

La gestion des données

Emmanuel DUBOIS

I- Les grandeurs quantifiant les données

Bit : Acronyme de Binary digiT, plus petite grandeur informatique pouvant avoir, de
 Octet : Association de 8 bits accolés les uns derrière les autres. Son sigle est la lettre *o*
 Mot : Association de 8, 16, 32 ou 64 bits correspondant à la taille de mémorisation des données à l'intérieur d'un processeur (cerveau d'un ordinateur)

Unités binaires

1 <i>kilo</i> octet binaire (kibiocet)	Kio	=1024 octets	2^{10} octets	1 024 octets
1 <i>méga</i> octet binaire (mébioctet)	Moi	=1024 Kio	2^{20} octets	1 048 576 octets
1 <i>giga</i> octet binaire (Gibiocet)	Gio	=1024 Moi	2^{30} octets	1 073 741 824 octets
1 <i>tera</i> octet binaire (Tébioctet)	Tio	=1024 Gio	2^{40} octets	1 099 511 627 776 octets
1 <i>péta</i> octet binaire (Pébioctet)	Pio	=1024 Tio	2^{50} octets	1 125 899 906 842 624 octets
1 <i>exa</i> octet binaire (Exbioctet)	Eio	=1024 Pio	2^{60} octets	1 152 921 504 606 846 976 octets
1 <i>zetta</i> octet binaire (Zébioctet)	Zio	=1024 Eio	2^{70} octets	1 180 591 620 717 411 303 424 octets
1 <i>yotta</i> octet binaire (Yobioctet)	Yio	=1024 Zio	2^{80} octets	1 208 925 819 614 629 174 706 176 octets

Unités usuelles

1 <i>kilo</i> octet	Ko	=1000 octets	10^3 octets	1 000 octets
1 <i>méga</i> octet	Mo	=1000 Ko	10^6 octets	1 000 000 octets
1 <i>giga</i> octet	Go	=1000 Mo	10^9 octets	1 000 000 000 octets
1 <i>tera</i> octet	To	=1000 Go	10^{12} octets	1 000 000 000 000 octets
1 <i>péta</i> octet	Po	=1000 To	10^{15} octets	1 000 000 000 000 000 octets
1 <i>exa</i> octet	Eo	=1000 Po	10^{18} octets	1 000 000 000 000 000 000 octets
1 <i>zetta</i> octet	Zo	=1000 Eo	10^{21} octets	1 000 000 000 000 000 000 000 octets
1 <i>yotta</i> octet	Yo	=1000 Zo	10^{24} octets	1 000 000 000 000 000 000 000 000 octets

II – Les unités de stockage de données

- Les unités de stockage de données sont les supports matériels permettant la mémorisation et la conservation des données entre deux sessions sur un ordinateur.
- Ils doivent être préparés afin de pouvoir recevoir un **système de fichiers** leur permettant de rendre effectif leur service de conservation de données.
- Un **système de fichier** est ce qui permet de transformer un stockage linéaire des données sur le disque en une arborescence de fichiers et de dossiers matérialisée par des noms symboliques auxquels sont associées des propriétés et un contenu. Il existe plusieurs systèmes de fichiers qui se différencient par la taille des unités de disques et la taille des fichiers supportés mais également par les propriétés étendues supportés sur les fichiers (listes de contrôle d'accès, chiffrement, compression) : FAT(*File Allocation Table*) pour les cartes mémoires et les clés USB, NTFS (*New Technology File System*), APFS (Apple File System) ou EXT4 pour les disques fixes.

