Inserm - Institut national de la santé et de la recherche médicale: Inserm - Modèle de PGD (français) - Informations générales sur les données

1. RENSEIGNEMENTS SUR LES RESSOURCES ET LES CONTRIBUTEURS

1.1. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ? *****

Recommandations:

● Expliquer comment les ressources nécessaires (par exemple le temps) à la préparation des données pour le partage/préservation (curation des données) ont été chiffrées. Examiner et justifier soigneusement toutes les ressources nécessaires pour diffuser les données.
● Il peut s'agir de frais de stockage, de coût matériel, de temps de personnel, de coûts de préparation des données pour le dépôt, de frais d'entrepôt et d'archivage. Les coûts liés à la gestion des données concernent principalement les ressources matérielles (serveur de stockage, logiciel d'analyse, stockage dans un entrepôt...) et les ressources humaines (embauche d'un data manager par exemple). Au début du projet, indiquez le budget prévisionnel. S'il vous est inconnu, vous pourrez répondre à cette question en fin de projet.
● Indiquer si des ressources supplémentaires sont nécessaires pour préparer les données en vue de leur dépôt ou pour payer tous les frais demandés par les entrepôts de données. Si oui, précisez le montant et comment ces coûts seront couverts.

Exemple de réponse:

Un budget de XXX euros est prévu pour le stockage des données dans un entrepôt de données ouvert. Les coûts seront financés par la Commission Européenne : "costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions)"

1.2. Contributeurs de la gestion des données du Projet *

Recommandations:

● Décrire les rôles et les responsabilités concernant les activités de gestion des données, par exemple : saisie des données, production des métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données. Nommer la(es) personne(s) responsable(s) impliquée(s) dans la mesure du possible.
● Pour les projets menés en collaboration, expliquer comment s'effectue la coordination des responsabilités de gestion des données entre partenaires.
● Indiquer qui est responsable de la mise en œuvre du PGD, et qui s'assure qu'il est examiné et, si nécessaire, révisé.
● Envisager des mises à jour régulières du PGD.

1.3. Quelle est la tutelle principale du projet ?

2. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

2.1. Quel est l'objectif de la collecte/génération des données ? *****

Recommandations:

Expliquez le lien entre les données générées/collectées et les objectifs du projet

Exemple de réponse:

Nous allons réaliser des expériences de séquençage de l'ARN de cellules tumorales issues de patients atteint colorectaux afin de déterminer les mécanismes moléculaires responsables de la transformation de ces cancers…

2.2. Combien de jeu(x) de données/produits de recherche allez-vous générer durant ce projet ? (Dans l'onglet "Fiche spécifique par jeu de données" vous devrez faire la description spécifique de chaque jeu de donnée / produit de recherche) ***** *(la réponse peut varier au cours du projet)*

Recommandations:

• Les produits de recherche sont les jeux de données mais aussi les logiciels, workflow et protocoles produits lors du projet
• Un jeu de données est un ensemble de données, brutes ou dérivées, rassemblées pour former un ensemble cohérent. On peut considérer que des données qui sont gérées de la même façon (même méthode de traitement, de stockage, de partage...) forment à un seul jeu de données. Le nombre de jeux de données peut évoluer au cours du projet.

2.3. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ? Si réutilisées, précisez leur origine. ***** *(la réponse peut varier au cours du projet)*

Recommandations:

● Avant de générer des données, il est recommandé de vérifier s'il n'est pas possible de réutiliser des jeux de données produits par d'autres scientifiques.
● Expliquer quelles méthodologies ou quels logiciels seront utilisés si de nouvelles données sont recueillies ou produites.
● Énoncer les éventuelles restrictions à la réutilisation des données préexistantes. Avant de réutiliser des données, vous devez vous assurer que vous avez le droit de les réutiliser :
- En vérifiant qu'elles ne sont pas protégées par une réglementation nationale ou internationale, un copyright ou un droit de propriété intellectuelle
- En vérifiant que les personnes sont informées de la réutilisation de leurs données personnelles. En cas de doute, vous devez vous rapprocher de XXX (définir les services pouvant être contactés)
● Expliquer comment la provenance des données sera documentée.
● Indiquer brièvement le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée mais écartée.

