

Université de Montpellier: Université de Montpellier - modèle de PGD (français)

1. INFORMATIONS SUR LE PROJET

1.1. Membres du projet et coordonnées

1.2. Dates, durée et calendrier (prévisionnel ou défini) du projet

1.3. Thématique de l'appel à projet

Exemple de réponse:

Domaine Sciences du Vivant (nomenclature HAL).

Recommandations:

- Utiliser des mots-clés.
- Préciser la nomenclature utilisée.

2. DESCRIPTION DES DONNEES 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 ?

- 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)
- Période de recueil des données
- Autres données et publications associées

Exemple de réponse:

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

Format de données : cela correspond à 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).

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

Indiquer la volumétrie prévisionnelle.

Recommandations:

- Indiquer les logiciels et les méthodologies utilisés pour le recueil des nouvelles données.
- Décrire le type de données produites (base de données, tableau, texte, image, audio, vidéo...).
- Il existe des standards de formats ouverts pour les données (https://doranum.fr/wp-content/uploads/FS2_liste_indicative_formats_V1.pdf).

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.

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...

Recommandations:

- Préciser si les données existantes et le matériel utilisés sont protégées par des droits spécifiques ou des restrictions (contraintes d'accès et de réutilisation...).
- Indiquer si vous combinez vos nouvelles données avec des anciens jeux de données.

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

Recommandations:

- Mentionner si la qualité et la conformité de la collecte des données sont contrôlées et documentées.
- Préciser si toutes les données documentées dans le cahier de laboratoire ont été exploitées.
- Retranscrire les données référencées dans le cahier de laboratoire (heure du relevé, calibration, répétition des échantillons et des mesures, validation de la saisie des données, description des processus expérimentaux...).

3. DESCRIPTION DES METADONNEES DES DONNEES DU PROJET

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

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.

Recommandations:

- Indiquer les standards et les formats de métadonnées utilisés (DDI, TEI, EML, MARC, CMDI...).

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

Recommandations:

- Indiquer s'il existe une méthodologie spécifique pour collecter les données.
- Préciser l'organisation des données au cours du projet (convention de nommage, contrôle de version, structures des dossiers...).

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

Recommandations:

- Mentionner s'il existe une documentation pour que les données puissent être réutilisées par d'autres chercheurs ou dans un but commercial.
- Expliquer qu'elles sont les informations utiles pour reproduire les procédures et méthodes d'analyse et pour comprendre les variables ou les unités de mesure.

4. RESPONSABILITES, EXIGENCES ETHIQUES ET DROITS DE PROPRIETE INTELLECTUELLE

4.1. Quel établissement détiendra les droits de propriété intellectuelle pour les données créées lors du projet ? Quelle est la législation applicable en la matière ?

Exemple de réponse:

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

Recommandations:

- Indiquer le propriétaire des données (mention dans l'accord de consortium...).
- Décrire le projet collaboratif (quelles institutions y prennent part, de quelles manières...).

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 ?

Exemple de réponse:

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

Recommandations:

- Indiquer si la réglementation sur la protection des données est appliquée (RGPD <https://www.cnil.fr/fr/comprendre-le-rgpd>).
- Préciser si les données ont besoin d'être anonymisées ou "pseudonymisées".
- Mentionner s'il existe une procédure d'accès spécifique mise en place pour les personnes autorisées à accéder aux données personnelles récoltées.
- Le DPO de votre unité de recherche peut vous aider dans la rédaction de cette partie et répondre à vos questions.

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

Recommandations:

- Indiquer si les questions d'éthique ont une incidence sur le stockage et la conservation des données.
- Consultez les codes de conduite nationaux/internationaux et le code d'éthique institutionnel pour vérifier vos pratiques au cours de votre projet de recherche.
- Vous pouvez utiliser les services et l'expertise du [comité d'éthique](#) de l'université de Montpellier.

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) ...

5. STOCKAGE, PARTAGE ET REUTILISATION DES DONNEES AU COURS DU PROJET

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

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) ...

Recommandations:

- Indiquer où sont stockées et sauvegardées les données (il est recommandé d'avoir deux lieux distincts et d'utiliser les services de stockages fournis par le service informatique de l'institution).
- Préciser comment les données sont sauvegardées (fréquence, automatisme...).
- Mentionner si les données sont sécurisées (il n'est pas recommandé de les stocker sur le cloud, des clés UBS, des disques durs externes, des ordinateurs portables...).
- Evoquez les ressources nécessaires (budget et temps) pour le stockage des données.

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

Exemple 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.

Recommandations:

- Indiquer les personnes ayant accès aux données pendant le processus de recherche (qui en contrôle l'accès...).
- Préciser comment le partage de données est organisé entre les différents partenaires pendant le processus de recherche.

6. DIFFUSION, PARTAGE ET REUTILISATION DES DONNEES A L'ISSUE DU PROJET

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

Recommandations:

- "Aussi ouvert que possible, aussi fermé que nécessaire".
- Préciser s'il existe des restrictions au partage des données (embargo, délais de publication, pour certaines communautés...).
- Indiquer si la réutilisation, la redistribution, l'utilisation commerciale et la création de nouveaux outils, services et ensembles de données ou produits dérivés sont autorisées.

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

Déposer les données issues de la recherche dans un entrepôt de données permet de les stocker de manière pérenne pour y avoir accès ensuite et pouvoir les réutiliser. Ils sont nombreux et thématiques (disciplinaires, institutionnels, généralistes, propres à un éditeur...).

Exemple de réponse:

En génomique, la communauté scientifique utilise notamment Unitprot ou GenBank comme base de référence.

Recommandations:

- Le Plan national pour la science ouverte (2018, <https://scienceouverte.couperin.org/category/politiques/pnso/>) invite les chercheurs à déposer les données dans des entrepôts de données certifiés (Coretrustseal permet d'en trouver un, <https://www.coretrustseal.org/>)
- Indiquer quel entrepôt est choisi pour stocker les données de votre projet de recherche.
- Préciser si cet entrepôt est recommandé par l'une des parties prenantes de votre projet (financeur, institution, éditeur...).

7. ARCHIVAGE ET CONSERVATION A LONG TERME

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

Recommandations:

- Préciser qu'elles sont les données à conserver (utilisations prévisibles dans un cadre de recherche, focus sur les méthodes d'échantillonnage des données) et à détruire (raisons légales, contractuelles, réglementaires...).
- Mentionner si les données seront anonymisées ou converties en formats de fichiers plus stables (si oui, lesquels).
- Mentionner le moment où les données seront disponibles et si leur partage sera limité ou différé (brevet, délais de publication, propriété intellectuelle...).

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:

- Définir le plan de préservation et de partage des données (choix de l'entrepôt, pérennité du dépôt : 3-5ans, 10 ans...)
- Indiquer l'entrepôt choisi pour l'archivage pérenne (<https://coop-ist.cirad.fr/gerer-des-donnees/deposer-des-donnees-dans-un-entrepot/5-comment-choisir-un-entrepot-de-donnees>).
- Évoquer les ressources nécessaires (budget et temps) pour l'archivage des données.

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

Exemple de réponse:

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..

Recommandations:

- Évoquer les outils et logiciels nécessaires pour avoir accès et pouvoir réutiliser les données.
- Indiquer si les données seront partagées via un portail de données (institutionnel...). Si oui, préciser qui y aura accès et comment.
- Mentionner les identifiants uniques et pérennes (DOI) des données si elles en possèdent.