II- Les unités de stockage de données

A - Types d'unités de disque :

- Disques locaux

- Supports magnétiques : Disques durs, bandes magnétiques
- Supports optiques : Blue Ray Disk, DVD, CD
- Supports à base de circuits : Solid State Disk (SSD), Clés USB, Carte mémoires

Mise en place :

Branchement sur un Bus de donnée relié au processeur de l'ordinateur

- Bus Serial ATA,
- Bus PCI
- Universal Serial Bus

- Disques distants

Unités de disque appartenant à un autre système informatique accessibles à distance (exemples espace Cloud)

Mise en place :

- Procédure de connexion avec authentification

II- Les unités de stockage de données

B - Préparation :

- 1. Initialisation** : Attribution d'un identifiant unique destiné à différencier les disques lorsque ces derniers ne sont pas fixes et initialiser les clé d'un éventuel chiffrement.
- 2. Partitionnement** : on décide si le disque est utilisé dans son ensemble ou si on le décompose en plusieurs unités indépendantes. Chaque ensemble est appelé une partition. Sur chaque disque existe au moins une partition dite principale, qui sera la première reconnue par l'ordinateur. Si une partition principale est activée, alors l'ordinateur cherchera à charger le système d'exploitation (Windows, OSX, Linux) depuis cette partition au démarrage. Un multi-partitionnement permet également d'installer plusieurs systèmes d'exploitation sur une même unité de disque.

II- Les unités de stockage de données

B - Préparation

3. **Formatage :**

Chaque partition est indépendante dans le sens où elle peut contenir un système de fichier différent et autonome.

Pour chaque partition on choisit le système le plus adapté et on amorce le système de fichier utilisé par l'unité de stockage.

4. **Montage ou attachement :**

On attache l'unité à un dossier dans l'arborescence de l'ordinateur.

Sous Windows on lui assigne une lettre entre C et Z accessible dans le dossier *Ce PC*.

III- Entretien d'un système de fichier

Les systèmes de fichiers sont des structures complexes et à ce titre doivent être entretenues par :

1. **Formatage** : Permet de réinstaller un nouveau système de fichier, au prix d'un effacement total du contenu du disque.
2. **Retrait** ou **Ejection** : Lorsqu'un disque amovible accessible en écriture est retiré il convient de s'assurer que le retrait laissera le système de fichier dans un état cohérent en terminant proprement les opérations en cours de transferts de données. Le risque d'un retrait trop rapide est de laisser le système de fichier dans un état incohérent avec risque de perte de données et perte d'espace de stockage jusqu'à la prochaine vérification du système de fichier.
3. Le **nettoyage** : suppression de fichiers temporaires et inutiles (corbeille)
4. La **défragmentation** : Les cycles d'ajouts et suppressions de fichiers accentuent la fragmentation du système : les fichiers ont tendance à ne plus être écrits d'un seul tenant sur une zone continue du disque mais répartis en plusieurs morceaux dans différents emplacements du disque, l'ordinateur reconstituant les fichiers de manière transparente lorsque l'utilisateur y accède au prix d'un très léger ralentissement des performances. Il est recommandé d'avoir une unité peu fragmentée moins pour les performances que pour la sécurité : en cas d'altération partielle du disque des logiciels spécifiques permettent de récupérer les fichiers non fragmentés qui n'ont pas été altérés et sont d'un moins bon pronostic sur une unité fragmentée.

