



**HAL**  
open science

## Pratiques et outils de stockage des données pendant un projet de recherche

Clément De Belsunce, Nathalie Leménager, Marie-Laure Trémélo

### ► To cite this version:

Clément De Belsunce, Nathalie Leménager, Marie-Laure Trémélo. Pratiques et outils de stockage des données pendant un projet de recherche. Semaine Data-SHS, Clément De Belsunce, Dec 2023, Aix-en-Provence, France. hal-04458305

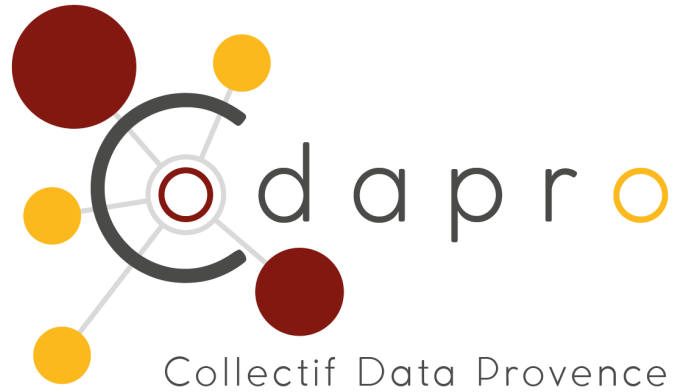
**HAL Id: hal-04458305**

**<https://amu.hal.science/hal-04458305>**

Submitted on 14 Feb 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

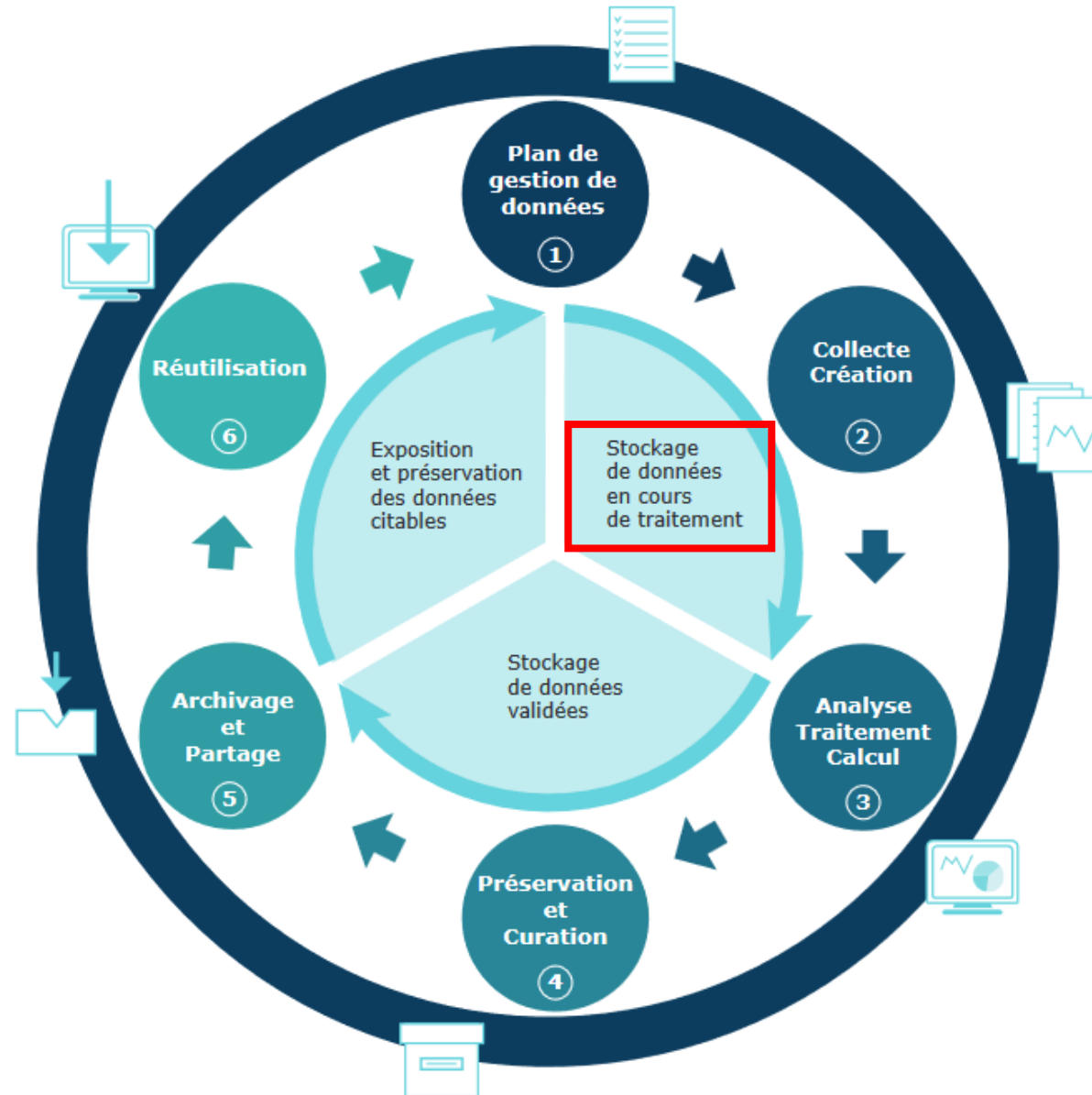


# Matinées « Gestion des données et bonnes pratiques de recherche » Épisode 3

## Pratiques et outils de stockage des données pendant un projet de recherche

Clément De Belsunce, Nathalie Leménager, Marie-Laure Trémélo  
13 décembre 2023

# Chronologie d'un projet de recherche



# Quelques définitions

- Stockage / sauvegarde / archivage

**Stockage** : c'est le moyen, le support, l'endroit où les données sont conservées.

**Sauvegarde** : elle concerne la réplication de données sur au moins deux supports distincts afin de pouvoir restaurer des données endommagées ou perdues. Elle peut aussi s'apparenter à un stockage à moyen terme de données liées à des projets terminés (donc froides) mais que l'on souhaite conserver pour réutilisation ultérieure éventuelle.

**Archivage** : c'est la conservation à long terme de données numériques de façon sécurisée et pérenne. Pour cela, il s'appuie sur des processus de vérification de l'intégrité des données (et des métadonnées associées), de relecture du contenu et des métadonnées associées, et de maintien des données sur un support exploitable.

# Quelques définitions

- Données froides / chaudes

Les **données froides** sont des données qui ne sont que très peu utilisées. On peut potentiellement un jour en avoir besoin, mais on n'y accède quasiment jamais.

