

Cycle de **Vie** des Données

Il comporte globalement **6 étapes**

planification

création / collecte

traitement / analyse

accès / partage

préservation

réutilisation

**A chaque phase du cycle,
des actions de gestion des données
doivent être entreprises.**



Les étapes





<p>Etape 1 1- concevoir et conceptualiser la recherche 2- mettre en place un Plan de Gestion de Données (PGD)</p>	<p>Etape 4 1- publier les articles et lier les data papers 2- appliquer une licence d'utilisation 3- choisir un entrepôt</p> <p> mettre à jour le PGD (cf. étape 1)</p>
<p>Etape 2 1- préciser l'origine des données (collectées et/ou créées) 2- décrire les données et les métadonnées</p> <p> mettre à jour le PGD (cf. étape 1)</p>	<p>Etape 5 Archiver les données à long terme</p> <p> mettre à jour le PGD (cf. étape 1)</p>
<p>Etape 3 1- vérifier, nettoyer et valider les données 2- interpréter les données 3- gérer et stocker les données</p> <p> mettre à jour le PGD (cf. étape 1)</p>	<p>Etape 6 1- permettre de nouvelles recherches 2- réexaminer les résultats / données 3- enseigner et apprendre 4- respecter la licence d'utilisation</p>

Réflexion, organisation en amont du projet de recherche

Plan de gestion de données (PGD)

Planification

Le **plan de gestion des données** est un outil de gestion de projet. Il se présente sous forme d'un document structuré en rubriques. Il a pour objectif de synthétiser la description et l'évolution des jeux de données de votre projet de recherche. Il prépare le partage, la réutilisation et la pérennisation des données.



+ info

[Cliquer ici](#)

DMP OPIDoR est un **outil d'aide à la création en ligne de plans de gestion de données** mis à disposition de l'Enseignement Supérieur et de la Recherche.

Il est hébergé et géré par l'Inist-CNRS. Basé sur le code open source DMPRoadmap, il a été adapté aux besoins de la communauté scientifique française.

+ info

[Cliquer ici](#)

Préciser l'origine des données Décrire les données et les métadonnées

2 Collecte
Création

Dans votre PGD, vous devez **préciser l'origine de vos données**. S'agit-il de données collectées ? D'en créer de nouvelles ? Ou de réutiliser des données existantes ?

Dans tous les cas, vous devez détailler la méthodologie de travail utilisée, et décrire le plus précisément possible vos données.

+ info

[Cliquer ici](#)



A partir de cette étape et pendant tout le cycle de vie des données, respecter les **aspects éthiques et juridiques** mettre **à jour** le plan de gestion de données

+ info

[Cliquer ici](#)

Décrire les données ou les jeux de données (privilégier les systèmes utilisant des standards de **métadonnées**, et à défaut documenter ses descriptions).

+ info

[Cliquer ici](#)

Privilégier les **formats de fichiers** non propriétaires, d'un usage très répandu au sein d'une communauté de recherche pour prévenir l'obsolescence des fichiers.

+ info

[Cliquer ici](#)

Vérifier, nettoyer et valider les données

Interpréter les données

Gérer et stocker les données

3 Traitement Analyse

Vérifier les erreurs dans les données présentes comme des erreurs de frappe, des informations manquantes, des imprécisions etc.

Définir une convention de nommage des fichiers.

[+ info](#)

[Cliquer ici](#)

Sauvegarder les données (3 copies sur 2 supports différents, dont 1 copie à distance).

Sécuriser l'accès aux données confidentielles ou sensibles.

[+ info](#)

[Cliquer ici](#)



Pendant tout le cycle de vie des données, respecter les aspects éthiques et juridiques mettre à jour le plan de gestion de données

[+ info](#)

[Cliquer ici](#)



Publier les articles et lier les data papers

Appliquer une licence d'utilisation

Choisir un entrepôt dans une logique de partage



Data paper : publication qui décrit un jeu de données scientifiques avec les métadonnées. Il doit être accessible, soit sous forme de fichier annexé, soit par un lien pérenne (URL, DOI) vers l'entrepôt de données

+ info

[Cliquer ici](#)

Licence : contrat par lequel le ou les auteur(s) autorise(nt) l'exploitation par un tiers d'utiliser les données et précise(nt) dans quels buts elles peuvent être utilisées.

Favoriser la réutilisation des données : privilégier les licences ouvertes.

+ info

[Cliquer ici](#)

Entrepôt : pour stocker les données de recherche, y accéder et les réutiliser. Il en existe plusieurs types : disciplinaires, multidisciplinaires, propres à un éditeur, institutionnels, spécifiques d'un projet de recherche

+ info

[Cliquer ici](#)



Pendant tout le cycle de vie des données, respecter les aspects éthiques et juridiques mettre à jour le plan de gestion de données

+ info

[Cliquer ici](#)



Archiver à long terme

5 Préservation



A long terme...

L'archivage pérenne a 3 objectifs principaux : conserver le document, le rendre accessible, et en préserver l'intelligibilité.

Ces trois services sont conçus sur le très long terme (plus de **30 ans**).

[+ info](#)

[Cliquer ici](#)

Il **ne faut pas confondre** stockage sécurisé (pour assurer la continuité de l'exploitation durant le projet) et archivage pérenne (conservation à moyen et long terme).

[+ info](#)

[Cliquer ici](#)



Pendant tout le cycle de vie des données, respecter les aspects éthiques et juridiques
mettre à jour le plan de gestion de données

[+ info](#)

[Cliquer ici](#)



Permettre de nouvelles recherches Réexaminer les résultats / données Enseigner et apprendre Respecter la licence d'utilisation



Les **données** collectées, observées, créées ou compilées dans le cadre de projets de recherche sont **précieuses** : elles valident ou invalident des hypothèses de recherche, elles détaillent les résultats de recherche présents dans une publication, elles sont le point de départ de nouvelles recherches, elles constituent un complément à d'autres travaux.

+ info

[Cliquer ici](#)



Trouver des jeux de données susceptibles d'enrichir les connaissances dans un domaine est essentiel à l'avancée des sciences.

Réutiliser des jeux de données est un facteur d'efficacité en recherche.

+ info

[Cliquer ici](#)