

COMMISSION EUROPÉENNE
Direction Générale de la Recherche et de l'Innovation

Lignes directrices pour la gestion des données dans Horizon 2020

Version 2.1
15 février 2016

Titre original : [Guidelines on Data Management in Horizon 2020, Version 2.1, 15 February 2016, European Commission, Directorate General for Research and Innovation](#)

© Union européenne, 1995-2016

Traduction française : INIST-CNRS, France

Avertissement : La traduction française ne saurait engager la responsabilité de la Commission européenne et de ses services. La diffusion et la réutilisation de l'original de cette traduction sont régies par la [décision de la Commission européenne du 12 décembre 2011 relative à l'utilisation des documents de la Commission \(2011/833/UE\)](#)

Introduction

Le programme Horizon 2020 comprend **une action pilote** limitée et cependant flexible **sur le libre accès aux données de la recherche**. Les projets participants devront élaborer un **plan de gestion des données (PGD)** précisant les données qui seront librement accessibles.

Le présent document a pour objectif d'aider les porteurs et bénéficiaires de projets au titre du programme Horizon 2020 à prendre leurs responsabilités quant à la **qualité de leurs données de la recherche, leur partage et leur sécurité**.

PLANS DE GESTION DES DONNÉES

Un nouvel élément dans Horizon 2020 est **l'utilisation de plans de gestion de données (PGD)**. Le PGD décrit le cycle de vie de gestion des données pour tous les jeux de données qui doivent être collectés, traités ou générés par un projet de recherche. Il doit englober :

- le traitement des données de la recherche pendant et après le projet
- les données qui seront collectées, traitées ou générées
- la méthodologie et les standards qui seront appliqués
- la possibilité ou non de partage/libre accès et selon quel mode
- la manière d'assurer la curation et la conservation.

L'utilisation d'un plan de gestion des données est **nécessaire pour les projets participant à l'action pilote libre accès aux données issues de la recherche**. D'autres projets sont invités à soumettre un PGD s'il est applicable à leurs prévisions de recherche.

Pour une description détaillée des prérequis de l'action pilote libre accès aux données issues de la recherche et de son champ d'application, veuillez consulter les [lignes directrices pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020](#).

PRÉSENTATION DES SOUMISSIONS ET ÉVALUATION

Toutes les propositions de projets soumises comme « actions de recherche et d'innovation » ou comme « actions d'innovation » comprennent une section sur la gestion des données de la recherche, évaluée sur la base du critère d'« impact ». Lorsque cela se justifie, comme porteur de projet, vous devez **présenter votre politique de gestion des données**, en répondant aux questions suivantes :


- Quels standards seront appliqués ?
- Comment ces données seront-elles exploitées et/ou partagées/rendues accessibles pour vérification et réutilisation ? Si les données ne peuvent être rendues disponibles, pour quelles raisons ?
- Comment les données seront-elles administrées et conservées ?

Votre politique devrait :

- être conforme aux accords actuels au sein du consortium concernant la gestion des données
- être en adéquation avec l'exploitation et les exigences des droits de propriété intellectuelle.

La section « gestion des données » est un aide-mémoire auquel il sera possible à l'avenir de se reporter pour les allocations de ressources et de budget consacrées à la gestion des données.

Les porteurs de projets prenant part à l'action pilote libre accès aux données issues de la recherche et ceux souhaitant présenter un PGD volontairement devraient inclure dans leur offre un livrable pour une première version de leur PGD dans les 6 premiers mois. (Consultez ci-dessous davantage de détails sur le processus de soumission de PGD).

 **Important** : la section sur la politique de gestion des données n'est pas un plan de gestion des données. Des plans de gestion des données totalement élaborés ne sont pas prévus à l'étape de la soumission et ne sont pas pris en compte lors de l'évaluation.

