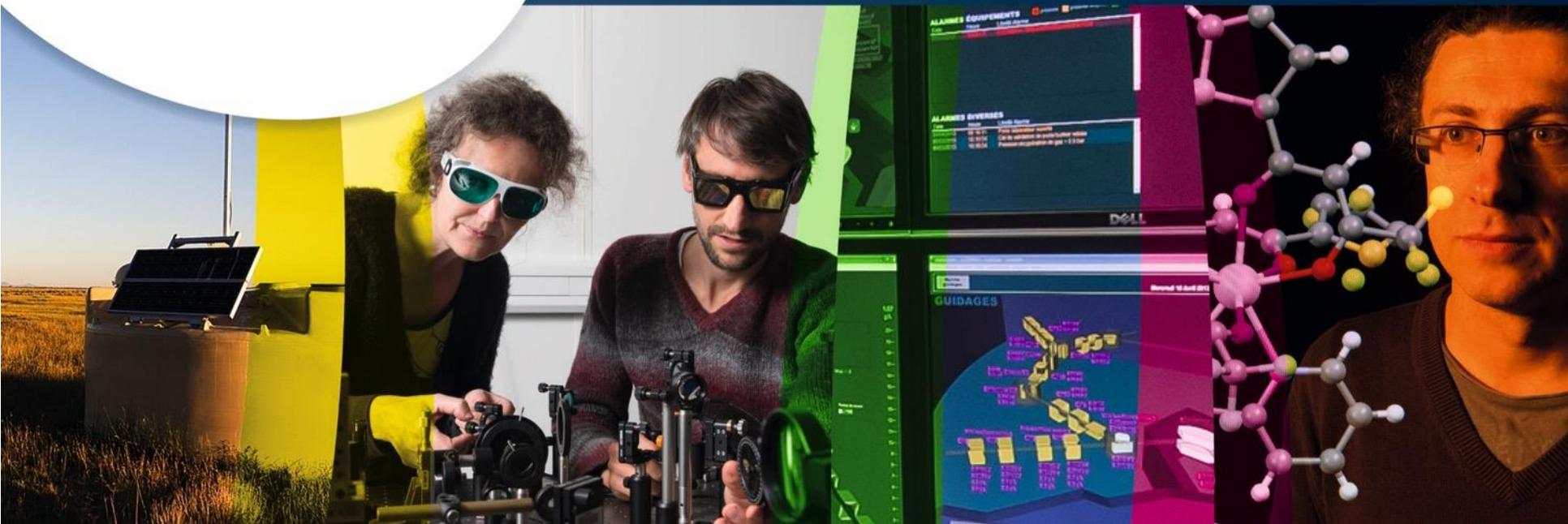




www.cnrs.fr

# Gestion et diffusion des données de la recherche

Présenté par JACQUEMOT-PERBAL M.-C. et CIOLEK-FIGIEL A.  
Inist - CNRS, Vandœuvre-lès-Nancy



Sauf mention contraire

Formation URFIST Lyon

22 mars 2016



[www.cnrs.fr](http://www.cnrs.fr)

# - 1 - INTRODUCTION





[www.cnrs.fr](http://www.cnrs.fr)

# INTRODUCTION

---

## Définitions - Contexte

# Définitions : données de recherche

## Données financées par des fonds publics

*« Enregistrements factuels (chiffres, textes, images, sons) utilisés comme source principales pour la recherche scientifique et généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche »*

(OCDE, 2007)

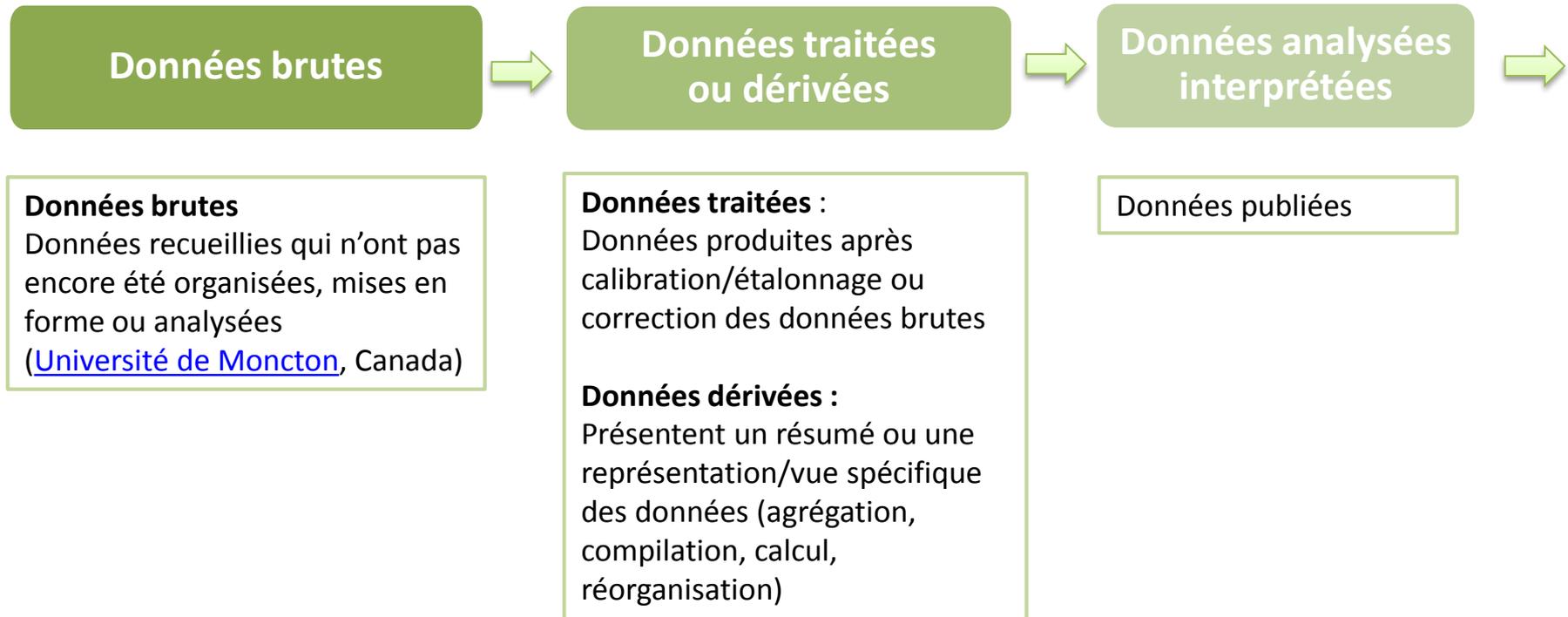
Cette définition exclut : *« carnets de laboratoire, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris) »*

## Point de vue des archivistes

« Les données de la recherche sont l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. A ce titre, elles constituent une partie des archives de la recherche. »

(AAF Section Aurore, 2014)

# Continuum données brutes ---> analysées



- **Données primaires** « données collectées par la personne qui effectue la recherche », **Données secondaires** « données collectées par une personne différente de l'utilisateur » ([Wikipédia](#), 2016)
- **Données sources** « données utilisées par les chercheurs pour leur recherche » **Données résultats** « données produites comme résultats de recherche » (Prost, 2015)

# Définitions : jeux de données

## Jeu de données (dataset)

« *Peut être défini comme l'agrégation, sous une forme lisible, de données brutes ou dérivées présentant une certaine « **unité** », rassemblées pour former un **ensemble cohérent** »*

(Gaillard R, 2014)

Collection d'informations et données connexes, généralement numériques, texte, son et/ou images, organisés pour permettre leur recherche, récupération ou traitement et réorganisation

(Traduit de CODATA-ICSTI, 2013)

# Déluge des données

Développement de nouvelles technologies

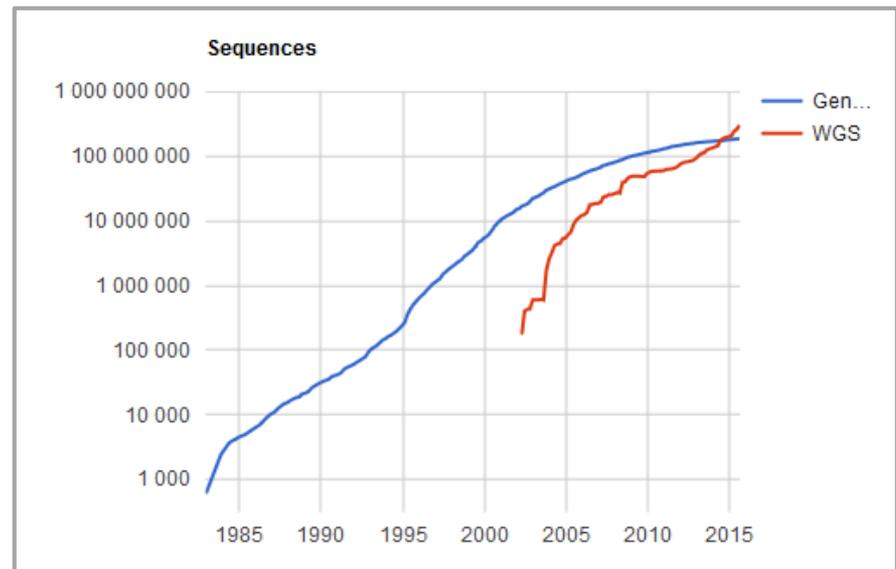


Production de grandes quantités de données numériques



Comment gérer, stocker, mettre à disposition et partager ces données ?

Technologies de séquençage d'ADN de plus en plus performantes



Croissance des séquences d'ADN dans GenBank

<http://www.ncbi.nlm.nih.gov/genbank/statistics>

# Open data - Open access

**Open data**

Données  
ouvertes

Données qu'un organisme met à la disposition de tous sous forme de fichiers numériques afin de permettre leur réutilisation. (JORF, 2014)

Mise à disposition de l'information scientifique pour un utilisateur final sans barrières financières, légales ou techniques. (Openaccess.inist.fr)

Libre accès

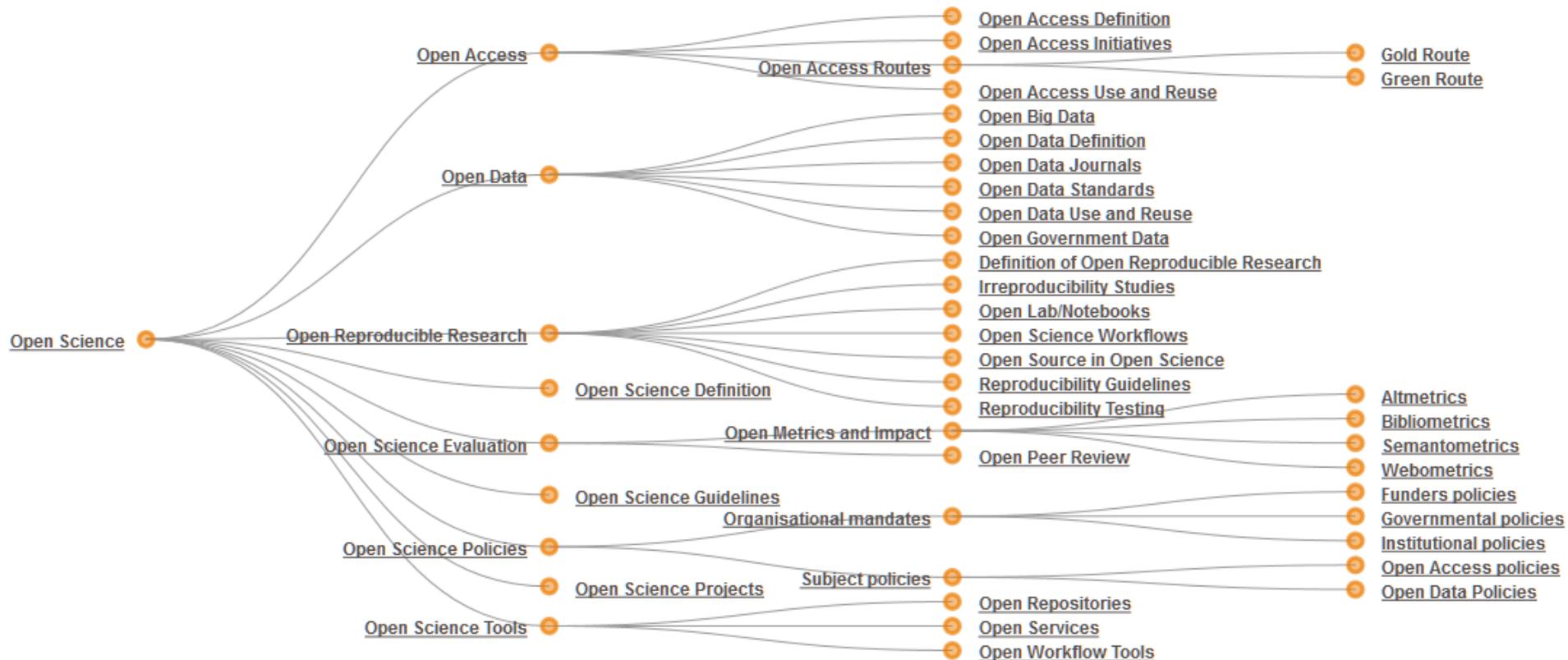
**Open access**

# Open data, Open access, Open research data



# Vers l'Open Science

Movement to make scientific research, data and dissemination accessible to all levels of an inquiring society (FOSTER)



« Tree Taxonomic » <https://www.fosteropenscience.eu/>

# Références

- ANDS (2011). Research Data Australia Guide. <http://ands.org.au/guides/research-data-australia.pdf>
- CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. Data Science Journal 12 ; p. CIDCR1-CIDCR75. doi.org/10.2481/dsj.OSOM13-043
- Gaillard, R. (2014). De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? (Mémoire de fin d'études de conservateur de bibliothèque, Enssib, Lyon, France). <http://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf>
- Prost H, Schöpfel J. Les données de la recherche en SHS. Une enquête à l'Université de Lille 3. : Rapport final. [Rapport de recherche] Université de Lille 3. 2015. <http://hal.univ-lille3.fr/hal-01198379>
- Section Aurore de l'Association des archivistes français (2014). Vers une définition des données de la recherche? <http://archivesfmsh.hypotheses.org/1209>

# Références

- OCDE (2007). Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. [www.oecd.org/fr/sti/sci-tech/38500823.pdf](http://www.oecd.org/fr/sti/sci-tech/38500823.pdf)
- The Royal Society (2012). Science as an open enterprise (p. 104). London: The Royal Society. <https://royalsociety.org/policy/projects/science-public-enterprise/Report/>
- Thessen AE, Patterson DJ (2011). Data issues in the life sciences. In: Smith V, Penev L (Eds) e-Infrastructures for data publishing in biodiversity science. ZooKeys 150: 15–51. doi: 10.3897/zookeys.150.1766
- European Commission (2013). Guidelines on Data Management in Horizon 2020. [Traduction française]  
[http://openaccess.inist.fr/IMG/pdf/14081\\_lignes\\_directrices\\_pgd\\_horizon\\_2020\\_tr\\_fr\\_versionavril2015-2.pdf](http://openaccess.inist.fr/IMG/pdf/14081_lignes_directrices_pgd_horizon_2020_tr_fr_versionavril2015-2.pdf)



[www.cnrs.fr](http://www.cnrs.fr)

# INTRODUCTION

---

## Enjeux

# Ouverture des données : bienfaits pour la société

- **Intégrité scientifique**, reproductibilité/validation, éviter la fraude
- Nouvelles approches scientifiques
  - **Science des données** (4<sup>ème</sup> paradigme)
    - Analyse big data
    - Intégration des données : linked data ou web de données et/ou moteurs de recherche
  - **Science collaborative** (nouvelles technologies ) et sans frontière
  - **Enseignement**
- **Retour sur investissement - Innovation**
- **Défis sociétaux**
  - Santé, sécurité alimentaire, énergie, environnement...
- **Transparence** : confiance des citoyens en la recherche

# Freins et obstacles à surmonter

- Juridique et éthique
  - Droit de propriété intellectuelle
  - Protection des données à caractère personnel
- Infrastructures
  - Infrastructures matérielles et outils encore insuffisants
  - Gestion des données à renforcer
- Financiers
  - Coût RDM (ressources humaines et techniques)
- Aspects socioculturels
  - Crainte de mésusage, interprétation erronée, « scooped » des données par les chercheurs
  - Absence de reconnaissance dans le processus d'évaluation des chercheurs
  - Temps nécessaire à la gestion (tâche supplémentaire)...

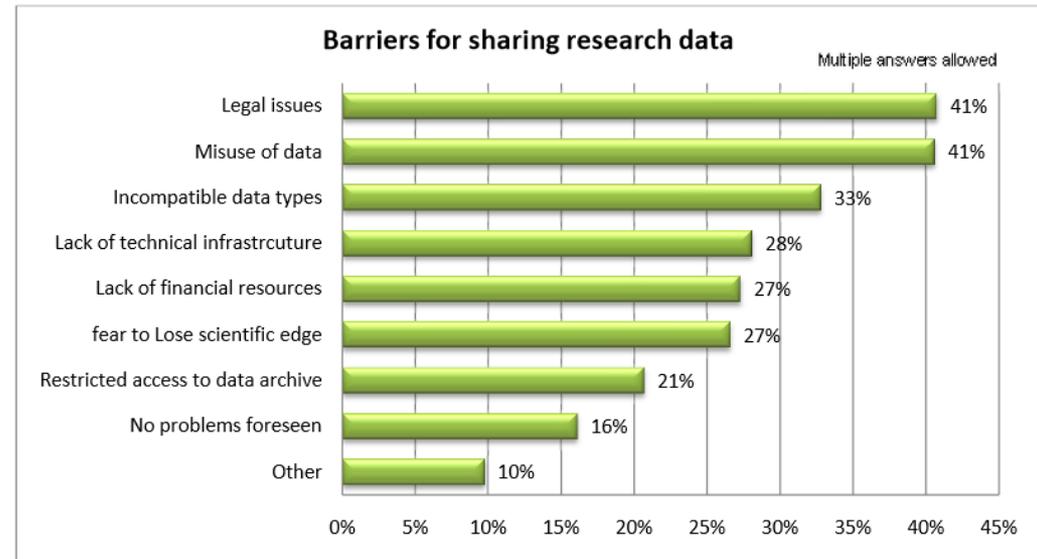


Figure 20: barriers for sharing research data, n = 1270

(Kuipers and Van der Hoeven, 2009)

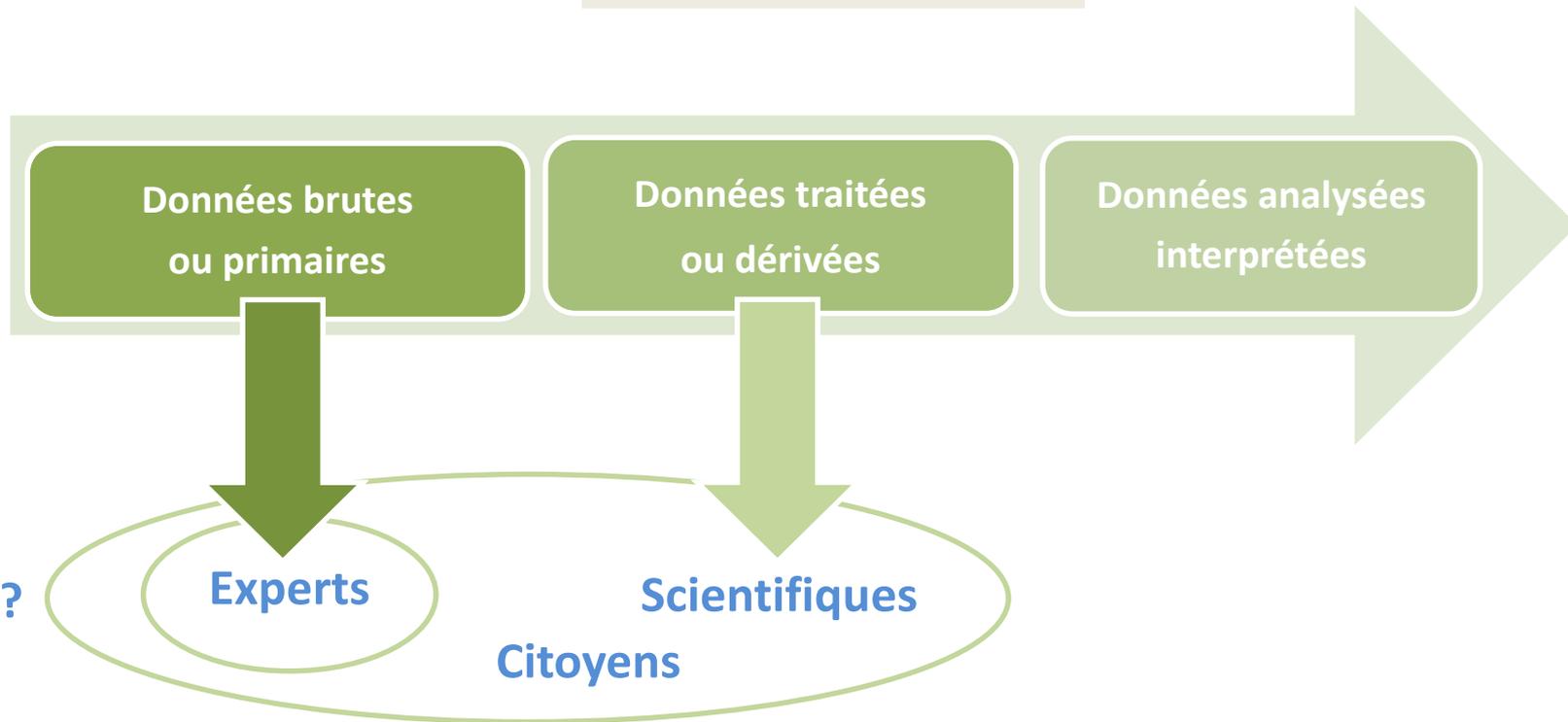
# Des principes

- Ouverture et réutilisation des données
  - Une « ouverture intelligente » (The Royal Society, 2012)
  - Quelles données?
  - Information contextuelle, provenance
  - Interopérabilité
  - Culture des données ouvertes
  - Esprit de partage
- Construire un écosystème propice au partage et à la réutilisation des données

# Quelles données partager ?

Données brutes seulement ?

« *Raw Data Now* »  
Tim Berners Lee, 2009



Utilisateurs ?

Des limites

Données à caractère personnel  
Intérêts commerciaux ...

