentués** sont à leur **place**.
- Il n'y a que les **caractères accentués** qui sont **mal interprétés**.

Cela correspond bien à ce qui est décrit dans le tableau : les caractères non accentués demeurent les mêmes comme tous les ASCII étendus sont basés sur l'ASCII pur, mais les caractères accentués situés dans les 128 caractères suivants ne sont pas placés aux mêmes endroits.



Problème d'interprétation d'un fichier en format texte

Il est parfois possible d'observer ce problème sur le Web lorsque le navigateur n'utilise pas le même jeu de caractères que celui utilisé pour créer des pages HTML. Ce problème d'interprétation est en fait possible lorsque les personnes qui ont conçu les pages HTML n'ont pas pris soin d'expliquer en clair le jeu de caractères utilisé (nous verrons comment le faire plus tard dans la session).

Être en mesure de reconnaître ce type de problèmes nous permet d'y apporter une **solution**. La solution consiste à utiliser un **logiciel** qui offre la possibilité de **changer le jeu de caractères** utilisé. Pour le problème illustré ci-dessus, une des manières de le corriger est, dans Bloc-notes, d'utiliser la police *Terminal* (menu *Format > Police...*) qui est basée sur le jeu de caractères du mode commande de Windows. Si vous croisez ce problème avec une page web, certains navigateurs permettent de changer de jeu de caractères. On retrouve dans Firefox, par exemple, une option *Encodage du texte* dans le menu *Affichage*.

3.5. Pistes pour l'identification du format d'un fichier



Il est possible que vous receviez un fichier dont le format vous est inconnu. Voici quelques **pistes** pour vous aider à l'identifier :

- L'**extension** du fichier est une bonne indication, même si elle n'est pas absolue.
 - On retrouve des ressources sur Internet pour reconnaître les formats de fichiers (voir, par exemple, les signets sur le site de l'EBSI [<https://ebsi.umontreal.ca/ressources-services/ressources-pedagogiques/sites-exploration-internet/#util3>]).
- Si l'extension est correcte, mais que vous n'avez aucun logiciel compatible, le fichier sera illisible; il faut soit vous procurer un **logiciel compatible** ou un **outil de conversion**.
- Si l'extension a été modifiée/perdue, vous pouvez l'**ouvrir** avec différents logiciels de lecture en fonction de la **nature** de l'information attendue. Un autre test simple consiste à ouvrir le fichier avec un **éditeur de texte simple** comme *Bloc-notes*
 - S'il s'agit d'un fichier de format texte ou basé sur un format texte, il sera fort probablement reconnaissable.
 - S'il s'agit d'un fichier binaire (donc illisible dans *Bloc-notes*), plusieurs incluent un en-tête reconnaissable en début de fichier (par ex. PDF, JPEG, MP3, etc.).

Le **seul moyen infallible** pour connaître le format d'un fichier demeure que le **créateur** indique le format dans la documentation d'accompagnement du fichier qu'il vous envoie.

4. Nommage des fichiers et des dossiers

Après avoir abordé la question des formats de fichiers, nous nous attarderons maintenant à une **métadonnée** très importante d'un fichier, soit son **nom**. On ne soupçonne pas toujours l'**importance de bien nommer** un fichier ou un dossier. Cette importance est bien réelle comme un fichier ou un dossier mal nommé peut causer des problèmes techniques, voire même de repérage.

Trois principaux aspects vont influencer le choix du nom d'un fichier ou d'un dossier :

1. Des **contraintes techniques** imposées par le système d'exploitation
2. Des **bonnes pratiques** pour faciliter l'échange de fichiers ou de dossiers
3. Des **considérations organisationnelles** pour faciliter la gestion des fichiers et des dossiers

Contraintes techniques imposées par le système d'exploitation

Tous les systèmes d'exploitation **excluent** certains **caractères spéciaux** des noms de fichiers comme, par exemple, la barre oblique ou le point d'interrogation sous Windows 11. Si vous essayez d'inclure ces caractères dans le nom, le système vous retournera un message d'erreur. Même si on travaille dans un système d'exploitation en particulier, il ne faut pas oublier que nous sommes dans un environnement hyperconnecté. Il est donc prudent d'éviter les caractères spéciaux qui pourraient être problématiques pour d'autres systèmes d'exploitation. Pour **éviter tout problème potentiel**, il peut être sage de se contenter de **lettres**, de **chiffres** et du **tiret** ou **barre de soulignement** dans les noms de nos fichiers et dossiers.

