

cnrs

Plan de Gestion de Données
Bonnes pratiques de rédaction et outils

2 avril 2024

→ Institut de l'information scientifique et technique
DVDR | Service Formation-DoRANum

VALORISER LES
DONNÉES DE LA
RECHERCHE

Ce webinaire a pour objectif de vous expliquer ce qu'est un Plan de Gestion de Données, à quoi il sert, ce qu'il contient et comment le rédiger.

Sommaire

- 01 Enjeux et objectifs d'un PGD
- 02 Contenu et rédaction
- 03 Présentation rapide de DMP OPIDoR



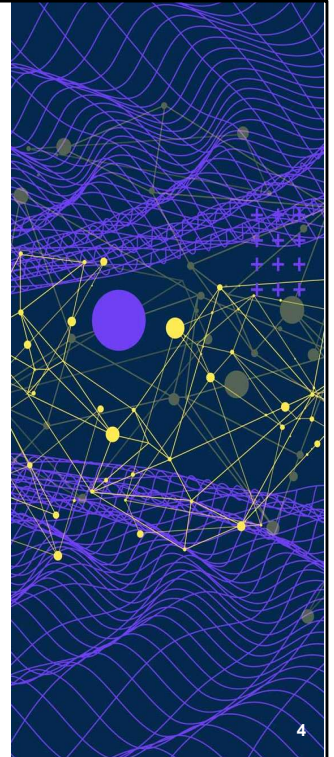
01

Enjeux et objectifs d'un PGD

Le PGD qu'est-ce que c'est ?

PGD (Plan de Gestion des Données) = **DMP** (Data Management Plan)

- Explique **comment sont gérées les données**, depuis leur création ou collecte, jusqu'à leur partage et leur archivage
- Aide à **organiser** et à **anticiper** toutes les étapes du cycle de vie des données
- Est **évolutif** : la rédaction commence dès le début du projet, mis à jour tout au long du projet



- Le DMP ou PGD est un document de quelques pages, qui permet de définir et de mettre en place les actions indispensables pour gérer les données qui vont être collectées ou produites au cours d'un projet.
- Il explique comment sont gérées les données depuis leur création ou collecte jusqu'à leur partage et leur archivage.
- Le DMP est évolutif : sa rédaction démarre dès le début d'un projet, mais il va être mis à jour régulièrement en fonction de l'évolution du projet. Vous n'avez pas à répondre à toutes les questions du DMP dans sa phase initiale mais il est bon de réfléchir à tous les points abordés dans le DMP. Ce document est là pour faciliter la gestion des données de recherche de votre projet.
- Surtout, le DMP aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée.

Ressource :

- *DoRANum. Cours introductif sur le Plan de Gestion de Données (PGD). 27 avril 2022.*
https://doranum.fr/plan-gestion-donnees-dmp/cours-introductif-sur-le-plan-de-gestion-de-donnees_10_13143_t3j4-vn03/

Les bénéfices du PGD



Permet de se poser les bonnes questions pour rendre les données **FAIR** (**F**aciles à trouver, **A**ccessibles, **I**nteropérables, **R**éutilisables)



Aide à évaluer les **ressources nécessaires**



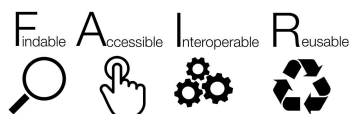
Définit les **responsabilités**

- L'élaboration d'un plan de gestion permet de se poser les bonnes questions afin d'obtenir des données FAIR, c'est-à-dire faciles à trouver, accessibles, interopérables et réutilisables.
- Le DMP va aussi vous aider à réfléchir aux ressources nécessaires afin de mettre en œuvre les actions décrites dans le DMP :
 - temps nécessaire à la préparation des données pour le stockage, le partage et l'archivage des données,
 - coûts de matériel, rémunération des personnels,
 - frais de stockage (serveurs dédiés, traitement, maintenance, sécurité, accès...), partage (site web, publication...) et d'archivage des données.
- Dans le DMP, il faut désigner nominativement la ou les personne(s) responsable(s) de la gestion des données pour toutes les étapes du projet et au sein du partenariat s'il y a lieu :
 - saisie des données,
 - production des métadonnées,
 - contrôle de la qualité des données,
 - mise à jour du DMP.

Principes FAIR

4 principes à respecter pour garantir une utilisation optimale des données de recherche et des métadonnées associées, à la fois **par les hommes et par les machines**

- **F** (Findable) = Facile à trouver
- **A** (Accessible) = Accessible
- **I** (Interoperable) = Interopérable
- **R** (Reusable) = Réutilisable



Applicables tout au long du cycle de vie des données

Principes admis par les différentes communautés scientifiques au niveau international, ainsi que par les financeurs (ex : Commission européenne, ANR, etc.)

Ressource :

DoRANum. Les principes FAIR. 4 décembre 2019. https://doranum.fr/enjeux-benefices/principes-fair_10_13143_z7s6-ed26/

Le PGD : obligations et recommandations

Incitation voire obligation :

- Organismes de recherche / universités



INRAE



- Agences de financement de la recherche



horizon
europe



European Research Council
Established by the European Commission

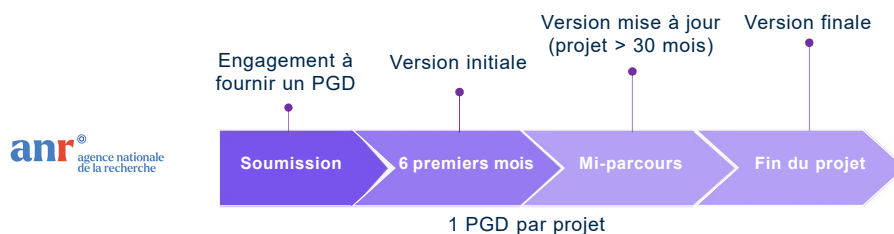
Le PGD répond aussi à une incitation voire à une obligation institutionnelle :

- Certains organismes de recherche et universités comme le CNRS, l'INRAE, l'Institut Pasteur ou l'Université de Montpellier incitent ou encouragent leurs chercheurs à élaborer un PGD en proposant leur propre modèle institutionnel (pour faciliter l'adoption de bonnes pratiques mais aussi permettre de répertorier les données produites au sein de l'institution).
- Pour d'autres, comme les agences de financement de la recherche, le PGD constitue un livrable qui devra être fourni dans les 6 mois qui suivent le démarrage scientifique du projet : ANR, Commission européenne (Horizon Europe), European Research Council...

