

Université de Lille: ANR - Modèle de PGD (français)

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.
- Énoncer 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.

Exemple de réponse:

"Les données utilisées seront strictement nouvelles et seront issues de X entretiens qualitatifs. Les entretiens seront retranscrits au format .txt, et enregistrés au format .mp3 à l'aide d'un micro à pieds de la marque X, via le logiciel X."

Recommandations:

Nouvelles données : Indiquer les protocoles mis en place pour leur création (nombre d'entretiens envisagés, support d'enregistrement des réponses, etc.).

Données réutilisées : Indiquer la source des données ainsi que les droits qui y sont associés ([licences](#)).

RGPD - L'utilisation des données personnelles issues d'une recherche est possible à condition que le responsable scientifique respecte ces dispositions :

- Les participants ont été informés de la possible réutilisation des données lors de la collecte initiale ;
- La réutilisation a pour objectif de répondre à une question précise et ponctuelle ;
- La durée de recherche est limitée et connue.

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).

Recommandations:

Volumétrie : Dans la version initiale du PGD, la volumétrie demandée n'est qu'une estimation qui pourra être réévaluée lors des versions ultérieures.

Formats de fichier : Les formats ouverts correspondent à des fichiers encodés de façon transparente, leur code source fait partie du domaine public. Ils sont interopérables, c'est à dire qu'ils peuvent être créés, lus et modifiés par tous les logiciels destinés à traiter le type du fichier (image, texte, audio, etc.). Les formats ouverts garantissent la pérennité des données et leur archivage.

L'outil FACILE du CINES permet de vérifier que les formats utilisés sont pérennes et archivables.

Exemple de réponse:

"Les prévisions indiquent que les données récupérées auront une volumétrie aux alentours de la dizaines de Go, mais cela peut être grandement variable en fonction des résultats.

Les données seront majoritairement des textes au format .txt, des comptes rendus au format .pdf et des tables de calculs au format .xls. Le format X sera aussi utilisé bien que non ouvert, mais faisait office de référence dans notre domaine."

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.

Recommandations:

Métadonnées : « Données qui décrivent d'autres données. » Elles permettent d'identifier et de comprendre la ressource décrite (exemple en imagerie).

Préciser quels types de métadonnées seront fournies. Il existe deux types de métadonnées :

- Les métadonnées externes : Métadonnées ajoutées manuellement. En général, lors du dépôt des jeux de données dans un entrepôt, il est demandé de remplir un formulaire qui permet d'enrichir les informations descriptives. Ces métadonnées peuvent également être précisées dans des fichiers descriptifs annexes ajoutés aux jeux de données (ex : les fichiers *readme*).
- Les métadonnées embarquées : Métadonnées générées automatiquement. Certains logiciels et matériels en génèrent nativement, mais il est possible de les enrichir manuellement.

Schémas de métadonnées (Dublin Core, DataCite Metadata Schema, etc.). Schéma ou standard, sont des modèles de données destinés à établir une compréhension commune de la signification ou de la sémantique des données, afin d'assurer un usage correct et approprié, et une interprétation des données par ses propriétaires et utilisateurs. Ces standards peuvent être imposés par une discipline ou une communauté. Ils permettent de :

- Décrire les données de façon normée et reconnue par une communauté
- Favoriser le partage et la réutilisation
- Favoriser la visualisation sur les divers supports

Convention de nommage : Elle permet d'explicitier et d'homogénéiser les intitulés des fichiers. Respecter la même organisation facilite le traitement automatique des fichiers dans un système informatique.

<p>Le nom des dossiers doit permettre d'identifier :</p> <ul style="list-style-type: none"> • Le projet ou son acronyme • Les initiales du/des chercheurs • La date de constitution des données • Le type de données • Les conditions de collecte. <p>Exemple de nom de dossier : EPC_SC_focus-groupes_DB</p>	<p>Le nom d'un fichier doit :</p> <ul style="list-style-type: none"> • Expliciter son contenu (sans répéter le nom d'un dossier) • Utiliser des abréviations normalisées si possible • Écrire les dates au format AAAAMMJJ (conserve l'ordre chronologique) • Indiquer les versions des fichiers (V1, VP pour version provisoire, VF pour version finale...). • Proscrire les caractères spéciaux, les accents et les espaces n(privilégier majuscules, tirets du 6 et underscores du 8). • Écrire les chiffres selon des séquences : 1 à 10 = 01-10 et 1 à 100 = 001-100. <p>Exemple de nom de fichier : 20220911_GR1_VP.</p>
---	---

Versioning :

- Utiliser des outils collaboratifs permettant à chaque membre de visualiser et de travailler sur la même version d'un fichier. Cela favorise aussi la récupération de versions antérieures en cas d'incidents.
- Toujours modifier des fichiers en ayant activé l'option « suivie des modifications ».

