



**HAL**  
open science

## Publier un Data Paper pour valoriser ses données

Dominique L'Hostis

► **To cite this version:**

Dominique L'Hostis. Publier un Data Paper pour valoriser ses données. Vers de nouveaux paradigmes pour l'édition scientifique? Dynamiques de validation et de légitimité, Université Fédérale Toulouse Midi-Pyrénées. Bordeaux, FRA., Mar 2016, Bordeaux, France. pp.47 slides. hal-02801694

**HAL Id: hal-02801694**

**<https://hal.inrae.fr/hal-02801694>**

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.

# Publier un Data Paper pour valoriser ses données

**Journée d'étude - URFIST Bordeaux**  
**Les nouveaux paradigmes éditoriaux – 18 Mars 2016**



# Plan

## Partager vs Publier ses données

## Différentes voies de diffusion des données

- Dépôt dans un entrepôt
- Matériel supplémentaire d'un article
- Data Paper

## Data Paper et Data Journal

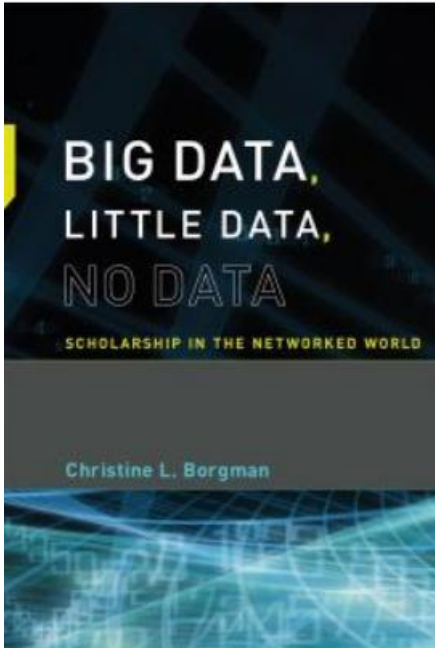
## Rôles et enjeux, freins et leviers pour les différents acteurs

# Data sharing – Data Publication ?

# Big Data, Little Data, No Data

Scholarship in the Networked World

By Christine L. Borgman



## Publier : 3 objectifs

Légitimation

Qualité, confiance

→ via le Peer Review, citation

Dissémination

Fonction essentielle

→ assurée par les éditeurs et les auteurs

Accès, préservation, curation

Disponibilité, accessibilité

→ rôles partagés : auteurs, éditeurs, bibliothèques

« Data Publication » ?

Data by itself has no value.

**The value of data lies in their use**

ist@inra

# Données partagées vs Données publiées



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

## F1000Research

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

# Différentes voies pour diffuser ses données

*ist@inra*

# Donner accès à ses données : plusieurs stratégies ...

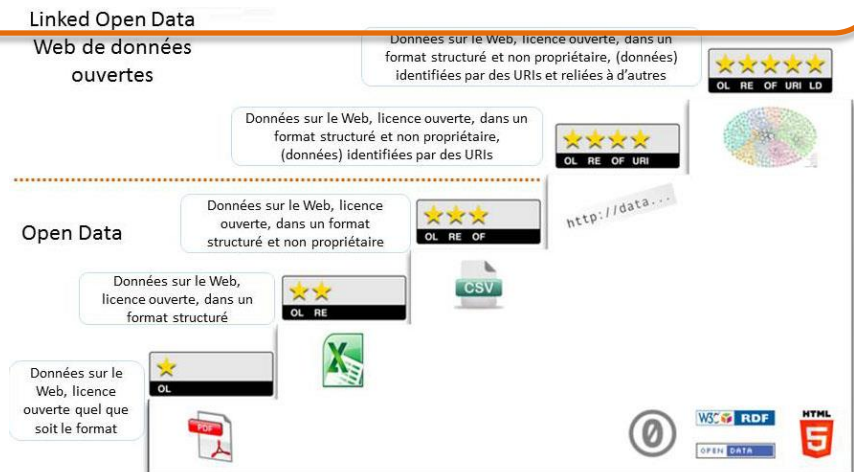


🌸 Déposer ses données dans un **entrepôt**

🌸 Fournir ses données sous la forme de **matériel supplémentaire** à la publication

🌸 Rédiger un **Data Paper**, publication scientifique spécifique décrivant les données, lié aux données et publié dans un journal.

🌸 Publier dans le **web des données** (linked data)



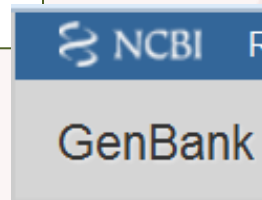
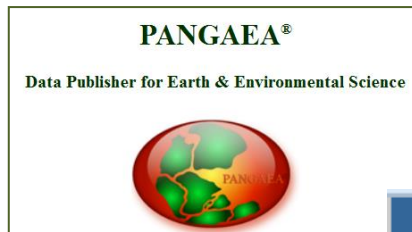
d'après

<http://5stardata.info/>

ist@inra



# Déposer dans un entrepôt



ist@inra

# Comment trouver un entrepôt ? (1/2)

Many many options for open'ing data



(Mounce, 2014)

## Entrepôts thématiques

- [Biosharing](#) (Biosciences)
- [Open Context](#) (Archéologie)
- [Data Ifremer](#) (Données marines)
- [DataOne](#) (Earth observation)
- [Pangaea](#) (Earth & Environmental Science)
- [Global Biodiversity Information Facility](#) (Biodiversité)
  - GBIF France : <http://portail.gbif.fr/>

## Entrepôts généralistes

- [Dryad](#)
- [Zenodo](#)
- [Figshare](#)
- [Dataverse](#)

# Comment trouver un entrepôt (2/2)

 Conseil : utiliser des annuaires ou répertoires



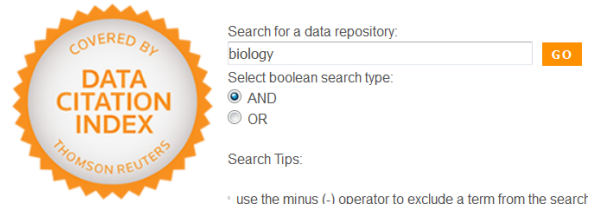
- [RE3DATA Registry of Research Data Repositories](#) (plus de 1400 enregistrements, 71 en France)
- [OpenDoar](#)
- [NIH Data Sharing Repositories](#)
- [Scientific Data Repositories and Datasets](#)
- [Wiki - Open Access Directory](#)
- [Base](#)
- [Research Data Australia](#)
- [Data Literature Interlinking Service](#)

Accès gratuit



- [Data Citation Index](#) (Thomson)

MASTER DATA REPOSITORY LIST



[Master Data Repository List](#)  
~ 300 entrepôts (mars 2016)

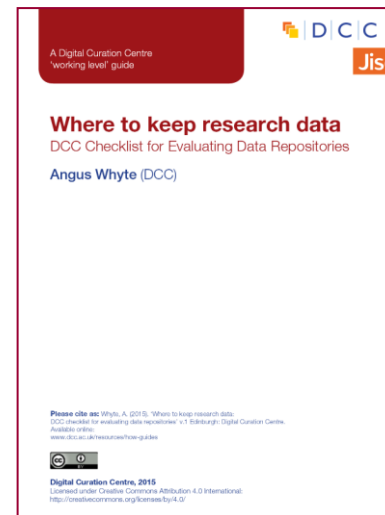
Accès payant



# Comment choisir un entrepôt ?

- 🌸 Selon les recommandations d'un financeur, d'un éditeur, de son organisme de rattachement ...
- 🌸 Selon les types d'entrepôts et leurs caractéristiques
  - Discipline,
  - Modèle économique (organisation à but lucratif ou non),
  - Type d'identification,
  - Licence,
  - Partenariat éditeurs,
  - Certification...

