

Script de la ressource : Le coût de la gestion des données

<https://doranum.fr/enjeux-benefices/le-cout-de-la-gestion-des-donnees/>



DONNÉES COLLECTÉES ET PRODUITES
Acquisition, transcription...

1 **Acquisition** : prévoyez-vous d'utiliser une base de données existante ? Si oui, est-elle payante ?
Exemple : une licence sur une base de données pour les analyses macro-économiques : 18.000€/an

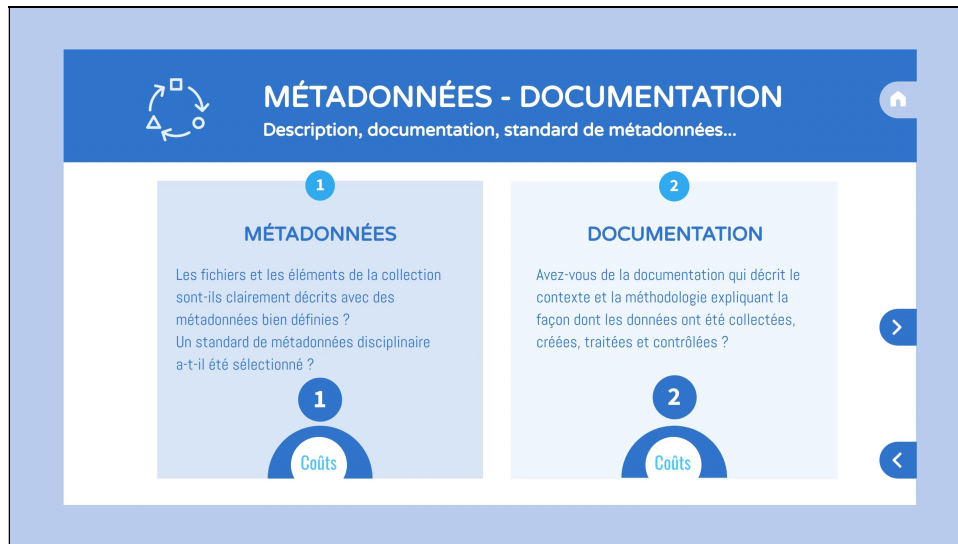
2 **Données produites** :

- vont-elles être dans un format ou modèle uniforme ?
- les éléments de vos données sont-ils clairement nommés et bien organisés ?
- des logiciels, matériel et formation sont-ils nécessaires ?

 La mise en forme et l'organisation des fichiers peuvent être confiées à un assistant étudiant (~18 à 20€/h) ou un gestionnaire de données (~80€/h)

3 une transcription est-elle nécessaire (ex. : entretiens enregistrés, interviews...) ? Si oui, devez-vous élaborer un modèle de transcription ou des lignes directrices pour assurer une cohérence du formatage ? une traduction est-elle nécessaire ?
 Si vous intégrez la transcription dans le cadre de votre pratique de recherche, des frais supplémentaires très faibles ou nuls s'appliqueront.
 Exemple : le temps nécessaire à la transcription est de 4 à 8 heures par heure d'enregistrement.
 Consulter [le calculateur de transcription](#) pour estimer le temps nécessaire à votre projet.

- **Acquisition** : prévoyez-vous d'utiliser une base de données existante ? Si oui, est-elle payante ?
(1) votre documentaliste peut vous aider à l'acquérir. Dans les entrepôts, les données peuvent être disponibles gratuitement ou à faible coût.
- **Données produites** :
 - vont-elles être dans un format ou modèle uniforme ? les éléments de vos données sont-ils clairement nommés et bien organisés ?
 - des logiciels, matériel et formation sont-ils nécessaires ?
 - (2)** vous pouvez prévoir à l'avance des modèles ou formulaires de saisie des données. Cela engendre un gain de temps et un coût moindre que si cette action est faite ultérieurement.
 - une transcription est-elle nécessaire (ex. : entretiens enregistrés, interviews...) ? Si oui, devez-vous élaborer un modèle de transcription ou des lignes directrices pour assurer une cohérence du formatage ? une traduction est-elle nécessaire ?
 - (3)** tenir compte du temps nécessaire pour élaborer des procédures, des modèles et pour les transcrire. Des prestations payantes existent pour retranscrire votre audio automatiquement et en toute sécurité, ce qui constitue un gain de temps.