2.4. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ? * *(la réponse peut varier au cours du projet)*

Recommandations:

● Donner des détails sur le type de données : par exemple numérique (bases de données, tableurs), textuel (documents), image, audio, vidéo, et/ou médias composites.
● Détailler 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).
● Justifier l'utilisation de certains formats. Par exemple, les choix d'un format peuvent être guidés par l'expertise du personnel de l'organisme, ou 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.
● Privilégier 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").
● Donner des détails sur les 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 volume prévisionnel des données générées pour ce projet. Le volume de données générées peut évoluer au cours du projet, donner une estimation globale et mentionner si cette volumétrie peut être amenée à changer.

Exemple de réponse:

- Images de microscopie confocale aux formats .tiff et .PNG
- Feuilles de calcul au format Excel et documents texte en .docx.
- Données de cytométrie en flux en format .FCS.
- Profils de chromatogrammes HPLC de purification Akta au format .csv.
- Images de Western-Blot au format .jpeg et .PNG
- Vidéos de microscopie en format .AVI
- Données épidémiologiques stockées sous forme de base de donnée REDCap

3. DOCUMENTATION ET QUALITE DES DONNEES

3.1. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ? * *(la réponse peut varier au cours du projet)*

Recommandations:

● Indiquer quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données.
● Indiquer quels standards de métadonnées seront utilisés (par exemple DDI, TEI, EML, MARC, CMDI).
● Utiliser les standards de métadonnées des communautés scientifiques lorsque ceux-ci existent.
Exemple de réponse :
Les métadonnées sont basées sur les métadonnées de XXX : titre, auteur, date, contributeur, description, mots-clés, format, type de ressources, etc.
● 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. Préciser si un plan de classement spécifique des données a été mis en place.
● Penser à la documentation qui serait nécessaire pour permettre une réutilisation des données. 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.
● Tenir 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:

Chaque unité a créé un plan de classement au début du projet. Ce plan de classement est organisé par méthode de collecte des données (microscopie, phénotypage, séquençage...) puis de façon chronologique. Les données brutes et les données traitées sont stockées dans des dossiers différents. Ci-dessous un aperçu du plan de classement :
I. Méthode de collecte 1 (ex : microscopie)
I.1 Date de la première collecte (ex : 2021-01-12)
I.1.1. Étape de traitement 1 (ex : quality check)
I.1.2. Étape de traitement 2 (ex : données brutes)
I.1.3. Étape de traitement 3 (ex : données analysées)
I.2. Date de la deuxième collecte (ex : 2021-01-19)
II. Méthode de collecte 2 (ex : séquençage)

3.2. 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.

Exemple de réponse:

Pour garantir la qualité des données, différentes mesures ont été mises en place :
- Répétition des expériences de manière indépendante (minimum de trois répétitions sur trois jours différents)
- Standardisation de la collecte des données (tous les animaux élevés dans les mêmes conditions, contrôle de la température, mêmes conditions de stimulation)
- Revue régulière des données avec le PI

4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE

4.1. 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 ? *****

Si vous répondez "oui" à l'une des 2 questions ci-dessous, alors votre projet comporte des données à caractère personnel :
- Est-ce que tout ou partie des données permettent d'identifier une personne humaine ?
(Ex : nom, photo, adresse...)
- Est-ce que tout ou partie des données permettraient d'identifier une personne humaine si on les associait avec d'autres informations détenues par vous ou un tiers ?
(Ex : un numéro d'identifiant, un code avec une table de correspondance détenue par un tiers, des données de localisation, une caractéristique physique, physiologique ou génétique particulière, des éléments propres à une situation psychique, économique, culturelle ou sociale....)

Recommandations:

Lorsque vous manipulez 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 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.

Si votre projet comporte des données à caractère personnel, vous devez contacter le Délégué à la Protection des Données

4.2. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Recommandations:

● Expliquer qui sera le propriétaire des données, c'est-à-dire qui aura le droit d'en contrôler l'accès :
● Expliquer quelles conditions d'accès s'appliqueront aux données. Les données seront-elles librement accessibles, ou des restrictions seront-elles appliquées ? Si oui, lesquelles ? Envisager l'utilisation de licences concernant l'accès et la réutilisation des données.
● S'assurer de couvrir, dans l'accord de consortium, ces questions de droits de contrôle d'accès aux données pour les projets multipartenaires et en cas de propriété partagée des données.
● Indiquer si les droits de propriété intellectuelle (par exemple la directive bases de données, droits sui generis) sont affectés. Dans l'affirmative, expliquer lesquels et comment cela sera traité.
● Indiquer s'il y a des restrictions sur la réutilisation des données fournies par des tiers.