Whyte, A. (2015). 'Where to keep research data: DCC checklist for evaluating data repositories' v.1  
Edinburgh: Digital Curation Centre.  
[www.dcc.ac.uk/resources/how-guides](http://www.dcc.ac.uk/resources/how-guides)



# Publier des données comme matériel supplémentaire d'un article

Author Manuscript

NIH-PA Author Manuscript

**Detailed Methods**

**Plasmids and Cell Lines**

Sequence verified ORF clones (Supplementary Table S1) in pDO-NR223 were recombined into either the Gateway destination vector MSCV-N-Flag-HA-IRES-PURO (LTR-driven expression) or pHAGE-N-Flag-HA (lentiviral vector) using  $\lambda$  recombinase<sup>9</sup>. After packaging in 293T cells, viruses were used to infect the indicated cell lines and selection accomplished using 1  $\mu$ g/ml puromycin. The pHAGE-N-Flag-HA vector was employed in transient transfections (293T cells, Lipofectamine 2000 (Invitrogen)) for a subset of AIN proteins that were toxic when expressed constitutively from the LTR promoter (Supplementary Table S1).

**Protein Purification**

For standard purifications, cells from four 15-cm tissue culture dishes at ~80% confluence (~ $10^7$  cells) were lysed in a total volume of 4 ml of lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% Nonidet P40, Roche complete EDTA-free protease inhibitor cocktail) for 1 hour with gentle rocking at 4°C. In some experiments, cells were incubated with 200 nM Torin1<sup>37</sup> (a gift from N. Gray, Dana Farber Cancer Institute) for 6 h prior to harvesting. Lysates were cleared using

## Part 1. Supplementary Methods

**Data Processing and Analysis.** Mass spectral data was processed using *CompPASS*, as previously described<sup>1</sup> with modifications discussed below. Briefly, Sequest summary files were processed into a high threshold dataset based on a 2% protein false-positive rate by keeping the XCorr thresholds for each charge state constant while varying the  $\Delta Cn$  (thresholds: XCorr 2+  $\geq$  2.5; XCorr 3+  $\geq$  3.2; XCorr 4+  $\geq$  3.5; +1 charge states were not collected). These processed data sets were merged for each duplicate run and used to populate a "stats table" consisting of each dataset for the AIN as well as 102 unrelated proteins (Dubs and their selected HCIPs<sup>1</sup>; [https://harper.hms.harvard.edu/CompPASS\\_Dubs.html](https://harper.hms.harvard.edu/CompPASS_Dubs.html)). The  $D^N$ -score and Z-score are calculated from total spectral counts (TSCs) for each protein found in association with each bait.

Because *CompPASS* was originally designed for analysis of mostly non-reciprocal datasets, we devised a new weighted  $D^N$ -score (WD<sup>N</sup>-score) (Supplementary Fig. S2), which aids in the identification of HCIPs that are associated with multiple baits in a network. WD<sup>N</sup>-scores were calculated as:

$$WD_{i,j} = \sqrt{(\lambda \omega_j)^p (x_{i,j})} \quad (\text{Eq. 1})$$

$$\lambda = \left( \frac{k}{\sum_i x_{i,j}} \right), \quad f_{i,j} = \begin{cases} 1; & x_{i,j} > 0 \\ x_{i,j} & \end{cases} \quad (\text{Eq. 2})$$

$$\omega_j = \left( \frac{\sigma_j}{\bar{x}_j} \right), \quad \bar{x}_j = \frac{\sum_{i=1}^m x_{i,j}}{k}; \quad n = 1, 2, \dots, m \quad (\text{Eq. 3})$$

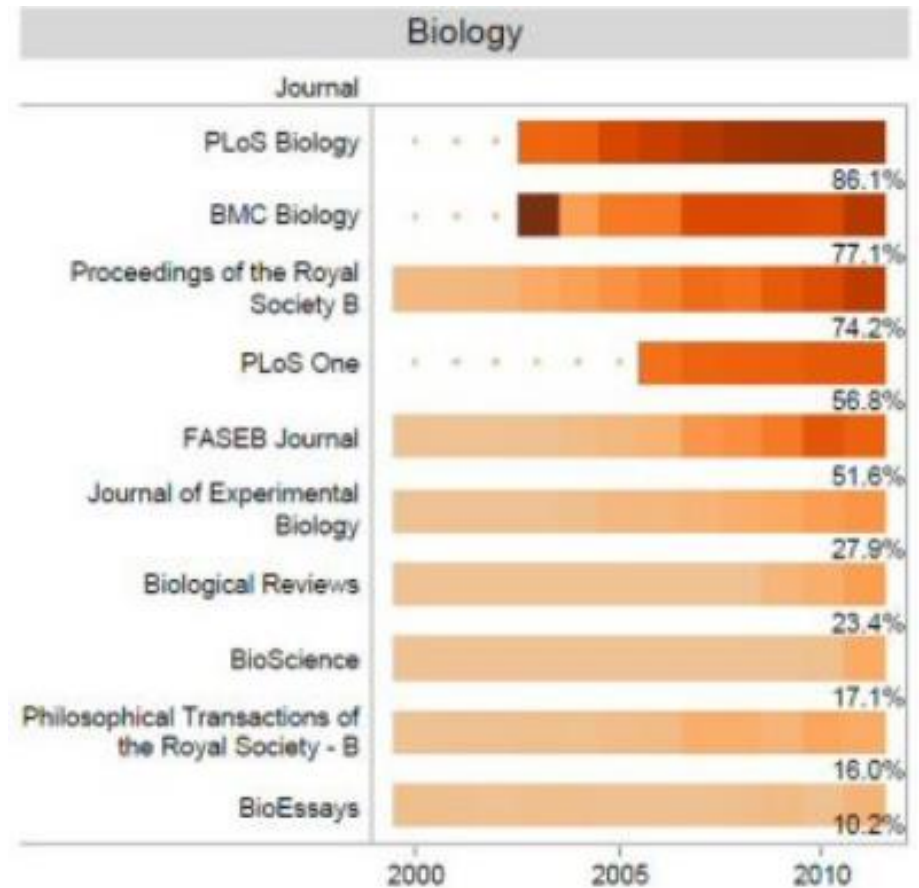
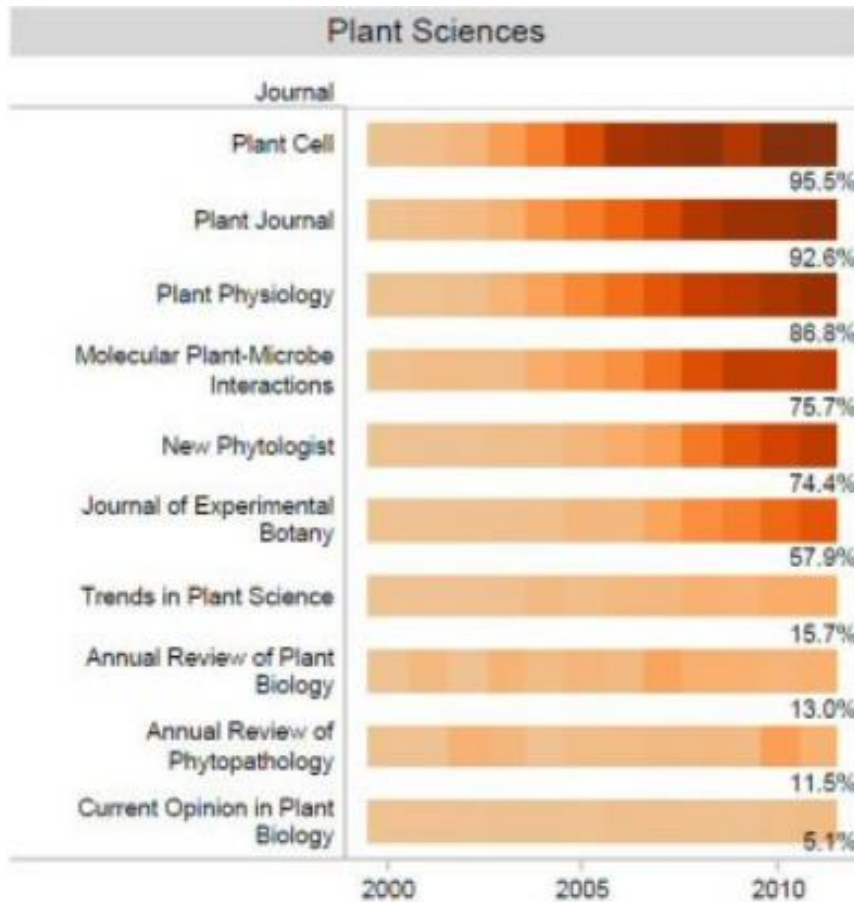
$x_{i,j}$  = total peptides for interactor  $j$  from bait  $i$

# Tendances...

## De plus en plus d'articles sont publiés avec du « matériel supplémentaire »

- Libellés variés : Supplemental material, Supplemental data, Auxiliary information, Supporting information, Supplementary content, Additional content ...
- Contenus variés : fichiers audio, vidéo, images à haute résolution, analyses statistiques, explications méthodologiques approfondies ...
- Mis à disposition par les auteurs et/ou à la demande des reviewers

# The percentage of articles with supplementary materials by journal.



The percentage of all articles with supplementary material for each journal over the study period of 2000-2011. Journals are ordered according to 2011 values. 2011 values are given at the end of each row.