Métadonnées : Les fichiers et les éléments de la collection sont-ils clairement décrits avec des métadonnées bien définies ? Un standard de métadonnées disciplinaire a-t-il été sélectionné ?

(1) A savoir : les logiciels utilisés pour les données créent automatiquement des métadonnées facilement exportables.

Exemples : pour les documents (date, auteur, poids...), pour les photos (appareil, localisation...).

Coûts : si la description des données est effectuée dans le cadre de la création de données, de la saisie ou de la transcription des données, le coût est faible voir nul. Si elles sont ajoutées ou harmonisées par la suite, le coût est plus élevé.

Coûts : si la description des données est effectuée dans le cadre de la création de données, de la saisie ou de la transcription des données, le coût est faible voir nul. Si elles sont ajoutées ou harmonisées par la suite, le coût est plus élevé.

Exemples :

- 4 heures (120 mesures) pour remplir 60 champs de métadonnées obligatoires, avec l'aide d'un gestionnaire de données (~80 €/heure).

- deux à trois semaines sont estimées dans une demande de subvention pour un projet de recherche de deux ans afin de préparer et rassembler les documents à déposer.

Source : [UK Data Service](#)

Documentation : Avez-vous de la documentation qui décrit le contexte et la méthodologie expliquant la façon dont les données ont été collectées, créées, traitées et contrôlées ?

(2) La documentation contextuelle et méthodologique est le plus souvent rédigée dans des publications et/ou data papers.

Coûts : si toutes les étapes de création des données sont bien documentées et que la documentation est bien organisée pendant la recherche, le coût est faible.

Si la documentation doit être écrite ou compilée spécifiquement par la suite, le coût est élevé. Cette activité peut être confiée à un gestionnaire de données (~80 €/heure).

STOCKAGE - SAUVEGARDE
Sécurité, accès, transfert

1. Durant le projet, disposez-vous d'un serveur sécurisé de votre institution ou partenaire(s) ?
2. Quel volume de stockage est nécessaire pour toute la durée du projet ?
3. Des personnes externes ont-elles besoin d'accéder aux données de recherche ?
4. Les données sont-elles protégées contre tout accès ou utilisation non autorisé ?

Coûts

Les coûts d'une sauvegarde institutionnelle sont considérés comme des frais généraux ou coûts indirects. En cas de sauvegarde supplémentaire, il est nécessaire d'évaluer le nombre de copies à conserver, la fréquence des sauvegardes et les supports de stockage.

Exemples :

- [Université de La Rochelle](#) : stockage gratuit jusque 500Go - stockage fichier 13 € - sauvegarde historisée 30 € (option). Le seuil minimal de location : 500€
- [Université de Lorraine](#) : stockage jusque 30To

Durant le projet, disposez-vous d'un serveur sécurisé de votre institution ou partenaire(s) ?

Quel volume de stockage est nécessaire pour toute la durée du projet ?

Des personnes externes ont-elles besoin d'accéder aux données de recherche ?

Les données sont-elles protégées contre tout accès ou utilisation non autorisé ?

Coûts : les coûts d'une sauvegarde institutionnelle sont considérés comme des frais généraux ou coûts indirects. En cas de sauvegarde supplémentaire, il est nécessaire d'évaluer le nombre de copies à conserver, la fréquence des sauvegardes et les supports de stockage.

