

GUIDE POUR L'ÉLABORATION DU PLAN DE GESTION DES DONNÉES (PGD) À DESTINATION DES ÉQUIPES LAURÉATES DU PEPR PREZODE

L'objectif de l'action *data* au sein du **PEPR PREZODE** est de garantir que les données de vos projets lauréats s'inscrivent dans les principes FAIR. Idéalement, à la fin de votre étude, vos objets de recherche (données, codes sources, publications, etc.), devraient être faciles à trouver, interopérables, réutilisables et accessibles.

Pour nous aider à atteindre cet objectif, le **Plan de Gestion des Données (PGD)** joue un rôle essentiel, car il nous permet d'anticiper une planification efficace dès le début de notre gestion des données. Il ne doit pas être perçu comme une simple formalité administrative, mais plutôt comme « un carnet de bord » et précisément comme un outil essentiel permettant d'anticiper la gestion des données dès le début du projet afin d'assurer leur FAIRisation. L'essentiel est que vous mettiez en place de bonnes pratiques de gestion des données, ce qui facilitera cette FAIRisation.

Le PGD n'est pas difficile à remplir, mais il est crucial **d'impliquer tous les collègues travaillant sur les données du projet** afin d'obtenir un document aussi complet et pertinent que possible. Il doit être vu comme une invitation à **collaborer en équipe** autour de la gestion des données.

Peut-être qu'au début du projet, vous n'aurez pas toutes les réponses pour compléter votre PGD, et ce n'est pas un problème : le PGD peut être complété progressivement tout au long de la durée du projet. L'important est de se sensibiliser à ces enjeux, et pour cela, nous avons toujours des solutions à disposition.

Basé sur le modèle fourni par **DMP Opidor**, ce document vous propose des recommandations pour remplir votre PGD. Il inclut des suggestions du **PEPR Data Working Group PREZODE**, qui a préparé ce guide afin de vous aider dans son élaboration.

COMMENT LIRE CE DOCUMENT ? Chaque question du PGD est décrite afin que vous puissiez y répondre en fonction des produits de recherche générés par votre projet. Vous devez définir combien de **produits de recherche** seront produits – ce ne sont pas toutes les données du projet qui doivent être considérées comme des produits de recherche. Sélectionnez les plus pertinents, ceux qui auront un impact direct sur les résultats de votre projet. Il convient de noter que l'ensemble des termes en *italique* possèdent une définition, disponible au glossaire à la fin du document.

SOUHAITEZ-VOUS APPORTER VOTRE EXPERIENCE ? Vous pouvez nous aider à améliorer ce document et rejoindre cette communauté œuvrant à l'amélioration de la gestion des données. Si des points restent à éclaircir, si vous souhaitez partager des exemples pratiques ou si vous souhaitez simplement partager votre satisfaction concernant ce document, n'hésitez pas à nous contacter à l'adresse paula.caceres@inrae.fr. Pour l'équipe DWG, votre feedback est important.

UN DOCUMENT EN CONSTANTE EVOLUTION : Au fur et à mesure de la réception des commentaires, le document sera mis à jour en fonction de vos besoins. Les différentes versions seront identifiées, vous permettant ainsi de savoir quelle version est la plus récente.

QUESTIONS

1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

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

Chaque projet lauréat du PEPR PREZODE aura une administration indépendante pour la gestion de ses données. Cependant, il est recommandé que :

- Les jeux des données produites dans chaque projet suivent les principes définis dans la Charte de principes pour les projets lauréats PEPR PREZODE, la CHARTE, incluant la conformité aux standards FAIR (Facile à trouver, Accessibles, Interopérables, Réutilisables).
- La collecte et la production des données doivent être documentées de manière détaillée pour assurer leur traçabilité et leur qualité, en décrivant les *métadonnées* associées dès les premières étapes du cycle de vie des données.
- Un Plan de Gestion de Données (PGD) doit être élaboré, le plus tôt possible, pour chaque projet afin de planifier et de structurer la gestion des données.
- Bien que chaque projet conserve son autonomie, la gouvernance des données doit être guidée par les bonnes pratiques établies dans la CHARTE.
- Veuillez noter que les données produites doivent être accompagnées de **licences ouvertes**, comme la licence CC BY 4.0 ou ETALAB 2.0, pour garantir leur réutilisation tout en respectant les restrictions légales définies par chaque consortium.

