

Formation

Plan de Gestion des Données (PGD)

Laurence Dedieu

laurence.dedieu@cirad.fr

Délégation à l'Information Scientifique et Technique (DIST)

Plan de la formation

Accueil, présentation des participants

Le PGD : c'est quoi, pourquoi ?

Données de recherche: définitions

Contenu du PGD:

Décrire des données

Documenter des données

Stocker et sauvegarder des données

Respecter les exigences légales et éthiques

Partager des données

Responsabilités et Ressources

PGD en pratique : modèles, structures, outils, sources d'exemples

Conclusion / Discussion



Vers 10h



Entre 12h30
et 14h

Plan de gestion des données

C'est quoi ; pourquoi ?



PGD : de quoi parle t'on ?



- Livrable de projet
 - 1 seul PGD par projet
 - 1 version au mois 6, + V2 à mi-projet, V3 en fin de projet
- Feuille de route pour décrire les données produites pendant le projet et :
 - leur documentation
 - leur stockage, sauvegarde et sécurisation
 - leur conformité au cadre juridique et éthique
 - leurs modalités de mise à disposition (partage)
 - les coûts de gestion
 - la répartition des responsabilités entre partenaires.

Pourquoi faire un PGD ?

➤ Répond à la demande des financeurs

Constat : Trop peu de données accessibles et réutilisables



*Perte de
10 milliards
d'€ / an*

→ Incitation au **partage et à la réutilisation** des données

➤ Répond à un besoin de transparence et d'intégrité

Constat : Trop de fraudes, de manipulation de données
crise de reproductibilité des études publiées

→ Incitation aux **bonnes pratiques de gestion des données**
(traçabilité, rigueur, qualité, fiabilité, responsabilité, ouverture)

PGD: Attendus des financeurs

■ Bonnes pratiques de gestion des données

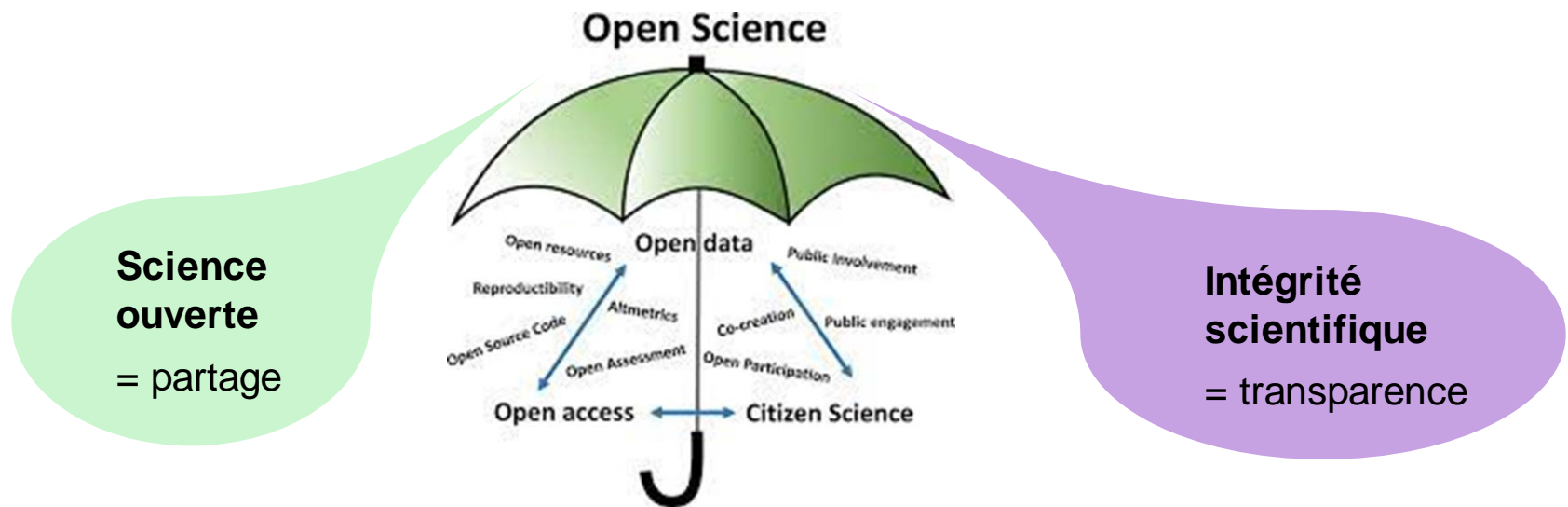
- Le PGD décrit les bonnes pratiques de gestion des données :
 - documentation et description des données
 - suivi des pratiques disciplinaires : standards, normes, formats, ...
 - sécurisation des données
 - prise en compte des contraintes juridiques et éthiques.

■ Partage de données pour le bénéfice de la science et de la société

- Le PGD décrit :
 - la stratégie du projet en terme de partage des jeux de données produits
 - les modalités de partage : quels jeux de données seront accessibles ?, quand ?, où ?, comment ?
 - les raisons justifiant que certaines données ne seront pas partagées.

PGD et Science ouverte

- S'inscrit dans la démarche mondiale de science ouverte
- Répond aux exigences des bailleurs
- Répond à la politique nationale de science ouverte
- à l'engagement du Cirad en faveur de la Science Ouverte



➔ Le PGD répond à ces 2 enjeux

Science ouverte : dynamique mondiale

- Diffuser les résultats et matériaux de recherche, **sans obstacles technique, juridique, géographique ou commercial**, et idéalement **sans aucun délai**.
- **Pour que tous les pays puissent avoir librement accès** aux connaissances, données, codes, méthodes, réactifs, ...pour faire face aux grands défis planétaires
- Enjeux: **reproductibilité et transparence de la recherche, intégrité scientifique et confiance** des citoyens dans la science.



“ Research data that are FAIR and open have the potential to boost impact, quality, efficiency, transparency, and integrity of R&I ”

[Lund Declaration on Maximising the Benefits of Research Data](#). Conseil de l'Union européenne. Juin 2023

SO: dynamique mondiale

- **Appliquée par les financeurs:** Europe (Horizon Europe, Conseil européen de la recherche), ANR, Wellcome Trust, Belmont Forum, B&M Gates, ...
 - Obligation de rendre accessibles ses **publications** immédiatement
 - **PGD:** Version 1 à livrer au mois 6,
V2 à mi-projet, V3 en fin de projet
 - Dépôt des **données** dans 1 entrepôt
le + tôt possible, à la fin du projet au + tard
 - Forte incitation au partage des autres produits de recherche
logiciels, codes sources, modèles, réactifs, ...



« *As open as possible as closed as necessary* »

= Aussi ouvert que possible / aussi fermé que nécessaire



https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf

SO : enjeu national



- 2eme Plan National Science ouverte 2021-2024
Comité Science Ouverte (COSO)
- Loi pour une République numérique (2016)
 - *Les données de recherche sont soumises par défaut aux principes d'ouverture et de libre réutilisation, lorsqu'elles ne sont pas protégées par un droit spécifique (PI, brevets, données personnelles)... »*
- Décret du 3/12/2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics de recherche (qui doivent):
 - *Promouvoir la diffusion des publications en accès ouvert et la mise à disposition des méthodes, protocoles, données et codes sources associés aux résultats de recherche afin d'en garantir traçabilité et reproductibilité*
 - *Veiller à la mise en œuvre de PGD*
 - *Contribuer aux infrastructures qui permettent la conservation et réutilisation des données et codes sources.*

SO : enjeu national



- Mise en application dans les projets ANR
pour encourager les bonnes pratiques de recherche

Le coordinateur s'engage en cas de financement à :

- 1: garantir le libre accès immédiat aux publications scientifiques
- 2: livrer la V1 du PGD au mois 6 du projet (V2 mi-projet, V3 fin de projet)
- 3: faciliter le partage et la réutilisation des données en particulier pour les données liées aux publications:
 - en déposant les données dans des entrepôts de confiance
 - en adoptant une démarche FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) conforme au principe « aussi ouvert que possible, aussi fermé que nécessaire ».
- 4: mettre à disposition, si possible, les logiciels et codes sources.

As open as possible **as closed as necessary**

- **Obligatoire** pour les données à l'origine des articles de recherche
- **Obligatoire** pour certaines données publiques géographiques et environnementales (directive Inspire)
- Ouverture recommandée pour les autres données
 - à détailler dans les versions du PGD
- **Pas d'ouverture si les données ne peuvent pas être partagées**
 - Nuit à l'exploitation ou aux publications prévues dans le projet
 - Contiennent des données confidentielles, non anonymisées,...
 - Sont couvertes par des droits de PI, d'auteurs, secrets, ...
 - dans ce cas: expliquez les raisons dans le PGD

PGD : enjeu pour la publication

- Une politique d'accès aux données est intégrée dans les revues scientifiques pour publier un article de recherche
 - De + en + de revues demandent que vos données soient accessibles et sinon pourquoi elles ne le sont pas.
 - Certaines revues exigent que les données à l'origine des résultats publiés dans l'article soient déposées dans un entrepôt de données.



PGD = gain de temps pour publier parce qu'il vous oblige à bien préparer vos données

PGD : enjeu pour la publication

- Outil d'aide : site internet Où publier <https://ou-publier.cirad.fr/>

Agricultural Systems

Par

Washington State University (Etats-Unis)
Elsevier (Pays-Bas)

[Site web de la revue](#) [Informations aux auteurs](#)

Agricultural Systems is an international journal that deals with interactions - among the components of agricultural systems, among hierarchical levels of agricultural systems, between agricultural and other land use systems, and

- Systèmes de culture
- Zootechnie, syst. d'élevage
- Economie des filières
- Gestion de l'espace foncier
- Environnement, durabilité : multidiscipl.

Libre accès	Auto-archivage	Langues	Notoriété
Libre accès optionnel payant	Revue absente dans Sherpa-Romeo	Anglais	FI - JCR SJR - SCImago Indicateurs

Informations générales	Types d'articles	Frais de publication	Données de recherche	Articles Cirad (Agritrop)
Politique d'accès aux données de recherche		Dépôt recommandé		
Entrepôts de données recommandés par la revue		https://www.elsevier.com/authors/author-resources/research-data/data-base-linking#repositories		

Behavioral Ecology

Par

International Society for Behavioral Ecology (Royaume-Uni)
Oxford University Press (Royaume-Uni)

[Site web de la revue](#) [Informations aux auteurs](#) [Autre lien](#)

The journal publishes studies on all aspects of the field of behavioral ecology in the whole range of behaving organisms, including plants, invertebrates, vertebrates, and humans. Areas include: Empirical and theoretical studies

- Ecologie : multidisciplinaire
- Biologie : multidisciplinaire
- Entomologie

Libre accès	Auto-archivage	Langues	Notoriété
Libre accès optionnel payant	Sherpa Romeo	Anglais	FI - JCR SJR - SCImago Indicateurs

Informations générales	Types d'articles	Frais de publication	Données de recherche
Politique d'accès aux données de recherche		Dépôt obligatoire	
Entrepôts de données recommandés par la revue		https://academic.oup.com/behco/pages/information_for_authors	

- Si vous soumettez à de telles revues: vos données doivent être prêtes à être déposées dans un entrepôt si acceptation du manuscrit

➔ Le PGD vous y aide !

PGD : Atout pour publier

- Outil d'aide : site internet Où publier <https://ou-publier.cirad.fr/>

Où Publier ?

Une sélection de revues et d'éditeurs d'ouvrages en sciences appliquées à l'agriculture



Choisir le support adapté à sa stratégie de publication

2289 Revues 70 Éditeurs d'ouvrages

Rechercher une revue Rechercher un éditeur d'ouvrages

Thématiques

AGRICULTURE, AGRO-ALIMENTAIRE, FORÊTS SCIENCES SOCIALES SCIENCES DE L'ENVIRONNEMENT SCIENCES DU VIVANT SCIENCES ET TECHNIQUES

Conseils pour publier

Participer

- Plus de 2300 revues
- Classées par thèmes, sous-thèmes,...
- + critères de recherche

Rechercher une revue

17 revues identifiées

Réinitialiser la recherche

Titre de la revue

Sélectionner et commencer à taper

Thèmes / Sous-thèmes ET OU

Mots ou expression ET OU

Libre accès ET OU

Langues OU

Notoriété ET OU

Types d'articles OU

- Sélectionner toutes les revues
- Acta Agriculturae Scandinavica Section B-Soil and Plant Science
- Agricultural Systems
- Agroecology and Sustainable Food Systems
- Agronomy for Sustainable Development
- Biological Agriculture and Horticulture
- Centro Agricola
- Communications in Soil Science and Plant Analysis
- International Journal of Phytoremediation
- Journal of Crop Improvement
- Journal of Plant Nutrition
- Journal of Plant Nutrition and Soil Science
- Organic Agriculture
- Philippine Agricultural Scientist
- Precision Agriculture
- Renewable Agriculture and Food Systems
- Systèmes Alimentaires/Food Systems
- Tropical and Subtropical Agroecosystems

Pourquoi faire un PGD ?

➤ Pour le bailleur

- Vous êtes financé par 1 projet, donc soumis aux engagements du contrat qui exige la rédaction du PGD au mois 6 du projet

➤ Pour le projet et sa réussite

- PGD = travail collectif qui permet d'harmoniser les pratiques, clarifier les objectifs, anticiper les problèmes et valoriser les données.

➤ Pour vous car vous publierez plus facilement

- Le PGD vous aide à bien préparer vos données.

➤ Pour faire avancer la science et la société

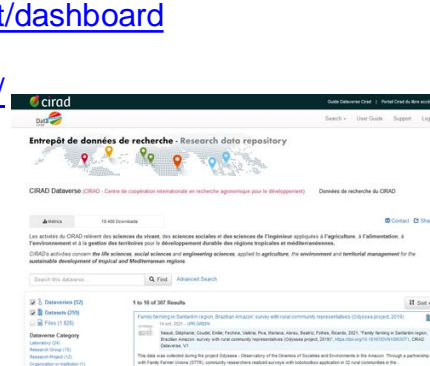
- Le PGD aide à produire des données compréhensibles et réutilisables
- Il répond aux enjeux de reproductibilité, transparence, intégrité scientifique (garantie contre fraude scientifique, confiance des citoyens,...), partage et préservation,....

Peut être simple

- Modèles et exemples disponibles
- Certaines parties déjà rédigées dans votre projet + livrables
- Vous savez rédiger : décrire les données, les méthodes, ...
- Pas obligé de répondre à toutes les questions de la version 1
- Vous disposez d'appui : DIST, DIMS, DJ, DPO, DSI, cellule APA
 - Conseils, aide à la rédaction du PGD (Laurence.dedieu@cirad.fr)
 - Formations PGD (Laurence Dedieu et Pauline Corbière)
- Sites d'information et d'outils
 - Intranet Data <https://intranet-data.cirad.fr/> → modèles PGD format Word et Excel
 - CoopIST <https://coop-ist.cirad.fr/gerer-des-donnees>
 - RGPD <https://intranet-rgpd.cirad.fr/>
 - Gestion de projets <https://collaboratif.cirad.fr/share/page/site/GestionProjet/dashboard>
 - DMP OPIDoR : outil de rédaction en ligne <https://dmp.opidor.fr/>



- Entrepôt de données institutionnel <https://dataverse.cirad.fr/> ;
dataverse@cirad.fr



A retenir sur le contexte du PGD

■ Un document :

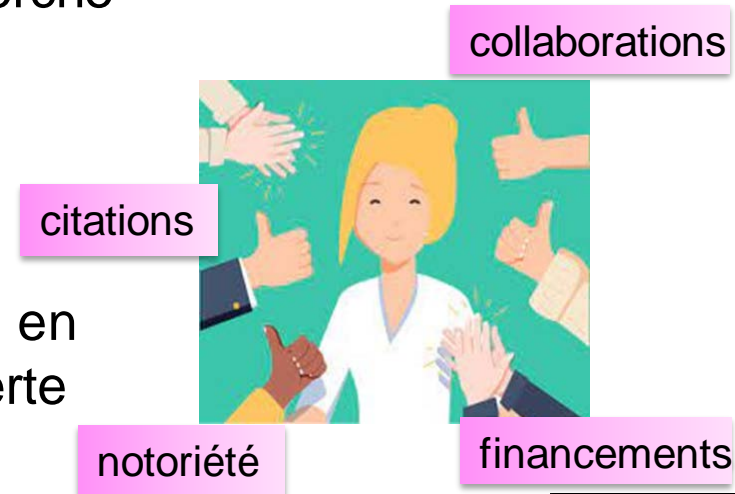
- Décrit les données produites pendant un projet
- Détaille tous les aspects de gestion et valorisation des données
- Montre l'importance accordée aux données du projet

■ Un outil :

- De collaboration
- De rigueur scientifique et de transparence
- Qui améliore vos pratiques de recherche

■ Un atout :

- Pour réussir un projet
- Pour publier des articles
- Pour développer des compétences en gestion de données et science ouverte



PGD et Science ouverte



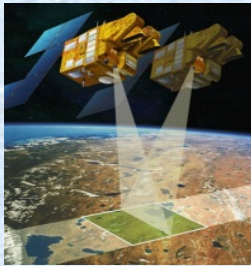
Données de recherche

Quelques définitions

Données

Jeux de données

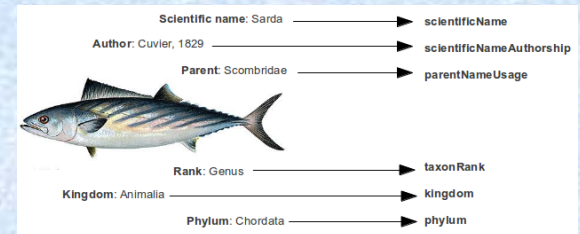
Métadonnées



Téledétection



Filière lait



Darwin Core

Données de recherche



■ Définition de référence

OCDE - Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics. 2007. <https://www.oecd.org/fr/science/inno/38500823.pdf>

« *enregistrements factuels* (chiffres, textes, images et sons), utilisés comme *sources principales pour la recherche* scientifique et généralement reconnus par la communauté scientifique comme *nécessaires pour valider des résultats de recherche* »

Version amendée 2021: <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0347>

« *Autres objets numériques pertinents : modèles, codes, logiciels, algorithmes, flux de travail, issus de la recherche financée en tout ou partie sur fonds publics et utilisés dans le cadre de travaux de recherche et de développement.* ».

Données de recherche

- **Données d'observation**

Capturées en temps réel, souvent uniques et impossibles à reproduire. Ont vocation à être conservées de façon pérenne.

- **Données expérimentales** (labo ou terrain)

Obtenues à partir d'équipements selon 1 méthodologie bien définie. Potentiellement reproductibles, mais à des coûts parfois prohibitifs.

- **Données de simulation** issues de modèles

Généralement facilement reproductibles

- **Données d'enquêtes**

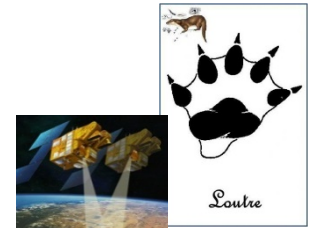
Souvent uniques, difficiles à reproduire
Données quantitatives ou qualitatives

- **Données extraites de documents existants**

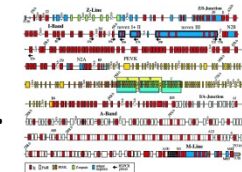
Text and data mining. Dans ce cas, les « données » sont des documents.

- **Codes sources, logiciels, modèles,...**

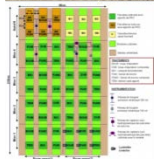
- **Pas des documents de projets, articles, carnets de laboratoire, ...**



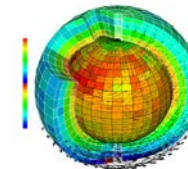
téledétection



Séquence de gènes



Résultats agronomiques



Modèle climatique



enquêtes



Données biblio - Data mining

Métadonnées

- Métadonnées = informations qui accompagnent une donnée et permet de la comprendre et de l'interpréter



Titre	Oiseau roux
Auteur	Emile Lucien
Appareil / modèle	Canon EOS
Format	jpg
Temps d'exposition	1/120 secondes
Focale	F/4.5
Distance focale	46 mm
Date	2020/06/12

Métadonnées

Localisation ??

Type d'habitat ??

Taxonomie ??

....

Métadonnées disciplinaires

- Standards de description des données définies par des communautés scientifiques



Phénotypage de plantes
[MIAPPE](#)

Localisation, période

Biosource

Environnement

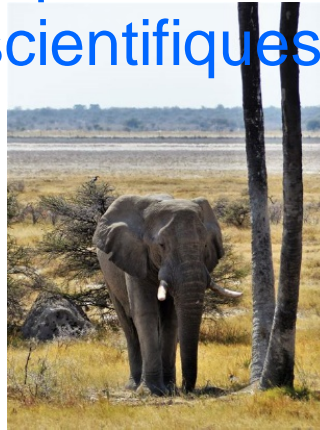
Serre, jardin, champs,
CO2, humidité, lumière, ...

Mode de culture

Traitements

Echantillonnage

Variables observées



Biodiversité
[Darwin Core](#)

Occurrences:

Taxonomie

Classe-Ordre-Famille-Genre-Espèce

Localisation

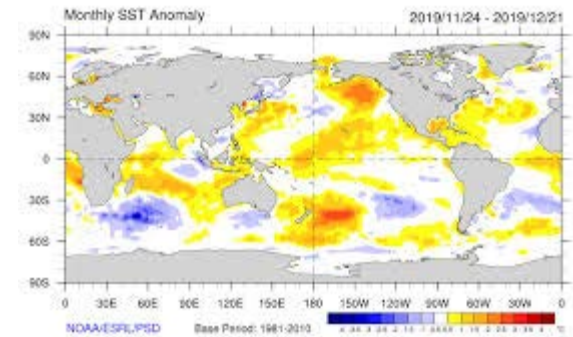
Pays, latitude, longitude, altitude

Période / date

Type d'échantillon

Spécimen vivant, fossilisé, ...

