



Une introduction aux données ouvertes de la recherche

Pascal Aventurier

► To cite this version:

Pascal Aventurier. Une introduction aux données ouvertes de la recherche. Journée d'étude " Les enjeux de l'Open Access pour la recherche ", Jul 2013, Montpellier, France. 23 p. hal-02803716

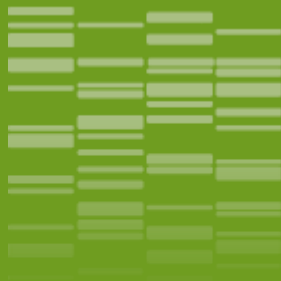
HAL Id: hal-02803716

<https://hal.inrae.fr/hal-02803716>

Submitted on 5 Jun 2020

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

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



Une introduction aux données ouvertes de la Recherche



Prêt pour un Data Cake ?

Data



Information



Presentation



Knowledge



EpicGraphic.com

<http://epicgraphic.com/data-cake/>



The delay in sharing research data is costing lives

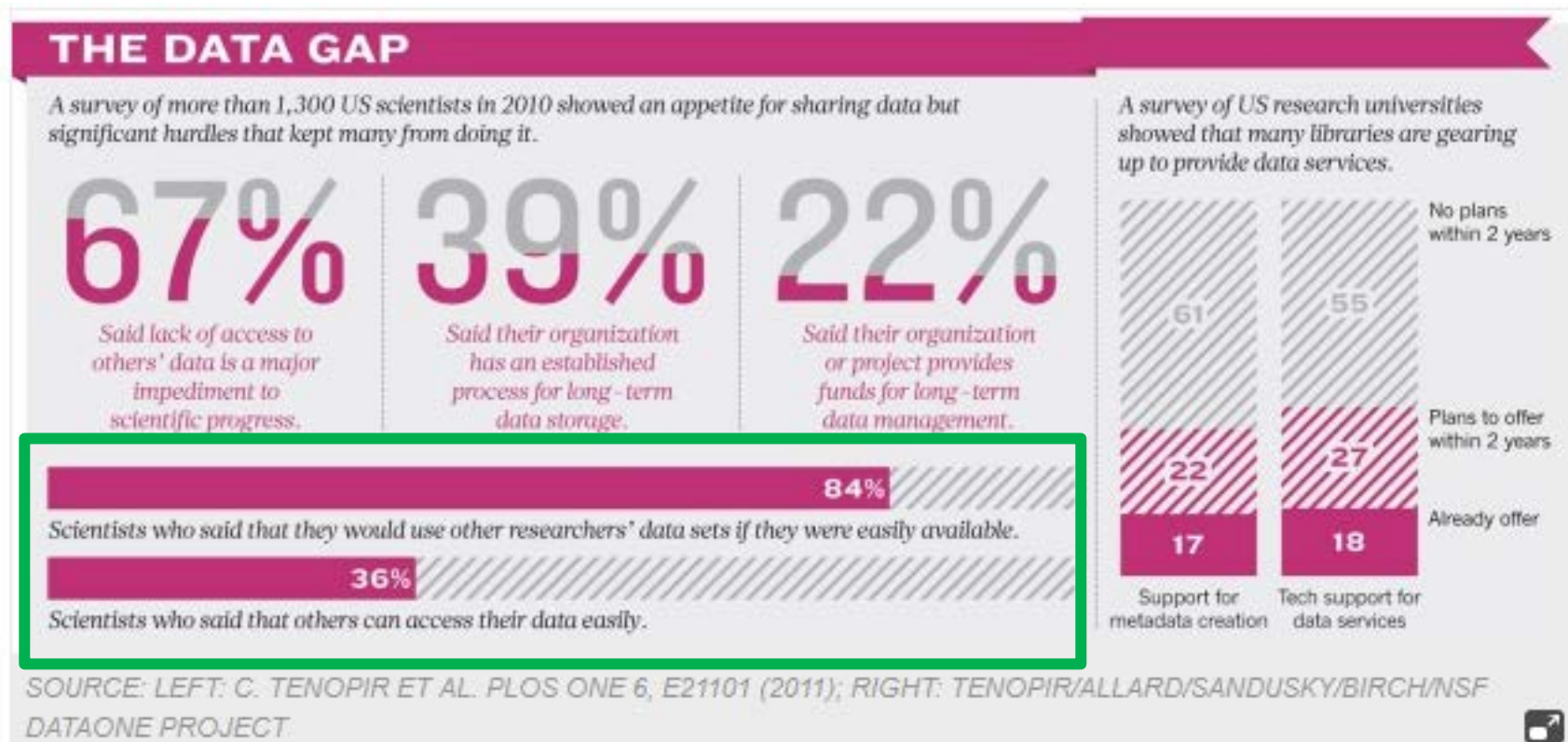
Josh Sommer

Nature Medicine **16**, 744 (2010) | doi:10.1038/nm0710-744

[PDF](#)[Citation](#)[Reprints](#)[Rights & permissions](#)[Article metrics](#)

