

# Recommandations IRD pour les Plans de Gestion de Données rédigés selon le modèle ANR<sup>1</sup>

GROUPE DE TRAVAIL DMP IRD\*

## ACCOMPAGNEMENT A LA REDACTION DE PLANS DE GESTION DE DONNEES

\*Mission culture scientifique et technique (MCST) : Pascal Aventurier, Luc Decker, Caroline Doucouré (animatrice) et Hanka Hensens

Mission science ouverte (MSO) : Jean-Christophe Desconnets, François Sabot

Département des relations internationales et européennes (DRIE) : Catherine Lauranson, Laurence Lombard

Département mobilisation de la recherche et de l'innovation pour le développement (DMOB) : Laetitia Berg, Laetitia Mahé

Direction des affaires juridiques (DAJ) : Zoé Simon

Direction du développement des usages numériques innovants (DDUNI) : Damien Alline

## Table des matières

Informations sur le projet de recherche.....	1
Produits de recherche.....	1
1. Description des collections et des données .....	2
Data Collection / Collecte ou réutilisation des données .....	2
Data Description / Type des données .....	3
Data Format / Format des données.....	3
Data Volume / Volume des données .....	3
2. Documentation et qualité des données .....	4
Metadata & Documentation / Documentation et métadonnées.....	4
Data Collection / Qualité des données .....	5
3. Stockage et sauvegarde pendant le processus de recherche .....	5
Storage & Security / Stockage et sauvegarde.....	5
Storage & Security / Sécurité des données sensibles.....	6
4. Exigences légales et éthiques, codes de conduite.....	6
Ethics & Privacy / Respect de la législation .....	6
Intellectual Property Right / Propriété intellectuelle .....	7
Ethics & Privacy / Déontologie et éthique.....	7
5. Partage des données et conservation à long terme.....	8
Data Sharing / Partage .....	8
Data Repository / Entrepôt .....	8
Preservation / Conservation .....	9
6. Responsabilités et ressources en matière de gestion des données .....	10
Roles & Responsibilities / Rôles et responsabilités .....	10
Budget / Ressources .....	11

Les recommandations IRD complètent les [recommandations de l'ANR](#).

Elles reprennent en partie certaines recommandations de l'INRAE, remercié ici pour son aimable autorisation.

La rédaction des Plans de Gestion de Données sera facilitée par l'utilisation des formulaires en ligne du site [DMP-Opidor](#), où ces recommandations sont affichées en face de chaque question.

<sup>1</sup> Le modèle de DMP ANR est basé sur les recommandations de Science Europe ([Science Europe. Guide pratique pour une harmonisation internationale de la gestion des données de recherche, juillet 2019](#))



## Informations sur le projet de recherche

- Acronyme du projet
- Nom/Titre du projet de recherche
- Financeur(s) du projet
- Numéro de subvention et/ou référence de la convention de financement
- Résumé du projet
- Dates et durée du projet
- Personne responsable du projet (chercheur principal) : nom, identifiant (ORCID <https://orcid.org/> ou autre), courriel, téléphone
- Personne responsable des données (si différente du responsable du projet) : nom, courriel, téléphone
- Institution leader du projet, coordinateur bénéficiaire (nom, pays)
- Autres partenaires (nom, pays, rôle de chacun des partenaires en dehors de l'institution leader)

## Produits de recherche

Avant tout, lister les produits de recherche du projet (publications exceptées) en leur trouvant un nom abrégé significatif et en sélectionnant leur type : collection, jeu de données, image, ressource interactive, modèle, objet physique (échantillon, ...), service, texte, logiciel, audiovisuel, son, workflow, autre (protocole, ...).

