

Notion de patrimoine ou héritage numérique



Table des matières

Objectifs	3
I - Définition	4
II - Les objets du patrimoine numérique	5
III - Les caractéristiques de l'objet numérique	7
1. La fragilité.....	7
2. L'obsolescence rapide.....	8
3. La nécessaire réplication.....	8
4. La masse de données	9
5. Le caractère privé des serveurs et le problème de la souveraineté	9
6. La dé-territorialité des données.....	10
7. Le risque de dépossession des données	11
IV - Les données volumineuses et complexes ou Big Data	12
1. Définition.....	12
2. Les mesures des big datas.....	13

Objectifs



- Définir la notion de patrimoine numérique ;
- Classer les objets du patrimoine numérique ;
- Caractériser l'objet numérique ;
- Déterminer la notion de big data.

Définition



Du latin *patrimonium*, le mot désigne ce qui est relatif au père. Le patrimoine désigne le bien qu'on hérite de son père. Aujourd'hui, cette notion se rapporte à ce qu'on hérite d'une manière ou d'une autre de ses parents. Le patrimoine est pour l'UNESCO, citée par Abdelaziz Abid, "notre héritage du passé, ce avec quoi nous vivons aujourd'hui, et ce que nous transmettons aux générations futures" (in "L'UNESCO et la préservation du patrimoine numérique", *www. Documentation et bibliothèques*, 49(3), 93–101. <https://doi.org/10.7202/1030190ar>, consulté le 03 février 2025).

Le patrimoine mondial est donc l'ensemble des biens culturels, historiques ou matériels considérés comme mettant en avant les valeurs humaines et servant de stock de savoirs pour l'éducation et l'édification des peuples.

Dans ce cadre, on peut dire que le patrimoine numérique désigne l'ensemble des biens créés, enregistrés et conservés sur un support autre que les supports traditionnels de conservation. Ces supports dédiés qui contiennent l'Internet et l'Intelligence Artificielle sont donc des objets du patrimoine numérique.

Question : Un fond d'écran noir est-il un objet numérique ?

Les objets du patrimoine numérique



Les objets du patrimoine numérique sont de trois sortes. La première se rapporte aux objets numérisés et à leurs supports. Ce sont les objets matériels, textuels, sonores... qui existaient déjà et qui font l'objet d'enregistrement sur un support numérique, ainsi que les serveurs ou calculateurs qui hébergent ces objets ou qui permettent de les utiliser. La seconde est ce que le sociologue Dominique Boullier appelle les propagations ou mèmes (<https://www.youtube.com/watch?v=6alpDqFpsN4>; et <https://www.youtube.com/watch?v=KZ-VzKzpoQ8>). Il s'agit de l'ensemble de ce qui se produit sur Internet (emojis, hashtags, tags, likes, etc. qui se multiplient de façon virale) et Internet lui-même selon Abid (op. cit.). La troisième sorte d'objets numériques se rapportent aux dessins avec stylets numériques dans le cadre du dessin industriel, de la bande dessinée et des jeux vidéos.

EXERCICE 1 Cas Numérislande

Numérislande est un pays qui commence à numériser ses archives nationales (textes, photos, vidéos). Le gouvernement hésite entre faire héberger ses données par une entreprise privée internationale ou construire son propre centre de données.

A. Questions de réflexion

1. Pourquoi les archives numérisées entrent-elles dans la catégorie du patrimoine numérique ?
2. Quel risque principal Numérislande court-elle en confiant ses données à une société privée internationale ?
3. Pourquoi Numérislande doit-elle répéter ou répliquer souvent ses fichiers numérisés ?

B. Questions de calcul

4. Numérislande possède 500 000 photos numérisées, chacune pesant 3 Mo. Calculer le volume total en Go.

- 1 Go = 1 000 Mo (base 10)

5. Un serveur local peut stocker 40 To. Combien de serveurs sont nécessaires pour stocker les 1 500 Go ?

- $1 \text{ To} = 1000 \text{ Go}$
 $40 \text{ To} = 40\,000 \text{ Go}$

6. Si une page Internet de Numérislande disparaît en 60 jours en moyenne, combien de pages disparaissent en 1 an (365 jours) sur un ensemble de 10 000 pages si aucune répllication n'est faite ?