Tous les systèmes d'exploitation imposent aussi une limite quant à la **longueur**, non pas du nom du fichier uniquement, mais de son **chemin absolu** (c'est-à-dire la mise bout-à-bout de tous les dossiers et sous-dossiers, à partir de l'unité de stockage, pour se rendre au fichier, par ex. C:\documents\cours\sci6005\cours1.pdf; nous reviendrons sur cette notion un peu plus tard). Sous Windows 11, le chemin absolu complet d'un fichier ne doit pas dépasser 260 caractères. Ainsi, bien que nous soyons débarrassés de la limite de 8 caractères qui prévalait à une certaine époque pour les noms des fichiers, il n'en demeure pas moins une **bonne pratique d'éviter des noms trop longs**. Pour raccourcir un nom de fichier, on peut par exemple **utiliser des abréviations** (pourvu qu'elles soient bien connues) ainsi qu'**enlever les mots vides de sens** (par exemple, les articles comme "le" et "la").

Bonnes pratiques pour faciliter les échanges

En sus des contraintes propres au système d'exploitation de son ordinateur, il faut aussi prendre en considération que nos fichiers pourraient être **partagés** par exemple avec nos collègues qui, peut-être, n'utilisent pas le même système d'exploitation. Nos fichiers peuvent aussi être déposés sur différentes plateformes de stockage en ligne qui sont sur des systèmes d'exploitation potentiellement différents. Ainsi, afin d'éviter des problèmes lors d'échange ou de transfert de fichiers, il est recommandé :

- D'**éviter** de mettre des **caractères accentués** dans les noms de fichiers ou de dossiers
- De **remplacer les espaces** dans les noms de fichiers ou de dossiers par un **tiret (-)** ou une **barre de soulignement (_)**

Certains vont même jusqu'à conseiller d'éviter d'utiliser des **lettres majuscules**, certains systèmes d'exploitation pouvant être sensibles à la casse (majuscule/minuscule) des lettres.

Considérations organisationnelles pour la bonne gestion des fichiers et dossiers

Finalement, certaines pratiques de nommage de fichiers peuvent être établies au sein des organisations afin d'**uniformiser** la manière de les nommer et en **faciliter** la gestion. Cela est particulièrement utile pour les espaces partagés de dépôt de fichiers. Par exemple, une bonne pratique est d'identifier dans le nom d'un fichier la version par une date (en privilégiant une seule forme pour les dates). Cela permettra facilement d'identifier la version la plus à jour d'un fichier.

Convention de nommage

Il n'est pas rare, et cela est même souhaitable, de consigner dans une **convention de nommage** l'ensemble des règles de nommage que l'on veut implanter dans notre organisation. Sur certains points, les conventions de nommage d'une organisation à l'autre pourront différer, mais certains principes de base s'y retrouveront assez systématiquement. On peut identifier facilement sur Internet des conventions de nommage. Vous retrouverez un exemple de règles de nommage pour la Ville de Brossard dans l'article de C. Couture (2009-2010)^{COUTURE 2009-2010 p.24} à l'annexe 1 (section D - Identification des fichiers et des répertoires numériques, p. 86-87). Vous en utiliserez une dans le cadre du TP sur les principes de base en gestion de l'information numérique.

