

Programme H2020

Lignes directrices pour la gestion des données FAIR dans Horizon 2020

Version 3.0
26 juillet 2016

Titre original : [H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 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 \(2011/833/UE\)](#)

Historique des modifications

Version	Date	Modification	Page
2.1	15/02/2016	<ul style="list-style-type: none">▪ Ce guide a également été publié dans le cadre du Manuel en ligne avec des contenus mis à jour et simplifiés.	1-13
3.0	26/07/2016	<ul style="list-style-type: none">▪ Cette version a été mise à jour dans le cadre de l'extension de l'action pilote libre accès aux données de la recherche.▪ Un nouveau plan de gestion des données (PGD) a été inclus.	1-13 7

1. Le contexte – Extension de l’action pilote libre accès aux données de la recherche dans Horizon 2020

 Il convient de faire la distinction entre le libre accès aux **publications** scientifiques évaluées par des pairs et celui aux **données** de la recherche :

- **publications** – le libre accès est une *obligation* dans Horizon 2020 ;
- **données** – la Commission mène une action pilote flexible, décrite ci-dessous, qui a été élargie.

Veillez consulter également les [lignes directrices pour le libre accès aux publications scientifiques et aux données de la recherche dans Horizon 2020](#).

Le présent document aide les bénéficiaires d’Horizon 2020 à rendre leurs données de recherche **trouvables, accessibles, interopérables et réutilisables (FAIR¹)**, pour s’assurer qu’elles sont bien gérées. Une bonne gestion des données de la recherche n’est pas une fin en soi, mais plutôt le passage clé qui mène à la découverte des connaissances et à l’innovation, aux données ultérieures ainsi qu’à l’intégration et à la réutilisation des connaissances.

Notez que ces lignes directrices ne s’appliquent pas dans leur totalité aux actions financées par l’ERC². Pour des informations et des prescriptions concernant le libre accès et l’action pilote libre accès aux données de la recherche à l’ERC, veuillez contacter erc-open-access@ec.europa.eu.

La Commission mène une action pilote flexible dans le cadre d’Horizon 2020 dénommée « action pilote libre accès aux données de la recherche » (pilote ORD). Le pilote ORD a pour objectif d’améliorer et de maximiser l’accès aux données de la recherche générées par les projets Horizon 2020 et leur réutilisation, et prend en compte le besoin d’équilibrer l’ouverture et la protection des informations scientifiques, la commercialisation et les droits de propriété intellectuelle, la protection de la vie privée, la sécurité ainsi que la gestion des données et les questions de conservation.

Dans les programmes de travail 2014-2016, le pilote ORD comprenait uniquement une sélection de domaines d’Horizon 2020. Dans le cadre de la version révisée du programme de travail 2017, [l’action pilote libre accès aux données de la recherche a été élargie pour couvrir tous les domaines thématiques d’Horizon 2020](#).

Alors que le libre accès aux données de la recherche devient ainsi applicable par défaut dans Horizon 2020, la Commission reconnaît également qu’il existe de bonnes raisons pour maintenir non accessible une partie voire même la totalité des données de la recherche générées dans un projet.

La Commission fournit par conséquent des possibilités bien établies de désengagement (*opt-out*) à n’importe quelle étape, à savoir :

- pendant la phase de candidature ;
- pendant la phase de préparation de convention de subvention (GAP³) ;
- après la signature de la convention de subvention.

¹ FAIR est l’acronyme anglais de « Findable, accessible, interoperable and reusable »


² European Research Council

³ Grant Agreement Preparation

Le pilote ORD s'applique essentiellement aux données nécessaires pour valider les résultats présentés dans les publications scientifiques. D'autres données peuvent également être fournies par les bénéficiaires à titre volontaire, comme il est précisé dans leurs plans de gestion des données. Les coûts liés au libre accès aux données de la recherche peuvent être déclarés comme des coûts éligibles à toute subvention d'Horizon 2020.

La participation au pilote ORD ne fait pas partie de l'évaluation des propositions. En d'autres termes, les propositions ne sont pas évaluées de manière plus favorable en raison de leur participation au pilote et elles ne sont pas pénalisées si elles n'y participent pas.

Pour en savoir plus sur le libre accès aux données de la recherche, veuillez également consulter le [Modèle de convention de subvention annoté](#) d'Horizon 2020.

 Participer au pilote ORD ne signifie pas nécessairement rendre accessible toutes vos données de recherche. Le pilote ORD suit plutôt le principe « **aussi ouvert que possible, aussi fermé que nécessaire** » et vise à promouvoir une saine gestion des données comme part essentielle des bonnes pratiques de la recherche.