# Big data *versus* small data

## Big data :

Données qui requièrent une quantité massive de puissance informatique pour les traiter

(The Royal Society, 2012)

Volume, Vélocité, Variété (3V)

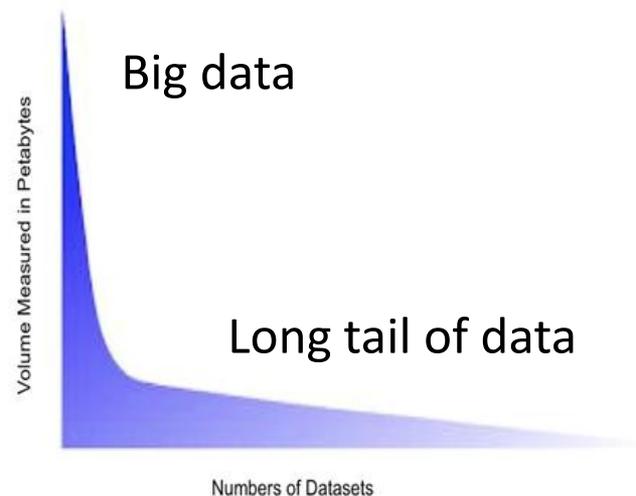
## Small data :

« Quantité de données que vous pouvez aisément stocker et utiliser sur une seule machine et plus précisément sur un seul ordinateur portable ou serveur de haute qualité »

(Traduit de Pollock R, 2013)

« L'enjeu est le même pour tous les producteurs de données (big et small) : recevoir le crédit pour la production des données et assurer la réutilisation et reproductibilité des données »

(The Royal Society, 2012)



(Image from Chuck Humphrey: [OpenAIRE-COAR Conference](#), Athens 2014)



## Forget big data, small data is the real revolution

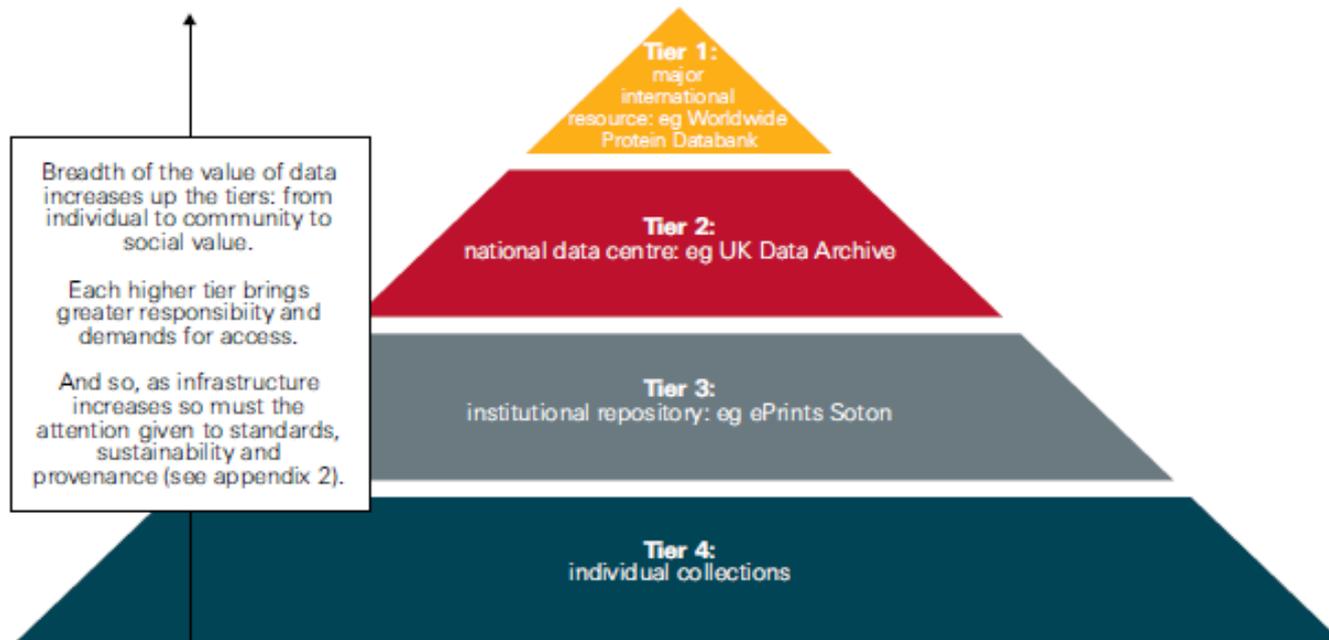
Rufus Pollock, Founder and Co-Director of the Open Knowledge Foundation, discusses why current hype around big data misses the point, and why the real revolution is small data, loosely joined

(The Guardian, 2013)

# Potentiel de réutilisation

## Box 4.1 The Data Pyramid – a hierarchy of rising value and permanence

Details of examples given in appendix 3.



(The Royal Society, 2012)

# Importance des métadonnées

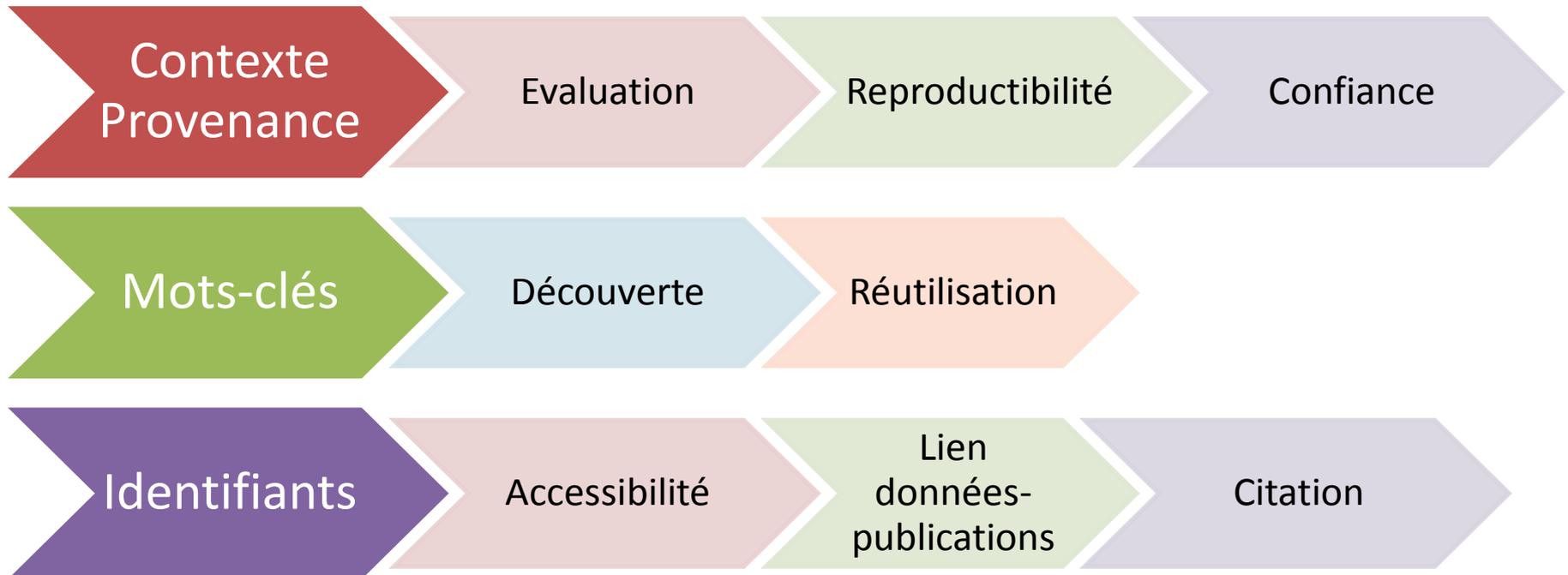
« Collections of data and annotated data (data + metadata) may have considerably enhanced value over the individual items »

(Murray-Rust P, 2009)

« **Poor quality of metadata!** »

« Qualité pauvre des métadonnées! »

Trop lu, trop entendu ...



# Cap sur l'interopérabilité

## Utiliser des langages communs

- Standards de métadonnées
- Terminologies : ontologies, thesaurus, classification...
- Identifiants pérennes
- Formats ouverts
- Outils
- Protocoles ...



# Développer une culture des données

- Bonnes pratiques de gestion et conservation  
A mettre en œuvre dès le début d'un projet !
- Responsabilisation

Démarche scientifique rigoureuse et intègre

Charte nationale de déontologie des métiers de la recherche



*« La politique d'éthique et d'intégrité de l'ANR met en œuvre les principes internationaux édictés par la Déclaration de Singapour (juillet 2010), par le « European Code of Conduct for Research Integrity » publié par l'ESF en 2010, et par le Global Research Council en mai 2012 »*

(ANR, 2014)

# Développer un esprit de partage

- **Contraintes ou incitations ?**

- Financement, réglementations



- Citation --> visibilité --> évaluation



Elliot Haney 2009, CC-BY-NC-ND

S'il n'y a que des contraintes,  
les chercheurs n'auront pas le « spirit of the law »

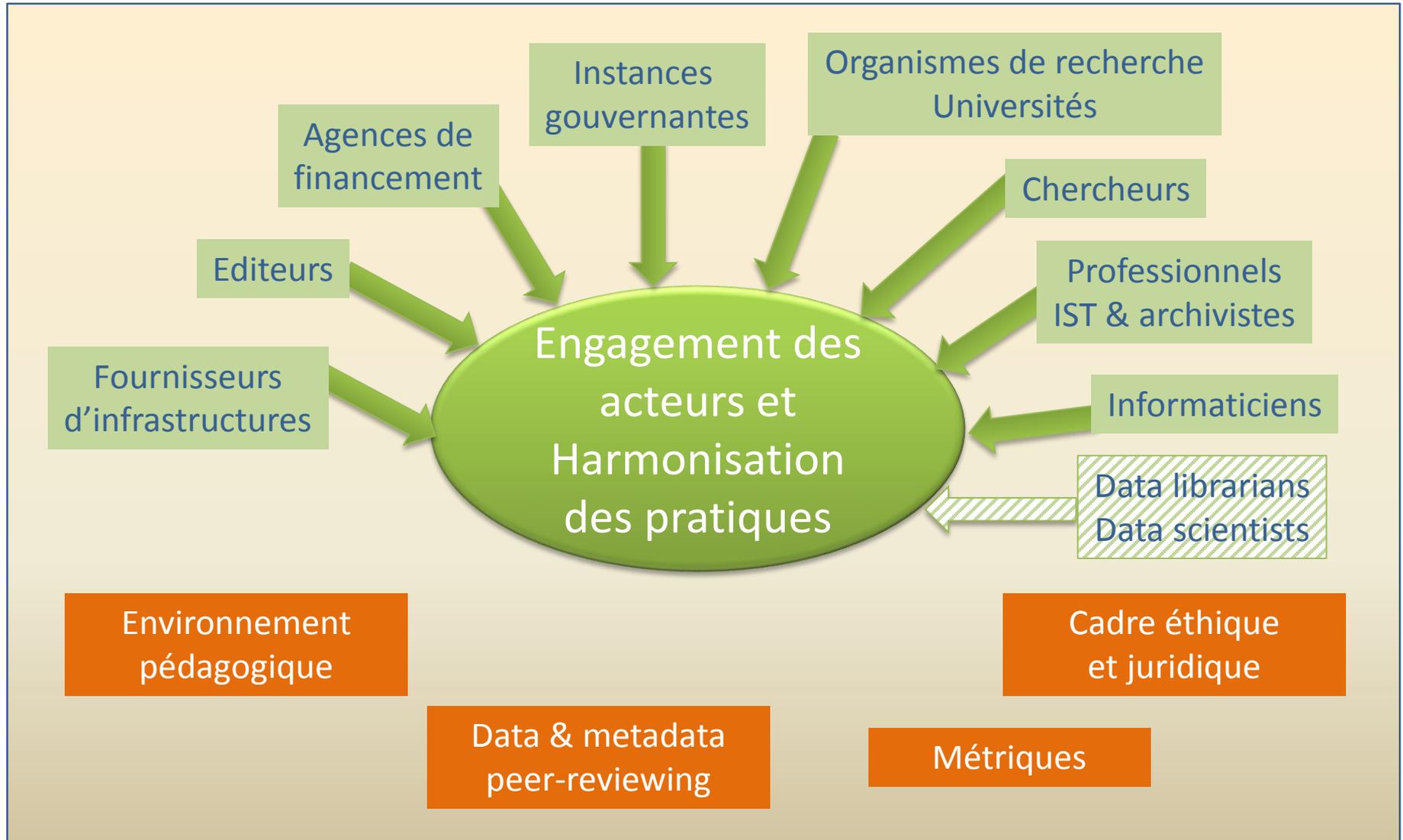
(Borgman CL, 2014)

- **Accompagnement et formation des chercheurs**

- Vers une culture du partage, développement de nouvelles connaissances et compétences

- Soutien de la part des professionnels de l'information et des informaticiens : montée en compétences

# Construire un écosystème propice au partage et à la réutilisation



# Un aperçu du paysage national

## Politiques de données



## Accompagnement

[Université Paris Diderot](#)  
[Université Nice Sophia Antipolis](#)  
[Copist](#) (HumaNum, Inist, RBDD...)

## Services

### Formation

[Université européenne de Bretagne](#)  
 [DoRANum](#)

## Feuille de route nationale « IR »



## Entrepôts/Centres de données

[SEANOE](#)



Université de Strasbourg

[IDRIS](#)

## Directive INSPIRE

[Loi pour une République numérique](#)

[Mission relative à l'intégrité scientifique](#) (MENESR)

Charte nationale de déontologie des métiers de la recherche

Politique d'éthique et d'intégrité de l'ANR

[Avis du Comets](#)

[Code de la recherche](#)

## Environnement juridique et éthique

## Infrastructures de données

### Catalogues/Portails

[Portail des données marines](#)

OSU



[Centre de calcul de l'IN2P3](#)

Mésocentres EXPLOR

**Centres de calcul** Lyon

# Références (1)

- Agence nationale de la recherche (2014). Politique en matière d'éthique et d'intégrité scientifique. <http://www.agence-nationale-recherche.fr/fileadmin/documents/2014/Politique-ethique-integrite-scientifique-aout-2014.pdf>
- Bernes-Lee T (2009). The next web. In TED Conference, Feb 2009. [http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web)
- Borgman CL (2014). Keynote: Data, Data, Everywhere, Nor Any Drop to Drink. In : RDA Fourth Plenary Meeting 22 - 24 September, Amsterdam, The Netherlands. <https://collegerama.tudelft.nl/Mediasite/Play/55eb502f04aa4124ac94b75584c62abc1d?catalog=3984a02f-bf33-4c70-a080-94a04d3e8112>
- Charte nationale de déontologie des métiers de la recherche (2015). [http://www.cnrs.fr/comets/IMG/pdf/charte\\_nationale\\_deontologie\\_signe\\_e\\_janvier2015.pdf](http://www.cnrs.fr/comets/IMG/pdf/charte_nationale_deontologie_signe_e_janvier2015.pdf)
- Confederation of Open Access Repositories (2015). COAR Roadmap. Future Directions for Repository Interoperability. [https://www.coar-repositories.org/files/Roadmap\\_final\\_formatted\\_20150203.pdf](https://www.coar-repositories.org/files/Roadmap_final_formatted_20150203.pdf)
- CODATA-ICSTI Task Group on Data Citation Standards and Practices (2013). Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data. Data Science Journal 12 ; p. CIDCR1-CIDCR75. doi.org/10.2481/dsj.OSOM13-043

# Références (2)

- Humphrey C (2014). Long tail of Science - Making the link from long tail to libraries. In : OpenAIRE-COAR Conference “ Open Access Movement to Reality - Putting the pieces together” May 21-22, Athens , Greece. [https://www.coar-repositories.org/files/2\\_OpenAIRE-COAR\\_session2\\_ChuckHumphrey.pdf](https://www.coar-repositories.org/files/2_OpenAIRE-COAR_session2_ChuckHumphrey.pdf)
- Kuipers T and Van der Hoeven J (2009). Insight into digital preservation of research output in Europe. Survey Report PARSE. Insight: INSIGHT into issues of Permanent Access to the Records of Science in Europe 1-83. <http://www.parse-insight.eu/publications/>
- Murray-Rust P (2009). BioIT 2009 - What is data? -1. [Billet de blog]. <http://blogs.ch.cam.ac.uk/pmr/2009/04/29/bioit-2009-what-is-data-1/>
- The Guardian (2013). Forget big data, small data is the real revolution. <http://www.theguardian.com/news/datablog/2013/apr/25/forget-big-data-small-data-revolution>
- The Royal Society (2012). Science as an open enterprise (p. 104).London: The Royal Society. <https://royalsociety.org/policy/projects/science-public-enterprise/Report/>



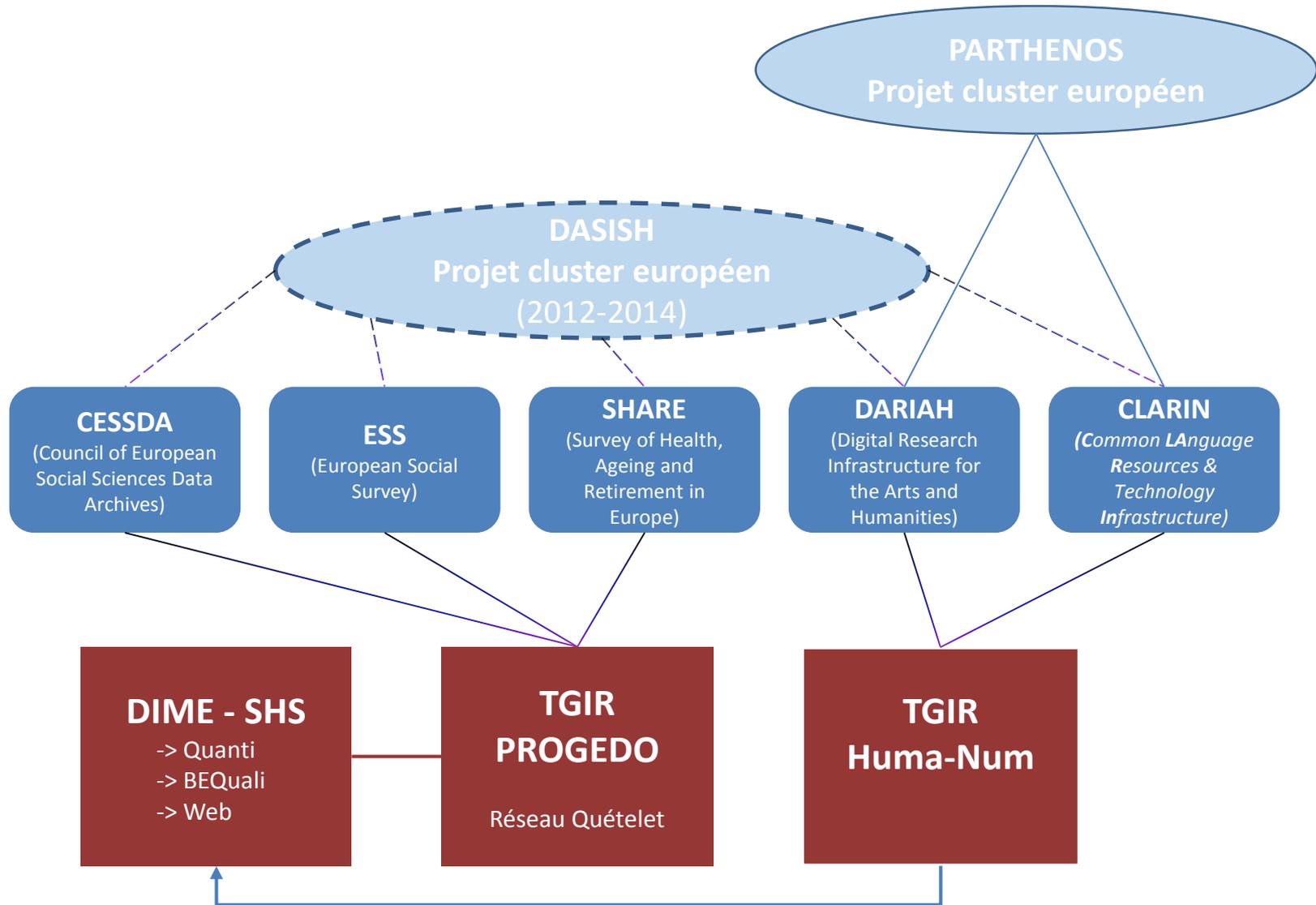
[www.cnrs.fr](http://www.cnrs.fr)

# INTRODUCTION

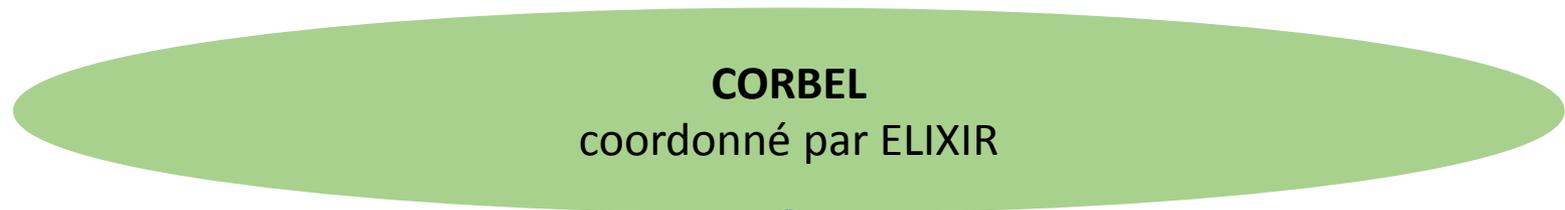
---

## Infrastructures

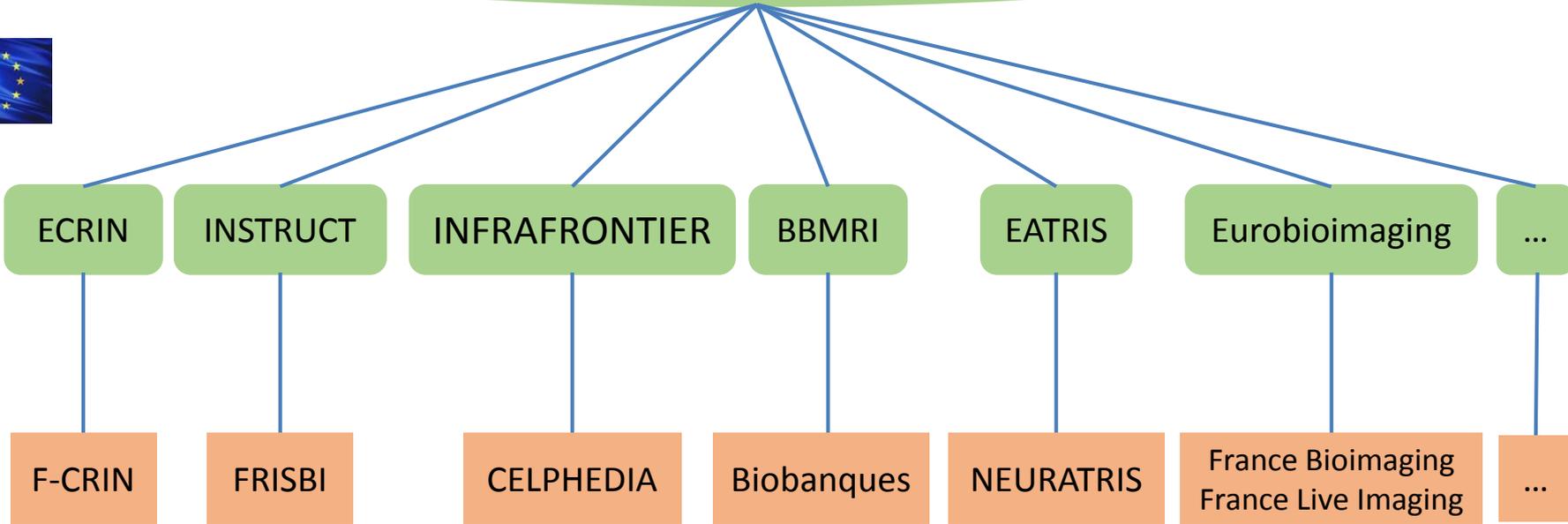
# Coordination des infrastructures en SHS



