



**HAL**  
open science

## Créer son plan de gestion des données

Marie Puren

► **To cite this version:**

Marie Puren. Créer son plan de gestion des données. École thématique. Lille, France. 2021. hal-03183724

**HAL Id: hal-03183724**

**<https://hal.science/hal-03183724v1>**

Submitted on 28 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Créer son plan de gestion des données

Marie Puren

MESHS Lille Nord de France - 29 mars 2021



# Les données de la recherche

Quelle définition ?

# Les données de la recherche OCDE

« Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. »

*Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics (2006)\**

\*A noter : c'est à cette définition que renvoie l'ANR

# Les données de la recherche OCDE

« Les données de la recherche désignent les informations, et en particulier les faits ou chiffres collectés pour être analysés et traités pour alimenter des réflexions, discussions ou calculs. Dans un contexte de recherche, ces données sont, par exemple, des statistiques, des résultats d'expériences, des mesures, des observations sur le terrain, des résultats d'enquêtes, des enregistrements d'entretiens ou des images. Il s'agit plus spécifiquement de données disponibles sous forme numérique. »

*Lignes directrices sur les règles pour le libre accès aux publications scientifiques et pour le libre accès aux données de la recherche dans Horizon 2020 (2016)*

# Les données de la recherche

## OCDE

« Aux fins de la présente directive, on entend par « données de la recherche », des documents se présentant sous forme numérique, autres que des publications scientifiques, qui sont recueillis ou produits au cours d'activités de recherche scientifique et utilisés comme éléments probants dans le processus de recherche, ou dont la communauté scientifique admet communément qu'ils sont nécessaires pour valider des conclusions et résultats de la recherche. »

Directive UE 2019/2014 concernant les données ouvertes  
(2019), art. 289

# Les données de la recherche

- En fonction des disciplines ou des thèmes de recherche
  - Linguiste : des écrits, des discours, des enregistrements
  - Géologue : coupes et observations de terrain consignées sur un carnet, des résultats de carottage, des analyses d'échantillons, des données sismographiques
- Au sein d'une même discipline:
  - Médiéviste : sources archivistiques, épigraphiques, iconographiques, littéraires, voire même archéologiques
  - Contemporanéiste : transcription de documents d'archives, ou des entretiens enregistrés et/ou filmés.
- Pour résumer, données de la recherche :
  - Pas seulement des données quantitatives
  - Pas seulement des séries homogènes, faciles à manipuler, échanger, compiler

# Les données de la recherche

- Frontière entre données et publications parfois difficile à distinguer
- Blogs, tweets... Données de recherche?
  - Pas des matériaux bruts mais dépourvus de tout appareil critique et non validés par les pairs (contrairement aux publications)
- Données publiées dans des articles, monographies sous forme de résultats
- Place des carnets de laboratoire, des carnets de recherche, des projets, des rapports

=> Question pas encore tranchée!



# Les données de la recherche

« Ces données sont des faits, des observations, ou des expériences sur lesquels on base un argument, une théorie ou un texte. Les données peuvent être numériques, descriptives ou visuelles. Les données peuvent être brutes ou analysées, expérimentales ou observationnelles.

Les données incluent : carnets de laboratoire (ou cahiers de laboratoire) ; carnets de fouilles ; données de la recherche primaires (incluant les données de la recherche sur copie papier ou dans une forme lisible par l'ordinateur) ; des questionnaires ; cassettes audio ; cassettes vidéo ; modèles ; photographies ; réponses à des tests. »

“What is research data ?”, Australian National Data Service

# Les données de la recherche

Selon Thierry Fournier (pilote de la commission recherche de l'Association des directeurs et des personnels de direction des bibliothèques universitaires (ADBU), ces données ne sont pas « nécessaires à la validation des résultats de la recherche », et se sont pas des données de la recherche « Les données de la recherche, définition et enjeux », Arabesques, 2015

## Les données de la recherche

« Les données de la recherche sont créées comme des résultats directs de “faire de la recherche”. En ce sens, cela exclut les supports destinés à l'enseignement et les documents administratifs (comme les fiches de poste, les mails ou les rapports financiers). Les données de la recherche peuvent être disponibles dans une infinie variété de formats. »

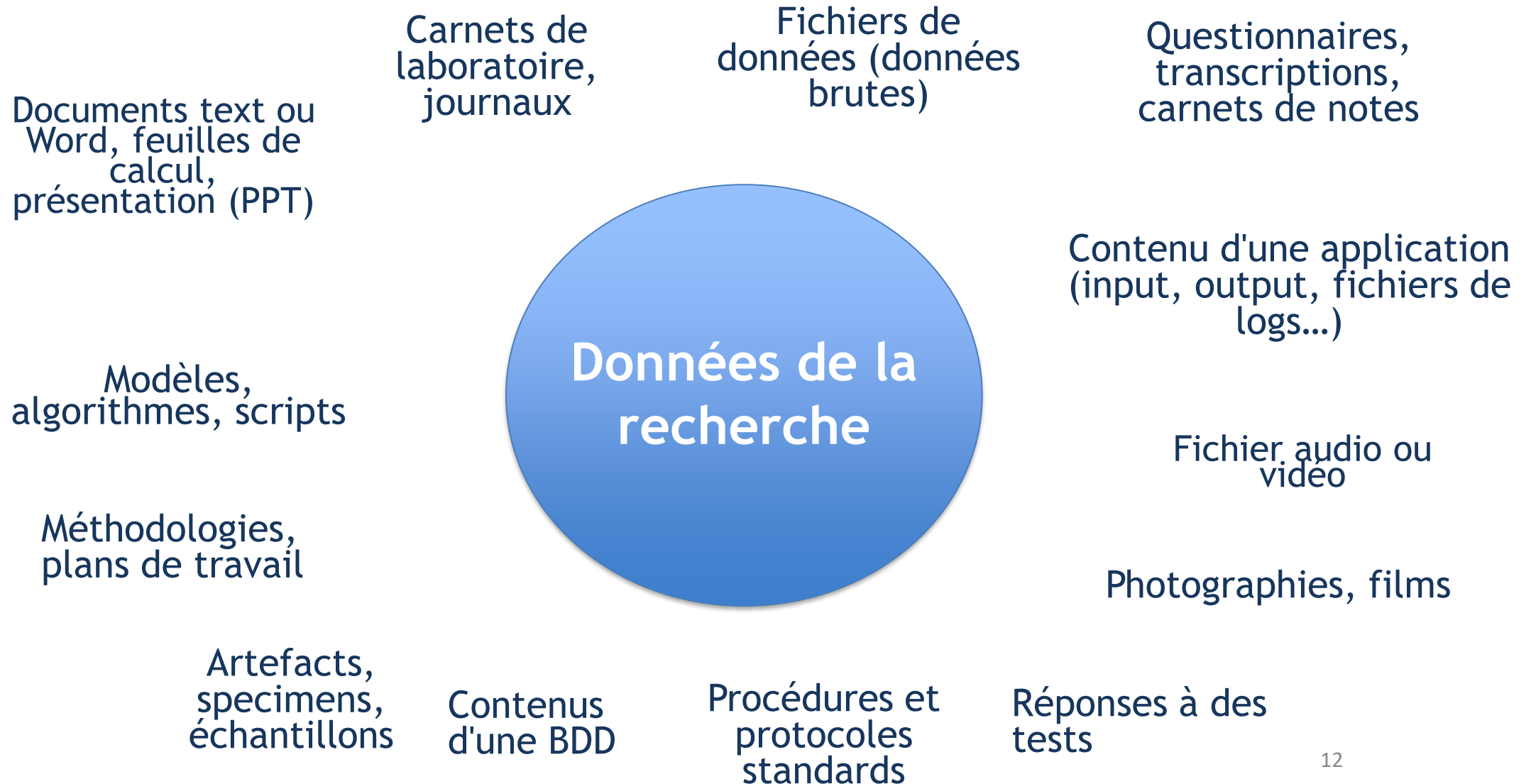
University of Bristol, [“A brief guide to research data”](#)

# Les données de la recherche

Les données de la recherche peuvent être :

- Observationnelles
- Expérimentales
- Générées au moyen de simulation
- Dérivées ou compilées (comme avec la fouille de textes et de données ou *text and data mining*)

# Les données de la recherche



# Les données de la recherche

- Jeu de données ou dataset : “agrégation, sous une forme lisible, de données brutes ou dérivées présentant une certaine 'unité', rassemblées pour former un ensemble cohérent”. Rémi Gaillard, [De l'Open data à l'Open research data : quelle\(s\) politique\(s\) pour les données de la recherche?](#)
  - Série de données compilées dans plusieurs documents différents, ou série de données compilées dans un document unique - par exemple, dans un fichier CSV
  - en fonction de la discipline, mais aussi des projets et des objectifs de la recherche, “l'échelle à laquelle ces données assemblées acquièrent leur unité pour former un 'jeu'” va beaucoup varier. (*Ibid.*)

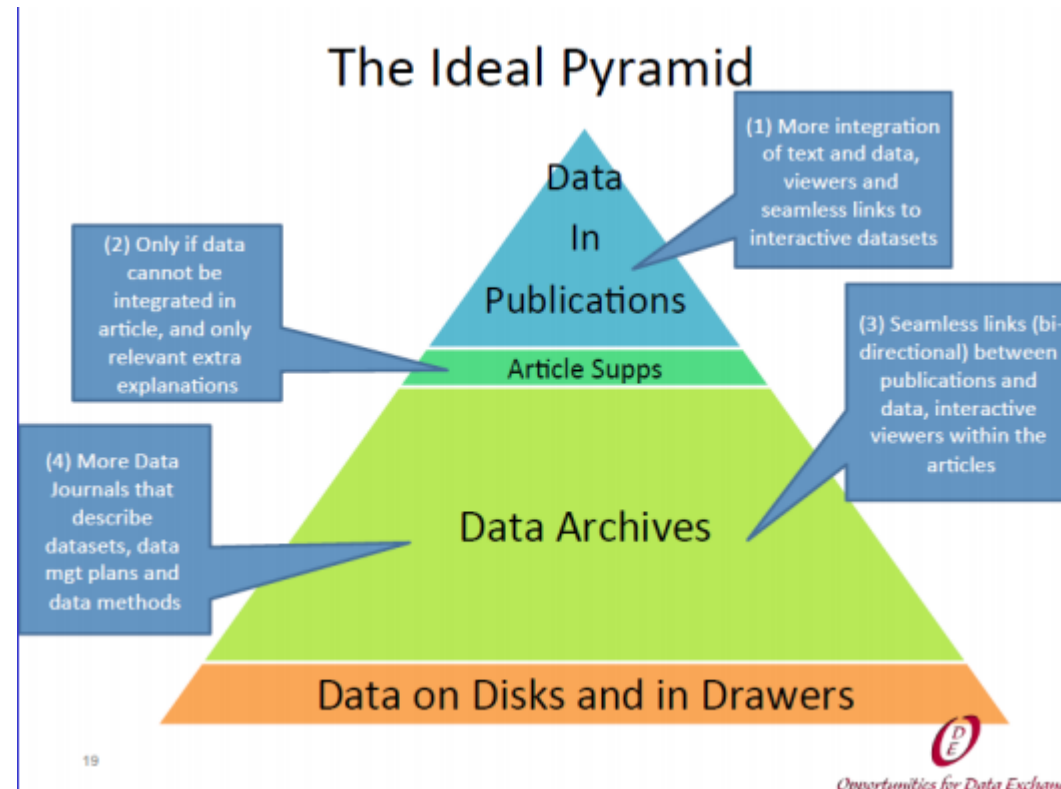
# L'ouverture des données

Un cadre nouveau

# Une nouvelle demande

La « pyramide idéale » :

- très peu de données inaccessibles
- la grande majorité d'entre elles dans des archives et des articles

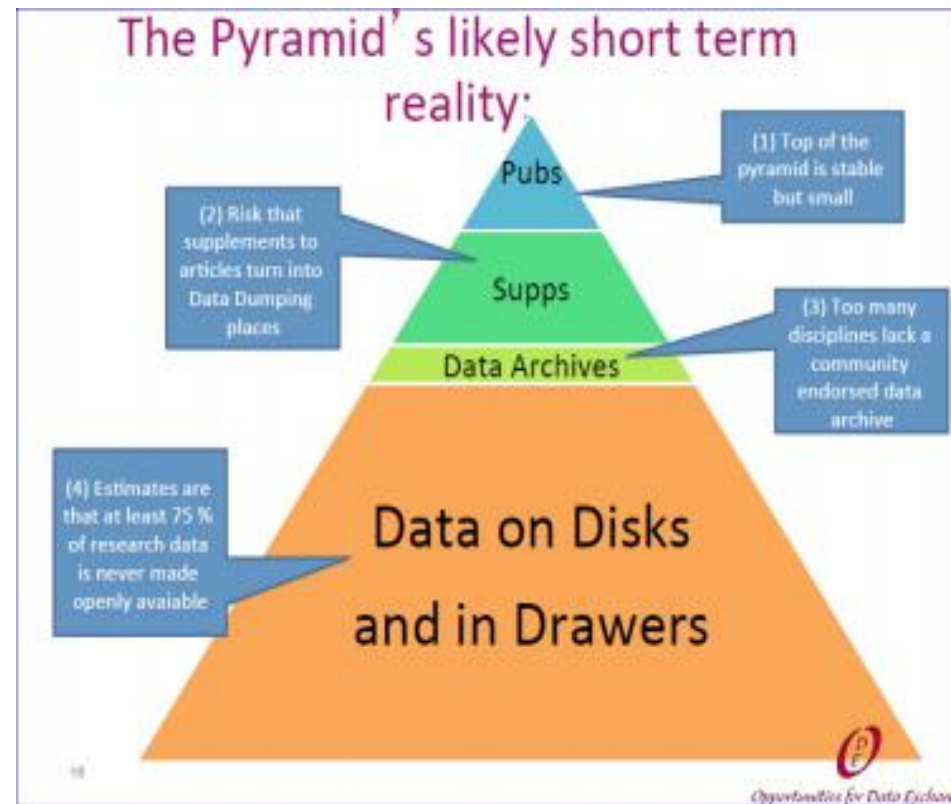




# Une nouvelle demande

La « pyramide réelle » :