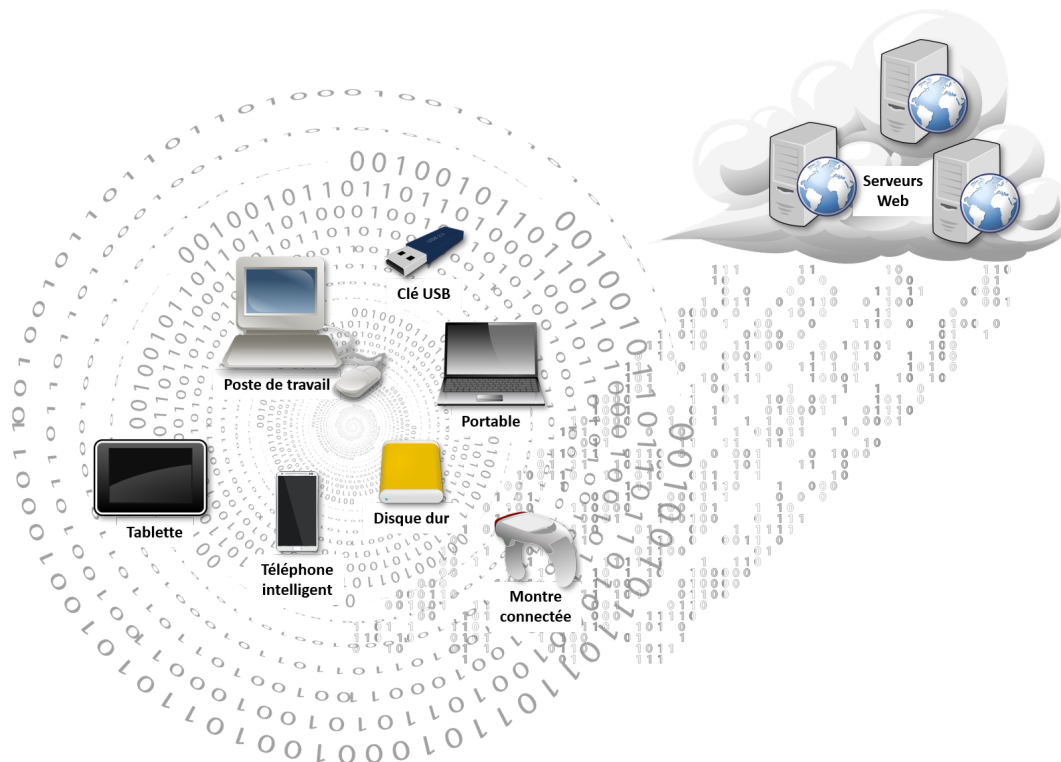
5. Organisation des fichiers

Nous aborderons l'organisation des fichiers sous deux perspectives : (1) leur **organisation physique** sur un support de stockage, et (2) leur **organisation intellectuelle** (c'est-à-dire la manière de définir une arborescence de dossiers).

5.1. Organisation physique des fichiers

Pensez à votre journée d'hier et aux données numériques que vous avez créées, modifiées, consultées et essayez d'en faire la cartographie... Courriel (sur votre poste de travail, votre portable, votre téléphone intelligent, votre tablette et/ou sur un serveur de courriel)? Navigation sur Internet? Édition d'un document (en ligne ou sur votre ordinateur)? Vous ferez probablement le constat, (1) que **vous avez utilisé de l'information numérique** dans votre journée, (2) que cette information numérique

venait de **sources** et d'**environnements multiples**. Et si vous essayiez maintenant de penser à où, **physiquement** ces informations résidaient, à comment elles sont organisées, il est possible que vous ne soyez pas nécessairement capable de répondre à cette question. Du fait de la multiplicité des espaces de stockage que l'on utilise, le portrait peut rapidement devenir assez complexe!



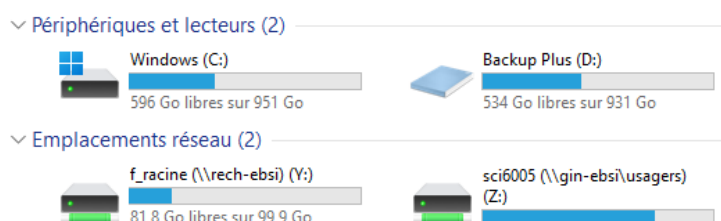
L'humain n'a jamais été aussi **branché**, n'a jamais **créé autant d'information**, n'a jamais été aussi **dispersé sur de multiples plateformes et unités de stockage**, et n'a probablement jamais été aussi **peu conscient de ce que cela implique** concrètement : Où sont nos données? Comment sont-elles organisées? *Et ce n'est pas uniquement la faute de l'humain.*

L'**évolution** des technologies fait que la **complexité technologique** sous-jacente est de plus en plus **cachée** aux utilisateurs – créant parfois un certain effet de "boîte noire". Par exemple, la structure de fichiers et de dossiers sur une tablette est bien souvent cachée; les fichiers n'existent que par le biais des applications/logiciels. Pourtant, cette information numérique est bien structurée sur le support de stockage. Pour un utilisateur "grand public", il n'est peut-être pas nécessaire de comprendre comment l'information est stockée. *Par contre, la professionnelle ou le professionnel de l'information doit le savoir sinon comment gérer cette information numérique?*

Nous avons dans la section précédente abordé la notion de format pour observer les fichiers d'un point de vue très micro. Il est maintenant temps de le faire d'une **perspective plus macro** pour comprendre comment les fichiers sont **stockés** et **organisés physiquement** sur un support de stockage.

