

CNRS: PGD CNRS

1. Description des données et collecte ou réutilisation de données existantes

1.1 Description générale du produit de recherche

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.
- Préciser les objectifs pour lesquels le jeu de données a été produit.
- Indiquer s'il sera envisagé d'attribuer aux données un identifiant persistant (PID). Les PID devraient être appliqués de manière à ce que les données puissent être localisées et référencées de façon fiable et efficace. Ils aident aussi à comptabiliser les citations et les réutilisations. Typiquement, un entrepôt pérenne digne de confiance attribuera des identifiants persistants. L'Inist- CNRS est le relai français de Datacite pour l'attribution de DOI. Si l'entrepôt ne propose pas d'identifiant, il est possible, par l'intermédiaire de la plateforme [PID OPIDoR](#), d'obtenir un identifiant pérenne pour vos données.

1.2 Est-ce que des données existantes seront réutilisées ?

Recommandations:

- Exploiter des données de recherche déjà existantes peut représenter un réel avantage en particulier si elles sont coûteuses à produire. Pour identifier les entrepôts où elles peuvent être archivées, vous pouvez consulter des annuaires comme [Cat OPIDoR](#), [re3data](#), [FAIRsharing](#). Il est recommandé à ce stade de se rapprocher d'un [atelier de la donnée](#) de proximité qui pourra vous accompagner dans cette tâche.
- Indiquer la nature et les caractéristiques des données. Si des données ont été réutilisées, expliquer la provenance de ces données – indiquer le cas échéant leur URL ou identifiant de provenance.
- Enoncer les éventuelles contraintes à la réutilisation des données préexistantes (par ex : licences contraignantes, coûts, etc.).
- Indiquer brièvement le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée, mais écartée.
- Pour vous aider sur les aspects juridiques liés à la réutilisation des données, vous pouvez consulter le logigramme mis à votre disposition par l'[Institut Pasteur](#).

1.3 Comment seront produites/collectées les nouvelles données ?

Recommandations:

- Expliquer quelles méthodologies ou quels logiciels seront utilisés si de nouvelles données sont recueillies ou produites. Indiquer pour les logiciels quelle sera la dernière version utilisée. Préciser le cas échéant comment les données seront générées ou collectées et avec quels outils, selon quels processus.
- Expliquer comment la provenance des données sera documentée.
- L'acquisition et la collecte des données, pour être fiable, doit être bien documentée. Elle peut être, dans certains cas, facilitée par l'usage de tablettes de terrain ou de cahiers de laboratoire électroniques (CLE). Le CNRS propose une offre de service de CLE basé sur la solution [eLabFTW](#), pour plus d'informations, consultez la [page intranet du CNRS](#). D'autres outils comme [Collec-Science](#) permettent également de consigner les relevés et métadonnées observés.

2. Documentation et qualité des données

2.1 Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Recommandations:

- Une attention particulière doit être portée sur le choix et la qualité des standards et métadonnées qui sont essentiels pour assurer la compréhension, la fiabilité et la réexploitation des données. Pour l'ensemble de ces tâches, il est recommandé de se rapprocher des [ateliers de la donnée](#) et des [centres de références thématiques](#). Les équipes scientifiques peuvent également s'appuyer sur les compétences et expertises développées au sein des [réseaux technologiques transverses du CNRS](#).
- Indiquer quelles métadonnées seront fournies pour aider à la recherche, l'identification, la compréhension et la réutilisation des données. Préciser si des métadonnées sont générées automatiquement ou manuellement.
- Indiquer quels standards de métadonnées seront utilisés (par exemple DDI, TEI, EML, MARC, CMDI). Utiliser les standards de métadonnées de communauté lorsque ceux-ci existent. L'outil [FAIRsharing](#) et la plateforme [Loterre](#) donnent accès et permettent le téléchargement de thésaurus, ontologies, vocabulaires contrôlés et autres référentiels terminologiques.
- Indiquer comment les données seront organisées au cours du projet, en mentionnant par exemple conventions de nommage, contrôle des versions et structures des dossiers. Pour guider votre démarche vers une gestion cohérente de vos fichiers, vous pouvez consulter la ressource « [Comment bien nommer ses fichiers](#) » disponible sur la plateforme DoRANum.
- 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.

