

Stage Urfist

Pourquoi et comment construire un Plan de Gestion de Données (PGD/DMP) ?

26 mars 2025

Florence Thiault

Rémi Joinville





Programme

- 1. Eléments de contexte : politiques de Science ouverte, préconisations / exigences des financeurs
- 2. Identifier les données de la recherche
- 3. Les principes d'un Plan de Gestion de Données (PGD)
- 4. Comment rédiger un PGD en appliquant les principes FAIR ?

Partie 1

Eléments de contexte : politiques de Science ouverte, préconisations et exigences des financeurs

Contexte et politiques

Open Research Data

Problématique de l'accès aux données de la recherche financée sur fonds publics

Le rapport de l'OCDE (2007)

Principes et lignes directrices de l'OCDE pour l'accès aux données

« La valeur des données réside dans leur exploitation. L'accès total et ouvert aux données scientifiques devrait devenir la norme internationale pour l'échange des données scientifiques issues de la recherche financée sur fonds publics »

Programme de recherche européen <u>Horizon 2020</u>



- □ Projet pilote lancé en 2013 sur les données de recherche
- Fait « du libre accès aux données de recherche la règle générale »



- Programme cadre <u>Horizon Europe</u> pour 2021-2027 dans la continuité
- => généralisation pratiques SO + incitation au partage des données (PGD)

Contexte et politiques

- Loi pour une République numérique (octobre 2016)
 - Article 30 : conditions de libre réutilisation des données de recherche

-« II.- Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre. »

Les politiques actuelles

Plan national pour la science ouverte (PNSO) : 2018 et 2021

DEUXIÈME AXE:

STRUCTURER, PARTAGER ET OUVRIR LES DONNÉES DE LA RECHERCHE





PNSO 1

2018

- 4. Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics
- Créer la fonction d'administrateur des données et le réseau associé au sein des établissements
- Créer les conditions et promouvoir l'adoption d'une politique de données ouvertes associées aux articles publiés par les chercheurs.

PNSO 2

2021

- 4. Mettre en œuvre l'obligation de diffusion ouverte ou de partage des données de recherche financées sur fonds publics
- Créer Recherche Data Gouv, la plateforme nationale fédérée des données de la recherche
- 6. Promouvoir l'adoption d'une politique de données prenant en compte l'ensemble du cycle des données de la recherche avec pour ambition de rendre les données Faciles à trouver, Accessibles, Interopérables et Réutilisables
 - Développer et structurer les offres d'accompagnement et les outils à disposition des chercheurs
 - · Reconnaître et amplifier la réutilisation des données de la recherche
 - · Coordonner et promouvoir la politique d'ouverture des données

Décret n° 2021-1572 du 3 décembre 2021 relatif au

respect des exigences de l'intégrité scientifique dans les établissements de recherche



Article 2

Les établissements publics et fondations reconnues d'utilité publique mentionnés au <u>troisième alinéa de l'article L. 211-2 du code de la recherche</u> :

(...) 3° **Promeuvent** la diffusion des publications en accès ouvert et **la mise à disposition des méthodes et protocoles, des données et des codes sources associés aux résultats de la recherche** afin d'en garantir la traçabilité et la reproductibilité. Ils incitent à la publication des résultats de recherche dits négatifs ; (...)

Article 6

[Ils] définissent une politique de conservation, de communication et de réutilisation des résultats bruts des travaux scientifiques menés en son sein. A cet effet, ils veillent à la mise en œuvre par leur personnel de plans de gestion de données et contribuent aux infrastructures qui permettent la conservation, la communication et la réutilisation des données et des codes sources.

Voir l'article de l'Institut Pasteur : <u>Décret du 3 décembre 2021 : la science ouverte et la gestion des données sont reconnus comme des</u> questions d'intégrité scientifique

Le PGD : obligations et recommandations

Incitation voire obligation:

 Organismes de recherche / universités







 Agences de financement de la recherche









Plan d'action 2019



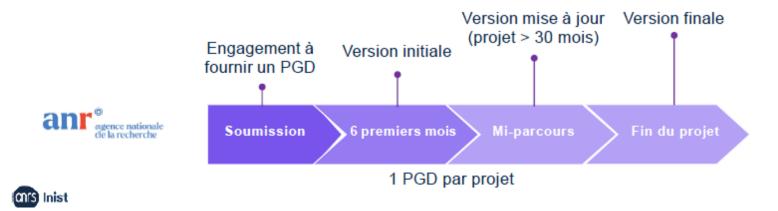
Modèle de DMP / PGD ANR sur DMP Opidor



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



source : Inist Cnrs, webinaire *Plan de Gestion de Données*

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

PGD (Plan de gestion des données) ou DMP (Data Management Plan)

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

Partie 2

Identifier les données de recherche

Quelles seraient les types de données que pourriez intégrer dans votre Plan de Gestion de Données ?

Indiquer les catégories de données dans le chat



Diversité des données de la recherche

"un concept parfois difficile à définir à cause de leur grande variété"

<u>Joachim Schöpfel, Eric Kergosien, Hélène Prost. « Pour commencer, pourriez-vous définir 'données de la recherche' ? » Une tentative de réponse (2017)</u>

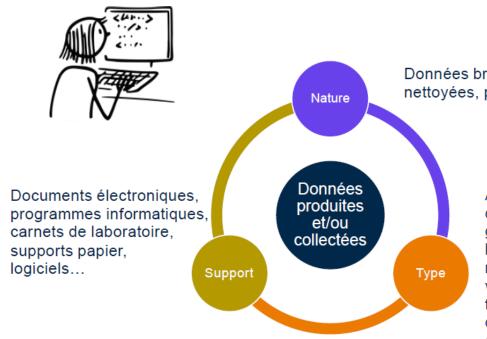
La définition de l'OCDE :

« Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche » (OECD Principles and Guidelines for Access to Reseach Data from Public Funding, 2007)



Diversité des données de la recherche

Définition des données de recherche



Données brutes, dérivées, formatées, nettoyées, primaires, secondaires, traitées....

Archives, audio, vidéos, bases de données, codes sources, données géospatiales, images, photographies, langages de programmation, matérielles et physiques, modèles, visualisations, 3D, numériques, textuelles, numérisations, scans, qualitatives, quantitatives, statistiques...



