

Données de la recherche : s'initier au plan de gestion de données

Emmanuel Collier

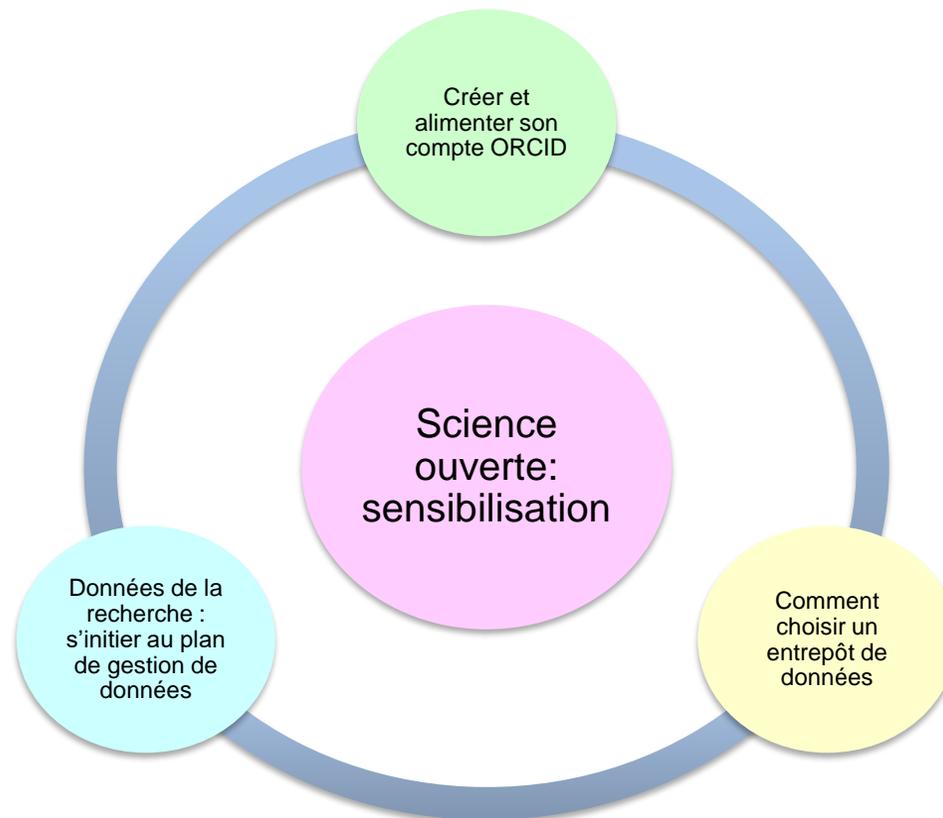
Humathèque Condorcet – Service formations des usagers

2023



Parcours de formations

**Science
ouverte**



Sommaire

- 1) Le plan de gestion de données (PGD)
- 2) Les attendus d'un plan de gestion de données
- 3) Des ressources pour commencer la rédaction d'un plan de gestion de données

Qu'est-ce-qu'un plan de gestion de données ?

Des données de recherche

➤ **Sources => fichiers informatiques => données de recherche**

Données qualitatives / Données quantitatives

- Notes de terrain
- Photos, enregistrements audio ou vidéo
- Questionnaires, transcriptions d'entretiens, notes d'entretiens, réponses à des tests, enregistrements de groupes
- Logiciels, programmes, codes source
- Données textuelles structurées (tableaux...)
- Texte brut
- Références bibliographiques
- [...]

➤ **Tout objet dont se saisit le chercheur à des fins de recherche**

Définition : [OCDE](#), 2007

Qu'est-ce-qu'un Plan de gestion de données ?

La gestion des données de recherche

Une pratique intrinsèque au processus de recherche

- Produire, collecter, réutiliser des données
- Traiter et analyser des données
- Stocker des données
- [...]

La gestion des données de recherche fait référence aux processus appliqués tout au long d'un projet de recherche concernant la collecte, la documentation, le stockage, le partage et la préservation des données de recherche.

Qu'est-ce-qu'un Plan de gestion de données ?

Un document

- **de formalisation** : des pratiques en matière de gestion des données de recherche
- **de synthèse** : décrit, pour les [jeux de données](#) produits ou réutilisés au cours du projet, comment ceux-ci seront gérés, depuis leur création ou collecte, jusqu'à leur partage et éventuellement leur archivage.
- **obligatoire** : dans le cadre des projets financés sur AAP et plus récemment pour toute recherche financée sur fonds publics (cf. [Décret n° 2021-1572 du 3 décembre 2021](#))
- **évolutif** : pour les projets financés par AAP : différentes versions (3 si projet > 30 mois). Pour les autres projets de recherche, il suit l'évolution du projet et celle de la gestion des données.
- **dont la rédaction est libre** au sein de modèles proposés.

