

# Découvrir des plans de gestion de données de la recherche en 4 points

1. Pourquoi gérer les données de la recherche ?
2. Définissez un plan de gestion de données (PGD), ou *Data Management Plan* (DMP)
3. Exemple de trame d'un plan de gestion de données (PGD)
4. Découvrez des modèles en anglais de plans de gestion de données

## 1. Pourquoi gérer les données de la recherche ?

Gérer des données de recherche fait partie intégrante du projet de recherche. La gestion des données de recherche (ou *Research data management*, RDM) requiert une organisation, une planification, et un suivi rigoureux tout au long de la vie de projet et au-delà pour assurer la pérennité, l'accessibilité et la réutilisation de ces données. La gestion des données de recherche répond aux objectifs suivants :

- elle accroît l'efficacité de la recherche, en facilitant l'accès et l'analyse des données par le chercheur qui a conduit la recherche ou par tout nouveau chercheur ;
- elle assure la continuité de la recherche par la réutilisation des données, tout en évitant la duplication des efforts ;
- elle favorise la diffusion élargie et accroît l'impact : des données de recherche correctement formatées, décrites et identifiées gardent une valeur à long terme ;
- elle permet d'assurer l'intégrité de la recherche et la validation des résultats. Des données de recherche exactes et complètes permettent également la reconstruction des événements et des processus qui ont conduit aux résultats ;
- elle réduit le risque de perte et renforce la sécurité des données par l'utilisation de dispositifs de stockage robustes et adaptés ;
- elle accompagne l'évolution actuelle de la publication : les revues scientifiques proposent de plus en plus que les données qui constituent la base d'une publication soient partagées et déposées dans un entrepôt de données accessible. De ce fait, la gestion des données de recherche facilite la soumission aux revues scientifiques d'articles s'appuyant sur des jeux de données documentés ;
- elle satisfait aux conditions de financement du projet par les bailleurs : ceux-ci s'intéressent de plus en plus à ce que les chercheurs font des données produites au cours d'un projet et conditionnent souvent le financement à l'ouverture de ces données pour qu'elles soient accessibles librement et gratuitement ;
- elle atteste de votre responsabilité : en gérant vos données de recherche et en les rendant disponibles, vous faites preuve d'une utilisation responsable du financement public de la recherche.

## **2. Définissez un plan de gestion de données (PGD), ou *Data Management Plan* (DMP)**

La gestion des données de recherche est facilitée par la définition d'un plan de gestion de données (ou *Data Management Plan*, DMP).

Un plan de gestion de données est un document formel explicitant la façon dont vous obtenez, documentez, analysez et utilisez vos données à la fois au cours de votre recherche et une fois le projet terminé. Il décrit dans le détail les méthodes et processus de création, de fourniture, de maintenance, de conservation et de protection des données.

## **3. Exemple de trame d'un plan de gestion de données (PGD)**

La trame de plan présentée ci-dessous est une adaptation de : **Checklist for a Data Management Plan**. V.4.0. Edinburgh, UK: Digital Curation Centre (DCC), 2013. Téléchargeable au format pdf à <http://www.dcc.ac.uk/resources/data-management-plans>, cette trame correspond au formulaire en ligne développé par **Digital Curation Centre's DMPonline tool** (<https://dmponline.dcc.ac.uk/>).

### **1) Informations administratives**

- nom et identifiant du projet
- description du projet
- agence(s) de financement
- nom et identifiant éventuel du responsable principal de projet
- contact pour les données de projet
- date de la 1<sup>re</sup> version
- date de la dernière mise à jour
- politiques associées au projet, incluant les instructions ou recommandations de l'agence de financement et de l'institution

### **2) Collection de données**

- description des données, incluant le type de données, le format et le volume
- jeux de données existants qui seront utilisés
- méthodes de collecte et de création des données
- système d'organisation, de nommage et de gestion des répertoires et des fichiers
- processus d'assurance qualité mis en œuvre

### **3) Documentation et métadonnées**

- informations nécessaires pour lire et interpréter les données
- organisation de la collecte et de la documentation
- standards de métadonnées adoptés

### **4) Ethique, cadre légal**

#### **41- Ethique**

- détails de l'accord de conservation et de partage des données
- étapes pour la protection de l'identité des participants

- étapes pour la sécurité du stockage et du transfert de données sensibles

#### **42- Droits de propriété intellectuelle et copyright**

- nom de(s) propriétaire(s) des données
- licence(s) pour la réutilisation des données (par exemple, une des licences *Creative Commons* ou *Open Data Commons*)
- restrictions d'utilisation par les tierces parties
- délai requis pour le partage de données (embargo lié à la publication dans une revue ou délai d'application d'un brevet)

