

# Institut Pasteur: Institut Pasteur - Trame de DMP (FR) - Informations générales

## 1. Informations sur le Plan de Gestion de Données (PGD)

Auteur(s) du PGD

*Exemple de réponse:*

Nom, prénom, e-mail(s)

Date de la première version du PGD

*Exemple de réponse:*

1er février 2017

Version actuelle du PGD et date

*Exemple de réponse:*

V2, 21 juillet 2017

Lieu de conservation du PGD

*Exemple de réponse:*

Les versions initiales et intermédiaires seront stockées sur un serveur interne et mises en ligne sur le site du projet. La version finale sera déposée sur l'entrepôt européen ZENODO et sur le site du projet.

## 2. Informations sur le projet

Coordinateur du projet

*Exemple de réponse:*

Nom, prénom, e-mail

Institution et unité de rattachement du coordinateur du projet

*Exemple de réponse:*

Institut Pasteur, Unité Interactions Bactéries-Cellules

Date de début du projet

*Exemple de réponse:*

01/02/2017

Date de fin du projet

*Exemple de réponse:*

31/01/2021

## 3. Présentation générale des données

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

*Exemple de réponse:*

Nous allons acquérir des images microscopiques des infections par le virus de la fièvre jaune pour effectuer une analyse en super résolution de l'infection cellulaire. Des images microscopiques, des vidéos de l'infection en direct ainsi que des images de Western-Blot sont nécessaires pour une compréhension totale et la réalisation du projet.

*Recommandations:*

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

Combien de jeu(x) de données allez-vous générer durant ce projet ?

*Recommandations:*

Un jeu de données peut être décrit comme l'agrégation de données brutes ou dérivées présentant une certaine "unité", rassemblées pour former un ensemble cohérent. Le nombre de jeux de données pourra être renseigné à la fin du projet.

*Exemple de réponse:*

3 jeux de données

Quelle est la nature et le format des données générées/collectées ?

*Recommandations:*

Précisez la nature des données, qu'elles soient générées ou réutilisées : données de génomique, protéomique, glycomiques, lipidomiques, transcriptomiques, données de structures moléculaires, données d'interactions moléculaires, données cellulaires, physiologiques, phénotypiques, cliniques, épidémiologiques...

Pour vous aider, un document présentant des exemples de types de données qui peuvent être générées à l'Institut Pasteur est accessible [ici](#).

*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

Indiquez le volume prévisionnel des données générées pour ce projet

*Exemple de réponse:*

3 To

*Recommandations:*

Au début du projet, indiquez le volume prévisionnel. S'il vous est inconnu, vous pourrez répondre à cette question en fin de projet.

Des données existantes seront-elles également réutilisées ? Si oui, précisez leur origine.

- Non applicable
- Non
- Oui

*Recommandations:*

Avant de réutiliser des données, vous devez vous rapprocher :

- du centre de recherche translationnelle (CRT) pour les données concernant la personne humaine : crt-opendesk@pasteur.fr
- de la Direction Juridique pour les autres types de données

En effet, vous devez vous assurer que vous avez le droit de réutiliser ces données :

- 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...

*Exemple de réponse:*

Aucune donnée ancienne ne sera utilisée dans ce projet. Tous les livrables du projet proviendront des données originales acquises au cours de cette période.

Mettez en évidence le potentiel de réutilisation de vos données : à qui seront-elles utiles ?

*Recommandations:*

Cette question doit être prévue dès le protocole dans le cas de recherches sur la personne humaine. En cas de besoin, n'hésitez pas à contacter votre chef de projet CRT ou l'Open-Desk du CRT : crt-opendesk@pasteur.fr

*Exemple de réponse:*

Les scientifiques du traitement d'images peuvent utiliser ce jeu de données d'acquisition en microscopie à super-résolution de virus individuels afin de comparer les nouveaux outils de traitement d'images.

#### **4. Ressources nécessaires à gestion des données**

De quelles ressources matérielles avez-vous besoin pour la gestion de vos données ?