Pourquoi rédiger un Plan de gestion de données ?

Anticiper, Identifier, Prévoir, Clarifier, Réunir, Planifier, Formaliser

- **Anticiper** d'éventuels questions juridiques, éthiques liées aux données de recherche
- **Identifier des solutions techniques** pour sécuriser les données et faciliter leur partage au sein d'une équipe de recherche
- **Prévoir et budgéter** les coûts liés à la gestion des données
- **Clarifier les rôles**, responsabilités et droits de chaque contributeur de données et acteur impliqué dans la gestion des données
- **Planifier et formaliser** la gestion des données du projet
- **Gagner du temps** plutôt que de travailler dans l'urgence
- **Identifier et réunir les acteurs et compétences** utiles pour la gestion des données du projet

Le Plan de gestion de données

Acteurs et compétences utiles pour la rédaction du plan de gestion de données

- **Chercheur** : coordinateur du PGD ; réutilisation, collecte, analyse, traitement des données
- **Ingénieur** : réutilisation, collecte, traitement, analyse, documentation des données
- **Informaticien** : sécurisation et stockage des données au cours du projet
- **Délégué à la protection des données (DPD)** : données à caractère personnel
Juriste : droits sur les données
- **Professionnel IST** (bibliothécaire, documentaliste) : entrepôts de données, standards et schémas de métadonnées, référentiels
- **Archiviste** : plan de conservation des données, solutions techniques d'archivage, coûts associés

Le Plan de gestion de données

Mouvement pour la science ouverte et demandes des financeurs

« La **science ouverte** est la **diffusion** sans entrave des **résultats**, des **méthodes** et des **produits de la recherche scientifique**. Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'**accès ouvert aux publications** et – autant que possible – **aux données, aux codes sources et aux méthodes de la recherche**. »

[Deuxième Plan national pour la science ouverte, 2021](#)

Dates clés

2017 : ERC : PGD obligatoire pour les projets financés par AAP

2019 : ANR : PGD obligatoire pour les projets financés par AAP

2021 : [Art 6, décret n° 2021-1572 du 3 décembre 2021](#)

Les établissements publics [...] **veillent à la mise en œuvre par leur personnel de plans de gestion de données** et contribue aux infrastructures qui permettent la conservation, la communication et la réutilisation des données et des codes sources.

Plan de gestion de données : attendus

1. Responsabilités et ressources en matière de gestion des données
2. Description des données
3. Stockage et sauvegarde des données pendant le processus de recherche
4. Documentation et qualité des données
5. Dépôt, accès et partage des données
6. Exigences légales et éthiques
7. Conservation à long terme

[Plan de gestion de données, modèle ANR](#)
[Grille de relecture de PGD – Modèle ANR \(Inist-CNRS\)](#)

Plan de gestion de données : attendus

- 1. Responsabilités et ressources en matière de gestion des données**
2. Description des données
3. Stockage et sauvegarde des données pendant le processus de recherche
4. Documentation et qualité des données
5. Dépôt, accès et partage des données
6. Exigences légales et éthiques
7. Conservation à long terme

Plan de gestion de données : attendus

1. Responsabilités et ressources en matière de gestion des données

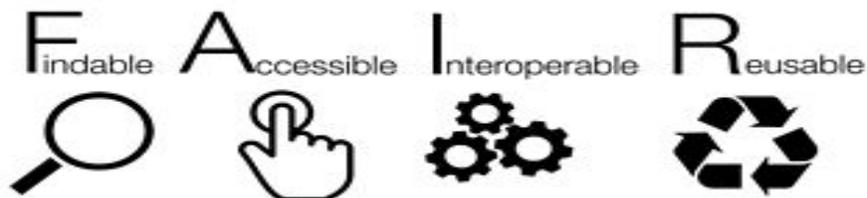
Coordonner la gestion des données

Définir les ressources (budget et temps alloués) dédiées à la gestion des données et permettant de s'assurer que les données seront FAIR.

ANR : les dépenses liées à la gestion des données sont éligibles durant le projet : acquisition, collecte, stockage, gestion des données [...]