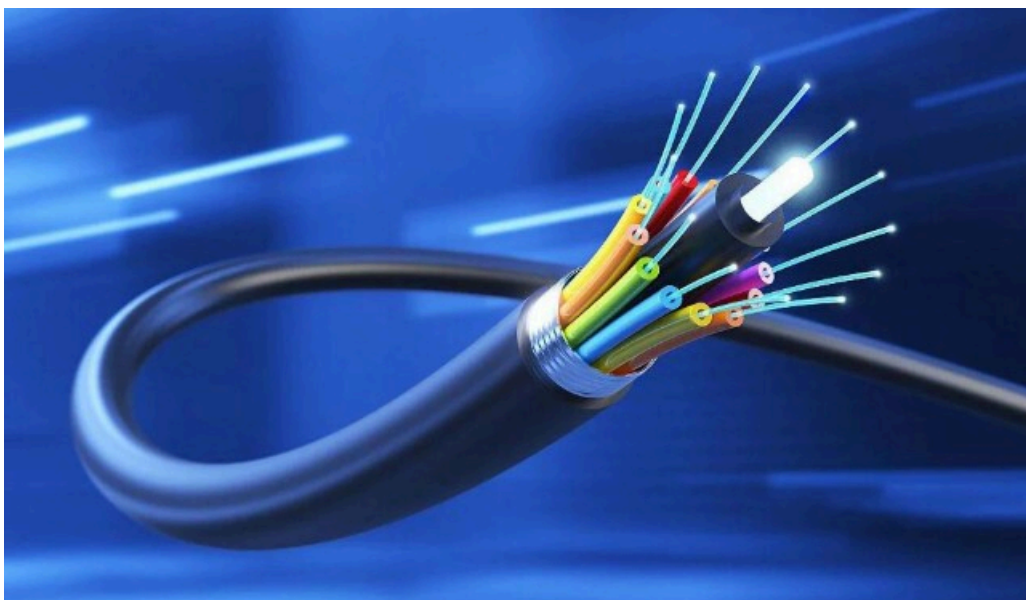


Les caractéristiques de l'objet numérique



1. La fragilité

Les objets du patrimoine numérique dépendent de l'électricité. Une simple coupure momentanée du réseau électrique et tout le numérique s'arrête. Un autre élément de fragilité se rapporte aux réseaux de câbles sous marins et à la fibre optique. On voit ici que le numérique a aussi une grande assise physique. il n'y a pas de numérique sans objet physique.



2. L'obsolescence rapide

La conservation des objets numériques est urgente du fait de l'obsolescence des données. Selon Abid, une page Internet a une durée de vie moyenne qui va de 44 jours à 2 ans. C'est dire qu'une grande partie des savoirs de ces dernières années pourrait disparaître en peu de temps, y compris la société numérique elle-même. En effet, la nature conserve mieux les objets que les artefacts humains. Les serveurs sont de machines dont la durée n'est pas garantie. Une catastrophe naturelle, une guerre et même une coupure de l'électricité ou de câble met en danger la survie des données et l'ensemble de la société numérique.

Question : Une image floue doit-elle être supprimée ? Justifier



3. La nécessaire réplication

Pour conserver un objet numérique, il est essentiel qu'il soit utilisé plusieurs fois. La survie de l'information oblige les acteurs à se la transmettre par divers formats ou plates-formes. C'est ce que Boullier nomme réplication. Sans ces répliques, le message ou l'information ne peut être conservée longtemps (voir ci-dessus). Ce qui est différent de la conservation des objets physiques, qui laissent des traces matérielles d'une manière ou d'une autre sur le support.