*Recommandations:*

Des ressources matérielles peuvent être nécessaires pour la collecte, le stockage, l'analyse et le transfert de vos données. Par exemple, des serveurs de stockage, des ordinateurs, tablettes, téléphones, écrans de sécurité...

*Exemple de réponse:*

Un espace de stockage supplémentaire sera nécessaire. De plus, une base de données devra être mise en place pour gérer les données de spectrométrie au cours du projet.

Qui assure la gestion opérationnelle des données au cours du projet de recherche ?

*Exemple de réponse:*

Dans l'équipe de recherche X (à répéter pour chaque équipe de recherche) :

- A est responsable de la collecte, traitement et analyse des données
- B est responsable de la génération des métadonnées et de la documentation accompagnant les données
- C est responsable du stockage des données
- D est responsable de l'archivage et du partage des données

Quelles formations ou quel accompagnement vous semblent nécessaires pour vous aider à la gestion de vos données ?

*Recommandations:*

Indiquez si une formation est nécessaire (rédaction du PGD, génération des métadonnées, génération des données...) : combien de jours ? Combien de personnes ?

Indiquez également tous les documents ou supports qui vous seraient utiles pour gérer vos données, assurer la qualité de vos données ou de la gestion des données...

*Exemple de réponse:*

Le responsable de projet aimerait des conseils juridiques et organisationnels sur les sujets suivants : les données à caractère personnel et les licences de réutilisation des données. L'équipe (5 personnes) aura également besoin d'une formation sur des questions techniques : métadonnées, standards de métadonnées et archivage.

Quel budget avez-vous prévu pour la gestion de vos données ? Comment comptez-vous financer ces coûts ?

*Recommandations:*

Au début du projet, indiquez le budget prévisionnel. S'il vous est inconnu, vous pourrez répondre à cette question en fin de projet.

*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)"

## 5. Aspects juridiques et éthiques

Votre projet comporte-t-il des données à caractère personnel ?

- Oui
- Non
- Non applicable

*Exemple de réponse:*

Oui, le projet comporte des données à caractère personnel. Les démarches légales pour pouvoir gérer ce type de données ont été effectuées.

*Recommandations:*

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....)

Si votre projet comporte des données à caractère personnel, vous devez contacter le Délégué à la Protection des Données (dpo@pasteur.fr) ou présenter votre projet au CRT (crt-opendesk@pasteur.fr)

Vous n'avez pas besoin de détailler les démarches entreprises dans la réponse à cette question. Indiquez simplement que les mesures pour gérer ce type de données ont été effectuées. Si cela est pertinent, incluez les références aux livrables éthiques et au chapitre d'éthique de votre projet de recherche.

Votre projet comporte-t-il d'autres données soumises à une obligation contractuelle, réglementaire ou légale ? Si oui, de quel type ?

- Oui
- Non
- Non applicable

*Exemple de réponse:*

Le projet comporte des données liées à un contrat avec un industriel qui ne pourront pas être rendues librement accessibles.

*Recommandations:*

Ci-dessous, la liste des types de données que votre projet peut comporter :

- données secret défense

- données soumises à un régime restrictif, qui peuvent porter préjudice aux potentiels scientifiques et techniques de la Nation

- données protégées par un secret professionnel

- données qui feront l'objet d'une demande de brevet

- données à caractère personnel

- données issues d'une extraction substantielle d'une base de données

- données issues de services statistiques comme l'INSEE

- données protégées par le droit d'auteur ou par un copyright
- données liées à un contrat avec un tiers

Consultez notre [outil de décision en ligne](#) pour savoir si vos données sont soumises à une obligation réglementaire, contractuelle ou légale et pour savoir quelles mesures doivent être prises

Toutes les informations à connaître sont résumées dans ce [logigramme](#)

## 6. Gestion des données au cours du projet

Quel est le lieu de stockage de vos données au cours du projet ?

*Recommandations:*

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.

*Exemple de réponse:*

Les données sont stockées sur l'espace de stockage proposé par la DSI.

Utilisez-vous un plan de classement pour gérer vos fichiers de données ? Indiquez brièvement comment il est organisé.

