

## PPR Autonomie

### Synthèse du webinaire d'information autour des possibilités de financement des dépenses liées à la gestion des données de la recherche dans le cadre de l'appel à projet sur les défis 3 et 4 du programme

#### Table des matières

Préambule .....	1
Cadre général .....	1
Comment estimer les coûts de gestion des données ? .....	2
Retour sur le Plan de gestion des données (PGD) .....	3
Le PGD, qu'est-ce que c'est ? .....	3
Avec qui rédiger un PGD ? .....	4
Quand rédiger un PGD ? .....	4
Pour aller plus loin .....	6
Comment traiter les données à caractère personnel ? .....	5
Définitions .....	5
Les acteurs de la gestion des données .....	5
Les droits des personnes .....	5
Qu'est-ce qu'une analyse d'impact relative à la protection des données (AIPD) ? .....	6

#### Préambule

Le 16 février 2023 l'équipe projet du PPR Autonomie réunissait les personnes intéressées par le dépôt d'un projet dans le cadre de l'appel en cours sur les défis 3 et 4 du programme pour évoquer les possibilités de financement des dépenses liées à la gestion des données de la recherche.

Ce webinaire était organisé en partenariat avec l'Humathèque Condorcet et l'Unité d'appui et de recherche 2028 (UAR Condorcet du CNRS). Il était animé par la directrice de l'UAR 2028, Camille Oloa-Biloua.

25 personnes se sont connectées à ce temps de présentation et d'échange.

#### Cadre général

La science ouverte est une nouvelle approche de la démarche scientifique basée sur la production collaborative des produits de science, sur leur partage, sur leur libre circulation et sur leur réutilisation. La France s'inscrit pleinement dans ce mouvement, à travers l'élaboration de textes successifs qui établissent

de nouvelles obligations d'ouverture auxquelles les établissements publics doivent se soumettre pour les publications, les données et les résultats qu'ils produisent.

Dans ce contexte, on notera l'évolution du cadre juridique :

- Loi Valter (2015) : principe de gratuité dans la réutilisation des informations publiques (interdiction de rendre payant l'accès aux bases de données produites par un établissement public à caractère scientifique et technique - EPST)
- Loi pour une République numérique (2016) : mise en place d'un principe d'ouverture par défaut pour les données publiques, y compris les données de la recherche
  - o Obligation pour les EPST de mettre en ligne spontanément les données qu'ils produisent dans un format ouvert
  - o Obligation de permettre la libre réutilisation de leurs données, y compris à des fins commerciales
- Plan national pour la science ouverte (2018) qui rend obligatoire :
  - o La publication en libre accès des articles et livres issus de recherches financées par appels d'offres sur fonds publics
  - o La diffusion ouverte des données de la recherche issues de programmes financés par appels à projets sur fonds publics.

Dans le cadre des appels à projets de l'ANR, le coordinateur ou la coordinatrice et les partenaires s'engagent en cas de financement de leur projet par l'ANR :

- A déposer les publications scientifiques (texte intégral) issues du projet de recherche dans une archive ouverte, soit directement dans HAL soit par l'intermédiaire d'une archive institutionnelle locale, dans les conditions de l'article 30 de la Loi « Pour une République numérique » 13 ;
- A fournir dans les 6 mois qui suivent le démarrage du projet un plan de gestion des données (PGD) selon des modalités communiquées dans l'acte attributif d'aide et le règlement financier de l'ANR. Par ailleurs, l'ANR recommande de privilégier la publication dans des revues ou ouvrages nativement en accès ouvert.

## Comment estimer les coûts de gestion des données ?

La gestion et l'ouverture des données peut générer trois types de dépenses :

- Prestations de service (*subcontracting costs*)
- Coûts de personnels (*personnel costs*)
- Frais généraux (*travel & subsistence costs*)

Voici les principales activités qui doivent être prises en compte au moment de la construction du budget car elles sont susceptibles d'engendrer des coûts supplémentaires qui peuvent être pris en charge par la subvention ANR dans le cadre du projet.

Du côté des coûts de prestation (*subcontracting costs*) :

- Organisation d'accès sécurisés aux données pour des chercheurs ou des intervenants extérieurs ;
- Numérisation, digitalisation des données ;
- Accès à une base de données externe, acquisition de données existantes ;
- Stockage des données sur le cloud, des serveurs ;
- Sécurisation, sauvegarde des données ;
- Licences de logiciels propriétaires (outils d'analyse, etc.).
- Etc.

