



Fabrique de la science ouverte – 24 juin 2024

Sobriété numérique et données de recherche : place à la pratique !

Ce document vous propose des fiches de prise en main des outils qui ont été présentés et testés ensemble lors de la journée du 24 juin. Il s'agit de pistes pour démarrer avec ces outils. Nous vous invitons à consulter les sites des différents logiciels qui proposent de la documentation plus complète.

L'équipe de la Fabrique de la science ouverte reste à votre disposition pour tout échange : scd-fabso@univ-lille.fr

Les fiches proposées :

Archifiltre : visualiser l'arborescence des répertoires et identifier les anomalies.....	2
Free File Sync : synchroniser ses fichiers pour les sauvegarder	4
GIT : gérer les versions de fichiers.....	5
ReadMe : un fichier pour décrire vos données.....	7
VeraCrypt : crypter des données	8

Archifiltre : visualiser l'arborescence des répertoires et identifier les anomalies



Archifiltre <https://archifiltre.fr/>

Les produits Archifiltre sont développés au sein de la Fabrique Numérique des Ministères Sociaux. Des outils libres et gratuits.

Que permet archifiltre ?

- Visualiser l'arborescence complète des répertoires en un coup d'oeil
- Identifier les anomalies, les documents obsolètes, redondants, à éliminer
- Générer un rapport d'audit
- Exporter un fichier qui liste les dossiers et les traitements décidés (éliminer, renommer, déplacer etc...) utile pour opérer des nettoyages hors de l'outil

Découverte de l'environnement : quelques étapes pour débuter

- Installation à partir du site archifiltre.fr
- Charger un répertoire
- Visualiser l'organisation des répertoires et des éléments (arborescence colorée cliquable, fil d'ariane à droite)
- Onglet Audit : identifier les caractéristiques des répertoires et la nature des éléments qui y sont stockés
- Onglet enrichissement : ajouter description, tag (à éliminer, à déplacer, renommer ...) (aucune opération de suppression ou de modification ne se fait à partir d'archifiltre directement).
- Onglet redondance : identifier et taguer les fichiers redondants
- Exporter : rapport d'audit (caractéristiques, nombre de fichiers en doublons et volume gagné si suppression...)
- Export excel/csv : liste des fichiers avec les caractéristiques et les tags choisis dans archifiltre. A partir de là, « il n'y a plus qu'à » se servir de cette liste pour opérer les modifications directement dans les répertoires (export d'un script possible).

Quelques idées de tri pour se lancer dans un « nettoyage des fichiers » : repérer...

- Les plus gros dossiers, les plus gros fichiers
- Les dossiers qui ont la même image de structure (doublons ?)
- Les dossiers / fichiers obsolètes
- Les dossiers zip

Tutoriels sur la chaîne <https://www.youtube.com/@archifiltre>

A noter : la suite Archifiltre propose également un outil pour le nettoyage des mails.

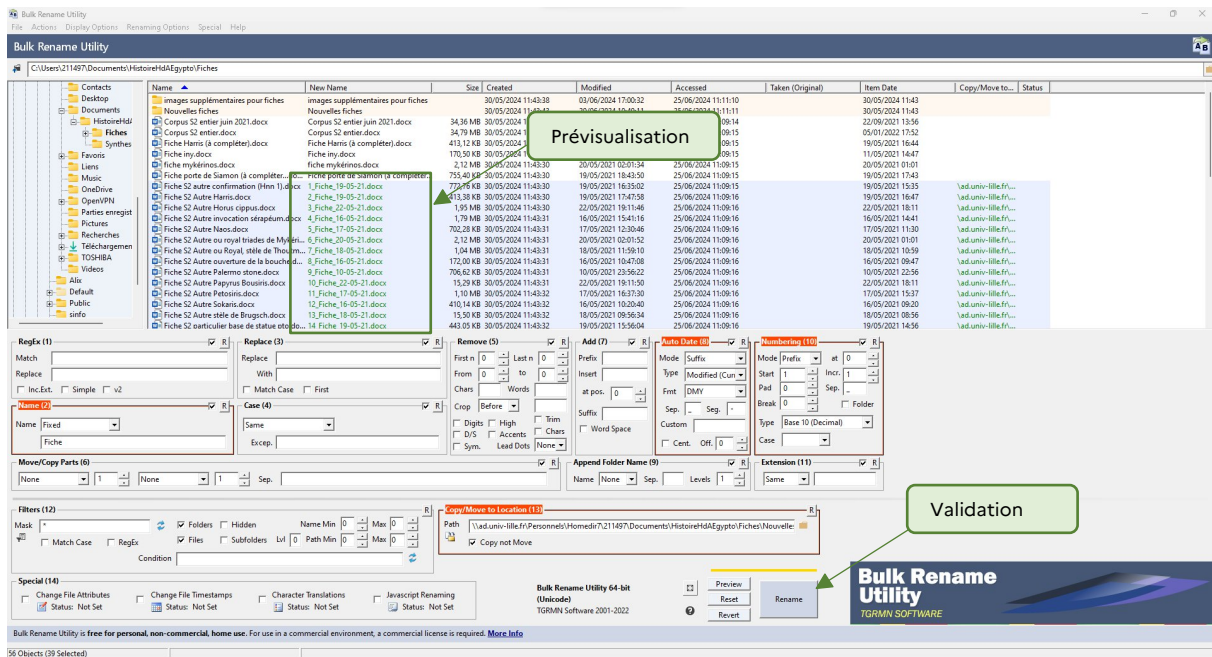
Bulk Rename Utility : renommer des fichiers en masse



