

Section 3 - Les multiples visages du document numérique

version 1

FRANÇOISE BANAT-BERGER
CLAUDE HUC

22 novembre 2011

Table des matières

Les multiples visages du document numérique	3
Chapitre 1. Objet de la section.....	3
Chapitre 2. Notions de base sur le document numérique.....	3
Chapitre 3 L'ordinateur et ses composants.....	8
Chapitre 4 Analogique et numérique.....	10
Chapitre 5 Les différents types de document numérique.....	11
Chapitre 6 Le document numérique dans l'ordinateur.....	11
Chapitre 7 Le document numérique voyage.....	13
Chapitre 8 La compression des données.....	14
Chapitre 9 Les principaux problèmes à résoudre pour l'archivage numérique.....	15
Questions générales sur le numérique : vrai ou faux	17
Exercice : Réviser vos connaissances	20
Solution des exercices	21
Glossaire	24
Bibliographie	26
Webographie	27

Les multiples visages du document numérique

Chapitre 1. Objet de la section

Avant d'entrer plus en détail dans le vif du sujet sur **les différents visages** que peut prendre le document numérique, il nous apparaît important :

- de décrire la nature particulière de ce document numérique,
- de donner quelques éléments terminologiques essentiels,
- de commencer à démystifier sa structure interne,
- de montrer comment il naît et circule à l'intérieur de l'ordinateur,
- de comprendre de quelle façon ce document se situe par rapport à ce que sont les fichiers et les répertoires que nous manipulons quotidiennement.

Nous voulons également montrer comment le document circule à travers les réseaux et à quelles conditions.

Nous parlerons de la compression de données, fonction fort utile pour les fichiers volumineux qui ne peuvent circuler facilement en raison des limites techniques des réseaux actuels.

Nous terminerons par une première synthèse des différents problèmes à résoudre pour assurer la conservation à long terme d'un document numérique.

Objet spécifique :

- Comprendre de quelle façon l'information est organisée dans un ordinateur.
- Comprendre la réalité des réseaux de communication et les technologies de transfert de l'information

Chapitre 2. Notions de base sur le document numérique

Aujourd'hui, un ordinateur, c'est devenu banal. Que ce soit pour le travail ou une utilisation plus ludique, la plupart des gens ont déjà eu affaire à cet outil technologique. Beaucoup l'utilisent même sur une base quotidienne. Mais connaît-on vraiment cet outil dont on ne saurait plus ce passer ? Comment fonctionne-t-il ?

L'ordinateur est le nerf de notre société de l'information. Mais connaît-on vraiment cet outil si polyvalent et si précieux qu'on ne saurait plus s'en passer ?

2.1 - Le système binaire

Entre humains, l'information peut être transmise de différentes façons. Nous utilisons, pour ce faire, différentes formes de « langages », comme la parole, l'écrit et même le geste. Pour échanger de l'information avec un ordinateur, ces mêmes formes de langages ne suffisent pas : l'information que nous désirons transmettre doit être codée de manière à ce qu'elle soit comprise par l'ordinateur. Car l'ordinateur parle son propre langage, basé sur ce que l'on appelle « le système binaire ».

À la fin des années 30, Claude Shannon démontra qu'une machine pouvait manipuler de l'information. Comment ? En exécutant des opérations logiques à l'aide de « contacteurs » fermés pour « vrai » et ouverts pour « faux », associant à « vrai » le nombre « 1 » et à « faux » le nombre « 0 ». Pensez à un interrupteur ouvert ou éteint. En effet, le signal de base d'un ordinateur est matérialisé par l'absence ou non d'un courant électrique pendant un laps de temps donné :

- si vous faites passer un courant électrique dans le fil de votre radio pendant une seconde, celle-ci s'allumera; on peut traduire cette action par le chiffre 1,
- si vous ne faites pas passer de courant électrique dans le fil pendant le même laps de temps, la radio demeure éteinte et cette information sera traduite par le chiffre 0.

C'est là la base du langage avec lequel on stocke et utilise toute l'information contenue dans un ordinateur.

2.2 - Bits et octets

Le bit[⊖] (pour « binary digit ») est **la plus petite unité d'information manipulable par un ordinateur**. Un bit vient préciser si le courant passe ou non. Ainsi, il ne peut avoir que la valeur de 0 ou de 1.

Un bit tout seul ne signifie pas grand chose, mais **groupés par huit, les bits forment des octets** (« bytes » en anglais) qui sont à même de représenter différentes informations[⊖] comme les lettres de l'alphabet et les chiffres.

4	0100	C	1100
5	0101	D	1101
6	0110	E	1110
7	0111	F	1111

Tableau 1 La représentation hexadécimale des bits

Ainsi, il sera plus simple de représenter la lettre Z par le code hexadécimal "5A" que par l'octet "1011010". Nous réutiliserons la représentation hexadécimale dans plusieurs parties du présent module.

2.5 - Premières notions de codage

Le codage consiste à établir une correspondance entre un élément ayant un premier niveau de sémantique et une représentation de cet élément dans un système conventionnel.

Ainsi la représentation du phonème « A » de notre alphabet latin par le symbole graphique « A » que nous utilisons constitue un codage qui est le plus souvent perçu comme implicite.

Exemple : Le système d'écriture tactile Braille standard

Dans ce système, un caractère est représenté par la combinaison de 1 à 6 points en relief, disposés sur une matrice de 2 points de large sur 3 points de haut.

Titre image : La représentation des lettres A et B dans l'alphabet Braille

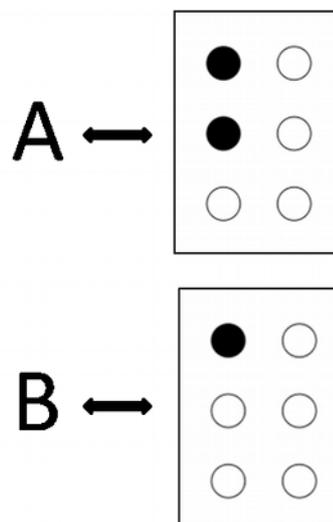


Image 1 Braille



Image 2 La représentation des lettres A et B dans l'alphabet Morse

Exemple

L'alphabet Morse est un autre type de codage :

L'alphabet morse est quant à lui, un code permettant de transmettre un texte à l'aide de séries d'impulsions courtes et longues. Ce codage des caractères assigne à chaque lettre, chiffre et signe de ponctuation une combinaison unique de signaux intermittents. C'est un précurseur des communications numériques qui n'est pratiquement plus utilisé aujourd'hui. Voici la représentation des lettres A et B dans cet alphabet.

Il paraîtra évident, pour un archiviste qui doit conserver des documents en code Braille ou en code Morse, de conserver en même temps la description du code en question. Dans la représentation binaire, c'est le même principe.

Représentation A et B

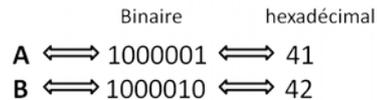


Image 3 Représentation normalisée des lettres A et B sous forme binaire ainsi que la représentation hexadécimale correspondante.