Sexe, Stade biologique



Climatologie
[Climate and Forecast \(CF\) metadata](#)

Couverture spatiale

Couverture temporelle

Station météo (nom, ID)

Pression

T°

Humidité

Précipitations

Rayonnement global

Jeu de données

- Un jeu de données (*dataset*) est défini par un ensemble de :
 - Données produites de la même façon : techniquement homogène
 - Données dont la valeur est liée à la combinaison de jeux de données
 - obtenus par diverses techniques : techniquement hétérogènes
 - mais intellectuellement cohérent.

```
GTCTGGAAACCCAGCCCGTGT  
GTCTGGAAACCCAGCCCGTGT  
GTCTGGATAACCCAGCCCGTGT  
GTCTGGAAACCCAGCCCATGT  
GTAGGAAACCATGCTCATGT  
GTAGGAAACCATGCTCATGT  
GTAGGAAACCATGCTCATGT
```

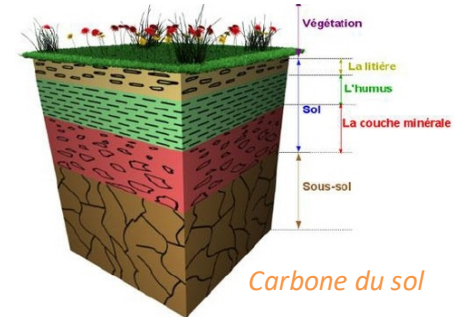
Séquences génétiques



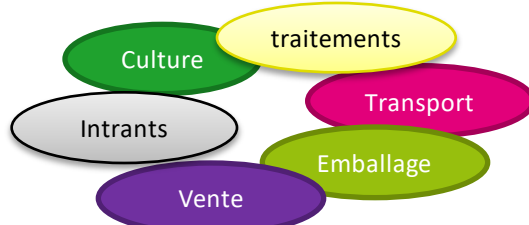
enquêtes



Production agricole



Carbone du sol



Impact environnemental



Filière



Cartographie d'acteurs

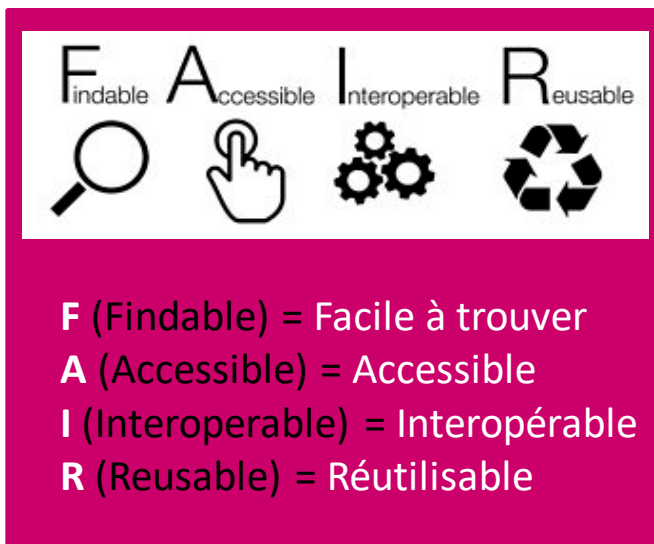


Thème

Recensement agricole (FAO)
Entreprises (BM)

Partage et principes FAIR

- Partager des données (= *Data sharing*)
 - Mettre à disposition des données dans un entrepôt
 - Accès gratuit à la communauté scientifique internationale
- Principes FAIR
 - 4 principes pour garantir la réutilisation des données à la fois par les hommes et par les machines
 - Ils énoncent que les données doivent être



PAUSE CAFE



Plan de Gestion des Données

Contenu



Contenu d'un PGD

Pages d'intro: Présentation du projet, résumé, contexte : Facultatif

1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données
6. Responsabilités et ressources

Schémas, illustrations, workflow

Annexes (liste des abréviations, des variables, ...)

Attendus des financeurs

1. Description des données

- Liste des types de données qui seront produits + méthodes
- Liste des données préexistantes qui seront réutilisées + contraintes

2. Documentation des données

- Liste des documents qui accompagneront les fichiers de données pour faciliter leur compréhension
- Noms des standards disciplinaires utilisés

3. Stockage et sauvegarde

- Supports de stockage, modalités de sauvegarde, de transfert

4. Exigences légales et éthiques

- Procédures de conformité aux législations et codes éthiques
- Clarification des droits des partenaires dans l'accord de consortium

5. Partage des données

- Précision sur quelles données seront partagées, où, quand, comment ?

6. Responsabilités et ressources

- Noms des responsables et budget pour la gestion des données.

Contenu d'un PGD

1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données
6. Responsabilités et ressources

1. Description des données

➤ Cette partie du PGD concerne 3 points:

- 1.1 - Lister les types de données produits dans le projet et autres produits de recherche (codes, logiciels, modèles, ...)
- 1.2 - Décrire les méthodes d'obtention
- 1.3 - Préciser si des données préexistantes sont utilisées



1.1 Liste des types de données

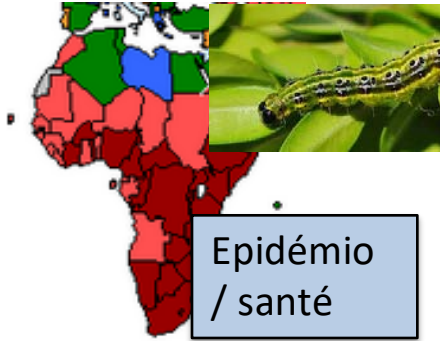


GTCTGGAAACCCAGCCCGTGT
GTCTGGAAACCCAGCCCGTGT
GTCTGGATAACCCAGCCCGTGT
GTCTGGAAACCCAGCCCGTGT
GTAGGAAACCATGCTCATGT
GTAGGAAACCATGCTCATGT
GTAGGAAACCATGCTCATGT

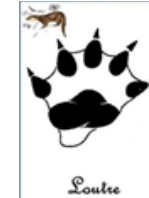
OMICS



Biochimiques /
sensorielles



Epidémiologie
/ santé



Biodiversité /
écologie

Par
discipline



Le PGD doit décrire tous les jeux de données
produits dans un projet



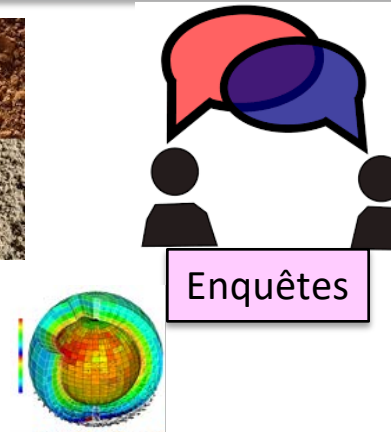
Agro / physio / phéno



Climatiques



Sols / hydro



Enquêtes

Simulations

ou par:
pays/sites
espèce/filière
traitement
work package

1.1 Liste des types de données



➤ 1ère étape du PGD

Lister tous les jeux de données produits dans le projet

en donnant qqs précisions:

objet d'étude, objectif de la collecte de données
localisation géographique, période temporelle

Exemples:

- Données d'enquêtes collectées auprès de femmes en Tunisie centrale en 2020 pour explorer les liens entre agriculture et sécurité alimentaire.
- Données sur le rendement de cacaoyers sélectionnés dans 150 cacaoyères réparties sur 15 zones de cacaoculture en Côte d'Ivoire et représentatives du gradient de complexité pédo-climatique. Collecte de données entre juillet 2020 et décembre 2021.
- Données sur le rendement et la valeur alimentaire pour des ruminants de différentes associations de graminées et légumineuses, cultivées sur le site de Sicalait à La Réunion, entre juin 2021 et mai 2023.

1.1 Description des données



- Précisez les types de données que vous produisez dans votre projet

- Types de données (génomique, enquête, biochimique, biodiversité, épidémiologique, ...) : *(de quoi s'agit-il ?)*
- Objets d'étude et objectifs *(pourquoi ?)*,
- Localisation géographique *(où ?)*,
- Période temporelle *(quand ?)*.

➔ **10 minutes**

1.1 Exemple du projet IMPRINT



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

Les données qui seront produites dans le cadre du projet IMPRINT sont de quatre types : (i) des **données microclimatiques recueillies in-situ** sur les différents sites d'étude envisagés dans le projet ; (ii) des **données d'observations du milieu réalisées in-situ** (p.ex. données de localisation géographique, relevés dendrométriques et botaniques, etc.) ; (iii) des **données issues de prélèvements réalisés in-situ puis analysés en laboratoire** (p.ex. pièges Barber pour inventorier les communautés d'arthropodes du sol, échantillons de sols, etc.) ; et (iv) des **données issues de la télédétection** (imagerie LiDAR).

(i) Les données microclimatiques (p.ex. T°C et humidité relative) seront recueillies à l'aide de capteurs de données environnementales posés directement in-situ par les scientifiques impliqués dans le projet ou bien posés à distance par l'intermédiaire des agents du Réseau National de suivi à long terme des ECOSystèmes FORestiers ([RENECOFOR](#)) de l'Office National des Forêts (ONF), auxquels les capteurs seront envoyés par courrier accompagné d'un protocole d'installation in-situ. En effet le projet s'appuie en partie sur le dispositif expérimental créé et géré par le réseau RENECOFOR pour la collecte et la mise à disposition de données existantes. Les capteurs seront posés in-situ autant que faire se peut dès la première année du projet (2020). Plusieurs types de capteurs seront utilisés : (i) des capteurs HOBO UA-001-64 pour les données de T°C de l'air à 1 m au dessus du sol (une fréquence d'un enregistrement toute les heures est visée) ; (ii) des capteurs HOBO UA-001-08 pour les données de T°C à 8 cm sous la surface du sol (une fréquence d'un enregistrement toute les 2 heures est visée) ; et (iii) des capteurs TMS4 qui permettent de prendre la T°C à 10 cm au dessus du sol, au niveau du sol et à 8 cm sous la surface du sol ainsi que l'humidité relative à 8 cm sous la surface du sol (une fréquence d'un enregistrement toute les 15 mins est visée). Les données microclimatiques seront déchargées au moins une fois par an (2021 et 2022), lors des campagnes de terrain, à l'aide des logiciels dédiés [HOBOWare](#) et [Lolly Manager](#) puis analysées sous le logiciel libre [R](#) de traitements statistiques.

(ii) Les données d'observations du milieu (p.ex. localisation des placettes de mesure in-situ, relevés dendrométriques et botaniques, etc.) seront collectées par chaque scientifique impliqué dans le projet. La "provenance des données" ou localisation géographique (latitude, longitude, altitude) des placettes de mesures sera documentée dès l'installation des placettes de suivie du microclimat, la première année (2020), à l'aide d'un GPS Trimble et d'une antenne pour une meilleure précision. Les données dendrométriques (e.g. diamètre des arbres à 1,30 m) de structure du peuplement forestier seront mesurées in-situ dès l'installation des placettes en 2020, avec prises de photographies hémisphériques pour quantifier l'ouverture du couvert forestier sur chaque placette. La composition en espèces végétales dans les différentes strates de végétation (muscinale, herbacée, arbustive et arborée) sera inventoriée sur des placettes de surface fixe (400 m²) au cours de la deuxième année du projet (2021). La saisie des données d'observation du milieu se fera sur un tableur Excel à l'aide d'une tablette endurcie et étanche pour faire face aux aléas climatiques liés aux conditions du terrain. L'analyse des données se fera à l'aide du logiciel libre [R](#).

<https://dmp.opidor.fr/plans/5082/export.pdf>

1.1 Exemple du projet ATLAS



Données d'enquêtes qualitatives

Il s'agit d'entretiens individuels et de groupe. Les entretiens seront enregistrés (audio, format MP3 ou équivalent) avec le consentement des participants. Pour les entretiens individuels, si l'enquêté-e refuse l'enregistrement audio, l'enquêteur-riche procédera à une prise de notes manuelles.

Chaque entretien sera numéroté en fonction de l'enquête ou de la sous-enquête à laquelle il correspond et du pays de collecte. Le numéro d'entretien sera reporté sur le formulaire de consentement.

Les entretiens seront menés par des anthropologues seniors avec l'appui d'assistant-e-s de recherche formées aux sciences sociales et à la conduite d'entretien. L'équipe qualitative se réunira à intervalles réguliers pour coordonner les activités de terrain et vérifier la qualité des entretiens.

Les enregistrements audio et les prises de note feront l'objet d'une retranscription (fichier texte avec mise en forme) en français (avec traduction le cas échéant pour les entretiens menés dans une autre langue). Lors de la retranscription, les entretiens seront codés (changement des noms et prénoms par exemple) afin de les pseudonymiser.

Données d'enquêtes quantitatives téléphoniques

L'enquête téléphonique (~~WF~~ Enquête Coupons) sera réalisée par la société Ipsos Côte d'Ivoire disposant d'une plateforme CATI (Collecte Assistée par Téléphone et Informatique) à Abidjan. Trois numéros verts seront mis en place (Côte d'Ivoire, Mali, Sénégal). Tous les appels seront routés vers le plateau CATI via des passerelles mises en place dans chaque pays (voix sur IP). Lors des phases de réception d'appels (phase 1 de l'enquête), les appels entrants seront automatiquement orientés vers un enquêteur ivoirien, malien ou sénégalais en fonction de la provenance de l'appel. Avant le démarrage de la collecte, les enquêteurs auront suivis une formation spécifique à l'enquête et à son questionnaire, menée à la fois par les superviseurs d'Ipsos Côte d'Ivoire et l'équipe de recherche.

Le programme PACCI, institution de recherche ivoirienne basée à Abidjan et partenaire du projet ATLAS, développera une plateforme dédiée à la capture des données, la gestion des rendez-vous téléphoniques, la production de tableaux de bord de suivi de l'enquête et l'exportation des données collectées dans un format textuel (CSV).

Des fichiers analytiques dédiés à l'analyse scientifique seront générés à partir de cette plateforme de collecte. Ces fichiers seront générés dans un format ouvert (fichiers texte de type CSV). Ces fichiers analytiques de l'enquête coupons ne contiendront pas de données personnelles.

La contrôle qualité de la collecte des données sera assuré par la mise en place de contrôles de cohérence au moment de la saisie, la double-écoute par les superviseurs d'entretiens conduits par les enquêteurs et des vérifications sur la base analytique en cours d'enquête. L'enquête coupons étant organisée en 3 vagues, une analyse poussée des données de la première vague sera réalisée avant la conduite des vagues 2 et 3 et les procédures opérationnelles seront ajustés en fonction.

Données de l'enquête des temps et mouvements

Cette enquête sera menée par deux économiste juniors supervisés par deux économistes seniors. Les observations de terrains et le remplissage des questionnaires seront effectués par les deux économiste juniors sur support papier. Les données collectées seront par la suite saisies par ces derniers au format Excel.

L'équipe d'économistes se réunira à intervalle régulier pour vérifier le bon déroulement des enquêtes et la qualité des données collectées.

Données financières

La même équipe d'économistes ira collecter des données financières auprès d'un échantillon de structures sanitaires et communautaires engagées dans la distribution de l'autodépistage du VIH dans le cadre du programme ATLAS. Les données collectées seront elles aussi au format Excel.

1.1 Exemple du projet STRADIV



List of the datasets produced by the project

Dataset reference number	Pays	Dataset name	Description	Format	Data storage and access	Dataset leader	Related WP(s)
M1	Madagascar	Stradiv 1 Experimental Design	Data collected on the STRADIV 1 experimental design (96 * 9 * 5.1) diversified rotation and effect on performance of cropping systems in the Vak in Ivory. Agronomic data, plot itk, plant material, test plan, early soil analysis, manure data, weed biomass data, flowering rice LAI, soil faunas (monoliths, pit fall trap), SPAD data (Rice, peanut, sorghum), yield component rice, biomass (rice, groundnut, sorghum, mucuna, nut, C. spectabilis, cowpea) At present data for 2015-2016. Data sets related to WP2 and WP3.	Excel	Database PostgreSQL, Access on request restricted to project partners	Aude Ripoche, CIRAD	WP 2, WP 3
M2	Madagascar	Stradiv 2 Experimental Design	Data collected on the STRADIV 2 device (60 * 9 * 15m) in the Vak in Ivory. + 17 farmers plots agronomic data, Plot itk, plant material, test plan, manure data, weed biomass data, soil faunas (monoliths, pit fall trap), data on white grubs (species, estimated yield (Rice, groundnut, sorghum), yield component in rice, biomass (rice, pen) Weed surveys ° + recoveries For now data 2015-2016.	Excel	Database PostgreSQL, Access on request restricted on project partners	Aude Ripoche, CIRAD	WP 2, WP 3
M3	Madagascar	Monitoring of reference farms MO 2015-2019	20 farms : multi-year structure, , resource map flows, livestock feedings, incomes and labor time for rainfed and lowland crops, local practices and innovations performance	Microsoft Access and Excel, QGIS	Access on request restricted on DP SPAD	Patrice Autfray, Jean-François Bélières, CIRAD	WP4
M4	Madagascar	Investigation diagnosis of farms MO 2014-2015	240 farms : Structure, practices including biodiversity, performance (income)	Microsoft Access and export	Access request restricted on DP SPAD	Jean-François Belleres, CIRAD	WP1, WP4

https://agritrop.cirad.fr/595724/1/CIRAD-PGD-STRADIV-V1.5_20020603.pdf



1.1 Exemple du projet EcoStack



Table 1 provides an overview of the main data type

Data types	Origin	Related WPs	Format
Actor contacts	Actor group databases and contact lists (non-public)	2 & 3	.xls, .csv + .odt
Actor interview data	Primary data	2 & 9	.mp3 + .docx + .txt
Actor workshop data	Primary data	2 & 9	.mp3 + .docx + .txt
Actor quantitative survey data	Primary data	2 & 9	.xls, + .csv
Crop yield data	Primary data	3	.xls, .csv + .odt
Diversity and abundance sample data (insects, plants and microbes)	Primary data	3, 4 & 5	.xls, .csv + .odt
Pest, plant & microbial interaction / response data	Primary data	5 & 6	.xls, .csv + .odt
Genomic data	Primary data	5 & 6	.xls, .csv, .txt, .odt, fasta, fastq + .bam
Eco-toxicological data	Primary data	6 & 7	.xls + .csv
Crop husbandry data	Primary data (expert interviews) and Secondary literature and data sets (public)	8 & 9	.docx, .xls + C++ code
Model scenario input data	Secondary data sets (public)	8 & 9	.txt + .csv
Economic and policy data	Secondary literature and data sets (public)	9 & 10	.docx + .xls

1.1 Description des données



1. Description des données

➤ Cette partie du PGD concerne 3 points:

1.1 - Lister les types de données produits dans le projet et autres produits de recherche (codes, logiciels, modèles, ...)

1.2 - Décrire les méthodes d'obtention

1.3 - Préciser si des données préexistantes sont utilisées



1.2 Exemple du projet TERRI4SOL



Dataset 1 : Base de données des propriétés physiques et chimiques des sols de la zone post-forestière en Côte d'Ivoire

Principales méthodes d'obtention :

- Pipette de Robinson
(Granulométrie (argile (0-2 μm), limon (2-50 μm) sable (50-2000 μm))
- COS (méthode de Walkley & Black et spectrométrie à infrarouge) ;
- Azote total (méthode de Kjeldahl ; CHN) ;
- Phosphore soluble (méthode Olsen) ;
- Bases échangeables (spectrométrie à infrarouge) ;
- CEC (distillation Kjeldahl) ;
- Oligoéléments (méthode AFNOR) ;
- pH (suspension de sol dilué à 1:25)

https://agritrop.cirad.fr/604742/1/DMP_du_projet_TERRI4SOL_produits_recherche_20230522_v1_finale.pdf

1.2 Exemple du projet MISTIC



Metabolomic data:

Methodologies or software: New metabolomic data will be acquired using state-of-the-art targeted and untargeted metabolomics at the Bordeaux metabolome facility ([10.15454/1.5572412770331912E12](https://doi.org/10.15454/1.5572412770331912E12)) following protocols for LCMS-based metabolomics and targeted biochemical phenotyping described previously (Luna et al., 2020, [10.3390/metabo10030096](https://doi.org/10.3390/metabo10030096); Dussarrat et al., 2022, [10.1111/nph.18095](https://doi.org/10.1111/nph.18095)). MS/DIAL will be used for the deconvolution of MS spectra.

Documentation of data provenance: Metadata identifying the samples and generated metabolomics variables will be documented.

DNA sequencing:

Methodologies or software: DNA sequences will be acquired using medium- and high-throughput sequencing technologies providing long and short reads using Illumina NextSeq2000, Illumina MiSeq, and Oxford Nanopore Technologies P2 solo sequencers.

Documentation of data provenance: Depending on the sequencing machine, the [NextSeq 1000/2000 Sequencing System](#) and the [PromethION Data Acquisition Unit A100](#) instrument documentation will be used.

Soil samples:

Methodologies: Soil samples will be acquired from crop cultures and processed following [Genosol protocols](#). Additional processes for sample treatment may be developed in MISTIC. DNA extracted from soil samples will be processed as above in ¶ *DNA Sequencing*.

Documentation of data provenance : Metadata identifying the samples will be documented, including date, host plant, and location. Samples may be acquired from existing crop culture experiments performed by partner projects, such as MetaNema. Horticultural samples are collected from three different agricultural greenhouses (Crop Lambesc, Crop Le Thor, and Crop Pernes-les-Fontaines) with five different types of plants (eggplant, eggplant rootstock, salad, cucumber and peppers).

Leaf samples:

Methodologies: Leaf samples will be acquired from crop cultures. Grapevine leaf samples are collected in Aquitaine as described in Paola Fournier's PhD thesis (PPR-CPA VITAE project), and analyzed using culture-dependent methods as described in Aarti Jaswa's PhD thesis (VITAE/BSA) and metabarcoding approaches (VITAE). Strains will be used to build synthetic microbial communities (SynComs) antagonists of downy mildew (VITAE). Time series data of strain abundance within SynComs will be collected using a quantitative microfluidic chip (VITAE). DNA extracted from leaf samples will be processed as above in ¶ *DNA Sequencing*.

Documentation of data provenance: A brief description of the sampling design and the culture collection can be found here: Fournier et al., 2023, <https://ives-openscience.eu/34866/>

1. Description des données

➤ Cette partie du PGD concerne 3 points:

1.1 - Lister les types de données produits dans le projet
et autres produits de recherche (codes, logiciels, modèles, ...)

1.2 - Décrire les méthodes d'obtention

1.3 - Préciser si des données préexistantes sont utilisées



1.3 Données pre-existantes



- Données préexistantes utilisées dans le projet : issues de
 - **Projets précédents** : données produites par vous ou par vos collègues, produites dans le cadre de contrats de projet,...
 - **Entrepôts** : FAOSTAT, WorldClim, ISRIC, GenBank, Dataverse...



- Êtes-vous sûrs d'avoir le droit de les réutiliser ?
 - Voir les « *terms of use* », licences, contrats précédents,...



<https://www.fao.org/3/ca7570en/ca7570en.pdf>



<https://www.gbif.org/dataset/656fa586-ec8a-4d27-aaa2-a54629c82296>



<https://doi.org/10.18167/DVN1/TLELPO>

1.3 Données pre-existantes



- Avant d'utiliser des données préexistantes : vérifiez que vous avez le droit de les réutiliser
- Dans le PGD il faut:
 - Lister les données préexistantes qui seront réutilisées
 - Préciser les éventuelles contraintes à leur réutilisation
 - Expliquer comment vous levez ces freins à la réutilisation.