- En 2025, une fois sélectionnés tous les projets lauréats, une formation spécifique sera coordonnée et proposée par l'équipe du groupe de travail sur les données du PEPR PREZODE afin de renforcer les capacités des équipes à gérer leurs données conformément aux principes FAIR, à ouvrir leurs données et à créer leur PGD. Ce travail sera effectué une fois que tous les projets auront été sélectionnés.
- En cas de questions spécifiques, vous pouvez contacter Paula Caceres à l'adresse paula.caceres@inrae.fr.

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

Les projets lauréats du PEPR PREZODE produiront divers types de données. Chaque projet décrira, dans cette section 1b, **de manière générique**, ses jeux de données, sans entrer dans le détail de chaque donnée individuelle. Cette description inclura :

- Les types de données (*collectées ou produites*).
- Les formats (*ouverts et standardisés*).
- Les volumes estimés.
- Les standards utilisés pour décrire les métadonnées.

Par exemple, si un produit de recherche correspond à des modèles ou des codes, ceux-ci seront accompagnés de licences appropriées et disponibles dans des formats compatibles. En outre, chaque projet ou consortium devra documenter et estimer ses besoins en termes de stockage et d'organisation, tout en précisant si les données soulèvent des enjeux juridiques ou éthiques spécifiques.

L'équipe Data Working Group de PEPR PREZODE invitera chaque consortium à identifier les familles de données qu'il produira, afin d'établir une cartographie globale des données de l'ensemble du projet PEPR PREZODE. Cette démarche permettra non seulement d'identifier des opportunités de mutualisation entre les équipes, mais aussi de valoriser les résultats de recherche. Les métadonnées associées à cette cartographie pourront être valorisées dans un futur portail PREZODE ou dans la plateforme internationale que PREZODE prévoit de développer. Les équipes lauréates s'engageront à partager leurs métadonnées pour soutenir cette cartographie.

2. DOCUMENTATION ET QUALITE DES DONNEES

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

Il est recommandé que toutes les *métadonnées*, selon leur nature, suivent des *standards* reconnus pour garantir leur interopérabilité et leur réutilisation, en conformité avec les principes FAIR.

Le *Data Working Group* (DWG) PEPR PREZODE recommande l'utilisation du [RDA Metadata Standards Catalog](#), une plateforme collaborative et ouverte qui propose un répertoire exhaustif de *standards* de métadonnées applicables aux données de recherche. Cette ressource permet de rechercher et de sélectionner les *standards* les plus adaptés à différents types de données et contextes de recherche.

En complément, chaque projet devra :

- Documenter clairement la méthodologie utilisée pour la collecte et l'organisation des données.
- Fournir des *métadonnées* détaillées, incluant des éléments tels que :
 - Titre
 - Description
 - Mots-clés
 - Format
 - Provenance
 - Conditions d'accès
 - Droits d'utilisation
- Identifier les *standards* spécifiques adoptés en fonction des types de données produites.

Ces recommandations visent à assurer une gestion harmonisée et standardisée des métadonnées dans tous les projets lauréats PEPR PREZODE, facilitant leur identification, leur accessibilité et leur valorisation à long terme.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Les instituts de recherche associés aux consortiums de chaque projet lauréat PEPR PREZODE **sont responsables de la qualité des données**, y compris de la définition des protocoles et de leur mise en œuvre. De nombreux projets PEPR PREZODE bénéficieront du soutien d'un gestionnaire de données qui accompagnera la gestion des données tout au long du projet (cf. cycle de vie des données), garantissant ainsi leur qualité.

Au-delà de ces mesures institutionnelles et de ces protocoles, le DWG PEPR PREZODE propose de mettre en œuvre des actions spécifiques axées sur la qualité, la transparence et la traçabilité des données :

1. Mise en œuvre de processus de validation des données depuis leur collecte : les données doivent être vérifiées dès leur collecte pour garantir leur conformité avec les définitions, protocoles et standards adoptés par chaque projet. En s'assurant que les processus de collecte respectent les standards définis dans le PGD de chaque projet. Cela inclut une documentation claire sur les outils utilisés, les méthodes d'acquisition, et les conditions dans lesquelles les données ont été générées.