- Non applicable
- Non
- Oui

*Exemple de réponse:*

Oui, chaque unité a créé un plan de classement au début du projet. Ce plan de classement chronologique reprend les différentes étapes du projet. Les données brutes et les données traitées seront stockées dans des dossiers différents.

*Recommandations:*

Pour les projets de recherche sur la personne humaine, nous vous conseillons de suivre le plan de classement proposé par le CRT. Pour utiliser ce plan de classement, contactez [crt-clincore@pasteur.fr](mailto:crt-clincore@pasteur.fr).

Voir le [document rédigé par le service des archives](#)

Quelles règles de nommage utilisez-vous pour vos données ? Quelles règles utilisez-vous pour un versionnage explicite ?

*Recommandations:*

Voir le [document rédigé par le service des archives](#)

Les règles pour un bon nommage sont :

- un nom succinct (maximum 30/40 caractères)
- un nom informatif : sujet\_TypeDoc\_date\_version
- un nom interopérable : pas d'espace (underscore uniquement), pas de ponctuation, pas de caractères spéciaux, date écrites de la façon suivante : AAAAMMJJ

*Exemple de réponse:*

Chaque fichier est nommé de la façon suivante : Sujet\_auteur\_TypeDocument\_date\_version.

Les différentes versions sont nommées de la façon suivante : V01, V02, VBR (version brouillon), VF (version finale)

Quelles sont les mesures mises en place pour garantir la qualité des données ?

*Recommandations:*

Expliquer comment la qualité de la collecte des données sera contrôlée et documentée. 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 ou la revue par les pairs.

Pour plus d'informations, merci de contacter [qualite@pasteur.fr](mailto:qualite@pasteur.fr)

Comment vous assurez-vous que les données de votre projet de recherche sont correctement gérées ? A quelle fréquence le vérifiez-vous et que faites-vous en cas d'écart ?

*Recommandations:*

Une check-list pour vérifier que les données de votre projet sont correctement gérées est disponible [ici](#).

Pour tout conseil ou information complémentaire, vous pouvez contacter le service qualité : [qualite@pasteur.fr](mailto:qualite@pasteur.fr)

*Exemple de réponse:*

Au début du projet, puis régulièrement au cours du projet (tous les 6 mois), le coordinateur du projet vérifie que les réponses à la check-list proposée par l'Institut Pasteur (ci-jointe) sont positives. En cas de réponse négative, des actions d'amélioration sont mises en place rapidement.

## 7. Sélection des données pour l'archivage à long terme

Existe-t-il des contraintes réglementaires de conservation de vos données ? Si oui, lesquelles ?

- Non applicable
- Non
- Oui

### *Recommandations:*

Pour connaître les contraintes applicables, se rapprocher de la Direction Juridique.

Pour les recherches sur la personne humaine ou utilisant des données de santé, se rapprocher du CRT (crt-opendesk@pasteur.fr).

Quels sont les jeux de données ayant une valeur à long-terme et qui devraient être conservés ? Quels sont les jeux de données à détruire ?

### *Recommandations:*

Pour vous aider à répondre à toutes les questions liées à l'archivage des données de recherche, consultez le [document rédigé par le CeRIS](#)

### *Exemple de réponse:*

Ces données sont uniques et devraient donc être conservées.

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 ?

### *Exemple de réponse:*

A la fin du projet, les données seront transférées sur l'entrepôt ZENODO qui assure un archivage durable des données de recherche finales.

### *Recommandations:*

Attention, les données dont il est question ici sont les données finales. Au cours de la recherche, vous devez privilégier les outils sécurisés mis à disposition par la DSI pour stocker vos données.

Précisez les formats choisis pour l'archivage.

### *Recommandations:*

Choisissez un format ouvert et stable dans le temps si possible. Evitez les formats propriétaires ou les formats qui dépendent de l'environnement technologique.

Voir le [document rédigé par le service des archives](#)

### *Exemple de réponse:*

XML, CSV, PDF/A, RDF, etc...

Pendant combien de temps les données seront-elles conservées ?

### *Recommandations:*

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-le.

Quel est le volume final de données archivées ?

### *Recommandations:*

Au début du projet, indiquez le volume prévisionnel. S'il vous est inconnu, vous pourrez répondre à cette question en fin de projet.

