

Créer un plan de gestion des données

Marie Puren

Université du Mans – 12 février 2020



Créer un 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.

Créer un 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.

Créer un plan de gestion des données

Pourquoi?

Les « FAIR Data »

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

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

Créer un plan de gestion des données

Pourquoi?

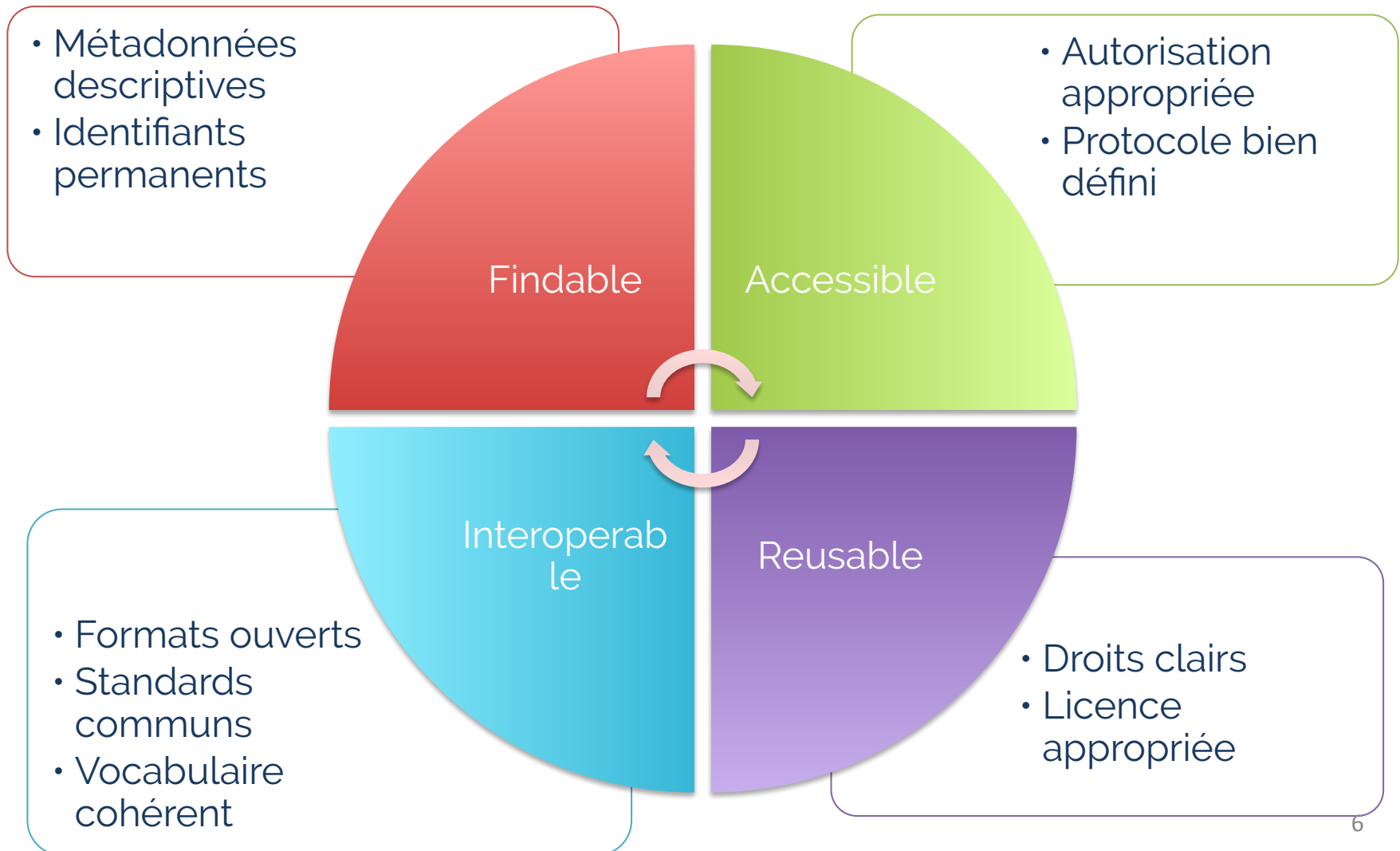
FAIR Data

- 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](#))

Créer un plan de gestion des données

Pourquoi?

FAIR Data



Créer un plan de gestion des données

Pourquoi?

- Rendre ses données **visibles** et montrer qu'elles sont **réutilisables**
- Faire naître une **culture de la donnée** = maîtriser la gestion des données
- Améliorer la **qualité** de ses données
- Reconnaître un **travail** - notamment **collectif**
- Faire gagner du **temps**!

