

Université de Montpellier: PGD structure EuroMov Digital Health in Motion

1. INFORMATIONS SUR LE PROJET

1.1. Membres du projet et coordonnées

2. DESCRIPTION DES DONNEES DU PROJET : COLLECTE ET/OU REUTILISATION DE DONNEES EXISTANTES

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
- Période de recueil des données

Recommandations:

- Les données produites sont des données créées, élaborées ou générées lors d'activités de recherche (observations, mesures...).
- Les données collectées sont des données déjà existantes (corpus, archives...) qui sont utilisées pour le projet.
- Indiquer les logiciels et les méthodologies utilisés pour le recueil des nouvelles données.
- Indiquer la volumétrie prévisionnelle.
- Décrire le type de données produites (tableau, texte, image, audio, vidéo...).
- Il est important de privilégier des [formats ouverts](#) pour pérenniser les données.
- Préciser l'organisation des données au cours du projet (convention de nommage, contrôle de version, structures des dossiers...).

Exemple de réponse:

Les données produites sont des données créées, élaborées ou générées lors d'activités de recherche (observations, mesures...). Toutes les données sont horodatées et recueillies selon un protocole de mesure validé par l'IRB interne de l'unité. Toutes les données numériques sont enregistrées en format texte ouvert (CSV) pour un volume prévisionnel de xx giga octets.

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 et les conditions de réutilisation définies par la ou les licence(s) (conversion / transformation de données ; partage / échange de données ; achat de données).

Recommandations:

- Préciser si les données existantes et le matériel utilisés sont protégées par des restrictions (contraintes d'accès et de réutilisation...). Référez-vous à l'accord de consortium ou à la réglementation si besoin (dipa-anr@umontpellier.fr).

2.3. Quelles mesures de contrôle de la qualité des données sont mises en œuvre ?

Exemple de réponse:

Toutes les données recueillies sont horodatées. La qualité et la conformité des données est contrôlée et documentée par le chercheur (métadonnées de contrôle qualité). Le contrôle de la qualité des données est fait par le chercheur (en particulier la qualité métrologique des appareils de mesure, le respect des protocoles de mesure, le respect des formats de données). Les données sont stockées au format texte (car c'est le plus durable et accessible) dans des fichiers de type Comma Separated Values (*.csv).

Recommandations:

- Mentionner si la [qualité et la conformité de la collecte des données sont contrôlées et documentées.](#)
- Avez-vous pensé à uniformiser et à enrichir 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...) ?
- Préciser si toutes les données documentées dans le cahier de laboratoire ont été exploitées.

3. DESCRIPTION DES METADONNEES DES DONNEES DU PROJET

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

Recommandations:

- Indiquer les standards et les formats de métadonnées utilisés. Les standards sont généralistes (Dublin Core) ou disciplinaires (DDI en SHS). Le format commun est xlm.
- Premier niveau de rédaction de métadonnées : ajouter aux données un fichier Read-me.
- Second niveau de rédaction de métadonnées : ajouter aux données un fichier xlm comportant des métadonnées retranscrites en respectant les standards existants.
- [Voici un outil](#) pour vous aider à rédiger des métadonnées en Dublin Core.
- Les métadonnées créées permettent-elles de rendre les données faciles à trouver, librement accessibles, interoperables et réutilisables ([principes "FAIR"](#)) ?
- Préciser l'organisation des données au cours du projet ([convention de nommage](#), contrôle de version, structures des dossiers...).

Exemple de réponse:

Le descriptif du protocole approuvé par l'institutionnal review board (IRB) est associé aux données (dans le répertoire DOC). Un fichier

readme.md est présent dans chaque répertoire. La structure type de stockage des données est la suivante :

RépertoireDuProjet

1. DAT (données)

- a. Readme.md (expliquant le contenu du répertoire)
- b. Fichiers et/ou répertoires contenant les données originales

2. DOC

- a. Readme.md (expliquant le contenu du répertoire)
- b. Document IRB (descriptif général de la recherche)
- c. Fichiers et/ou répertoires contenant les explications sur les données (méta données)

3.2. Comment les métadonnées ont-elles été produites (saisie manuelle, logiciel) ? Par qui ? A quel moment ?

Recommandations:

- Indiquer s'il existe une méthodologie spécifique pour collecter les données.
- Pour les communautés de recherche sans standard de métadonnées établi, préciser les informations pertinentes liée à la méthode de collecte de données.

Exemple de réponse:

Le chercheur est le responsable de la production des métadonnées.