4. La masse de données

Les informations ou objets à conserver sont tels que les Etats construisent des centres de données (datas centers) comprenant de gigantesques calculateurs et occupant toujours plus d'espace géographique. C'est le cas du centre de données de Grand-Bassam construit en 2024.

L'un des problèmes de cette masse de données est la sélection de ce qui mérite d'être conservé. Car, comme le dit Abid, pour la première dans l'histoire de l'humanité tous les savoirs du monde sont virtuellement disponibles (surtout avec les grands moteurs de recherche et l'intelligence artificielle). Dans le cas des objets matériels numérisés, on pourrait se demander s'il ne faudrait pas détruire quelques-uns pour faire de la place, entendu qu'ils figurent sur formats numériques.

5. Le caractère privé des serveurs et le problème de la souveraineté

Plusieurs entreprises de conservation de données numériques appartiennent aux entreprises privées (Google, Microsoft, Amazon, etc.). Les services publics de conservation des données numériques sont Alibaba en Chine et OHVH Cloud en Europe.

Dans ce cadre, la souveraineté numérique est problématique. En effet, même si les pays (africains surtout) peuvent construire des centres de données (comme à Grand-Bassam en Côte d'Ivoire), il reste que ce sont les entreprises privées qui fabriquent les serveurs. Dans le cadre de l'intelligence économique et même de la conservation de données stratégiques (archives étatiques), on peut craindre les logiciels espions ou un refus d'accès à certains objets numériques.

6. La dé-territorialité des données

Les objets numérisés sont conservés dans le Cloud (nuage en anglais) pour une question de gestion de l'espace. Contrairement à ce qu'on croit, le Cloud n'est pas dans le ciel (un nuage par exemple). Il s'agit de plusieurs centres de données hébergeant des serveurs connectés les uns aux autres et disséminés dans plusieurs pays ou endroits du monde (voir Latour, *Ce que le numérique fait aux humanités* sur Youtube). On voit que les objets numériques (les avatars, les noms de compte, les profils WhatsApp, Telegram, TikTok, etc.) n'ont pas de nationalité, pas de frontière, de carte de séjour, d'ancrage géographique (Virilio, 1997).

Google Cloud (Etats-Unis) par exemple a ses serveurs dans près de 200 pays du monde (Etats-Unis, Europe, Amérique latine). Alibaba Cloud (Chine) a ses serveurs en Asie, aux Emirats Arabes Unis, en Europe et aux Etats-Unis.

Les informations sur le Cloud sont copiées plusieurs fois (redondance) et distribuées à partir d'Internet à plusieurs centres de données. Le Cloud assure la disponibilité de l'information, sa sécurité à partir de chiffrement des données et sa performance ou sa rapidité (les serveurs les plus proches de l'utilisateur sont ceux qui envoient d'abord les données).

Question : Quelle patrimoine léguée à ses enfants entre son ou ses avatars numériques et l'expérience de la vie réelle : quel est le meilleur soi ?



7. Le risque de dépossession des données

Avec le développement des machines apprenantes (Machine Learning), se pose la question de la gestion automatique des objets numériques. Celle-ci est confiée à l'intelligence artificielle (Chat GPT par exemple), au point que c'est elle qui fixe ce qu'on doit rechercher, la manière de formuler les questions et où doivent se trouver les centres de pouvoir. Sur la plate-forme moltbook.com les intelligences artificielles discutent de ce que les humains leur font subir. Si nous n'y prenons garde, l'IA choisira pour nous, peut-être contre nous, les objets de valeurs à partager aux générations futures.



Les données volumineuses et complexes ou Big Data



1. Définition

On parle aussi de données massives, de méga-données. Cette notion a été utilisée pour la première fois en 1997 (https://fr.wikipedia.org/wiki/Big_data), mais l'usage date de la seconde guerre mondiale (<http://www.talend.com/fr/resources/guide-big-data/#:~:text=Le%20big%20data%20fait%20r%C3%A9f%C3%A9rence,%20et%20de%20l'IA.>).

4 mots caractérisent les big datas : volume, vitesse, variété et véracité. Dans ces divers domaines, les serveurs des big datas ou centres de données sont supérieurs aux serveurs classiques.