Ces dépenses nécessitent de solliciter des infrastructures existantes, en interne auprès de l'établissement gestionnaire du projet ou auprès d'une infrastructure extérieure.

Du côté des coûts de personnel (*personnel costs*) :

- Harmonisation, standardisation des formats et des structures des fichiers ;
- Organisation de la collecte de données personnelles (formulaires de consentement, etc.) ;
- Préparation des données en vue de l'archivage (conversion, sélection, etc.) ;
- Documentation des données et des métadonnées ;
- Coordination des activités de gestion des données ;
- Anonymisation des données, transcription d'entretiens ;
- Gestion, paiement des droits d'auteur ;
- Préparation des données en vue de leur publication ;
- Nettoyage des données en vue de leur dépôt dans un entrepôt ;
- Etc.

Afin d'estimer correctement les coûts de personnel pour ces actions, il convient d'estimer le temps de travail qui sera engendré et d'identifier le profil de la ou des personnes qui pourront les réaliser (étudiant, chercheur confirmé, data manager, etc.). Il faudra ensuite intégrer ce temps de travail supplémentaire dans les prévisions de recrutement des personnels du projet.

Du côté des frais généraux (*travel & subsistence costs*) :

- Montée en compétence des personnels, ce qui peut passer par la présence sur des journées d'étude, des séminaires, des écoles d'été, etc., ce qui engendre des dépenses éligibles ;
- Frais de mission liés à des réunions entre partenaires pour organiser la gestion des données ;
- Etc.

## Retour sur le Plan de gestion des données (PGD)

### Le PGD, qu'est-ce que c'est ?

Le PGD (ou DMP) est un document texte non contractuel qui vise à expliciter et décrire la manière dont les données seront obtenues, traitées, documentées, diffusées, archivées, échangées, etc. tout au long d'un projet.

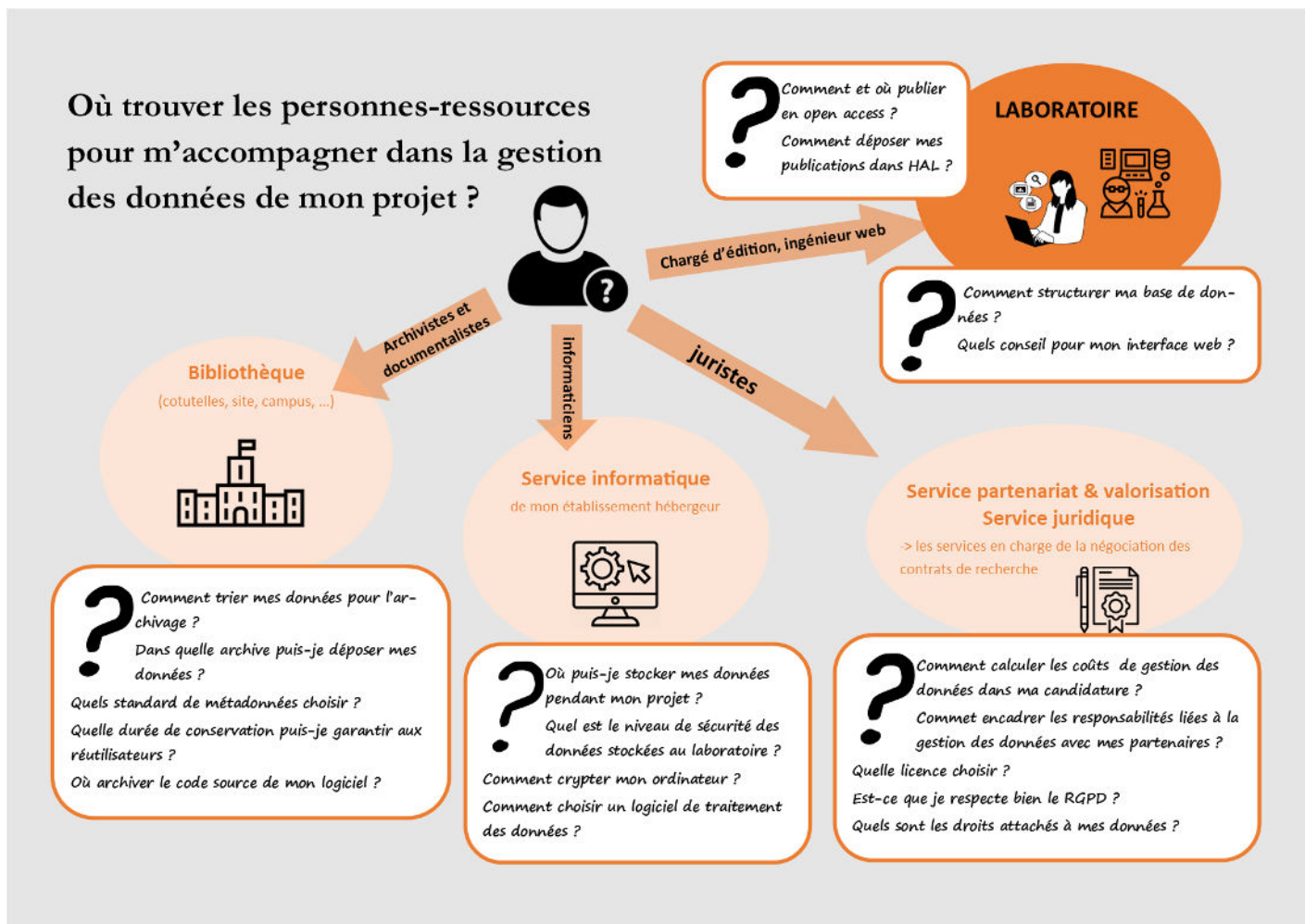
Dans le modèle ANR les sections sont les suivantes :

