

**Programme prioritaire de recherche (PPR) Autonomie****Synthèse de l'atelier « Prise en main du PGD » à destination des projets  
lauréats du 1<sup>er</sup> appel à projets du programme**

## Table des matières

|  |   |
|--|---|
| Préambule .....  | 1 |
| Quelques définitions .....   | 1 |
| L'obligation de diffusion des données de recherche, jusqu'où ?.....                                | 2 |
| La discussion autour des données de la recherche, au cœur de la vie scientifique d'un projet ..... | 2 |
| Le PGD, un outil évolutif et non-contraignant .....  | 3 |
| Un jeu de données comme ensemble cohérent de données .....   | 3 |
| Pour aller plus loin.....  | 3 |
| Contacts.....  | 4 |

## Préambule

Le **2 février 2023** l'équipe projet du PPR Autonomie réunissait **les projets lauréats du premier appel à projets<sup>1</sup>** pour un **webinaire de prise en main du plan de gestion de données**. Organisé en **partenariat avec l'Humathèque Condorcet**, et **animé par Emmanuel Collier**, coordinateur de l'activité de conseil à la rédaction des PGD<sup>2</sup>, ce webinaire poursuivait **l'objectif principal d'offrir un espace de discussion aux personnes en charge de la coordination de la rédaction du PGD** à remettre à l'ANR, dans sa première version, dans un délai de 6 mois après le démarrage du projet.

Les quatre projets lauréats étaient représentés.

Est proposée ci-après une synthèse des échanges et discussions qui se sont tenus à des fins de partage au plus grand nombre.

## Quelques définitions

Une **donnée** est une source transformée en fichier informatisé utilisée par un chercheur dans le cadre de ses travaux (retranscriptions d'entretiens, observations, codes source, programmes statistiques, bases de données, bibliographie, etc.).

Un **jeu de données** est un ensemble cohérent de données (un lot).

---

<sup>1</sup> [Télécharger le résumé des projets lauréats \(PDF, 411Ko\).](#)

<sup>2</sup> [Consulter le profil d'Emmanuel Collier sur le site internet ORCID.](#)

Les **métadonnées** sont un ensemble de champ d'informations qui décrit une donnée ou un jeu de données déposé dans un entrepôt. Quelques exemples de standards pour les sciences humaines et sociale : [DDI](#), [DublinCore](#), [DataCite](#).

Les **entrepôts de données** sont des espaces qui rendent accessibles selon les conditions et cadres définis par les créateurs des données, et dans le respect des textes juridiques, les données, ainsi que les métadonnées qui y sont associées. Ce sont les créateurs de données et jeux de données qui sont chargés de les y déposer. Certains entrepôts ont mis en place une curation, un contrôle a posteriori de la qualité des métadonnées renseignées lors du dépôt, c'est le cas de DataIned-Progedo et de l'entrepôt national recherche.data.gouv.

### L'obligation de diffusion des données de recherche, jusqu'où ?

Il existe une **« obligation d'ouverture » des données** de recherche (cf. Loi pour une République numérique). Elle concerne les données des recherches financées au minimum à 50 % sur fonds publics. Il est également fait mention de « mettre en œuvre l'obligation de diffusion des données de recherche » dans le 2ème Plan national pour la science ouverte de 2021.

Il n'y a **pas d'opposition entre diffusion d'un côté, et protection des données de la recherche** de l'autre. Diffuser les données ne signifie pas rendre accessible à tous toutes les données.

Par ailleurs, **l'obligation minimale fait référence au dépôt des données et à la publication**, ce qui n'implique pas une ouverture à tous. C'est le signalement public (d'où la notion de « publication ») qui est minimal : description des données et jeux de données avec des métadonnées.

Cela peut participer **de la visibilité des travaux de recherche** sur une thématique particulière.

Par ailleurs, de la même manière qu'une sélection est réalisée dans des documents papiers au moment de constituer des archives, il s'agit de **sélectionner les données et jeux de données qui seront finalement déposés dans un entrepôt**.

Ce sont bien **les chercheurs qui fixent les règles de sélection, d'accès et de réutilisation des données qui sont signalées et déposées dans un entrepôt**. Ils peuvent par exemple décider de les **mettre sous embargo**, de n'en donner l'accès qu'à une partie, etc.

