



**RECOMMANDATIONS  
pour FAIRiser  
vos données**

**PARTHENOS**

Facile  
à trouver

Accessible

Interopérable

Réutilisable

## A PROPOS DE CE GUIDE

Ce guide présente une série de recommandations visant à harmoniser les efforts des producteurs de données, des archivistes et des utilisateurs de données en sciences humaines et sociales pour rendre les données de la recherche aussi réutilisables que possible.

Ces recommandations sont le fruit du travail de plus de cinquante membres du projet PARTHENOS chargés de trouver des points communs à la mise en œuvre des politiques et des stratégies de gestion des données de la recherche en Europe. Pour ce faire, ils ont utilisé les résultats de recherches documentaires, de questionnaires et d'entretiens avec une sélection d'experts afin de rassembler une centaine de politiques actuelles de gestion des données (incluant des guides sur les formats privilégiés, les politiques de révision des données et les bonnes pratiques, aussi bien officielles qu'officieuses).

En portant une attention particulière à la qualité des (méta)données et des entrepôts de données, l'équipe de PARTHENOS a recensé vingt recommandations communes à différentes disciplines.

Pour plus de lisibilité, l'équipe a réparti chacune des recommandations parmi les catégories suivantes: Facile à trouver, Accessible, Interopérable, Réutilisable. Cette répartition est basée sur les principes FAIR initialement publiés par FORCE11 (2016) et destinés à accompagner ceux qui souhaitent améliorer la réutilisabilité des données de la recherche. Chacune des recommandations de PARTHENOS est complétée par des recommandations spécifiques, d'une part pour les producteurs et utilisateurs de données et d'autre part pour les archivistes. Les icônes sous les recommandations indiquent à quel acteur elles s'adressent



*L'icône représentant une lampe indique des recommandations pour les producteurs et les utilisateurs de données comme les chercheurs et les communautés de recherche en histoire, archéologie, linguistique et sciences sociales.*



*L'icône de la roue indique les recommandations pour les infrastructures de recherche et les centres d'archivage de données des centres de recherche et des institutions du patrimoine culturel.*



**PARTHENOS est un consortium de seize institutions et infrastructures de recherche européennes. Les membres de PARTHENOS ont pour but d'accroître la réutilisabilité des données de la recherche en construisant des ponts entre le cycle de vie des données des communautés de recherche, les entrepôts de données, les infrastructures de recherche et les institutions du patrimoine culturel dans les domaines connexes des sciences humaines et sociales.**

# 20 recommandations pour FAIRiser la gestion des données et rendre les données réutilisables



1

## Investir dans les personnes et l'infrastructure

Un prérequis important pour réussir à mettre en œuvre les dix-neuf autres recommandations de ce guide est d'investir dans les infrastructures de données et de recruter et former des experts en données.



*Familiarisez-vous avec les bonnes pratiques de gestion des données de la recherche. Découvrez les modules de formation PARTHENOS sur la gestion des données ou consultez le guide CESSDA pour les spécialistes de la gestion des données.*



*Recrutez et formez des experts en données et prévoyez un budget destiné aux investissements dans l'infrastructure technique et le personnel.*



# Facile à trouver

Les données de la recherche doivent être faciles à trouver tant par les humains que par les systèmes informatiques et doivent être obligatoirement décrites à l'aide de métadonnées qui facilitent la découverte des jeux de données pertinents.

2

## Utiliser des identifiants pérennes

Localiser les données est une condition préalable à toute autre opération, de l'accès à la réutilisation. Pour être facile à trouver, tout jeu de données doit être identifiable de façon unique et constante dans le temps à l'aide d'un identifiant pérenne (PID). Un PID est toujours valide même si l'adresse Web de la ressource change. Les PID peuvent prendre différentes formes, comme un Handle, un DOI, un PURL, un URN ou un ARK.



*Mentionnez le PID attribué à votre jeu de données dans vos travaux de recherche.*



*Sélectionnez le format de schéma d'identification pérenne adéquat et affectez un PID à chaque ressource. Utilisez le guide PID du NCDD (The Netherlands Coalition for Digital Preservation) pour choisir le PID le plus adapté à votre infrastructure de recherche.*

3

## Citer les données de la recherche

Si les données de la recherche ont un identifiant pérenne et sont citées conformément aux normes des communautés, les données ou les jeux de données correspondants sont plus faciles à trouver.