# Coordination des infrastructures en SDV

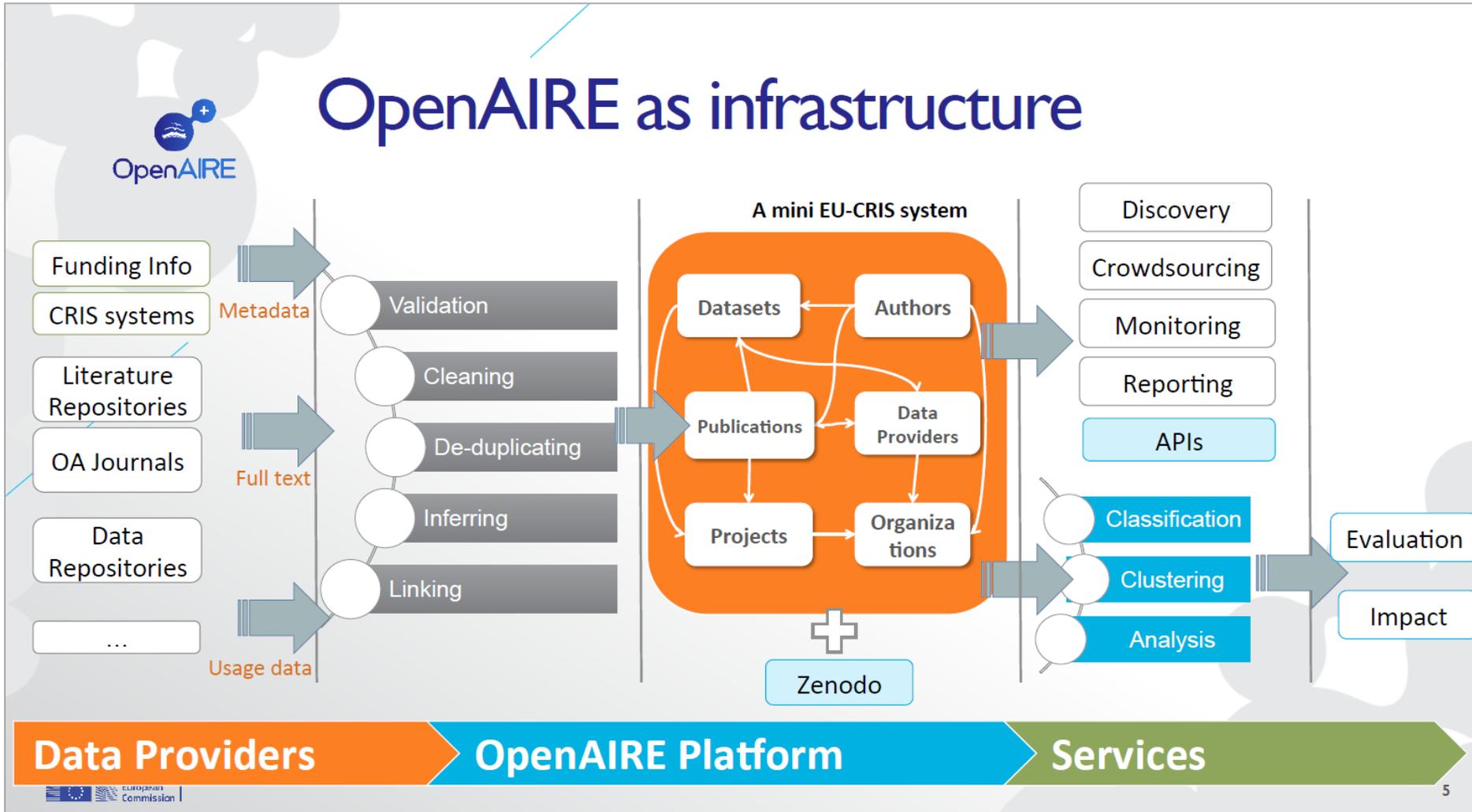


11  
ESFRIs



Nœuds ou coordinateurs français

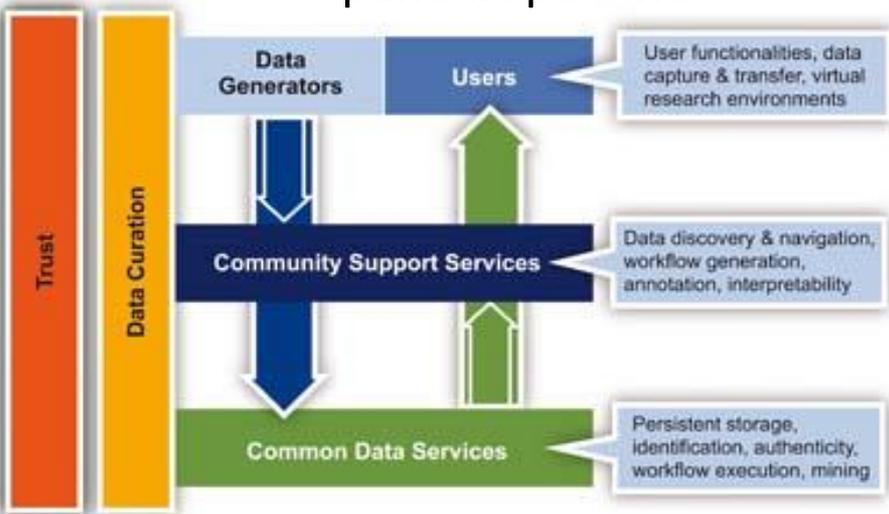
# OpenAIRE



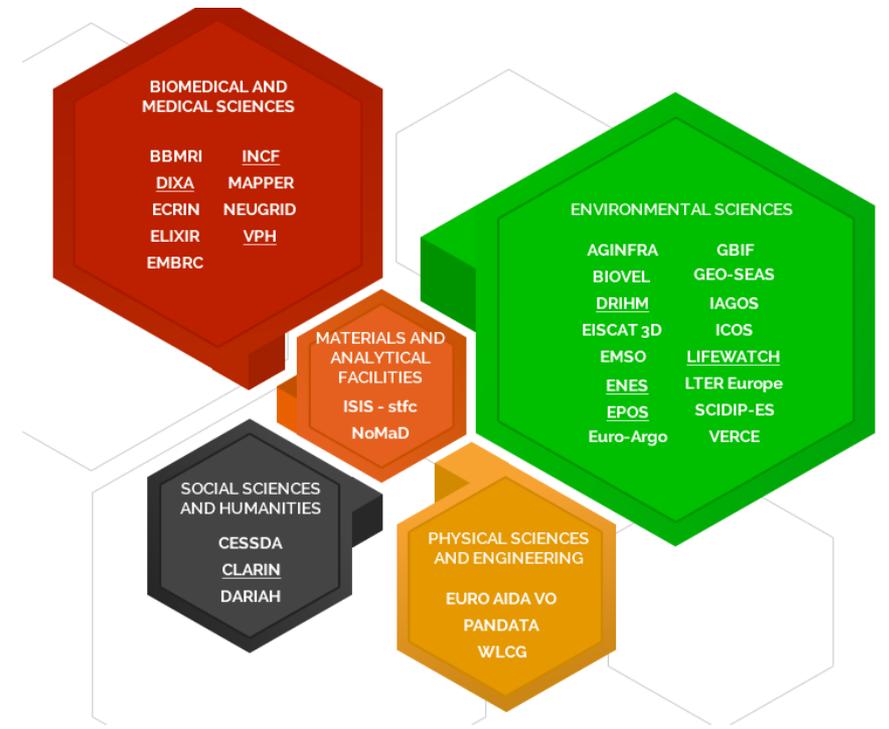
(Hermans, 2016)

# EUDAT

Une infrastructure collaborative paneuropéenne



[www.eudat.eu](http://www.eudat.eu)

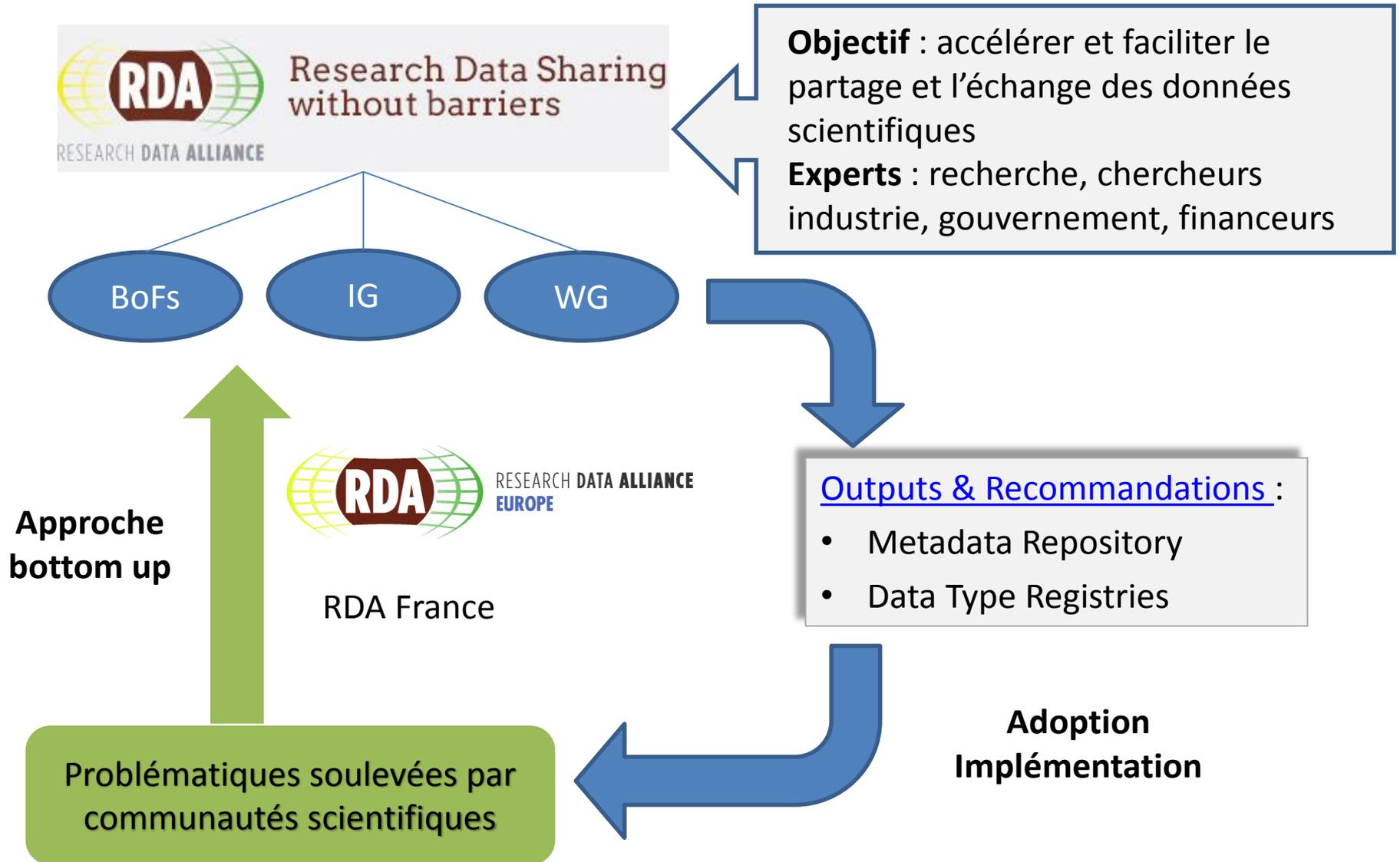


- B2DROP**  
Sync and Exchange Research Data
- B2SHARE**  
Store and Share Research Data
- B2SAFE**  
Replicate Research Data Safely
- B2STAGE**  
Get Data to Computation
- B2FIND**  
Find Research Data

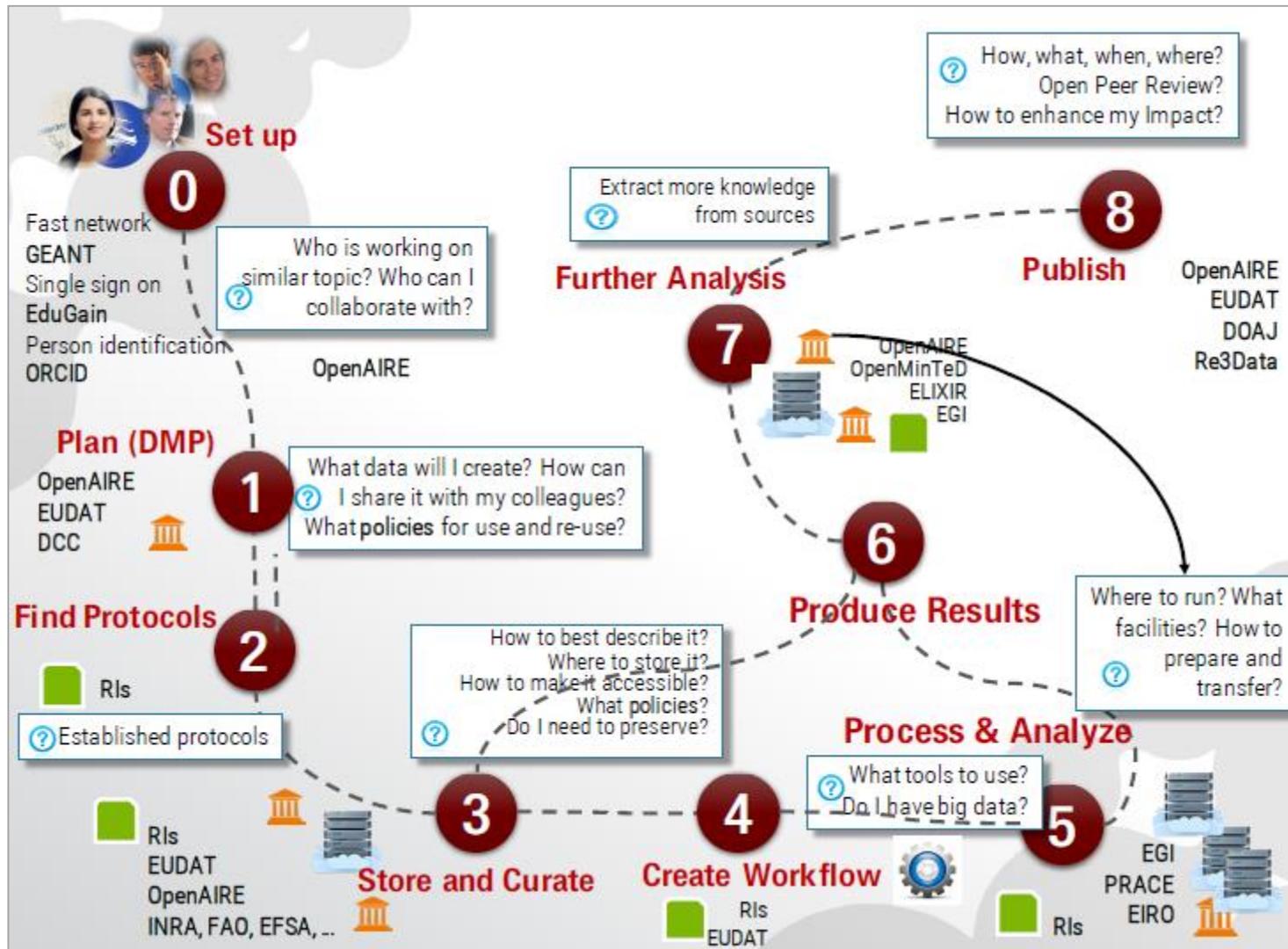
Services

En interaction avec des communautés

# Coordination internationale



# European Open Science Cloud



(Manola, 2016)

# Références

- EUDAT. EUDAT overview. <http://eudat.eu/eudat-overview>
- Hermans E (2016). OpenAire Sessions - An Open Knowledge & Research Information Infrastructure. Open Belgium 2016.  
<http://fr.slideshare.net/OpenKnowledgeBE/openaire-sessions-an-open-knowledge-research-information-infrastructure>
- Manola N (2016). e-Infrastructures for Open Science. European Open Science Cloud for Research (EOSC) Workshop, Rome, Italy.  
[https://www.eudat.eu/sites/default/files/Integration%40EUDAT OSC Feb 2016.pdf](https://www.eudat.eu/sites/default/files/Integration%40EUDAT%20OSC%20Feb%202016.pdf)
- OpenAIRE. Open Access Infrastructure for Research in Europe.  
<https://www.openaire.eu/>
- RDA Europe. <https://europe.rd-alliance.org/>
- RDA. Research Data Alliance. <https://rd-alliance.org/>

## **Pour en savoir plus :**

- Site d'information sur les données de la recherche.  
<http://www.donneesdelarecherche.fr/>



[www.cnrs.fr](http://www.cnrs.fr)

## - 2 - Méthodologie





# MÉTHODOLOGIE

---

## Cycle de vie des données

# Cycle de vie des données

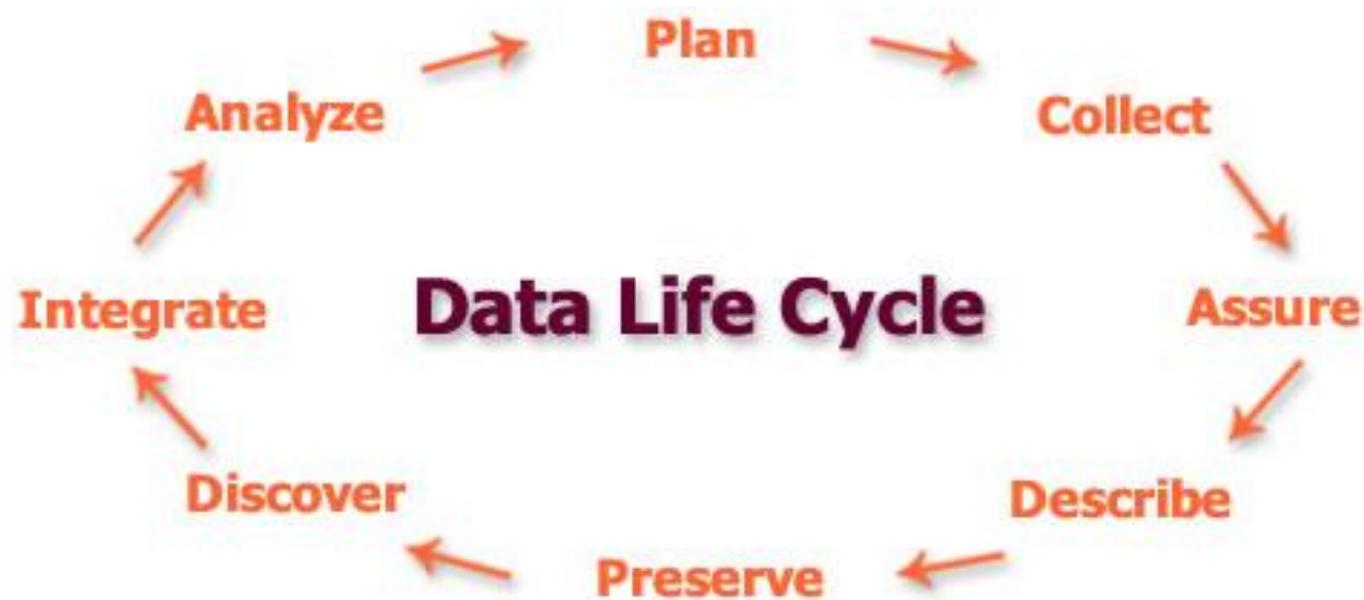
- Research data lifecycle
- C'est l'ensemble des étapes de gestion, de conservation, de diffusion et de réutilisation des données scientifiques liées aux activités de recherche (Cirad, 2016)
- Nombreuses représentations du cycle de vie des données (CEOS - Working Group on Information Systems and Services, 2012 ; Ball A, 2012)



# Quelques exemples

## DataONE

<https://www.dataone.org/best-practices>

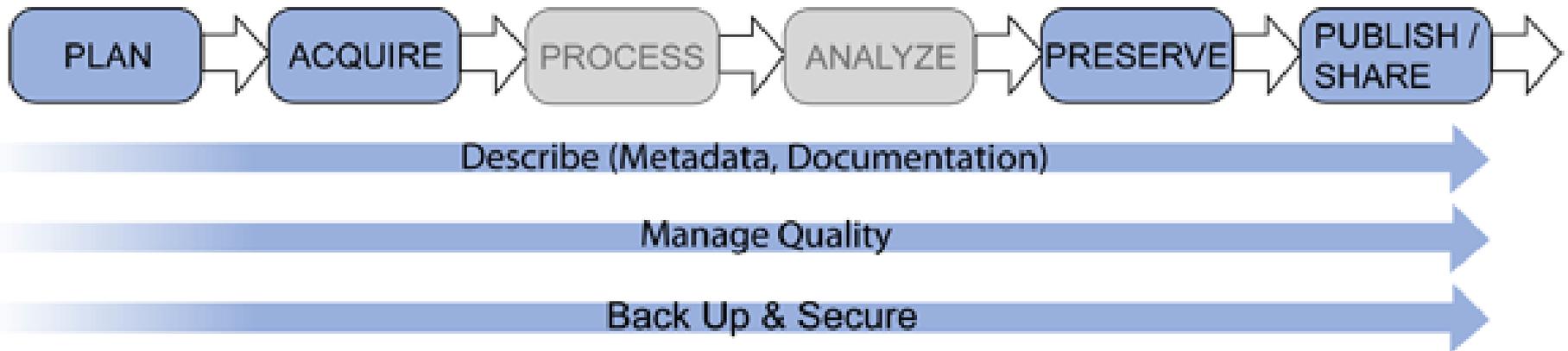


DataONE Data Observation Network for Earth's (USA)

# Quelques exemples

## USGS Data Lifecycle Diagram

<http://www.usgs.gov/datamanagement/why-dm/lifecycleoverview.php>

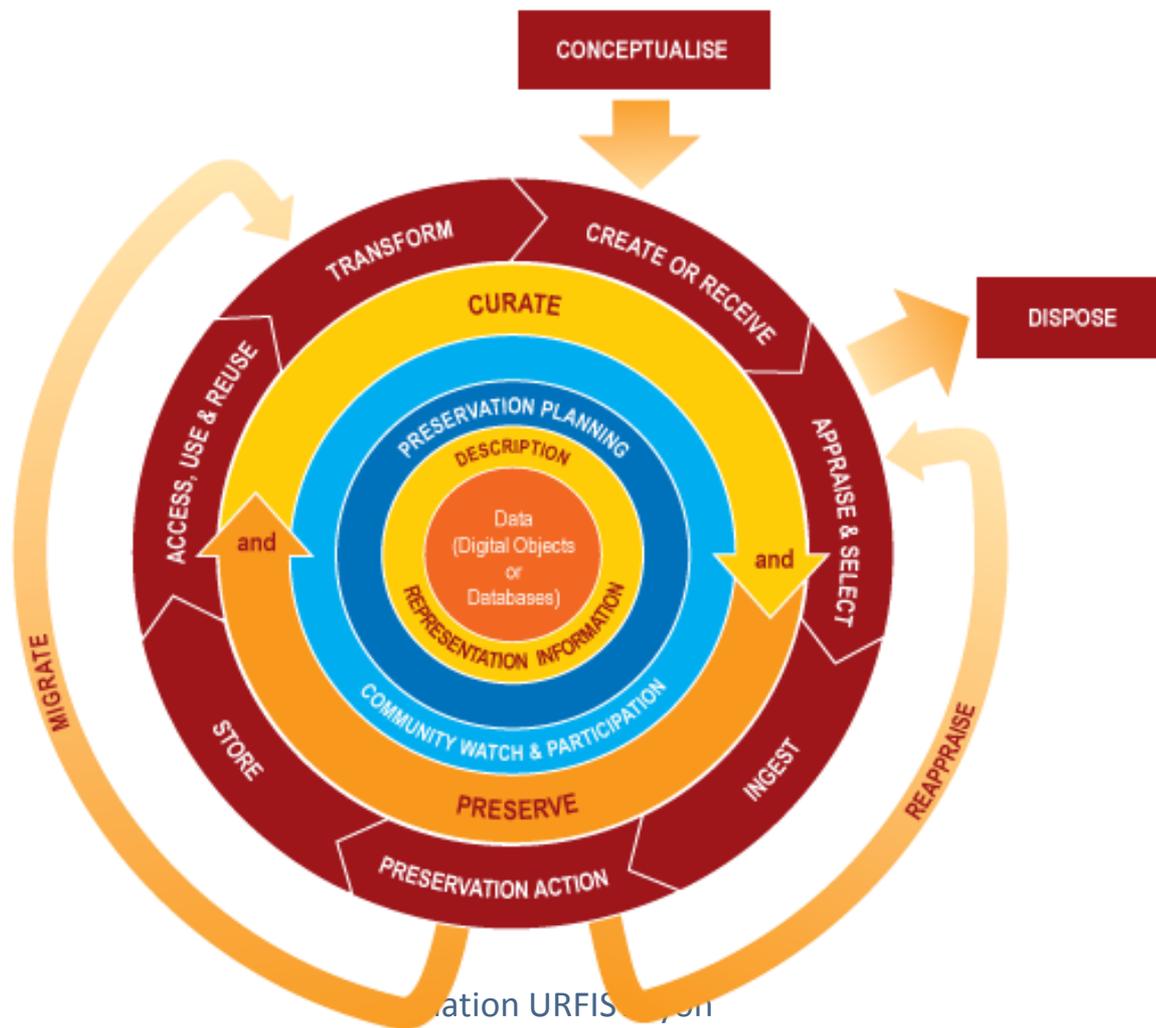


USGS : US Geological Survey

# Quelques exemples

## DCC Curation Lifecycle Model

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>



# Quelques exemples



(Jones S, 2015)

- **Créer ou collecter** des données (Create)
- **Documenter** les données (Document)
- **Accéder** aux données et utilisation (Use)
- **Stocker et sauvegarder** des données (Store)
- **Partager et publier** les données (Share)
- **Conserver** les données (Preserve)

Les bonnes pratiques de gestion des données s'appliquent à chaque étape du cycle de vie