Créer un plan de gestion des données

Comment

- Le DMP doit être régulièrement **mis à jour**.
- Deux versions du DMP : au début et à la fin du projet (cf. exigences H2020)
 - Encore mieux une version intermédiaire entre ces 2 versions!

Créer un plan de gestion des données

Pourquoi?

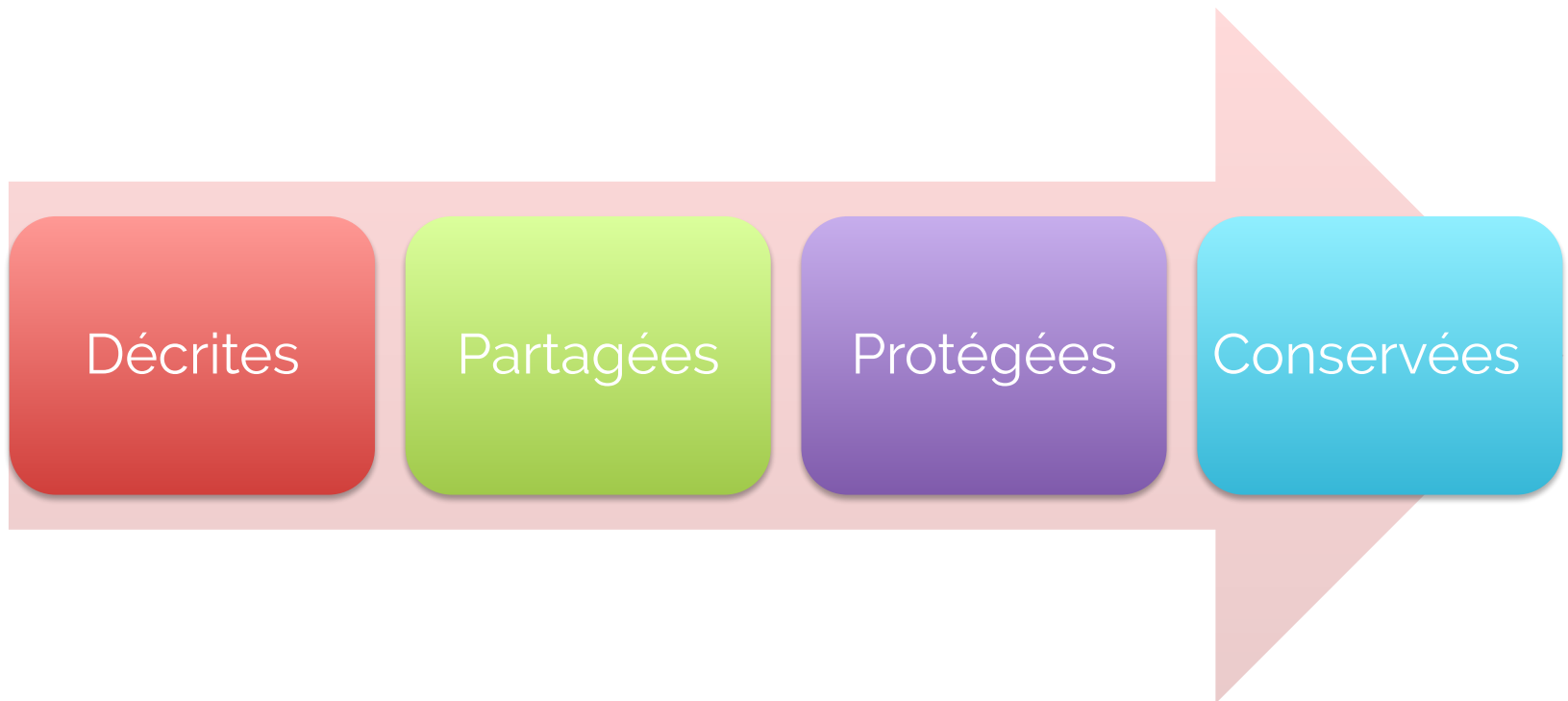
Exigences minimales (1^{ère} version du DMP):

- Une **description des données** qui seront générées ou collectées ;
- Les **standards** et les **métadonnées** qui seront utilisées ;
- Le **partage** des données ;
- L'**archivage et la préservation**

Créer un plan de gestion des données

En résumé

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



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 [ORCID](#) par exemple)
- Le(s) responsables de la gestion des données (nom, coordonnées)
- Dates :
 - De la 1^{ère} 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:

Il faut indiquer :

- Un nom
- Un identifiant

Créer un plan de gestion des données

Les grandes parties d'un DMP

Pour chaque jeu de données, il faut enfin respecter les 4 grandes parties d'un DMP:

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

Origine des données

- Origine des données:
 - 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

- 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)
Photoshop (.psd)	TIFF (.tif, .tiff)
Quicktime (.mov)	MPEC-4 (.mp4)

* [Correspondance formats ouverts / formats fermés](#) (Article Wikipédia)

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?