### *Exemple de réponse:*

2 To

Si une conservation à long terme est requise, comment comptez-vous couvrir ces coûts ?

### *Exemple de réponse:*

Les coûts associés à la conservation à long terme seront pris en charge par l'Institut Pasteur.

# Institut Pasteur: Institut Pasteur - Trame de DMP (FR) - Jeu de données

## 1. Description des données

Identifiant et nom du jeu de données

Qui est le fournisseur ou le producteur des données ?

*Recommandations:*

Cette question est importante dans le cas de projets collaboratifs. Vous devez indiquer ici lequel des partenaires va produire ou fournir le jeu de données.

Quelle est la nature et le format des données de ce jeu de données ?

*Recommandations:*

Pour vous aider, un document présentant des exemples de types de données qui peuvent être générées à l'Institut Pasteur est accessible [ici](#).

Choisissez un format ouvert et stable dans le temps si possible. Evitez les formats propriétaires ou les formats qui dépendent de l'environnement technologique.

Voir le [document rédigé par le service des archives](#).

*Exemple de réponse:*

Données tabulées au format CSV, images au format PNG, données de neuro-imagerie au format NIFTI, structures moléculaires au format FCS, données de diffraction au format ASCII.

Décrivez de façon plus détaillée les données de ce jeu de données

*Exemple de réponse:*

Ce jeu de données comprend 20 échantillons biologiques (vache et mouton) qui ont été analysés 4 fois par spectrométrie de masse ainsi que des échantillons contrôle.

Décrivez la méthode de collecte et/ou de génération de ces données

*Recommandations:*

Indiquez comment les données sont générées ou collectées : données générées par une machine, enquête, observation, simulation, analyse... Précisez si les données sont générées pendant le projet ou réutilisées.

*Exemple de réponse:*

Les données sont générées par un spectromètre de masse et ensuite analysées avec le logiciel MassChroQ.

Décrivez votre jeu de données avec des mots-clés

*Recommandations:*

Nous vous conseillons de décrire le jeu de données par 3 mots-clés au minimum. Un jeu de données décrit précisément sera plus facilement trouvé et donc réutilisé.

Indiquez l'URL ou l'identifiant pérenne permettant d'accéder à votre jeu de données

*Recommandations:*

Certains entrepôts de données et certains éditeurs attribuent des identifiants pérennes aux jeux de données. Si c'est le cas, indiquez l'identifiant pérenne ici. Sinon, indiquez l'URL permettant d'accéder au jeu de données.

Exemples d'identifiants pérennes : Handle, DOI (Digital Object Identifier), Ark...

Quel est volume de données dans ce jeu de données ?

*Exemple de réponse:*

1 To

## 2. Rendre les données librement accessibles

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

- Non applicable
- Non
- Oui

*Exemple de réponse:*

Toutes les structures protéiques seront librement disponibles car stockées dans RCSB Protein Data Bank sous licence CC0.

Quel entrepôt de données avez-vous choisi pour stocker et rendre accessible ce jeu de données ?

*Recommandations:*

La Commission Européenne demande que les porteurs de projet H2020 déposent leurs données (celles permettant de

valider les résultats présentés dans les publications) dans un entrepôt de données en ligne. L'Institut Pasteur développe actuellement un entrepôt nommé OWEY. Si vous souhaitez déposer des données dans cet entrepôt, contactez [owey@pasteur.fr](mailto:owey@pasteur.fr)

Pour le cas où l'entrepôt interne ne serait pas adapté, le CeRIS vous propose [un document pour vous aider dans votre choix](#).

*Exemple de réponse:*

Entrepôt interne (OWEY), Entrepôt externe (GenBank, RCSB Protein Data Bank, Zenodo...)

Ce jeu de données fera-t-il l'objet d'une demande de brevet ? Si oui, il devra rester confidentiel.