➡ Objectif: être sûr d'avoir les droits de réutilisation

- Pensez à faire une recherche de données dans votre domaine

DataciteCommons : <https://commons.datacite.org/>

Google Dataset Search : <https://datasetsearch.research.google.com/>

- Car si des données préexistantes similaires aux vôtres existent et que vous ne les utilisez pas: vous devez expliquer pourquoi !

1.3 - 3 exemples de réponses



- **Aucune donnée externe ou préexistante n'a été utilisée.**
L'ensemble du jeu de données a été collecté pendant le projet.
- Le jeu de données contient des données préexistantes de type :
 - **données climatiques** (Températures, Précipitations) **en libre accès** dans les entrepôts [WorldClim](#) et [CHELSA](#).
 - **données forestières** (parcellaire, type de peuplement, âge du peuplement, gestion, etc) obtenues auprès de **l'Office National des Forêts** grâce à une **convention d'échanges de données**.
- **Les données proviennent de l'entrepôt national [BeQuali](#) :**
Enquêtes qualitatives réalisées en 2000-2003 dans la ville de Mexico dans le cadre du projet ANR «Villes du Sud» (Références, institutions).
Données numérisées et mises à disposition de la communauté scientifique dans l'entrepôt [BeQuali](#).
Données accessibles **sur demande motivée et contrôlée** (Références du contrat de réutilisation garantissant l'obligation de confidentialité).

1.3 Exemple du projet C4F



Tableau 2 : Liste et modalités d'accès des jeux de données externes utilisés par le projet C4F.

Entrepôt	Source	Description du jeu de données	Origine des données	Modalités d'accès
CHELSEA	https://chelsea-climate.org	Free climate data at high resolution	Swiss Federal Institute for Forest, Snow and Landscape Research WSL	Full access allowed through https://chelsea-climate.org
SoilGrids	https://www.isric.org/explore/soilgrids	A system for digital soil mapping based on global compilation of soil profile data and environmental layers	ISRIC World Soil Information	The SoilGrids maps are publicly available under the CC-BY 4.0 License.
WorldClim 2	https://www.worldclim.org	Maps, graphs, tables, and data of the global climate	https://doi.org/10.1002/joc.5086	Free to download. Reuse will respect any licensing restrictions
GenBank	https://www.ncbi.nlm.nih.gov/genbank/	Séquences d'ADN	National Library of Medicine National center for biotechnology information	NCBI places no restrictions on the use or distribution of the GenBank data. However, some submitters may claim patent, copyright, or other intellectual property rights in all or a portion of the data they have submitted.
Pleiades ESA archive	https://earth.esa.int/eogateway/catalog/pleiades-esa-archive	Images satellites	ESA	The collection is available to users located in ESA Member States (including Canada) and European Commission Member States upon approval of the data access request, authorized after successful identity or eligibility check (within 2 days).

1.3 Données pre-existantes



- Utilisez-vous des données préexistantes ?

Si « Oui » : précisez:

- Types de données,
- Sources des données
- Droits de réutilisation:
 - les données sont-elles librement réutilisables ? ou sous conditions ?
 - sont-elles sous licence, droits d'auteurs, ...?

➔ **10 minutes**

1.3 Données pre-existantes



Contenu d'un PGD

1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données
6. Responsabilités et ressources

2. Documentation des données



➤ Cette partie du PGD concerne 4 points:

1. La documentation qui accompagnera vos données
2. Les métadonnées scientifiques décrivant vos données
3. Les fichiers : formats, organisation, conventions de nommage, contrôle de version,...
3. Les mesures de contrôle qualité

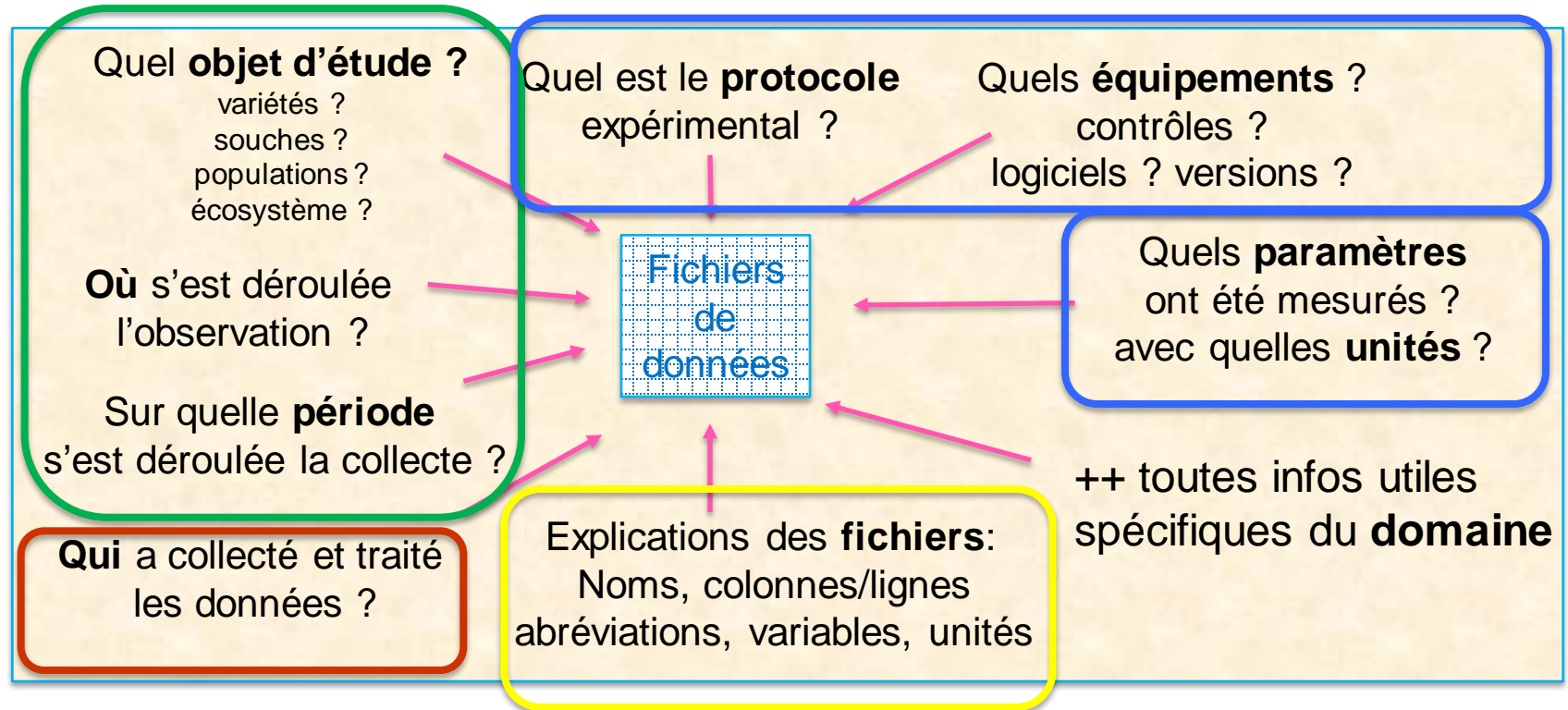


+ documentation complète et + métadonnées enrichies
→ + vos données seront comprises et réutilisables

2.1 Documentation des données



- Quelles informations peuvent-être utiles ?



Toutes les informations sont pertinentes

Car vous ne savez pas quels sont les besoins d'un futur utilisateur .

2.1 Documentation des données



➤ Informations sur les méthodes et matériels

- Protocoles, plan échantillonnage
- Questionnaires, guides d'enquêteurs, technique d'anonymisation
- Equipement utilisé, méthode de calibration, contrôles
- Dictionnaires des variables, unités de mesure
- Explication des colonnes et lignes, abréviations
- Format des fichiers et liens entre fichiers d'un même jeu de données
- Infos rassemblées dans 1 fichier « Lisez moi » (« *Read me* »)

➤ Informations contextuelles

- Description et objectifs du projet, objectifs de la collecte
- Problèmes rencontrés, détails sur des erreurs de données
- Liens vers les articles publiés à partir de ces données,....

➡ **Tout ce qui aide à comprendre les données**

2.1 Documentation des données

- Important d'expliquer les tableaux

	A	B	C	D	E
1		GTT date	GTT weight	time	glucose mg/dl
2	321	2/9/15	24.5	0	99.2
3				5	349.3
4				15	286.1
5				30	312
6				60	99.9
7				120	217.9
8	322	2/9/15	18.9	0	185.8



Signification des colonnes ?? abréviations ??

Date: JJ/MM/AA ? MM/JJ/AA ?

Temps : jours ? heures ? minutes ?

Pourquoi certaines cases sont vides ??

2.1 Documentation des données



- Décrivez la documentation qui sera associée à vos données



➔ 10 minutes

2.1 Exemple du projet IMPRINT



■ Documentation associée aux données

Chaque fichier de données sera accompagné au minimum d'un fichier "Read_me" indiquant :

- Les protocoles et conditions dans lesquelles les données sont collectées,
- la liste des variables
- Le type d'appareillage pour la prise de mesure de chaque variable
- l'unité de la variable mesurée
- toutes autres informations permettant la compréhension et la réutilisation des données.

<https://dmp.opidor.fr/plans/5082/export.pdf>

2.1 Exemple du projet ATLAS



- Documentation associée aux données

Données d'enquêtes qualitatives

Chaque retranscription d'entretien sera accompagnée d'un rapport d'entretien incluant :

- le contexte de passation de l'entretien,
- Les caractéristiques sociodémographiques
- une présentation de la ou des personnes interrogées
- un résumé.

<https://dmp.opidor.fr/plans/3354/export.pdf>

2.1 Documentation des données



- Dans l'entrepôt:

Associer fichiers de données et documentation

Exploitations cotonnières et pratiques culturelles dans quatre villages de la zone CMDT au Mali, campagne 1997-98

1 to 4 of 4 Files Download ▾

<input type="checkbox"/>	 List_variables.tab Tabular Data - 9.7 KB - 24 mars 2021 - 0 Downloads 3 Variables, 115 Observations - UNF:6:3EbheRvxTye6OB3BuVnABQ== Liste des variables avec signification en français et en anglais	
<input type="checkbox"/>	 ML_97_parcelles.xlsx MS Excel Spreadsheet - 155.7 KB - 24 mars 2021 - 2 Downloads MD5: d534a174a61f2cba7533a57589cb3aef Données de suivi des parcelles de culture	
<input type="checkbox"/>	 ML_97_Struc.xlsx MS Excel Spreadsheet - 19.4 KB - 24 mars 2021 - 0 Downloads MD5: c276e26c851d95a45a44ac4c41cc890c Données sur les caractéristiques structurelles des exploitations	
<input type="checkbox"/>	 Readme.txt Plain Text - 550 B - 24 mars 2021 - 0 Downloads MD5: 2bc7f46d64efe883907547d8adb54e23	

Liste des variables

Données

Read me

2. Documentation des données



➤ Cette partie du PGD concerne 4 points:

1. La documentation qui accompagnera vos données
2. Les métadonnées scientifiques décrivant vos données
3. Les fichiers : formats, organisation, conventions de nommage, contrôle de version,...
3. Les mesures de contrôle qualité



+ documentation complète et + métadonnées enrichies
→ + vos données seront comprises et réutilisables

2.2 Métadonnées scientifiques



■ Questions du PGD :

- Quelles métadonnées seront associées aux données ?
- Quels standards seront utilisés ?
- Suivez-vous les pratiques courantes de votre discipline ?



- Qu'est-ce qu'une métadonnée ?
- Connaissez-vous une norme ou un standard de métadonnées disciplinaires ?

2.2 Métadonnées scientifiques



- Ensemble d'informations (= métadonnées) définies par une communauté scientifique pour décrire les données dans cette discipline
- Standards de métadonnées disciplinaires
= Règles de description des données adaptées à une discipline
 - Assure la compréhension et l'interprétation des résultats sans ambiguïté
 - Permet de comparer, intégrer, agréger de multiples expériences
 - Facilite la reproduction des expériences.
- Exemples :
 - Normes d'écriture de dates ou de localisation géographique
 - Standard en écologie, biodiversité, enquête,

Norme pour écrire une date

➤ Quelle est la norme internationale pour écrire une date ?



1. **AAAA-MM-JJ** (année-mois-jour)

2023-04-25

2. **MM-JJ-AAAA** (mois-jour-année)

04-25-2023

3. **JJ-MM-AAAA** (jour-mois-année)

25-04-2023

Norme pour écrire une date



- Norme ISO-8601: <https://www.iso.org/fr/iso-8601-date-and-time-format.html>

AAAA-MM-JJ

- Pour lever l'ambiguïté quand les dates sont exprimées en chiffres
- Permet d'exprimer la date de manière claire et compréhensible par les personnes et les machines.

Départ le 05-01-23

5 janvier 2023
DD-MM-YYYY

Vol Abidjan – New York



Arrivée le 05-01-23

1 mai 2023
MM-DD-YYYY



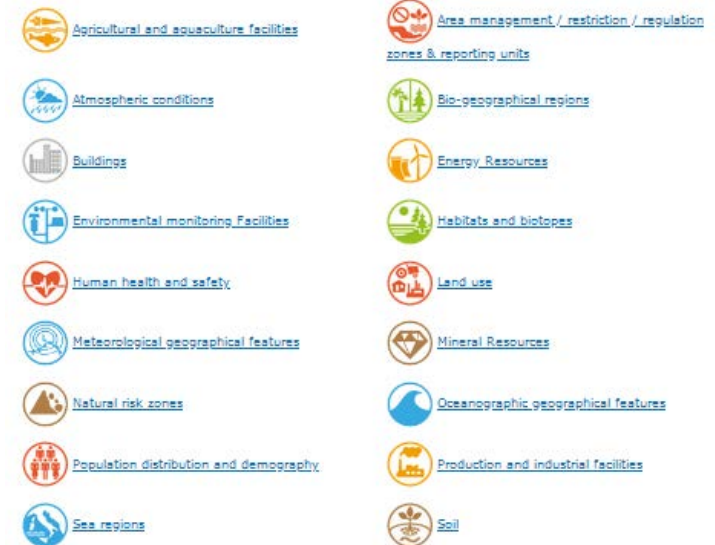
Départ le 2023-01-05



Standards pour données géographiques

- **Norme ISO-19115** : Geospatial metadata standard
= Description normalisée de données géographiques
(Représentation spatiale, Emprise géographique, Résolution,...)

- **Norme européenne INSPIRE**
Thèmes: hydro, agro, sol, habitats...



- **GeoNames**:  GeoNames

Base de données gratuite couvrant tous les pays;
contient 25 millions noms géographiques standardisés.

Métadonnées

- Métadonnées = informations qui accompagnent une donnée et permet de la comprendre et de l'interpréter



Titre	Oiseau roux
Auteur	Emile Lucien
Appareil / modèle	Canon EOS
Format	jpg
Temps d'exposition	1/120 secondes
Focale	F/4.5
Distance focale	46 mm
Date	2020/06/12

Métadonnées

Métadonnées embarquées : automatiquement générées

Métadonnées enrichies : complémentaires, ajoutées a posteriori

Localisation ? Type d'habitat ? Taxonomie ?....

PGD: harmonisation des protocoles

- Le PGD incite les partenaires qui collectent le même type de données à définir ensemble des règles pour faciliter l'intégration, combinaison, comparaison des données.
- Harmonisation:
 - Protocoles
 - T° mesurées par thermomètre électronique à 50 cm du sol et sous ombrage
 - écriture des dates : AAA-MM-JJ
 - abréviations (ex : T° ou temp)
 - noms des variables (ex: "carbon dioxide flux" ou "CO2 flux")
 - unité des variables selon un standard international
 - langue utilisée



Gain de temps !!



En l'absence de PGD



Données de navigation
système métrique
mètres; millimètres ; Kg



Données d'accélération
système anglo-saxon
feet; inches; pounds

**Clarifier toute ambiguïté
entre collaborateurs**



125 Mn \$

Remember the Mars Climate Orbiter incident from 1999?

<https://www.simscale.com/blog/2017/12/nasa-mars-climate-orbiter-metric/>



Standards de métadonnées



- Ecologie, Biodiversité
 - EML : Ecological Metadata Language <https://eml.ecoinformatics.org/>
 - DwC : Darwin Core <http://rs.tdwg.org/dwc/>
 - Corine Biotopes: référentiel européen des habitats naturels et semi-naturels
[2 référentiels d'habitats pour La Réunion](#)
- Météorologie, Hydrologie
 - [SANDRE](#) : Service français des données et référentiels sur l'eau
- Phénotypage de plantes
 - MIAPPE: [Minimum Information About a Plant Phenotyping Experiment](#)
- Sciences Humaines et Sociales
 - DDI : [Data Documentation Initiative](#)
- Genomique
 - MlxS standards = Minimum Information about any (x) Sequence)
<https://www.gensc.org/pages/standards/checklists.html>

DDI: standard pour données d'enquêtes

- **DDI = Data Documentation Initiative** <https://ddialliance.org/>

pour décrire les données et toutes les informations contextuelles

Description des données

- Unité d'observation
- Couverture géographique
- Couverture temporelle
- Univers

Méthodologie et traitement

- Méthode d'échantillonnage
- Mode et Fréquence de collecte
- Opérations de contrôle
- Nettoyage des données
-
- Collecteur des données
- Producteur
- ...



➔ DDI : permet la réexploitation des données d'enquêtes grâce à la précision des données de contexte.

Dataverse et norme DDI

- Jeu de données d'enquêtes dans le Dataverse avec métadonnées DDI

<https://doi.org/10.18167/DVN1/LWT7BG>

The screenshot shows the Dataverse interface for a dataset. The dataset title is "Exploration des liens entre agriculture et sécurité alimentaire : Une enquête auprès des femmes du gouvernorat de Sidi-Bouzyd, en Tunisie centrale". The dataset is associated with the UMR MOISA (Univ Montpellier, CIHEAM-IAMM, CIRAD, INRAE, Institut Agro, IRD) and is categorized under "Marchés, Organisations, Institutions, Stratégies d'Acteurs". The dataset version is 2.0. The description states: "Enquête réalisée en zone rurale du Centre-est de la Tunisie auprès des femmes, des chefs de ménages et des exploitations agricole. (2017-06-19)". The subject is "Social Sciences". The keywords are "Insécurité alimentaire, Food insecurity, Attribution de pouvoirs et de moyens aux femmes, Women's empowerment, Moyens d'existence, Livelihoods, Exploitations agricoles familiales, Family farms". The related publication is "Gaillard Cédric, Martin Sofyan, Bosc Pierre-Marie, El-Ali Jailla, Dop Marie-Claude, Trabelsi Tarek, Azzouli Hedi, (2020), Dury Sandrine, 2018, Explorations des liens entre agriculture et sécurité alimentaire en Tunisie, Cahiers...". The interface includes tabs for "Files", "Metadata", "Terms", and "Versions". A pink box highlights the "Social Science and Humanities Metadata" dropdown menu.

The screenshot shows a list of 8 files associated with the dataset. The files are:

- Diversité production _ alimentation DV.tab (Tabular Data - 34.2 KB - 20 janv. 2022 - 1 Download) - 24 Variables, 290 Observations - UNF:6:JPG0xGQZDXvHJhaTV08/yjg== - Files with diversity dietary and production diversity scores.
- fichier chef de menage anonymisé.tab (Tabular Data - 44.4 KB - 20 janv. 2022 - 1 Download) - 22 Variables, 575 Observations - UNF:6:GX0gg642TAZ37N1owgaw==
- fichier corrigé femmes pour dv anonym.dta (Stata Binary - 144.7 KB - 20 janv. 2022 - 1 Download) - MD5: 54c185fbc9e2c0bc57400aa2b137c4b
- Fichier Exploitations agricoles DV anonymisation.dta (Stata Binary - 316.5 KB - 20 janv. 2022 - 1 Download) - MD5: 1d7624424821e5e1b5921ca6805c29a7
- q2-femme-final.pdf (Adobe PDF - 1.0 MB - 19 jun 2017 - 26 Downloads) - MD5: 5903a99cae45d74f4cbee7cde77bc552 - Questionnaire adressé aux femmes
- q3-chefmenage-final.pdf (Adobe PDF - 753.9 KB - 19 jun 2017 - 25 Downloads) - MD5: caba5459290cb025da41c7e402bae01 - Questionnaire adressé au chef de ménage
- q4-exploit-final.pdf (Adobe PDF - 1.1 MB - 19 jun 2017 - 25 Downloads) - MD5: 48ec0974c6917bd071231bae8c6b255e - Questionnaire adressé aux chefs d'exploitation
- schemas-conceptuel-hbdd.pdf (Adobe PDF - 435.4 KB - 26 jun 2018 - 26 Downloads) - MD5: a18513c24bc99356d575ed902277331 - Le schéma conceptuel simplifié de la base de données

Dataverse et norme DDI



<https://doi.org/10.18167/DVN1/LWT7BG>

CIRAD Dataverse > UMR MOISA >

Exploration des liens entre agriculture et sécurité alimentaire : Une enquête auprès des femmes du gouvernorat de Sidi-Bouzyd, en Tunisie centrale

Social Science and Humanities Metadata ^

Unit of Analysis ?	Femmes Exploitations agricoles (chef d'exploitation) si la femme vit sur une exploitation Ménages (chef de ménage)
Universe ?	Femmes de 20 à 50 ans vivant dans le gouvernorat de Sidi-Bouzyd et son environnement familiale (chef d'exploitation si elle vit au sein d'un ménage agricole et chef de ménage)
Time Method ?	enquête transversale
Data Collector ?	INNTA (Institut National Nutrition et de la Technologie Alimentaire en Tunisie)
Sampling Procedure ?	Echantillon aléatoire en grappe basée sur la population féminine de 20 à 50 ans et sur leur répartition au sein des districts du gouvernorat
Target Sample Size ?	575
Collection Mode ?	Entretien en face-à-face
Type of Research Instrument ?	Entretien structuré
Characteristics of Data Collection Situation ?	Les temps d'administration des différents questionnaires « Femme », « Ménage » et « Exploitation », sont respectivement en moyenne de trente minutes, dix minutes et d'environ quarante minutes. Il a été convenu de commencer la collecte au niveau des terrains dits difficiles, de par les conditions d'accessibilités et climatiques.
Actions to Minimize Losses ?	2 superviseurs ont suivi les enquêteurs durant toute la phase de collecte
Control Operations ?	Double-saisie des données
Weighting ?	Un coefficient de pondération est appliqué à chaque individu de l'échantillon pour extrapoler les résultats à la population du gouvernorat (voir fichier)
Cleaning Operations ?	Recherche des valeurs aberrantes sur les variables quantitatives. Lien entre les différentes questions qualitatives
Response Rate ?	79,8% de réponses par rapport à la population ciblée

2.2 Exemple du projet ATLAS



- Quelle documentation et quelles métadonnées seront associées aux données ?

