

Organisation et nommage des fichiers de données de recherche

Rendez-vous de l'info scientifique

0. Organisation des données

Organiser ses fichiers de données est une tâche qui prend du temps et qui peut paraître compliquée, mais qui s'avère utile tant pour le travail individuel que collaboratif. Elle se révèle également payante à long terme, puisque des données bien structurées et bien nommées sont plus aisées à localiser et identifier.

- A. Investir du temps dès le départ, pour en gagner ensuite**
- B. Songer aux personnes devant accéder aux données lorsqu'on met en place son organisation**
- C. Etre précis mais réaliste dans les choix retenus (organisation, convention de nommage, etc.) et se tenir à ses décisions**

1. Structure des fichiers – méthode hiérarchique

Le système hiérarchique, utilisé dans presque tous les types de classification systématique, est un système familier et largement utilisé. L'arborescence est particulièrement adaptée pour représenter la structure de l'information. De plus, les éléments similaires sont généralement stockés ensemble.

Cependant, l'établissement d'un système hiérarchique peut être étonnamment difficile. Sans compter que les dossiers et sous-dossiers se multiplient parfois, induisant un grand nombre de clics avant d'atteindre le document souhaité. On est parfois tenté de mettre un fichier à plusieurs endroits et les versions d'un même document peuvent se multiplier. Enfin, la réorganisation d'un arbre obsolète ou la création ou modification d'un arbre à posteriori prend énormément de temps.

- A. Eviter les catégories qui se chevauchent** : des sous-dossiers portant le même nom deviendront illisibles, sortis de leur contexte.
- B. Pondérer la profondeur et la largeur de l'arbre** : Eviter les listes interminables de fichiers dans un dossier ou les sous-dossiers « poupées-russes ».
- C. Associer des métadonnées à vos dossiers** (p.ex. README.txt)
- D. Penser large et anticiper le tri des fichiers/dossiers** (ex. YYYYMMDD pour les dates ou 01 au lieu de 1)
- E. Ne pas multiplier les copies d'un fichier** : utiliser des raccourcis pointant vers celui-ci
- F. Envisager des logiciels dédiés pour les références bibliographiques** (ex. EndNote ou Zotero)

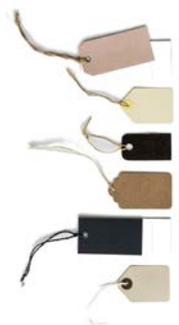
2. Nommage des fichiers

Les conventions de nommage permettent d'apporter logique et cohérence à la structure. Concrètement, il s'agit de définir à l'avance la syntaxe qui gouvernera le libellé des fichiers. Une convention de nommage doit être :

A. Descriptive

Le nom du fichier est plus lisible et permet mieux de le trouver en faisant une recherche lorsqu'il contient certains éléments comme :

- Un identifiant unique (ex. Nom du projet ou #Financement dans le nom du dossier)
- Les conditions de production des données (Instrument, producteur, etc.)
- La référence de l'expérience ou de la version (séquentielle)
- Une date



B. Cohérente et appliquée de manière uniforme

Il est nécessaire de maintenir une cohérence et une logique d'ordonnancement. Par exemple, de toujours utiliser la même syntaxe pour les dates (YYYYMMDD ou YYYY-MM-DD, etc.) ou de mettre systématiquement les mêmes informations dans les noms de fichier. Si le choix de la syntaxe est libre, son utilisation devrait être systématique.

C. Bonnes pratiques :

- × **Eviter les caractères spéciaux ou accentués** : & , * % # ; * () ! @ \$ ^ ~ ' { } [] ? < > -
- × **Eviter les espaces** (privilégier plutôt les _ - ou l'écriture « en chameau »)
- × **Eviter les points** (les réserver à l'extension indiquant le type de fichier (.docx, .csv, etc))
- × **Eviter les noms trop longs** (ils s'affichent de manière incomplète)
- × **Eviter les intitulés trop lacunaires** (si le fichier est sorti de son contexte, il sera illisible)
- ✓ **Choisir un format de date sans ambiguïté et permettant un tri chronologique automatique** : YYYYMMDD ou YYYY-MM-DD
- ✓ **Privilégier des numérotations permettant de traiter et de trier de gros volumes** (001 et non 1)

Exemples :

- × Présentation_étudeDMP@HUG_fév.18.ppt ✓ 201802_DMP_HUGpresentation.ppt
- × Resultats|enquête|v.corr ✓ 2018_QualiteVie_EnqueteGE_Results_v02.csv

Voici un exemple de convention de nommage :

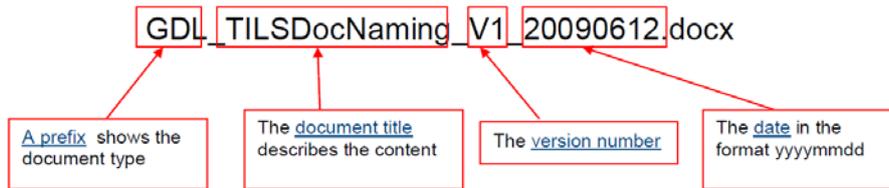


Figure 1 Convention de nommage de la Queensland University of Technology.
<https://www.library.qut.edu.au/about/management/documents/QUTTILSDocNamingConvention.pdf>

3. Versions de fichiers

Le nombre de versions d'un fichier pouvant rapidement augmenter, il est nécessaire de pouvoir les distinguer aisément, y compris bien plus tard.

Voici quelques conseils à ce sujet :

A. Numéroté les fichiers comme suit :

- a. Modification majeure : v02, v03
- b. Modification mineure : v02-01, v02-02

B. Eviter l'intitulé « final » ou le réserver à la version diffusée publiquement

C. Conserver les fichiers sources dans un dossier séparé

D. Créer un dossier archives avec les anciens fichiers.

E. Se poser la question de ce qu'il faut vraiment garder comme versions intermédiaires ?

4. Sources et Ressources

- MIT Libraries Data Management Services. (2017) *Data Management: File Organization*. <https://libraries.mit.edu/data-management/services/workshops/> CC-BY-NC
- EDINA and Data Library, University of Edinburgh. (s.d.) *Research Data MANTRA [online course, Organising Data module]* <https://mantra.ed.ac.uk/>. CC-BY.
- CORNELL University RESEARCH DATA MANAGEMENT SERVICE GROUP (s.d.) *Guide to writing "readme" style metadata* <https://data.research.cornell.edu/content/readme> CC-BY