Le PGD : obligations et recommandations

Exemple de l'ANR

- Demande un PGD pour tous les projets qu'elle finance
- Le PGD est un livrable à fournir dans les 6 mois



Le PGD répond aussi à une incitation voire à une obligation institutionnelle :

- Pour d'autres, comme les agences de financement de la recherche, le PGD constitue un livrable qui devra être fourni dans les 6 mois qui suivent le démarrage scientifique du projet : ANR, Commission européenne (Horizon Europe), European Research Council...
- L'ANR demande ainsi l'élaboration d'un plan de gestion de données pour les projets qu'elle finance.

02

Contenu et rédaction

Éléments communs à tous les PGD

Informations administratives

Sécurité des données

Description des données

Stockage des données durant le projet

Documentation, métadonnées, standards

Partage des données avec le dépôt dans un entrepôt

Aspects juridiques

Archivage pérenne

Responsabilités

Coûts

Il n'existe pas de trame unique.

Il faut choisir le modèle de DMP qui vous convient le mieux parmi tous ceux qui existent, mais on y retrouve les mêmes éléments, que voici :

- Les informations administratives
- La description de vos différents jeux de données
- Tout ce qui concerne la documentation de vos données avec le renseignement des métadonnées, selon quel standard...
- Les procédures mises en place pour assurer la qualité des données
- Les exigences légales et éthiques
- La sécurité des données
- Les modalités de stockage et de sauvegarde sécurisée des données durant le projet
- Les modalités de partage : quels jeux seront partagés, dans quel entrepôt, selon quelles conditions...
- Éventuellement, les modalités d'archivage pérenne si certains jeux de données le nécessitent
- La répartition des responsabilités
- L'évaluation des coûts.

En fonction du modèle de DMP choisi, certains paragraphes seront plus ou moins détaillés et les informations seront agencées différemment, mais à chaque fois, cela se présente sous forme de questions à renseigner.

Nous allons détailler chaque rubrique en mettant en exergue certains points et en vous proposant des exemples (tirés de 3 DMP publics) et des ressources.

Un exemple de PGD : **TRANSFUNERAIRE**

DMP du projet "TRANSFUNERAIRE: approche comparative des rituels collectifs de ré-inhumation en Europe et en Amérique-Latine XXe-XXIe siècles "

Plan de gestion de données créé à l'aide de DMP OPIDOR, basé sur le modèle "ANR - Modèle de PGD (simple)" fourni par Agence nationale de la recherche (ANR).

Renseignements sur le plan

Titre du plan DMP du projet "TRANSFUNERAIRE: approche comparative des rituels collectifs de ré-inhumation en Europe et en Amérique-Latine XXe-XXIe siècles "

Langue fra

Date de création 2020-06-26

Date de dernière modification 2020-07-22

Identifiant

Renseignements sur le projet

Titre du projet TRANSFUNERAIRE: approche comparative des rituels collectifs de ré-inhumation en Europe et en Amérique-Latine XXe-XXIe siècles

Résumé Notre programme est réalisé par deux équipes d'anthropologues sociaux travaillant sur l'Europe (sous la direction d'Elisabeth Anstett) et sur l'Amérique-Latine (sous la direction de Valérie Robin Azevedo) ; il propose d'effectuer une analyse comparative des pratiques funéraires qui ont actuellement lieu en Europe et en Amérique-Latine, dans des contextes marqués par la violence extrême et la mort de masse.

Il vise à éclairer leurs transformations les plus contemporaines mais aussi les brotages rituels, religieux et politiques mis en œuvre pour faire avec des corps fragmentés ou une absence inéluctable de corps. Au-delà d'une interrogation sur l'adaptabilité des grammaires et des lexiques funéraires, et d'un questionnement sur l'impact des pratiques médico-légales, notre objectif est de saisir les enjeux de ces transformations et les modalités de reconfiguration de la relation aux défunts ré-inhumés. Dans ces situations mortuaires inédites, une compréhension empirique et précise des procédures matérielles et symboliques entourant le destin posthume des restes humains ou des corps absents, ambitionne ainsi de mieux cerner les mécanismes socio-culturels de la résilience en temps de crise.

Sources de financement • Agence nationale de la recherche (ANR) : ANR-19-CE27-0022

Produits de recherche :

1. Données collectées par les membres du programme Transfunéraire (Collection)
2. Données produites par les membres, dans le cadre du programme Transfunéraire (Collection)
3. Fichier croisant les informations relatives aux données collectées et produites (Texte)

DMP du projet "TRANSFUNERAIRE: approche comparative des rituels collectifs de ré-inhumation en Europe et en Amérique-Latine XXe-XXIe siècles "

1. Description des données et collecte ou réutilisation de données existantes

Données collectées par les membres du programme Transfunéraire

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données sont collectées et produites lors d'enquêtes ethnographiques réalisées sur le terrain à partir d'entretiens non directif et d'observations (données produites). Ces enquêtes permettent aussi de collecter des corpus documentaires (archives presse ou médias, et archives personnelles).

La situation sanitaire mondiale induite par la crise du Covid-19 (maladie à coronavirus) a empêché que la plupart des enquêtes de terrain prévues à être réalisées en 2020 pour produire des données qualitatives inédites ne se tassent dans les délais prévus initialement.

Cette situation a amené les membres à réorienter leurs travaux vers l'analyse de matériaux préexistants. Ces données sont issues de recherches antérieures financées par les institutions respectives des membres du projet, et dont ils conservent un usage libre. Ces données sont utilisées pour les analyses forgées par les chercheurs membres du programme de la même façon que les données collectées et les données produites.

Ces données peuvent être :

- soit issues de recherches antérieures des membres du programme,
- soit issues d'une collecte réalisée à distance du terrain, sur des sources numérisées disponibles en ligne (articles de presse, images photographiques ou filmées).