Exemple :

Titre_projet_Jeudedonnées1(identifiant)_aa
aammjj

Créer un plan de gestion des données

1. La description des données

Système 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
- Choisir des noms homogènes et pertinents
- Supprimer des versions obsolètes si nécessaire

Créer un plan de gestion des données

2. Standards et 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.

Créer un plan de gestion des données

2. Standards et 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

Créer un plan de gestion des données

2. Standards et métadonnées

Des métadonnées, pourquoi ?

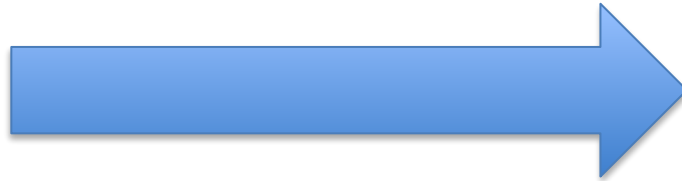
Quand vous créez des métadonnées, vous documentez vos données. Cela vous permet de:

- **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.

Créer un plan de gestion des données

2. Standards et métadonnées

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



Fournit des informations sur



Créer un plan de gestion des données

2. Standards et 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



Créer un plan de gestion des données

2. Standards et métadonnées

Raw hyperspectral imaging data of Baltic Sea algae cultures

Eskelinen, Matti; Aleksanteri, Salmi, Pauliina

Researcher(s)

Pölonen, 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
- *Scorpsella 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)

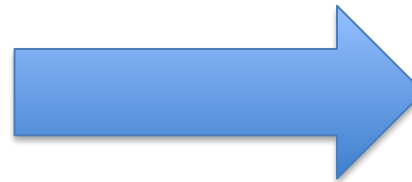
Keyword(s):
[Baltic Sea algae](#) [Hyperspectral imaging](#)

License (for files):
[Creative Commons Attribution 4.0](#)

Versions

Version 1	10.5281/zenodo.822338	Jul 3, 2017
-----------	-----------------------	-------------

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.822337. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)



Files (11.0 GB)

Name

[baltic_algae_raw.tar.gz](#)
md5:3ab7170c696af3eb6e889daf77f01059 ⓘ

Fournit des
informations sur

Créer un plan de gestion des données

2. Standards et métadonnées

Les métadonnées dans un DMP





Dans un DMP, il faut:

- Expliciter **comment vous allez décrire** les données,
- Inclure les **standards** applicables au contenu et au format des métadonnées,
- Préciser **les procédures, les outils, les logiciels** utilisés pour recueillir et éditer les métadonnées.

Créer un plan de gestion des données

2. Standards et métadonnées

Standards and metadata

B *I*    

Save

Not answered yet

Guidance **Share note**

CE Guidance +

Digital Curation Centre guidance on
Metadata -

Questions to consider:

- How will you capture / create the metadata?
- Can any of this information be created automatically?
- What metadata standards will you use and why?

Guidance:

Metadata should be created to describe the data and aid discovery. Consider how you will capture this information and where it will be recorded e.g. in a database with links to each item, in a 'readme' text file, in file headers etc.

Researchers are strongly encouraged to use community standards to describe and structure data, where these are in place. The DCC offers a [catalogue of disciplinary metadata standards](#).

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

Créer un plan de gestion des données

2. Standards et 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>)

Créer un plan de gestion des données

2. Standards et métadonnées

Pour les SHS

- [DDI - Data Documentation Initiative](#)
- [OLAC](#) (Open Language Archives Community)
- [CMDI](#)
- [MIDAS-Heritage](#)
- [CARARE metadata schema](#)
- [CESSDA MLI - Council of European Social Science Data Archives Minimum Level of Information](#)

Créer un plan de gestion des données

2. Standards et métadonnées

Par types de données

- Pour les images : [IPTC](#) (International Press Telecommunications Council)

Créer un plan de gestion des données

2. Standards et métadonnées

Pour trouver le standard qu'il vous faut

- Ressources interdisciplinaires
 - [RDA Metadata Standards Directory Working Group](#)
 - [Digital Curation Center](#)
- [Huma-Num](#) pour les SHS

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:
 - La page wiki “Data repositories” fait partie du projet [Open Access Directory](http://oad.simmons.edu/oadwiki/Data_repositories) . Elle fournit une liste d'entrepôts pour les données ouvertes. Classé par discipline, chaque entrepôt est aussi décrit:
http://oad.simmons.edu/oadwiki/Data_repositories
 - [re3data](http://re3data.org/) ou “Registry of Research data Repositories” est un catalogue d'entrepôts de données.