Kenyon, J., & Sprague, N. R. (2014). Trends in the Use of Supplementary Materials in Environmental Science Journals. *Issues in Science and Technology Librarianship*.

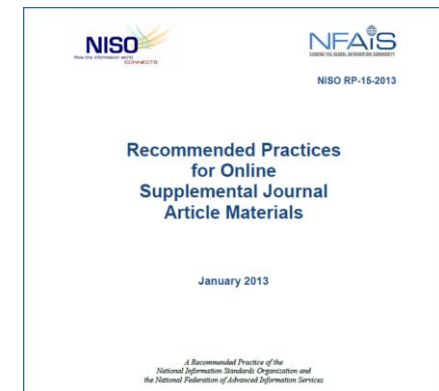
<http://dx.doi.org/10.5062/F40Z717Z>

# L'édition scientifique s'adapte...

- Les politiques éditoriales intègrent de plus en plus le dépôt et le partage de données mais dans un contexte d'augmentation des volumes des données :
  - imposent parfois des limites de taille aux données fournies comme matériel supplémentaire des articles,
  - externalisent la gestion la curation et le stockage des données en recommandant souvent des entrepôts spécifiques
  - certaines font de la mise à disposition des données une condition préalable à l'acceptation de l'article (Nature ou PLoS)

## Guide de bonnes pratiques

- Recommandations pour les éditeurs, les auteurs, les relecteurs et les **documentalistes** (NISO & NFAIS, 2013)





# Exemple : Groupe Nature (1)



[authors & referees](#) > [Policies](#) > Availability of data, material and methods

**Site content**

- Homepage
- Policies
  - Publication ethics
  - Bioethics
  - Availability of data & materials
  - Peer-review policy
  - Embargo
  - Corrections
  - License to publish
  - Feedback
- Author resources
- Peer review
- Open access

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. **A condition of publication in a Nature journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications.** Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must also be disclosed in the submitted manuscript.

- [experimental protocols](#)
- [clinical trials](#)
- [further reading](#)

# Exemple : Checklist / Groupe Nature

## Reporting Checklist For Life Sciences Articles

This checklist is used to ensure good reporting standards and to improve the reproducibility of research. Please read [Reporting Life Sciences Research](#).

► Data deposition

Reported in section

Mandatory deposition	Suitable repositories
Protein sequences	<a href="#">Uniprot</a>
DNA and RNA sequences	<a href="#">Genbank</a>
	<a href="#">DNA DataBank of Japan (DDBJ)</a>
	<a href="#">EMBL Nucleotide Sequence Database (ENA)</a>
DNA and RNA sequencing data	<a href="#">NCBI Trace Archive</a>
	<a href="#">NCBI Sequence Read Archive (SRA)</a>
Genetic polymorphisms	<a href="#">dbSNP</a>
	<a href="#">dbVar</a>
	<a href="#">European Variation Archive (EVA)</a>
Linked genotype and phenotype data	<a href="#">dbGAP</a>
	<a href="#">The European Genome-phenome Archive (EGA)</a>
Macromolecular structure	<a href="#">Worldwide Protein Data Bank (wwPDB)</a>
	<a href="#">Biological Magnetic Resonance Data Bank (BMRB)</a>
	<a href="#">Electron Microscopy Data Bank (EMDB)</a>
Microarray data (must be MIAME compliant)	<a href="#">Gene Expression Omnibus (GEO)</a>
	<a href="#">ArrayExpress</a>
Crystallographic data for small molecules	<a href="#">Cambridge Structural Database</a>

17. Provide accession codes for deposited data.

Data deposition in a public repository is mandatory for:

- a. Protein, DNA and RNA sequences
- b. Macromolecular structures
- c. Crystallographic data for small molecules
- d. Microarray data

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available [here](#). We encourage the provision of other source data in supplementary information or in unstructured repositories such as [Figshare](#) and [Dryad](#). We encourage publication of Data Descriptors (see [Scientific Data](#)) to maximize data reuse.

18. If computer code was used to generate results that are central to the paper's conclusions, include a statement in the Methods section under "**Code availability**" to indicate whether and how the code can be accessed. Include version information as necessary and any restrictions on availability.

### About Scientific Data

Scientific Data is an open-access, peer-reviewed journal for descriptions of scientifically valuable datasets. Our primary article-type, the **Data Descriptor**, is designed to make your data more discoverable, interpretable and reusable.

[Browse publications by date](#) ►

[Browse publications by subject](#) ►

<http://www.nature.com/authors/policies/checklist.pdf>

# Publier un Data Paper

« Data as the subject of a paper »



ist@inra

## Data article

A **data article** is a ‘data publishing’ product, also known as a ‘**data descriptor**’, that may appear in a **data journal** or any other journal. When publishers refer to ‘data publishing’ they usually mean a data article rather than the underlying dataset.

Data articles focus on making data discoverable, interpretable and reusable rather than testing hypotheses or presenting new interpretations (by contrast with traditional journal articles).

Whether linked to a dataset in a separate repository, or submitted in tandem with the data, **the aim of the data article is to provide a formal route to data-sharing.**

//

The length of such articles can vary from micro papers (focused on one table or plot) to very detailed presentation of complex datasets.

<https://rd-alliance.org/key-components-data-publishing-using-current-best-practices-develop-reference-model-data-publishin-0>

*ist@inra*

# Caractéristiques & VA

## Informer la communauté scientifique

- existence et disponibilité d'un dataset

## Valoriser les données avec objectif d'ouverture

- Bonne visibilité
- Documentation
- Potentiel de réutilisation explicite

## Obtenir la reconnaissance du travail réalisé pour producteurs /éditeurs de données

**DOI : indexation & citation**  
(indexé dans : WoS, Scopus, Zoological Record, Google Scholar, CAB Abstracts, DOAJ, Ebsco...)

**Description des données sous une forme structurée, lisible par un humain**  
**Métadonnées rigoureuses**

**Lien vers entrepôt**

ZooKeys 489: 15–24 (2015)  
doi: 10.3897/zookeys.489.9292  
<http://zookeys.pensoft.net>

DATA PAPER



**The Jean Gutierrez spider mite collection**

Alain Migeon<sup>1</sup>

<sup>1</sup> INRA, UMR 1062 CBGP, F-34988 Montpellier-sur-Lez, France

Corresponding author: *Alain Migeon* ([alain.migeon@supagro.inra.fr](mailto:alain.migeon@supagro.inra.fr))

Academic editor: *V. Pesic* | Received 28 January 2015 | Accepted 10 March 2015 | Published 23 March 2015

<http://zoobank.org/0719C382-988D-4DCC-8B24-00E65E3C8CC5>

**Citation:** Migeon A (2015) The Jean Gutierrez spider mite collection. *ZooKeys* 489: 15–24. doi: 10.3897/zookeys.489.9292

**Abstract**  
The family Tetranychidae (spider mites) currently comprises 1,275 species and represents one of the most important agricultural pest families among the Acari with approximately one hundred pest species, ten of which considered major pests. The dataset presented in this document includes all the identified spider mites composing the Jean Gutierrez Collection hosted at the CBGP (Montpellier-sur-Lez, France), gathered from 1963 to 1999 during his career at the Institut de Recherche pour le Développement (IRD). It consists of 5,262 specimens corresponding to 1,564 occurrences (combination species/host plant/date/location) of 175 species. Most specimens were collected in Madagascar and other islands of the Western Indian Ocean, New Caledonia and other islands of the South Pacific and Papuaia. The dataset constitutes today the most important one available on Tetranychidae worldwide.