Au contraire, les **données chaudes** sont des données actives, auxquelles on accède souvent, voire de façon intensive.

- Volumétrie des données

Le **volume** fait référence à la quantité de données générées et stockées. Les volumétries des données à stocker sont extrêmement différentes d'un projet à l'autre. Actuellement, on parle de « grosses volumétries » à partir de quelques dizaines de To. Ces tailles de données complexifient en particulier les transferts des données par le réseau.

# Questions préliminaires

- Que dois-je stocker ?
  - Des données scientifiques, des codes, des documents ?
  - Des fichiers binaires ou ascii ?
- Quelle est / sera la volumétrie des données ?
- De quel niveau de sécurisation ai-je besoin ?
- Une sauvegarde est-elle nécessaire ?
- Les données ont-elles un caractère sensible ou confidentiel ?
- Qui devra avoir accès aux données ?
- De quelle manière accéder aux données ?

# Questions préliminaires

- Quand vais-je avoir besoin de ces données ?
  - Rapidement, de façon régulière
  - Dans quelques mois
  - Peut-être dans quelques années ...
- À la fin du projet, quel volume de données devra être préservé et sur quelle durée ?
- Quel est le financement prévu pour la gestion et le stockage des données ?

*le stockage des données (au sens large) a un coût matériel et humain qu'il est impératif de prendre en compte lors du dépôt des projets de recherche et dans le plan de gestion de données (PGD)*

# Des règles à mettre en place...

- Assurer une sécurité informatique
  - avoir un **antivirus à jour** sur son ordinateur est primordial pour minimiser les attaques informatiques
  - une solution plus radicale que l'antivirus : stocker ses données sur un ordinateur qui n'est connecté à aucun réseau
  - protéger l'accès à vos documents avec un **mot de passe** d'un niveau de protection élevé (longue chaîne de caractères, majuscules, chiffres, etc.)
  - **chiffrer** l'espace disque (Bitlocker, VeraCrypt...)