L'outil **Bulk Rename Utility** sert à renommer des groupes de fichiers de manière systématique en y appliquant des règles de renommage.

On peut l'installer ici : <https://www.bulkrenameutility.co.uk/Download.php>

Pour une version qui ne nécessite pas d'installation, le lien « Portable version » sur le lien ci-dessus donne accès à un fichier zip à décompresser.



L'interface comprend un explorateur de fichier pour sélectionner les noms à modifier, puis, en dessous, des modules qui créent des « règles » pour renommer et déplacer les fichiers, numérotés dans l'ordre où ils vont s'appliquer. On peut prévisualiser les nouveaux titres dans la deuxième colonne, et annuler les modifications après avoir validé. Les modules qui apportent des modifications sont encadrés en orange. Ceux qui ne servent pas peuvent être ignorés : tout n'a pas vocation à être rempli.

Les modules permettent par exemple :

- Numérotation ou datation automatique
- Suppression ou ajout en masse de préfixe/suffixe, y compris avec les métadonnées, de manière automatique
- Déplacement ou copie des fichiers renommés vers un autre emplacement

Le manuel complet (en anglais) peut-être trouvé ici : https://www.s3.tgrmn.com/bru/BRU_Manual.pdf

Attention :

- Le programme est susceptible de ne pas voir les documents des sessions personnelles sur les ordinateurs de l'Université de Lille qui ne sont pas sur le disque dur local. Pas de problème en revanche avec les clients Nextcloud ou les dossiers partagés.
- Il n'existe pas de version Mac ou Linux

Free File Sync : synchroniser ses fichiers pour les sauvegarder

Free File Sync est un utilitaire permettant la sauvegarde et la synchronisation des fichiers pour éviter la perte des données. Pratique pour sauvegarder simultanément sur deux supports différents et conserver la version d'un fichier la plus à jour. Gratuit, open source et ergonomique, il est disponible pour Windows, macOS et Linux.

Free File Sync peut comparer les différentes versions d'un même fichier pour en effectuer la sauvegarde, la suppression, la restauration ou l'exclusion. Il fonctionne sur les disques durs externes, les clés USB et permet également d'effectuer des sauvegardes sur des serveurs en ligne et des dossiers en réseau. Il est possible de paramétrer des synchronisations automatiques pour ne pas avoir à le faire manuellement et selon des périodes déterminées.

Le site web disponible ici : <https://freefilesync.org/> propose une documentation complète : on y trouve notamment des tutoriels pour se former à l'utilisation de l'outil, une FAQ et un forum pour poser des questions et faire remonter des bugs.

Un point de vigilance tout de même à garder en tête : l'outil ne reconnaît pas toujours un fichier qui a été renommé ensuite. L'algorithme l'identifiera comme un nouveau fichier. Ainsi, il proposera de le dupliquer dans le second dossier de sauvegarde et considérera la version précédente comme manquante.

GIT : gérer les versions de fichiers



git

L'outil **GIT** est un logiciel de gestion de versions décentralisé (notre pc est utilisé comme serveur d'enregistrement du suivi des modifications) libre et gratuit.

Que permet GIT ?

- Assurer l'intégrité des fichiers
- Assurer la disponibilité des fichiers pour une édition collaborative
- Assurer le suivi de historique des modifications dites « commits » (modifications d'un script, ajouts des nouveaux fichiers, etc) des fichiers dans un espace de travail (dossier GIT)
- Collaborer avec des autres membres des projets via la création des « clones » (copies des espaces de travail) locales dans les espaces locaux des collaborateurs ou sur des forges

Quel public ?

