

# SCIENCE EUROPE

## GUIDE PRATIQUE POUR UNE HARMONISATION INTERNATIONALE DE LA GESTION DES DONNÉES DE RECHERCHE

Ce « Guide pratique pour une harmonisation internationale de la gestion des données de recherche » est la traduction française de « Practical Guide to the international alignment of research data management » publié en novembre 2018 par Science Europe, dont la production a été coordonnée par le Groupe de travail de Science Europe sur les données de recherche.

Pour de plus amples informations, veuillez contacter [office@scienceeurope.org](mailto:office@scienceeurope.org)

Le document original en langue anglaise est publié sous licence Creative Commons Attribution 4.0. Il est accessible sur le site de Science Europe :

[https://www.scienceeurope.org/wp-content/uploads/2018/12/SE\\_RDM\\_Practical\\_Guide\\_Final.pdf](https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf)

La traduction française a été réalisée dans le cadre des travaux du Collège Données du Secrétariat Permanent pour la Science Ouverte. Cette traduction est publiée sous licence Creative Commons Attribution 4.0

Le texte de l'annexe au présent document emprunte la traduction des principes FAIR réalisée par l'Inra et disponible à l'adresse suivante :

<https://www6.inra.fr/datapartage/Produire-des-donnees-FAIR>

## Contenu

Avant-propos du Professeur Stan Gielen .....	3
Introduction.....	4
EXIGENCES MINIMALES POUR LES PLANS DE GESTION DES DONNÉES .....	6
Introduction.....	7
Exigences minimales pour les plans de gestion de données.....	8
CRITÈRES DE SÉLECTION DES ENTREPÔTS DE CONFIANCE .....	10
Introduction.....	11
Critères de sélection des entrepôts de confiance.....	12
CONSEILS DE MISE EN OEUVRE .....	13
Traduction de ces exigences minimales dans un modèle de DMP .....	14
Conseils pour la sélection des entrepôts dignes de confiance.....	20
Notes et références .....	23

## Avant-propos du Professeur Stan Gielen

Membre du Conseil d'administration de Science Europe et Président de l'Organisation néerlandaise pour la recherche scientifique

Lors du Sommet européen Open Science Cloud (EOSC) en juin 2017, je me suis engagé à promouvoir l'harmonisation de la gestion des données de recherche (RDM) entre les organismes de financement de la recherche en Europe. Cet engagement a été à l'origine d'une initiative lancée à cette fin par Science Europe et l'organisation pour la recherche scientifique aux Pays-Bas (NWO) en janvier 2018. L'objectif de l'initiative était d'élaborer un ensemble de spécifications de base pour les plans de gestion des données (DMP), ainsi qu'une liste de critères pour la sélection d'entrepôts dignes de confiance dans lesquels les chercheurs peuvent déposer leurs données à des fins de partage.

À la lumière de la mise en œuvre d'EOSC et dans le contexte d'une tendance croissante au partage des données, ces spécifications et critères devraient contribuer à l'harmonisation des règles de gestion des données dans toute l'Europe. Cela facilitera la conformité aux exigences en matière de gestion de données, même lorsque les chercheurs travaillent avec différents financeurs et organismes de recherche.

Moins d'un an après son lancement, j'ai le plaisir de vous présenter les résultats de cette initiative. Ces spécifications de base pour les DMP et les critères de sélection des entrepôts dignes de confiance ont été mis au point par des experts d'organisations membres de Science Europe, qui ont sollicité des contributions supplémentaires de la part de parties prenantes externes pour assurer un large consensus.

Science Europe et NWO vont promouvoir ces exigences et critères de base afin de s'assurer qu'ils sont adoptés et soutenus par le plus grand nombre possible de parties prenantes.

Je suis fier que NWO fasse partie, au sein de Science Europe, des organisations pionnières qui ont l'intention de mettre en œuvre ces exigences et critères de base dans leurs politiques, dans le courant de l'année 2019. Avec le soutien d'autres organismes de recherche, tant à l'intérieur qu'à l'extérieur de Science Europe, ce guide peut servir de document de référence pour l'élaboration ou la révision des politiques de GDR dans toute l'Europe et au-delà.

Novembre 2018

Professeur Dr C.C.A.M. Gielen

## Introduction

Agences de financement de la recherche, organismes de recherche<sup>i</sup> et chercheurs individuels ont des exigences et des besoins différents en matière de gestion des données de recherche (RDM). Les exigences de base des plans de gestion des données (PGD) et les critères de sélection d'entrepôts dignes de confiance présentés dans ce guide fournissent aux organisations et aux communautés une base commune à partir de laquelle elles peuvent élaborer des politiques de RDM. Ce sont là des exigences minimales qui peuvent être complétées par d'autres exigences plus spécifiques en fonction des besoins des communautés ou des organisations.