2. Plan de gestion des données – définition générale

Les plans de gestion des données (PGD) constituent un *élément clé* d'une bonne gestion des données. Un PGD décrit le cycle de vie de gestion des données pour celles qui doivent être collectées, traitées et/ou générées par un projet Horizon 2020. Afin de rendre les données de recherche trouvables, accessibles, interopérables et réutilisables (FAIR), un PGD devrait comprendre des informations sur :

- le traitement des données de la recherche pendant et après la fin du projet ;
- les données qui seront collectées, traitées et/ou générées ;
- la méthodologie et les standards qui seront appliqués ;
- la possibilité ou non de partage/libre accès ;
- la manière d'assurer la curation et la conservation des données (y compris après la fin du projet).

Un PGD est requis pour tous les porteurs de projets participant au pilote ORD élargi, à moins qu'ils ne se désengagent du pilote ORD. Cependant, les porteurs de projets qui se désengagent sont encore encouragés à soumettre un PGD à titre volontaire.

3. Proposition, soumission et évaluation

 Le fait qu'un projet proposé participe ou non au pilote ORD n'affecte pas son évaluation. En d'autres termes, les propositions ne seront pas pénalisées si elles sont retirées du pilote ORD élargi.

Puisque la participation au pilote ORD n'est pas un critère d'évaluation, la proposition **ne** devrait **pas** contenir un PGD complètement élaboré. Cependant, une bonne gestion des données de la recherche en tant que telle devrait être évaluée selon le

critère d'impact pertinent par rapport au projet. Votre dossier de candidature devrait répondre 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 gérées et conservées ?

Votre politique devrait également :

- 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.

Vous devriez également veiller à la planification budgétaire et des ressources concernant la gestion des données et inclure dans votre proposition un livrable pour une première version de votre PGD au plus tard dans les 6 premiers mois.

4. Les plans de gestion des données de la recherche dans le cadre du cycle de vie d'un projet

Une fois un projet approuvé quant à son financement et qu'il a démarré, vous **devez soumettre une première version de votre PGD** (comme un livrable) dans les 6 premiers mois du projet. La Commission fournit un modèle de PGD en annexe, dont l'usage est recommandé mais facultatif.

Il est nécessaire de **mettre à jour** le PGD pendant la durée du projet chaque fois que des modifications significatives interviennent, telles que (mais sans s'y limiter) :

- de nouvelles données ;
- des changements dans les orientations du consortium (p. ex. un potentiel d'innovation, une décision de déposer une demande de brevet) ;
- des changements dans la composition du consortium et des facteurs externes (p. ex. des nouveaux membres intégrant le consortium ou d'anciens membres le quittant).

Le PGD devrait au minimum être mis à jour à temps pour l'évaluation périodique du projet. S'il n'y a pas d'autres examens périodiques prévus dans le cadre de la convention de subvention, alors une telle mise à jour doit être réalisée à temps pour l'examen final au plus tard. En outre, le consortium peut définir un calendrier d'examen indiqué dans le PGD.

Rapports périodiques

Pour obtenir des informations générales sur les rapports périodiques, veuillez consulter les chapitres suivants du manuel en ligne :

- comment remplir [les tableaux de reporting pour les publications, les livrables](#) ;
- la [procédure de reporting continu](#) dans le système de gestion de subvention.

5. Assistance

Remboursement des coûts

Les coûts liés au libre accès aux données de la recherche dans Horizon 2020 sont éligibles au remboursement pendant la durée du projet selon les conditions définies dans la convention de subvention d'Horizon 2020, en particulier dans [l'article 6](#) et [l'article 6.2.D.3](#), mais également dans d'autres articles pertinents pour la catégorie de coût choisi.

Plan de gestion des données

Un **modèle** de PGD est fourni dans l'annexe 1. Bien que la Commission ne propose pas actuellement son propre outil en ligne pour les plans de gestion de données, les bénéficiaires peuvent générer des PGD en ligne, en utilisant des outils qui sont compatibles avec les exigences énoncées dans l'annexe 1 (cf. également la section 7 de l'annexe 1).