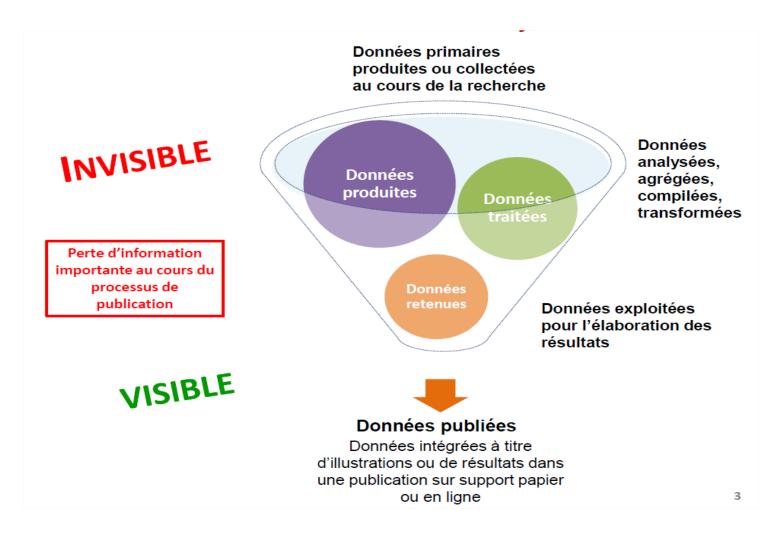
Voir Doranum:
L'origine et la
description des
données de la
recherche





source : Inist Cnrs, webinaire *Principes FAIR et cycle de vie des données*

Que deviennent les données ?



Source : Magali Moysan. Les plans de gestion de données, 2016

Pourquoi ouvrir ses données?

Dans quels objectifs partager ses données ? indiquer votre réponse dans le chat.



Les enjeux actuels de l'ouverture des données de la recherche

- Transparence et obligation de rendre des comptes
- renforce la confiance des citoyens
- meilleure implication (sciences participatives)
- Nouveaux terrains de recherche

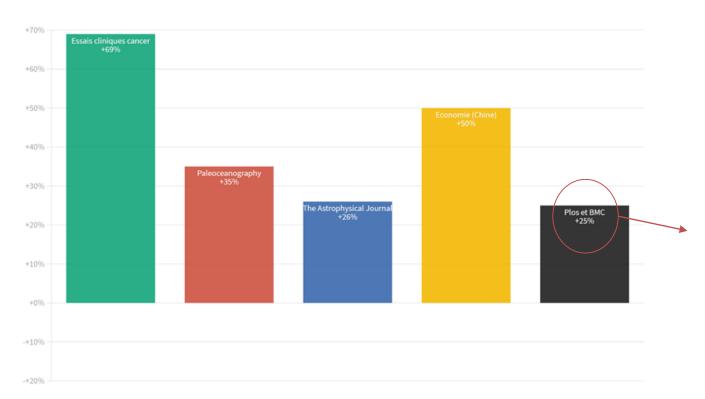
source : Chloé Fabre, Gestion et diffusion des données de recherches, Callisto 2022 Un meilleur retour pour la SOCiété De nouvelles exigences et opportunités pour le chercheur

De nouvelles perspectives pour la science

- Condition pour l'obtention de projets (ex : ANR)
- Recommandation ou exigence des éditeurs
- Meilleure <u>visibilité et</u>
 citabilité du travail du
 chercheur = données
 réutilisées
- Nouvelle forme de publication scientifique (*data paper*)
- Une garantie contre la fraude scientifique
- Qualité du travail de recherche avec la mise en place de bonnes pratiques grâce au partage
- Meilleure gestion des coûts de la recherche
- De nouvelles directions de recherche

à consulter Fiche Doranum : enjeux et bénéfices

Focus : meilleures visibilités et citabilités du travail du chercheur avec le partage des données



Impact de l'ajout de données associées aux articles en termes de citations des articles concernés

Colavizza, Giovanni, Iain Hrynaszkiewicz, Isla Staden, KirstieWhitaker, et Barbara McGillivray. 2020. « The Citation Advantage of LinkingPublications to ResearchData ». PLOS ONE 15 (4): e0230416.

https://doi.org/10.1371/journal.pone.0230416

source du graphique : <u>Ouvrons la Science</u>, <u>Les données de la recherche : un sujet stratégique</u>

=> croissance forte des citations lorsqu'on associe des données à un papier : opportunité pour la recherche de renforcer sa visibilité



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

Principes applicables tout au long du cycle de vie des données

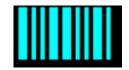
SOURCE: https://doranum.fr/enjeux-benefices/principes-fair 10 13143 z7s6-ed26/

Principe F: Données faciles à trouver par des humains et par des systèmes informatiques

Exemples de bonnes pratiques :



- déposer ses données dans un entrepôt de données
- leur attribuer un identifiant pérenne (PID) : mécanisme permettant d'identifier de façon stable et unique des ressources sur le Web) => ex : DOI



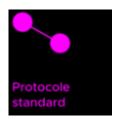
 les décrire avec des métadonnées riches (pour faciliter la recherche d'une donnée)



Principe A : faciliter l'accès et/ou le téléchargement des données et métadonnées, en spécifiant les conditions d'accès (accès ouvert ou restreint)

Exemples de bonnes pratiques :

- Protocoles standardisés et ouverts
- Authentification et autorisation si besoin
- Accès permanent aux métadonnées







Principe I : Données interopérables. Les données sont téléchargeables, utilisables, intelligibles et combinables avec d'autres données, par des humains et des systèmes informatiques

Exemples de bonnes pratiques :

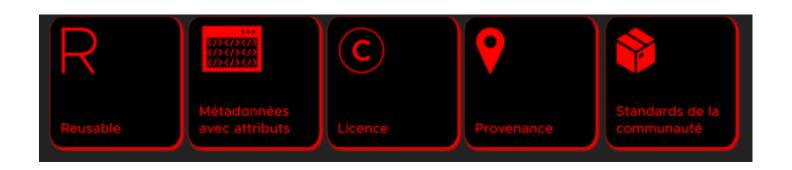
- Utiliser des formats ouverts et indépendants
- Utiliser des vocabulaires standards, des thésaurus, des ontologies... Le but étant de « parler le même langage ».
- Contextualiser les données en indiquant les liens vers d'autres données et métadonnées ou vers des publications.



Principe R: rendre vos données réutilisables pour de futures recherches ou d'autres finalités (enseignement, innovation, reproduction et transparence de la science)

Exemples de bonnes pratiques :

- Description riche (métadonnées)
- Provenance (auteurs, labos, objectifs, outil utilisé, méthode, etc.)
- Standards communautaires
- Licences d'utilisation



Qu'apportent les principes FAIR ?

Pour le chercheur

- Bonnes pratiques à adopter
- Un gain de temps dans la gestion du projet de recherche
- Meilleure visibilité et citation
- Ses données sont plus accessibles
- L'intégrité scientifique est facilitée
- Cela favorise les collaborations

Pour la communauté scientifique

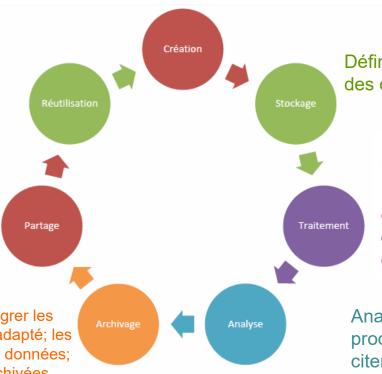
- Accès facile à des données publiques et réutilisables
- Gain de temps et d'argent : ne pas recréer des données déjà existantes
- Meilleure reproductibilité de la recherche

Le cycle de vie des données

Réaliser des analyses secondaires, un suivi de recherche, une nouvelle recherche, des évaluations de la recherche...

