



Publier un Data Paper pour valoriser ses données

Dominique L'Hostis, Marjolaine Hamelin, Virginie Lelievre, Pascal Aventurier

► To cite this version:

Dominique L'Hostis, Marjolaine Hamelin, Virginie Lelievre, Pascal Aventurier. Publier un Data Paper pour valoriser ses données. Support de formation Infodoc Express (octobre 2016) (Infodoc Express), 2016, pp.66 slides. hal-02801638

HAL Id: hal-02801638

<https://hal.inrae.fr/hal-02801638>

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

L'Hostis, D., Hamelin, M., Lelievre, V., Aventurier, P.
Pôle Digitalist



Octobre 2016



L'Hostis, D., Hamelin, M., Lelievre, V., Aventurier, P. (2016). Publier un Data Paper pour valoriser ses données (Cours).

[10.15454/1.478247389988942E12](https://doi.org/10.15454/1.478247389988942E12)

<http://prodinra.inra.fr/record/375633>

ist@inra

Plan

Introduction

- Contexte Open Science, Contexte Inra

Pourquoi publier un Data Paper ?

- Définition, objectifs, et valeur ajoutée du Data Paper

Où ?

- Dans quelle revue publier un Data paper ?

Comment ?

- Structure du Data paper
- Métadonnées
- Processus éditorial

Après la publication

- Comment citer un jeu de données ?
- Un « Open data citation advantage » ?

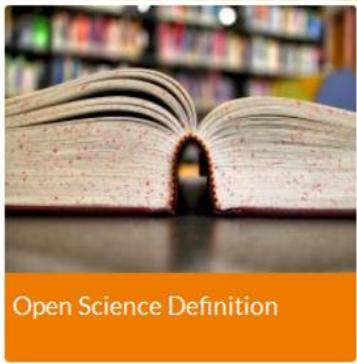
Conclusion

- Freins et leviers
- Un “paysage” éditorial en évolution...

Introduction - Contexte

ist@inra

Open Science



Open Science is the practice of science in such a way that others can collaborate and contribute, where **research data**, lab notes and other research processes are **freely available**, under terms that **enable reuse**, **redistribution** and **reproduction** of the research and its **underlying data and methods**.

<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

La politique de l'Inra



01

Le libre accès aux informations issues de travaux de recherche doit être systématiquement recherché

02

L'Inra organise la gestion et la valorisation de son information scientifique

Les données sont un patrimoine qu'il faut gérer et valoriser

La publication d'articles sur les données (**data papers**) est encouragée de même que le référencement des jeux de données par des identifiants numériques.

ist @ Inra

Outre le libre accès aux publications, l'engagement de l'Inra en faveur de l'**ouverture des données de la recherche** s'inscrit dans une double dynamique :

- La **mise à disposition des données** dans un objectif de transparence et de preuve en appui aux publications scientifiques, pour les citoyens, les journalistes, les acteurs de la société civile, les ONG, ou dans le contexte de recherches participatives...
- La **réutilisation des données**, dans un objectif de création de valeur et d'accélération de l'innovation.

Pourquoi publier un Data Paper ?

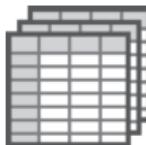
ist@inra

Données partagées vs Données publiées

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)

SHARED



AVAILABLE

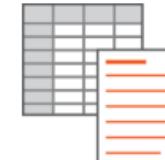
PUBLISHED



AVAILABLE



CITABLE



DOCUMENTED



VALIDATED

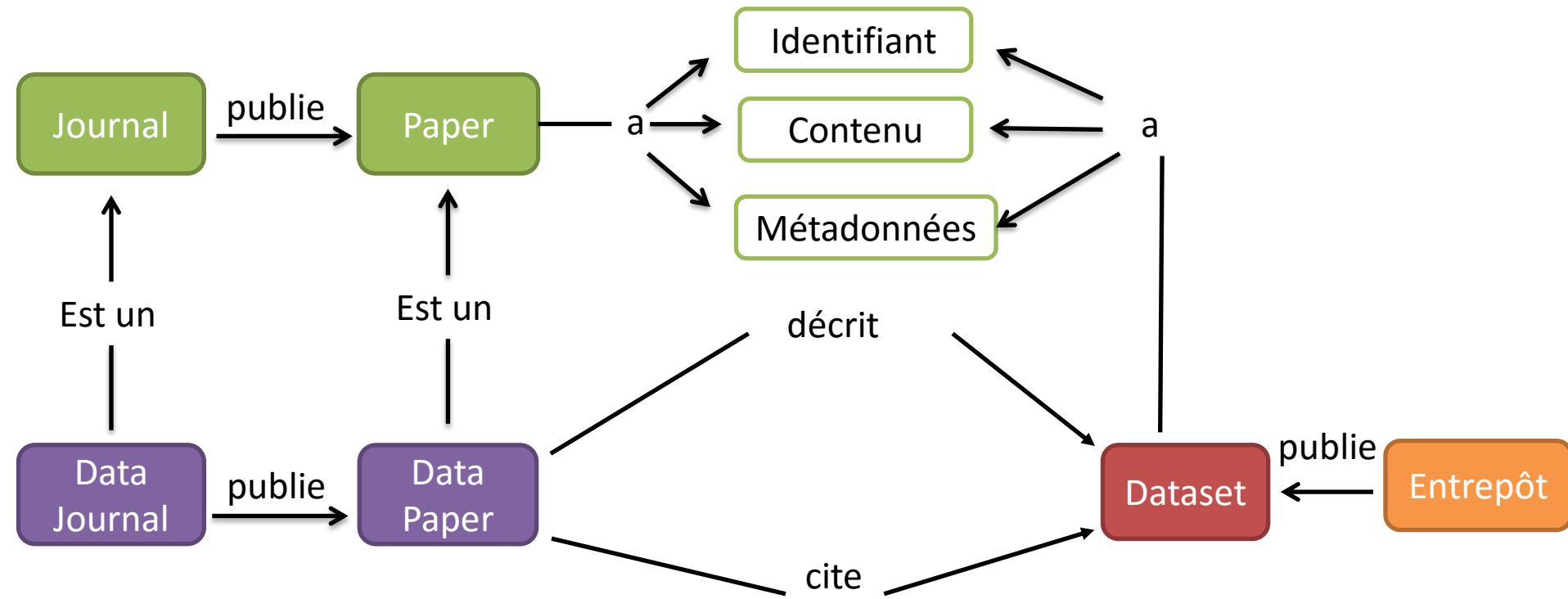
La publication de données traduit le fait que les données sont rendues disponibles au public (car déposées dans un entrepôt) et citables (grâce à un identifiant). Cela peut aussi impliquer que les données sont validées par un processus d'évaluation (Peer Review).

ist@inra

Plusieurs stratégies de publication des données



- **Données intégrées** dans un article classique
- **Données en matériel supplémentaire** d'un article classique
- **Données publiées** dans (ou associées à) un **Data paper**



Représentation graphique du concept de data paper. Traduit de :

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>

Voies de publication des données	Avantages	Limites
Données intégrées dans un article classique	<ul style="list-style-type: none"> • Intégration maximale des données et de l'article : citable, recherchable • Paternité des données / crédits aux auteurs 	<ul style="list-style-type: none"> • Données difficiles à trouver indépendamment de l'article et dans une forme peu ou pas réutilisable
Données en matériel supplémentaire d'un article	<ul style="list-style-type: none"> • Format des données libéré des contraintes de rédaction de l'article • Paternité des données / crédits aux auteurs 	<ul style="list-style-type: none"> • Taille souvent limitée • Peu de standardisation sur le signalement des fichiers «supplémentaires» • Identification des données indépendamment de l'article possible (via DOI) mais rare • Données difficiles à trouver indépendamment de l'article et dans une forme peu ou pas réutilisable

Voies de publication des données	Avantages	Limites
Données publiées dans (associées à) des Data Papers Liens réciproques entre l'article et les données (dépôt soit dans un entrepôt interne à la revue, soit dans un système externe)	<ul style="list-style-type: none"> • Paternité des données / crédits aux auteurs • Citation aisée • Réutilisation des données facilitée • Entrepôts reconnus par une communauté disciplinaire • Données normalisées, standardisées, conservées de façon pérenne • Pas de restriction en volume • Liens réciproques sécurisés 	<ul style="list-style-type: none"> • Interrogation sur la qualité du Peer-Review ? • Choix de l'entrepôt : disciplinaire (biologie, sciences de la vie, sciences du sol, chimie), ou générique (Zenodo ...).

