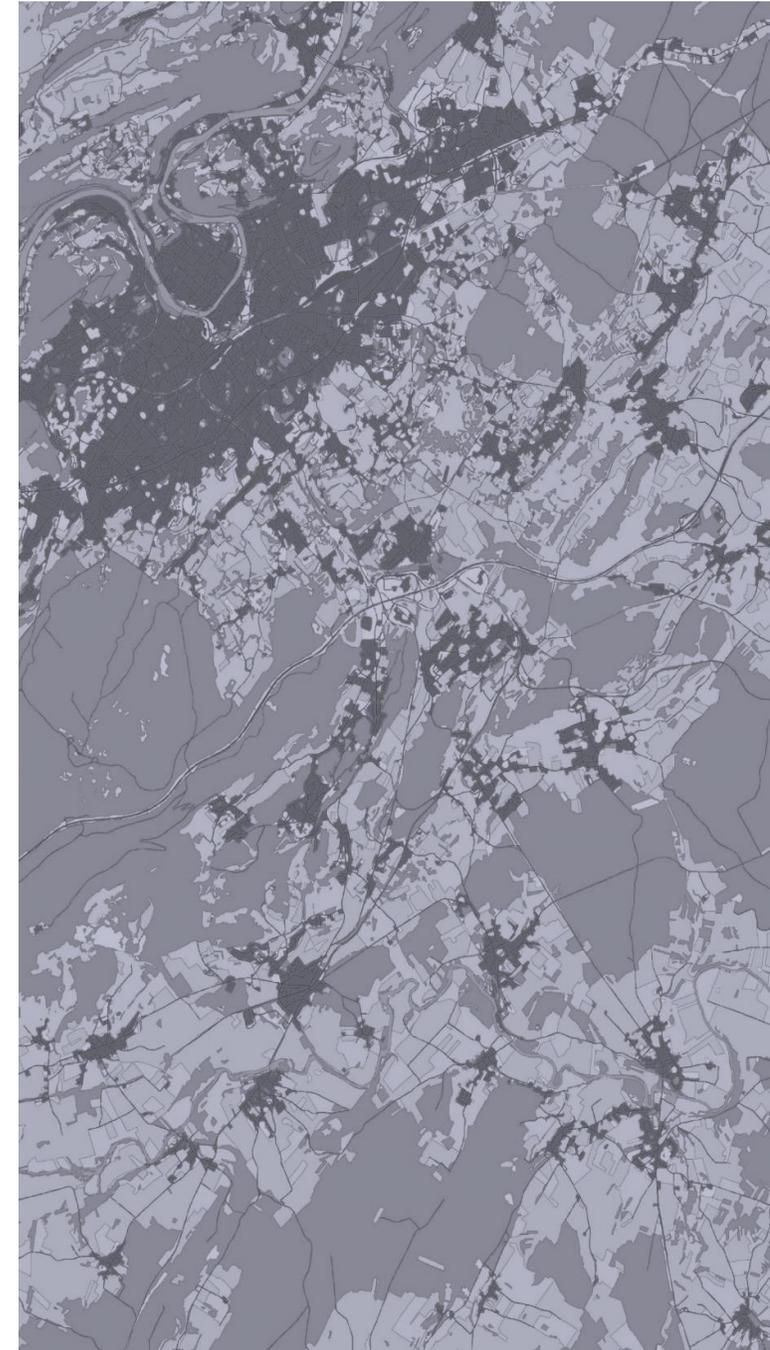


« Optimiser la gestion des données de votre structure avec le DMP entité »

La rédaction d'un *plan de gestion de données entité* pour le laboratoire Théma



La gestion des données à Théma ?

1 groupe de travail :

- 2 informaticiens
- 1 géomaticien
- 1 Ingénieur en production et analyse des données
- 1 documentaliste

Quels objectifs ?