Définir les droits d'auteur et d'utilisation; contrôler les métadonnées et la documentation ; faire des copies de sauvegarde; publier et partager les données.

Collecter / créer les données ; capturer les données avec les métadonnées; acquérir des données existantes d'un tiers.



Définir un espace de stockage des données

Entrer, numériser, transcrire, traduire les données ; vérifier, valider, nettoyer, anonymiser ; dériver des données ; décrire et documenter les données ; gérer et sauvegarder les données.

Préparer les données pour la conservation ; migrer les données vers le meilleur format et un support adapté; les sauvegarder et les décrire dans un entrepôt de données; définir les accès et promouvoir les données archivées

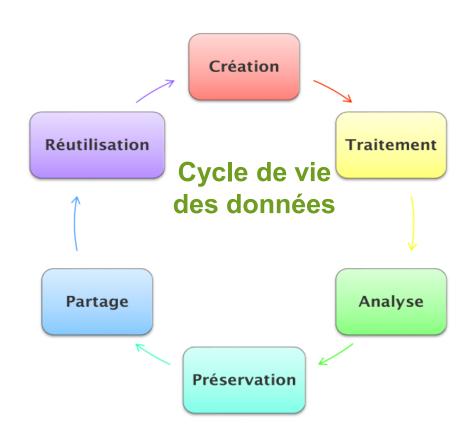
Archivage

Archiva

Le PGD : présenter ses données

S'appuie sur le cycle de vie des données

Document formalisé qui décrit la façon dont les données seront obtenues, traitées, documentées, analysées, organisées, stockées, sécurisées, préservées, partagées... au cours et à l'issue d'un projet.



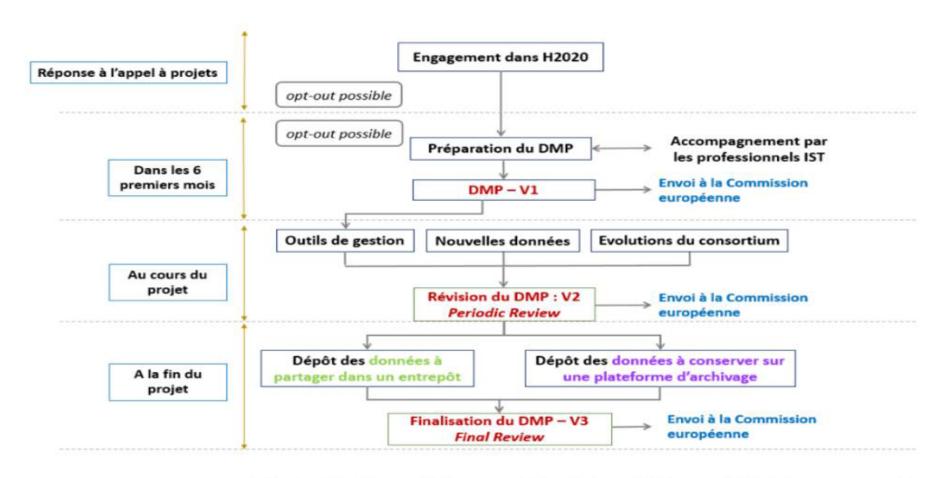
Aide à la mise en place de bonnes pratiques de gestion à toutes les étapes du cycle de vie des données

Partie 3

Les principes d'un PGD

Un document évolutif

- Document qui se complète au fur et à mesure du projet
- A envoyer aux financeurs à différents stades du projet



A quelles questions répond un PGD?



Comment la gestion et le partage des données sont-ils financés, en particulier à long terme?

Ressources

- Qui pourra accéder aux données?
- Les données seront-elles publiées ?
- Comment?
- Dans quel délai?
- Sous quelle licence?

Accès et partage des données

- En quoi consiste le projet ?
- Qui sont les partenaires ?
- Quelle est la politique de gestion des données ?
- Qui est responsable de la gestion des données ?

Responsabilités dans le projet

- Quelles données seront produites/utilisées au cours du projet ? (type, format, volume et accroissement...).
- Comment seront-elles produites ou transformées ?

Collecte des données

- Comment les données seront elles identifiées, décrites ?
- Quels standards de métadonnées utilisera t'on ?
- Comment seront générées les métadonnées ?

Documentation des données

What's plan?

^{3/1}s le cadre d'un projet de recherche ou n^{on.}

Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?

Sauvegarde des données

- Qui sera propriétaire des données produites?
- Des données externes serontelles utilisées ?

Propriété intellectuelle

- Des données sensibles serontelles produites ou utilisées ?
- Comment sera assurée leur anonymisation?

Ethique

Quel est le plan d'archivage et de préservation à long terme ?

Archivage et préservation des données

Les plans de gestion de données - S. Cocaud et D. L'Hostis, INRA, URFIST Paris - 05 avril 2019

PLAN DE GESTION DES DONNÉES MODÈLE ANR

 Modèle composé de 6 grandes thématiques illustrant les bonnes pratiques de gestion et de partage :

> Description des données et collecte ou réutilisation des données existantes

Documentation et qualité des données

Stockage et sauvegarde pendant le processus de recherche

Exigences légales et éthiques, codes de conduite

Partage des données et conservation à long terme

Responsabilités et ressources en matière de gestion des données

Modèle de DMP pour l'ANR

Contenu et rédaction de PGD

Eléments communs à tous les PGD

Informations administratives

Description des données

Documentations, métadonnées, standards

Aspects juridiques

Responsabilités

Sécurité des données

Stockage des données durant le projet

Partage des données avec le dépôt dans un entrepôt

Archivage pérenne

Coûts

1. Informations administratives

Un exemple de PGD : TRANSFUNERAIRE

https://dmp.opidor.fr/plans/6619/export.pdf

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 (français)" 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

siècle

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 bricolages 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 réinhumés. Dans ces situations mortuaires inédites, une compréhension empirique et précise des procédures matérielles et symboliques entourant le destin posthume des restes humains

et symboliques entourant le destin posthume des restes hum ou des corps absents, ambitionne ainsi de mieux cerner les mécanismes socioculturels 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)
- Données produites par les membres, dans le cadre du programme Transfunéraire (Collection)
- Fichier croisant les informations relatives aux données collectées et produites (Texte)

Contributeurs

Nom	Affiliation	Rôles
Clara Duterme		Responsable du plan
Elisabeth Anstett - https://orcid.org/0000-0002- 1548-4146		 Coordinateur du projet Personne contact pour les données (Données collectées, Données produites, Base de données)

Les responsabilités : différents acteurs impliqués

Source : Parcours interactif sur la gestion des données de la recherche, <u>DoRANum</u>

Archiviste

Communauté scientifique



Ingénieur projet

Informaticien



Service juridique

Professionnel de l'IST





Délégué à la Protection des Données

Les acteurs impliqués : exemple

Un exemple de PGD : TRANS FUNERAIRE

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

L'ingénieure de recherche du programme sera responsable de la gestion des données pour la durée du programme de recherche, de la mise en place du PGD et de ses mises à jour régulières, en lien avec les deux coordinatrices du programme.