Pour limiter les erreurs systématiques, il est recommandé d'adopter d'outils automatisés pour détecter les incohérences des données, de champs obligatoires dans les questionnaires, ou de listes déroulantes pour standardiser les réponses.

2. Documentation, provenance et traçabilité des données : comme mentionné dans la section 2a, chaque *jeu de données* doit être accompagné de *métadonnées* détaillées. Parallèlement, il convient de noter le type de documentation existante afin de garantir la *provenance* et la *traçabilité* des données à

l'avenir. Sans cette *provenance*, les utilisateurs potentiels des données seront réticents à les exploiter, compromettant ainsi leur valeur scientifique. La *provenance* des données ne se limite pas à l'ensemble des informations permettant de retracer leur origine, mais englobe également le type de vocabulaire, les ontologies à utiliser et leur identification unique par des *identifiants pérennes*. Garantir une provenance rigoureuse est le premier pas vers une traçabilité complète.

3. **Harmonisation et standardisation des données** : l'application de formats ouverts et de standards internationaux reconnus pour les données et métadonnées (comme recommandé dans la section 2a) est essentielle pour faciliter leur comparabilité, leur interopérabilité, et leur réutilisation.

4. **Revue par des pairs et collaboration avec le *Data Sharing Committee (DSC)***: le PEPR DWG PREZODE, en collaboration avec le PEPR MIE, mettra en place un comité international composé d'experts en gestion des données. Ce comité fournira un soutien aux lauréats pour résoudre divers défis liés à l'accessibilité et à la FAIRisation des données (TBC).

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 ?

Chaque consortium lauréat de projet est responsable de définir et de mettre en œuvre les mécanismes nécessaires pour le stockage et la sauvegarde des données et des métadonnées tout au long du processus de recherche. Ces mécanismes doivent répondre aux critères suivants :

- **Sécurité** : les données doivent être stockées dans des environnements sécurisés qui garantissent leur intégrité et leur protection contre les accès non autorisés. Les solutions choisies doivent inclure des options de sauvegarde régulières pour minimiser les risques de perte de données.
- **Facilité de collaboration** : les outils de stockage doivent permettre un échange efficace et sécurisé des données entre les membres du consortium, en respectant les normes de confidentialité et les exigences éthiques.
- **Documentation** : les consortia sont encouragés à documenter clairement leurs processus de *stockage*.
- **Transition vers les dépôts de confiance** : à la fin du projet, les données doivent être transférées vers des *dépôts de confiance* reconnus comme recommandé dans la *CHARTRE*. Ces dépôts garantissent l'attribution d'identifiants uniques (DOI) et offrent des fonctionnalités FAIR, facilitant ainsi la réutilisation des données par la communauté scientifique.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Chaque projet lauréat du PEPR PREZODE devra définir et mettre en œuvre des mesures adaptées à la nature des données sensibles qu'il pourrait traiter. Exemples de données sensibles :

- **Données personnelles** : Informations identifiables comme noms, adresses email, affiliations institutionnelles (RGPD).
- **Informations médicales des patients.**
- **Secrets industriels** et données confidentielles liées aux contrats de recherche.
- **Données réglementées** : Conformité avec les cadres législatifs spécifiques (sécurité biologique, sécurité publique, sûreté des personnes).
- **Données stratégiques** : Données pouvant impacter la sécurité nationale, la cyber sécurité ou les intérêts techniques/scientifiques avant publication.

RECOMMANDATIONS GENERALES POUR LA SECURITE DES DONNEES SENSIBLES :

1. Stockage sécurisé :

- Utiliser des systèmes de stockage sécurisés avec des accès restreints pour limiter les risques de fuite ou d'accès non autorisé.
- Adapter le niveau de sécurité aux types de données (e.g., données médicales vs. données ouvertes).

2. Sauvegardes régulières :

- Effectuer au moins deux sauvegardes sur des supports sécurisés distincts.
- Éviter de conserver des données sensibles sur des ordinateurs personnels ou des dispositifs non sécurisés.

3. Sensibilisation des équipes :

- Former les équipes de recherche aux bonnes pratiques en matière de gestion des données sensibles, y compris la manipulation et le partage de ces données.