Définition d'un data paper



Research Data Sharing
without barriers

Un **data article** aussi appelé *data descriptor* est un produit de publication de données, qui peut apparaître dans un **data journal** ou dans une revue académique classique. Quand les éditeurs parlent de *data publishing*, ils se réfèrent généralement au **data article** et non au jeu de données sous-jacent.

Contrairement aux articles de recherche classiques, les « data articles » ont pour but de rendre les données **accessibles, interprétables et réutilisables**, plutôt que tester des hypothèses ou présenter de nouvelles analyses.

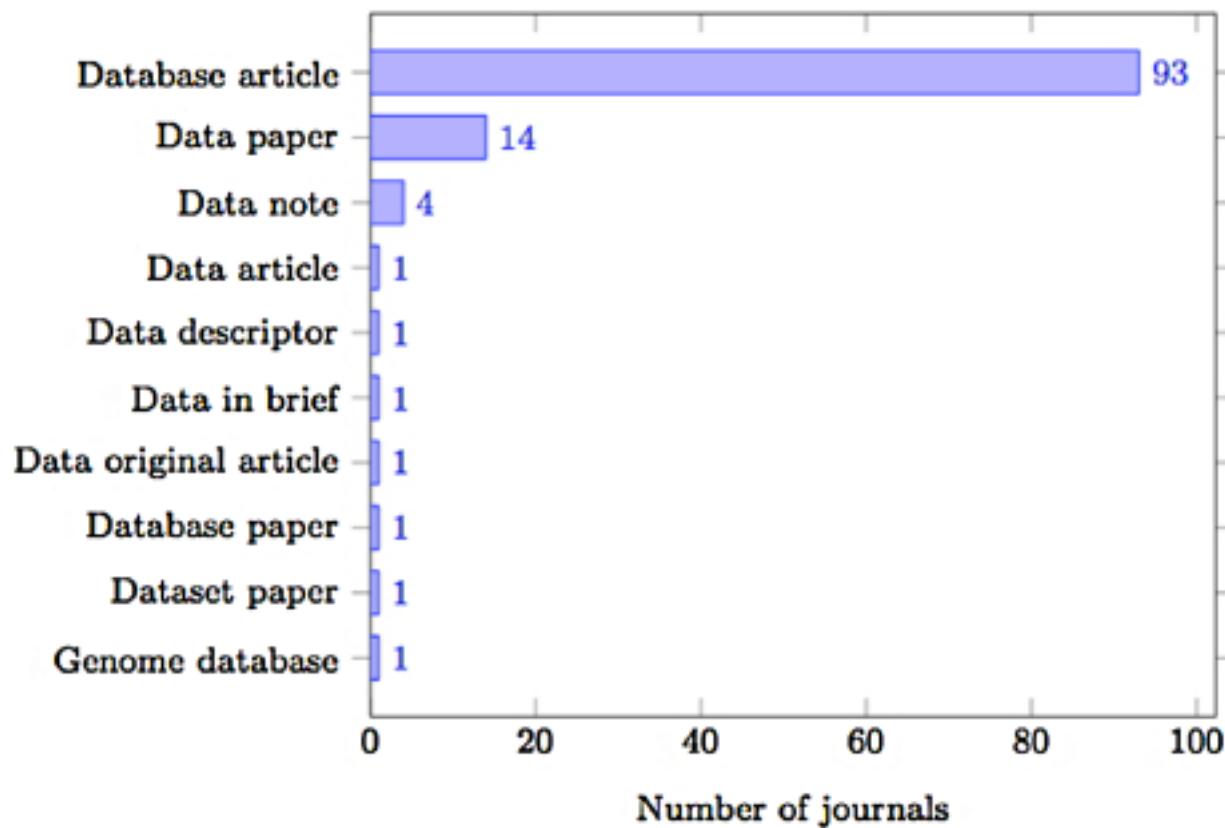
Qu'il soit lié à un jeu de données dans un entrepôt séparé ou soumis avec les données, **l'objectif du data article est de fournir une voie formalisée au partage des données.**

Traduit de

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

ist@inra

Différentes appellations



D'après Candela, L., Castelli, D., Manghi, P. and Tani, A. (2015), Data journals: A survey. J Assn Inf Sci Tec, 66: 1747–1762. doi: [10.1002/asi.23358](https://doi.org/10.1002/asi.23358)

Caractéristiques et valeur ajoutée

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

I INRA, UMR 1062 CBGP, F-34988 Montferrier-sur-Lez, France

Corresponding author: Alain Migeon (alain.migeon@supagro.inra.fr)

Academic editor: V. Petic | 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 (Montferrier-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 Papuaasia. The dataset constitutes today the most important one available on Tetranychidae worldwide.

Keywords

Acari, Tetranychidae, World, Madagascar, Western Indian Ocean, New Caledonia, South Pacific, Papuaasia

Data published through GBIF

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

ist@inra



GBIF Global Biodiversity Information Facility

Spider Mites Collection of Jean Gutierrez

Summary

PARTICLE: Spider Mites Collection of Jean Gutierrez

DESCRIPTION: The spider mites collection (Acari: Tetranychidae) has been set up by Jean Gutierrez, scientific director of the Institute of Research for Development from 1963 to 1999. The collection is deposited in CBGP (INRA, UMR 1062, Montferrier-sur-Lez, France). The collection is managed by Alain Migeon (IRD Montpellier Europe). The actual curator is Alain Migeon (IRD Montpellier Europe). The collection is located in the Western Indian Ocean, New Caledonia and Pacific Islands.

HOST: INRA, UMR 1062, Montferrier-sur-Lez, France

HOSTED BY: GBIF France

FORMAT: XML

ALTERNATIVE IDENTIFIERS:

Où publier un Data Paper ?

ist@inra

Où publier un Data Paper (1/2)

Dans un « Data Journal », dédié à ce type de publication

Data journal



Un « Data Journal » est un journal (toujours en libre accès) qui publie des articles de données « data articles ». Il fournit habituellement des **modèles** (templates) de description des données et guide les chercheurs sur les lieux de dépôt et sur la façon de décrire et de présenter leurs données.

Selon les cas, ces modèles peuvent être génériques ou spécifiques d'une discipline. Certaines revues ou leurs éditeurs maintiennent également leurs propres entrepôts.

Les Data Journals valorisent la liaison réciproque entre un article de données et le(s) jeu(x) de données (s) correspondant(s), développent et facilitent l'utilisation d'identifiants pérennes, mettent en place des procédures pour garantir la qualité des données (processus d'examen par les pairs – Peer review). Ils peuvent également fournir des lignes directrices éditoriales pour l'évaluation de la qualité des données.

Traduit de

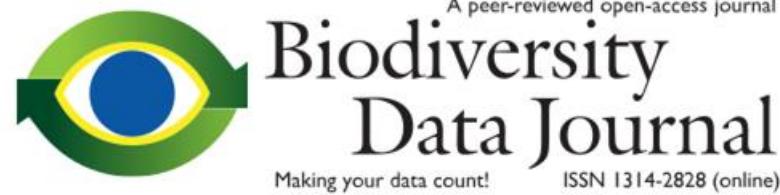
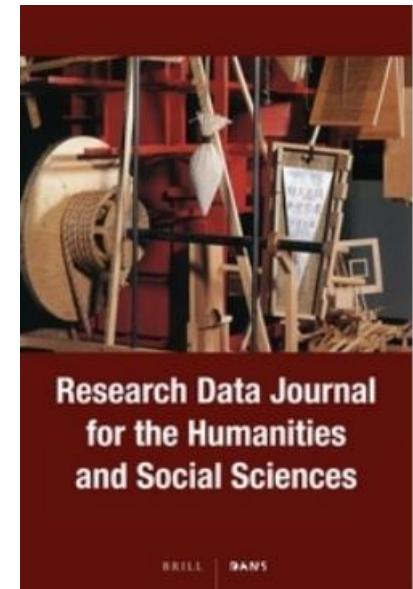
<https://rd-alliance.org/key-components-data-publishing-using-current-best-practices-develop-reference-model-data-publishin-0>
ist @ inra

Eléments clés de la publication de données

d'après Austin, C.C., Bloom, T., Dallmeier-Tiessen, S. et al. Int J Digit Libr (2016).
doi: [10.1007/s00799-016-0178-2](https://doi.org/10.1007/s00799-016-0178-2)

Services requis pour la publication	Services additionnels pour améliorer le contexte, la qualité, la visibilité			
<p>Données primaires avec un identifiant pérenne</p> <ul style="list-style-type: none">Entrepôt proposant un identifiant pérenneGénération de métadonnéesCurationDiffusion	<p>Contexte / objets liés</p> <ul style="list-style-type: none">Documentation richeLien vers l'article de rechercheLien vers le data paperLiens vers le logiciel, le code...	<p>Qualité</p> <ul style="list-style-type: none">Processus éditorialCuration par des experts du domaineRelecture par les pairsAide à la soumission	<p>Visibilité / accessibilité</p> <ul style="list-style-type: none">Indexation possibleServices à valeur ajoutée pour accéder aux donnéesAméliorer l'impact : plateforme de gestion de l'évaluation, agrégateurs de métriques, pages d'auteurs...	

