

Université de Picardie Jules Verne: UPJV - Plan de gestion des données

A. Description des données collectées ou réutilisées

1. Quelles sont les données que vous allez produire ou collecter ? (*Exemples de type de données : tableur, base de données, images, visée, vidéo, génome, spectre, phénotype, nuage de point, liste, biblio...*) Expliquez votre méthodologie de recueil de données. Comment allez-vous réutiliser des données préexistantes, comment vous seront-elles transmises ?

Recommandations:

Si vous récupérez des données à caractère personnel ou industriel appartenant à un partenaire académique ou industriel, veillez à mettre en place une convention de transfert de données dans laquelle les conditions de réception, d'utilisation et de stockage seront explicitées.

2. Expliquez quels sont les logiciels que vous allez utiliser et sous quel format vos données seront enregistrées.

Recommandations:

Il est important que vous justifiiez le format d'enregistrement. Nous vous recommandons d'utiliser des formats ouverts (libres ou propriétaires comme le PDF) permettant la bonne diffusion de vos données entre partenaires.

3. Où allez-vous stocker vos données ? Comment allez-vous échanger vos données avec les partenaires pendant la durée du projet ? Pensez à expliciter le lien avec les objectifs du projet. Énoncer les éventuelles restrictions à la réutilisation de données préexistantes.

Recommandations:

Il est impératif de définir des limites pour les données personnelles et de les stocker sur un serveur sécurisé. Pour les données physiques : bureau sécurisé, armoire forte, ZRR...

B. Documentation et Métadonnées

1. Quelles métadonnées et quelle documentation (*par exemple : méthodologie de collecte et mode d'organisation des données*) accompagneront les données que vous allez collecter ?

2. Indiquez comment les données seront organisées au cours du projet, en mentionnant par exemple les conventions de nommage (*document interne au projet en accord avec les partenaires spécifiant les formats et la manière de nommer les documents*), le contrôle de version et les structures des dossiers.

Recommandations:

Pensez 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.

3. Allez-vous utiliser un cahier de laboratoire ?

Recommandations:

Des cahiers de laboratoires sont disponibles au secrétariat de la Direction de la Recherche.

4. Précisez comment vous allez veiller à la bonne utilisation des données en termes d'enregistrement et d'échange.

Recommandations:

Par exemple, lors des réunions de projets, il est préférable d'identifier un responsable « données ». Selon la caractéristique du projet (en cas de transdisciplinarité notamment) vous pouvez identifier un responsable des données par partenaire.

5. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ? En particulier, 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.

C. Stockage des données

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

Recommandations:

N'oubliez pas de vous rapprocher du délégué à la protection des données de votre établissement.

2. Expliquez qui aura accès aux données au cours du processus de recherche et comment l'accès aux données sera contrôlé, en particulier dans le cadre de recherches menées en collaboration. Précisez où seront stockées vos données.

Recommandations:

Évitez les échanges type Google drive, privilégiez les services de Cloud universitaire et entreprise .

3. Avez-vous pris en compte la protection des données, en particulier si vos données sont sensibles (*par exemple données à caractère personnel, données politiquement sensibles, informations commerciales, secrets industriels...*) ? Décrivez les principaux risques et la façon dont ils seront gérés. Veillez à réaliser en amont votre analyse d'impact* (pour les données personnelles).

4. Expliquez comment vous appliquez le RGPD (si pertinent).

D. Éthique et propriété intellectuelle

1. Préciser si votre projet a fait l'objet d'un examen en CERNI ou en Comité de Protection des Personnes (CPP)

1.a. Si vous collectez des données personnelles : expliquez votre méthodologie en reprenant par exemple le dossier CERNI. Si vous n'avez pas sollicité l'avis du CERNI, merci de préciser si ces aspects vont être abordés : consentement éclairé pour la préservation et/ou partage de données personnelles

1.b. Envisagez-vous l'anonymisation des données personnelles pour la préservation et/ou le partage (des données correctement anonymisées ne sont plus considérées comme des données personnelles) ?

1.c. Allez-vous envisager la pseudonymisation des données personnelles (la principale différence avec l'anonymisation est que la pseudonymisation est réversible) ?

1.d. Envisagez-vous le chiffrement des données (considéré comme un cas particulier de pseudonymisation - la clé de cryptage doit alors être stockée séparément des données, par exemple chez un tiers de confiance).

1.e. Expliquez si une procédure d'accès spécifique a été mise en place pour les utilisateurs autorisés à accéder aux données personnelles.

2. Espérez-vous obtenir des résultats brevetables ? Comment allez-vous gérer la confidentialité des données et des résultats entre les partenaires ? Comment allez-vous gérer la répartition de la propriété intellectuelle entre les partenaires ? Y a-t-il des conditions législatives particulières à votre thématique de recherche ?