- Apporter une aide technique aux membres du laboratoire
- Harmoniser les pratiques
- Sensibiliser aux bonnes pratiques

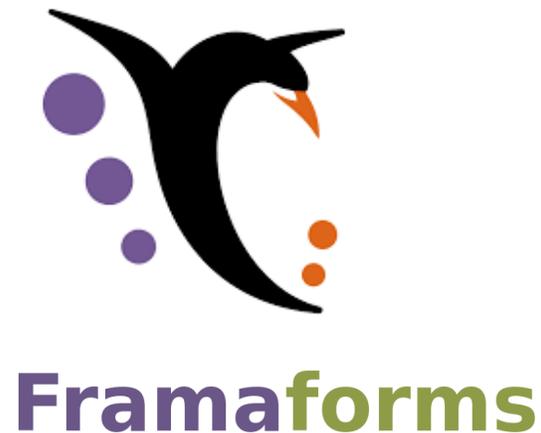


- Aider à la rédaction des métadonnées pour décrire des données
- Indiquer des espaces de stockage et de sauvegarde des données
- Signaler les jeux de données sur le portail Dat@UBFC
- Rédiger un plan de gestion de données entité pour le laboratoire

Le DMP entité à ThéMA ?

- Formaliser la gestion des données au laboratoire et donner des directives communes
 - Faciliter la rédaction pour les chercheurs du laboratoire
 - Encourager la rédaction de DMP au sein du laboratoire
- 
- Une trame commune
 - Adaptée aux recherches et aux types de données produites à ThéMA
 - Transposable à chaque projet de recherche

Etape 1 : réalisation d'une enquête à destination des membres du laboratoire



Pouvez-vous citer quelques formats utilisés pour vos données?

Ex : TIFF, JPEG, CSV, XLSX, Shapefile, GeoPackage...

Avez-vous déjà collecté au moins une fois des données à caractère personnel ?

- Oui
 Non

Habituellement, pendant le temps des recherches, où stockez-vous vos données ?

- PC professionnel
 Disque dur externe
 Serveur du laboratoire
 Cloud de l'Université ou du CNRS
 Plateforme privée (ex : Google drive, One drive, Arcgis online)
 PC personnel

Autre :

Présentation du questionnaire pour le public (Framaforms).

VOS DONNEES

Quels types de données produisez-vous ? (Choix multiples)

- Données déclaratives (ex : enquêtes)
- Données textuelles
- Images aériennes (ex : drones, télédétections)
- Mesures physiques (ex : température, distance, lidar, comptage)
- Données dérivées ou compilées (data mining, text mining...)
- Données de simulation
- Données spatiales
- Photographies (ex : paysage)
- Sons et vidéos
- Autre :

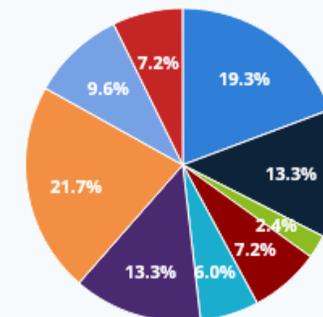
Pouvez vous préciser le format de vos données (TIFF, JPEG, CSV, xlsx, Shapefile, GeoPackage...)

Avez-vous déjà collecté au moins une fois des données à caractère personnel ?

- OUI ou NON

Quels types de données produisez-vous ?

[Chart options »](#)



Données déclaratives (ex : enquêtes)	16
Données textuelles	11
Images aériennes (ex : drones, télédétections)	2
Mesures physiques (ex : température, distance, lidar, comptage)	6
Données dérivées ou compilées (data mining, text mining...)	5
Données de simulation	11
Données spatiales	18
Photographies	8
Sons et vidéos	6

Résultat extrait de l'outil Framiforms

SAUVEGARDE ET STOCKAGE

Habituellement, où stockez-vous vos données ? Pendant le temps des recherches / A la fin du projet.

- PC professionnel
- Disque dur externe ou un autre support
- Serveur du laboratoire
- Cloud de l'Université ou du CNRS
- Plateforme privée (type google Drive, onedrive, Arcgis Online)
- PC personnel
- Autre :

Comment évaluez-vous le risque de perdre des données ?

> Très probable – impossible (1 à 5)

Avez-vous déjà perdu des données ?

