

Du plan de gestion au Data Paper...

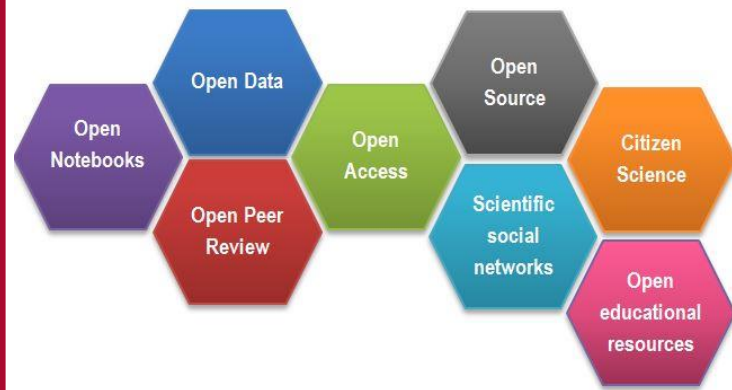
Dominique L'Hostis – INRA Pays de la Loire

Plan

- Open science / Open data : éléments de contexte
- Pour le chercheur
 - Cycle de vie des données et démarche de recherche
 - Questions pratiques
- Éléments-clés
 - Principes FAIR
 - Entrepôts de données
 - Plan de gestion de données
 - Qu'est ce que c'est ?
 - Pourquoi, comment rédiger un PGD ?
 - Voies de diffusion des données
 - Déposer dans un entrepôt
 - Rédiger un data paper

Open Science Open Data

Éléments de contexte



Open Science

Open Science:

Making **process**
outputs
of research available to all

thought process
workflows
code
output generation

process

publications
datasets
images
software/code
proposals
reviews
other

outputs



From Flickr by Ninja M.

Open data

certain data should be **freely available** to everyone to **use & republish** as they wish, **without restrictions** from copyright, patents or other mechanisms of control

Strasser, C. (2016). Scientific data management & open science.
https://figshare.com/articles/Scientific_Data_Management_and_Open_Science/3503411

Pourquoi partager les données ?

Un contexte favorable à l'Open Data

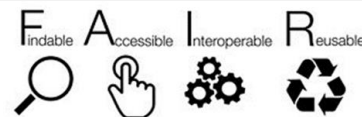
- De plus en plus de « mandats » pour rendre accessibles les données liées aux publications en particulier
 - de l'édition scientifique,
 - des organismes de financement,
 - des politiques nationale et institutionnelle.
- Glissement de la problématique

« Les chercheurs devraient-ils publier leurs données? »



« Quelles données publier ? »
(volet juridique)

« Comment publier les données de manière fiable? »
(volet technique et scientifique)



- Revendication de considérer les données comme un produit de recherche de première classe

Priorités pour les différents acteurs de la recherche

- Réutilisabilité des données,
- Retour sur investissement,
- Innovation

- Gestion des ressources
- Respect des mandats
- Ethique et reproductibilité en recherche
- Collaborations



- Injonction des financeurs
- Qualité du peer-reviewing
- Reproductibilité de la recherche

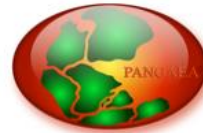
- Meilleure gestion des données
- Sécurisation, préservation des données
- Production académique supplémentaire
- Reconnaissance et crédit,

D'après : Narayana Varma, A. (2017). Data publications as a means to FAIR.
https://figshare.com/articles/Data_publications_as_a_means_to_FAIR/6389126

L'offre d'hébergement se développe ...

- Services en ligne = **entrepôts** (repository) permettant la collecte, la description, la conservation, la recherche et la diffusion des jeux de données,

4TU.Centre for Research Data



<https://www.inra.fr/datapartage/Partager-Publier/Valoriser-ses-donnees/Choisir-un-entrepot>

Financeurs de la recherche

Commission Européenne – Cadre H2020

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

Grantees have the right to opt-out, but need to say why



RESEARCH DATA - OPEN BY DEFAULT



HORIZON 2020 GRANTEES ARE REQUIRED

take measures to ensure open access to the data underlying their scientific publications

provide open access to any other research data of their choice

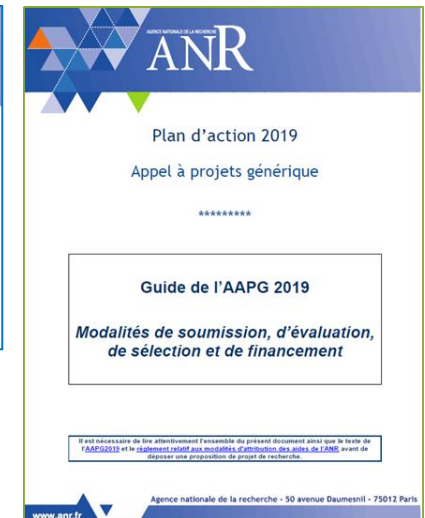
Horizon 2020 grantees are encouraged to also share datasets beyond publication



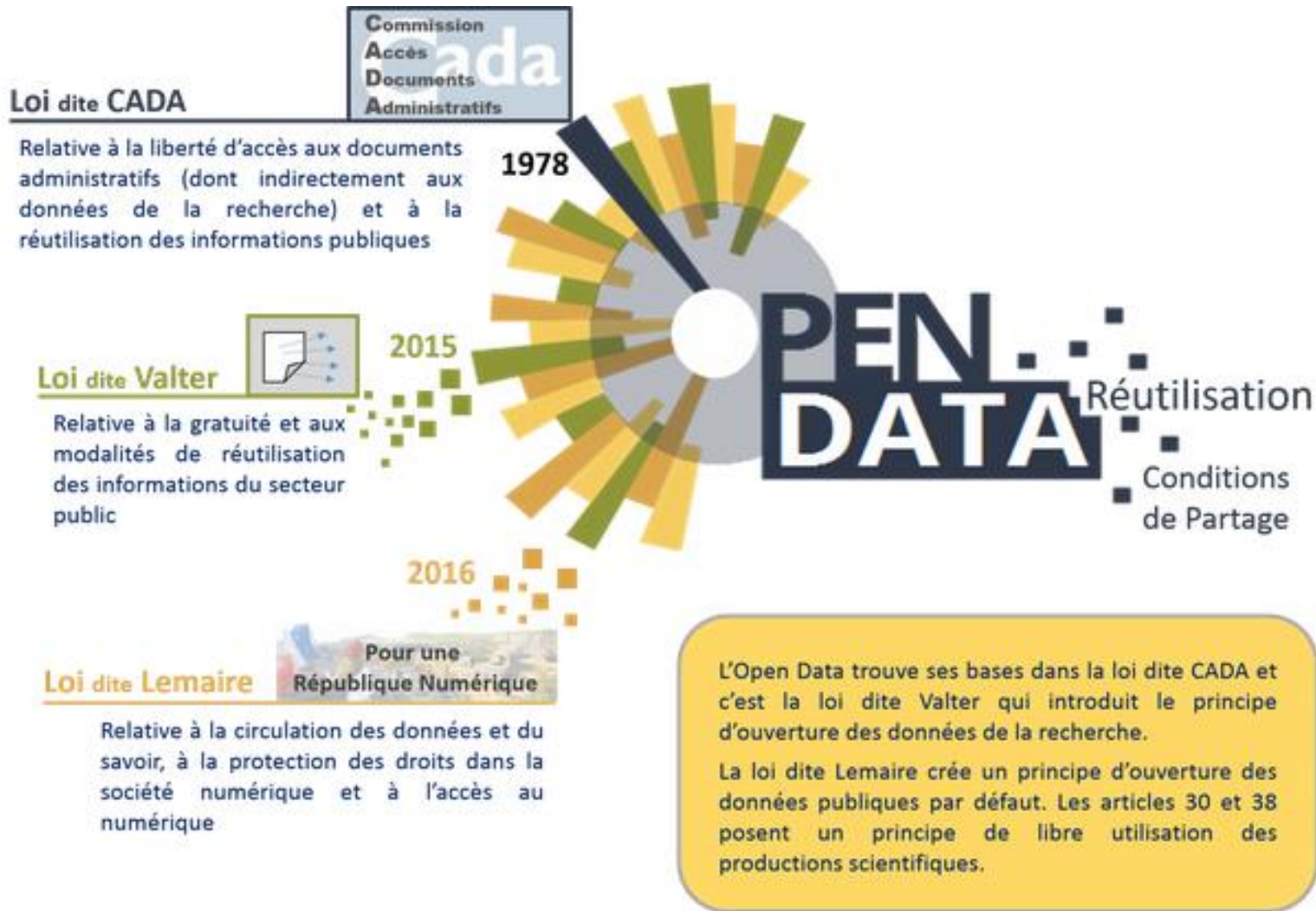
https://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf

ANR

- « Afin de favoriser la diffusion ouverte des données de recherche, l'ANR attire l'attention des déposants sur l'importance de considérer la question des données de recherche au moment du montage et tout au long du projet ».
- Elle impose un plan de gestion des données (DMP) pour les projets financés à partir de 2019.



Le contexte juridique y contribue également



Communicabilité des données

Open – Not Open ?

Open or not Open?

Données
à diffusion interdite

- Les documents réalisés en exécution d'un **contrat** de prestation de services pour une(des) **personnes déterminées privées**;
- Les données relevant du **secret défense**, de la **sécurité publique**;
- Les données relatives aux **secrets professionnels** : secret des procédés, secret des informations économiques et financières, secret des stratégies commerciales ou industrielles;
- Données portant atteinte à la **sécurité du SI de l'administration**

Données communicables
sous conditions

- Données relatives au potentiel scientifique et technique de la nation : **unités protégées** ou en **Zone à Régime Restrictif**;
- Données protégées par le **droit d'auteur** (textes, sons, images) : revenir vers les propriétaires (sauf Text Data Mining à des fins de recherche);
- **Données personnelles** dont les données de santé (de multiples exemples où il faut être attentif : attention aux Systèmes d'Information Géographique à déclarer à la CNIL);
- Données **provenant d'un tiers privé, co-propriété avec un tiers privé**;
- Données protégées par le **secret statistique**.