Ces données peuvent être : des photographies, des enregistrements sonores, des enregistrements vidéos, des documents écrits (articles de presse, documents d'archives, notes de terrain).

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Les données collectées et produites peuvent être :

- Des images
- Des vidéos
- Des documents

Pour en savoir plus : <https://dmp.opidor.fr/plans/6619/export.pdf>

Qui gère les données et qui peut m'aider ?



Images : "Designed by Freepik"

Le chercheur n'est pas seul face à la rédaction du PGD. La gestion des données demande en effet un effort collectif. Les responsabilités sont souvent partagées, le projet étant mené le plus souvent en équipe ou avec des partenaires.

D'autres contributeurs peuvent également être sollicités :

- **L'ingénieur-projet** peut vous aider à coordonner les actions autour du PGD mais aussi dans l'obtention d'un agrément, pour l'éligibilité des coûts...
- **Le Service juridique** peut notamment vous conseiller sur la propriété intellectuelle des données.
- **Le Délégué à la protection des données** est l'interlocuteur idéal pour les aspects RGPD (règlement général sur la protection des données).
- **Les Professionnel de l'Information Scientifique et Technique** peuvent notamment aider au choix des standards de métadonnées, conseiller sur les entrepôts mais aussi réaliser des alignements avec des référentiels existants...
- **L'informaticien** est l'interlocuteur privilégié en ce qui concerne le stockage et la sécurisation des données, les aspects infrastructure et les coûts associés.
- **L'Archiviste** peut aider à sélectionner les données pour la conservation, à définir les durées et les solutions techniques.
- **La communauté scientifique** : les chercheurs du même domaine peuvent être d'une grande aide grâce au partage d'expérience, par exemple dans le choix de standards de métadonnées ou celui d'un entrepôt de données...

Ressources :

- *DoRANum. Les acteurs de la gestion et du partage des données de la recherche. 29 mars 2021.* https://doranum.fr/enjeux-benefices/acteurs-gestion-donnees-recherche_10_13143_1gjk-hs77/
- *Cat OPIDoR : recense les services dédiés aux données de la recherche en France.* <https://cat.opidor.fr/>
- *SOS-PGD : répertoire des Services Opérationnels de Soutien à la rédaction des Plans de Gestion des Données. 15 janvier 2024.* <https://gtso.couperin.org/gtdonnees/sos-pgd/>

Responsabilités : exemple 1

Extrait du DMP public *IMPRINT*, Jonathan Lenoir (CNRS) :

« La mise en œuvre du plan de gestion des données (PGD) a été réalisée par [...] le porteur du projet IMPRINT. Il endossera la responsabilité du PGD et la gestion globale de ce dernier sur la durée totale du projet [...]. Il s'assurera de la [...] mise à jour régulière du PGD sur la durée du projet. [...] »

Exemple tiré du DMP public IMPRINT : C'est le porteur du projet qui a la responsabilité du PGD.

Responsabilités : exemple 2

Extraits du DMP public *Hospitam* Lauriane Locatelli (ENS Lyon) :

« Le gestionnaire de données sera l'ingénieur recruté. [...] La saisie des données et la production des métadonnées sera assurée par les chercheurs du projet. L'ingénieur s'occupera de la qualité des données, du stockage et de la sauvegarde ainsi que de l'archivage et du partage des données. Il sera responsable de la mise en œuvre du plan de gestion de données. [...] »

Exemple tiré du DMP public Hospitam : Les rôles sont bien définis.

Responsabilités : exemple 3

Extrait du DMP public Transfunéraire, Clara Duterme (AMU) :

« Après la fin du programme de recherche, la responsabilité des données sera partagée par la Directrice de recherche au CNRS, coordinatrice du programme Transfunéraire, la Professeure à l'université de Paris, coordinatrice du programme Tranfunéraire et Clara Duterme (IGR, programme de recherche Transfunéraire). [...] »

Exemple tiré du DMP public Transfunéraire : La responsabilité des données est précisée également pour l'après-projet.

Description des données

Vous devez décrire tous les types de données (préexistantes et/ou produites) que vous allez « manipuler » pendant le projet

Un jeu de données à décrire = un « **lot** » de données pour un type de données défini

Vous devez décrire :

- Le type de données
- Les formats
- La provenance des données
- Les méthodes et outils employés

Description des données : Quelles données (types, formats, volumes, qualité...) seront collectées ou produites, et comment ?

Il s'agit de la partie la plus scientifique du PGD, dans laquelle vous allez devoir décrire les données que vous allez « manipuler » durant le projet (données préexistantes et/ou produites pour le projet).

Il ne s'agit pas de décrire toutes les données du projet, mais uniquement de décrire des jeux de données en fonction de leurs formats, de leurs types.

Concrètement, vous devez décrire :

- Le type de données (données numériques, textuelles, images, audio, vidéos etc.),
- Les formats, ouverts de préférence,
- La provenance des données (collecte de nouvelles données, conversion ou transformation des données, partage/échange de données, achat de données etc.),
- Les méthodes et outils.

Ressources :

DoRANum. Définition(s) des données de la recherche. 15 mars 2022. https://doranum.fr/plan-gestion-donnees-dmp/definitions-des-donnees-de-la-recherche_10_13143_b4zc-8e79/

DoRANum. L'origine et la description des données de la recherche. 15 mars 2022.

https://doranum.fr/plan-gestion-donnees-dmp/origine-description-donnees-recherche_10_13143_e9zh-w908/

Description des données : exemple 1

Extrait du DMP public Hospitam Lauriane Locatelli (ENS Lyon) :

« Les données produites seront les suivantes :

- Documents **textuels** des textes antiques non lemmatisés et lemmatisés en langue grec (**alphabet grec ancien avec accent, format Unicode**) en **.txt** regroupés en dossiers par auteurs, c'est-à-dire **6 dossiers de maximum 30 Mo** [...] »

Exemple tiré du DMP public Hospitam :

On voit que le type de données, les formats (ouverts), les volumes sont bien décrits. Pour information, la lemmatisation désigne un traitement lexical apporté à un texte en vue de son analyse.