<https://dmp.opidor.fr/plans/3354/export.pdf>

Données d'enquêtes quantitatives

Les fichiers seront documentés selon la **norme DDI** (*Data Documentation Initiative*: <http://www.ddialliance.org/>) dédiée à la documentation de données d'enquêtes quantitatives en SHS.

Données de l'enquête des temps et mouvements

Les fichiers de l'enquête des temps et mouvements seront également documentés selon la **spécification DDI**.

EML : standard en écologie

EML = Ecological Metadata Language

<https://eml.ecoinformatics.org/>

DataONE

SANParks

South African National Park Data Repository

The Knowledge Network for Biocomplexity

KNB

Class II. Research origin descriptors

Site Description

- Site type
- Geography (location, size)
- Habitat
- Geology, Landform
- Climate

Experimental or sampling design

- Design characteristics
- Variables included
- Species sampled
- Data collection period, frequency

Research methods

- Field/Laboratory
- Instrumentation

Class III. Data set status and accessibility

Status

- Latest update
- Latest archive date
- Metadata status
- Data verification

Accessibility

- Storage location and medium
- Contact person(s)
- Copyright restriction
- Costs

Class IV. Data structural descriptors

Data Set Files

- Data set Identity
- Size
- Format and storage mode



**EML : permet l'interopérabilité des données en écologie
facilite la recherche, interprétation, sélection et exploitation.**

2.2 Exemple du projet IMPRINT



- Quelles métadonnées accompagneront les données ?

<https://dmp.opidor.fr/plans/5082/export.pdf>

Le standard de métadonnées envisagé est le **standard EML** (*Ecological Metadata Language* : <https://eml.ecoinformatics.org/>) avec une implémentation sous le logiciel R grâce au package EML.

Guides de standards de métadonnées

- La pratique de sa communauté est le meilleur guide mais des guides existent

The image shows a screenshot of the Digital Curation Centre (DCC) website. At the top, there is a logo with 'D|C|C' and the text 'Digital Curation Centre'. Below the logo is a search bar labeled 'Search by Discipline'. Underneath the search bar, there are six categories with corresponding images: Biology (microscopic view of cells), Earth Science (coastal landscape), General Research Data (circuit board), Physical Science (molecular model), and Social Science & Humanities (group of people raising hands). Each category is represented by a small image and a text label.

<http://www.dcc.ac.uk/resources/metadata-standards>

The logo for the Research Data Alliance (RDA) is a word cloud. The most prominent word is 'Data', followed by 'Standard' and 'Metadata'. Other visible words include 'Model', 'Information', 'Document', 'File', 'Ivoa', 'Describe', 'Observ', 'Use', 'Astronomical', 'Formal', 'Exchange', 'Initiality', 'Sponsored', 'Support', 'Om', 'Net', 'Develop', 'Such', 'Iso', 'Tiff', 'Virtual', 'Dataset', 'Aeron', 'Image', 'Schema', 'Tagging', 'Interchangeable', and 'Standard'. The word cloud is set against a dark background.

<http://rd-alliance.github.io/metadata-directory/>



<https://www.iso.org/fr/home.html>



<https://fairsharing.org/>



<https://www.elixir-europe.org/communities>

2.2 Exemple du projet PIE News



- Si les standards de métadonnées scientifiques ne sont pas identifiés en début de projet

The PIE News research team is not able at the present moment to specify (the content of) all the data that will be collected.

Given the interdisciplinary nature of the PIE News project,

the research team aims at choosing the standard of metadata that is more suitable for each discipline simultaneously to improve interoperability and make data finding easier.

The research team **expects to define the standard of metadata when more information will be gathered about the data** that are needed, for example after the preliminary analysis of EUS survey data.

<https://cordis.europa.eu/project/id/687922/results>

2.2 Exemples STIRRER - OPTIMA



- Si aucun standard n'existe: expliquez dans le PGD quelles seront les métadonnées utilisées

PGD du projet **STIRRER**: https://dmp.opidor.fr/public_plans

Les métadonnées correspondront :

- au matériel végétal (génotypes utilisés, rang foliaire, conditions de culture et homogénéité des réplicas biologiques),
- protocoles wet-lab (kits, conditions expérimentales, composition des réactifs),
- protocoles d'analyse des données (outils, versions, workflows, options utilisées).

PGD du projet **OPTIMA**: <http://optima-h2020.eu/wp-content/uploads/2019/11/Deliverable-8.2-Data-Management-Plan-Support-Pack.pdf>

Metadata will be : time and date of data creation or data amendments, bounding box, confidence score, owners of actions that took place, lab and field experiments set-up.

2.2 Métadonnées scientifiques



▪ Quelles métadonnées utilisez-vous ?

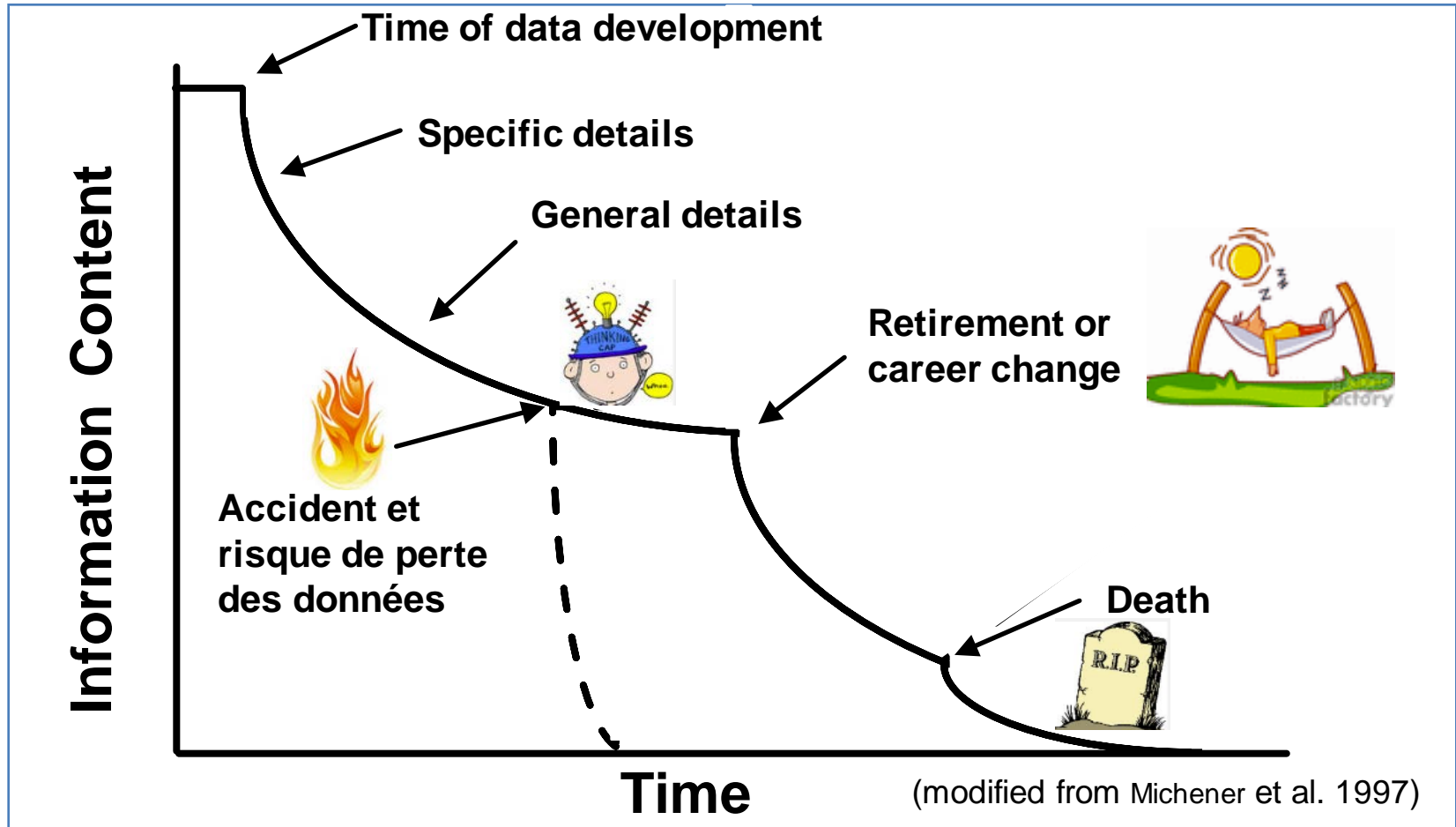
- Suivez-vous un protocole classique dans votre discipline ?
- Utilisez-vous une norme ou un standard de métadonnées ?
- Savez-vous quelles métadonnées sont importantes dans votre discipline ?

➔ **10 minutes**

2.2 Métadonnées scientifiques



- Les noter le + tôt possible: plus on attend, plus on oublie !



A retenir sur les métadonnées



- Le PGD permet d'anticiper ces questions en réfléchissant dès le début du projet à la documentation et description des données
- Nombreux standards de métadonnées et normes
- Pour choisir:
 - Guides de standards de métadonnées
 - Pratiques de votre communauté scientifique
 - Entrepôts de données choisis ou revues visées
- Utiliser un standard disciplinaire facilite l'interprétation de vos données pour un futur utilisateur (et pour les machines)
 - c'est ce qu'attend le bailleur / reviewer
 - Mentionner ce standard suffit pour le PGD
- Renseigner les métadonnées le + tôt possible
- + vous ajoutez de métadonnées, mieux vous documentez vos données et + elles seront trouvables, comprises et réutilisables (principes FAIR)

2.2 Métadonnées scientifiques



2. Documentation des données



➤ Cette partie du PGD concerne 4 points:

1. La documentation qui accompagnera vos données
2. Les métadonnées scientifiques décrivant vos données
3. Les fichiers : formats, organisation, conventions de nommage, contrôle de version,...
3. Les mesures de contrôle qualité



+ documentation complète et + métadonnées enrichies
→ + vos données seront comprises et réutilisables

2.3 Format des fichiers

- Différents format des fichiers

Le format peut être le principal frein au partage des données, s'il ne peut pas être ouvert par la personne qui souhaite y accéder.



Privilégiez les **formats reconnus** dans vos disciplines et les **formats ouverts** (lisibles par tous)

Format non propriétaire :

For spreadsheets: Comma Separated Values (.csv)

For text: plain text (.txt), or PDF/A (.pdf)

For images: TIFF (.tif, .tiff), or PNG (.png)

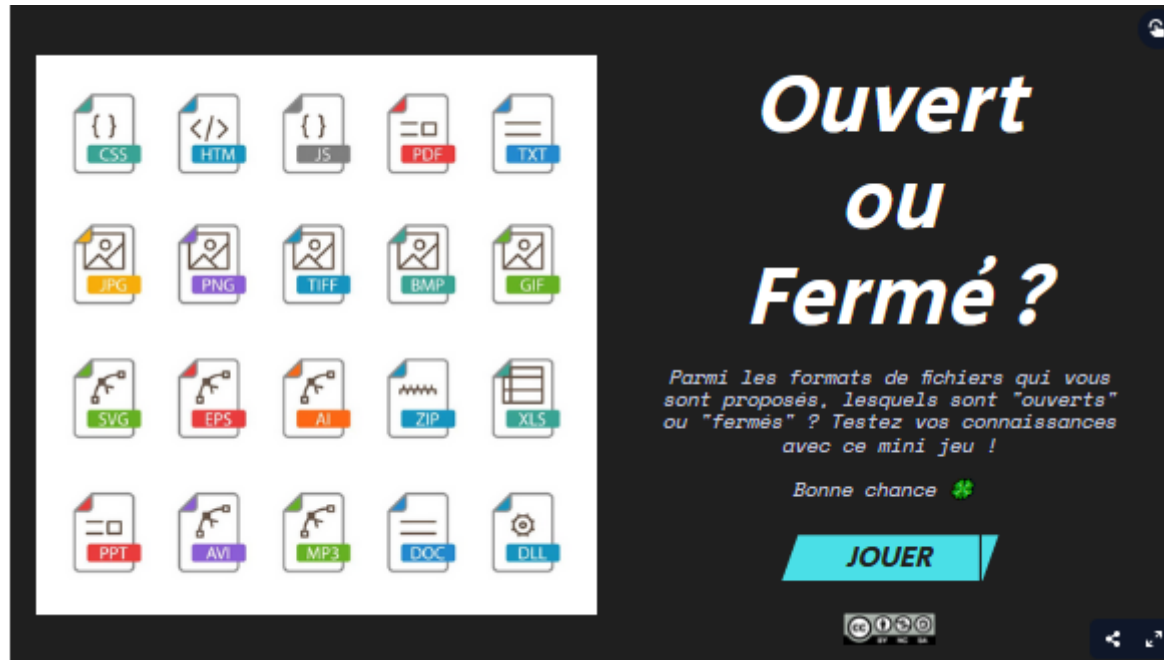
For videos: MPEG-4 (.mp4)



le format de l'open data

2.3 Format des fichiers

- QUIZZ : Format ouvert ou fermé ?



https://doranum.fr/stockage-archivage/quiz-format-ouvert-ou-ferme_10_13143_mcwq-qs64/

Tableau indicatif de formats ouverts et fermés en bas de page.

2.3 Format des fichiers

- Exemple du PGD du projet AfriCultuRes

<https://cordis.europa.eu/project/id/774652/results>

Vector data: Geospatial vector data will be converted into a standard OGC format or well-known format. This list includes, but is not limited to, shape file, KML, GML. Additional formats may be used with approval. Where possible and appropriate, styling information will be provided in SLD format. All will include projection parameters.

Raster data: Geospatial raster data, such as maps or EO derived products, will be converted into a standard OGC or well-known formats. This list includes, but is not limited to, GeoTIFF, JPEG, JPEG2000 or Binary grid. Additional formats may be used upon agreement. Where possible and appropriate, styling information will be provided in SLD format. All will include projection parameters.

Tabular data: Tabular data will be converted into a readily accessible or well-known format. This list includes, but is not limited to, CSV, tab-delimited text file, or Excel spreadsheet.

Documents: Project documentation or outputs open to the public, such as research papers, brochures, presentations, slides, public deliverables, etc. will be converted into a readily accessible or well-known format, like Portable Document Format (PDF) or MS Word (.doc or .docx extensions)

Scripts: Automation of the data integration into models will be achieved through the use of customized scripts that will be developed. By the end of the operational chain task, a library with these scripts will be available. These scripts will be made public in standard scripting languages, like Python or R.

2.3 Format des fichiers

- Exemple de réponse en tableau

Organisation Acronym	category	subcategory	kind of data, unit of measure, eventually broad method, which will be later defined	Format
UNITO	SOIL	BIOLOGICAL PROPERTIES - macrofauna	bateria and fungi population, classic medium-based method, DNA-based method (PCR)	csv and/or jpg
UNITO	PLANT	AGRONOMIC PARAMETERS	growth and yield, disease incidence and severity	csv
UCPH	ENVIRONMENT	CLIMATIC	Local meterological station; data-loggers in field experiment plots	excel
FOEKO	ENVIRONMENT	CLIMATIC	data available by meterological stations near the trial locations	excel
FOEKO	ENVIRONMENT	GEOGRAPHIC	location description	word
FOEKO	PLANT	AGRONOMIC PARAMETERS	Measure vigour of the shoots and harvest in comparison to control. Measure efficacy of the treatment (if scab, control of infestation rate of 50 shoots per treatment)	excel
FOEKO	PLANT	PHENOLOGICAL PARAMETERS	classic phenological phases, BBCH scale	
NIOO-KNAW	SOIL	BIOLOGICAL PROPERTIES - microfauna	AMF root colonization (microscope, staining)	excel
NIOO-KNAW	PLANT	AGRONOMIC PARAMETERS	growth and yield	Excel

2.3 Format des fichiers

■ Exemple de réponse en chek list

Type of data collected

- | | | | |
|-------------------------------------|-------------------------|-------------------------------------|-------------------------|
| <input checked="" type="checkbox"/> | Field measurements | <input type="checkbox"/> | Surveys |
| <input checked="" type="checkbox"/> | Observations | <input checked="" type="checkbox"/> | Models |
| <input checked="" type="checkbox"/> | Laboratory measurements | <input type="checkbox"/> | Personal communications |
| <input type="checkbox"/> | Software | <input type="checkbox"/> | Others |

Format of data collected

- | | | | | | |
|-------------------------------------|--------------|-------------------------------------|---------------|-------------------------------------|--------------|
| <input checked="" type="checkbox"/> | .doc / .docx | <input checked="" type="checkbox"/> | Photo formats | <input checked="" type="checkbox"/> | .txt |
| <input checked="" type="checkbox"/> | .xls / .xlsx | <input type="checkbox"/> | Video formats | <input checked="" type="checkbox"/> | .csv, .tab |
| <input type="checkbox"/> | .ppt / .pptx | <input checked="" type="checkbox"/> | .xml | <input type="checkbox"/> | .sgm / .sgml |
| <input checked="" type="checkbox"/> | .pdf | <input type="checkbox"/> | Others | <input type="checkbox"/> | .nc, .cdf |

Estimated size of data

<10 GB

Software needed to create/process/visualize the data

Microsoft Office, Matlab, SCADA

2.3 Organisation des fichiers

- A la fin du projet, le contenu de votre ordinateur ressemble plutôt à ?



Une bonne organisation des fichiers fait gagner beaucoup de temps dans son projet de recherche.

2.3 Nommage de fichiers

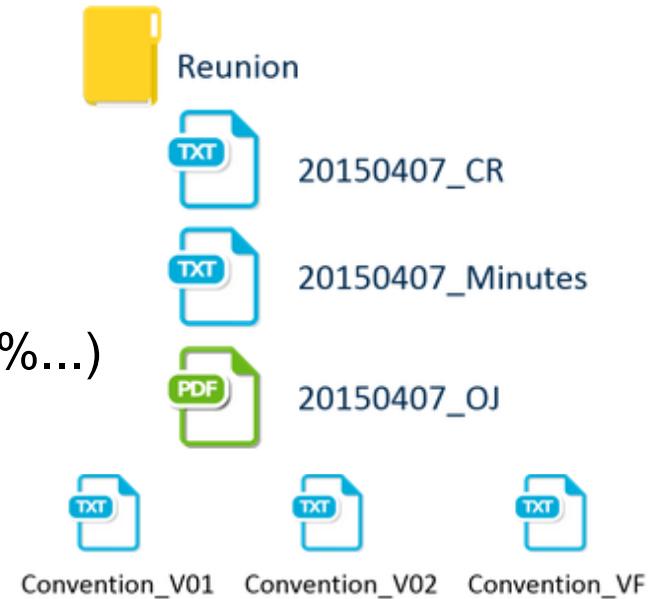
➤ Qqs règles de nommage de fichiers

- Nom bref et explicite selon le contenu
(*CR, rapport d'activité, résultat, bilan,...*)
- Pas d'espace ni caractères spéciaux (! ? # %...)
- Eviter les mots vides : le, la, les, un, une, ...
- Indiquer les versions des documents
- Indiquer la date au bon format : AAAAMMJJ
Mettre l'année en 1^{er} permet de classer les fichiers par ordre chronologique.

[Comment bien nommer ses fichiers ?](#)

➤ Organiser les dossiers de manière logique

- Garder la même structure tout au long du projet
- Préciser les règles (convention de nommage) dans le PGD



2.3 Exemple du projet AfriCultuRes

<https://cordis.europa.eu/project/id/774652/results>

- The naming convention will adhere to the following structure:

<AfriCRS>_<producer>_<wpNo>_<AOI>_<theme>_<date>_<version>_<productionDate><.fileExt>

AfriCRS_GMV_wp42_ET1_NDVI_201801_201803_v11_2018041

correspond to

- an AfriCultuReS data product (AfriCRS)
- produced by GMV (GMV)
- in the context of the WP4.2 (wp42)
- for an area of interest in Ethiopia (ET1)
- The product is the Normalized Difference Vegetation Index (NDVI) in that area
- for the period January to March 2018 (201801_201803)
- GMV produced the NDVI after a minor modification
- on April-15th 2018 (20180415)
- The product format is JPEG2000 (.jp2)

<AfriCRS>	Stands for the shortened acronym of the project
<producer>	Corresponds to the short-name of the AfriCultuReS partner in charge of the generation of the product or dataset. Short names as given in the Grant Agreement. In case of nested products, resulting out from the fusion of various producers' inputs, the acronym will identify the entity responsible for the generation of the final product
<wpNo>	Work package identifier, set to determine in which phase of the project the data has been produced
<aoi>	Geographic location of the area of interest, according to the ISO 3166-1 alpha-2 two-letter country codes defined in ISO 3166-1, part of the ISO 3166 standard to represent countries, dependent territories, and special areas of geographical interest. To ensure unique identification of the AOI, the ISO 3166-1 alpha-2 codes will be followed by a correlative number set to identify the AOIs in case of multiple areas within a single country, e.g.: ET1, ET2... denote AOI1 and AOI2 in Ethiopia
<theme>	Acronym or short name of the parameter, theme or phenomena addressed by the product. Acronyms will be maintained in a dedicated dictionary (set for the project) to ensure discoverability. As far as possible, the definition of specific acronyms and abbreviations for the AfriCultuReS data will be avoided, instead the project will adopt as much as possible standard acronyms and abbreviations from dictionaries and thesauri (like AGROVOC or the GEOSS Earth Observation Dictionary, refer to Search keywords for details on standard keywords).
<date>	Date and/or time, including time-range, of the product, as defined by date-time of the input(s) involved in its generation. Date/and or time will be in compliance with ISO 8601, Data elements and interchange formats - Information interchange - Representation of dates and times, that is an international standard covering the exchange of date- and time-related data. Double entry indicates time interval, e.g. 201801_201803 (January-February-March 2018)
<version>	Two digits version correlative numbering, the first digit corresponds to a major version, the second one correspond to minor versions
<productionDate>	Date/and or time will be designated following the ISO 8601 encoding
<.fileExt>	Optional, file extension

2. Documentation des données



➤ Cette partie du PGD concerne 4 points:

1. La documentation qui accompagnera vos données
2. Les métadonnées scientifiques décrivant vos données
3. Les fichiers : formats, organisation, conventions de nommage, contrôle de version,...
4. Les mesures de contrôle qualité



+ documentation complète et + métadonnées enrichies
→ + vos données seront comprises et réutilisables

2.4 Contrôle qualité

➤ Procédures de contrôle qualité :

- Plan d'assurance qualité (ex: ISO 9001)
- Protocoles de métrologie pour les appareils et équipements de mesures
- Calibration des équipements, témoins de contrôle
- Duplication des échantillons et expériences
- Répétition des mesures et des analyses;
- Capture automatisée des données
- Vérification systématique de la saisie des données par échantillonnage aléatoire de 2 % du jeu des observations.
- Validation de saisie des données (ex: transcrits) par une autre personne
- Contrôle des valeurs manquantes ou aberrantes
- Contrôle de la cohérence des formats de date, de localisation, de taxonomie, des unités de mesure, ...
- Utilisation de vocabulaires contrôlés

2.4 Exemple du projet STIRRER

https://dmp.opidor.fr/public_plans

➤ Quelles mesures de contrôle qualité ?