### **5) Stockage, sauvegarde, et sécurité**

#### **51- Stockage et sauvegarde**

- lieu de stockage des données
- plan de sauvegarde
- personne ou équipe responsable de la sauvegarde
- procédures de récupération

#### **52- Sécurité**

- risques et leur gestion
- dispositif d'accès
- dispositif éventuel pour le transfert sûr et intègre des données collectées sur le terrain

### **6) Sélection et conservation**

- informations détaillées sur les données qui seront retenues, partagées et/ou conservées, et référence aux obligations contractuelles, légales ou réglementaires
- utilisations de recherche prévues des données
- durée de conservation des données au-delà du projet
- entrepôt ou archive de conservation des données et responsabilités associées
- temps et effort nécessaires à la préparation des données pour leur conservation et leur partage

### **7) Partage des données**

- étapes à mener pour faciliter la prise de connaissance (*discovery*) des données par les autres
- conditions de restriction du partage des données et détails de leur application dans l'accord de partage de données
- mécanisme de partage de données (via un entrepôt, sur demande expresse ou tout autre processus)
- délai de publication
- procédure éventuelle d'obtention d'un identifiant persistant pour les données

### **8) Responsabilités et moyens**

- nom de la personne responsable de la mise en œuvre du plan de gestion de données
- nom de la personne responsable de chaque activité de gestion des données
- équipements et logiciels requis (en addition à ceux existants fournis par l'institution)
- besoins additionnels d'expertise ou de formation
- charges imposées par les entrepôts de données

#### 4. Découvrez des modèles en anglais de plans de gestion de données

Voici une sélection non exhaustive de modèles de plans de gestion de données de recherche. Cette sélection comprend uniquement des modèles en anglais, aucun modèle en français n'ayant été identifié à la date du 17 juillet 2014.

Le classement retenu est celui de l'intérêt (par ordre décroissant) que représentent les différents supports selon le contexte lié par exemple au financement de recherches, selon leur niveau de précision, ou encore selon le champ thématique dans lequel ils s'inscrivent.

##### **Guidelines on data management in Horizon 2020 research projects: Version 1.0**

European Commission, 11 December 2013, 6 p.

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

Le guide de la Commission européenne rappelle les principes de libre accès qui sous-tendent la politique de l'Union européenne en matière de diffusion des résultats de la recherche. Il insiste sur la responsabilité des porteurs de projets du programme Horizon 2020 pour assurer la qualité, le partage et la sécurité des données de recherche.

Dans Horizon 2020, la gestion de données est évaluée selon des critères d'impact. Les candidats doivent décrire la stratégie de gestion de données de leur projet en répondant aux 4 questions : 1) Quels types de données le projet créera-t-il ou collectera-t-il ? 2) Quels standards de données seront-ils utilisés ? 3) Comment les données seront-elles exploitées et/ou partagées et/ou rendues accessibles pour leur vérification et leur réutilisation, et si les données ne peuvent être rendues disponibles, pourquoi ? 4) Comment les données seront-elles traitées et conservées ?

##### **Data Management Plan for Post-Graduate Research Projects**

University of Edinburgh, UK. 14 May 2014, 2 p.

[http://datalib.edina.ac.uk/xerte/USER-FILES/2-datalib-Nottingham/media/Data%20management%20plan\\_template\\_University\\_of\\_Edinburgh.pdf](http://datalib.edina.ac.uk/xerte/USER-FILES/2-datalib-Nottingham/media/Data%20management%20plan_template_University_of_Edinburgh.pdf)

Six questions structurent le plan de gestion proposé pour un projet de recherche : 1) Quelles données seront produites ? 2) Comment seront-elles documentées et décrites ? 3) Comment les données seront structurées et stockées ? 4) Quels sont les plans de partage de données et d'accès après la soumission de la thèse ? 5) Existe-t-il des conditions particulières liées aux données produites ? 6) Quels sont les plans d'archivage à long terme des données supportant la recherche menée ?

##### **Digital Curation Centre's DMPonline tool**

Edinburgh, UK : Digital Curation Centre (DCC)

<https://dmponline.dcc.ac.uk/>

DMPonline est une interface web développée par Digital Curation Centre (UK) pour la rédaction en ligne de plans de gestion de données. Après enregistrement gratuit et connexion, l'utilisateur est invité à répondre via un formulaire en ligne à une série de questions correspondant à la trame du plan de gestion de données fourni par DCC. L'ensemble des réponses saisies en mode texte peut être exporté à tout moment en format web ou en texte.