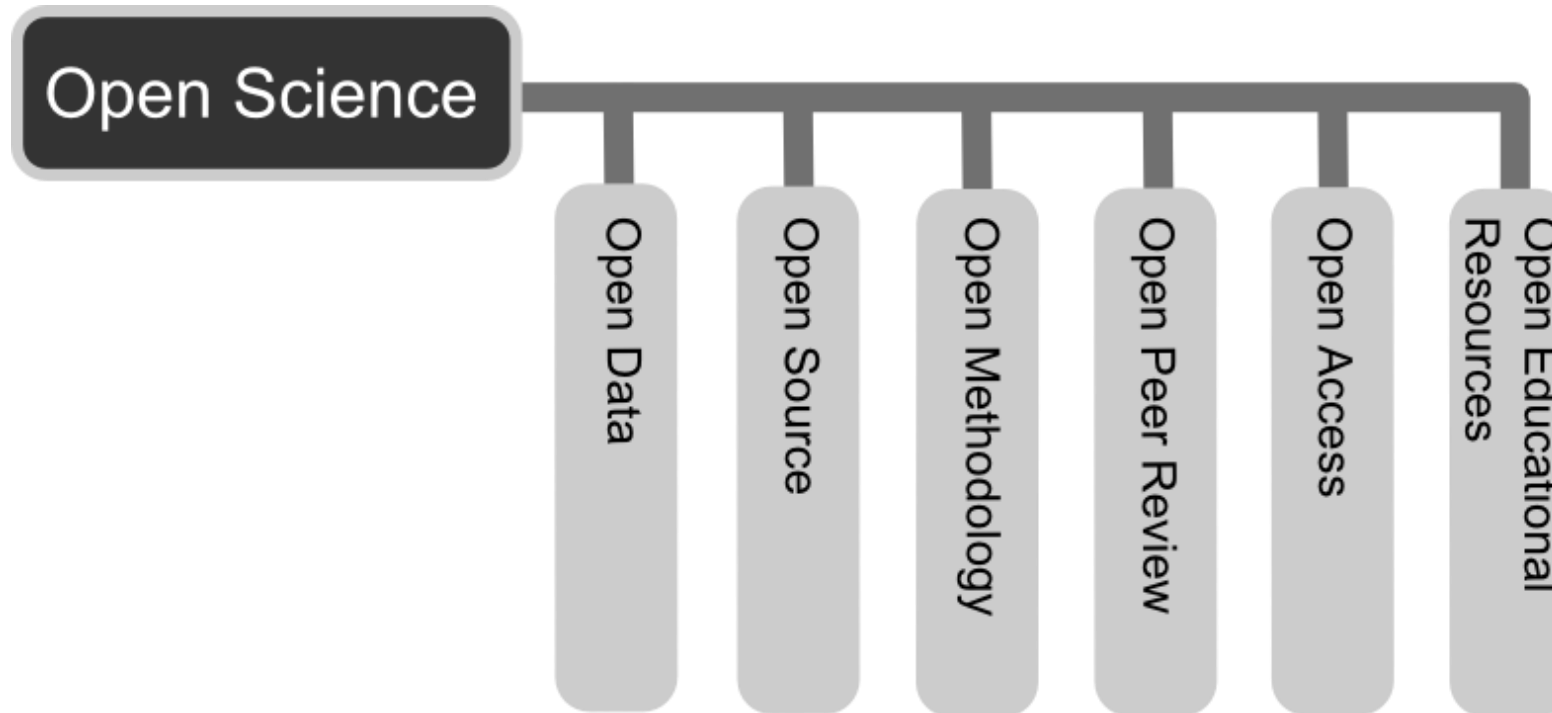
- Peu de données accessibles
- la grande majorité d'entre elles dans des « tiroirs » ou des disques durs



# Un nouveau modèle d'ouverture pour les données de la recherche

- La Science Ouverte : « mouvement destiné à promouvoir plus de partage et de transparence » dans la science
  - Ouverture des résultats de la sciences
  - Nouvelle façon de faire de la recherche : ouverture des protocoles de recherche, des méthodes, des codes
- Promeut un nouveau modèle d'ouverture pour les données de la recherche.
- Principaux aspects:
  - La disponibilité et l'accès
  - La réutilisation et la redistribution
  - La participation de tous

# Un nouveau modèle d'ouverture pour les données de la recherche



Par [Andreas E. Neuhold](#), travail personnel- basé sur “[The taxonomy tree](#)”, FOSTER (Facilitate Open Science Training for European Research)

Pour plus d'infos sur la Science ouverte: Michael Nielsen, [Reinventing Discovery: The New Era of Networked Science](#), Princeton University Press, 2011.

# Un nouveau modèle d'ouverture pour les données de la recherche

Ouvert = « Quiconque peut librement avoir accès, utiliser, modifier et partager pour quelque motif que ce soit. »

Open Knowledge International, "[The Open definition](#)"

# Un nouveau modèle d'ouverture pour les données de la recherche

- **2002** : [Budapest Open Access Initiative](#) : définition du mouvement en faveur du libre accès
- **2003** : [Déclaration de Berlin](#) en faveur de l'Open Access : les données de la recherche sont assimilées aux publications.
- **2006** : Déclaration sur l'accès aux données de la recherche financée par des fonds publics par l'OCDE. L'OCDE insiste sur la nécessité d'ouvrir les données produites dans le cadre de recherches financées sur fonds publics.

# Un nouveau modèle d'ouverture pour les données de la recherche

- **2007** : Le Conseil scientifique du Conseil européen de la recherche (CER) publie des recommandations demandant la mise en accès libre des résultats de recherches financées par le CER, dans un délai de six mois suivant leur publication, dans des archives ouvertes disciplinaires appropriées (research repositories) ou des archives ouvertes institutionnelles (institutional repository).
- **Fin 2009** : Création de l'infrastructure OpenAIRE (Open Access Infrastructure for Research in Europe) pour les publications scientifiques en libre-accès

# Un nouveau modèle d'ouverture pour les données de la recherche

- **Fin 2011**, OpenAIREplus lui succède et étend son périmètre aux données de la recherche.
- **2012** : Recommandation de la Commission européenne aux États-membres “relative à l'accès aux informations scientifiques et à leur conservation”. Elle pose comme principal objectif qu'avant 2014, “des politiques de libre-accès aux articles et données scientifiques [soient] établies dans tous les États membres à tous les niveaux pertinents”. Cette ambition rejoint l'idée de la création d'un espace européen de la recherche, qui se veut une “innovation ouverte”.

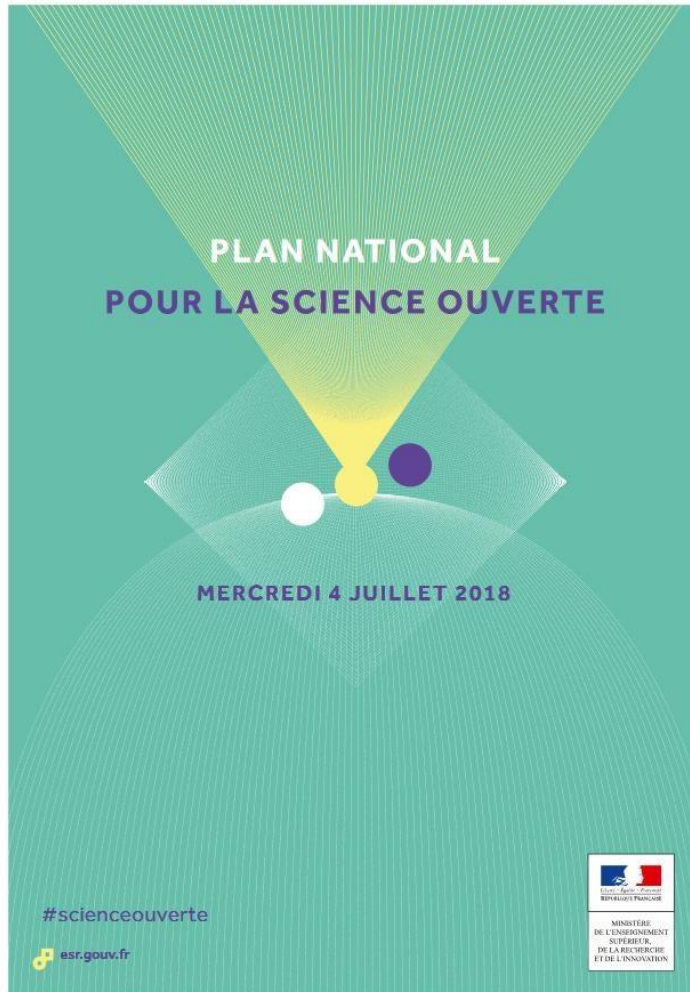
# Un nouveau modèle d'ouverture pour les données de la recherche

- **2013** : Open Research Data Pilot : mise en place d'une politique européenne en matière de libre-accès aux données
- **2016**: Extension de l'Open Research Data Pilot à tous les projets Horizon 2020 ( => les projets du Work Programme 2017 font par défaut partie de l'ORD)
- **2018** : Plan national pour la Science Ouverte (France)
- **2019** : Plan de gestion des données obligatoires pour les projets ANR



# La Science Ouverte

## En France

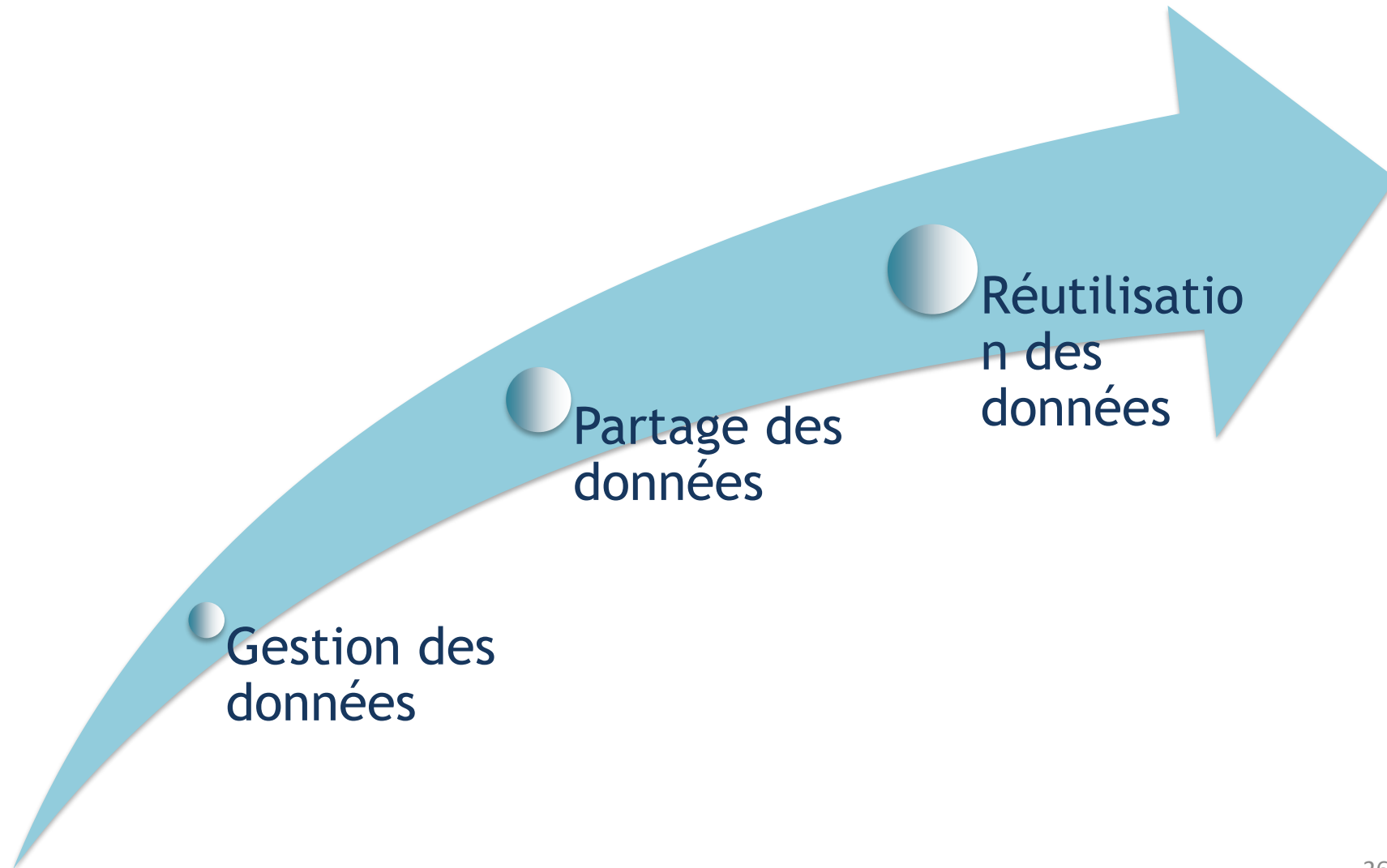


- Evolutions juridiques :
  - Loi Valter relative à la gratuité et aux modalités de la réutilisation des informations du service public (28 décembre 2015) = gratuité des informations publiques
  - Article 30 de la loi pour une République numérique (7 octobre 2016) = garantie diffusion en OA des publications scientifiques
- => Principe d'ouverture des données publiques par défaut
- 3 axes principaux Plan national pour la Science Ouverte en juillet 2018 :
  - Généraliser l'accès ouvert aux publication
  - Ouvrir les données de la recherche
  - Participer au mouvement de la Science Ouverte (niveau européen et international)
- Engagements de l'ANR pour la Science Ouverte depuis Appel à projets 2019
  - Publications en libre accès
  - Ouverture des données de la recherche => PGD

# La gestion des données de la recherche

Gérer pour partager

# Pourquoi gérer ses données?



# Gestion des données de la recherche

## Définition

“La gestion des données fait partie intégrante du processus de recherche.”

