

Modèle de Plan de Gestion de Données

Université de Montpellier

Préambule

Le Plan de Gestion des Données (PGD) ou Data Management Plan (DMP) est un document qui définit quelles données seront créées ou collectées, comment elles seront structurées, documentées, partagées et conservées pendant et après le projet. Il s'inscrit dans le cadre des démarches de science responsable, science reproductible et science ouverte.

Pour ce faire, il doit être initié dès le montage du projet et être mis à jour régulièrement. Il prend en compte toutes les étapes du cycle de vie des données et se construit collectivement (scientifiques, personnels de bibliothèque, informaticiens, juristes, chargés de projet...). Toutes les données n'ont pas vocation à être conservées au-delà du projet. Pour les données qui le seront, l'accent doit être mis sur la préparation de données faciles à trouver, accessibles, interopérables et réutilisables (c'est-à-dire respectant les principes "FAIR").

Le PGD devient un livrable obligatoire notamment dans le cadre des projets subventionnés par l'ANR et l'Union Européenne. Le présent modèle est mis à disposition de la communauté scientifique par l'Université de Montpellier.

Le PGD peut être révisé, toutes les versions feront l'objet d'un suivi et d'une conservation attentifs. Les différents champs proposés peuvent être complétés au fur et à mesure des différentes versions du PGD au cours du projet.

TABLE DES MATIÈRES

I. INFORMATIONS SUR LE PROJET	4
▪ Acronyme projet.....	4
▪ Titre complet du projet.....	4
▪ Nom de l'institution coordinatrice.....	4
▪ Nom, prénom et email du responsable scientifique / du contact PGD (insérer une ligne s'il ne s'agit pas de la même personne)	4
▪ Membres du projet et coordonnées (responsabilités, rôles, noms, prénoms, emails des différents membres du projet et ressources utilisées dans la gestion des données).....	4
▪ Dates, durée et calendrier (prévisionnel ou défini) du projet.....	4
▪ Financeur(s) du projet (développer les sigles)	4
▪ Identification de l'appel à projet.....	4
▪ Partenaires du projet.....	4
▪ Nom du programme de recherche	4
▪ Thématique de l'appel à projet (utilisez des mots-clés et précisez la nomenclature utilisée)	4
▪ Objectifs et attendus du projet.....	4
II. DESCRIPTION DES DONNÉES DU PROJET	4
2.1 Présentation des données générées et collectées.....	4
2.2 Réutilisation des données existantes.....	5
2.3 Quelles mesures de contrôle de la qualité des données seront-elles mises en œuvre ?.....	6
III. DESCRIPTION DES MÉTADONNÉES DES DONNÉES DU PROJET	6
3.1 Quelles métadonnées (ontologies, taxonomies...) et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront-elles les données ?..	6
3.2 Comment les métadonnées ont-elles été produites (GPS, type d'instrument, entrée manuelle etc.) ? Par qui ?.....	6
3.3 Les données sont-elles faciles à trouver, librement accessibles, interopérables et réutilisables (principes "FAIR") ?	6
IV. RESPONSABILITÉS, EXIGENCES ÉTHIQUES ET DROITS DE PROPRIÉTÉ INTELLECTUELLE	7
4.1 Quel établissement détiendra les droits de propriété intellectuelle pour les données créées lors du projet ?.....	7
Quelle est la législation applicable en la matière ?	7
4.2 Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?	8
4.3 Comment les questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?	8
V. STOCKAGE, PARTAGE ET RÉUTILISATION DES DONNÉES AU COURS DU PROJET	9
5.1 Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?	9

VI. DIFFUSION, PARTAGE ET RÉUTILISATION DES DONNÉES A L'ISSUE DU PROJET	10
6.1 Comment les données et les métadonnées seront-elles diffusées à l'issue du projet de recherche ?	10
6.2 Où seront stockées les données en vue de leur diffusion ?.....	11
Le dépôt des données dans un entrepôt de données de la recherche permet de les stocker à l'issue du projet pour y accéder et de les réutiliser. Il existe des milliers d'entrepôts répartis en plusieurs types : disciplinaires, multidisciplinaires, propres à un éditeur, institutionnels, spécifiques d'un projet de recherche.....	
	11
VII. ARCHIVAGE ET CONSERVATION A LONG TERME.....	11
7.1 Quelles sont les données à conserver à moyen et long terme ?.....	11
7.2 Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?.....	11
7.3 Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?.....	12
Contacts et assistance.....	13