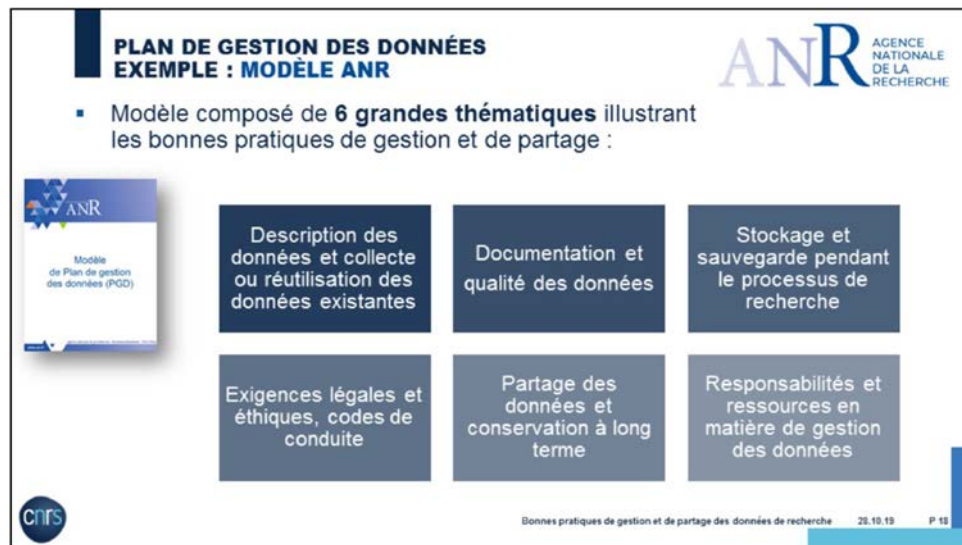
Il est conseillé de réfléchir, avant toute saisie ou rédaction, à la granularité la plus adaptée pour ces produits de recherche :

- Limiter autant que possible le nombre de produits de recherche, pour ne pas multiplier les onglets/sous parties, et alourdir la rédaction et la lecture.
- Créer des produits de recherche relativement homogènes, pour ne pas avoir à détailler trop de cas différents dans chaque rubrique.

Avec DMP-Opidor:

- Dans l'onglet Rédiger, au début de chaque section, vous pouvez cocher une case si les réponses aux questions sont communes à tous les produits de recherche, afin de fusionner ponctuellement les différents produits de recherche.
- Dans l'onglet Télécharger, vous pourrez choisir d'exporter votre plan hiérarchisé par produits de recherche (détaillés par sections), par sections (détaillées par produits de recherche), ou par questions (détaillées par produits de recherche)

Le PGD ANR comporte 6 sections de 2 à 4 questions correspondant à chaque étape du cycle de vie des données :



## 1. Description des collections et des données

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

**ANR :**

- 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.

### Data Collection / Collecte ou réutilisation des données

- Détailler les méthodologies (outils, procédures, conditions expérimentales, etc.), les raisons du choix de ces méthodologies et l'origine des données reprises.
- Présenter l'ensemble des jeux de données en précisant pour chaque jeu, de préférence dans un tableau :
  - le mode d'obtention : données générées, données produites par un tiers (conversion / transformation de données, partage / échange de données ou achat de données)
  - l'origine des données, par exemple : Analyse, Agrégation, Code, Corpus audiovisuel, Corpus textuel, Enquête, Expérimentation, Observation, Simulation/Modélisation
- Vous rapprocher éventuellement de votre observatoire des sciences de l'univers (OSU) de rattachement pour une aide à la conception de systèmes de collecte de données : <https://www.insu.cnrs.fr/fr/les-observatoires-des-sciences-de-lunivers>.

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

### ANR :

- 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).

### Data Description / Type des données

Présenter l'ensemble des jeux de données du produit de recherche (conseil : sous forme de tableau) en précisant pour chaque jeu :

- son nom (unique et explicite pour faciliter sa citation)
- le type de données (Audiovisuel, Collection, Jeu de données, Image, Ressource interactive, Méthode, Modèle, Objet physique, Service, Logiciel, Son, Texte, Workflow,...)
- la nature des données (précise le type, exemple : Analyses génomiques, Images satellites, Relevés de températures, Transcription d'entretiens ...)
- le périmètre thématique et géographique des données en précisant si les données sont géoréférencées

### Data Format / Format des données

On choisira par ordre de préférence :

- les enregistrements directs dans des formats non-propriétaires, ouverts, pérennes, réutilisables et gratuits,
- des conversions vers des formats ouverts en conservant également les fichiers originaux,
- un format standard de la discipline, largement utilisé dans sa communauté scientifique.