Créer un plan de gestion des données

3. Le partage des données

Quelques exemples d'entrepôts

Nom	URL	Type	Description
Zenodo	https://zenodo.org/	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	http://datadryad.org/	Organisation à but non lucratif	Entrepôt de données à vocation pluridisciplinaire, mais dans les faits, il contient surtout des données issues du monde médical.

Créer un plan de gestion des données

3. Le partage des données

Quelques exemples d'entrepôts

Nom	URL	Type	Description
Datahub	https://datahub.io/fr/	Organisation à but non lucratif	Entrepôt interdisciplinaire. Il est le fruit du travail de l'Open Knowledge Foundation
Figshare	https://figshare.com/	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
 - 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] »

Créer un plan de gestion des données

3. Le partage des données

Les licences

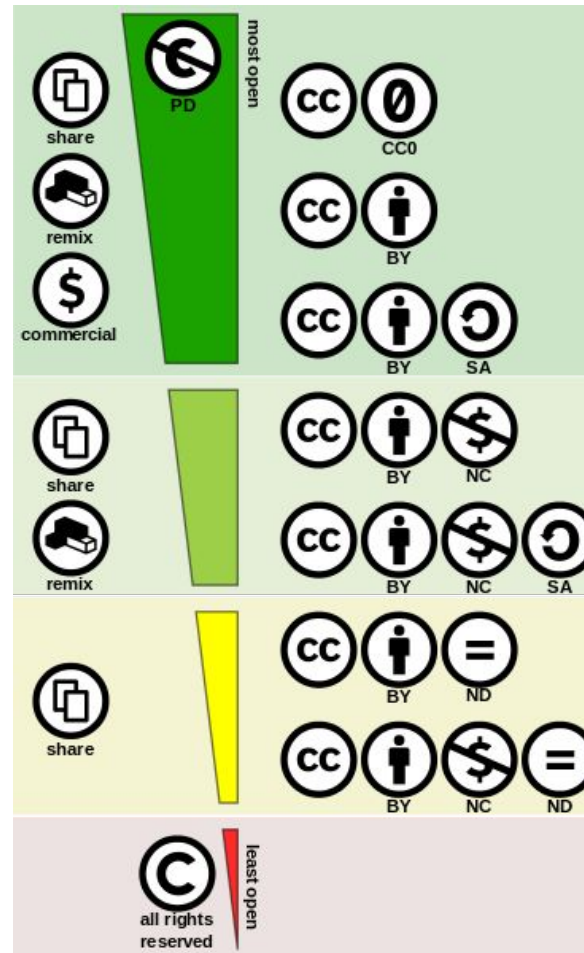
- Les licences [Creative Commons](#) créées en 2002 pour la diffusion de contenus numériques.
 - Combinent 4 clauses afin de préciser les droits auxquels l'utilisateur renonce ou non
 - Attribution (sigle BY) : paternité.
 - Non commercial (sigle NC)
 - Pas de modification (sigle ND - No Derivative Works)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

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

Créer un plan de gestion des données

3. Le partage des données

Les licences
Creative Commons



Créer un plan de gestion des données

3. Le partage des données

Les exceptions

- Du point de vue légal, certains jeux de données ne sont pas librement communicables:
 - les données à caractère personnel => anonymisation (si possible)
 - 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 secrets 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.

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)

Créer un plan de gestion des données

4. L'archivage des données

Où archiver 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 (utilise les services du CINES)

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

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

Titre du projet

☐ projet de test, d'entraînement ou à des fins de formation

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 **Horizon 2020 FAIR DMP (anglais)**.

[Retrouvez la liste des modèles disponibles](#)

CNRS (Votre organisme)

Autre organisme

Financier

Souhaitez-vous utiliser le modèle d'un financeur ?

Commencez à saisir du texte pour afficher une liste déroulante

Merci de sélectionner un organisme de financement valide de la liste.

Utiliser le modèle par défaut

Les plans de gestion de données : bilan et retours d'expérience

- Pour l'instant, pas tant de littérature sur l'expérience française => trop tôt
- Dans le monde anglo-saxon : plus de recul
 - depuis quelques années : retours d'expérience sur l'usage et els apports des DMP

Les plans de gestion de données : bilan et retours d'expérience

- Jenny Delasalle. 2013. "Research Data Management at the University of Warwick: recent steps towards a joined-up approach at a UK university. ". *LIBREAS. Library Ideas*, 23.
<https://libreas.eu/ausgabe23/10delasalle/>
- 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. <http://dx.doi.org/10.7191/jeslib.2016.1101>
- Judith E. Pasek. 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