ANNEXE 1

Modèle de plan de gestion de données FAIR dans Horizon 2020

Version: 26 juillet 2016

Introduction

Ce modèle de PGD FAIR dans Horizon 2020 a été conçu pour être applicable à tout projet d'Horizon 2020 qui produit, collecte ou traite les données de la recherche. Vous devriez élaborer **un seul PGD pour votre projet** afin de couvrir son approche globale. Cependant, si des problèmes particuliers se présentent pour des jeux de données individuels (p. ex. concernant leur ouverture), vous devriez l'expliquer clairement.

Gestion des données FAIR

D'une manière générale, vos données de recherche devraient être « FAIR », c'est-à-dire trouvables, accessibles, interopérables et réutilisables. Ces principes précèdent les choix de mise en œuvre et n'impliquent pas nécessairement une technologie spécifique, un standard ou une solution de mise en pratique.

Ce modèle n'est pas conçu pour une mise en œuvre technique stricte des principes FAIR, il s'inspire plutôt de FAIR en tant que concept général.

Plus d'informations sur FAIR :

[Principes des données FAIR \(forum de discussion de FORCE 11\)](#)

[Principes FAIR \(article dans la revue *Nature*\)](#)

Structure du modèle

Le modèle est une **série de questions** auxquelles vous devriez répondre avec un niveau de détails approprié au projet.

Il n'est pas exigé de fournir des réponses détaillées à toutes les questions dans la première version du PGD qui doit être soumis dans les 6 premiers mois du projet. Le PGD est plutôt conçu pour être un *document vivant* dans lequel les informations peuvent être mises à disposition à un degré plus fin de granularité au moyen de mises à jour lorsque la mise en œuvre du projet progresse et quand des modifications significatives se produisent. Par conséquent, les PGD devraient disposer d'un numéro de version visible et comprendre un calendrier de mises à jour. Le PGD devrait au minimum être mis à jour dans le contexte de l'évaluation périodique du projet. S'il n'y a pas d'autres examens périodiques prévus dans le cadre de la convention de subvention, alors une mise à jour doit être réalisée à temps pour l'examen final au plus tard.

Dans la suite du document, les sections principales qui doivent être couvertes par le

PGD sont exposées. A la fin du document, le tableau 1 présente un résumé de ces éléments sous forme de liste à puces.

Le modèle lui-même peut être mis à jour selon l'évolution des orientations.

1. Résumé des données

Quel est l'objet de la collecte/génération des données et son rapport avec les objectifs du projet ?

Quels types et formats de données le projet va générer/collecter ?

Allez-vous réutiliser des données existantes et de quelle manière ?

Quelle est l'origine des données ?

Quel est le volume prévu des données ?

À qui pourraient-elles être utiles (« utilité des données ») ?

2. Données FAIR

2.1. *Rendre les données trouvables, y compris la fourniture des métadonnées*

Les données produites et/ou utilisées dans le projet sont-elles découvrables avec les métadonnées, identifiables et localisables par un dispositif d'identification standard (p. ex. des identifiants pérennes et uniques tels que les DOI⁴) ?

Quelles conventions de nommage appliquez-vous ?

Des mots-clés de recherche seront-ils fournis pour optimiser les possibilités de réutilisation ?

Attribuez-vous des numéros de version visibles ?

Quelles métadonnées seront créées ? Au cas où des standards de métadonnées n'existent pas dans votre discipline, veuillez exposer quel type de métadonnées sera créé et de quelle manière.

2.2. *Rendre les données librement accessibles*

Quelles données produites et/ou utilisées dans le projet seront rendues librement accessibles par défaut ? Si certains jeux de données ne peuvent pas être partagés (ou nécessitent d'être partagés sous restrictions), expliquer pourquoi, en distinguant clairement les raisons légales et contractuelles des restrictions volontaires.

Notez que dans les projets multi-bénéficiaires, il est également possible pour des bénéficiaires particuliers de maintenir leurs données confidentielles si des dispositions appropriées sont établies dans la convention de subvention et sont conformes aux motifs de *désengagement*.

Comment les données seront-elles rendues accessibles ? (p. ex. en les déposant dans un entrepôt) ?

Quels outils logiciels ou méthodes sont nécessaires pour accéder aux données ?

La documentation sur le logiciel nécessaire pour accéder aux données est-elle incluse ?

Est-il possible d'inclure le logiciel adéquat (p. ex. en code open source) ?

⁴ Digital Object Identifier

Où seront déposés les données et métadonnées associées, la documentation et le code ? Il convient d'accorder la préférence à des entrepôts certifiés qui soutiennent le libre accès si possible.

Avez-vous envisagé des modalités adéquates avec l'entrepôt identifié ?