Université de Leicester, [“When do you need to think about RDM”](#)

Pendant un projet de recherche et après l'achèvement du projet:

- Rassembler,
- Organiser,
- Gérer,
- Stocker,
- Sauvegarder,
- Préserver,
- Partager vos données.

# Gestion des données de la recherche

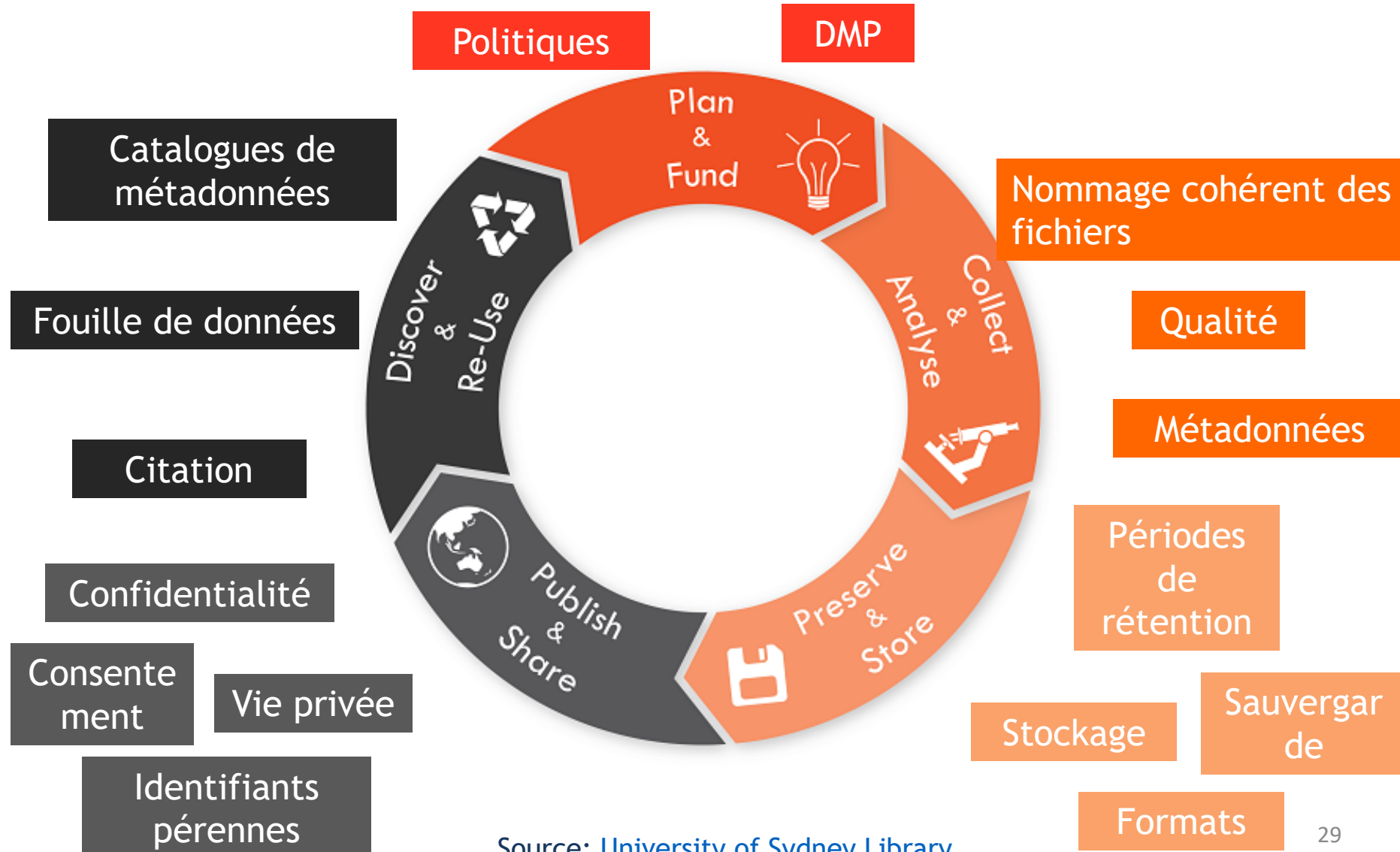
## Définition

La gestion des données implique généralement:

- De créer un **Plan de Gestion des Données (DMP)**, soumis avec la demande de financement, afin d'expliquer comment les données seront gérées pendant et après l'achèvement du projet ;
- De stocker des données de la recherche durant le projet et de les partager avec les collègues autorisé(e)s ;
- A la fin du projet, de déposer ces données dans un entrepôt pérenne, et de les rendre disponibles pendant plusieurs années.

# Gestion des données de la recherche

## Le cycle de vie des données de la recherche



# Le plan de gestion des données (PGD) / Data Management Plan (DMP)

Qu'est-ce que c'est ? Pourquoi le créer ?

# Plan de gestion des données

## Définition

« Les plans de gestion des données sont des éléments-clé d'une bonne gestion des données. Un plan de gestion des données décrit le cycle de vie de la gestion des données qui seront recueillies, traitées et/ou générées [...]. »

*H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0, 26 juillet 2016,  
p.4.



# Plan de gestion des données

## Définition

Un document formel qui décrit:

- Les données que vous espérez acquérir ou générer au cours du projet de recherche,
- Comment vous allez gérer, décrire, analyser et stocker ces données,
- Quels mécanismes vous utiliserez à la fin de votre projet pour partager et préserver ces données.

# Plan de gestion des données

## Définition

- Une feuille de route régulièrement mise à jour ;
- Un document normalisé ;
  - Son contenu varie en fonction des exigences du projet et des demandes des agences de financement ;
  - Il se concentre sur les données et jeux de données recueilli(e)s, créé(e)s, analysé(e)s par le projet de recherche.

# Plan de gestion des données

## Définition

**DMP = livrable du projet**

- Il matérialise la politique du projet en matière de données ;
- Il résume les buts et les actions qui seront mis en place ;
- Il répond aux exigences des financeurs.

# Plan de gestion des données

## Pourquoi?

### L'étape principale de la gestion des données de la recherche

« Gestion » active des données numériques (=> le « M » de DMP):

- Une gestion continue (sauvegarde, migration, conversion...), tout au long du cycle de vie des données ;
- Un plan d'action en termes de qualité des données, de faisabilité technique et de viabilité financière.

# Plan de gestion des données

## Pourquoi?

### L'étape principale de la gestion des données de la recherche

#### Gestion active et curation numérique

« Les activités de curation de données permettent de faciliter la découverte et la récupération de données, de maintenir la qualité des données, de leur ajouter de la valeur et d'en fournir pour de futures réutilisations. Ce nouveau champ inclut la représentation, l'archivage, l'authentification, la gestion, la préservation, la récupération, et l'utilisation. »

*Digital Humanities Data Curation*

# Plan de gestion des données

## Pourquoi?

L'étape principale de la gestion des données de la recherche

Curation numérique ≠ Stockage de données

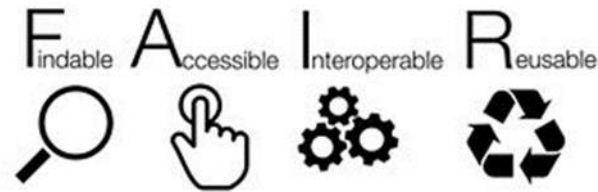
« Le stockage de données se résume à garder les données existantes et d'assurer qu'on pourra y accéder en fonction des besoins. Il n'implique pas nécessairement **des pratiques de rafraichissement ou de migration de format** (essentiel pour maintenir les données dans une forme utilisable), ni des pratiques de curation **comme l'amélioration des données par l'addition de métadonnées, ou la migration d'un standard vers un autre**. La curation de données va donc beaucoup plus loin que le stockage des données. »

*Digital Humanities Data Curation*

# Plan de gestion des données

## Pourquoi?

### Les « FAIR Data »



Un DMP « aide les bénéficiaires du programme Horizon 2020 à rendre leurs données de la recherche 'findable' (trouvables), 'accessible' (accessibles), 'interoperable' (intéropérables) et 'reusable' (réutilisables) (FAIR) [...]. »

[H2020 Programme Guidelines on FAIR Data Management in Horizon 2020](#), Version 3.0, 26 July 2016, p.3.

# Plan de gestion des données

## Pourquoi?

### FAIR Data

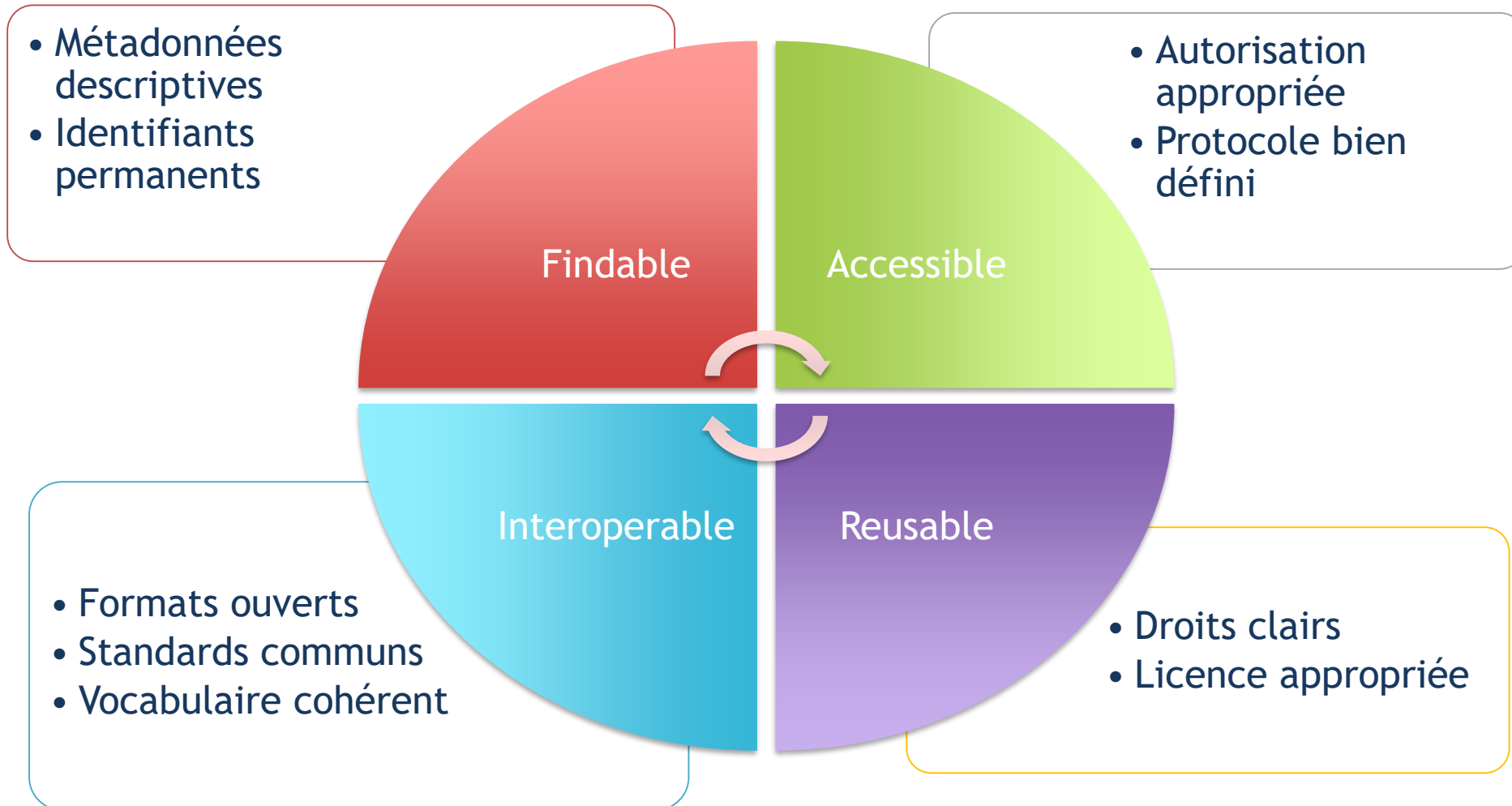
- Janvier 2014: une rencontre est organisée par le Netherlands eScience Center et le Dutch Techcentre for the Life Sciences (DTL) au Lorentz Center à Leyde.
- **Les principes FAIR:** “les producteurs et les consommateurs de données - à la fois machines et humains - peuvent plus facilement découvrir, avoir accès, interopérer et réutiliser intelligemment, avec des citations adéquates, les vastes quantités d'information qui sont générées par la science contemporaine et son usage intensif des données. » ([Force 11](#))



# Plan de gestion des données

Pourquoi?