Exemples de Data Journals



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

Où publier un Data Paper ? (2/2)

☁ Dans des revues classiques qui publient aussi des Data Papers

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:
SAGE
<http://www.sagepublications.com>



PLOS' New Data Policy: Public Access to Data
Posted February 24, 2014 by Liz Silva in Aggregators, Open Access

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

Issue

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

ist@inra

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](#)



Exemple : Annals of Forest Science

ANNALS OF FOREST SCIENCE CONTRIBUTES TO THE DEVELOPMENT OF OPEN SCIENCE AND OPEN DATA

By Erwin Dreyer On 15 August 2016 In Editorial News, For Authors, For Referees, Opinion

Erwin DREYER

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



GUIDELINES FOR REVIEWING A DATA PAPER

By Marianne Peiffer On 23 November 2015 In Editor's Choice, Editorial News, For Authors, For Referees

Un choix éditorial qui s'inscrit dans la démarche de l'Open Science et de l'Open Data

Un workflow dédié à la publication des Data Papers, et pour les auteurs, une procédure en 3 étapes :

- Déposer des données dans un entrepôt public, permettant l'identification des jeux de données avec un DOI (le dépôt d'un dataset dans une plateforme de base de données Inra, associé à un DOI Inra est une procédure innovante encouragée par AFS)
- Télécharger des modèles (templates) pour décrire les métadonnées (fichier « data-set_paper.xlsx » à compléter obligatoirement) et rédiger le manuscrit (<https://metadata-afs.nancy.inra.fr/ressources/>).
- Soumettre en ligne le manuscrit et la description des métadonnées via Editorial Manager

ist@inra

Une utilisation encore modeste...

What tools/sites do you use to publish?

787 out of 1000 people answered this question



Enquête internationale - Université Utrecht - 2016
Résultats sur les 1000 premières réponses
<https://101innovations.wordpress.com/tag/updates-insights/>

ist@inra

...mais en progression !

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 Data
Papers publiés, dont 23,5%
sur la dernière année
étudiée.

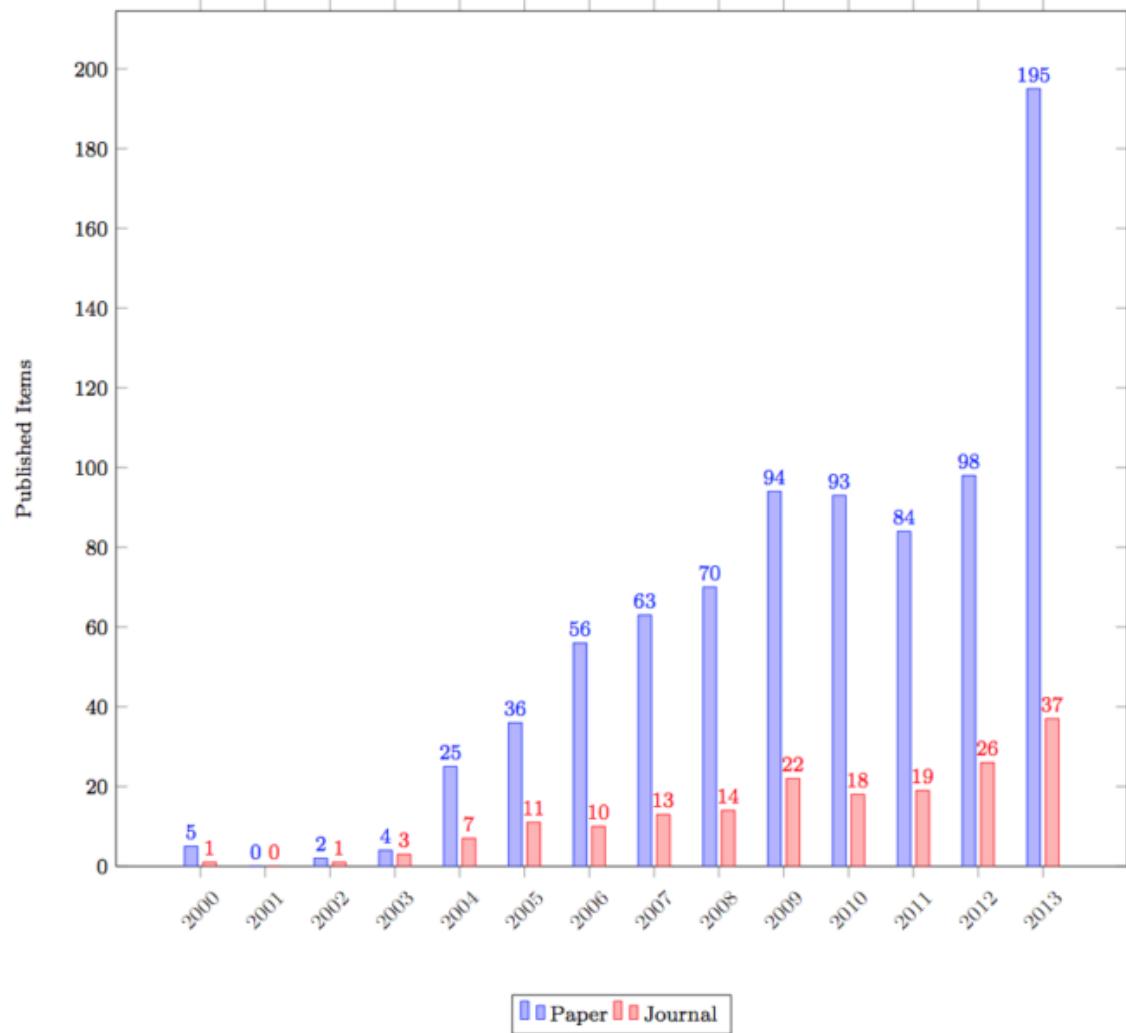


Figure 3. Number of published papers and operational data journals, by year.



Data Papers Inra

[Data Papers "INRA"](#)

mise à jour 10/11/2016

- 42 Références bibliographiques de 2009 à 2016
- 13 titres de journaux
- 26% dans "Data in Brief" - Elsevier

Titre	Nb	WOS	Pubmed
Data in Brief	11		
Ecology	7	+	+
Biodiversity Journal	4		
Earth System Science Data	3		
Genomics Data	3	+	+
Genome Announcements	2		+
Scientific Data	2		+
Zookeys	2	+	+
BMC Bioinformatics	1	+	+
Open Data Journal for Agricultural Research	1		
Phytokeys	1	+	+
Proteomics	1	+	+
Renewable and Sustainable Energy Reviews	1		+

ist@inra

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

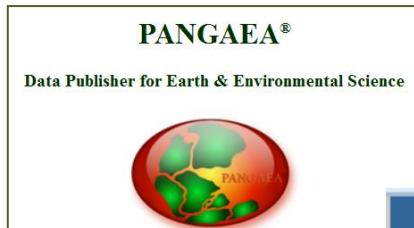
Candela, L., Castelli, D., Manghi, P., & Tani, A. (2014). Data journals: A survey (preprint). Journal of the Association for Information Science and Technology, 1-20. [10.1002/asi.23358](https://doi.org/10.1002/asi.23358)



[Une compilation INRA en cours d'élaboration](#)

ist@inra

Où déposer les données ?



ist@inra

Où déposer les données ?