**Keywords**  
Acari, Tetranychidae, World, Madagascar, Western Indian Ocean, New Caledonia, South Pacific, Papuaia

**Data published through GBIF**

<http://www.gbif.org/dataset/ac60a288-fcc9-43fe-a7d4-e732b748a981>



# Data Paper - Structure

🌸 Une structure spécifique parfois minimale

Data Paper



## General structure

- Title
- Authors, affiliations
- Abstract
- Keywords
- Context
  - Spatial coverage, temporal coverage
- Methods
  - Steps, sampling strategy, quality control, constraint ethical considerations
- Dataset description
  - Object names, data type, format names & versions, creators, creation dates, language, license, location (DOI), publication date
- Reuse potential
- Acknowledgements
- References

## Liens pérennes



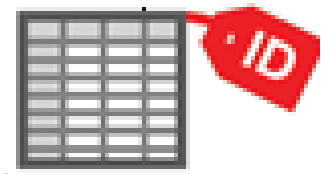
- Dépôt au sein du journal ou dans un entrepôt externe (recommandé ou au choix de l'auteur)
- En libre accès ou avec une restriction d'accès temporaire


Dataset

Citation vers  
les données



Citation vers  
l'article



 **« Lorsque cela est possible, il est recommandé que les contributeurs soient identifiés et les données attribuables, les deux de manière unique, grâce à des identifiants pérennes, non-propriétaires, ouverts et interopérables »**

(Extrait des « Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020- 2013»)

## Exemples d'identifiants pérennes

- ORCID pour les contributeurs
- DOI-DataCite pour les identifiants de données
  - Digital Object Identifier-DOI : identifiant pérenne et unique permettant de référencer, citer et fournir un lien stable vers un objet numérique et sa citation.

# Data Paper - Peer review

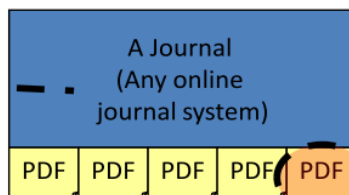
## The traditional online journal model

1) Author prepares the paper using word processing software.

Word processing software with journal template



2) Author submits the paper as a PDF/Word file.



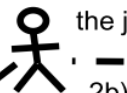
3) Reviewer reviews the PDF file against the journal's acceptance criteria.



## Overlay journal model for publishing data

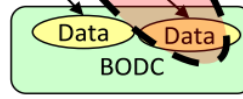
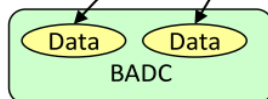
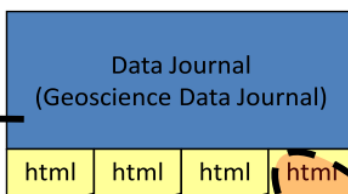
1) Author prepares the data paper using word processing software and the dataset using appropriate tools.

Word processing software with journal template



2a) Author submits the data paper to the journal.

2b) Author submits the dataset to a repository.



3) Reviewer reviews the data paper and the dataset it points to against the journal's acceptance criteria.

Effort d'intégration des processus entrepôts

(Bloom, Dallmeier-Tiessen, & Newbold, n.d.)



# Les limites du partage ...

## Limites éthiques

- Protection des données personnelles
- Données sensibles

## Limites juridiques

- Propriété industrielle
- Propriété intellectuelle
  - Données détenues par un tiers
  - Droit d'auteur sur les données brutes ?
  - Droit des bases de données

### **Diapo reprise de :**

Cosserat, F. – Une introduction à la gestion et au partage des données de la recherche (octobre 2015)

# Les conditions du partage : définir/choisir une licence?

 « **il faut lui (jeu de données) apposer une licence de diffusion fixant les conditions de son utilisation :**

- *droits d'utilisation et de modification de la donnée,*
- *droits de réutilisation commerciale et non commerciale,*
- *obligations éventuelles comme la mention de la source des données ou le partage à l'identique. »*

 « *le type de licence préconisé est donc un critère de sélection important pour choisir un entrepôt ou un éditeur* ». (Dedieu, 2015)

 **Outil d'aide**

- <http://ufal.github.io/public-license-selector/>

# Data Citation

🌸 La citation : une méthode connue, éprouvée et reconnue des scientifiques pour relier des documents...

## 🌸 Citation des données

- 8 principes définis dans une déclaration commune (2014)  
<https://www.force11.org/group/joint-declaration-data-citation-principles-final>



- |                           |                                     |
|---------------------------|-------------------------------------|
| 1. Importance             | 5. Access                           |
| 2. Credit and attribution | 6. Persistence                      |
| 3. Evidence               | 7. Specificity and verifiability    |
| 4. Unique identification  | 8. Interoperability and flexibility |

- DataCite Metadata Schema for the Publication and Citation of Research Data (2015)  
modèle "minimal" - 5 éléments

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

Data Paper

The GBS dataset: measurements of satellite site diversity at 20.7 GHz in the UK

S. A. Callaghan\*, J. Waight, J. L. Agnew, C. J. Walden, C. L. Wrench and S. Ventouras

Article first published online: 17 MAR 2013

DOI: 10.1002/gdj3.2

© 2013 The Authors. Geoscience Data Journal published by John Wiley & Sons Ltd and Royal Meteorological Society.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes

Alt score 12

Additional Information (Show All)

How to Cite | Author Information | Publication History | Funding Information

The research presented in this paper was funded by the UK's Ofcom as part of the Spectrum Efficiency Scheme and the support of Ofcom in providing the funding for the GBS experiment is greatly appreciated.

Abstract | Article | References | Cited By

Enhanced Article (HTML) | Get PDF (1849K)

Keywords: site diversity; radio propagation; fade mitigation techniques

Abstract

The GBS (Global Broadcast Service) dataset is a series of radio attenuation measurements made at three sites in the UK: Chilbolton and Sparsholt, both in southern UK, and Dundee in Scotland. The aim of the experiment was to make long term measurements of the signal strength received from a 20.7 GHz beacon on the US Department of Defense satellite UFO-9 at multiple sites, in order to determine whether the use of site diversity as a fade mitigation technique would be effective. The dataset spans a period of 3 years, from August 2003 to August 2006 with signal attenuation sampled once per second.

Dataset

The GBS (Global Broadcast Service) dataset comes as 3 separate data streams:

- Identifier: doi:10.5285/639A3714-BC74-46A6-9026-64931F355E07  
 Creator: Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J. Walden, J. Agnew and S. Ventouras].  
 Title: GBS 20.7 GHz slant path radio propagation measurements, Chilbolton site  
 publisher: NERC British Atmospheric Data Centre  
 Publication year: 2009  
 Resource type: Metadata document  
 Version: 1.0

SEARCH

In this issue

Advanced > Saved Searches >

ARTICLE TOOLS

- Get PDF (1849K)
- Save to My Profile
- E-mail Link to this Article
- Export Citation for this Article
- Get Citation Alerts
- Request Permissions

Share | Facebook | Twitter | LinkedIn | Email

References

Jump to...

Callaghan SA, Waight J, Agnew JL, Wrench CL. 2005. Medium and long range site diversity measurements from the three site GBS experiment in the UK, 3rd International Workshop Cost Action 280 PM9-101, June 2005.

Callaghan SA, Boyes B, Couchman A, Waight J, Walden CJ, Ventouras S. 2008. An investigation of site diversity and comparison with ITU-R recommendations. *Radio Science* 43: RS4010, doi:10.1029/2007RS003793.  
[Abstract](#) | [Article](#) | [PDF\(663K\)](#) | [References](#) | [Web of Science® Times Cited: 4](#)

Castanet L, Bolea-Alamañac A, Bousquet M. 2003. Interference and Fade Mitigation Techniques for Ka and Q/V Band Satellite Communication Systems, COST 272-280 Int'l. Wksp. Satellite Communication from Fade Mitigation to Service Provision, Noordwijk, The Netherlands, May 2003.