Données diffusables
en Open Data

- Données financés sur fonds publics (50% +) qui ne sont pas dans les deux autres catégories
- Obligation de diffusion des bases de données et données et leurs mises à jour et présentant, notamment, un intérêt économique, social, environnemental et sanitaire : mise en application pour octobre 2018

Esther Dzalé / AG CNUE – 05/06/2019

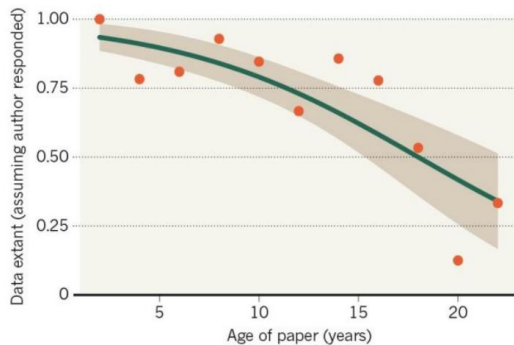


- [Guide juridique](#)
- Andro M., Morcrette, N., Gandon, N. Système expert d'aide à la decision pour diffuser les donnees de la recherche.
<http://www.bibliotheque-numerique.fr/DonneesDiffusables.php>
- Infographie ENPC : https://espacechercheurs.enpc.fr/sites/default/files/logigramme_a_plat.pdf

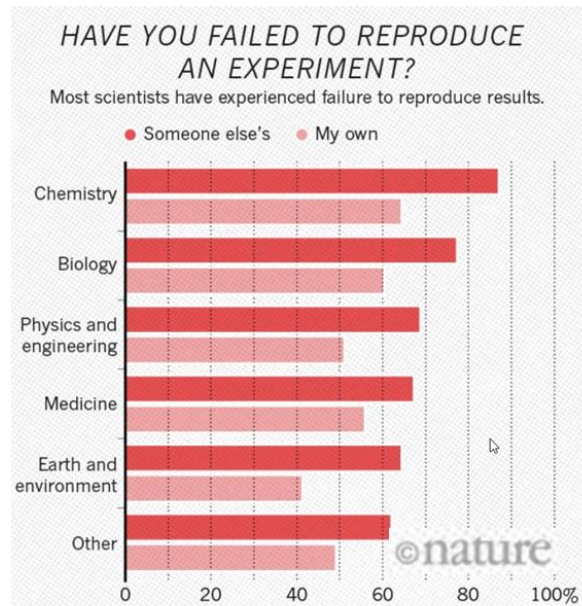
Contexte scientifique

Risques de perte des données, « crise » de la reproductibilité, de l'intégrité scientifique

80% of ecology data irretrievable after 20 years
(516 studies)



Vines TH *et al.* (2013) *Current Biology* DOI: 10.1016/j.cub.2013.11.014



Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, 533(7604). <https://doi.org/10.1038/533452a>

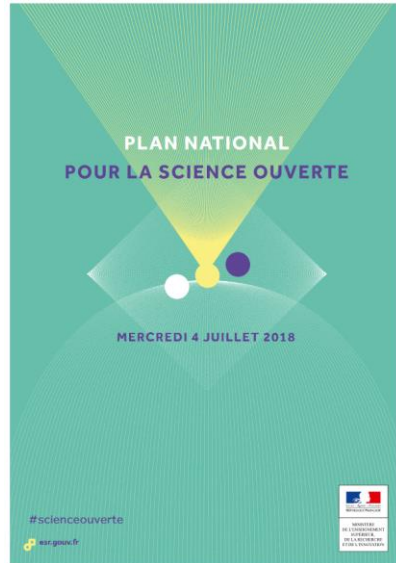
Colloque « Intégrité scientifique et science ouverte » Une synthèse du colloque organisé par l'OFIS le 4 avril 2019



2. Intégrité scientifique et science ouverte, entre opportunités, risques et tensions

Cadre national favorable

Plan national pour la science ouverte (07/18)



1. Généraliser l'accès ouvert aux publications

2. Structurer et ouvrir les données de la recherche

3. S'inscrire dans une dynamique durable, européenne et internationale

[Accès au document](#)



DEUXIÈME AXE : STRUCTURER ET OUVRIR LES DONNÉES DE LA RECHERCHE

Notre ambition est de faire en sorte que les données produites par la recherche publique française soient progressivement structurées en conformité avec **les principes FAIR** (Facile à trouver, Accessible, Interopérable, Réutilisable), préservées et, quand cela est possible, ouvertes. Lors de l'annonce par décret pour toutes les données produites dans le cadre d'appels à projet sur fonds publics. Cette obligation sera limitée par les exceptions légitimes et l'annulation des données, seront généralisées. Un prix des données de la recherche sera mis en place afin de mettre en valeur et récompenser les équipes

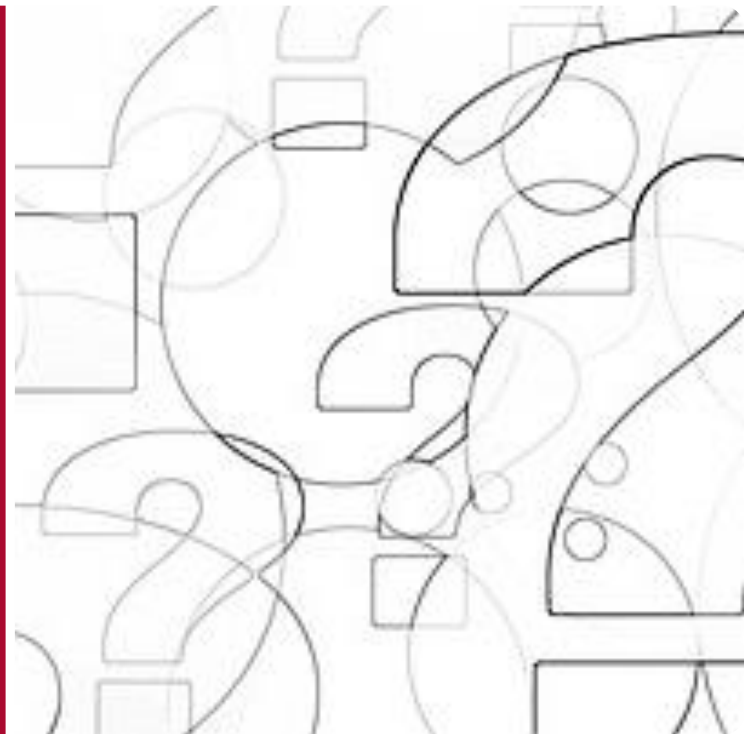
Les plans de gestion des données, instrument de définition des règles de construction, conservation et diffusion des données, seront généralisés. Un

supérieur, de la Recherche et de l'Innovation se dotera d'un administrateur des données de la recherche. Dans le cadre du soutien public aux revues, la

Dans le cadre du soutien public aux revues, la France recommandera l'adoption d'une politique de données ouvertes associées aux articles et le développement des **data papers**. Une politique similaire concernant les thèses sera également mise en place.

Pour le chercheur ?

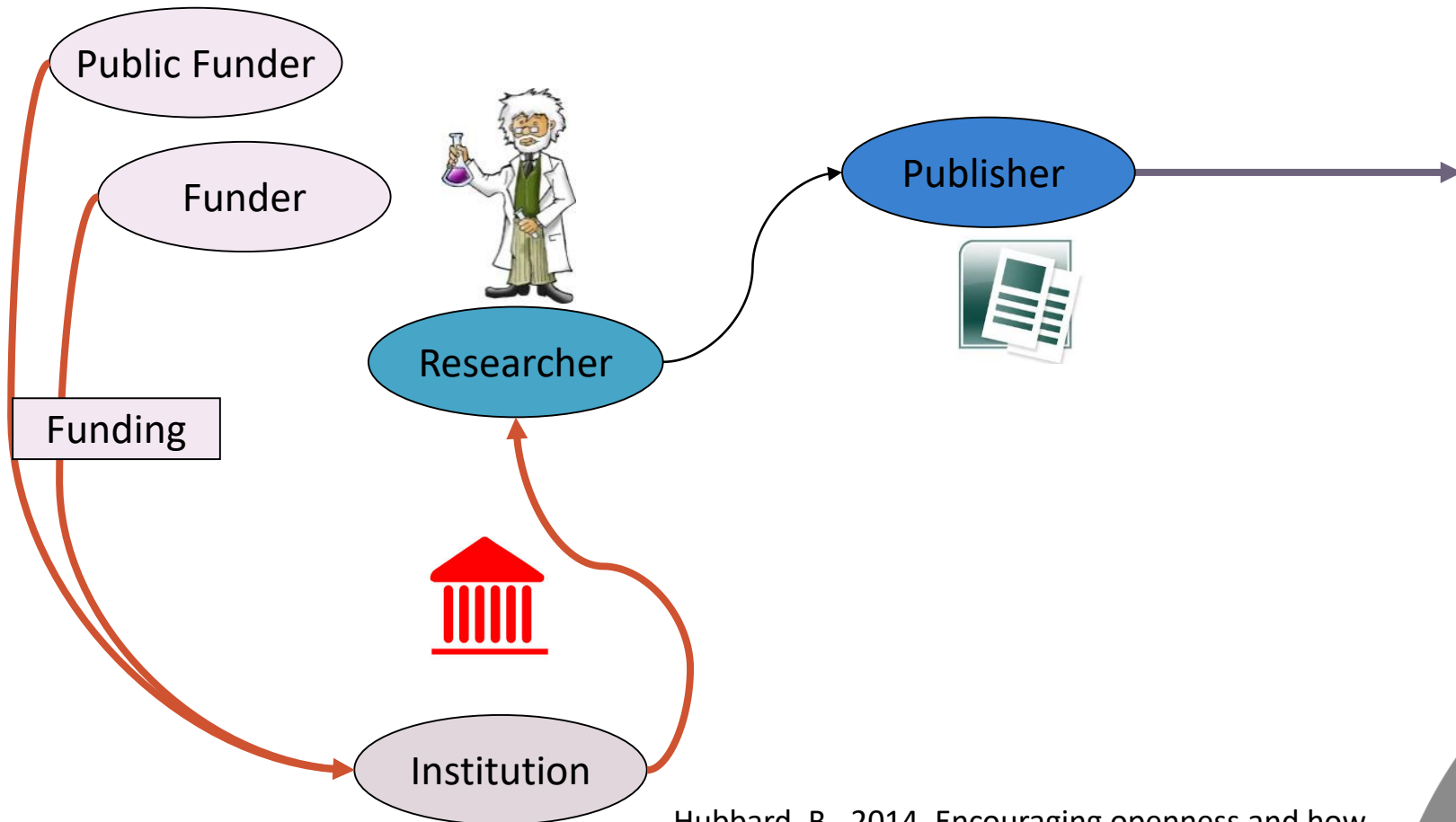
En pratique ...



<https://pixabay.com/illustrations/banner-header-question-mark-1090829/>

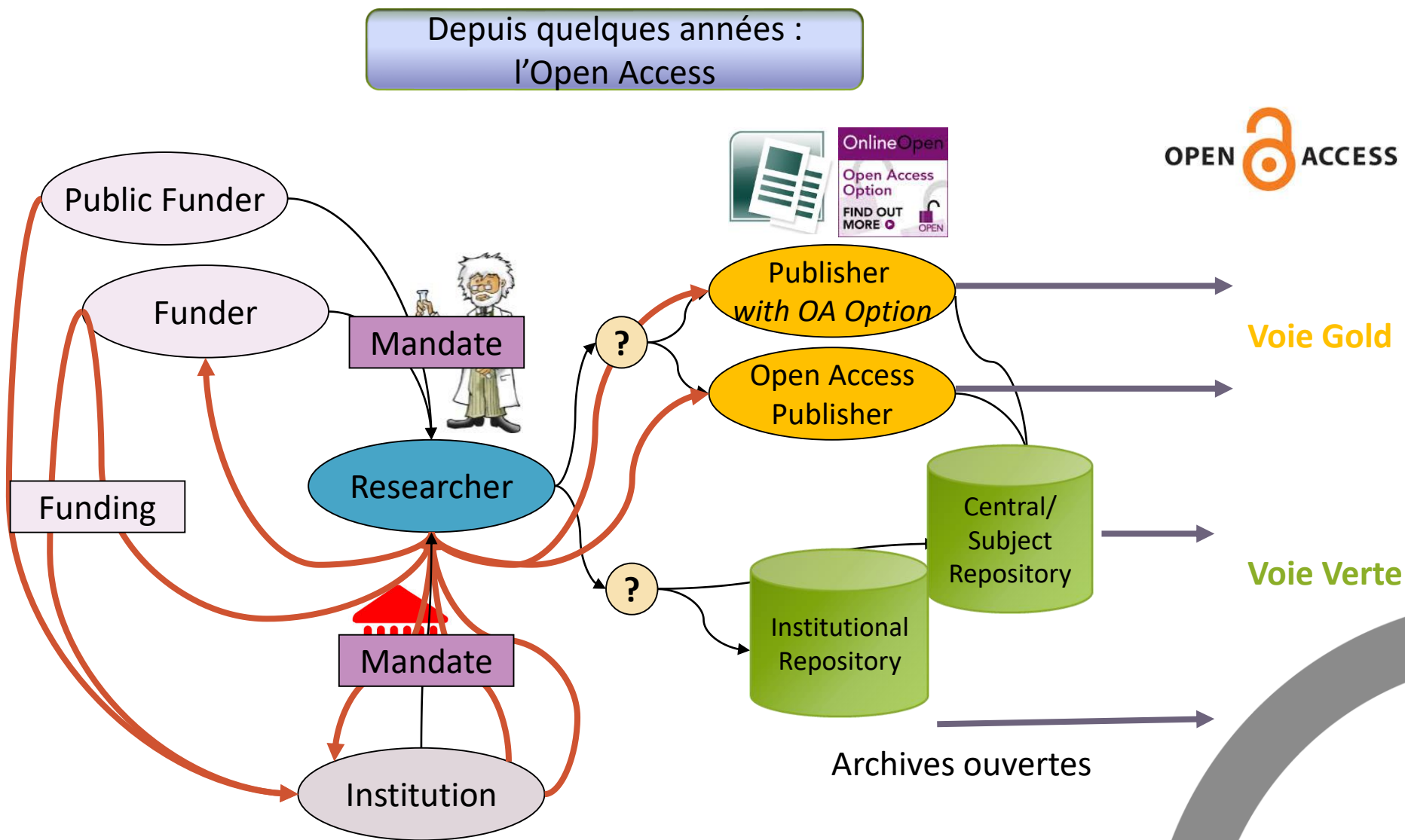
Une situation qui se complexifie pour le chercheur ...

