

## AGENCE NATIONALE DE LA RECHERCHE (ANR): ANR - MODELE DE PGD (FRANÇAIS)

### 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 ?

#### **Exemple 1 :**

*Les données, compte-rendus (.doc, .pdf) et fichiers de données brutes dument référencés (.xls, .jpg, .wav), seront essentiellement recueillis lors des expériences et des réunions qui auront lieu dans le cadre du projet XX. Les données seront échangées entre les partenaires sur un espace d'échange sécurisé de type cloud (« La Box du Laboratoire XLIM »). Toutes les données produites ou recueillies, consignées dans des rapports ou présentations, seront destinées à être utilisées uniquement dans le périmètre du consortium de MEAP et toute communication des données (publication scientifique par exemple) sera soumise à l'accord des membres du consortium. Il n'y a pas à proprement parler de données réutilisées.*

--

#### **Exemple 2 :**

*Les données sont collectées au travers de questionnaires envoyés aux instances compétentes. Les questionnaires sont divisés en plusieurs parties, visant à collecter des données stratégiques d'une part, et techniques d'autre part. Les données stratégiques seront collectées grâce à des questionnaires (questions à choix multiples, questions ouvertes ou fermées) envoyés à XXX et XXX. Les questions seront par exemple :*

- *Quels sont vos outils politiques pour appliquer le plan stratégique XXX ?*
- *Quels éléments sont pris en compte pour la mesure de XX dans le cadre du plan XX ?*

*Les données techniques seront collectées auprès de personnels en charge de la mise en œuvre des plans et politiques de territoire, comme les chargés de développement, et fourniront une base de bonnes pratiques et outils qui pourront être réutilisés. Les questions seront par exemple :*

- *Quels sont les coûts engendrés par la politique XX et dans sa mise en œuvre sur le territoire XX ?*
- *Quelles sont selon vous les bonnes pratiques à retenir ou encourager en termes de XX ?*

*Recommandations:*

- Expliquer quelles méthodologies ou quels logiciels seront utilisés si de nouvelles données sont recueillies ou produites.
- Enoncer les éventuelles restrictions à la réutilisation des données préexistantes.
- Expliquer comment la provenance des données sera documentée.
- Indiquer brièvement le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée mais écartée.

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

#### **Exemple 1 :**

*Les données sont numériques :*

- *Documentations, spécifications, rapport d'études et compte-rendu d'expérience : extensions .doc, .pdf.*
- *Fichiers de données expérimentales : tableurs (.xls), images (.tif, .jpg) et vidéos (.wav, .mp4)*

*Le volume de données n'est pas connu à l'avance. Il peut être de quelques ko à plusieurs Go, pour les vidéos notamment.*

—

### **Exemple 2 :**

*Les données sont principalement numériques :*

- *Tableaux incluant les réponses aux questionnaires, agrégés sous format .txt, .xls, .xlsx, .csv or .pdf*
- *Rapports en format .pdf*
- *Notes ou comptes rendus en .pdf*

### **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 de colonnes).

## **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 ?

### **Exemple 1 :**

*Les documents sont archivés sur un cloud partagé sécurisé avec la nomenclature suivante :*

- *Numero\_Nom\_Work\_Package/Numero\_Nom\_Task/Documentation*
- *Numero\_Nom\_Work\_Package/Numero\_Nom\_Task/Calendrier*
- *Numero\_Nom\_Work\_Package/Numero\_Nom\_Task/Fichiers\_Résultats*
- *Numero\_Nom\_Work\_Package/Numero\_Nom\_Task/Rapports*
- *Numero\_Nom\_Work\_Package/Numero\_Nom\_Task/Valorisation*
- *Etc. ...*

*Pas de métadonnées autres que celles générées automatiquement ou accessibles via le système d'exploitation de l'ordinateur.*

*Les rapports contiennent une page d'historique des versions.*

### **Exemple 2 :**

*Les documents sont archivés sur l'intranet sécurisé du projet. Seuls les membres du projet munis d'un mot de passe pourront y accéder.*

*Les données seront classées par work package, activité selon la nomenclature suivante :*

- *Numéro WP/ Numéro activité/ Nom fichier/ Date/ version*

### **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 entêtes de fichiers, dans un livre de référence (« code book ») ou dans les cahiers de laboratoire.