I. INFORMATIONS SUR LE PROJET

- Version du PGD et date d'édition

Exemple de réponse : v2, éditée le 02-03-2019

- Acronyme projet
- Titre complet du projet
- Nom de l'institution coordinatrice
- Nom, prénom et email du responsable scientifique / du contact PGD (insérer une ligne s'il ne s'agit pas de la même personne)
- Membres du projet et coordonnées (responsabilités, rôles, noms, prénoms, emails des différents membres du projet et ressources utilisées dans la gestion des données)
- Dates, durée et calendrier (prévisionnel ou défini) du projet
- Financeur(s) du projet (développer les sigles)
- Identification de l'appel à projet

Exemple de réponse : AAP ACC 2017

- Référence de la convention de financement
- Partenaires du projet
- Nom du programme de recherche
- Thématique de l'appel à projet (utilisez des mots-clés et précisez la nomenclature utilisée)

Exemple de réponse : Domaine Sciences du Vivant (nomenclature HAL).

- Objectifs et attendus du projet

II. DESCRIPTION DES DONNÉES DU PROJET

2.1 Présentation des données générées et collectées

Quelles données (types, formats et volumes par exemple) seront collectées ou produites au cours du projet ?

- Nature, type, échelle
- Méthodes de production et traitement des données : collecte de nouvelles données ou réutilisation de données (conversion / transformation de données ; partage / échange de données ; achat de données)
- Volumétrie prévisionnelle
- Période de recueil des données
- Autres données et publications associées
- Contrôle qualité

Recommandations :

- *Donnez des détails sur le type de données : par exemple base de données, tableur, texte, image, audio, vidéo...*
- *Détaillez le format des données : la manière selon laquelle les données sont codées pour le stockage, généralement reflétée par l'extension du nom de fichier (par exemple pdf, xls, doc, txt, ou rdf).*
- *Mentionnez la provenance et l'origine des nouvelles données et/ou de celles réutilisées.*
- *Justifiez l'utilisation de certains formats. Par exemple, les choix d'un format peuvent être guidés par l'expertise du personnel de l'organisme, par une préférence pour les formats ouverts, par les standards de format acceptés par les entrepôts de données, par l'usage largement répandu dans une communauté de recherche ou par le logiciel ou l'équipement qui sera utilisé.*
- *Privilégiez si possible les formats standards et ouverts car ils facilitent le partage et la réutilisation à long terme des données (plusieurs catalogues fournissent des listes de ces "formats préférés" – par exemple : https://doranum.fr/wp-content/uploads/FS2_liste_indicative_formats_V1.pdf).*
- *Donnez des détails sur une prévision des volumes (qui peuvent être exprimés en espace de stockage requis (octets), et/ou en quantités d'objets, de fichiers, de lignes, et colonnes).*
- *Indiquez le ou les pays où les données ont été collectées, ainsi que la période de recueil des données.*

Exemple de réponse :

Types de données : données expérimentales, données d'observation, données d'enquêtes, données textuelles, données génomiques, échantillons, images, données audiovisuelles, modèles...

Outils de traitement des données : Excel, Stata, SPSS, Matlab v8 avec Signal Processing toolbox...

2.2 Réutilisation des données existantes

Les données utilisées dans le projet s'appuient-elles sur la réutilisation de données existantes ? Si oui, citez leur origine.

Recommandations :

- *Expliquez quelles méthodologies ou quels logiciels seront utilisés si de nouvelles données sont recueillies.*
- *Énoncez les éventuelles restrictions à la réutilisation des données préexistantes. Indiquez brièvement le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée mais écartée.*

- *Si du matériel protégé par des droits spécifiques est utilisé au cours du projet, précisez les éventuelles contraintes d'accès et/ou réutilisation.*
- *Les données déjà existantes que vous allez utiliser sont-elles protégées par une licence qui limite leur réutilisation ? Précisez si besoin les modalités de réutilisation et les moyens utilisés pour vous mettre en conformité avec ces modalités.*
- *Allez-vous combiner vos nouvelles données avec ces anciens jeux de données ? Là encore vérifiez qu'une licence de réutilisation permet ce cas de figure.*

Exemple de réponse : données statistiques FAOSTAT sur la production de sorgho, livres d'utilisation, données Land Cover Map 2000, Center for Ecology and Hydrology, Forest Research has licence agreement...

2.3 Quelles mesures de contrôle de la qualité des données seront-elles mises en œuvre ?

Recommandations :

- *Expliquez comment la qualité et la conformité de la collecte des données seront contrôlées et documentées. Il s'agit là de préciser les processus comme l'heure du relevé, la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la description des processus expérimentaux même s'ils ne sont pas à but commercial, la revue par les pairs, ou la représentation basée sur des vocabulaires contrôlés.*
- *Exploitez les données documentées dans le cahier de laboratoire, non pas dans une optique de certification (ex norme ISO-9001), mais qualitative*

6

Exemple de réponse : retranscrire les données référencées dans le cahier de laboratoire

III. DESCRIPTION DES MÉTADONNÉES DES DONNÉES DU PROJET

3.1 Quelles métadonnées (ontologies, taxonomies...) et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront-elles les données ?

