

## Archivage / Conservation à long terme des données de recherche

A la fin d'un projet de recherche, se pose la question de l'archivage des données, c'est-à-dire de leur préservation sur le long terme en toute sécurité. L'objectif de l'archivage est que vos données soient disponibles et lisibles lorsque d'autres chercheurs ou les autorités auront besoin de les consulter pour les réutiliser ou les contrôler dans l'avenir. Pour bien archiver vos données, vous devez vous poser plusieurs questions : Quelles données conserver ? Où ? Dans quelles conditions ? Avec cette fiche pratique, le CeRIS vous propose de vous aider à y répondre.

### Quelles données conserver à long terme ?

Pour faire un tri parmi vos données et déterminer quelles données doivent être conservées sur le long terme, vous devez prendre en compte 3 critères :

#### Utilité

Conserver les données :

- Qui pourront être réutilisées pour de futures recherches
- Qui sont uniques, non reproductibles, ou difficilement reproductibles
- Qui permettent de comprendre l'ensemble du projet de recherche
- Qui apportent la preuve de la bonne conduite de la recherche
- Qui permettent de valider les résultats présentés dans les publications scientifiques

#### Obligation juridique

Exemple : pour les recherches biomédicales portant sur un médicament dérivé du sang, les documents et données relatifs à la recherche doivent être conservés pendant quarante ans après la fin de la recherche biomédicale ou son arrêt anticipé. (Arrêté du 8 novembre 2006)

Exemple : Les données personnelles doivent être supprimées une fois que les besoins pour lesquels elles ont été collectées sont satisfaits (droit à l'oubli consacré par la « loi informatique et libertés »). Ainsi, les données non anonymisées ont une durée de conservation déterminée par la CNIL. Contactez [rgpd@pasteur.fr](mailto:rgpd@pasteur.fr) pour plus d'information.

#### Intérêt historique / patrimonial

Exemple : Les données épidémiologiques relatives à la pandémie de grippe espagnole en 1918 sont encore analysées 100 ans plus tard pour en tirer des leçons.



Cette œuvre est mise à disposition selon les termes de la [Licence Creative Commons CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)

## Où archiver les données de recherche ?

Les données de recherche doivent être archivées dans une base de données :

- **Certifiée** pour une conservation et une gestion à long terme : pour préserver la lisibilité, l'intégrité et l'intelligibilité des données
- **Interrogeable** : pour faciliter la découverte et donc la réutilisation des données.

Trois solutions s'offrent à vous :

### Le CINES : Centre Informatique National de l'Enseignement Supérieur

#### Qu'est-ce que c'est ?

Le CINES est la seule plateforme d'archivage numérique française pour l'Enseignement Supérieur et la Recherche. Cette plateforme est certifiée pour une conservation à long terme et s'engage à :

- Conserver plusieurs copies des données, dont une sur site distant
- Convertir les données en d'autres formats si la technologie change
- Migrer les données sur d'autres supports en cas de nécessité
- Associer aux données des métadonnées et un identifiant unique et pérenne.

#### Comment archiver ses données ?

Pour mener un projet d'archivage au CINES, vous devez :

- Convertir vos données dans un format pérenne, accepté par le CINES
- Envoyer au CINES une lettre d'intention pour décrire le projet d'archivage : données à archiver, volumétrie, formats, métadonnées
- Prévoir éventuellement des développements informatiques pour l'interfaçage avec la plateforme du CINES
- Prévoir une tarification qui dépend de la volumétrie et du nombre de sauvegardes.

### EOSC-Hub : European Open Science Cloud

#### Qu'est-ce que c'est ?

Le projet EOSC-Hub est un projet financé par la Commission Européenne dans le cadre du programme H2020. L'objectif est de mettre en place d'ici 2020 un environnement virtuel pour le stockage, la gestion, l'analyse et la réutilisation des données de recherche européennes. L'EOSC-Hub rassemble plusieurs fournisseurs de services et notamment un service d'archivage des données.