Goldstein O, Messer H, Zinevich A. 2009. Rain rate estimation using measurements from commercial telecommunications links. *Signal Processing, IEEE Transactions*, 57: 1616-1625, April 2009, doi:10.1109/TSP.2009.2012554.  
[CrossRef](#). [Web of Science® Times Cited: 19](#). [ADS](#)

Panagopoulos AD, Arapoglou P-DM, Cottis PG. 2004. Satellite communications at KU, KA, and V bands: propagation impairments and mitigation techniques. *Communications Surveys & Tutorials, IEEE*, 6: 2-14.  
[CrossRef](#). [Web of Science® Times Cited: 40](#)

Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan SA, Waight J, Walden CJ, Agnew J, Ventouras S]. 2009a. GBS 20.7 GHz slant path radio propagation measurements, Sparsholt site. NERC British Atmospheric Data Centre. doi: 10.5285/E8F43A51-0198-4323-A926-FE69225D57DD.

Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan SA, Waight J, Walden CJ, Agnew J, Ventouras S]. 2009b. GBS 20.7 GHz slant path radio propagation measurements, Chilbolton site. NERC British Atmospheric Data Centre. doi: 10.5285/639A3714-BC74-46A6-9026-64931F355E07.

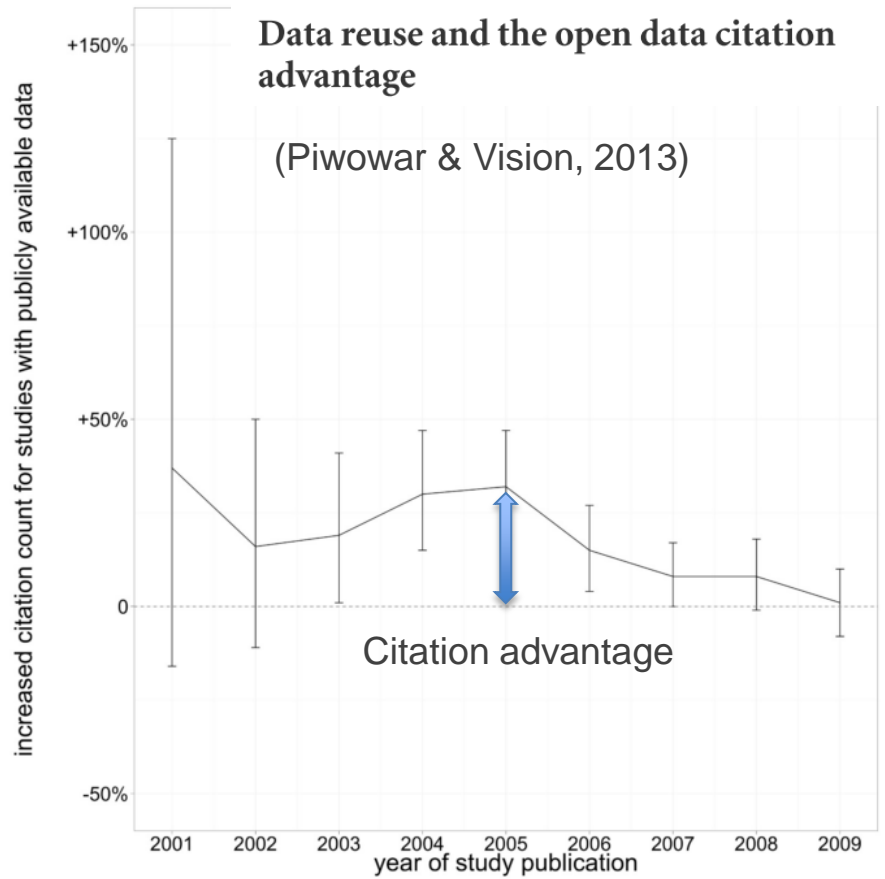
Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan SA, Waight J, Walden CJ, Agnew J, Ventouras S]. 2009c. GBS 20.7 GHz slant path radio propagation measurements, Dundee site. NERC British Atmospheric Data Centre. doi: 10.5285/db8d8981-1a51-4d6e-81c0-coed9b921390.

Ventouras S, Callaghan SA, Wrench CL. 2006. Long-term statistics of tropospheric attenuation from the Ka/U band ITALSAT satellite experiment in the United Kingdom. *Radio Science* 41: RS2007, doi:10.1029/2005RS003252.  
[Abstract](#) | [Article](#) | [PDF\(1315K\)](#) | [References](#) | [Web of Science® Times Cited: 2](#)



« I found 377 variant methods of citing different versions of a single data set, WOA/WOD, suggesting that researchers are not consistently using the citation formats provided for these data sets” (Belter, 2014)

# Open Data citation advantage ?



« Studies that made data available in a public repository received 9% more citations than similar studies for which the data was not made available »



(Belter, 2015)

**Global-level data sets may be more highly cited than most journal articles**



Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. Plos One

“My results suggest that all three data sets are more highly cited than most journal articles. Each data set has probably been cited more often than 99% of the journal articles in oceanography that were published during the same years as the data sets”

ist@inra



# Data Journals - Exemples

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

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

*Calling for submissions in Fall 2013, launching in Spring 2014*  
[nature.com/scientificdata](http://nature.com/scientificdata)

Journal of **open** health data

Ecological Archives

Research Data Journal  
 for the Humanities  
 and Social Sciences

BRILL | OANS

A peer-reviewed open-access journal

**Biodiversity Data Journal**

Making your data count! ISSN 1314-2828 (online)

Journal of **open** psychology data

A peer-reviewed open-access journal

**ZooKeys**

Launched to accelerate biodiversity research

(GIGA)<sup>n</sup> SCIENCE

**Data in Brief**

[Open Access](#) | [About this Journal](#) | [Submit your Article](#)

- Get new article feed
- Subscribe to new volume alerts
- Add to Favorites

Copyright © 2016 Elsevier Inc. All rights reserved

ist@inra

# Data Journals - Paysage

## Data journals: A survey

(Candela, Castelli, Manghi, & Tani, 2014)

- 116 journaux examinés  
« promoting data papers »  
(7 « dédiés », 109 «hybrides»)
- issus de 15 publishers

Entre 2000 et 2013 :  
le nombre de journaux  
publiant des DP augmente

Sur la période : 826 DP  
publiés, dont 23,5% sur la  
dernière année étudiée.