2.6 - Données, information et documents

Afin d'établir une terminologie commune au présent module, nous proposons quelques définitions essentielles. Les définitions des termes "Information" et "Données" sont celles de **la norme ISO 14721: 2003 "Modèle de référence pour un système ouvert d'archivage"**, norme essentielle pour l'archivage des documents numériques et dont nous donnerons les caractéristiques principales dans la partie 5 de ce module, dédiée en majeure partie à ce Modèle.

Un texte sur une feuille de papier, un texte écrit en braille, un fichier issu d'un logiciel de bureautique, une image, un tableau de chiffres, un graphique sont des conteneurs d'information, c'est-à-dire des données : lorsqu'ils sont transmis à un individu, celui-ci pourra, sous certaines conditions, en dériver une information qui modifiera un capital de connaissances personnel ou collectif.

La dérivation de l'information à partir d'une donnée pourra imposer aussi le recours à un appareil.

Ce qu'il importe de retenir ici, c'est que l'information est indépendante de la manière dont elle sera représentée par des données.

Nous parlerons également souvent de Document.

Définition

Information : L'information est définie comme une connaissance pouvant être échangée. En pratique, l'information est donc un élément de connaissance susceptible d'être codé pour être conservé, traité ou communiqué.

Donnée : Représentation formalisée de l'information, adaptée à la communication, l'interprétation ou le traitement. Par exemple : une séquence de bits, un tableau de nombres, les caractères d'une page, un enregistrement audio, etc. Une donnée est donc un conteneur porteur d'une information ou d'un fragment d'information.

Document : Tout écrit ou enregistrement considéré comme une unité documentaire (ISO 15489-1 "Records Management").

Le document est donc un « ensemble d'informations enregistrées », considéré comme une unité qui pourra être utilisable à des fins de consultation ou comme preuve.

Document numérique : C'est un document constitué sous forme numérique, soit de façon native/initiale, soit par numérisation de document non numérique.

Dans le domaine du numérique, l'association entre le support et l'information qu'il porte perd son sens. Le document pourra au cours du temps être stocké sur des supports différents sans pour autant que cela ne change sa nature ni n'altère son intégrité.

2.7 - Un document numérique très simple

Création d'un tout petit **document** : Nous allons regarder de plus près ce que peut contenir un document numérique très simple. Créons, pour cela, à l'aide d'un éditeur de texte disponible sur tous les ordinateurs (bloc-notes, textedit...), un fichier contenant simplement la phrase « **Les feuilles mortes** » (1), titre du célèbre poème de Jacques Prévert.

Enregistrons ce fichier sur un répertoire de notre ordinateur. Le nom du fichier sera suivi d'un point et de l'extension "txt" signifiant qu'il s'agit d'un fichier de type texte intégral. (2)

Du point de vue de l'information qu'il contient, ce fichier constitue un document très élémentaire.

Précisons en outre qu'en matière de codage, **l'espace entre deux mots est considéré comme un caractère à part entière** : le caractère "espace" est porteur d'une information particulière, à savoir qu'il faut insérer un espace pour séparer deux mots. En prenant cette considération en compte, nous pouvons compter que notre fichier contient 19 caractères : 17 caractères alphabétiques et deux caractères "espace".

En pointant avec la souris sur le nom du fichier et en appuyant sur le bouton droit, nous aurons accès à un menu dit contextuel nous permettant d'accéder aux propriétés du fichier. Ces propriétés sont résumées dans une fenêtre qui s'ouvre lorsqu'on pointe sur "propriétés" et qu'on appuie sur le bouton gauche de la souris. (3)

Dans les propriétés, nous pouvons observer que la taille du fichier est de 19 octets, ce qui est très intéressant car cela nous montre que ce fichier n'est qu'une simple suite d'octets, chaque octet correspondant à un caractère.

Il est difficile d'imaginer plus simple.

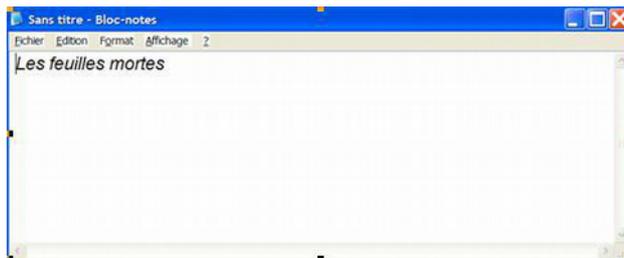


Image 4 1 - les feuilles mortes

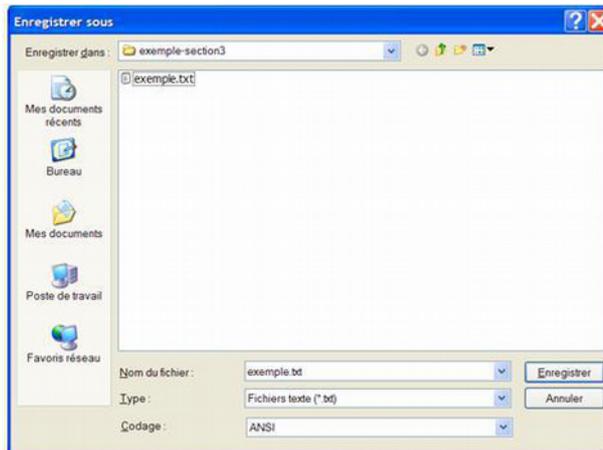


Image 5 2.txt

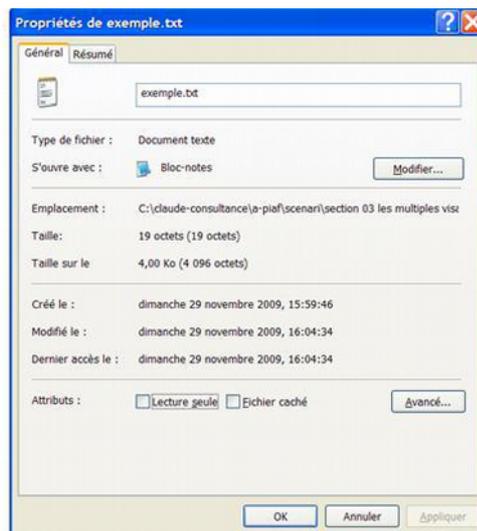


Image 6 3- Propriétés



Complément : Décortiquons notre petit document

Le recours à un éditeur hexadécimal nous permet de voir ce que le fichier contient réellement (et pas seulement ce que l'éditeur de texte nous avait montré). Cet éditeur hexadécimal nous permet de voir en quoi consiste réellement la suite des octets contenus dans le fichier :

4c 65 73 20 66 65 75 69 6c 6c 65 73 20 6d 6f 72 74 65 73 (en représentation hexadécimale)

Ou encore :

01001100 01100101 01110011 00100000 01100110 01100101 01110101 01101001 01101100 01101100 01100101 01110011 00100000 01101101 01101111 01110010 01110100 01100101 01110011 (en représentation binaire)

Sachant que dans le codage normalisé utilisé, 4c correspond à la lettre "L" (majuscule), 65 correspond à la lettre "e" minuscule, etc., nous pouvons ainsi décortiquer le contenu de notre petit document et constater qu'il n'y a rien d'obscur ni de magique dans tout cela.