2.a. Expliquez quelles conditions d'accès s'appliqueront aux données. Qui sera le propriétaire des données, qui aura le droit d'en contrôler l'accès ? Les données seront-elles librement accessibles, ou des restrictions seront-elles appliquées ? Si oui, lesquelles ? Envisagez-vous l'utilisation de licences concernant l'accès et la réutilisation des données ? (Creative Commons)

Recommandations:

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.

2.b. Indiquez si les droits de propriété intellectuelle (par exemple la directive bases de données, droits sui generis) sont affectés. Si oui, expliquez lesquels et comment cela sera-t-il traité.

2.c. Indiquez s'il y a des restrictions sur la réutilisation des données fournies par des tiers.

E. Partage des données et stockage à long terme

1. Quand et comment les données seront-elles partagées ? Y a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

Recommandations:

Pensez à 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).

2. Définissez un plan de préservation des données. Quelle sera la durée d'archivage pérenne des données ?

3. Expliquez à quel moment les données seront rendues disponibles. Indiquer les délais de publication prévus. Expliquez si une utilisation exclusive des données est revendiquée et, dans l'affirmative, pour quelle raison et pour combien de temps. Indiquez si le partage des données sera différé ou limité, par exemple pour des raisons de publication, pour protéger la propriété intellectuelle, ou en vue d'un dépôt de brevets.

4. Indiquez qui pourra utiliser les données. S'il s'avère nécessaire de restreindre l'accès pour certaines communautés (industriels ou tiers extérieurs au projet par exemple) ou d'imposer un accord pour le partage de données, expliquez comment et pourquoi. Expliquez les mesures qui seront prises pour dépasser ou minimiser ces restrictions.

5. 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:

Pensez à indiquer quelles données ne doivent pas être divulguées ou doivent être détruites pour des raisons contractuelles, légales, ou réglementaires.

6. Indiquez comment il sera décidé quelles données garder. Décrire les données qui seront à préserver à long terme, c'est-à-dire sur une durée supérieure à 5 ans.

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

8. Indiquez 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 normes de métadonnées, et coûts mis en œuvre) ont été vérifiées.

9. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder aux données et les utiliser sur le long terme ?

Recommandations:

Essayez de trouver un format qui pourra être toujours lisible sur le long terme

F. Responsabilités et ressources en matière de gestion des données

1. Qui (rôle, position et institution de rattachement) sera responsable de la gestion des données dans le cadre de votre projet (c'est-à-dire le gestionnaire des données) ? Nommez les personne(s) responsable(s) impliquées.

Recommandations:

Décrivez les rôles et les responsabilités concernant les activités de gestion des données : saisie de données, production de métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données.

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

3. Indiquez qui est responsable de la mise en œuvre du PGD, et qui s'assure qu'il est examiné et, si nécessaire, révisé. Envisagez-vous des mises à jour régulières du PGD ? (C'est plutôt un NB qu'un point de questionnement)

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

Recommandations:

Expliquez comment les ressources nécessaires (par exemple, le temps) à la préparation des données pour le partage/préservation (curation des données) ont été chiffrées. Examinez et justifiez soigneusement toutes les ressources nécessaires pour distribuer 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. Pensez à 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-en le montant et indiquez comment ces coûts seront couverts.

Université de Picardie Jules Verne: UPJV - Lexique

Analyse d'impact relative à la protection des données (AIPD) :

L'AIPD est un outil qui permet de construire un traitement conforme au RGPD et respectueux de la vie privée. Elle concerne les traitements de données personnelles qui sont susceptibles d'engendrer un risque élevé pour les droits et libertés des personnes concernées.

Archives ouvertes (ou archives OAI / archives en libre accès / archives électroniques ouvertes (open archive)) :

Désigne un espace de stockage numérique disciplinaire. Selon la politique éditoriale de l'archive ouverte, le document qui y est déposé peut être :

- Une publication à l'état de projet, par exemple un document non encore soumis à une revue ;
- Une publication soumise à un éditeur, par exemple un manuscrit d'article dans sa première version avant relecture par les pairs (on parle de preprint) ;
- Une publication évaluée par les pairs et corrigée, par exemple un article accepté pour publication dans une revue et corrigé par l'auteur pour répondre aux demandes des relecteurs (on parle de postprint) ;
- Une publication publiée dans sa forme finale, par exemple un article accepté par une revue, relu par les pairs, corrigé par l'auteur, et mis en forme par l'éditeur de la revue pour être publié (on parle de version éditeur ou version PDF) ;
- Tout autre document, tel qu'un jeu de données correctement décrit, un rapport de recherche, un document de travail, une fiche, une présentation de congrès, une thèse, un ouvrage, un chapitre d'ouvrage, etc.