Structure des dossiers :

- Séparer les données brutes des données traitées. Exemple : L'arborescence peut être thématique (gouvernance, expérimentation, publication...).
- Éviter de créer trop de sous-dossiers afin de limiter le nombre de clics pour accéder à l'information.

Documentation :

- Il existe des modèles de rédaction de fichiers Readme. Les *readme* peuvent être ajoutés aux jeux de données lors du dépôt dans un entrepôt. Ils contiennent les informations essentielles sur le jeu de données permettant d'en comprendre le contexte de production, la structure et le contenu.
- Préciser si du vocabulaire et des ontologies spécifiques sont utilisées pour vos données, et si oui, lesquels ?

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.

Recommandations:

Stockage des données : Cette partie évoque les enjeux liés à la sauvegarde et au partage des données auprès de l'équipe pendant la durée de vie du projet de recherche.

- Comment prévenir les risques de perte des données ?
- Comment contrôler les politiques d'accès aux données protégées ?
- Comment transmettre des fichiers lourds de manière sécurisée ?
- Attribuer différents niveaux d'usages des fichiers (lecture seule, écriture, suppression, accès limité dans le temps...).

Pour en apprendre plus, l'université de Lille propose un [guide des bonnes pratiques](#) à ce sujet.

Espace de stockage à l'université de Lille :

- **Nextcloud**, dans une limite initiale de 50 Go. Il est possible de demander un agrandissement de l'espace disponible auprès de la direction du numérique.
- **Serveurs virtuels de grande capacité** jusqu'à 10 To (service payant).
- Le **Mésocentre de Calcul Scientifique Intensif** permet d'effectuer de la simulation et de la modélisation de données. Les capacités de stockage et de traitement sont paramétrables auprès de la DGDNum (Direction Générale Déléguée au Numérique) de l'université.

Espace de stockage national et travail collaboratif :

<ul style="list-style-type: none">• Sharedocs , pour les sciences humaines et sociales.• Ods MyCore du CNRS.• Osmose plateforme des communautés professionnelles de l'État.• Filesender , pour le partage sécurisé et chiffré de fichiers lourds jusqu'à 100 Go gratuitement.	<ul style="list-style-type: none">• OnlyOffice , outil collaboratif Open Source, une alternative à Google Docs• Code.gouv propose une liste des logiciels recommandés.• Google Drive est déconseillé par la CNIL dans l'Enseignement supérieur, car il n'assure pas la sécurité des données et la conservation des droits sur les données.
--	--

Règle de stockage :

- La règle du **3-2-1** : **3** copies des données sur au moins **2** supports différents (exemple : sauvegarde sur un disque dur externe et sur une clé USB), dont au moins **1** copie à distance avec synchronisation (cloud).
- Adapter la fréquence des sauvegardes en fonction de l'état d'avancement du projet : accentuer la fréquence lors de phases fortement changeantes telle que celle de la collecte.

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.

Recommandations:

Récupération : Mettre en place des procédures de récupérations fiables et qui soient testées par plusieurs membres de l'équipe avant de rencontrer un premier incident.

Accès : Protéger l'accès aux données via des rôles, des droits et des mots de passe (possible aisément avec les outils de travail collaboratif). Les données stockées sur des supports physiques peuvent être sécurisées en utilisant des chiffrements (avec une clé de cryptage séparée des données). L'outil VeraCrypt permet de crypter des données.

Réglementation : L'outil Amnesia en ligne permet d'anonymiser des données.

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.

Recommandations:

Réglementation : Si des données personnelles ou sensibles sont utilisées dans la recherche, il est obligatoire de contacter le **DPO** (délégué à la protection des données) de l'université de Lille via le guichet unique ou par mail (dpo@univ-lille.fr). Le **référent RGPD** (Règlement Général sur la Protection des Données) du laboratoire peut être contacté également.