It is not uncommon for potentially life-saving research data to be published years after being generated. But the setback to progress caused by the delay in releasing data is troublesome for people who selflessly participate in trials and desperately await new therapies. Scientists need to feel greater urgency to share their findings quickly, and they need additional avenues to facilitate this process.

84% des chercheurs aimeraient utiliser des jeux de données d'autres chercheurs



Source : <http://www.nature.com/news/publishing-frontiers-the-library-reboot-1.12664>

Définition de l'OCDE 2007

[Accueil de l'OCDE](#)[A propos](#)[Pays](#)[Thèmes](#)[Statistiques](#)[Salle de presse](#)

[OECD Home](#) › [Direction de la Science, de la technologie et de l'industrie](#) › [Politiques scientifiques et technologiques](#) › Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics

Politiques scientifiques et technologiques

> L'innovation dans la
science, la technologie et

Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics

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

<http://www.oecd.org/fr/sti/sci-tech/principesetlignesdirectricesdelocdepourlaccesauxdonneesdelarecherchefinanceesurfondspublics.htm>

- 
- **Données primaires ou brutes** : l'information enregistrée comme des notes, images, données d'observation, appartenant à un projet de recherche spécifique
 - **Données issues d'un traitement** : Analyses, descriptions, conclusions arrangées sur un support ou dans une publication
 - **Published data**: L'information distribuée aux personnes impliquées dans l'acquisition de données et leur administration

Site d'information sur les

Données de la Recherche

Rechercher

OK

[ACCUEIL](#) [NORMES](#) [POLITIQUES](#) [PROJETS](#) [MÉTIERS](#) [WEBOGRAPHIE](#) [ARCHIVES](#)

> Accueil

Partager Imprimer

Ce site, créé à l'initiative du ministère de l'Enseignement supérieur et de la Recherche, est une source d'informations sur les données de la recherche.

[Lire la suite](#)

Actualités

Call for expressions of interest : EC public consultation on open research data

2 juillet 2013 - Bruxelles, Belgique

Dear Sir, dear Madam, Open enquiry is at the heart of scientific endeavour, and rapid technological change has profound implications for the way that science is both conducted and communicated. (...)

[Lire la suite](#)

Charte Éthique & Big Data

La charte Éthique & Big Data est une co-construction d'acteurs majeurs, académiques et industriels, pour faciliter la création, la diffusion et l'utilisation des grands volumes de données (Big Data) (...)

Événements

International Information Conference on Search, Data Mining and Visualization
>14 et 15 Avril 2014 - Nice, France

Commodity, catalyst or change-agent ?
Data-driven transformations in research, education, business & society
>Du 24 au 27 février 2014 - San Francisco, États-Unis

APA International Conference on Digital Preservation and Development of Trusted Digital Repositories
>5 et 6 février 2014 - New Delhi, Inde

MTSR 2013 : 7th Metadata and Semantics Research Conference
>Du 19 au 22 novembre 2013 - Thessalonique, Grèce



MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE



Inist



Données de la recherche

Enjeux

- Mettre à disposition des données pour reproduire les expériences
- (éthique , qualité des recherches)
- Valorisation supplémentaire des travaux
- Utiliser librement les données produites par d'autres laboratoires
- Conditions d'archivages et de stockage des données

Reproduction des expériences

Ethique

Réutilisation

Valorisation

Bonnes conditions de production

Partage collaboration

Archivage

Interopérabilité

Data mining

- Revues sur les jeux de données

Data Journal



Scientific Data (Nature)
Geoscience Data Journal
CODATA Data Science
Journal Earth System
Science Data
Journal of Chemical &
Engineering Data (ACS
Publications)
Big Data
Biodiversity Data Journal
Genomics Data (Elsevier)
Journal of Open Archaeology
Data

Ne contiennent pas les données mais des articles de recherche décrivant les données

Les données doivent être accessibles dans des « répertoire de données »

Data journal (groupe Nature)



SCIENTIFIC DATA

Home | About | For Authors | Advisory and Editorial Board | Open Access | Contacts | FAQ