NGS: **3 réplicats** biologiques correspondant à des feuilles de boutures du **même génotype**. L'ensemble des génotypes ont été **cultivés au même moment** dans la **même chambre** de culture en **milieu contrôlé**. Tous les échantillons ont été **prélevé le même jour** au **même stade de développement**, sur le **même rang foliaire**.

Avant séquençage, tous les échantillons suivent un contrôle qualité.

Durant le séquençage, **contrôles qualités effectués par la plateforme de séquençage**.

A réception des données, **contrôles qualités bio-informatique** effectués par nos soins: qualité de l'appel des bases, qualité de mapping (quantité de reads mappés, fraction de reads mappant de manière unique), vérification de l'origine des banques (homogénéité des réplicats via ACP après mapping).

2.3 Format des fichiers



Pause déjeuner



Contenu d'un PGD

1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données
6. Responsabilités et ressources

3. Stockage et sauvegarde



➤ Points à traiter

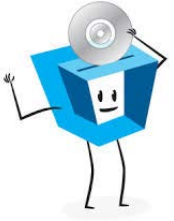
1. Stockage, sauvegarde, sécurisation

- Nommer les supports de stockage de données utilisés
Ordinateurs, serveurs, disque dur externe, entrepôts, cloud, etc
Outils : tablette, smartphone, capteur (sonde, automate ...)
- Décrire les procédures de sauvegarde : fréquence, supports
- Décrire les procédures pour sécuriser les données, surtout dans cas de données contenant des données à caractère personnel
Accès physique aux locaux et équipements protégé par badge et gardiennage, bureau fermé à clé,
Ordinateur protégé par mot de passe et antivirus, fichiers cryptés,...

2. Partage de données entre partenaires

- Décrire les modalités

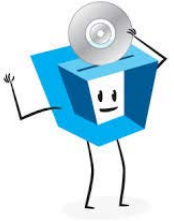
3.1 Stockage et sauvegarde



- Qui a déjà perdu des données ?
- Que signifie la règle d'or pour sauvegarder ses données ?



3.1 Stockage et sauvegarde

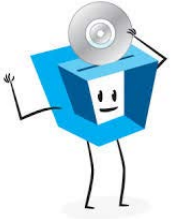


■ Règle du 3 - 2 - 1



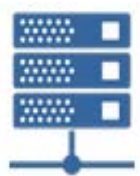
- Faire au moins **3 copies** de ses données
 - Stocker les copies sur au moins **2 supports** différents
 - Avoir au moins **1 copie de sauvegarde hors site**
- **Zéro inquiétude** pour la sauvegarde des données !!

3.1 Stockage et sauvegarde



➤ Bonnes pratiques en matière de stockage

- Ne pas stocker uniquement sur son portable !
- Conserver des copies multiples (règle 3-2-1)
- Diversifier les supports de stockage
- Utiliser des sauvegardes automatiques



Serveur
institutionnel



SUR VOTRE ORDINATEUR



Service cloud



Renater



Capteur



Support de
stockage externe



Tablette

Comparatif des systèmes

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 <p>Ordinateur professionnel</p>	<p>★ ★ ☆ ☆</p> <p>Sujet au piratage informatique, aux détériorations et pannes</p>	<p>★ ☆ ☆ ☆</p> <p>Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...)</p>	<p>★ ★ ★ ★</p> <p>Pas de coût supplémentaire ou coût peu important</p>	<p>- Pour un stockage temporaire</p> <p>- Nécessité de crypter les données confidentielles et sensibles</p>
 <p>Support externe</p>	<p>★ ☆ ☆ ☆</p> <p>- Sujet au vol, à la perte du support - Durée de vie limitée (dégradation du matériel)</p>	<p>★ ★ ★ ☆</p> <p>Facilement transportable, il permet de transférer les données vers un autre ordinateur</p>	<p>★ ★ ★ ★</p> <p>Pas de coût supplémentaire ou coût peu important</p>	<p>- Pour un stockage temporaire</p> <p>- Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles</p>
 <p>Serveur institutionnel</p>	<p>★ ★ ★ ★</p> <p>Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)</p>	<p>★ ★ ☆ ☆</p> <p>La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures</p>	<p>★ ★ ☆ ☆</p> <p>Coût assez important mais pas forcément répercuté sur l'utilisateur</p>	<p>- Pour un stockage plus pérenne</p> <p>- Adapté pour le stockage de données sensibles et des versions « stables » de vos données</p> <p>- Toutes les institutions ne proposent pas ce service</p>
 <p>Serveur Cloud</p>	<p>★ ★ ☆ ☆</p> <p>On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent</p>	<p>★ ★ ★ ★</p> <p>Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage</p>	<p>★ ★ ☆ ☆</p> <p>Payant à partir d'une certaine limite de stockage</p>	<p>- Pour un partage avec des personnes externes à l'institution</p> <p>- Ne pas y mettre de données sensibles ou confidentielles</p> <p>- Pas de contrôle sur la procédure de sauvegarde des données</p>

[Stocker ses données de façon sécurisée – DoRANum](#)

Stockage, partage, archivage

Stockage, partage et archivage : quelles différences ?

STOCKAGE ET SAUVEGARDE



Serveur sécurisé de l'institution

Espace collaboratif de travail



Le chercheur



L'équipe de recherche

PARTAGE



Entrepôt

Entrepôt généraliste ou disciplinaire



L'équipe de recherche



Autre équipe de recherche (même domaine)

Autre équipe de recherche (autre domaine)

ARCHIVAGE PÉRENNE



Plateforme d'archivage (ex : CINES)



L'équipe de recherche



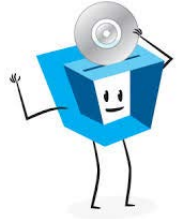
Autre équipe de recherche (même domaine)

Autre équipe de recherche (autre domaine)

Interne projet

Externe projet

Services Cirad et Renater



Dataverse <https://dataverse.cirad.fr/>

- Données et documents associés
- Nouvelle version fin 2023 : plus de limite de taille de fichier
- Conservation à moyen terme : 10 ans minimum



Alfresco-Share

- Principalement pour des documents ; 10 ans minimum



Forge logicielle <https://intranet-data.cirad.fr/outils-et-ressources/forge-logicielle/presentation>

- Ouverte depuis juin 2023



Plateforme Meso@LR <https://meso-lr.umontpellier.fr/> *MESO@LR*

- Stockage, calcul, cloud
- Accessible à toutes les UR

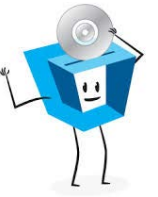
Renater: Réseau numérique de la communauté enseignement recherche.

- Offres : FileSender, Evento, Rendez-Vous
- DRIVE : cloud sécurisé <https://drive.demo.renater.fr/>



Cybersécurité: <https://cybersecurite.cirad.fr/fr/boite-a-outils>

3.1 Stockage et sauvegarde



- Décrivez vos procédures de stockage et de sauvegarde

- Où sont stockées vos données
- Combien de sauvegarde faites-vous par mois ?
- Sur quels supports ?

➔ **10 minutes**

3.1 Exemple du projet IMPRINT

- Procédure de stockage et de sauvegarde

Sur toute la durée du projet, les données et métadonnées générées seront hébergées sur les **postes de travail chiffrés** des membres du projet IMPRINT.

Pour les données les plus volumineuses (données LiDAR), il est prévu l'achat d'un **disque dur externe**, qui sera également **chiffré**.

Pour les données microclimatiques et les données des inventaires dendrométriques, floristiques et faunistiques et pour toutes les infos relatives au projet, il est prévu d'utiliser un **dossier partagé "IMPRINT"** sur la **plateforme de sauvegarde et partage sécurisé du CNRS** : myCoRe.

Des **sauvegardes régulières** (tous les 6 mois) de l'ensemble des données (inclues les données LiDAR) seront effectuées **sur le serveur NAS** de l'UR EDYSAN.

A noter également que les données LiDAR seront partagées avec chaque agence territoriale de l'ONF concernée par le projet (FD de l'Aigoual, FD de Blois et FD de Mormal). Cela assurera une **sauvegarde supplémentaire**.

3.1 Exemple du projet ZERO BRINE

Décliné pour les 10 WPs

Table 4-4: Expected data storage and back up in WP3

Data	Storage medium and location	Backup location	Backup frequency
Raw data	Researcher's computer, task leader's computer, portable USB drives, laboratory notebooks	<ol style="list-style-type: none"> 1. Task leader's office at SUT 2. Task leader's home outside the SUT 	Backup drives are connected for a few hours every day and incremental backup is performed every 1 h; hourly snapshots are kept for a week, daily snapshots are kept for a month, weekly snapshots are kept indefinitely. Previous versions of files can be recovered when required
Processed data			
Models/code			
Informed consents	If applicable, paperwork is kept at the SUT's archives. Scanned version is kept with the other data and subjected to the same treatment		Paperwork is kept indefinitely in the SUT's archives, electronic version - see raw/processed data and models/code

3.2 Partage entre partenaires

➤ Points à traiter

- Avec qui les données sont-elles échangées ? Qui aura accès ?
 - Tous les partenaires du projet ? Uniquement certains pour certains WP ?
 - Quelles procédures de sécurisation si données non anonymisées ?
- Comment ?
 - Plateforme dédiée et sécurisée : Alfresco (pour les documents)
 - Outil FileSender du réseau RENATER (communauté Education-Recherche)
 - Entrepôt Dataverse Cirad du projet (pour les données)
 - E-mails (pour fichiers n'ayant pas besoin d'être sécurisés et de petite taille)
- Selon quelles modalités ?
 - Accord de consortium ? Contrat de collaboration entre partenaires ?
 - Modalités définies dans le PGD ?

3.2 Exemple du projet IMPRINT

- Procédures de partage entre partenaires

L'ensemble des scientifiques impliqués dans le projet aura accès aux données et métadonnées les moins volumineuses (données microclimatiques et données issues des inventaires dendrométriques, floristiques et faunistiques) via la **plateforme de sauvegarde et de partage sécurisée du CNRS** (myCoRe).

Pour les données plus volumineuses (données LiDAR), stockées sur disque dur externe chiffré, l'accès se fera **au cas par cas** et suivant les besoins.

L'ensemble des agences territoriales de l'ONF concernées par le projet (Aigoual, Blois et Mormal) aura également accès aux données et métadonnées générées par le projet.

3. Stockage et sauvegarde



Contenu d'un PGD

1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données
6. Responsabilités et ressources

4. Exigences légales et éthiques



- Objectif : Lever tous les freins pour permettre la réutilisation des données du projet
 - Données **à caractère personnel**
ex: données d'enquête, socio-économiques, santé, ...
 - Données issues de **ressources génétiques**
ou de savoirs traditionnels associés
 - Données qui posent des questions **éthiques**
ex: expérimentation animale ou impactant l'homme ou l'environnement, ...
 - Données avec des droits de **propriété intellectuelle**
ex: données obtenues lors collaboration avec un partenaire privé,
données avec droits d'auteurs (ex: photos)
données obtenues à partir de données fournies par un partenaire,...

4. Exigences légales et éthiques



génétiques



enquêtes



santé



climatiques



biodiversité



Partenariat privé

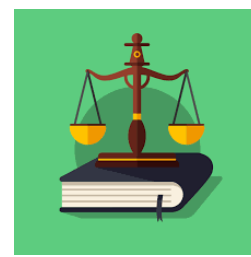
- Montrer que vous respectez le cadre éthique et juridique
 - Données à caractère personnel : [RGPD](#) + Lois des pays où sont réalisées les enquêtes
 - Données issues de ressources génétiques : [Nagoya](#) / [APA](#)
 - Données issues d'études ayant un impact sur l'animal, l'homme ou l'environnement : [Comité éthique](#)
 - Données avec droits de propriété intellectuelle : [AC](#), [contrat](#)
- ➡ Freins levés pour permettre à d'autres de réutiliser les données du projet

4.1 Données personnelles

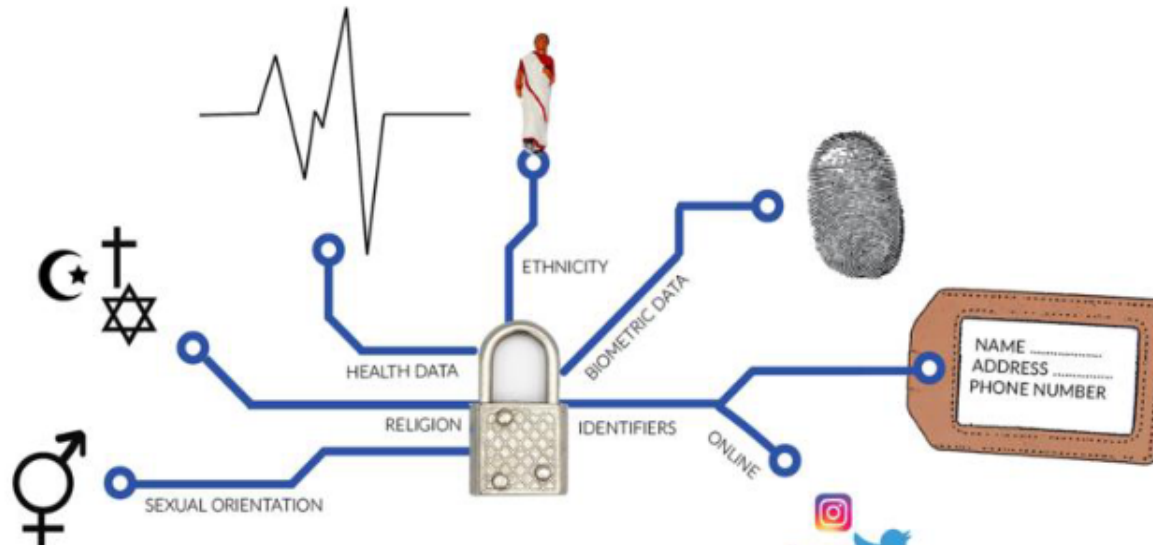


- Donnée à caractère personnel = toute information permettant d'identifier directement ou indirectement une personne physique.
- Dans la liste suivante: quelles informations correspondent à une donnée personnelle donc identifiante ?
 1. une adresse ou un point GPS
 2. les types de cultures dans une exploitation
 3. la ville de résidence
 4. une donnée sur la qualité du sol
 5. la taille du ménage
 6. la surface d'une cacaoyère
 7. la voix d'une personne
 8. la démarche d'une personne
 9. la photo floutée d'une personne

4.1 Données personnelles

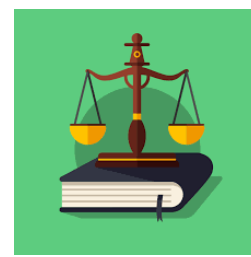


- Cas particulier des données « sensibles »



- Par principe la collecte et le traitement de ces données sont interdits
- Quatre exceptions à cette interdiction prévues pour la recherche
 - consentement libre, éclairé et explicite
 - données sensibles rendues publiques par la personne concernée
 - recherche nécessaire pour motifs d'intérêt public importants
 - données nécessaires à la recherche publique

4.1 Données personnelles



➤ Cadre législatif à prendre en compte

- Règlement Général sur la Protection des Données (RGPD)
<https://www.cnil.fr/fr/reglement-europeen-protection-donnees>
- Loi du pays où sont réalisées les enquêtes



Trouver les lois: <https://www.cnil.fr/fr/la-protection-des-donnees-dans-le-monde>

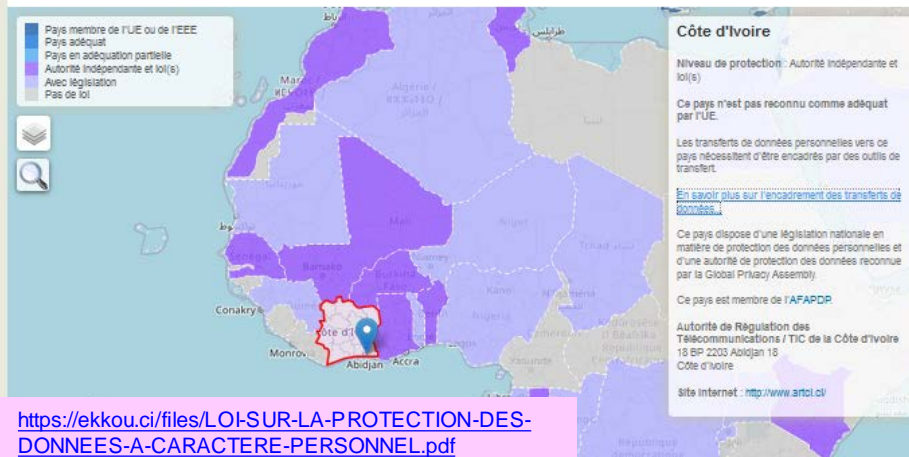
La protection des données dans le monde

Dans quel pays transférer des données personnelles et à quelles conditions ?

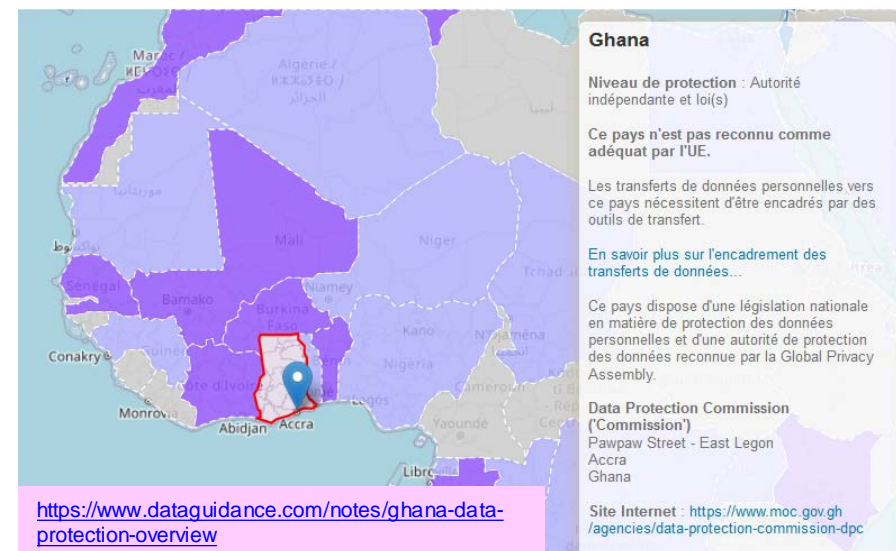
Quel pays dispose d'une législation spécifique ou d'une autorité de protection des données personnelles ?

Cette carte vous permet de visualiser les différents niveaux de protection des données des pays dans le monde. Vous pouvez afficher les autorités de protection des données à l'aide de l'icône « calque » située en haut à gauche de la carte.

L'icône « loupe » vous permet de rechercher un pays et de le positionner sur la carte.

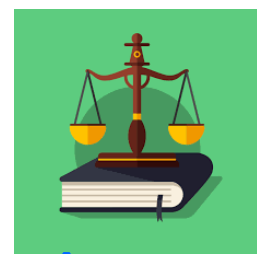


<https://ekkou.ci/files/LOI-SUR-LA-PROTECTION-DES-DONNEES-A-CARACTERE-PERSONNEL.pdf>



<https://www.dataguidance.com/notes/ghana-data-protection-overview>

4.1 Données personnelles



➤ Contraintes à respecter : RGPD + loi du pays d'enquête

- **Obtenir le consentement libre et éclairé** des participants
- **Minimiser** la collecte de données
- Collecter les données pour une **finalité précise**
- Assurer la **confidentialité et la sécurité** des données
- **Limiter la durée de conservation** des données personnelles
- **Anonymiser** les données le + tôt possible

➤ Dans certains pays, en plus, il faut obtenir l'accord de :

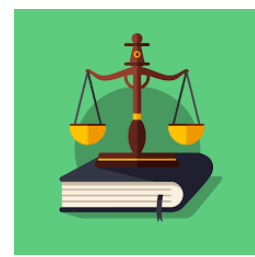
- Autorité de protection en Côte d'Ivoire: <https://www.autoritedeprotection.ci/>
- Data Protection Commission au Ghana: <https://www.dataprotection.org.gh/>

➤ Appui au Cirad

- Site intranet RGPD : <https://intranet-rgpd.cirad.fr/>
- DPO : Data Protection Officer : dpo@cirad.fr

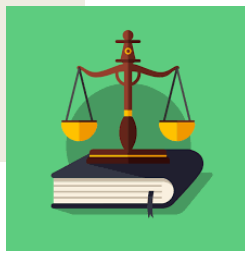


4.1 Données personnelles



- Le formulaire de consentement doit préciser :
 - les objectifs de la recherche d'une manière compréhensible, transparente et précise
 - les informations qui seront demandées aux enquêtés
 - les modalités de diffusion envisagées pendant et après le projet (ex: articles, rapports, colloques,...)
 - les mesures de protection et la durée de conservation des données perso
 - informer les participants de leurs droits: être informé, avoir accès aux données, droits de rectification, de retrait à tout moment,...
 - que toute information identifiante sera supprimée avant que les fichiers ne soient partagés ou rendus publics.

4.1 Exemple du projet ATLAS



Le protocole de recherche a été évalué et autorisé par 4 comités d'éthique :

- OMS : Ethical Research Committee, date :
- Côte d'Ivoire : Comité National d'Éthique des Sciences de la vie et de la santé, date : 28 mai 2019
- Mali : Comité d'Éthique de la Faculté de Médecine et de Pharmacie de l'université de Bamako , date :
- Sénégal : Comité National d'Éthique pour la Recherche en Santé, date : 26 juillet 2019

L'ensemble des formulaires de consentement et des notices d'informations sont disponibles en annexe du protocole, sur le site <https://atlas.solthis.org/>

Les formulaires de consentement seront conservés pendant 5 ans après la fin du projet (soit jusque fin 2026) dans une armoire sous clé dans les locaux de Solthis situés à Abidjan pour les formulaires de consentement collectés en Côte d'Ivoire, Bamako pour les formulaires collectés au Mali et Dakar pour les formulaires collectés au Sénégal.

Les enregistrements audio des entretiens seront conservés le temps de leur transcription et de leur pseudonymisation. Chaque chercheur-e s'engage à ce que les enregistrements audios soient détruits au plus tard à la fin du projet (fin 2021).

Comme mentionné sur les notices d'information, les données pseudonymisées seront conservées pour une durée maximum de 5 ans après la fin du projet, soit jusque fin 2026 au plus tard. Seules des données anonymisées feront l'objet d'un archivage de longue durée (voir le chapitre "Sélection et préservation").