# Références

- Ball, A., 2012. Review of Data Management Lifecycle Models. Other. Bath, UK: University of Bath. <http://opus.bath.ac.uk/28587/>
- CEOS - Working Group on Information Systems and Services (2012). Data Life Cycle Models and Concepts. CEOS Version 1.2. <http://wgiss.ceos.org/dsig/whitepapers/Data%20Lifecycle%20Models%20and%20Concepts%20v12.docx>
- DataONE. Best Practices. <https://www.dataone.org/best-practices>
- Digital Curation Centre. DCC Curation Lifecycle Model. <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
- Cirad. Le cycle de vie des données de la recherche – Poster. <http://coop-ist.cirad.fr/actualites/integrer-la-gestion-des-donnees-aux-activites-de-recherche-poster>
- Jones S (2015). Managing research data and Horizon 2020. In: ConsorcioMadróno conference on Data Management Plans and Horizon 2020, ETSI Industriales, 25th February, Madrid, Spain. [http://www.consorcioamadrono.es/noticias\\_eventos/2015/JornadaPGD/sarah.pdf](http://www.consorcioamadrono.es/noticias_eventos/2015/JornadaPGD/sarah.pdf)
- USGS. USGS Data Lifecycle Diagram. <http://www.usgs.gov/datamanagement/why-dm/lifecycleoverview.php>



# MÉTHODOLOGIE

---

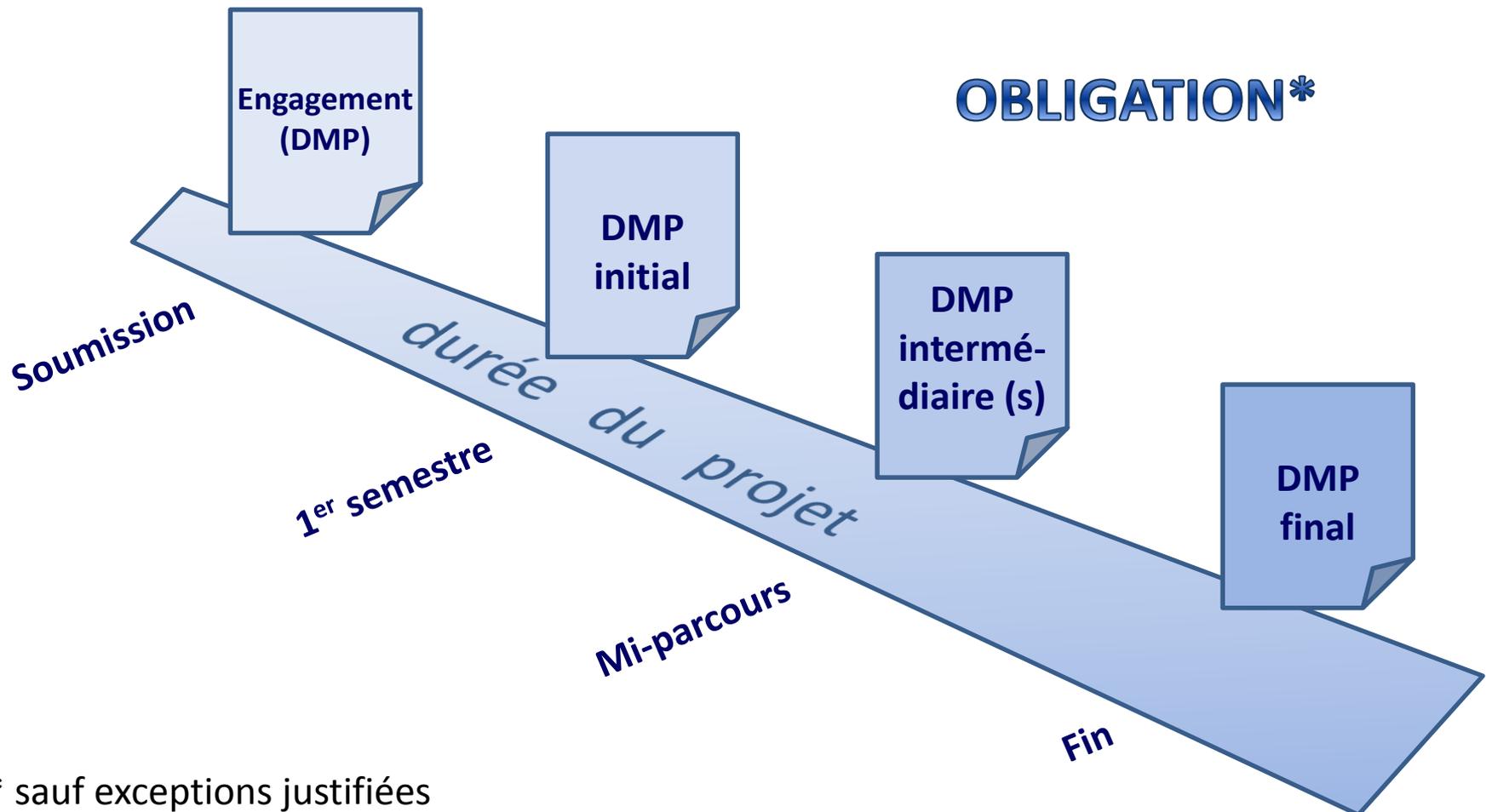
## Plan de gestion des données

# Définition

- Point commun des cycles de vie  
Plan de gestion des données = Data Management Plan (DMP)
- Document rédigé **au commencement** d'un projet de recherche et qui définit ce que les chercheurs feront de leurs données pendant et après le projet, explicitant notamment la mise à disposition des données
- Doit être continuellement mis à jour tout au long du projet
- Incitation voire obligation par les organismes de recherche et agence de financement de la recherche

# Horizon 2020 – Commission européenne

## Projet pilote « Open Research Data »



\* sauf exceptions justifiées

# Éléments de contenu

- Qui seront les personnes responsables de chaque étape de la gestion ?
- Quelle sera la politique appliquée aux données : celle des agences de financement, celle de l'institution... ?
- Quels types de données seront collectés ou générés au cours du projet ?
- Comment seront organisés les données et les fichiers ?
- Comment seront décrites les données (documentation et standards de métadonnées) ?
- Comment et où seront stockées, sauvegardées et sécurisées les données ?
- Comment seront partagées les données ? Propriété intellectuelle ? Licence de réutilisation ?
- Comment seront préservées ces données à long terme ?
- Quels seront le coût et les ressources nécessaires à la gestion et au partage des données ?

(DCC, 2013 ; European Commission, 2015 ; ICPSR)

# Modèles & Outils

- France
  - [Université Paris Diderot – Paris Descartes](#)  
Guide de rédaction pour « Réaliser un plan de gestion de données »
  - [Cirad](#) Découvrir des plans de gestion des données de la recherche, en 4 points. (Deboin, 2014)
  - [Inist-CNRS](#) Cinq tutoriels sur la rédaction d'un DMP, dans le cadre H2020



S'initier au concept et à la gestion des données de la recherche <http://coop-ist.cirad.fr>  
Public cible : chercheurs, professionnels de l'IST

**Découvrir des plans de gestion de données de la recherche en 4 points**

1. Pourquoi gérer les données de la recherche ?
2. Définissez un plan de gestion de données (PGD), ou *Data Management Plan* (DMP)
3. Exemple de trame d'un plan de gestion de données (PGD)
4. Découvrez des modèles en anglais de plans de gestion de données

DMP - Présentation générale Plan | Narration | Glossaire | Ressources



Une série de tutoriels sur le libre accès aux résultats de recherche dans le cadre d'Horizon 2020

Data Management Plan  
Plan de gestion des données

**Tutoriel**  
**<< Présentation générale >>**

<http://www.inist.fr/>



SUIVANT >

# Modèles & Outils

- International
  - [Digital Curation Center](#) Checklist for a Data Management Plan (UK)
  - Data Management Plan Format [Wageningen UR](#) (Pays-Bas)
  - Template [3TU. Datacentrum](#) (Pays-bas)
  - Monash University - [Research Data Planning Checklist](#) (Australie)
  - [ICPSR](#) – Inter-university Consortium for Political and Social Research (USA)
- Institutions - Universités
  - Australie :  
[https://docs.google.com/spreadsheets/d/1WOR5RAS\\_IMx95eeoJxqN2spGX6II-6G9y8cRaitKL3Q/edit?pli=1#gid=0](https://docs.google.com/spreadsheets/d/1WOR5RAS_IMx95eeoJxqN2spGX6II-6G9y8cRaitKL3Q/edit?pli=1#gid=0)
- Agences de financement
  - [Agences de financement du Royaume-Uni](#)
  - [NSF](#) National Science Foundation (USA)
  - [Commission européenne](#) (Horizon 2020)

# Projet pilote « Open Data Research »

## • Référence et nom du jeu de données

Identifiant du jeu de données qui va être produit.

## • Description du jeu de données

Celle-ci décrit les données qui seront générées ou collectées, leur origine (s'il y a collecte), leur nature et leur échelle, à qui elles pourraient être utiles et si elles viennent en appui d'une publication scientifique, mais aussi les informations sur l'existence (ou la non-existence) de données similaires et leurs possibles intégrations et réutilisations.

## • Normes et métadonnées

Indication des normes existantes relatives à la discipline. À défaut, un résumé expliquant quelles métadonnées seront créées et comment.

## • Partage des données

Description des modalités de partage, y compris les procédures d'accès, les périodes d'embargo (le cas échéant), exposé succinct des dispositifs techniques de diffusion et logiciels nécessaires ainsi que les autres outils permettant la réutilisation de données, et indications sur le caractère librement accessible des données ou si celles-ci seront limitées à des groupes spécifiques. Identification de l'entrepôt dans lequel les données seront stockées, s'il existe déjà et est identifié, en indiquant son type (institutionnel, ou de référence dans la discipline considérée, etc.).

Les raisons d'une éventuelle impossibilité de partage d'un jeu de données devront être indiquées (par ex. motifs de nature éthique, liés à la protection des données personnelles, de la propriété intellectuelle, commerciale, à la protection de la vie privée, et de sécurité, etc.).

## • Archivage et conservation (y compris stockage et sauvegarde des données)

Description des procédures mises en place pour une conservation à long terme des données. Indication de la durée de conservation souhaitable des données, leur volume approximatif, les coûts associés et la façon donc ces coûts seront couverts.



Pour chaque  
jeu de données

**Annexe 1** : modèle de plan de gestion de données (PGD) - (European Commission, 2015, v2.0 [Traduction française])

# Modèles - Outils



DMPonline

Digital Curation Centre - UK (2010)

<https://dmponline.dcc.ac.uk/>



DMPTool

University of California Curation Centre of CDL - US (2011)

<https://dmptool.org/>

# Exemples

- Sciences de la vie et de la terre
  - L. Vermeulen, Environmental Systems Analysis Group, Wageningen University  
[www.wageningenur.nl/web/file?uuid=4c3b23ee-02ee-4819-b555-988c23deca44&owner=7e5add26-d3a7-4118-aed0-a55dbb7d892e](http://www.wageningenur.nl/web/file?uuid=4c3b23ee-02ee-4819-b555-988c23deca44&owner=7e5add26-d3a7-4118-aed0-a55dbb7d892e)
  - B. Ramírez, Earth System Research group, Wageningen University  
<https://www.wageningenur.nl/web/file?uuid=0346c29a-ea91-4881-b612-211989044ca3&owner=7e5add26-d3a7-4118-aed0-a55dbb7d892e>
- Sciences humaines et sociales
  - University of Virginia's Odum Institute for Research in Social Science  
<http://www.odum.unc.edu/odum/contentSubpage.jsp?nodeid=570>
  - <https://docs.google.com/file/d/0B9Ccd1URnalvYmEyNGU4MDQtZDk4MS00MDYyLWFIMDEtYWYwMTM2NzU2MzUx/edit?pli=1>

# Exemples

- DataONE (Data Observation Network for Earth) <https://www.dataone.org/data-management-planning>
- DCC <http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>
- UC San Diego (USA) <http://libraries.ucsd.edu/services/data-curation/data-management/dmp-samples.html> (NSF)
- Purdue University (USA) <https://purr.purdue.edu/dmp/dmpexamples>

# Coûts associés

- Anticiper le coût
  - Ressources humaines
  - Logiciels/équipement pour préparation des données et documentation
  - Frais pour archivage des données



[Pixabay](#), CC0

- Comment financer ces coûts ?

**Exemple :** Projet pilote « Open Research Data » Horizon 2020 (2016, v2.1)

- Eligible au remboursement pendant la durée du projet pilote
- Lignes directrices pour la gestion des données dans Horizon 2020 (Annexe 1)  
« *Archivage et conservation (y compris stockage et sauvegarde des données)*  
*[...] Indication de la durée de conservation souhaitable des données, leur volume approximatif, les coûts associés et la façon donc ces coûts seront couverts* »

# Coûts associés

- Outil de calcul des coûts associés à la gestion des données en sciences sociales (UK Data Archives, 2013)

ACTIVITY	COMMENTS AND SUGGESTIONS	√	COST
<b>Data description</b> <ul style="list-style-type: none"> <li>• Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions, etc.?</li> <li>• Are labels consistent?</li> <li>• Do textual data like interview transcripts need description of context, e.g. included as a heading page?</li> </ul>	<ul style="list-style-type: none"> <li>• if data description is carried out as part of data creation, data input or data transcription – low or no additional cost</li> <li>• if needed to be added afterwards – higher cost</li> <li>• codebooks for datasets can often be easily exported from software packages</li> </ul>		
<b>Data cleaning</b>			

- Extrait de “Data in Real Life: A DMP Example” (DataOne, 2012)

*“A tablet computer will be used for data collection in the field, which will cost approximately \$500. Data documentation and preparation for reuse and storage will require approximately one month of salary for one technician. The technician will be responsible for data entry, quality control and assurance, and metadata generation. These costs are included in the budget in lines 12-16”*

# Références

- Cartier A, Moysan M, Reymonet N (2015). Réaliser un plan de gestion de données. Université Paris Diderot et Université Paris Descartes, 30 p. [http://www.univ-paris-diderot.fr/DocumentsFCK/recherche/Realiser\\_un\\_DMP\\_V1.pdf](http://www.univ-paris-diderot.fr/DocumentsFCK/recherche/Realiser_un_DMP_V1.pdf)
- DataONE (2012). DataONE Education Module: Data Management Planning. [http://www.dataone.org/sites/all/documents/L03\\_DataManagementPlanning.pptx](http://www.dataone.org/sites/all/documents/L03_DataManagementPlanning.pptx)
- Digital Curation Centre (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/data-management-plans>
- Deboin MC (2014). Découvrir des plans de gestion des données de la recherche, en 4 points. Montpellier (FRA): CIRAD, 6 p. <http://coop-ist.cirad.fr/content/download/5435/40362/version/4/file/CoopIST-plan-gestion-donnees-recherche-20140717.pdf>
- European Commission (2015). Guidelines on Data Management in Horizon 2020. v2.0 [Traduction française]. [http://www.donneesdelarecherche.fr/IMG/pdf/lignes\\_directrices\\_gestion\\_des\\_donnees\\_horizon\\_2020\\_tr\\_fr.pdf](http://www.donneesdelarecherche.fr/IMG/pdf/lignes_directrices_gestion_des_donnees_horizon_2020_tr_fr.pdf)
- ICPSR. Guidelines for Effective Data Management Plans. <http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf>
- Inist-CNRS (2015). Tutoriels « Le libre accès aux résultats de la recherche dans le cadre d'Horizon 2020 ». <http://www.inist.fr/?-Tutoriels-multimedias-H2020->
- Romier G (2014). Une préoccupation partagée : plan de gestion des données et projets Horizon 2020 (France Grilles). In : *Journée de sensibilisation à la sécurisation et à la pérennisation des données* [en ligne]. Institut d'Astrophysique de Paris. [Consulté le 29 mai 2015]. Disponible à l'adresse : <http://rddb.cnrs.fr/IMG/pdf/francegrilles-rddb6nov2014v2.pdf?104/d9e53228e07433f0a0f081a31cf232b100b0c050>
- UK Data Service (2013). Data management costing tool. UK Data Archive, University of Essex. [www.data-archive.ac.uk/media/247429/costing\\_v3.docx](http://www.data-archive.ac.uk/media/247429/costing_v3.docx)



[www.cnrs.fr](http://www.cnrs.fr)

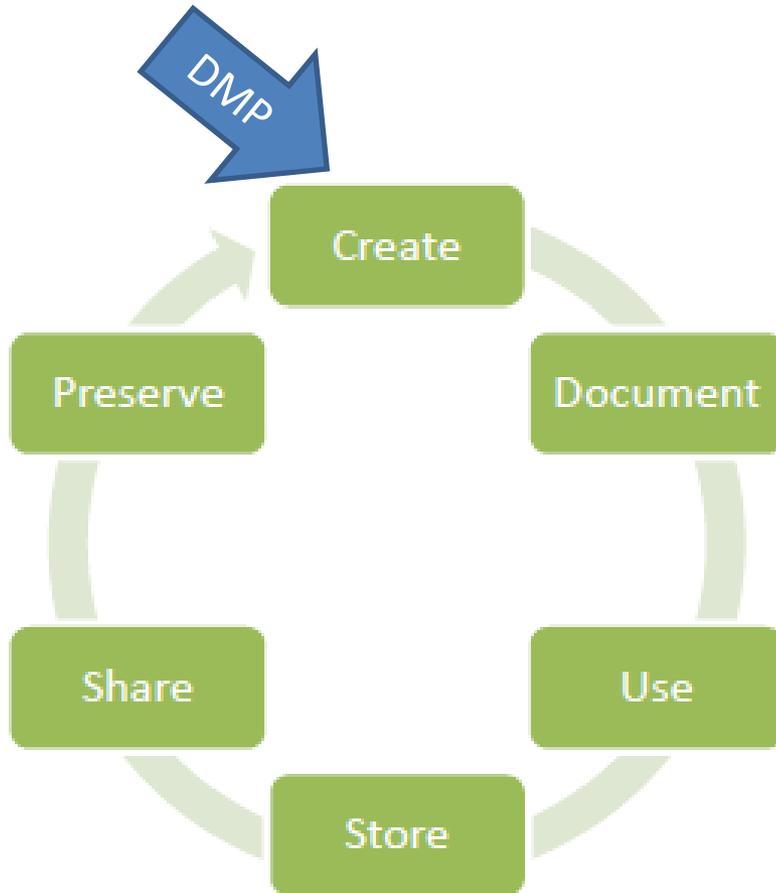
- 3 -

# Bonnes pratiques Elaboration d'un DMP



# Bonnes pratiques et cycle de vie

Jones S (2015). [Managing research data and Horizon 2020](#)



Les bonnes pratiques de gestion des données s'appliquent à chaque étape du cycle de vie



# BONNES PRATIQUES ELABORATION D'UN DMP

---

## Rôles et responsabilités



# Rôles et responsabilités

- Responsable DMP (mise en œuvre, révision, mise à jour)
- Pour chaque étape de gestion des données
  - Collection/création des données
  - Production de la documentation et des métadonnées
  - Qualité des données
  - Stockage et sauvegarde des données
  - Archivage & Partage des données
- Inclus les partenaires
- Politique appliquée aux données : agences de financement, institution...

*« Ce projet est conforme à la politique de données de l'Université de Bath et celle de l'EPSRC»  
Le projet est financé par un Partenaire industriel et il est couvert par un accord de consortium et ma convention de stage »*

*Traduit de University of Bath, Postgraduate Data Management Plan template,  
[www.bath.ac.uk/research/data/planning/dmp-templates/PGR\\_DMP\\_Template\\_20150324.docx](http://www.bath.ac.uk/research/data/planning/dmp-templates/PGR_DMP_Template_20150324.docx)*



# BONNES PRATIQUES ELABORATION D'UN DMP

---

## Créer/collecter des données



# Types de données

## Données d'observation

- capturées en temps réel
- habituellement uniques, impossible à reproduire

*Relevés météo, images*  
*Enquêtes sociales*  
*Fouilles archéologiques*



## Données expérimentales

- obtenues à partir d'équipements de laboratoire
- souvent reproductibles, parfois coûteuses

*Poids biomasse,*  
*Séquence peptide*

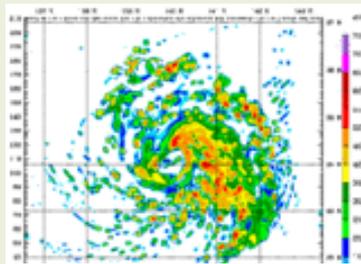


[Pixabay](#), CC0

## Données de simulation numérique

- générées par des modèles informatiques
- souvent reproductibles si le modèle est correctement documenté

*Modèle climatique*  
*Modèle économique*



[Wikimedia](#), CC-BY-Sa 3.0

## Données dérivées ou compilées

- issues du traitement ou de la combinaison de données "brutes"
- souvent reproductibles mais coûteuses

*Base de données compilées*  
*Fouille de texte*



[Heiti Paves](#), CC-BY-SA 3.0

## Données de référence

Séquence gènes ,TP53, Structures chimiques



# Types de données

- Données générées, collectées ? Données existantes (source) ?
- Données brutes, dérivées, analysées ?
- Estimation de volumétrie

*« Each experiment produces about 50MB of data, so over the course of my PhD I expect this to add up to about 700GB. [...]*

*Each simulation generates 4TB of temporary data but I will only retain the output file, which is 100GB.*

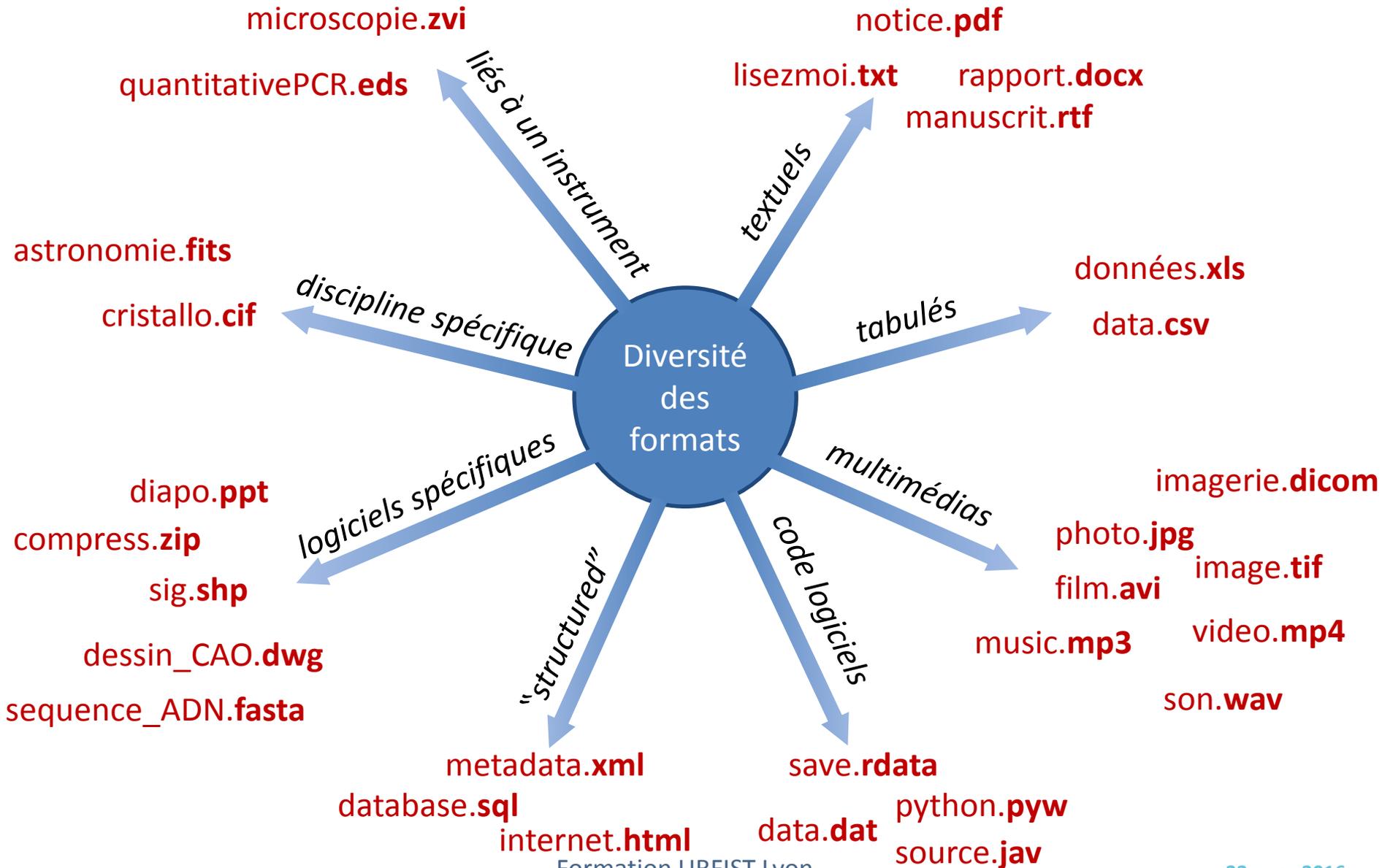
*Based on other members of my research group, I expect to fill 5 lab notebooks during my PhD »*