4. EXIGENCES LEGALES ET ETHIQUES, 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 ?

Le Data Working Group (DWG) PEPR PREZODE recommande que tous les projets respectent les dispositions du RGPD lorsqu'ils traitent des données à caractère personnel. Plus spécifiquement, les recommandations suivantes s'appliquent :

1. Consentement et information préalable :

- Les participants doivent être clairement informés de l'utilisation prévue de leurs données personnelles, incluant :
 - Le temps de conservation des données.
 - Les processus d'anonymisation ou de pseudonymisation appliqués.
- Leur consentement explicite doit être obtenu avant toute collecte de données personnelles, et ce consentement doit être enregistré de manière formelle.

2. Anonymisation et pseudonymisation :

- Si possible, appliquer une anonymisation systématique des données personnelles, en remplaçant les informations identifiables (telles que noms ou adresses email) par des codes uniques garantissant la confidentialité des participants.
- Considérer l'utilisation de règles de cryptage pour permettre une pseudonymisation réversible lorsque cela est nécessaire.

3. Partage des données anonymisées :

- Les données correctement anonymisées ne sont plus considérées comme des données personnelles selon le RGPD, ce qui facilite leur partage pour des études secondaires tout en respectant les règles de confidentialité.

4. Sécurité des systèmes de stockage :

- Les données personnelles doivent être stockées dans des systèmes conformes aux normes de sécurité, avec un accès strictement limité aux membres autorisés.
- Il est essentiel d'expliquer si une procédure spécifique a été mise en place pour réguler l'accès des utilisateurs autorisés à ces données.

5. En cas de doutes ou de situations spécifiques : il est recommandé de consulter un expert en protection des données, tel qu'un *Délégué Informatique et Libertés* (DIL) ou un référent RGPD, disponible normalement dans les instituts de recherche. Ces experts peuvent fournir des conseils adaptés aux cas particuliers, notamment pour :

- Identifier les **obligations légales** spécifiques liées au traitement des données personnelles.

- Valider les processus d'anonymisation ou de pseudonymisation mis en place.
- S'assurer de la conformité des procédures de stockage, partage et destruction des données sensibles.

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 ?

Le Data Working Group (DWG) PEPR PREZODE recommande que les équipes discutent en amont ces points, notamment lorsqu'il y a plusieurs instituts impliqués dans la conduite de la recherche. À cet égard, il est important de :

- Établir clairement dans **l'accord de consortium**, lorsqu'il existe, **la titularité ou les droits de propriété intellectuelle sur les données**, ainsi que les modalités de contrôle d'accès et les conditions de partage définies pour le projet. Cette étape doit être réalisée avant toute collecte de données, avec les services juridiques des partenaires. Il est à noter que, généralement, les données produites dans le cadre des projets appartiennent à l'institut ou aux institutions responsables de leur production.
- En l'absence d'accord de consortium, ces aspects doivent être définis dans le cadre du **Plan de Gestion des Données (PGD, point 4b)** et validés par les partenaires.
- Accompagner les données d'une **licence adaptée** (par exemple, CC-BY 4.0) définissant clairement les modalités d'accès et de réutilisation. Il est recommandé d'utiliser des licences ouvertes pour maximiser l'impact des données tout en respectant les restrictions éthiques, légales ou contractuelles.
- S'assurer que les **données fournies par des tiers respectent les accords ou droits préexistants**, afin de garantir leur utilisation conforme.
- Prendre en compte les **droits spécifiques** attachés à certains matériels, comme les échantillons humains nécessitant des formalités spécifiques, et les mentionner dans les documents juridiques associés au projet.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Le DWG PEPR PREZODE recommande les mesures suivantes pour garantir que les projets respectent les normes éthiques et déontologiques :

