Stockage, partage et archivage : quelles différences ?

STOCKAGE ET SAUVEGARDE

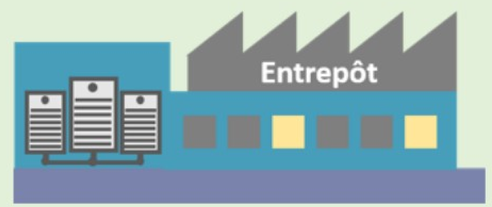
Serveur sécurisé de l'institution

Espace collaboratif de travail

Le chercheur

L'équipe de recherche

PARTAGE



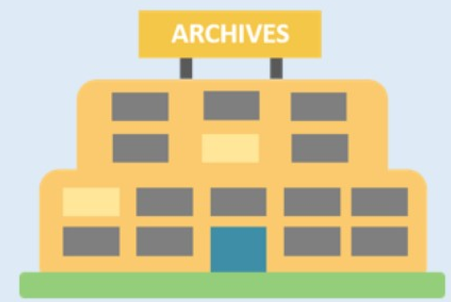
Entrepôt généraliste ou disciplinaire

L'équipe de recherche

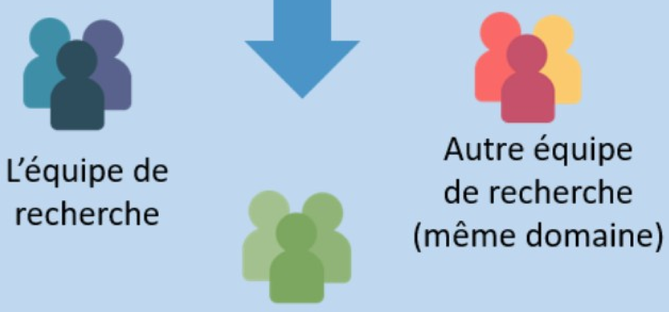
Autre équipe de recherche (même domaine)

Autre équipe de recherche (autre domaine)

ARCHIVAGE PÉRENNE



Plateforme d'archivage (ex : CINES)