Des données de recherche de qualité sont des éléments clés du processus de recherche. Les données de recherche devraient être disponibles en permanence, publiquement et gratuitement pour une réutilisation éventuelle. Au cours des dernières années, divers intervenants, des financeurs de la recherche aux éditeurs, ont approuvé un ensemble concis de principes, connus sous le nom de principes des données FAIR<sup>ii</sup>, pour améliorer la réutilisation des données. Les exigences minimales pour les DMP et les critères de sélection des entrepôts dignes de confiance présentés dans ce guide sont conformes aux principes FAIR, et vont même au-delà sur certains aspects tels que stockage et archivage des données en cours de projet et préservation à long terme<sup>iii</sup>. Une gestion des données basée sur ces exigences et critères minimaux aideront donc les chercheurs à s'assurer que les données sont FAIR. Il peut y avoir des raisons légitimes (liées à des spécificités du projet ou relatives à des données personnelles) pour en différer ou limiter l'accès, qui impose une approche équilibrée en matière d'ouverture des données de recherche.

Ce guide a été élaboré par des experts appartenant à des organisations membres de Science Europe. Tout au long du processus d'élaboration de ce guide, ils ont confronté les modèles existants aux pratiques actuelles de gestion de données. Des acteurs du milieu de la recherche ont également été consultés pour prendre compte les différents besoins<sup>iv</sup>.

Ce guide a été conçu de manière à être utile pour un grand nombre d'organisations et de chercheurs. Il met l'accent sur les questions de fond et ne fait pas référence aux aspects procéduraux de l'utilisation des DMP<sup>v</sup> et de choix des entrepôts, car ceux-ci peuvent varier considérablement d'une organisation à l'autre.

Ce guide est divisé en trois parties :

**Exigences minimales pour les plans de gestion des données** : six points que tout DMP devrait couvrir tous les aspects, illustrés avec des questions clés.

**Critères de sélection des dépôts dignes de confiance** : quatre points détaillant les critères que tout entrepôt de confiance doit respecter.

**Conseils de mise en œuvre** : renseignements détaillés et exemples pour guider la mise en œuvre des exigences et des critères dans les politiques d'établissement.

## COMMENT UTILISER CE GUIDE

Agences de financement de la recherche, organismes et communautés de recherche sont encouragés à utiliser **ces exigences minimales pour les plans de gestion des données** comme une base pour établir leurs propres modèles de DMP. Dans un deuxième temps, ce guide peut également servir de document de référence pour l'évaluation des DMP.

Les organismes de recherche et les chercheurs individuels peuvent se référer à ce guide pour la rédaction de leurs DMP.

Pour les autres acteurs du secteur de la recherche, ce guide sert de document de référence sur la façon dont un DMP devrait être structuré et utilisé.

Les **critères de sélection des entrepôts dignes de confiance** aideront les agences de financement, les organismes de recherche et les chercheurs individuels à identifier des entrepôts adéquats pour le stockage et le partage des données.

Les exigences minimales pour les DMP et les critères de sélection des entrepôts dignes de confiance peuvent être considérés comme deux documents autonomes et utilisés indépendamment. Il est toutefois recommandé de tenir compte de ces deux ensembles lors de l'élaboration ou de la mise à jour d'une politique de données institutionnelle ou disciplinaire, de façon à aligner au mieux les positions des institutions.

## **EXIGENCES MINIMALES POUR LES PLANS DE GESTION DES DONNÉES**

## Introduction

Les organismes de financement de la recherche et les organismes de recherche exigent de plus en plus souvent des plans de gestion des données. Ces plans aident le chercheur à prendre en compte tous les aspects pertinents de la gestion des données dès l'entame d'un projet de recherche. Un DMP devrait inciter les chercheurs à réfléchir à des solutions optimales en matière de traitement, d'organisation, de documentation et de stockage de leurs données.

À l'heure actuelle, les politiques de gestion des données de recherche varient beaucoup. Nombreux sont les agences de financement, organismes et communautés de recherche ayant élaboré leurs propres règles et modèles. Cela peut prêter à confusion pour les chercheurs et est d'autant plus problématique que de nombreux chercheurs obtiennent leur financement de sources différentes ; ils sont ainsi de plus en plus confrontés à des exigences différentes entre sources de financement et politiques institutionnelles. Il est urgent d'harmoniser les politiques de gestion des données afin d'apporter plus de clarté aux chercheurs. Les DMP ne devraient pas apparaître comme une tâche administrative supplémentaire, mais un moyen utile d'aide à la planification et la réalisation d'un projet de projet de recherche.