Dans le passé



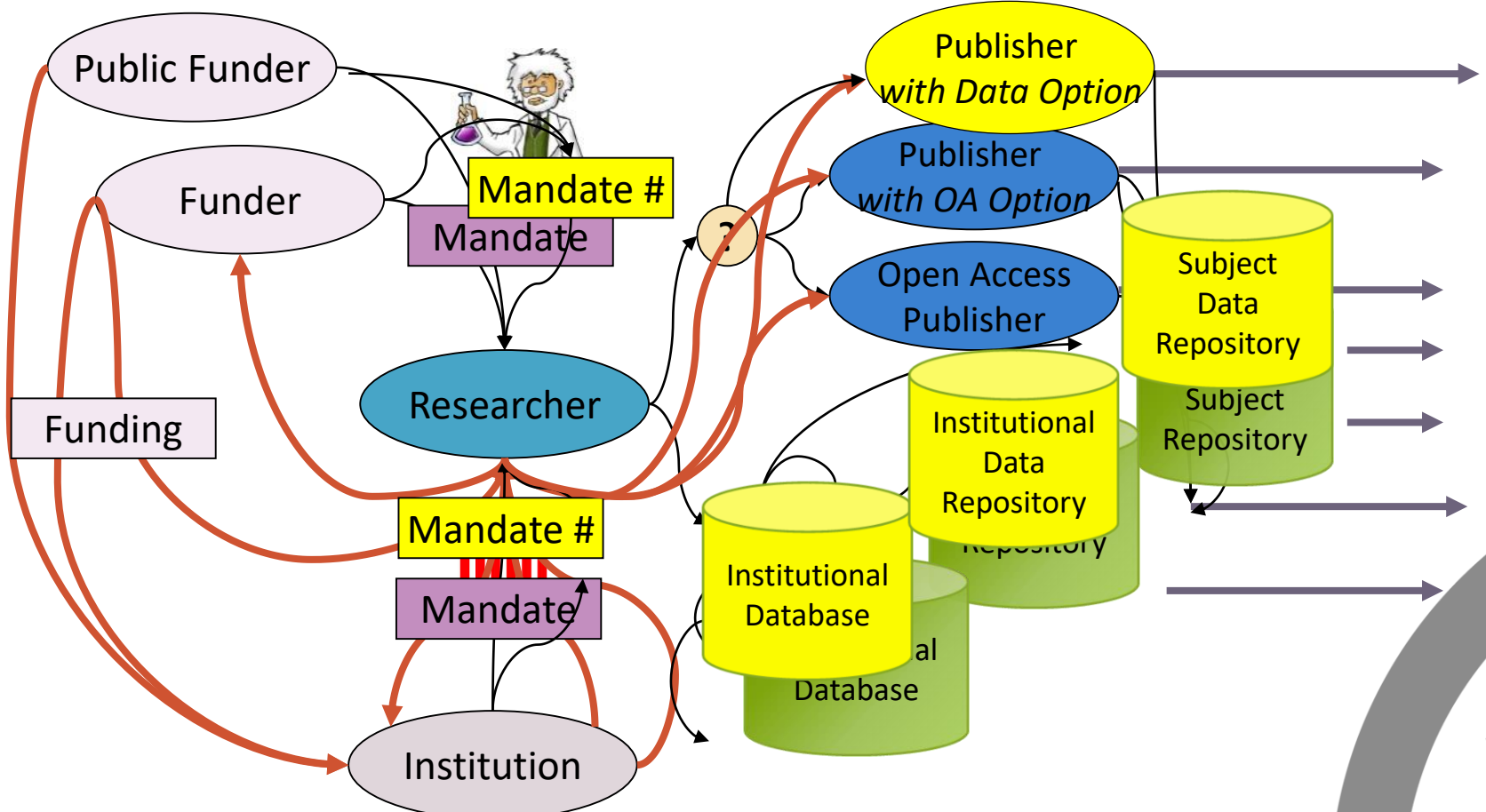
Hubbard, B., 2014. Encouraging openness and how stakeholder policies can support or block it ! (webinar support) 1-28 p.

Une situation qui se complexifie pour le chercheur ...



Une situation qui se complexifie pour le chercheur ...

Et aujourd'hui l'Open data...



Bien gérer ses données pour pouvoir les partager...

« L'ouverture des données ne peut constituer un point de départ, elle ne doit être envisagée que comme le résultat nécessaire d'une bonne politique de gestion de données, qui en constitue le préalable indispensable ».

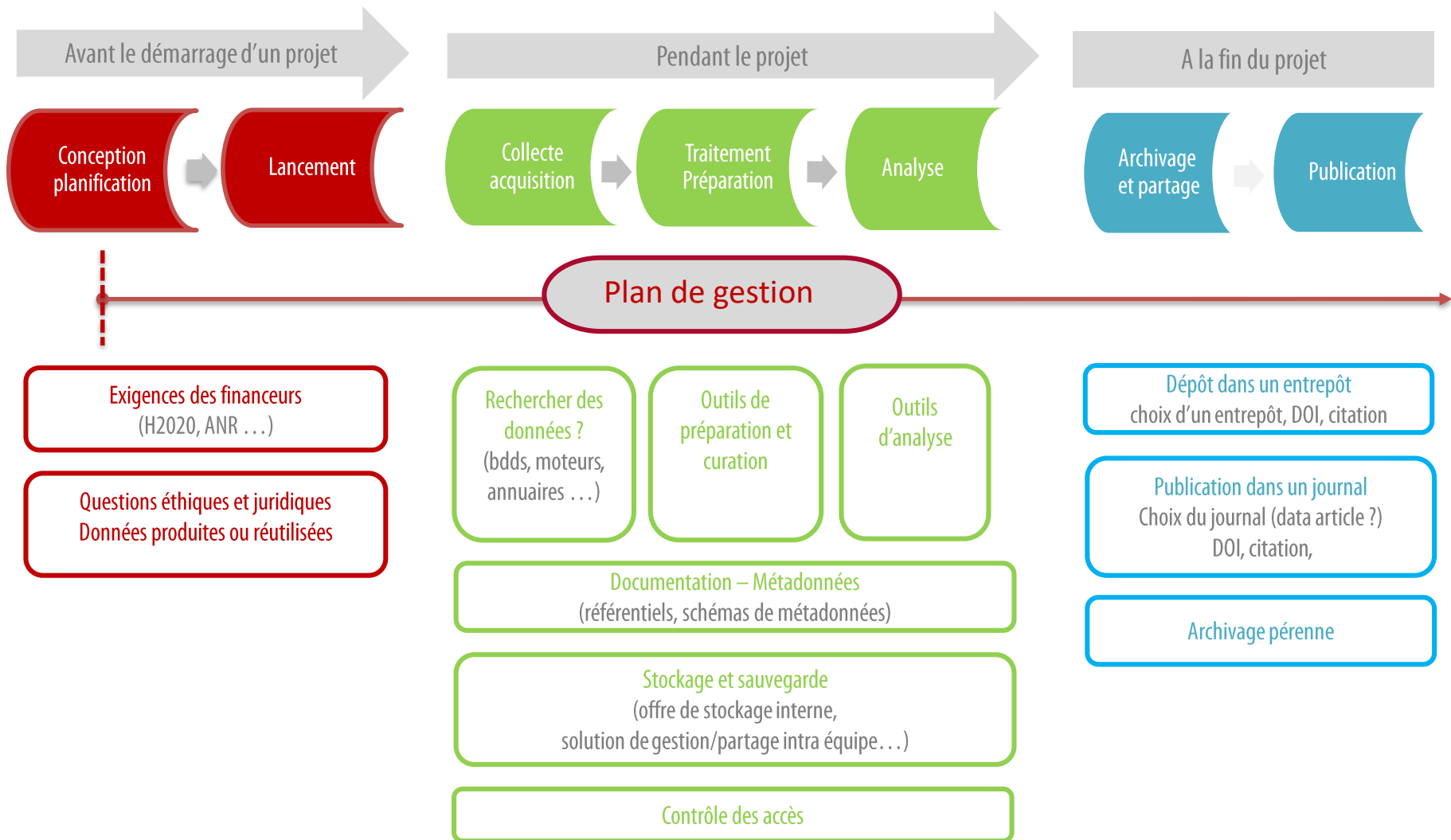
Odile Hologne (2014)

“
Open data needs to
be embedded in the
research process from
start to finish.
”

Open Data. The researcher perspective

<https://www.elsevier.com/about/open-science/research-data/open-data-report>

Cycle de vie des données



Principes FAIR

Questions pratiques



Juridiques

- Quelles données produisons-nous ?
- Qui est propriétaire des données produites ?
- Quel pourcentage de financement public?
- Puis-je - ou suis-je obligé de diffuser les données obtenues ?

[Guide juridique](#)



Techniques

- Comment améliorer la gestion des données au quotidien ?
- Comment implémenter les principes FAIR ?
 - Quelles métadonnées utiliser ? Comment sont-elles générées ?
 - Quels standards utiliser pour structurer les données et les métadonnées ?
 - Quelle(s) autre(s) documentation(s) ?
 - Obtenir un DOI ? À quelle granularité pour le DOI ?
 - Quelle licence de réutilisation ?
 - Où diffuser ?
 - Qui, quoi et comment citer ?



Scientifiques

- Comment répondre au mieux aux exigences des financeurs, aux demandes des éditeurs ?
- Quelles données partager ?
- Quelle stratégie adopter : partage ou publication des données ?
- Quels types de publications (article classique, data article ?)
- Dans quelles revues ?

