Comment s’assurer de l’intégrité de ses données de recherche ?

Pour accéder à la ressource : https://doranum.fr/stockage-archivage/comment-sassurer-de-lintegrite-de-ses-donnees-de-recherche_10_13143_myya-mx96/

Date de publication : 16/06/2025

Sommaire

Quel que soit votre domaine de recherche, lorsque vous souhaitez analyser des données que vous avez produites ou recueillies, vous devez en premier lieu vous assurer de leur **intégrité** avant de les étudier.
En effet, les données peuvent avoir été altérées par de mauvaises manipulations ou des transferts informatiques défectueux durant leurs phases de sauvegarde et de stockage. Il est donc important de vérifier que vos données ne sont pas altérées par ces opérations.
Afin de vérifier l’intégrité de vos données, cette ressource pédagogique vous propose des exemples d’outils et de bonnes pratiques qui assurent la **pérennité** et l’**analyse** de vos données dans de bonnes conditions.

1. S’assurer de l’intégrité de ses données de recherche au moyen de sommes de contrôle

Vos données peuvent être corrompues par exemple par des modifications générées par un **logiciel malveillant** (ransomware) ou éventuellement suite à une **erreur de manipulation**.
Elles peuvent également être corrompues lors d'un **transfert défectueux** ou lors du **plantage d’un disque dur**.
Dans tous les cas, pour vérifier l'intégrité de vos données de recherche, il faut **pouvoir contrôler** qu'elles n'ont **pas** été accidentellement **modifiées**.
Pour cela, il faut faire appel aux **sommes de contrôle**.
Les sommes de contrôle sont générées par des algorithmes. Elles attribuent une chaîne de lettres et de nombres qui sont uniques à un fichier. Cette chaîne de caractères est une sorte d’**empreinte numérique de la donnée**.
En pratique, vous devez avoir la même empreinte numérique avant et après toute opération de manipulation ou de transfert informatique de vos données (exemples : opérations de migrations de formats de fichiers, opérations de copies, de transferts sur d’autres supports, etc).

2. S’assurer de l’intégrité de ses données de recherche avec le contrôle d’accès

Pour protéger efficacement vos données durant votre projet, vous devez également mettre en place un **contrôle d’accès**. Ce dernier doit permettre de voir qui a accédé à vos données et qui les a modifiées.
Il est essentiel d’accorder des **permissions d’accès** uniquement au regard de ce qui est **strictement nécessaire** au bon déroulement de votre projet de recherche. Pour cela :

- Pensez à limiter le nombre d'utilisateurs ayant accès à vos données.
- Limitez également la visibilité de vos données, notamment quand elles sont sur un réseau interne ou a fortiori sur Internet.
- N’utilisez jamais de système de partage public sans le chiffrement de vos données.

Par ailleurs, il est primordial de mettre vos **données brutes**, recueillies au début du projet, **en lecture seule** afin de garantir qu’elles ne soient/seront pas modifiées. Il

vous faut donc utiliser un système de partage qui propose cette fonctionnalité (par exemple, les espaces collaboratifs Nextcloud ou RESANA).
Pour en savoir plus, consultez la ressource pédagogique « Les espaces de travail collaboratifs sécurisés ».
Enfin, n'oubliez pas que si vous donnez accès à des données sensibles, vous devez documenter l'accès et les conditions d'accès. L'accès à ces données est strictement contrôlé et limité aux membres de l'équipe de recherche autorisés.

3. S'assurer de l'intégrité de ses données de recherche en limitant le nombre de copies

Il faut garder à l'esprit que multiplier le nombre de copies n'améliore pas l'intégrité de vos données.
En effet, si vous multipliez les copies, vous ne saurez plus par exemple quelle est la dernière version mise à jour.
Il faut donc **limiter les copies** mais également les placer sur des **systèmes de stockage sécurisés** qui eux-mêmes font appel à la **règle « 3-2-1 »** : 3 exemplaires stockés sur 2 supports différents dont 1 exemplaire sur un serveur distant.
Pour en savoir plus sur la règle « 3-2-1 », consultez la ressource pédagogique DoRANum « La sauvegarde 3-2-1 ».

L'utilisation de systèmes garantissant l'intégrité de vos données de recherche lors de leurs manipulation, transfert et stockage vous permet d'analyser vos données de recherche dans de bonnes conditions et, à terme, d'obtenir des résultats scientifiques solides.