Concernant les données qualitatives, considérant que le processus d'anonymisation décrit précédemment pourrait être source d'erreur et que l'identification de certaines des personnes enquêtées pourrait toujours être théoriquement possible par croisement d'informations, l'accès aux données archivées sera restreint. L'accès, via la plateforme Zenodo, se fera sur demande, par des équipes académiques et sera soumis à la signature d'un engagement de confidentialité.

4.1 Exemple du projet MOOD



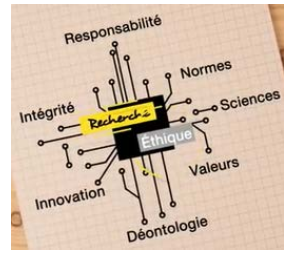
<https://agritrop.cirad.fr/596545/>

Ethical aspects

MOOD has a WP dedicated to Ethical issues.

- Procedures to identify/recruit participants and to collect their consent, templates of the Consent forms, and Information sheets are [available as deliverable D8.1 of the project](#).
- Measures to safeguard the rights and freedoms of the data subjects/research participants are also [available as deliverable D8.2 of the project](#).

4.2 Données et Ethique



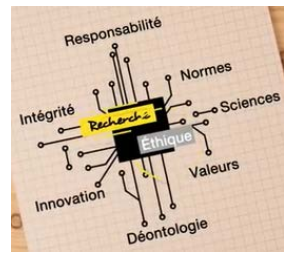
➤ Données dont la collecte, l'utilisation et le partage pourraient avoir un impact sur:

- les personnes, la société, les communautés, la santé
 - ✓ Recherche menées sur des peuples autochtones
 - ✓ Collecte de données de petits agriculteurs
 - ✓ Collecte de données avec usage de techniques invasives
- les animaux, la biodiversité, l'environnement, ...

- Dans le PGD, précisez si vos données pourraient avoir un impact
- si oui, précisez les mesures prises

➔ **L'absence de validation par un comité d'éthique peut vous empêcher de publier !**

4.2 Données et Ethique



➤ Appui CIRAD

- Déléguée à la déontologie et à l'intégrité scientifique

Estelle Jaligot (estelle.jaligot@cirad.fr)

- Comité Éthique INRAE-Cirad-Ifrémer-IRD <https://www.ethique-en-commun.org/>
- Charte de déontologie du Cirad. <https://intranet.cirad.fr/les-rubriques-du-portail/les-actualites/2018/actus/integrite.-une-charte-de-deontologie-explicite-l-engagement-ethique-du-cirad>

➤ Guides

- Ethics self-assessment. Commission Européenne. 2021. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/how-to-complete-your-ethics-self-assessment_en.pdf
- Data Ethics Canvas. ODI. 2021 <https://theodi.org/article/the-data-ethics-canvas-2021/>
Version courte en français: <https://theodi.org/wp-content/uploads/2021/07/Data-Ethics-Canvas-French-Colour.pdf>
- Respect des peuples et savoirs autochtones : [CARE](#), [OCAP](#)
- Protection de l'environnement, biodiversité, animaux,... <https://docs.gbif.org/sensitive-species-best-practices/master/fr/>

4.2 Données et Ethique

➤ Questions à se poser pour évaluer le risque d'impact

- Certains individus, groupes ou organisations pourraient-ils être affectés négativement par ce projet ?

Si oui: quelles mesures pouvez-vous prendre pour minimiser les dommages ?

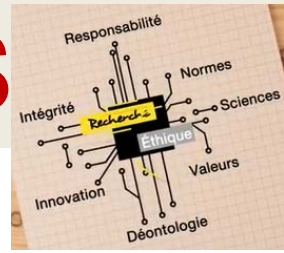
- Est-ce que vous améliorez les choses pour la société ou l'environnement ?

Si oui: comment et pour qui ?

- Quels sont les individus, groupes ou organisations qui seront positivement touchés par ce projet ?

Comment mesurez-vous l'impact positif ?

4.2 Exemple du projet SOLIDUS



- Ethics self-assessment
 - the SOLIDUS Consortium is taking into account the ethical criteria described in the Respect Code of Practice on “Professional and ethical code for socio-economic research in the information society”.
- Respect of national and international legislation on discrimination
- Personal data protection
- National legal frameworks of personal data collection
- Procedures and criteria to identify and recruit participants
- Informed Consent form and information sheet

<https://cordis.europa.eu/project/id/649489/results>

4.3 Données et protocole de Nagoya



- Données issues de **ressources génétiques (RG)** ou **connaissances traditionnelles associées (CTA)**
- Obligation de respecter le protocole de Nagoya :
 - Accord international adopté en 2010, entré en vigueur en 2014
 - Règle l'accès aux RG et aux CTA
 - Sauf**
 - RG et CTA acquises avant octobre 2014
 - RG couvertes par le TIRPAA - [Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture](#).
 - Vise un **partage juste et équitable** des avantages issus de l'utilisation des RG et des CTA entre fournisseur et utilisateur
 - Les avantages doivent être utilisés au bénéfice de la biodiversité.

4.3 Données et protocole de Nagoya



- Règles d'Accès et Partage des Avantages (APA) varient selon le pays fournisseur :

Centre d'échange d'informations (CEE-APA): site interactif : <https://absch.cbd.int/fr/>

Fiches pays : <https://www.fondationbiodiversite.fr/fiches-pays-relatives-a-la-reglementation-apa/>

- Plus d'infos sur Nagoya et APA

- Fondation pour la recherche sur la biodiversité (FRB) : <https://www.fondationbiodiversite.fr/les-enjeux-de-la-biodiversite/biodiversite-et-reglementation/zoom-apa/>
- Ministère de la Transition Ecologique : https://www.ecologie.gouv.fr/acces-et-partage-des-avantages-decoulant-lutilisation-des-ressources-genetiques-et-des-connaissances#scroll-nav_1

- Appui Cirad

- Cellule APA au Cirad: <https://intranet-dgdrs.cirad.fr/instances-interne/corebio/cellule-apa>

4.3 Exemple du projet TROPIBIO



Compliance with the **Nagoya Protocol** on Access and Benefit Sharing, EU Regulation No 511/2014 and Portuguese Decree-Law 1222/2017, related to the collection and analysis of genetic samples and the generation of the TropiBio data is all thoroughly **outlined in the TropiBio Grant Agreement.**

Contacts have been made with **Namibia** regarding Ethics Clearance/Permits and ABS issues.

Informal contacts have also been established regarding these issues in **Guinea Bissau** with Instituto da Biodiversidade e Áreas Protegidas.

4.4 Données et propriété intellectuelle



➤ Données avec droits de **propriété intellectuelle**

- obtenues en partenariat
- obtenues avec 1 acteur privé pour être exploitées
- images protégées par droit d'auteur,
- données préexistantes sous licence, issues d'anciens projets

➤ Vous devez prendre en compte et respecter

- les droits d'auteurs, les licences, les brevets
- les accords, les contrats, les accords de consortium

➤ Pour compléter :

Les données : Les questions à se poser pour leur diffusion. 2021 : https://doranum.fr/aspects-juridiques-ethiques/les-donnees-les-questions-a-se-poser-pour-leur-diffusion_10_13143_8f0g-6491/

Cours introductif sur les aspects juridiques et éthiques. 2022 : https://doranum.fr/aspects-juridiques-ethiques/cours-introductif-sur-les-aspects-juridiques-et-ethiques_10_13143_eanf-pz90/

4.4 Données et propriété intellectuelle



➤ Question à se poser:

- Du matériel protégé par des droits spécifiques sera t-il utilisé dans le cadre du projet ?
- Qui aura le droit de contrôler l'accès aux jeux de données ?

➤ Dans le PGD:

- Mentionnez les jeux de données concernés et les droits de PI à considérer
- Spécifiez ce qui est décrit dans le contrat de partenariat ou l'accord de consortium du projet
- Précisez pourquoi ces données ne seront pas partagées ou ce que vous mettez en œuvre pour qu'elles le soient.

4.4 Exemple du projet LandMark

<https://landmark2020.eu/>



2.1/ PLANS FOR INTELLECTUAL PROPERTY RIGHTS (BACKGROUND AND FOREGROUND)	31
2.1.1/ GRANT AGREEMENT (BACKGROUND)	31
2.1.2/ CONSORTIUM AGREEMENT (BACKGROUND)	31
2.1.3/ FOREGROUND	32
2.2/ SHARING DATA AND METADATA	33
2.3/ HANDLING, SHARING, PROTECTING, MAKING DATA OPENLY ACCESSIBLE	36
2.4/ MAKING DATA INTEROPERABLE	36
2.5/ INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES) 37	

2.5.1/ Data ownership: **2.5.1/ Data ownership:**

2.5.2/ Joint ownership agreement:

1. **Collected data:** following the **Consortium Agreement**, data are subjected to legal restrictions or limits, including those imposed by the rights of third parties.

2. **Produced data:** third parties and partners who have participated in the production of data will share the property rights of the derived data.

4. Exigences légales et éthiques

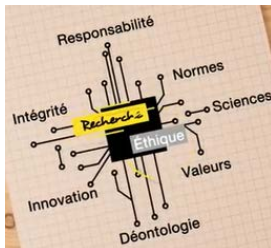


▪ Quelles procédures avez-vous mis en œuvre pour respecter les exigences ?

- Certaines de vos données sont-elles soumises à l'une des exigences décrites (RGPD, Nagoya, Ethique, PI) ?
 - Vos données contiennent-elles des données identifiantes ?
 - Posent-elles des questions éthiques ?
 - Sont-elles soumises à des droits ?
 - Prélevez-vous des échantillons biologiques dans des pays soumis au protocole de Nagoya ?
- Si « Oui » : quelles mesures avez-vous mis en place ?

➔ **10 minutes**

4. Exigences légales et éthiques



Contenu d'un PGD

1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données
6. Responsabilités et ressources

5. Partage des données



- **Partage des données (*Data sharing*) = au-delà du projet**
 - être utile à d'autres chercheurs et d'autres acteurs
 - accroître l'impact du projet
 - faire avancer la science et la société.

- **Questions posées dans le PGD**

- Quelles données seront partagées ?
- Où ? = dans quels entrepôts de données
- Quand ?
- Comment ?

→ Souvent réponses non connues au début du projet

➔ Dans la V1: montrez que le consortium du projet réfléchit à une stratégie de partage des données

Plus de détails dans la V2 puis complet dans la V3

5.1 Quelles données partager ?

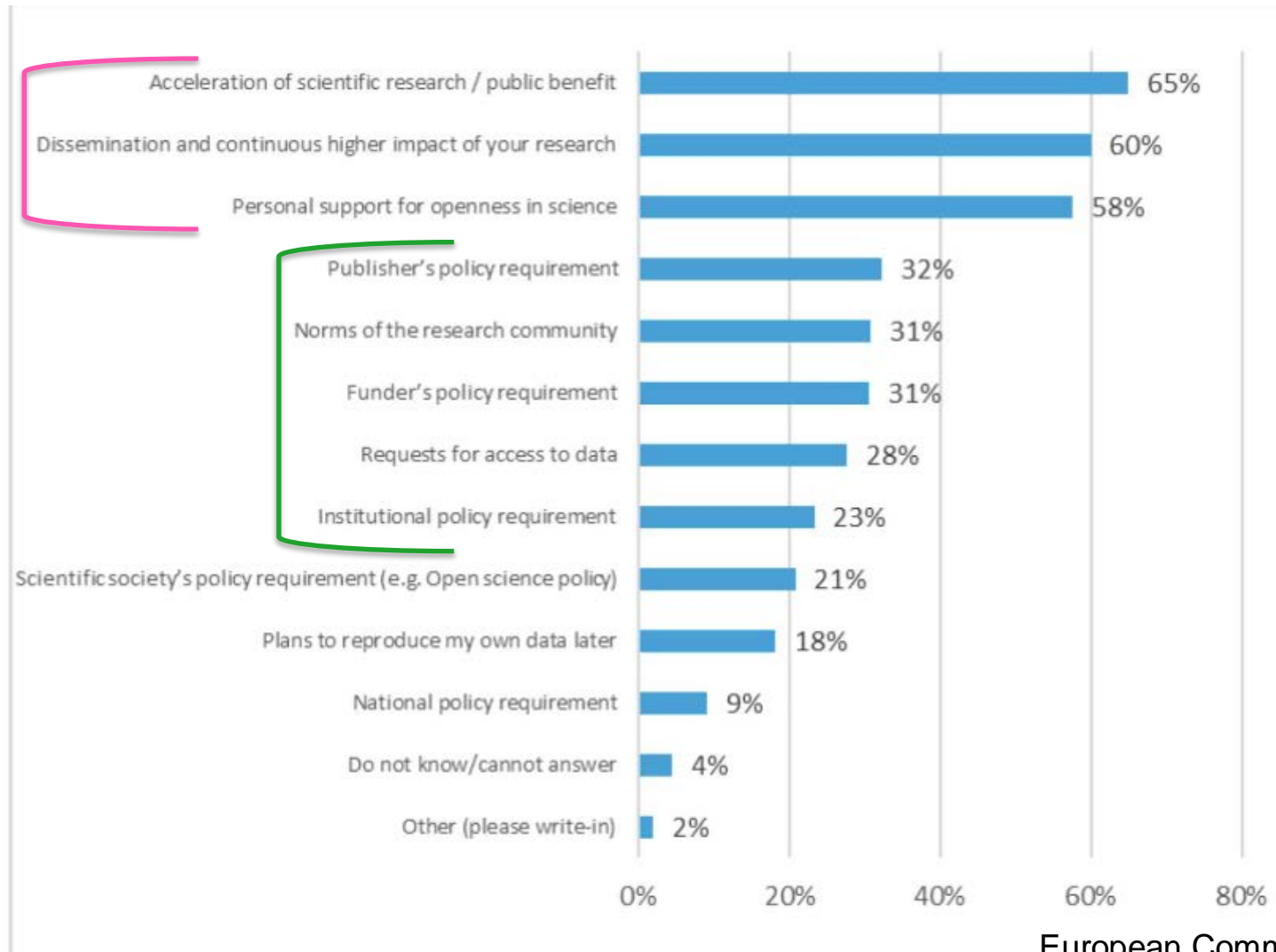


- A minima, les jeux de données à l'origine des articles de recherche publiés
- Recommandé pour les autres jeux de données
 - s'ils ont un potentiel de réutilisation
 - S'ils ont de la valeur scientifique, économique, historique, environnementale, sociétale, ...
 - S'ils sont uniques, originaux, innovants, couteux
 - S'ils présentent un intérêt pour d'autres acteurs: scientifiques, citoyens, communautés locales, décideurs, ONGs, privés
 - S'ils peuvent faire avancer la science ou la société

5. Partage des données



➤ Pourquoi les chercheurs déposent dans un entrepôt



5. Partage des données



➤ Avantages du dépôt des données dans 1 entrepôt

- Vous êtes identifié en tant que créateur des données → **visibilité sur votre travail et vos données**
- Vos données sont **préservées sur le long terme**
- Vous contribuez à la transparence et à l'intégrité des processus de recherche → **rigueur scientifique et bonnes pratiques de recherche**
- Vous mettez vos données à disposition de la communauté et contribuez à la science ouverte et aux avancées scientifiques et sociétales
→ **Satisfaction, fierté**
- Vos données sont visibles, accessibles et réutilisées
→ **vous êtes cité, reconnu**



collaborations

citations

notoriété

financements

5.1 Quelles données partager ?



- **Exceptions acceptées pour raisons :**
 - Ethiques, données non anonymisées, de santé,
 - Protection des droits humains, des savoirs autochtones
 - Risque d'impact sur des personnes, des communautés
 - Protection d'écosystèmes, d'espèces rares, menacées ou en danger
 - Risque d'impact sur l'environnement, la biodiversité,...
 - Protection des droits de PI, partenariat privé, confidentialité,
 - Droits d'auteurs, droit du secret, sécurité nationale....
 - Données stratégiques que vous souhaitez exploiter
- **Restriction d'accès et de réutilisation**
 - Accès seulement sur demande dans un premier temps
 - Accès après embargo pour publication ou brevet
 - Attribution de licences restrictives (ex: pas d'utilisation commerciale)

➡ **Doit être justifié dans le PGD**

5.1 Quelles données partager ?



- Préciser si certaines données ne seront pas conservées Et selon quels critères de choix
- ou doivent être détruites pour des raisons contractuelles, légales ou réglementaires.

Exemples de données à détruire

- données à caractère personnel selon recommandations CNIL et RGPD
- données non réutilisables car de mauvaise qualité
- données intermédiaires qui n'ont pas besoin d'être conservées
- données de simulation facilement reproductibles.

- Contribuez à la sobriété énergétique

Pensez au coût que représente le stockage de données !!

« Aujourd'hui, le numérique représente déjà environ 4 % des émissions de gaz à effet de serre (GES) mondiales »

5.2 Où déposer ses données ?



➤ Dans un entrepôt de données

= infrastructure prévue pour accueillir des données de recherche.
Vous pouvez y déposer vos données et télécharger d'autres données.

➤ Choisir 1 entrepôt recommandé par sa communauté, la revue, bailleur, ... “deposit data with a community-recognized repositories”

- Délivrant un identifiant numérique pérenne et unique
- Adapté à vos données et à la taille de vos fichiers
- Attribuant la licence de diffusion voulue
- Gratuit (la plupart) ou coût raisonnable

➤ Entrepôt institutionnel, entrepôt de projet Entrepôt disciplinaire, thématique ou généraliste



Ni un site web de projet , ni Research Gate !!!!!

5.2 Où déposer ses données ?



Connaissez-vous des **entrepôts de données dans votre domaine ?**

pour déposer des données ?

pour trouver des données fiables ?

5.2 Où déposer ses données ?



➤ Entrepôts généralistes

Dryad, FigShare, Dataverse, Open Science Framework

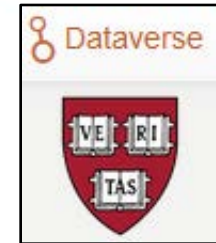


➤ Institutionnels, Nationaux, Européens

Dataverse (Universités, Cirad, IRD, CEA, ...)

Recherche Data Gouv (entrepôt national)

Zenodo, Eudat, B2Share



➤ Portés par des éditeurs de revues

Oxford Univ Press (GigaDB), Ubiquity Press (Dataverse), Elsevier (Mendeley Data)



➤ Thématiques ou disciplinaires: recommandés par les revues !

5.2 Où déposer ses données ?



➤ Entrepôts thématiques

Biodiversité : Global Biodiversity Information Facility



Phenotypage de plantes:



Sciences de la terre :



Microbio alimentaire:



OMICs :



GenBank



Pastoralisme – Mobilité animale :



Biologie



Enquêtes:



SHS:



Codes, Logiciels :



Software Heritage



L'entrepôt Dataverse du Cirad

➤ Entrepôt de données institutionnel



- Ouvert aux UR, aux projets et aux partenaires de projets
- Tous types de données + documentation associée
- Dépôt gratuit
- Espace privé pour partager des données entre partenaires
- Choix du moment et de la licence de diffusion des données
- Appui DIST : Céline Barthelemy et Gwenaël Doux.

Guides d'entrepôts de données



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

<https://www.re3data.org/search>

3152 entrepôts

Home Search Browse Suggest FAQ About Schema API Contact Imprint

Search for Repositories (1335 reviewed repositories)

Search

Subject Content Type Country (of the responsible institutions)

Add subjects Add content types Add countries

Certificates Open Access Persistent Identifier

Reset filter

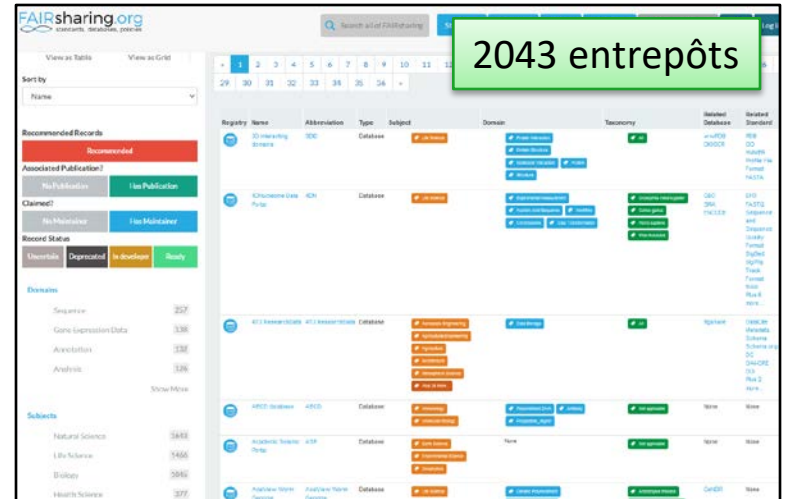
1335 results (1 - 25)

Sort by Weight



Informe aussi sur des entrepôts portés par des organismes internationaux (ex: FAOStat, WorldClim, WorldPop, ISRIC...) donnant accès à des données de référence.

<https://fairsharing.org/search?fairsharingRegistry=Database>



FAIRsharing.org

2043 entrepôts

Registry Name Abbreviation Type Subject Domain Security Related Database

Registry	Name	Abbreviation	Type	Subject	Domain	Security	Related Database
Recommended	Recommended						
Associated Publication?	No Publication	Yes Publication					
Claimed?	No Publication	Yes Publication					
Record Status	Unusable	Discontinued	In circulation	Ready			
Reviews	Signature	357					
	Gene Expression Data	338					
	Abundance	338					
	Analysis	338					
Subjects	Natural Science	348					
	Life Science	348					
	Biology	348					
	Health Science	348					

Entrepôts en biologie moléculaire



<https://elixir-europe.org/platforms/data/elixir-deposition-databases>



<https://ngdc.cncb.ac.cn/databasecommons/>

Répertoire re3data

re3data.org

Sea

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
<https://www.re3data.org/search>

Filter

- Subjects
- Content Types
- Countries
- AID systems
- API
- Data access
- Data access restrictions
- Database access
- Database access restrictions
- Database licenses
- Data licenses
- Data upload
- Data upload restrictions
- Enhanced publication
- Institution responsibility type
- Institution type
- Keywords
- Metadata standards
- PID systems
- Provider types
- Quality management
- Repository languages
- Software
- Syndications
- Repository types
- Versioning

Mot-clé

pathogen

Previous 1 Next

Found 18 result(s)

Nombre de résultats

Pathogen Portal

the Bioinformatics Resource Centers Portal

Subject(s)

Basic Biological and Medical Research Microbiology, Virology and Immunology Bioinformatics and Theoretical Biology

Content type(s)

Standard office documents Scientific and statistical data formats Raw data Structured text Software applications

Country

United States

Pathogen Portal is a repository linking to the Bioinformatics Resource Centers (BRCs) sponsored by the National Institute of Allergy and Infectious Disease. The BRCs are providing web-based resources to scientific community conducting basic and applied research on organisms causing emerging or re-emerging diseases. The Pathogen Portal supports and links to five Bioinformatics Resource Centers: different group of pathogens, focusing on, but not limited to, pathogens causing (Re-)Emerging Infectious Diseases, and those in the NIAID Category A-I research. The scope of the BRCs also includes Invertebrate Vectors of Human Disease. Pathogen Portal covers EuPathDB, IRD, PATRIC, VectorBase and

Host - Pathogen Interaction Database

HPIDB

Subject(s)

Basic Biological and Medical Research Microbiology, Virology and Immunology Biology Life Sciences Medicine

Content type(s)

Standard office documents Raw data

Country

United States

HPIDB is a public resource, which integrates experimental PPIs from various databases into a single database. The Host-Pathogen Interaction Database is a resource for understanding molecular interactions between key organisms and the pathogens to which they are susceptible.