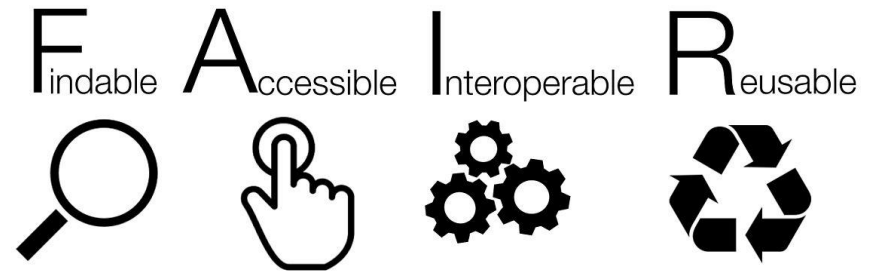
Éléments clés

Principes FAIR, Plans de gestion de données, entrepôts, DOI, Data paper

FAIR !
Soyez FAIR ...



Produire des données FAIR



■ Les « FAIR Guiding Principles »



The Future of Research Communications and e-Scholarship

<https://www.force11.org>



Comment | OPEN | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Phillip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao & Barend Mons - Show fewer authors

Scientific Data 3, Article number: 160018 (2016) | Download Citation

Objectifs

Volume
Variété, complexité
Vitesse de création } des données



- Faciliter / Optimiser la « réutilisabilité » des données
- Les rendre exploitables par les machines et pas seulement par l'homme

What is... FAIR ?

Findable:

- F1.** (meta)data are assigned a globally unique and persistent identifier;
- F2.** data are described with rich metadata;
- F3.** metadata clearly and explicitly include the identifier of the data it describes;
- F4.** (meta)data are registered or indexed in a searchable resource;

Interoperable:

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles;
- I3.** (meta)data include qualified references to other (meta)data;

Accessible:

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol;
 - A1.1** the protocol is open, free, and universally implementable;
 - A1.2.** the protocol allows for an authentication and authorization procedure, where necessary;
- A2.** metadata are accessible, even when the data are no longer available;

Reusable:

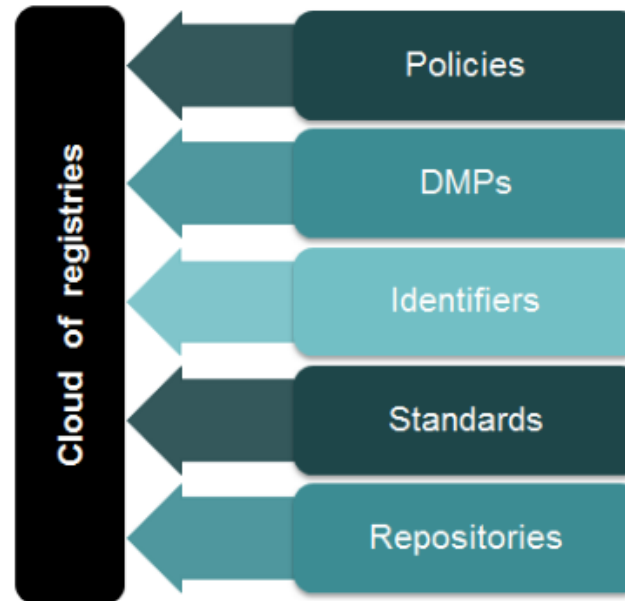
- R1.** meta(data) are richly described with a plurality of accurate and relevant attributes;
 - R1.1.** (meta)data are released with a clear and accessible data usage license;
 - R1.2.** (meta)data are associated with detailed provenance;
 - R1.3.** (meta)data meet domain-relevant community standards;

15 caractéristiques associées à 4 principes

Comment faire FAIR ?

- La mise en œuvre des principes FAIR repose sur différents composants qui doivent être reliés et connectés

- Politique
- Plan de gestion de données
- Identifiants
- Standards
- Entrepôts



Hodson, Jones et al. (2018) Turning FAIR data into reality. Interim report of the European Commission Expert Group on FAIR data. <https://doi.org/10.5281/zenodo.1285272>

Mes données sont-elles FAIR ?

■ Outils d'auto-évaluation

- FAIR data assessment tool du DANS (Data Archiving and Networked Services) <https://www.surveymonkey.com/r/fairdat>
- 5 ★ DATA RATING TOOL du CSIRO : <http://oznome.csiro.au/5star/>

Findable	★★★★★
Accessible	★★★★★
Interoperable	★★★★★
Reusable	★★★★★

■ RDA [FAIR Data Maturity Model WG](#)



<https://rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/results-analysis-existing-fair-assessment-tools>

Mise en œuvre FAIR : exemples



Workflows Allowing Creation of Journal Article Supporting Information and Findable, Accessible, Interoperable, and Reusable (FAIR)-Enabled Publication of Spectroscopic Data

Agustin Barba,[†] Santiago Dominguez,[†] Carlos Cobas,^{*,†} David P. Martinsen,[‡] Charles Romain,^{*,§} Henry S. Rzepa,^{*,§} and Felipe Seoane[†]

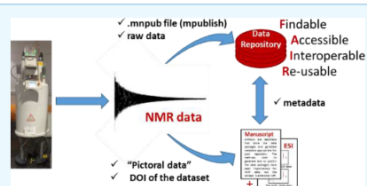
[†]Mestrelab Research, S.L., Feliciano Barrera 9B – Bajo, 15706 Santiago de Compostela, Spain

[‡]David Martinsen Consulting, Rockville, Maryland 20850, United States

[§]Department of Chemistry, MSRH, Imperial College London, 80 Wood Lane, London W12 0BZ, U.K.

Supporting Information

ABSTRACT: There is an increasing focus on the part of academic institutions, funding agencies, and publishers, if not researchers themselves, on preservation and sharing of research data. Motivations for sharing include research integrity, replicability, and reuse. One of the barriers to publishing data is the extra work involved in preparing data for publication once a journal article and its supporting information have been completed. In this work, a method is described to generate both human and machine-readable supporting information directly from the primary instrumental data files and to generate the metadata to ensure it is published in accordance with findable, accessible, interoperable, and reusable (FAIR) guidelines. Using this approach, both the human readable supporting information and the primary (raw) data can be submitted simultaneously with little extra effort. Although traditionally the data package would be sent to a journal publisher for publication alongside the article, the data package could also be published independently in an institutional FAIR data repository. Workflows are described that store the data packages and generate metadata appropriate for such a repository. The methods both to generate and to publish the data packages have been implemented for NMR data, but the concept is extensible to other types of spectroscopic data as well.



<http://doi.org/10.1021/acsomega.8b03005>

Plant Phenomics

A SCIENCE PARTNER JOURNAL

About

RESEARCH ARTICLE

Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS

C. Pommier^{1,†}, C. Michotey¹, G. Comut¹, P. Roumet², E. Duchêne³, R. Flores¹, A. Lebreton¹, M. Alaux¹, S. Durand¹, E. Kimmel¹, T. Letellier¹, G. Merceron¹, M. Laine¹, C. Guerche¹, M. Loaec¹, D. Steinbach¹, M. A. Laporte⁴, E. Arnaud⁴, H. Quesneville¹, and A. F. Adam-Blondon¹

+ Show affiliations

Volume: 2019
Article ID: 1671403
DOI: 10.34133/2019/1671403

Received: 8 January 2019
Accepted: 8 April 2019
Published: 30 April 2019

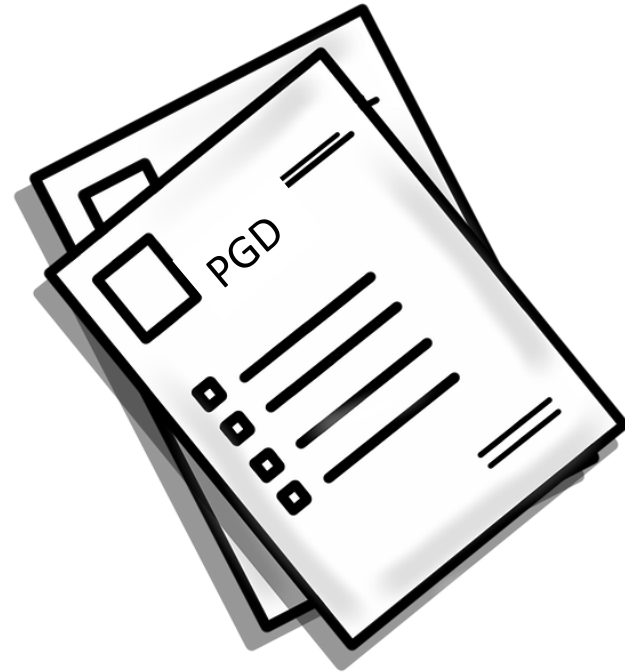
Abstract

GnpIS is a data repository for plant phenomics that stores whole field and greenhouse experimental data including environment measures. It allows long-term access to datasets following the FAIR principles: Findable, Accessible, Interoperable, and Reusable, by using a flexible and original approach. It is based on a generic and ontology driven data model and an innovative software architecture that uncouples data integration, storage, and querying. It takes advantage of international standards including the Crop Ontology, MIAPE, and the Breeding API. GnpIS allows handling data for a wide range of species and experiment types, including multiannual perennial plants experimental network or annual plant trials with either raw data, i.e., direct measures, or computed traits. It also ensures the integration and the interoperability among phenotyping datasets and with genotyping data. This is achieved through a careful curation and annotation of the key resources conducted in close collaboration with the communities providing data. Our repository follows the Open Science data publication principles by ensuring citability of each dataset. Finally, GnpIS compliance with international standards enables its interoperability with other data repositories hence allowing data links between phenotype and other data types. GnpIS can therefore contribute to emerging international federations of information systems.

<http://doi.org/10.34133/2019/1671403>

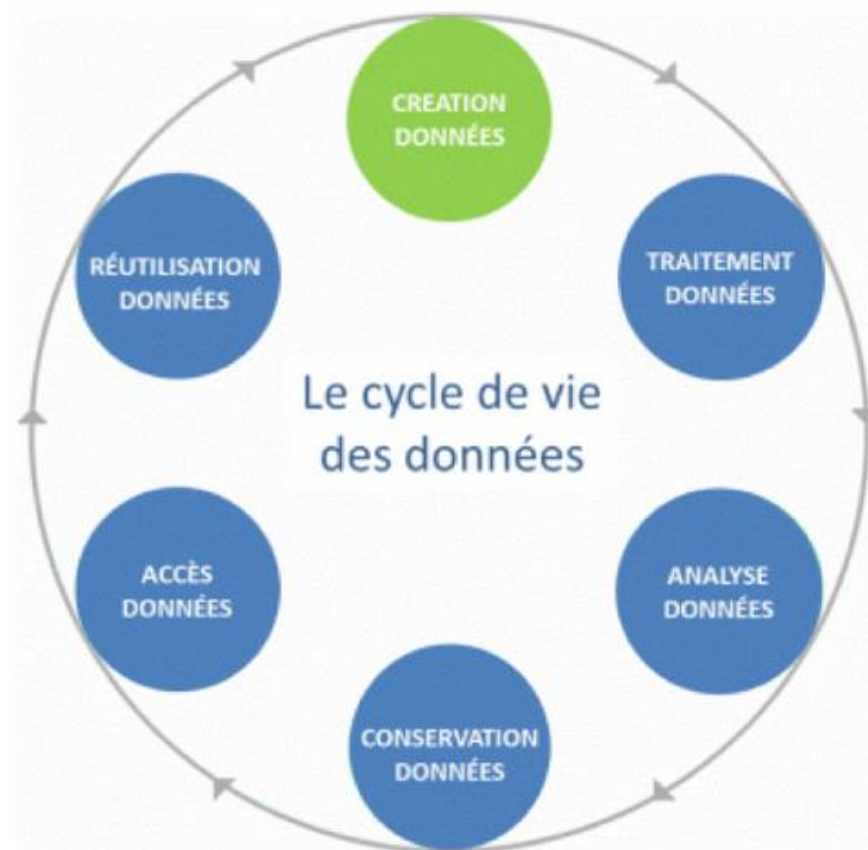
Plans de gestion de données

Élément clé pour produire des données FAIR



Un PGD : qu'est ce que c'est ?