Centre de données (en anglais Data Center) :

Infrastructure immobilière et technique qui sert à stocker des informations de manière centralisée. Cette infrastructure héberge d'importants équipements informatiques (serveurs) dans des salles techniques spécifiques et sécurisées (alimentation électrique, refroidissement, accès au réseau). Les data center sont un groupement de serveurs eux-mêmes stockant des données

Comité d'Éthique sur les Recherches Non Interventionnelles (CERNI) :

Le CERNI a pour vocation de fournir un avis sur des protocoles de recherche non interventionnelles, soit toutes les recherches qui concernent directement ou indirectement la personne humaine et qui ne sont ni à visée thérapeutique ni impliquant des moyens physiquement invasifs. Le CERNI est une instance pluridisciplinaire et pluraliste composée de chercheurs, enseignants, personnels de l'UPJV ainsi que de personnes choisies pour leurs compétences et leur intérêt pour les questions éthiques. Le CERNI examine et rend des décisions contraignantes sur les aspects éthiques, juridiques et déontologiques concernant les projets de recherche.

Recommandations:

Plus d'informations sur <https://www.u-picardie.fr/cerni/>.

Data Paper :

Article scientifique examiné par les pairs (peer-reviewing) décrivant les données, leurs méthodes d'obtention et leur potentiel pour de futures recherches. Le data paper (data article, data descriptor) est une publication qui décrit un jeu de données scientifiques, notamment à l'aide d'informations structurées appelées métadonnées. Contrairement aux articles de recherches classiques, les data papers fournissent une voie formalisée au partage des données plutôt que tester des hypothèses ou présenter de nouvelles analyses.

Recommandations:

Pourquoi publier un data paper ?

- Pour communiquer sur l'existence des données et permettre de les trouver ;
- Pour créditer les auteurs (reconnaissance, référence citable) et valoriser les données ;
- Pour faciliter la réutilisation des données (en les rendant intelligibles).

Où publier un data paper ?

- Dans un data journal, revue dédiée à ce type de publication ;
- Dans une revue classique qui publie des data papers en plus des articles classiques.

Que contient un data paper ? Les data papers ont une structure très variable selon les revues. Ils ont cependant des [composantes communes](#). Voir <https://doranum.fr/data-paper-data-journal/fiche-synthetique/>

Délégué à la protection des données (en anglais Data Protection Officer (DPO)) :

Ses missions sont définies dans les articles 38 et 39 du Règlement général sur la protection des données. Le DPO doit s'assurer et contrôler que le RGPD (Règlement général sur la protection des données)est bien respecté dans l'organisation (publique ou privée) où il exerce. Il conseille de manière indépendante le responsable du traitement des données et l'informe quant aux obligations en matière de protection des données personnelles. Enfin, il gère les interactions avec la CNIL (ou toute autre autorité de contrôle).

Données de la recherche (ou Données scientifiques ou Research data) :

Ensemble des informations collectées, observées ou créées sous une forme numérique dans le cadre d'un projet de recherche. Exemples de données :

- Données d'observation (capturées en temps réel) : relevés météo, images, enquêtes sociales, fouilles archéologiques (dessin), témoignages

oraux, plan...

· Données expérimentales (obtenues à partir d'équipement et de manipulation en laboratoire) : séquence peptide, poids, biomasse, analyses moléculaires...

· Données de simulations informatiques : modèle climatique, modèle économique, ...

· Données dérivées ou compilées (issues de traitement de données brutes) : bases de données compilées, fouille de textes, tableaux ou matrices, par exemple : la population des villes médiévales françaises etc...

· Considérez tous vos actifs de recherche comme des données de recherche qui pourraient être potentiellement réutilisées par d'autres chercheurs et leur être utiles, afin de rédiger par un exemple l'état de l'art de leur recherche

Entrepôt de données (ou Data Share House ou Data Warehouse) :

Base de données rassemblant une grande quantité d'informations issues de différentes applications de production avec des formats divers (Excel, XML...).

Exploration de données (ou Data Mining) :

Exploration et extraction de tendances à partir de larges quantités de données par des méthodes automatiques ou semi-automatiques fonctionnant au moyen d'algorithmes. Cette exploration permet ainsi de construire des modèles scientifiques. Cette fouille peut s'appliquer aussi à du texte : on parlera alors de Text and Data Mining ou TdM.