III- Entretien d'un système de fichier

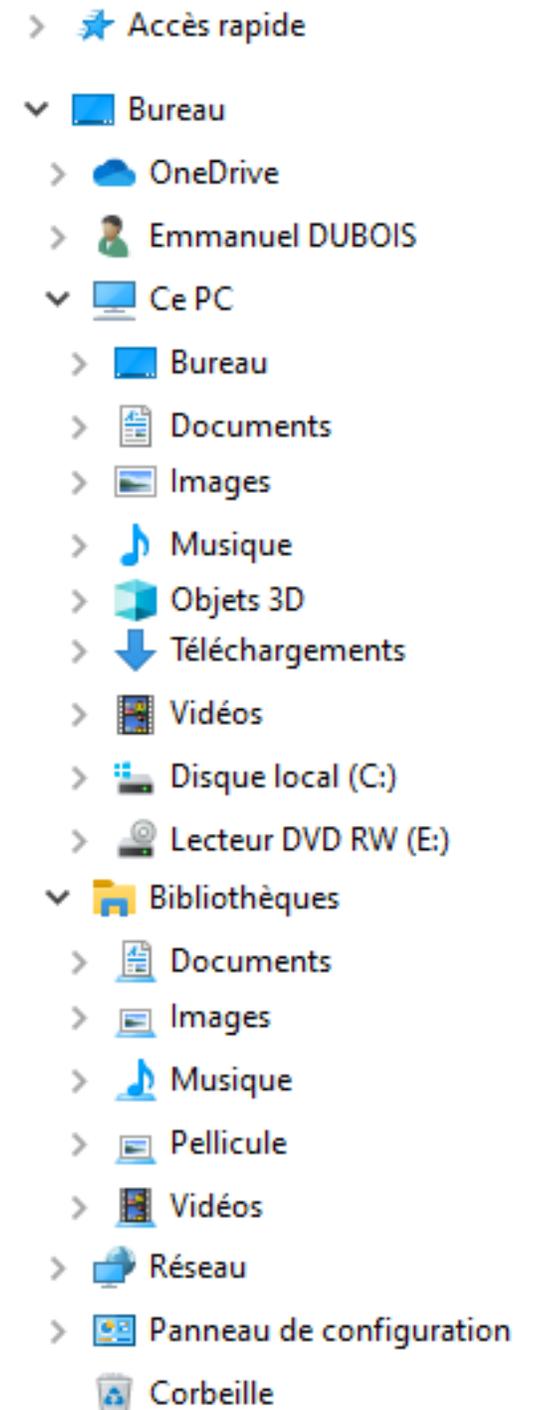
5. Vérification : Des incidents d'écritures de fichiers peuvent altérer le système de fichier (altération de l'arborescence ou plus généralement non libération de l'espace associé à des fichiers pourtant inaccessibles), voire l'obsolescence peut entraîner l'apparition de secteurs défectueux sur le disque : emplacement qui ne sont plus aptes à mémoriser des données. Plus couramment, le retrait inopiné d'un support de stockage amovible est la cause la plus probable des problèmes qui peuvent survenir sur son système de fichier. Lors de la vérification, les erreurs sont corrigés et les blocs physiquement défectueux sont marqués comme inutilisables en vue de prévenir des erreurs ultérieures.

III – Entretien d'un système de fichier

6. **Sauvegarde** : Un disque est un matériel fragile qui peut être victime de nombreux aléas tant matériels que logiciels. Il est donc recommandé de le sauvegarder régulièrement sur un disque externe ou sur un disque distant au travers du réseau.
7. **Contrôle antivirus** : Les infections informatiques (virus, bombe logique, cheval de Troie, ver) peuvent être la cause d'indisponibilité des données, d'altération ou de destruction de données, ou de vol d'informations. Sauf absence totale de comportement à risque, il convient de faire régulièrement parcourir les disques par un automate qui va rechercher des signatures (octets caractéristiques) d'infections informatiques connues pour soit supprimer les fichiers infectés, soit les réparer en supprimant l'infection lorsque cela est possible, soit mettre le fichier en quarantaine (le fichier est indisponible au niveau du système de fichiers mais peut être récupéré par une fonction de l'outil antivirus). Le contrôle antivirus d'une unité de stockage peut être programmé, réalisé en temps réel lors de chaque accès à un fichier, ou lancé au cas par cas. Comme de nouvelles infections informatiques apparaissent en permanence, il est nécessaire de tenir à jour la base de signatures de l'antivirus par chargement manuel ou automatique depuis le site de son éditeur.

IV – L'arborescence des dossiers

- Les informations ne gagnent pas à être laissées en vrac dans l'ordinateur. Les documents sont regroupés dans des dossiers.
- Les dossiers forment une arborescence dans le sens où un dossier peut contenir des fichiers mais également des sous-dossiers, pouvant contenir eux-mêmes d'autres dossiers, *etc.*
- Les dossiers sont identifiés par un nom, semblable aux noms de fichiers. Ils ont en outre des attributs comme les fichiers.



V – Les fichiers

Les informations contenues dans un ordinateur, qu'elles concernent des données ou des applications, sont contenues dans des fichiers, rangés dans les dossiers au sein des unités de stockage.

