Université de Montpellier: Université de Montpellier - modèle de PGD (français) 2

I. DESCRIPTION DES DONNÉES PRODUITES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES

1.1. Quelles données seront collectées ou produites au cours du projet ? Quelles données seront réutilisées ?

Exemple de réponse:

Exemple de citation de données réutilisées : Grabcad , 2021, base de données en ligne, NIrma : Aadarsh S Chandran, <https://grabcad.com/library/knee-joint-model-1>, consulté le 7/04/2022

Recommandations:

- Détailler le type de données : numérique (bases de données, tableurs), textuel (documents), image, audio, vidéo, médias composites, etc.
- Indiquer le format des données : .pdf, .xls, .doc, .txt, .rtf, etc.
- Privilégier les formats standards et ouverts car ils facilitent le partage et la réutilisation des données. Pour en savoir + sur les formats
- Indiquer la volumétrie des données du projet (qui peut être exprimée en espace de stockage requis (octets), et/ou en quantité d'objets, de fichiers, de lignes, et colonnes) : une estimation dans la version initiale qui sera à préciser dans les versions ultérieures.

1. 2. Comment les données seront-elles produites ou collectées ?

Recommandations:

- Indiquer les logiciels et les méthodologies utilisés pour le recueil des nouvelles données.
- Préciser la date de création ou de collecte des données.
- Préciser si les données existantes ou le matériel utilisé sont protégées par des droits spécifiques ou des restrictions (contraintes d'accès et de réutilisation...).

II. DOCUMENTATION ET QUALITÉ DES DONNÉES

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

Recommandations:

- La documentation accompagnant les données apporte aux utilisateurs les informations nécessaires à leur bon usage et leur bonne interprétation. A minima, un fichier de type "lisez-moi" (ou readme) peut être rédigé pour rassembler les informations de base sur les données (nom de la source, format du fichier, identifiant, description du contenu...).
- Exemple de fichier readme
- Utiliser les standards de métadonnées des communautés scientifiques lorsque ceux-ci existent. Outils pour trouver un standard : https://www.dcc.ac.uk/guidance/standards/metadata et https://rdamsc.bath.ac.uk/
- Présenter le vocabulaire contrôlé choisi pour l'indexation. Si aucun n'est choisi, expliquer pourquoi (absence de solution adaptée pour les données traitées, etc.) et décrire la solution adoptée le cas échéant. Outils pour trouver un vocabulaire contrôlé : https://www.lexicool.com/dictionnaires-en-ligne-par-sujet.asp et https://skosmos.loterre.fr/fr/ Indiquer 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.
- Pour en savoir + sur le nommage des fichiers

2.2. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Recommandations:

- Expliquer comment la qualité et la conformité de la collecte des données seront contrôlées et documentées. Il s'agit là de préciser les processus comme la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la revue par les pairs, ou la représentation basée sur des vocabulaires contrôlés.
- Mentionner les éventuelles certifications qualité.

III. EXIGENCES LÉGALES ET ÉTHIQUES

3.1. Qui détiendra les droits de propriété intellectuelle sur les données du projet ?

Exemple de réponse:

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

Recommandations:

- Les données de la recherche se situent dans le cadre général d'ouverture par défaut des données publiques. Toutefois, leur diffusion peut être soumise à conditions, notamment dans les cas des données personnelles et sensibles, des données protégées par le droit d'auteur ou encore en cas de contrat avec un tiers.
- Déterminer qui sont les titulaires des droits sur les données et de quels droits ils disposent. https://opendatacanvas.org/pgd-etape-1-acquisition-1-5
- Pour toute question sur les accords de consortium, contacter la DIPA

3.2. Si des données à caractère personnel sont traitées, comment le respect de la législation sur les données à caractère personnel 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 pseudonymisées.

Les données à caractère personnel sont stockées sur papier et archivées pendant 10 ans dans une salle sécurisée du bâtiment XXX. Les données sont pseudonymisées 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.

Les données publiées ne permettent en aucun cas l'identification des personnes, même par recoupement ou re-analyse des données avec des méthodes futures.

Procédure systématique de pseudonymisation

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:

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

- Obtenir un consentement éclairé pour la conservation 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). https://datascience.etalab.studio/pseudo/

Pour toute question sur la gestion des données personnelles, contacter le Délégué à la protection des données (DPO)

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

Exemple de réponse:

Les données récoltées lors d'expérimentations animales ont été validées par le comité d'éthique de la recherche de l'université de Montpellier dans sa décision yyyy. Le projet de recherche regroupe des partenariats avec des pays hors Europe. Le protocole de Nagoya (APA) est respecté tout au long du projet de recherche…

Recommandations:

- Consultez les codes de conduite nationaux et internationaux et le code d'éthique institutionnel. 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.
- Certains projets requièrent des avis de comité d'éthique préalables, par exemple :
- Comités de protection des personnes pour la Loi Jardé - 5 mars 2012.
- 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 :
- 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.
- Importation ou exportation d'échantillons biologiques humains destinés à la recherche.
- Contact comité d'éthique de la recherche de l'UM

IV. STOCKAGE ET SAUVEGARDE DES DONNÉES

4.1. Comment les données et les métadonnées seront-elles stockées et sécurisées tout au long du processus de recherche ? Comment seront-elles partagées entre les partenaires du projet ?