Description des données : exemple 2

Extraits du DMP public Transfunéraire, Clara Duterme (AMU) :

« Les données collectées peuvent être :

- soit issues de recherches antérieures des membres du programme.
- soit issues d'une collecte réalisée à distance du terrain, sur des sources numérisées disponibles en ligne (articles de presse, images photographiques ou filmées). [...] »

« Les données produites incluent :

- des photographies
- des enregistrements sonores
- des enregistrements vidéos
- des documents écrits tels que des notes de recherches, des schémas ou des cartes [...] »

Exemples tirés du DMP public Transfunéraire :

Une bonne distinction est faite entre les données collectées et les données produites. L'origine des données collectées est mentionnée.

Les formats (le plus possible ouverts et standards) et les volumes sont décrits, en précisant que l'évaluation du volume peut évoluer.

Documentation et qualité des données

Vous devez expliquer comment vous allez documenter vos données (à l'aide de **métadonnées enrichies**) afin de les rendre compréhensibles, et éventuellement réutilisables

Vous devez expliquer :

- Comment les métadonnées sont produites
- Quels sont les standards de métadonnées utilisés
- Quels sont les vocabulaires spécifiques utilisés
- Quelle procédure de contrôle qualité des données est mise en place

Documentation et qualité des données : Quelles métadonnées, standards et formats ? Quel contrôle qualité ?

Comment les données seront identifiées, décrites, quels standards de métadonnées seront utilisés ?

Vous devez expliquer comment vous allez documenter vos données (à l'aide de métadonnées enrichies) afin de les rendre compréhensibles, et éventuellement réutilisables.

Rappel : les métadonnées **permettent de décrire plus précisément les données**. Elles facilitent la recherche et l'identification des données.

Elles peuvent donner des informations sur les conventions de nommage, les procédures et méthodes d'analyse, sur la définition de variables, d'unités de mesure etc...

Dans le PGD, vous devez expliquer :

- Comment les métadonnées seront produites ?
- Quels sont les standards de métadonnées utilisés ? Il est recommandé d'utiliser des standards de métadonnées de votre discipline (voir ressources).
- Si une procédure de contrôle qualité des données est mise en place, afin d'assurer la qualité de la collecte. Préciser alors les processus (comme la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la revue par les pairs...)

Documentation et qualité des données : exemple 1

Extrait du DMP public ***IMPRINT***, Jonathan Lenoir (CNRS) :

« Chaque fichier de données sera accompagné au minimum d'un **fichier texte (.txt)** de type "Read_me.txt" indiquant les **métadonnées** telles que la liste des noms (header) des variables enregistrées, avec pour chaque variable (header), le type d'appareillage utilisé pour la prise de mesure, les conditions dans lesquelles les données ont été collectées, l'unité de la variable mesurée ou toutes autres informations permettant une réutilisation des données. [...]

Le standard de métadonnées envisagé dans le cadre du projet est le **standard "Ecological Metadata Language" (EML)** [...] »

Exemples tiré du DMP public IMPRINT :

Le fichier Read_me.txt permet de préciser des métadonnées complémentaires : noms de variables, type d'appareillage utilisé, conditions de collecte... Cela facilite la réutilisation des données. Le standard de métadonnées EML est un standard disciplinaire.

Ressources :

- DoRANum. *Métadonnées, standards, formats : Fiche synthétique*. 27 novembre 2017. https://doranum.fr/metadonnees-standards-formats/metadonnees-standards-formats-fiche-synthetique_10_13143_vbjs-6288/
- DoRANum. *Les standards de métadonnées : pourquoi et lequel ?* 13 avril 2021. https://doranum.fr/metadonnees-standards-formats/standard-metadonnees_10_13143_y5py-w521/
- *Répertoires de standards de métadonnées :*
 - DCC. *Disciplinary Metadata*. <https://www.dcc.ac.uk/guidance/standards/metadata>
 - RDA. *Metadata Standards Catalog*. <https://rdamsc.bath.ac.uk/>

Documentation et qualité des données : exemple 2

Extrait du DMP public ***IMPRINT***, Jonathan Lenoir (CNRS) :

« Concernant la qualité et la conformité de la collecte des données microclimatiques, il est prévu une **phase d'intercalibration des capteurs** HOBO UA-001-08, HOBO UA-001-64 et TMS4 **en conditions contrôlées**. L'installation des capteurs de T°C et d'humidité relative du sol, in-situ, suivra un **protocole standardisé** pour l'ensemble des sites étudiés. Concernant les données saisies sur le terrain (inventaires dendrométriques et floristiques) ou en laboratoire (détermination des espèces de carabes et analyses de sol éventuellement), elles **seront validées par l'ensemble des scientifiques impliqués dans le projet**. [...] »

Exemple tiré du DMP public **IMPRINT** :

Les procédures de contrôle qualité et de validation sont très précises.

Exigences légales et éthiques

Vous devez indiquer si vos données soulèvent des questions éthiques et/ou juridiques

Vous devez préciser :

- Qui sera titulaire des droits de propriété intellectuelle sur les données
- Si un accord de consortium a été rédigé
- Si des données protégées par des droits spécifiques seront utilisées au cours du projet
- Quelles mesures seront prises pour garantir la confidentialité des données personnelles

Un autre point très important qui vient jalonner tous les aspects du DMP : il s'agit des exigences légales et éthiques.

Exigences légales et éthiques : Quelles questions éthiques et/ou juridiques les données soulèvent-elles ? Quelles éventuelles restrictions ?

Dans le DMP vous devez préciser :

- Qui sera titulaire des droits de propriété intellectuelle sur les données (qui aura le droit d'en contrôler l'accès) ?
- Si un accord de consortium a été rédigé, si votre projet se déroule dans le cadre d'un partenariat,
- Si des données protégées par des droits spécifiques seront utilisées au cours du projet : par exemple, des données personnelles, des bases de données etc.
- Quelles mesures seront prises pour garantir la confidentialité des données personnelles ? Pour des données personnelles par exemple, vous pouvez envisager la pseudonymisation ou l'anonymisation, vous pouvez aussi envisager le chiffrement des données.

Dans le plan, mentionnez l'existence d'éventuels comités d'éthique pour votre institution ou votre discipline.