Un fichier est une succession linéaire d'informations au sein d'un même conteneur, identifié par un certain nombre de propriétés (accessibles par un clic droit puis *propriétés* sur un fichier sélectionné):

1. Un **nom**, composé de moins 260 lettres ou chiffres ou ponctuation (à l'exception notable des caractères \ : * ? « < > | interdits).
2. Un **emplacement** ou **chemin d'accès** : nom de tous les dossiers gigognes dans lesquels il faut successivement entrer pour atteindre le fichier.
3. Des **dates** de création, de dernière modification, de dernier accès en lecture.
4. Un **type**, permettant d'identifier l'application qui a servi à créer ce fichier ou qui peut être amenée à l'utiliser. Ce type peut apparaître de deux manières :
 - une extension de quelques lettres suivant le nom de fichier. Elle est séparée de ce dernier par un point.
 - un motif graphique (nommé *icône*) précédant le nom de fichier.

V – Les fichiers

5. une **taille**, exprimée en octets, kilo octets (Ko), Méga octets (Mo) ou Giga Octets (Go) Tera Octets (To) ou Peta Octets (Po). Il y a un rapport de 1000 entre chaque unité (1024 en unités binaires, fréquemment rencontré chez les informaticiens et dans la littérature et notamment dans l’affichage de la taille des fichiers sous Windows).
6. des **attributs** au nombre de quatre : ce sont des drapeaux qui, s’ils sont positionnés, portent les informations décrites ci-après.
 - **Lecture seule** : Le fichier ne peut pas être modifié. En cas d’effacement, une confirmation sera systématiquement demandée.
 - **Archive** : Utilisé pour reconnaître les fichiers devant être sauvegardés. Un fichier créé ou modifié a cet attribut placé et les logiciels de sauvegarde le retirent au moment de la sauvegarde.
 - **Caché** : Le nom du fichier n’apparaîtra pas dans les listes de fichiers.
 - **Systeme** : Le fichier est un fichier absolument nécessaire au démarrage de l’ordinateur.

V – Les fichiers

7. des métadonnées

- documents bureautiques : titre sujet mots-clés, auteurs, organisation,...)
- multimédia : sujet, auteur, caractéristiques techniques et de classement

8. des propriétés étendues :

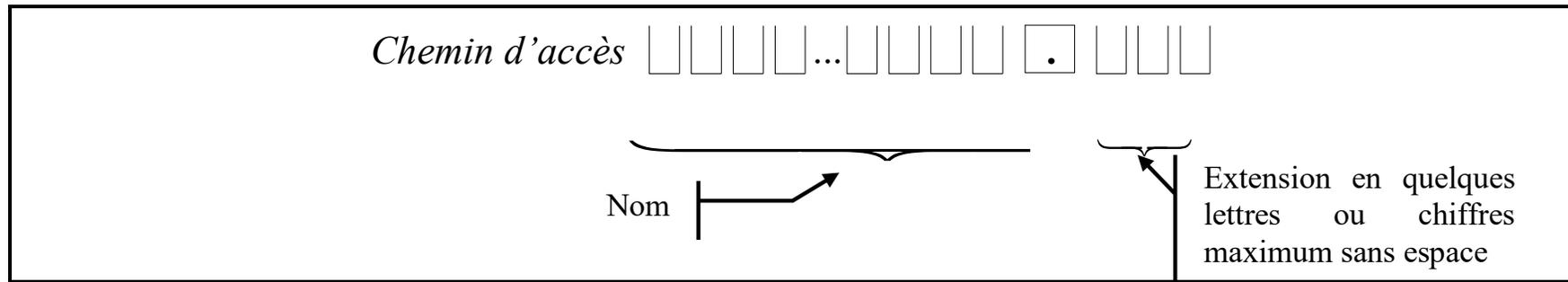
- **Indexation** : Précise si le contenu du document doit être accessible dans le cadre d'une recherche en texte plein, sinon la recherche ne s'intéressera qu'au nom de fichiers et ses métadonnées.
- **Chiffrement** : le chiffrement est géré de manière transparente et ne se fait pas par mot de passe mais avec une clé stockée au sein de votre compte utilisateur et générée d'après le mot de passe du compte. Un autre utilisateur, y compris l'administrateur de l'ordinateur, n'ayant pas le même compte ne peut lire le fichier faute de clé de déchiffrement. En cas de perte du compte utilisateur, voire de changement du mot de passe associé au compte, les données ne vous seraient plus accessibles, aussi, avant d'activer le chiffrement sur des fichiers il convient de créer et sauvegarder sur un disque externe un certificat protégé par mot de passe contenant la clé de déchiffrement.