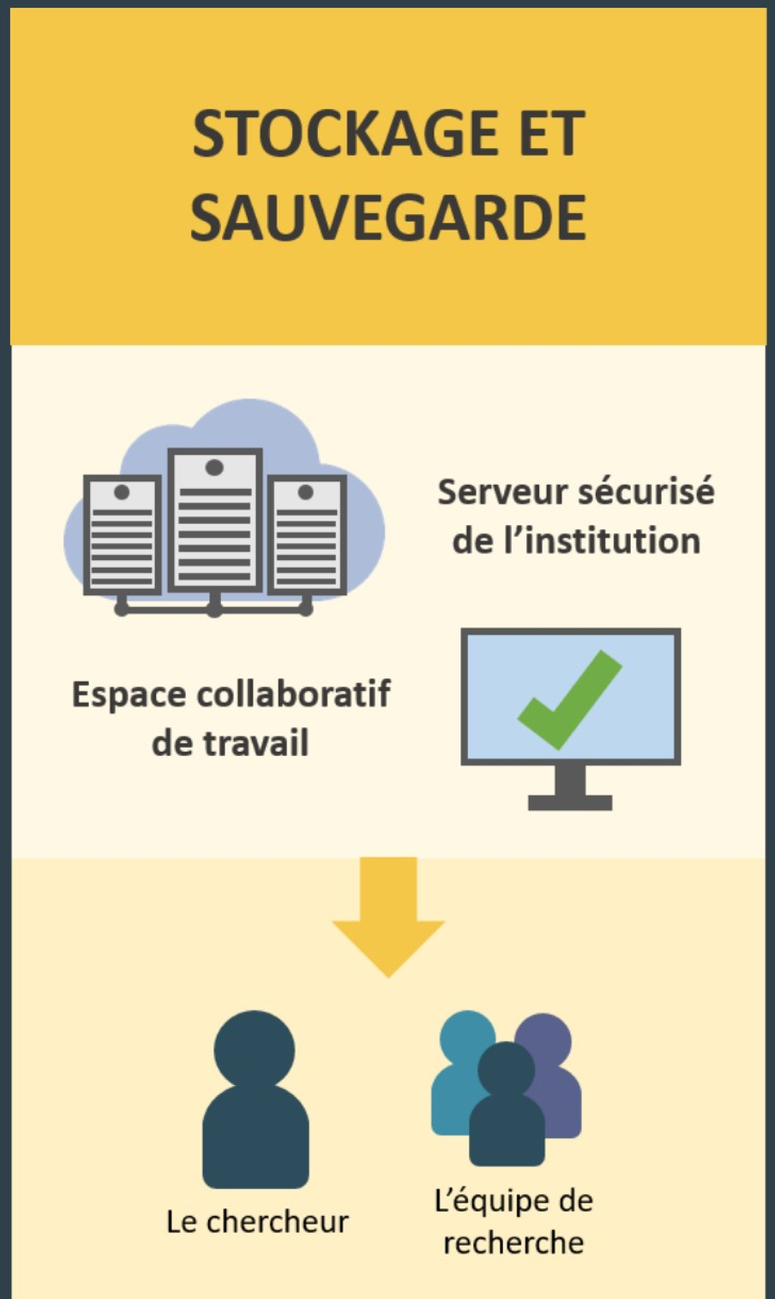


L'équipe de recherche

Autre équipe de recherche (même domaine)

Autre équipe de recherche (autre domaine)

Stockage, partage et archivage : quelles différences ?



La sauvegarde et le stockage sécurisé durant le projet ont pour objectifs de garantir la sécurité des données, et d'en faciliter l'accès pour l'ensemble des collaborateurs du projet.

Comment faire ?

- Dans l'idéal, dupliquer, stocker les données à différents endroits sur différents supports.

> **Règle du 3-2-1 :**
> - garder 3 exemplaires des données
> - sur 2 supports ou technologies différents
> - dont 1 se trouve hors site

- Déterminer l'hébergement : serveurs locaux (machines virtuelles), cloud institutionnel avec accès sécurisé...

- Estimer la volumétrie des données.

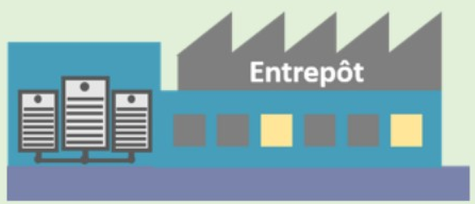
- Organiser et planifier la sauvegarde des données :
  - À chaque point d'étape du projet, sélectionner les données à sauvegarder, à supprimer.
  - Définir leur durée de conservation.

- Gérer les versions :
  - Les différents états des données sont conservés en corrélation avec les différentes étapes de traitement.
  - Cela permet de revenir à une version antérieure si besoin.

Source : Université de Lille Sciences et Technologies -
https://indico.math.cnrs.fr/event/2317/contributions/1314/attachments/692/773/problematique_des_sauvegardes-journees_mathrice-20170928.pdf

Stockage, partage et archivage : quelles différences ?