#### Comment archiver ses données ?

Pour mener un projet d'archivage, vous devez :

- Indiquer le volume de données à archiver
- Indiquer la durée pendant laquelle vous souhaitez les conserver
- Prévoir une tarification qui dépend de la volumétrie et de la durée de l'archivage.

## Un entrepôt de données certifié pour la conservation à long terme

### Comment trouver un entrepôt ?

Pour trouver un entrepôt de données qui corresponde à votre thématique de recherche, vous pouvez consulter :

- [FAIRsharing](#) : un répertoire d'entrepôts en sciences de la vie
- [Re3data](#) : un répertoire d'entrepôts multidisciplinaire
- [La liste d'entrepôts dans le domaine biomédical proposée par le CeRIS](#)

### Comment savoir s'il est certifié pour la conservation à long terme ?

Vous devez consulter la politique d'archivage de l'entrepôt sur son site.

Exemples :

- [La politique d'archivage de Harvard Dataverse](#)
- [La politique d'archivage de Dryad](#)

## Pendant combien de temps conserver les données ?

Il n'existe pas de tableau d'archivage pour les données de recherche comme il existe [pour les documents](#). Toutefois, la même logique peut être utilisée pour déterminer la durée de conservation des données.

**Exemple** : à l'Institut Pasteur, les cahiers de laboratoire sont conservés pendant 25 ans minimum car ils permettent de valider les résultats de recherche. Si l'on suit la même logique, les données de recherche associées aux publications qui valident les résultats publiés devraient être conservées pendant 25 ans minimum.

Ainsi, pour déterminer la durée de conservation des données, il faut prendre en compte :

- **Les obligations légales et réglementaires**
- **Les exigences normatives**, si cela est pertinent dans votre contexte
- **Les risques en cas de non-disponibilité des données**. Par exemple, pour les données uniques, impossibles ou difficiles à reproduire, le risque est une perte de temps et un coût non négligeable pour retrouver/reconstituer les informations.

En cas de doute, vous devez toujours prendre la durée la plus longue.

## Quels formats d'archivage choisir ?

Pour que vos données restent lisibles et intelligibles dans l'avenir, vous devez choisir un **format de fichier pérenne**, c'est-à-dire :

- **Ouvert** (voir ci-après la définition)
- **Largement utilisé**
- **Normalisé** (si possible)

Si vous prévoyez d'archiver vos données au CINES, vous devez convertir vos données dans un format accepté par le CINES. La liste des formats acceptés est disponible [ici](#). Si aucun des formats acceptés n'est adapté pour vos données, l'équipe du CINES travaillera avec vous à trouver un format pérenne qui convienne à vos données.

### Formats ouverts et formats fermés : de quoi s'agit-il ?

Format ouvert	Format fermé
Les spécifications fonctionnelles et techniques sont <b>publiques</b> et disponibles <b>gratuitement</b> (ou à faible coût)	Les spécifications fonctionnelles et techniques ne sont <b>pas publiques</b>
Aucune entrave légale ne s'oppose à sa <b>libre utilisation</b>	Des <b>restrictions légales</b> s'opposent à son utilisation (droit d'auteur, brevet, copyright).
Un document enregistré dans un format ouvert sera <b>indépendant du logiciel</b> utilisé pour le créer, le modifier, le lire et l'imprimer. Les formats ouverts permettent donc l' <b>interopérabilité</b> des données.	Ce format n'est lisible qu'avec un <b>logiciel particulier</b> ou une famille de logiciels
Il est maintenu par une <b>organisation à but non lucratif</b>	Les formats fermés correspondent souvent à des <b>formats propriétaires</b> .

### Exemples d'alternatives ouvertes aux formats fermés

	Format fermé	Format ouvert
Texte	DOC, PAGES	ODT, TXT
Tableur	XLS, NUMBERS	ODS, CSV
Vidéo	AVI, WMV	MP3, MP4, MKV
Image	PSD, JPEG 2000	PNG, BMP, JPG, GIF
Archivage – compression	RAR	ZIP, TAR
Présentation	PPT	ODP