

Organisation des données

Midi de l'info scientifique

0. Organisation des données

L'organisation des fichiers de données demande du temps et paraître compliqué, mais elle s'avère réellement efficace, tant pour le travail individuel que collaboratif.



- A. Passer un peu de temps à l'avance, peut gagner beaucoup de temps plus tard**
- B. Être réaliste: trouver un équilibre entre faire trop et trop peu**
- C. Il n'y a pas le bon moyen de le faire; Établir un système qui fonctionne pour vous**
- D. Pensez à savoir à qui votre système doit travailler: Juste vous? Vous et votre groupe de laboratoire? Des collaborateurs?**

1. Structure des fichiers – méthode hiérarchique

Le système hiérarchique, utilisé presque dans tous les genres de classement systématique, est un système familier et largement utilisé. L'arborescence est particulièrement adéquate pour représenter la structure de l'information.



Les éléments similaires sont stockés ensemble et les sous-dossiers peuvent fonctionner comme des listes de tâches. Cependant la mise en place d'un système hiérarchique peut être étonnamment difficile ; il faut trouver la bonne balance entre la largeur et la profondeur de l'arbre, il faut prendre en compte qu'un élément ne peut se trouver qu'à un seul emplacement de l'arbre que la réorganisation d'une arborescence obsolète peut demander énormément de temps.

Voici quelques conseils et bonnes pratiques pour créer au mieux un système hiérarchique et trouver plus facilement vos dossiers :

- A. Éviter les catégories qui se chevauchent**
- B. Ne laisser pas vos dossiers devenir trop grands**
- C. Ne laisser pas votre structure devenir trop profonde**
- D. Ordonner les dates en les faisant commencer par l'année, de sorte qu'un tri par date soit possible** (par ex., AAAAMMJJ)
- E. Intégrer si possible des métadonnées à vos fichiers**
- F. Ajouter des raccourcis aux fichiers dans d'autres dossiers pertinents**

2. Nommage des fichiers

Il est important de respecter certaines conventions de nommage, élaborées pour créer une logique et une consistance au sein d'une infrastructure. Il est nécessaire que ces conventions soient publiées et connues de tous les utilisateurs afin de permettre un travail collaboratif efficient.



Une convention de nommage devrait être :

A. Descriptive

Pour cela, il faut considérer l'utilisation de certains éléments

- Un identifiant unique
- Conditions (instrument de laboratoire, solvant, température, etc.)
- Expérience (séquentielle)
- Date
- Version

B. Consistante

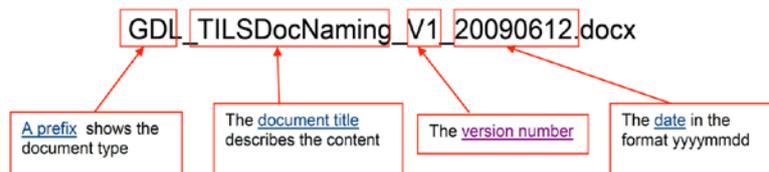
Il faut maintenir un ordre. Par exemple, lors du nommage d'une date, il faut utiliser toujours le même patron (AAAAMMJJ ou MMJJAAAA, etc.). Le choix du patron est libre, mais son utilisation doit être systématique.

Il faut aussi inclure à chaque fois la même information (DateProjetID ou TempsDate, etc). Une fois de plus, le choix du patron du nommage est libre, mais son utilisation doit être systématique.

Il existe certaines bonnes pratiques quant au nommage des fichiers, que vous pouvez trouver dans le tableau ci-dessous :

Bonnes pratiques	Exemples
Limiter le nom du fichier à 32 caractères (de préférence moins!)	32CharactersLooksExactlyLikeThis.csv
Lorsque la numérotation séquentielle est utilisée, use leading zeros to allow for multi-digit versions For a sequence of 1-10: 01-10 For a sequence of 1-100: 001-010-100	NON ProjID_1.csv ProjID_12.csv OUI ProjID_01.csv ProjID_12.csv
Ne pas utiliser de caractères spéciaux & , * % # ; * () ! @ \$ ^ ~ ' { } [] ? < > -	NON name&date@location.doc
Utiliser un seul point et l'utiliser juste avant l'extension de fichier.	NON name.date.doc OUI name_date.doc
Eviter d'utiliser des noms génériques de fichier de données qui peuvent entrer en conflit lorsqu'ils sont déplacés d'un endroit à un autre.	NON MyData.csv OUI ProjID_date.csv

Enfin, selon la TILS Document Naming Convention, le nommage d'un document devrait suivre cette convention :



3. Versions des fichiers

Le nombre de différentes versions d'un document peut rapidement augmenter au sein d'un projet de recherche. Le nommage et l'organisation de ces versions sont alors très importants afin d'éviter la perte ou la confusion des documents.

Voici quelques conseils pour gérer au mieux les versions de documents :

- A. Eviter les labels « final » imprécis
- B. Ranger les anciennes versions dans un dossier à part
- C. Enregistrer les nouvelles versions
- D. Etablir une convention consistante
- E. Documenter votre convention de nommage des versions
- F. Considérer vos besoins de contrôle des versions



4. Les 4 C de l'organisation des fichiers

L'organisation des fichiers peut finalement se résumer en cinq mots :

- A. Clair
- B. Concis
- C. Consistant
- D. Correct



E. Conforme