3.2 Comment les métadonnées ont-elles été produites (GPS, type d'instrument, entrée manuelle etc.) ? Par qui ?

3.3 Les données sont-elles faciles à trouver, librement accessibles, interopérables et réutilisables (principes "FAIR") ?

Recommandations :

- *Indiquez quels standards et quels formats de métadonnées seront utilisés (par exemple DDI, TEI, EML, MARC, CMDI).*
- *Indiquez quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données.*

- *Utilisez les standards de métadonnées des communautés scientifiques lorsque ceux-ci existent. Utilisez-vous les standards de métadonnées pour votre domaine ?*
- *Indiquez comment les données seront organisées au cours du projet, en mentionnant par exemple les conventions de nommage, le contrôle de version et les structures des dossiers. Des données bien classées et gérées de façon cohérente seront plus faciles à retrouver, à comprendre et à réutiliser.*
- *Pensez à la documentation qui serait nécessaire pour permettre une réutilisation des données par d'autres chercheurs ou dans un but commercial. Il peut s'agir notamment de l'information sur la méthodologie utilisée pour collecter les données, sur les procédures et méthodes d'analyse utilisées, sur la définition des variables, des unités de mesure, etc.*
- *Tenez compte de la façon dont ces informations seront obtenues et enregistrées par exemple dans une base de données avec des liens vers chacun des fichiers, dans un fichier texte de type « lisez-moi », dans les en-têtes de fichiers, dans un livre de référence (« code book ») ou dans les cahiers de laboratoire.*

Exemple de réponse : les métadonnées sont spécifiées par un scientifique et s'organisent selon les champs recommandés par le standard MIAPPE fourni en OWL : General metadata, Timing and location, Biosource, Environment, Treatments, Experimental design, Sample collection, processing, management, Observed variables.

IV. RESPONSABILITÉS, EXIGENCES ÉTHIQUES ET DROITS DE PROPRIÉTÉ INTELLECTUELLE

7

Nom Prénom, institution, collectif d'appartenance, ville, pays, email du responsable de la collecte des données

Nom Prénom, institution, collectif d'appartenance, ville, pays, email du responsable de l'analyse des données

Nom Prénom, institution, collectif d'appartenance, ville, pays, email du responsable de l'exécution du plan de gestion des données

4.1 Quel établissement détiendra les droits de propriété intellectuelle pour les données créées lors du projet ?

Etablissement détenteur des droits de propriété intellectuelle

Quelle est la législation applicable en la matière ?

Recommandations :

- *Explicitez le projet collaboratif : mentionner toutes les institutions qui y prennent part, de quelle(s) manière(s)...*
- *Les conditions de propriété des données ont-elles été étudiées ?*

En tout état de cause, la propriété des données est stipulée dans l'accord de consortium dont les termes doivent être négociés en amont du projet et du PGD : donnez le lien vers l'accord de consortium si possible.

Exemple de réponse : Voir l'article 5 "Propriété intellectuelle" dans l'accord de consortium du projet XXX (LIRMM)

4.2 Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Recommandations :

Lorsque vous manipulez des données à caractère personnel, veillez à ce que la réglementation sur la protection des données (par exemple, RGPD) soient appliquées, notamment :

- Obtenir un consentement éclairé pour la préservation et/ou le partage de données personnelles.*
- Envisager l'anonymisation des données personnelles pour la préservation et/ou le partage (des données correctement anonymisées ne sont plus considérées comme des données personnelles).*
- Envisager la pseudonymisation des données personnelles (la principale différence avec l'anonymisation est que la pseudonymisation est réversible).*
- Envisager le chiffrement des données, qui est considéré comme un cas particulier de pseudonymisation (la clé de cryptage doit alors être stockée séparément des données, par exemple par un tiers de confiance).*
- Expliquer si une procédure d'accès spécifique a été mise en place pour les utilisateurs autorisés à accéder aux données personnelles.*

Le DPO de votre unité de recherche peut vous aider dans la rédaction de cette partie et répondre à vos questions.

Exemple de réponse : Données personnelles (récoltés lors d'enquêtes) anonymisées ; utilisation du logiciel PIA de la CNIL pour décrire la sensibilité des données....