Arborescence de fichiers et de dossiers

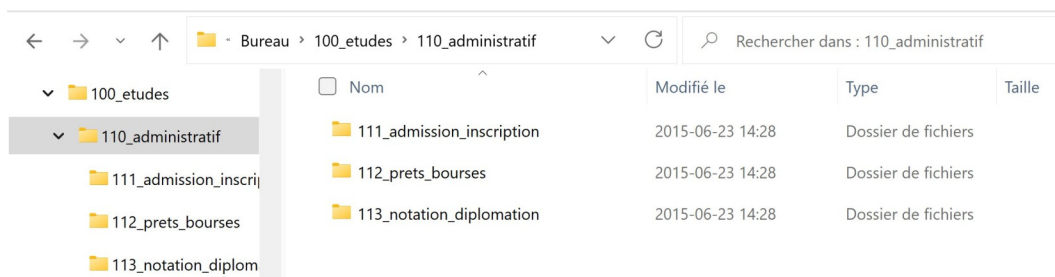
Sur un poste de travail, on peut retrouver de multiples **unités de stockage**. Selon les systèmes d'exploitation, la manière de les représenter et de les nommer peut varier, mais les principes de base demeurent. Les exemples donnés ci-dessous proviennent de Windows, système d'exploitation très présent dans les organisations.



Exemple d'unités de stockage sur un poste sous Windows 11

En cliquant sur *Ce PC*, on pourra voir la liste des unités de stockage « visibles » sur le poste de travail. Sous Windows, chacune est représentée par une lettre. On y retrouvera minimalement le **disque dur** du poste (ici C:) qui pourrait être partitionné (c'est-à-dire découpé en plusieurs unités). On pourra aussi y voir des **périphériques de stockage amovibles** comme un lecteur de disquette, une clé USB (ici D:). Finalement, on pourrait y retrouver des **emplacements réseau**, c'est-à-dire des unités de stockage accessibles via un réseau (ici Y: et Z:).

Si on ouvre une unité de stockage, on y retrouvera une arborescence, c'est-à-dire une **hiérarchie de dossiers** (nommés aussi *répertoires*), qui sont des *contenants*, et de **fichiers**, qui sont des *contenus*. Les **explorateurs de fichiers graphiques** des systèmes d'exploitation, comme l'*Explorateur de fichiers* (Windows), représentent visuellement cette arborescence sous forme de hiérarchie navigable par l'utilisateur. Il est possible d'entrer dans un dossier, pour voir ce qu'il contient, soit d'autres dossiers et/ou des fichiers.



Exemple d'une hiérarchie de dossiers dans l'Explorateur de fichiers (Windows 11)

Il est possible aussi de naviguer dans l'arborescence en utilisant le **mode commande** du système d'exploitation qui est un mode de représentation uniquement **textuel**.

```
C:\Users\Christine Dufour\Documents\cours\sci6005\ebis_plan_classe1>tree
Structure du dossier pour le volume Windows
Le numéro de série du volume est D061-9A7F
C:.
├── 100_etudes
│   ├── 110_administratif
│   │   ├── 111_admission_inscription
│   │   ├── 112_prets_bourses
│   │   └── 113_notation_diplomation
│   ├── 120_organisation
│   ├── 130_academique
│   ├── 140_parascolaire
│   │   ├── 141_activites_etudiantes
│   │   ├── 142_benevolat
│   │   └── 143_autres
```

Exemple d'une hiérarchie de dossiers en mode commande sous Windows

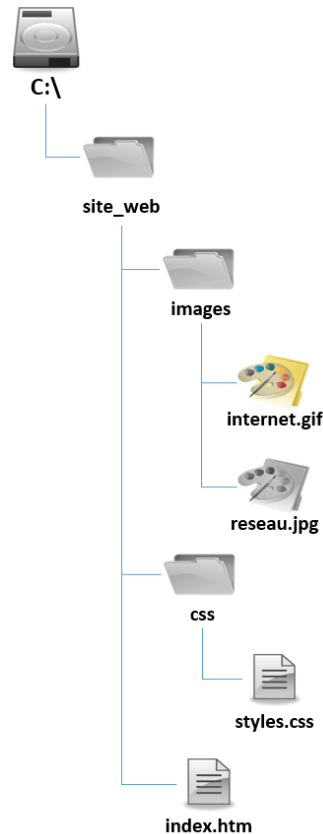
Fut un temps où, pour certains systèmes d'exploitation, seule l'interface textuelle existait. L'interface graphique présentée initialement par Apple a rapidement été adoptée par l'ensemble des systèmes d'exploitation, cette dernière étant plus intuitive pour les « non informaticiens ». L'interface textuelle est maintenant peu connue du grand public, mais elle existe toujours – même Mac OS a son mode commande avec l'application *Terminal*. Il y a même des tâches que le mode graphique ne peut facilement exécuter rendant fort utile le mode commande (vous en découvrirez une dans le travail pratique sur les principes de base en gestion de l'information numérique).