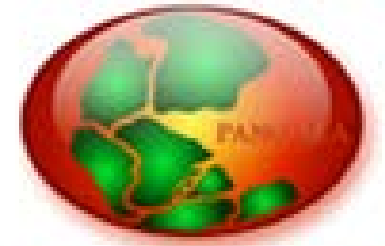
Helping you publish, discover,
and reuse research data

 Credit Credit, through a citable publication, for depositing & sharing your data	 Reuse Complete, curated & standardized descriptions enable the reuse of your data	 Quality Rigorous community based peer review
 Discovery Find datasets relevant to your research	 Open Promotes & endorses open science principles & available to all through a Creative Commons license	 Service In-house curation, rapid peer review & publication of your data descriptions

Welcome to *Scientific Data*

Data Repositories

- [Dryad](#)
- [Figshare](#)
- [Pangea](#) (données sur la terre)
- [ICPSR](#) (Inter-university Consortium for Political and Social Research)



-
- Dryad : [Joint Data Archiving Policy \(JDAP\)](#)

Dryad Joint Data Archiving Policy (JDAP)

- Accord avec des journaux “classiques” qui exigent le dépôt des données dans Dryad

- Au minimum les données ayant servi à l'article

Journaux - 30

environ dont

BMC Ecology

BMC Evolutionary

Genetics

Journal of Heredity

Nature

Nucleic Acids

Research

Paleobiology

PLOS

Science

Métadonnées - Licence CCO

DOI

titre

Journal (vol Issue), Année

Auteurs – auteur contact

Résumé - Mots clés

Nom taxonomique

Localisation géo

Titre et description des données

Embargo



DataCite

*Recense plus de 480
sites de stockage de
données*

Registrations by Allocators

Registrations by Datacentres

Registrations by Prefixes

Resolutions by Month

Allocator	DOI Registrations				Metadata Uploads				Metadata Ratio
	All Time	This Year	Last 30 Days	Last 7 Days	All Time	This Year	Last 30 Days	Last 7 Days	
ANDS - Australian National Data Service	2 617	102	7	0	1 711	102	7	0	65%
BL - The British Library	28 162	4 445	1 480	483	28 055	4 446	1 480	483	99%
CDL - California Digital Library	420 092	174 293	129 559	22 717	416 689	174 414	129 563	22 719	99%
CISTI - Canada Institute for Scientific and Technical Information	2 084	158	80	77	2 066	158	80	77	99%
CRUI - CRUI2011	7 697	4 596	4 107	0	4 570	1 469	980	0	59%
DATA CITE - DataCite	7	0	0	0	7	0	0	0	100%
DELFT - TU Delft Library	7 190	4 498	123	30	6 914	4 500	125	32	96%
DK - Technical Information Center of Denmark	2 539	514	130	2	2 538	514	130	2	99%
ETHZ - ETH Zurich	411 106	32 404	1 270	440	411 103	79 837	17 044	15 665	99%
GESIS - GESIS - Leibniz Institute for the Social Sciences	7 455	1 400	1 034	21	7 194	1 534	1 034	21	96%
INIST - Institute for Scientific and Technical Information	394	294	29	16	394	294	29	16	100%
OSTI - Office of Scientific and Technical Information (OSTI), US Department of Energy	474	30	7	6	474	30	7	6	100%
PURDUE - Purdue University Library	2 162	1 724	39	3	2 161	1 724	39	3	99%
SND - Swedish National Data Service	2	1	0	0	2	1	0	0	100%

Re3data.org

<http://www.re3data.org/> (20 mai 2013)



Home Search Suggest FAQ About Schema Contact Imprint

re3data.org launched

Posted on May 28, 2013 by re3data.org team

An increasing number of universities and research organisations are starting to build research data repositories to allow permanent access in a trustworthy environment to data sets resulting from research at their institutions. Due to varying disciplinary requirements, the landscape of research data repositories is very heterogeneous. This makes it difficult for researchers, funding bodies, publishers, and scholarly institutions to select an appropriate repository for storage of research data or to search for data.

The re3data.org registry allows the easy identification of appropriate research data repositories, both for data producers and users. The registry covers research data repositories from all academic disciplines. Information icons display the principal attributes

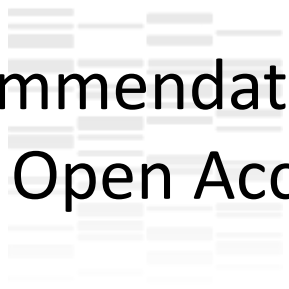
4

PARTNERS



- Australian National Data Service; iCordi - European Commission (FP7); RDA/US activity (National Science Foundation)
- Accélération internationale de l'innovation "data-driven" et de leur exploitation (discover) en facilitant le partage des données
- Utilisation et ré-utilisation des données, harmonisation des standards et "discoverability"

Livrables : infrastructure, règles, utilisation, standards...