### La discussion autour des données de la recherche, au cœur de la vie scientifique d'un projet

**Le PGD est un support de discussion scientifique** qui va accompagner le déroulé de la recherche. C'est au niveau du collectif de travail que doivent être abordées les questions de partage, de stockage, de diffusion, etc. des données. Il s'agit de **constituer une équipe en charge de sa rédaction**, qui intègre, en plus des chercheurs, des juristes, des informaticiens, le ou la déléguée à la protection des données, etc. qui vont se répartir les tâches, et **travailler en collaboration rapprochée avec le rédacteur principal en charge de la synthèse**.

Par ailleurs, il **peut être intéressant de choisir assez tôt dans le déroulé de la recherche le ou les entrepôts sur lesquels seront diffusés les jeux de données**, afin d'anticiper sur les renseignements nécessaires à la description des données par des standards de métadonnées, et l'accompagnement éventuel dispensé par les services de soutien à la recherche dans les établissements, dont l'Humathèque Condorcet.

Il est possible de mobiliser l'expertise des documentalistes dans ce choix qui identifieront, parmi les standards de métadonnées, celui qui décrit le mieux les données produites. Les documentalistes

pourront également se charger de la description et du dépôt des données et jeux de données dans les entrepôts.

### Le PGD, un outil évolutif et non-contraignant

**Le contenu du PGD va évoluer tout au long du déroulé de la recherche.** Ses différentes versions (trois pour les projets de recherche de plus de 30 mois) reflètent et synthétisent les différentes étapes de la gestion des données. Elles ne sont pas contraignantes et les différents chantiers (rubriques du PGD) peuvent être investis au fur et à mesure de l'avancée du projet et de la gestion de ses données. Ainsi, il ne s'agit pas, notamment dans sa première version, de régler d'emblée tous les sujets. Devant être remis dans les 6 mois après le démarrage du projet, il est normal que des questions restent en suspens. En revanche, la **version finale doit être complète, et faire état de tout ce qui aura été réalisé.**

### Un jeu de données comme ensemble cohérent de données

**La cohérence d'un jeu de données peut à la fois reposer sur le format des données rassemblées** (enregistrements vidéo ou sonores, retranscriptions d'entretiens, etc.) ou sur **la cohérence intellectuelle des données** (pour une monographie par exemple, il peut s'agir de retranscriptions d'entretiens, de compte-rendu d'observation, de croquis, de documentation institutionnelle, etc.).

**C'est aux chercheurs de choisir l'organisation des données en un ou plusieurs jeux de données** qui font sens.

Chaque **jeu de données est décrit par des métadonnées et peut être accompagné d'un document de type « Lisez-moi »** qui synthétise toutes les étapes et les choix opérés qui aboutissent au jeu de données signalé. Cela reste une recommandation. Autrement dit, c'est un **document de contextualisation du jeu de données**, qui permet d'aller plus loin que les métadonnées (standardisées).

Quelles que soient les règles encadrant l'accès au jeu de données, ce document est important car il permet aux producteurs et aux potentiels réutilisateurs des données de comprendre le contexte de production des données afin de pouvoir, le cas échéant, reproduire la recherche documentée par les données.

À **chaque jeu de données déposé** dans un entrepôt est **automatiquement attribué un identifiant unique et pérenne** comme le DOI (Digital object identifier) par exemple. Citer un jeu de données en spécifiant le DOI dans une publication scientifique (article, ouvrage, etc.) permet de renforcer la visibilité du ou des jeux de données sur lesquels se sont appuyés les chercheurs pour développer leur argumentation. Cette visibilité est renforcée par les métadonnées qui sont lisibles par les machines et donc identifiées par les moteurs de recherche. Les lecteurs de ces publications (chercheurs ou non) peuvent alors accéder aux données, ou à défaut à leur documentation.

### Pour aller plus loin

[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 » \(PDF, 191Ko\).](#)

[Fiche synthétique DoRANum « Métadonnées, standards et formats ».](#)

[Vidéo « ANR Tour 2023 - Accords de Consortium » sur YouTube.](#)

[Support de présentation de la vidéo « ANR Tour 2023 – Accords de Consortium » \(PDF, 3,7Mo\).](#)

[Vidéo « ANR Tour 2023 - Transfert des données, RGPD et collaborations internationales » sur YouTube.](#)

[Support de présentation de la vidéo « ANR Tour 2023 - Transfert des données, RGPD et collaborations internationales » \(PDF, 813Ko\).](#)

### Contacts

Pour contacter les équipes de l'Humathèque Condorcet et être accompagné dans ses démarches :  
services.humathèque[at]campus-condorcet.fr.

Pour contacter l'équipe du PPR Autonomie : ppr-autonomie[at]cnrs.fr.