- Dans le cas de données impliquant la personne humaine, contacter le Comité d'éthique pour la recherche. Plus d'informations sur [ce lien](#).
- Dans le cas de données issues d'expériences biomédicales sur l'être humain, contacter le Comité de protection des personnes. Plus d'informations sur [ce lien](#).

Indiquer si un comité éthique a été consulté : lequel, à quelle date, conclusions de ce dernier.

Préciser si des autorisations et protocoles spécifiques à certaines disciplines et certains objets de recherche sont requis.

Est-ce qu'un recueil de consentement a été établi ? Si oui, toujours en garder une trace.

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.

Recommandations:

Données créées : Dans la majorité des cas, les données produites sont publiques (pas de droit d'auteur par défaut) et la propriété en revient à l'établissement ou à l'État. Cela implique un droit d'accès sur demande, une diffusion gratuite et une libre réutilisation. L'accent est mis sur le côté accessible plus qu'ouvert des données : il faut pouvoir les trouver, et savoir qui contacter pour demander une consultation en cas de non ouverture.

Dans le cas de recherches réalisées avec des institutions privées, les questions de propriété des données sont précisées dans les accords de consortium. Reprendre les points qui y sont abordés en précisant à qui appartiennent les données produites et collectées, ainsi que les droits régissant les différentes bases de données impliquées le cas échéant.

Licences de réutilisation : Favoriser l'utilisation de licences ouvertes Creative Commons ou Etalab pour les données de recherche. Plus d'informations [ici](#).

Données réutilisées : Les droits liés aux données réutilisées doivent être mentionnés : accès libre ou sous licences / régimes spécifiques ?

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.

Recommandations:

Indiquer si un comité éthique a été consulté : lequel, à quelle date, conclusions de ce dernier.

Préciser si des autorisations et protocoles spécifiques à certaines disciplines et certains objets de recherche sont requis.

Est-ce qu'un recueil de consentement a été établi ? Si oui, toujours en garder une trace

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.

Recommandations:

Signalement et diffusion des données : Le partage, la diffusion des données et la conservation à long terme concernent la durée de vie des données après le projet. Dans la version initiale du PGD, il suffit d'émettre des suppositions d'entrepôts de données et de délais. Il sera ensuite possible d'affiner les réponses au fil des différentes versions du PGD.

Lieu de diffusion des données :

Privilégier le dépôt dans un entrepôt de données de confiance (assurant sécurité et préservation, critères consultables sur Ouvrir la Science).

- CoreTrustSeal propose une liste des entrepôts certifiés.
- Re3data permet de trouver un entrepôt disciplinaire ou pluridisciplinaire s'il n'y a pas de recommandation claire dans votre laboratoire ou par vos partenaires
- En l'absence d'entrepôt disciplinaire reconnu, l'université conseille de déposer les données sur Recherche Data Gouv, la plateforme nationale fédérée des données de la recherche.

Licences de réutilisation des données : Indiquer qui pourra utiliser les données signalées en précisant la licence qui sera appliquée.

Pour les données non partagées, tenter de justifier cette décision (données personnelles, confidentielles, contractuelles...).

- Vérifier dans le guide du financeur vos obligations en matière de diffusion des données.
- Selon la loi pour une république numérique de 2016, le code de la recherche et le code régissant la relation entre le public et l'administration il est précisé que : si rien ne s'y oppose (hors exceptions réglementaires), les données financées à hauteur de 50 % par des fonds publics doivent être partagées afin de sortir d'une logique de propriété privée et de constituer ces données en biens communs. Les données sont alors soumises au principe d'ouverture par défaut et de réutilisation libre et gratuite.

Exemples d'exceptions permettant le non partage des données : Les données uniquement communicables à l'intéressé (vie privée, secret médical), secret des affaires, les documents réalisés en exécution d'un contrat de prestation de services exécuté pour le compte d'une ou plusieurs personnes déterminées, les données couvertes par des droits de propriété intellectuelle, etc.

Voir notamment : Article L. 533-4 du Code de la recherche, article 6 de la Loi pour une république numérique, articles L. 311 s, L.312, L. 322-1 du CRPA.

Voir aussi, le guide d'application de la loi pour une république numérique pour les données de la recherche.