Chemin absolu ou relatif d'un fichier ou d'un dossier

Nous allons prendre quelques instants pour examiner l'idée de **référence à un dossier ou à un fichier**. En effet, en sus de savoir où se trouve un fichier dans une arborescence de dossiers, il faut connaître la manière de *représenter cette position*. Cette référence s'appelle un **chemin**, chemin qui nous permet d'atteindre un fichier ou un dossier. On retrouve deux types de chemin : **chemin absolu** et **chemin relatif**.

Le **chemin absolu** est le **chemin complet** donnant la route à suivre à partir de l'*unité de stockage* jusqu'au *fichier ou au dossier que l'on veut atteindre*. Le chemin fonctionnera peu importe notre point de départ. C'est un peu comme une adresse physique que l'on donne à quelqu'un. En donnant le numéro civique, le nom de la rue, de la ville, de la province et du pays, peu importe où on se retrouve sur le globe, on pourra vous retrouver.

Par exemple, le chemin absolu du fichier `internet.gif` dans l'arborescence ci-dessous est `C:\site_web\images\internet.gif`. Remarquez la *syntaxe* : on débute par l'**unité de stockage** pour se rendre jusqu'au **fichier** en nommant, dans l'ordre, **tous les dossiers parcourus** pour s'y rendre. Sous Windows, on sépare les différents niveaux par le symbole de barre oblique inversée (`\`). Sous Mac OS et Linux, c'est le symbole de la barre oblique (`/`) qui est plutôt utilisé.



Exemple d'une arborescence de dossiers et fichiers

Revenons à l'exemple d'une adresse physique. Si au lieu de l'adresse complète vous dites simplement à votre interlocuteur, par rapport à l'endroit où il se trouve (sa localisation), de marcher deux blocs d'immeubles, de tourner à droite, et de marcher jusqu'à la maison jaune, ce chemin, qui est **relatif** à sa **localisation**, ne sera utile qu'à lui, mais il sera beaucoup **plus efficace** que de lui donner l'adresse complète. On retrouve cette notion de **chemin relatif** aussi au niveau d'une arborescence de fichiers, la localisation de départ étant le dossier d'où l'on part.

Par exemple, dans l'arborescence ci-dessus, si je me situe dans le dossier `site_web` et que je veux me rendre jusqu'au fichier `internet.gif`, le chemin relatif sera simplement de rentrer dans le dossier `images` et ensuite d'y retrouver le fichier : `images\internet.gif`. Si je me situe dans le dossier `css`, il faudrait que je commence par remonter d'un niveau – ce qui est représenté par deux points consécutifs `..` – pour ensuite entrer dans le dossier `images` et y retrouver le fichier `internet.gif` : `..\images\internet.gif`. Ici aussi, chacun des niveaux est séparé soit par la barre oblique inversée sous Windows, ou par la barre oblique sous Mac OS et Linux.

Vous aurez l'occasion dans le cadre du *TP sur les pratiques de base en gestion de l'information numérique* de vous pratiquer à composer des chemins absolus et relatifs. Prenez le temps de bien comprendre ces notions qui sont **importantes** afin de bien intégrer comment concrètement les fichiers et dossiers sont stockés sur un disque dur et comment il est possible de les référencer. Nous verrons un peu plus tard dans la session que ces mêmes notions se retrouvent lorsque l'on développe un site web et que l'on veut y insérer une image par exemple.

5.2. Organisation intellectuelle des fichiers

Nous nous sommes attardés dans un premier temps aux considérations techniques de l'organisation des fichiers. Vous vous demandez fort probablement comment procéder pour décider de l'**organisation intellectuelle des fichiers**, c'est-à-dire de la **définition de la hiérarchie** des dossiers sur un support de stockage. Comment décide-t-on du nom des dossiers et de leur nombre? Sur votre propre ordinateur, vous avez défini une organisation qui correspond à une logique qui vous est propre et qui vous permet, souhaitons-le!, de facilement organiser et retrouver vos fichiers. Si nous travaillons uniquement en vase clos, sur nos ordinateurs personnels, cette méthode est efficace. Dans un contexte entre autres organisationnel où l'on travaille dans des espaces partagés de stockage de fichiers et où il se peut qu'une autre personne que vous ait à accéder à votre ordinateur (par exemple lors d'une absence), il est souhaitable d'arriver à concevoir une hiérarchie de dossiers qui soit efficace pour tout le monde.