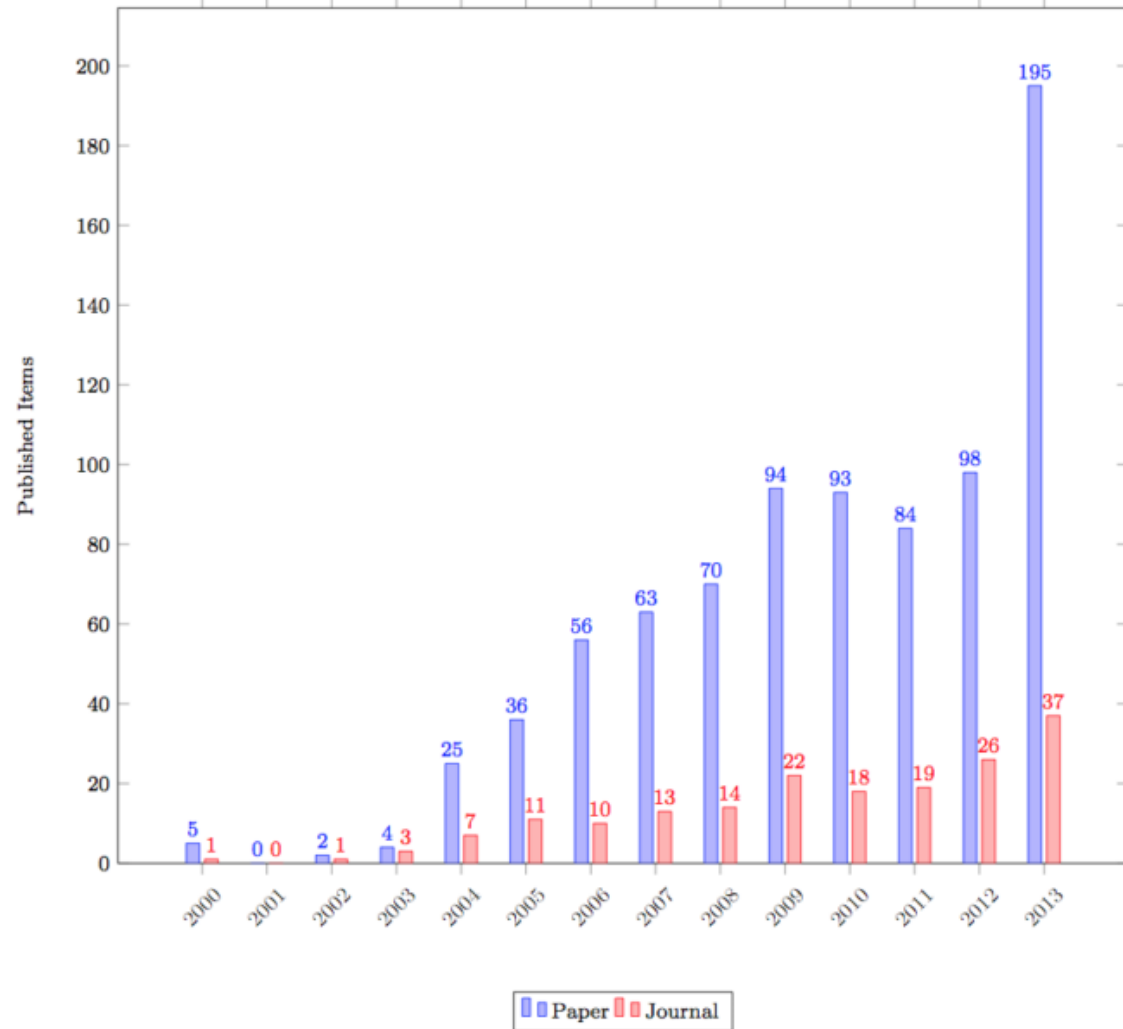


Figure 3. Number of published papers and operational data journals, by year.

# What tools/sites do you use to publish?



787 out of 1000 people answered this question



<https://101innovations.wordpress.com/tag/updates-insights/>



# Data Journals - Listes

-  Une liste maintenue par l'université d'Edinburgh  
<https://www.wiki.ed.ac.uk/display/datashare/Sources+of+dataset+peer+review>
-  Liste établie dans le cadre du projet PREPARDE (UK)  
<http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList>
-  Une liste de 116 titres fournie par les auteurs de l'article Candela, L., Castelli, D., Manghi, P., & Tani, A. (2014). Data journals: A survey (preprint). Journal of the Association for Information Science and Technology, 1-20. [10.1002/asi.23358](https://doi.org/10.1002/asi.23358)



# **Rôles et enjeux ? Freins et leviers ? pour les différents acteurs**

*ist@inra*

# Pour le chercheur...



Producteur/Utilisateur  
de données

7 critères ~qualités à atteindre pour rendre visibles, partager ses données...

Disponibilité

Visibilité

Interprétabilité

Réutilisabilité

Citabilité

Curation

Préservation

## Décrire et documenter ses données ?

- Quelles métadonnées?
- Plan de gestion des données ?

## Partager ou ne pas partager ?

## Où et comment diffuser ses données?

- Quelles voies utiliser ? Quel entrepôt ?
- Publier un Data Paper ?

## Comment faciliter la réutilisation ?

- Accès pérenne, format durable
- Identifiant - Citation
- Licences

Rapport ODE (Reilly, Schallier, Schrimpf,  
Smit, & Wilkinson, 2011)

ist@inra

# Publier un Data Paper – Retours d'expérience



Gray, E. A. (2015)

Best practice open data is time consuming  
(but still worth the extra effort!)

Emilio M. Bruna recently provided an estimate of the amount of time it took him to prepare & upload open data related to publication to figshare & dryad.

11 Hours & \$90  
(for Dryad)

Providing open-source code was the most time consuming part (25.5 hours), and Open Access publication the most expensive (\$600).

<http://brunalab.org/blog/2014/09/04/the-opportunity-cost-of-my-openscience-was-35-hours-690/>

Mounce, R. (2014)



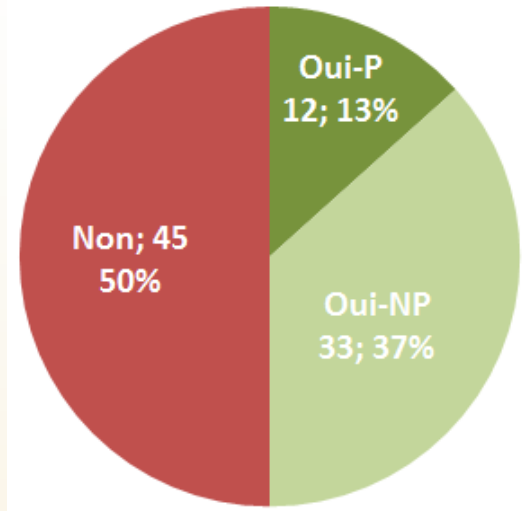
# « Publiez-vous, ou pourriez-vous publier des Data Papers ? »

**ENQUETE Inra  
2015**

## Objectifs

- Compléter une liste de Data Papers « Inra » établie de façon empirique,
- Avoir une vision de l'utilisation actuelle de cette voie de publication et de son intérêt pour les scientifiques
- Échanger sur les difficultés éventuelles rencontrées dans la démarche, les points de vigilance à signaler, les suites éventuelles de ce type de publication (contact, réutilisation des données)

## CONNAISSEZ-VOUS LE TYPE DE PUBLICATION DATA PAPER ?



90 réponses analysables

## PARMI LES 12 PUBLIANTS

- Aucune difficulté particulière sur le processus éditorial (7/12)
- Difficultés mentionnées
  - Rédaction des métadonnées
  - Choix de l'entrepôt
  - Choix de la licence associée aux données
- Une publication suite à la demande d'un éditeur (dans « Data in Brief » suite à une première soumission à « Journal of Proteomics »)

## Parmi les autres répondants

(72%) **Connaisseurs**

**Non Publiants**

(80%)

**Non connaisseurs**

intéressés par les DP soit comme lecteur-utilisateur, soit comme auteur-fournisseur de données, soit pour les 2 raisons

59 personnes (soit 64% des répondants) sont intéressées par la création d'un **service d'accompagnement à la publication de Data Papers**

# Intérêt du Data Paper ? Avis partagés

Avis Positifs	Plus mitigés
<ul style="list-style-type: none"><li>• Intéressant pour les unités expérimentales comme voie de valorisation des données produites</li><li>• Intéressant pour l'aspect réutilisation des données</li></ul>	<ul style="list-style-type: none"><li>• Activité chronophage</li><li>• Non suffisamment reconnue</li><li>• Uniquement sous la pression des journaux classiques</li><li>• Coût des DP et plus largement des articles en OA</li><li>• Quelle valeur ajoutée du DP ?<ul style="list-style-type: none"><li>– par rapport à un article classique avec des SM et des données expérimentales (« sauf pour l'évaluation : 2 articles au lieu d'un ... »)</li><li>– Plutôt favorable à un renforcement de la pression pour la mise à disposition des données dans les articles classiques</li></ul></li><li>• Aspects juridiques (quelle licence choisir ?)</li></ul>

