



Huma-Num
la TGIR des humanités numériques

Comment perdre ses données : quels sont les risques ?

Michel Jacobson
TGIR Huma-Num

Journées "Préserver ses données numériques"
Campus Condorcet - 6-7 février 2020



Aix-Marseille
université

CAMPUS
CONDORCET
Paris-Aubervilliers

Comment bien perdre ses données ?

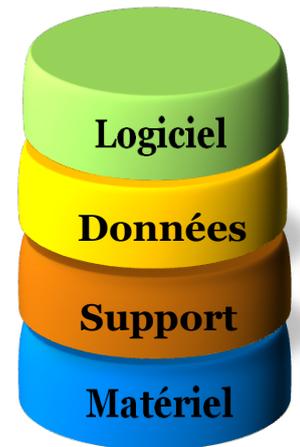
- C'est facile, fréquent, c'est à la portée de tout le monde

Les particularités du numérique

L'accès à la donnée numérique dépend d'un empilement de technologies

L'information est codée

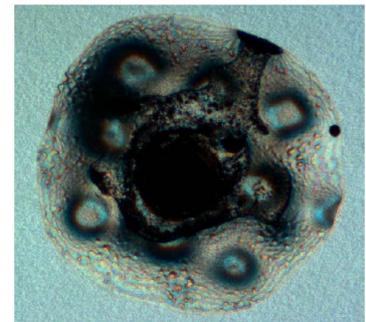
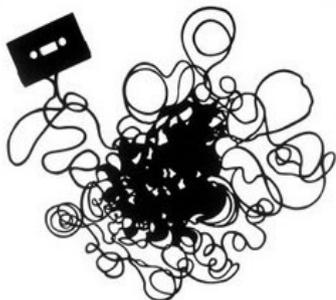
- Les principaux risques sont
 - L'obsolescence
 - Le vieillissement
 - L'oubli
 - La destruction



Les problèmes liés aux supports

Vieillessement des supports

- Perturbations magnétiques
- Interactions chimiques entre les couches d'un CD, oxydation de la partie métallique
- Accidents de manipulation



→ Besoin de locaux, d'équipement et de conditionnement adaptés

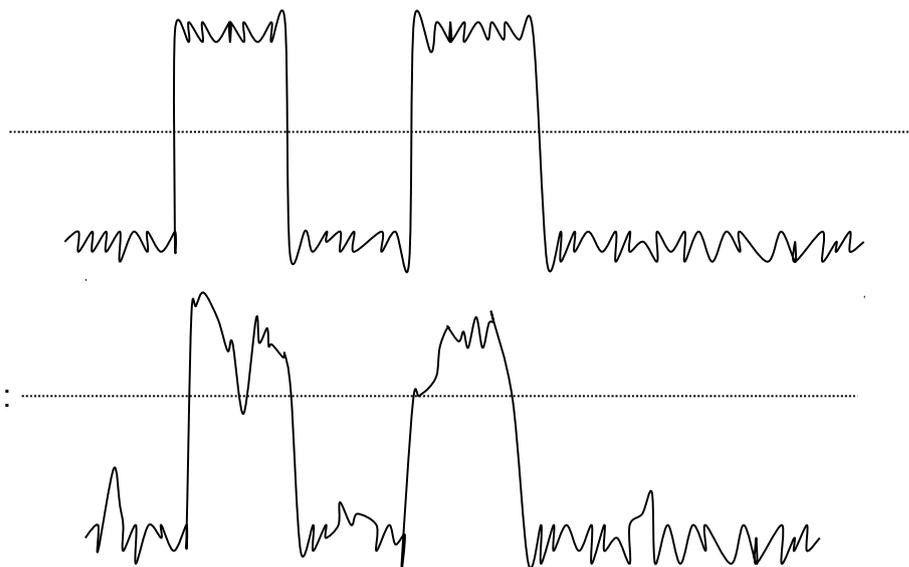
Vieillissement des supports



Les problèmes liés aux supports

Érosion du signal

- information binaire est le résultat de l'interprétation d'un signal qui, sur le support, s'exprime sur un continuum et a tendance à se dégrader, s'affaiblir

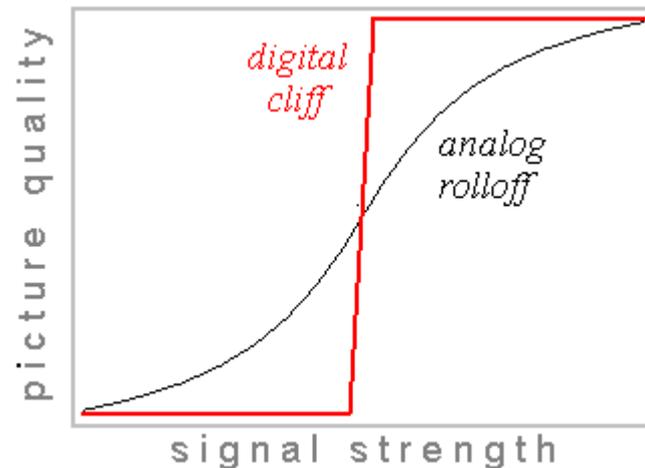


- Besoin d'auditer l'état des supports
- Besoin de faire des opérations périodiques de rafraîchissement des supports