Nous ne nous attarderons que brièvement sur la notion d'organisation intellectuelle des fichiers, cette notion relevant plus d'un cours sur la classification, en particulier en contexte archivistique. Il demeure tout de même important d'y être sensibilisé.e. Nous exploiterons pour ce faire le compte-rendu d'expérience de C. Couture (2009-2010)^{COUTURE 2009-2010 p.24} pour l'élaboration d'un cadre de classification pour la Ville de Brossard.

Pour Couture (2009-2010, p. 73) "[l]a première étape, à notre avis, de tout projet de gestion de documents numériques [doit] en être une d'organisation de l'information". L'auteure fait ainsi ressortir **toute l'importance**, dans un **contexte organisationnel**, que revêt la **définition d'un plan (cadre) de classification commun** en raison, entre autres, des mouvements de personnel et des manières différentes que les personnes peuvent avoir d'organiser les documents. Les informations sont éparpillées, parfois sur des ordinateurs personnels, d'autres fois dans des espaces partagés. Cet éparpillement et ce foisonnement dans les manières d'organiser les fichiers causent des problèmes de repérage qui peuvent être minimisés par l'adoption d'un cadre de classification commun (p. 73).

L'auteure identifie **trois principales contraintes** à l'élaboration d'un plan de classification pour les documents numériques qui rejoignent certaines des préoccupations techniques abordées précédemment (p. 74-75) :

1. **Nombre de niveaux de la structure de classification** : le chemin absolu d'un fichier ne pouvant dépasser 260 caractères, il est conseillé de limiter au minimum le nombre de niveaux. L'auteure suggère de ne pas dépasser quatre niveaux. De plus, l'auteure souligne le côté expérience-utilisateur : "pour qu'il [cadre de classification] soit bien accepté pour les documents numériques, le cadre de classification doit imposer le moins de « clics » possible, donc le moins de niveaux possible" (p. 74), un autre argument en faveur d'une arborescence qui ne comporte qu'un *nombre restreint de niveaux*.
2. **Taille des noms de fichiers** : Toujours en raison du nombre de caractères maximal du chemin absolu d'un fichier, il est important que *chacun des termes qui composent le nom d'un fichier soit court et significatif*. Nous ajouterons à ce conseil fort judicieux un argument complémentaire : la signification d'un fichier repose sur son nom, mais aussi sur le nom des dossiers qui le contiennent. La hiérarchie apporte un contexte qu'il n'est pas nécessaire de dédoubler dans le nom du fichier comme tel. Par exemple, si vous conservez les procès-verbaux de votre unité dans un dossier nommé "proces_verbaux", il est inutile d'inclure le terme "proces_verbaux" dans les fichiers eux-mêmes à moins que ces fichiers se retrouvent parfois "hors contexte".
3. **Règles de nommage des fichiers** : Finalement, Couture souligne la pertinence d'imposer des **règles** quant à la manière de **nommer** les fichiers et les dossiers, pour normaliser la présentation des informations, ce qui en facilite la lecture, mais aussi pour éviter des problèmes techniques. Les règles adoptées pour la Ville de Brossard sont présentées en annexe à titre d'exemples.

Pour l'élaboration du cadre de classification, les applications antérieures utilisées à la Ville de Brossard ont permis de recueillir des informations quant aux activités de la ville - le cadre se basant sur ces activités - et aux types de documents produits. Il est en effet important de **bien connaître le milieu** pour lequel un cadre de classification est développé pour qu'il y soit bien adapté. De plus, plusieurs

références ont été exploitées pour entre autres identifier des **bonnes pratiques**, dont des plans de classification d'autres villes, la norme *ISO-15489 - Records management*, et une thèse de doctorat sur la classification (Mas, 2007^{MAS 2007 p.24}).

Pour préciser le vocabulaire utilisé pour nommer les dossiers, les documents de la ville ont été consultés ainsi que les employés. La consultation des employés remet à nouveau de l'avant l'importance de **bien adapter** un cadre de classification au milieu pour lequel il est développé. Cela en facilitera l'adoption en sus de le rendre plus efficace.

La structure du cadre défini comporte, d'une part, une structure corporative qui peut être utilisée par toutes les unités de la Ville de Brossard, et, d'autre part, des classes personnalisées pour les unités. On y retrouve donc à la fois une **structure commune** dans un but d'**uniformisation** ainsi que de la **flexibilité** pour prendre en compte les **particularités** des unités.