- Non applicable
- Non
- Oui

*Recommandations:*

Si vous prévoyez de protéger une invention et de déposer une demande de brevet, veillez à respecter les points suivants :

- Ne publiez pas vos données tant que vous n'avez pas vérifié auprès du Service des Brevets et Inventions si elles ne peuvent pas faire l'objet d'un dépôt de demande de brevet.
- Veillez à ne pas mentionner de données sur le site internet de votre projet car cela compromettrait la brevetabilité de votre invention.
- Les données ne seront pas publiées jusqu'à 18 mois après le dépôt de la demande de brevet.
- Marquez vos données comme confidentielles (cf. directive de classification de l'information en cours de validation)

Pour toute information complémentaire, n'hésitez-pas à contacter le Service des Brevets et Inventions : [sbi@pasteur.fr](mailto:sbi@pasteur.fr)

Des formulaires de Déclaration d'Invention sont à votre disposition [ici](#).

Si ce jeu de données ne peut pas être mis à disposition pour une autre raison, expliquez pourquoi.

*Exemple de réponse:*

Ex 1 : Ce jeu de données contient des données à caractère personnel non anonymisées et non pseudonymisées. Ce jeu de données ne peut donc pas être rendu public.

Ex 2 : Ce jeu de données a été produit en collaboration avec une entreprise privée. Le contrat avec cette entreprise prévoit que les données ne pourront pas être rendues publiques.

*Recommandations:*

Certaines données de recherche ne peuvent pas être rendues publiques car il s'agit de données soumises à une obligation réglementaire, contractuelle ou légale.

Consultez notre [outil de décision en ligne](#) pour savoir si vos données sont soumises à une obligation réglementaire, contractuelle ou légale.

Consultez [ce logigramme](#) pour un rappel des données qui ne doivent pas être publiées

Précisez comment l'accès à ce jeu de données sera fourni au cas où des restrictions s'appliquent.

*Recommandations:*

Pour savoir si des restrictions d'accès s'appliquent à votre jeu de données, contactez l'expert en classification de l'information à [rssi@pasteur.fr](mailto:rssi@pasteur.fr)

*Exemple de réponse:*

Les données de ce jeu de données sont de type "données relatives au potentiel scientifique et technique de la Nation". L'accès aux données nécessite une autorisation auprès de l'Institut Pasteur, propriétaire des données. Après autorisation, un login et un mot de passe seront fournis au demandeur pour accéder aux données.

Quelles techniques ou outils logiciels sont nécessaires pour accéder aux données ? Fournissez-vous une documentation ou le code open source du logiciel ?

*Exemple de réponse:*

L'accès aux données nécessite un logiciel développé par notre service. Pour rendre nos données accessibles, nous fournissons le code open source de ce logiciel.

### 3. Rendre les données 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.

- Non applicable
- Oui
- Non

*Exemple de réponse:*

Oui, ce jeu de données est identifié par un DOI. Il est constitué d'un ensemble de fichiers numériques identifiés par un nom explicite : sujet\_typeDoc\_equipe\_date\_version

*Recommandations:*

Exemples d'identifiants pérennes : système Handle, DOI, Ark. Le choix de l'identifiant pérenne dépend généralement de l'entrepôt dans lequel vous avez déposé vos données.

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.

*Recommandations:*

Les métadonnées sont définies comme des données fournissant des informations sur les données. A l'inverse de la documentation, les métadonnées sont structurées et lisibles par une machine et par les humains. Pour la description de vos données, donnez la préférence aux standards de votre discipline.

Pour plus d'informations sur les métadonnées et standards de métadonnées, consultez [ce document](#).

*Exemple de réponse:*

Les métadonnées sont basées sur les métadonnées de ZENODO : titre, auteur, date, contributeur, description, mots-clés, format, type de ressources, etc.

Ce jeu de données est-il décrit par des mots-clés permettant de le retrouver facilement ?

- Non applicable
- Oui
- Non

*Exemple de réponse:*

Oui, ce jeu de données est décrit par 3 mots-clés minimum

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

- Non applicable
- Oui
- Non

*Recommandations:*

La documentation est uniquement lisible par un humain, alors que les métadonnées doivent être lisibles par une machine.

*Exemple de réponse:*

Un fichier est disponible pour chaque donnée pour récapituler les analyses effectuées.