Exemples :

- [Université de La Rochelle](#) : stockage gratuit jusque 500Go - stockage fichier 13 € - sauvegarde historisée 30 € (option). Le seuil minimal de location : 500€

<https://services-numeriques.univ-larochelle.fr/Stockage-de-donnees-scientifiques>

- [Université de Lorraine](#) : stockage jusque 30To

<https://numerique.univ-lorraine.fr/catalogue-des-services/stockage-prive-filer>

CONTRÔLE - QUALITÉ - RESPONSABILITÉ
Responsable de la gestion des données, temps alloué...

Coûts

1. Qui sera responsable de la création et mise à jour du Plan de Gestion des Données ?
2. Qui sera responsable du contrôle de la qualité des jeux de données (qualité des métadonnées, données anonymisées...)?
3. Qui détient les droits d'auteurs ? Des conseils juridiques sont-ils nécessaires ?

Le coût est nul si ces activités sont effectuées par les membres du projet. Si ce n'est pas le cas, il faut évaluer le coût de l'aide d'un gestionnaire de données (~80 €/heure)

Pour un conseil juridique, le coût est de ~180 €/heure

- Qui sera responsable de la création et mise à jour du Plan de Gestion des Données ?
- Qui sera responsable du contrôle de la qualité des jeux de données (qualité des métadonnées, données anonymisées...)?
- Qui détient les droits d'auteurs ? Des conseils juridiques sont-ils nécessaires ?

Coûts : Le coût est nul si ces activités sont effectuées par les membres du projet. Si ce n'est pas le cas, il faut évaluer le coût de l'aide d'un gestionnaire de données (~80 €/heure)
Pour un conseil juridique, le coût est de ~180 €/heure



Déposer des données de recherche demande d'être vigilant sur plusieurs critères. Un travail de gestion et de préparation des données est nécessaire pour pouvoir effectuer le dépôt.

Les jeux de données à partager ont-ils été sélectionnés et nettoyés ?

Des données devront-elles être numérisées pour favoriser le partage ?

Les fichiers de données sont-ils dans un format standard ou ouvert ?

Les principes juridiques et éthiques sont-ils respectés ?

Le volume des fichiers dépasse-t-il la limite autorisée par l'entrepôt ?

Coûts :

Les jeux de données à partager ont-ils été sélectionnés et nettoyés ? Le nettoyage des données prend du temps. S'il est effectué lors de la saisie et de la préparation des données, le coût est faible. Par contre, le coût est élevé s'il est fait a posteriori.

Exemple : le nettoyage peut varier entre 270€ à plus de 1 800€ ou s'il est confié à un gestionnaire de données au taux horaire de ~80 €.

Des données devront-elles être numérisées pour favoriser le partage ? Certaines données comme les cartes, images, photographies... doivent être numérisées pour accroître le partage. Pour évaluer le coût :

- estimer si un logiciel et des équipements sont nécessaires,
- définir le type de numérisation (coût faible si les documents sont composés de texte et d'image, coût élevé pour une reconnaissance optique des caractères auquel il faut ajouter le coût de la relecture, coût élevé aussi pour la saisie de données en cas de données tabulaires).

Les fichiers de données sont-ils dans un format standard ou ouvert ? Des logiciels ou matériel supplémentaires sont-ils nécessaires pour la conversion ? Pour les données audiovisuelles, la conversion aux formats numériques ouverts peut prendre du temps et nécessite un équipement et/ou un logiciel spécial(aux). Pour les bases de données, les conversions peuvent nécessiter la vérification des troncatures, la perte des métadonnées... La conversion des formats peut être confiée à un gestionnaire de données au taux horaire de ~80 €.

Les principes juridiques et éthiques sont-ils respectés ? Si les données nécessitent une anonymisation pour des raisons éthiques comme les données médicales, consulter la rubrique "Cas particuliers" de cette ressource.

Le volume des fichiers dépasse-t-il la limite autorisée par l'entrepôt ? En fonction du volume des fichiers, vérifier les conditions tarifaires ou la taille maximale des fichiers autorisée pratiquées par l'entrepôt.