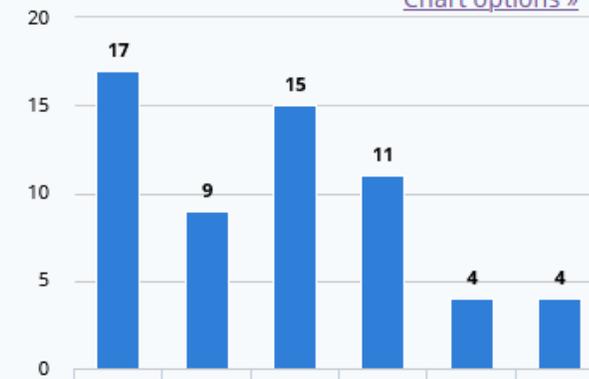
- Oui ou Non

Pensez-vous qu'un dispositif mutualisé d'archivages pour le laboratoire serait intéressant ?

- Oui ou non

Habituellement, à la fin de votre projet, où stockez-vous vos données ?

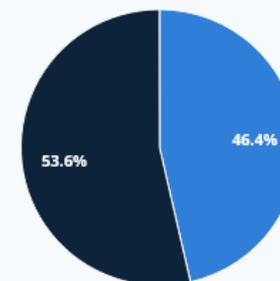
[Chart options »](#)



PC professionnel	17
Disque dur externe	9
Serveur du laboratoire	15
Cloud de l'Université ou du CNRS	11
Plateforme privée (ex : Google drive, One drive, Arcgis online)	4
PC personnel	4

Avez-vous déjà perdu des données ?

[Chart options »](#)



Oui	13
Non	15

DESCRIPTION DES DONNEES

Utilisez-vous une méthode « normalisée » pour nommer vos fichiers et dossiers ?

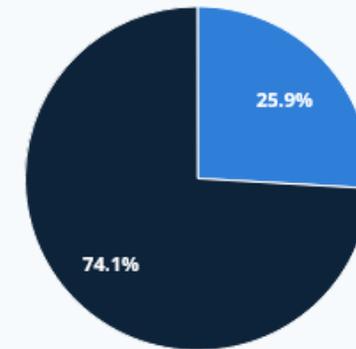
- Oui ou Non

Quand vous produisez un jeu de données, rédigez-vous un document de métadonnées de type « lisez-moi » pour les décrire ?

- Oui ou Non

Utilisez-vous une méthode « normalisée » pour nommer vos fichiers et dossiers ?

[Chart options »](#)



Oui	7
-----	---

Non	20
-----	----

Résultat extrait de l'outil Framiforms

Avez-vous déjà rédigé un plan de gestion de données¹ (DMP) ?

- Oui ou Non

Si oui :

Avez-vous eu des difficultés à le faire ?

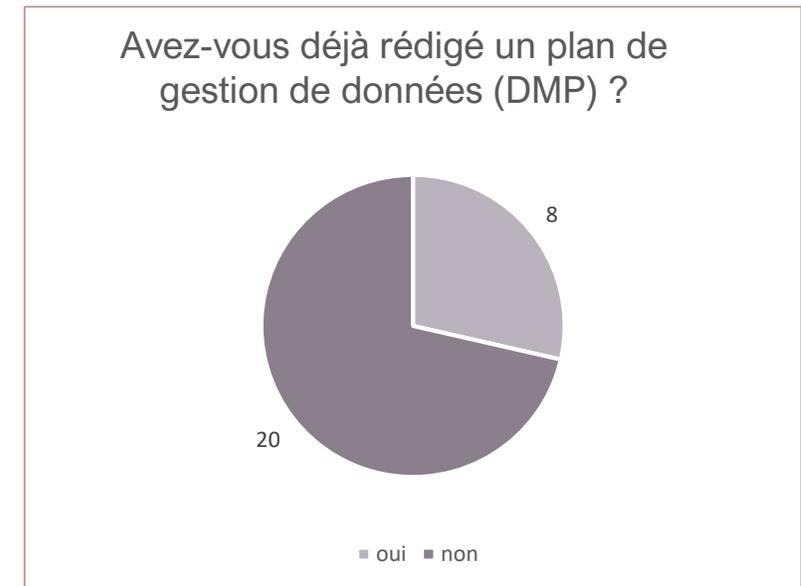
- Oui ou Non

Avez-vous utilisé l'outil DMP Opidor ?

- Oui ou Non

Pensez-vous qu'un « DMP modèle » adapté au laboratoire vous serait utile ?

- Oui ou Non



DIFFUSION ET PARTAGE DES DONNEES

Actuellement, pensez-vous que vos données seraient réutilisables par quelqu'un d'autre ?

- Oui ou Non

Avez-vous l'habitude de diffuser vos données ?