Gestionnaire de données (ou Data Manager) :

Désigne la personne qui recueille et organise les informations de l'entité publique ou privée dont il est l'employé en vue de leur exploitation optimale. Il crée et administre les bases de données. Il a généralement des compétences en statistiques et/ou informatique.

Jeu de données (ou Dataset) :

« Agrégation (...) de données brutes ou dérivées présentant une certaine 'unité', rassemblées pour former un ensemble cohérent » d'après la définition donnée par l'INIST. Un jeu de données peut avoir une structure tabulaire (CSV type Excel), une structure d'arbre (type JSON ou XML) une structure de graphe (type RDF).

Identifiant persistant (PID) :

Un identifiant pérenne (Persistent identifier ou PID) est un code alphanumérique associé à un objet ou ressource de façon permanente. Il est disponible et gérable à long terme ; il ne changera pas si l'objet est renommé ou déplacé (changement de site, d'entrepôts de données...). Cet identifiant persistant (PID) constitue une référence durable à une ressource numérique, contrairement aux URL, qui peuvent casser. Un identifiant ORCID est un exemple d'identifiant persistant pour une personne. ORCID travaille en étroite collaboration avec de nombreuses autres organisations PID pour établir des liens de confiance entre ORCID ID et autres identifiants. Les DOI (identificateurs d'objets numériques), ont des identificateurs persistants pour des choses ou des entités telles que des articles de revues, des livres et des ensembles de données. CrossRef et DataCite sont des exemples de registre attribuant des DOI pour la communication scientifique. ORCID prend en charge les DOI et de nombreux autres PID qui sont utilisés pour identifier les résultats scientifiques.

Libre accès :

Par « accès libre », il faut entendre mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et l'utilisation d'Internet. La seule contrainte sur la reproduction et la distribution, et le seul rôle du copyright dans ce domaine devraient être de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être correctement reconnus et cités via une licence creative commons.

Licence CREATIVE COMMONS :

Toutes les licences Creative Commons ont de nombreuses caractéristiques importantes communes. Chaque licence Creative Commons assure également aux titulaires de droits qu'ils seront crédités et cités de la manière dont ils le souhaitent. Toutes les licences Creative Commons fonctionnent dans le monde entier et s'appliqueront aussi longtemps que dureront les droits d'auteur (parce qu'elles sont fondées sur les droits d'auteur). Ces caractéristiques communes servent de point de référence, à partir duquel les titulaires de droits peuvent choisir d'accorder des permissions complémentaires quand ils décident de la manière dont ils souhaitent que leur œuvre soit utilisée.

Métadonnées :

En termes simples, les métadonnées sont des « données sur les données », ou des informations utilisées pour identifier et décrire les données, elles sont l'un des éléments clés de la pratique FAIR. Selon les principes FAIR, l'identifiant persistant (PID) et « des métadonnées suffisamment riches » sont suffisants pour permettre à vos données d'être trouvées, utilisées et citées de manière fiable. Les métadonnées sont inhérentes aux documents que vous avez créés comme celles présentes dans un document texte, dans une photographie, ou d'image produite par un microscope avec les données constructeur. Votre travail sera de les enrichir avec des mots clés pour qu'elles puissent être trouvées par une simple recherche. Pour cela, vous devez créer vos métadonnées en utilisant un vocabulaire contrôlé.

Plan de gestion de données (PGD) (en anglais Data Plan Management (DMP)) :

Document formel explicitant la façon dont vous obtenez, documentez, analysez et utilisez vos données à la fois au cours de votre recherche et une fois le projet terminé.

Il décrit dans le détail les méthodes et processus de création, de fourniture, de maintenance, de conservation et de protection des données. (Définition Cirad)

Voie dorée (ou golden road, gold road, gold OA, journal-mediated OA) :

S'applique à la publication d'articles dans des revues en libre accès, quel que soit leur mode de financement. Elle correspond à la deuxième stratégie recommandée dans l'Initiative de Budapest pour l'Accès Ouvert : « Revues alternatives : en second lieu, les savants ont besoin des moyens pour lancer une nouvelle génération de revues alternatives engagées dans le libre accès et pour aider les revues existantes qui choisissent d'opérer la transition vers l'accès libre. »

Voie verte (ou green road, green OA, green road to OA, repository-mediated OA) :

Auto-archivage par les chercheurs ou archivage par une tierce personne des articles dans des archives ouvertes. Elle correspond à la première stratégie préconisée dans l'Initiative de Budapest pour l'Accès Ouvert : « Auto-archivage : en premier lieu, les savants ont besoin d'outils et d'assistance pour déposer leurs articles de revues à comité de lecture dans des archives électroniques ouvertes, une pratique communément appelée auto-archivage »