La suite de ce document présente six thèmes qui devraient être abordés dans les DMP, chacun d'eux accompagné de plusieurs questions d'orientation. Ces sujets et questions pour la mise en place d'un DMP constituent les exigences minimales que tout organisme de financement de la recherche doit demander au chercheur en vue d'élaborer un DMP utile. L'ordre des sujets à traiter peut être adapté en fonction de besoins spécifiques et de l'organisation souhaitée. Mais les six thèmes doivent être abordés dans le DMP.

Un exemple de modèle fournissant des conseils sur les aspects à approfondir dans le cadre d'un DMP se trouve à la page 14 du présent guide.

## Exigences minimales pour les plans de gestion de données

Dans l'élaboration de robustes plans de gestion des données, il est demandé aux chercheurs de traiter les points suivants et de répondre aux questions suivantes :

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

- a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?
- b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

### 2. Documentation et qualité des données

- a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?
- b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

### 3. Stockage et sauvegarde pendant le processus de recherche

- a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?
- b. 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

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

### 5. Partage des données et conservation à long terme

- a. 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 ?
- b. 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) ?
- c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?
- d. Comment l'application d'un identifiant unique et pérenne (comme le DOI) sera réalisée pour chaque jeu de données ?



## **6. Responsabilités et ressources en matière de gestion des données**

- a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c.-à-d. le gestionnaire des données) ?
- b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données soient FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

## **CRITÈRES DE SÉLECTION DES ENTREPÔTS DE CONFIANCE**

## Introduction

L'accès aux données est l'un des piliers d'une science saine et reproductible. De plus en plus d'organismes, d'institutions et d'agences de financement de la recherche exigent que les chercheurs déposent les données de recherche dans des entrepôts. Les chercheurs ont besoin d'identifier des entrepôts dignes de confiance où ils peuvent stocker leurs données à des fins de partage. Il n'existe actuellement aucune liste de référence pour de tels entrepôts de confiance, alors que les catalogues d'entrepôts existants en répertorient plus de 2 000. Toutefois, la maturité et la fiabilité de ces dépôts sont difficiles à évaluer. De nombreux entrepôts n'ont pas encore demandé à être certifiés par un organisme de certification reconnu. Identifier un entrepôt approprié peut donc représenter une tâche difficile pour les chercheurs, leurs institutions d'appartenance, et les agences de financement.

Dans certaines disciplines, les chercheurs travaillent avec des entrepôts disciplinaires qui ont déjà mis en place certaines politiques et normes qui répondent aux besoins spécifiques d'une communauté. D'autres entrepôts de données servent un public de chercheurs plus large, et en conséquence leurs politiques et leurs normes sont plus génériques.

Certains dépôts ont été certifiés comme entrepôts dignes de confiance par un ou plusieurs organismes de certification reconnus. Afin de faciliter le repérage par les chercheurs des entrepôts de confiance, il est fortement recommandé que les entrepôts non encore certifiés demandent leur labellisation.

Il est toujours recommandé de se référer en premier lieu à des entrepôts disciplinaires ou certifiés largement reconnus. Quand de tels entrepôts ne peuvent être identifiés, alors les critères de sélection d'entrepôts présentés dans ce guide peuvent être utilisés ?

La liste de critères présentée dans ce guide se compose d'un certain nombre de critères minimaux, organisé autour de quatre thèmes majeurs, que tous les entrepôts de confiance doivent satisfaire. Cette liste n'accorde pas la priorité à un critère plutôt qu'à un autre.

Des explications plus détaillées sur les critères de sélection des entrepôts de confiance trouvent à la page 20 du présent guide.

## Critères de sélection des entrepôts de confiance

Les entrepôts dignes de confiance doivent répondre aux critères minimaux suivants :

### 1. Attribution d'identifiants uniques et pérennes (PIDs)

- a. Permettre le repérage et l'identification des données
- b. Permettre la recherche, la citation et la récupération de données
- c. Permettant la gestion des versions des jeux de données

### 2. Métadonnées

- a. Permettre le repérage de données
- b. Permettre de référencer des informations pertinentes connexes, telles que d'autres jeux de données et publications
- c. Fournir de l'information librement accessible et mise à jour, y compris sur des données non publiées, protégées, rétractées, ou supprimées
- d. Utiliser des normes de métadonnées qui sont largement acceptées (par la communauté scientifique)
- e. S'assurer que les métadonnées sont récupérables par programme

### 3. Accès aux données et licences d'utilisation

- a. Permettre l'accès aux données dans des conditions bien définies
- b. Garantir l'authenticité et l'intégrité des données
- c. Permettre la récupération des données
- d. Fournir de l'information sur les licences et les autorisations d'accès aux données (idéalement sous forme lisible par programme)
- e. Garantir la confidentialité et respecter les droits des personnes concernées et des créateurs de données.