Après la fin du programme de recherche, la responsabilité des données sera partagée par Misabeth Anstett (Directrice de recherche au CNRS, coordinatrice du programme Transfunéraire), Valérie Robin Azevedo (Professeure à l'université de Paris, coordinatrice du programme Transfunéraire) et Clara Duterme (IGR, programme de recherche Transfunéraire).

2. Description des données

Décrire tous les types de données (préexistantes et/ou produites) que l'on va "manipuler" pendant le projet.

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

Eléments à décrire :

- le type de données
- les formats
- la provenance des données
- les méthodes et les outils employés

2. Description des données : exemple

Extraits du DMP public <u>Transfunéraire</u>, 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 [...] »

3. Documentation et qualité des données

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

3. Documentation et qualité des données : exemple

Extraits du DMP public: PGD IMPRINT Jonathan Lenoir (CNRS)

"Certains fichiers sont également accompagnés 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 utilisé dans le cadre du projet est le standard "Ecological Matadata Language" (EML) avec une implémentation possible sous le logiciel R grace au package EML. [...]

4. Aspects juridiques : Exigences légales et éthiques

Vous devez indiquer dans le PGD 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

4. Aspects juridiques : Exigences légales et éthiques

Propriété intellectuelle des données

=> Se faire accompagner par un juriste est recommandé pour déterminer qui a le droit d'accéder aux données

Règle générale

Attribution de la propriété intellectuelle des données à l'établissement de tutelle des producteurs de données

Si partenariat

- Nécessité d'établir au préalable un accord de consortium
- Cas de collaboration entre secteurs public et privé (déploiement industriel, commercialisation, etc.)
- Cas de collaborations internationales

Penser à préciser la propriété intellectuelle des données et les responsabilités dans les PGD

Focus : La protection des données personnelles

Si des données protégées par des droits spécifiques seront utilisées au cours du projet : cas des données personnelles

- Principe 1 la Finalité : définir les objectifs du fichier
- Principe 2 La transparence : informer les individus de l'utilisation des données
- Principe 3 La conservation : limiter la conservation des données
- **Principe 4 Les droits** : respecter les droits des personnes
- Principe 5 La sécurité : sécuriser les données

Un contact dans les universités : le DPO Délégué à la protection des données (DPD) /Data Protection Officer (DPO)

- Le RGPD (Règlement général sur la protection des données)
 - Applicable dans les pays de l'Union européenne depuis mai 2018



Source: https://www.cnil.fr/fr/comprendre-vos-obligations/les-principes-cles

Démarches juridiques et éthiques

Logigrammes interactifs : questions juridiques et éthiques liées à mon projet de recherche

L'objectif de ces trois logigrammes est de vous guider dans les démarches juridiques et éthiques de votre projet de recherche.

Cliquez sur la question de votre choix pour commencer à parcourir le logigramme correspondant.

Quelles démarches lorsque je traite des données personnelles?

Dois-je demander le consentement des personnes concernées par mon projet de recherche?

Dois-je consulter le comité d'éthique de la recherche de mon établissement ?



Diapositive nº 1.1

Un outil pour vous aider à identifier les démarches juridiques et éthiques à entreprendre lors du traitement de données personnelles.

- Quand demander le consentement des personnes concernées par un projet de recherche;
- Reconnaître les situations dans lesquelles consulter un comité d'éthique de la recherche.







Comité d'éthique de la recherche de l'Université Rennes 2



Exigences légales et éthiques : exemple

Extraits du DMP public <u>Transfunéraire</u>

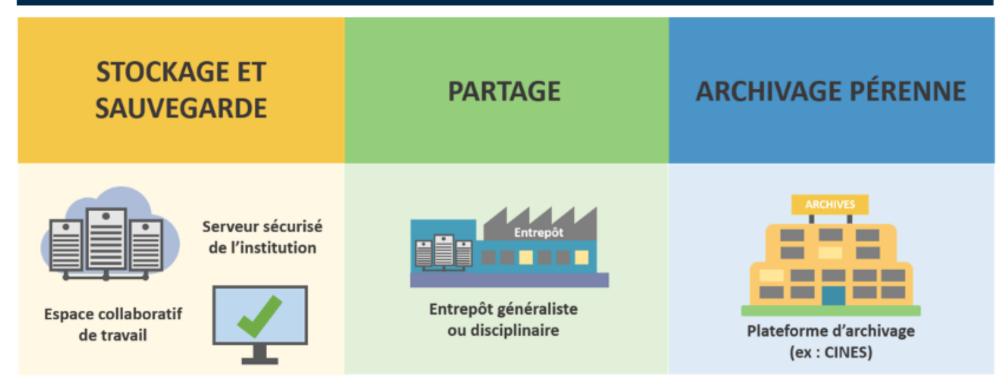
Partie 4 du PGD

"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 responsables, 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."

5. Stocker, sauvegarder, archiver ses données : 3 étapes distinctes

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



Introduction aux données de recherche, aux principes FAIR et aux PGD – <u>INIST</u> 2019

5. Stocker, sauvegarder au cours du projet



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

Vous devez préciser :

- Les supports de stockage utilisés
- Le système de nommage utilisé
- La volumétrie des données
- Comment sera organisée la sauvegarde des données
- Comment seront gérées les versions

5. Stocker, sauvegarder au cours du projet : nommage de fichiers



Bien nommer ses fichiers

- Se repérer entre les types de fichiers / les versions
- Regrouper les fichiers similaires et classer automatiquement ses fichiers de façon pertinente dans les dossiers
- Faciliter l'archivage et la préservation à long terme des fichiers

• Quelques règles et bonnes pratiques

- Définir un plan de nommage et une arborescence de fichiers
- Noms courts et significatifs (30 caractères maximum)
- Pas de caractères spéciaux ni accentués : lettres et chiffres seulement
- Bannir les espaces, vive les _ (tiret du 8)!