4.3 Comment les questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Recommandations :

- Déterminer si les questions d'éthique auront une incidence sur la façon dont les données seront stockées et transférées, qui pourra les voir ou les utiliser et quelles durées de conservation leur seront appliquées. Démontrer que ces aspects sont bien pris en compte et planifiés.*
- Adopter les codes de conduite nationaux et internationaux et le code d'éthique institutionnel et vérifier si une revue des pratiques (par exemple par un comité*

d'éthique) est requise pour ce qui concerne la collecte de données dans le cadre du projet de recherche.

Exemple de réponse : données récoltées lors d'expérimentations animales validées par le comité d'éthique xxxx dans sa décision yyyy, recherche effectuée en partenariat avec des pays hors Europe : les recommandations du manuel H2020 ont été respectées ; respect du protocole de Nagoya (APA) ...

V. STOCKAGE, PARTAGE ET RÉUTILISATION DES DONNÉES AU COURS DU PROJET

5.1 Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Recommandations :

- Décrivez l'endroit où les données seront stockées et sauvegardées au cours du processus de recherche et la fréquence à laquelle la sauvegarde sera effectuée. Il est recommandé de stocker les données dans au moins deux lieux distincts.*
- Privilégiez l'utilisation de systèmes de stockage robustes, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution d'origine. Le stockage des données sur des ordinateurs portables, des disques durs externes, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.*
- Mettez l'accent sur l'accès aux données et les moyens mis en œuvre pour les sécuriser.*
- Évoquez les ressources (budget et temps alloués) dédiées au stockage des données*

Exemple de réponse :

- pour l'hébergement physique des données : les données seront stockées sur l'infrastructure Données du mésocentre Meso@LR avec réplication. L'environnement est hébergé au Cines en zone ZRR. Le système est nativement protégé contre les ransomwares.*
- sur la localisation géographique des données : en France, dans l'UE, hors UE, ne sait pas (cloud sans précision de stockage géographique) ...*

5.2 Comment le partage des données sera assuré entre les partenaires tout au long du processus de recherche ?

Recommandations :

- Expliquez qui aura accès aux données au cours du processus de recherche et comment l'accès aux données est contrôlé, en particulier dans le cadre de recherches menées en collaboration.*

- Mettez l'accent sur le partage des données entre les différents partenaires du projet de recherche, notamment sur l'organisation afin de fluidifier le travail collaboratif.

Exemples de réponse :

- sur le partage des données : les données collectées seront partagées avec le Groupe de recherche [...] de l'Université X... en charge de la thématique Y.
- sur les droits d'accès aux données pendant les recherches, politique de contrôle d'accès ; convention entre partenaires établie ; processus formalisé ; demande d'autorisation d'accès ponctuelle ...
- sur les droits d'accès aux données par les chercheurs partenaires : accès libre ; déclaration d'identification ; mot de passe ; certificat ; clé privée ...
- pour suivre la production et l'analyse des données au cours du projet : cahiers de laboratoire, protocoles de recherche, registres de consignation, horodatage des données et résultats, authentification personnelle sur les outils, etc.

VI. DIFFUSION, PARTAGE ET RÉUTILISATION DES DONNÉES A L'ISSUE DU PROJET

6.1 Comment les données et les métadonnées seront-elles diffusées à l'issue du projet de recherche ?

Recommandations :

- Le principe qui s'applique est le suivant : « Aussi ouvert que possible, aussi fermé que nécessaire ».
- Expliquez s'il existe des restrictions au partage des données ou des raisons de définir un embargo ?
- Prenez en compte les obligations des financeurs en matière de dissémination, publication, d'open access et open data. Les articles dédiés dans les conventions de subvention imposent des délais pour rendre les données accessibles
- Expliquez quand et comment les données pourront être retrouvées et partagées (par exemple, par le dépôt dans un entrepôt de données de confiance, l'indexation dans un catalogue, par l'utilisation d'un service de données sécurisé, par le traitement direct des demandes de données, ou l'utilisation de tout autre mécanisme).
- Indiquez qui pourra utiliser les données. S'il s'avère nécessaire de restreindre l'accès pour certaines communautés ou d'imposer un accord pour le partage de données, expliquer comment et pourquoi. Expliquez les mesures qui seront prises pour dépasser ou minimiser ces restrictions.
- Cette recherche sera-t-elle publiée dans une revue qui a besoin des données sous-jacentes pour accompagner les articles ?
- Autoriserez-vous la réutilisation, la redistribution ou la création de nouveaux outils, services, ensembles de données ou produits (dérivés) ? L'utilisation commerciale sera-t-elle autorisée ?