Remarque

Nous avons utilisé pour cela, l'éditeur hexadécimal gratuit « Hex editor neo » disponible sur Internet et que vous

Chapitre 3 L'ordinateur et ses composants

Maintenant que nous comprenons un peu mieux comment l'information est codée par l'ordinateur, il reste à voir comment ce dernier est constitué.

3.1 - Architecture de base

Comme nous l'avons vu précédemment, l'ordinateur ne fait qu'une chose : manipuler des nombres binaires. La partie de l'ordinateur qui effectue tous ces calculs se nomme **le processeur ou microprocesseur**. Nous verrons plus loin les composants internes à l'ordinateur plus en détail mais retenons principalement que le **processeur** lit des instructions dans la **mémoire** de l'ordinateur (appelée mémoire vive) et les exécute.

Un programme est une séquence d'instructions. Notons que la mémoire de l'ordinateur contient aussi des données sur lesquelles l'ordinateur doit travailler.

Le processeur est relié à la mémoire par des dispositifs de communication appelés des **bus** de données, qui transportent, comme l'indique le nom, toutes les données de l'un à l'autre. Les bus sont eux-mêmes gérés par plusieurs composants, regroupés en général en une seule puce : le chipset ou jeu de composants

Tous ces composants sont architecturés autour d'une carte mère (un circuit imprimé) qui constitue le cœur de l'ordinateur.

3.2 – Les composants du cœur de l'ordinateur

Voici plus en détails les composants internes à l'ordinateur, l'important n'étant pas d'en retenir tous les éléments mais plutôt de mieux comprendre son fonctionnement.

Processeur ou microprocesseur : il est constitué de circuits électroniques miniatures rassemblés sur une seule puce. Le processeur interprète et exécute les instructions.

Bus de données : il transporte les données entre les différents blocs fonctionnels et composants d'un ordinateur. Un bus qui n'interconnecte que deux dispositifs est appelé un port.

Chipset (jeu de composants) : il est le cœur de la carte-mère et gère les échanges de données sur les différents bus. Il s'agit d'un ensemble de circuits intégrés qu'on appelle généralement « Puce ».

Carte-mère : c'est la carte principale d'un ordinateur qui comporte tous les composants nécessaires au fonctionnement de celui-ci. On y retrouve entre autres la mémoire centrale, les bus et les connecteurs d'extensions.

Horloge système : contrôle et synchronise le microprocesseur et les composants associés. Sa « fréquence » ou vitesse est exprimée en mégahertz (MHz).

La mémoire "cache" : zone de mémoire ultra-rapide intégrée dans les processeurs où sont conservées les instructions et données qui reviennent le plus souvent.

Connecteurs d'extensions : sur la carte-mère, emplacements disponibles destinés à recevoir des cartes d'usages divers.

Entrées/sorties : les ordinateurs comportent des ports série et parallèle pour la liaison avec des périphériques de type imprimante, modem, ...

Mémoire vidéo : le mémoire, dite VRAM, où se stockent les images à afficher à l'écran.

Les cartes vidéos et cartes son : branchées sur un connecteur d'extension, les cartes vidéos permettent de décharger le microprocesseur des calculs de l'affichage (c'est-à-dire qu'elles prennent en charge tout ce qui concerne l'affichage; c'est un circuit électronique indépendant qui n'est pas intégré à la carte-mère), tandis que les cartes son permettent d'enregistrer et de reproduire du son et de piloter un lecteur de CD-ROM.

3.3 – Les périphériques

Les périphériques sont des dispositifs extérieurs à l'ordinateur. Ils vont jouer un rôle essentiel dans le transfert d'informations vers l'ordinateur ou à partir de celui-ci.

Les périphériques d'entrée :

La souris : permet de sélectionner et d'activer ce qui est affiché à l'écran.

Le clavier : permet de communiquer des caractères ou des fonctions au programme ou au microprocesseur, clavier et souris ont à peu près les mêmes qualités, disons qu'ils sont complémentaires.

Le scanner, le caméscope ou l'appareil photo numérique, ou encore la webcam sont aussi des périphériques d'entrée puisqu'ils permettent l'introduction de données dans l'ordinateur.

Les périphériques de sortie :

L'écran : permet la transmission d'informations visibles à l'utilisateur (texte, images, vidéo...)

L'imprimante : permet le transfert de documents depuis l'ordinateur vers un support papier.

Les périphériques d'entrée-sortie :

Le disque dur : lieu de stockage principal des données. La taille d'un disque-dur se compte en gigaoctets (GO)

Le lecteur de CD-Rom et de DVD : périphérique de lecture qui permet de lire les informations se trouvant sur un CD-Rom ou sur un DVD

Le lecteur de clés USB, le lecteur de disquette : (de plus en plus rare aujourd'hui), le **Modem** (pour les connexions au réseau téléphonique), la carte réseau (pour les connexions à un réseau numérique) sont également des périphériques d'entrée-sortie.

Chemin d'accès : Pour chaque périphérique de lecture/écriture, le système d'exploitation attribue un chemin d'accès. Windows, par exemple, représente ces chemins d'accès par des lettres : A pour le lecteur de disque 8 :

pour le disque dur et D pour le lecteur de CD-Rom.

3.4 – Les supports de stockage

Les documents que l'on crée peuvent être enregistrés par l'ordinateur pour pouvoir être utilisés plus tard. On dispose, pour ce faire, de différents supports de stockage appelés lecteurs par l'ordinateur.

Ne nous trompons pas, il s'agit ici des supports de stockage de l'information utilisables dans le poste de travail informatique habituel. Ce n'est que plus tard, dans ce module, que nous analyserons la question des supports de stockage vus du point de vue de la conservation à long terme de l'information.

Voici les principaux supports de stockage utilisés :

Disque dur : Ils ont une capacité d'environ qui varie de 100 Go à 1 To. Cette capacité augmente de façon continue avec les progrès de la technologie. Le système d'exploitation et les logiciels qu'on utilise sont stockés le disque dur, et il ne doit pas être plein pour que le système d'exploitation fonctionne correctement.

CD-R : le CD-R a une capacité de 650 Mo. C'est le support intéressant pour stocker des documents textuels. Le CD-R peut être réinscriptible ou non.

Support de stockage



Image 7 CD-R

DVD (Digital versatil disc) : c'est un disque optique dont la capacité de stockage se situe entre 4,7 et 17 Go en fonction du nombre de faces et du nombre de couches utilisables par face. Les lecteurs DVD peuvent lire les CD-Rom. C'est un support qui devient fiable et qui offre plus de possibilités que le CD-Rom, notamment pour stocker du son ou de la vidéo. Sa capacité reste cependant très inférieure à celle des disques durs.

Le Blu-Ray: C'est un disque optique à haute résolution, breveté par Sony, et dont la capacité varie entre 25 et 50 Go (simple ou double couche). Il est principalement utilisé pour stocker des vidéogrammes en haute-définition mais il permet en pratique de stocker tout type de document numérique. C'est un support émergent et les équipements de lecture ne sont pas encore vendus de façon standard en même temps que les ordinateurs.

