**La gestion des données au sein de votre projet**

Avant d’envisager un dépôt de projet auprès d’un financeur ayant potentiellement des exigences en matière de gestion et de diffusion de données, il est important de s’interroger sur la maturité de sa réflexion sur le sujet et au besoin de se faire accompagner en vue de la réalisation de son plan de gestion des données (PGD), livrable désormais obligatoire de nombreux appels à projets.

Outils de positionnement autant que support d’échange avec les équipes du SGR et de CADOR, le présent questionnaire a vocation à identifier les éventuels points de blocage ou de vigilance à lever en termes de gestion de données pour garantir le succès de votre projet. Il offre par ailleurs les prémisses d’une trame de plan de gestion permettant d’amorcer, dès à présent, sa co-construction et sa consolidation entre les équipes du SGR et de CADOR et les porteurs de projets.

# Informations générales sur le projet

### Description du projet

Prévoir un résumé du projet de recherche en explicitant la méthode, les enjeux et les impacts attendus. Pour cette section, n’hésitez pas à reprendre les éléments figurant déjà dans votre projet de réponse à l’AAP ciblé. Pensez à expliciter la manière dont la gestion et la valorisation des données serviront les objectifs d’impact de votre projet et des différents partenaires.

### Qui pourrait utiliser ou réutiliser les données produites au cours du projet ?

Indiquer comment les données pourraient être réutilisées dans d'autres contextes : par d’autres chercheurs de la discipline ou d’autres disciplines, comme matériel pédagogique, à des fins économiques ou sociétales etc.

### Qui seront les partenaires du projet ? Les partenaires sont-ils des entités privées ?

Préciser si d’autres universités, institutions de recherches publiques ou privées, associations ou entreprises seront partenaires ou impliquées dans le projet ? Attention, si le projet prévoit la création d’infrastructure numérique (corpus, base de données, logiciels etc.) l’inclusion d’un partenaire public ou privé spécialisé peut se révéler nécessaire (prestataire, MSH, BNF, archives etc.).

### Les collaborateurs font-ils partie de l’université ou d’autres établissements ?

Si le projet prévoit des collaborateurs extérieurs à l’établissement, il sera nécessaire de s’interroger d’une part sur la propriété des données et résultats, d’autre part sur les conditions de stockage et de partage des données durant le projet pour garantir un accès sécurisé à tous les partenaires.

### Avez-vous identifié un ou des financeurs pour le projet, si oui, a-t-il / ont-ils des exigences en matière de gestion de données et en matière de diffusion des données et des résultats ?

Exemples : appels à projet ANR, ERC, Horizon Europe… En cas de doute sur les politiques science ouverte et données de recherche des financeurs, vous pouvez consulter la plateforme [Sherpa Juliet](https://v2.sherpa.ac.uk/juliet/).

### Avez-vous déjà envisagé la propriété des données et résultats et sa répartition entre les partenaires ?

Dans le cadre des projets européens, un accord de consortium sera obligatoire. S'assurer de couvrir, dans l’accord les problématiques de contrôle d'accès aux données 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.

# Description des données, collecte ou réutilisation de données existantes

### Quelles seront les méthodes de collecte ou de production de données mises en œuvre ?

Si de nouvelles données sont recueillies ou produites, préciser les méthodologies ou logiciels utilisés. Expliquer comment la provenance des données sera documentée.

### Envisagez-vous la réutilisation de données existantes ?

Exemples : données issues de précédents projets, moissonnage de données publiques ou diffusées etc. Le cas échéant en indiquer la source et la finalité. A l’inverse, justifier les éventuelles limites ou restrictions à la réutilisation de données préexistantes.

### De quelle nature seront les données collectées ou produites ? Quels seront les formats utilisés ?

Exemples : types de données (observation, simulation, expérimentales, dérivées ou compilées, canoniques, logiciels), nature des données (enquêtes, bases de données, corpus numérique ou numérisés, cartographie etc.), formats des données (textes, images, sons, vidéos, tableurs, calques, médias composites etc. en précisant si possible les extensions). Le cas échéant, justifier de l'utilisation de certains formats (gratuité, expertise des équipes, prérequis de l’entrepôt, disponibilité du matériel etc.). D’une manière générale privilégier les formats libres e ouvertes pour la diffusion et l’archivage. Préciser les volumes de données (en octets ou en quantités d'objets, de fichiers, de lignes, et colonnes).

# Documentation et qualité des données

### Avez-vous connaissance d’un standard de métadonnées utilisables pour décrire vos données produite ou collectées ?

Utiliser de préférence les standards de métadonnées de votre communauté disciplinaire lorsque ceux-ci existent (consultez le [catalogue international des standards de métadonnées](https://rdamsc.bath.ac.uk/) ou référez-vous idéalement à des standards génériques plus communément utilisés comme le Dublin Core par exemple.