☁ A partir du **choix de la revue de publication**, vérifier les recommandations

- les instructions aux auteurs préconisent souvent des entrepôts reconnus en fonction de la nature des données (génétique, biodiversité, écologie, géoscience, sciences humaines et sociales, etc.),
- certaines revues disposent de leur propre entrepôt de données
- certaines revues autorisent dans certaines conditions l'intégration des données au sein du Data Paper (exemple : Data In Brief, Biodiversity Data Journal)

☁ A défaut, le choix revient aux producteurs des données et rédacteurs du Data Paper

Comment trouver un entrepôt ? (1/2)

Many many options for open'ing data



(Mounce, 2014)

Entrepôts généralistes

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

ist@inra

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/>
- [Master Data Repository List](#)

En projet..., un site pour rassembler les politiques éditoriales et répondre aux questions du type :
“What, when, how, and where to deposit data”
[Journal Research Data Policy Registry \(JRDPR\)](#)

Comment trouver un entrepôt (2/2)

- Conseil : utiliser des annuaires ou répertoires



OpenDOAR

- [RE3DATA Registry of Research Data Repositories](#)
(~1728 enregistrements, 80 en France.
Chiffres 24/10/2016)
- [OpenDoar](#)
- [NIH Data Sharing Repositories](#)
- [Scientific Data Repositories and Datasets](#)
- [Wiki - Open Access Directory](#)
- [Base](#)
- [Research Data Australia](#)
- [Data Literature Interlinking Service](#)

Accès gratuit

ist@inra



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

MASTER DATA REPOSITORY LIST



Search for a data repository:
biology

Select boolean search type:
 AND
 OR

Search Tips:

* use the minus (-) operator to exclude a term from the search

Accès payant

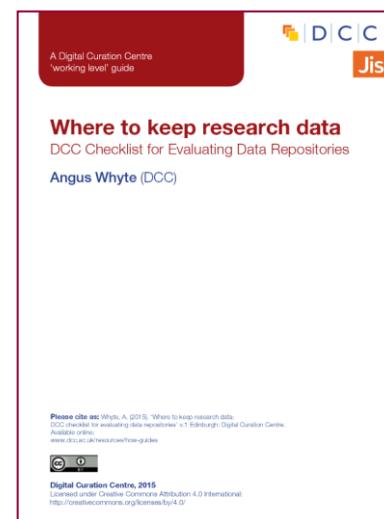
- [Master Data Repository List](#)
~ 322 entrepôts (sep 2016)

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



ist@inra

Une solution : Déposer ses données dans Zenodo-Inra

En l'absence d'un entrepôt spécialisé recommandé, et dans l'attente d'un portail des données INRA, le dépôt dans **Zenodo-Inra** est possible

- ☁ Zenodo : entrepôt de données européen géré par le CERN
<https://zenodo.org/about>
- ☁ gratuit, utilisable pour tous les résultats de la recherche (positifs et négatifs), dans tous les domaines scientifiques



Mémento à décharger

The screenshot shows a PDF document titled "Mémento à décharger" (Downloadable memento) for depositing data in Zenodo. It includes a green header bar with the INRA logo and a green footer bar. The main content area contains a bulleted list of steps for depositing data, followed by two screenshots of the Zenodo web interface: one showing the upload process and another showing the dataset details page.

ist@inra

Comment rédiger un Data Paper ?

ist@inra

Conseils généraux de rédaction

☁ Comme pour les autres types d'articles :

- Appliquer les instructions aux auteurs !
- Utiliser les articles déjà publiés dans la revue cible comme modèles

☁ Certaines revues proposent des **modèles** (templates)

- Exemple: Data in Brief, Nature Scientific Data...

http://www.elsevier.com/inca/publications/misc/dib_data%20article%20template.docx

http://www.nature.com/uploads/ckeditor/attachments/2963/Scientific_Data_Templates.zip



Data in Brief

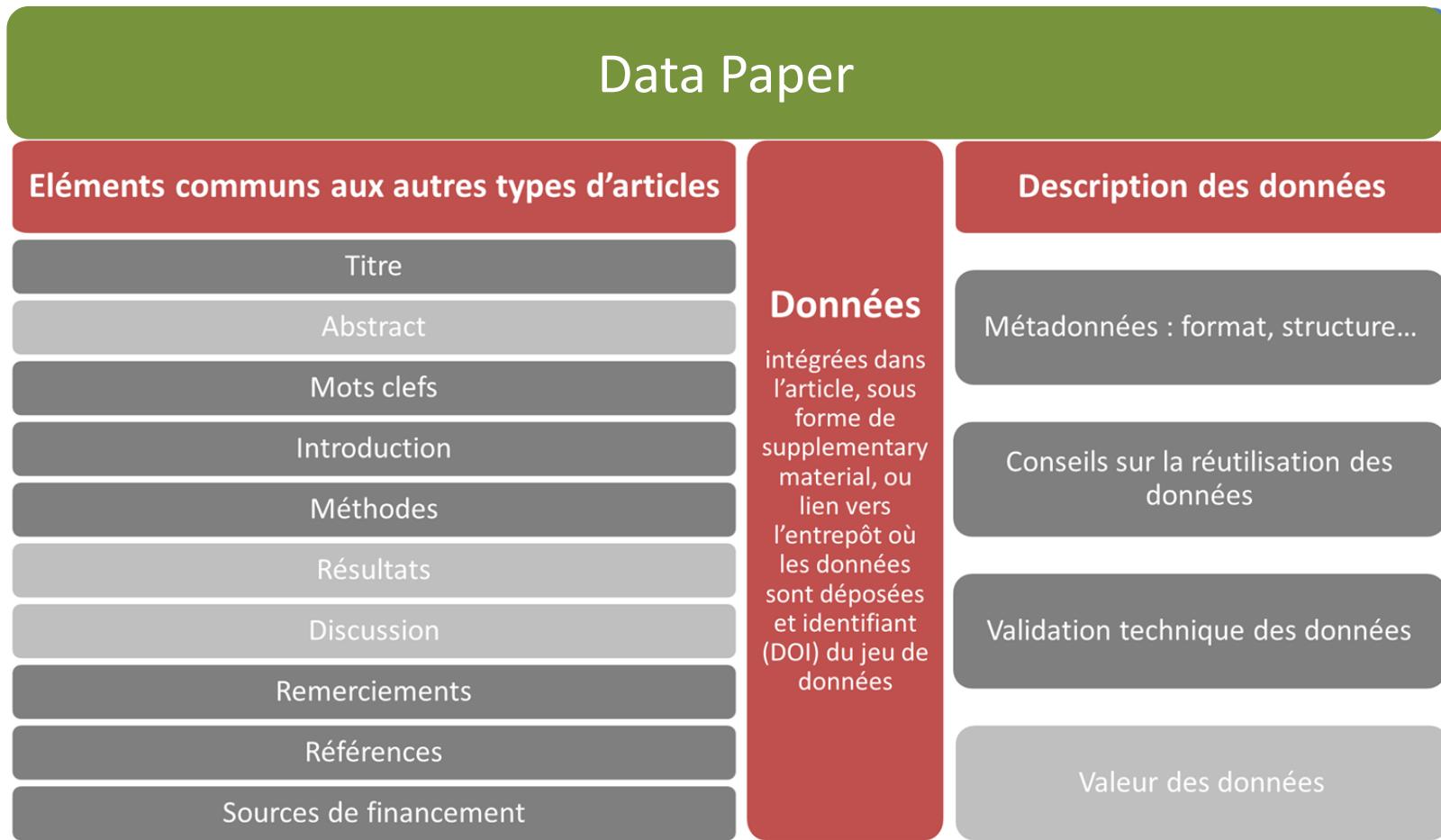
☁ Certaines revues (i.e. Biodiversity Data Journal) n'acceptent pas les manuscrits en Word et proposent **leur propre outil de rédaction en ligne** (Pensoft Writing Tool) ou demandent de soumettre depuis des **plateformes externes** (Scratchpads, GBIF Integrated Publishing Toolkit)



ist@inra

Structure type du data paper

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



Structure globale du data paper

Eléments communs aux articles classiques



Abstract : Sa structure est propre à chaque revue. Il peut contenir :

- Une **présentation du contexte** d'obtention des données (front de recherche, question de recherche)
- **Le lien éventuel du data paper avec un article de recherche** analysant le même jeu de données
- La **méthode d'obtention** des données (matériels et méthodes)
- La **description des données**
- (Le **résultat de l'analyse** de ces données)
- **Les informations d'accès** aux données si elles ne sont pas intégrées dans l'article

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

DATA PAPER

ZooKeys
Launched to accelerate biodiversity research

The Jean Gutierrez spider mite collection

Alain Migeon¹

¹ INRA, UMR 1062 CBGP, F-34988 Montferrier-sur-Lez, France

Corresponding author: Alain Migeon (alain.migeon@supagro.inra.fr)

Academic editor: V. Pešic | Received 28 January 2015 | Accepted 10 March 2015 | Published 23 March 2015
<http://zoobank.org/0719C382-98BD-4DCC-BB24-00E65E32C9C5>

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 (Montferrier-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 Papuaasia. The dataset constitutes today the most important one available on Tetranychidae worldwide.

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

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

→ Se conformer aux instructions, et prendre des abstracts publiés comme modèle.