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

**Exemple 1 :**

*Pas de processus qualité spécifique dans le domaine d'activité concerné.*

**Exemple 2 :**

*Les données seront collectées sur la base du volontariat. Les données personnelles des répondants seront fournies sur la base du volontariat et traitées selon les règles RGPD en vigueur. Aucune donnée personnelle ne sera partagée sans l'accord expresse des propriétaires.*

*Les données personnelles collectées seront :*

- *Nom et prénom*
- *Nom de la structure/organisme de rattachement*
- *Poste ou rôle dans la structure*
- *Secteur d'activité de la structure*
- *Pays/ville*
- *Adresse email*

*Les données collectées ne seront accessibles qu'aux membres du projet, et ne seront ni publiées, ni accessibles à des institutions externes.*

*A tout moment, les participants pourront mettre un terme à leur coopération et échanges avec le projet. Ils pourront se désinscrire des listes de diffusion sur simple demande. Une fois désabonnés ou retirés du projet, leurs données personnelles seront effacées.*

*Les données personnelles ne seront pas réutilisées à d'autres fins que celles décrites pour ce projet. Elles ne seront pas transmises à des parties tierces.*

*Aucune publication scientifique ne sera basée sur les données personnelles collectées.*

*La collecte des données sera contrôlée par XXX et XXX*

**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 ?

**Exemple 1 :**

*Les données stockées sur le cloud sécurisé seront sauvegardées tous les jours, avec un historique des versions restaurables en cas de problème.*

*Les données de recherche des enseignants-chercheurs et chercheurs du projet sont stockées sur le réseau du laboratoire et sauvegardées avec la même périodicité.*

**Exemple 2 :**

*Les données seront stockées sur l'intranet du projet pour toute la durée du projet, géré par XX. L'accès est limité aux personnes identifiées, munies d'un identifiant et d'un mot de passe.*

*Toutes les données seront effacées 5 ans après la fin 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. Il est recommandé de stocker les données dans au moins deux lieux distincts.
- Privilégier l'utilisation de systèmes de stockage robustes, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution d'origine. Le stockage des données sur des ordinateurs portables, des disques durs externes, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.

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

*La restauration des données est sous la responsabilité des services informatiques des partenaires, qui administrent ces procédures suivant les normes et processus des établissements.*

*L'accès aux données est réservé aux personnes déclarées dans les activités du consortium XX. Les espaces de partages sont avec accès nominatifs et sécurisés par mot de passe.*

*Dans le cas de partage de données nécessaire avec des tiers, la décision est prise au niveau du consortium.*

*La politique de protection des résultats est définie par le consortium.*

**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 mis en œuvre.

**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 ?

**Exemple 1 :**

*Pas de collecte de données ou d'actions nécessitant l'application du RGPD*

**Exemple 2 :**

*Les données seront collectées sur la base du volontariat. Les données personnelles des répondants seront fournies sur la base du volontariat et traitées selon les règles RGPD en vigueur. Aucune donnée personnelle ne sera partagée sans l'accord expresse des propriétaires.*

*Les données personnelles collectées seront :*

- *Nom et prénom*
- *Nom de la structure/organisme de rattachement*
- *Poste ou rôle dans la structure*
- *Secteur d'activité de la structure*
- *Pays/ville*
- *Adresse email*

*Les données collectées ne seront accessibles qu'aux membres du projet, et ne seront ni publiées, ni accessibles à des institutions externes.*

*A tout moment, les participants pourront mettre un terme à leur coopération et échanges avec le projet. Ils pourront se désinscrire des listes de diffusion sur simple demande. Une fois désabonnés ou retirés du projet, leurs données personnelles seront effacées.*

*Les données personnelles ne seront pas réutilisées à d'autres fins que celles décrites pour ce projet. Elles ne seront pas transmises à des parties tierces.*

*Aucune publication scientifique ne sera basée sur les données personnelles collectées.*

*Recommandations:*

Lorsque vous manipulez des données à caractère personnel, veillez à ce que les lois sur la protection des données (par exemple, RGPD) soient appliquées, notamment :

- Obtenir un consentement éclairé pour la préservation et/ou le partage de données personnelles.
- Envisager l'anonymisation des données personnelles pour la préservation et/ou le partage (des données correctement anonymisées ne sont plus considérées comme des données personnelles).
- Envisager la pseudonymisation des données personnelles (la principale différence avec l'anonymisation est que la pseudonymisation est réversible).
- Envisager le chiffrement des données, qui est considéré comme un cas particulier de pseudonymisation (la clé de cryptage doit alors être stockée séparément des données, par exemple par un tiers de confiance).
- Expliquer si une procédure d'accès spécifique a été mise en place pour les utilisateurs autorisés à accéder aux données personnelles.