FAIR Data



# Plan de gestion des données

## Pourquoi?

### Pour satisfaire les exigences de votre financeur

	Horizon Europe	ANR
Obligatoire	Oui	Oui
Courte description de la politique de gestion des données dès le épôt de la proposition	Obligatoire	Pas obligatoire mais vivement conseillé
Organisation des lots de travail (work packages ou WP) et des tâches	<ul style="list-style-type: none"> <li>• Dans le WP 1 « Gestion du projet » ou dans le WP consacré à la diffusion des résultats (s'il existe)</li> <li>• Une ou deux tâches dédiées</li> </ul>	<ul style="list-style-type: none"> <li>• Dans le WP 1 « Gestion du projet »</li> <li>• Une ou deux tâches dédiées</li> </ul>
Première version	Dans les 6 premiers du projet	Dans les 6 premiers du projet
Version à mi-projet	Pas obligatoire mais vivement conseillée	Obligatoire pour les projets > 30 mois Autrement : pas obligatoire mais vivement conseillée
DMP final à la fin du projet	Obligatoire	Obligatoire

# Plan de gestion des données

## Pourquoi?

Parce que c'est une bonne pratique de recherche qui permet de

- Formaliser la gestion de son travail
- Gérer son investissement/budgétiser
- Donner de la visibilité aux données en montrant qu'elles peuvent être réutilisées
- Ecrire des [data papers](#)
- Gagner du temps par la suite car un DMP peut améliorer la gestion des données

# Plan de gestion des données

## En résumé

- Quand on crée un DMP, on définit comment les données du projet seront:

Décrites

Partagées

Protégées

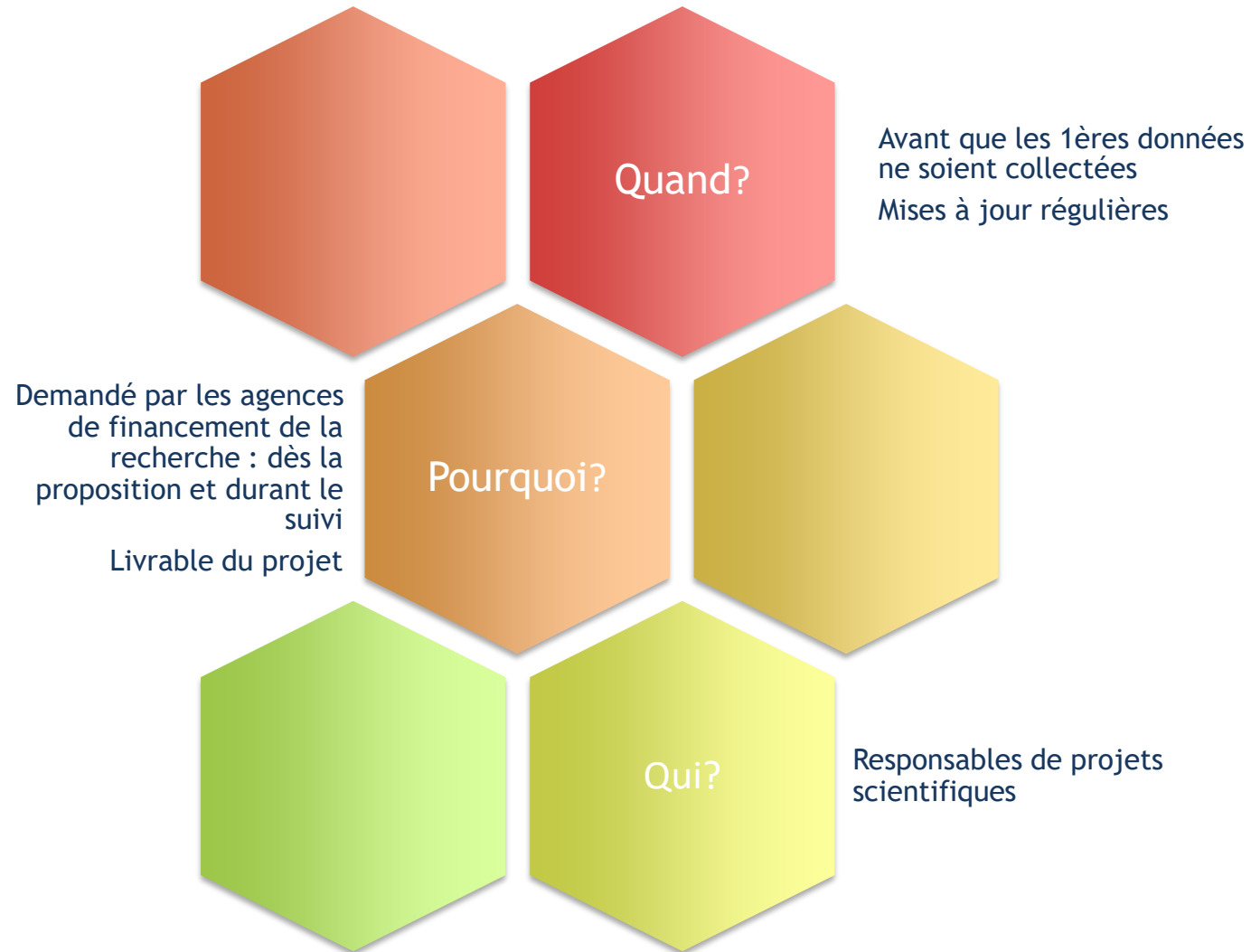
Conservées

# Comment le créer ?

Les points-clé

# Créer un plan de gestion des données

## Les premières questions



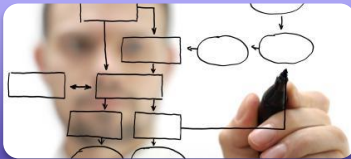
# Créer un plan de gestion des données

## Définir les responsabilités



### Chercheur(se)

- Responsable des données, coordinateur(trice) du DMP
- Description des données et découpage en différents jeux de données



### Ingénieur(e)-projet

- Coordonne les actions autour du DMP



### Informaticien(ne)

- Interlocuteur pour le stockage et la sécurisation des données, les aspects infrastructures et les coûts



### Spécialiste de l'IST

- Propose des standards, des alignements avec des référentiels existants



### Documentaliste et/ou archiviste

- Aide le(la) chercheur(se) à sélectionner les données, définir les durées de conservation et les solutions techniques

# Créer un plan de gestion des données

## Comparaison Horizon Europe et ANR

Horizon Europe	ANR
1. Résumé descriptif des données	1. Description des données et collecte ou réutilisation de données existantes
2. Données FAIR (4 sous parties montrant que les données sont 1. trouvables, 2. accessibles, 3. interopérables, 4. réutilisables°)	2. Documentation et qualité des données
3. Allocation de ressources	3. Stockage et sauvegarde pendant le processus de recherche
4. Sécurité des données	4. Exigences légales et éthiques, codes de conduite
5. Aspects éthiques	5. Partage des données et conservation à long terme
6. Autres	6. Responsabilités et ressources en matière de gestion des données



# Créer un plan de gestion des données

## Comment se structure un DMP?

### Les informations administratives du projet

Il faut d'abord indiquer l'identité du projet :

- Son identifiant
- Son nom
- Une description (court résumé)
- Les agences de financement
- Le(s) responsable(s) du projet (nom, coordonnées, identifiants - identifiant [idHAL](#) ou [ORCID](#) par exemple)
- Le(s) responsables de la gestion des données (nom, coordonnées)
- Dates :
  - De la 1<sup>ère</sup> version du DMP
  - De la dernière mise à jour

# Créer un plan de gestion des données

## Les grandes parties d'un DMP

Pour chaque jeu de données, vous allez donner les informations suivantes :

1. La description des jeux de données : types de données recueillies ou générées par le projet?
2. Les standards et métadonnées : quel(s) standard(s) avez-vous utilisé pour les métadonnées?
3. Le partage des données : comment ces données seront-elles exploitées, partagées, rendues accessibles?
4. L'archivage et la conservation des données: comment ces données seront-elles administrées et conservées?

# Créer un plan de gestion des données

## 1. La description des données

**Quels seront les types de données générées ou recueillies par le projet?**

- L'origine des données
- Leur nature et leur échelle
- Qui pourraient les utiliser ?
- Appui d'une publication scientifique ?
- Y a-t-il des données similaires ?

# Créer un plan de gestion des données

## 1. La description des données

### Origine des données

- Source:
  - Collecte de jeux de données existants?
  - Création de jeux de données nouveaux?
  - Les deux ?

# Créer un plan de gestion des données

## 1. La description des données

### Origine des données

- Préciser en détail la méthodologie adoptée pour recueillir ces jeux de données
  - Données existantes : accès à ces données (citation de la source, modalités d'accès, description synthétique)
  - Données nouvelles : processus de création ou le mode opératoire mis en place pour les obtenir

# Créer un plan de gestion des données

## 1. La description des données

### Description formelle et technique

- ***Type de données*** : source, forme, stabilité, volumétrie
- ***Format***, ouverture, encodage, compression
- ***Système d'organisation, de nommage et de gestion*** des répertoires et des fichiers

# Créer un plan de gestion des données

## 1. La description des données

### *Type de données*

- Source :
  - données d'observation,
  - expérimentales,
  - de simulation,
  - dérivées ou compilées (ex. : fouille de texte)
- Forme:
  - données textuelles (notes, carnets de laboratoire)
  - données numériques (tableaux)
  - données audiovisuelles
  - modèles/ codes informatiques
  - données spécifiques liées à certains instruments

# Créer un plan de gestion des données

## 1. La description des données

### *Type de données*

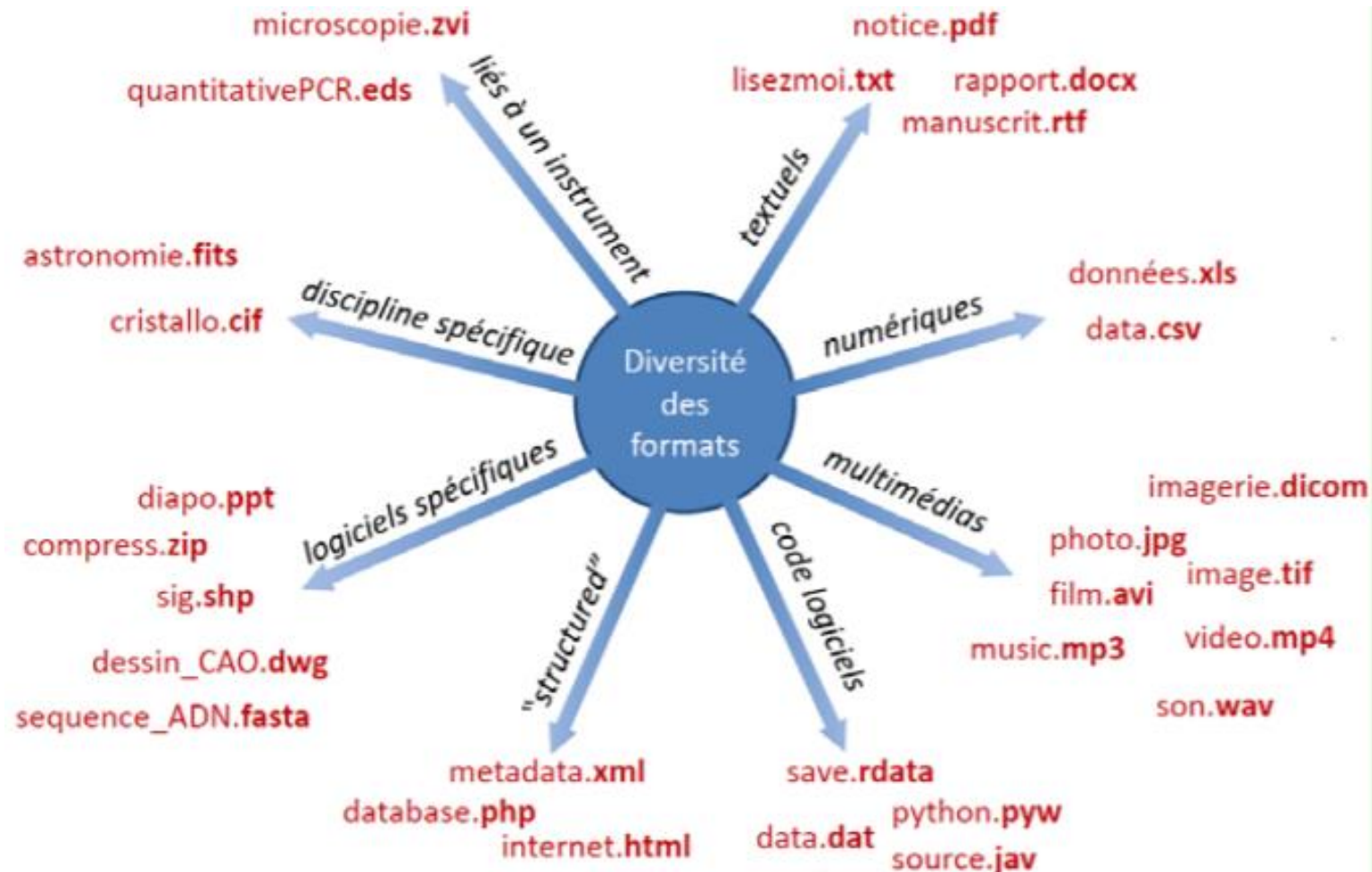
- Stabilité des données :
  - données fixes : elles ne changent pas après avoir été recueillies
  - croissantes : de nouvelles données peuvent être ajoutées, mais les anciennes ne seront pas modifiées ou supprimées
  - révisables : de nouvelles données peuvent être ajoutées, et les anciennes peuvent être modifiées ou supprimées
- Versions: organiser, identifier et sauvegarder les différentes versions (V1, V2, V3...etc.)
- Volume : préciser le volume envisagé pour chacun des jeux de données:
  - Capacités de stockage?
  - Budget nécessaire à prévoir