NB : Dans certaines revues comme Frontiers, pas d'abstract

ist@inra

Abstract du data paper

Exemple 1 : Data paper contenant le jeu de données

Contexte
(front de recherche)



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 (Montferrier-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 Papuasia. The dataset constitutes today the most important one available on Tetranychidae worldwide.

Description
des données



Couverture
spatiale des
données



Nouveauté
- intérêt du jeu de
données



Migeon A (2015) The Jean Gutierrez spider mite collection. ZooKeys 489: 15-24. doi: [10.3897/zookeys.489.9292](https://doi.org/10.3897/zookeys.489.9292)

ist@inra



Abstract du data paper

Exemple 2 : Data paper ne contenant pas le jeu de données

Contexte
(front de recherche)



Méthode
d'obtention
des données



Description
des données



Résultats obtenus
par l'analyse des
données



Nouveauté/intérêt
du jeu de données



Information sur
l'accès aux données
(ici non incluses
dans l'article)



Aspergillus flavus is one of the predominant causative organisms of mycotic keratitis in tropical parts of the world. Extracellular proteins are the earliest proteins that come in contact with the host and have a role in the infection process. Exoproteins of *A. flavus* isolated from infected cornea, sputum and a saprophyte were pooled and identified using high resolution mass spectrometry in order to get the total exoproteome from cultures isolated from different sources. A total of 637 proteins was identified from the pooled *A. flavus* exoproteome. Analysis based on GO annotations of the 637 identified proteins revealed that hydrolases form the predominant class of proteins in the exoproteome. Interestingly, a greater proportion of the exoproteins seem to be secreted through the non-classical pathways. This data represent the first in-depth analysis of the representative *A. flavus* exoproteome of a large set of isolates from distinct sources. This data have been deposited to the ProteomeXchange with identifier PXD001296.

Selvam, R. M. et al. (2015). Data set for the mass spectrometry based exoproteome analysis of *Aspergillus flavus* isolates. Data in brief, 2, 42-47.
<http://dx.doi.org/10.1016/j.dib.2014.12.001>



Abstract du data paper

Exemple 3 : Data paper lié à un article de recherche

Lien avec l'article de recherche associé au jeu de données



The data presented in this article are related to the research article entitled “Ant abundance and *Cosmopolites sordidus* damage in plantain fields as affected by intercropping” (A.G. Dassou, D. Carval, S. Dépigny, G.H Fansi, P. Tixier, 2015) [1].

Description des résultats liés à l'analyse des données



This article describes how associated crops maize (*Zea mays*), cocoyam (*Xanthosoma sagittifolium*) and bottle gourd (*Lagenaria siceraria*) intercropped in the plantain fields in Cameroun modify ant community structure and damages of banana weevil *Cosmopolites sordidus*. The field data set is made publicly available to enable critical or extended analyzes.

Information sur l'accès aux données (ici incluses dans l'article)

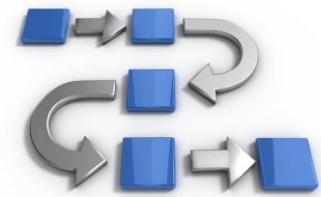


Dassou, A. G., Carval, D., Dépigny, S., Fansi, G., & Tixier, P. (2016). Dataset on the abundance of ants and *Cosmopolites sordidus* damage in plantain fields with intercropped plants. Data in Brief, 9, 17-23.

<http://dx.doi.org/10.1016/j.dib.2016.08.027>

Structure globale du data paper

Eléments communs aux articles classiques



Introduction

- Présentation du contexte de l'obtention des données, du front de recherche, de la question de recherche, des objectifs de l'étude...

Methods

- Description de la méthode d'obtention des données
- Peut se diviser en sous-sections: étapes de la méthode, procédures utilisées, design expérimental, méthode d'échantillonnage, contrôle qualité effectué, incertitudes...

References

Acknowledgements

Funding information

Dans certains cas : Results, Discussion

(Frontiers in Microbiology, Frontiers in Plant Science...)

ist@inra

Structure globale du data paper

Données

- Intégrées à l'article : sous forme de tableaux, de liste bibliographique, ou encore [au format XML pour permettre la fouille de texte](#)
- Jointes à l'article sous forme de matériel supplémentaire

Table 1. Tetranychidae species described by Jean Gutierrez with respective title and source of the publication and present nomenclature (Blommers and Gutierrez 1975; Gutierrez 1966; 1967a; b; 1968a; b; 1969; 1970; 1972; 1977; 1978; 1982; Gutierrez and Bolland 1973; 1986; Gutierrez and Etienne 1981; Gutierrez and Helle 1971). The presence of the type specimen and the number of types and paratypes in the dataset is also indicated.

Original genus	Species	Author	Present combination	Publication title	Publication source	Type	Number of types and paratypes specimens
<i>Eonychus</i>	<i>grewiae</i>	Gutierrez, 1969	<i>Eonychus grewiae</i>	Tetranychidae nouveaux de Madagascar (Cinquième note)	Acarologia, 11: 43–64	yes	17
<i>Eotetranychus</i>	<i>befandrianae</i>	Gutierrez, 1967	<i>Eotetranychus befandrianae</i>	Huit nouvelles espèces du genre <i>Eotetranychus</i> Oudemans (Acariens : Tetranychidae) de Madagascar	Acarologia, 9: 370–394	yes	15
<i>Eotetranychus</i>	<i>borbonensis</i>	Gutierrez, 1968	<i>Eotetranychus borbonensis</i>	Note sur quelques acariens phytophages de l'Île de la Réunion avec description d'une nouvelle espèce du genre <i>Eotetranychus</i> Oudemans (Tetranychidae)	Acarologia, 10: 444–446	yes	47
<i>Eotetranychus</i>	<i>botryanthae</i>	Gutierrez, 1970	<i>Eotetranychus botryanthae</i>	Tetranychidae nouveaux de Madagascar (Sixième note)	Acarologia, 12: 714–731	yes	23
<i>Eotetranychus</i>	<i>capricorni</i>	Gutierrez, 1967	<i>Eotetranychus capricorni</i>	Huit nouvelles espèces du genre <i>Eotetranychus</i> Oudemans (Acariens : Tetranychidae) de Madagascar	Acarologia, 9: 370–394	yes	8
<i>Eotetranychus</i>	<i>friedmanni</i>	Gutierrez, 1968	<i>Eotetranychus friedmanni</i>	Tetranychidae nouveaux de Madagascar (Quatrième note)	Acarologia, 10: 13–28	yes	105
<i>Eotetranychus</i>	<i>garnieri</i>	Gutierrez, 1978	<i>Eotetranychus garnieri</i>	Cinq nouvelles espèces de Tetranychidae (Acariens) de Nouvelle-Calédonie	Acarologia, 20: 351–364	yes	19
<i>Eotetranychus</i>	<i>grandis</i>	Gutierrez, 1969	<i>Eotetranychus grandis</i>	Tetranychidae nouveaux de Madagascar (Cinquième note)	Acarologia, 11: 43–64	yes	10
<i>Eotetranychus</i>	<i>greveanae</i>	Gutierrez, 1970	<i>Eotetranychus greveanae</i>	Tetranychidae nouveaux de Madagascar (Sixième note)	Acarologia, 12: 714–731	yes	10
<i>Eotetranychus</i>	<i>limoni</i>	Blommers & Gutierrez, 1975	<i>Eotetranychus limoni</i>	Les tétranyques vivant sur agrumes et avocatiers dans la région de Tamatave (Madagascar-est) et quelques-uns de leurs prédateurs	Fruits, 30: 191–200	yes	23
<i>Eotetranychus</i>	<i>paracybelus</i>	Gutierrez, 1967	<i>Eotetranychus paracybelus</i>	Huit nouvelles espèces du genre <i>Eotetranychus</i> Oudemans (Acariens : Tetranychidae) de Madagascar	Acarologia, 9: 370–394	yes	50

Migeon A (2015) The Jean Gutierrez spider mite collection. ZooKeys 489: 15–24. doi:
[10.3897/zookeys.489.9292](https://doi.org/10.3897/zookeys.489.9292)

Structure globale du data paper

Données

- Cloud icon: Déposées dans un entrepôt ou une base de données, dans le cas de grands jeux de données.
- Cloud icon: Dépôt obligatoire pour certaines revues (e.g. Frontiers, Annals of Forest Science...), mais optionnel pour d'autres (e.g. revues Pensoft, Data in Brief, Scientific Data...)
- Cloud icon: L'identifiant des données (DOI) permet d'établir le lien vers les données

The diagram illustrates the relationship between a journal article abstract and its corresponding data set. On the left, a screenshot of the journal 'ECOLOGY' shows an article titled 'Grassland species composition and biogeochemistry in 153 sites along environmental gradients in Europe'. An orange arrow points from this abstract to the right side of the diagram. On the right, a screenshot of the figshare platform shows the same article as a data set. The title 'Grassland species composition and biogeochemistry in 153 sites along environmental gradients in Europe' is highlighted with an orange box. Another orange arrow points from the figshare page back to the journal abstract. Both pages include the logo of the Ecological Society of America (ESOA) in the bottom right corner.