#### **4. Rendre les données interopérables**

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

- Non applicable
- Non
- Oui

*Exemple de réponse:*

Oui, les photographies de microscopie sont au format PNG. Les tableaux de résultats accompagnant les photographies sont au format CSV. Les formats PNG et CSV sont des formats ouverts et recommandés par le document de référence français sur l'interopérabilité (Référentiel Général d'Interopérabilité).

*Recommandations:*

L'interopérabilité est la capacité que possède un système, à fonctionner avec d'autres systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre. L'interopérabilité technique dépend essentiellement du format des données.

Un format de données est interopérable si :

- il est ouvert
- il est accessible, largement diffusé et de nombreux logiciels peuvent l'exploiter
- il existe des outils permettant de le faire migrer vers un autre format
- il ne dépend pas de l'environnement technologique ou économique.

Pour plus d'informations sur l'interopérabilité et sur les formats recommandés, consulter le [référentiel général d'interopérabilité](#)

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

*Exemple de réponse:*

Nos données sont dans un format lisible uniquement avec un logiciel développé par notre service. Toutefois, nous fournissons le code open source du logiciel nécessaire pour accéder aux données (documentation annexe).

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 ?



*Exemple de réponse:*

Ex 1 : Nous avons décrit nos données suivant le format CIF (Crystallographic Information File). Il s'agit d'un standard permettant l'archivage et l'échange de données en cristallographie : <http://www.iucr.org/resources/cif>.

Ex 2 : Nos métadonnées sont spécifiques de notre projet mais nous les alignerons avec le Dublin Core et l'EML (Ecological Markup Language).

Ex 3 : Comme notre projet concerne des produits médicaux à usage humain, nous avons utilisé le medDRA (Medical Dictionary for Regulatory Activities) pour décrire nos données.

*Recommandations:*

Une ontologie définit un vocabulaire commun pour les chercheurs qui ont besoin de partager l'information dans un domaine. Elle inclut des définitions lisibles en machine des concepts de base de ce domaine et de leurs relations.

## 5. Accroître la réutilisation des données

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.

*Recommandations:*

Si un jeu de données qui peut intéresser le public est mis à disposition de tous, n'hésitez pas à contacter la Direction de la Communication, ce qui permettra de valoriser les recherches de l'Institut Pasteur.

La réutilisation des jeux de données ayant fait l'objet d'une demande de brevet est restreinte. En effet, même si les données sont rendues publiques 18 mois après le dépôt de la demande de brevet, elles ne peuvent pas être réutilisées par des parties tierces sans une licence d'exploitation de la demande de brevet ou du brevet.

Pour toute information complémentaire, n'hésitez pas à contacter le service des Brevets et Inventions : [sbi@pasteur.fr](mailto:sbi@pasteur.fr)

*Exemple de réponse:*

Ex 1 : Les données de ce jeu de données pourront être réutilisées en interne (par les autres entités de l'Institut Pasteur), par des instituts tiers et par des industriels.

Ex 2 : Le jeu de données fera l'objet d'une demande de brevet. Les données ne pourront donc pas être réutilisées sans l'accord du titulaire du brevet : toute réutilisation nécessite une licence d'exploitation de la demande de brevet ou du brevet. Par exception, les données pourront uniquement être réutilisées pour vérifier expérimentalement que le brevet fonctionne (exemption de recherche).

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

*Recommandations:*

Pour savoir comment appliquer une licence à vos données de recherche et quelle licence est la plus adaptée, consultez le [guide du Digital Curation Centre](#).

En cas de besoin, n'hésitez pas à contacter la Direction juridique

*Exemple de réponse:*

Ex 1 : Licence CC-BY (<https://creativecommons.org/licenses/by/2.0/>)

Ex 2 : Open Data Commons Attribution License (<http://opendatacommons.org/licenses/by/{version}>)

Ex 3 : Licence libre Etalab

Quand 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.

*Exemple de réponse:*

Les données pourront être réutilisées dans un an, après la soumission et la publication de l'article.

*Recommandations:*

Vous pouvez choisir de ne pas permettre la réutilisation de vos données pendant une certaine durée (embargo). Par exemple, si vous souhaitez mener d'autres recherches à partir de ces données.