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 ? Quel est le projet collaboratif (quelles institutions y prennent part, de quelle manière) ?

Recommandations:

- Vous pouvez présenter le projet collaboratif sous forme de tableau (institutions, rôles ou actions...).
- Est-ce qu'une loi relative à la propriété des données est mentionnée dans l'accord de consortium ? Si oui laquelle ?
- Si besoin, vous pouvez contacter la DIPA ou vous référer au logigramme du [Guide sur l'Ouverture des données](#) du MESRI.
- S'il existe, il est recommandé d'ajouter l'accord de consortium en annexe du PGD.

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:

La réglementation sur la protection des données est appliquée. Premièrement par la signature d'un formulaire de consentement éclairé dans lequel le participant est averti que toutes les données à caractère personnel resteront strictement confidentielles (consultable seulement par les chercheurs de cette expérimentation). D'autre part toutes les données qui pourront faire l'objet d'un traitement informatisé seront rendues anonymes.

Les données à caractère personnel sont stockées sur papier et archivées pendant 10 ans dans une salle sécurisée du bâtiment EuroMov. Les données publiées ne permettent en aucun cas l'identification des personnes, même par recoupement ou re-analyse des avec des méthodes futures. Les données sont rendues anonymes par un identifiant unique par volontaire. Le lien avec les informations personnelles n'est possible qu'en ayant accès aux archives papier qui ne sont jamais numérisées.

Procédure systématique d'anonymisation

Les informations nominatives sont stockées sur papier uniquement, dans un local fermé à clef.

Les données numérisées sont associées à un identifiant unique à chaque volontaire.

L'identifiant unique a la forme ci-dessous :

• 000NnPp

o 3 digits = numéro unique du participant dans la recherche

o Nn = 2 premières lettres du nom de famille

o Pp = 2 premières lettres du premier prénom

Recommandations:

- Indiquer si le [RGPD](#) est appliqué.
- Préciser si les données personnelles ont besoin d'être anonymisées ou pseudonymisées (au contraire de l'anonymisation, la pseudonymisation est réversible).
- Le DPO de votre unité de recherche ou de l'université peut vous aider dans la rédaction de cette partie et répondre à vos questions (dpo@umontpellier.fr).
- L'outil [Amnesia](#) permet d'obtenir des données anonymisées en trois étapes.
- L'outil [PIA](#) développé par la CNIL offre une analyse d'impact sur la protection des données.
- Indiquer qui peut accéder aux données et de quelle manière ?

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

Exemple de réponse:

La collecte des données est effectuée dans le respect de ce qui a été annoncé dans l'IRB d'EuroMov Digital Health in Motion validé par le comité d'éthique de la recherche de l'université de Montpellier.

Rappel : l'IRB est une procédure qualité interne : toute recherche dans EuroMov doit être évaluée par le comité IRB interne. Avec la création du comité d'éthique de la recherche de l'université de Montpellier, toutes les recherches dans EuroMov sont maintenant évaluées par ce comité (en plus et/ou à la place de l'IRB).

Recommandations:

- Le stockage et la conservation des données sont-ils conformes aux règles d'éthique et de déontologie ? Détaillez.

- Consultez les codes de conduite nationaux/internationaux et les comités d'éthique de la recherche institutionnels pour vérifier vos pratiques en amont et au cours de votre projet de recherche.
- Certains projets requièrent des avis de comité d'éthique préalables, par exemple :
 - o Comités de protection des personnes pour la Loi Jardé - 5 mars 2012.
 - o Comités d'éthique en expérimentation animale (CEEA) pour l'utilisation des animaux à des fins expérimentales.
- Certains protocoles de recherche requièrent des déclarations, autorisations ou agréments préalables par exemple :
 - o Protocole de Nagoya sur l'accès aux ressources génétiques et le partage juste et équitable des avantages découlant de leur utilisation à la Convention sur la diversité biologique.
 - o Importation ou exportation d'échantillons biologiques humains destinés à la recherche.
- Certains protocoles de recherche relevant de tous les domaines scientifiques et en lien, notamment, avec des personnes humaines, des éléments de la vie privée et des données relatifs à des personnes humaines, hors réglementation spécifique, peuvent bénéficier d'un avis éthique consultatif du [comité d'éthique de la recherche](#) de l'université de Montpellier (recommandé mais non obligatoire).