20150405_CR



20160310_CR



20160515 CR

- Dates au format ISO américain (AAAAMMJJ)
- Nombres sur 2 unités (ex : 01, 09...)
- Numéros ou codes de versions

5. Stocker, sauvegarder au cours du projet : exemple

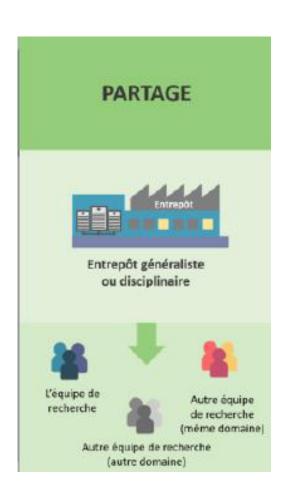
Extraits du DMP public <u>Transfunéraire</u>

Partie 3 Stockage et sauvegarde :

"Au cours des terrains de recherche, 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 passé quand il s'agit de données à caractère particulièrement confidentiel.[...]

Les chercheurs peuvent également déposer leurs données sur les platesformes 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 [...] "

5. Partager ses données



Garder le principe "aussi ouvert et FAIR que possible, aussi fermé que nécessaire" au coeur de votre démarche.

Vous devez préciser :

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

5. Partager ses données

Communicabilité des données

Cela dépend de :

- la nature ou le type des données
- l'origine des données
- leur (ré)utilisation

Peut être empêchée temporairement ou définitivement. Toute restriction : à expliquer dans le PGD.

Communication obligatoire pour certaines disciplines

- Données géographiques
- Données environnementales

Communication sous conditions

- Données protégées par le droit d'auteur ou par contrat
- Données personnelles
- · Statistiques...

Communication interdite par principe

- Secrets professionnels
- · Secrets défense
- Sécurité de l'établissement

source : Inist Cnrs, webinaire Plan de Gestion de Données

2 exemples de licences standard



1 <u>Licence ouverte</u>Etalab

=> pour faciliter la réutilisation des données publiques mises à disposition gratuitement.

Attribuer une licence à son jeu de données

2 <u>Licence CC creative</u> commons



Licences Creative Commons (CC-By)

→ notamment 3 possibilités de combinaisons :

• NC : pas d'utilisation commerciale

SA : partage dans les mêmes conditions

ND : pas de modifications

3 Aussi des Licences spécifiques pour les logiciels et les bases de données

voir fiche CIRAD : https://coop-ist.cirad.fr/etre-auteur/utiliser-les-licences-cc creative-commons/4-les-6-licences-cc

5. Partager ses données : un exemple



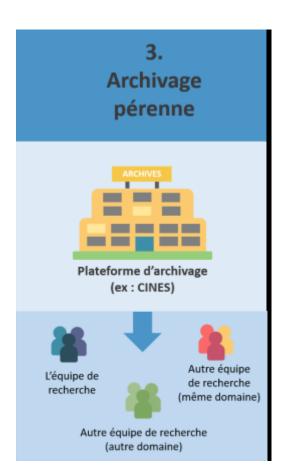
Extraits du DMP public <u>Transfunéraire</u>

Partie 5 Partage des données :

"Les données seront partagées à la fin du programme de recherche sur 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. [...]

5. Préservation à long terme des données

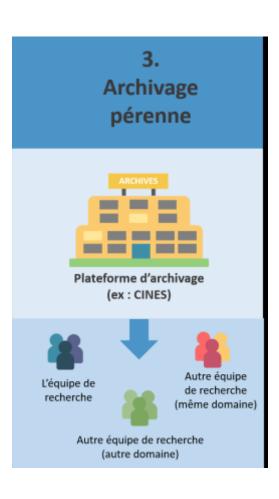


Vous devez dire si un plan de conservation des données sur le long terme a été établi.

Vous devez préciser :

- Quelles sont les données (+ métadonnées associées) sélectionnées pour l'archivage
- Quelle est la volumétrie pour ces données
- La durée de conservation à long terme
- Le budget prévu pour l'archivage

5. Préservation à long terme des données : un exemple



Extraits du DMP public <u>Transfunéraire</u>

Partie 5b Partage des données sur le long terme

"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.[...]"

6. Evaluer le coût de la gestion des données

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 et/ou la préservation

Exemples de coûts : frais de stockage, coût de matériel, temps de personnel, frais d'entrepôt et d'archivage, etc.

à noter : certains coût sont éligibles dans le cadre des projets financés.

6. Evaluer le coût de la gestion des données : un exemple

Extraits du DMP public: PGD IMPRINT

« 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). [...] »

PGD de projet ou PGD entité?

PGD de projet

- Périmètre thématique
- Durée déterminée
- Organisé par « produit de recherche » (données, logiciels, protocole, échantillon)
- S'applique à l'ensemble des jeux de données du projet

PGD Entité

- Périmètre plus large
- Durée non déterminée
- Organisé par mode d'obtention, nature des données, mode de gestion des données, type d'activité



INRAE trame projet
INRAE trame entité

Modèles de PGD

Quelques modèles, issus de Sylvie Cocaud (INRAE), <u>Rédiger un plan de gestion de données</u>, 2022

S'appuyer sur des modèles et les aides associées

INRAO

- Informations sur le PGD
- Informations sur le projet ou la structure
- Présentation générale des données
- Organisation et documentation des données
- Droits de propriété intellectuelle
- Confidentialité
- Stockage et sécurité des données
- Partage des données à l'issue du projetArchivage et conservation des données après la fin du projet

Trame projet v2 43 questions Trame structure 39 questions



- Description des données et collecte ou réutilisation de données existantes
- Documentation et qualité des données
- Stockage et sauvegarde pendant le processus de recherche
- Exigences légales et éthiques, codes de conduite
- Partage des données et conservation à long terme
- Responsabilités et ressources en matière de gestion des données

ANR Modèle de PGD 15 questions Guide en français



- Résumé descriptif des données
- Données FAIR
- Allocation de ressources
- Sécurité des données
- Aspects éthiques
- Autres

horizon europe

- Data Summary
- FAIR data
- Other research outputs
- Allocation of resources
- Data security
- Ethics
- Other issues

H2020 FAIR DMP (v.f.)

9 questions Horizon Europe Data

this template is recommended for Horizon Europe beneficiaries Verburg, M., & Grootveld, M. (2022, février 22). FAIR-Aware Additional guidance to the Science Europe DMP assessment rubric. https://doi.org/10.5281/zenodo.6088215

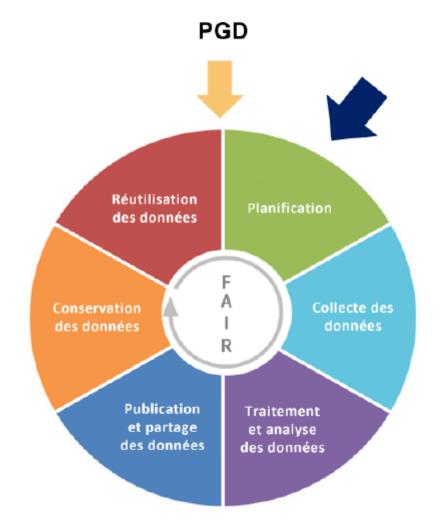


- Description des données et collecte ou réutilisation de données existantes
- Documentation et qualité des données
- Exigences légales et éthiques, codes de conduite
- Traitement et analyse des données
- Stockage et sauvegarde pendant le processus de recherche
- Partage des données et conservation à long terme

Science Europe
« modèle structuré »
(+ version anglaise) 21

Planifier la gestion des données

Place dans le cycle de vie des données





Partie 4

Comment rédiger un PGD en appliquant les principes FAIR

?