Dans tous les cas, pour les questions RGPD, il existe dans chaque organisme/institution un/une déléguée à la protection des données DPO ou DPD (en français).

Liste des Organismes ayant désigné un(e) délégué(e) à la protection des données (DPD/DPO) :

<https://www.data.gouv.fr/fr/datasets/organismes-ayant-designe-un-e-delegue-e-a-la-protection-des-donnees-dpd-dpo/#/resources>

Exigences légales et éthiques : exemple 1

Extrait du DMP public *Transfunéraire*, Clara Duterme (AMU) :

« Le programme de recherche « Transfunéraire » respecte le règlement général sur la protection des données (RGPD) de l'Union Européenne. Ses membres respectent les principes directeurs d'une recherche responsable, tels qu'énoncés dans le guide "Pratiquer une recherche intègre et responsable" publié par le Comité d'éthique du CNRS (COMET) en 2017.

[...] avant tout entretien en face à face, un formulaire de consentement sera signé par la personne enquêtée. Un consentement spécifique sera demandé pour les enregistrements audio. La signature du formulaire de consentement sera accompagnée d'une présentation claire des objectifs de la recherche, de la pratique de l'anonymisation et de l'usage éventuel des données à caractère personnel. »

Exemples tirés du DMP public Transfunéraire :

Ici, il est bien précisé que les règles sur la protection des données sont respectées. Un formulaire de consentement est proposé. Les objectifs de la recherche, de la pratique de l'anonymisation et de l'usage des données à caractère personnel sont présentés à la personne enquêtée.

Ressources :

- DoRANum. Aspects juridiques, éthiques, intégrité scientifique : Fiche synthétique. 23 juillet 2019. https://doranum.fr/aspects-juridiques-ethiques/aspects-juridiques-et-ethiques-fiche-synthetique_10_13143_wz7a-hh81/
- DoRANum. Questions juridiques liées aux données de recherche : interview de Lionel Maurel. 22 juillet 2022. https://doranum.fr/aspects-juridiques-ethiques/questions-juridiques-liees-aux-donnees-de-la-recherche_10_13143_xjgm-hb78/
- DoRANum. Webinaire des Tuto@mate « A qui appartiennent les données ? » - Présentation de Lionel Maurel. 2 septembre 2021. https://doranum.fr/aspects-juridiques-ethiques/webinaire-des-tutoamate_10_13143_mna5-ys10/
- data.gouv.fr. Organismes ayant désigné un(e) délégué(e) à la protection des données (DPD/DPO). <https://www.data.gouv.fr/fr/datasets/organismes-ayant-designe-un-e-delegue-e-a-la-protection-des-donnees-dpd-dpo/>

Exigences légales et éthiques : exemple2

Extrait du DMP public *Hospitam*, Lauriane Locatelli (ENS Lyon) :

« Nos données sont des **textes** antiques **libres de droit** comme la bible ou le texte des auteurs antiques bruts, n'appartenant pas aux maisons d'éditions. **Nous n'avons pas de données à caractère personnel.** [...] »

Exemple tirés du DMP public Hospitam :

Le projet n'est pas concerné par les exigences légales (traitement de données libres de droit et non sensibles) : les auteurs l'ont précisé dans la section.

Stockage, partage et archivage : 3 étapes distinctes

Principe « Aussi ouvert que possible, aussi fermé que nécessaire »

STOCKAGE ET
SAUVEGARDE



Espace collaboratif
de travail

Serveur sécurisé
de l'institution



PARTAGE



Entrepôt généraliste
ou disciplinaire

ARCHIVAGE PÉRENNE



Plateforme d'archivage
(ex : CINES)

- Le **stockage** et la **sauvegarde** sécurisés des données se font durant le projet. L'objectif est de garantir la sécurité des données et d'en **faciliter l'accès** pour l'ensemble des collaborateurs du projet.
- Le **partage** consiste à déposer les données dans un entrepôt de données afin de les rendre **accessibles facilement** et de permettre leur **réutilisation** par des chercheurs du même domaine ou d'un autre domaine, selon les principes FAIR et sur le court et le moyen terme. Le partage se fait souvent à l'issue du projet.
- L'**archivage pérenne** a pour objectif de conserver les données, d'en garantir l'**accès** et d'en préserver l'intelligibilité sur le long terme, c'est-à-dire plus de 30 ans. En réalité, l'archivage pérenne concerne peu de données. Seulement celles qui présentent une grande valeur scientifique reconnue par la communauté dont elles proviennent. Soit parce qu'elles sont très coûteuses, soit parce qu'elles sont uniques, non reproductibles.

Les **bonnes pratiques** de **sauvegarde**, d'ajout de **métadonnées**, attribution d'**identifiants pérennes** et de **licences** participent tous au respect des **principes FAIR**.

Ressource :

DoRANum. *Stockage, partage et archivage : quelles différences ?* 1^{er} juillet 2021.

<https://doranum.fr/stockage-archivage/stockage-partage-archivage-quelles-differences-10-13143-5dax-qp58/>

Stockage et sauvegarde des données au cours du projet

Vous devez expliquer quel processus de stockage des données sera mis en place durant le projet, afin d'en garantir l'accès pour l'ensemble des collaborateurs du projet

Vous devez préciser :

- Quels supports de stockage seront utilisés
- Quel système de nommage sera utilisé
- Quelle sera la volumétrie des données
- Comment sera organisée la sauvegarde des données
- Comment seront gérées les versions

Stockage et sauvegarde des données : Comment les données seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Vous devez expliquer quel processus de stockage des données sera mis en place durant le projet, afin d'en garantir l'accès pour l'ensemble des collaborateurs du projet.

Vous devez préciser :

- Quels seront les supports de stockage (serveur sécurisé de votre institution, espace de travail collaboratif etc...) ?
- Quel système de nommage sera utilisé ? La fiabilité d'accès aux données pour l'ensemble d'une équipe passe par un système de nommage unique et précis des fichiers.
- Quelle sera la volumétrie des données ? Une estimation suffit.
- Comment sera organisée la sauvegarde des données ? À chaque point d'étape du projet, sélectionner les données à sauvegarder, à supprimer. Définir leur durée de conservation.
- Comment seront gérées les versions ? Les différents états des données sont conservés en corrélation avec les différentes étapes de traitement. Permet de revenir à une version antérieure si besoin.