2.2 Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?

Recommandations:

- Montrer que l'ensemble des activités qui conduit à l'obtention d'une donnée est maîtrisé : par exemple maîtrise de la chaîne fonctionnelle d'une analyse (pipettes, balance, équipements d'analyse) ou du respect d'un protocole d'enquête (échantillon

représentatif).

- 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.
- S'assurer également du bon niveau de structuration de la donnée (structure et format du fichier), du respect des standards et référentiels reconnus et de la présence d'informations complémentaires (documentation) qui garantissent également la qualité d'une donnée.
- Pour l'adoption de bonnes pratiques, vous pouvez consulter le guide « [Traçabilité des activités en recherche et gestion des connaissances](#) » et vous rapprocher du [réseau Qualité en Recherche](#) (QeR).

3. Exigences légales et éthiques, code de conduite

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Recommandations:

Se référer aux recommandations du Service Protection des Données du CNRS pour le traitement des données à caractère personnel.

3.2 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 les conditions légales d'utilisation et d'accès aux données :

- Expliquer quelles conditions d'accès s'appliqueront aux données. Les données seront-elles librement accessibles, ou des restrictions seront-elles appliquées ? Si oui, lesquelles ? Le [logigramme](#) de l'École des Ponts ParisTech ou celui proposé par [INRAe](#), vous aideront à répondre à ces questions.
- Les délais relatifs au dépôt d'un brevet justifient la mise en place d'un embargo. Justifier le motif d'un embargo éventuel et sa durée.
- Assurez-vous 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. Définissez distinctement les responsabilités de chaque partenaire pour tous les produits de recherche.
- Indiquer si les droits de propriété intellectuelle (par exemple la directive bases de données, droits sui generis) sont affectés. Dans l'affirmative, veuillez expliquer lesquels et comment cela sera-t-il traité. Pour en savoir plus sur le cadre juridique applicable aux données, consultez le « [Guide d'application de la Loi pour une République numérique pour les données de la recherche](#) ».
- Indiquer s'il y a des restrictions sur la réutilisation des données fournies par des tiers.

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

Recommandations:

- Déterminer si les questions d'éthique ont une incidence sur la façon dont les données sont stockées et transférées, qui peut 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.
- Pour en savoir plus sur les pratiques éthiques et responsables, vous pouvez vous rapprocher du [comité d'éthique du CNRS](#) et consulter des guides de référence comme « [Pratiquer une recherche intègre et responsable](#) » et « [The European Charter for Researchers](#) ».
- Pour vous guider sur les bonnes pratiques en matière de déontologie et d'intégrité scientifique, consultez la « [Charte déontologique du CNRS](#) », le « [Code de conduite européen pour l'intégrité en recherche](#) ».

4. Traitement et analyse des données

4.1 Comment et avec quels moyens seront traitées les données ?

Recommandations:

- Expliquer quels sont les traitements opérés sur les données brutes issues des acquisitions et des collectes.
- Préciser si des conversions de format ont été opérées.
- Indiquer quels outils ou infrastructures logicielles sont utilisés.

5. Stockage et sauvegarde des données pendant le processus de recherche

5.1 Comment les données seront-elles stockées et sauvegardées tout au long 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 (quotidienne, hebdomadaire, mensuelle). Stocker les données dans au moins deux lieux distincts. Le principe de base est de dupliquer les données sur des supports différents de ceux de l'équipement utilisé (si possible, 2 copies sur 2 supports différents, dont 1 sur un lieu déporté).
- Privilégier l'utilisation de systèmes de stockage robustes, avec sauvegarde automatique dans des infrastructures mutualisées, tels que data centres, mesocentres ou les services informatiques de l'institution hôte. 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.
- Évaluer la périodicité de mise à jour et d'accès aux données (pour déterminer la fréquence optimale des sauvegardes).
- Évaluer les besoins et moyens nécessaires au stockage et sauvegarde des données en intégrant dans la réflexion les possibles impacts et coûts environnementaux (ou financiers) engendrés par le traitement et la manipulation des données. Il est conseillé de

travailler en amont avec une équipe informatique.