SilkPathDB

Silkworm Pathogen Database

Subject(s)

Biology Life Sciences Basic Biological and Medical Research Pathology and Forensic Medicine Microbiology, Virology and Immunology

Content type(s)

Standard office documents Scientific and statistical data formats Images Raw data Software applications other

Country

China

Silkworm Pathogen Database (SilkPathDB) is a comprehensive resource for studying on pathogens of silkworm, including microsporidia, fungi, bacteria not only genomic data including functional annotation of genes and gene products, but also extensive biological information for gene expression data and will be help with researches on pathogens of silkworm as well as other Lepidoptera insects.

Répertoire re3data

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
<https://www.re3data.org/search>

SEANOE

General

Institutions

Terms

Standards

Name of repository

SEANOE

Additional name(s)

Sea scientific open data publication
Publication de données scientifiques marines

Repository URL

<https://www.seanoe.org/>

Subject(s)

Geosciences (including Geography) Life Sciences Natural Sciences Biology
Agriculture, Forestry, Horticulture and Veterinary Medicine Chemistry Physics

Description

"Seanoe (SEA scientific Open data Edition) is a publisher of scientific data in the field of marine sciences. It is operated by Ifremer (<http://wwz.ifremer.fr/>). Data published by SEANOE are available free. They can be used in accordance with the terms of the Creative Commons license selected by the author of data. Seance contributes to Open Access / Open Science movement for a free access for everyone to all scientific data financed by public funds for the benefit of research. An embargo limited to 2 years on a set of data is possible; for example to restrict access to data of a publication under scientific review. Each data set published by SEANOE has a DOI which enables it to be cited in a publication in a reliable and sustainable way. The long-term preservation of data filed in SEANOE is ensured by Ifremer infrastructure. "

Contact

frederic.merceur@ifremer.fr
data@seanoe.org

Content type(s)

Raw data other Standard office documents Structured graphics

Keyword(s)

marine sciences

Persistent identifier(s) of the repository

ISSN 2491-1836
FAIRsharing_doi:10.25504/FAIRsharing.bb5rna

Repository type(s)

disciplinary

Mission statement for designated community

<https://www.seanoe.org/html/about.htm>

modalités et licences

embargo

Répertoire re3data

SEANOE

General Institutions **Terms** Standards

modalités et licences

Policies (1)

Policy Name [About SEANOE](#)

Database access

Type of access to research data repository open

Modalités d'accès aux données

Data access (3)

Type of access to data open

Type of access to data embargoed

Type of access to data restricted

Data access restriction type(s) other

Modalités de dépôt de données

Data licenses (1)

DataLicense [CC](#)

Data upload (1)

Type of data upload open

Data upload license (1)

Data upload license name [Publish your marine data](#)

5.3 Modalités de partage



- Dans le PGD: préciser les modalités de partage

= quand et comment vos données seront accessibles et réutilisables

ex : embargo, accès ouvert ou restreint, réutilisation libre ou limitée, ...

- Les modalités dépendent :

- des exigences du bailleur, de votre institution, de lois, de la revue dans laquelle vous voulez publier,
- du type de données : ex: données OMICs
- du respect du cadre éthique et juridique
 - Avez-vous l'accord des contributeurs ?
 - Avez-vous tous les droits pour choisir ?
 - Avez-vous respecté les réglementations ? RGPD, Nagoya, ...
 - Avez-vous anonymisé les données ?, ...

5.3 Modalités de partage

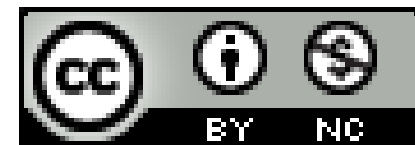
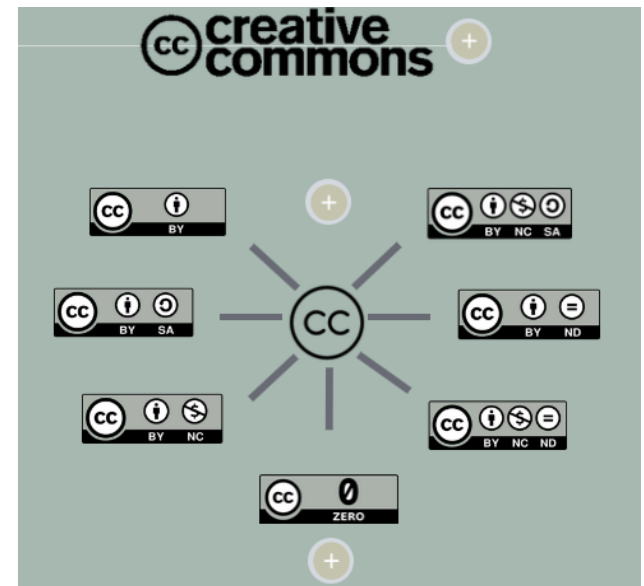


- **Attribuer une licence de diffusion à ses données**
 - La licence sert à préciser les droits dont disposent les futurs réutilisateurs pour réutiliser les données
 - les auteurs ont donc la possibilité de définir comment et jusqu'où on peut réutiliser leurs données.
 - Créées en 2002 aux USA par un organisme à but non lucratif
 - Dispositif juridique très simple et très utilisé
 - Plusieurs licences possibles

- **CC-BY : la + recommandée**



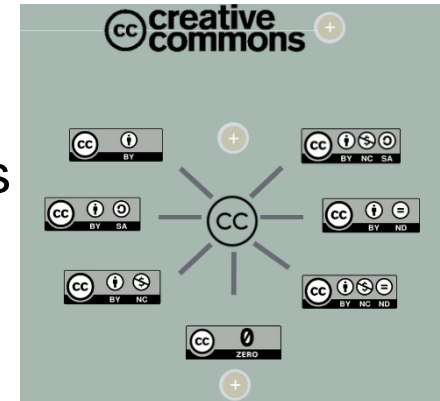
Vous pouvez réutiliser les données
A condition de citer les auteurs




5.3 Modalités de partage



- **Choix de la licence de diffusion**
 - Généralement l'entrepôt propose plusieurs licences
→ Vous choisissez



- **Attention à certains entrepôts**

- Qui impose une seule licence: la licence CC0 
= domaine public, aucune restriction d'usage
et sans obligation de citer les créateurs des données



GenBank



5.3 Modalités de partage



➤ Modalités identiques pour tous les jeux de données

Tous les jeux de données seront déposés dans le Dataverse Cirad dédié au projet, avec accès libre à tous sous licence CC-BY

➤ Modalités différentes selon le type de données

Données génomiques → GenBank → accès immédiat, licence CC0

Données d'enquêtes → Dataverse Cirad avec un embargo d'un an pour anonymisation puis accès aux données sous licence CC-BY.

Données biochimiques (issu d'un partenariat privé) → entrepôt non défini, licence CC-BY-NC excluant toute utilisation commerciale qui sera réservée au partenaire privé pour développer des tests de diagnostic.

➡ Le PGD aide à clarifier, dès le début du projet, les objectifs et droits des partenaires sur les données produites

5. Partage des données



■ Envisagez-vous de partager vos données ?

- Pensez-vous que vos données puissent intéresser d'autres scientifiques ou d'autres personnes ?
- Pensez-vous qu'elles ont un potentiel de réutilisation ?
- Sont-elles prêtes pour être déposées dans un entrepôt ? selon les principes FAIR ?
- Envisagez-vous de les rendre disponibles ?

Si « Oui » : comment : dans un entrepôt, dans un article, ...

➔ **10 minutes**

5. Exemple du projet POdePO Bourget

Modalités de partage des données après le projet

L'ensemble des données collectées dans le cadre de cette ANR seront disponibles en libre accès (licence CC-BY) après embargo de publication finale des données traitées. Les données seront mises à disposition par l'intermédiaire de l'entrepôt [Zenodo](#), accessibles aussi sur le portail de données du réseau des observatoires des Lacs de Montagnes Savoyards ([OLMS](#)) et dans la banque internationale [Genbank](#) pour les données génétiques

5. Exemple du projet EnvMetaGen



Modalités de partage des données après le projet

Nucleic acid sequence data

Raw reads will be archived in **NCBI Short Reads Archive** or similar archives of un-analysed sequence data or a generic archive such as **DRYAD**.

Sequences that have been inspected and associated with a clearly defined taxon verified by a taxonomic expert will be deposited in **GenBank**, **Barcoding of Life Database** or **DRYAD**.

Taxonomic data

Taxonomic information relating to DNA sequences will be archived in custom databases with links to the sequences and organisms associated with them. Where appropriate, sub-sets of this data will also be deposited in **BOLD** or **GenBank**.

Biodiversity data

Where appropriate the data on species occurrences will be added to the **GBIF** (Global Biodiversity Information Facility). Associated metadata such as collection locations for type specimens or photographs will be archived according to the requirements of the data repositories.

<https://cordis.europa.eu/project/id/668981/results>

5. Exemple du projet TERRI4SOL



Modalités de partage des données après le projet

DATASET N°1 "BASE DE DONNÉES DES PROPRIÉTÉS PHYSIQUES ET CHIMIQUE, CAS DES SOLS DE LA ZONE POST-FORESTIÈRE EN CÔTE D'IVOIRE"

- Les jeux de données issus du dispositif de l'analyse en laboratoire des échantillons de sol seront déposés sur le **Dataverse du Cirad** et **disponibles sous demande aux auteurs les 3 premières années** avant d'être en Open access.

DATASET N°2 "TRANSITIONS D'USAGE DES TERRES DANS LA RÉGION DE LA MÉ ET DE BLÉ"

- Les jeux de données anonymisées, issus de l'enquête dans le village de Mébifon, seront publiés dans le cadre de Policy Brief et **publications scientifiques**. Les **données anonymisées seront mis en accès libre trois ans** après la fin du projet.

DATASET N°3 "BASES DE DONNÉES EXTERNES EN ACCÈS LIBRE"

- En accès libre sur le site des bases de données correspondantes

DATASET N°4 "ENQUÊTE AUPRÈS DES INSTITUTIONS PUBLIQUES IVOIRIENNES"

- Les données collectées et générées au cours du projet ainsi que les métadonnées qui y sont associées seront mises en accès libre **trois ans** après la fin du projet, dans le respect des obligations spécifiques à la gestion des données à caractère personnel (Anonymisation).

https://agritrop.cirad.fr/604742/1/DMP_du_projet_TERRI4SOL_produits_recherche_20230522_v1_finale.pdf

5. Exemple du projet EcoStack



Repository name	Short description	Type	Web link
Zenodo	EC digital library connected with OpenAIRE project (research data network & exchange)	General	https://zenodo.org/
The Knowledge Network for Biocomplexity (KNB)	Data repository for ecological, environmental and earth science research	Research field	https://knb.ecoinformatics.org/
Dryad	Digital data repository for research and education	General	https://datadryad.org/stash/
The European Nucleotide Archive (ENA)	Data repository for nucleotide sequencing information	Research field	https://www.ebi.ac.uk/ena
The European Bioinformatics Institute (EMBL-EBI)	Data repository for biological data and life sciences	Research field	https://www.ebi.ac.uk/
Sequence Read Archive (SRA)	Repository for high throughput sequencing data	Research field	https://www.ncbi.nlm.nih.gov/sra
Newcastle University data repository	Institutional repository for university research	Institutional	https://data.ncl.ac.uk/

<https://cordis.europa.eu/project/id/773554/results>

5. Données et Entrepôts



- **Quiz** : jeu sur les entrepôts de données

https://doranum.fr/depot-entrepots/deposez-un-jeu-sur-les-entrepots-de-donnees_10_13143_aabd-hs57/

Dép'Osez ! Le jeu sur les entrepôts de données

COMMENCER LE MODULE

A retenir sur le partage



- En pratique dans la 1ere version du PGD, montrez :
 - que le collectif du projet réfléchit à la meilleure stratégie de partage des données après le projet, pour le bénéfice de tous.
- Les données à partager sont :
 - à *minima* celles à l'origine d'articles de recherche
 - pour les autres: privilégiez le partage des données de qualité, qui ont un potentiel de réutilisation et qui peuvent contribuer aux avancées scientifiques et sociétales
- Dépôt des données dans un entrepôt de confiance :
 - le plus utilisé dans votre discipline
 - si possible au maximum à la fin du projet
 - avec une licence clarifiant les conditions de diffusion et réutilisation.

5. Partage des données



Contenu d'un PGD

1. Description des données
2. Documentation et qualité des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Partage des données
6. Responsabilités et ressources

6. Responsabilités et ressources



➤ Responsabilités

- Préciser la répartition des responsabilités

➤ Ressources

- Préciser comment les coûts de gestion des données sont couverts.

6. Responsabilités



- Nommer les personnes qui auront 1 responsabilité dans:
 - Gestion des données
 - Animation et rédaction du PGD
 - Collecte
 - Traitement
 - Stockage
 - Anonymisation
 -
 - ➔ nom(s), équipe, institution et e-mail(s)

- Impliquer vos partenaires

- Expliquer comment s'effectue la coordination des responsabilités de gestion des données entre partenaires.

6. Exemple du projet OPERA



The **Project Coordinator** has a particular responsibility to ensure that data shared through the OPERA website are easily available, but also that backups are performed and that proprietary data are secured.

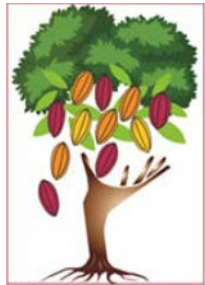
WP1 leader will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.

Validation and registration of datasets and metadata is the **responsibility of the partner that generates the data in the WP.**

Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data.

Quality control of these data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

6. Exemple du projet Cocoa4Future



Cocoa4Future

Tableau 3 : Répartition de la mise en œuvre de la gestion des données (en particulier des données personnelles) générées dans le cadre du projet C4F

Module d'activités	Activité de traitement (code)	Partie/entité responsable (institution partenaire)
1.1.	Caractérisation de la complexité des systèmes de cacaoculture (1.1.1)	- Peuplements cacaoyers : Wascal (UFHB) - Espèces associées : laboratoire de botanique (UFHB), FOREN (ESA/INP-HB) et UR Forêts et Sociétés (Cirad)
	Evaluation des SE fournis par les systèmes de cacaoculture (1.1.2)	- Stockage carbone et Biodiversité : laboratoire de botanique (UFHB) et UR Forêts et Sociétés (Cirad) - Fourniture de bois et PFNL : FOREN (ESA/INP-HB) et UR Forêts et Sociétés (Cirad) - Rendement cacao : Wascal (UFHB) - Qualité du sol : LEDD (UNA)
	Evaluation socio-économique des systèmes de cacaoculture (1.1.2)	JRC Séville
	EMC ² des performances des systèmes de cacaoculture (1.1.3)	UMR ABSys (Cirad)
	Modélisation éco-physiologique de la réponse du cacaoyer à la sécheresse et à l'ombrage en Côte d'Ivoire (1.1.4.)	Dispositif expérimental SAO : UMR ABSys (Cirad) Dispositif milieu réel : Wascal (UFHB)
	Modélisation éco-physiologique de la réponse du cacaoyer à la sécheresse et à l'ombrage au Ghana (1.1.4.)	CRIG
	Simulation de l'évolution des SE sous différents scénarios (1.1.5.)	UR Forêts et Sociétés (Cirad)
Co-conception de SCAF ³ (1.1.6.)	UMR ABSys (Cirad)	
1.2.	Activités sur la maladie du CSSV ⁴ (1.2.1. à 1.2.5.)	Etudes de diversité (collecte, tri, identifications) : Wascal (UFHB) et UMR PHIM (Cirad) Séquençage : UMR CBGP (Cirad) Téledétection : CURAT (UFHB) Ecologie du CSSV : Wascal et UMR PHIM (Cirad)
2.1.	Dispositif de suivi d'exploitations cacaoyères en Côte d'Ivoire (2.1.1.)	GRIEPE (UJLoG)
	Dispositif de suivi d'exploitations cacaoyères au Ghana (2.1.1.)	School of Agriculture (University of Ghana)
	Dispositif de suivi d'exploitations piscicoles en Côte d'Ivoire (2.1.2. et 2.1.3.)	APDRA Pisciculture paysanne
	Dispositif d'accompagnement de cacaoculteurs en Côte d'Ivoire (2.1.4. et 2.1.5.)	Nitidae : filières et territoires
2.2.	Activités sur qualité du cacao (2.2.1. et 2.2.2)	Département LBMA (UNA)
	Activités sur innovations organisationnelles au Ghana	ISSER (Université du Ghana) UMR Innovation (Cirad)



6. Ressources



- Ressources dédiées à la gestion des données
 - Coûts éligibles
 - Matériel, traitement, personnel
 - La gestion des données du projet représente t-elle un coût ?
 - Ce coût a-t-il été prévu dans le budget du projet ?
 - temps de personnel
 - achat matériel
 - frais technique :
 - Collecte, stockage, dépôt en entrepôt, traitement, anonymisation,...
 - Si non prévu dans le budget :
 - Y a-t-il un coût ?
 - Quelles ont les sources du financement

6. Ressources



➤ Guide et outils d'évaluation des coûts

[Le coût de la gestion des données.](#)

<https://www.jisc.ac.uk/guides/rdm-toolkit/costing>

<https://ukdataservice.ac.uk/learning-hub/research-data-management/plan-to-share/costing/>

<https://www.ukdataservice.ac.uk/media/622368/costingtool.pdf>

<https://bbf.enssib.fr/consulter/bbf-2021-00-0000-057>

6. Exemple du projet POdePO Bourget



Type de coût (Titre)	Montant	Etape du cycle de vie
Matériel informatique (Matériel de prelevements)	2000 EUR	Coûts liés à la collecte/production des données - Carnet et tableur
Stockage (Archivage à longue terme)	3000 EUR	Coûts liés à la conservation à long terme des données - Carnet et tableur
Stockage (Hébergement)	5000 EUR	Coûts liés au stockage et à la sauvegarde des données - BDD
Stockage (Estimation du coût de stockage et calcul auprès du mésocentre MESO@LR)	165.0 EUR	Coût total de stockage (coût calculé automatiquement) - Carnet et tableur
Frais de personnel (Recrutement des personelles)	10000 EUR	Coûts liés à la collecte/production des données - Carnet et tableur

6. Exemples REPAIR - OPTIMA



PGD du projet **REPAIR**: <https://zenodo.org/record/1438301#.YZPTkLrjKUn>

1.6 Allocation of Resources

Long term storage. The costs for depositing data at the 4.TU Center for Research Data for 15 years are €4,50 per GB. In total about €6.000 are allocated for long term data storage. This cost is estimated to be sufficient for storing all preservable REPAIR research data of all case studies.

Journal open access costs. €30.000 are allocated for OpenAccess Publishing, cost for open access publishing vary significantly dependent on the scientific journal but this should be sufficient for at least 10 peer reviewed articles.

PGD du projet **CIRC4Life**: <https://cordis.europa.eu/project/id/776503/results>

4 Allocation of Resources

The costs of using Arkivum Archiving-as-a-Service solution available at NTU at the end of the project is **£0.50 per GB of archived data**. As the projected final amount of data is expected to be in the order of 100 Gigabytes, this will cost approximately **£50 per year**, or £500 for ten years.

6. Responsabilités et ressources



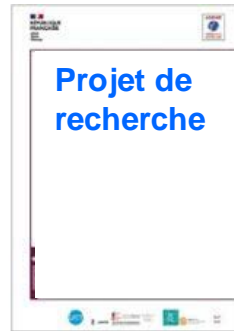
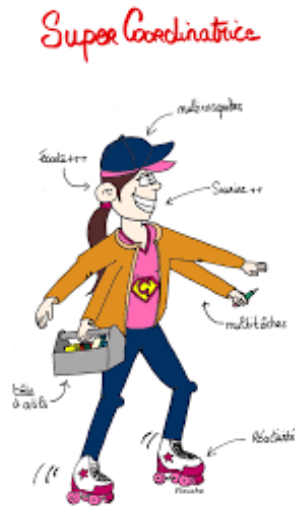
Plan de gestion des données

Comment faire ?



Rédiger un PGD de projet

- Vous êtes coordinateur d'un projet multi-partenaires



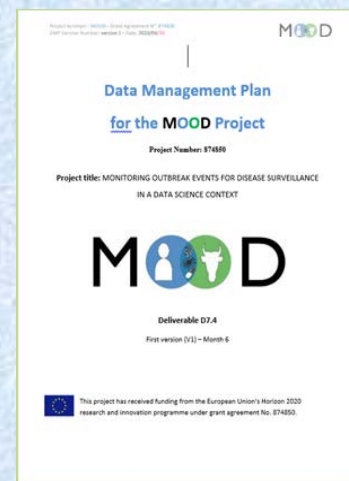
- Comment faites-vous pour produire un PGD de qualité 6 mois après le début du projet ?

PGD : comment faire ?

1. Identifier le modèle de PGD proposé par le bailleur
Horizon Europe, ANR, Biodiversa, Wellcome Trust,
Modèles Cirad : <https://intranet-data.cirad.fr/modeles-cirad-de-pgd>
DMP OPIDoR (accès à nombreux modèles): <https://dmp.opidor.fr/>
2. Nommer 1 responsable pour coordonner le PGD du projet +
1 responsable par work package, type de données ou pays
3. Collecter les informations auprès des responsables ou des
partenaires en utilisant les outils que vous voulez:
fichier doc ou Excel, DMP OPIDoR, ...
4. Renseigner les différentes parties du PGD
5. Faites valider par la coordination du projet + partenaires

Plan de Gestion des Données

Modèles



Modèles de PGD

■ Modèles des bailleurs

- ANR: <https://anr.fr/fileadmin/documents/2019/ANR-modele-PGD.pdf>
- Horizon Europe: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf accès au template pages 46 et 49.