University of Bath, Postgraduate Data Management Plan template,

[www.bath.ac.uk/research/data/planning/dmp-templates/PGR\\_DMP\\_Template\\_20150324.docx](http://www.bath.ac.uk/research/data/planning/dmp-templates/PGR_DMP_Template_20150324.docx)

- Démarche qualité
  - Assurance qualité / Contrôle qualité
  - Tout au long du cycle de vie
  - DataONE Education module: Data Quality Control and Assurance (2012)

# Formats des fichiers



# Choix du format des fichiers

- Privilégier des formats ouverts, non-propriétaires, d'un usage très répandu au sein d'une communauté de recherche.
  - Prévenir obsolescence des fichiers
  - Données réutilisables et durables
- Quelques exemples (d'après [https://dmptool.org/dm\\_guidance#formats](https://dmptool.org/dm_guidance#formats) )

Format déconseillé	Format recommandé
Excel (.xls, .xlsx)	Comma Separated Values (.csv)
Word (.doc, .docx)	Plain text (.txt) If formatting is needed, PDF/A (.pdf)
PowerPoint (.ppt, .pptx)	PDF/A (.pdf)
Photoshop (.psd)	TIFF (.tif, .tiff)
Quicktime (.mov)	MPEG-4 (.mp4)
Access	MySQL (.sql)

- Et plus : [Tableau des formats de fichiers du UK Data Archive](#)

# Transformations de fichiers

- Transformations de fichiers (Mantra)
  - Conversion, migration : risque pour intégrité  
*« Une description détaillée des formats de conversion est obligatoire dans le contexte du DMP » (Cartier, 2015)*
  - Compression (.zip de facto standard)

# Organisation des fichiers et dossiers

- Bonnes pratiques de convention de nommage
  - Unique, court et descriptif
  - Éviter caractères accentués ou spéciaux tels que \ / : \* ? " < >
  - Pas d'espace et utiliser « \_ » ou utiliser une majuscule en début de mot
  - Ne pas utiliser de mots vides : le, la, les, une...
- Attention aux versions
  - Identifier version avec date (AAAAMMJJ) et numéro de version (\_v01, \_v01-01)
  - Garder version originale, éliminer les versions obsolètes
- Créer un plan de classement des dossiers :  
Type données (textes, images, modèles..), temps (années, mois, session...),  
caractéristiques du sujet d'étude (espèces, groupe d'âge...), activité de recherche  
(espèces, groupe d'âge...)
- Important pour projet collaboratif

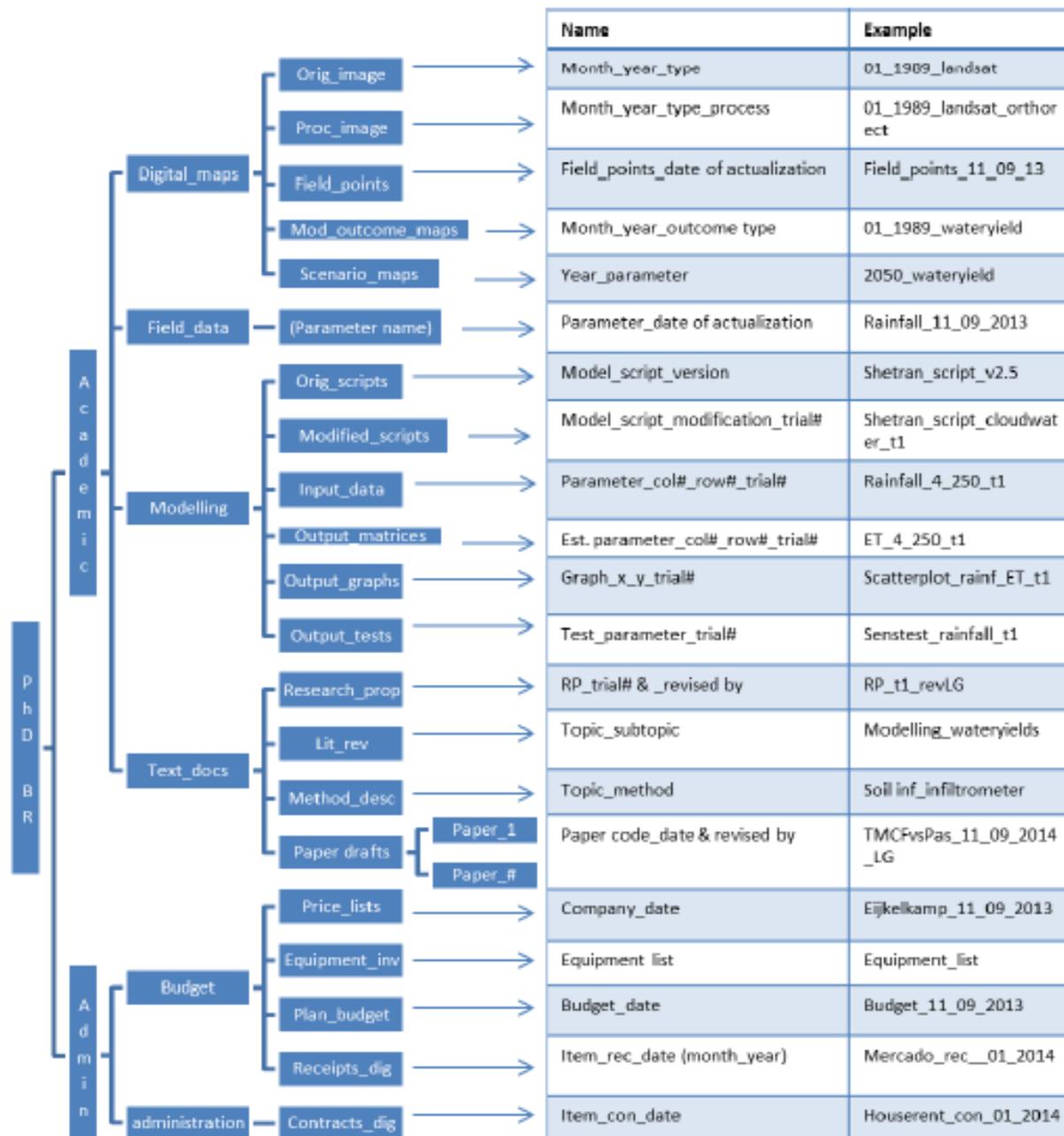
# Exemples

Sans convention de nommage	Avec convention de nommage
Tab 1.XLSX	pathogenMeasurement20130218_v01.xlsx
Design for project.doc	20130503_DOEProject_DesignDocument_Smith_v2-01.docx
Meeting Notes Oct 23	20141023_DOEProject_ProjectMeetingNotes_v1-00.docx
Lab_work_Eric.xlsx	ECOLIRI_DepeAulneData_v20150510.csv <i>Projet : ECOLIRI / Type de données : niveau de dépérissement des aulnes</i> <i>Date version du fichier : 10 oct 2015 / Format fichier : csv</i>

(d'après DataOne, 2012 ; [Purdue University Libraries](#))

Pour en savoir plus :

- University of Edimburgh, [Naming Conventions](#)
- DISC - Inserm (2014) : [Règles de nommage des fichiers électroniques](#)



Extrait de DMP  
 Ramirez B,  
 Wageningen University  
<https://www.wageningenur.nl/web/file?uid=25cb4622-f3a0-4628-8a38-c3c19084e877&owner=0164641e-ab10-40db-b4a7-876f39d24a28>

Figure 1. Proposed folder arrangement and file naming strategy.

# Références

- Cartier A, Moysan M, Reymonet N (2015). Réaliser un plan de gestion de données. Université Paris Diderot et Université Paris Descartes, 30 p. [http://www.univ-paris-diderot.fr/DocumentsFCK/recherche/Realiser\\_un\\_DMP\\_V1.pdf](http://www.univ-paris-diderot.fr/DocumentsFCK/recherche/Realiser_un_DMP_V1.pdf)
- DataONE (2012). DataONE Education Modules. <https://www.dataone.org/education-modules>
- DCC (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. <http://www.dcc.ac.uk/resources/data-management-plans>
- DMPTool. Data Management General Guidance. [https://dmptool.org/dm\\_guidance#formats](https://dmptool.org/dm_guidance#formats)
- EDINA and Data Library, University of Edinburgh (2015). File formats & transformation. Research Data MANTRA [online course, <http://datalib.edina.ac.uk/mantra>]
- Gaillard, R. (2014). De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? (Mémoire de fin d'études de conservateur de bibliothèque, Enssib, Lyon, France). <http://www.enssib.fr/bibliotheque-numerique/documents/64131-de-l-open-data-a-l-open-research-data-quelles-politiques-pour-les-donnees-de-recherche.pdf>
- Scott M (2013). Introducing Research Data. University of Southampton, 31p. <eprints.soton.ac.uk/360442/1/researchdataintro.pdf>
- TGIR Huma-Num (2015). Le guide de bonnes pratiques numériques. [http://www.huma-num.fr/sites/default/files/guide\\_des\\_bonnes\\_pratiques.pdf](http://www.huma-num.fr/sites/default/files/guide_des_bonnes_pratiques.pdf)
- UK Data Archive. File formats table. <http://www.data-archive.ac.uk/create-manage/format/formats-table>
- University of Bath. The Library. Postgraduate Data Management Plan template. [www.bath.ac.uk/research/data/planning/dmp-templates/PGR\\_DMP\\_Template\\_20150324.docx](http://www.bath.ac.uk/research/data/planning/dmp-templates/PGR_DMP_Template_20150324.docx)



# BONNES PRATIQUES ELABORATION D'UN DMP

---

## Documenter les données



# Documenter les données

- Documenter les données = Description des données
- Etape primordiale
- Afin que les données de la recherche soient réutilisables, le contexte de leur production doit être documenté de manière précise et intelligible :
  - Avec une documentation
  - Avec des métadonnées

**Est-ce que les « réutilisateurs » peuvent comprendre mes données ?**

# Avec une documentation

- Pour qui ? **Vous et les autres**
- Pourquoi ? Trouver, **comprendre**, répliquer, vérifier, réutiliser, archiver
- Documentation
  - **Qui, Quoi, Où, Quand, Comment et Pourquoi ?**
  - Informations sur le projet : hypothèse, méthodologie, échantillonnage, instruments, ...
  - Informations sur les fichiers ou base de données et sur les paramètres : unités de mesure, vocabulaire, abréviations...
  - Lisible par l'humain sous forme de fichiers :  
Readme file.txt, document.pdf

# Exemples

Documentation en ligne pour une collection de données dans l'archive UK Data Archive ([Van den Eynden, 2011, p10](#))

FORMAT	NAME	SIZE IN KB	DESCRIPTION
PDF	6713dataset_documentation.pdf	1403	Dataset Documentation (variable list, derived variables, variables used in report tables)
PDF	6713project_instructions.pdf	1998	Project instructions (interviewer, nurse and coding and editing instructions)
PDF	6713questionnaires.pdf	2010	Questionnaires (CAPI and self-completion questionnaires and showcards)
PDF	6713technical_report.pdf	6056	Technical Report
PDF	6713userguide.pdf	256	User Guide
PDF	UKDA_Study_6713_Information.htm	19	Study information and citation

<http://discover.ukdataservice.ac.uk/catalogue/?sn=7576&type=Data%20catalogue#documentation>

# Avec des métadonnées

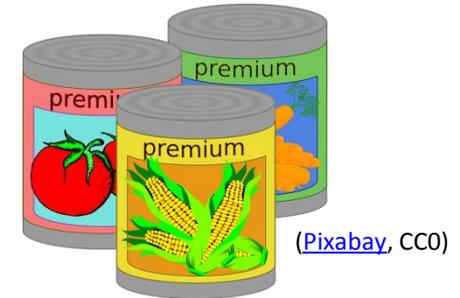
- Donnée à propos d'une autre donnée
- « Ensemble de données structurées décrivant des ressources physiques ou numériques, ou, sur un plan plus fonctionnel, “de **l'information structurée** qui décrit, explique, localise la ressource et en facilite la recherche, l'usage et la gestion” » (Morel-Pair, 2005)
  - **QUI** a créé les données ?
  - **QUE** contiennent-t-elles ?
  - **OÙ** sont-elles ?
  - **QUAND** ont-elles été créées ?
  - **COMMENT** ont-elles été créées ?
  - **POURQUOI** ont-elles été créées ?



([Rainer Zenz](#), CC BY-SA 3.0)

# Avec des métadonnées

- Types de métadonnées
  - métadonnées descriptives du contenu : titre, sujet, description libre, mots clés, qualité des données...,
  - métadonnées techniques et de structure : format, date ...
  - métadonnées administratives: droits de propriété et d'usage, responsabilité (contact), préservation de la ressource
- Interprétables par des machines
  - Permettent l'interopérabilité entre machines
  - Facilitent la **découverte** des données par le biais des moteurs de recherche
- Compléter les métadonnées tout au long du cycle de vie



# Avec des standards de métadonnées

- Utilisés par une large communauté et documentés
  - Schéma/format de métadonnées : structure/plan des métadonnées
  - Renseignés par des valeurs homogènes : codes, thésaurus, ontologies, dates ([YYYY-MM-DDThh:mmTZD](#)), langue (ex : [norme ISO 639-2](#))
- Types
  - Générique : DublinCore, DataCite
  - Spécifique d'une discipline : ISO19115, Ecological Metadata Language (EML)
  - Spécifique d'un thème ou d'une technique : DarwinCore, MAGE-TAB, ISA-TAB

# Où trouver des standards de métadonnées ?

- Choix selon discipline



<http://www.dcc.ac.uk/resources/metadata-standards>



<http://rd-alliance.github.io/metadata-directory/>



<https://www.biosharing.org/standards/>

- Choix selon les objectifs
  - EML/Darwin Core pour les données en écologie

« Adopt, Adapt or Act »

<http://ands.org.au/guides/metadata-working.html>



Si Act, alors renseigner les « informations minimales »

# Exemples - DataCite Schema

[doi:10.4231/D3P26Q35R](https://doi.org/10.4231/D3P26Q35R)

This page represents DataCite's metadata for *doi:10.4231/D3P26Q35R*.

For a landing page of this dataset please follow <http://dx.doi.org/10.4231/D3P26Q35R>

Citation	Fosmire, Michael; Wertz, Ruth; Purzer, Senay; (2013): Critical Engine
Descriptions	
Abstract	The CELT download includes the blueprint for writing assessment ite
Resource type	
Dataset	Dataset
Rights	CC0 - Creative Commons
Language	en
Dates	
Valid	2013-04-08
Accepted	2013-04-08
Version	1.0
Contributors	
ProjectLeader	Michael Fosmire

#### Other formats

[text/html](#)

[application/x-datacite+xml](#)

[application/vnd.datacite.datacite+xml](#)

[application/x-datacite+text](#)

[application/vnd.datacite.datacite+text](#)

[application/rdf+xml](#)

[text/turtle](#)

[application/x-bibtex](#)

[application/x-research-info-systems](#)

[application/citeproc+json](#)

[application/vnd.citationstyles.csl+json](#)

[text/x-bibliography](#)

This metadata record was deposited by Purdue University in co-operation with Purdue Universit

```
<?xml version="1.0" encoding="UTF-8"?>
- <resource xsi:schemaLocation="http://datacite.org/schema/kernel-2.1 http://
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://
    <identifier identifierType="DOI">10.4231/D3P26Q35R</identifier>
  - <creators>
    - <creator>
      <creatorName>Fosmire, Michael</creatorName>
    </creator>
    - <creator>
      <creatorName>Wertz, Ruth</creatorName>
    </creator>
    - <creator>
      <creatorName>Purzer, Senay</creatorName>
    </creator>
  </creators>
- <titles>
  <title>Critical Engineering Literacy Test (CELT)</title>
</titles>
<publisher>PURR</publisher>
<publicationYear>2013</publicationYear>
- <contributors>
  - <contributor contributorType="ProjectLeader">
    <contributorName>Michael Fosmire</contributorName>
  </contributor>
</contributors>
- <dates>
  <date dateType="Valid">2013-04-08</date>
  <date dateType="Accepted">2013-04-08</date>
</dates>
<language>en</language>
<resourceType resourceTypeGeneral="Dataset">Dataset</resourceType>
<version>1.0</version>
<rights>CC0 - Creative Commons</rights>
- <descriptions>
  <description descriptionType="Abstract">The CELT download include
    (4.1) includes 2 parts, one on an engineering topic, and one on
  </description>
</descriptions>
</resource>
```

[Exemple pour un jeu de données en XML \(langage « balisé »\)](#)

# Examples - DataCite Schema

Table 1: DataCite Mandatory Properties

<i>ID</i>	<i>Property</i>	<i>Obligation</i>
1	Identifier (with type sub-property)	M
2	Creator (with name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M

Table 2: DataCite Recommended and Optional Properties

<i>ID</i>	<i>Property</i>	<i>Obligation</i>
6	Subject (with scheme sub-property)	R
7	Contributor (with type, name identifier, and affiliation sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
10	ResourceType (with general type description sub-property)	R
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point and box sub-properties)	R

<https://schema.datacite.org/meta/kernel-3.1/index.html>

# EML : Ecological Metadata Language

knb ABOUT DATA SHARE TOOLS Search for data SIGN IN

Northeastern Ecosystem Research Cooperative (NERC). **Compilation of surface water chemistry data for the northeastern United States and southeastern Canada** (doi:10.5063/AA/NERC.14.4)

Files in this dataset

Name	File type	Size	Downloads
Metadata: Compilation of surface water chemistry data for the northeastern United States and southeastern Canada	xml (EML)	12 KB	49 views

Download

General

Title: **Compilation of surface water chemistry data for the northeastern United States and southeastern Canada**

Identifier: NERC.14.4

Abstract: The surface water chemistry component of the Northeastern Ecosystem Research Cooperative (NERC) regional database was assembled from five prior NERC synthesis projects designed to investigate regional environmental trends in forested and aquatic ecosystems of northeastern North America. These individual NERC synthesis projects compiled surface water chemistry data from independent studies across the northeastern United States and southeastern Canada. The combined foliar, soil, and surface water chemistry data in the full NERC regional database thus provide an opportunity to analyze environmental trends across ecosystem compartments at a regional scale. Surface water chemistry data in the database are from lakes or streams and include measures of pH, water clarity, nutrient concentrations or fluxes, other element concentrations or fluxes, organic and inorganic forms of carbon and nitrogen, streamflow, and others.

Keywords

Keyword	Type
Aquatic ecosystem	
Element concentration	
Element flux	
Lake	
Nutrients	
Stream	
Surface water chemistry	

Pourquoi?

Quoi ?

# EML : Ecological Metadata Language

<https://knb.ecoinformatics.org/#view/doi:10.5063/AA/NERC.14.4>, consulté en juin 2015

Geographic Region

Où ?

Geographic Description

Surface water chemistry data included in the NERC regional database are predominantly from sites in the northeastern United States (CT, MA, MD, ME, NH, NY, RI, VT, WV) and southeastern Canada (NB, NF, NS, PQ). Approximately 85% of the data were collected in New England or New York.

Bounding Coordinates

North 49.6 degrees

Sampling, Processing and Quality Control Methods

Methods

Step 1

Comment?

Description

Compilation of data from NERC synthesis projects

Surface water chemistry data were obtained from the following individual NERC synthesis project databases. For each of these synthesis projects, investigators compiled surface water chemistry data from existing studies conducted in northeastern North America, in order to facilitate the examination of regional environmental trends:

- Experimental Nitrogen Addition Synthesis Project (ExpN): Synthesis of forested and associated aquatic ecosystem responses to experimental nitrogen addition. Analysis in progress, led by Lindsey Rustad, US Forest Service Northern Research Station. Citation: Rustad, L.E. et. al. In prep. A meta-analysis of forest ecosystem response to experimentally elevated nitrogen deposition.
- Mercury (Water) Synthesis Project (Hg Water): Analysis of physical factors controlling mercury concentrations in lakes and streams of the northeastern United States. Citation: Shanley, J. B., N. C. Kamman, T. A. Clair, and A. Chalmers. 2005. Physical controls on total and methylmercury concentrations in streams and lakes of the northeastern USA. *Ecotoxicology* 14:125-134.
- Nitrogen Saturation (Water) Synthesis Project (N Sat S): Synthesis of nitrogen cycling data for lakes and streams of the northeastern United States, to contribute to an analysis of whether nitrogen deposition is altering the nitrogen status of this region. Citation: Aber, J. D., C. L. Goodale, S. V. Ollinger, M. L. Smith, A. H. Magill, M. E. Martin, R. A. Hallett, and J. L. Stoddard. 2003. Is nitrogen deposition altering the nitrogen status of northeastern forests? *Bioscience* 53:375-389.
- University of Maine Mitchell Center Lakes Database (Lakes): Repository of lake chemistry data for the northeastern United States compiled from the Temporally Integrated Monitoring of Ecosystems (TIME), Regionalized Long Term Monitoring (RLTM), High Elevation Lake Monitoring, (HELM), and Aquifer Lakes (ALPS) projects. Citations: Citations using data from these studies are listed in the SurfaceWaterChemistryData and Publications tables.

Data Set Usage Rights

Intellectual Rights

The NERC and its data contributors make these data freely accessible to the public under the following conditions: (1) Those wishing to use these data to generate broad ranges, averages, or other summary statistics should cite the NERC regional database using the Data Set Citation provided. (2) Those wishing to publish or otherwise extensively use these data agree first to contact the contributing scientists of the specific data of interest regarding the proposed use. The primary contact information and other coauthors for each study are listed in the SurfaceWaterChemistryData table, and email addresses for primary contacts are provided in the Primary Contacts table. (3) Those wishing to use these data agree to credit the contributing scientists at the level appropriate to the data use. Depending on the agreement reached between the data user(s) and the data contributor(s), appropriate credit may include review and citation of associated publications, citation of funding sources, and acknowledgment in the user's publications. (4) While the NERC regional database has been compiled from many contributions by independent sources. While efforts have been made to ensure data accuracy, the NERC and its data contributors will not be held liable for any errors in the compiled data. (5) Data are available for new information or further data review.

Comment ?

Access Control

Auth System

knb

Order

allowFirst

Allow

Permission

Principal

[read]

public

Quand ?

Temporal Coverage

Date Range

# Exemple : DDI (Data Documentation Initiative)

Jeu de données: Espaces de vie et environnement (1991-1992)

---

## Entité responsable de la collecte des données

Institut national d'études démographiques (INED)

---

## Procédure d'échantillonnage

L'échantillon a été stratifié selon les critères suivants puis affiné selon l'appartenance rurale ou urbaine :

- critère régional : découpage du territoire en 8 régions (Bretagne, Ouest du bassin parisien, Nord Est, Sud du bassin parisien, Rhône-Alpes, Auvergne-Limousin, Méditerranée, Sud-Ouest)
- critère rural-urbain, avec pour les villes découpage en classe de taille et localisation au sein de l'agglomération (centre ou périphérie)
- critère lié au milieu physique : situation de la commune du lieu résidence en zone littorale, de montagne ou de plaine

Les critères conduisent à définir 53 strates dans l'échantillon rural. Les unités géographiques sont tirées indépendamment dans chaque strate et à taux fixe dans chacune d'elles : 1 enquêté sur 2000 habitants pour le littoral rural, 1 sur 4000 pour la montagne rurale, 1 sur 6000 pour la plaine rurale.

Pour l'échantillon urbain, les taux de sondage sont de 1 enquêté pour 12 000 en zone littorale ou de montagne et 1 sur 15 000 en plaine.