*Familiarisez-vous avec les règles de citation des données spécifiques à votre domaine ou discipline et citez les données de la recherche selon celles-ci.*



*Fournissez aux communautés de recherche des informations sur les bonnes pratiques pour citer les données et facilitez la citation des données par les utilisateurs de données, avec par exemple un bouton standardisé indiquant "Comment citer ce jeu de données".*

## 4 Utiliser des identifiants auteur pérennes

Un identifiant auteur pérenne (par exemple VIAF, ISNI ou ORCID) aide à établir des liens entre les jeux de données, les activités de recherche, les publications et les chercheurs et favorise l'identification des auteurs et ressources ainsi que la communicabilité de celles-ci.



*Ne soyez pas confondu avec d'autres chercheurs ou groupes de recherche. Demandez un identifiant auteur si vous n'en avez pas déjà un et indiquez-le dans vos jeux de données.*



*Mentionnez les identifiants auteur dans les métadonnées.*

## 5 Choisir un schéma de métadonnées approprié

Les métadonnées sont essentielles pour rendre les données faciles à trouver, notamment les métadonnées utilisées pour citer et décrire les données. Un schéma de métadonnées est une liste d'éléments normalisés permettant de fournir des informations sur une ressource, comme par exemple un titre, un identifiant, un nom d'auteur ou une date. L'utilisation de schémas de métadonnées existants garantit le respect des normes internationales pour l'échange de données.



*Afin de permettre la découverte du contenu, décrivez les données de recherche de la façon la plus cohérente et la plus complète possible. Fournissez suffisamment d'informations pour que les données puissent être accessibles et compréhensibles plus tard. Si possible, utilisez un schéma de métadonnées existant qui correspond au type de donnée ou de jeu de données que vous décrivez.*



*Indiquez clairement le schéma de métadonnées utilisé et recommandez-le aux communautés de recherche concernées. Pour enrichir les jeux de données lors du dépôt des données, prévoyez un formulaire de soumission de données qui recueille des métadonnées supplémentaires, par exemple sur la provenance des données.*

# ACCESSIBLE

Les données de la recherche doivent être facilement accessibles et récupérables avec des conditions d'accès bien définies et des protocoles de communication normalisés.



## Choisir un entrepôt fiable

Un entrepôt certifié assure un stockage fiable pour les jeux de données. La certification garantit que les données sont stockées de manière sûre et qu'elles seront disponibles, faciles à trouver et accessibles sur le long terme. CoreTrustSeal, nestor Seal, AFNOR et la certification ISO 16363 sont des exemples de normes de certification.



*Rendez vos données accessibles via un entrepôt fiable. De plus, en suivant les recommandations des entrepôts (sur les formats de fichiers, les schémas de métadonnées, etc.), vous vous assurez que toutes les conditions sont remplies pour rendre vos données FAIR.*



*Indiquez clairement le niveau de certification sur votre site Web. Si vous n'êtes pas (encore) certifié, indiquez comment vous prévoyez d'assurer la disponibilité, la facilité de trouver, l'accessibilité et la réutilisabilité des données à long terme*

7

## Indiquer clairement les conditions d'accès

Les informations d'accès précisent comment un utilisateur peut accéder à un jeu de données. Lors du dépôt dans un entrepôt de données, il doit être clairement indiqué quelles options d'accès un déposant de données peut choisir.



*Lors du choix d'une option d'accès, tenez compte des obligations légales, des pratiques propres à la discipline et des principes éthiques, selon le cas. Dans la mesure du possible, optez pour le libre accès. Lors de la collecte de données personnelles, demandez-vous si elles contiennent des informations susceptibles de révéler l'identité des participants, à quoi les participants ont consenti et quelles mesures vous avez prises pour protéger vos données. Si vos données ne peuvent pas être publiées en accès libre, les métadonnées devraient l'être, pour permettre la découverte des données.*



*Encouragez la publication des (méta)données en accès libre. Indiquez clairement les restrictions d'accès aux (méta)données sensibles qui ne doivent pas faire partie des (méta)données accessibles au public. Dans ce cas, veillez à rendre les (méta)données disponibles au moyen d'une procédure d'accès contrôlée et documentée.*





8

## Appliquer un embargo sur les données au besoin

Pendant la période d'embargo sur les données, seule la description du jeu de données est publiée. Les données elles-mêmes ne sont pas accessibles. La totalité des (méta)données sera disponible après un certain temps.