Recommendations of the participants of the workshop on Open Access to Research Data to the European Commission (21.2.2013)

We recommend that the European Commission should introduce mandatory data management plans which include a description of the data sets, information on data handling beyond project lifetime and possible repositories for depositing the data sets.



24

Aug

2018

Ten recommendations for libraries to get started with research data management

...

Offrir un plan de gestion des données

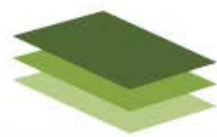
- S'engager dans le développement de standards de métadonnées
- Participer activement aux plans de gestion des données de votre institut
- Supporter le cycle de vie des données en fournissant des services pour le stockage, la « découverte » et l'accès permanent
- Mettre en place un catalogue des données
- Mettre en place des Identifiants pérennes

-...

Plan de gestion des données

- **MIT** <http://libraries.mit.edu/guides/subjects/data-management/>
- **Digital Curation Center**
<http://www.dcc.ac.uk/resources/data-management-plans>
- [DMP Online](#) un outil pour créer des plans personnalisés selon le contexte du financer de la recherche. Pour les chercheurs : exemples, guides et bonnes pratiques
- [Funders' data plan requirements](#) : résumés des attentes des financeurs
Summary of funders' expectations for data management and sharing plans
- [Checklist for a Data Management Plan](#)
- **Monash Univ**
<http://monash.edu/library/researchdata/index.html>

Outil pour la création d'un plan de gestion des données <https://dmp.cdlib.org>



DMPTool

Guidance and Resources for your Data Management Plan

[Contact Us](#) | [Get Started](#) | [Login](#)

[Home](#) [About DMP Tool](#) [DMP News](#) [My Plans](#) [Funder Requirements](#) [Help](#)



Create ready-to-use data management plans for specific funding agencies.

Photo courtesy of Argonne National Laboratory

The DMP Tool allows you to: 1 2 3 4

[Get Started!](#)

Data Management Plan: Sample Plan Created at the DataONE Best Practices Workshop - Santa Fe NM 7/2011
Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013

1. Types of data produced

All samples at Mauna Loa Observatory will be collected continuously from an intake located at five towers, a central tower and four towers located at compass quadrants. These data files will contain continuously measured CO2 concentrations, calibration standards, reference standards, daily check standards, and baselines. The sample files located at compass quadrants were used to examine the influence of source effects associated with wind directions. In addition to the CO2 data, we will record real-time data (wind speed and direction, temperature, humidity, precipitation, and cloud cover). Site conditions at Mauna Loa Observatory will also be coded and stored. The first data product will consist of 5-minute, 15-minute, hourly, daily, and monthly average atmospheric concentration of

[See a plan created with the DMP Tool](#)

Recent DMP News

[DMPTool's Top 10](#)

[Data Management Resources: Libguides](#)

[DMPTool2 Project - May 2013 Report](#)

[More news >](#)

Que contient un plan de gestion des données ?

<http://monash.edu/library/researchdata/advice/data-managment-brochure.pdf>

1. Copyright, Propriété intellectuelle (IP), éthique : comment sont-ils reliés à vos données de recherche ?
2. Où stocker les données numériques et les sauvegarder ?
3. Méthodes pour organiser les données et la documentation les décrivant
4. Partager vos données: au bon moment et avec les bonnes personnes
5. Optimiser l'impact de vos recherches en diffusant vos données et des informations de base
6. Prendre des dispositions pour les données après la fin de votre projet de recherche



Lier les données

- Utiliser des formats qui permettent de lier ses données avec les données d'autres sources
- Exemple : Données de recherche et thesaurus, lui-même relia à d'autres ressources (catalogue de documents)
- Pour
 - Enrichir son jeu de données
 - Construire de nouvelles interfaces intelligentes
 - Nouveaux modes de recherche



Conclusion

- Données n'appartiennent pas aux éditeurs : contexte différent des publications
- Des compétences à développer
- Droit et licence : un élément clé pour les données
- However, libraries are particularly well positioned to fill the data-management gap, says Michener. They have long experience helping faculty members and they are not likely to disappear, he says. “Libraries have a significant amount of trust capital.”*

* <http://www.nature.com/news/publishing-frontiers-the-library-reboot-1.12664>