



Trier et conserver ses données

Cellule Data Grenoble Alpes – mai 2022

Pourquoi trier ?



Cycle de vie des données, Recherche Data Gov

- Lisibilité de l'arborescence.
- Facilité pour retrouver une donnée.
- Aspect éthique : prendre le temps de réfléchir à sa recherche

<http://slow-science.org/>

- Impact environnemental : limiter le volume de données.
- Contraintes juridiques.





- Pendant le projet :
 - de l'ordre d'une fois pas an,
 - à la fin d'un livrable du projet.
- À la fin du projet.
- Après le projet : tous les 5 ans.

- **Rappels :**
 - mettre à jour le PGD au besoin,
 - vérifier la complétude des métadonnées avant la suppression de données,
 - vérifier les aspects juridiques.



- **Trier dès le départ** : ne pas stocker les données non nécessaires au projet de recherche
- Quel est le but du projet ? Quelles sont les données nécessaires ?
- Ne pas stocker car disponible. Positionnement inverse de certaines domaines liés aux "big data".



- Est-ce que ces données existent déjà ?
Attention à la pérennité de la source des données.
- ✗ Si oui, ne pas les garder.
- Sinon, est-ce que ces données sont uniques ?
 - ✓ Si oui, les garder.
 - Sinon, déterminer s'il vaut mieux stocker les données ou le processus les ayant fournies (par exemple simulation vs. code source).



- Est-ce que les données nécessitent un pré-traitement ?
- Si oui, est-ce que celui-ci est coûteux en temps ou en ressources ?
- Si oui, déterminer s'il faut garder les données pré-traitées et/ou les données brutes.
- ✓ Sinon, garder simplement le processus de pré-traitement.



- Est-ce que l'analyse des données est coûteuse en temps ou en ressources ?
 - ✓ Si oui, déterminer la durée de conservation.
 - ✓ Sinon, garder simplement le processus d'analyse.

Besoin de reproductibilité !



- Tri par finalité :

Tests et exploration

- *données tièdes*
- *à curer*
- *à supprimer*

Diffusion

- *données froides*
- *curées*
- *temps court*

Article / Data paper

- *description approfondie*
- *temps moyen ou long*

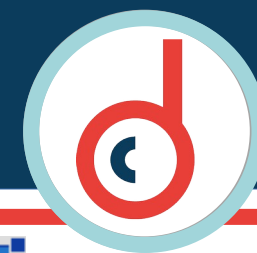
Archivage

- *intérêt historique*
- *temps très long*



Remarques :

- durée de conservation à adapter :
 - à chaque domaine scientifique,
 - à la fréquence d'accès : intérêt scientifique et réutilisation,
- conserver les métadonnées et la documentation associées aux données supprimées.



Grille d'analyse pour identifier les durées applicables

Contexte

- Le document, le dossier ou l'applicatif contient-il des données personnelles ?
- Quel est le traitement concerné ? Quel est l'objectif (finalité) de ce traitement ?
- Quels sont les éléments qui peuvent être utiles à la détermination de la durée ?
P. ex : présence de données sensibles, environnement juridique ou réglementaire applicable, etc.
- Les données relèvent-elles du régime des archives publiques, c'est-à-dire des archives produites par une structure publique ou une structure privée chargée d'une mission de service public ?



Utilisation courante

- Un texte (législatif ou réglementaire) impose-t-il une durée pour ce traitement ?
 - **Si oui** : quel est le périmètre de cette obligation ? Quelles sont les données personnelles concernées par cette obligation ? Pendant combien de temps doivent-elles être conservées en base active ? Est-ce une durée minimale ou maximale ? S'agit-il d'une obligation d'effacement ou de conservation ?
 - **Si non** : jusqu'à quand les données sont-elles nécessaires pour atteindre l'objectif (finalité) fixé ? Existe-t-il des préconisations de la CNIL (voir référentiels de durées) ou des recommandations sectorielles pour ce traitement ?

Archivage intermédiaire

- A l'issue de la période d'utilisation courante, un texte législatif ou réglementaire impose-t-il d'archiver ces données ?
 - **Si oui** : quelles sont les données personnelles ou catégories de données concernées par cette obligation d'archivage ? Quelle est la durée prévue par le texte ?
 - **Si non** : les données présentent-elles un intérêt administratif ou juridique (ex : se prémunir d'un contentieux spécifique) ?
 - **Si non** : procéder à l'anonymisation ou à la destruction de manière sécurisée de toutes les données (pour les archives publiques : après l'autorisation de la personne chargée du contrôle scientifique et technique sur les archives).
 - **Si oui** : quelle est la durée à retenir en fonction de l'objectif de cet archivage et des modalités du traitement ? Quelles sont les données pertinentes par rapport à l'objectif justifiant cet archivage ?
- Pendant cette période, quelles seront les personnes spécifiquement habilitées à accéder aux données archivées en base intermédiaire ?

Archivage définitif

- À l'issue de l'archivage intermédiaire, les données présentent-elles un intérêt justifiant qu'elles ne fassent l'objet d'aucune destruction conformément aux dispositions du code du patrimoine ? Quelles sont les données concernées par cet archivage définitif ?
 - **Si oui** : se rapprocher du service d'archives territorialement compétent pour le versement obligatoire.
 - **Si non** : destruction des données après l'obtention de l'autorisation de la personne chargée du contrôle scientifique et technique sur les archives publiques.





- **En groupe :**

- Dans quelle catégorie (exploration, diffusion, etc.) va chaque type de données manipulées ? En particulier, avez-vous des données à archiver ?
- Déterminer les durées de conservation des données : à diffuser, associées à un article, et à archiver.

Restitution :

- Avez-vous des données à archiver ?
- Quelle durée principale de conservation a été retenue ? Pourquoi ?