*Indiquez clairement pourquoi et pour combien de temps un embargo sur les données est nécessaire. Rendez les (méta)données publiquement disponibles dès que possible.*



*Précisez si un embargo sur les données est autorisé et sous quelles conditions il s'applique.*

9

## Utiliser des protocoles d'échange normalisés

En utilisant des protocoles d'échange normalisés, les infrastructures de recherche peuvent rendre les (méta)données publiquement accessibles et moissonnables, par exemple par les moteurs de recherche, ce qui améliore considérablement leur accessibilité.



*Utilisez des protocoles normalisés tels que SWORD, OAI-PMH, ResourceSync et SPARQL. Convertissez les schémas de métadonnées en XML ou RDF. Constituez un répertoire des endpoints, des chemins d'accès aux données de recherche et publiez-les.*

Pour accélérer la découverte et ouvrir de nouvelles perspectives, les données de la recherche doivent pouvoir être facilement intégrées à d'autres jeux de données par les humains et les systèmes informatiques.

## Interopérable

10

### Mettre en place des API bien documentées et exploitables par une machine

Des API bien documentées et exploitables par une machine - un ensemble de définitions de sous-programmes, de protocoles et d'outils pour construire des logiciels d'application - permettent l'indexation, l'extraction et la compilation automatiques de (méta)données depuis différents entrepôts de données.



*Documentez correctement les API et fournissez le schéma du modèle de (méta)données. Pensez à donner des exemples sur la façon d'extraire avec succès des données à partir de différents endpoints et de les intégrer dans de nouveaux jeux de données utilisables pour de nouvelles recherches.*

11

### Utiliser des vocabulaires ouverts et précis

La description des éléments de métadonnées doit suivre les recommandations de la communauté en utilisant des vocabulaires ouverts, précis et reconnus. Ces vocabulaires décrivent la signification exacte des concepts et des caractéristiques que les données représentent.



*Utilisez des vocabulaires pertinents dans votre domaine, enrichissez et structurez vos travaux de recherche en conséquence dès le début de votre projet de recherche.*



*Donnez des exemples de vocabulaires que le monde de la recherche peut utiliser, selon les particularités du domaine de recherche.*

12

### Documenter les modèles de métadonnées

Documenter avec précision les modèles de métadonnées aide les développeurs à comparer et à établir des correspondances entre les métadonnées.



*Publiez les modèles de métadonnées utilisés dans votre infrastructure de recherche. Documentez les spécifications techniques et définissez les classes (groupes d'éléments ayant des propriétés communes) et les propriétés (éléments qui expliquent les attributs d'une métadonnée ainsi que les relations entre les différentes métadonnées). Pour la modélisation des métadonnées, précisez les propriétés obligatoires et celles seulement recommandées.*



13

## Préconiser et utiliser des formats de données interopérables

L'utilisation d'un format de données soutenu par une communauté solide accroît la possibilité de partager, réutiliser et réunir les données collectées.



*Vérifiez quels formats de données sont utilisés par l'entrepôt où vous voulez déposer vos données. Structurez votre ensemble de données dans ce(s) format(s) dès le début de votre projet de recherche.*



*Précisez clairement quels standards votre institution utilise, réunissez une communauté autour d'eux et maintenez-les à jour, notamment dans une perspective d'interopérabilité. Le CMDI (pour les études linguistiques), le standard SIKBO102 (pour l'archéologie aux Pays-Bas), la TEI (pour l'encodage des sources textuelles) et le CIDOC CRM (pour les données patrimoniales internationales) en sont de bons exemples.*

14

## Élaborer des procédures pour améliorer la qualité des données

Pour améliorer la qualité des (méta)données et, par conséquent, l'interopérabilité, il faut élaborer des procédures (automatiques) pour nettoyer, extraire et enrichir ces (méta)données.



*Instaurez des procédures pour minimiser le risque d'erreurs dans la collecte des données. Par exemple, sélectionnez une date dans un calendrier au lieu de la saisir manuellement.*



*Investissez dans des outils pour aider à nettoyer les (méta)données et à les convertir dans des formats normalisés et interopérables. Combinez les efforts pour développer des méthodes de travail et des solutions logicielles pour ces traitements automatiques, par exemple en utilisant des outils d'apprentissage automatique.*

15

## Préconiser et utiliser des formats de fichiers pérennes

Toutes les données hébergées dans un entrepôt de données doivent être dans un format ouvert, international et normalisé afin de garantir une interopérabilité à long terme sur le plan de l'ergonomie, de l'accessibilité et de la pérennité.