➤ **Personne(s) ressource(s) : chercheur, coordinateur de la gestion des données**

Coûts dédiés à la gestion des données peuvent être portés par les agences de financements



Facile à trouver
Accessible
Interopérable
Réutilisable

Les principes FAIR – DoRANum. https://doranum.fr/enjeux-benefices/principes-fair_10_13143_z7s6-ed26/

Plan de gestion de données : attendus

1. Responsabilités et ressources en matière de gestion des données
- 2. Description des données**
3. Stockage et sauvegarde des données pendant le processus de recherche
4. Documentation et qualité des données
5. Dépôt, accès et partage des données
6. Exigences légales et éthiques
7. Conservation à long terme

2. Description des données

Décrire les données du projet : collectées/produites, réutilisées

➤ **Personnes ressources : chercheur, ingénieur**

- **Provenance des jeux de données** : collecte de nouvelles données, réutilisation de données, conversion ou transformation des données, partage/échange de données, achat de données, etc.
- **Méthodologie de collecte, traitements effectués, logiciels utilisés** pour produire et traiter les données
- **Nature des données** (types, formats, volumes)

Cas des données à caractère personnel

« Toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres »

- **Le RGPD** (2016 : CE et Parlement européen ; 2018 : applicable dans tous les États membres)

➤ **Personne ressource : Délégué à la protection des données (DPD)**

Plan de gestion de données : attendus

1. Responsabilités et ressources en matière de gestion des données
2. Description des données
- 3. Stockage et sauvegarde des données pendant le processus de recherche**
4. Documentation et qualité des données
5. Dépôt, accès et partage des données
6. Exigences légales et éthiques
7. Conservation à long terme

STOCKAGE ET SAUVEGARDE



Serveur sécurisé
de l'institution

Espace collaboratif
de travail



Le chercheur



L'équipe de
recherche

3. Stockage et sauvegarde des données pendant le processus de recherche

Protéger les données et assurer leur sécurité

➤ **Personne ressource** : informaticien (service informatique)

Stockage sécurisé : [ShareDocs \(Huma-Num\)](#)

Envoi sécurisé de données : [FileSender](#) de RENATER

Doranum (2021). Stockage, partage et archivage : quelles différences ? DOI : 10.13143/5dax-qp58/

Plan de gestion de données : attendus

1. Responsabilités et ressources en matière de gestion des données
2. Description des données
3. Stockage et sauvegarde des données pendant le processus de recherche
- 4. Documentation et qualité des données**
5. Dépôt, accès et partage des données
6. Exigences légales et éthiques
7. Conservation à long terme

Plan de gestion de données : attendus

4. Documentation et qualité des données

- Contextualiser les jeux de données : la documentation des données

Ex : **fichier** texte de type « **lisez-moi** », dans les en-têtes des jeux de données et fichiers
=> Méthodologies utilisées : collecte des données, procédures et méthodes d'analyse utilisées, définition des variables, des unités de mesure, etc.

- Assurer la qualité des données

Axes :

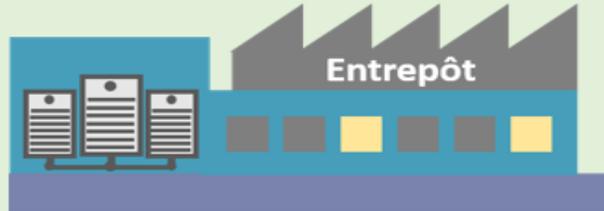
- Formats de fichiers : adaptés aux utilisateurs potentiels
- Structure des fichiers : noms de propriété explicite, compréhensibles et identifiables
=> Convention de nommage
- Si nécessaire, procédures de contrôle : validation de saisie des données, contrôle par des experts, utilisation d'outils de vérification...

➤ **Personnes ressources** : chercheur, ingénieur, professionnel IST

Plan de gestion de données : attendus

1. Responsabilités et ressources en matière de gestion des données
2. Description des données
3. Stockage et sauvegarde des données pendant le processus de recherche
4. Documentation et qualité des données
- 5. Dépôt, accès et partage des données**
6. Exigences légales et éthiques
7. Conservation à long terme

PARTAGE



Entrepôt généraliste
ou disciplinaire



L'équipe de
recherche



Autre équipe
de recherche
(même domaine)



Autre équipe de recherche
(autre domaine)

5. Dépôt, accès et partage des données

Déposer les données de recherche

- Sélection des jeux de données
- Choix d'un ou de plusieurs entrepôts de données

Décrire les jeux de données à l'aide de métadonnées : renforcer leur « découvrabilité »

- Standards de métadonnées (DDI, DataCite, DC...)
- Référentiels (ex : vocabulaires contrôlés)

Définir les conditions d'accès et de partage des données

- Libre-accès
- Embargo, restriction d'accès, accès conditionné
- Licences de diffusion/réutilisation (CC)
- Identifiant unique et pérenne pour chaque jeu de données (DOI, ...)