# Freins et leviers du partage des données

Bénéfices	Barrières et freins
<ul style="list-style-type: none"><li>• Favorise la reproduction / validation de la recherche</li><li>• Réduction des erreurs et des fraudes</li><li>• Réduction de la duplication de travaux de recherche</li><li>• Respect des mandats des financeurs et des éditeurs</li><li>• Meilleur retour sur investissement de la recherche</li><li>• Favorise l'émergence de nouvelles hypothèses de recherche</li><li>• Utilisation pour l'enseignement</li><li>• Gain de citation</li></ul>	<ul style="list-style-type: none"><li>• Activité chronophage</li><li>• Coûts associés au partage de données</li><li>• Problèmes de confidentialité (données sensibles ...)</li><li>• Interrogation sur la notion de propriété des données ...</li><li>• Inquiétudes sur une réutilisation inappropriée</li><li>• Manque de reconnaissance académique et institutionnelle</li><li>• Méconnaissance des procédures (rédaction des métadonnées, choix entrepôts ...)</li></ul>

(Hrynaszkiewicz & Shintani, 2014)  
(Tenant, 2105)



# Editeurs scientifiques

## The International Journal of Robotics Research

<http://ijr.sagepub.com/>

**Editorial: Data Papers — Peer Reviewed Publication of High Quality Data Sets**  
 Paul Newman and Peter Corke  
*The International Journal of Robotics Research* 2009 28: 587  
 DOI: 10.1177/0278364909104283

The online version of this article can be found at:  
<http://ijr.sagepub.com/content/28/5/587.citation>

Published by:  
  
<http://www.sagepublications.com>

# 2009



## PLOS' New Data Policy: Public Access to Data

Posted February 24, 2014 by Liz Silva in [Aggregators](#), [Open Access](#)

# 2014

## BJET | British Journal of Educational Technology

Editorial  
**Editorial: Data papers**

Nick Rushby Editor

Article first published online: 22 JUL 2015  
 DOI: 10.1111/bjet.12337

© 2015 British Educational Research Association



**British Journal of Educational Technology**  
 Special Issue: Open Data in Learning Technology  
**Volume 46, Issue 5, pages 899–903, September 2015**

# 2015

## PUBLISHING DATA PAPERS IN ANNALS OF FOREST SCIENCE: DETAILED GUIDELINES FOR A SMOOTH PREPARATION AND SUBMISSION

By [Erwin Dreyer](#) On 20 November 2015 In [Editor's Choice](#), [Editorial](#), [Editorial News](#), [For Authors](#)



# 2015

# Publishers vs Journal editors ?

OPEN ACCESS Freely available online

PLOS BIOLOGY

Perspective

## Recommendations for the Role of Publishers in Access to Data

Jennifer Lin<sup>1\*</sup>, Carly Strasser<sup>2</sup>

<sup>1</sup> PLOS, San Francisco, California, United States of America, <sup>2</sup> California Digital Library, University of California, Oakland, California, United States of America

Lin J, Strasser C (2014) Recommendations for the Role of Publishers in Access to Data. PLoS Biol 12(10): e1001975. [10.1371/journal.pbio.1001975](https://doi.org/10.1371/journal.pbio.1001975)

### Box 2. Recommendations for Publishers to Increase Access to Data

1. Establish and enforce a mandatory data availability policy.
2. Contribute to establishing community standards for data management and sharing.
3. Contribute to establishing community standards for data preservation in trusted repositories.
4. Provide formal channels to share data.
5. Work with repositories to streamline data submission.
6. Require appropriate citation to all data associated with a publication—both produced and used.
7. Develop and report indicators that will support data as a first-class scholarly output.
8. Incentivize data sharing by promoting the value of data sharing.

<http://ist.blogs.inra.fr/afs/2014/11/01/open-data-a-role-for-publishers-or-for-journal-editors/>

### A role for publishers... or for journal editors?

Posted by edreyer on 30 Oct 2014 at 12:28 GMT

I fully agree with the content of this paper. Institutions and journal editors are getting fully aware of the importance of open data as a full scientific production that needs to be made available for re use, and for quality control as companions to published papers. Unfortunately, I see an important ambiguity in the whole paper that requires some clarification. It is that we never know whether the actors should be the "publishers" (i.e., Elsevier, Springer, Francis and Taylor, Nature publishing group, etc) or the journal editors (i.e., the scientists in charge of the editorial policy of the journal and the warrants of the scientific quality of the published material). I strongly believe that this issue, as well as the issue of the ethics in scientific publication, is a central duty of the journal editors and not of the publishers. The procedures for submission and evaluation, the guidelines and rules for access and for the data repositories should be a major concern for the scientists in charge of the editorial policy of the journals, and the publishers should act in support of this policy. Otherwise, we would rapidly run into the classical conflict of interest which is already visible today for the publication of papers: publishers need to publish to make their economic model run. Scientists need to publish relevant and solid stuff including data sets. I will not list again the concerns raised by many scientists about the policy of publishers towards the publication business. This confusion is a real problem for me, and I would urge the authors of the present "recommendations" to clarify this point. Actually, the solution would rely on a tight cooperation between editors (who should take the lead and develop editorial policies for their journals) and the publishers (who may bring an added value due to their expertise in publication techniques and data management systems).

Indeed, the solution is a clear and transparent cooperation between editors and publishers, each with his aims and responsibilities to avoid the conflict of interest that will arise if we do not clarify this matter.

Erwin Dreyer, Editor of Annals of Forest Science, Inra Nancy (France).

# Des collaborations existent...

## Consortium

- Research facilities
- Data repositories
- Universities
- Libraries
- Industry

(Blum et al., 2014)



Rassemblant les différents acteurs :

- Institutions de recherche
- Entrepôts
- Editeurs scientifiques
- Editeurs commerciaux
- Chercheurs ...

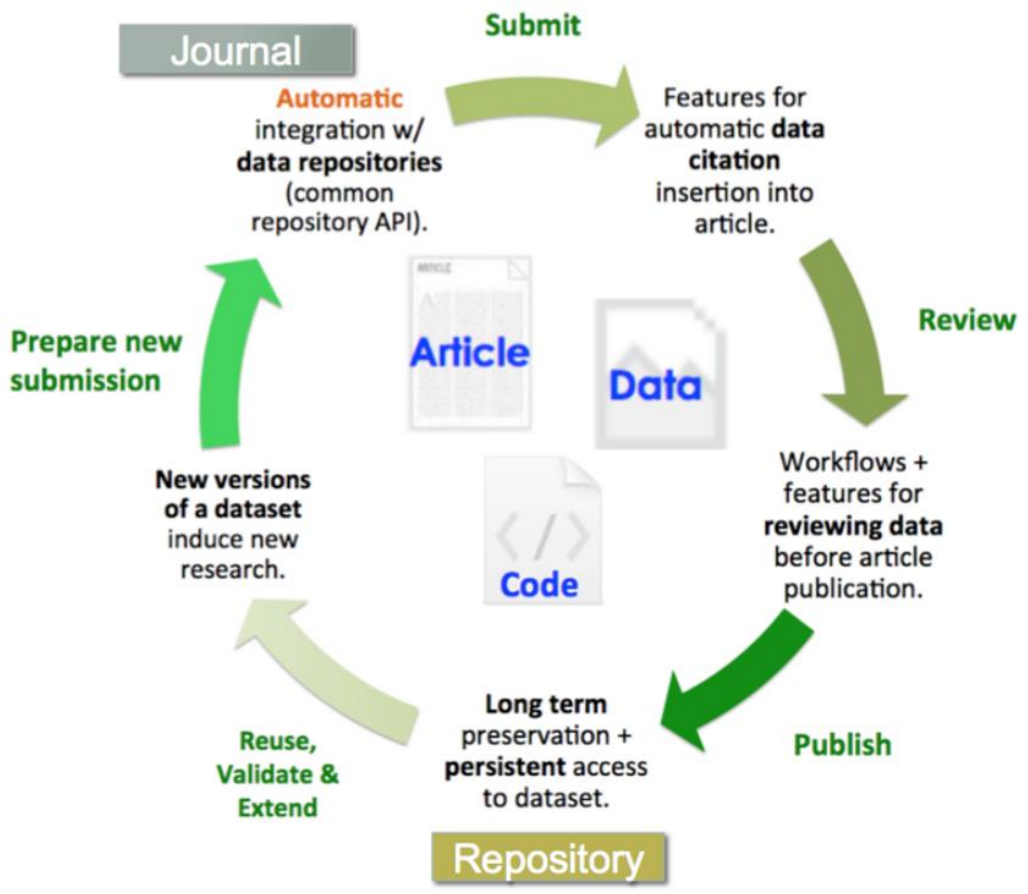
## Groupes de travail RDA-Alliance

<https://rd-alliance.org/groups/working-groups>

- RDA/WDS Publishing Data Bibliometrics WG
- RDA/WDS Publishing Data Services WG
- RDA/WDS Publishing Data Workflows WG



# De nouveaux workflows, plus d'intégration...



Intégration de la soumission des articles et des datasets

Exemples

- Dryad
- Open Journal Systems et Dataverse

Dispositifs facilitant la création des Data Papers

Exemple :

GBIF et Pensoft et Nature ...