RGPD : Si des données personnelles ou sensibles, sont impliquées dans le processus de recherche, il devient obligatoire de contacter le référent **RGPD** (Règlement général sur la protection des données) de l'université de Lille à l'adresse suivante : **dpo@univ-lille.fr**.

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.

Recommandations:

Sélection des données : Il est recommandé de conserver les données de recherche pendant 15 ans, sauf s'il s'agit de données personnelles. Dans ce dernier cas, la limite de conservation est de 2 ans après la dernière publication. Passé cette période il est impératif de détruire les données identifiantes, ou de les anonymiser.

Exemple : « Les données personnelles seront détruites à l'issue du projet, le reste des données sera conservé pendant 15 ans sur les serveurs de l'université ».

Il est en effet recommandé de conserver les données de recherche pendant 15 ans, sauf s'il s'agit de données personnelles. Dans ce dernier cas, la limite de conservation est de 2 ans après la dernière publication. Passé cette période il est impératif de détruire les données identifiantes, ou de les anonymiser.

Réutilisation : Il est demandé de montrer que la question de la réutilisation des données a été réfléchie. Il faut donc déterminer dans quel contexte vos données pourraient être utiles à d'autres chercheurs dans le futur, que ce soit dans votre discipline ou une autre.

Recherche Data Gouv : En dehors d'une option spécialisée, conseillée par le laboratoire ou par la discipline, il est recommandé de sélectionner un entrepôt pluridisciplinaire tel que Recherche Data Gouv . Il s'agit d'une solution gratuite, fiable et nationale, qui permet une gestion des métadonnées standardisée et compatible avec toutes les disciplines.

Archivage : L'archivage pérenne (plus de 25 ans) est un processus précis et complexe nécessitant un savoir faire spécifique. Pour toute question, contacter le service des archives de l'Université.

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.

Recommandations:

Logiciel(s) : Indiquer les logiciels utilisés pour créer et traiter les données, ils sont nécessaires à leur réutilisation (mentionner si des alternatives libres existent). Ils garantiront l'accès et l'intelligibilité des données.

Code(s) source(s) : Mentionner si un logiciel a été développé dans le cadre du projet. Préciser si les codes sources et logiciels, avec leur documentation, seront conservés et partagés. Si oui, expliquer comment. Utiliser l'entrepôt Software Heritage pour conserver et partager les codes sources.

Plan de gestion de logiciel : Préciser si le logiciel et le code source développés au cours du projet seront documentés dans un plan de gestion de logiciel.

Un plan de gestion de logiciel décrit les bonnes pratiques mises en place lors du développement d'un logiciel. Le Plan de gestion de logiciel et le plan de gestion des données doivent se référencer mutuellement. Des modèles sont disponibles dans DMPOPIDoR, comme le modèle de projet PRESOFT et celui du SoftwareSustainability Institute.

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.

Recommandations:

Les identifiants : La plupart des entrepôts assignent un identifiant pérenne aux jeux de données. Cette option est à vérifier dans la politique de l'entrepôt sélectionné.

Lorsque les données sont mises à disposition sur une base de données qui ne dispose pas d'une attribution de DOI automatisée, l'université de Lille dispose des ressources nécessaires à la mise en place d'une attribution automatique des DOI au sein d'une base.

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.

Recommandations:

Il faut indiquer pour l'ensemble des membres du projet jouant un rôle dans la gestion des données de recherche les nom, prénom, contact et rôle (collecte, traitement, stockage, rédaction du PGD...).

Pour les projets menés en collaboration, expliquer comment s'effectue la coordination des responsabilités de gestion des données entre partenaires.

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.

Recommandations:

Données FAIR : pour en apprendre plus sur ce que sont les données FAIR, nos pages dédiées sur le site du SCD de l'Université de Lille peuvent servir de références.

Coûts : L'évaluation des coûts humains et financiers fait partie des demandes de l'ANR. Aussi, les coûts liés à la gestion des données peuvent être éligibles à un remboursement selon les agences de financement.

Ne pas hésiter à mentionner les équipes de l'Université de Lille : **SCD, DPO, DGDNum, archivistes**, etc. pour manifester les ressources humaines mises à disposition du projet.

La plupart des services de l'université ainsi que la majorité des entrepôts de confiance sont gratuits. Ne pas hésiter à le mentionner pour montrer que la question a été réfléchie. Cependant, ne pas sous estimer le facteur temps que peut représenter ces actions.