Les ménages enquêtés ont été sélectionnés à partir du fichier des abonnés de France Télécom (listes rouge et orange exclues). Un adulte du ménage a été désigné selon la procédure de Kish.

Pour plus d'informations sur le plan de sondage, voir le document "Plan de sondage et redressement de l'échantillon" dans "Autre matériel pour la description de l'étude / Matériel relié".

---

## Mode de collecte des données

Enquête principale : Face à face

Enquête téléphonique : Téléphone

---

## Type d'instrument de collecte

Questionnaire majoritairement directif comprenant 18 modules.

Dans ce questionnaire, les personnes étaient tout d'abord interrogées sur leur passé, en accordant une importance toute particulière à leurs lieux de résidences successifs, aux souvenirs liés à ces lieux, aux paysages dans lesquels elles ont vécu (modèles A et B du questionnaire). Ensuite,

Catalogue Nesstar des enquêtes de l'INED

<http://nesstar.ined.fr/webview/?=2&study=http://nesstar.ined.fr:80/obj/fStudy/IE0180&mode=documentation&submode=ddi&node=0&top=yes>

# Métadonnées de la Directive INSPIRE

Tableau 1: Métadonnées pour les séries de données géographiques et les ensembles de séries de données géographiques

Référence	Éléments de métadonnée	Cardinalité	Condition
1.1	Intitulé de la ressource	1	
1.2	Résumé de la ressource	1	
1.3	Type de ressource	1	
1.4	Localisateur de la ressource	0..*	Obligatoire s'il existe un URL permettant d'obtenir davantage d'informations sur la ressource et/ou les services liés à l'accès.
1.5	Identificateur de ressource unique	1..*	
1.7	Langue de la ressource	0..*	Obligatoire si la ressource inclut des informations textuelles.
2.1	Catégorie thématique	1..*	
3	Mot-clé	1..*	
4.1	Rectangle de délimitation géographique	1..*	
5	Référence temporelle	1..*	
6.1	Généalogie	1	
6.2	Résolution spatiale	0..*	Obligatoire pour les séries de données et les ensembles de séries de données pour lesquels une échelle équivalente ou une distance de résolution peuvent être indiquées.
7	Conformité	1..*	
8.1	Conditions applicables à l'accès et à l'utilisation	1..*	
8.2	Restrictions concernant l'accès public	1..*	
9	Organisation responsable	1..*	
10.1	Point de contact des métadonnées	1..*	
10.2	Date des métadonnées	1	
10.3	Langue des métadonnées	1	

(CNIG, 2014: p81)

Utilisé par :

- [CARGOS](#) (catalogue de données géographiques en SHS). Obligation d'attribuer un mot-clé issu du thésaurus [INSPIRE-GEMET Spatial Data Themes](#)
- [Portail des données marines](#) (Ifremer)
- [Indigeo](#) (Infrastructure géographique de données géospatialisées sur l'environnement)

# Références

- BioSharing. Standards registry. <https://www.biosharing.org/standards/>
- Maison des Sciences de l'Homme de Dijon. CARGOS Catalogue de données géographiques en SHS. <http://cargos.huma-num.fr/>
- CNIG (2014). Guide de saisie des éléments de métadonnées INSPIRE. 103p. <http://cnig.gouv.fr/wp-content/uploads/2014/07/Guide-de-saisie-des-%C3%A9l%C3%A9ments-de-m%C3%A9tadonn%C3%A9es-INSPIRE-v1.1.1.pdf>
- DataCite. DataCite Metadata Schema Repository. <https://schema.datacite.org/>
- Digital Curation Centre. Disciplinary Metadata. <http://www.dcc.ac.uk/resources/metadata-standards>
- Morel-Pair C (2005). Panorama des métadonnées pour les ressources électroniques. In Ateliers des réseaux de la documentation scientifique, Arcachon, 11-13 octobre 2005. Format PDF. <http://www.enssib.fr/bibliotheque-numerique/notice-1841>
- Ifremer. Portail des données marines. <http://data.ifremer.fr/>
- RDA Metadata Standards Directory Working Group. Metadata Repository. <http://rd-alliance.github.io/metadata-directory/>
- Van den Eyden V, Corti L, Woollard B et al (2011). Managing and sharing data. Best practice for researchers. [www.data-archive.ac.uk/media/2894/managingsharing.pdf](http://www.data-archive.ac.uk/media/2894/managingsharing.pdf)



# BONNES PRATIQUES ELABORATION D'UN DMP

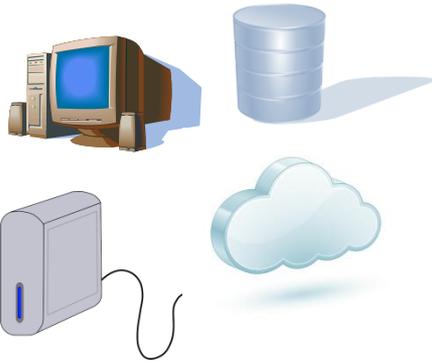
---

Stocker - Utiliser



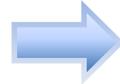
# Stocker - Utiliser

## Solutions de stockage



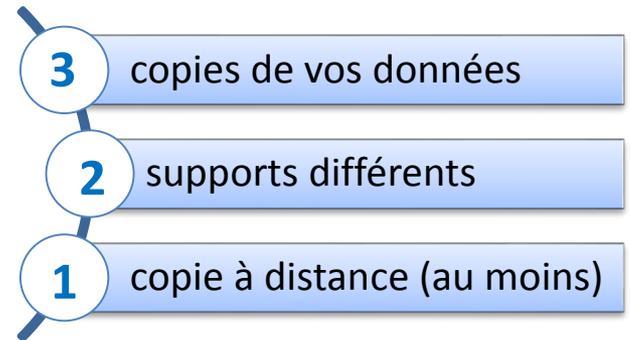
### Avantages et inconvénients

(Wageningen UR/Pays- Bas)



## Préférer le stockage centralisé

- Accès sécurisé aux fichiers en tout lieu
- Règle du 3...2...1... Sauvegarde !



Données confidentielles  
Données à caractère personnel



Accès contrôlé sur site  
Encryptage lors du stockage  
et transfert des données



# BONNES PRATIQUES ELABORATION D'UN DMP

---

## Partager les données



# Où partager?

- En les publiant dans les **fichiers supplémentaires** associés à un article (supplementary materials)
- En les publiant dans des **data papers** ou **data journals**
- En les déposant dans des **entrepôts de données de recherche**
- En les publiant *via* le site web d'un projet
- Dans le web de données
  - Ex : Dzale Yeumo WE (2015). Publication en Linked Open Data de données expérimentales sur la chenille processionnaire du pin.  
<http://prodinra.inra.fr/?locale=fr#!ConsultNotice:195427>

# Data papers & data journals

## Qu'est-ce qu'un data paper ?

- Publication dont le but est de décrire un ou plusieurs jeux de données scientifiques, notamment à l'aide d'informations précises, appelées métadonnées (qui, quoi, où, pourquoi, comment)
- Pas d'hypothèse/conclusion/interprétation issues de l'analyse des données
- Accès en ligne aux données décrites dans le data paper
  - par lien pérenne (DOI, URL) vers l'entrepôt de données où elles ont été déposées
  - sous forme de fichiers annexés
- Processus de peer-review variables

(Archambeau, 2013; Dedieu, 2014)

# Data papers & data journals

Data papers publiés dans deux types de revues (Dedieu, 2014)

- **Data journals** - revues contenant exclusivement des data papers

Journal of  
**open** archaeology data

Journal of  
**open** psychology data

**open** health data

Ubiquity Press

Research Data Journal  
for the Humanities  
and Social Sciences

Brill

SCIENTIFIC DATA

Nature Publishing Group

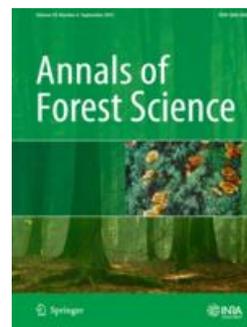
- Revues classiques



ESA



Elsevier



Springer

Autres exemples :

- Akers K (2014), University of Michigan  
[A growing list of data journals](#)

# Bénéfices

- Promouvoir, faire connaître et rendre accessibles et réutilisables ses données de recherche
- Décrire les données sous forme structurée et lisible par un humain
- Augmenter la visibilité et l'impact des données de la recherche grâce à une meilleure citabilité des données
- Apporter du crédit aux chercheurs *via* une publication scientifique citable

(Archambeau, 2013; Dedieu, 2014)

# Structure d'un data paper

- Template "Journal of Open Archaeology Data (Ubiquity Press)"

**Data paper for submission to the Journal of Open Archaeology Data**

To complete this template, please replace the blue text with your own. The paper has four main sections: (1) Overview; (2) Methods; (3) Dataset description; (4) Reuse potential. Please submit the completed paper to: [joad.editor@ubiquitypress.com](mailto:joad.editor@ubiquitypress.com).

---

**(1) Overview**

**Title**  
The title of the data paper should focus on the data, e.g. "Geographic information system data from the X project". If the data is closely linked to a specific project, then "Data from Paper Title" is appropriate.

**Paper Authors**  
1. Last name, first name; (*Lead/corresponding author first*)  
2. Last name, first name; etc.

**Paper Author Roles and Affiliations**  
1. First author role and affiliation  
2. Second author role and affiliation etc.

**Abstract**  
A short (ca. 100 word) summary of the dataset being described: what it covers, how it was collected, how it is stored, and its reuse potential.

**Keywords**  
[keyword 1](#); [keyword 2](#); etc.

**Context**

**Spatial coverage**  
(Boundary coordinates can be found using the [itouch map](#) tool.)

Description: Place names including country, region, town, of where collected  
Northern boundary: +/- [x.x](#)  
Southern boundary: +/- [x.x](#)  
Eastern boundary: +/- [x.x](#)  
Western boundary: +/- [x.x](#)

**Temporal coverage**  
The start and end dates of the period to which the data is related ([dd/mm/yyyy](#)). Data from excavation of a Roman site in England might have the dates AD250-AD350.

**(2) Methods**  
Describe the methods used to create the dataset (ca. 100-200 words), including the following sub-headings:

**Steps**  
The series of procedures followed to produce the dataset. This should include any data used, as well as software and instrumentation involved.

**Sampling strategy**  
If relevant, please outline the sampling strategy used to produce the data.

**Quality Control**  
Please list the methods used for quality control in the production of the data.

**Constraints**  
Please list any constraints involved in data production.

**(3) Dataset description**

**Object name**  
Typically the name of the file or file set in the repository.

**Data type**  
Please enter one or more from primary data, secondary data, processed data, image of data, or final report.

**Format names and versions**  
E.g. ASCII, CSV, [Autocad](#), EPS, JPEG, Excel, SQL, etc.

**Creation dates**  
The start and end dates of when the data was created ([dd/mm/yyyy](#)).

**Dataset Creators**  
Please list anyone who helped to create the database (who may also not be an author of the data paper), including their roles and affiliations.

**Language**  
Language.

**License**  
[The](#) open license under which the data has been deposited (e.g. CC0).

**Repository location**  
If already available, please include a permanent identifier such as a DOI that points to the online location of the dataset.

**Publication date**  
If already known, the date the dataset was published in the repository ([dd/mm/yyyy](#)).

**(4) Reuse potential**  
(ca. 50-200 words) Please describe the ways in which your data could be reused by other researchers both within and outside of your field. This might for example include aggregation, further analysis, reference, validation, teaching or collaboration.

**Acknowledgements**  
Please add any relevant acknowledgements to anyone else who supported the project in which the data was created, but did not work directly on the data itself.

**Funding statement**  
If the data resulted from funded research please give the funder and grant number.

---

**References**

Please enter references in the Harvard style and include a DOI where available, citing them in the text with a number in square brackets, e.g.

[1] [Piwowar, H A](#) 2011 Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. *PLoS ONE* 6(7): e18657. DOI: <http://dx.doi.org/10.1371/journal.pone.0018657>.

---

**Copyright Notice**  
Authors who publish with this journal agree to the following terms:

Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a [Creative Commons Attribution License](#) that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.

Authors are able to enter into separate, additional contractual arrangements for the non-exclusive distribution of the journal's published version of the work (e.g., post it to an institutional repository or publish it in a book), with an acknowledgement of its initial publication in this journal.

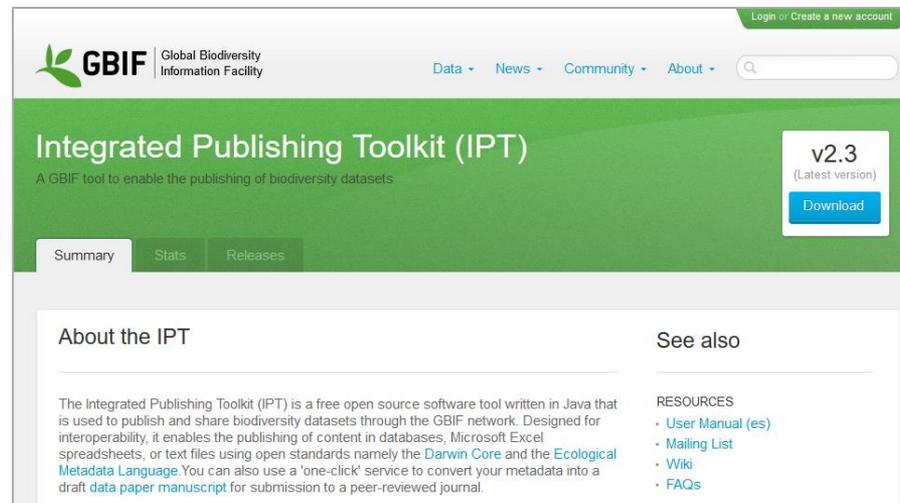
By submitting this paper you agree to the terms of this Copyright Notice, which will apply to this submission if and when it is published by this journal.

<http://openarchaeologydata.metajnl.com/about/submissions/>

- Divers noms : data article, dataset paper, data descriptors, data note, data in brief articles...(Candela, 2014)

# Exemples d'outils d'aide à la publication

- Arpha writing tool / Pensoft
  - [http://arpha.pensoft.net/login.php?back\\_uri=%2F](http://arpha.pensoft.net/login.php?back_uri=%2F)
  - Outil d'écriture collaborative en ligne
- Integrated Publishing Toolkit / GBIF (Parmelon, 2015)
  - <http://www.gbif.org/ipt>
  - Plateforme logicielle libre
  - Publication des données au format Darwin Core et des métadonnées en EML



GBIF | Global Biodiversity Information Facility

Data - News - Community - About -

## Integrated Publishing Toolkit (IPT)

A GBIF tool to enable the publishing of biodiversity datasets

v2.3  
(Latest version)  
[Download](#)

Summary Stats Releases

### About the IPT

The Integrated Publishing Toolkit (IPT) is a free open source software tool written in Java that is used to publish and share biodiversity datasets through the GBIF network. Designed for interoperability, it enables the publishing of content in databases, Microsoft Excel spreadsheets, or text files using open standards namely the [Darwin Core](#) and the [Ecological Metadata Language](#). You can also use a 'one-click' service to convert your metadata into a draft [data paper manuscript](#) for submission to a peer-reviewed journal.

### See also

RESOURCES

- [User Manual \(es\)](#)
- [Mailing List](#)
- [Wiki](#)
- [FAQs](#)

# A noter

Dans le contexte de l'Open Science :

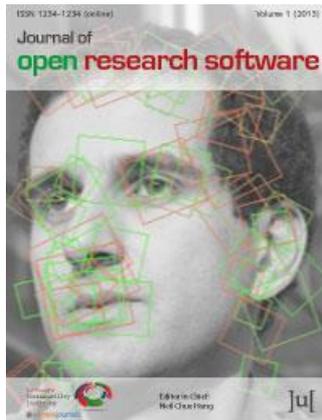
- Publication de Bioresource papers

*Open Journal of Bioresources*  
(Ubiquity Press)

<http://openbioresources.metajnl.com/>



- Publication de Software metapapers / Software papers



*Journal of open research software*  
(Ubiquity Press)

<http://openresearchsoftware.metajnl.com/>

# Références

- Archambeau AS, Lecoq ME, GBIF : Global Biodiversity Information Facility Système Mondial d'Information sur la Biodiversité. In : *Journées rBDD 2013* [en ligne]. S.l. [Consulté le 28 mai 2015]. Disponible à l'adresse : [http://rbdd.cnrs.fr/spip.php?action=accéder\\_document&arg=54&cle=2a15d54c40fef3ee84220e4b913b4a6260413a41&file=pdf%2F10-11.pdf](http://rbdd.cnrs.fr/spip.php?action=accéder_document&arg=54&cle=2a15d54c40fef3ee84220e4b913b4a6260413a41&file=pdf%2F10-11.pdf)
- Candela L, Castelli D, Manghi P et Tani A (2014). Data journals: A survey. *J Assoc Inf Sci Technol*. doi: 10.1002/asi.23358  
[http://www.niso.org/apps/group\\_public/download.php/14938/DataJournalsSurvey%20%281%29.pdf](http://www.niso.org/apps/group_public/download.php/14938/DataJournalsSurvey%20%281%29.pdf)
- Dedieu L (2014). Rédiger et publier un data paper dans une revue scientifique en 5 points. Montpellier (FRA) : CIRAD, 7 p. <http://url.cirad.fr/ist/data-paper>
- Parmelon S (2015). Le GBIF et les identifiants persistants : Application des DOI aux jeux de données. In : *Journées rBDD 2015* [en ligne]. Disponible à l'adresse : <http://rbdd.cnrs.fr/spip.php?article186>

# Entrepôts de données

- Data repository, digital repository
- Un **entrepôt de données** est un réservoir de données de recherche, brutes ou dérivées, qui peuvent être retrouvées et réutilisées grâce à une description par des métadonnées. Un identifiant pérenne ou numéro d'accès est attribué à chaque jeu de données (d'après Aventurier P, 2013)
- Types d'entrepôts
  - Disciplinaire: [SEANOE](#), [PANGAEA](#), [Réseau Quetelet](#)
  - Multidisciplinaire : [Zenodo](#), [Figshare](#), [Dryad](#)
  - Institutionnel : [Edinburgh DataShare](#), [Merritt](#), [3TU Datacentrum](#)
  - Spécifique d'un projet de recherche: [Scientific Drilling Database](#)
  - Banque de données de référence : [GenBank](#), [UniProt](#)

# Entrepôts de confiance

- Trusted repository
- Entrepôt de données répondant aux critères de qualité exigés pour obtenir une certification (format des données, qualité des métadonnées, conditions d'accès et de réutilisation, identifiant pérenne, archivage à long terme, ...)
- Certification « niveau basique » de [Data Seal of Approval](#)
  - Les données sont disponibles sur Internet
  - Les données sont accessibles (licences et droits)
  - Les données sont disponibles dans un format utilisable
  - Les données sont fiables
  - Les données sont citables *via* un système d'identifiant unique et pérenne
- [Groupe de travail RDA](#) : harmonization des critères de certifications entre DSA et ICSU-WDS



# Pourquoi déposer dans un entrepôt ?

- Mise à disposition des données
- Réutilisation des données par le déposant et ses pairs
- Citabilité des données
- Visibilité accrue
- Transparence des données
- Lien bidirectionnel entre des jeux de données et un article (ou data paper)
- Conservation des données

# Programme-Cadre Horizon 2020

## RESEARCH DATA MANAGEMENT IN THE MODEL GRANT AGREEMENT

References to research data management as part of the Pilot on Open Research Data in the Horizon 2020 are included in [Article 29.3](#) of the Model Grant Agreement, which applies to all projects participating in the Pilot:

*'Regarding the digital research data generated in the action ('data'), the beneficiaries must:*

- (a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:*
  - (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;*
  - (ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan (see Annex I);*
- (b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and —*

These provisions are further explained in the [Open Access Guidelines](#) and in the [Annotated Model Grant Agreement](#).

(European Commission, 2016)

# Comment trouver un entrepôt ?

- Multidisciplinaires
  - [re3data](#) (Registry of Research Data Repositories) (mentionné dans les lignes directives de la Commission européenne - H2020)
  - [Open Access Directory](#)
- Spécialisés en sciences de la vie et biomédecine
  - [BioSharing](#)
- Archives en libre accès
  - [OpenDOAR](#) Directory of Open Access Repositories
  - [ROAR](#) Registry of Open Access Repositories

# Quelques critères de sélection

## Politique(s)

- Financeurs :  
[NERC](#) (NERC data centre)
- Editeurs :  
[Elsevier](#)  
NPG ([Scientific Data](#))

Politique(s)

Caractéristiques  
de l'entrepôt

Disciplines

## Disciplines

- [Réseau Quetelet](#) (sc. sociales)
- [CDS](#) (données astronomiques)

## Caractéristiques de l'entrepôt

- Formats de fichiers acceptés
- Modalités d'accès
- Facilité au dépôt
- Entrepôt de confiance
- Localisation du serveur
- Coût...

Entrepôt

(d'après Swauger S, 2015)

# Quand partager ?

- Variable selon discipline et type de données (« période propriétaire »)
- Dépend des recommandations des agences de financement
  - Horizon 2020 / CE : pour les données sous-jacentes, idéalement en même temps que le dépôt des publications scientifiques
  - [NSF-Earth Sciences Division](#) : *“no later than two (2) years ears after the data were collected”*

# Comment partager ?

- Définir les permissions d'accès
  - Restreint ou ouvert : possibilité de limiter l'accès aux données à des groupes restreints de personnes (communauté scientifique, ...)  
<http://discover.ukdataservice.ac.uk/> (voir filtre Access)  
<http://reshare.ukdataservice.ac.uk/851776/>
  - Embargo : période pendant laquelle les données de la recherche (déposées dans un entrepôt de données) ne sont pas accessibles librement

# Exemples « Période d'embargo »

The image shows a screenshot of a Zenodo record page. The header features the Zenodo logo and the tagline "Research. Shared." with navigation links for Search, Communities, Browse, Upload, and Get started, along with Sign In and Sign Up buttons. The record details include the date "08 May 2015", a "Dataset" label, and an "Embargoed access" label. The title is "Amber fossils demonstrate deep-time stability of Caribbean lizard communities". The authors listed are Sherratt, Emma; Castañeda, María del Rosario; Garwood, Russell; Mahler, D. Luke; Sanger, Thomas J.; Herrel, Anthony; de Queiroz, Kevin; and Losos, Jonathan B. The abstract discusses the stability of ecological communities over macroevolutionary timescales. The right sidebar provides metadata: Publication date (08 May 2015), Embargoed status (files available as Open Access after 12 July 2015), DOI (10.5281/zenodo.17442), Keywords (Anolis, micro-Computed Tomography, amber, Adaptive radiation, Ecomorph, Hispaniola), Collections (Communities, Datasets, Open Access), License (Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International), and Uploader (esherratt, on 08 May 2015). Two red boxes highlight the "Embargoed access" label and the "Embargoed" status with the open access date.

08 May 2015

Dataset Embargoed access

## Amber fossils demonstrate deep-time stability of Caribbean lizard communities

Sherratt, Emma ; Castañeda, María del Rosario ; Garwood, Russell ; Mahler, D. Luke ; Sanger, Thomas J. ; Herrel, Anthony ; de Queiroz, Kevin ; Losos, Jonathan B.

(show affiliations)

This is the dataset for the paper Sherratt et al. 2015 (PNAS, doi: ). Included are the X-ray microtomography (micro-CT) TIFF stacks for the amber encased anole lizard fossils studied. Also included are the morphometric data for the fossil and 100 modern specimens, and the phylogenetic character matrix for the fossils and 181 species (91 morphological characters, 4873 DNA bases).

Abstract: Whether the structure of ecological communities can exhibit stability over macroevolutionary timescales has long been debated. The similarity of independently evolved *Anolis* lizard communities on ecologically similar Greater Antillean islands supports the notion that community evolution is deterministic. However, a dearth of Caribbean *Anolis* fossils - only 3 have been described to date - has precluded direct investigation of the stability of anole communities

**Publication date:**  
08 May 2015

**Embargoed**  
Files available as **Open Access** after 12 July 2015

**DOI**  
DOI 10.5281/zenodo.17442

**Keyword(s):**  
Anolis micro-Computed Tomography amber Adaptive radiation Ecomorph Hispaniola

**Collections:**  
Communities  
Datasets  
Open Access

**License (for files):**  
Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

**Uploaded by:**  
esherratt (on 08 May 2015)

<https://zenodo.org/record/17442?ln=en#.VWrQy0bATAF>

Data from: Evidence for non-allopatric speciation among closely related sympatric *Heliotropium* species in the Atacama Desert

Consulté en février 2014

### Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Terms of Service](#). To the extent possible under law, the author or neighboring rights to this data.  