### 4. Préservation

- a. Assurer la pérennité des métadonnées et des données
- b. Faire preuve de transparence au sujet des missions, périmètre, politique de préservation, et des plans (y compris la gouvernance, la pérennité budgétaire, et plan de maintenance)

## **CONSEILS DE MISE EN OEUVRE**

## Traduction de ces exigences minimales dans un modèle de DMP

L'exemple suivant d'un modèle de plan de gestion des données est basé sur les exigences minimales pour les DMP<sup>vi</sup>. Ce minimum requis devrait être considéré comme un standard de base, tout en laissant la possibilité de formuler des spécifications additionnelles en fonction de besoins particuliers de certains domaines ou liées à des législations locale ou nationale.

Le modèle présenté ci-dessous se réfère aux 15 questions couvrant les six points principaux exigés pour une bonne gestion des données. Les présents conseils de mise en œuvre et explications supplémentaires ont pour but d'aider les chercheurs à renseigner le modèle et pour s'assurer que tous les aspects pertinents de la gestion des données sont effectivement couverts. Le tableau ci-dessous est un exemple de la façon dont les exigences minimales définies auparavant peuvent être traduites en un modèle de DMP. Il appartiendra à chaque organisation et à chaque discipline d'élaborer des modèles qui correspondent à leurs besoins.

### INFORMATIONS GÉNÉRALES

Renseignements administratifs

- Fournir des renseignements d'identification tels que le nom du déposant, numéro du projet, du programme de financement, version du DMP.

### 1 DESCRIPTION DES DONNÉES ET COLLECTE OU RÉUTILISATION DE DONNÉES EXISTANTES

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

- Expliquer quelles méthodologies ou quels logiciels sera utilisés si de nouvelles données sont recueillies ou produites.
- Énoncer les éventuelles contraintes à la réutilisation des données préexistantes.
- Expliquer comment la provenance des données sera documentée.
- Indiquer brièvement le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée mais écartée.

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

- Donnez des détails sur le type de données : par exemple numérique (bases de données, tableurs), textuel (documents), image, audio, vidéo, et/ou médias composites.
- Détailler le format des données : la manière selon laquelle les données sont codées pour le stockage, généralement reflétée par l'extension du nom de fichier (par exemple pdf, xls, doc, txt, ou rdf).
- Justifier l'utilisation de certains formats. Par exemple, les choix d'un format peuvent être guidés par l'expertise du personnel de l'organisme, ou par une préférence pour les formats ouverts, par les standards de format acceptés par les entrepôts de données, par l'usage largement répandu dans une communauté de recherche ou par le logiciel ou l'équipement qui sera utilisé.

- Privilégier les formats standards et ouverts car ils facilitent le partage et la réutilisation à long terme des données (plusieurs catalogues fournissent des listes de ces "formats préférés").
- Donnez des détails sur les volumes (qui peuvent être exprimés en espace de stockage requis (octets), et/ou en quantités d'objets, de fichiers, de lignes, et colonnes).

## 2 DOCUMENTATION ET QUALITÉ DES DONNÉES

### 2 a Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

- Indiquez quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données.
- Indiquez quelles standards de métadonnées seront utilisés (par exemple DDI, TEI, EML, MARC, CMDI).
- Utiliser les standards de métadonnées de communauté lorsque ceux-ci existent.
- Indiquez comment les données seront organisées au cours du projet, en mentionnant par exemple conventions de nommage, contrôle de version et structures des dossiers. Des données bien classées et gérées de façon cohérente seront plus faciles à retrouver, à comprendre et à réutiliser.
- Pensez à la documentation qui serait nécessaire pour permettre une réutilisation des données. Il peut s'agir notamment de l'information sur la méthodologie utilisée pour collecter les données, sur les procédures et méthodes d'analyse utilisées, sur la définition des variables, des unités de mesure, etc.
- Tenez compte de la façon dont ces informations seront obtenues et enregistrées par exemple dans une base de données avec des liens vers chacun des fichiers, dans un fichier texte de type « lisez-moi », dans les en-têtes de fichiers, dans un livre de référence (« code book ») ou dans les cahiers de laboratoire.