➤ **Personnes ressources** : chercheur, professionnel IST, juriste

Doranum (2021). Stockage, partage et archivage : quelles différences ? DOI : 10.13143/5dax-qp58/

Plan de gestion de données : attendus

1. Responsabilités et ressources en matière de gestion des données
2. Description des données
3. Stockage et sauvegarde des données pendant le processus de recherche
4. Documentation et qualité des données
5. Dépôt, accès et partage des données
- 6. Exigences légales et éthiques**
7. Conservation à long terme

Plan de gestion de données : attendus

6. Exigences légales et éthiques, codes de conduite

"Aussi ouvert que possible, aussi fermé que nécessaire"

Définir les conditions d'accès aux données :

- **Données à caractère personnel et RGPD** (cf. Description des données)

Accès aux données : respecter les mesures pour assurer la confidentialité des données (anonymisation, pseudonymisation) ; procédures d'accès conditionné s'il y a lieu.

CNIL : Analyse d'impact relative à la protection des données ([AIPD](#))

- **Cadre juridique :**

Licences de diffusion et de réutilisation des données (Creative Commons, Etalab, etc.)

Droits et propriété des données

➤ **Personnes ressources : DPD, juriste**

Aspects éthiques

➤ **Personnes ressources : membres du comité d'éthique** (Comité d'éthique de la recherche => [Fédération des Comités d'Éthique de la Recherche](#))

Plan de gestion de données : attendus

1. Description des données et collecte ou réutilisation de données existantes
2. Stockage et sauvegarde pendant le processus de recherche
3. Dépôt, accès et partage des données
4. Documentation et qualité des données
5. Exigences légales et éthiques, codes de conduite
6. Responsabilités et ressources en matière de gestion des données
- 7. Conservation à long terme**

ARCHIVAGE PÉRENNE



Plateforme d'archivage
(ex : CINES)



L'équipe de
recherche



Autre équipe
de recherche
(même domaine)



Autre équipe de recherche
(autre domaine)

7. Conservation à long terme

Établir un plan de conservation des données

➤ **Personne ressource : Archiviste**

Doranum (2021). Stockage, partage et archivage :
quelles différences ? DOI : 10.13143/5dax-qp58/

Exemples de PGD

- PGD du projet ANR "Decolonizing Southeast Asian Sound Archives" -
https://dmp.opidor.fr/plans/10610/export.pdf?export%5Bquestion_headings%5D=true
- PGD du projet "TRANSFUNERAIRE: approche comparative des rituels collectifs de ré-inhumation en Europe et en Amérique-Latine XXe-XXIe siècles "
https://dmp.opidor.fr/plans/6619/export.pdf?export%5Bquestion_headings%5D=true



Durée : 10 minutes

Ressources pour commencer la rédaction d'un PGD

Le service d'accompagnement à la rédaction de plan de gestion de données de l'[Humathèque](#)

Offre de services :

- Entretien(s) avec les responsables de projets
- Relecture de PGD : suggestions, conseils
- Suivi des PGD tout au long des projets :
 - Projets financés par AAP : Version initiale, version à mi- projet (projets > 30 mois), version finale à l'issue du projet
 - Projets de recherche financés sur fonds publics : suivi de l'évolution de la gestion des données du projet
 - Thèses financées

Ressources pour commencer la rédaction d'un PGD

Des plateformes d'apprentissage

- [Dorandum](#) (Données de la recherche : apprentissage numérique)
- [Coopist](#) (Coopérer en information scientifique et technique)

Ressources pour commencer la rédaction d'un PGD

- Des modèles de PGD

- [ANR](#)
- [Modèles dans DMP-OPIDoR](#)

Et

- [Grille de relecture de PGD, modèle ANR \(Inist-CNRS\)](#)

- Des exemples de PGD

- [DMPs publics](#) (DMP-OPIDoR)
- [PGD primés](#) (DMPTool)

- Une plateforme d'aide à la rédaction de PGD

- [DMP-OPIDoR](#)

Conclusion : le plan de gestion de données

Données (jeux de données) – Gestion des données – Planification (PGD)

Anticiper, Identifier, Prévoir, Clarifier, Réunir, Planifier, Formaliser

Science ouverte et demande des financeurs

Attendus du PGD

Ressources pour commencer la rédaction d'un PGD

Suite proposée :