ECOLOGY
A PUBLICATION OF THE ECOLOGICAL SOCIETY OF AMERICA
Explore this journal >

Data Paper
Grassland species composition and biogeochemistry in 153 sites along environmental gradients in Europe
Ecological Archives E092-128

Carly J. Stevens, Cecelia Dupré, Edu Dorland, Cassandra Gaudnik, David J. Gowing, Martin Diekmann, Didier Alard, Roland Bobbink, Emmanuel Corcket, J. Owen Mountford, Vigdís Vandvik, Per Arild Arrestad, Serge Muller, Nancy B Dise

First published: 1 July 2011 Full publication history
DOI: 10.1890/11-0115.1 View citation
Cited by: 5 articles Citation tools

Abstract

This data set consists of vascular plant and bryophyte species composition and plant and soil biogeochemical data from 153 acid grasslands located in the Atlantic biogeographic region of Europe. Data were collected between 2002 and 2007. The grasslands all belong to the *Violion caninae* association and were managed by grazing or cutting but had not received fertilizer inputs. These data provide plant composition from five randomly located 2 x 2 m quadrats at each site with all vascular plants and bryophytes identified to species level with cover estimates for each species. Topsoil and subsoil were collected in each quadrat, and data are provided for pH, metal concentrations, nitrate and ammonium concentrations, total carbon and N, and Olsen extractable phosphorus. Aboveground plant tissues were collected for three species (*Rhytidiodelphus squarrosus*, *Galium saxatile*, and *Agrostis capillaris*), and data are provided for percentage N, carbon, and phosphorus. These data have already been used in a number of research papers focusing on the impacts of atmospheric N deposition on grassland plant community and biogeochemistry. The unique data set presented here provides the opportunity to test theories about the effect of environmental variation on plant communities, biogeochemistry, and plant-soil interactions, as well as spatial ecology and biogeography.

The complete data sets corresponding to abstracts published in the Data Papers section of the journal are published electronically in Ecological Archives at (<http://esapubs.org/archive>). (The accession number for each Data Paper is given directly beneath the title.)

figshare
search on figshare
Browse Upload Sign up Log in

Grassland species composition and biogeochemistry in 153 sites along environmental gradients in Europe
Published on 09 Aug 2016 - 12:30 (GMT)
0 views
READ THE ORIGINAL, PEER-REVIEWED ARTICLE →
Grassland species composition and biogeochemistry in 153 sites along environmental gradients in Europe

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

AUTHORS
Carly J. Stevens
Cecelia Dupré
Edu Dorland
Cassandra Gaudnik
David J. Gowing
Martin Diekmann
Didier Alard
Roland Bobbink
Emmanuel Corcket
J. Owen Mountford
Vigdís Vandvik
Per Arild Arrestad
Serge Muller
Nancy B Dise

Description des données

Sections spécifiques au Data paper



On appelle **métadonnées** l'ensemble des éléments descriptifs permettant de chercher et trouver le jeu de données, juger de sa qualité/fiabilité, et pouvoir le comprendre ou le réutiliser dans un autre contexte :

- **Description des données :**

- type de données, structure de la base de données, couverture géographique, temporelle, taxonomique, licence...

- **Potentiel de réutilisation des données, et limites**

- instructions pour faciliter la réutilisation des données.

- **Valeur des données**

- intérêt du jeu de données, nouveauté

- **Validation technique / contrôle qualité**

- méthodes utilisées pour valider la technique d'obtention des données

Focus sur les métadonnées (2)

Standards de description des jeux de données

- génériques, disciplinaires ou technologiques
- exemples :
 - Données en Ecologie :
[Ecological metadata language](#)
 - Données en Sciences humaines et sociales:
[Data documentation initiative](#)

Contact us Search

 because good research needs good data

Home Digital curation About us News Events Resources Training Projects Community Tailored support

Home > Resources > Metadata Standards

In this section

- Briefing Papers
- How-to Guides & Checklists
- Developing RDM Services
- Curation Lifecycle Model
- Curation Reference Manual
- Policy and legal
- Data Management Plans
- Tools
- Case studies
- Repository audit and assessment

Standards

- Disciplinary Metadata**
- DIFFUSE
- Publications and presentations
- Roles
- Curation journals
- Informatics research
- External resources
- Online Store

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Please note that a community-maintained version of this directory [has been set up under the auspices of the Research Data Alliance](#).

Search by Discipline


Biology


Earth Science


General Research Data


Physical Science


Social Science & Humanities

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

Description des données

Tableaux récapitulatifs : description synthétique du jeu de données

Placés souvent après l'abstract

Exemples : Data in Brief, Scientific Data

Specifications Table [please fill in right-hand column of the table below]

Subject area	<i>Physics, Chemistry, Biology, Economics, Psychology</i>
More specific subject area	<i>Describe narrower subject area</i>
Type of data	<i>Table, image (x-ray, microscopy, etc), textfile, graph, figure</i>
How data was acquired	<i>Microscope, survey, SEM, NMR, mass spectroscopy, etc. If an instrument was used, please provide the model and make of the instrument</i>
Data format	<i>Raw, filtered, analyzed, etc</i>
Experimental factors	<i>Brief description of any pretreatment of samples</i>
Experimental features	<i>Very brief experimental description</i>
Data source location	<i>City, Country and/or Latitude & Longitude (& GPS coordinates) for collected samples/data if applicable</i>
Data accessibility	<i>State if data is with this article or in public repository. If public repository, please explicitly name repository and data identification number and provide a direct URL to data</i>

<https://www.elsevier.com/connect/can-data-be-peer-reviewed>

Design Type(s)	database creation objective • data integration objective
Measurement Type(s)	crop production measures
Technology Type(s)	data item extraction from journal article
Factor Type(s)	
Sample Characteristic(s)	Fabaceae

[Cernay, C. et al. A global experimental dataset for assessing grain legume production. Sci. Data 3:160084 doi: 10.1038/sdata.2016.84 \(2016\).](https://doi.org/10.1038/sdata.2016.84)

Description des données

Exemples (1/2)

[Ecology / Ecological Archives \(Ecological Society of America, ESA\)](#)



- **Introduction** : contexte, question de recherche et objectifs de l'étude,
- **Data set descriptors** : description du jeu de données
- **Data structural descriptors** : description du format des fichiers
- **Data set status and accessibility** : description du statut, de la qualité et de l'accessibilité aux données (libre, restrictions d'accès, coût),
- **Research origin descriptors** : contexte de l'étude, méthodes et procédures

[Biodiversity Data Journal \(Pensoft Publishers\)](#)



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

- **Introduction** : contexte, question de recherche et objectifs de l'étude,
- **Taxonomic Coverage, Spatial Coverage, Temporal Coverage** : taxons, zone géographique et période couvertes par les données
- **Project Description** : titre du projet, personnes impliquées, sources de financement, description de la zone d'étude et du design,
- **Methods** : étapes, échantillonnage, contrôle qualité
- **Dataset Descriptions** : nom du jeu de donné, encodage, nom et version du format, lien (URL), date de publication, langue, propriété intellectuelle

Description des données

Exemples (2/2)

Genomics Data (Elsevier)

- Specification table : tableau de Spécifications des données,
- Experimental design, Materials and Methods,
- Discussion : discussion courte mettant en valeur la portée du jeu de données.



Scientific Data (Nature Publishing Group)

SCIENTIFIC DATA

- **Background & Summary** : contexte scientifique, question de recherche, objectifs de l'étude, valeur et potentiel des données pour leur réutilisation,
- **Methods** : méthodes, procédures et analyses. Le protocole expérimental peut être présenté sous forme de figure et les variables dans un tableau,
- **Data Records** : description des données et de l'entrepôt où elles sont déposées, format des fichiers. La présentation sous forme de tableaux est recommandée,
- **Metadata Records** : métadonnées présentées dans des tableaux (modèle fourni) : processus expérimental, méthode d'échantillonnage, description du site, conditions d'expériences, protocoles suivis, données obtenues...
- **Technical Validation** : explication de la rigueur scientifique de l'étude et de la qualité technique des données,
- **Usage Notes** : facultatives, indications facilitant la réutilisation des données par d'autres scientifiques.