PARTAGE



Entrepôt généraliste ou disciplinaire

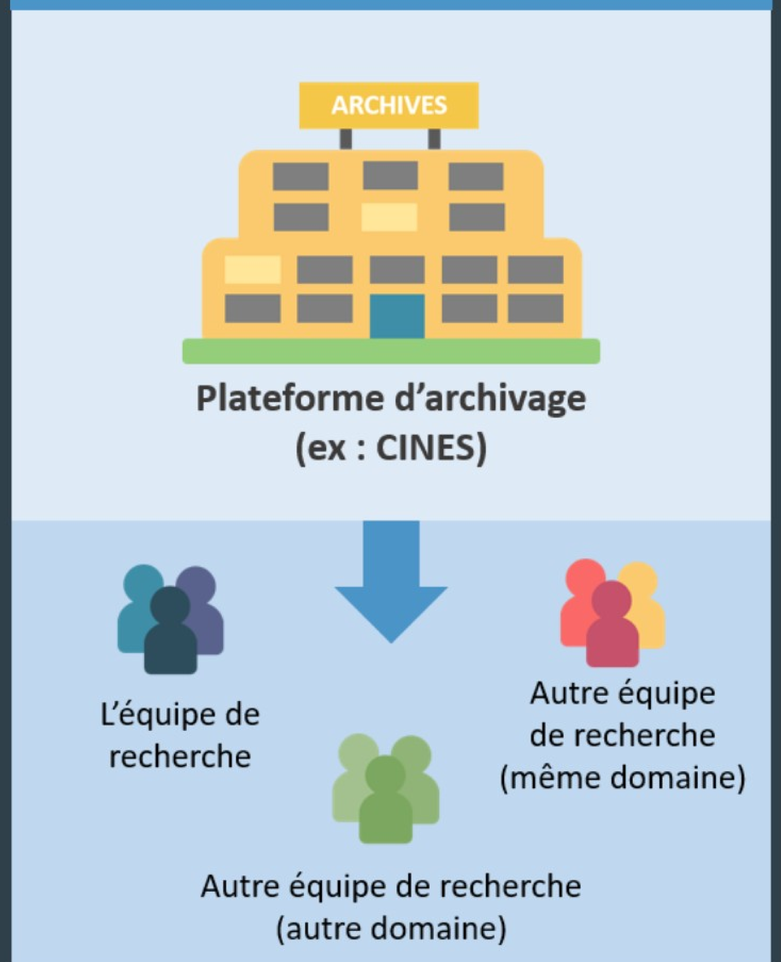
Le partage consiste à déposer vos données dans un entrepôt de données afin de les rendre accessibles facilement et de permettre leur réutilisation (principes FAIR) sur le court et le moyen terme (5 à 10 ans), pour des chercheurs du même domaine, ou d'un autre domaine.

Comment faire ?

- Sélectionner les données à partager.
- Estimer la volumétrie des données et éventuellement le budget.
- Vérifier la compatibilité et l'interopérabilité des formats de données en se posant par exemple ces questions :
  - Le contenu est-il valorisé ou toujours valorisable ?
  - Le fichier et son contenu sont-ils compréhensibles en l'état ?
  - Le format de fichier utilisé nécessite-t-il un logiciel pour être lisible ?
- Migrer si nécessaire vers un format adapté, le plus ouvert possible.
- Préparer si nécessaire les codes sources (ex : scripts) qui permettront de lire et traiter les données.
- Compléter et enrichir les métadonnées (en fonction de l'entrepôt choisi).
  - Si ce n'est pas déjà fait, choisir un standard de métadonnées.
  - S'il n'en existe pas d'adapté, créer un schéma de métadonnées.
  - Compléter les champs pour chaque jeu de données, suivant le standard adopté.

Stockage, partage et archivage : quelles différences ?

ARCHIVAGE PÉRENNE



L'archivage pérenne a pour objectif de conserver les données, d'en garantir l'accès et d'en préserver l'intelligibilité sur le long terme. L'archivage pérenne garantit une conservation des données pour plus de 30 ans. Les données à archiver doivent présenter une valeur scientifique reconnue par la communauté dont elles proviennent.

En France : le CINES est l'opérateur mandaté par le Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation pour opérer la mission d'archivage pérenne.

Comment faire ?

- Sélectionner les jeux de données (et métadonnées associées) à conserver à long terme (peuvent être différents des jeux de données partagés).
- Prévoir la volumétrie des données et le budget nécessaire.
- Traiter les données si cela est nécessaire (ex : données personnelles à anonymiser).
- Vérifier la validité des formats de fichiers de données avec l'outil FACILE mis en place par le CINES.
- Documenter les logiciels permettant l'accès aux données. Documenter les logiciels permettant l'accès aux données.
- Compléter et enrichir si besoin les métadonnées (les données doivent posséder une description minimale imposée par le CINES).

Sources : Huma-Num - Archivage à long terme - https://www.huma-num.fr/services-et-outils/archiver
CINES – FACILE - https://facile.cines.fr/