- Document qui décrit la façon dont les données seront :
obtenues, traitées, organisées, stockées, sécurisées, préservées, partagées,...
au cours et à l'issue d'un projet.
- Aide à la mise en place de bonnes pratiques de gestion à toutes les étapes du cycle de vie des données



Cycle de vie des données (source : inist.fr)

A quelles questions répond un PGD ?



- Comment la gestion des données est-elle financée, en particulier à long terme ?

Ressources

- En quoi consiste le projet / Quelles sont les activités de la structure ?
- Qui sont les partenaires ?
- Quelle est la politique en matière de gestion des données ?
- Qui est responsable de la gestion des données ?

Le projet / la structure

- Quelles données seront produites ou réutilisées ? (type, format, volume, accroissement...).
- Comment seront-elles produites et/ou transformées ?

Collecte des données

- Comment les données seront-elles identifiées et décrites ?
- Quels standards de métadonnées utilisera-t-on ?
- Comment seront générées les métadonnées ?

Documentation des données

- Y a-t-il un plan pour l'archivage et la préservation à long terme ?

Archivage et préservation des données

What's the plan?

Dans le cadre d'un projet de recherche ou non...

- Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?

Sauvegarde des données

- Qui pourra accéder aux données ?
- Les données seront-elles partagées ? publiées ?
- Avec qui ? Comment ? Dans quel délai ?
- Sous quelle licence ?

Accès et partage des données

- Qui sera propriétaire des données produites ?
- Des données externes seront-elles utilisées ?

Propriété intellectuelle

- Des données sensibles seront-elles produites ou utilisées ?
- Comment sera assurée leur anonymisation ?

Ethique

Pourquoi rédiger un PGD ?

1- Planifier la gestion des données

- Se poser les bonnes questions en amont d'un projet ou dans le cadre d'une structure pour...
 - **identifier les risques** liés à la gestion des données, assurer la sécurité et la préservation des données sur le long terme,
 - **identifier les responsabilités**, les rôles de chacun dans la gestion des données, planifier les ressources et compétences nécessaires à cette gestion,
 - **garantir des données** fiables et bien gérées, compréhensibles, disponibles et préservées sur le long terme pour une réutilisation future (**démarche FAIR**)

"Plan ahead to create high-quality shareable research data"



<https://www.ukdataservice.ac.uk/manage-data/plan.aspx>

Pourquoi rédiger un PGD ?

1- Élément clé pour produire des données FAIR

Core requirements for DMPs

1. Data description and collection or re-use of existing data

F2, R1.2. How will new data be collected or produced and/or how will existing data be re-used?

I1, R1.2, R1.3. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. Documentation and data quality

I1, I2, I3, R1, R1.3. What metadata and documentation (for example the methodology of data collection and way of organizing data) will accompany data?

I2, R1, R1.2. What data quality control measures will be used?

3. Storage and backup during the research process

a. How will data and metadata be stored and backed up during the research process?

b. How will data security and protection of sensitive data be taken care of during the research?

4. Legal and ethical requirements, codes of conduct

a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

A1.2, R1.1. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

A2. How will possible ethical issues be taken into account, and codes of conduct followed?

5. Data sharing and long-term preservation

A1.2, A2, I3, R1.1. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

A1.1, A1.2, I3. What methods or software tools will be needed to access and use the data?

F1, F3, A2. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

6. Data management responsibilities and resources

a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?



Science Europe. (2018). [Practical guide to the international alignment of research data management](#) (pdf).

Pourquoi rédiger un PGD ?

3- Répondre aux exigences d'un financeur



Comment rédiger un PGD ?

En utilisant un modèle...

■ Financeurs

H2020 FAIR DMP

1. Résumé descriptif des données
2. Données FAIR
 - 2.1. Rendre les données faciles à trouver, en incluant la mise à disposition des métadonnées :
 - 2.2. Rendre les données librement accessibles :
 - 2.3. Rendre les données interopérables :
 - 2.4. Accroître la réutilisation des données (au moyen de licences)
3. Allocation de ressources
4. Sécurité des données
5. Aspects éthiques
6. Autres

ANR (à venir)

■ Organismes de recherche



Comment rédiger un PGD ?

En utilisant un outil d'aide à la rédaction



Optimisation du **P**artage et de l'**I**nteropérabilité
des **D**onnées de la **R**echerche

<https://dmp.opidor.fr/>

- Déployé par l'Inist-CNRS pour l'Enseignement Supérieur et Recherche français
 - gratuité pour l'ESR en France
 - Basé sur le code open source DMP Roadmap développé par le Digital Curation Centre (DCC) et l'University of California Curation Center (UC3).

roadmap

Principales fonctionnalités



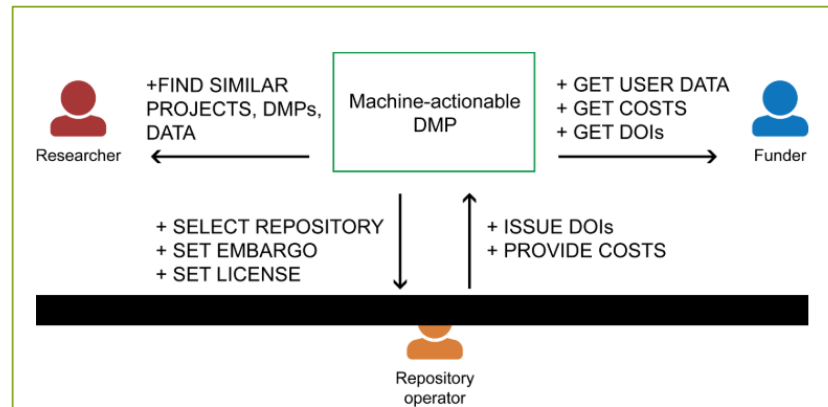
- **Créer et rédiger** un PGD en fonction des exigences d'un financeur ou d'un organisme
 - Mises à jour aisées jusqu'à la version finale
 - Copie possible d'un plan existant
 - Aides et exemples spécifiques en ligne

- **Partager un PGD** avec des collaborateurs
 - Permissions : copropriétaire / éditeur/ lecteur seul
 - Définir la visibilité d'un PGD :
 - privé / administrateur / organisme / public

- **Exporter** un PGD dans différents formats
 - Plan complet ou une sélection de questions

Le plan de gestion

- Un outil pour la **gouvernance des données**
 - processus pour mieux gérer et valoriser les données, qui définit des rôles et des responsabilités et qui renvoie à des questions Techniques, Juridiques, Scientifiques, Economiques, Politiques
- Un **révélateur** de la prise en compte des données au sein d'un projet
- Evolution vers des « Machine Actionable DMP » ([Active DMP](#))
Miksa, T (et al.) [10.1371/journal.pcbi.1006750](https://doi.org/10.1371/journal.pcbi.1006750)



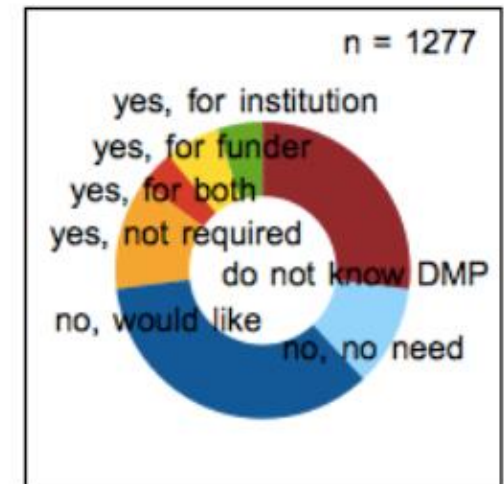
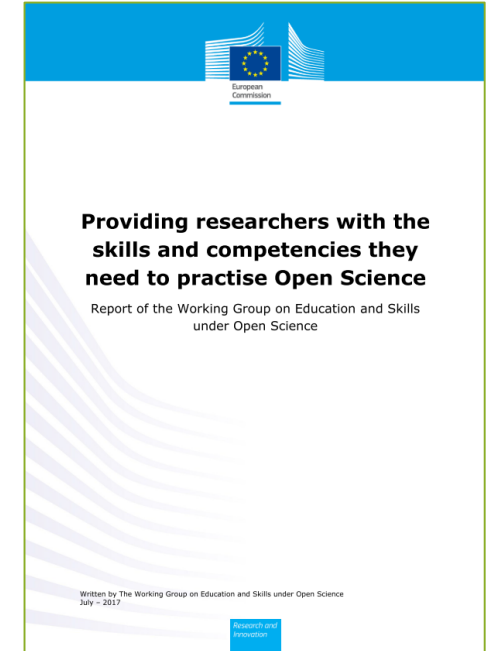
- « Software Management plan » : plan de gestion de logiciel
<https://hal.archives-ouvertes.fr/hal-01802565>

Enquête Europe 2017

Encore beaucoup de méconnaissance

(21) Have you used a data management plan in your research?

		Answers	Ratio
yes, it was required by my institution		67	5.25 %
yes, it was required by my funding body		79	6.19 %
yes, it was required by my institution and funding body		49	3.84 %
yes, it was not required		151	11.82 %
no, I would like to		440	34.46 %
no, I do not need to		145	11.35 %
no, I do not know what a data management plan is		346	27.09 %
No Answer		0	0 %



(O'Carroll, Kamerlin et al. 2017)

Voies de partage des données



Données partagées vs Données publiées



“published” traduit le fait que les données sont **disponibles** au public (car déposées dans un **entrepôt**) et **citables** (grâce à un identifiant), **validées** par l’existence d’un processus d’évaluation (Peer Review)

F1000Research

Kratz J and Strasser C 2014 Data publication consensus and controversies [v2; ref status: indexed, <http://f1000r.es/3hi>]
F1000Research 2014, 3:94
[10.12688/f1000research.3979.2](https://doi.org/10.12688/f1000research.3979.2)

Exemple / dépôt - entrepôt

Entrepôt ouvert
exemple : Data Inra

URGI Plant and Fungi Dataverse (www.urgi.versailles.inra.fr) URGI Website

Portail Data Inra > Omics Dataverse > URGI Plant and Fungi Dataverse > Transposable element annotation of Botrytis cinerea B05.10

Metrics 0 Downloads

Transposable element annotation of Botrytis cinerea B05.10 Version 1.1

URGI, 2018, "Transposable element annotation of Botrytis cinerea B05.10" Portal Data Inra, V1 <https://doi.org/10.15454/TFYH9N> Portal

Cite Dataset

Learn about Data Citation Standards.

Description

The TEdenovo was launched on Botrytis cinerea strain B05.10 whole genome (18 chromosomes, NCBI bioProject PRJNA264284, assembly ASM83294v1). The TE Consensus library from the TEdenovo (36 consensus) was filtered out for consensus sequences classified as SSR, or noCat built from less than 10 fragments (HSPs). The library of the remaining 26 consensus sequences was used to annotate TE copies in the whole genome using TEannot pipeline. Using results of this first TEannot pipeline, we manually eliminate redundancy by filtering out 5 consensus classified class I LTR, one noCat and 6 consensus classified class II MITE containing more than 30 % of micro/mini satellites. We finally kept 15 consensus sequences used to launched a second TEannot pipeline. Genome assembly website: http://fungi.ensembl.org/Botrytis_cinerea/Info/Index Genome annotation files and TE consensus ID mapping between previous and new classification are available for download. (2016-09-01)

Subject

Omics; Plant Health and Pathology

Keyword

transposable_element, Molecular Sequence Annotation

Link to

<http://urgi.versailles.inra.fr/repedb/begin.do#search?taxonGroup=332648>

Terms of Use

Waiver

Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse.