S'il existe des restrictions d'utilisation, comment l'accès sera-t-il fourni ?

Un comité d'accès aux données est-il nécessaire ?

Les conditions d'accès sont-elles bien décrites ? (c.-à-d. une licence lisible par une machine) ?

Comment l'identité de la personne ayant accès aux données sera-t-elle vérifiée ?

2.3. *Rendre les données interopérables*

Les données produites dans le projet sont-elles interopérables, c'est-à-dire permettant l'échange de données et leur réutilisation entre chercheurs, établissements, organisations, pays, etc. (c.-à-d. en respectant des standards pour les formats, aussi conformes que possible aux applications logicielles disponibles (ouvertes), et en particulier en facilitant des recombinaisons avec des jeux de données de diverses origines) ?

Quels vocabulaires de données et de métadonnées, standards ou méthodologies appliquerez-vous pour rendre vos données interopérables ?

Utiliserez-vous des vocabulaires standards pour tous les types de données présents dans votre jeu de données, afin de permettre une interopérabilité interdisciplinaire ?

Au cas où il est inévitable que vous utilisiez ou génériez des ontologies ou des vocabulaires inusuels spécifiques à un projet, fournirez-vous les alignements avec les ontologies les plus fréquemment utilisées ?

2.4. *Accroître la réutilisation des données (au moyen de licences explicites)*

Quelles licences seront attribuées aux données afin de permettre la réutilisation la plus large possible ?

Quand les données seront-elles rendues accessibles pour leur réutilisation ? Si un embargo est demandé pour s'octroyer du temps afin de publier ou déposer des brevets, précisez pourquoi et pendant combien de temps il s'appliquera, en gardant à l'esprit que les données de la recherche devraient être rendues accessibles dès que possible.

Les données produites et/ou utilisées dans le projet sont-elles utilisables par des parties tierces, en particulier après la fin du projet ? Si la réutilisation de certaines données est restreinte, expliquez pourquoi.

Pendant combien de temps est-il prévu que les données demeurent réutilisables ?

Les procédures d'assurance qualité des données sont-elles décrites ?

Outre les principes FAIR, les PGD devraient également aborder :

3. Allocation de ressources

Quels sont les coûts pour rendre FAIR les données dans votre projet ?

Comment ces coûts seront-ils couverts ? Notez que les coûts liés au libre accès aux données de la recherche sont éligibles dans le cadre de la subvention d'Horizon 2020 (s'ils sont conformes aux conditions de la convention de subvention).

Qui sera responsable de la gestion des données dans votre projet ?

Les ressources pour une conservation à long terme sont-elles examinées (coûts et valeur potentielle, qui décide de conserver quelles données de quelle manière et pendant combien de temps) ?

4. Sécurité des données

Quelles dispositions sont en vigueur pour la sécurité des données (y compris la récupération des données ainsi que le stockage sécurisé et le transfert des données sensibles) ?

Les données sont-elles stockées en toute sécurité dans des entrepôts certifiés pour une conservation et une gestion à long terme ?

5. Aspects éthiques

Y a-t-il des questions d'ordre éthique ou juridique qui peuvent avoir un impact sur le partage de données ? Ces questions peuvent être examinées dans le contexte de l'évaluation éthique. Si cela est pertinent, incluez les références aux livrables éthiques et au chapitre d'éthique dans la Description de l'action (DoA).

Le consentement éclairé pour le partage de données et la conservation à long terme est-il inclus dans des questionnaires traitant des données à caractère personnel ?

6. Autres questions

Faites-vous usage d'autres procédures nationales/sectorielles/départementales/de financeurs pour la gestion des données ? Si oui, lesquelles ?

7. Assistance supplémentaire dans l'élaboration de votre PGD

L'Alliance RDA (Research Data Alliance) fournit un [annuaire des standards de métadonnées](#) qui peut être interrogé pour trouver des standards spécifiques à une discipline et les outils associés.

L'outil [EUDAT B2SHARE](#) comprend un assistant de licence intégré qui facilite la sélection d'une licence appropriée pour les données de la recherche.

Des listes utiles d'entrepôts de données comprennent :

- le répertoire [Registry of Research Data Repositories](#) ;
- quelques entrepôts comme [Zenodo](#), fruit d'une collaboration entre OpenAIRE et le CERN, permettent aux chercheurs de déposer à la fois des publications et des données, tout en fournissant des outils pour les relier ;
- d'autres outils utiles comprennent [DMPonline](#) et des plateformes pour effectuer des observations scientifiques individuelles disponibles comme [ScienceMatters](#).