Support de stockage 2



Image 8 Blu-ray

Clé USB : sa capacité varie aujourd'hui que quelques centaines de Mo à 10 Go. C'est un support de stockage temporaire utile pour transférer des données d'un ordinateur à un autre.

Disquette : la disquette de 3"1/2 a une capacité de stockage de 720 Ko (disquette basse densité) ou 1,44 Mo (disquette haute densité). C'est un support de stockage en voie de disparition comme l'a été dans le passé, la disquette de 5"1/4.



Image 9 La disquette 3"1/2

Chapitre 4 Analogique et numérique

Puisque l'on traite dans ce module de document numérique, il convient d'expliquer en quoi consiste le numérique.

4.1 – Deux formes de mémoire de l'information

On appelle signal la représentation physique d'une information. En fait, un signal est la variation d'une grandeur (tension, courant électrique par exemple) en fonction de la durée. Les phénomènes qui nous entourent, lorsqu'ils sont quantifiables, passent généralement d'une valeur à l'autre sans discontinuer : on dit alors de ces phénomènes qu'ils sont continus. Pour mémoriser les valeurs de ces phénomènes, il faut les enregistrer sur un support physique qui peut prendre des valeurs continues

L'analogique et le numérique sont deux formes différentes de mémoire de l'information :

Analogique : la mémorisation peut se faire par une analogie au phénomène. On parle de données analogiques :

- le gravage d'un disque vinyle reproduit la vibration du son dans la galette plastique. Il en est de même pour l'enregistrement du son sur un cylindre phonographique,
 - un thermomètre enregistreur va tracer les variations de la température au cours du temps sur un rouleau de papier,
- Les données analogiques présentent un caractère continu.

Numérique : l'information est transformée par un processus qui la rend numérique – on parle de données numériques

- la photo est scannée (ou numérisée par l'appareil photo) pour mise en mémoire,
- le son est échantillonné et stocké dans un format propre à la musique (ex : MP3),
- l'enregistrement des variations de la température se présentera alors comme une suite de couples de valeurs (temps1, température1), (temps2, température2), etc.

Dans ce cas, le signal ne peut prendre que des valeurs bien définies, en nombre limité, on parle alors de signal numérique. Un signal numérique se traduit ainsi par des « 0 » et des « 1 » ; c'est donc un nombre binaire ou une collection de nombres binaires.

Les données numériques se présentent toujours comme une suite ou un ensemble de valeurs. Elles ont donc un caractère discontinu, discret.

L'univers analogique existe depuis de nombreux siècles, il s'est fortement développé avec le début de l'électricité, son correspondant numérique est apparu plus récemment avec l'ère de l'informatique.

Représentation schématique des deux formes de mémoire de l'information.

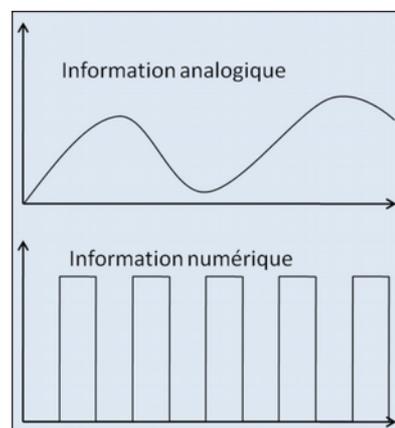


Image 10 formes de mémoire

4.2 – Avantages et limites de la représentation numérique

Un avantage principal du numérique par rapport à l'analogique est la possibilité de stockage, de transport, et de restitution des données sans que celles-ci soient altérées. Il permet également d'effectuer aisément des transformations, des traitements de toute nature qui permettent de créer des informations nouvelles.

En effet, le signal analogique peut être copié (sur un ruban magnétique, sur un disque de vinyle, etc.) mais il se dégrade à chaque étape. Une copie d'un signal analogique est donc moins bonne que l'original, une copie d'une copie encore moins bonne et ainsi de suite, jusqu'à ce que le signal soit complètement détérioré et inexploitable.

En comparaison, la copie numérique est comme un clone de l'original. Le signal, codifié une fois pour toute, ne se dégrade plus et peut en principe passer sur plusieurs supports et canaux différents sans qu'il y ait de pertes ou d'altération d'informations.



Remarque

On appelle numérisation la transformation d'un signal analogique en signal numérique.

Par contre, nos oreilles ne savent entendre en numérique ! Un signal audio numérisé pour son stockage devra à nouveau être converti en analogique pour pouvoir être écouté.

Voir sur ce sujet le site Web ¹

Chapitre 5 Les différents types de document numérique

Les documents numériques pourront être classés en différentes catégories en fonction de leur mode de création

5.1 - Le document « né numérique »

C'est un document créé directement sur ordinateur et qui ne contient que des éléments générés par le logiciel qui l'a créé, que ce soit du texte ou des dessins. On parlera aussi de document numérique natif.

On parle en ce cas d'un document né numérique.

Les photographies issues de l'appareil numérique sont de la même nature.

5.2 - Le document numérisé

La numérisation peut se faire à partir de toute forme de support, mais le plus souvent, le document « numérisé » est à l'origine un document papier (ou photo ou film, etc.) « numérisé » par l'intermédiaire d'un numériseur ou scanner.

Celui-ci transforme l'image papier en un quadrillage très fin dont chaque élément est appelé pixel. A chaque pixel correspondra une valeur sous forme numérique. Ainsi « numérisé », le document papier peut être exploité par un logiciel. Il sera alors possible de générer des copies ou des variantes sans qu'il perde de sa qualité.

De la même façon, on peut « numériser » un document sonore ou audiovisuel, afin d'exploiter plus largement son contenu.

5.3 - Le document hybride

Le document hybride est une combinaison des deux premiers.

Il peut s'agir, par exemple, d'un document texte créé par ordinateur auquel on associera une image numérisée. Il peut aussi s'agir d'un document né numérique qui a été re-numérisé à partir d'une version papier. Cette situation peut intervenir lorsque le support électronique est illisible et qu'il n'existe pas de copies ou encore lorsque le format du document né numérique initial ne permet pas sa conservation à long terme.



Remarque

On peut classer les documents numériques selon d'autres critères que leur mode de création. On pourra distinguer par exemple les documents numériques « statiques », comme une image et les documents numériques « dynamiques » dont la visualisation à l'écran fait appel à plusieurs programmes ou sources d'informations. C'est typiquement le cas d'une page Web dynamique, par exemple, ou d'un fichier de tableur type « Excel » qui contient des données provenant d'autres fichiers.

Chapitre 6 Le document numérique dans l'ordinateur

Qu'est-ce qui se passe lorsque l'on crée un document dans l'ordinateur, à l'aide d'un traitement de texte par exemple? Où s'enregistre-t-il et comment le retrouve-t-on ?

6.1 - Le fichier

Dans le cas le plus simple, le document que vous créez dans votre ordinateur à l'aide d'un logiciel ou d'une application est appelé « fichier ». Il peut s'agir d'un fichier texte ou d'un autre type de fichier (audio, vidéo,...).

Un fichier peut être enregistré sur n'importe quel support de stockage.