- Respect des codes déontologiques internationaux et institutionnels : Les projets doivent aligner leurs pratiques sur les principes éthiques reconnus, ainsi que les directives institutionnelles. Si nécessaire, elles doivent vérifier si leurs pratiques requièrent une revue ou une validation par un comité d'éthique, notamment pour les activités de collecte de données.
- Consentement éclairé et volontariat : Comme mentionné précédemment, il est essentiel d'informer les participants de manière transparente sur les objectifs du projet, l'utilisation prévue de leurs données et leurs droits. La participation doit toujours être volontaire, formalisée par un accord explicite, en veillant à ce que les participants soient pleinement protégés et ne subissent aucune pression pour participer.
- Supervision éthique continue : Toute question éthique soulevée dans le cadre d'un projet doit être discutée au sein du consortium concerné, avec le soutien des référents éthiques de chaque institut de recherche. Par ailleurs, le DWG PEPR PREZODE, en collaboration avec le PEPR MIE, mettra en place en 2025 un *Data Sharing Committee* composé d'experts internationaux, pouvant agir comme comité d'éthique pour apporter un soutien en cas de besoin.
- Prise en compte des questions éthiques dans la gestion des données : Les aspects éthiques doivent être intégrés dans la planification des données, notamment :
 - Comment elles seront stockées et transférées.
 - Qui pourra y accéder ou les utiliser.
 - Quelle durée de conservation leur sera appliquée, etc.
- Afin de garantir que la collecte, la gestion et le partage des données respectent non seulement les standards techniques internationaux, mais aussi les droits, les intérêts et la souveraineté des communautés concernées, nous recommandons de prendre en compte les principes **CARE** (*Collective*

Benefit, Authority to Control, Responsibility, Ethics). Ceux-ci mettent l'accent sur la collectivité, l'autorité de contrôle, la responsabilité et l'éthique, en soulignant particulièrement le respect des droits, de la souveraineté et des valeurs des peuples autochtones en matière de données. Pour plus d'informations, voir : [Doranum – Les principes CARE](#)

5. PARTAGE DES DONNEES ET CONSERVATION A 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 ?

Le Data Working Group (DWG) PEPR PREZODE recommande que cette section du PGD explique clairement comment les données produites par le projet, **une fois publiées**, pourront être retrouvées et partagées pour leur réutilisation.

Nous vous recommandons notamment :

- **Dépôt dans des [entrepôts](#) des données de confiance** : Les données produites dans le cadre des projets PEPR PREZODE doivent être déposées dans un **entrepôt de confiance thématique ou institutionnel** à l'issue du projet. Cela garantit leur accessibilité à long terme et leur conformité avec les principes FAIR.
- **Gestion des périodes d'exclusivité ou d'embargo** : Si une période d'exclusivité est nécessaire pour protéger les droits de publication ou d'exploitation des données, **elle doit être clairement définie dans cette section**. Par exemple, les données peuvent être déposées immédiatement dans un entrepôt public, mais leur accès peut être différé via un embargo prédéfini.
- **Archivage à long terme** : Cette section doit indiquer si les données seront archivées de manière pérenne. L'**archivage** à long terme consiste en un *cold storage* sécurisé, chiffré, et rarement consulté. Il est recommandé d'utiliser des [archives](#) publiques de confiance avec des financements stables à long terme, telles que le [CINES](#).
 - Une stratégie d'archivage pourrait inclure une révision systématique des jeux de données tous les 10 ans pour évaluer leur pertinence à être conservés. À ce stade, les données pourraient être transférées en *cold storage* avec un index permettant de les retrouver facilement si nécessaire.
- **Délais de publication et disponibilité** : Cette section doit préciser à quel moment les données seront rendues disponibles. Les délais de publication prévus doivent être clairement indiqués, notamment si des restrictions temporaires sont mises en place.
- **Modalités d'accès** : Les modalités d'accès aux données doivent être détaillées, en indiquant si elles sont libres d'accès ou soumises à des restrictions. Si un accès restreint est nécessaire (par exemple, pour des raisons éthiques, juridiques ou de sécurité), des accords spécifiques doivent être établis pour garantir un partage sécurisé.

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) ?

Le Data Working Group (DWG) PEPR PREZODE recommande :

- Indiquer dans cette section les **critères de sélection** des données destinées à un [entrepôt de données](#) ou à une archive. Par exemple, les données ayant une valeur scientifique, permettant la validation des résultats, ayant un potentiel de réutilisation pour d'autres recherches, présentant une valeur juridique ou patrimoniale, ou dont la perte pourrait compromettre la reproductibilité des résultats doivent être identifiées.
- Préciser dans quels [entrepôts de données de confiance](#) (thématiques ou institutionnels) ces données seront déposées.
- Il convient de noter que tous les jeux de données produits dans un projet **ne nécessitent pas une conservation à long terme**. Les **modalités d'accès aux données** doivent être définies en fonction des besoins des utilisateurs prévus.
- Décrire comment les autres données, jugées non pertinentes, seront **détruites** de manière sécurisée.
- Prendre en compte une **évaluation périodique** après la fin du projet et la décrire dans cette section. Cette évaluation permettra de réviser les [jeux de données](#) pour décider lesquels conserver ou transférer en archive à long terme. Les critères d'analyse peuvent inclure la question de savoir si les

données sont encore exploitables ou reproductibles, le coût de leur conservation, et si celui-ci est justifié par leur potentiel de réutilisation.