TABLEAU RÉCAPITULATIF n° 1

La gestion des données FAIR en bref : les questions à examiner dans votre PGD Horizon 2020

Ce tableau fournit un résumé des questions à traiter dans le Plan de Gestion des Données (PGD), comme indiqué à l'Annexe 1. Vous devriez consulter l'annexe et le texte principal des lignes directrices pour obtenir de plus amples indications.

Élément du PGD	Questions à examiner
1. Résumé des données	<ul style="list-style-type: none">• Indiquer l'objectif de la collecte/génération de données.• Expliquer la relation par rapport aux objectifs du projet.• Préciser les types et formats des données générées/collectées.• Préciser si des données existantes sont réutilisées (le cas échéant).• Préciser l'origine des données.• Indiquer le volume prévu des données (s'il est connu).• Mettre en évidence le potentiel de réutilisation des données : à qui seront-elles utiles ?
2. Données FAIR 2.1. Rendre les données trouvables, y compris la fourniture des métadonnées	<ul style="list-style-type: none">• Souligner la trouvabilité des données (fourniture des métadonnées).• Décrire la procédure d'identification des données et faire référence à un système standard d'identification. Utilisez-vous des identifiants pérennes et uniques tels que les DOI ?• Définir les conventions de nommage utilisées.• Présenter la démarche de recherche par mot-clé.• Décrire l'approche utilisée pour un versionnage explicite.• Préciser les standards de création de métadonnées (le cas échéant). Si aucun standard n'existe dans votre discipline, décrire comment les métadonnées seront créées ainsi que leur type.

<p>2.2 Rendre les données librement accessibles</p>	<ul style="list-style-type: none"> • Préciser quelles données seront librement disponibles. Si certaines données ne sont pas mises à disposition, le justifier. • Préciser comment les données seront rendues disponibles. • Préciser quelles techniques ou outils logiciels sont nécessaires pour accéder aux données. La documentation sur le logiciel nécessaire pour accéder aux données est-elle fournie ? Est-il possible d'intégrer le logiciel adéquat (p. ex. en code open source) ? • Préciser où les données et les métadonnées associées, la documentation et le code sont déposés. • Préciser comment l'accès sera fourni au cas où des restrictions s'appliquent.
<p>2.3. Rendre les données interopérables</p>	<ul style="list-style-type: none"> • Évaluer l'interopérabilité de vos données. Préciser quels vocabulaires décrivent les données et métadonnées, quels standards ou méthodologies seront appliqués pour faciliter l'interopérabilité. • Préciser si vous utiliserez du vocabulaire standard pour tous les types de données présents dans votre jeu de données pour permettre une interopérabilité interdisciplinaire. Dans le cas contraire, fournirez-vous un alignement avec les ontologies les plus fréquemment utilisées ?
<p>2.4. Accroître la réutilisation des données (au moyen de licences explicites)</p>	<ul style="list-style-type: none"> • Préciser quelle licence sera attribuée à vos données afin de permettre la réutilisation la plus large possible. • Préciser quand les données seront accessibles pour leur réutilisation. Le cas échéant, préciser pour quelle raison et pendant combien de temps un embargo sur les données est nécessaire. • Préciser si les données produites et/ou utilisées dans le projet sont exploitables par des parties tierces, en particulier après la fin du projet. Si la réutilisation de certaines données est restreinte, expliquer pourquoi. • Décrire les processus d'assurance qualité des données. • Préciser la durée pendant laquelle les données resteront réutilisables.

3. Allocation de ressources	<ul style="list-style-type: none"> • Estimer les coûts pour rendre vos données conformes aux principes FAIR. Décrire comment vous comptez financer ces coûts. • Identifier clairement les responsabilités de gestion des données dans votre projet. • Décrire les coûts et la valeur potentielle de la conservation à long terme.
4. Sécurité des données	<ul style="list-style-type: none"> • Examiner la récupération des données ainsi que le stockage sécurisé et le transfert de données sensibles.
5. Aspects éthiques	<ul style="list-style-type: none"> • Pour être couvert dans le cadre de l'évaluation éthique, la section éthique de la Description de l'action et les livrables éthiques. Inclure des références et les aspects techniques liés s'ils ne sont pas couverts par la précédente évaluation.
6. Autres	<ul style="list-style-type: none"> • Mentionner les autres procédures nationales/sectorielles /départementales/de financeurs pour la gestion des données que vous utilisez (le cas échéant).