GESTION DES DONNÉES DE LA RECHERCHE DANS LE MODÈLE DE CONVENTION DE SUBVENTION

Il est fait référence à la gestion des données de la recherche dans le cadre de l'action pilote libre accès aux données issues de la recherche d'Horizon 2020 dans [l'article 29.3](#) du Modèle de convention de subvention, article qui s'applique à tous les projets participant à l'action pilote :

En ce qui concerne les données de la recherche numériques générées dans cette action (« données »), les bénéficiaires doivent :

- (a) *déposer dans un entrepôt de données de la recherche et prendre des mesures afin de rendre possibles l'accès, l'exploration, l'exploitation, la reproduction et la diffusion par un tiers – et ce gratuitement pour tout utilisateur – des éléments suivants :*
 - (i) *les données, y compris les métadonnées associées, nécessaires à la validation des résultats présentés dans des publications scientifiques le plus tôt possible ;*
 - (ii) *les autres données, y compris les métadonnées associées, comme indiqué, et dans les délais stipulés dans le plan de gestion des données (voir annexe 1) ;*
- (b) *fournir des informations – via l'entrepôt – sur les outils et instruments à la disposition des bénéficiaires et nécessaires à la validation des résultats (et, si possible, fournir les outils et instruments eux-mêmes).*

Ces dispositions sont expliquées en détail dans les [lignes directrices pour le libre accès](#) et dans le [Modèle de convention de subvention annoté](#).

PLANS DE GESTION DES DONNÉES DANS LE CADRE DU CYCLE DE VIE D'UN PROJET

Première version : les porteurs de projets prenant part à l'action pilote libre accès aux données issues de la recherche et ceux souhaitant présenter un PGD volontairement doivent fournir une première version du plan de gestion des données **dans les 6 premiers mois** du projet. Ces porteurs de projet doivent s'assurer d'inclure dans leur offre une première version de PGD en tant que livrable dans les 6 premiers mois.

Mises à jour : le plan de gestion des données devrait être mis à jour, si nécessaire, pendant la durée de vie du projet (*sous la forme de livrables*). Veuillez créer et soumettre une nouvelle version du PGD chaque fois que des changements importants interviennent dans votre projet comme :

- des nouveaux jeux de données
- des évolutions des orientations du consortium
- des facteurs externes.

Échéance des mises à jour :

- au minimum, mettre à jour le PGD à temps pour son évaluation à mi-parcours puis à son terme dans le cadre des revues de projets concernant ceux avec planification de revues
- des mises à jour de PGD sont requises au moins au terme des projets, pour les projets n'ayant pas planifié de revues
- à d'autres moments tels que définis par les consortiums du projet.

Rapports périodiques : consultez plus de détails dans le manuel en ligne sur la manière de remplir les [tableaux de reporting pour les publications, les livrables](#) et la [procédure de reporting continu](#) dans le système de gestion de subvention du portail du participant.

ASSISTANCE

Les **coûts** liés à la gestion des données dans Horizon 2020 sont **remboursables** pendant la durée du projet (voir [l'article 6.2.D.3](#) du Modèle de convention de subvention annoté).

Pour le moment la Commission ne propose pas d'outil de PGD. Les bénéficiaires peuvent générer leur propre plan de gestion des données ou utiliser des **outils qui sont compatibles avec les exigences** énoncées dans l'Annexe 1 – le modèle de plan de gestion des données – tel [l'outil DMPonline du Digital Curation Centre \(Royaume-Uni\)](#).

Une **assistance** pour la gestion des données de la recherche peut être obtenue auprès d'[OpenAIRE2020](#). Ces services d'assistance sont offerts essentiellement aux projets de recherche financés par Horizon 2020, avec une préférence donnée à ceux prenant part à l'action pilote libre accès aux données de la recherche.

ANNEXE 1 : modèle de plan de gestion de données (PGD)

Le plan de gestion des données (PGD) a pour but de permettre une analyse des principaux éléments de la politique de gestion des données que les demandeurs suivront, en prenant en compte tous les jeux de données qui seront générés par le projet.

Le PGD n'est pas un document figé. Il évolue au fur et à mesure du projet.

Le PGD doit aborder les points ci-dessous en traitant unitairement chaque jeu de données, et il doit refléter l'état de la réflexion du moment au sein du consortium concernant les données qui seront produites.

• Référence et nom du jeu de données

Identifiant du jeu de données qui va être produit.