6.2 Où seront stockées les données en vue de leur diffusion ?

Le dépôt des données dans un entrepôt de données de la recherche permet de les stocker à l'issue du projet pour y accéder et de les réutiliser. Il existe des milliers d'entrepôts répartis en plusieurs types : disciplinaires, multidisciplinaires, propres à un éditeur, institutionnels, spécifiques d'un projet de recherche...

Recommandations

- *Un entrepôt vous est-il recommandé par l'une des parties prenantes de votre projet (votre financeur, éditeur, institution, etc.) ? Si aucun entrepôt ne vous est recommandé, il vous faut en trouver un adapté à vos besoins. Vous pouvez utiliser des annuaires d'entrepôts (re3data, OAD, OpenDOAR...) en fonction de vos critères de recherche (entrepôt disciplinaire, européen...)*
- *Le Plan national pour la science ouverte (2018) invite les chercheurs à déposer les données dans des entrepôts de données certifiés : vous pouvez consulter le site CoreTrustSeal pour en trouver.*

Exemple : En génomique, la communauté scientifique utilise notamment UniProt ou GenBank comme base de référence.

VII. ARCHIVAGE ET CONSERVATION A LONG TERME

7.1 Quelles sont les données à conserver à moyen et long terme ?

Recommandations :

- *Définissez le plan de préservation et de partage des données et fournir l'information sur la durée d'archivage pérenne des données (3-5 ans, 10 ans...). Précisez le coût de la solution retenue et les financements prévus ou à prévoir.*
- *Expliquez à quel moment les données seront rendues disponibles. Indiquez les délais de publication prévus. Expliquez si une utilisation exclusive des données est revendiquée et, dans l'affirmative, pour quelle raison et pour combien de temps. Indiquer si le partage des données sera différé ou limité, par exemple pour des raisons de publication, pour protéger la propriété intellectuelle ou le dépôt de brevets.*
- *Évoquez les ressources (budget et temps alloués) dédiées à l'archivage des données*

7.2 Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Recommandations :

- *Indiquez quelles données ne doivent pas être divulguées ou doivent être détruites pour des raisons contractuelles, légales, ou réglementaires. Indiquer comment il sera décidé quelles données garder. Décrire les données qui seront à préserver à long*

terme (notamment après la fin du projet de recherche). Décrivez les méthodes d'échantillonnage des données.

- Décrivez les utilisations (et/ou les utilisateurs) prévisibles des données dans un cadre de recherche.
- Comment allez-vous préparer les données pour la préservation ou le partage de données ? Les données devront-elles être anonymisées ou converties en formats de fichiers plus stables ?
- Indiquez où les données seront déposées. Dans un entrepôt ou un référentiel spécifique à la discipline ? Si aucun entrepôt reconnu n'est proposé, démontrer dans le plan de gestion des données que les données pourront être prises en charge efficacement au-delà de la durée de financement du projet. Il est recommandé de démontrer que les politiques des entrepôts et les procédures de dépôts (y compris les standards de métadonnées, et coûts mis en œuvre) ont été vérifiés.

7.3 Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Recommandations :

- Indiquez si les utilisateurs potentiels auront besoin d'outils ou de logiciels spécifiques pour l'accès et la (ré) utilisation des données. Si oui, seront-ils archivés ? Tenir compte de la durée de vie des logiciels nécessaires pour accéder aux données.
- Indiquez si les données seront partagées via un portail de données (portail institutionnel par exemple), si les demandes d'accès seront traitées en direct, ou si un autre mécanisme sera utilisé ?
- Indiquez si l'attribution d'un identifiant unique et pérenne (comme le DOI) sera assurée ?

Exemple : Les données stockées au sein de l'infrastructure Meso@LR seront accessibles via l'outil NextCloud du laboratoire xxx à tout utilisateur présent dans l'annuaire LDAP de la structure. L'ensemble des personnes associées au projet auront un compte. A l'issue du projet, les données seront déversées dans l'entrepôt de données ouvertes via le portail Dataverse de l'Université de Montpellier et se verront attribuer un DOI. Autant que possible, des data papers seront publiés.

Les logiciels associés au traitement des données seront versés au projet Software Heritage..

Contacts et assistance

L'équipe inter-directions de l'UM est à votre disposition pour vous accompagner dans la rédaction d'un PGD et répondre à vos questions.

Contact : donnees-recherche@umontpellier.fr

Modèle de PGD rédigé par le groupe inter-directions « Science ouverte à l'UM » (Service commun de documentation, Direction de l'innovation et des partenariats, Direction de la recherche et des études doctorales, Direction des services informatiques et numériques et Institut de Science des Données de Montpellier) en lien avec la Vice-Présidente déléguée à la Science ouverte et aux données de la recherche.