On reconnaît le type de fichier par son extension. L'extension est ce qui suit le point « . ».

- **Exemples**

- nom_du_fichier.doc pour un fichier de type Microsoft Word (jusqu'à Word 2003),
- nom_du_fichier.txt pour un fichier créé avec un éditeur de texte (Bloc note, textedit...)
- nom_du_fichier.odt pour un fichier créé avec OpenOffice Writer, l'équivalent de Microsoft Word dans le monde des logiciels libres,
- nom_du_fichier.docx pour un fichier créé avec Microsoft Word 2007
- nom_du_fichier.aiff pour un fichier audio créé sur un ordinateur Macintosh (Audio Interchange File Format)
- nom_du_fichier.wav pour un fichier audio créé sur un PC,
- Etc.

L'extension permet de connaître le type de fichier. Le fichier est créé par une application, c'est-à-dire par un programme qui crée et qui interprète tel ou tel type de fichier.

Un fichier « .doc » aura le plus souvent été créé par l'application Microsoft Word, mais d'autres applications comme OpenOffice sont capable de créer et de lire des fichiers avec l'extension « .doc »



Complément

Un fichier peut avoir une structure interne très simple ou très complexe :

- Un fichier de type « .txt » n'est constitué que d'une suite de caractères comme nous l'avons vu dans la section qui aborde les notions de base sur le document numérique,
- Les fichiers de type « .docx », « .odt », « .wav »...ont des structure internes plus complexes. Ce sont en fait des fichiers « conteneurs », c'est-à-dire des boîtes à l'intérieur desquelles on peut trouver plusieurs fichiers et répertoires.

6.2. Le répertoire

Un répertoire ou « dossier » est l'objet informatique qui peut contenir des fichiers et d'autres répertoires. Il est lui aussi stocké sur un support (disque dur, DVD, etc.).

Un répertoire peut contenir des fichiers mais aussi d'autres répertoires: on utilise alors le terme de sous-répertoire; la plus grande entité est celle qui ne se trouve pas dans un répertoire, mais qui contient tous les répertoires: on la nomme la racine; elle est notée «.» dans la plupart des systèmes d'exploitation. Enfin, on appelle chemin d'accès ou adresse la position d'un fichier ou d'un répertoire dans l'ordinateur. On retrouve cette information dans la barre d'adresse de l'outil explorateur du système d'exploitation, qui fonctionne de la même façon qu'un navigateur web. Dans Windows, on accède à cet outil à partir du menu Démarrer, option Programmes, option Accessoires. On reconnaît l'Explorateur de Windows à son icône montrant une loupe sur un dossier.

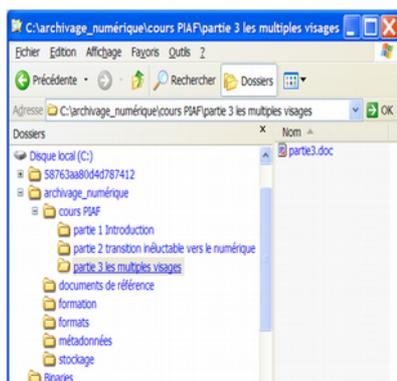
Par exemple :

Dans le système d'exploitation Windows, « C: » est la racine du lecteur C (le disque dur) qui contient des répertoires, des sous-répertoires et des fichiers. La structure suivante illustre la position d'un document dans l'ordinateur, ou son chemin d'accès.

- Support de stockage : disque dur C:
- Répertoire (ou dossier) : « archivage_numérique »
- Sous-répertoire (ou dossier) : « cours PIAF »
- Sous-sous-répertoire (ou dossier) : « partie 3 les multiples visages »
- Fichier : « partie3.doc »

Le chemin d'accès au fichier est :

C: \archivage_numérique.cours PIAF.partie 3 les multiples visages.partie3.doc



Représentation, avec internet Explorer, de l'arborescence de répertoires décrite plus haut

Image 11 Une arborescence

6.3 Système d'exploitation

Le système d'exploitation est le logiciel maître de votre ordinateur, il :

- fait fonctionner les programmes,
- permet de gérer les fichiers et les répertoires en établissant des liens entre le matériel, l'utilisateur et les applications,
- sert à effectuer les commandes de base sur les informations entrées dans l'ordinateur qui y sont stockées ou qui doivent en sortir,
- assure aussi de multiples autres tâches comme la gestion de périphériques (par exemple, le pilotage d'une imprimante) et l'initialisation du système.

Le plus connu des systèmes d'exploitation est sans contredit Windows (avec de multiples versions, les dernières étant Windows 2000, Windows XP, Vista), mais il en existe bien d'autres, comme le MacOS qui est réservé aux Macs, et Unix dans ses différentes variantes (Linux, Solaris, HP-UX ...).

Quand on parle de systèmes d'exploitation, on parle souvent de systèmes multi-tâches ou encore de systèmes multiprocesseurs.

Un système multi-tâches est un système qui permet de partager le temps du processeur pour plusieurs programmes.

Les programmes semblent alors s'exécuter en simultané : on peut dès lors travailler avec plusieurs applications à la fois (un traitement de texte, un tableur et un logiciel de création de pages Web, par exemple).

Un système multiprocesseur est un système multi-tâches qui permet d'organiser l'exécution simultanée de plusieurs applications sur les différents processeurs du système.

Ceux-ci peuvent être architecturés autour d'un processeur central ou être indépendants les uns des autres et posséder leur propre système d'exploitation. Ils communiqueront entre eux par l'intermédiaire de protocoles.

Chapitre 7 Le document numérique voyage

Le document  créé dans un ordinateur n'a pas à y rester confiné ! En fait, on peut l'envoyer à quelqu'un ou le transporter avec soi sans problème.

Les opérations de transfert de documents d'un utilisateur à un autre, d'un ordinateur à un autre, sont rendues possibles par le moyen des réseaux et des protocoles de communication.

7.1 - Les réseaux

Le document numérique peut voyager d'un ordinateur à un autre. Comment ? Par le biais de supports de stockage (CD, clé USB, DVD), bien sûr, mais aussi par le moyen de réseaux.

Les réseaux numériques permettent d'échanger de l'information de différentes façons : par le transfert de fichiers, le courrier électronique, l'échange d'informations personnelles en temps réel (le « chat »), etc.

Les réseaux s'appuient le plus souvent sur des infrastructures de communication existantes comme la téléphonie (avec ou sans fil). Les réseaux peuvent être publics : ils sont alors mis en place et administrés par l'État pour ses propres besoins : administration, éducation, recherche. Ils peuvent être privés et propres à une entreprise.

7.2 - Les protocoles

Pour se comprendre, les ordinateurs, comme les humains, doivent parler le même langage.

Un protocole est l'ensemble des spécifications décrivant les conventions et les règles à suivre pour permettre la communication entre deux machines.

Il existe plusieurs protocoles mais le plus connu est probablement le TCP/IP

- TCP (Transmission Control Protocol) pour la « langue »,
- IP (Internet Protocol) pour la méthode d'acheminement des données.
- TCP/IP est le protocole utilisé par tout ordinateur qui se sert d'Internet pour parler à un autre.