- Oui ou Non

Dans quelle mesure êtes-vous d'accord ou pas d'accord avec ce qui suit ?

Tout à fait d'accord / Plutôt d'accord / Sans opinion / Plutôt pas d'accord / Pas du tout d'accord

- Je suis prêt.e à partager mes données de recherche avec d'autres
- Je souhaiterais pouvoir réutiliser les données d'autres chercheurs ou chercheuses
- Je pense que le partage de mes données donnera une meilleure visibilité à mes travaux
- Je pense que le partage de mes données peut être l'occasion de nouveaux contacts, de nouvelles collaborations

Avez-vous déjà publié un data paper² ?

² Un *Data paper* (ou *Data descriptor*, *Data article*, *Data Briefs*, *Resource Announcements*, *Data Resource Profile*) est un article qui décrit un jeu de données (*data*, *data set*), la méthode ayant permis de l'obtenir et le potentiel de réutilisation de ce jeu. Le *Data paper* ne décrit pas de résultats de recherche et ne contient ni discussion, ni conclusion

- Oui ou Non

Vous arrive-t-il de consulter des données existantes via des catalogues dédiés ? Si oui, sur quel catalogue ?

- Oui ou Non

Nom du ou des catalogues :

Avez-déjà réutilisé un jeu de données produits par un autre membre du laboratoire ?

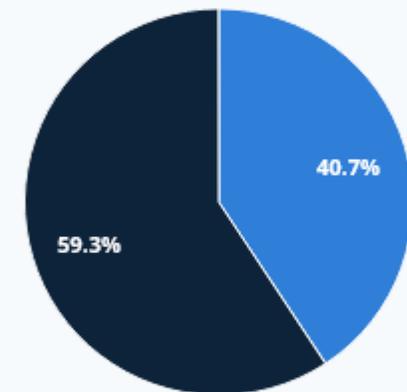
- Oui ou non

Quel(s) besoin(s) en matière d'accompagnement et de formation souhaiteriez-vous ?

Commentaires, remarques et suggestions

Avez-déjà réutilisé un jeu de données produits par un autre membre du laboratoire ?

[Chart options »](#)



Oui	11
Non	16

Bilan de l'enquête

- 28 réponses complètes
- Le sujet ne semble pas maîtrisé par les répondants
- Un bon point de départ pour orienter le service d'appui à la gestion des données et pour sensibiliser les collègues

« Une super idée :), merci pour cette initiative. Je gère mes données à ma manière je suis très intéressé d'améliorer mes pratiques et l'uniformisation de toutes les données que j'ai.. ».

« Un accompagnement serait le bienvenu si besoin de rédiger un DMP (maintenant quasi systématiquement demandé dans le cadre des AAP). »

Etape 2 : la recherche de DMP existants

FICHE PRATIQUE : faire un PLAN de GESTION de DONNÉES (PGD ou DMP) à l'échelle d'une structure (établissement, laboratoire, équipe de recherche)

3 QUESTIONS pour éclairer le sujet :

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

C'est un document de référence pour la structure (établissement, laboratoire, équipe) sur la gestion des données qu'elle produit dans le cadre de son activité de recherche. Il est élaboré de manière collective, dans une démarche de définition d'une politique de gestion des données de recherche pour la structure. Le « PGD structure » a un périmètre plus large que le « PGD projet », une durée de vie plus longue.

Que disent les textes officiels ?

Les établissements publics et fondations reconnues d'utilité publique [...] 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.

Décret n°2021-1572 du 03/12/2021 relatif au respect des exigences de l'intégrité scientifique, article 6

Pourquoi élaborer un plan de gestion de données structure ?

- ☞ Pour harmoniser les pratiques de gestion/ouverture des données au sein de la structure
- ☞ Pour monter en qualité sur la gestion des données
- ☞ Pour faciliter la rédaction des « PGD projets » pour les chercheurs·ses et personnels de recherche, et respecter les exigences des financeurs de la recherche publique
- ☞ Pour préserver le patrimoine scientifique de la structure :
 - conserver et valoriser les données produites au cours du temps
 - documenter les données produites
 - ne pas « perdre » les données produites par les doctorant·e·s à leur départ
- ☞ Pour adopter une démarche « science ouverte » et respecter les principes FAIR
- ☞ Pour faciliter la réutilisation des données de recherche : le 1er ré-utilisateur potentiel est le chercheur·se à l'origine des données, ou bien ses collaborateur·rice·s

Par où COMMENCER ?