No waiver has been selected for this dataset.

Terms of Use

LICENCE OUVERTE / OPEN LICENCE Licence Ouverte / Open Licence Version 2.0 compatible CC BY

Files

Metadata

Terms

Versions

Licence

Exemple – Data Paper

Data Descriptor | OPEN | Published: 21 August 2018

A global yield dataset for major lignocellulosic bioenergy crops based on field measurements

Wei Li, Philippe Ciais, David Makowski & Shushi Peng

Scientific Data 5, Article number: 180169 (2018) | Download Citation

Abstract

Reliable data on biomass produced by lignocellulosic bioenergy crops are essential to identify sustainable bioenergy sources. Field studies have been performed for decades on bioenergy crops, but only a small proportion of the available data is used to explore future land use scenarios including bioenergy crops. A global dataset of biomass production for key lignocellulosic bioenergy crops is thus needed to disentangle the factors impacting biomass production in different regions. Such dataset will be also useful to develop and assess bioenergy crop modelling in integrated assessment socio-economic models and global vegetation models. Here, we compiled and described a global biomass yield dataset based on field measurements. We extracted 5,088 entries of data from 257 published studies for five main lignocellulosic bioenergy crops: eucalypt, Miscanthus, poplar, switchgrass, and willow. Data are from 355 geographic sites in 31 countries around the world. We also documented the species, plantation practices, climate conditions, soil property, and managements. Our dataset can be used to identify productive bioenergy species over a large range of environments.

<http://doi.org/10.1038/sdata.2018.169>

Metadata summary

Design Type(s)	data integration objective
Measurement Type(s)	total biomass yield
Technology Type(s)	data item extraction from journal article
Factor Type(s)	
Sample Characteristic(s)	Eucalyptus • Populus • Salix • Miscanthus • Panicum virgatum • Earth (Planet)

[Download metadata file](#)

Machine-accessible metadata file describing the reported data (ISA-tab format)

Download Citations

Li, W., Ciais, P., Makowski, D., & Peng, S. Figshare <https://doi.org/10.6084/m9.figshare.c.3951967> (2018)

Cite Download (1.05 MB) Share Embed + Collect (you need to log in first)

Bioenergy crop yield data and reference

57 views | 21 downloads | 0 citations

READ THE PEER-REVIEWED PUBLICATION

A global yield dataset for major lignocellulosic bioenergy crops based on field measurements

SCIENTIFIC DATA

CATEGORIES

- Crop and Pasture Biomass and Bioproducts
- Crop and Pasture Nutrition
- Plant Biology

KEYWORD(S)



bioenergy crop plantings biomass C

LICENCE

CC BY

https://figshare.com/articles/Bioenergy_crop_yield_data_and_reference/6165809/1

Stratégies de partage des données

Voies de diffusion /données	Avantages	Limites & freins
Dépôt dans un entrepôt	<ul style="list-style-type: none"> • Nombreux entrepôts connus et reconnus • Citation directe des données possible • Pas(peu) de limite de taille 	<p>Selon le choix de l'entrepôt</p> <ul style="list-style-type: none"> ○ Métadonnées plus ou moins riches ○ Diffusion orientée vers une communauté d'intérêt spécifique (mais moteurs de + en + plus performants ?)
Intégration dans un article classique	<ul style="list-style-type: none"> • Intégration maximale des données et de l'article : citable, recherchable 	<ul style="list-style-type: none"> • Données difficiles à trouver indépendamment de l'article et dans une forme peu ou pas réutilisable
Matériel supplémentaire d'un article classique	<ul style="list-style-type: none"> • Format des données libéré des contraintes de rédaction de l'article (format, volume, nature des données...) 	<ul style="list-style-type: none"> • Données plus difficiles à trouver indépendamment de l'article et dans une forme peu ou pas réutilisable • Présentation hétérogène et taille de fichier) souvent limitée 
Associées à un Data article	<ul style="list-style-type: none"> • Visibilité accrue : double citation (du Data Paper et du jeu de données)  • Recherche et réutilisation des données facilitée par la richesse des métadonnées • Pas de restriction de taille • Peer Review, crédit / auteurs 	<ul style="list-style-type: none"> • Choix de l'entrepôt pour le jeu de données : disciplinaire (biologie, sciences de la vie, sciences du sol, chimie), institutionnel (Data Inra), générique (Zenodo ...) • Temps de rédaction • Coût de la publication

Entrepôts de données

Pour diffuser ses données



Fonctionnalités des entrepôts

■ Dépôt et conservation

- Téléchargement / import de fichiers
- Organisation des données en collections
- Conservation (stockage sécurisé et archivage à long terme)

■ Identification pérenne des données

- Attribution d'un identifiant au moment du dépôt (dataset, version...)
- Intégration d'identifiants préexistants (Zenodo ...)

■ Description des données

- Métadonnées
 - Schémas de métadonnées –génériques, spécifiques)
 - Qualité des métadonnées (contrôlés ou non)
- Documentation complémentaire

■ Droit d'accès aux données

- téléchargement libre (open)
- embargo
- demande d'accès (restricted), guestbook
- pas de téléchargement possible (closed)
- génération d'URL privé parfois proposé

■ Attribution d'une licence à chaque jeu de données

- Choix ouvert ou liste fermée

■ Recherche, affichage, export des (méta)données

- Recherche
- Affichage (des métadonnées et des données)
- Exports
- Création de template de data article (à partir des métadonnées)

■ Exploration et visualisation des données

- Outils d'analyse ou de visualisation

■ Découverte, visibilité des données

- Exposition des métadonnées
- API pour la recherche et l'accès
- Lien avec d'autres ressources
- Statistiques d'usage des données

■ Gestion des utilisateurs et des droits

Data sharing : recommandations pour le choix de l'entrepôt



D'après : Grootveld, M. (2019, 2019/02). [Data management planning : AMID PROFS].



Arbre de décision

<https://www6.inra.fr/datapartage/Partager-Publier/Choisir-un-entrepot>

Comment choisir un entrepôt ?

(1/2)

Recommandations O/N ?

- Financeur
- Éditeur
Exemple - Scientific Data
<http://www.nature.com/sdata/data-policies#repo-suggest>
- Institution
- Partenaires
- Discipline



Critères de choix ?

- **Caractéristiques liées aux données**
 - Type de données acceptées
 - Formats de fichiers
 - Modifications et versions
- **Caractéristiques liées au partage**
 - Modalités d'accès (restriction d'accès, embargo possibles O/N ?)
 - Lien avec la publication
 - Identifiant pérenne
 - Licences de diffusion
- **Caractéristiques liées aux entrepôts**
 - Disciplinarité
 - Certification
 - Pérennité des données
 - Localisation du serveur
 - Coût du dépôt
 - Facilité du dépôt
 - Statut
 - Statistiques d'utilisation

Comment choisir un entrepôt ?

(2/2)

■ Certification des entrepôts

- Depuis Janvier 2018, les procédures Data Seal of Approval (DSA) et World Data System (WDS) sont alignées et la certification de base devient le CoreTrustSeal (<https://www.coretrustseal.org/>)
- peu d'entrepôts certifiés actuellement



■ Science Europe



Practical Guide to the International Alignment of Research Data Management- [Rapport 2018](#)

Table of Contents	
Foreword by Professor Stan Gielen	2
Introduction	4
CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS	7
CRITERIA FOR THE SELECTION OF TRUSTWORTHY REPOSITORIES	11
GUIDANCE	15
Translating the Core Requirements into a DMP template	
Guiding the Selection of Trustworthy Repositories	
Notes and References	31
Annex: Compatibility with the FAIR Data Principles	32

FAIRSharing

<https://fairsharing.org>



View as Table | View as Grid

Sort by: Best Match

Recommended Records

Recommended

Associated Publication?

No Publication | Has Publication

Claimed?

No Maintainer | Has Maintainer

Record Status

Uncertain | Deprecated | In development | Ready

26 records in view

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	In Collection/Recommendation	Status
MathML	Mathematics Markup Language	MathML	Standard	Mathematical	None	All	None	FuGEFlow, CellML, XML, JATS	None	FAIRDOM Community Standards	R
MAMO	Mathematical Modelling Ontology	MAMO	Standard	Computational Biology, Mathematical, Ontology And Terminology, Systems Biology	None	All	BioModels, OBO	SED-ML	None	OBO Foundry, Computational Modeling	R
ProbOnto	Ontology and Knowledge Base of Probability Distributions	ProbOnto	Database	Computational Biology, Mathematical, Statistics And Probability, Systems Biology	Mathematical Model	None	None	None	None	None	R
PMR	Physiome Model Repository	PMR	Database	Biomedical Science, Life Sciences, Mathematics	Mathematical Model	All	None	None	None	None	R

CML
Chemical Markup Language
MODEL/FORMAT

Implementing databases: 3

Publications: 1

In Collections: 2

Recommended: 0

1 Taxa types, including:

All

19 Data types, including:

Assay, Biological Model, Coordinate Model, Enzyme Kinetics Annotation, Mathematical Model

Exemple : 26 items sur requête : Mathematics

- 14 standards
- 9 databases
- 2 policies
- 1 project

Search...

Search

Filter

Reset all

Subjects

- Life Sciences (1)
 - Biology (1)
 - Basic Biological and Medical Research (1)
 - Biophysics (1)
 - Bioinformatics and Theoretical Biology (1)
 - Agriculture, Forestry, Horticulture and Veterinary Medicine (1)
 - Agriculture, Forestry, Horticulture and Veterinary Medicine (1)
 - Soil Sciences (1)
- Natural Sciences (1)
 - Chemistry (1)
 - Analytical Chemistry, Method Development (Chemistry) (1)
 - Physics (1)
 - Mathematics (1)
 - Geosciences (including Geography) (1)
 - Geophysics and Geodesy (1)
 - Geophysics (1)
 - Geochemistry, Mineralogy and Crystallography (1)
 - Geography (1)
 - Water Research (1)
 - Engineering Sciences (1)
 - Materials Science and Engineering (1)
 - Materials Science (1)
 - Computer Science, Electrical and System Engineering (1)
 - Systems Engineering (1)
 - Traffic and Transport Systems, Logistics (1)
 - Computer Science (1)
 - Construction Engineering and Architecture (1)
 - Construction Engineering and Architecture (1)
 - Urbanism, Spatial Planning, Transportation and Infrastructure Planning, Landscape Planning (1)

Content Types

Countries

AID systems

API

Certificates

DSA (1)

mathematics

Search

Toggle short help

← Previous 1 Next →

Sort by ▾

Found 1 result(s)

4TU.Centre for Research Data
4TU.ResearchData



Subject(s)

- Construction Engineering and Architecture
- Engineering Sciences
- Natural Sciences
- Agriculture, Forestry, Horticulture and Veterinary Medicine
- Agriculture, Forestry, Horticulture and Veterinary Medicine
- Basic Biological and Medical Research
- Life Sciences
- Urbanism, Spatial Planning, Transportation and Infrastructure Planning, Landscape Planning
- Construction Engineering and Architecture
- Computer Science
- Traffic and Transport Systems, Logistics
- Systems Engineering
- Computer Science, Electrical and System Engineering
- Materials Science
- Materials Science and Engineering
- Water Research
- Geography
- Geochemistry, Mineralogy and Crystallography
- Geophysics
- Geophysics and Geodesy
- Geosciences (including Geography)
- Mathematics
- Physics
- Analytical Chemistry, Method Development (Chemistry)
- Chemistry
- Soil Sciences
- Bioinformatics and Theoretical Biology
- Biophysics
- Biology

Content type(s)

- Archived data
- Structured text
- Plain text
- Raw data
- Scientific and statistical data formats
- Audiovisual data
- Images
- Standard office documents

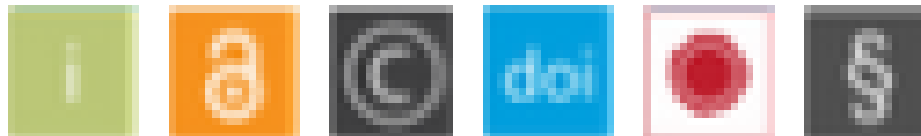
Country

- Netherlands

4TU.ResearchData, previously known as 3TU.Datacentrum, is an archive for research data. It offers the knowledge, experience and the tools to share and safely store scientific research data in a standardized, secure and well-documented manner. 4TU.Centre for Research Data provides the research community with: Advice and support on data management; A long-term archive for scientific research data; Support for current research projects; Tools for reusing research data.