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

- Expliquer comment la qualité et la conformité de la collecte des données seront contrôlées et documentées. Il s'agit là de préciser les processus comme la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la revue par les pairs, ou la représentation basée sur des vocabulaires contrôlés.

## 3 STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

### 3 a Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

- Décrire l'endroit où les données seront stockées et sauvegardées au cours du processus de recherche et la fréquence à laquelle la sauvegarde sera effectuée. Il est recommandé de stocker les données dans au moins deux lieux distincts.
- Privilégier l'utilisation de systèmes de stockage robustes, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution d'origine. Le stockage des données sur des ordinateurs portables, des disques durs externes, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.

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

- Expliquez comment les données seront récupérées en cas d'incident.
- Expliquer qui aura accès aux données au cours du processus de recherche et comment l'accès aux données est contrôlé, en particulier dans le cadre de recherches menées en collaboration.
- Tenez compte de la protection des données, en particulier si vos données sont sensibles (par exemple données à caractère personnel, politiquement sensibles des informations ou secrets commerciaux). Décrire les principaux risques et la façon dont ils seront gérés.
- Expliquer quelle politique institutionnelle de protection des données institutionnelles est mis en œuvre.

## 4 EXIGENCES LÉGALES ET ÉTHIQUES, CODES DE CONDUITE

### 4 a 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 ?

- Lorsque vous manipulez des données à caractère personnel, veillez à ce que les lois sur la protection des données (par exemple, RGPD) soient appliquées, notamment:
  - Obtenir un consentement éclairé pour la préservation et/ou le partage de données personnelles.
  - Envisager l'anonymisation des données personnelles pour la préservation et/ou le partage (des données correctement anonymisées ne sont plus considérées comme des données personnelles).
  - Envisager la pseudonymisation des données personnelles (la principale différence avec l'anonymisation est que la pseudonymisation est réversible).
  - Envisagez le chiffrement des données, qui est considéré comme un cas particulier de pseudonymisation (la clé de cryptage doit alors être stockée séparément des données, par exemple par un tiers de confiance).
  - Expliquez si une procédure d'accès spécifique a été mise en place pour les utilisateurs autorisés à accéder aux données personnelles.

### 4 b 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 ?

- Expliquez qui sera le propriétaire des données, c'est-à-dire qui aura le droit d'en contrôler l'accès :
  - Expliquez quelles conditions d'accès s'appliqueront aux données ? Les données seront-elles librement accessibles, ou des restrictions seront-elles appliquées? Dans Si oui, lesquelles ? Envisager l'utilisation de licences concernant l'accès et la réutilisation des données.
  - Assurez-vous de couvrir, dans l'accord de consortium, ces questions de droits de contrôle d'accès aux données pour les projets multipartenaires et en cas de propriété partagée des données.



- Indiquez si les droits de propriété intellectuelle (par exemple la directive bases de données, droits *sui generis*) sont affectés. Dans l'affirmative, veuillez expliquer lesquels et comment cela sera-t-il traité.
- Indiquez s'il y a des restrictions sur la réutilisation des données fournies par des tiers.

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

- Déterminer si les questions d'éthique ont une incidence sur la façon dont les données sont stockées et transférées, qui peut les voir ou les utiliser, et quelle durée de conservation leur seront-elles appliquées. Démontrer que ces aspects sont bien pris en compte et planifiés.
- Adoptez les codes de conduite nationaux et internationaux et le code d'éthique institutionnel et vérifiez si une revue des pratiques (par exemple par un comité d'éthique) est requise pour ce qui concerne la collecte de données dans le cadre du projet de recherche.

### 5 PARTAGE DES DONNÉES ET CONSERVATION À LONG TERME

#### 5 a 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 ?

- Expliquez comment les données pourront être retrouvées et partagées (par exemple, par dépôt dans un entrepôt de données de confiance, indexé dans un catalogue, par l'utilisation d'un service de données sécurisé, par traitement direct des demandes de données, ou l'utilisation de tout autre mécanisme).
- Définissez le plan de préservation des données et fournir l'information sur la durée d'archivage pérenne des données.
- Expliquez à quel moment les données seront rendues disponibles. Indiquer les délais de publication prévus. Expliquez si une utilisation exclusive des données est revendiquée et, dans l'affirmative, pour quelle raison et pour combien de temps. Indiquez si le partage des données sera différé ou limité, par exemple pour des raisons de publication, pour protéger la propriété intellectuelle, ou dépôt de brevets.
- Indiquez qui pourra utiliser les données. S'il s'avère nécessaire de restreindre l'accès pour certaines communautés ou d'imposer un accord pour le partage de données, expliquez comment et pourquoi. Expliquez les mesures qui seront prises pour dépasser ou minimiser ces restrictions.