■ Modèles Cirad

- Sur Intranet Data : <https://intranet-data.cirad.fr/modeles-cirad-de-pgd>

■ Outils de rédaction et sources de modèles

- DMP OPIDoR : <https://dmp.opidor.fr/>

→ Accès à nombreux modèles institutionnels (dont Cirad) et bailleurs

Modèles Cirad de PGD

Site Intranet Data <https://intranet-data.cirad.fr/>

Data Cirad
Gérer et valoriser les données scientifiques

GÉRER SES DONNÉES | DROITS | PRÉSERVER ET PUBLIER | RÉUTILISER | OUTILS ET RESSOURCES

Gérer ses données > Plan de gestion des données (PGD)

Plan de gestion des données (PGD)

Un PGD (DMP - Data Management Plan) est un outil pour aider les scientifiques à gérer leurs données dans le cadre d'un projet. Rédiger un PGD en début de projet permet de mettre en place de bonnes pratiques de gestion des données, facilite les échanges entre partenaires et assure un gain de temps pour la publication et le partage raisonné des données en fin de projet.

Un plan de gestion des données évolue dans le temps. Il est vérifié, revu et complété à différentes étapes du projet. A minima : une première version en début de projet ou mois 6 pour les projets H2020 et ANR, une seconde version à mi-parcours et une troisième en fin de projet.

Le PGD répond aux questions :

- Quels sont les types de données qui seront générés ?
- Quels sont les jeux de données existants qui seront mobilisés ?
- Quelles sont les méthodologies qui seront utilisées ?
- Comment les données seront-elles décrites et documentées ?
- Comment les données seront-elles exploitées, partagées et conservées ?
- Quel cadre juridique et éthique s'appliquera aux données ?
- Comment les responsabilités seront-elles distribuées ?

Bonnes pratiques

- Réaliser le PGD dès le début du projet
- Prévoir les jeux de données à produire
- Identifier les métadonnées adaptées
- Clarifier les responsabilités des contributeurs
- Mettre au point des bases de données et des jeux de données propres et valorisables
- Réfléchir aux modes de valorisation

Voir aussi

Les métadonnées

Modèles Cirad de PGD

Rédiger un plan de gestion des données

Formation

La minute Plan de gestion des données (DMP)

DMP OPIDeR

Contact

data-pgd@cirad.fr

Pourquoi rédiger un PGD ?

- Mettre en œuvre de bonnes pratiques de gestion et documentation des données
- Garantir la qualité des recherches et la production de données fiables et compréhensibles
- Contribuer à la transparence, à l'intégrité scientifique et à la reproductibilité des recherches
- Réduire le risque de perte de données ou de données non réutilisables.
- Clarifier les rôles, responsabilités et droits de chaque contributeur.
- Anticiper les problèmes juridiques, éthiques ou techniques
- Assurer la sécurité des données personnelles, sensibles ou stratégiques
- Faciliter le partage des données au sein du collectif
- Prévoir les besoins et coûts pour générer, traiter, conserver et partager les données.
- Répondre à la demande des bailleurs.

Rédiger un Plan de Gestion des Données

- 1 - Qu'est ce qu'un PGD
 - 2 - Pourquoi rédiger un PGD
 - 3 - Structure globale du PGD
 - 4 - Décrire les données
 - 5 - Clarifier le cadre éthique et juridique
 - 6 - Garantir la compréhension et l'accessibilité des données
 - 7 - Assurer le stockage, la sécurité et la sauvegarde des données
 - 8 - Définir les responsabilités
 - 9 - Spécifier les modalités de diffusion et de partage des données
- Liens utiles

<https://coop-ist.cirad.fr/gerer-des-donnees/rediger-un-pgd/1-qu-est-ce-qu-un-pgd>

Site internet d'informations CoopIST

CoopIST
COOPERER EN INFORMATION SCIENTIFIQUE ET TECHNIQUE

La lettre de CoopIST
Archives
Cirad

<https://coop-ist.cirad.fr/>

Trouver l'information | Etre auteur | Rédiger | Publier et diffuser | Evaluer | Gérer des données

Coop IST > Gérer des données

Gérer les données de la recherche

S'initier en ligne aux données de la recherche

Mettre en œuvre une gestion des données

Les métiers liés aux données de la recherche

Se familiariser avec les PGD

Rédiger un PGD

► **S'initier en ligne aux données de la recherche et à leur gestion**

Sélection de formations en ligne en français ou en anglais pour se familiariser avec la gestion des données de recherche et adopter de bonnes pratiques.

► **Mettre en œuvre une gestion des données de la recherche dans une institution**

Les données de la recherche sont un sujet récent et stratégique. La mise en place d'une gestion des données de la recherche doit associer les scientifiques et les services d'appui. Communiquer, former les personnels et identifier les compétences permettent d'ancrer le sujet des données de la recherche dans l'établissement.

Voir aussi

Identifier et rechercher une publication ou un jeu de données par son DOI






Accès aux modèles en format Word et Excel

Modèles Cirad de PGD

Modèles Cirad de PGD

<https://intranet-data.cirad.fr/modeles-cirad-de-pgd>

3 modèles Cirad de PGD sont à votre disposition, à choisir en fonction de votre contexte.

- Les 2 premiers sont adaptés à la majorité des projets, notamment de type  ANR et pour la Version 1 des projets H2020.
- Un  modèle simplifié en français
- Un  modèle en version bilingue français/anglais (2022-02) et en format tabulaire.
- Le  3eme modèle est spécifique des projets financés par l'Europe . Il suit les recommandations du  guide Horizon Europe (février 2022). Il est structuré par les principes *FAIR* qui précisent que les données doivent être : Faciles à trouver, Accessibles, Interopérables et Réutilisables. Son utilisation peut être complexe pour la V1, il peut donc être réservé aux versions ultérieures.

Téléchargez la  notice explicative des 3 modèles.

1 modèle de fichier Excel est aussi proposé pour collecter les informations nécessaires au PGD.

PGD: modèle classique ou FAIR

Modèle classique

1. Description des données
2. Documentation des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Responsabilités et ressources
6. Partage des données

Modèle FAIR projets Horizon Europe

1. Data summary
2. FAIR Data
 - 2.1. Making data Findable
 - 2.2. Making data Accessible
 - 2.3. Making data Interoperable
 - 2.4. Increase data Re-use
3. Other research outputs
Protocoles, codes, logiciels, modèles, ...
4. Allocation of resources
5. Data security
6. Ethics

Modèles Cirad de PGD

- Tableau disponible en annexe des modèles Cirad de PGD

<https://intranet-data.cirad.fr/modeles-cirad-de-pgd>

Nom du jeu de données	Description du jeu de données	Source des données	Modalités d'accès et d'utilisation
NBN Gateway	<i>Species and habitat data, including National Vegetation Classification, Phase 1 Habitat Classification</i>	<i>National Biodiversity Network</i>	<i>Full access allowed through www.searchnbn.net</i>
<i>Land Cover Map 2000 (LCM2000)</i>	<i>Satellite derived remote-sensed datasets providing broad habitat definitions. GIS vector (polygon) dataset at 0.1m resolution.</i>	<i>Centre for Ecology & Hydrology</i>	<i>Forest Research have licence agreement</i>
<i>National Inventory of Woodland and Trees</i>	<i>Derived from LCS88 dataset plus updated to 1995 from Forestry Commission sources; provides info on broadleaved/conifer woodland > 2ha and small woods and trees (0.1-2ha)</i>	<i>Forestry Commission</i>	<i>Full access as part of Forestry Commission</i>
<i>Incidence of reported zoonotic disease</i>	<i>Actual cases of Lyme disease and unconfirmed cases; if possible to be used as patient sample</i>	<i>Health Protection Agency</i>	<i>Potential confidentiality issues</i>

Modèles Cirad de PGD

Intranet Data : <https://intranet-data.cirad.fr/modeles-cirad-de-pgd>

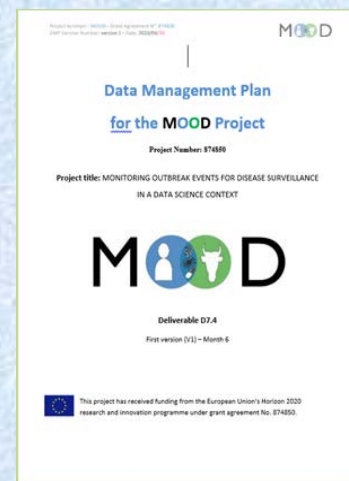
Nom du jeu de données ou identifiant	Type de données	Description du jeu de données	Pays dans lequel est collecté le jeu de données		Work package/Activité/Tache du projet	Mise à jour (format YYYYMMDD)	Commentaires
				Site de collecte (si pertinent)			
Dataset-1 : Pratiques d'alimentation des ruminants dans les systèmes sylvo-pastoraux et agrosylvo-pastoraux recensées par enquêtes auprès des éleveurs et agroéleveurs	Données d'enquêtes	A l'aide d'enquêtes, des informations (données qualitatives) sur les pratiques d'alimentation utilisées par les éleveurs et agroéleveurs seront recueillies afin d'identifier la diversité des régimes alimentaires du bétail	Niger	commune de Dantchandou	WP 1	2021-09-01	Données qualitatives
Dataset-2 : Distribution du papillon <i>Vanessa cardui</i> (Belles-Dames) dans la réserve naturelle de la Somone au Sénégal	Données de biodiversité / données entomologiques	La distribution de <i>Vanessa cardui</i> sera déterminée par observations directes sur le terrain, dans la réserve naturelle de la Somone au Sénégal. Trois séries annuelles d'observations auront lieu du 1er février au 30 juin en 2020, puis en 2021 et puis en 2022. Les données de présence (occurrence) de <i>Vanessa Cardui</i> seront accompagnées d'informations caractérisant : la localisation géographique (latitude, longitude, altitude), la date d'observation, la température et le type d'habitat (plaines, forêts, rivière, etc.). Des photographies de chaque observation de Belles-Dames seront aussi réalisées. Le jeu de données final sera constitué de ces 3 jeux de données annuelles, nommés: BelDamSo_2020, BelDamSo_2021, BelDamSo_2022.	Sénégal	Réserve naturelle de la Somone	WP3	2021-09-22	Espèce commune, pas en danger
Dataset-3 : Caractérisation des séquences génétiques de 4 variétés de riz pluvial	Données de séquences	séquences génétiques de 4 variétés de riz pluvial originaires de Madagascar	Sénégal		WP2	2021-09-01	Ces variétés de riz originaires de Madagascar sont issues de la collection de l'Université Cheikh Anta Diop (UCAD) de Dakar.
Dataset-4 : Recueil de connaissances traditionnelles sur des plantes ayant une efficacité anti-bactérienne	Données d'enquêtes	Les informations seront recueillies lors d'entretiens individuels et de groupes avec les guérisseurs et connaisseurs de la médecine et pharmacie traditionnelle dans différentes communautés au Burkina Faso	Burkina Faso		WP3	2021-09-22	Données qualitatives; données de savoir traditionnel



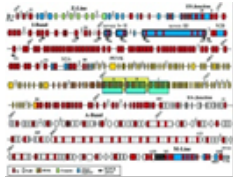
Informations	Jeux de données	Droits-Ethique	Responsabilités	Documentation	Stockage-Sauvegarde	Partage	Publications
--------------	------------------------	----------------	-----------------	---------------	---------------------	---------	--------------

Plan de Gestion des Données

Structure



Rédiger le PGD



génétiques



enquêtes



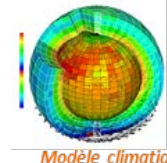
santé



climatiques



biodiversité



simulations

1. Description des données
2. Documentation des données
3. Stockage et sauvegarde
4. Exigences légales et éthiques
5. Responsabilités et ressources
6. Partage des données

Rédiger le PGD



génétiques



enquêtes



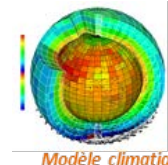
santé



climatiques



biodiversité



simulations

Description des données
Documentation des données
Stockage et sauvegarde
Exigences légales et éthiques
Responsabilités
Partage des données

Description des données
Documentation des données
Stockage et sauvegarde
Exigences légales et éthiques
Responsabilités
Partage des données

Description des données
Documentation des données
Stockage et sauvegarde
Exigences légales et éthiques
Responsabilités

Description des données
Documentation des données
Stockage et sauvegarde
Exigences légales et éthiques
Responsabilités
Partage des données

Description des données
Documentation des données
Stockage et sauvegarde
Exigences légales et éthiques
Responsabilités
Partage des données

Description des données
Documentation des données
Stockage et sauvegarde
Exigences légales et éthiques
Responsabilités
Partage des données

- Certaines parties sont spécifiques de certains types de données
- Certaines parties sont génériques et peuvent être traitées pour l'ensemble des données

Rédiger le PGD

➤ Pour être utile au projet et à vous en priorité

- Combinez textes, tableaux, schémas, liens, ...: Facilitez la lecture !
- Ajoutez illustrations, listes, annexes,...: Faciliter la compréhension
- 3 ou 30 pages....

Finding predictive Features for countries living in extreme poverty. 3 pages:
<https://zenodo.org/record/3757216#.YZKGWrrjKUI>

AfriCultuRes : 35 pages dont 2 pages d'infos administratives + intro au projet + sommaire + listes des tables, figures, abréviations + nombreux tableaux et illustrations + annexes:
http://www.africultures.eu/afbeeldingen/pdf/AfriCRS-D02-GMV-DataManagementPlan_v1.0_31_07_2018.pdf

PGD du projet INVITE : https://cordis.europa.eu/project/id/817970/results	17
ANNEX 1 – LIST OF INVITE DELIVERABLES	17
ANNEX 2 – LIST OF VCU AGRO-ENVIRONMENTAL VARIABLES REQUESTED BY WP4	20
ANNEX 3 – LIST OF VCU VARIABLES REQUESTED BY WP4 FOR WHEAT	24
ANNEX 4 – LIST OF VCU VARIABLES REQUESTED BY WP4 FOR MAIZE	25
ANNEX 5 – LIST OF VCU VARIABLES REQUESTED BY WP4 FOR SUNFLOWER	26
ANNEX 6 – PROPOSED PROCEDURE FOR ACCESS TO DATA AND MATERIAL (DEC 17TH 2019)	27

Plan de Gestion des Données

Outils et exemples



Outils de rédaction - exemples de PGD

- **DMP OPIDoR** (France) <https://dmp.opidor.fr/>



DMP publics

Modèles de DMP

Accès institutionnel

Nombreux modèles institutionnels (dont Cirad) et de bailleurs

> 150 PGD publics

- DMP Online** (UK) <https://dmponline.dcc.ac.uk/>; > 600 DMP publics



- DMP Tool** (USA) <https://dmptool.org/>; > 1000 DMP publics



- **CORDIS** (Europe) <https://cordis.europa.eu/projects/en>
> 6000 DMP publics

Horizon 2020 Advisory platform for small farms based on earth observation

Deliverables

- Documents, reports (18)
- Websites, patent filings, Videos etc. (3)
- Open Research Data Pilot (3)**

Project Information

APOLLO
Grant agreement ID: 687412
Closed project

Start date: 1 May 2018
End date: 28 February 2019

Funded under: H2020-EU 2.1.6
Overall budget: € 2 135 795,11
EU contribution: € 1 727 548,36

Coordinated by: ORNANO ENVIRONMENTAL SA
Orneco

- Archive ouverte du Cirad [Agritrop](#) : accès à qqs PGD du Cirad

Modèles Cirad de PGD

Site Intranet Data <https://intranet-data.cirad.fr/>

Data cirad
Gérer et valoriser les données scientifiques

GÉRER SES DONNÉES DROITS PRÉSERVER ET PUBLIER RÉUTILISER OUTILS ET RESSOURCES

intranet > Gérer ses données > Plan de gestion des données (PGD)

Plan de gestion des données (PGD)

Un PGD (DMP - Data Management Plan) est un outil pour aider les scientifiques à gérer leurs données dans le cadre d'un projet. Rédiger un PGD en début de projet permet de mettre en place de bonnes pratiques de gestion des données, facilite les échanges entre partenaires et assure un gain de temps pour la publication et le partage raisonné des données en fin de projet.

Un plan de gestion des données évolue dans le temps. Il est vérifié, revu et complété à différentes étapes du projet. A minima : une première version en début de projet (au mois 6 pour les projets H2ozo et ANR), une seconde version à mi-parcours et une troisième en fin de projet.

Le PGD répond aux questions :

- Quels sont les types de données qui seront générés ?
- Quels sont les jeux de données existants qui seront mobilisés ?
- Quelles sont les méthodologies qui seront utilisées ?
- Comment les données seront-elles décrites et documentées ?
- Comment les données seront-elles exploitées, partagées et conservées ?
- Quel cadre juridique et éthique s'appliquera aux données ?
- Comment les responsabilités seront-elles distribuées ?

Pourquoi rédiger un PGD ?

- Mettre en œuvre de bonnes pratiques de gestion et documentation des données
- Garantir la qualité des recherches et la production de données fiables et compréhensibles
- Contribuer à la transparence, à l'intégrité scientifique et à la reproductibilité des recherches
- Réduire le risque de perte de données ou de données non réutilisables.
- Clarifier les rôles, responsabilités et droits de chaque contributeur.
- Anticiper les problèmes juridiques, éthiques ou techniques
- Assurer la sécurité des données personnelles, sensibles ou stratégiques
- Faciliter le partage des données au sein du collectif
- Prévoir les besoins et coûts pour générer, traiter, conserver et partager les données.
- Répondre à la demande des bailleurs.

Bonnes pratiques

- Réaliser le PGD dès le début du projet
- Prévoir les jeux de données à produire
- Identifier les métadonnées adaptées
- Clarifier les responsabilités des contributeurs
- Mettre au point des bases de données et des jeux de données propres et valorisables
- Réfléchir aux modes de valorisation

Voir aussi

Les métadonnées

Modèles Cirad de PGD

Rédiger un plan de gestion des données

Formation

La minute Plan de gestion des données (DORANum)

DMP OPIDoR

Contact

data-pgd@cirad.fr

Accès aux modèles
en format Word et Excel

Accès à DMP OPIDoR

Outil de rédaction – DMP OPIDoR

DMP OPIDoR

DMPs publics

Modèles de DMP

Aide Plus ▾





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

Nombreux modèles de PGD

> 150 DMP publics

Bienvenue !

DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.

-  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

DMP OPIDoR évolue grâce à vos retours. Les développements s'inscrivent dans le cadre d'une collaboration internationale autour du logiciel open source DMPRoadmap

CIRAD

Connexion S'inscrire

Accès institutionnel

Connexion

Accès individuel

* Courriel

* Mot de passe

Mot de passe oublié ?

Se souvenir de moi

Connexion

Outil de rédaction – DMP OPIDoR

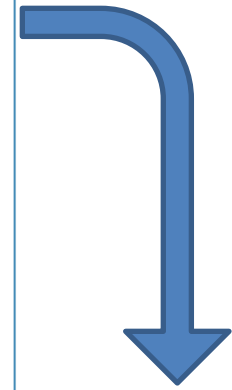
Modèle Cirad simplifié

Renseignements sur le projet Produits de recherche Modèle choisi **Rédiger** Partager Télécharger

tout développer | tout réduire

- 1. Décrire les données (1 question) +
- 2. Clarifier le cadre éthique et juridique (5 questions) +
- 3. Garantir la compréhension et l'accessibilité des données (4 questions) +
- 4. Assurer le stockage, la sécurité et la sauvegarde des données (4 questions) +
- 5. Définir les responsabilités (1 question) +
- 6. Spécifier les modalités de diffusion et de partage des données (2 questions) +

tout développer | tout réduire



1. Décrire les données (1 question)

Identifiez les types de jeux de données qui seront collectés ou générés dans votre projet

B *I* [Liste à puces] [Liste numérotée] [Liens] [Tableau]

Enregistrer

Recommandations Commentaires

Cirad

Listez les jeux de données sous forme de texte ou en utilisant un tableau Excel.

Pour les ciradiens, un exemple de tableau (annexe 1) est proposé dans le modèle simplifié sous format Word: <https://intranet-data.cirad.fr/modeles-cirad-de-pgd>.

Précisez l'objet d'étude, l'origine géographique, la période temporelle, et expliquez l'objectif de la collecte/génération de ces données.

Plan de gestion des données

A RETENIR



Conclusion

➤ PGD : livrable de projet

- 1 seul par projet; 3 versions: 6 mois, mi-projet, fin de projet
- Pour décrire tous les jeux de données produits dans 1 projet et leur gestion pendant et après le projet
- À préparer le plus tôt possible
- En nommant des responsables et impliquant les partenaires

➤ PGD : document structuré mais flexible et adaptable

- Choix de la forme la + cohérente par rapport à votre projet
- Format de rédaction libre : texte, tableaux, graphiques, annexes,....
- Enrichissez le avec tout ce qui aide sa lecture et sa compréhension

➤ Pour rédiger le PGD

- Modèles et exemples disponibles
- Outils de rédaction
- Formations et ressources utiles

Conclusion

- 1 PGD par projet → pour le bailleur, pour montrer :
 - La valeur que le consortium accorde aux données produites
 - L'application de bonnes pratiques disciplinaires, harmonisées entre partenaires
 - Le respect des lois et principes éthiques pour lever tous freins au partage et à la publication des données
 - La volonté de produire des données FAIR, si possible déposées dans un entrepôt de confiance
 - L'intention d'en faire bénéficier d'autres acteurs: scientifiques, décideurs, société civile, ...et d'accroître l'impact du projet.



Répond aux exigences du bailleur, à l'engagement du Cirad et à la politique nationale de science ouverte

Répond aux enjeux de recherche responsable, reproductible, transparente et intègre.

Conclusion

- PGD: outil d'animation de projet → pour le collectif, pour :
 - Aborder (et régler) ≠ points dès le début du projet
 - Préciser les responsabilités
 - Anticiper les contraintes éthiques et juridiques
 - Harmoniser les protocoles et les pratiques entre partenaires
 - Assurer la qualité des données et leur mise en commun
 - Clarifier les objectifs des partenaires en terme de partage des données après le projet
 - Valoriser les données en facilitant leur diffusion.

→ Outil pour faciliter la collaboration, gagner du temps et de l'efficacité pour réussir le projet et publier

Conclusion

■ PGD: un outil et un atout pour vous

- Vous appliquez des bonnes pratiques de recherche
- Vos données sont bien décrites, documentées, conservées...
- Vous êtes transparent sur vos méthodes et protocoles
- Vous savez évaluer et respecter le cadre juridique et éthique
- Vous savez comment valoriser aux mieux vos recherches
 - Déposer des données dans un entrepôt
 - Publier un article de recherche et un *Data paper*
- Collaborer avec vous est facile et efficace.



Vous avez acquis des compétences en gestion de données de recherche et en science ouverte

Vous publiez plus facilement.

MERCI POUR
VOTRE ATTENTION



Laurence Dedieu

laurence.dedieu@cirad.fr