Stockage et sauvegarde : exemples

Extrait du DMP public *Transfunéraire*, Clara Duterme (AMU) :

« [...] les données seront sauvegardées sur des cartes mémoires fournies aux participants (concerne les images photographiques et les films), [...] chaque membre conserve une copie des données qu'il/ elle a produites sur son ordinateur personnel, protégé par mot de passe quand il s'agit de données à caractère particulièrement confidentiel. [...] Les chercheurs peuvent également déposer leurs données sur les plates-formes numériques individuelles mises à disposition par leurs institutions pour en garantir la pérennité en cas de perte, ou procéder à leur sauvegarde sur les disques dur externes qui leur ont été fournis dans le cadre du programme de recherche [...] »

Exemple tiré du DMP public Transfunéraire :

Dans cet exemple, les différents supports et lieux de sauvegarde sont bien précisés. Les données destinées à être partagées entre membres du programme sont stockées via une solution de stockage de type cloud sécurisé. L'accès est strictement réservé aux membres du programme.

Ressources :

- DoRANum. Stocker ses données de façon sécurisée. 26 juillet 2022.
https://dorum.fr/stockage-archivage/stockage-donnees_10_13143_z0ge-nc29/
- DoRANum. Comment bien nommer ses fichiers ? 12 octobre 2018.
https://dorum.fr/stockage-archivage/comment-nommer-fichiers_10_13143_wqqw-aa59/
- DoRANum. Conseils pour l'organisation des données. 14 septembre 2021.
https://dorum.fr/stockage-archivage/conseils-pour-lorganisation-des-donnees_10_13143_h6qx-e249/

Stockage et sauvegarde : exemples

Extrait du DMP public Hospitam, Lauriane Locatelli (ENS Lyon) :

« Concernant la gestion du versioning, nous utiliserons un logiciel de gestion des versions : Git. Git est un logiciel libre. »

Exemple tiré du DMP public Hospitam :

Nous voyons bien comment sera géré le versioning.

Ressources :

- DoRANum. Stocker ses données de façon sécurisée. 26 juillet 2022.
https://doranum.fr/stockage-archivage/stockage-donnees_10_13143_z0ge-nc29/
- DoRANum. Comment bien nommer ses fichiers ? 12 octobre 2018.
https://doranum.fr/stockage-archivage/comment-nommer-fichiers_10_13143_wgqw-aa59/
- DoRANum. Conseils pour l'organisation des données. 14 septembre 2021.
https://doranum.fr/stockage-archivage/conseils-pour-lorganisation-des-donnees_10_13143_h6gx-e249/

Stratégie de partage et d'ouverture des données

Garder le principe « aussi ouvert et FAIR que possible, aussi fermé que nécessaire » au cœur de votre démarche

Vous devez préciser :

- Si vous êtes concerné par une obligation de partage
- S'il y a une restriction ou interdiction de partage
- Avec qui les données seront-elles partagées
- Dans quel(s) entrepôt(s) les données seront-elles déposées
- Quelles licences seront appliquées aux jeux de données
- Si des identifiants pérennes seront attribués aux données

Stratégie de partage et d'ouverture des données : Quelle politique d'accès, de partage et de réutilisation des données ?

Le PGD doit montrer que vous menez une réflexion à propos de l'ouverture et du partage de vos données, tout en gardant le principe « **aussi ouvert et FAIR que possible, aussi fermé que nécessaire** » au cœur de votre démarche.

Vous devez préciser :

- S'il y a une obligation de partage (par exemple pour les données géographiques dans le cadre de la directive Inspire).
- S'il y a une restriction ou interdiction de partage, par exemple pour des raisons éthiques, si les données sont confidentielles ou personnelles, soumises à une propriété intellectuelle, commerciale, ou pour des raisons de sécurité...
- Avec qui les données seront partagées (accès ouvert ou restreint) ? Sachant que même si vos données sont libres d'accès, vous pouvez décider quand même d'un embargo sur ces données avant de les rendre publiques.
- Dans quel(s) entrepôt(s) les données seront-elles déposées ? Avez-vous à disposition un entrepôt disciplinaire ou institutionnel ? Sinon vous pouvez déposer vos données dans l'entrepôt national Recherche Data Gouv.
- Quelles licences seront appliquées aux jeux de données ?
- Un identifiant pérenne (ex : DOI) sera-t-il attribué ?

Partage et ouverture des données : exemples

Extrait du DMP public [Transfunéraire](#), Clara Duterme (AMU) :

« Les données seront partagées à la fin du programme de recherche sur la plateforme [Didomena](#), l'entrepôt de données de recherche de l'EHESS.

(<https://didomena.ehess.fr>). Cette plateforme est dédiée aux sciences sociales et permet de partager et valoriser les données de recherche. Il n'existe pas de raison motivant un embargo, les données partagées seront immédiatement accessibles. [...] »

Exemple tiré du DMP Transfunéraire :

Ici, on voit que l'entrepôt de données utilisé ainsi que les conditions d'accès aux données sont bien précisés.

Ressources :

- DoRANum. Les critères pour choisir un entrepôt de données. 30 août 2023. <https://doranum.fr/depot-entrepots/criteres-pour-choisir-entrepot-de-donnees-10-13143-zqpb-9449/>
- DoRANum. Les licences de réutilisation dans le cadre de l'Open data et de la loi pour une république numérique. 26 janvier 2023. <https://doranum.fr/aspects-juridiques-ethiques/les-licences-de-reutilisation-dans-le-cadre-de-lopen-data-2-10-13143-ssh2-zd93/>
- Re3data (répertoire international d'entrepôts). <https://www.re3data.org/>
- Cat OPIDor (répertoire des services français dédiés aux données de la recherche). <https://cat.opidor.fr/>
- Recherche Data Gouv : <https://recherche.data.gouv.fr/fr>
- Recherche Data Gouv, un entrepôt au service du partage et de l'ouverture des données de recherche : <https://doranum.fr/depot-entrepots/recherche-data-gouv-un-entrepot-au-service-du-partage-et-de-louverture-des-donnees-de-recherche/>

Partage et ouverture des données : exemples

Extraits du DMP public *IMPRINT*, Jonathan Lenoir (CNRS) :

« Les données utilisées pour la publication des résultats seront [...] publiées en **libre accès** au moment de la publication de l'article, **ainsi que les scripts** (lignes de codes) utilisés pour générer les résultats à partir des données. Ceci dans le but d'assurer la reproductibilité des résultats publiés. [...] »

Exemple tiré du DMP IMPRINT :

Les codes sources utilisés sont partagés également.