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

*Le consortium étant composé exclusivement de membres Français, la réglementation française s'appliquera. Il n'est pas possible de réutiliser les données hors du périmètre du consortium s'il s'agit de résultats partagés, sauf accord écrit préalable des partenaires. La demande devra être signée par tous les partenaires concernés et transmise au coordinateur de projet pour archives.*

*Recommandations:*

- Expliquer qui sera le propriétaire des données, c'est-à-dire qui aura le droit d'en contrôler l'accès :
  - Expliquer quelles conditions d'accès s'appliqueront aux données. Les données seront-elles librement accessibles, ou des restrictions seront-elles appliquées ? Si oui, lesquelles ? Envisager l'utilisation de licences concernant l'accès et la réutilisation des données.
  - S'assurer de couvrir, dans l'accord de consortium, ces questions de droits de contrôle d'accès aux données pour les projets multipartenaires et en cas de propriété partagée des données.
- Indiquer si les droits de propriété intellectuelle (par exemple la directive bases de données, droits *sui generis*) sont affectés. Dans l'affirmative, expliquer lesquels et comment cela sera traité.
- Indiquer s'il y a des restrictions sur la réutilisation des données fournies par des tiers.

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

*Exemple 1 :*

*Thématique non concernée par ces aspects.*

*L'université de Limoges a adopté la Charte Nationale de déontologie des métiers de la Recherche en 2015. Cette charte assure une démarche scientifique rigoureuse et intègre.*

### **Exemple 2 :**

*Tous les chercheurs impliqués sont formés aux questions d'éthique.*

*Ceux-ci répondent aux protocoles en place dans le labo XX.*

**Recommandations:**

- Déterminer si les questions d'éthique auront une incidence sur la façon dont les données seront stockées et transférées, qui pourra les voir ou les utiliser et quelles durées de conservation leur seront appliquées. Démontrer que ces aspects sont bien pris en compte et planifiés.
- Adopter les codes de conduite nationaux et internationaux et le code d'éthique institutionnel et vérifier si une revue des pratiques (par exemple par un comité d'éthique) est requise pour ce qui concerne la collecte de données dans le cadre du projet de recherche.

## **5. PARTAGE DES 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 ?

### **Exemple 1 :**

*La politique de publication des travaux, brevets et protection par secret est discutée au sein du consortium. La priorité est donnée aux publications dans des revues scientifiques ou congrès internationaux, après accord des partenaires.*

*Les embargos imposés par les éditeurs seront respectés dans le délai légal de 6 mois, sauf pour les publications publiées en gold open access.*

*En cas de rupture technologique, le brevet et la protection par secret seront une alternative discutée au niveau du consortium.*

### **Exemple 2 :**

*Les données produites seront facilement accessibles, car contenues dans les livrables du projet et publiés sur le site internet XXX.*

*Toutes les publications en lien avec le projet auront un DOI.*

*Les données seront également partagées lors d'événement scientifiques, congrès ou conférences. Des rencontres avec les politiques et les acteurs du secteur seront également organisées tout au long du projet pour partager avec eux les conclusions et résultats des enquêtes menées. Les données leur seront alors transmises.*

*Certaines données seront accessibles via les répertoires de données comme ZENODO. L'accès se fera via le nom du projet.*

**Recommandations:**

- Expliquer comment les données pourront être retrouvées et partagées (par exemple, par le dépôt dans un entrepôt de données de confiance, l'indexation dans un catalogue, par l'utilisation d'un service de données sécurisé, par le traitement direct des demandes de données, ou l'utilisation de tout autre mécanisme).
- Définir le plan de préservation des données et fournir l'information sur la durée d'archivage pérenne des données.
- Expliquer à quel moment les données seront rendues disponibles. Indiquer les délais de publication prévus. Expliquer si une utilisation exclusive des données est revendiquée et, dans l'affirmative, pour quelle raison et pour combien de temps. Indiquer si le partage des données sera différé ou limité, par exemple pour des raisons de publication, pour protéger la propriété intellectuelle ou le dépôt de brevets.
- Indiquer qui pourra utiliser les données. S'il s'avère nécessaire de restreindre l'accès pour certaines communautés ou d'imposer un accord pour le partage de données, expliquer comment et pourquoi. Expliquer les mesures qui seront prises pour dépasser ou minimiser ces restrictions.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

**Exemple 1 :**

*Chaque année, le consortium décide des données à archiver à long terme, détruire, publier, communiquer.*