- Personnes travaillant sur des scripts (python, R, markdown)
- Personnes ne souhaitant pas travailler avec des interfaces, familiarisés avec le code. L'ensemble des commandes est réalisé via une console dans un environnement dédiée.

Découverte de l'environnement : quelques étapes pour débiter et découvrir l'outil

- Installation à partir du site : <https://git-scm.com/>
- Création ou sélection de l'espace de travail (un dossier avec nos fichiers)
- Ouvrir Git BASH (console) et se placer dans les commandes sur le dossier/espace de travail



- Déclaration de l'espace de travail en tant que dossier GIT (pour l'enregistrement des modifications via des *commits* à *posteriori* via les commandes ci-dessous

```
> git init
```

- Renseigner le nom et le mail utilisateur pour la configuration de l'espace via les commandes ci-dessous :

```
$ git config --global user.name "John Doe"  
$ git config --global user.email johndoe@example.com
```

- Sauvegarde des fichiers existantes comme statut initiale des modifications dans le dossier GIT grâce à ma commande « git status »
- Réalisation d'une modification dans l'espace de travail : par exemple, modification d'un script ou ajout d'un nouveau fichier
- Vérification de le statut des modifications sur le dossier GIT (en vert les modifications prises en compte, en rouge les modifications pas déclarés) via la commande « git status »
- Déclaration des modifications via des commits documentés avec la commande « git commit -m « commentaire du commit »

Pour aller plus loin sur les fonctionnalités de GIT, il es possible de consulter l'espace de documentation du projet sur <https://git-scm.com/doc>. Avec un manuel et des vidéos pour démarrer.

Pour visualiser et tester la construction des branches des versions consulter : https://learngitbranching.js.org/?locale=fr_FR&demo=

ReadMe : un fichier pour décrire vos données

Un document readme fournit des informations sur un jeu de données et assure que ce dernier puisse être correctement interprété, par vous-même à une date ultérieure, ou par des personnes extérieures lorsque les données sont partagées ou publiées. Des métadonnées suivant un standard sont préférables, mais lorsqu'aucun standard ne semble correspondre entièrement à vos besoins, la rédaction de métadonnées de type « readme » est une stratégie appropriée.

Consultez les bonnes pratiques en matière de rédaction de ReadMe et téléchargez un modèle sur le site des Bibliothèques universitaire et Learning center de l'Université de Lille : <https://bu.univ-lille.fr/chercheurs-doctorants/science-ouverte/donnees-de-recherche/guide-de-redaction-de-metadonnees-de-type-readme>

VeraCrypt : crypter des données



VeraCrypt est un logiciel utilitaire **sous licence libre** utilisé pour le **chiffrement à la volée** (OTFE). Il est développé par la société française IDRIX2 et permet de **créer un disque virtuel chiffré dans un fichier ou une partition**. L'ensemble du dispositif de stockage demande une authentification avant de monter le disque virtuel.

Intégré dans le **SILL**, la liste des logiciels libres préconisés par l'État dans le cadre de la modernisation globale de ses systèmes d'informations

Objectif : garantir la confidentialité des données facilement et sans ralentir votre système.

Méthode : création de volumes chiffrés, lisibles uniquement par leur propriétaire. On peut chiffrer une partition complète ou un répertoire seulement => permet de ne crypter qu'une partie des fichiers, et pas la totalité d'un disque dur par exemple.

Ces espaces sécurisés peuvent ensuite être montés, et devenir accessibles seulement à l'aide du mot de passe choisi. Il prend en charge les disques durs internes, mais aussi les supports de stockage amovibles tels que les clés USB ou les disques externes.

L'utilisateur peut choisir le mode de chiffrement qui correspond à ses besoins : AES, Serpent, Twofish ou une combinaison des trois. Enfin, l'application embarque des outils qui permettent d'évaluer l'efficacité de votre protection.

VeraCrypt est disponible en plusieurs langues et est compatible avec les systèmes d'exploitation les plus répandus, tels que Windows, macOS et Linux. Vous pouvez télécharger le logiciel gratuitement via nos liens directs vers son site officiel.

<https://www.veracrypt.fr>

<https://doc.ubuntu-fr.org/veracrypt>

Guide pas à pas : <https://veracrypt.fr/en/Beginner%27s%20Tutorial.html>