1. Responsabilités et ressources en matière de gestion des données
2. Description des données
3. Stockage et sauvegarde des données pendant le processus de recherche
4. Documentation et qualité des données
5. Dépôt, accès et partage des données
6. Exigences légales et éthiques
7. Conservation à long terme

[Découvrir d'autres modèles de PGD sur le site DMP OPIDoR.](#)

## Avec qui rédiger un PGD ?

Toutes les tâches liées à la gestion des données demandent un suivi rigoureux et un travail nécessairement collaboratif, car de nombreuses compétences-métier doivent être mobilisées (documentaliste, informaticien, juriste, etc.). Le schéma ci-dessous à vocation à aiguiller les chercheurs vers des interlocuteurs pertinents en fonction de la problématique qu'elles rencontrent :



## Quand rédiger un PGD ?

Une première version doit être remise à l'ANR 6 mois après le début du projet. Dans l'idéal, il s'agirait d'y réfléchir dès le montage de projet, pour pouvoir notamment anticiper les coûts et les indiquer dans le budget.

Cette première version n'a pas à répondre à l'ensemble des questions, dans la mesure où certaines d'entre elles ne peuvent trouver de réponse : terrain non encore exploré, inconnues liées à des événements externes, etc.

A mi-projet, une version révisée doit être produite en fonction des avancées de la recherche.

La dernière version, à la fin du projet détaille tout ce qui aura été fait. Remarque : ce document peut être associé aux données déposées dans un entrepôt ou publiées dans un data paper pour expliciter les contextes de production et de traitement des jeux de données.

## Comment traiter les données à caractère personnel ?

### Définitions

Les données à caractère personnel sont toutes les informations qui permettent d'identifier directement ou indirectement la personne (Article 4 du RGPD) :

- Les données directement identifiantes : nom, prénom, adresse, photo, voix, etc. ;
- Les données indirectement identifiantes : un numéro de téléphone, le croisement d'informations tel le fils du directeur de recherche, ce dernier habitant sur l'île de Batz, etc.

A noter :

- Les données anonymisées de manière irréversible, qui ne permettent plus la ré-identification d'une personne, ne sont pas soumises à la réglementation sur la protection des données personnelles ;
- Les données pseudonymisées sont les données à caractère personnel qui ne peuvent plus directement être attribuées à la personne concernée. Mais le recours à des d'informations supplémentaires, par exemple une table de correspondance, permet de ré-identifier cette dernière. Dans ce cas, la réglementation sur la protection des données personnelles s'applique.

### Les acteurs de la gestion des données

Le responsable du traitement (RT) est la personne, l'autorité publique ou l'organisme qui détermine la finalité et les moyens du traitement mis en œuvre. Pour les unités mixtes de recherche, le directeur d'unité est responsable de traitement (RT). Il doit donc s'assurer du respect de la réglementation sur la protection des données à caractère personnel et désigner un délégué à la protection des données. Il s'appuie pour cela sur les responsables scientifiques des projets conduits au sein de l'unité. Le plus souvent, lorsque le directeur d'unité est employé par le CNRS, il désigne la Déléguée à la protection des données du CNRS.