<http://coop-ist.cirad.fr/aide-a-la-publication/rediger/data-paper/4-exemples-de-structure-de-data-papers-en-sciences-du-vivant>

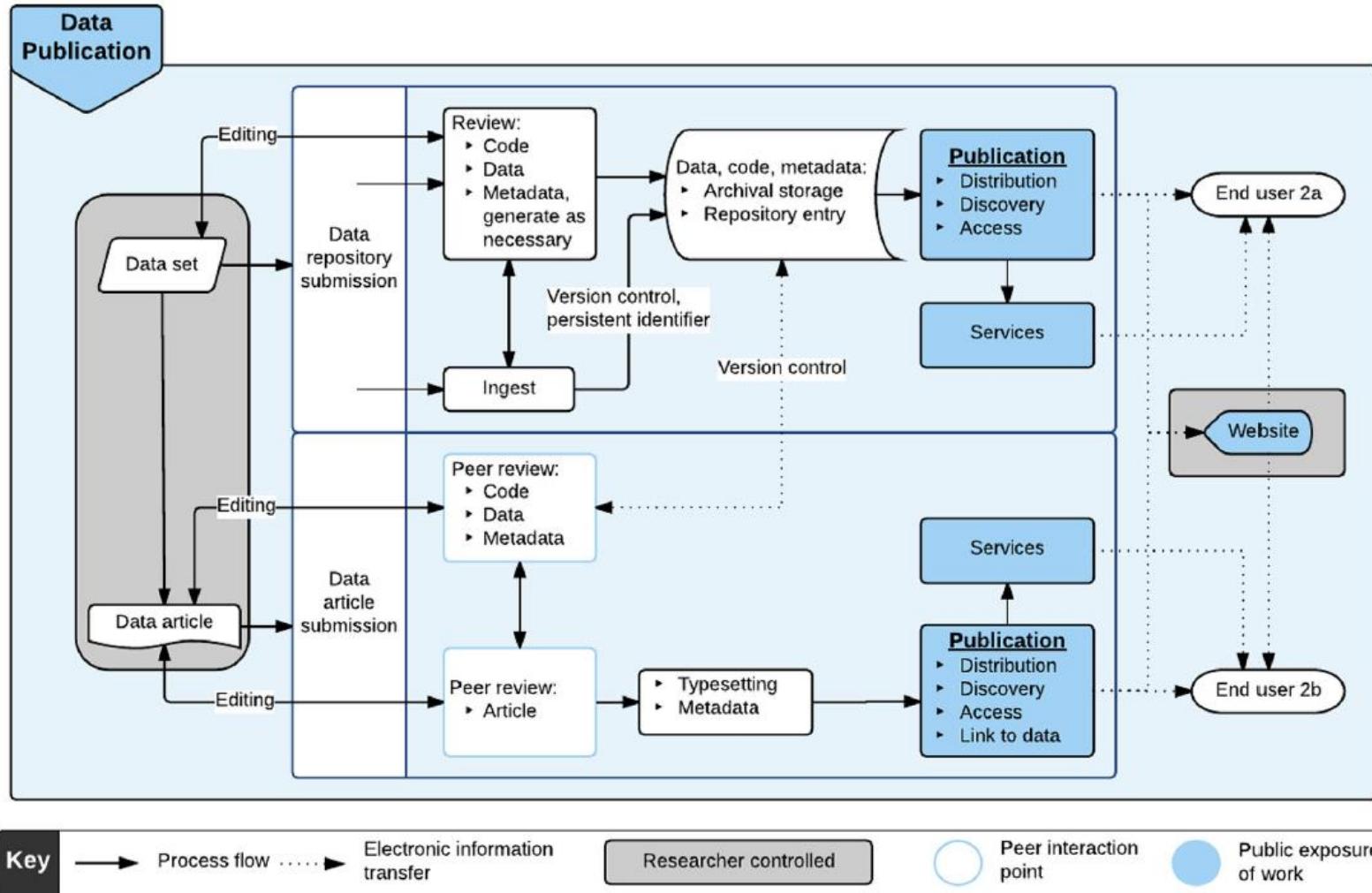
Soumission du data paper

Même processus que pour la soumission d'un article de recherche :

1. Choisir la **revue cible**. S'assurer que la revue effectue le **peer-review des données**
2. Identifier l'**entrepôt cible** pour le dépôt des données * (recommandé par la revue ou adapté au type de données)
3. Déposer ses données dans l'**entrepôt** *
4. Rédiger le data paper
5. Préparer la **cover letter** (instructions spécifiques, le cas échéant, cf Frontiers)

* étapes facultatives dans le cas où les données sont intégrées à l'article

Publication du data paper



[Austin, C.C., Bloom, T., Dallmeier-Tiessen, S. et al. Int J Digit Libr \(2016\). doi:10.1007/s00799-016-0178-2](https://doi.org/10.1007/s00799-016-0178-2)

ist@inra

Critères d'évaluation du data paper



- **Importance** et **originalité** des données
- **Potentiel** et **valeur de réutilisation** des données
- **Qualité** et **fiabilité** des données : structure de la base de données, organisation logique des données, intégrité des données (vérification des erreurs potentielles)
- **Accès** aux données : point important. L'auteur doit s'assurer que ses données sont toujours accessibles dans l'entrepôt. Si les données sont retirées de l'entrepôt, la rétraction du data paper pourra être décidée par l'éditeur
- Qualité et rigueur de la **méthode de collecte** des données : méthode appropriée, actuelle, suffisamment claire pour permettre la reproductibilité
- Choix des **métadonnées descriptives** et **formats** : présentation, complétude, degré de précision...
- Autres critères “classiques” : qualité générale du manuscrit, citations appropriées, respect des instructions...

Certaines revues affichent leurs critères d'évaluation de data papers :

- <https://www.elsevier.com/journals/data-in-brief/2352-3409>
- http://www.esapubs.org/archive/instruct_d.htm

Procédure d'évaluation du data paper

Exemple : Annals of Forest Science



Critères d'évaluation de la qualité du manuscrit

- Le DOI du jeu de données est-il fourni ?
- Le fichier de métadonnées est-il complété ?
- Le titre et le message clef reflètent-ils bien le contenu de l'article ?
- Les éléments non textuels (tableaux, figures...) sont-ils pertinents ?
- Les abréviations et symboles sont-ils définis ?
- Le contexte des recherches antérieures est-il bien défini (citation de jeux de données et articles pertinents antérieurs) ?
- Les conflits d'intérêt, autorisations et autres questions éthiques sont-ils abordés de façon appropriée ?

Critères d'évaluation de la qualité du jeu de données

- Les indications apportées dans le manuscrit permettent-elles de reproduire les données ?
- Les méthodes utilisées pour obtenir les données sont-elles suffisamment décrites ?
- Les sources d'erreur potentielles ont-elles été présentées dans le protocole ou l'article ?
- Les données sont-elles cohérentes et décrites à l'aide de normes applicables (format de fichier, nom de fichier, taille de fichier, unités, métadonnées) ?
- Les données couvrent-elles un domaine scientifique, géographique, un groupe taxonomique ou une période suffisamment importante pour faire l'objet d'une publication séparée ?
- Les méthodes d'analyse des données brutes sont-elles suffisamment décrites pour être répliquables ?
- Les données sont-elles plausibles au vu du protocole ?
- Le jeu de données est-il complet ?

Procédure d'évaluation du data paper

Exemple : Annals of Forest Science (2)



Critères d'appréciation du caractère réutilisable du jeu de données

- Le manuscrit fournit-il une description précise du jeu de données ?
- Le manuscrit décrit-il correctement la façon d'accéder aux données ?

Valeur et utilité des données

- Le jeu de données est-il suffisamment original pour mériter une publication dans un data paper ?
- Les utilisations du jeu de données décrites dans l'article sont-elles cohérentes avec les données ?
- Les affirmations établies dans l'article sont-elles toutes bien justifiables par le jeu de données ?

Après la publication ...

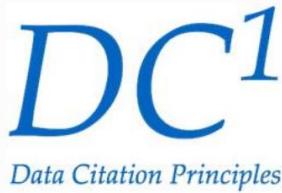
ist@inra

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 |

<http://www6.inra.fr/datapartage/Reutiliser/Citer-des-donnees>



DataCite Metadata Schema for the Publication and Citation of Research Data (2015)

http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf

- modèle "minimal" - 5 éléments
Creator (PublicationYear): Title. Publisher. Identifier.
- modèle incluant des propriétés optionnelles
Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier.



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

ist@inra

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.