# Créer un plan de gestion des données

## 1. La description des données

*Format, ouverture, encodage, compression*



Source : [Gianni Sergi, Le Data Management Plan, DMP ou le Plan de Gestion de données, PGD et OPIDoR, 2018](#)

# Choix des formats

- Liste des formats à privilégier : <https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-formats>
- Multitude de formats possibles, pour une multitude de types de données

Formats fermés	Formats ouverts
Excel (.xls, .xlsx)	Comma separated value (.csv)
Word (.doc, .docx)	Plain text (.txt) ou si le formatage est nécessaire PDF/A (.pdf)
PowerPoint (.ppt, .pptx)	Alternative Format PDF/A (.pdf)
Photoshop (.psd)	TIFF (.tif, .tiff)
Quicktime (.mov)	MPEC-4 (.mp4)

# Créer un plan de gestion des données

## 1. La description des données

### *Systeme d'organisation, de nommage et de gestion des répertoires et des fichiers*

#### Quelques principes de base

- Harmoniser les dates
- Identifier les différentes versions
- Mettre en place des sauvegardes
- Organisation pertinente et cohérente des dossiers et sous-dossiers
  - Ressource : <https://www.ukdataservice.ac.uk/manage-data/format/organising>
- Choisir des noms homogènes et pertinents
- Supprimer des versions obsolètes si nécessaire

# Créer un plan de gestion des données

## 1. La description des données

### *Systeme d'organisation, de nommage et de gestion des répertoires et des fichiers*

- Quelle(s) convention(s) avez-vous choisi?
  - Recommandations : <https://sciencespo.libguides.com/donnees-de-la-recherche/organiser>
  - Décrire précisément comment cette convention est structurée

Exemple :

Titre\_projet\_Jeudedonnées1(identifiant)\_aaaammjj

# Créer un plan de gestion des données

## 2. Standards et métadonnées

### Les métadonnées dans un DMP

Dans un DMP, il faut:

- Expliquer comment les métadonnées seront créées : automatiquement, manuellement, automatiquement + complétés manuellement ([Outils de création de métadonnées](#))
- Expliciter comment vous allez décrire les données,
- Inclure les standards applicables au contenu et au format des métadonnées, (standards de métadonnées+ vocabulaires contrôlés ([exemples](#)))
- Préciser les procédures, les outils, les logiciels utilisés pour recueillir et éditer les métadonnées.

# Les métadonnées

## Définition des métadonnées

- Métadonnées = données sur les données
- Décrivent certaines caractéristiques des données :
  - qui a créé les données,
  - ce que le fichier de données contient,
  - quand les données ont été générées,
  - où les données ont été générées,
  - pourquoi les données ont été générées,
  - comment les données ont été générées.

# Les métadonnées

## Des métadonnées, pourquoi ?

- Elles permettent de décrire des ressources numériques en précisant leurs propriétés
  - Contenu intellectuel (résumé, mots-clés...)
  - Contexte de production (date, producteur...)
  - Caractéristiques des fichiers et des données (formats, taille...)
  - Propriété intellectuelle et droits d'utilisation

# Les métadonnées

## Des métadonnées, pourquoi ?

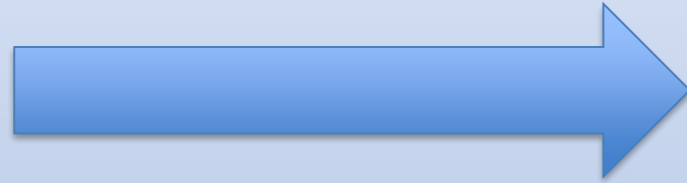
Quand vous créez des métadonnées, vous documentez vos données. Cela vous permet de:

- **retrouver et réutiliser** facilement vos propres données;
- découvrir, évaluer, et réutiliser les **données produites par d'autres**;
- **aider les autres** à découvrir, reproduire, réutiliser, et citer vos données;
- faciliter la **préservation** des données numériques, alors que les logiciels et les formats ne cessent d'évoluer au fil du temps.



# Les métadonnées

Nutrition Facts	
Valeur nutritive	
Per 250 mL / par 250 mL	
Amount	% Daily Value
Teneur	% valeur quotidienne
<b>Calories / Calories 110</b>	
<b>Fat / Lipides 0 g</b>	<b>0 %</b>
Saturated / saturés 0 g	0 %
+ Trans / trans 0 g	0 %
<b>Cholesterol / Cholestérol 0 mg</b>	
<b>Sodium / Sodium 0 mg</b>	<b>0 %</b>
<b>Potassium / Potassium 470 mg</b>	<b>13 %</b>
<b>Carbohydrate / Glucides 27 g</b>	<b>9 %</b>
Fibre / Fibres 0 g	
Sugars / Sucres 23 g	
<b>Protein / Protéines 2 g</b>	
Vitamin A / Vitamine A	0 %
Vitamin C / Vitamine C	120 %
Calcium / Calcium	2 %
Iron / Fer	0 %
Folate / Folate	25 %



Fournit des informations sur



# Les métadonnées

**Notice bibliographique**

Notice Au format public

Auteur(s) : [Colette \(1873-1954\)](#)

Titre(s) : Sido [Texte imprimé] ; suivi de Les vrilles de la vigne / Colette ; préf. d'Alain Brunet

Publication : Paris : Librairie générale française, 2000

Impression : impr. en Allemagne

Description matérielle : 224 p. : couv. ill. en coul. ; 18 cm

Collection : Le livre de poche ; 373

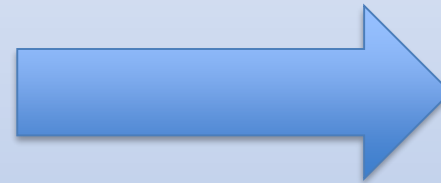
Lien à la collection : [Le Livre de poche](#)

Autre(s) auteur(s) : [Brunet, Alain \(1950-2005\)](#). Préfacier

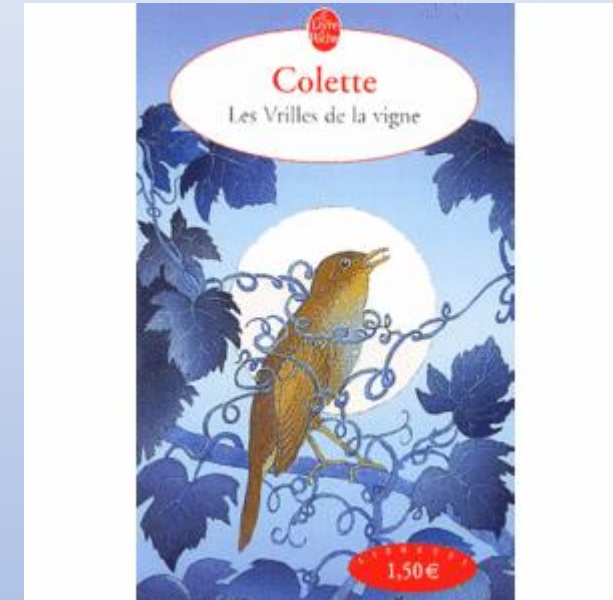
Numéros : ISBN 2-253-00523-1 (br.) : 26 F

Notice n° : FRBNF37108727

 Cette notice appartient à [l'univers jeunesse](#)



Fournit des  
informations sur



# Les métadonnées

## Raw hyperspectral imaging data of Baltic Sea algae cultures

Eskelinen, Matti Aleksanteri; Salmi, Pauliina

### Researcher(s)

Pölönen, Ilkka; Kremp, Anke; Eskelinen, Matti; Salmi, Pauliina

This file archive contains the raw data from hyperspectral imaging of Baltic sea algae cultures performed on 16th of August, 2018 at the hyperspectral imaging laboratory of the Faculty of Information Technology, University of Jyväskylä, Finland.

The dataset contains images of cultures of the following algal species in various dilutions and mixes:

- *Diatoma tenuis* DTTV-1401
- *Melosira arctica* MATV-1402
- *Scirppsiella hangoei* (aka *Apocathium Malmogiense*) SHTV-1
- *Kryptopenidium foliaceum* KFF-1001
- *Monoraphidium* TV70
- *Chlorella pyrenoidosa* TV216

In addition, the dataset includes images of pure water samples, empty petri dishes and millimeter paper useful for transmittance calculations and size measurement.

The imaging setup consisted of living samples pipetted on glass Petri dishes, with a halogen light source illuminating the dish from the bottom towards the camera on top.

The signal in each image contains slight fluctuation in the spectral dimension due to the AC current light source used.

OpenAIRE

### Publication date:

July 3, 2017

### DOI:

DOI: [10.5281/zenodo.822338](https://doi.org/10.5281/zenodo.822338)

### Keyword(s):

[Baltic Sea algae](#) [Hyperspectral imaging](#)

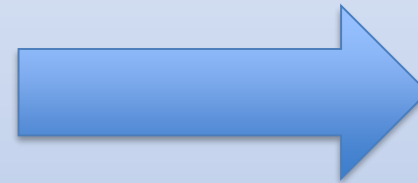
### License (for files):

[CC BY](#) Creative Commons Attribution 4.0

### Versions

Version 1 [10.5281/zenodo.822338](https://doi.org/10.5281/zenodo.822338) Jul 3, 2017

**Cite all versions?** You can cite all versions by using the DOI [10.5281/zenodo.822338](https://doi.org/10.5281/zenodo.822338). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)



Fournit des informations sur

Files (11.0 GB)

### Name

[baltic\\_algae\\_raw.tar.gz](#)

[md5:3ab7170c696af3eb6e889daf77f01059](#) ⓘ

# Créer un plan de gestion des données

## 2. Standards et métadonnées

### Standards des métadonnées

- Il faut indiquer dans le DMP le(s) standard(s) que vous avez choisi(s) pour les métadonnées
- Ressources :
  - [Introduction à la gestion des métadonnées](#)

# Standards de métadonnées

## Dublin Core

- Le **Dublin Core** est certainement le standard de métadonnées le plus utilisé, et il constitue un bon point de départ lorsque l'on commence à travailler sur les métadonnées.
- Le Dublin Core comporte 15 éléments constitutifs (version française : <https://who.rocq.inria.fr/Anne-Marie.Vercoustre/METADATA/DC-fr.1.1.html>)

# Standards de métadonnées

## Des standards interdisciplinaires

- MARC est un format de métadonnées pour le catalogage, très utilisé par les bibliothèques. Il est basé sur le format MARC (MACHINE-Readable Cataloging).
- DataCite Metadata Schema fournit une liste de propriétés choisies pour l'identification précise et cohérente des données en vue de leur citation et de leur découverte.
- Project Open Data Metadata Schema : schéma utilisé pour décrire les jeux de données publiés sur [agency.gov/data](https://www.data.gov/). (données produites dans le cadre de projets de recherche financés par le gouvernement américain).

# Standards de métadonnées

## Pour les SHS

- DDI - Data Documentation Initiative
- OLAC (Open Language Archives Community)
- CMDI
- MIDAS-Heritage
- CARARE metadata schema

# Standards de métadonnées

Pour trouver le standard qu'il vous faut

- Ressources interdisciplinaires
  - [RDA Metadata Standards Directory Working Group](#)
  - [Digital Curation Center](#)
- Doranum : [Les standards de métadonnées : pourquoi et lequel ?](#)



# Créer un plan de gestion des données

## 3. Le partage des données

### Les modalités de dépôt

- Où déposer?
  - Dans un entrepôt de données de la recherche
  - A préciser dans le DMP : quel entrepôt vous allez utiliser?
  - Pour trouver un entrepôt de données, des répertoires et des catalogues sont à votre disposition:
    - Open DOAR, répertoire d'entrepôt en Open Access : <https://v2.sherpa.ac.uk/opensoar/>
    - [re3data](#) ou “Registry of Research data Repositories”
    - La page wiki “Data repositories” fait partie du projet [Open Access Directory](#) . Elle fournit une liste d'entrepôts pour les données ouvertes : [http://oad.simmons.edu/oadwiki/Data\\_repositories](http://oad.simmons.edu/oadwiki/Data_repositories)
    - Liste d'entrepôts recommandés par *Nature* : <https://www.nature.com/sdata/policies/repositories>

# Choisir un entrepôt de données

- Si possible, entrepôt spécialisé pour un type de données
- Recommandé par une communauté
- En fonction des types de données acceptées (discipline, formats)
- Durée de préservation des données
- Qualité de préservation des données
- Identifiants pérennes associés aux données
- Niveau de précision des métadonnées
- Mode d'accès
- Licences acceptées

# Choisir un entrepôt de données

## Quelques exemples d'entrepôts

Nom	URL	Type	Description
Zenodo	<a href="https://zenodo.org/">https://zenodo.org/</a>	Institution publique	Fruit de la coopération entre le CERN et OpenAire. il accueille des jeux de données très différents et fournit un DOI (Digital Object Identifier).
Dryad	<a href="http://datadryad.org/">http://datadryad.org/</a>	Organisation à but non lucratif	Entrepôt de données à vocation pluridisciplinaire, mais dans les faits, il contient surtout des données issues du monde médical.