Consulter Inserm Transfer sur ces questions

4.3. Comment les éventuelles 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.

Consulter référent éthique/intégrité scientifique Inserm sur ces questions (Ghislaine Filliatreau)

5. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

5.1. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ? * *(la réponse peut varier au cours du projet)*

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. Indiquez si vos données sont stockées :
- Sur votre ordinateur,
- Sur un serveur propre à de l'unité,
(- Sur un espace de stockage partagé fourni par la DSI)
Attention, vous ne devez surtout pas stocker vos données sur un espace de stockage sur internet (Dropbox, Google drive, one drive, etc.) car ces espaces ne sont pas sécurisés.
● 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, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.

5.2. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ? *****

Recommandations:

● 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 est contrôlé, en particulier dans le cadre de recherches menées en collaboration.
● Tenir compte de la protection des données, en particulier si vos données sont sensibles (par exemple données à caractère personnel, politiquement sensibles des informations ou secrets commerciaux). Décrire les principaux risques et la façon dont ils seront gérés.
● Expliquer quelle politique institutionnelle de protection des données est mise en œuvre. **Contacter le délégué à la protection des données**

6. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

6.1. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ? Quelle est l'utilité des données ? *****

Recommandations:

● Expliquer 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).
● Définir le plan de préservation des données et fournir l'information sur la durée d'archivage pérenne des données. Des contraintes réglementaires existent essentiellement dans le cas de recherche sur la personne humaine ou utilisant des données de santé. Dans ce cas, les durées de conservation des données doivent être définies dès la définition du protocole.
● Expliquer à quel moment les données seront rendues disponibles. Indiquer les délais de publication prévus. Expliquer 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.
● Indiquer 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. Expliquer les mesures qui seront prises pour dépasser ou minimiser ces restrictions.
● Indiquer l'utilité des données, à qui elles peuvent éventuellement servir

6.2. Comment sera effectué le choix des données à conserver et sur quelle durée ? *(la réponse peut varier au cours du projet)*

Recommandations:

a) Choix :

- Indiquer 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.

b) Durée :

- S'il existe une durée légale de conservation des données, citez la réglementation applicable. Si vous considérez que les données doivent être conservées pendant une durée plus longue que la durée légale, justifiez-le.
  S'il n'y a pas de réglementation mais que vous pensez que vos données ont une valeur à long-terme, indiquez pendant combien de temps les données seront conservées.

c) Décrire les utilisations (et/ou les utilisateurs) prévisibles des données dans un cadre de recherche.

Exemple de réponse:

a) Choix :

Ex : Les jeux de données 1 et 2 doivent être conservées en raison des difficultés de reproductibilité et du temps nécessaire à leur régénération. Leur conservation est essentielle pour assurer la reproductibilité des résultats présentés dans les publications et pour pouvoir les comparer aux données qui seront générées ultérieurement.

b) Durée :

Ex 1 : Les jeux de donnés seront conservés pendant la durée maximale permise par l'entrepôt. Pour XXX, cela correspond à la durée de vie du laboratoire hôte, qui dispose actuellement d'un programme expérimental défini pour les 20 prochaines années au moins.

Ex 2 : Les données seront conservées pour une durée illimitée tant que l'espace alloué au sein de l'Inserm est disponible.

Ex 3 : Le jeu de données 4 inclut des données à caractère personnel. Il sera supprimé après la publication du dernier article lié à ce projet.

Ex 4 : les données brutes de séquençage (jeu de données 3) seront supprimées après leur dépôt sur GEO ou EBI, afin de gagner de l'espace de stockage.

6.3. Où les données seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ? *(la réponse peut varier au cours du projet)*

Recommandations:

● Indiquer où les données seront déposées. Sur quelle plateforme ou dans quel entrepôt les jeux de données à conserver seront-ils archivés ? Cette plateforme est-elle certifiée pour une conservation et une gestion à long terme ? Cette plateforme est-elle certifiée pour une conservation de données de santé ?

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.