6. Conservation des fichiers

Comme pour l'organisation des fichiers, on peut aborder leur conservation sous deux angles : (1) la **dimension technique** de la conservation (conservation physique), et (2) la **dimension archivistique** de la conservation. En effet, pour toute information numérique que vous créez, il faut réfléchir aux conditions de sa conservation physique sur un support, ainsi qu'il faut se demander combien de temps il est nécessaire de la conserver. Cette réflexion sera d'autant plus critique si l'information est très importante. Il s'agit de **gestion de risques** : à quel point peut-on prendre la chance de perdre cette information? Par exemple :

- *Le fichier Word d'un travail de session important en cours de rédaction* : à moins d'être prêt.e à refaire tout le travail, il est crucial de vous assurer d'en avoir une copie de sauvegarde idéalement sur un espace de stockage qui ne se retrouve pas dans la même pièce que votre ordinateur. Un serveur comme *OneDrive Entreprise* représente un environnement de sauvegarde sécuritaire; il faut toutefois réaliser que vous ne pourrez y conserver vos fichiers éternellement. Il faut aussi vous demander combien de temps il est nécessaire et utile de le conserver!
- *La facture pour l'achat d'un ordinateur* : si vous n'avez qu'une version numérique de la facture, il est très important de vous assurer de bien la conserver si on veut, par exemple, avoir recours à la garantie. Il est ainsi risqué de la conserver uniquement dans le dossier "téléchargement" de votre ordinateur ou de la conserver sur un serveur à distance plus ou moins fiable. S'il s'agissait de la facture pour l'achat d'une plante verte, vous assumeriez plus facilement ce risque!
- *Le fichier d'installation d'un logiciel* : si vous pouvez sans difficulté récupérer ce fichier sur le site du producteur du logiciel, il n'est pas nécessaire d'en garder une copie.

6.1. Conservation physique des fichiers

La durée de vie d'un fichier numérique repose sur plusieurs facteurs, les plus importants étant les suivants :

- D'une part, il y a la **pérennité du format** du fichier comme tel.
- D'autre part, il y a la **pérennité du support** sur lequel il réside.

Pérennité du format de fichiers

Certains formats peuvent avoir la vie plus longue que d'autres, c'est-à-dire qu'il y aura dans les années à venir des logiciels en mesure d'en comprendre la sémantique. Comme nous l'avons vu précédemment, les **formats normalisés**, du fait d'être chapeautés par des instances de normalisation, ainsi que les **formats ouverts**, comme leurs spécifications sont publiques et accessibles, offrent une **meilleure garantie de pérennité**. Un format fermé et propriétaire est en effet dépendant d'un producteur commercial qui pourrait disparaître ou décider de ne plus le maintenir.

Pérennité du support de stockage

En sus de la pérennité du format, il faut aussi prendre en considération la pérennité du support sur lequel les fichiers sont enregistrés.

Cette **pérennité** peut se comprendre sur le **plan physique des supports de stockage** : Combien de temps un disque dur demeure-t-il lisible et fonctionnel? Combien de temps une clé USB peut conserver des fichiers? Qu'en est-il des supports de stockage comme les disquettes, les CD, les DVD?

Leur durée de vie est tributaire des **matériaux** qui les composent, mais aussi de leurs **conditions d'utilisation et de conservation**. Des **stratégies de migration** - c'est-à-dire du transfert de données d'un support à un autre - doivent être envisagées pour les contenus qui doivent être conservés sur de longues durées afin de s'assurer de ne pas les perdre en raison d'un support de stockage qui a dépassé sa durée de vie. Nous n'entrerons pas dans les détails techniques de la conservation des supports numériques, comme il s'agit d'un domaine d'expertise très vaste qui dépasse les objectifs de ce cours. Vous retrouverez sur le Web des ressources intéressantes en la matière, comme par exemple sur le site de l'Institut canadien de conservation du Gouvernement du Canada : <https://www.canada.ca/fr/institut-conservation/services/soin-objets/supports-electroniques.html>. Pour celles et ceux qui s'intéressent à ces questions, le cours SCI6358 Préservation audiovisuelle et numérique¹ permet d'aller plus loin sur ces notions.

Il faut aussi penser à l'**accès** à l'information contenue sur ces supports. Même si un support a une durée de vie de 200 ans, il devient inutilisable si les **lecteurs** pour accéder à son contenu n'existent plus ou si l'**accès à la plateforme en infonuagique** sur laquelle l'information est conservée n'est plus possible (par exemple, si vous n'avez plus de compte actif sur leurs serveurs).