Un traitement de données est toute opération qui porte sur des données à caractère personnel, quel que soit le procédé, le support utilisé, informatisé ou non. Les données sont utilisées pour répondre à des objectifs/des finalités. Le traitement de données au sens « protection des données à caractère personnel » dépasse l'analyse ou l'exploitation de la donnée, il concerne la collecte, l'analyse, la réutilisation des données, l'archivage, ...

La finalité du traitement fait partie des principes essentiels de la réglementation. Tout traitement de données se réalise en fonction d'une finalité déterminée, explicite et légitime. Les données ne peuvent pas être traitées d'une manière incompatible avec la finalité définie.

Les données peuvent néanmoins être utilisées ultérieurement à des fins de recherche en apportant des garanties pour protéger la vie privée des personnes concernées par les données collectées.

Pour les recherches en sciences sociales, la problématique de recherche est souvent la finalité du traitement de données.

### Les droits des personnes

Toute personne a :

- Le droit d'accès à ses données ;
- Le droit d'être informée d'une violation des données en cas de risque élevé pour les personnes concernées ;
- Le droit d'opposition (elle peut s'opposer, pour des motifs légitimes, à l'utilisation de ses données personnelles) ;
- Le droit de rectification ;

- Le droit à l'effacement (demande de la suppression des données) ;
- Le droit à la portabilité (elle peut demander à recevoir les données qui la concernent dans un format structuré et lisible par machine) ;
- Le droit à une utilisation restreinte de ses données.

L'une des principales évolutions du RGPD (Règlement général sur la protection des données) porte sur l'obligation, pour le responsable de traitement (et le sous-traitant) de définir et d'organiser les mesures permettant de démontrer à tout moment la conformité à la réglementation.

### Qu'est-ce qu'une analyse d'impact relative à la protection des données (AIPD) ?

Les informations ci-après sont issues du [site internet de la CNIL](#).

L'AIPD un outil qui permet de construire un traitement conforme au RGPD et respectueux de la vie privée, lorsqu'un traitement de données personnelles est susceptible d'engendrer un risque élevé pour les droits et libertés des personnes concernées. Une AIPD doit obligatoirement être menée quand le traitement est « susceptible d'engendrer un risque élevé pour les droits et libertés des personnes concernées ».

Pour savoir s'il est nécessaire de réaliser une AIPD :

- Soit le traitement envisagé figure dans la liste des types d'opérations de traitement pour lesquelles la CNIL a estimé obligatoire de réaliser une analyse d'impact relative à la protection des données.
- Soit le traitement remplit au moins deux des neuf critères issus des lignes directrices du G29 :
  - o Evaluation/scoring (y compris le profilage) ;
  - o Décision automatique avec effet légal ou similaire ;
  - o Surveillance systématique ;
  - o Collecte de données sensibles ou données à caractère hautement personnel ;
  - o Collecte de données personnelles à large échelle ;
  - o Croisement de données ;
  - o Personnes vulnérables (patients, personnes âgées, enfants, etc.) ;
  - o Usage innovant (utilisation d'une nouvelle technologie) ;
  - o Exclusion du bénéfice d'un droit/contrat.

Exemple. Une entreprise met en place un traitement publicitaire visant à collecter les données de géolocalisation de plusieurs millions d'individus pour créer des profils publicitaires et leur afficher de la publicité ciblée en fonction de leurs déplacements. Ce traitement remplit le critère de la collecte à grande échelle et celui de la collecte de données sensibles (données de localisation), donc la réalisation d'une AIPD sera nécessaire.

### Pour aller plus loin

[Faire entrer la science ouverte dans son projet ANR : un guide pratique.](#)

[Fiche repère PPR Autonomie « La gestion des données » \(PDF, 159Ko\).](#)

[Fiche repère PPR Autonomie « La diffusion des données » \(PDF, 179Ko\).](#)

[Fiche repère PPR Autonomie « La réutilisation des données » \(PDF, 148Ko\).](#)

[Fiche repère PPR Autonomie « Accéder aux données du CASD » \(PDF, 191Ko\).](#)