La trame de plan *Checklist for a Data Management Plan* qui correspond au formulaire en ligne est téléchargeable au format pdf à <http://www.dcc.ac.uk/resources/data-management-plans>. Elle est découpée en 8 sections : 1) Informations administratives ; 2) Collection de données ; 3) Documentation et métadonnées ; 4) Ethique et cadre légal ; 5) Stockage, sauvegarde, et sécurité ; 6) Sélection et conservation ; 7) Partage des données ; 8) Responsabilités et moyens.

### **University of California's DMPTool**

University of California Curation Center, US, 2014

<https://dmptool.org/>

DMPTool est une interface web développée par l'Université de Californie (US) pour rédiger en ligne des plans de gestion de données selon différents modèles. Après enregistrement gratuit et connexion, l'utilisateur est invité à sélectionner un modèle (*Template*) parmi 10 modèles d'agences de financement américaines : Gordon and Betty Moore Foundation, Gulf of Mexico Research Initiative, Institute of Education Sciences, Institute of Museum and Library Services, Joint Fire Science Program, National Institutes of Health, National Science Foundation, NEH-ODH, NEH-ODH: Office of Digital Humanities, U.S. Geological Survey DMP Guidance.

A l'issue de la saisie guidée par des questions, le plan peut être exporté aux formats html, pdf et txt. Il est organisé en 5 sections : 1) Types de données produites ; 2) Données et standards de métadonnées ; 3) Politiques d'accès et de partage des données ; 4) Politiques pour la réutilisation et la redistribution des données ; 5) Plans d'archivage et de conservation.

### **BBSRC Data Sharing Policy: Version 1.1**

Biotechnology and Biological Sciences Research Council, June 2010, 15 p.

[www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx](http://www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx)

Le plan de gestion de données préconisé par le BBSRC et à renseigner par les porteurs de projet s'articule autour de 9 sections : 1) Types et volumes de données produites ; 2) Standards et métadonnées adoptés pour la gestion des données ; 3) Relation avec d'autres données accessibles dans des entrepôts publics ; 4) Réutilisation prévisible ou envisagée des jeux de données ; 5) Partage des données ; 6) Mécanismes prévus pour rendre les données accessibles via des bases de données publiques ou sur demande ; 7) Restrictions de partage liées aux données propriétaires ou aux données brevetables ; 8) Délai de publication des données ; 9) Format des jeux de données finaux.

### **ESRC Research Funding Guide**

Economic and Social Research Council (ESRC), February 2014, 61 p.

[www.esrc.ac.uk/about-esrc/information/data-policy.aspx](http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx)

Trois guides au format pdf sont fournis, respectivement pour les porteurs de projet, pour les évaluateurs (*peers reviewers*), et pour citer des données de recherche.

Les plans de gestion de données proposés intègrent 4 sections : 1) La réutilisation des données ; 2) La gestion et le partage des données avec les responsabilités associées ; 3) Le partage de données sensibles ; 4) Toute information relative à la propriété intellectuelle, la confidentialité, la sécurité, la protection des données et l'éthique.

### **Full Data Management Plan Template (.doc)**

Natural Environment Research Council (NERC), 16 August 2012, 3 p., (.doc)

[www.nerc.ac.uk/research/sites/data/dmp/](http://www.nerc.ac.uk/research/sites/data/dmp/)

Le plan de gestion de données proposé aux porteurs de projet ne doit pas excéder une page une fois complété. La trame (format Word) comporte 5 sections : 1) Nom du projet ; 2) Porteur principal du projet ; 3) Production ou non de données ; 4) Centre(s) de données désigné(s) ; 5) Description des jeux de données produits et volume.

### **Marie-Claude Deboin**

Délégation à l'information scientifique et technique, Cirad

17 juillet 2014

#### **Informations**

*Comment citer ce document :*

*Deboin, M.C. 2014. Découvrir des plans de gestion de données de la recherche, en 4 points,. Montpellier (FRA): CIRAD, 6 p. <http://url.cirad.fr/ist/donnees-de-la-recherche>*

*Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons : Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International, disponible en ligne.: <http://creativecommons.org/licenses/by-nc-sa/4.0/deed.fr> ou par courrier postal à : Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.*

*Cette licence vous permet de remixer, arranger, et adapter cette œuvre à des fins non commerciales tant que vous créditez l'auteur en citant son nom et que les nouvelles œuvres sont diffusées selon les mêmes conditions.*