Identifier les données générées par la structure de recherche...

- Données d'observation ? (capturées en temps réel : imagerie médicale, données d'enquête, relevés de capteurs...)
- Données expérimentales ? (obtenues à partir d'équipements de laboratoire : chromatogrammes, séquençage ADN...)
- Données computationnelles ? (générées par des modèles informatiques ou de simulation)
- Données dérivées ou compilées ? (issues du traitement de données brutes : feuille de texte, bases de données compilées...)
- Données de référence ? (collection de jeux de données revus par les pairs et mis à disposition : GenBank, base de données de l'IGN, archives d'images historiques...)

... et en déduire les éventuels points de vigilance :

- Volumétrie importante ?
- Accès fréquents et/ou traitements intensifs sur les données ?
- Données à caractère personnel ? Voir données sensibles ? (données de santé, relatives à l'orientation sexuelle, politique ou syndicale, relatives à l'appartenance à un groupe ethnique, relatives à des infractions ou condamnations judiciaires)
- Données qui relèvent de la PPST ? (protection du potentiel scientifique et technique de la nation)
- Données qui peuvent donner lieu à des innovations, des brevets ? Générées dans le cadre de partenariats avec des entreprises privées ?
- Réutilisation de données produites par d'autres, donc éventuels problèmes de droits d'auteur ou de propriété intellectuelle ?
- Données qui mettent en jeu le droit à l'image des personnes ?

Choisir un mode d'organisation pour le PGD :

- Par type d'activités de la structure ?
- Par type de données produites ?
- Selon le mode d'obtention, la nature ou le format des données ?

Exemple : les données produites en ZRR ne relèvent pas toutes de la PPST, elles peuvent donc faire l'objet de gestion différenciée



PRODIG - UMR: PRODIG- Modèle de DMP

Identifiants du DMP

Identification du document livré - Modèle du DMP

Exemple de réponse:
ANR - Modèle de DMP (français)

Identification du document livré - Historique des mises à jour du DMP

Préciser les dates de transmission au bailleur du DMP aux différentes étapes (en général au début, au milieu, à la fin)

Exemple de réponse:
- Version 3 : 27/04/2020
- Version 2 : 5/12/2019
- Version 1 : 1/3/2018 - date de dépôt effectif à la structure exigeant le DMP en tant que livrable (en général à remettre dans les 6 mois suivant le début du financement, cf. cas ANR et H2020)

Identification du document livré - Date de dernière modification du DMP

Identification du document livré - Créateur(s) du DMP

Prénom Nom, affiliation principale (Institution, Laboratoire, Ville, Pays OU identifiant type ORCID)

Exemple de réponse:
Jean-Yves Menut, CNRS, Prodig, UMR 8586, Paris, France
OU Jean-Yves Menut, ORCID 0000-0001-9010-500X

Identification du document livré - Droits d'auteur

Exemple de réponse:
Les créateurs de ce plan acceptent que tout ou partie du texte de ce plan soit réutilisé et personnalisé si nécessaire pour constituer un autre plan.

Identification du document livré - Contact pour les Données : Adresse mail du créateur du DMP

Identification du projet de recherche concerné - Intitulé du projet de recherche : Titre et acronyme du projet

Exemple de réponse:
Caractérisation des Risques de submersion sur des Sites Sensibles, CRISSIS

Identification du projet de recherche concerné - Financement(s)

Exemple de réponse:
Agence nationale de la recherche (ANR)

Identification du projet de recherche concerné - Référence ou Numéro de subvention du projet ou de la convention de financement

Exemple de réponse:
ANR-29-CE35-0008, AAP ACC 2017

Identification du projet de recherche concerné - Dates de début et de fin du projet : Dates de début et fin du financement ou de début et fin des travaux de recherche en l'absence de convention de financement

Identification du projet de recherche concerné - Chercheur(s) Principal(aux) ou coordinateurs scientifiques : Prénom Nom, (Institution d'affiliation, Laboratoire, Ville, Pays OU identifiant type ORCID)

