

Programme-cadre européen pour la recherche et l'innovation

Horizon 2020

Lignes directrices pour la gestion des données
dans Horizon 2020

Version 1.0
11 décembre 2013

Titre original : [Guidelines on Data Management in Horizon 2020, Version 1.0, 11 December 2013, European Commission, Research and Innovation](#)

© Union européenne, 1995-2014

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 \(2011/833/UE\)](#)

Lignes directrices pour la gestion des données dans Horizon 2020

Version du 16 décembre 2013

Introduction

Le programme Horizon 2020 mettra en œuvre une action pilote sur le libre accès des données de recherche.¹ Il sera exigé des projets participants qu'ils élaborent un plan de gestion des données (PGD) précisant les données qui seront libres.

Le présent document fournit les lignes directrices concernant la gestion des données dans Horizon 2020 et s'adresse à tous les porteurs et bénéficiaires de projets au titre du programme-cadre.² Ces lignes directrices sont censées leur indiquer comment assumer leurs responsabilités quant à la qualité de leurs données de recherche, leur partage et leur sécurité.

Présentation des soumissions et évaluation

Toutes les propositions de projets soumises dans le cadre des « actions de recherche et d'innovation » ainsi qu'aux « actions d'innovation » comprennent une section sur la gestion des données de recherche, évaluée sur la base du critère d'« impact ». Lorsque cela se justifie, les porteurs de projet doivent présenter brièvement et dans les grandes lignes leur politique de gestion des données, en répondant aux questions suivantes :

- Quels seront les types de données générées ou recueillies par le projet ?
- Quelles normes seront utilisées ?
- Comment ces données seront-elles exploitées ou partagées/rendues accessibles pour vérification et réutilisation ? Si les données ne peuvent être rendues disponibles, expliquer pourquoi.
- Comment les données seront-elles administrées et conservées ?

La description de cette politique devra être conforme aux accords au sein du consortium concernant la gestion des données et découler de ceux traitant de l'exploitation et la protection des résultats. Veuillez vous reporter au modèle B de proposition pour disposer d'instructions structurelles et autres.

La section gestion des données peut aussi être considérée comme un futur aide-mémoire auquel il sera possible de se reporter pour les allocations de ressources et de budget consacrées à la gestion des données.

¹ Voir les "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020". L'étendue de l'action pilote est également indiquée dans l'introduction du programme de travail d'Horizon 2020. Le projet pilote Libre accès aux données issues de la recherche (*Open Research Data*) sera suivie tout au long du programme Horizon 2020 dans une optique de déploiement accru de la politique de l'EU en faveur d'une recherche ouverte.

² D'autres informations générales sur les orientations en matière de gestion des données de recherche sont présentes dans la Déclaration des ministres des Sciences du G8 du 12 juin 2013 à cette adresse : <https://www.gov.uk/government/news/g8-science-ministers-statement>. La gestion des données de recherche répond par exemple au principe suivant, adopté dans cette déclaration : « Les données de recherches en libre accès devraient être facilement découvrables, accessibles, évaluables, compréhensibles, utilisables et partout où cela est possible, interopérables suivant des normes qualitatives spécifiques ».

Les PGD sont présentés comme suit dans le programme de travail d'Horizon 2020 pour 2014-2015 :

L'utilisation de plans de gestions des données (PGD) est un nouvel élément du programme Horizon 2020. Ceux-ci décrivent en détail les données qui seront générées par un projet, si et comment elles seront exploitées ou rendues accessibles pour leur vérification et réutilisation, et la façon dont elles seront administrées et conservées. L'utilisation d'un plan de gestion des données est exigée dans le cadre de projets présentés dans le cadre du projet pilote Libre accès aux données issues de la recherche (Open Research Data).

Le portail destiné aux participants offre une description détaillée du projet pilote Libre accès aux données issues de la recherche ainsi que du champ d'application de ses prescriptions.

Les projets prenant part au projet pilote Libre accès aux données issues de la recherche doivent fournir une première version de leur PGD dans les six premiers mois. Les projets soumis pour cette action pilote aussi bien que les projets présentant volontairement un PGD parce qu'il est dans la logique de leurs recherches, devraient veiller à ce que ce type d'élément livrable soit mentionné dans leur projet. Puisque les PGD sont censés évoluer tout au long du projet, des versions plus élaborées du plan pourront être incluses en tant qu'éléments livrables supplémentaires lors de phases ultérieures. Le PGD a pour but le financement du cycle de vie suivi par la gestion appliquée aux données que le projet sera amené à collecter, traiter ou générer.

Convention de subvention/projet pilote Libre accès aux données issues de la recherche dans Horizon 2020

Il est fait référence à la gestion des données de recherche dans l'article 29.3 du Modèle de convention de subvention (article qui s'applique à tous les projets participant au projet pilote Libre accès aux données issues de la recherche dans Horizon 2020).