# Choisir un entrepôt de données

## Quelques exemples d'entrepôts

Nom	URL	Type	Description
Datahub	<a href="https://datahub.io/fr/">https://datahub.io/fr/</a>	Organisation à but non lucratif	Entrepôt interdisciplinaire. Il est le fruit du travail de l'Open Knowledge Foundation
Figshare	<a href="https://figshare.com/">https://figshare.com/</a>	Organisation à but non lucratif	Les chercheurs peuvent déposer gratuitement leurs données dans l'entrepôt. Figshare propose aussi une solution payante aux institutions pour les aider à gérer leurs données.

# Créer un plan de gestion des données

## 3. Le partage des données

### Les modalités de dépôt

- Quand déposer?
  - Jeux de données générés ou recueillis dans le cadre du projet : dès que possible, mais pas d'exigence d'immédiateté
  - Autres données? Possible, à préciser dans le DMP en même temps la date limite de dépôt des données

# Créer un plan de gestion des données

## 3. Le partage des données

### Les modalités de partage

- Procédures d'accès aux données
  - Accès immédiat ou embargo ?
  - Accès ouvert à tous ou restreint à des groupes spécifiques ?
- Comment peut-on y accéder ?
  - Outils et instruments nécessaires pour vérifier et réutiliser les données (logiciels, protocoles d'analyse...)
  - Associer un identifiant unique et pérenne à vos données. Exemple : [DOI](#) pour les objets, [ORCID](#) pour les contributeurs/producteurs

# Créer un plan de gestion des données

## 3. Le partage des données

### Les licences

- Pour que les données soient réutilisables, il est nécessaire de les associer à des licences.
- Ajouter la déclaration de ce type aux données : « Ces données/  
Cette base de données est mise à disposition sous la [licence choisie]. Le texte complet de [nom de la licence choisie et version si nécessaire] est disponible ici : [url correspondante] »

# Les licences de réutilisation

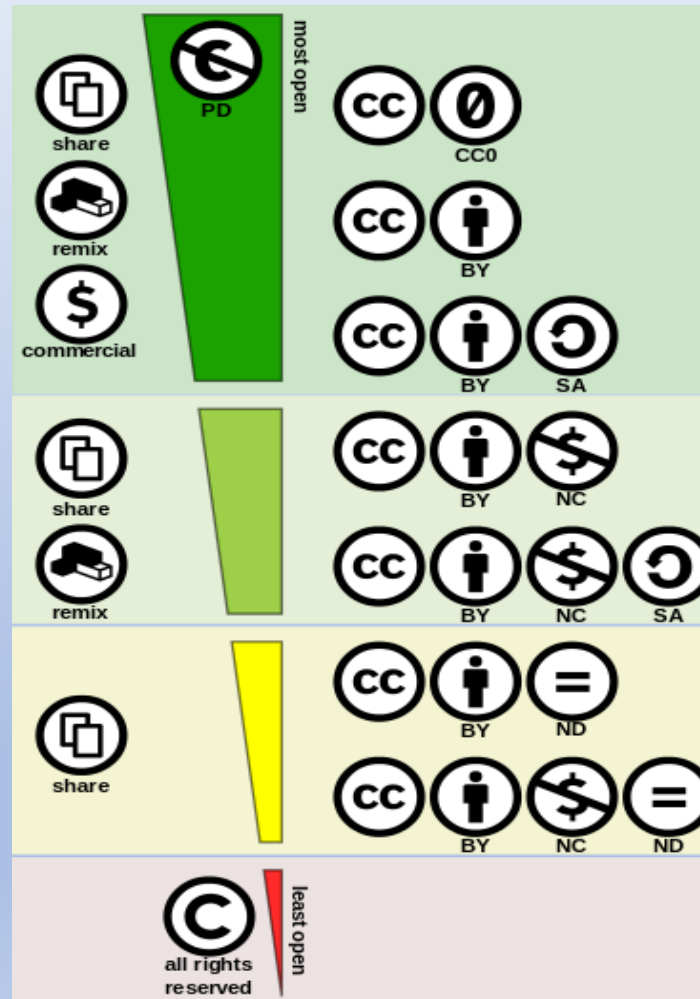
- Les licences Creative Commons créées en 2002 pour la diffusion de contenus numériques.
  - Combinent 4 clauses afin de préciser les droits auxquels l'utilisateur renonce ou non
  - Attribution (sigle BY) : paternité.
  - Non commercial (sigle NC)
  - Pas de modification (sigle ND - No Derivative Works)
  - Partage dans les même conditions (sigle SA - Share Alike)

**Recommandation : utiliser une licence CC-BY (respect de l'attribution).**



# Les licences de réutilisation

## Les licences Creative Commons



# Les licences de réutilisation

## Les licences

- La licence [CC-by 4.0](#) (Creative Commons Attribution) impose seulement de créditer l'les auteur(s) des données.
- La licence [CC0](#) (Creative Commons Public Domain Dedication) de type domaine public. Imposée par l'entrepôt de données pluridisciplinaire Dryad.

# Les licences de réutilisation

## Les licences

- La licence [ODC-by](#) (Open Database Commons) est très proche de la licence CC-by.
- La licence [ODC-ODBL](#) (Open database License) est très proche de la licence Creative Commons by-sa.

# Les licences de réutilisation

## Les licences

- La licence PDDL (Public domain dedication and license) de type domaine public
- La Licence ouverte (LO) créée en 2011 par la mission Etalab pour faciliter la diffusion des données publiques françaises sur la plateforme [data.gouv.fr](http://data.gouv.fr)

# Le cadre juridique

- Du point de vue légal, certains jeux de données ne sont pas librement communicables:
  - les données à caractère personnel
  - les données relatives à l'exécution d'un [contrat de prestation de services](#) exécuté pour le compte d'une ou de plusieurs personnes déterminées (non publiques)
  - les données relatives à la sécurité publique ou au secret
  - les données relatives au secret professionnel.
- Du point de vue éthique, il peut aussi être nécessaire de ne pas divulguer certaines données.
- Dans le DMP, il faut expliquer les raisons qui vous empêchent de partager vos données
- Ressources :
  - [Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France \(2020\)](#)
  - Cf. [La gestion des risques](#)

# Créer un plan de gestion des données

## 4. L'archivage des données

### Archiver ≠ diffuser

- Diffuser:
  - Accès facile aux données
  - Explorer les données
  - Navigation riche
- Archiver:
  - Conserver sur le long terme
  - Accès pérenne dans le temps
  - Navigation simple (en général)

# Les métadonnées pour l'archivage pérenne

Cela requiert aussi de conserver toutes les **données d'accompagnement en plus des données brutes**

- Avec les métadonnées d'accompagnement, on assure :
  - La compréhension des données
  - Le contexte de leur collecte
  - Les conditions de restitution
- Deux normes pour ces métadonnées :
  - [Metadata Encoding and Transmission Standard \(METS\)](#) permet de décrire des ressources numériques.
  - [PREMIS](#) (Preservation Metadata: Implementation Strategies) fournit un ensemble de métadonnées que la plupart des entrepôts d'objets numériques auraient besoin d'enregistrer et d'utiliser pour préserver ces objets sur le long terme.
- [En savoir plus sur METS et PREMIS](#)

# Où archives ses données ?

- Evaluation et labellisation des sites d'archivage pérenne:
  - [Data Seal of Approval](#)
  - [Drambora](#) (Digital Repository Audit Method Based on Risk Assessment) : méthodologie pour l'auto-évaluation (DCC)
  - [TRAC](#) (Trustworthy Repositories and Certification Checklist) : outils pour l'audit de sites d'archivage (DCC)
- En France:
  - CINES (Centre Informatique National de l'Enseignement supérieur) : [plateforme d'archivage](#)
  - Huma-Num : [service d'archivage](#) pour les données en SHS



# Le plan de gestion des données

## Les exigences de l'ANR

Demandé dans le cadre du plan national pour la Science Ouverte  
([accéder au document](#))

« Les plans de gestion des données, instrument de définition des règles de construction, conservation et diffusion des données, seront généralisés. Un prix des données de la recherche sera mis en place afin de mettre en valeur et récompenser les équipes qui réalisent un travail exemplaire dans ce domaine. » (p.6)

# Le plan de gestion des données

## Modèle de l'ANR

Modèle disponible en ligne :

<https://anr.fr/fileadmin/documents/2019/ANR-modele-PGD.pdf>

- Adoption du modèle proposé par Science Europe et recommandé par le Comité pour la Science Ouverte
- Utilisation de ce modèle et rédaction avec DMP Opidor
- Possible d'utiliser le modèle de PGD proposé par l'établissement de rattachement - s'il existe.

# Le plan de gestion des données

## Modèle de l'ANR

Y penser avant même le début du projet, au moment du dépôt...  
Gestion des données ne fait pas de l'évaluation du projet **MAIS**:

- [Recommandations du Comité pour la Science Ouverte](#) adressées à l'ANR : Mentionner la gestion des données dans la proposition détaillée (n° 5), et l'évaluer dès la sélection des projets (n° 3)
- Important de mettre tous les atouts de son côté en tenant compte des principes de la Science Ouverte lorsque l'on répond à l'AAP de l'ANR
  - Anticiper toutes les éventualités, de façon à ne pas avoir d'imprévus que le budget ne pourrait pas couvrir
  - Meilleure faisabilité des projets
  - Plus globalement, développer une politique solide en matière de Science Ouverte
- Pour plus d'infos : [Améliorer les chances de succès de son projet ANR grâce à la Science Ouverte \(2020\)](#)

# Le plan de gestion des données

## Modèle de l'ANR

- Diffusion des données n'est pas obligatoire (notamment si cela présente un risque pour des personnes physiques ou morales), mais fortement recommandée
  - Principe : « aussi ouvert que possible, aussi fermé que nécessaire »
  - Toutes les données ne peuvent pas être ouvertes. Pour vous assurer de respecter le cadre juridique : guide sur [l'Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France \(2020\)](#)
- Aucun champ du PGD de l'ANR n'est obligatoire. Si besoin, indiquer « sans objet ».
- Frais de gestion et diffusion des données pris en charge s'ils sont facturés AVANT la fin du projet

# Le plan de gestion des données

## Modèle de l'ANR

**6 parties, déclinées en question pour faciliter la rédaction**

Ces 6 parties sont précédées par « Informations générales » :

1. Description des données - collecte ou réutilisation des données existantes
2. Documentation et qualité des données
3. Stockage et sauvegarde des données durant le projet
4. Exigences légales et éthiques, codes de conduite
5. Partage des données et conservation à long terme
6. Responsabilités et ressources en matière de gestion des données

# Le plan de gestion des données

## Modèle de l'ANR

### Informations générales

- Acronyme
- Nom du coordinateur
- Version du PGD
- Date... etc.

=> Tout ce qui permet d'identifier le projet et la version du PGD

# Le plan de gestion des données

## Modèle de l'ANR

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

#### 1a. Recueil et production des données

- Comment des données existantes seront recueillies ? Provenance, méthode utilisée
- Y a-t-il des restrictions sur ces données ?
- Méthodologie utilisée pour créer de nouvelles données

#### 1b. Formats et volumes des données

- Types de données : numériques, textuelles, tableau, images, sons, vidéos...
- Détailler et justifier les formats utilisés. Soyez très précis sur la qualification des formats. Exemple : .txt = format texte brut, avec le codage des caractères (UNICODE, ASCII)
- Privilégier les formats ouverts. Si pas possible, expliquer pourquoi
- Informations sur le volume (espace de stockage ou nombre de fichiers)

# Le plan de gestion des données

## Modèle de l'ANR

### 2. Documentation et qualité des données

#### 2a. Métadonnées et documentation

- Quelles métadonnées ? (Date, auteur, versions...)
- Standards de métadonnées utilisés et pourquoi
- Organisation des données : convention de nommage, versions, structure des dossiers
- Documentation pour faciliter la réutilisation (comment les données ont été produites ou recueillies)
- Comment les métadonnées seront mises à disposition. Ex. fichier « read me », [codebook](#) (Un livre de référence est un document d'accompagnement destiné à faciliter l'interprétation des données à des fins de reproduction et de réutilisation par un utilisateur final).

#### 2b. Contrôle de la qualité des données

- Production des données : échantillonnage, calibration, relecture par les pairs, utilisation de vocabulaires contrôlés



# La gestion des risques

## Quels risques et comment les gérer ?

### 1. Recueil de données sensibles

- *Description du risque* : décrire les données sensibles collectées
  - Données personnelles
  - Données de patients
  - Données biologiques
- *Gestion du risque* : identifier les personnes qui vont vous aider à gérer les données sensibles
  - Protection renforcée
    - Autorisation d'accès
    - Sécurité renforcée du stockage des données
  - Diffusion des données
    - Anonymisation réversible ou irréversible
    - Pseudonymisation
    - Limitation de la réutilisation à certains usages
    - Demande d'autorisation d'accès

# La gestion des risques

## Quels risques et comment les gérer ?

### 2. Gestion des données

- *Description du risque* :
  - Difficultés pour collecter des données sensibles (autorisations à obtenir)
  - Délais pour collecter des données sensibles
  - Difficultés à gérer l'échange de données entre partenaires
  - Difficultés de curation : du nettoyage des données à leur stockage
- *Gestion du risque* : identifier les personnes qui vont pouvoir vous aider à gérer ces risques
  - Décrire les démarches à effectuer pour obtenir des autorisations
  - Evaluer les délais
  - Mettre en place un protocole d'échange détaillé en collaboration avec les partenaires
  - Envisager toutes les difficultés éventuelles dans la curation des données avec les services supports

# La gestion des risques

## Quels risques et comment les gérer ?

### Les services supports pour la gestion des données sensibles :

- Délégué à la protection des données : [DPO](#)
- Comité de protection des personnes : [CPP](#)
- Comité éthique de l'établissement
- Service juridique de l'établissement
- SCD ou bibliothèque universitaire : aide à la diffusion des données

### Ressources :

- [Guide](#) de la bibliothèque de Sciences Po
- Doranum : [aspects juridiques et éthiques](#)
- Partager des enquêtes qualitatives : [bibliographie](#)
- Anonymisation : [UK Data Service](#) et [National Centre for Research Methods](#)

# Le plan de gestion des données

## Modèle de l'ANR

### 4. Exigences égales et éthiques, codes de conduite

#### 4a. Traitement des données à caractère personnel

- TOUJOURS appliquer les lois sur la protection des données. Ex. [RGPD](#)
- Consentement éclairé des participants, anonymisation, accès bien contrôlé aux données (cf. 3b), etc.