# Des règles à mettre en place...

- Règle de sauvegarde 3-2-1

*préconisée par le CINES (Centre informatique national de l'enseignement supérieur)*

consiste à :

- conserver **3 copies** d'une même donnée
- sur au moins **2 supports** différents
- avec **1 copie à distance**

*Par exemple, pour des données courantes auxquelles on aurait recours régulièrement, il conviendrait de stocker une première version en local sur l'ordinateur, une copie sur un stockage externe local (un disque dur ou une clé USB) et une copie sur un serveur institutionnel.*

# Le stockage des données au quotidien

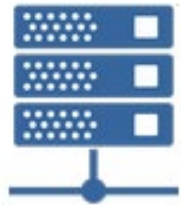
## Différents supports de stockage disponibles



votre ordinateur



support de stockage externe



serveur institutionnel



service cloud institutionnel

*Chaque support a ses avantages et inconvénients, notamment en termes de coût, de sécurité, de capacité de stockage et d'accès.*

# Le stockage des données au quotidien

| Support de stockage   | Sécurité   | Accès  | Coût  | Remarque d'utilisation   |
|---|--|--|---|--|
|  <p>Ordinateur professionnel</p> | <p>★★★★☆</p> <p>Sujet au piratage informatique, aux détériorations et pannes</p>   | <p>★★★★☆</p> <p>Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...)</p> | <p>★★★★★</p> <p>Pas de coût supplémentaire ou coût peu important</p>                    | <ul style="list-style-type: none"> <li>- Pour un stockage temporaire</li> <li>- Nécessité de crypter les données confidentielles et sensibles</li> </ul>   |
|  <p>Support externe</p>          | <p>★★★☆☆</p> <ul style="list-style-type: none"> <li>- Sujet au vol, à la perte du support</li> <li>- Durée de vie limitée (dégradation du matériel)</li> </ul> | <p>★★★★☆</p> <p>Facilement transportable, il permet de transférer les données vers un autre ordinateur</p>             | <p>★★★★★</p> <p>Pas de coût supplémentaire ou coût peu important</p>                    | <ul style="list-style-type: none"> <li>- Pour un stockage temporaire</li> <li>- Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles</li> </ul>  |
|  <p>Serveur institutionnel</p>   | <p>★★★★★</p> <p>Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)</p>  | <p>★★★★☆</p> <p>La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures</p>   | <p>★★★★☆</p> <p>Coût assez important mais pas forcément répercuté sur l'utilisateur</p> | <ul style="list-style-type: none"> <li>- Pour un stockage plus pérenne</li> <li>- Adapté pour le stockage de données sensibles et des versions « stables » de vos données</li> <li>- Toutes les institutions ne proposent pas ce service</li> </ul>      |
|  <p>Serveur Cloud</p>          | <p>★★★★☆</p> <p>On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent</p>  | <p>★★★★★</p> <p>Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage</p>            | <p>★★★★☆</p> <p>Payant à partir d'une certaine limite de stockage</p>                   | <ul style="list-style-type: none"> <li>- Pour un partage avec des personnes externes à l'institution</li> <li>- Ne pas y mettre de données sensibles ou confidentielles</li> <li>- Pas de contrôle sur la procédure de sauvegarde des données</li> </ul> |