Le service d'accompagnement à la rédaction des PGD

Pour tout renseignement : emmanuel.collier@campus-condorcet.fr



Bibliographie sélective

Plan de gestion de données

ANR. Modèle de plan de gestion de données, 2019. <https://anr.fr/fileadmin/documents/2019/ANR-modele-PGD.pdf>

Cocaud, Sylvie, et Dominique l'Hostis. *Pourquoi et comment rédiger un plan de gestion de données ?* 2019, p. 128 slides. *hal.inrae.fr*, <https://hal.inrae.fr/hal-02791507>

Coso. *Plan de gestion de données – Recommandations à l'ANR*. <https://www.ouvrirlascience.fr/plan-de-gestion-de-donnees-recommandations-a-lanr>

Dedieu, Laurence. *1 - Qu'est ce qu'un PGD / Rédiger un PGD - Coopérer en information scientifique et technique - Cirad*. <https://coop-ist.cirad.fr/gerer-des-donnees/rediger-un-pgd/1-qu-est-ce-qu-un-pgd>

Féret, Romain, et al. *Rédiger un Plan de Gestion de Données (PGD) : supports d'atelier (JNSO)*. Zenodo, 22 novembre 2019. Zenodo, <https://doi.org/10.5281/zenodo.3550572>

INIST-CNRS. Grille de relecture de plans de gestion de données – modèle ANR. <https://doranum.fr/wp-content/uploads/Grille-relecture-PGD-Modele-ANR-V3.pdf>

Reymonet, Nathalie, et al. *Réaliser un plan de gestion de données « FAIR » : modèle*. janvier 2018. https://archivesic.ccsd.cnrs.fr/sic_01690547

Bibliographie sélective

Protection des données - RGPD

CNIL. *RGPD : de quoi parle-t-on ?* <https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on>

Delplanque Catherine, et al. « Guide : Règlement général pour la protection des données ». UPL, <https://www.u-plum.fr/guide-reglement-general-pour-la-protection-des-donnees/>

INSHS. Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte : guide pour la recherche. Version 2, février 2021. <https://www.ouvrirlascience.fr/les-sciences-humaines-et-sociales-et-la-protection-des-donnees-a-caractere-personnel-dans-le-contexte-de-la-science-ouverte-v2/>

Ginouvès Véronique, Gras Isabelle (2018). La diffusion numérique des données en SHS : guide des bonnes pratiques éthiques et juridiques. <https://hal.archives-ouvertes.fr/hal-01903040>

SupDPO (2020). Quinze recommandations aux chercheurs sur la protection des données dans le cadre de leurs activités de recherche. <https://reseau.supdpo.fr/wp-content/uploads/2020/01/SupDPO-Recommandations-chercheurs-v1.pdf>

Bibliographie sélective

Principes FAIR

Wilkinson, Mark D., et al. « The FAIR Guiding Principles for Scientific Data Management and Stewardship ». Scientific Data, vol. 3, n° 1, mars 2016, p. 160018. www.nature.com, <https://doi.org/10.1038/sdata.2016.18>

Coûts liés à la gestion des données

ANR. Règlement financier. <https://anr.fr/fr/rf/>

ANR. Fiche Règlement financier n°3 : les coûts admissibles (dépenses éligibles) – V2 MAJ février 2020. <https://anr.fr/fileadmin/documents/2017/ANR-RF-Fiche-COUTS.pdf>

Bibliographie sélective

Diffusion des données de recherche

Cirad. 2021. Avez-vous le droit ou l'obligation de diffuser vos données ? Parcourez l'arbre de décision du Cirad. Montpellier (FRA) : CIRAD, 7 p. <https://doi.org/10.18167/coopist/0075>

Emilie Cotte, Fanny Sébire. Modèle de logigramme de l'Institut Pasteur relatif aux questions juridiques liées à la diffusion des données de la recherche. 2022. [⟨pasteur-03587216⟩](#)

Véronique Ginouvès, Isabelle Gras. La diffusion numérique des données en SHS - Guide des bonnes pratiques éthiques et juridiques. Véronique Ginouvès; Isabelle Gras. Presses universitaires de Provence, 2018, Digitales, 9791032001790

Anne-Laure Stérin. Diffuser des données de la recherche dans le respect du droit et de l'éthique : Comment faire lorsqu'on n'est pas juriste ?. Véronique Ginouvès; Isabelle Gras. La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques, Presses universitaires de Provence, pp.299-302, 2018, Digitales, 9791032001790. [⟨hal-02050510⟩](#)

Pour tout renseignement :
formations.humatheque@campus-condorcet.fr