Découverte de DMP OPIDOR

DMP OPIDOR - Guide d'utilisation



https://dmp.opidor.fr

Outil d'aide à la création en ligne de DMP, permettant d'anticiper la gestion des données dans le cadre de projets de recherche

- Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers
- Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données
- Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche
- Collaboratif : il facilite les échanges entre les partenaires d'un même projet et les services d'accompagnement
- Outil open source, développé par une communauté internationale, basé sur le code commun <u>DMPRoadmap</u> du Digital Curation Center et de l'UC3



DMP OPIDOR



18/04/2024 DMP OPIDoR

https://opidor.fr/dmp-opidor-un-nouveau-parcours-de-creation-de-plan/



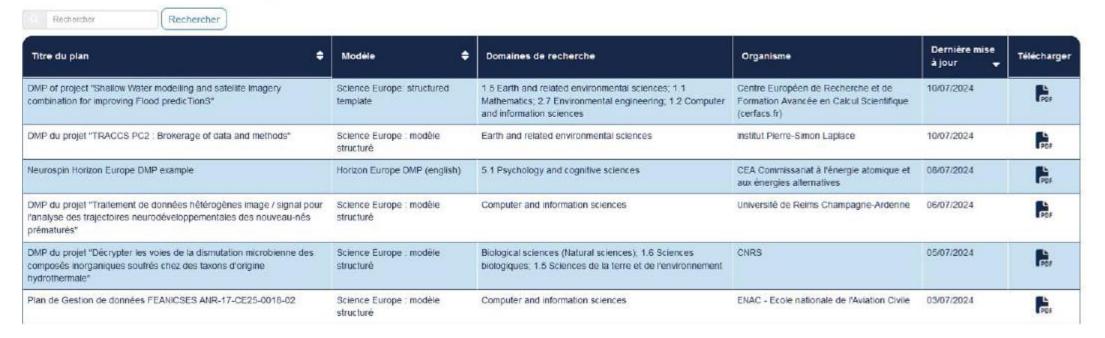
		■ FR ∨	MON ESPACE
	Accès institutionn	el e connecter	
Créer un compte	Accès individuel Pas ancore inscrit ? Cre * Courriel	bar mon compti	2
* Prénom	^ Mot de passe		
* Nom de famille	□ Se souvenir de m	OI	
* Courriel	Se connecter Mot de passe oublié ?		
* Organisme Commencaz à taper pour voir une liste de suggestions.			SUPPORT TECHNIQUE
* Mot de passe			
□ J'accepte les conditions générales d'utilisation Les informations requellées vois concernant vont faire l'objet d'un traitement d'estré à l'inter-CNRS (UURTS) et au correspondent Informatique et Literrés. Pour la finable airverte : gestion des complets des utilisateurs. Les destinateires de ses données sent les administrateurs DMP OPIDER de l'inité CNRS. Le durée de sonservetion des données est de 5 ans suite à la demaine commune. Vous bénéficies d'un drait d'asvès, de residination de portabilité, d'efficientent de cofficer ou une limitation du traitement Vous pour ce vous opposer au Indiament des gonnées sous nonventer d'éléposes ou deut de celtrar vetre consententent à tout moment en vous acrossant à dings opidonigment. Vous avez le possibilité l'infrinduire une réclamation augrés d'une autorité de contrôle. Créer un compte			





PGD publics

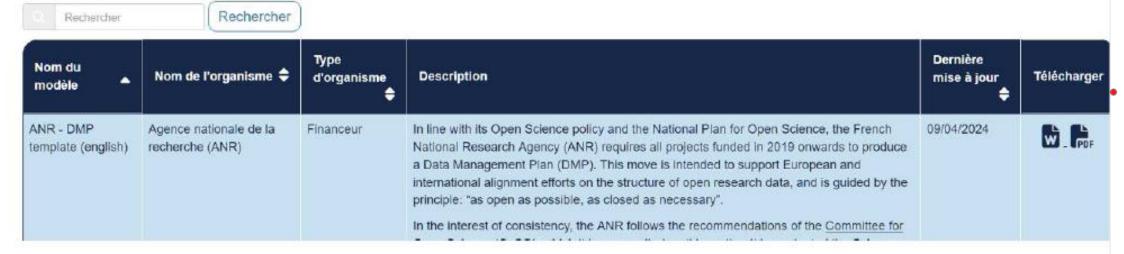
Les PGD publics sont créés à l'aide de DMP OPIDOR et partagés publiquement par leurs propriétaires. Ils n'ont pas été vérifiés pour leur qualité, leur exhaustivité ou leur adhésion aux lighes directrices des financeurs.





Modèles de PGD

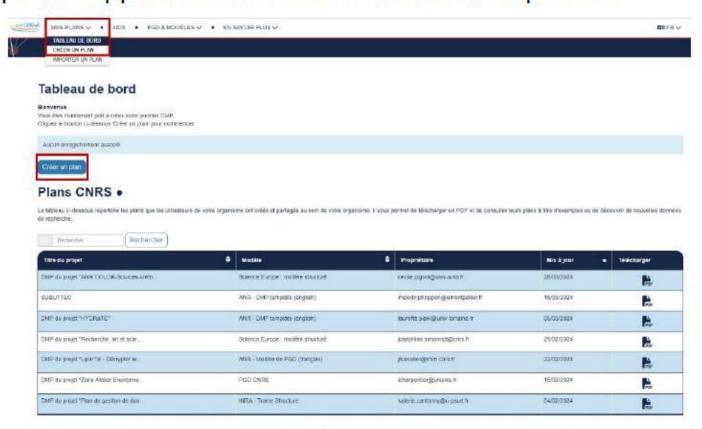
Modèles de DMP proposés par les financeurs ou par les organismes de recherche, disponibles dans DMP OPIDoR. Vous pouvez télécharger ces modèles et les recommandations associées.



Modèles consultables librement



Quand vous vous connectez, vous êtes redirigé vers votre tableau de bord personnel et l'onglet "Mes plans" apparait dans la barre de menu supérieur :



Création d'un plan de gestion de données à partir du tableau de bord

Cliquez sur " Créer un plan ". Vous accédez alors au formulaire de création de plan.



Créez votre plan en 3 étapes



Création d'un plan de gestion de données à partir de l'onglet " Créer/Importer "

Créez votre plan en 3 étapes



3. Choisissez le modèle parmi les propositions

Créez votre plan en 3 étapes





Onglet 'Informations générales'

Cet onglet est commun à tous les plans quel que soit le modèle choisi.