Figure 1. Lifecycle of an automated and integrated journal and data publishing workflow.

Altman et al., 2015



# L'avenir ...?

Une  
vision  
optimiste

## Predictions for the (near) future

- **Research funding bodies** will tighten-up their rules to ensure **immediate post-publication data sharing**. No embargoes, no bullshit.
- If no published data comes from your funded research, it will negatively effect your future chances of funding
- **Research institutions will** significantly improve **research data management training** for ALL staff & students, old and new alike
- **Good journals** will **strictly enforce mandatory data sharing**. Journals that don't will get a bad reputation for irreproducible research
- **CC0** for data will become the de facto standard. Everyone will realise that legal protection under **copyright is completely the wrong tool** for ensuring the ethical use of data & appropriate authorship assignment.

Mounce, R. (2014) The state of Open Research Data, pp 1-29.

# Merci de votre écoute

Dominique L'Hostis  
[dolhostis@nantes.inra.fr](mailto:dolhostis@nantes.inra.fr)  
INRA Angers-Nantes

*ist@inra*

# Bibliographie (1/2)

- Altman M., Castro E. , Crosas M., Durbin P , Garnett A., and Whitney J. (2015). Open Journal Systems and Dataverse Integration– Helping Journals to Upgrade Data Publication for Reusable Research . Code {4}lib Journal. <http://journal.code4lib.org/articles/10989>
- Archambeau, A. S. (2015). Data paper. Une incitation à la publication des données de biodiversité. Presented at the Formation GBIF-Ecoscope, Paris - France, 2015/09/16. retrieved from <http://slideplayer.fr/slide/9059893/>.
- Belter, C. (2015). Global-level data sets may be more highly cited than most journal articles. Retrieved from <http://blogs.lse.ac.uk/impactofsocialsciences/2014/05/15/global-level-data-sets-highly-cited/>
- Belter, C. (2014). Measuring the Value of Research Data: A Citation Analysis of Oceanographic Data Sets. *PLoS ONE*, 9(3), e92590. <http://dx.doi.org/10.1371/journal.pone.0092590>
- Bloom, T., Dallmeier-Tiessen, S., & Newbold, E. (n.d.). Publishing Data. Workflows. <https://www.icsu-wds.org/community/webinars/webinar-3/RDAWDSPublishingDataIGWebinar3Workflows.pdf>
- Blum, T., Burton, A., Callaghan, S., Dallmeier-Thiessen, S., Diepenbroek, M., Dillo, I., Hodson, S., Koers, H., Lehnert, K., Mokrane, M., Newbold, E., Smit, E., & Tedds, J. (2014). [Publishing Data: Scientific Data as Integral Part of Scholarly Publishing]. <http://docslide.us/documents/publishing-data-scientific-data-as-integral-part-of-scholarly-publishing-theodora.html>
- Borgman, C. L., *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press: 2015; p 1-383.
- Candela, L., Castelli, D., Manghi, P., & Tani, A. (2014). Data journals: A survey (preprint). *Journal of the Association for Information Science and Technology*, 1-20. <http://dx.doi.org/10.1002/asi.23358>
- Cosserat, F. (2015/10). Une introduction à la gestion et au partage des données de la recherche] [http://www.u1027.inserm.fr/medias/fichier/florence-cosserat\\_1444897589645-pdf?INLINE=FALSE](http://www.u1027.inserm.fr/medias/fichier/florence-cosserat_1444897589645-pdf?INLINE=FALSE)
- Dallmeier-Tiessen, S. (2015). Data Publishing models. <http://fr.slideshare.net/datascienceiqss/dallmeier-tiessen-datapublishingmodelssdt-dataverse>
- Dedieu, L. (2014). *Rédiger et publier un Data Paper en 5 points*. <http://url.cirad.fr/ist/data-paper>
- Gorgolewski, K. J., & Poldrack, R. A. (2016). A practical guide for improving transparency and reproducibility in neuroimaging research. 1-9. Retrieved from bioRxiv website: [10.1101/039354](https://doi.org/10.1101/039354)
- Gray, E. A. (2015). Case study: Publishing a data paper. 1-6. <http://data.bris.ac.uk/files/2015/05/Publishing-a-data-paper.pdf>
- Hrynaszkiwicz, I., & Shintani, Y. (2014). Scientific Data : An open access and open data publication to facilitate reproducible research. *Journal of Information Processing and Management*, 57(9), 629-640. <http://dx.doi.org/10.1241/johokanri.57.629>

# Bibliographie (2/2)

- Hole, B. (2015). Research Data Publishing. [http://fr.slideshare.net/brianhole/research-data-publishing?qid=0ac4b7d4-e1b3-4017-94ee-1d5bc482fe27&v=default&b=&from\\_search=29](http://fr.slideshare.net/brianhole/research-data-publishing?qid=0ac4b7d4-e1b3-4017-94ee-1d5bc482fe27&v=default&b=&from_search=29)
- Kenyon, J., & Sprague, N. R. (2014). Trends in the Use of Supplementary Materials in Environmental Science Journals. *Issues in Science and Technology Librarianship*. [10.5062/F40Z717Z](https://doi.org/10.5062/F40Z717Z)
- Kratz, J., & Strasser, C. (2014). Data publication consensus and controversies. *F1000Res*, 3, 94. [10.12688/f1000research.3979.2](https://doi.org/10.12688/f1000research.3979.2)
- Lin, J., & Strasser, C. (2014). Recommendations for the role of publishers in access to data. *PLoS Biol*, 12(10), e1001975. [10.1371/journal.pbio.1001975](https://doi.org/10.1371/journal.pbio.1001975)
- MacMillan, D. (2014). Data Sharing and Discovery: What Librarians Need to Know. *The Journal of Academic Librarianship*, 40(5), 541-549. [10.1016/j.acalib.2014.06.011](https://doi.org/10.1016/j.acalib.2014.06.011)
- Mounce, R. (2014). The state of Open Research Data. Presented at the OpenCon 2014: the Student and Early Career Researcher Conference on Open Access, Open Education and Open Data Washington, DC - USA, 2014/11/15-17. Retrieved from <http://fr.slideshare.net/RightToResearch/open-con-mouncedata-41594350>
- NISO, & NFAIS. (2013). *Recommended practices for online Supplemental Journal Article Materials*. [http://www.niso.org/apps/group\\_public/download.php/10055/RP-15-2013\\_Supplemental\\_Materials.pdf](http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf)
- Parsons, M. A. F., P.A. (2013). Is Data Publication the Right Metaphor ? <http://www2.nict.go.jp/isd/ISDS-contents/wds-kyoto-2011.org/pdf/IS703.pdf>
- Piwowar, H. A. (2007). Sharing detailed research data is associated with increased citation rate. <https://impactstory.org/HeatherPiwowar/product/bgt0op6c7pfx7cxmhen48fe8/fulltext>
- Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. <http://dx.doi.org/10.7717/peerj.175>
- Reilly, S., Schallier, W., Schrimpf, S., Smit, E., & Wilkinson, M. (2011). *Report on integration of data and publications* 1-87. [http://www.stm-assoc.org/2011\\_12\\_5\\_ODE\\_Report\\_On\\_Integration\\_of\\_Data\\_and\\_Publications.pdf](http://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf)
- Tenant, J. (2015) Open Sesame! The Risks and Rewards of Open Data for Researchers. <http://www.digital-science.com/blog/guest/open-sesame-the-risks-and-rewards-of-open-data-for-researchers/>