# Opérateurs de services de stockage - National



- Centre de calculs - IN2P3

UAR CNRS, pilotée par le CC-IN2P3, localisée à Lyon

**stockage de masse**, préférable pour de gros volumes qui ne nécessitent pas un accès fréquent ou rapide

Coût en fonction des projets

Accès, via un client spécifique, après accord pour les projets autorisés

<https://cc.in2p3.fr/>

<https://doc.cc.in2p3.fr/fr/Data-storage/mass-storage.html>

# Opérateurs de services de stockage - National



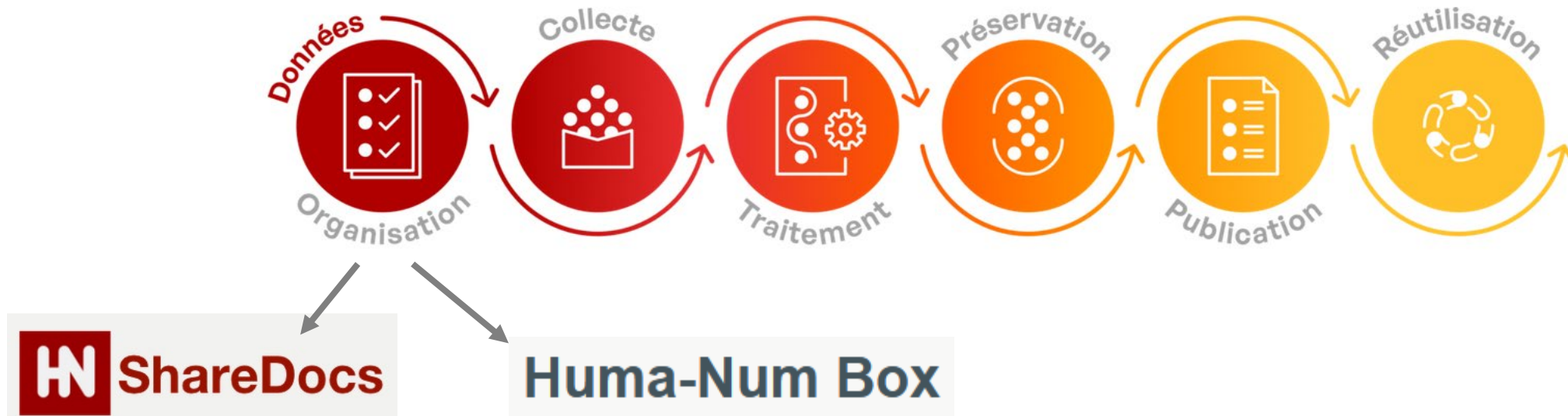
- Huma-Num

Pour stocker, échanger, partager des données de petite, moyenne ou grande taille

Pour des données chaudes ou froides

Accès limité à la communauté SHS

Différents types de stockage



# Opérateurs de services de stockage - National



- ShareDocs

gestionnaire de fichiers qui permet :

- de stocker les données d'un projet de recherche sur un serveur hautement sécurisé
- d'organiser les données et de les partager entre les participants.

Il peut être utilisé via un navigateur web, un client WebDAV ou un logiciel de synchronisation de fichiers.

Il est adapté au travail quotidien et à la mise à jour régulière des fichiers Son volume peut aller jusqu'à environ 1 To de données

# Opérateurs de services de stockage - National



- Huma-Num Box
  - service de stockage sécurisé pour la préservation de volumes importants de données (plusieurs To), par réplication et distribution sur plusieurs sites en France.
  - s'utilise comme un disque réseau ou via le protocole SFTP (avec un client comme Filezilla).
  - accessible depuis tout Internet, indépendamment de la présence ou non d'un serveur du dispositif sur son site géographique.
  - destiné aux données dites froides ou tièdes, qui ne sont pas fréquemment accédées.

# Opérateurs de services de stockage - National

- France Grilles

Stockage sur le cloud (disponible et personnalisable)

Pour des données chaudes ou froides



<https://www.france-grilles.fr/presentation/gis/>

Accès aux membres de l'organisation virtuelle France Grilles. Accès à cette VO sur demande.

Accès libre, participation à l'infrastructure si volumétrie importante

France Grilles est une infrastructure de Recherche structurée en GIS soutenue par différents organismes de recherche (MESRI, CNRS, CEA, CPU, INRA, INRIA, INSERM, RENATER).



# Opérateurs de services de stockage - en local



- MyCORE

Pour tous les personnels des unités CNRS

- permet :

- d'accéder à ses fichiers depuis n'importe quel terminal via le web
- de synchroniser des documents de travail présents sur un ou plusieurs terminaux de travail
- De synchroniser automatiquement toute modification apportée à un fichier sur les terminaux désignés
- de partager facilement et rapidement des fichiers volumineux avec des collaborateurs ou des partenaires externes

*My CoRe n'est pas prévu pour la synchronisation de données scientifiques ou de code*

[https://ods.cnrs.fr/my\\_core.php](https://ods.cnrs.fr/my_core.php)

# Opérateurs de services de stockage - en local



- AmuBox, permet de :
  - Stocker de manière sécurisée vos documents (60 Go d'espace)
  - Partager facilement en quelques clics des fichiers avec des personnes d'AMU ou ne faisant pas partie d'AMU,
  - Accéder à vos documents de manière synchronisée, depuis n'importe quel poste ou appareil mobile connecté à Internet

# Opérateurs de services de stockage - en local



- Data Center  
l'offre

## Hébergement physique =

Au U, à la demi-baie ou baie  
sécurisée

## Hébergement virtuel =

Des serveurs avec un haut  
niveau de service

## Hébergement Web =

LAMP en mode PAAS

## Solutions de stockage =

Orientées blocs, fichiers et  
bientôt objets

<https://datacentersud.univ-amu.fr/>

Et vous,  
quelle(s) solution(s) de stockage  
utilisez-vous ?