INCa - Institut national du cancer: INCa - Modèle de PGD (français)

RENSEIGNEMENT ADMINISTRATIF COMPLEMENTAIRE

Indiquer le numéro de l'acte attributif (tel qu'il vous a été fourni lors du conventionnement)
Ce numéro est différent du numéro de subvention.

Exemple de réponse:

Ce numéro se compose de l'année suivie de trois chiffres. Par exemple pour un projet financé en 2013, le numéro de l'acte attributif a le format suivant : 2013-033.

1. DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Recommandations:

- Expliquer quelles méthodologies ou quels logiciels seront utilisés si de nouvelles données sont recueillies ou produites.
- Enoncer les éventuelles restrictions à la réutilisation des données préexistantes.
- 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.

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

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).

2. DOCUMENTATION ET QUALITÉ DES DONNÉES

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

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.
- 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.
- 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.

2b. 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.

3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

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, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.

3b. 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.

4. EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE

4a. 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 ?

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.

4b. 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.

4c. 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.

5. PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME

5a. 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 ?

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.
- 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.

5b. 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:

- 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.
- Décrire les utilisations (et/ou les utilisateurs) prévisibles des données dans un cadre de recherche.
- Indiquer où les données seront déposées. 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.

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

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.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

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.

6. RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

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.

6b. 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.
- 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.