Pour la plupart des bailleurs, la norme est d'utiliser des formats ouverts : dans le cas où l'utilisation de formats ouverts ne seraient pas possible, le justifier.

Pour en savoir plus : [https://data.ird.fr/gerer/quelles-donnees/#Les\\_formats\\_de\\_donnees](https://data.ird.fr/gerer/quelles-donnees/#Les_formats_de_donnees)

### Data Volume / Volume des données

Donner un ordre de grandeur pour chaque jeu de données, à préciser au fur et à mesure du projet et à détailler d'autant plus que les données sont volumineuses, donc complexes et coûteuses à gérer, à traiter (temps de calcul), à sauvegarder, à transmettre (bande passante).

## 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 ?

ANR :

- 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.

#### Metadata & Documentation / Documentation et métadonnées

La documentation accompagnant les données permet aux utilisateurs de les repérer facilement et apporte les informations nécessaires à un bon usage et une bonne interprétation :

- A minima, un fichier de type "lisez-moi" peut être rédigé pour rassembler les informations de base sur les données : nom de la source, format du fichier, identifiant, description du contenu ...
- Documenter la signification de chaque champ ou variable, préciser les unités de mesures (data dictionary) et la signification de tous les codes utilisés (code dictionary) pour rendre les fichiers de données compréhensibles et réutilisables.
- Utiliser de préférence des standards de métadonnées, généraux ou spécifiques à une discipline scientifique, et des outils d'aide à la production de métadonnées (<https://data.ird.fr/documenter-ses-donnees/>).
- Indiquer les principales publications associées aux jeux de données avec leur DOI. La rédaction d'un Data Paper à partir des métadonnées est recommandée (<https://data.ird.fr/datapapers/>).

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

ANR :

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.

## Data Collection / Qualité des données

Préciser en 2b l'organisation des données pendant le projet : architecture des dossiers et règle de nommage des fichiers.

Vérifier la conformité avec les standards/normes de la discipline (codage de certains types de données ...) et le respect des processus qualité de vos organismes de tutelle, s'ils existent.

S'inspirer du guide des bonnes pratiques de DoRANum pour tenir un cahier de labo :

<https://www.dataacc.org/bonnes-pratiques/utiliser-un-cahier-de-laboratoire-numerique/ressource-dorandum-cahiers-de-labo/>

Pour en savoir plus : <https://data.ird.fr/qualite-des-donnees/>

## 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 ?**

ANR :

- 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.

### Storage & Security / Stockage et sauvegarde

Indiquer ici les outils utilisés pour stocker vos données pendant la durée du projet. Il est préférable d'utiliser des outils institutionnels, pérennes et régulièrement mis à jour. L'utilisation d'ordinateurs personnels, de disques durs externes ou de clefs USB seuls est particulièrement déconseillée.

Préciser la procédure de sauvegarde au cours du projet : périodicité, automatisation, nombre de copies et localisation. En principe : sauvegarde au moins 3 fois sur au moins 2 supports différents, dont 1 en ligne, de 3 versions successives. En fonction de l'importance des données, considérer la possibilité d'effectuer des sauvegardes occasionnelles hors réseau (off-line).

La DDuni propose une offre de stockage de données partagées et de construction collaborative de documents en ligne : <https://ent.ird.fr/> et <https://drive.ird.fr/>. Pour plus de renseignements s'adresser à [assistance-informatique@ird.fr](mailto:assistance-informatique@ird.fr), tél. +33 (0)4 67 41 62 19.

Pour en savoir plus : <https://data.ird.fr/sauvegarder-ses-donnees/>

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

ANR :

- 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.

## Storage & Security / Sécurité des données sensibles

Les données sensibles nécessitent une vigilance accrue.

Pour en savoir plus : [https://data.ird.fr/cadre-juridique/#Cas\\_particuliersPoints\\_de\\_vigilance](https://data.ird.fr/cadre-juridique/#Cas_particuliersPoints_de_vigilance) et <https://www.cnil.fr/fr/definition/donnee-sensible>.

Détailler les personnes/fonctions à qui sont attribués des droits d'accès aux données (en lecture, édition, suppression... à distance ou en local uniquement).