- Veiller à ne stocker que les données qui méritent de l'être (il est parfois moins coûteux de régénérer les données à la demande que de les stocker).
- 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.
- Tenez 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. Se référer à la [politique de sécurité des systèmes d'information du CNRS](#) pour les données hébergées au CNRS.

6. Partage des données et conservation à long terme

6.1 Comment les données seront-elles partagées ?

Recommandations:

- 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, CSV, DOC, TXT, HDF, 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 »). Pour anticiper la conservation à long terme, une [liste des formats éligibles à l'archivage](#) au CINES est disponible sur son site institutionnel.
- Donner des détails sur les volumes (qui peuvent être exprimés en espace de stockage requis (octets), et/ou en quantité d'objets, de fichiers, de lignes et colonnes).
- Expliquer comment les données pourront être retrouvées et partagées (par exemple, par dépôt dans un entrepôt de données de confiance, indexées dans un catalogue, par l'utilisation d'un service de données sécurisé, par traitement direct des demandes de données ou l'utilisation de tout autre mécanisme).
- 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.
- Envisager l'utilisation de licences concernant l'accès et la réutilisation des données. Pour identifier la licence correspondant à vos besoins, vous pouvez consulter cette [liste](#). Les outils [License Selector](#) et [Choose a license](#) peuvent également vous aider à faire un choix.
- Décrire les utilisations (et/ou les utilisateurs) prévisibles des données dans un cadre de recherche.
- Indiquer où les données seront déposées. Des catalogues comme [Cat OPIDoR](#), [re3data](#) ou [FAIRsharing](#), permettent d'identifier des entrepôts de données. Il est recommandé de bien choisir son entrepôt en fonction des données que vous déposez, d'éventuelles recommandations de votre financeur et de privilégier le choix d'un entrepôt conforme aux [critères CoreTrustSeal](#). Vous pouvez vous rapprocher des [ateliers de la donnée](#) pour vous accompagner dans cette démarche. Démontrer que les politiques des entrepôts et les procédures de dépôts (y compris les normes de métadonnées et coûts mis en œuvre) ont été vérifiés.
- Si aucun entrepôt thématique de confiance n'est identifié, il est recommandé de déposer ses données dans l'espace institutionnel CNRS de la plateforme Recherche Data Gouv. Le logigramme «[Où publier vos données ?](#)» vous assurera que vos données sont éligibles.
- Indiquer si les utilisateurs potentiels ont 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.

6.2 Comment les données seront-elles conservées à long terme ?

Recommandations:

- Les données créées au sein d'un établissement public comme le CNRS sont des archives publiques et doivent faire l'objet d'un tri, d'une sélection en vue de leur conservation aux Archives nationales ou départementales.
- Fournir l'information sur la durée d'archivage pérenne des données. Vous pouvez vous appuyer sur le [support de formation DoRANum](#) et le «[Référentiel de gestion des archives de la recherche](#)» pour avoir une idée des durées de conservation préconisées. Indiquer les modalités de financement prévues.
- Identifier les données à conserver à long terme. Certaines données doivent être conservées pour assurer leur traçabilité, justifier l'antériorité d'une découverte ou préserver leur valeur patrimoniale et scientifique. D'autres ont vocation à être éliminées selon la réglementation en vigueur.
- Les entrepôts n'offrent pas la possibilité de pérenniser les données sur le long terme. Pour cela, il convient de choisir une plateforme d'archivage pérenne comme le CINES. Préciser, si les données ne sont pas déposées sur ce type de plateforme, comment elles seront préservées sur le long terme.
- Pour l'archivage pérenne définitif, il existe quelques grands principes de sélection des données à appliquer et il est recommandé de se rapprocher des services d'archives du CNRS.
- Indiquer comment il sera décidé quelles données garder. Préciser les critères de sélection opérés.