<https://www.re3data.org>

■ Filtres de recherche et fiche descriptive



Data Paper

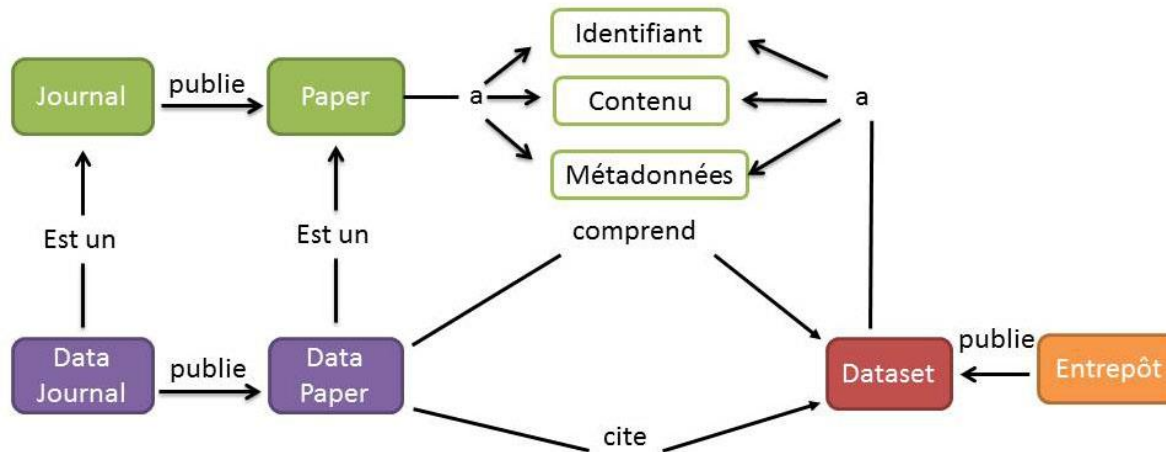


«Data publishing: A Process for Bridging the Gap between Data Author and Data User»

Thanos, C. (2017). [10.3390/publications5010002](https://doi.org/10.3390/publications5010002)

Data Paper + dépôt/données dans un entrepôt: meilleure approche de valorisation...

“A data paper is a searchable metadata document, describing a particular dataset or a group of datasets, published in the form of a peer-reviewed article in a scholarly journal” (GBIF)



Un concept hétérogène

De nombreuses appellations différentes...

- **Data paper**
(*Annals of Forest Science*,
Zookeys ...)
- **Data notes**
(*F1000 Research*,
Gigascience)
- **Data article**
(*Data in Brief*)
- **Data descriptor**
(*Scientific Data*)
- **Dataset brief**
(*Proteomics*)
- **Data in brief**
(*Genomics Data*)
- **Database article – Software article**
(*BMC*, *Chemistry Central Journal*)
- **Data reports**
(*Frontiers in Plant Science*)
- **Resource article**
(*Plant Journal*)
- **Database paper**
(*Plant & Cell Physiology*)
- **Methods, software, databases, and tools**
(*Plos One*)
- ...

Une structure variable...

Des composantes communes mais une **structure très variable** selon les revues

Data Paper

Éléments communs aux autres types d'articles

Titre

Abstract

Mots-clés

Introduction

Méthodes

Résultats

Discussion

Remerciements

Références

Sources de financement

Données

intégrées dans l'article, ou déposées dans un entrepôt, accessibles et identifiées (DOI)

Description des données

Conseils sur la réutilisation des données

Métadonnées

Validation technique

Valeur des données

Où les publier ?

Journaux classiques
intégrant les « **data articles** »

PUBLISHING DATA PAPERS IN ANNALS OF FOREST SCIENCE: DETAILED GUIDELINES FOR A SMOOTH PREPARATION AND SUBMISSION
By **Erwin Dreyer** On 20 November 2015 In *Editor's Choice, Editorial, Editorial News, For Authors*



PLOS ONE
TENTH ANNIVERSARY

PLOS' New Data Policy: Public Access to Data
Posted February 24, 2014 on [Lip Shih](#) in [Aggregators](#), [Open Access](#)

“
Data journals are still a relatively small-scale phenomenon.
”

Data Journals : Journaux spécialisés, dédiés « **data articles** »



SCIENTIFIC DATA
Helping you publish, discover, and reuse research data

(GIGA)ⁿ SCIENCE



A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

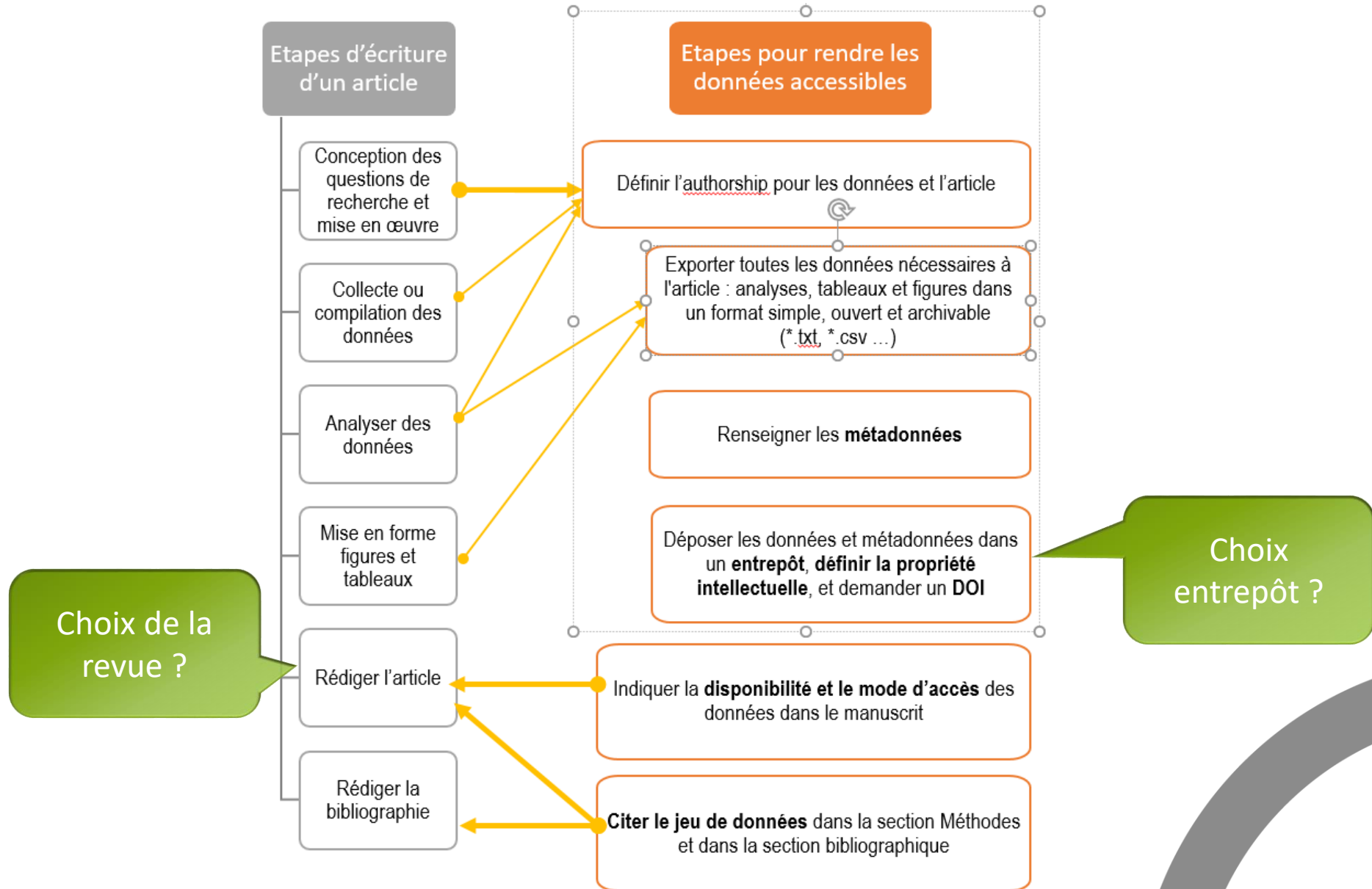


A peer-reviewed open-access journal
Biodiversity Data Journal
Making your data count! ISSN 1314-2828 (online)

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

“
The popularity of data journals and citations to these journals is growing rapidly.
”

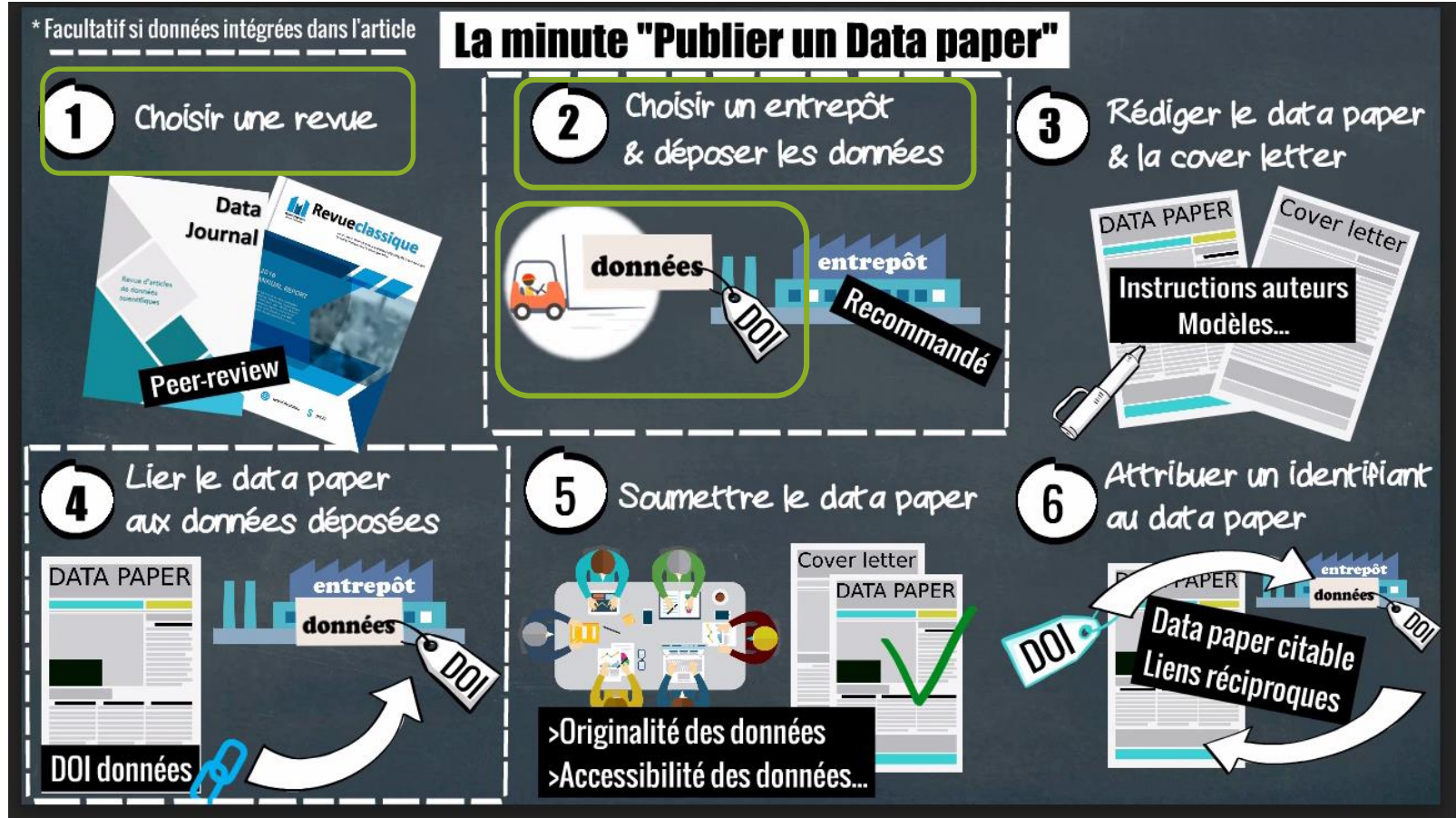
Processus de rédaction : article + données