Title	<b>Heliotropium_AFLP_Data</b>
Downloaded	8 times
Description	AFLP data in genpop format
Download	<a href="#">Mcheza table AFLP Heliotropium.txt</a>
Details	<a href="#">View File Details</a>

Title	<b>Species tree of Heliotropium sect. Cochranea</b>
Description	Species tree of Heliotropium sect. Cochranea generated with star-BEAST
Download	<a href="#">14422 (0bytes)</a>
Details	<a href="#">View File Details</a>

At the request of the author, this item is embargoed until 2014-12-29

Data from: Evidence for non-allopatric speciation among closely related sympatric *Heliotropium* species in the Atacama Desert



Consulté en mai 2015

### Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the [Dryad Terms of Service](#). To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.  

Title	<b>Heliotropium_AFLP_Data</b>
Downloaded	4 times
Description	AFLP data in genpop format
Download	<a href="#">Mcheza table AFLP Heliotropium.txt (89.16Kb)</a>
Details	<a href="#">View File Details</a>

Title	<b>Species tree of Heliotropium sect. Cochranea</b>
Description	Species tree of Heliotropium sect. Cochranea generated with star-BEAST
Download	<a href="#">S14422 (0bytes)</a>
Details	<a href="#">View File Details</a>

Title	<b>Additional DNA sequences of Heliotropium sect. Cochranea</b>
-------	---

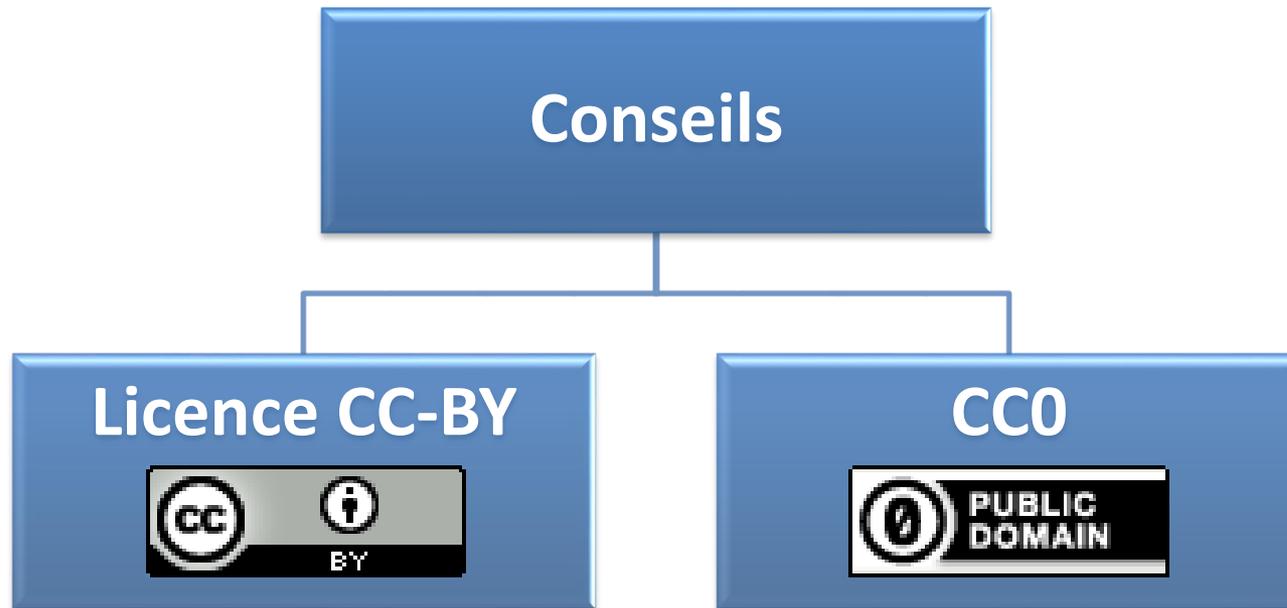
<http://datadryad.org/resource/doi:10.5061/dryad.ps46g>

# Comment partager ?

- Propriété littéraire et artistique (propriété intellectuelle)
  - Droit d'auteur : autorisation pour des « données détenues par un tiers »
  - Droit sui generis des bases de données
- Attribuer des licences aux données (Dedieu, 2015 ; Fily, 2015)
  - [Licences de l'Open Knowledge Foundation](#)
    - ODC-by : Open Database Commons
    - ODC-ODbL : Open Database License ([traduction française](#))
    - PDDL : Public domain dedication and License
  - [Licences Creative Commons](#) version 4.0
    - Internationale
    - Droit des bases de données pris en compte
    - Modulables (4 options, 6 licences et Déclaration CC0)
  - [Licence ouverte](#) (Open Licence)



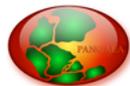
# Projet Pilote Horizon 2020



« [...] autorise toute exploitation de l'œuvre, y compris à des fins commerciales, ainsi que la création d'œuvres dérivées, [...], à condition de l'attribuer à son l'auteur en citant son nom [...] »

« [...] vous renoncez (dans la mesure du possible) à tous les droits d'auteur, droits voisins et/ou connexes que vous détenez sur votre œuvre [...] »

<http://creativecommons.fr/licences/>



Data Description

Citation: Malikova, IN et al. (2008): (Table 2) Contents of labile mercury forms in soils from the area of the Bol'shoye Yarovoe lake. doi:10.1594/PANGAEA.782758, Supplement to: Malikova, IN; Ustinov, MT; Anoshin, GN; Badmaeva, ZhO; Malikov, Yul (2008): Mercury in soils and plants in the area of Lake Bol'shoye Yarovoe. Translated from Geologiya i Geofizika, 2008, 49(1), 59-66, Russian Geology and Geophysics, 49(1), 46-51, doi:10.1016/j.rgg.2007.12.004

Abstract: We examined mercury distribution in soils and plants of the catchment area of the Bol'shoye Yarovoe lake with the Altaikhimprom chemical industrial complex localized on its shore. Data on mercury contents show considerable mercury flow into environment, though the industrial complex and its waste do not pose serious threat. Revealed pollution in the immediate vicinity of the industrial complex including that due to atmospheric transport, as well as elevated content of mercury in hydrochloric and alkaline extracts suggest its intense migration from the waste. Therefore, it is necessary to make an inventory of the industrial-complex waste and perform a detailed study of the lake ecosystem as a model unit.

Project(s): Archive of Ocean Data (ARCOD) 🔍

Coverage: Median Latitude: 52.884333 \* Median Longitude: 78.592000 \* South-bound Latitude: 52.832000 \* West-bound Longitude: 78.552000 \* North-bound Latitude: 52.915000 \* East-bound Longitude: 78.652000 Minimum DEPTH, sediment/rock: 0.0 m \* Maximum DEPTH, sediment/rock: 1.0 m

Event(s): MUABM08-03-12 🔍 \* Latitude: 52.915000 \* Longitude: 78.572000 \* Location: Altay Territory, Russia 🔍 MUABM08-03-4 🔍 \* Latitude: 52.832000 \* Longitude: 78.552000 \* Location: Altay Territory, Russia 🔍 MUABM08-03-8 🔍 \* Latitude: 52.906000 \* Longitude: 78.652000 \* Location: Altay Territory, Russia 🔍

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	Event label	Event				Metadata
2	LONGITUDE 🔍	Longitude				Geocode
3	LATITUDE 🔍	Latitude				Geocode
4	DEPTH, sediment/rock 🔍	Depth	m			Geocode
5	Depth, top/min 🔍	Depth top	m	Malikova, IN 🔍		
6	Depth, bottom/max 🔍	Depth bot	m	Malikova, IN 🔍		
7	Soil composition 🔍	Soil comp		Malikova, IN 🔍		
8	Soil horizon 🔍	Soil hori		Malikova, IN 🔍		
9	Mercury 🔍	Hg	mg/kg	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	total
10	Mercury 🔍	Hg	mg/kg	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by bidistillate H2O
11	Mercury 🔍	Hg	%	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by bidistillate H2O, of total
12	Mercury 🔍	Hg	mg/kg	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by 1N HCl
13	Mercury 🔍	Hg	%	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by 1N HCl, of total
14	Mercury 🔍	Hg	mg/kg	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by ammonium acetate buffer
15	Mercury 🔍	Hg	%	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by ammonium acetate buffer, of total
16	Mercury 🔍	Hg	mg/kg	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by 0.1N NaOH
17	Mercury 🔍	Hg	%	Malikova, IN 🔍	Atomic absorption spectrometry (AAS) of cold vapors 🔍	leached by 0.1N NaOH, of total

License: Creative Commons Attribution 3.0 Unported

Size: 143 data points

Download Data

Download dataset as tab-delimited text (use the following character encoding: ISO-8859-1: ISO Western (PANGAEA default) 🔽)

# Restrictions au partage

- Confidentialité des informations commerciales ou industrielles,
- Confidentialité des données à caractère personnel
  - cf [CNIL](#) / [CIL du CNRS](#) : consentement, anonymisation des données...
- Droits de propriété intellectuelle (des tiers),
- Secret de la défense nationale et sécurité publique
- Autres : protection de l'environnement...

(d'après Leobet, 2013 ; Commission européenne, 2016)

# Comment partager ?

- Attribuer des identifiants pérennes
  - Pour les données de recherche
  - Pour les chercheurs

## Projet Pilote Horizon 2020

*Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, version 2.1 fev 2016*

« Where possible, contributors should also be uniquely identifiable, and data uniquely attributable, through identifiers which are persistent, non-proprietary, open and interoperable (e.g. through leveraging existing sustainable initiatives such as ORCID for contributor identifiers and DataCite for data identifiers). »

# Comment partager ?

- Qu'est-ce qu'un identifiant pérenne (Persistent identifier ou PID) ?
  - Un code unique (une chaîne de caractères) qui est assigné à un objet ou une personne de façon permanente
  - Disponible et gérable à long terme ; il ne changera pas si l'objet numérique est renommé ou déplacé

<http://www.inist.fr/mesdonnees>

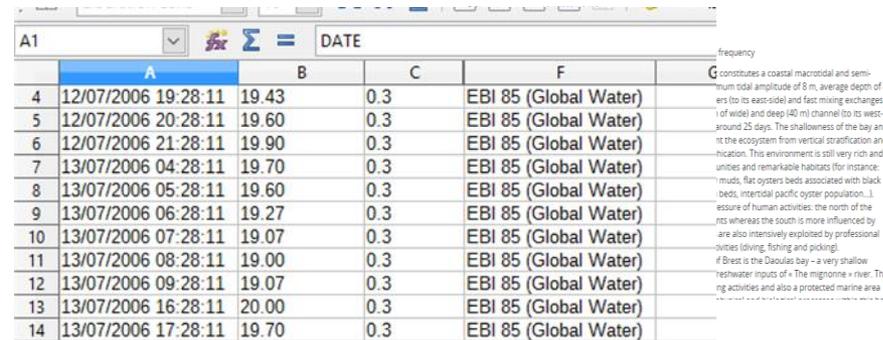
 The page cannot be found

The page you are looking for might have been removed, had its name changed, or is temporarily unavailable.

.....

HTTP 404 - File not found  
Internet Explorer

doi:10.17882/42493



A1	DATE	frequency
4	12/07/2006 19:28:11	19.43
5	12/07/2006 20:28:11	19.60
6	12/07/2006 21:28:11	19.90
7	13/07/2006 04:28:11	19.70
8	13/07/2006 05:28:11	19.60
9	13/07/2006 06:28:11	19.27
10	13/07/2006 07:28:11	19.07
11	13/07/2006 08:28:11	19.00
12	13/07/2006 09:28:11	19.07
13	13/07/2006 16:28:11	20.00
14	13/07/2006 17:28:11	19.70

SEANOE Sea scientific open data edition

Marine environmental station database of Daoulas bay

Publication date: 2016-02-18  
Temporal extent: 2006-07-12-2015-12-31  
Author(s): Petton Sébastien<sup>1</sup>, Le Roy Valérie<sup>1</sup>, Bellec Gwenaél<sup>1</sup>, Oukeau Isabelle<sup>1</sup>, Le Souchu Patrick<sup>1</sup>, Pouzineau Stéphane<sup>1</sup>  
Affiliation(s): 1: Ifremer, France

Click to download the data

DATA



# Comment partager ?

- Identifiants pour les données de recherche
  - Rendre visible et accessible les données de la recherche au même titre que les publications
  - Citer les données de la recherche
  - Lier les données de la recherche aux publications
- Principaux systèmes d'identifiants pour les données

Système	Nom développé	Exemple
<b>Handle</b>	Handle	<a href="http://hdl.handle.net/11403/prax000824">hdl.handle.net/11403/prax000824</a>
<b>PURL</b>	Persistent Uniform Resource Locator	<a href="http://purl.org/phylo/treebase/phyloids/matrix/TB2:M2610">purl.org/phylo/treebase/phyloids/matrix/TB2:M2610</a>
<b>DOI*</b>	Digital Object Identifier	doi:10.1594/PANGAEA.726855
<b>ARK</b>	Archival Resource Key	ark:/b7272/q6td9v7j
<b>ePIC*</b>	European Persistent Identifier Consortium	<a href="http://hdl.handle.net/11304/3339d230-b988-11e3-8cd7-14feb57d12b9">http://hdl.handle.net/11304/3339d230-b988-11e3-8cd7-14feb57d12b9</a>

\*identifiant pérenne basé sur le système Handle

## Autres identifiants pérennes

- URN : NBN National Bibliographic Numbers (ex : urn:nbn:nl:ui :32-424171)
- LSID, Life Science Identifier

# Comment partager ?

**Merritt** Logout | Contact UC3 | Help  
Logged in as Guest

Collection home   Change collection ▾ **ARK**

Object: [ark:/b7272/q6td9v7j](http://n2t.net/ark:/b7272/q6td9v7j)  
Merritt > Collection: [UCSF Datashare](#) > Object: [ark:/b7272/q6td9v7j](#)

Laurance, Megan. **Re-analysis of microarray data from rapamycin resistant DLBCL cell lines** (2013)

**permanent link:** <http://n2t.net/ark:/b7272/q6td9v7j>  
**primary identifier:** [ark:/b7272/q6td9v7j](http://n2t.net/ark:/b7272/q6td9v7j)  
**title:** Re-analysis of microarray data from rapamycin resistant cell lines  
**creator:** Laurance, Megan  
**date:** 2013  
**local id:** qzazilygds; [doi:10.7272/Q6TD9V7J](https://doi.org/10.7272/Q6TD9V7J)  
**last modified:** 2013-12-03 10:54 AM UTC  
**created:** 2013-10-22 06:45 PM UTC  
**total size** (all versions): 13.9 MB  
**total versions:** 5

[Download object](#)

**Object Lookup**  
 [Go](#)  
Search object author, title, date or identifier.

**Current Version**

Version 5: 2013-12-03 10:54 AM UTC

- [GSE27255\\_iReport\\_stats\\_QC\\_results.pdf](#)
- [GSE27255\\_stats\\_and\\_QC\\_details.pdf](#)
- [mrt-datacite.xml](#)
- [Rapamycin\\_v\\_Sensitive\\_DLBCL\\_DEGs.xlsx](#)

1 more files

**Prior Versions**

Version 4: 2013-11-20 01:59 PM UTC  
Version 3: 2013-11-01 04:14 PM UTC  
Version 2: 2013-10-31 09:50 AM UTC  
Version 1: 2013-10-22 06:45 PM UTC

<https://merritt.cdlib.org/m/ark%253A%252Fb7272%252Fq6td9v7j>

# Comment citer les données ?

Une option : Format standard de DataCite (« Joint Declaration of Data Citation Principles » FORCES 11-2013 <https://www.force11.org/datacitation>)

**Creator** (PublicationYear): Title. **Publisher**. Identifier

**Hogarth, Luke; Burkett, Brendan, J; McKean, Mark** (2015): Activity Profiles and Fitness in Tag Football; **University of the Sunshine Coast**. doi:[10.4227/39/55E7917B2C43E](https://doi.org/10.4227/39/55E7917B2C43E)

**Ewers, Rob** (2016): SAFE sampling stations. **Zenodo**. doi: [10.5281/zenodo.46173](https://doi.org/10.5281/zenodo.46173)

# Lien entre les jeux de données et la publication par les identifiants pérennes

Entrepôts de données

Journal of Biogeography (J. Biogeogr.) (2012) 39, 1217–1234

ORIGINAL  
ARTICLE

## Did geckos ride the Palawan raft to the Philippines?

Cameron D. Siler<sup>1\*</sup>, Jamie R. Oaks<sup>1</sup>, Luke J. Welton<sup>1</sup>, Charles John C. Swab<sup>1</sup>, Arvin C. Diesmos<sup>2</sup> and Rafe M. Brown<sup>1</sup>

<sup>1</sup>Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, USA,  
<sup>2</sup>Herpetology Section, Zoology Division, Philippine National Museum, Dinalup City, Palawan

### ABSTRACT

**Aim** We examine the genetic diversity within the lizard genus *Gerrhonotus* on Philippine islands to understand the role of geography and geology in shaping species diversity in this group. We test multiple biogeographic hypotheses.

#NEXUS



About

For researchers

For organizations

Nexus Alignment File

```
Seq Begin data;
Init      Dimensions ntax=172 nchar=1682;
200      Format datatype=dna gap=- missing=?;
(Ma      Matrix
gru      Cyrt_tioamanensis_WMalaysia_LSUHC3771
infe     ATAAACCCGCTTATCTGACTCACACTAATCACAGTCTGCTAACTAGCACAAATTATCACAAATATCGAGTACCCAT
like     AACAGAAGCCGCAACAAAATATTTCTTAATCCAAGTAAACCGCCGCCACTATACTCCTGCTAGCAAGCATTATAAG
no       CAACCCCAACAACAACACTTCTAATCACAAACGGCAATTATATTAATACTAGGCGCCGCCCTACATACATATGAT
con      CTATACATAACCCAAAACCACCTAAACCACCAAACCCTACTATCACTGGGCCTACTATCAGCCATCATTGGTGGAA
yses     AATTGCTCTTAGTCCAAGCCTTACCACCCTAACAATAGTAACATAACAATAACAATGGCCACCTTCCTCCC
lack     TGCTACTATAAAAAATATGACCACCATAGGAACTGCCTGACCGCTATCACCTGCATCACTCACTGCCACAATGCT
uals     CAGGACTCCCCTTTTCGCCACCGTACTGCTACTAGCAAGTCTGCCAAGCTTATTTTTCTATATCCGACTGGCCT
we       ATCGCCCAATAATCACAGCATCCACCCTACTTCTACCTATAACAGTAGTACTATATACCGTAACCTAG??????
uals
dat
```

Dryad (doi:10.5061/dryad.7f327q53).

Partitioned Bayesian analyses were conducted in MrBAYES 3.1.2 (Ronquist & Huelsenbeck, 2003). The mitochondrial

Checksum (MD5): 547193153065d8828657cb3ca320b473

### MOMARSAT2014

#### DONNÉES PUBLIÉES

Legrand Julien, Sarradin Pierre-Marie, Cannat Mathilde (2015). **Buoy management observatory, 2013-2014**. Seanoe. <http://dx.doi.org/10.17882/39755>

Legrand Julien, Sarradin Pierre-Marie, Cannat Mathilde (2015). **Buoy management observatory, 2013-2014**. Seanoe. <http://dx.doi.org/10.17882/39759>

Legrand Julien, Sarradin Pierre-Marie, Cannat Mathilde (2015). **COSTOF, technical the EMSO-Azores observatory, 2013-2014**. Seanoe. <http://dx.doi.org/10.17882/39>

#### BIBLIOGRAPHIE

Matabos Marjolaine, Cuvelier D., Brouard J., Shillito B., Ravaux J., Zbinden M., Barthelemy D., Sarradin Pierre-Marie, Sarradin Jozee (2015). **Behavioural study of two Hydrothermal crustacean decapods: Mirocaris fortunata and Segonzacia mesatlantica, from the lucky strike vent field (mid-Atlantic ridge)**. *Deep-sea Research Part II-topical Studies In Oceanography*, 121, 146-158. Publisher's official version : <http://dx.doi.org/10.1016/j.dsr2.2015.04.008> , Open Access version : <http://archimer.ifremer.fr/doc/00266/37714/>

Sarradin Jozee, Legendre Pierre, De Busserolles Fanny, Fabri Marie-Claire, Guilini Katja, Ivanenko Viatcheslav N., Morineaux Marie, Vanreusel Ann, Sarradin Pierre-Marie (2015). **Biodiversity patterns, environmental drivers and indicator species on a High-temperature Hydrothermal edifice, mid-Atlantic ridge**. *Deep-sea Research Part II-topical Studies In Oceanography*, 121, 177-192. Publisher's official version : <http://dx.doi.org/10.1016/j.dsr2.2015.04.013> , Open Access version : <http://archimer.ifremer.fr/doc/00266/37713/>

Laes Agathe, Legrand Julien, Sarradin Pierre-Marie, Rolin Jean-Francois, Tanguy Virginie, Cathalot Cecile, Blandin Jerome (2015). **Long term in situ monitoring of total dissolved iron concentrations on the MoMAR observatory**. EGU2015 - European Geosciences Union General Assembly, 12 - 17 April 2015, Vienne.

Opderbecke Jan, Brisset Laurent, Boy Emilie (2015). **Infrastructures Marines et Numériques. Unité Systèmes Sous-Marins. Rapport d'activité 2014**.

Laes-Huon Agathe (2014). **Compte Rendu de Réunion**. CHEMINI - A050802.

Laes Agathe, Legrand Julien, Tanguy Virginie, Cathalot Cecile, Blandin Jerome, Rolin Jean-Francois, Sarradin Pierre-Marie (2014). **Monitoring à long terme des concentrations en fer total dissous sur l'Observatoire MoMAR / Long term in situ survey of total dissolved iron concentrations on MoMAR observatory**. Sea Tech Week 2014 - 9th Marine Science Conference - From 13th to 17th October 2014 - Brest, France.