### Dans le cas contraire, comment vos données seront-elles décrites ?

Lister les informations (métadonnées) nécessaires pour garantir la lecture, la réplicabilité, l’interopérabilité, la réutilisation et la préservation de vos données. Ces informations doivent renseigner à minima sur l’origine et le contexte de production des données (métadonnées administratives), les conditions de production (métadonnées techniques), les objectifs de la collecte et les références à sa période et à sa localisation (métadonnées scientifiques), les conditions d’accès et les modalités d’utilisation ou de réutilisation des données (métadonnées juridiques). Vous pouvez par [exemple utiliser des outils de création de métadonnées](https://doranum.fr/metadonnees-standards-formats/outils-creation-metadonnees_10_13143_9sw6-z467/), tenir un tableur avec des liens aux données, regrouper des informations contextuelles dans un fichier « lisez-moi », réaliser un livre de référence (« code book ») ou tenir un cahier de laboratoire.

### Avez-vous défini des règes de nommage pour vos fichiers de données ?

Des données bien classées et gérées de façon cohérente seront plus faciles à retrouver, à comprendre et à réutiliser. Les conventions de nommage des dossiers et fichiers constituent souvent premier niveau de métadonnées renseignant sur l’origine et le contenu des fichiers. Préciser si des règles communes de nommages, tri ou versioning ont été définies au niveau du projet. Donner des exemples le cas échéant.

Exemples : Acronyme du projet\_Sujet\_Nature du document\_Date\_Version / ACCRO\_Etat-lieux\_CR\_2022-01-15\_v01 / SYTHIR\_Evaluation\_Fiche-Pratique\_2021-12-09\_V03

### Comment envisagez-vous de contrôler la qualité et l’intégrité des données collectées ou produites ?

Préciser les processus et protocoles métrologiques qui seront mis en place pour garantir l’intégrité, la cohérence et la qualité des données notamment lorsque plusieurs acteurs interviennent. Exemple : processus de la calibration, répétition des échantillons ou des mesures, capture standardisée, validation de saisie, revue par les pairs, utilisation de vocabulaires contrôlés etc.

# Stockage et sauvegarde pendant le processus de recherche

### Avez-vous déjà identifié où et sur quel support seront stockées les données et les métadonnées au cours du projet ?

Selon la règle « 3-2-1 », il est recommandé de disposer d’au moins 3 copies, sur 2 supports différents dont 1 copie hors site. La nature des données (volumétrie, structuration, sensibilité) peut induire des contraintes de stockage particulier. D’une manière générale, privilégier le recours à des systèmes de stockage robustes et publics, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution (OwnCloud Lyon 3), les TGIR (ShareDoc d’HumaNum) ou des plateformes collaboratives publiques liées à votre discipline (Ortolang, Simogih etc.)

### Le projet nécessitera-t-il un espace de stockage commun et/ou un outil collaboratif ? Si oui, devra-t-il être accessible à des chercheurs de différents établissements ?

# Exigences légales et éthiques, codes de conduite

### Le projet génèrera ou collectera-t-il des données sensibles ou à caractère personnel impliquant des garanties de confidentialité ?

Une donnée personnelle se rapporte à une personne physique identifiée ou identifiable. Les données sensibles forment une catégorie particulière des données personnelles touchant à la prétendue origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données des données de santé ou des données concernant la vie ou l'orientation sexuelle. Un échange avec le DPD de l’établissement est recommandé pour tout projet impliquant la gestion de données à caractère sensible ou personnel.

### Si oui, à quoi seront-elles destinées ?

Conformément aux dispositions du RGPD, il est nécessaire dans le cadre d’une recherche d’informer explicitement les personnes impliquées de la finalité des collectes de données les concernant et des traitements qui seront appliqués afin de leur permettre d’en conserver la maitrise. Pour les données sensibles, il est également nécessaire de recueillir le consentement formel des personnes impliqués pour l’usage prévu dans le cadre de la recherche. Selon le type de recherche pratiquée ou les modalités de publications envisagées des résultats, l’avis complémentaire d’un comité d’éthique de la recherche (CER), d’un comité de protection des personnes (CPP) et/ou de la CNIL pourra être requis. Pour savoir si vous êtes concernés : [questionnaire en ligne d’auto qualification](https://enquete.univ-reims.fr/limesurvey/index.php/786859?lang=fr)

### Si oui, avez-vous défini un protocole de stockage spécifique ?

Outre des précautions particulières en matière de diffusion, les données sensibles ou personnelles impliquent des protocoles de traitement et de stockage particuliers : recours à des espaces de stockage publics et sécurisés, si externalisation recours à des prestataires agréés notamment pour les données de santé), gestion rigoureuse des droits d’accès, si nécessaire mesures d’anonymisation ou de pseudonymisation (opération réversible), cryptage etc.