Exemple de réponse:

Les données recueillies sont stockées sur l'infrastructure Meso@LR avec réplication. L'environnement est hébergé au CINES, en Zone à Régime Restrictif (contrôle d'accès).

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 de la structure. L'ensemble des personnes associées au projet auront un compte.

Recommandations:

- Décrire 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.

- Privilégier 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, des périphériques de stockage tels que des clés USB ou des clouds commerciaux (Google drive, Dropbox, etc) n'est pas recommandé.
- Solution institutionnelle de stockage et calcul intensif : meso-lr@umontpellier.fr
- Expliquer comment les données seront récupérées en cas d'incident.
- Expliquer qui aura accès aux données au cours du processus de recherche et comment l'accès aux données sera contrôlé, en particulier dans le cadre de recherches menées en collaboration.
- 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.
- Expliquer si une procédure d'accès spécifique a été mise en place pour les utilisateurs autorisés à accéder aux données personnelles.

V. DIFFUSION, PARTAGE ET RÉUTILISATION DES DONNÉES

5.1. Quelles données seront partagées ? Sur quels critères de sélection ?

Recommandations:

- Quels sont les critères de sélection des données (et des métadonnées) à conserver et à diffuser ?
- Pendant la durée du projet, puis entre 5 et 10 ans après la fin de celui-ci, les données à conserver sont notamment celles permettant la validation des résultats et celles ayant une valeur de preuve ou un potentiel de réutilisation.
- Indiquer quelles données ne doivent pas être divulguées ou doivent être détruites pour des raisons contractuelles, légales, ou réglementaires.

5.2. Comment les données seront-elles partagées ?

Exemple de réponse:

A l'issue du projet, les données et les métadonnées associées seront déposées dans l'espace institutionnel UM de la plateforme recherche data gouv. Un DOI leur sera ainsi attribué, ainsi qu'une licence de diffusion (CC-0).

Recommandations:

- Expliquer où et dans quelles conditions les données seront partagées.
- Il est recommandé de déposer les données sélectionnées dans un entrepôt disciplinaire certifié ou de confiance. Pour trouver un entrepôt
- En l'absence de solution disciplinaire, il est possible de déposer les données dans l'espace institutionnel de l'UM au sein de l'entrepôt national multidisciplinaire recherche data gouv.
- Accorder une licence claire et explicite aux données, même si elles sont en accès restreint. Une liste de licences pouvant être utilisées par les administrations publiques en France a été fixée par décret. Il est également possible d'utiliser une licence CC-BY ou CC-BY-SA parmi les Creative Commons.
- L'utilisation d'un identifiant pérenne est fortement recommandée afin d'améliorer la visibilité et l'accessibilité de votre jeu de données. Il facilite la citation du jeu de données et permet de le lier à une publication ou à tout autre produit de recherche. Le dépôt dans un entrepôt de données donne souvent lieu à l'attribution automatique d'un identifiant pérenne, notamment d'un DOI, mais il convient de s'en assurer.
- Si une période d'exclusivité d'accès aux données est nécessaire (par exemple pour la valorisation scientifique des données), indiquer sa durée (notion d'embargo : le producteur des données dépose ses données dans un entrepôt public mais celles-ci ne seront accessibles qu'après une période définie).
- Définissez dès le départ les niveaux d'accès à vos données (public, restreint, embargo, fermé, ou une combinaison).
- Indiquer si les utilisateurs potentiels auront besoin d'outils spécifiques pour l'accès et la (ré)utilisation des données.

VI. ARCHIVAGE ET PÉRENNISATION DES DONNÉES

6.1. Si certaines données sont concernées, comment seront-elles sélectionnées et archivées ?

Recommandations:

- Les questions suivantes peuvent être posées afin de sélectionner les données :
- Sont-elles probantes ou révélatrices de l'apport scientifique ? Les données sont-elles reproductibles ? A quel coût ? Pourront-elles faire l'objet de traitements complémentaires ? Peuvent-elles être exploitées ultérieurement pour d'autres usages ? Quelles seraient les conséquences d'une perte des données ?
- Dans le cadre des données personnelles, en cas de traitements de données complémentaires ou en cas de traitements ayant des finalités différentes que celles pour lesquels les données ont été collectées il convient d'en informer les personnes concernées.
- Pour préserver ses données et en assurer la lisibilité sur le très long terme, un archivage pérenne peut être envisagé en plus du dépôt dans un entrepôt, mais il s'agit d'une prestation payante. Le principal service d'archivage français est le CINES.
- Le CINES a développé un service de validation de formats pour vérifier que les formats utilisés sont pérennes et pourront être archivés.
- Pour les logiciels et codes sources, le service de validation de format pour la conservation des données recommandé est Software Heritage.
- Pour en savoir + sur l'archivage

Exemple de réponse:

Exemple de logiciel archivé dans Software Heritage et signalé dans HAL