

# Note d'orientation : le plan de gestion de données

## Appel à projets collaboratifs CollEx-Persée 2021\_2022

Un plan de gestion de données (PGD ou, en anglais, DMP pour Data Management Plan) est un outil de gestion de projet qui permet de garantir **une meilleure sécurité juridique et technique sur les données**, particulièrement dans le cas d'un projet géré par différents partenaires institutionnels.

Dans un PGD, les différents partenaires sont amenés à :

- identifier les grands types de données et métadonnées qui seront produites et/ou utilisées dans le cadre du projet ;
- identifier le statut juridique et les risques éventuellement associés ; préciser clairement le statut juridique des données à l'issue du projet (et le cas échéant, la licence à employer) ;
- décrire leur origine, leur mode de production et d'utilisation, leurs caractéristiques techniques (volume, format) ;
- indiquer quelles données seront diffusées, où et sous quelle(s) forme(s) ;
- indiquer quelles données ont vocation à être archivées, où et sous quelle(s) forme(s).

Un PGD doit aussi, le plus précisément possible, indiquer quel partenaire est responsable de quelle action sur tel ou tel type de données : production, récupération, diffusion, pérennisation... Cette **répartition claire des responsabilités**, menée le plus tôt possible dans le projet, est un facteur de succès. C'est aussi un élément permettant à l'organisme financeur de s'assurer du bon déroulement du projet. Ainsi, c'est un outil utile quel que soit le type de projet (numérisation enrichie, production d'un référentiel, etc.).

L'élaboration d'un PGD ne doit pas représenter une charge administrative supplémentaire, mais plutôt l'occasion de **se poser des questions** et d'identifier d'éventuels angles morts : des données dont le statut juridique n'est pas clarifié, dont le responsable n'est pas identifié... Il doit aider les porteurs de projet, pas les contraindre.

À ce titre, même s'il a vocation à recenser toutes les données utilisées ou produites dans le cadre du projet, le PGD requis par le dispositif des AAP CollEx-Persée ne vise pas nécessairement l'exhaustivité. Il doit, en revanche, être le plus complet et le plus détaillé possible pour toutes les données qui ont vocation à être conservées à long terme.

C'est aussi pour cela que le PGD peut être élaboré quelques mois après le début du projet (pas plus de six mois) et peut être régulièrement mis à jour au cours du projet. Le PGD final est un livrable du projet.

En définitive, c'est un instrument au service de la science ouverte en ce qu'il permet de produire des données FAIR (faciles à trouver, accessibles, interopérables et réutilisables).

Pour l'élaboration et la diffusion du PGD, l'utilisation de DMP Opidor (<https://dmp.opidor.fr/>) est recommandée. Cet outil gratuit et accessible à toute la communauté scientifique de l'ESR permet de choisir un modèle de PGD, de répondre aux différentes questions, d'attribuer des droits de lecture ou d'édition à différents participants (partenaires, financeurs...), et le cas échéant de le diffuser en ligne. Il est mis à disposition par l'INIST, un opérateur du GIS CollEx-Persée.

Comme modèle de PGD, l'utilisation du modèle ANR, qui est interdisciplinaire, est recommandé. Cependant, une institution pourra préférer l'utilisation d'un modèle différent, s'il est plus adapté.

À titre d'information, ce modèle était le suivant au 14/12/2020 :

#### **1. Description des données et collecte ou réutilisation de données existantes**

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

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

#### **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 ?

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

#### **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 ?

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

#### **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 ?

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 ?

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

#### **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 ?

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) ?

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

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

#### **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) ?

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) ?

#AAPCollExPersée21\_22