Fiche pratique : faire un plan de gestion de données.
Atelier de la donnée. Dat@UBFC. 2023

DMP Structure de l'UMR Prodig. Mis à disposition sur DMP
Opidor

Etape 3 : rédaction du DMP entité de ThéMA

- **En noir** : la structure du document – *une proposition de rubriques qui doivent figurées dans le DMP*
- *En gris* : Les explications - les informations attendues dans chaque rubrique
- **En bleu foncé** : les propositions - des exemples qui correspondent aux données de ThéMA
- **En rose** : les commentaires - *des précisions sur les attendus*

2. ORIGINES DES DONNÉES PRODUITES ET/OU REUTILISÉES

a. Données produites par la structure

Nature, type de support, volumétrie prévisionnelle, formats, outils de collecte, auteur(s) de la collecte...

Type d'acquisition : *Données d'enquêtes / Mesures physiques (ex : température, distance, lidar, comptage) / Données de simulation*

Type de données : *textuelles, spatiales, photos, vidéos, sons, tabulaires*

Format :

Tabulaire : CSV

Image : JPG, PNG, PDF, TIFF

Son : MP3, FLAC

Texte : TXT (encodage UTF8), PDF, XML

Vidéo : MPG, AVI, MKV

Spatial : GPKG, GeoTIFF

Ex : les enregistrements audios et les prises de notes effectuées durant les entretiens feront l'objet d'une retranscription en français.

Préciser la part des données brutes et dérivées

Préciser les méthodes et outils utilisés pour acquérir et traiter les données.

Préciser le périmètre, l'échelle, la couverture spatio-temporelle, la période et la fréquence de collecte.

Renvoyer éventuellement au protocole de recherche accessible en ligne.

Privilégier des formats standards, ouverts, non-propriétaires et répandus au sein de votre communauté de recherche pour leur préservation et accessibilité ultérieure.

Extrait du DMP structure ThéMA. (2023)

4. STOCKAGE INTERMEDIAIRE DES DONNÉES

Le stockage intermédiaire concerne le stockage des données pendant le processus de recherche, de travail, contrairement aux données finalisées, sur lesquelles on ne travaille plus (=données « froides »).

Pendant toute la durée du projet les données seront stockées sur un espace sécurisé (au choix) :

- un serveur interne (préciser) qui est accessible par le réseau ou par VPN. Ces dossiers sont automatiquement sauvegardés quotidiennement.

- sur un Nextcloud hébergé par la Maison des Sciences de l'Homme et de l'Environnement, Université de Franche-Comté. / le service Mycore (Owncloud) mis en place par le CNRS.

- Sur une base de données PostgreSQL (pgisdata à la Maison des Sciences de l'Homme et de l'Environnement)

Cette solution permet des sauvegardes régulières et un accès protégé.

Il est recommandé de ne pas déposer les données chez un éditeur privé en « supplementary material », car dans ce cas, les données ne sont pas systématiquement libres d'accès ni correctement documentées, de plus il y a un risque de perdre la souveraineté de ses données.

RECOMMANDATIONS pour le choix d'un autre entrepôt :

- Privilégier un entrepôt disciplinaire (reconnu dans la discipline)

- Privilégier un entrepôt de confiance (qui offre un service de qualité en termes de sécurité et de pérennité des accès), par exemple une solution institutionnelle : Pour le laboratoire ThéMA le service proposé par l'Atelier de la donnée Dat@UBFC accessible depuis leur portail de description des données.

- Choisir un entrepôt qui permet d'attribuer des identifiants pérennes type DOI et des licences pour encadrer la réutilisation des données

...Un DMP
à tester !

Ressources utiles

Atelier de la données dat@UBFC. (2023). *Fiche pratique : Faire un plan de gestion de données (PGD ou DMP) à l'échelle d'une structure.*

Bertram, M. (2022). *Gestion et ouverture des données de la recherche : Pratiques, représentations, besoins. Rapport de l'enquête réalisée en 2022 au Conservatoire national des arts et métiers.* Zenodo. <https://doi.org/10.5281/zenodo.7234643>

INIST CNRS. (2023). Le Plan de Gestion de Données pas à pas. DoRANum. https://doranum.fr/plan-gestion-donnees-dmp/le-plan-de-gestion-de-donnees-pas-a-pas_10_13143_t94g-9j96/