*Dès le début de votre projet de recherche, pensez à des formats de fichiers pérennes. Privilégiez des formats recommandés par l'entrepôt de données et indépendants de logiciels, de développeurs ou de fournisseurs spécifiques.*



*Favorisez l'utilisation de formats considérés comme appropriés pour la préservation à long terme, tels que les fichiers PDF-A, CSV et MID/MIF. Donnez un aperçu détaillé et facilement accessible des formats de fichiers acceptés.*

Les données de la recherche doivent être disponibles pour de futurs travaux de recherche et traitements, de façon à ce que les résultats puissent être reproduits et que de nouvelles recherches s'appuient sur les résultats déjà obtenus.

## Réutilisable

16

### Documenter systématiquement les données

Afin de déterminer clairement ce qui peut ou non être contenu dans un jeu de données ou un entrepôt, les données doivent être systématiquement documentées. Faire preuve de transparence sur le contenu des données favorise la confiance et, par conséquent, la réutilisation des données.



*Fournissez des guides d'utilisation, incluant une description de la méthodologie, une liste des abréviations, une description des lacunes dans les données, la configuration de la base de données, etc.*

17

### Respecter les conventions de nommage

Suivre une convention de nommage précise et cohérente - un modèle communément accepté pour nommer les fichiers - permet aux futures générations de chercheurs de retrouver, d'accéder et de comprendre plus facilement les données et les jeux de données.



*Consultez les normes et les bonnes pratiques de votre discipline ou domaine de recherche pour trouver la convention de nommage la plus appropriée.*



*Indiquez clairement les bonnes pratiques pour créer et utiliser les conventions de nommage spécifiques aux fichiers.*

18

## Utiliser des formats de fichier courants

Utiliser des formats de fichiers normalisés qui sont largement utilisés dans votre communauté permet d'augmenter leur réutilisabilité



*Utilisez des formats de fichiers répandus en plus des formats d'archivage pour partager vos données, par exemple Excel (XLSX) et CSV ou un fichier de formes ESRI (SHP) en plus des fichiers MID/MIF.*



*Publiez les données dans des formats répandus en plus du format d'archivage s'ils sont différents.*

19

## Garantir l'intégrité des données

Les données de la recherche collectées doivent être identiques aux données de la recherche qui seront accessibles ultérieurement. Pour garantir leur authenticité, il faut contrôler l'intégrité des données.



*Implémentez une méthode de contrôle de version. Garantir que chaque changement effectué sur une version d'un jeu de données est correctement documenté est de la plus grande importance pour l'authenticité de chaque jeu de données.*



*Pour déterminer si un fichier a été modifié, il est essentiel de noter la provenance - l'origine des données et tout changement apporté au fil du temps - et de comparer les copies à l'original. Un contrôle d'intégrité des données peut être effectué grâce à une empreinte comme une somme de contrôle ou par une comparaison directe de deux fichiers. Proposez un mécanisme pour gérer différentes versions, par exemple en ajoutant la version à l'identifiant comme paramètre de recherche.*

20

## Autoriser la réutilisation

Pour permettre une réutilisation la plus large possible des (méta)données, il convient de préciser qui détient les droits sur les (méta)données et quelle licence s'applique.

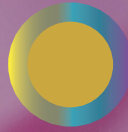
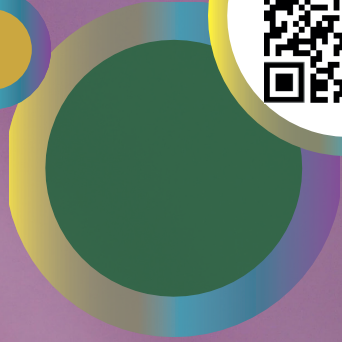
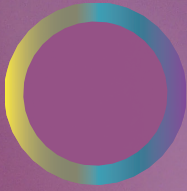


*Assurez-vous de savoir qui détient les droits sur les (méta)données avant de publier vos données de recherche.*



*Diffusez la licence sur les (méta)données et les options de réutilisation de manière transparente et dans un format lisible par une machine. Pour favoriser l'interopérabilité, essayez de rattacher vos licences à des modèles déjà largement adoptés, tels que Creative Commons ou ODBL.*





PARTHENOS est un projet Horizon 2020 financé par la Commission européenne. Les opinions exprimées dans ce document relèvent de la seule responsabilité de l'auteur et ne reflètent pas nécessairement celles de la Commission européenne.

Ce guide (version octobre 2019) est sous licence Creative Commons CC BY 4.0. Design: Verbeeldingskr8. DOI: 10.5281/zenodo.3463521.