# Partage et conservation des données

### Les données auront-elles vocation à être diffusées ? Si non, pourquoi ? Si oui, quand ?

Une majorité de financeurs de la recherche impose aujourd’hui le dépôt en accès ouvert des données de recherche. Cette diffusion doit cependant se faire dans le respect du principe « aussi ouvert que possible aussi fermé que nécessaire » dans le respect des obligations de protection des données à caractère sensible, personnel ou des enjeux de valorisation. Toute dérogation au principe de diffusion devra être motivé. Des délais d’embargo peuvent être envisagés pour concilier les deux exigences.

### Comment envisagez-vous la diffusion des données ?

Exemple : par le dépôt dans un entrepôt de données de confiance, l'indexation dans un catalogue, la création d’une base de données ouvertes etc. Précisez à quel moment les données seront rendues disponibles et dans quelle mesure la solution retenue permettra de garantir que les données repérables et accessibles. Toutes les données n’ont pas forcément vocation à être diffuser, certaines peuvent également l’être selon des modalités différentes. Il conviendra dans ce cas de distinguer plusieurs « jeux de données » au sein du projet.

### Comment envisagez-vous la FAIRisation des données ?

La valorisation des données passe par leur documentation et leur diffusion selon les principes FAIR (Faciles à trouver, Accessibles, Interopérables et Réutilisables). La mise en œuvre de ces principes repose sur plusieurs actions : recourir à l’usage des standards de métadonnées génériques disciplinaires ; privilégier le dépôt dans des entrepôts de données publics - disciplinaires, institutionnels, nationaux ou internationaux ; s’assurer de l’attribution aux jeux de données diffusés d’un identifiant pérenne permettant l’identification unique d’un objet physique ou numérique (DOI ou équivalent) ; privilégier le choix du format ouvert et de licences de diffusion libres permettant le partage et la réutilisation des données (licence Ouverte, Etalab, Creative Commons etc.).

### L’ensemble de vos données a-t-il vocation à être conservé ? Quels seront les critères de tri appliqués ?

Il n’est ni possible ni recommandé de conserver l’intégralité des données produites ou collectées. Pour guider le processus de tri et de sélection, il est important de s’interroger sur la valeur des données (sont-elles reproductibles ? à quel coût ? faut-il conserver les données brutes, les résultats finaux ou les modèles de traitement ? etc.).

### Quelle(s) durée(s) de conservation envisagez-vous selon les types ou jeux de données ?

Certaines données ont vocation à être à préserver à long terme, notamment celles venant en appui direct de la justification des résultats publiés. D’autres ont vocation à n’être conservé que de manière transitoire. Pour les différents jeux de données, préciser les utilisations (et/ou les utilisateurs) attendus et en conséquence les durées de conservation préconisées. Une partie de la réflexion peut également être guidée par le [référentiel de gestion des archives de la recherche](https://www.archivistes.org/IMG/pdf/referentiel_recherche_intro_septembre2012_corrige_.pdf?2969/9934949897cbdb7904889234afaded8f3c7d9035).

### Comment envisagez-vous la conservation à moyen et long terme de vos données ?

Indiquer sur quels plateformes ou outils les données ont vocation à conserver. Les supports peuvent varier d’un jeu de données à l’autre en fonction des durées de conservation retenues, de la nature et des formats de données. Certains entrepôts de données offrent des garanties d’archivage ou de pérennisation à long terme des données Si ce n’est pas le cas, il conviendra de démontrer comment les données pourront être prises en charge efficacement au-delà de la durée de financement du projet (stockage et garantie d’intégrité à long terme, migration de formats etc.)

# Responsabilités et ressources en matière de gestion des données

### Qui sera responsable de la rédaction et de la mise à jour du plan de gestion des données ?

*Le plan de gestion est un livrable contractuel demandé par de plus en plus de financeurs de la recherche dans les six mois suivant l’acceptation du projet. Document unique mais évolutif, il reprend la majorité des interrogations évoquées au cours de ce document et doit faire l’objet d’une actualisation régulière (au moins 3 versions : version initiale, version à mi projet, version finale)*

### Comment se définira la responsabilité de la gestion et du traitement des données ou des jeux de données au cours du projet ? (Une répartition entre plusieurs acteurs est envisageable)