VI – Manipulations des fichiers et dossiers

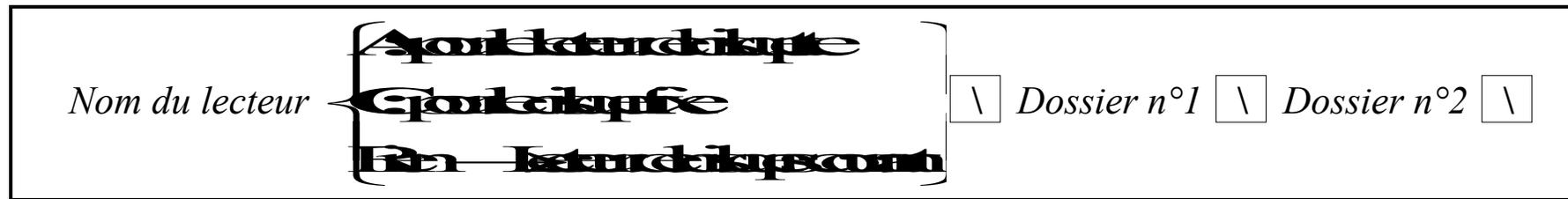
- Un fichier ou dossier doit être sélectionné en préalable à toute action le concernant
 - Sélection simple
 - Sélection multiple
 - Recherche
- Opérations sur les dossiers :
 - Création, Suppression, Renommage, Copie, Déplacement
- Opérations sur les fichiers :
 - Création, Suppression, Renommage, Edition des propriétés et métadonnées, Copie et déplacement

VII – Expression des chemins d'accès aux fichiers et dossiers (sur PC)

Les noms de fichiers ou de dossiers sont formés de la manière suivante :



Le chemin d'accès est formé de la manière suivante :



Exemple de nom de fichier avec chemin d'accès :
C:\TRAVAIL\TEXTES\RAPPORTS\RAPPORT1.TXT

VII – Expression des chemin d'accès aux fichiers et dossiers (sur PC)

- Le chemin d'accès n'est pas absolument nécessaire : s'il est omis, le nom du fichier est compris comme étant dans le dossier courant.
- Chemin d'accès absolu :
Dans l'exemple C:\TRAVAIL\TEXTES\RAPPORTS\RAPPORT1.TXT, le chemin d'accès est précisé de manière absolue en citant tous les dossiers présents jusqu'au lecteur de disque.
- Chemin d'accès relatif :
Il peut être également défini de manière relative, c'est à dire exprimé par rapport au dossier courant. Pour cela :
 - on ne nomme que les dossiers contenus à partir du dossier courant
 - on utilise la séquence •• qui signifie que l'on remonte au dossier père
 - On utilise • qui représente le dossier courant.