• Description du jeu de données

Description des données qui seront générées ou collectées, leur origine (dans le cas d'une collecte), leur nature et leur échelle, à qui elles pourraient être utiles et si elles viennent en appui d'une publication scientifique, mais aussi les informations sur l'existence (ou la non-existence) de données similaires et leurs éventuelles intégrations et réutilisations.

• Standards et métadonnées

Indication des standards existants relatifs à la discipline. À défaut, un résumé expliquant quelles métadonnées seront créées et comment.

• Partage des données

Description des modalités de partage, y compris les procédures d'accès, les périodes d'embargo (le cas échéant), exposé succinct des dispositifs techniques de diffusion et logiciels nécessaires ainsi que les autres outils permettant la réutilisation de données, et indications sur l'accès libre ou restreint des données vis-à-vis de groupes spécifiques. Identification de l'entrepôt dans lequel les données seront stockées, s'il existe déjà et est identifié, en indiquant son type (institutionnel, ou de référence dans la discipline considérée, etc.).

Les raisons d'une éventuelle impossibilité de partage d'un jeu de données devront être indiquées (par ex. motifs de nature éthique, commerciale, de propriété intellectuelle, de sécurité, liés à la protection des données personnelles ou à la vie privée etc.).

• Archivage et conservation (y compris stockage et sauvegarde des données)

Description des procédures mises en place pour une conservation à long terme des données. Indication de la durée de conservation souhaitable des données, de leur volume final approximatif, des coûts associés et de la façon dont il est prévu de couvrir ces coûts.

ANNEXE 2 : prescriptions supplémentaires concernant les plans de gestion des données

Les présentes prescriptions peuvent s'appliquer à n'importe quel projet qui produit, collecte ou traite des données de la recherche, et ces indications sont ajoutées au présent document comme support de référence lors de l'élaboration de plans de gestion des données dans le cadre de projets d'Horizon 2020. Ce guide est structuré sous la forme d'une série de questions qui en principe devraient être explicitées pour tous les jeux de données produits par le projet.

Les données de la recherche scientifique devraient être facilement :

1. Découvrables

a. Questionnement PGD : les données et logiciels associés produits et/ou utilisés dans le projet sont-ils découvrables (et facilement repérables), et identifiables par un dispositif d'identification standard (p. ex. : *Digital Object Identifier*) ?

2. Accessibles

a. Questionnement PGD : les données et logiciels associés produits et/ou utilisés dans le projet sont-ils accessibles, sous quelles modalités, avec quel champ d'application, sous quelles licences (p. ex. : système de licences pour la recherche et l'enseignement, périodes d'embargo, exploitation commerciale, etc.) ?

3. Évaluables et compréhensibles

a. Questionnement PGD : les données et logiciels associés produits et/ou utilisés dans le projet sont-ils évaluable et compréhensibles par des tiers dans des contextes tels qu'une analyse scientifique ou une évaluation par des pairs ? (p. ex. : les jeux de données minimaux sont-ils examinés en même temps que les articles scientifiques au cours de la procédure d'évaluation par des pairs ? Les données sont-elles fournies de telle manière qu'il soit possible de juger de leur fiabilité et des compétences de leurs auteurs ?)

4. Utilisables au-delà du but premier de leur collecte

a. Questionnement PGD : les données et logiciels associés produits et/ou utilisés dans le projet sont-ils utilisables par des tiers, même longtemps après la collecte des données ? (p. ex. : les données sont-elles stockées en toute sécurité dans des entrepôts certifiés pour une conservation et une administration à long terme ? Les données sont-elles stockées avec le minimum de logiciels, de métadonnées et de documentations les rendant utilisables ? Les données sont-elles utiles pour les besoins du grand public et utilisables éventuellement à des fins intéressant des non-spécialistes ?)

5. Interopérables selon des standards qualitatifs spécifiques

a. Questionnement PGD : les données et logiciels associés produits et/ou utilisés dans le projet sont-ils interopérables, permettant des échanges de données entre chercheurs, institutions, organismes, pays, etc. (p. ex. : respect de standards relatifs à l'annotation et à l'échange des données, compatibilité avec les applications logicielles disponibles, recombinaisons possibles avec des jeux de données de diverses origines) ?