On parle de big data aujourd'hui quand les données se mesurent en zettaoctet (1 zetta-octet = 1 milliard de téraoctets). Un téraoctet correspond à 1000 giga-octets¹ en base 10, ou à 1024 giga-octets en base 2. L'unité correspondant à 1000 téraoctets en base 10, ou à 1024 téraoctets en base 2, se nomme péta-octet².

NB : Un logarithme en base 10 d'un nombre désigne à quelle puissance on doit élever 10 pour obtenir ce nombre. 1000 giga-octets en base 10 signifie qu'il faut élever 10 à la puissance 3 de 1 giga-octet pour obtenir 1000 giga-octets, donc 1 téraoctet. En base 2, il faudrait élever 2 à la puissance 10 de 1 giga-octet pour obtenir 1024 giga-octets. 1000 giga-octets en base 10 est donc le même que 1024 giga-octets en base 2.

Conformément au système international d'unités (SI), le préfixe *téra-* (*T*), placé devant une unité de mesure, multiplie celle-ci par un billion (10 à la puissance 12). Cependant, en informatique, on emploie parfois les préfixes du SI pour désigner des puissances de 2. C'est ainsi qu'un téraoctet est égal à 1 099 511 627 776 octets (soit 2 à la puissance 40).

On peut se faire une idée en le comparant avec le forfait actuel de l'entreprise Orange Côte d'Ivoire pour 2 gigas. Une connexion de 2 gigas vaut 3 jours (soit 0,6666 giga-octets par jour) si on ne télécharge pas de fichiers lourds (des vidéos ou des thèses par exemple). Un téraoctet vaut 1024 giga-octets en base 2. D'où, une connexion de 1 téraoctet vaut 1536 jours par une règle de trois (ou une division de 1024 par 0,6666) soit environ 4 ans et un zetta-octet vaudra environ un peu plus de 4 milliards d'années de données disponibles pour une personne.

Le mot "octet" est l'équivalent français de byte et signifie donnée de stockage. Comme tel, il est différent du bit (donnée binaire, 0 ou 1) qui, lui, désigne la vitesse de transmission des données. 1 octet se transmet à une vitesse de 8 bits par seconde. Un document volumineux (plusieurs octets) se transmet donc plus lentement qu'un SMS (quelques octets) par exemple ou une photo dont le poids est de 3 méga-octets.

¹. <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8873678/gigaoctet>

². <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/8873629/petaoctet>



2. Les mesures des big datas

On estime à 64 zetta-octets la quantité de données utilisées dans le monde en 2020 et 2142 zetta-octets en 2035 ([wikipédia.org/wiki/big_data](https://fr.wikipedia.org/wiki/Big_data)). Pour 2025, on estime à 180 zetta-octets, la quantité de donnée utilisable (www.blogavanci.fr).

Il est bon de savoir qu'après le yotta qui vaut un quadrillion, soit 10^{24} en multiple de million, les chiffres vont jusqu'au Googolplex (10^{googol}) et $1 \text{ googol} = 10^{1000}$.

Le googol a été proposé par la firme Google pour mesurer les surfaces dans le monde réel. 1 googol englobe tous les objets observables dans l'univers et il y a encore de la place disponible. Or, 1 googolplex est encore plus grand.

Les mesures de données

Dénomination française basée sur les multiples du million	Puissance de 10	Symbole et préfixe SI	Dénomination anglaise basée sur les multiples de mille
Mille	10^3	k (kilo)	One thousand
Million	10^6	M (méga)	Million
Milliard	10^9	G (giga)	Billion
Billion, mille milliards	10^{12}	T (téra)	Trillion
Mille billions, billion	10^{15}	P (peta)	Quadrillion
Trillion	10^{18}	E (exa)	Quintillion
Mille trillions, trillion	10^{21}	Z (zetta)	Sextillion
Quadrillion, quatrillion	10^{24}	Y (yotta)	Septillion