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

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

Exemple de réponse:

Pour l'hébergement physique des données : elles seront sauvegardées sur des serveurs différents : l'ordinateur du chercheur, la sauvegarde de l'ordinateur du chercheur, le serveur de données EuroMov. Toutes les données numériques sont sauvegardées en format texte (CSV) et archivées dans le répertoire dédié dont l'accès est restreint par mot de passe avec validation en deux étapes.

Sur la localisation géographique : les données seront stockées et sauvegardées en France, à Montpellier (34000).

Recommandations:

Au moins 3 copies, dont une hors du site.

- Ordinateur du chercheur
- Sauvegarde de l'ordinateur du chercheur
- Serveur de données EuroMov (euromov.fr)

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

Exemple de réponse:

Toutes les données numériques sont enregistrées en format texte (CSV) et archivées dans le répertoire dédié dont l'accès est restreint par mot de passe avec validation en deux étapes (serveurs d'EuroMov). Seuls les chercheurs du projet ont l'autorisation d'accès à ce répertoire et uniquement pendant la durée nécessaire.

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

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

Exemple de réponse:

Dans un objectif de publication, toutes les données et les métadonnées ainsi que les codes de calculs seront diffusés en OpenAccess sur osf.io. Tout le monde pourra utiliser les données et les codes de calcul. De plus, nous autorisons toute réutilisation des données selon la licence Creative Commons CC4 : <https://choosealicense.com/licenses/cc-by-4.0/> et la réutilisation des codes de calcul selon la licence GNU GPL-V3 <https://choosealicense.com/licenses/gpl-3.0/>

Recommandations:

- La maxime concernant l'ouverture des données de la recherche est **"Aussi ouvert que possible, aussi fermé que nécessaire"**.
- Quels sont les critères de sélection des données et des métadonnées à conserver et à diffuser, à l'issue du projet ?
- Existe-t-il des restrictions au partage des données (autorisations de recherche spécifiques, embargo, délais de publication, pour certaines communautés...) ?
- A noter qu'il est possible de préciser l'embargo au moment du dépôt des données dans un entrepôt.
- Mentionner si les données personnelles seront anonymisées ou converties en formats de fichiers plus stables (si oui, lesquels).
- Il existe [différentes licences](#) pour les données et les codes source de logiciel. Quelle licence a été choisie (Creative Commons, Etalab ou Open Database Licence) et pourquoi ?
- Évoquer les outils et logiciels nécessaires pour avoir accès et pouvoir réutiliser les données.

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

Déposer les données issues de la recherche dans un entrepôt de données certifié permet de les stocker pour y avoir accès ensuite et pouvoir les réutiliser. Ils sont nombreux et peuvent être disciplinaires, institutionnels, généralistes, propres à un éditeur...

Recommandations:

- Indiquer quel est l'entrepôt choisi pour stocker les données et métadonnées de votre projet de recherche.
- Il est recommandé de déposer les données et les métadonnées associées à une publication dans des entrepôts.
- Quelques exemples d'entrepôts généralistes FAIR : [Zenodo](#), [OSF](#), Recherche Data Gov.
- Il est recommandé d'avoir un [identifiant unique et pérenne](#) (DOI ou autre PID) pour chaque jeux de données.

Exemple de réponse:

Les données et les métadonnées seront stockées sur la plateforme [Osf.io](http://osf.io) pendant 30 ans. Un DOI leur sera ainsi attribué.

A l'issue du projet, les données et les métadonnées associées seront déversées dans l'entrepôt X. Un DOI leur sera ainsi attribué. Les logiciels associés au traitement des données seront versés au projet Software Heritage. Autant que possible, des data papers seront publiés.

7. ARCHIVAGE ET CONSERVATION A LONG TERME

7.1. Comment les données et les métadonnées seront-elles sélectionnées et conservées à long terme ?

Exemple de réponse:

Les données sont rendues publiques sur OSF.io avec un stockage en Europe pour 30 ans.

Recommandations:

- Préciser quelles sont les données à conserver (utilisations prévisibles dans un cadre de recherche, focus sur les méthodes d'échantillonnage des données) et à détruire (critères réglementaires, raisons légales, contractuelles, volumétrie...).
- Définir le plan de préservation et de partage des données.
- Indiquer qui va payer pour la solution d'archivage sélectionnée et avec quelles preuves d'engagement.
- Actuellement en France, seul le [CINES](#) propose un archivage pérenne des données et des métadonnées (le coût est important).
- Pour l'archivage des logiciels, Software Heritage est une solution pérenne et fiable.