Le cas échéant, préciser :

- les mécanismes de traçabilité qui s'appliquent automatiquement pour tout changement effectué,
- les restrictions d'accès (login/mot de passe),
- la gestion spécifique des données sensibles : répertoire sécurisé différent, entrepôt certifié HDS (<https://data.ird.fr/certification-hds-hebergement-de-donnees-de-sante/>) pour les données de santé ...
- les éventuels processus de chiffrement et de contrôles automatiques de l'intégrité des données (lors des transferts, du stockage, etc.), utilisés pour éviter la divulgation accidentelle ou malintentionnée de données sensibles, avant l'anonymisation (si on ne peut pas anonymiser dès la collecte de données, à la source),
- les procédures de pseudonymisation, floutage par agrégation ou anonymisation prévues pour les données personnelles.

Préciser quelles données seront conservées ou détruites, à quels moments et qui décidera (<https://data.ird.fr/cadre-juridique/>).

## 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 ?*

ANR :

*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.*

## Ethics & Privacy / Respect de la législation

Répertorier ici les jeux de données nécessitant des mesures particulières, à préciser, du fait de la législation des pays concernés par la recherche : données scientifiques à risques (sécurité de l'Etat, sécurité des populations, etc.), données personnelles, données sensibles et données de santé, données liées à l'intelligence économique (secret industriel et commercial), etc.

Pour en savoir plus : [https://data.ird.fr/cadre-juridique/#Cas\\_particuliersPoints\\_de\\_vigilance](https://data.ird.fr/cadre-juridique/#Cas_particuliersPoints_de_vigilance)

Le responsable de mise en œuvre du traitement des données aura rempli un registre de traitement des données personnelles, afin de se mettre en conformité (il peut être fourni en annexe du plan de gestion de données).

Pour en savoir plus : <https://www.ird.fr/reglement-general-sur-la-protection-des-donnees-rgpd> (page intranet) et <https://data.ird.fr/cadre-juridique/>.

**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 ?**

**ANR :**

*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.*

### **Intellectual Property Right / Propriété intellectuelle**

Avant tout, consulter la synthèse du cadre juridique : <https://data.ird.fr/cadre-juridique/>

Détailler ici les droits de propriété et d'utilisation des différents partenaires du projet pour chaque jeu de données. Ces droits seront utilement définis dès la conception du projet dans les conventions et accords de consortiums signés par les différents organismes partenaires, de même que les futures licences standard de diffusion ou les accords de transmission des données (« data transfer agreements »)

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

**ANR :**

*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.*

### **Ethics & Privacy / Déontologie et éthique**

De manière générale, se poser la question « Est-ce que la collecte ou/et la divulgation de ces données est susceptible de causer du tort à des personnes ou à des espèces ? » (préciser si nécessaire la balance entre les risques à collecter et/ou partager les données et les bénéfices attendus).

Lister les types de données (données sensibles, données de santé, données personnelles, etc.) qui pourraient poser des problèmes d'éthique, en précisant pourquoi. Si aucun jeu de données ne soulève a priori de problème d'éthique, le dire aussi.

Préciser également si une formation ou une certification est requise pour les personnes autorisées à accéder/traiter des données sensibles.



Déterminer quel(s) comité(s) d'éthique sera(ont) contacté(s) si les données posent des questions éthiques ou de protection de la confidentialité, dont les comités d'éthiques des éventuels pays partenaires. Dans le cas où un comité d'éthique a revu le projet, le préciser.

A l'IRD, le [Comité consultatif éthique pour la recherche en partenariat \(CCERP\)](#) peut examiner d'un point de vue éthique les protocoles de recherche. Contact [chloe.desmots@ird.fr](mailto:chloe.desmots@ird.fr)

Pour en savoir plus : [Guide IRD des bonnes pratiques de la recherche pour le développement](#)

## 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 ?**

ANR :

- 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.

### Data Sharing / Partage

Détailler les licences ou les contrats qui seront appliqués aux données partagées, voir [https://data.ird.fr/cadre-juridique/#Choix\\_de\\_licence](https://data.ird.fr/cadre-juridique/#Choix_de_licence).