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

- (a) Déposer dans un entrepôt de données de recherche et prendre des mesures afin de rendre possible l'accès, l'exploitation, la reproduction et la diffusion par un tiers – et ce gratuitement pour tous – 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).*

Un PGD, en tant que document définissant la façon dont les données de recherche sont utilisées pendant un projet de recherche, comme à son terme, est très important, et ce pour tous les aspects des projets répondant au projet pilote Libre accès aux données issues de la recherche que pour d'autres projets de recherche. Le projet devrait toujours contenir des descriptions claires et précises quant aux régimes d'accès prévus pour les ensembles de données qui ont été collectés, et plus particulièrement quand le projet répond à cette action pilote. Le paragraphe ci-dessous, issu du Modèle de convention de subvention, explicite davantage ce principe :

À titre exceptionnel, les bénéficiaires ne sont pas tenus de garantir une liberté d'accès à certaines parties spécifiques de leurs données de recherche si la réalisation de l'objectif

principal de l'action pilote, comme décrit en l'annexe 1, était par le fait mise en péril. En ce cas, le plan de gestion des données doit énoncer les raisons pour une mise sous embargo.

Évaluation des projets

Un PGD décrit le cycle de vie de la gestion des données de tous les ensembles de données collectés, traités ou générés par un projet de recherche. C'est un document décrivant la façon dont les données de recherche doivent être gérées pendant un projet de recherche, mais aussi à son terme, tout en expliquant quelles données doivent être collectées, traitées ou générées, et en suivant quelles méthodologies et normes, si et comment ces données doivent être partagées ou rendues accessibles, ainsi que la manière dont elles seront administrées et conservées. Le PGD n'est pas un document figé. Il évolue au fil du temps et gagne en précision et en substance au fur et à mesure que le projet évolue.

La première version du PGD est censée être envoyée dans les 6 mois qui suivent le début du projet. L'objectif du PGD est de suivre le modèle fourni par la Commission (voir annexe 1). Des versions plus élaborées du PGD pourront être envoyées à des stades plus avancés du projet. Il est nécessaire que le PGD soit mis à jour au moins deux fois : lors de son évaluation à mi-parcours puis à son terme pour l'adapter plus finement aux données générées et aux utilisations identifiées par le consortium, puisque toutes les données ou utilisations potentielles ne ressortent pas clairement dès le début du projet. De nouvelles versions du PGD devront être créées chaque fois qu'un changement important résultera de l'inclusion de nouveaux ensembles de données, de changements dans les orientations du consortium ou de facteurs externes. L'annexe 2 propose des informations complémentaires concernant le PGD dans ces versions plus élaborées.

Le soutien accordé à la gestion des données de recherche dans le cadre de projets financés par Horizon 2020 est planifié via des projets financés au titre du programme de travail *Infrastructures de recherche 2014-15* (appel intitulé *3 e-Infrastructures*). Les plans exacts pour la livraison de services seront disponibles sur le portail internet des participants fin 2014. Un soutien aux services d'aides complet est attendu fin 2015, en coordination avec d'autres initiatives nationales similaires. Ces services seront uniquement disponibles aux projets de recherche financés par Horizon 2020, avec une préférence pour les projets participants aussi au projet pilote Libre accès aux données issues 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 les ensembles 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 elle doit refléter l'état de la réflexion du moment à l'intérieur 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

Celle-ci décrit les données qui seront générées ou collectées, leur origine (s'il y a 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 possibles intégrations et réutilisations.

• Normes et métadonnées

Indication des normes existantes relatives 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 le caractère librement accessible des données ou si celles-ci seront limitées à des 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, liés à la protection des données personnelles, de la propriété intellectuelle, commerciale, à la protection de la vie privée, et de sécurité, 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, leur volume approximatif, les coûts associés et la façon dont ces coûts seront couverts.

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 recherche, et ces indications sont ajoutées au présent document comme support de référence lors de l'élaboration de plans de gestion de 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 ensembles de données produits par le projet.

Les données de recherche scientifiques 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 (ex. : un identifiant d'objet numérique comme le DOI, ou *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 quelle étendue, éventuellement sous quelles licences

(ex. : cadre de référence de licences pour la recherche et l'enseignement, les périodes d'embargo, l'exploitation commerciale, etc.) ?

3. Évaluables et compréhensibles

a. Questionnement PGD : les données et logiciels associés produits et/ou utilisés dans le projet sont-ils évaluables et compréhensibles par des tiers dans des contextes tels qu'une analyse scientifique ou une évaluation par des pairs (ex. : les ensembles 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 façon à permettre de juger de leur fiabilité et des compétences de leurs créateurs ?) ?

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 avoir la collecte des données (ex. : les données sont-elles stockées et sécurisées 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 utilisables pour les besoins du grand public et éventuellement à des fins intéressant des non-spécialistes ?) ?

5. Interopérables selon des normes qualitatives spécifiques

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