Complément

Il existe en fait une série de protocoles spécialisés, conformes à TCP/IP. Citons les plus importants :

- FTP (File Transfert Protocol) destiné au transfert des fichiers
- HTTP (Hypertext Transfer Protocol) est lié à la navigation dans Internet
- TELNET permet à l'utilisateur d'accéder à un ordinateur hôte distant et d'y utiliser les ressources disponibles
- SMTP (Simple Mail Transfer Protocol) est le protocole utilisé pour le courrier électronique.
- SFTP (Secure File Transfer Protocol) et HTTPS (Hypertext Transfer Protocol Secure) sont utilisés pour les transferts sécurisés : documents confidentiels, transactions bancaires...).

Chapitre 8 La compression des données

Il est possible de compresser un fichier pour en réduire la taille.

Pourquoi réduire la taille d'un fichier? Deux raisons principales :

- faciliter son transfert sur un réseau. Les réseaux ont des capacités de transfert limitées (on exprimera souvent cette capacité en Mégabits par seconde). Plus le fichier aura une taille réduite, plus le temps de transfert sera court. Par ailleurs, de nombreux organismes imposent une limitation dans la taille des fichiers qui peuvent être attachés à un courriel sous forme de pièce jointe. Dans ce cas également, il peut être intéressant de compresser le fichier,
- le stocker sur un support dont la capacité de stockage est limitée. Ainsi, on peut souhaiter par exemple réduire un fichier dont la taille excède la capacité d'une clé USB. On peut vouloir économiser de la place disponible sur son disque dur. Le système d'exploitation Windows propose par exemple de compresser les fichiers qui n'ont pas été utilisés depuis longtemps.



Attention

La compression à des fins de conservation à long terme présente un certain nombre de risques qu'il conviendra d'analyser.

8.1 - La procédure de compression

Il existe de nombreuses techniques de compression, plus ou moins efficaces et plus ou moins coûteuses à mettre en œuvre. Le logiciel de compression va réécrire l'information de façon plus concise, il pourra par exemple remplacer les données répétitives du fichier par un code.



Exemple

Voici un cas tout simple :

Une image comportant plusieurs rangées de pixels de la même couleur rouge (250 pixels, par exemple).

Au lieu d'enregistrer chacun de ces pixels sous la forme d'un numéro pour représenter cette couleur (le numéro 110, par exemple), le logiciel de compression émet un code qui veut dire : mettre ici 250 fois la couleur n° 110.

Un code est beaucoup plus compact que 250 codes.

On peut ainsi arriver à réduire de moitié la taille des fichiers, parfois plus.

Plus généralement, les techniques de compression font appel à des méthodes mathématiques qui sortent du cadre de ce cours.

8.2 - Compression avec pertes ou sans perte

La compression sans perte permet de reconstituer le fichier d'origine avec une exactitude absolue. C'est en général ce que l'on emploie pour la plupart de fichiers.

La compression avec pertes est généralement utilisée pour des images, des fichiers sons ou vidéos. On peut compresser un document de manière à ne perdre quelques détails, souvent indécélables par un humain et ainsi réduire parfois de façon considérable la taille du fichier. Bien sûr, l'intégrité du document original n'est plus totalement respectée.

On pourra donc archiver des données en respectant totalement l'intégrité d'un document et utiliser une compression avec perte pour sa diffusion, dès lors que le document diffusé répond aux besoins des utilisateurs.



Exemple

Les images au format PNG font l'objet d'une compression sans perte, par contre, les images au format JPEG, font le plus souvent, l'objet d'une compression avec perte.

Il existe de nombreux formats de fichiers de compression. On les reconnaîtra en général avec le suffixe du nom de fichier : 7z, zip, ace, rar,...

8.3 - La décompression

La décompression se fait généralement très facilement.



Exemple

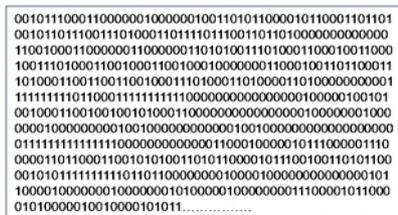
Avec le logiciel Win Zip, par exemple, il suffit d'ouvrir le fichier en cliquant dessus, puis d'extraire le fichier, avec la fonction « Extract ». Il est alors possible de choisir l'emplacement du fichier à extraire, c'est-à-dire l'endroit où l'on souhaite le placer.

Certains logiciels, tel WinRAR, permettent de décompresser des fichiers au format rar mais aussi zip, ace

Chapitre 9 Les principaux problèmes à résoudre pour l'archivage numérique

A la fin de cette partie sur les multiples visages du document numérique, nous pouvons identifier les principaux problèmes que nous aurons à résoudre pour l'archivage long terme des documents numériques. Ces problèmes seront repris en détail dans la suite de ce cours.

9.1 Le stockage et la préservation des bits



Le document numérique se présente toujours sous la forme d'un ou d'un ensemble de séquences de bits.

Image 12 un document numérique

Le premier problème à résoudre sera d'assurer au cours du temps la préservation et l'intégrité de ces séquences de bits. Ce point sera analysé dans la partie VI consacrée au stockage.

9.2 Extraire l'information sous forme intelligible

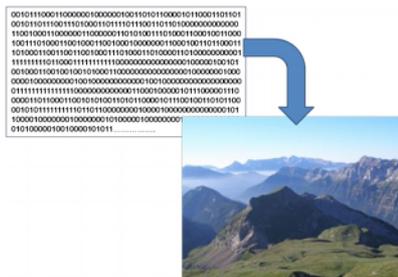
Comment passer du document sous sa forme numérique à une information intelligible que nous saurons lire, comprendre, interpréter ?

Nous pourrions penser qu'il suffit de connaître le format du fichier ? La question est beaucoup plus complexe que l'on peut l'imaginer :

- Le format est-il pérenne ou va-t-il disparaître des environnements techniques sur le court ou moyen terme ? Il existe des milliers de formats de données. Certains disparaissent, d'autres sont inventés. Les formats propriétaires non publiés vous ligotent aux logiciels des mêmes propriétaires,
- Disposerons-nous toujours, dans 10 ou 20 ans, de logiciels permettant de lire les données dans ce format ?
- La spécification du format dont nous disposons est-elle complète, précise, exacte ? Qui en assure la conservation à long terme ?
- Nos données sont-elles réellement conformes au format supposé ? Est-ce qu'on a pu le vérifier ?

En outre, la question du format ne résoudra qu'une partie du problème. Pour des informations scientifiques, économiques, statistiques complexes, nous aurons besoin d'informations complémentaires pour accéder réellement à la sémantique de l'information.

Telles sont les questions abordées dans la partie 7 consacrée aux formats.



Une question centrale, passer des bits à un document intelligible.

Image 13 Des bits au contenu

9.3 – Retrouver notre document au milieu de millions d'autres

Comment évaluer l'adéquation d'un document à un usage déterminé ? Ce sera l'objet des métadonnées descriptives qui permettront à l'utilisateur futur de découvrir les documents correspondants à ses besoins.