- Comme indiqué précédemment, toutes les données conservées ou archivées doivent être accompagnées de métadonnées complètes et enregistrées dans des formats pérennes afin de garantir leur accessibilité à moyen terme (10 ans) et à long terme (50 ans ou plus).
- Lorsqu'un partage public des données n'est pas possible, les restrictions spécifiques doivent être clairement justifiées (par exemple, confidentialité, sécurité ou restrictions légales).

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

Le Data Working Group (DWG) PEPR PREZODE recommande que cette section soit adaptée à la spécificité de chaque *jeu de données* produit dans le cadre des projets lauréats. Les modalités d'accès et les outils nécessaires dépendront du type et du format des données. Toutefois, tenez compte des éléments suivants:

- Accès aux données via un *entrepôt des données de confiance* (institutionnel ou thématique).
- Pour les données sensibles ou nécessitant des autorisations spécifiques, un **mécanisme de gestion des demandes d'accès** (par exemple, un formulaire en ligne) doit être mis en place pour garantir la sécurité et le respect des conditions d'utilisation.
- À noter que le PEPR PREZODE, ainsi que l'initiative internationale PREZODE, ont prévu de développer un portail, puis une plateforme internationale pour valoriser les données concernant la prévention des maladies zoonotiques issues des projets. Toutefois, si un projet prévoit de développer un portail dédié pour partager ses données, il doit s'assurer que ce portail est compatible avec les standards d'interopérabilité et qu'il restera accessible sur le long terme.
- Les utilisateurs potentiels doivent être informés des **outils ou logiciels nécessaires pour accéder aux données et les réutiliser** (par exemple, un logiciel spécifique pour lire des formats de données propriétaires ou encore un système d'information ou catalogue présentant certains types de données).
- Il est essentiel de **prendre en compte la pérennité des outils recommandés**. Lorsque des logiciels spécifiques sont nécessaires, des instructions claires doivent être fournies et il est recommandé d'utiliser des formats ouverts et standardisés afin de minimiser la dépendance aux logiciels propriétaires.
- Si un traitement supplémentaire des données est nécessaire, les scripts, codes ou modèles associés doivent être partagés avec les données et accompagnés de licences adaptées.
- **Les métadonnées documentant les formats et outils nécessaires doivent accompagner les données.** Cela inclut les spécifications sur les standards utilisés pour garantir leur interopérabilité dans différents environnements de recherche.
- Les projets doivent prévoir de **tester les outils recommandés** pour s'assurer de leur compatibilité avec les données fournies et de leur capacité à répondre aux besoins des utilisateurs.

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

Le Data Working Group (DWG) PEPR PREZODE recommande :

- L'attribution systématique d'identifiants uniques et pérennes (tels que le DOI) à chaque *jeu de données* jugé pertinent pour la réutilisation. Ces identifiants permettent de localiser et de référencer les données de manière fiable, facilitant ainsi leur citation et leur traçabilité dans des contextes scientifiques ou collaboratifs futurs.
- Privilégier l'utilisation d'entrepôts de confiance thématiques ou institutionnels, comme le portail Data INRAE, qui offrent des mécanismes intégrés pour générer et gérer ces identifiants.
- Inclure dans cette section du PGD une explication des raisons pour lesquelles certaines données pourraient ne pas être accessibles publiquement. Ces raisons peuvent inclure des restrictions éthiques, juridiques, contractuelles, ou des enjeux de confidentialité. Si des données ne sont pas partagées, il est crucial de documenter comment la validation des résultats sera garantie.