6.2. Conservation des fichiers du point de vue archivistique

Nous n'entrerons pas en détail dans les considérations archivistiques de la conservation des fichiers, qui relèvent plutôt d'un cours d'archivistique. Il importe toutefois d'être conscient.e que tous les documents (et donc les fichiers numériques qui les composent) **n'ont pas les mêmes besoins en termes de conservation en fonction de leur nature et du contexte**. Par exemple, les organismes publics ont l'obligation de mettre en place un calendrier de conservation de leurs documents qui doit être approuvé par Bibliothèque et archives nationales du Québec (BANQ). Il s'agit d'une **obligation légale** liée à la Loi sur les archives :

« Tout organisme public doit établir et tenir à jour un calendrier de conservation qui détermine les périodes d'utilisation et les supports de conservation de ses documents actifs et semi-actifs et qui indique quels documents inactifs sont conservés de manière permanente et lesquels sont éliminés. »

(Loi sur les archives, Article 7, <https://legisquebec.gouv.qc.ca/fr/ShowDoc/cs/A-21.1>)

Ce **calendrier de conservation** précisera pour **différents types de documents** ce qu'il faut **légalement** en faire dans des règles de conservation. Certains courriels, en fonction de la nature de l'information qu'on y retrouve, doivent ainsi être conservés pendant un certain temps, selon la règle de conservation qui leur est assignée. Il faut donc être très critique, à cet égard, lorsque l'on utilise un nouvel environnement numérique, d'une part sur les règles de conservation qui régissent les documents qu'on y crée et, d'autre part, sur le choix des formats et des supports afin de s'assurer de pouvoir respecter les délais de conservation impartis. Vous retrouverez en annexe du *TP sur les principes de base en gestion de l'information numérique* des exemples de règles de conservation pour certains documents qui peuvent se retrouver dans le plan de classification des activités académiques et professionnelles.

¹ <https://cours.ebsi.umontreal.ca/planscours/sci6358>

7. Ressources en lien avec le cours

Matériel de cours

- *Notes de cours [cf. sci6005_h2023_c2_notes_cours.pdf]*

Lectures suggérées

- *BOULÉTREAU & HABERT, 2014*^{BOULETREAUT & HABERT 2014 p.24}
- *COUTURE, 2009-2010*^{COUTURE 2009-2010 p.24}
- *MARCOUX, 2019*^{MARCOUX 2019 p.25}

Abréviations



OQLF : Office Québécois de la Langue Française

Bibliographie



[BOULETREAUT & HABERT 2014] BOULÉTREAU, Viviane; HABERT, Benoît. *Les formats*. In E. Sinatra Michael, Vitali-Rosati, Marcello (éds.), *Pratiques de l'édition numérique*, p. 145-159, 2014. Montréal : Les Presses de l'Université de Montréal. <<http://www.parcoursnumeriques-pum.ca/les-formats>>

[COUTURE 2009-2010] Couture, Cynthia. *L'élaboration et l'implantation d'un cadre de classification pour la gestion des documents numériques et sur support papier de la Ville de Brossard (notes et bilans d'expérience)*. *Archives*, 41(1), 71-91, 2009-2010. http://archivistes.qc.ca/cora/afficheFic.php?fic=vol41_1/41_1_couture.pdf

[MAS 2007] Mas, Sabine. *Schémas de classification et repérage des documents administratifs électroniques dans un contexte de gestion décentralisée des ressources informationnelles*. Thèse de doctorat. Montréal : Université de Montréal, École de bibliothéconomie et des sciences de l'information. 2007. <http://hdl.handle.net/1866/1433>

Webographie



[GDT] Office québécois de la langue française (OQLF). *Grand dictionnaire terminologique*. <http://gdt.oqlf.gouv.qc.ca/>

[MARCOUX 2019] MARCOUX, Yves. *Les jeux de caractères*. 2019.<<http://marcoux.ebsi.umontreal.ca/enseignement/6052/JeuxCar/JeuxCar.html>> (consultée le 2020-08-20).

Index



Chemin absolu	15
Chemin relatif.....	15
Fichiers en format binaire.....	4
Fichiers en format texte	4
Formats de fichiers.....	4
Formats fermés	4
Formats normalisés	4
Formats ouverts	4
Formats propriétaires	4
Jeux de caractères	4
Organisation intellectuelle des fichiers	15
Organisation physique des fichiers	15
Règles de nommage.....	14
Ressources en lien avec le cours	22

Crédits des ressources



p. 3

<http://creativecommons.org/licenses/publicdomain/4.0/fr/>, johnny_automatic

p. 4

<http://creativecommons.org/licenses/publicdomain/4.0/fr/>, maoriveros