Précisez la durée pendant laquelle le jeu de données sera réutilisable

*Exemple de réponse:*

Les données seront conservées pendant la durée maximale permise par la capacité de stockage de l'Institut Pasteur.

## 6. Sécurité des données

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 ?

*Recommandations:*

Préciser notamment si ce jeu de données doit rester confidentiel car il est soumis à une réglementation. Indiquer si l'accès à ce jeu de données doit être restreint (aux pasteuriens et Orex, aux chercheurs du projet, à certaines personnes du projet...) ou si les données sont publiques et n'ont pas besoin d'être sécurisées.

*Exemple de réponse:*

Les données brutes sont des données à caractère personnel qui doivent rester confidentielles et uniquement accessibles aux chercheurs du projet à l'Institut Pasteur. Les résultats intermédiaires seront accessibles plus largement aux chercheurs du projet (projet multi-partenaires). Des mesures de sécurité adaptées seront mises en place à chaque étape du projet : données sécurisées à l'Institut Pasteur pendant le projet (cf question suivantes), données conservées dans un entrepôt sécurisé externe à l'Institut Pasteur après le projet (cf question suivantes)

Pendant le projet (avant le dépôt des données dans un entrepôt), ce jeu de données est-il stocké de façon sécurisée ?

*Exemple de réponse:*

Ce jeu de données étant confidentiel, il sera stocké sur des serveurs internes sécurisés à l'Institut Pasteur pendant la durée de la recherche.

*Recommandations:*

*! - la sécurité consiste à répondre à un besoin légitime. Si les données sensibles, comme celles à caractère personnel doivent être protégées, il n'est pas nécessaire de protéger des données publiques.*

Merci de décrire toutes les mesures prises pour protéger vos données.

Vos données sont sécurisées si :

- Elles sont stockées sur un espace sécurisé répondant à la politique de sécurité des systèmes d'information de l'Institut Pasteur
- L'accès à vos données est restreint aux personnes concernées
- Les données stockées sur les supports mobiles (postes de travail, clé USB...) sont chiffrées
- Vous effectuez régulièrement des copies de sauvegarde de vos données. Ces copies ne doivent pas être conservées dans le même local que le support original

Pour toute question sur la sécurité des données, n'hésitez pas à contacter [rssip@pasteur.fr](mailto:rssip@pasteur.fr)

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 ?

*Recommandations:*

Cette question concerne l'ensemble des projets. Par défaut, les projets dont les données sont stockées dans le Datacenter de la DSI sont conformes. Pour les projets hébergeant tout ou partie des données en externe, des documents complémentaires peuvent être nécessaires. Si vos données sont sensibles et sont stockées sur un entrepôt tiers, vous devez :

- signer des clauses contractuelles portant sur la sécurité des données avec le tiers
- OU signer un Plan Assurance Sécurité avec le tiers
- OU vous assurer que le tiers dispose d'une certification ISO27001 depuis au moins 3 ans
- OU vous assurer que le tiers dispose et met en œuvre une Politique de sécurité de son système d'information

*Exemple de réponse:*

Ex 1 : Nous nous sommes assurés que nos données sont bien sécurisées dans l'entrepôt choisi. En effet, nous avons signé un Plan Assurance Sécurité avec l'organisme qui héberge nos données.

Ex 2 : Ce jeu de données n'étant pas confidentiel, il n'a pas besoin d'être protégé dans un entrepôt sécurisé. Il sera conservé sur un entrepôt public après la fin du projet.

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

*Recommandations:*

Les échanges de données sont sécurisés si :

- une plateforme d'échange sécurisée a été mise en œuvre
- les données sont transmises de manière chiffrée (https, ftp, pièce jointe chiffrée ou utilisation de pgp)

Pour vous aider à répondre à cette question, n'hésitez pas à contacter [rssip@pasteur.fr](mailto:rssip@pasteur.fr)

*Exemple de réponse:*

Notre projet ne comporte pas de collecte et d'échange de données. Aucune mesure de sécurité n'est nécessaire.