*Les données archivées le sont sur des serveurs informatiques sécurisés gérés par le service informatique des partenaires, pendant une durée minimum de XX ans après la fin du projet.*

*Chacun des partenaires est le propriétaire des données générées dans le cadre du projet XX et sera libre de les réutiliser dans de futures recherches. Les résultats générés par le consortium devront faire l'objet d'une autorisation écrite des partenaires impliqués pour être réutilisées pour de futures recherches.*

**Exemple 2 :**

*Le consortium s'accordera sur les données à rendre publiques. Elles le seront via un répertoire de données, en l'occurrence ZENODO pour le projet XX. Les données en lien avec une publication seront rattachées à son DOI, pour assurer leur visibilité et leur interopérabilité.*

**Recommandations:**

- Indiquer quelles données ne doivent pas être divulguées ou doivent être détruites pour des raisons contractuelles, légales, ou réglementaires.
- Indiquer comment il sera décidé quelles données garder. Décrire les données qui seront à préserver à long terme.
- Décrire les utilisations (et/ou les utilisateurs) prévisibles des données dans un cadre de recherche.
- Indiquer où les données seront déposées. Si aucun entrepôt reconnu n'est proposé, démontrer dans le plan de gestion des données que les données pourront être prises en charge efficacement au-delà de la durée de financement du projet. Il est recommandé de démontrer que les politiques des entrepôts et les procédures de dépôts (y compris les standards de métadonnées, et coûts mis en œuvre) ont été vérifiés.

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

*Utilisation d'outils informatiques commerciaux possédés par les partenaires ou formats standards (libres) accessibles via des outils connus de la communauté scientifique.*

**Recommandations:**

- Indiquer si les utilisateurs potentiels auront besoin d'outils spécifiques pour l'accès et la (ré)utilisation des données. Tenir compte de la durée de vie des logiciels nécessaires pour accéder aux données.
- Indiquer si les données seront partagées via un entrepôt, si les demandes d'accès seront traitées en direct, ou si un autre mécanisme sera utilisé ?

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

**Exemple 1 :**

*Pas d'attribution de DOI, il n'y a pas de démarche qualité de ce type. Seuls les travaux scientifiques publiés font l'objet d'un DOI par l'organisme de publication.*

**Exemple 2 :**

*Toutes les publications du projet auront un DOI, attribué par les éditeurs. Les données seront rattachées au DOI lorsque cela sera possible.*

**Recommandations:**

- Expliquer comment les données pourraient être réutilisées dans d'autres contextes. Les identifiants pérennes devraient être appliqués de manière à ce que les données puissent être localisées et référencées de façon fiable et efficace. Les identifiants pérennes aident aussi à comptabiliser les citations et les réutilisations.
- Indiquer s'il sera envisagé d'attribuer aux données un identifiant pérenne. Typiquement, un entrepôt pérenne de confiance attribuera des identifiants pérennes.

## 6. 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) ?

### Exemple 1 :

*La gestion des données est effectuée par les services informatiques (DSI) des établissements. Le PGD est sous la responsabilité du porteur de projet, associé aux services juridiques de l'établissement, si nécessaire. Le coordinateur est responsable de la mise à jour du PGD si besoin.*

### Exemple 2 :

*Le partenaire XX est responsable de la bonne exécution et de la mise à jour de ce DMP au cours de la vie du projet. Il agira sous la responsabilité du Responsable Science Ouverte nommé par XX.*

Recommandations:

- Décrire les rôles et les responsabilités concernant les activités de gestion des données, par exemple : saisie des données, production des métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données. Nommer la(es) personne(s) responsable(s) impliquée(s) dans la mesure du possible.
- Pour les projets menés en collaboration, expliquer comment s'effectue la coordination des responsabilités de gestion des données entre partenaires.
- Indiquer qui est responsable de la mise en œuvre du PGD, et qui s'assure qu'il est examiné et, si nécessaire, révisé.
- Envisager des mises à jour régulières du PGD.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

### Exemple 1 :

*Pas de budget chiffrable (pas de service dédié à la gestion d'un PGD) dans l'établissement du porteur. Le coût des infrastructures informatique associé et leur gestion sont globalisés à l'échelle de l'établissement.*

### Exemple 2 :

*Le projet XX privilégiera les publications en green open access, dans des journaux gratuits. Cependant, et pour assurer la plus grande visibilité possible, la publication en gold open access sera autorisée une fois par an par partenaire, et dans la limite de 1500euros. Ces frais seront couverts par le budget commun du projet/le budget de chaque établissement.*

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.