#### 4b. Propriété intellectuelle des données + législation appliquée

- Qui est le propriétaire des données = qui va en contrôler l'accès.
- Expliquer si oui ou non les données seront librement accessibles.
- Restrictions sur les données fournies par des tiers
- Appliquer des licence pour l'accès et la réutilisation des données
- Aspect important : [protection juridique des bases de données](#).  
Comment [droit d'auteur et droit \*sui generis\*](#) sont-ils gérés ?

# Le plan de gestion des données

## Modèle de l'ANR

### 4. Exigences égales et éthiques, codes de conduite

#### 4c. Gestion des questions éthiques et codes déontologiques

- Envisager les problèmes éthiques pour le stockage, l'accès, la réutilisation des données, etc. et montrer qu'on sait les gérer
- Appliquer les codes de conduite nationaux ou internationaux, code éthique institutionnel si nécessaire. Ex. codes déontologique sur la gestion des données appliqué par le Conservatoire d'Espaces Natures de Midi-Pyrénées, charte Ethique et Big Data

# Le plan de gestion des données

## Modèle de l'ANR

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

#### 5a. Partage des données

- Modalités de partage des données + comment on pourra les trouver
- Archivage pérenne des données (cf. [Cines](#))
- Eventuel embargo à justifier
- Expliquer si utilisation exclusive des données : propriété intellectuelle, brevets, etc.
- Réutilisation libre ou limitée des données ? Ex. nécessité d'un accord pour y accéder + pourquoi

#### 5b. Sélection des données à archiver et solution d'archivage

- Justifier l'éventuelle destruction de données (raison légales, contractuelles, réglementaires)
- Justifier le tri qui peut être réalisé
- Donner un aperçu des utilisations possibles de ces données (cadre de la recherche)
- Où les données seront-elles déposées ? Si pas d'entrepôt défini, montrer que tout a été prévu pour faciliter un futur dépôt.

# Le stockage des données pendant le projet

## Quelles solutions ?

- Eviter un hébergement local sur un ordinateur ou un disque dur externe : forts risques de perte des données (pas recommandé)
- Stockage en réseau / disque partagé en réseau sur un serveur du centre de recherche :
  - Sécurisé
  - Peut être difficile d'y accéder pour des personnes extérieures
- Google Drive :
  - Crypter les données personnelles avant de les charger
  - Besoin de bien gérer les accès au moment du partage
  - Attention : conflit avec le RGPD car hébergement dans l'UE pas garanti
- Stockage dans le cloud (ex. Dropbox) : risqué, pas recommandé
- Solutions institutionnelles
  - [Huma-Num Box](#) + [ShareDocs](#)
  - [CNRS My Core](#)
- Solutions privées et payantes comme [ownCloud](#)

# Le plan de gestion des données

## Modèle de l'ANR

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

#### 5c. Méthodes et logiciels nécessaires pour accéder et utiliser les données

- Besoin d'outils spécifiques pour accéder et utiliser les données
- Si données partagées via un entrepôt, expliquer comment seront gérées les demandes d'accès (en direct ou non)

#### 5b. Attribution d'un identifiant unique et pérenne pour chaque jeu de données

- Utilisation ou non d'un identifiant pérenne
- Comment les données pourront-elles être réutilisées dans d'autres contextes. Identifiants pérennes facilitent la découverte et réutilisation des données



# Le plan de gestion des données

## Modèle de l'ANR

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

#### 6c. Responsabilité des données

- Responsabilités dans les diverses activités qu'implique la gestion des données : saisie des données, création des métadonnées, vérification de la qualité...etc. Qui (nom, rattachement, fonction) fait quoi ?
- Responsable du DMP
- Indiquer le rythme des mises à jour du DMP

#### 6b. Ressources (budget et temps) allouées pour partager des données

- Explication du chiffrage pour préparer les données pour leur diffusion
- Donner une évaluation des frais de stockage, matériels, temps de personnel, coûts de préparation des données en vue de leur dépôt dans un entrepôt, les frais d'entrepôt et d'archivage

=> Ressources pour : curation des données + préparation des données en vue de leur partage (FAIRiser les données) + archivage pérenne

# Les coûts

## Quels coûts et comment les évaluer ?

- Les différents coûts :
  - Frais de personnel : prendre en compte les frais additionnels liés à la FAIRisation des données (ex. anonymisation, documentation, etc.)
  - Coûts matériels : production, stockage et archivage des données
  - Prestations de service : accès à un cloud, anonymisation, conseils juridiques, numérisation, transcription
  - Dépenses liées à l'acquisition et maintenance de logiciels nécessaires pour FAIRiser les données (ex. transcription, ou anonymisation comme [Amnesia](#))
  - Frais de diffusion en libre accès (publications et jeux de données - il peut y avoir des *Dataset Processing Charges* (DPC))
  - Eventuels frais généraux : formations et frais de déplacements des formateurs et formatrices
- Les outils pour les évaluer
  - Les évaluer tout au long du projet :
    - OpenAIRE : [identifier et évaluer les coûts de gestion des données de la recherche](#)
    - [Data management costing tool](#) (UK Data service)
    - [Guide de l'Université d'Utrecht](#)
    - [EPFL Cost Calculator for Data Management](#)
    - TU Delft [Data management costs](#)

# Les aides à la rédaction

- Au sein de vos institutions, vous avez des interlocuteurs et interlocutrices : il faut les repérer et s'adresser à eux (elles) **dès le dépôt du projet**
  - Pour vous aider à les identifier => **SOS-PGD** (répertoire des Services Opérationnels de Soutien à la rédaction des Plans de Gestion des Données) : <https://scienceouverte.couperin.org/sos-pgd/>. Recensement des services d'accompagnement à la rédaction de plans de gestion de données.
- Ressources : Doranum (Données de la Recherche Apprentissage numérique <https://doranum.fr/>)
- Grilles de relecture :
  - ANR : <https://doranum.fr/wp-content/uploads/Grille-relecture-PGD-Modele-ANR-V3.pdf>
  - Union Européenne : <https://wiki.surfnet.nl/download/attachments/10875098/Data%20Management%20Plan%20assessment%20grid%20v0%203.pdf?version=1&modificationDate=1479398494019&api=v2>
  - DDC Checklist : <https://www.dcc.ac.uk/DMPs/Checklist>

# Créer un plan de gestion des données

## Les outils disponibles

- DMPonline :
  - créé par le Digital Curation Center (DCC)
  - En accord avec les recommandations de la Commission Européenne (intégrer H2020)
  - Propose modèles et conseils
- DMP OPIDoR : adaptation de DMPonline par le CNRS
- DMPTool:
  - Créé par l'Université de Californie
  - Modèle simplifié de DMP

# Créer un plan de gestion des données

## DMP OPIDoR

- Data Management Plan pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche :  
<https://dmp.opidor.fr>
- Développé par l'INIST (CNRS)
- Pour l'instant, il propose le modèle pour les projets H2020 et les projets ERC (Conseil européen de la recherche)

# Créer un plan de gestion des données

## DMP OPIDoR

[DMPs publics](#)[Modèles de DMP](#)[Aide](#)[Plus ▾](#)[Français ▾](#)

## Bienvenue !

DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.



Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers



Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données



Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche



Collaboratif : il facilite les échanges entre les partenaires d'un même projet et les services d'accompagnement

DMP OPIDoR évolue grâce à vos retours. Les développements s'inscrivent dans le cadre d'une collaboration internationale autour du logiciel open source DMPRoadmap

[Connexion](#)[S'inscrire](#)

### Accès institutionnel

[Connexion](#)

### Accès individuel

\* **Courriel**

\* **Mot de passe**

[Mot de passe oublié ?](#)

Se souvenir de moi

[Connexion](#)

# Créer un plan de gestion des données

## DMP OPIDoR



Tableau de bord

Créer des plans

DMPs publics

Modèles de DMP

Aide

Plus ▾

 Français ▾

 Marie Puren ▾

## Choisissez un modèle

Vous pouvez choisir soit un modèle fourni par votre organisme soit par un autre organisme, ou un modèle financeur. Le modèle par défaut est **Science Europe - DMP template (english)**.

Retrouvez la liste des modèles disponibles

CNRS (Votre organisme)

Autre organisme

Financier

Souhaitez-vous utiliser le modèle d'un financeur ?

anr|

Agence nationale de la recherche (ANR)

# Créer un plan de gestion des données DMP OPIDoR

Renseignements sur le projet Produits de recherche Modèle choisi **Rédiger** Partager Demande d'assistance conseil Télécharger

**\* Titre du projet**  
Mon projet ANR  
 projet de test, d'entrainement ou à des fins de formation

**Financier**  
Agence nationale de la recherche (ANR)

**Numéro de subvention**

**Résumé du projet**  
B I [Liste à puces] [Liste à puces] [Liens] [Tableau]

### Sélection des recommandations du plan

Pour vous aider à rédiger votre plan, DMP OPIDoR peut vous proposer des recommandations provenant de différents organismes.

Choisir au maximum 6 organismes dont vous souhaitez afficher les recommandations.

Trouver les recommandations d'autres organismes ci-dessous

[Voir la liste complète](#)

**Enregistrer**

Créer des plans DMPs publics Modèles de DMP Aide Plus

### Sélectionner la recommandation

Pour vous aider à rédiger votre plan, DMP OPIDoR peut vous proposer des recommandations provenant de différents organismes. Dans la liste ci-dessous, choisissez jusqu'à 6 organismes proposant des recommandations adaptées à votre plan.

**N'oubliez pas de sauvegarder vos changements après avoir fait vos sélections**

- Digital Curation Centre - DCC
- INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement - INRAE
- Le Mans Université - LMU
- Sorbonne Université - Sorbonne U
- Université Rennes 2 - Université Rennes 2
- Université de Lorraine - Université de Lorraine
- Université de Montpellier - Recommandations UM
- Université de Paris - Université de Paris



# Créer un plan de gestion des données

## DMP OPIDoR

[Tableau de bord](#)[Créer des plans](#)[DMPs publics](#)[Modèles de DMP](#)[Aide](#)[Plus ▾](#)[Français ▾](#)[Marie Puren](#)

### Maître S Plan

[Renseignements sur le projet](#)[Produits de recherche](#)[Modèle choisi](#)[Rédiger](#)[Partager](#)[Demande d'assistance conseil](#)[Télécharger](#)

**Produit de recherche : jeu de données, logiciel, workflow, échantillon, protocole...**

A renseigner séparément pour des produits de recherche nécessitant une gestion spécifique à leur nature ou discipline.

\* **Nom abrégé (20 caractères max.)** MET

\* **Nom complet** silknow\_met1\_20190309


\* **Type**  Jeu de données

**Identifiant Pérenne**



\* **Nom abrégé (20 caractères max.)** MAD

\* **Nom complet** silknow\_mad1\_20190312

\* **Type**  Jeu de données

**Identifiant Pérenne**



[Ajouter un produit de recherche](#)

**Sauvegarder**

# Créer un plan de gestion des données DMP OPIDoR



Tableau de bord

Créer des plans

DMPs publics

Modèles de DMP

Aide

Plus ▾

Français ▾

Marie Puren ▾

Renseignements sur le projet

Produits de recherche

Modèle choisi

Rédiger

Partager

Demande d'assistance conseil

Télécharger

tout développer | [tout réduire](#)

1. Data description and collection or re-use of existing data ( 2 questions )



2. Documentation and data quality ( 2 questions )



3. Storage and backup during the research process ( 2 questions )



4. Legal and ethical requirements, code of conduct ( 3 questions )

5. Data sharing and long-term preservation ( 4 questions )

6. Data management responsibilities and resources ( 2 questions )

tout développer | tout réduire

1. Data description and collection or re-use of existing data ( 2 questions )

MET

MAD

Les réponses de cette section sont communes à tous les produits de recherche

**1a. How will new data be collected or produced and/or how will existing data be re-used?**

**B** *I*

Enregistrer

Recommandations

Commentaires

ANR

- Explain which methodologies or software will be used if new data are collected or produced.
- State any constraints on re-use of existing data if there are any.
- Explain how data provenance will be documented.
- Briefly state the reasons if the re-use of any existing data sources has been considered but discarded

# Créer un plan de gestion des données DMP OPIDoR

Renseignements sur le projet   Produits de recherche   Modèle choisi   Rédiger   Partager   Demande d'assistance conseil   Télécharger

tout développer | tout réduire

1. Data summary ( 1 question )

Summary ( 3 questions )

2. FAIR data ( 4 questions )

FAIR data and resources ( 5 questions )

3. Allocation of resources ( 1 question )

4. Data security ( 1 question )

*This section should describe how you plan to make the project data Findable, Accessible, Interoperable  
Each of the following five issues should be addressed with a level of detail appropriate to the project.*

5. Ethical aspects ( 1 question )

**1. Making data findable**

6. Other ( 1 question )

**2. Making data openly accessible**

Modèle H2020

Modèle ERC

**3. Making data interoperable**

**4. Increase data reuse**

**5. Allocation of resources and data security**

# Créer un plan de gestion des données

## DMP OPIDoR

[Renseignements sur le projet](#)[Produits de recherche](#)[Modèle choisi](#)[Rédiger](#)[Partager](#)[Demande d'assistance conseil](#)[Télécharger](#)

### Gérer les collaborateurs

Inviter des personnes à lire, modifier ou administrer votre plan. Les invités recevront une notification par courriel indiquant qu'ils ont accès à ce plan.