Il vous invite à renseigner des informations pour décrire le projet de recherche et le plan. Il comprend 3 parties.







DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"



Cliquez ici pour sélectionner les recommandations appliquées à votre plan

Vous trouverez ci-dessous la liste des organismes proposant des recommandations et des conseils pour vous quider dans la rédaction de votre plan tout en respectant leur politique de gestion de données. Vous avez la possibilité de sélectionner 6 organismes au maximum. Puis cliquez pour enregistrer votre sélection.



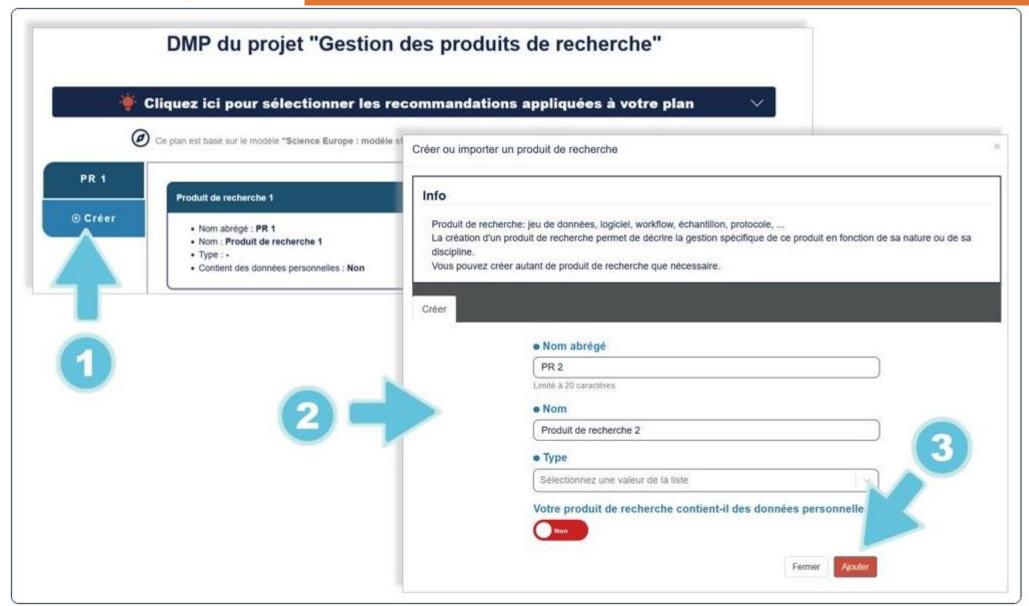














JEU DONNÉES

LOGICIEL 2

IMAGE 3

⊕ Créer







DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"

Liste des personnes contribuant à la gestion des produits de recherche au cours d'un projet et leurs rôles.

L'attribution d'un rôle à une personne se fait dans l'onglet "Informations générales", dans l'onglet "Rédiger" pour les modèles structurés ou dans l'onglet "Produits de recherche" pour les autres modèles.

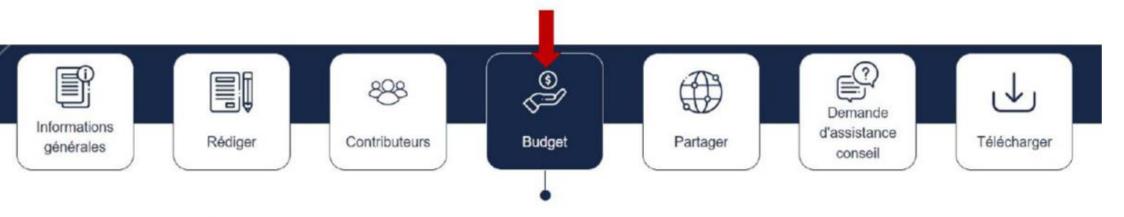
Nom	Affiliation	Rôles attribués (Produits de recherche associés)	Actions
Lafosse Yvette - https://orcid.org/0000-0002-3972-1687	Institut de l'Information Scientifique et Technique - https://ror.org/02mn0vt57	Membre du projet (Jeu données 1)	& ×
Poirot Léa Personne + ORCID	Institut de l'Information Scientifique et Technique - https://ror.org/02mn0vt57	Personne contact (Logiciel 2, Jeu données 1, Image 3) Responsable du plan	♂ ×
VAN HELDEN Jacques - https://orcid.org/0000-0002-8799-8584	Institut français de bioinformatique - 201321719F	Coordinateur du projet	♂ ×

Ajouter personne ou organisation

Produits de recherche concernés

Tableau des contributeurs





DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"

Liste des coûts engagés dans la gestion des données. La création d'un coût s'effectue dans l'onglet "Rédiger".

Type de coût (Titre)	Montant	ntant Etape du cycle de vie		
Stockage (Estimation du coût de stockage et calcul auprès du mésocentre MESO@LR)	137.5 EUR	Coût total de stockage (coût calculé automatiquement) - Jeu données 1		
Stockage (Frais d'archivage à long terme)	10000 EUR	Coûts liés à la conservation à long terme des données - Jeu données 1		
Matériel informatique (Frais d'hébergement de l'entrepôt)	1000 EUR	Coûts liès au partage des données - Jeu données 1		
Stockage (Estimation du coût de stockage et calcul auprès du mésocentre MESO@LR)	125.0 EUR	Coût total de stockage (coût calculé automatiquement) - Image 3		
Quel type de coût ?	collecte	ape du cycle de vie : , documentation, ockage?	Pour quel produit de recherche ?	

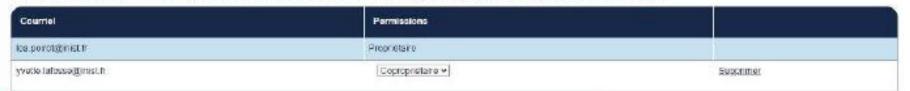




DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"

Gerer les collaborateurs

Invitor des passionnes à tre, modifier de administrar votre plan. Les invites recevrent une notification par message électronique indiquant qu'ils ont accès à ce plan.

