

# La minute du PSIR

13 octobre 2021 - #36

## Archivage pérenne

Cette semaine la Minute du PSIR fait le point sur la notion d'archivage pérenne. Ou bien comment conserver ses données publiques le plus longtemps possible en préparant le terrain pour qu'elles puissent être portées dans le temps et restent accessibles le plus longtemps possible.

Il s'agit de s'assurer que les données numériques produites par les projets de recherche resteront accessibles, lisibles et intègres sur une longue période (**30 ans et plus**). Cette démarche est devenue une obligation pour obtenir des financements de type ANR, ERC, Coll-Ex Persée, etc. depuis quelques années. Elle repose sur la définition et le suivi d'un plan de gestion de la donnée ou DMP (Data Management Plan) qui liste les stratégies de gestion et préservation des données pendant et après le projet scientifique.

Les problèmes qu'il est nécessaire de surmonter pour que les données de recherche puissent être qualifiées pour l'archivage :

- **Le choix des périphériques de stockage**, (les CD pourtant privilégiés il y a quelques dizaines d'années se sont avérés un mauvais candidat avec le temps) ;
- **Les formats de fichiers**, (choix de formats ouverts, qui ne nécessitent pas l'achat de logiciel propriétaire dont on ne peut préjuger de leur longévité) ;
- Et également **le manque d'informations** décrivant le set de données archivées ne permettant plus de comprendre le contenu hors contexte au bout de quelques années.

Notre partenaire, le CINES, est un bon candidat pour archiver les données. Il a obtenu l'agrément auprès du Service Interministériel des Archives de France (SIAF) en 2010 et a été renouvelé pour la dernière fois en 2020. Le CINES peut ainsi conserver, en tant que tiers-archivageur, des données publiques définitives. Il a un engagement de migration des formats des fichiers dont il a pris la charge si cela s'avérait nécessaire pour s'assurer de la lisibilité des données au sein des 30 ans sur lesquels il s'engage.

Il propose une procédure de versement au sein d'une chaîne qui peut notamment passer par l'infrastructure Huma-Num qui s'occupe de vérifier la conformité des jeux de données avant envoi au CINES en fonction des obligations de formalisation des données (formats de fichiers, documentation suffisante...).

Une interface de vérification, et une Api, de validation des fichiers pour l'archivage est disponible pour préparer votre documentation : <https://facile.cines.fr>

Dans un premier temps, il est également possible de déposer un ensemble de données le « mieux documentée » possible respectant les formats préconisés, sur une plateforme de type Zenodo gérée par le CERN (sur un engagement de 20 ans) qui permet d'assurer l'accessibilité et l'intégrité à long terme.

Plus d'informations : <https://www.cines.fr/archivage/un-concept-des-problematiques/le-concept-darchivage-numerique-perenne/>

NIVEAU DE  
DIFFICULTÉ



Données

PERSONNES  
CONCERNÉES