#### DONNÉES ARCHIVÉES AU SISMER

Doi: [10.17600/14000300](https://doi.org/10.17600/14000300) TGIR Flotte

#### DONNÉES ARCHIVÉES AU SISMER

##### TEMPERATURE DE SURFACE COLLECTÉES EN ROUTE - THERMOMETRE (1 Fichier)

Fichier	Niveau de traitement	Taille(Ko)	Format
<a href="#">Libre accès</a> 201400030045-intake-PP_SBE38.ths	DONNEES BRUTES	15284	THERMOMETRE NetCdf

##### DONNEES DE SALINITE DE SURFACE (1 Fichier)

Fichier	Niveau de traitement	Taille(Ko)	Format
<a href="#">Libre accès</a> 201400030045-hydrology-PP_SBE21.ths	DONNEES BRUTES	14617	TSG TECHSAS Netcdf

##### NAVIGATION TRAITEE DES NAVIRES IFREMER, SHOM, IRD ET IPEV (1 Fichier)

Fichier	Niveau de traitement	Taille(Ko)	Format
<a href="#">Libre accès</a> 201400030055.nvi	DONNEES CONTROLEES	43236	NAVIGATION .NVI 1999

##### BANQUE NATIONALE DE BATHYMETRIE MONOFAISCEAU (IFREMER) (3 Fichiers)

Fichier	Niveau de traitement	Taille(Ko)	Format
<a href="#">Accès protégé</a> 201400030045-sbdpth-PP_EA638.depth	DONNEES BRUTES	5812	BATHY. VERT. NetCDF
<a href="#">Accès protégé</a> 201400030045-sbdpth-PP_EA62C.depth	DONNEES BRUTES	5812	BATHY. VERT. NetCDF
<a href="#">Accès protégé</a> 201400030045-sbdpth-PP_EA612.depth	DONNEES BRUTES	5812	BATHY. VERT. NetCDF

Tout cocher

#### PRÉLEVEMENTS

Search:

Nom	Plongée	Equipement	Date	Localisation	Profondeur(m)	Description
<a href="#">MMR14-PL-01-TI1</a>	576-1	Seringue Titane	16/07/2014 07:16:41	37° 17.287' N 32° 16.532' W	1701	PRELEVEMENT M14flu001 déclenchée
<a href="#">MMR14-PL-01-TI2</a>	576-1	Seringue Titane	16/07/2014 07:22:13	37° 17.286' N 32° 16.533' W	1700	PRELEVEMENT M14flu002 déclenchée
<a href="#">MMR14-PL-01-TI3</a>	576-1	Seringue Titane	16/07/2014 07:25:50	37° 17.286' N 32° 16.532' W	1700	PRELEVEMENT M14flu003 déclenchée
<a href="#">MMR14-</a>	576-1	Seringue Titane	16/07/2014	37° 17.287' N	1700	PRELEVEMENT

# Lien entre les jeux de données et la publication par les identifiants pérennes



Russian Geology and Geophysics  
Volume 49, Issue 1, January 2008, Pages 46–51



**Mercury in soils and plants in the area of Lake Bol'shoe Yarovoe (*Altai Territory*)**  
I.N. Malikova<sup>a</sup>, M.T. Ustinov<sup>b</sup>, G.N. Anoshin<sup>a</sup>, Zh.O. Badmaeva<sup>a</sup>, Yu.I. Malikov<sup>a</sup>

<http://dx.doi.org/10.1016/j.rgg.2007.12.004> [Get rights and content](#)

**Abstract**  
We examined mercury distribution in soils and plants of the catchment area of Lake Bol'shoe Yarovoe with the Altaikhimprom chemical industrial complex situated on its shore. Data on mercury contents show a considerable mercury flow into the environment, though the industrial complex and its waste do not pose serious threat. The revealed pollution in the immediate vicinity of the industrial complex, including that due to atmospheric transport, as well as the elevated content of mercury in hydrochloric and alkaline extracts suggest its intense migration from the waste. Therefore, it is necessary to make up an inventory of the industrial-complex waste and perform a detailed study of the lake ecosystem as a model unit.

**Keywords**  
Mercury; catchment area of Lake Bol'shoe Yarovoe; soils; plants; waste

**References**  
Anoshin et al., 1996 G.N. Anoshin, I.N. Malikova, S.I. Kovalev  
**Mercury in soils of the Southern West Siberia**  
W. Bayens, R. Ebinghaus (Eds.), Global and Regional Mercury Cycles, Kluwer Academic Publishers, Dordrecht, the Netherlands (1996) pp. 475–489

**Mobile mercury species in soils of natural and natur...**  
2011, Russian Geology and Geophysics [more](#)

**Mercury contamination in the vicinity of a chlor-alkal...**  
2009, Science of The Total Environment [more](#)

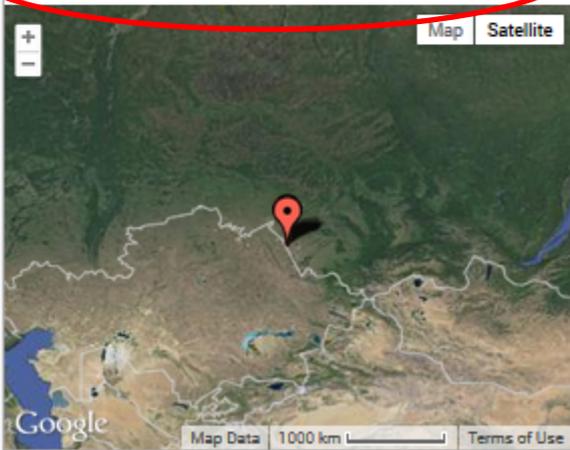
**Mercury spe**  
2008, Water Res [more](#)

[View more arti](#)

**Citing articles (2)**

**Related reference work articles**

 **PANGAEA<sup>®</sup> – Related Data**  
(Table 2) Contents of labile mercury forms in soils from the area of the Bol'shoe Yarovoe lake



Map Satellite

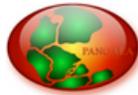
Google

Map Data 1000 km Terms of Use

Lien vers  
entrepôt de  
données

<http://www.sciencedirect.com/science/article/pii/S1068797107002362>

# Lien entre les jeux de données et la publication par les identifiants pérennes



**PANGAEA®**  
Data Publisher for Earth & Environmental Science

## Entrepôts de données

Always

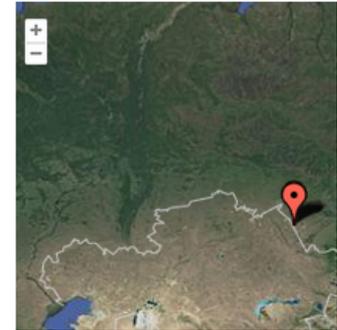
### Data Description

Show Map

**Citation:** Malikova, IN et al. (2008): (Table 2) Contents of labile mercury forms in soils from the area of the Bol'shoe Yarovoe lake. doi:10.1594/PANGAEA.782758.  
Supplement to: Malikova, IN; Ustinov, MT; Anoshin, GN; Badmaeva, ZhO; Malikov, Yul (2008): Mercury in soils and plants in the area of Lake Bol'shoe Yarovoe. *Translated from Geologiya i Geofizika*, 2008, 49(1), 59-66, *Russian Geology and Geophysics*, 49(1), 46-51. doi:10.1016/j.rgg.2007.12.004

**Abstract:** We examined mercury distribution in soils and plants of the catchment area of the Bol'shoe Yarovoe lake with the Altaikhimprom chemical industrial complex localized on its shore. Data on mercury contents show considerable mercury flow into environment, though the industrial complex and its waste do not pose serious threat. Revealed pollution in the immediate vicinity of the industrial complex including that due to atmospheric transport, as well as elevated content of mercury in hydrochloric and alkaline extracts suggest its intense migration from the waste. Therefore, it is necessary to make an inventory of the industrial-complex waste and perform a detailed study of the lake ecosystem as a model unit.

**Project(s):** [Archive of Ocean Data \(ARCOD\)](#)



### Data

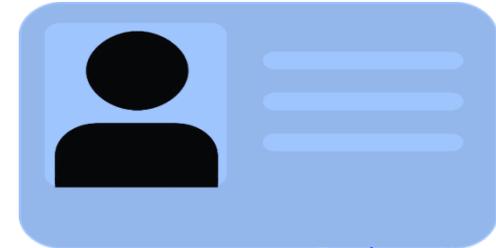
Download dataset as tab-delimited text (use the following character encoding: )

1 Event	2 Longitude	3 Latitude	4 Depth [m]	5 Depth top [m]	6 Depth bot [m]	7 Soil comp	8 Soil hori	9 Hg [mg/kg] (total)	10 Hg [mg/kg] (leached by bicarbonate H2O)	11 Hg [%] (leached by bicarbonate H2O, ...)	12 Hg [mg/kg] (leached by 1N HCl)	13 Hg [%] (leached by 1N HCl of total)	14 Hg [mg/kg] (leached by ammonium)
MUABM08-03-4	78.552	52.832	0.025	0.00	0.05	medium loamy chemozemlic solonetz	A1	0.071	0.0020		2.8	0.0027	3.8
MUABM08-03-4	78.552	52.832	0.090	0.05	0.13	same	A2	0.064	0.0016		2.5	0.0078	12.2
MUABM08-03-4	78.552	52.832	0.215	0.13	0.30	same	B	0.038	0.0026		6.8	0.0075	19.7
MUABM08-03-4	78.552	52.832	0.395	0.30	0.49	same	B2	0.022	0.0015		6.8	0.0084	38.2
MUABM08-03-4	78.552	52.832	0.645	0.49	0.80	same	Ck	0.047	0.0009		1.9	0.0064	13.4
MUABM08-03-8	78.652	52.906	0.020	0.00	0.04	meadow saltterra (anthropogenically disturbed)	A1	0.089	0.0049		5.5	0.0061	6.8
MUABM08-03-8	78.652	52.906	0.120	0.04	0.20	same	II	0.050	0.0033		6.6	0.0062	12.4
MUABM08-03-8	78.652	52.906	0.385	0.20	0.57	same	III	0.060	0.0005		<1.0	0.0061	10.2
MUABM08-03-8	78.652	52.906	0.950	0.90	1.00	same	IV	0.035	0.0017		4.8	0.0086	24.6
MUABM08-03-12	78.572	52.915	0.020	0.00	0.04	technogenic chemozemlic solonetz	Ad	0.320	<0.0005		<0.1	0.0051	1.6
MUABM08-03-12	78.572	52.915	0.120	0.04	0.20	same	B	0.045	0.0015		3.3	0.0047	10.4

<http://doi.pangaea.de/10.1594/PANGAEA.782758>

# Comment partager ?

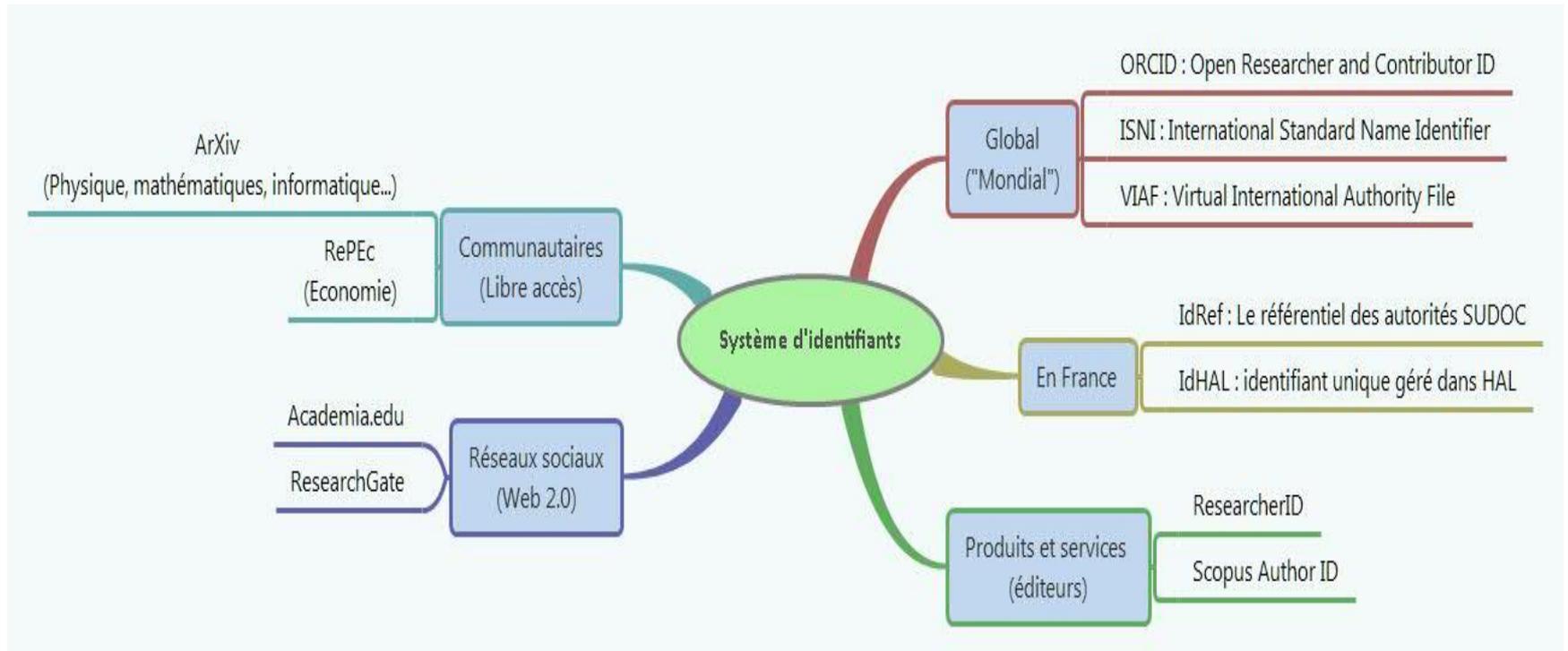
- Identifiant chercheur / auteur



[Pixabay](#), CC0

- Désambigüiser (homonymie) et associer à un identifiant toutes les formes d'un nom
- Trouver et citer toute la production d'un chercheur sous une seule entrée
  - ➡ augmenter sa visibilité
- Eviter au chercheur de ressaisir les informations le concernant (import ou export entre des bases)

# Comment partager ?



Quelques systèmes d'identifiants pour les chercheurs/auteurs

# Un exemple : ORCID

- Open Researcher and Contributor ID
- Organisation internationale à but non-lucratif (siège Etats-unis)
- Gestion d'un répertoire d'identifiants chercheurs (toutes disciplines) et données des profils des chercheurs inscrits
  - Inscription gratuite
  - Accès au profil du chercheur est public, restreint ou privé
- Possibilité d'établir une connexion (sous réserve de votre autorisation)
  - ResearcherID (Thomson-Reuters) et Scopus Author ID (Elsevier)
  - Agences de financement (Wellcome Trust, CE..)
  - Entrepôts de données (Dryad, Figshare)
- Service « auto-update » avec agences d'attribution de DOI, CrossRef et DataCite



<http://orcid.org/>

## Mogens Sandfær

### ORCID ID

 [orcid.org/0000-0001-8436-5346](http://orcid.org/0000-0001-8436-5346)

### Also known as

M. Sandfaer, M. Sandfær, Mogens Sandfaer

### Country

Denmark

### Email

[ms@dtic.dtu.dk](mailto:ms@dtic.dtu.dk)

[mosa@dtu.dk](mailto:mosa@dtu.dk)

[sandfaer@gmail.com](mailto:sandfaer@gmail.com)

### Other IDs

Technical University of Denmark C

WIS: 2225

Scopus Author ID: 34168257000

## Biography

Head of Bibliometrics and Data Management at the Technical University of Denmark

### Works (2)

↑↓ Sort

#### Approach for a joint global registration agency for research data

Information Services and Use

2009 | journal-article

DOI: [10.3233/ISU-2009-0595](https://doi.org/10.3233/ISU-2009-0595)

URL: <http://www.scopus.com/inward/record.url?eid=2-s2.0-69549113049&partnerID=M...>

Source: Scopus to ORCID

Preferred source

#### Charging, Paying and Copyright - Information Access on Open Networks

journal-article

DOI: [10.1629/0949](https://doi.org/10.1629/0949)

Source: Mogens Sandfær

Preferred source

<http://orcid.org/0000-0001-8436-5346>

# Références (1)

- **Entrepôts de données**

- Aventurier P. (2013). Données ouvertes de la recherche : nouvelles pratiques de publication et de partage. Cours Enssib (Cours). <http://prodinra.inra.fr/record/217160>
- European Commission (2016). Guidelines on Data Management in Horizon 2020. version 2.1. [http://www.ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://www.ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)
- Data Seal of Approval. <http://www.datasealofapproval.org/en/information/guidelines/>
- Pampel H et al (2013). Making Research Data Repositories Visible: The re3data.org Registry. PLoS ONE 8(11):e78080. <http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0078080&representation=PDF>
- re3data - Registry of Research Data Repositories. <http://www.re3data.org>
- Swauger S, Vision TJ (2015). What factors influence where researchers deposit their data? A survey of researchers submitting to data repositories. International Journal of Digital Curation 10(1): 68–81. <http://dx.doi.org/10.2218/ijdc.v10i1.289>

# Références (2)

- **Licences**

- Ball A (2012). How to License Research Data. DCC How-to Guides. Edinburgh / Digital Curation Centre. <http://www.dcc.ac.uk/resources/how-guides/license-research-data>
- Creative Commons. A propos des licences. <https://creativecommons.org/licenses/>
- Dedieu L, Fily MF (2015). Rendre publics ses jeux de données scientifiques en 6 points. Montpellier (FRA): CIRAD, 6 p. <http://url.cirad.fr/ist/rendre-publics-ses-donnees>
- Etalab (2011). Le blog de la mission Etalab. Licence ouverte. <https://www.etalab.gouv.fr/licence-ouverte-open-licence>
- Fily MF (2015). Connaitre et utiliser les licences Creative Commons en 6 points. Montpellier (FRA) : CIRAD, 11p. <http://coop-ist.cirad.fr/content/download/5804/42700/version/3/file/Coopist-licences-cc-vfinale.pdf>
- Open Data Commons. <http://opendatacommons.org/>

- **Identifiants données de la recherche**

- Lee DJ, Stvilia B (2012). Identifier Schemas and Research Data. Proceedings of the 75<sup>th</sup> American Society for Information Science and Technology, 28-31 october, Baltimore, USA. <https://www.asis.org/asist2012/proceedings/Submissions/311.pdf>

# Références (3)

- **Identifiants chercheurs**

- CCSD-CNRS (2014). Mon idHAL. <https://hal.archives-ouvertes.fr/page/mon-idhal>, consulté le 4 juin 2015
- Deboin MC (2015). Utiliser un identifiant chercheur pour gérer ses publications en 12 points. Montpellier (FRA): CIRAD, 8 p. <http://url.cirad.fr/ist/identifiant-chercheur>
- Elsevier. Scopus preview. <https://www.scopus.com/search/form/authorFreeLookup.url>
- Evrard AE, Erdmann C, Holmquist J et al (2015). Persistent, Global Identity for Scientists via ORCID. arXiv:1502.06274
- ORCID Connecting Research and Researchers. <http://orcid.org/>
- Thomson-Reuters (2015). Web of Science : ResearcherID. <http://www.researcherid.com/Home.action>

- **Autres**

- Leobet M (2013). Le triangle juridique autour de la donnée géographique. <http://rddb.cnrs.fr/spip.php?article124>



# BONNES PRATIQUES ELABORATION D'UN DMP

---

## Conserver les données



# Conserver les données

## Doit-on conserver à long terme toutes les données générées au cours d'un projet de recherche ?

- Quelques critères de sélection des données
  - Est-ce que les données peuvent être réutilisées ?
  - Doivent-elles être conservées pour des raisons juridiques ou politiques (financeurs ou institutions) ?
  - Doivent-elles être conservées pour leurs valeurs potentielles ?
  - Considérer le coût - bénéfice

(DCC, 2014 ; NERC ; University of Bristol, 2015)

Conservation des données **MAIS AUSSI** métadonnées et logiciels, algorithmes

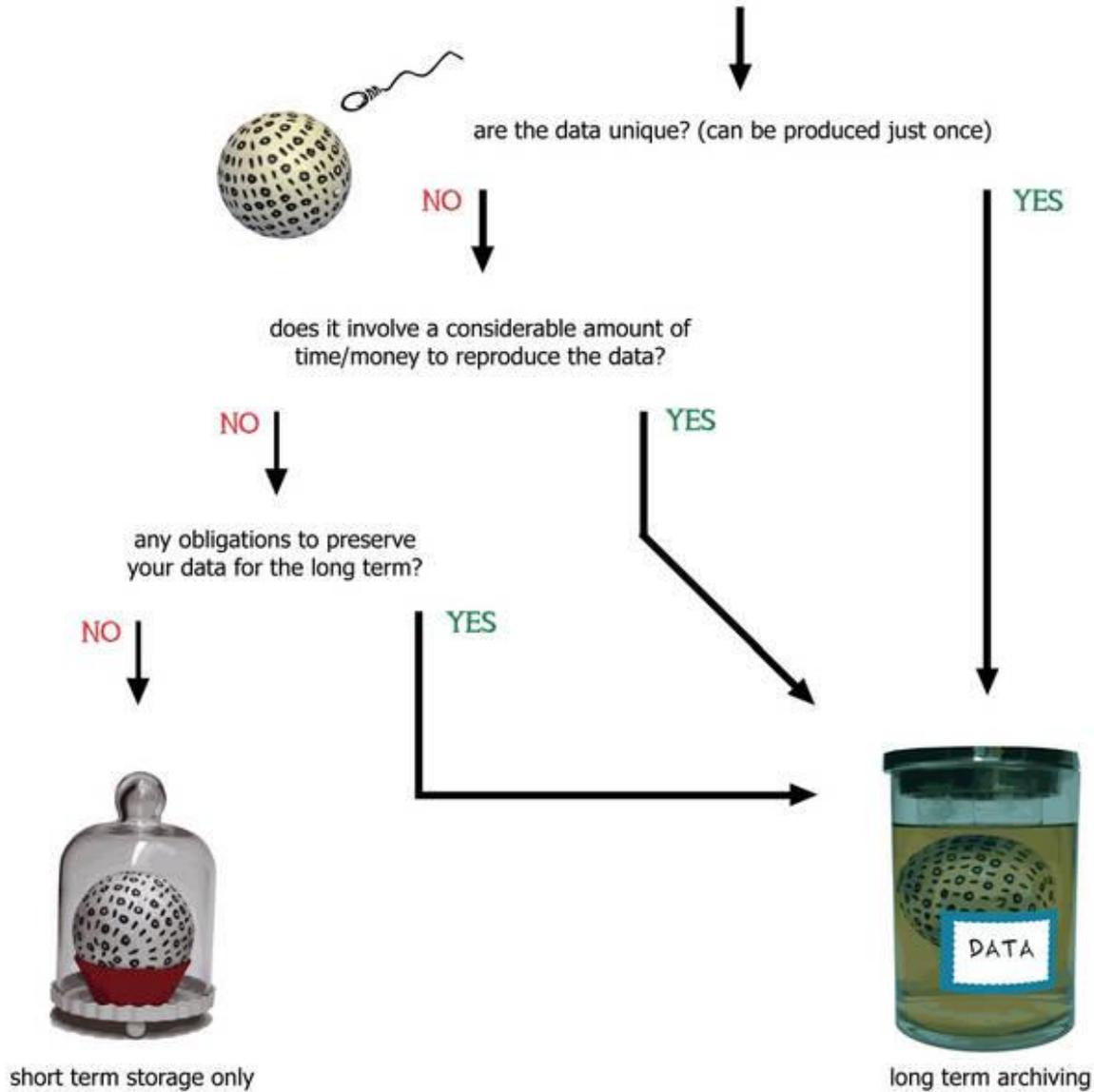
# Conserver les données

- Quels formats?
  - FACILE : Outil de validation du Format d'Archivage du CINES par analyse et Expertise (<http://facile.cines.fr/>)
  - [Guide Méthodologique pour l'archivage des bases de données](#) (Prat, 2013)



- Délai de conservation :
  - Association des archivistes français - Section Aurore. [Référentiel de gestion des archives de la recherche](#)
  - Recommandation ESF (2011) :  
« *Original scientific or scholarly research data should be documented and archived for a substantial period (at least 5 years, and preferably 10 years)* »

should I consider storing my data in a long term archive?



[Essentials 4 Data Support by Research Data Netherlands](#) (License CC-BY-SA)

# Références

- Digital Curation Centre (2014). Five steps to decide what data to keep: a checklist for appraising research data v.1. Edinburgh: Digital Curation Centre.  
<http://www.dcc.ac.uk/sites/default/files/documents/publications/Five%20Steps%20to%20decide%20what%20data%20to%20keep.pdf>
- European Science Foundation (2011). The European Code of Conduct for Research Integrity. <http://www.esf.org/coordinating-research/mo-fora/research-integrity.html>
- NERC. NERC Data Value Checklist. <http://www.nerc.ac.uk/research/sites/data/policy/data-value-checklist/>
- Prat P et Bécharde L (2013). Guide Méthodologique pour l'archivage des bases de données. [https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/ca323bcc-804a-43e3-822f-5d18fa304dff/GM\\_archivage\\_BDD.pdf](https://alfresco.cines.fr/alfresco/d/d/workspace/SpacesStore/ca323bcc-804a-43e3-822f-5d18fa304dff/GM_archivage_BDD.pdf)
- Research Data Netherlands. Essential 4 Data Support. Data selection. <http://datasupport.researchdata.nl/en/start-de-cursus/iv-gebruiksfase/data-archiveren/selectie-van-data/>
- University of Bristol (2015). Research Data Evaluation Guide <http://data.bris.ac.uk/files/2014/02/Research-data-evaluation.pdf>



www.cnrs.fr

# Merci de votre attention

## Equipe Valorisation des données de la recherche

Service Analyser Valoriser  
Département de l'Offre de Services  
Inist - CNRS  
2, allée du Parc de Brabois  
54519 Vandœuvre-lès-Nancy  
tél : 03 83 50 46 32  
mèl : [contact-donneesrecherche@inist.fr](mailto:contact-donneesrecherche@inist.fr)