- Lorsque des données sont mises à disposition via un site web dédié, il est recommandé de prévoir leur stockage dans un entrepôt de confiance à l'issue du projet pour garantir leur pérennité et leur accessibilité à long terme.
- Dans le cas où une période d'embargo est nécessaire pour protéger les droits de publication ou d'exploitation scientifique des données, cette période doit être clairement définie et justifiée dans le PGD. Par exemple, les données déposées dans un entrepôt public ne seront accessibles qu'après une période spécifiée, permettant une dissémination différée tout en préservant les droits des producteurs.
- Les équipes doivent démontrer que l'utilisation des identifiants pérennes contribue activement à la réutilisation des données dans d'autres contextes scientifiques, renforçant ainsi leur impact et leur visibilité au sein de la communauté scientifique.

6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

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) ?

Nous vous recommandons, en plus d'indiquer le nom de la personne, sa position et l'institution pour laquelle elle travaille, de décrire ses responsabilités et d'inclure ses coordonnées, telles que son adresse e-mail ou tout autre moyen de contact.

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) ?

En plus d'indiquer de manière générale le budget prévu dans votre projet pour la gestion des données, vous pouvez préciser les ressources humaines et organisationnelles impliquées. Essayez de les quantifier, par exemple : XXXX, data steward travaillant à XXX et consacrant 20 % de son temps aux activités liées à la gestion des données du projet. Ces ressources humaines, combinées aux initiatives organisationnelles (comme la participation aux formations proposées par l'équipe PEPR DWG), peuvent être prises en compte pour montrer que tous ces efforts sont déployés pour garantir que les données produites dans le cadre de votre projet respectent les principes FAIR et qu'elles soient gérées efficacement tout au long du projet.

GLOSSAIRE

Archive : Destiné à une **conservation pérenne ou cold storage**, où les données sont sécurisées, chiffrées et rarement consultées. L'archive est utilisée pour préserver les données sur le long terme, avec une gestion adaptée aux formats et métadonnées pour garantir leur réutilisation future. Il est recommandé d'utiliser des archives reconnues comme le CINES, qui garantissent la préservation numérique face à l'obsolescence des formats et des logiciels.

Entrepôts des données : permet un accès facile et immédiat aux données pour une période déterminée. Ces données sont accessibles pour des usages actifs de recherche ou de collaboration.

Identifiants pérennes (Persistent Identifier) : chaîne de caractères alphanumériques qui désigne une ressource indépendamment de son emplacement et garantit un lien stable à la ressource en ligne, en faisant correspondre en permanence l'identité de la ressource à sa localisation sur le web. Ref. disponible [ici](#). Consulté le 7 fév. 2025.

Jeu de données (dataset) : Agrégation, sous une forme lisible, de données brutes ou dérivées présentant une certaine "unité", rassemblées pour former un ensemble cohérent. Toutefois, l'échelle à laquelle ces données assemblées acquièrent leur unité pour former un « jeu » varie selon les disciplines, les types de données, les projets, etc. Sous l'angle spécifique de l'« ouverture » des données de recherche, on peut définir le jeu de données comme un enregistrement de données sous la forme d'un ou plusieurs fichiers électroniques, téléchargeables, citables (notamment par l'intermédiaire d'un DOI) et intelligibles – ce jeu étant accompagné des métadonnées descriptives suffisantes. Ref. disponible [ici](#). Consulté le 7 fév. 2025.

Métadonnées : Ensemble d'informations structurées qui décrit, explicite, localise une ressource informationnelle, dans le but d'en faciliter la recherche, l'usage et la gestion. Ref. disponible [ici](#). Consulté le 7 fév. 2025.

Provenance des données : l'ensemble des informations permettant de retracer leur origine de la donnée : qui a produit les données, à quel moment, dans quel contexte, etc.

Standards de métadonnées (ou schéma de métadonnées) : est un modèle qui précise toutes les métadonnées nécessaires pour décrire un certain type de données. Utiliser un standard de métadonnées permet de décrire les données de façon riche et précise, en utilisant le même vocabulaire que votre communauté (**interopérabilité sémantique**). Exemples : Dublin Core, DDI (*Data Documentation Initiative*), EML (*Ecological Metadata Language*), etc. Ref. disponible [ici](#). Consulté le 7 fév. 2025.

Traçabilité : garantit un suivi transparent et documenté des données tout au long de leur cycle de vie, en retraçant les modifications, les méthodes utilisées, et les intervenants à chaque étape. Cela permet de reconstituer leur historique complet.