Ressources :

- DoRANum. Les critères pour choisir un entrepôt de données. 30 août 2023. <https://doranum.fr/depot-entrepots/criteres-pour-choisir-entrepot-de-donnees-10-13143-zqpb-9449/>
- DoRANum. Les licences de réutilisation dans le cadre de l'Open data et de la loi pour une république numérique. 26 janvier 2023. <https://doranum.fr/aspects-juridiques-ethiques/les-licences-de-reutilisation-dans-le-cadre-de-lopen-data-2-10-13143-ssh2-zd93/>
- Re3data (répertoire international d'entrepôts). <https://www.re3data.org/>
- Cat OPIDor (répertoire des services français dédiés aux données de la recherche). <https://cat.opidor.fr/>
- Recherche Data Gouv : <https://recherche.data.gouv.fr/fr>
- Recherche Data Gouv, un entrepôt au service du partage et de l'ouverture des données de recherche : <https://doranum.fr/depot-entrepots/recherche-data-gouv-un-entrepot-au-service-du-partage-et-de-louverture-des-donnees-de-recherche/>

Préservation à long terme des données

Vous devez dire si un plan de conservation des données sur le long terme a été établi

Vous devez préciser :

- Quelles sont les données (et les métadonnées associées) que vous avez sélectionnées pour l'archivage
- Quelle est la volumétrie prévue pour ces données
- Quelle sera la durée de conservation à long terme
- Quel budget est prévu pour l'archivage

Préservation à long terme des données : Quelle stratégie d'archivage ? Dans quelles conditions ? Pour quelles données ?

Certains projets ne seront pas concernés par cette section. En effet, les données conservées à long terme seront les données coûteuses, non reproductibles, à haute valeur scientifique.

Dans le DMP vous aurez à aborder aussi l'aspect « préservation à long terme : Quel archivage à long terme pour quelles données ? Dans quelles conditions ? »

Lorsqu'on parle de « préservation à long terme », on parle d'archivage pérenne qui a pour objectif de conserver les données, et d'en garantir l'accès sur le long terme (pour + de 30 ans).

C'est différent du stockage durant le projet ou du partage dans des entrepôts de données.

Dans le DMP, vous devez indiquer :

- Quelles sont les données (et les métadonnées associées) que vous avez sélectionnées pour l'archivage. ...
- Quel est la volumétrie prévue pour ces données ?
- Quelle sera la durée de conservation à long terme ?
- Quel budget est prévu pour l'archivage : l'archivage étant assez coûteux, il faut indiquer quelles sont vos garanties de financement pour couvrir les frais ?

A noter que ce point en particulier peut être abordé dans le cadre d'une politique institutionnelle d'archivage (et plus largement une politique de gestion des données de recherche).

Préservation à long terme des données : exemples

Extrait du DMP public *Transfunéraire*, Clara Duterme (AMU)

« Le choix des données à conserver sur le long terme se fera en concertation commune entre les membres du programme ; la décision finale appartiendra aux deux coordinatrices du Programme, en fonction du caractère inédit des données, de leur valeur et de leur pertinence pour des démarches comparatives futures. [...] »

Exemple tiré du DMP public Transfunéraire :

Ici, on voit qu'au stade de rédaction de ce DMP, les questions d'archivage sont en cours de réflexion. Même si les actions ne sont pas encore bien définies, il est important de dire que la réflexion est en cours.

Ressources :

- DoRANum. *Stockage et archivage : fiche synthétique*. 6 août 2018. https://dorum.fr/stockage-archivage/stockage-et-archivage-fiche-synthetique_10_13143_0c4b-2743/
- DoRANum. *La minute Archivage pérenne des données de recherche*. 10 mars 2022. https://dorum.fr/stockage-archivage/la-minute-archivage-perenne-des-donnees-de-recherche_10_13143_77vj-0011/
- CINES. *FACILE – Service de validation de formats*. <https://facile.cines.fr/>

Évaluer le coût de la gestion des données

Vous devez **examiner et justifier** soigneusement toutes les ressources nécessaires pour gérer et diffuser les données :

- Coût de **production ou collecte des données**
- Coût de **préparation des données** pour le partage/préservation

Exemples de coûts : frais de stockage, coût du matériel, temps de personnel, coûts de préparation des données pour le dépôt, frais d'entrepôt et d'archivage

À noter : certains coûts sont éligibles dans le cadre de projets financés

- Examiner et justifier soigneusement toutes les ressources nécessaires pour gérer et diffuser les données.
- Comment les ressources nécessaires à la préparation des données pour le partage/préservation ont été chiffrées : temps, nettoyage des données...
- Exemples de coûts : frais de stockage, de coût matériel, de temps de personnel, coûts de préparation des données pour le dépôt, de frais d'entrepôt et d'archivage.

Précisez les montants et comment ces coûts seront couverts.

Par exemple, pour les projets financés par l'ANR, toutes les dépenses liées à la gestion des données sont éligibles durant le projet : acquisition, collecte, stockage, personnel dédié à la gestion des données... En cas de recours à un tiers, le coût de stockage des données est admissible jusqu'à 5 ans après la date de fin scientifique du projet, sous réserve que le contrat avec ce tiers soit conclu avant la fin scientifique du projet.

Gala Garcia Reategui. La politique science ouverte de l'Agence Nationale de la Recherche et le DMP.