#### 5 b 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) ?

- Indiquez quelles données ne doivent pas être divulguées ou doivent être détruites pour des raisons contractuelles, légales, ou réglementaires.
- Indiquez comment il sera décidé quelles données garder. Décrire les données qui seront à préserver à long terme.
- Décrire les utilisations (et/ou les utilisateurs) prévisibles des données dans un cadre de recherche.
- Indiquez où les données seront déposées. Si aucun entrepôt reconnu n'est proposé, démontrer dans le plan de gestion des données que les données pourront être prises en

charge efficacement au-delà de la durée de financement du projet. Il est recommandé de démontrer que les politiques des entrepôts et les procédures de dépôts (y compris les normes de métadonnées, et coûts mis en œuvre) ont été vérifiés.

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

- Indiquez si les utilisateurs potentiels ont besoin d'outils spécifiques pour l'accès et la (ré)utilisation des données. Tenez compte de la durée de vie des logiciels nécessaires pour accéder aux données.
- Indiquez si les données seront partagées via un entrepôt, si les demandes d'accès seront traitées en direct, ou si un autre mécanisme sera utilisé ?

#### 5 d Comment l'application d'un identifiant unique et pérenne (comme le DOI) sera réalisée pour chaque jeu de données ?

- Expliquez comment les données pourraient être réutilisées dans d'autres contextes. Les identifiants persistants devraient être appliqués de manière à ce que les données puissent être localisées et référencées de façon fiable et efficace. Les identifiants persistants aident aussi à comptabiliser les citations et les réutilisations.
- Indiquez s'il sera envisagé d'attribuer aux données un identifiant persistant. Typiquement, un entrepôt pérenne digne de confiance attribuera des identifiants persistants.

## 6 RESPONSABILITÉS ET RESSOURCES EN MATIÈRE DE GESTION DES DONNÉES

#### 6 a Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c.-à-d. le gestionnaire des données) ?

- Décrivez les rôles et les responsabilités concernant les activités de gestion des données par exemple : saisie de données, production de métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données. Nommez les personne(s) responsable(s) impliquées dans la mesure du possible.
- Pour les projets menés en collaboration, expliquez comment s'effectue la coordination des responsabilités de gestion des données entre partenaires.
- Indiquez qui est responsable de la mise en œuvre du DMP, et qui s'assure qu'il est examiné et, si nécessaire, révisé.
- Envisagez des mises à jour régulières du DMP.

#### 6 b Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données soient FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

- Expliquez comment les ressources nécessaires (par ex, le temps) à la préparation des données pour le partage/préservation (curation des données) ont été chiffrées. Examinez et justifiez soigneusement toutes les ressources nécessaires pour distribuer les données.
- Il peut s'agir de frais de stockage, de coût matériel, de temps de personnel, de coûts de préparation des données pour le dépôt, de frais d'entrepôt et d'archivage.

- Indiquez si des ressources supplémentaires sont nécessaires pour préparer les données en vue de leur dépôt ou pour payer tous les frais demandés par les entrepôts de données. Si oui, précisez-en le montant et comment ces coûts seront couverts.

## Conseils pour la sélection des entrepôts dignes de confiance

Le tableau suivant fournit des conseils pour la sélection des entrepôts dignes de confiance, structurés selon quatre thèmes principaux.

### 1 ATTRIBUTION D'IDENTIFIANTS UNIQUES ET PÉRENNES (PIDS)

Un dépôt digne de confiance devrait :

#### 1a Permettre le repérage et l'identification des données

- S'assurer que des PIDs sont effectivement présents dans les métadonnées associées.

#### 1b Permettre la recherche, la citation et la récupération de données

- affecter systématiquement des PID (par exemple un DOI<sup>vii</sup>, URN<sup>viii</sup>, ARK<sup>ix</sup>) aux données qu'il détient, ce qui permet aux données et métadonnées associées d'être localisées, référencées et récupérées, même en cas de changement de lieu de stockage des données.

#### 1c Permettant la gestion des versions des jeux de données

- s'assurer que la version des données stockées dans l'entrepôt est clairement spécifiée et documentée par le biais d'un système d'audit permanent pour permettre d'en retracer la provenance.