Adresse courriel	Permissions
marie.puren@cnrs.fr	Propriétaire

### Inviter des collaborateurs

**\* Courriel**

**\* Permissions**

- Co-propriétaire: peut modifier les détails du projet, changer la visibilité et ajouter des collaborateurs.
- Editeur: peut commenter et effectuer des changements
- Lecture seule: peut voir et commenter, mais ne peut pas faire de modifications

Sauvegarder

# Créer un plan de gestion des données

## DMP OPIDoR

Renseignements sur le projet

Produits de recherche

Modèle choisi

Rédiger

Partager

Demande d'assistance conseil

Télécharger

### Demande d'assistance conseil

Cliquez ci-dessous pour permettre au service d'accompagnement de votre organisme (CNRS) de lire et commenter votre plan.

L'équipe DMP OPIDoR prendra en charge votre demande dès que possible. Pour toute autre question, veuillez nous contacter à l'aide du [formulaire de contact](#).

The DMP OPIDoR team will respond to your request as soon as possible. For any other question, please contact us using [the contact form](#).

Vous pouvez continuer à modifier et à télécharger le plan en attendant.

**Demande d'assistance conseil** Une assistance conseil a été demandée.

# Créer un plan de gestion des données

## DMP OPIDoR

Renseignements sur le projet

Produits de recherche

Modèle choisi

Rédiger

Partager

Demande d'assistance conseil

Télécharger

### Paramètres de téléchargement

#### Sélectionner les produits de recherche à télécharger

Tous

MET

MAD

#### Mode d'export des produits de recherche

Export par produit de recherche

Export par section

Export par question

#### Éléments Du Plan

page de renseignements sur le projet

texte de la question et entête de la section

questions non répondues

html

pdf

docx

pdf



# Créer un plan de gestion des données

## DMP OPIDoR

[Tableau de bord](#)[Créer des plans](#)[DMPs publics](#)[Modèles de DMP](#)[Aide](#)[Plus ▾](#) [Français ▾](#) [Marie Puren](#)

Dans le tableau ci-dessous figurent les plans que vous avez créés, ainsi que ceux que vous partagez avec d'autres. Vous pouvez à tout moment les modifier, les partager, les télécharger, en faire une copie ou les supprimer.

Titre du projet	Modèle	Modifié	Rôle	Propriétaire	Test	Visibilité	Partagé	
Marie's Plan	ANR - DMP template (english)	28/03/2021	Propriétaire	Vous	<input type="checkbox"/>	Privé	Non	Actions▾
SILKNOW. Silk heritage in the Knowledge Society: from punche ...	Horizon 2020 FAIR DMP (anglais)	11/02/2021	Propriétaire	Vous	<input type="checkbox"/>	Privé	Oui	Actions▾
Exemple H2020 français	Horizon 2020 FAIR DMP (français)	11/02/2021	Propriétaire	Vous	<input type="checkbox"/>	Privé	Non	Actions▾

# Bibliographie

## Plan de gestion des données

- Cartier Aurore, Moysan Magalie et Reymonet Nathalie, 2018, « Chapitre 11. Gestion des données, partage et conservation pérenne avec le Data Management Plan », *Expérimenter les humanités numériques : Des outils individuels aux projets collectifs, Parcours numérique*, É. Cavalié et al. éd., Montréal, Presses de l'Université de Montréal, p. 159. URL : <http://books.openedition.org/pum/11132>
- Morlock Emmanuelle, Romier Geneviève, 2016, *Les enjeux du « Data Management Plan » pour les chercheurs*. URL : [https://hnlyon2016.sciencesconf.org/conference/hnlyon2016/pages/20160525\\_SensibilisationAuDataManagementPlanPourChercheurs\\_Morlock\\_Romier.pdf](https://hnlyon2016.sciencesconf.org/conference/hnlyon2016/pages/20160525_SensibilisationAuDataManagementPlanPourChercheurs_Morlock_Romier.pdf)
- Reymonet Nathalie, Moysan Magalie, Cartier Aurore et Délémontez Renaud, 2018, « Réaliser un plan de gestion de données « FAIR » : modèle », URL : [https://archivesic.ccsd.cnrs.fr/sic\\_01690547](https://archivesic.ccsd.cnrs.fr/sic_01690547)



# Bibliographie

## Guides et ressources

- Datapartage INRAE : <https://datapartage.inrae.fr/>
- Digital Curation Centre (DCC) : <https://www.dcc.ac.uk/guidance>
- Données de la recherche apprentissage numérique : <https://doranum.fr/>
- Données de la recherche, guide de la bibliothèque de Sciences Po : <https://sciencespo.libguides.com/donnees-de-la-recherche/home>
- Exemples de plans de gestion de données : [https://dmp.opidor.fr/public\\_plans](https://dmp.opidor.fr/public_plans)
- Inist : <https://www.inist.fr/>
- OpenAIRE, “How to create a Data Management Plan for H2020 projects”. URL : <https://www.openaire.eu/how-to-create-a-data-management-plan>

## Gestion des données de la recherche

- Burnette Margaret H., Sarah C. Williams, and Heidi J. Imker. 2016. “From Plan to Action: Successful Data Management Plan Implementation in a Multidisciplinary Project”, *Journal of eScience Librarianship* 5(1): e1101. URL : <http://dx.doi.org/10.7191/jeslib.2016.1101>
- Delasalle Jenny, 2013, “Research Data Management at the University of Warwick: recent steps towards a joined-up approach at a UK university”, *LIBREAS. Library Ideas*, 23. URL : <https://libreas.eu/ausgabe23/10delasalle/>

# Bibliographie

## Gestion des données de la recherche

- Pasek Judith E., 2017, “Historical Development and Key Issues of Data Management Plan Requirements for National Science Foundation Grants: A Review”, *Issues in Science and Technology Librarianship*, Summer 2017; Number 87, DOI:10.5062/F4QC01RP
- Science Europe, 2018, Science Europe Guidance Document. Presenting a Framework for Discipline-specific Research Data Management, URL : [https://www.scienceeurope.org/wp-content/uploads/2018/01/SE\\_Guidance\\_Document\\_RDMPs.pdf](https://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf)
- Wilkinson Mark D. et al., 2016, « The FAIR Guiding Principles for scientific data management and stewardship », *Scientific Data*, vol. 3, p. 160018

## Cadre juridique

- Becard Nicolas, Castets-Renard Céline, Chassang Gauthier, Dantant Martin, Freyt-Caffin Laurence, et al.. 2017, *Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France*. URL : <https://hal.inrae.fr/view/index/identifiant/hal-02791224>
- Castets-Renard Céline et Gandon Nathalie, 2016, « Open data des données de la recherche publique : entre réformes législatives et retour d'expérience sur un guide pratique à destination des chercheurs », *LEGICOM*, vol. N° 56, n° 1, p. 67-75.
- Maurel Lionel, 2018, *La réutilisation des données de la recherche après la loi pour une République numérique. La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques*, Presses Universitaires de Provence. URL : <https://hal.archives-ouvertes.fr/hal-01908766>

# Bibliographie

## Ouverture et partage des données de la recherche

- Borgman Christine L., 2012, « The conundrum of sharing research data », Journal of the American Society for Information Science and Technology, vol. 63, n° 6, p. 1059-1078, URL : <https://onlinelibrary.wiley.com/doi/full/10.1002/asi.22634>
- Dillaerts Hans, 2017, « Ouverture et partage des résultats de la recherche dans l'économie de la connaissance européenne : quelle(s) liberté(s) de circulation pour l'IST ? », Communication management, vol. Vol. 14, n° 1, p. 39-54.
- Gaillard Rémi, 2014, « De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? », URL : <http://www.enssib.fr/bibliotheque-numerique/notices/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche>
- Gérout Marie-Madeleine, 2014, « Sur le terrain des données de la recherche au SCD Lille 1 », Arabesques, n° 73, p. 14-15.
- High level Expert Group on Scientific Data, 2010, Riding the wave. How Europe can gain from the rising tide of scientific data,, rapport, URL : <https://ec.europa.eu/digital-single-market/en/news/digital-agenda-unlock-full-value-scientific-data-high-level-group-presents-report>

# Bibliographie

## Ouverture et partage des données de la recherche

- Jacquemin Bernard, Schöpfel Joachim, Chaudiron Stéphane et Kergosien Eric, 2018, « L'éthique des données de la recherche en sciences humaines et sociales. Une introduction », URL : <https://hal.univ-lille3.fr/hal-01958472>
- Leclere Felicia, 2010, « Too Many Researchers Are Reluctant to Share Their Data », The Chronicle of Higher Education. URL : <https://www.chronicle.com/article/Too-Many-Researchers-Are/123749>
- OCDE, 2007 Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics, URL : <http://www.oecd.org/fr/sti/inno/38500823.pdf>
- Odeh Souad, 2017, « Les données de la recherche : transformation ou transmission du métier de documentaliste », I2D Information, données & documents, vol. Volume 54, n° 4, URL : <https://www.cairn.info/revue-i2d-information-donnees-et-documents-2017-4-page-4>
- Prost Hélène et Schöpfel Joachim, 2015, Les données de la recherche en SHS. Une enquête à l'Université de Lille 3., URL : <https://hal.univ-lille3.fr/hal-01198379>
- Reymonet Nathalie, 2017, « Améliorer l'exposition des données de la recherche : la publication de data papers », URL : [https://archivesic.ccsd.cnrs.fr/sic\\_01427978](https://archivesic.ccsd.cnrs.fr/sic_01427978)

# Bibliographie

## Ouverture et partage des données de la recherche

- Schöpfel Joachim, 2018, « Hors norme ? Une approche normative des données de la recherche », Revue COSSI : communication, organisation, société du savoir et information, vol. , n° 5. URL : <https://hal.archives-ouvertes.fr/hal-01979798>
- Rosemberg Noémie, 2015, « De la définition des données de la recherche »? En quête des données, URL : <https://donneesshs.hypotheses.org/39>
- Schöpfel Joachim, 2018, Vers une culture de la donnée en SHS : Une étude à l'Université de Lille, Rapport, Université de Lille. URL : <https://hal.archives-ouvertes.fr/hal-01846849>
- Schöpfel Joachim, Kergosien Eric et Prost Hélène, 2017, « “ Pour commencer, pourriez-vous définir 'données de la recherche' ? ” Une tentative de réponse », Atelier VADOR : Valorisation et Analyse des Données de la Recherche; INFORSID 2017, Toulouse, France. URL : <https://hal.univ-lille3.fr/hal-01530937>
- Serres Alexandre, Malingre Marie-Laure, Mignon Morgane, Pierre Cécile et Collet Didier, 2017, Données de la recherche en SHS. Pratiques, représentations et attentes des chercheurs : une enquête à l'Université Rennes 2, report, URL : <https://hal.archives-ouvertes.fr/hal-01635186>

# Bibliographie

## Science Ouverte

- Direction de l'Information Scientifique et Technique - CNRS, 2016, Livre blanc – Une Science ouverte dans une République numérique, Marseille, OpenEdition Press. URL : <http://books.openedition.org/oep/1548>
- Saby Mathieu, 2020, « Ouverture Des Publications et Des Données : Les Demandes Des Financeurs (ANR et Commission Européenne) », OSF. URL : <https://osf.io/uk734/>

## Agence National de la recherche

- Agence Nationale de la Recherche, Modèle de Plan de gestion des données (PGD). URL : <https://anr.fr/fileadmin/documents/2019/ANR-modele-PGD.pdf>
- Agence Nationale de la Recherche, 2018, Plan d'action 2019. Appel à projets générique 2019 V1.2, URL : <https://anr.fr/fileadmin/aap/2019/aapg-anr-2019-v1.2.pdf>
- Agence Nationale de la Recherche, Webinaire - La politique science ouverte de l'ANR et le Plan de Gestion des Données (PGD), 2019. URL : [https://anr.fr/fileadmin/documents/2019/Slides\\_ANR\\_webinaire\\_science\\_ouverte\\_-\\_PGD.pdf](https://anr.fr/fileadmin/documents/2019/Slides_ANR_webinaire_science_ouverte_-_PGD.pdf)
- Férét Romain, Bracco Laetitia, Cheviron Stéphanie, Lehoux Elise, Arènes Cécile, & Li Ling, 2020, *Améliorer les chances de succès de son projet ANR grâce à la Science Ouverte*. URL : <http://doi.org/10.5281/zenodo.3741666>

# Bibliographie

## Commission Européenne

- Commission Européenne, H2020 DMP Template. URL : [https://dmp.opidor.fr/template\\_export/11.pdf](https://dmp.opidor.fr/template_export/11.pdf)
- H2020 Programme, 2016, Guidelines on FAIR Data Management in Horizon 2020, URL : [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- H2020 Programme, 2017, Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, URL : [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- Koulocheri Eleni, « What is the EC Open Research Data Pilot? », OpenAIRE. URL : <https://www.openaire.eu/what-is-the-open-research-data-pilot>