[https://octaviana.fr/document/VUN0041_02#?c=&m=&s=&cv=.](https://octaviana.fr/document/VUN0041_02#?c=&m=&s=&cv=)

Coûts de gestion des données : exemples

Extrait du DMP public IMPRINT, Jonathan Lenoir (CNRS) :

« Concernant les frais de stockage et de partage des données, nous utiliserons au maximum les plateformes de stockage gratuites, notamment pour les jeux de données de taille moyenne. Pour les jeux de données les plus volumineux (plus de 100 Go à plusieurs To), il sera éventuellement envisagé de recourir à des plateformes payantes pour la mise en libre accès (cf. Scientific Data). Pour cela, une partie du budget IMPRINT sera potentiellement alloué à la publication en libre accès (max 2000 EUR). [...] »

Exemple tiré du DMP public IMPRINT :

Les plateformes gratuites sont privilégiées. Les plateformes payantes sont envisagées pour des jeux de données volumineux et pour favoriser le libre accès.

Ressources :

- DoRANum. *Le coût de la gestion des données*. 12 avril 2023. https://doranum.fr/enjeux-benefices/le-cout-de-la-gestion-des-donnees_10_13143_hch2-h207/
- OpenAIRE. *Guides for Researchers. How to identify and assess Research Data Management (RDM) costs*. <https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>
- O'Connor Ryan, Delipalta Alexandra, Jones Sarah. *What will it cost to manage and share my data?* 21 mai 2020. <https://doi.org/10.5281/zenodo.3837717>
- EPFL Library. *Cost Calculator for Data Management*. <https://costcalc.epfl.ch/>

04

Présentation rapide de DMP OPIDoR

DMP OPIDOR, qu'est-ce que c'est ?



<https://dmp.opidor.fr/>

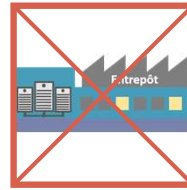
- Un outil **gratuit** développé par l'Inist-CNRS
- **Accessible** à l'ensemble de la communauté scientifique de l'ESR et à ses partenaires français ou étrangers
- **Collaboratif** : il permet les échanges entre les partenaires d'un même projet et les services d'accompagnement
- **Sécurisé** : hébergé sur un serveur à l'Inist-CNRS, en France
- **Évolutif** : l'outil évolue grâce aux retours des utilisateurs

Pour vous guider dans la rédaction d'un DMP, l'Inist-CNRS propose un outil d'aide en ligne à l'élaboration de DMP.

DMP OPIDoR, c'est un outil :

- Gratuit et développé par l'Inist-CNRS sur la base du code source DMP Roadmap (fusion du code DMP Online et DMPTool, outils britannique et américain),
- Accessible à l'ensemble de la communauté scientifique de l'ESR et à ses partenaires français ou étrangers,
- Collaboratif : il permet les échanges entre les partenaires d'un même projet et les services d'accompagnement,
- Sécurisé sur un serveur hébergé à l'Inist-CNRS, en France,
- Évolutif : l'outil évolue grâce aux retours des utilisateurs (chercheurs et personnels d'accompagnement).

À quoi ça sert ?



- Créer / rédiger un DMP à partir d'un modèle
- Bénéficier de recommandations de certains organismes
- Collaborer sur un DMP (partage du PGD, commentaires)
- Bénéficier d'une assistance conseil auprès des services d'appui de son organisme de recherche (s'il le propose)
- Télécharger son DMP sous différents formats (pdf, docx, html, JSON)

Attention : **DMP OPIDoR n'est pas un entrepôt de données ni une archive de DMP**

- DMP OPIDoR permet de créer/rédiger un plan de gestion à partir d'un modèle.
- Bénéficier de recommandations, d'exemples, de certains organismes
- Il permet de partager votre plan avec vos collaborateurs, d'ajouter des commentaires.
- Il permet de bénéficier d'une assistance conseil auprès des services d'appui de votre organisme s'il le propose.
- Télécharger son DMP sous différents formats.

Attention, DMP OPIDoR n'est pas un entrepôt de données. Il ne sert pas à déposer des données de recherche mais à les décrire via la rédaction d'un DMP.

DMP OPIDoR n'est pas non plus une archive de DMP.

Les modèles disponibles dans DMP OPIDoR

- Agences de financement : Science Europe, ANR, ERC, INCa (Institut national du cancer)...
- Organismes de recherche : CNRS, Cirad, CEA, ENSAM, INRAE, ICM, Institut Pasteur, Sciences Po, CC-IN2P3, MASA Consortium...
- Universités : Lille, Montpellier, Paris Dauphine, Paris Sorbonne, Toulouse Jean Jaurès...
- Autres : Projet PRESOFT, Digital Curation Centre (UK), EPFL...

Pour vous guider dans la rédaction d'un DMP, les financeurs et organismes de recherche proposent des modèles de DMP intégrés directement dans l'outil DMP OPIDoR.

Un modèle, c'est une trame qui reprend tous les aspects que nous avons vu dans la partie précédente (description des données, documentation, aspects éthiques etc).

Dans les modèles, vous pouvez trouver également des recommandations d'organismes de recherche et universités.

Ils peuvent proposer des liens vers des ressources supplémentaires, des recommandations spécifiques (par exemple l'entrepôt INRAE pour déposer les données de recherche projets), des contacts DPD (délégué à la protection des données), des liens vers le site du service des archives d'université, etc.

Pour en savoir +

- Pour tout savoir sur le PGD :
- [DoRANum – Le Plan de Gestion de Données pas à pas](#)
- Pour vous aider à améliorer le contenu de votre PGD :
- [DoRANum - Grilles de relecture de Plans de Gestion de Données](#)
- [DoRANum – Foire Aux Questions](#)
- Pour retrouver toutes nos ressources pédagogiques :
www.dorandum.fr



Merci de votre attention

cnrs-contactpgd@inist.fr

contact-formation@inist.fr

Merci aux propriétaires de ces PGD de les avoir rendus publics ☺

DMP Hospitam - Modèle ANR - ENS de Lyon - Lauriane Locatelli - 24/04/2020

DMP TRANSFUNERAIRE : approche comparative des rituels collectifs de ré-inhumation en Europe et en Amérique-Latine XXe-XXIe siècles - Modèle ANR - Université d'Aix-Marseille - Clara Düterme - 22/07/2020

DMP IMPRINT - Modèle ANR - EDYSAN (CNRS) - Jonathan Lenoir - 12/03/2020

cnrs Inist

DORA Num