*Nota : Tous les entrepôts n'utilisent pas un système de PID accepté et universel, comme ceux mentionnés ci-dessus. En lieu et place, ils utilisent un identifiant local ou un numéro administratif que l'entrepôt lui-même administre. Cela augmente le risque que les données ne puissent pas être localisées si elles viennent à être déplacées, ou si l'entrepôt cesse d'exister, se réorganise ou modifie sa gouvernance.*

### 2 MÉTADONNÉES

Les données doivent être décrites avec précision à l'aide d'un jeu métadonnées riche. Les métadonnées doivent renseigner sur la façon dont les données ont été générées, sous quelle licence et comment elles peuvent être réutilisées et doivent fournir le contexte pour une interprétation appropriée par d'autres chercheurs.

Un entrepôt digne de confiance devrait :

#### 2a Permettre le repérage des données

- assurer l'interopérabilité et la réutilisation des données par d'autres en fournissant données et métadonnées dans un langage accessible, basé sur un formalisme bien établi. Données et métadonnées doivent être décrites en utilisant des vocabulaires standards et des formats permettant aux systèmes informatiques de les rechercher, les combiner de manière automatique et de distinguer métadonnées et fichiers de données.

## 2b Permettre de référencer des informations pertinentes connexes, telles que d'autres jeux de données et publications

- s'assurer qu'il est possible de déclarer dans les métadonnées des liens vers des informations pertinentes ou connexes en fournissant un PID et une description de la relation scientifique. Un type particulier d'information concernerait les chercheurs associés, pour lesquels des identifiants pérennes existent (tels que ORCID<sup>x</sup>, ISNI<sup>xi</sup>, ou DAI<sup>xii</sup>).

## 2c Fournir de l'information librement accessible et mise à jour, y compris sur des données non publiées, protégées, rétractées, ou supprimées

- s'assurer que les métadonnées sont archivées sur le long terme et qu'elles restent toujours consultables, même si les données de recherche correspondantes ne sont pas ou plus disponibles (par exemple en raison d'une restriction de confidentialité, d'obligations légales, ou d'autres mesures de protection).
- s'assurer que les données rétractées en raison de mauvaises pratiques de recherche ou de méconduites peuvent encore être localisées grâce aux métadonnées et conservées dans des bases de données afin d'en permettre l'examen.

## 2d Utiliser des normes de métadonnées qui sont largement acceptées (par la communauté scientifique)

- s'assurer que les métadonnées conservées par l'entrepôt sont lisibles par programme et respectent des normes largement acceptées par la communauté scientifique.
- s'assurer que les normes des communautés ou les bonnes pratiques en matière de traitement des données sont, si elles existent. Notez que les entrepôts spécialisés dans un domaine de recherche particulier peuvent avoir des normes de communautés concernant les données et les métadonnées qui y sont déposées.

## 2e S'assurer que les métadonnées sont récupérables par programme

- encourager à la structuration des métadonnées de manière à permettre leur récupération par programme, par exemple en fournissant un formulaire avec des champs spécifiques à remplir.

## 3 ACCÈS AUX DONNÉES ET LICENCES D'UTILISATION

Un entrepôt digne de confiance devrait :

### 3a Permettre l'accès aux données dans des conditions bien définies

- préciser clairement les conditions dans lesquelles les données peuvent être réutilisées. De telles conditions (licence) sont généralement incluses dans les métadonnées.

### 3b Garantir l'authenticité et l'intégrité des données

- s'assurer que les métadonnées contiennent de l'information détaillée sur la provenance des données, y compris la façon dont elles ont été générées, traitées, dans quel contexte elles peuvent être réutilisées, et quel est leur degré de fiabilité.

### 3c Permettre la récupération des données

- permettre la récupération de données ou au moins de métadonnées à l'aide d'un protocole standardisé et ouvert (pas un protocole de communication propriétaire).

### 3d Fournir de l'information sur les licences et les autorisations d'accès aux données (idéalement sous forme lisible par programme

- permettre de faire référence de façon structurée à la licence d'utilisation, de façon à ce que les conditions d'utilisation soient claires, de préférence pour l'homme comme pour la machine.  
Dans la mesure du possible, les licences communes ou largement acceptés (comme Creative Commons) devraient être utilisés et référencés par URL.

### 3e Garantir la confidentialité et respecter les droits des personnes concernées et des créateurs de données.

- fournir un moyen d'authentification et d'autorisation des personnes et des machines utilisatrices, permettant de définir des droits d'accès spécifiques à l'utilisateur (ou au groupe) pour tenir compte des données présentant des problèmes de confidentialité ou d'autres restrictions.

## 4 PRÉSERVATION

Un dépôt digne de confiance devrait :

### 4a Assurer la pérennité des métadonnées et des données

- assurer la préservation, la disponibilité et l'accès sur la durée aux données et métadonnées qui lui sont confiées par les utilisateurs.