D'après Soranno, P.A. (2019)

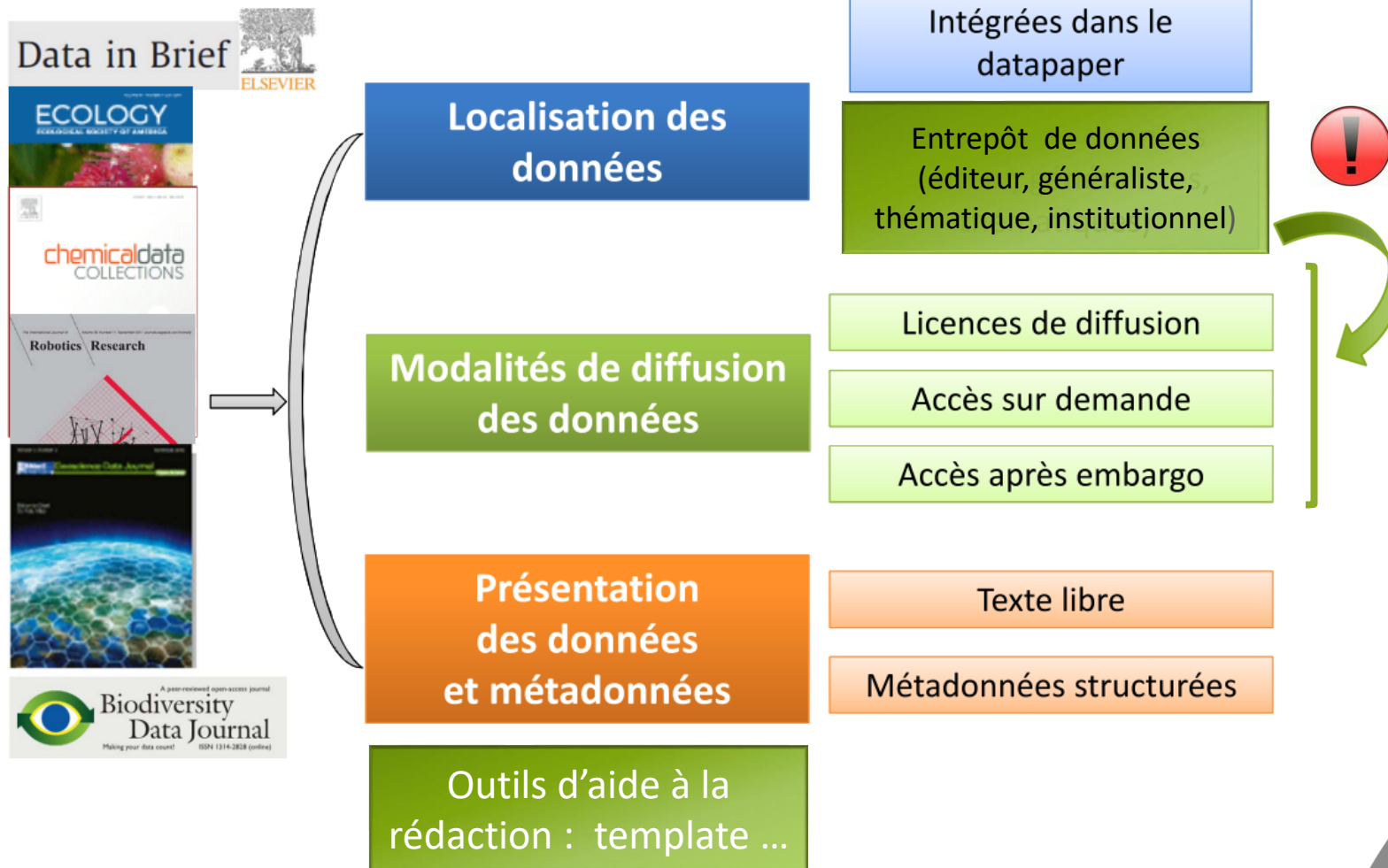
<https://aslopubs.onlinelibrary.wiley.com/doi/epdf/10.1002/lob.10303>

Processus éditorial du data paper



La minute vidéo - Réalisation Doranum & IST Inra
<http://doranum.fr/la-minute-publication-de-data-papers/>

Impact du choix de la revue





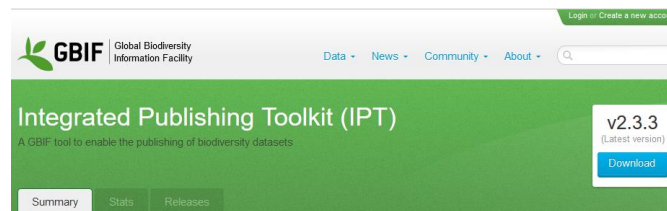
Outils d'aide à la création de Data Papers

■ Templates fournis par les éditeurs

- Data in Brief,
http://www.elsevier.com/inca/publications/misc/dib_data%20article%20template.docx
- Nature Scientific Data...
http://www.nature.com/uploads/ckeditor/attachments/2963/Scientific_Data_Templates.zip

■ Outils d'aide à la rédaction et à la publication

- Integrated Publishing Toolkit (IPT) du GBIF
<http://www.gbif.org/ipt>



- Pensoft : « Arpha writing tool » <https://arpha.pensoft.net/>



Exemple - Data Inra (Dataverse)



Génération d'un Data Paper

Cet outil vous permet de générer une ébauche de data paper (publication scientifique décrivant un jeu de données) à partir du DOI

DOI

Générer



Extraction des métadonnées et création d'une amorce de Data Paper

Ranking semantics for the choice of environment-friendly food packagings - Consumers' answers to poll for Wood packaging.

Patrice Buche [1]

Author's identifier(s) :

Corresponding author(s): Ummia (stephane.dervaux@versailles.inra.fr) www.inra.fr

Citation: Patrice Buche... (year) Ranking semantics for the choice of environment-friendly food packagings - Consumers' answers to poll for Wood packaging. . *Journal Name, Volume, (Issue number), doi of the data paper.*

Abstract

value unavailable

Kind of Data: Dataset, XML

Subject: Not specified

Keywords: Ranking

Series

- . Name:
- . Information:

Project details

Project title:

Personnel:

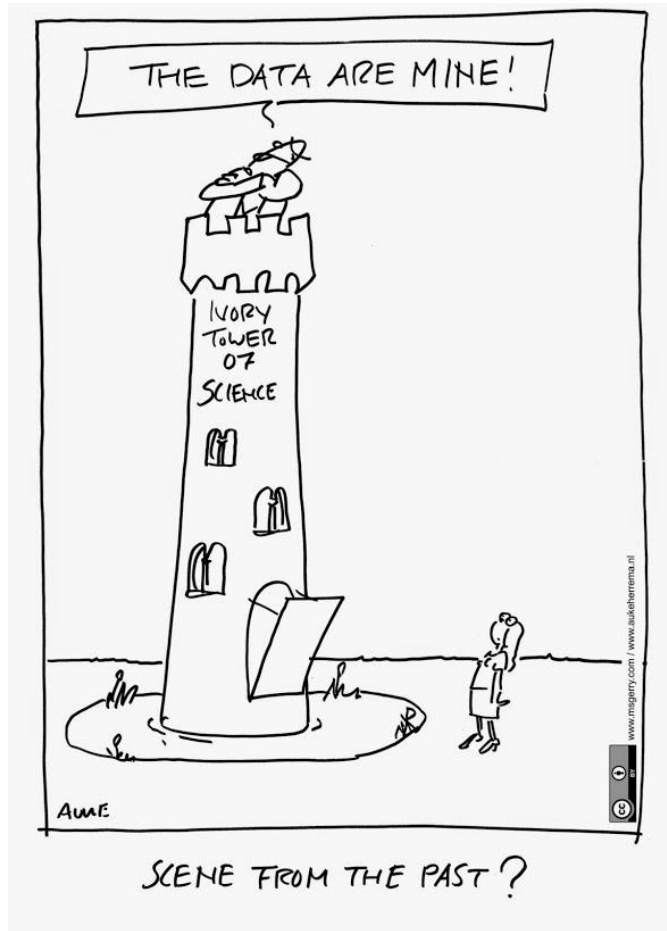
Funding:

Grant Agency:



Pour conclure

Un changement de culture...



<https://rd-alliance.org/plenary-meetings/fourth-plenary/plenary-cartoons.html>

Freins plus culturels que technologiques ?

“The main obstacles to better research data management and sharing are cultural”

“the main obstacles preventing people from using existing institutional tools and infrastructure are cultural – data management is not embedded in researchers’ everyday practice”

Teperek, M., & Dunning, A. (2018). Billet de blog LSE 14/11/2018
<https://blogs.lse.ac.uk/impactofsocialsciences/2018/11/14/the-main-obstacles-to-better-research-data-management-and-sharing-are-cultural-but-change-is-in-our-hands/>

Des dispositifs d'accompagnement innovants...

Université de Cambridge (UK) : Data Champions

Data Champions



What is a Data Champion?

Data Champions (DCs) are volunteers who advise members of the research community on proper handling of research data. In this, they promote good research data management (RDM) and support Findable, Accessible, Interoperable, and Re-usable (FAIR) research principles.

TU Delft (NL) : Data stewards



SPRINGER NATURE



« les Data stewards agissent comme des consultants et non comme des policiers (objectif = améliorer les cultures et non la conformité).
Rôle principal : discuter avec les chercheurs, premier point de contact pour toutes leurs questions liées aux données (solutions de stockage, outils de gestion des données, options d'archivage des données, plans de gestion des données, conseils sur le partage des données, budget pour la gestion des données dans les demandes de subvention, etc.)

Ce diaporama représente une compilation des diapositives issues de supports conçus par Sylvie Cocard, Esther Dzale, Odile Hologne et Dominique L'Hostis (Inra - pôle DigitalIST)

Merci de votre écoute