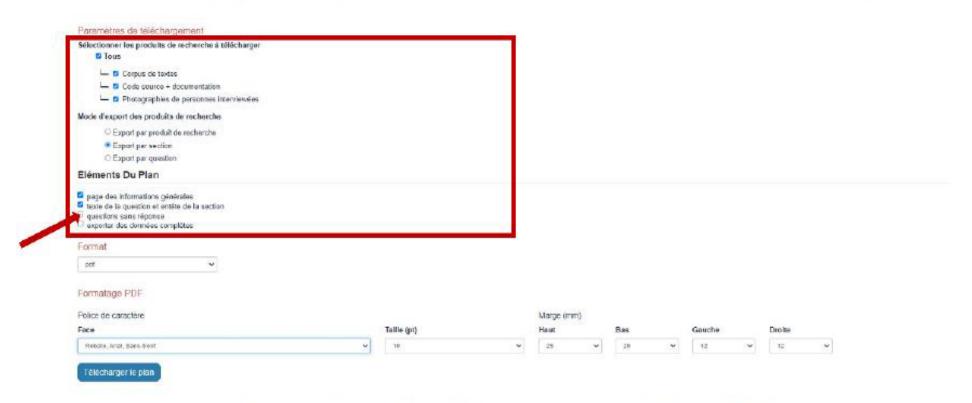








DMP du projet "Mettre le plan de gestion des données automatisé dans les mains des biologistes"



Paramètres de téléchargement d'un PGD

Des ressources



La plateforme Doranum



DES RESSOURCES POUR ACCOMPAGNER LA COMMUNAUTÉ SCIENTIFIQUE DANS LA GESTION ET LE PARTAGE DE LEURS DONNÉES





THÉMATIQUES ABORDÉES

PARCOUREZ LES RESSOURCES CLASSÉES PAR THÉMATIQUES

https://doranum.fr/

ENJEUX & BÉNÉFICES

Pourquoi partager les données ? Qu'est-ce que l'Open Science ?



ASPECTS JURIDIQUES, ÉTHIQUES, INTÉGRITÉ SCIENTIFIQUE

Que puis-je partager, réutiliser ? Quelles pratiques devrais-je respecter ?



PLAN DE GESTION DE DONNÉES

Pourquoi et comment rédiger un plan de gestion des données ?



MÉTADONNÉES

Comment décrire les données ?



IDENTIFIANTS PÉRENNES

Comment associer durablement des données à son auteur?



DÉPÔT & ENTREPÔTS

Comment et où déposer mes données?



STOCKAGE & ARCHIVAGE

Quelles données conserver à long terme et comment ?



DATA PAPERS & DATA JOURNALS

Comment publier mes données comme un article scientifique?



ACCÈS & VISUALISATION

Où et comment extraire et visualiser les données qui m'intéressent ?



Un jeu pour acquérir de bonnes pratiques de gestion et partage des données de la recherche



Jeu GopenDoRe : le jeu coopératif pour acquérir des bonnes pratiques de gestion et de partage des données de la recherche

Valérie Gadrat¹; Yvette Lafosse¹ (D); Claire Sowinski¹ (D); Coralie Wysoczynski¹ (D)

Show affiliations

Le partage des données va de pair avec une bonne gestion des données. En effet, pour pouvoir être diffusées, les données doivent avoir été au préalable décrites, documentées, conservées dans un format réutilisable, etc. Mais comment gérer et partager ses données ? Quels sont les bonnes pratiques à connaître ?

GopenDoRe a été créé par l'équipe DoRANum pour animer de façon ludique les formations sur la gestion et le partage des données de la recherche.

Cette version est mise à la disposition de tous. Elle est réutilisable dans le cadre de formations présentielles. Les questions et réponses proposées dans ce jeu peuvent être complétées par l'animateur de la formation.

Le jeu est aussi disponible sur DoRANum, rubrique "Supports de formation"

https://zenodo.org/records/3533025#.YgN6wurMJAU

Version du jeu en ligne



Jeu GopenDoRe

Bienvenue dans ce jeu

L'objectif est d'acquérir de bonnes pratiques concernant les données de la recherche, plus particulièrement leur gestion et leur partage.

Commencez







Une autoformation sur Callisto

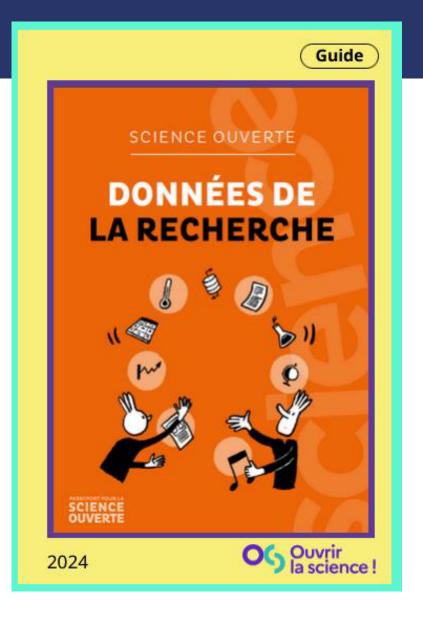


Parcours pédagogique Callisto

https://callisto-formation.fr/course/view.php?id=160

Science ouverte – Données de la recherche

Guide de la collection
Ouvrir la science (MESR),
2024



Un guide qui aborde les principales notions relatives à la gestion et à la diffusion des données de la recherche.

LES DONNÉES DE LA RECHERCHE : DE QUOI PARLE-T-ON ?

Les données de la recherche dans tous leurs états

Statut juridique des données de la recherche

POURQUOI LES DIFFUSER ? COMMENT LES DIFFUSER ?

Préparer la diffusion de ses données

Diffuser les données de la recherche

Questions pratiques

ET APRÈS ? PRÉPARER L'AVENIR

Valoriser ses données

Lier ses données au reste de ses travaux scientifiques

Identifier les différentes versions d'un jeu de données

Archiver de manière pérenne

POUR ALLER PLUS LOIN SOURCES
GLOSSAIRE



Ateliers de la donnée (Recherche Data Gouv)

Des services généralistes pour accompagner la gestion des données sur tout le territoire : https://recherche.data.gouv.fr/fr/ateliers-de-la-donnee

- Objectifs : proposer un maillage national de service de proximité géographique à destination des équipes de recherche et leur apporter une première expertise dans la gestion raisonnée des équipes de recherche
- 22 Ateliers de la donnée

exemple de l'<u>Atelier Rennais "ARDoISE</u>"

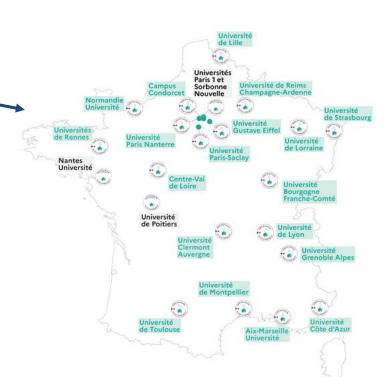
(Atelier rennais de la Donnée : Information et Soutien aux Équipes de recherche)



exemples d'<u>Ateliers ARDoISE</u>:

27 mars 2025 : Parcours juridique et éthique lié au traitement de données personnelles dans un projet de recherche

12 juin 2025 : Le partage du code pour la réplicabilité : Méthodes et outils pour rendre le code source réutilisable





Merci beaucoup pour votre attention!

florence.thiault@univ-rennes2.fr remi.joinville@univ-rennes2.fr



évaluation: https://enquetes.univ-rennes2.fr/limesurvey/index.php/791465?lang=fr