### 4b Faire preuve de transparence au sujet des missions, périmètre, politique de préservation, et de l'organisation (y compris la gouvernance, la pérennité budgétaire, et le plan de maintenance)

- gérer la préservation des données et des métadonnées de façon documentée. En particulier, l'entrepôt devrait disposer d'une politique de préservation qui détaille les missions et périmètre de l'entrepôt, les aspects relatifs à la gouvernance, la viabilité budgétaire, les partenaires sous-traitants, et les durées de conservation (le calendrier de mise en œuvre de la préservation).
- disposer d'un plan d'urgence accessible publiquement et assurer la préservation des données et métadonnées au-delà de l'existence même de l'entrepôt (par exemple en permettant l'extraction et le transfert de données et de métadonnées vers un autre entrepôt).

## Notes et références

---

<sup>i</sup> Le terme "organismes de recherche" désigne les organismes qui effectuent des recherches, les universités et les instituts de recherche.

<sup>ii</sup> Rendre les données Faciles à trouver, Accessibles, Interopérables et Réutilisables. Veuillez consulter l'annexe pour plus d'informations ou visitez le site :

<https://www.force11.org/group/fairgroup/fairprinciples>

<sup>iii</sup> Pour plus d'informations sur la manière dont les principes FAIR sont traduits dans les exigences minimales et les critères de sélection, veuillez consulter l'annexe.

<sup>iv</sup> Le concept a été présenté lors d'une manifestation publique le 30 janvier 2018, et deux consultations ont été organisées en avril 2018 et en août/septembre 2018 : <https://scieur.org/rdm-initiative>

<sup>v</sup> Pour les éléments procéduraux de la mise en œuvre des DMP : RDA Common Standard Working Group <https://www.rd-alliance.org/groups/dmp-common-standards-wg>

<sup>vi</sup> Les exigences minimales pour les plans de gestion de données ont été élaborées dans le cadre d'une initiative pour une harmonisation volontaire de la gestion des données de recherche à l'échelle internationale sous l'égide de Science Europe et de NWO (the Netherlands Organisation for Scientific Research). Des informations détaillées sur l'initiative sont disponibles à l'adresse

<http://scieur.org/rdm-initiative>

<sup>vii</sup> Digital Object Identifier

<sup>viii</sup> Uniform Resource Name

<sup>ix</sup> Archival Resource Key

<sup>x</sup> Open Researcher and Contributor ID

<sup>xi</sup> International Standard Name Identifier

<sup>xii</sup> Digital Author Identifier

## Annexe : Compatibilité avec les principes FAIR

		Exigence	Critère de sélection
<b>Pour être Facile à trouver</b>			
F1	Les données et les métadonnées sont identifiées par un identifiant global unique et pérenne	E 5d	Critère 1
F2	Les métadonnées décrivant les données sont riches	E 2a	Critère 2
F3	Les données et les métadonnées sont enregistrées et indexées dans un dispositif permettant de les rechercher	E 5d	Critère 1, Critère 2
F4	Les métadonnées spécifient l'identifiant de la donnée		Critère 2

<b>Pour être Accessible</b>			
A1	Les données et les métadonnées sont accessibles par leur identifiant via un protocole de communication standardisé	E 5c	Critère 1, Critère 2
A1.1	Le protocole utilisé est ouvert, libre et peut être implémenté de manière universelle	E 5c	Critère 2
A1.2	Le protocole utilisé permet l'authentification et l'autorisation si besoin	E 4b, E 5a, E 5c	Critère 3
A2	Les métadonnées sont accessibles même quand les données ne le sont plus	E 4c, E 5a, E 5d	Critère 2c



Pour être Interopérable			
I1	Les données et les métadonnées utilisent un langage formel, accessible, partagé et largement applicable pour la représentation des connaissances	E 1b, E 2a	Critère 2d
I2	Les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR	E 2a, E 2b	Critère 2
I3	Les données et les métadonnées incluent des liens vers d'autres (méta)données	E 2a, E 5a, E 5c	Critère 2b

Pour être Réutilisable			
R1	Les données et les métadonnées ont des attributs multiples et pertinents	E 2a, E 2b	Critère 2
R1.1	Les données et les métadonnées sont mises à disposition selon une licence explicite et accessible	E 4b, E 5a	Critère 3d
R1.2	Les données et les métadonnées sont associées à leur provenance	E 1a, E 1b, E 2b	Critère 1c, Critère 2, Critère 3b, Critère 4a
R 1.3	Les données et les métadonnées correspondent aux standards des communautés indiquées	E 1b, E 2a	Critère 2d