Exemple : Dryad, 20 \$ par soumission (pour un volume inférieur à 50 Go). Au-delà, 50 \$ par tranche supplémentaire de 10Go (cf. [conditions sur le site Dryad](#)).

Anonymisation

L'anonymisation doit être cohérente tout au long de la collecte des données.

A noter : les données anonymisées peuvent être conservées sans limitation de durée, car elles ne sont plus soumises au RGPD ([Règlement Général sur la Protection des Données de la Protection des Données de la CNIL](#)).

Par contre, les données pseudonymisées restent soumises aux respects du RGPD.

Coûts : le **coût dépend de la sensibilité ou de la complexité des données** et de la quantité d'informations d'identification enregistrées dans les données – si seule la suppression des noms est nécessaire, le coût est faible ; la pseudonymisation nécessitera plus de temps.

Si l'anonymisation est prévue **avant la collecte des données**, le coût peut être réduit.

1. Pour les **données audiovisuelles** – l'anonymisation (voix ou et/ou visages) peut être **très coûteuse** et pourrait réduire l'utilité des données.
2. Pour les **données quantitatives** (par exemple les données d'enquête), le coût est **faible** si les identificateurs sont a priori exclus des fichiers de données, faciles à supprimer ou si des variables identifiables sont codées pour éviter la divulgation ; le coût peut être plus élevé si les variables doivent être restructurées par la suite pour éviter la divulgation.
3. Pour les **données textuelles qualitatives** (par exemple, les transcriptions d'entretien), les coûts peuvent être réduits si l'anonymisation est effectuée pendant la transcription (ou au moins mise en évidence/codée pendant la transcription).

Des logiciels gratuits sont disponibles (voir la rubrique "Outils / Aides" de ce tutoriel).

Exemple : transcrire / anonymiser simultanément l'audio (parole) jusqu'à une heure par fragment de 5 minutes (selon le niveau de précision de la transcription). Ce travail peut être confié à un assistant étudiant (~18 à 20€/h).

Conservation à long terme

Les coûts de conservation ont une incidence sur les décisions prises à ce niveau.

En effet, la conservation à long terme des données (au-delà de 30 ans) implique des investissements importants, tant dans le stockage que dans les mesures déployées pour assurer la lisibilité des formats dans le futur pour compenser l'évolution technologique. Cette conservation ne concerne que les données à haute valeur scientifique et doit être décidée dans le cadre de la politique de votre établissement.

Coûts : le coût à long terme est variable en fonction des services souhaités mais aussi en fonction des prestataires. La facturation peut être mensuelle ou annuelle, par nombre de documents ou par taille des fichiers.

Exemples :

- à partir de 20 centimes pour un document conservé par tranche de 10 ans
- coût par Go ou To allant de 3 000 à 7 000 €/an.

OUTILS / AIDES
Liens, outils, aides...

LIENS

CAT OPIDoR
Annuaire des services dédiés aux données de la recherche (plus de 700 services français) : entrepôt, conservation à long terme, services d'accompagnement...

CINES : opérateur pour l'archivage à long terme des données et documents numériques produits par la communauté de l'Enseignement Supérieur et de la Recherche française (ESR).

OUTILS GRATUITS

Pour anonymiser les données, l'application **Amnesia** (OpenAIRE)

Outils pour calculer les coûts :