Les métadonnées joueront également un rôle essentiel pour aider l'utilisateur à comprendre et à interpréter correctement les documents extraits de des archives : pouvoir identifier les documents de façon unique et surtout pérenne est essentiel, savoir pourquoi et dans quelles conditions ce document a été produit, quels sont les relations qui existent entre un document donné et d'autres documents archivés, etc.

Ces questions seront analysées dans la partie 9 consacrée aux métadonnées.



Image 14 Quelle valeur probante pour le document numérique ?

Est-ce que nous pouvons faire confiance dans notre document numérique ? Quels éléments avons-nous pour démontrer son intégrité et son authenticité ?

Quelle pourra être la valeur de ce document devant un tribunal, dans le cadre d'un contentieux ?

Tous ces aspects juridiques seront examinés dans la section 10

9.5 Conclusions

La résolution de ces quatre problèmes essentiels : stockage, formats, métadonnées, valeur probante ne pose pas que des questions techniques, loin de là.

Nous verrons qu'il subsiste de nombreux besoins en matière normative et en matière juridique. Nous pouvons ajouter que les aspects organisationnels, les stratégies vont également jouer un rôle essentiel. Enfin les questions de coût pèseront lourd dans toutes les décisions.

Nous sommes prêt maintenant pour aborder tous ces sujets plus en profondeur.

Questions générales sur le numérique : vrai ou faux

Objectifs

Avez-vous compris tout ce qui vient de vous être enseigné ?

Si vous voulez le vérifier, faites les exercices proposés ci-dessous.

Si vous ne savez pas répondre, ne regardez pas trop vite le corrigé, travaillez à nouveau la (les) section(s) précédente(s) où vous découvrirez les solutions.

Bien sûr, si vous n'y arrivez vraiment pas, vous pouvez consulter les réponses. Ne les lisez pas avec précipitation mais avec une grande attention et surtout essayez de comprendre.

Remarque sur la limite de nos exercices

Nous vous proposons des exercices sur le numérique avec toutes les réserves que cela comporte en raison notamment des évolutions technologiques. Nous sommes en effet dans une discipline récente qui est en constante évolution. Les principes qui ont été posés dans ce module resteront sûrement durablement valides, par contre le nombre de solutions aujourd'hui valides ne le seront plus que partiellement demain ou plus du tout.

Par exemple : dans un contexte organisationnel donné, ce que nous pouvons recommander en matière de formats ou de moyens de stockage, pourra être remis en cause dans un ou deux ans.

Il est donc nécessaire que les partenaires d'un projet d'archivage numérique s'approprient les principes mais ne procèdent à l'analyse de la situation et des contraintes qu'au moment où le projet se met en marche et cela avec les partenaires concernés ; en effet, lorsqu'un archiviste peut se trouver seul au sein de son organisation, l'archivage numérique ne doit pas être une activité individuelle : ce sera toujours la mise en commun d'un ensemble de compétences complémentaires. Donc rien ne remplacera les exercices en vraie grandeur.

Nous rappelons que les exercices du PIAF sont à l'usage d'une auto-formation. Nous proposons à cet effet des types d'exercices de mémorisation, d'accompagnement, de positionnement afin de permettre à l'utilisateur de vérifier l'acquisition d'une culture minimale précise.

Les exercices ci-dessous porteront surtout sur les principes puisqu'il n'est pas possible d'aller trop loin en matière de solution.

Exercice 1

[Solution n°1 p 21]

Un bit permet de représenter 4 états différents : 0, 1, 01, 10

Vrai

Faux

Exercice 2

[Solution n°2 p 21]

Il faut 8 bits pour faire un octet

Vrai

Faux

Exercice 3

[Solution n°3 p 21]

Questions générales sur le numérique : vrai ou faux

Vrai

Faux

Exercice 4

[Solution n°4 p 21]

Un répertoire, ou dossier, peut contenir des fichiers mais aussi d'autres dossiers

Vrai

Faux

Exercice 5

[Solution n°5 p 21]

La compression permet de réduire la taille des fichiers

Vrai

Faux

Exercice 6

[Solution n°6 p 22]

L'échange de données numériques peut se faire à l'aide de supports amovibles (clé USB, CD, DVD...) ou par l'intermédiaire d'un réseau.

Vrai

Faux

Exercice 7

[Solution n°7 p 22]

Identifier les quatre problèmes principaux à résoudre pour l'archivage des documents numériques :

- | | |
|--------------------------|--|
| <input type="checkbox"/> | • Stocker les données et assurer la préservation des bits sur des supports adéquats |
| <input type="checkbox"/> | • Disposer des informations |
| <input type="checkbox"/> | • Avoir choisi les bons formats de représentation de l'information |
| <input type="checkbox"/> | • Pouvoir passer des fichiers stockés sous forme numérique à une information intelligible |
| <input type="checkbox"/> | • Choisir des supports stables et durables adaptés à la volumétrie |
| <input type="checkbox"/> | • Assurer la sécurité de l'infrastructure informatique utilisée |
| <input type="checkbox"/> | • Avoir choisi le format de métadonnées le plus pertinent |
| <input type="checkbox"/> | • Décrire les documents archivés avec des métadonnées suffisamment précises et pertinentes pour pouvoir retrouver ces documents au milieu d'une masse considérable de documents archivés |
| <input type="checkbox"/> | • Garantir l'intégrité des documents à l'aide d'une empreinte |
| <input type="checkbox"/> | • Disposer d'éléments apportant la preuve de l'intégrité et l'authenticité du document |

Exercice : Réviser vos connaissances

[Solution n°8 p 22]

Exercice

Je suis un support de stockage dont la capacité varie de 4,7 à 17 Go. Qui suis-je?

- Blu Ray
 - Disque dur
 - CD-Rom
 - DVD
 - Disquette
-

Exercice

La compression avec perte des fichiers est utile dans un contexte de conservation

- Vrai
 - Faux
-

Exercice

Quel est le principal problème à résoudre pour l'archivage numérique?

- La préservation et l'intégrité de la séquence de bits
 - L'extraction de l'information sous forme intelligible
 - Le repérage de documents pertinents aux besoins de l'utilisateur
 - Disposer d'éléments probants sur l'intégrité et l'authenticité du document
 - Tous ces problèmes sont à résoudre (1,2,3,4)
-

Solution des exercices

> Solution n°1 (exercice p. 17)

Vrai

Faux

Un bit permet de représenter uniquement deux états 0 ou 1

> Solution n°2 (exercice p. 17)

Vrai

Faux

Un octet est composé de 8 bits, par ex « 01111001 »

> Solution n°3 (exercice p. 17)

Vrai

Faux

Les 16 états sont 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E

> Solution n°4 (exercice p. 18)

Vrai

Faux

Vous pouvez constater, en parcourant les répertoires du disque local C : que les répertoires qu'il contient (par exemple « Program Files » ou « Documents and settings », contiennent d'autres répertoires qui contiennent eux même des répertoires et des fichiers, et ainsi de suite

> Solution n°5 (exercice p. 18)

Vrai

Faux

La compression est une technique de réorganisation des séquences de 0 et de 1 permettant de stocker la même quantité d'information en réduisant la longueur de la séquence binaire, et par conséquent, en diminuant la taille des fichiers.