Préciser à qui les données pourraient être utiles si on en a une idée.

Si une période d'embargo (exclusivité d'accès aux données réservée au producteur) est nécessaire, indiquer sa raison (par exemple pour la valorisation scientifique des données, le traitement, la publication) et sa durée.

Expliquer ici pourquoi certaines données ne seront pas partagées (raisons éthiques, données confidentielles ou personnelles, propriété intellectuelle, commerciale, raisons de sécurité...).

En effet, il existe des exceptions à l'obligation d'ouverture des données, voir <https://data.ird.fr/cadre-juridique/>, dont [https://data.ird.fr/cadre-juridique/#Aides\\_a\\_la\\_decision](https://data.ird.fr/cadre-juridique/#Aides_a_la_decision)

Si les données ne sont pas partagées (ou archivées), indiquer comment les résultats pourraient être validés en cas de sollicitation.

### Data Repository / Entrepôt

Les données sont parfois mises à disposition des utilisateurs via un site web dédié créé et maintenu par l'équipe de recherche : le préciser.

Dans tous les cas, il est recommandé de prévoir de préserver les données dans un entrepôt de confiance à l'issue du projet :

- Les données peuvent être déposées dans un entrepôt généraliste, thématique ou institutionnel, comme DataSuds <https://dataverse.ird.fr/> (pour en savoir plus : <https://data.ird.fr/l-entrepot-datasuds/>)
- Pour choisir un entrepôt, consulter : <https://data.ird.fr/entrepots-de-donnees-2/>
- L'attribution d'un DOI facilitant l'accès aux données et leur citation par l'entrepôt est un des critères à prendre en compte : cf. <https://data.ird.fr/obtenir-un-doi/>

**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) ?**

**ANR :**

- *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.*

### **Preservation / Conservation**

Préciser les principes qui permettront de choisir quelles données conserver/détruire à la fin du projet et qui en décidera.

Respecter les exigences légales et déontologiques de conservation ou de destruction de certains types de données.

Les données brutes sont intéressantes pour les réutilisations futures. Difficiles ou impossibles à reproduire, elles sont donc à conserver autant que possible.

Si les données occupent un espace de stockage négligeable (tel que 100 ko), considérer la solution simple de conserver indistinctement l'ensemble des données.

Pour en savoir plus : <https://www.cines.fr/archivage/typologies/donnees-scientifiques/>,  
<https://doranum.fr/wp-content/uploads/Fiche-Synth%C3%A9tique-Archivage.pdf>,  
<https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep>

Rechercher quelle plateforme pérenne pourrait accueillir les données et à quel coût.

Pour les données volumineuses, la DDuni de l'IRD propose une offre de stockage massif. Contacter [assistance-informatique@ird.fr](mailto:assistance-informatique@ird.fr)

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

ANR :

- *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 ?**

ANR :

- *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. 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) ?**

ANR :

- *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.*

### **Roles & Responsabilities / Rôles et responsabilités**

Identifier en particulier une personne responsable de la sécurisation des données et de leur sauvegarde.

Prévoir des suppléants pour chaque rôle pour assurer la continuité, en particulier pour les rôles critiques, les données à vérifier ou à surveiller en temps réel, par exemple.

Définir les modalités de décision dans un projet qui implique de nombreux partenaires, pour arbitrer une situation non-anticipée (un rôle ayant évolué ou n'ayant pas été prévu, etc.).

Voir <https://www.ird.fr/modeles-de-contrats> (page intranet).

**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) ?**

**ANR :**

- 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.

### **Budget / Ressources**

Ne pas oublier de chiffrer le salaire du/des gestionnaire(s) de données.

Le coût élevé de l'anonymisation des données est à anticiper le cas échéant.

Mentionner les services obtenus gratuitement (par exemple l'utilisation de l'entrepôt DataSuds).

Pour estimer le coût de la gestion des données, voir :

- UK Data Service - [Data management costing tool](#)
- OpenAIRE - [How to identify and assess Research Data Management \(RDM\) costs](#)

*Pour toute question ou assistance complémentaire, merci de contacter [dmp@ird.fr](mailto:dmp@ird.fr)*