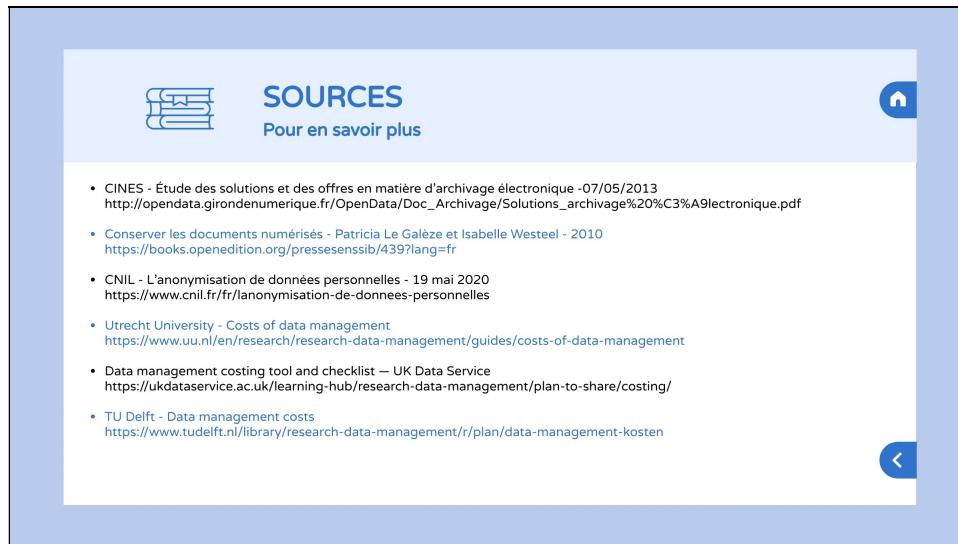
- Cost Calculator for Data Management - EPFL Library
- Management Tool - Delft University

AIDES

Pour évaluer les coûts, n'hésitez pas à consulter un informaticien (stockage et sécurité), les professionnels de l'Information Scientifique et Technique (IST), notamment les documentalistes (choix des standards de métadonnées, de l'entrepôt), un archiviste (aide pour la sélection des données à conserver), un service juridique (droit d'auteur), le délégué à la protection des données (RGPD)...

- **Liens**
[CAT OPIDoR](#) : Annuaire des services dédiés aux données de la recherche (plus de 700 services français) : entrepôt, conservation à long terme, services d'accompagnement...
[CINES](#) : opérateur pour l'archivage à long terme des données et documents numériques produits par la communauté de l'Enseignement Supérieur et de la Recherche française (ESR).
- **Outils gratuits**
 Pour anonymiser les données, l'application [Amnesia](#) (OpenAIRE)

 Outils pour calculer les coûts :
 - [Cost Calculator for Data Management](#) - EPFL Library
 - [Management Tool](#) - Delft University
- **Aides**
 Pour évaluer les coûts, n'hésitez pas à consulter un informaticien (stockage et sécurité), les professionnels de l'Information Scientifique et Technique (IST), notamment les documentalistes (choix des standards de métadonnées, de l'entrepôt), un archiviste (aide pour la sélection des données, à conserver), un service juridique (droit d'auteur), le délégué à la protection des données (RGPD).



- CINES - [Étude des solutions et des offres en matière d'archivage électronique](http://opendata.girondenumerique.fr/OpenData/Doc_Archivage/Solutions_archivage%20%C3%A9lectronique.pdf) - 07/05/2013
http://opendata.girondenumerique.fr/OpenData/Doc_Archivage/Solutions_archivage%20%C3%A9lectronique.pdf
- [Conserver les documents numérisés](https://books.openedition.org/pressesenssib/439?lang=fr) - Patricia Le Galèze et Isabelle Westeel - 2010
<https://books.openedition.org/pressesenssib/439?lang=fr>
- CNIL - [L'anonymisation de données personnelles](https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles) - 19 mai 2020
<https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles>
- Utrecht University - [Costs of data management](https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management)
<https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management>
- [Data management costing tool and checklist](https://ukdataservice.ac.uk/learning-hub/research-data-management/plan-to-share/costing/) - UK Data Service
<https://ukdataservice.ac.uk/learning-hub/research-data-management/plan-to-share/costing/>
- TU Delft - [Data management costs](https://www.tudelft.nl/library/research-data-management/r/plan/data-management-kosten)
<https://www.tudelft.nl/library/research-data-management/r/plan/data-management-kosten>