VII – Expression des chemins d'accès aux fichiers et dossiers (sur PC)

- si le dossier courant est C:\TRAVAIL, alors le fichier C:\TRAVAIL\TEXTES\RAPPORTS\RAPPORT1.TXT peut être exprimé par TEXTES\RAPPORTS\RAPPORT1.TXT
- si le dossier courant est C:\TRAVAIL\DEVOIRS\MATH alors le fichier C:\TRAVAIL\TEXTES\RAPPORTS\RAPPORT1.TXT peut être exprimé par ..\..\TEXTES\RAPPORTS\RAPPORT1.TXT
- si le dossier courant est C:\TRAVAIL\TEXTES\RAPPORTS alors le fichier C:\TRAVAIL\TEXTES\RAPPORTS\RAPPORT1.TXT peut être exprimé par RAPPORT1.TXT ou .\RAPPORT1.TXT.
- si le dossier courant est C:\DOS, alors le fichier C:\TRAVAIL\TEXTES\RAPPORTS\RAPPORT1.TXT peut être exprimé par \TRAVAIL\TEXTES\RAPPORTS\RAPPORT1.TXT car le nom du lecteur de disques, omis, est automatiquement remplacé par le lecteur courant, c'est à dire C:

VII – Expression des chemin d'accès aux fichiers et dossiers (sur PC)

- Filtrage optionnel des noms de fichiers dans un chemin d'accès :

Dans l'expression d'un nom de fichier, il est possible d'utiliser de définir un filtre à l'aide des caractères * et ? qui peuvent être utilisés pour filtrer les fichiers disponibles à un emplacement en fonction de leur nom, de leur extension ou des deux à la fois.

- * utilisé dans le nom du fichier signifie que l'on admet n'importe quelle séquence de caractères jusqu'à la fin du nom ou de l'extension.
- ? utilisé à la position d'un caractère permet d'admettre n'importe quel caractère, mais un et un seul, à une position donnée.
- N'importe quel autre caractère précisé doit être présent dans le nom ou l'extension du fichier à la position indiquée pour qu'un fichier soit affiché.

VII – Expression des chemin d'accès aux fichiers et dossiers (sur PC)

Supposons que nous ayons 5 fichiers dans le dossier courants, nommés :

AAAA.AAA ABCDEF.ABC ABC.CBA CAB.CBA CBADEF.AAA

- **AAAA.AAA** retournera seulement AAAA.AAA.
- ***.*** retournera les 5 fichiers car toutes les noms ou extensions sont autorisés.
- **A*.*** retournera AAAA.AAA, ABCDEF.ABC c'est à dire les seuls fichiers dont le nom débute par la lettre A.
- ***.CBA** retournera ABC.CBA, CAB.CBA c'est à dire les seuls fichiers ayant CBA pour extension du nom.
- ***.??A** retournera nommés AAAA.AAA, ABC.CBA, CAB.CBA, et CBADEF.AAA c'est à dire les seuls fichiers ayant la lettre **A** en 3^{ème} position de l'extension.
- **?B*.C*** retournera ABC.CBA, c'est à dire le seul fichier ayant un **B** en 2^{ième} position du nom et un **C** en 1^{ère} position du nom.
- **D*.*** ne retournera aucun fichier car aucun fichier n'a de nom commençant par la lettre *D*.

VIII- Raccourcis

Lors de l'organisation de ses dossier personnels, il est souvent souhaité qu'un fichier, un dossier, une page web, ou encore une application, que nous appellerons pour la suite *l'élément source*, soit disponible depuis plusieurs emplacement. En revanche, il n'est pas souhaitable de dupliquer cet élément pour des considérations telles que la place occupée, la volonté d'éviter d'avoir à gérer plusieurs versions non synchronisées lors des mises à jour et corrections etc...

La solution est le **raccourci**, qui ne duplique pas un document mais mémorise l'emplacement de l'élément afin que le système puisse le retrouver. Un raccourci peut ensuite être déposé :

- au sein de n'importe quel dossier (et notamment le dossier **Accès Rapide** (anciennement **Favoris**) dont le rôle est justement de collecter les raccourcis vers les dossiers et documents les plus utiles),
- au sein des menus et barre d'outils ou sur le bureau.

VIII- Raccourcis

Création d'un raccourci sur PC :

1. Dans le dossier de destination : clic droit + *Nouveau Raccourci*
2. Glisser déposer de l'élément source jusqu'à destination en maintenant la touche **Alt**.
3. Clic droit sur l'élément source puis *Epingler à Accès Rapide* ou *Envoyer vers ... /Bureau (Créer un raccourci)*