Les problèmes liés aux supports

L'effet de falaise

- Au delà d'un seuil, la dégradation du support qui passe complètement inaperçue se traduit par une perte de brutale de l'information
- Toutes les zones d'un support ne sont pas équivalentes
 - Par exemple la table des matières du support



Les problèmes liés aux supports

Obsolescence

- Je n'ai plus de lecteur du support ou mon lecteur est HS sans plus de maintenance par le fabricant
- Mon lecteur est OK mais le connecteur ne se branche pas sur mon PC
- Mes lecteur + connecteur sont OK mais je n'ai pas de pilote ou mon pilote est incompatible avec mon système d'exp



→ Besoin de veille sur les technologies

Les problèmes liés aux supports

Obsolescence, ça continue

- Le système de fichiers du support n'est pas compréhensible par mon système (NTFS, FAT, HFS...)

Les problèmes liés aux formats

La donnée binaire est opaque. On a besoin de connaître son organisation pour la décoder. Ce décodage se fait par un logiciel

- J'ai un document dwg mais je n'ai pas de licence AutoCAD
- Mon logiciel d'édition ne comprend pas le codec XXX de ma vidéo AVI
- Je n'ai pas la police de caractère utilisé par l'auteur du texte
- La lecture de mon fichier par la nouvelle version du logiciel me donne un autre résultat qu'avant
- Je ne peux plus installer mon unique outil de visualisation sur la nouvelle version de mon système d'exploitation

→ Besoin de décrire finement avec des métadonnées

→ Besoin de formats ouverts

→ Besoin de faire de la veille

```
10111011101111111100111010011100  
11001110...
```

L'oubli

- Je ne sais pas où j'ai mis mon fichier
 - J'avais un fichier, reçu en 2015, c'était le scan d'une page manuscrite, il était sur mon ancien portable dans un dossier « en cours » que j'ai du recopier sur mon nouveau portable dans le dossier « anciens fichiers ». Je ne me rappelle pas du nom que j'ai donné à ce fichier...
- Je ne sais pas quel est le bon fichier
 - J'ai 28 copies du même document sur des supports différents à des stades d'avancement différents mais je ne sais plus quel est le plus complet. Je ne suis pas à l'abri d'avoir fait des modifications sur une version très ancienne qui ne contenait pas des ajouts récents
- J'ai un fichier toto.dat, je ne sais pas ce que c'est
 - Besoin d'organisation

Le codage de l'information

- J'ai retrouvé mon fichier mais le contenu semble altéré
- J'ai retrouvé mon fichier mais je ne comprends plus ce qu'il contient
 - Il contient des mesures mais l'unité n'est pas indiquée
 - Je ne me souviens plus qui est sur la photo ni où elle a été prise

→ Besoin de sécurisation,
de contrôles, de
descriptions fines

La suppression

- J'ai jeté par mégarde mon document
- J'ai reformaté mon disque dur
- Je n'avais pas de copie de mes données
- Mon exemplaire de travail et mon disque de sauvegarde étaient dans la même pièce qui a pris l'eau

→ Besoin de
sécurisation

Les risques

- Risque de fonctionnement
- Risque juridique
- Risque scientifique
- Risque d'image

Les risques

- Risque de fonctionnement
 - Le fichier sur lequel vous travaillez n'est pas le plus récent
 - Vous avez détruit des informations nécessaires à votre travail
 - Vous n'arrivez pas à retrouver les données dont vous avez besoin pour votre cours de demain

Les risques

- Risque juridique
 - Vous avez perdu une facture, la fiche de consentement signée d'un interviewé...
 - Vous n'arrivez pas dans les temps demandés par le fisc à produire des pièces justificatives
 - Vous ne pouvez pas prouver que ce n'est pas l'image que vous aviez déposé

Les risques

- Risque scientifique
 - Vous ne pouvez pas prouver la paternité d'un document qu'un autre chercheur revendique
 - Vous citez des références qui ne sont pas vérifiables

Les risques

- Risque d'image
 - Votre manque d'organisation donne une mauvaise image de la qualité de votre travail ou du sérieux de votre institution

Conclusion

- Couverture des risques
 - Le plan de gestion de données : un outils pour exprimer comment on couvre les risques et plus globalement comment on gère le cycle de vie des données
 - HN : des outils pour la manipulation de données au service des SHS
 - Un circuit de préservation
 - Un cadre juridique, normatif et organisationnel