> **Solution n°6** (exercice p. 18)

<input checked="" type="radio"/>	Vrai
<input type="radio"/>	Faux

L'ordinateur peut recopier des fichiers de données sur un support amovible. Ce support pourra ensuite être lu par un autre ordinateur disposant d'un équipement de lecture correspondant. Ces fichiers pourront être transmis entre ordinateurs par l'intermédiaire d'un réseau : en pièce jointe à un message électronique, ou directement en utilisant un protocole de communication spécialisé pour le transfert de fichiers (par exemple FTP, File Transfer protocol).

> **Solution n°7** (exercice p. 19)

<input checked="" type="checkbox"/>	• Stocker les données et assurer la préservation des bits sur des supports adéquats
<input type="checkbox"/>	• Disposer des informations
<input type="checkbox"/>	• Avoir choisi les bons formats de représentation de l'information <i>faux, ceci n'est qu'une partie du point 2, il ne suffit pas de connaître le format des données pour que l'information contenue soit compréhensible</i>
<input checked="" type="checkbox"/>	• Pouvoir passer des fichiers stockés sous forme numérique à une information intelligible <i>faux, ceci n'est qu'une partie du point 1</i>
<input type="checkbox"/>	• Choisir des supports stables et durables adaptés à la volumétrie
<input type="checkbox"/>	• Assurer la sécurité de l'infrastructure informatique utilisée <i>faux, ceci n'est qu'une conséquence indirecte des quatre problèmes identifiés</i>
<input type="checkbox"/>	• Avoir choisi le format de métadonnées le plus pertinent <i>faux, ceci n'est qu'une partie du point 3</i>
<input checked="" type="checkbox"/>	• Décrire les documents archivés avec des métadonnées suffisamment précises et pertinentes pour pouvoir retrouver ces documents au milieu d'une masse considérable de documents archivés
<input type="checkbox"/>	• Garantir l'intégrité des documents à l'aide d'une empreinte <i>faux, ceci n'est qu'une réponse partielle au point 4</i>
<input checked="" type="checkbox"/>	• Disposer d'éléments apportant la preuve de l'intégrité et l'authenticité du document

> **Solution n°8** (exercice p. 20)

Exercice

Solution des exercices

- Blu Ray
 - Disque dur
 - CD-Rom
 - DVD
 - Disquette
-

Exercice

- Vrai
 - Faux
Il y a une perte de l'intégrité du document; la compression avec perte est utile pour la diffusion plutôt
-

Exercice

- La préservation et l'intégrité de la séquence de bits
 - L'extraction de l'information sous forme intelligible
 - Le repérage de documents pertinents aux besoins de l'utilisateur
 - Disposer d'éléments probants sur l'intégrité et l'authenticité du document
 - Tous ces problèmes sont à résoudre (1,2,3,4)
-

Glossaire

Analogie

La mémorisation peut se faire par une analogie au phénomène. On parle de données analogiques :

- le gravage d'un disque vinyle reproduit la vibration du son dans la galette plastique. Il en est de même pour l'enregistrement du son sur un cylindre phonographique,
- un thermomètre enregistreur va tracer les variations de la température au cours du temps sur un rouleau de papier.

Les données analogiques présentent un caractère continu.

Authenticité

un document authentique est un document dont on peut prouver

- a) qu'il est bien ce qu'il prétend être,
- b) qu'il a été effectivement produit ou reçu par la personne qui prétend l'avoir produit ou reçu, et
- c) qu'il a été produit ou reçu au moment où il prétend l'avoir été.

(ISO 15489 « records management »)

bit

Le bit (pour "binary digit") est la plus petite unité d'information manipulable par un ordinateur. Un bit vient préciser si le courant passe ou non. Ainsi, il ne peut avoir que la valeur de 0 ou de 1.

Document

Tout écrit ou enregistrement considéré comme une unité documentaire (ISO 15489-1 "Records Management").

Le document est donc un « ensemble d'informations enregistrées », considéré comme une unité qui pourra être utilisable à des fins de consultation ou comme preuve.

Dans le domaine du numérique, l'association entre le support et l'information qu'il porte perd son sens. Le document pourra au cours du temps être stocké sur des supports différents sans pour autant que cela ne change sa nature ni n'altère son intégrité.

Document numérique

Tout écrit ou enregistrement considéré comme une unité documentaire (ISO 15489-1 "Records Management").

Le document est donc un « ensemble d'informations enregistrées », considéré comme une unité qui pourra être utilisable à des fins de consultation ou comme preuve.

Dans le domaine du numérique, l'association entre le support et l'information qu'il porte perd son sens. Le document pourra au cours du temps être stocké sur des supports différents sans pour autant que cela ne change sa nature ni n'altère son intégrité.

Donnée

Représentation formalisée de l'information, adaptée à la communication, l'interprétation ou le traitement. Par exemple : une séquence de bits, un tableau de nombres, les caractères d'une page, un enregistrement audio, etc. Une donnée est donc un conteneur porteur d'une information ou d'un fragment d'information (glossaire du modèle de référence OAIS).

Données

Terme utilisé, en particulier en informatique, pour désigner une information.

Information

L'information est définie comme une connaissance pouvant être échangée. En pratique, l'information est donc un élément de connaissance susceptible d'être codé pour être conservé, traité ou communiqué.

Intégrité

L'intégrité d'un document renvoie au caractère complet et non altéré de son état, (ISO 15489 « records management »). Le document n'a subi aucune modification non tracée.

Métadonnées

Étymologiquement, « méta » provient du grec signifiant « après, au-delà de, avec » : « méta » données signifie « au-delà des données », « qui dépasse les données », « qui englobe les données ». Les métadonnées sont d

données sur les données, à propos des données, qui définissent, décrivent des données, leur contexte, leur contenu, leur structure des ainsi que leur gestion dans le temps.

Numérique

L'information est transformée par un processus qui la rend numérique. On parle de données numériques :

- la photo est scannée (ou numérisée par l'appareil photo) pour mise en mémoire,
- le son est échantillonné et stocké dans un format propre à la musique (ex : MP3),
- l'enregistrement des variations de la température se présentera alors comme une suite de couples de valeurs (temps1, température1), (temps2, température2), etc.

Dans ce cas, le signal ne peut prendre que des valeurs bien définies, en nombre limité, on parle alors de signal numérique. Un signal numérique se traduit ainsi par des « 0 » et des « 1 » ; c'est donc un nombre binaire ou une collection de nombres binaires.

Les données numériques se présentent toujours comme une suite ou un ensemble de valeurs. Elles ont donc un caractère discontinu, discret.

Bibliographie

[Premier ouvrage de synthèse sur l'archivage numérique en langue française.] • BANAT-BERGER F., HUC C., DUPLOUY L., L'Archivage numérique à long terme, les débuts de la maturité? Paris, La Documentation française, 2009.

Webographie

[Norme de référence essentielle pour comprendre le problème posé par l'archivage numérique]
[http://public.ccsds.org/publications/archive/650x0b1\(F\).pdf](http://public.ccsds.org/publications/archive/650x0b1(F).pdf)