6.4. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ? Précisez les formats choisis pour l'archivage *(la réponse peut varier au cours du projet)*

Recommandations:

● Indiquer si les utilisateurs potentiels auront besoin d'outils spécifiques pour l'accès et la (ré)utilisation des données. Tenir compte de la durée de vie des logiciels nécessaires pour accéder aux données.
● Indiquer si les données seront partagées via un entrepôt, si les demandes d'accès seront traitées en direct, ou si un autre mécanisme sera utilisé.
● Choisissez un format ouvert et stable dans le temps si possible. Evitez les formats propriétaires ou les formats qui dépendent de l'environnement technologique.

6.5. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ? *(la réponse peut varier au cours du projet)*

Recommandations:

● Expliquer comment les données pourraient être réutilisées dans d'autres contextes. Les identifiants pérennes devraient être appliqués de manière à ce que les données puissent être localisées et référencées de façon fiable et efficace. Les identifiants pérennes aident aussi à comptabiliser les citations et les réutilisations.
● Indiquer s'il sera envisagé d'attribuer aux données un identifiant pérenne. Typiquement, un entrepôt pérenne de confiance attribuera des identifiants pérennes. Exemples d'identifiants pérennes : Handle, DOI (Digital Object Identifier), Ark... Sinon, indiquez l'URL permettant d'accéder au jeu de données.

Inserm - Institut national de la santé et de la recherche médicale: Inserm - Modèle de PGD (français) - Fiche spécifique par jeu de données

1. DESCRIPTION DES DONNEES

Description des données

2. RENDRE LES DONNEES LIBREMENT ACCESSIBLES

Ce jeu de données sera-t-il librement disponible ? *

- Non applicable
- Non
- Oui

Localisation/entrepôt de données choisi pour stocker et rendre accessible ce jeu de données ?

Ce jeu de données fera-t-il l'objet d'une demande de brevet ?

Possibilités et mode d'accès au jeu de données au cas où des restrictions s'appliquent.

Quels logiciels sont nécessaires pour visualiser ou accéder aux données ? Fournissez-vous une documentation ou le code open source du logiciel ? *

Précisez la durée pendant laquelle le jeu de données sera accessible. *

3. RENDRE LES DONNEES TROUVABLES

Ce jeu de données est-il identifié par un identifiant pérenne unique tel que le DOI (Digital Object Identifier) ? Sinon, décrivez comment ce jeu de données et les données sont identifiés. *

Quels standards de métadonnées utilisez-vous ? Si vous n'utilisez pas de standard de métadonnées, précisez quel(s) type(s) de métadonnées seront créés et comment. *

Fournissez-vous une documentation complémentaire pour décrire plus précisément les données ?

4. RENDRE LES DONNEES INTEROPERABLES

Les données de ce jeu de données sont-elles interopérables du point de vue technique ?

Sinon, quelles méthodologies appliquerez-vous pour rendre vos données interopérables ?

Précisez si vous utilisez du vocabulaire standard dans votre jeu de données pour permettre une interopérabilité sémantique interdisciplinaire. Dans le cas contraire, fournirez-vous un alignement avec les ontologies les plus fréquemment utilisées ?

5. RENDRE LES DONNEES REUTILISABLES

A la fin du projet, ce jeu de données pourra-t-il être réutilisé par des parties tierces ? Si la réutilisation est restreinte, expliquez pourquoi (Brevet, données humaines...). *

Quelle licence sera attribuée à ce jeu de données afin de permettre la réutilisation la plus large possible ?

Recommandations:

Exemples de licence : https://creativecommons.org/licenses/?lang=fr-FR

A quelle date le jeu de données sera-t-il accessible pour la réutilisation ? Le cas échéant, précisez pour quelle raison et pendant combien de temps un embargo est nécessaire.

Précisez la durée pendant laquelle le jeu de données sera réutilisable. *

6. SECURITE DES DONNEES

Ce jeu de données doit-il rester confidentiel au cours de votre projet ? Si oui, pouvez-vous préciser auprès de qui il peut être diffusé ? *

Pendant le projet ce jeu de données est-il stocké de façon sécurisée ? *

L'entrepôt de données choisi pour conserver ce jeu de données après le projet met-il en œuvre une politique de sécurité pour son système d'information ?

Quelles sont les mesures de sécurité mises en œuvre pour la collecte et l'échange de données ?