Ani 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

Jump to...

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

Jump to...

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

- Identifier: doi:[10.5285/639A3714-BC74-46A6-9026-64931F355E07](https://doi.org/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. Weight, 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

ist@inra

References

- 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](https://doi.org/10.1029/2007RS003793).
- 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 Intl. Wksp. Satellite Communication from Fade Mitigation to Service Provision, Noordwijk, The Netherlands, May 2003.
- Goldshtein 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](https://doi.org/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-FE89225D57DD](https://doi.org/10.5285/E8F43A51-0198-4323-A926-FE89225D57DD).
- 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](https://doi.org/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-cced9b921390](https://doi.org/10.5285/db8d8981-1a51-4d6e-81c0-cced9b921390).
- Ventouras S, Callaghan SA, Wrench CL. 2006. Long-term statistics of tropospheric attenuation from the Ku/U band ITALSAT satellite experiment in the United Kingdom. *Radio Science* 41: RS2007, doi:[10.1029/2005RS003252](https://doi.org/10.1029/2005RS003252).
- [CrossRef](#). [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)

Freins et leviers du partage des données

Bénéfices

- Favorise la reproductibilité / validation de la recherche
- Réduction des erreurs et des fraudes
- Réduction de la duplication de travaux de recherche
- Favorise l'émergence de nouvelles hypothèses de recherche
- Utilisation pour l'enseignement
- Gain de citation
- Respect des mandats des financeurs et des éditeurs (organismes)
- Meilleur retour sur investissement de la recherche

Barrières et freins

- Coûts associés au partage de données
- Problèmes de confidentialité (données sensibles ...)
- Interrogation sur la notion de propriété des données ...
- Inquiétudes sur une réutilisation inappropriée
- Manque de reconnaissance académique et institutionnelle
- Méconnaissance des procédures (réécriture des métadonnées, choix entrepôts ...)

d'après (Hrynaszkiewicz & Shintani, 2014) et (Tenant, 2015)

Freins et leviers du Data paper

Bénéfices

- Valorise les jeux de données
- Notion de données sur les données “*data about data*”
- Présence et comptage dans les bases de données (Web of Science)
- Evalué comme une publication
- Peer review (data paper et jeu de données)

Barrières et freins

- Activité chronophage
- Coût de la publication ([voir liste](#))



- Publier un Data Paper n’empêche pas de publier un article de recherche
- C’est une façon de faire connaître ou reconnaître un travail de publication de données dans un système d’évaluation qui aujourd’hui prend peu en compte le travail sur les données.

ist@inra

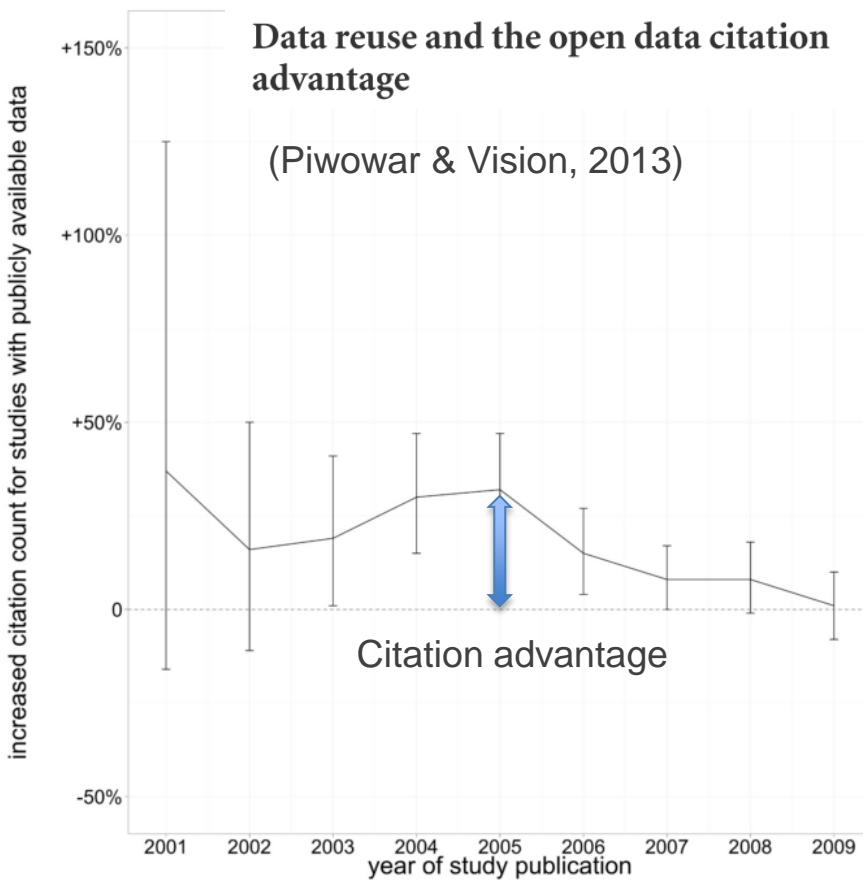
Intérêt du Data Paper ? Avis mitigés ...

Enquête INRA auprès des scientifiques - 2015
90 réponses analysables

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

ist@inra

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 »

ist@inra



(Belter, 2015)

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



(Belter, 2014).

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”

An advantage also in Astrophysics
Drachen, T.M. et al., (2016).

Des collaborations existent...

Consortium

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

(Blum et al., 2014)



Groupes de travail RDA-Alliance

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

Rassemblant les différents acteurs :

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

RDA/WDS Publishing Data Bibliometrics WG

RDA/WDS Publishing Data Services WG

RDA/WDS Publishing Data Workflows WG

De nouveaux worflows, plus d'intégration...

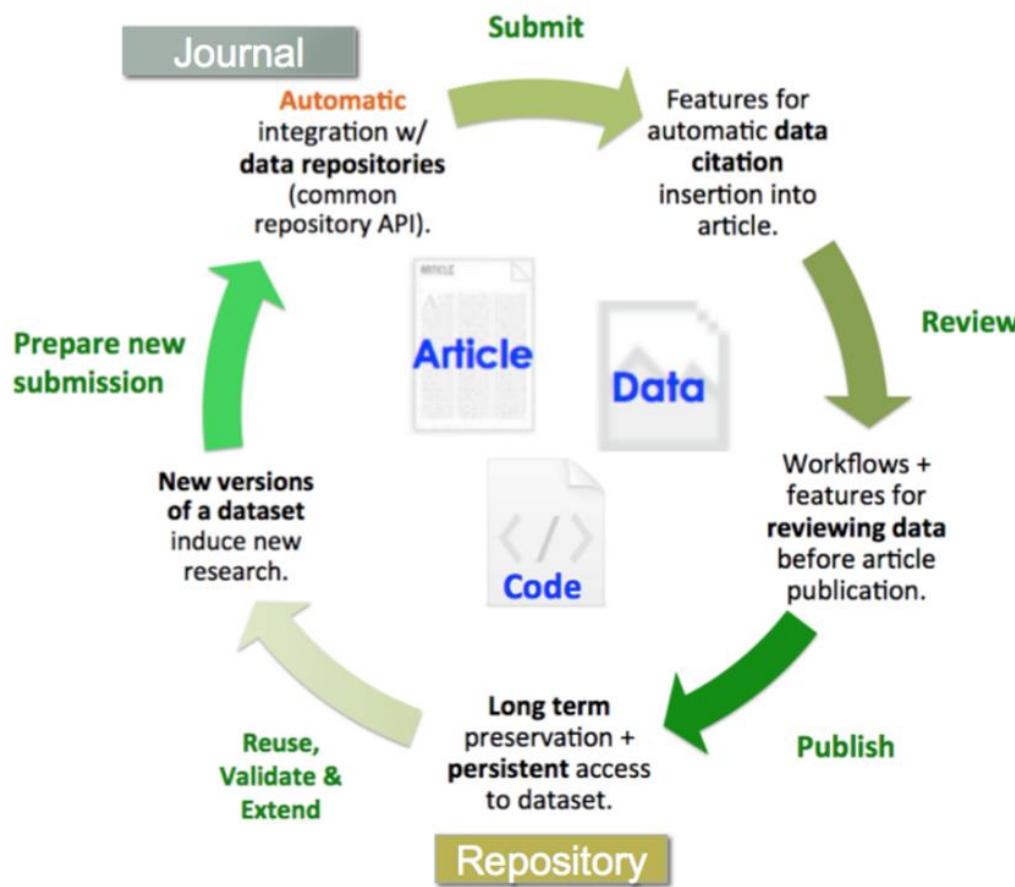


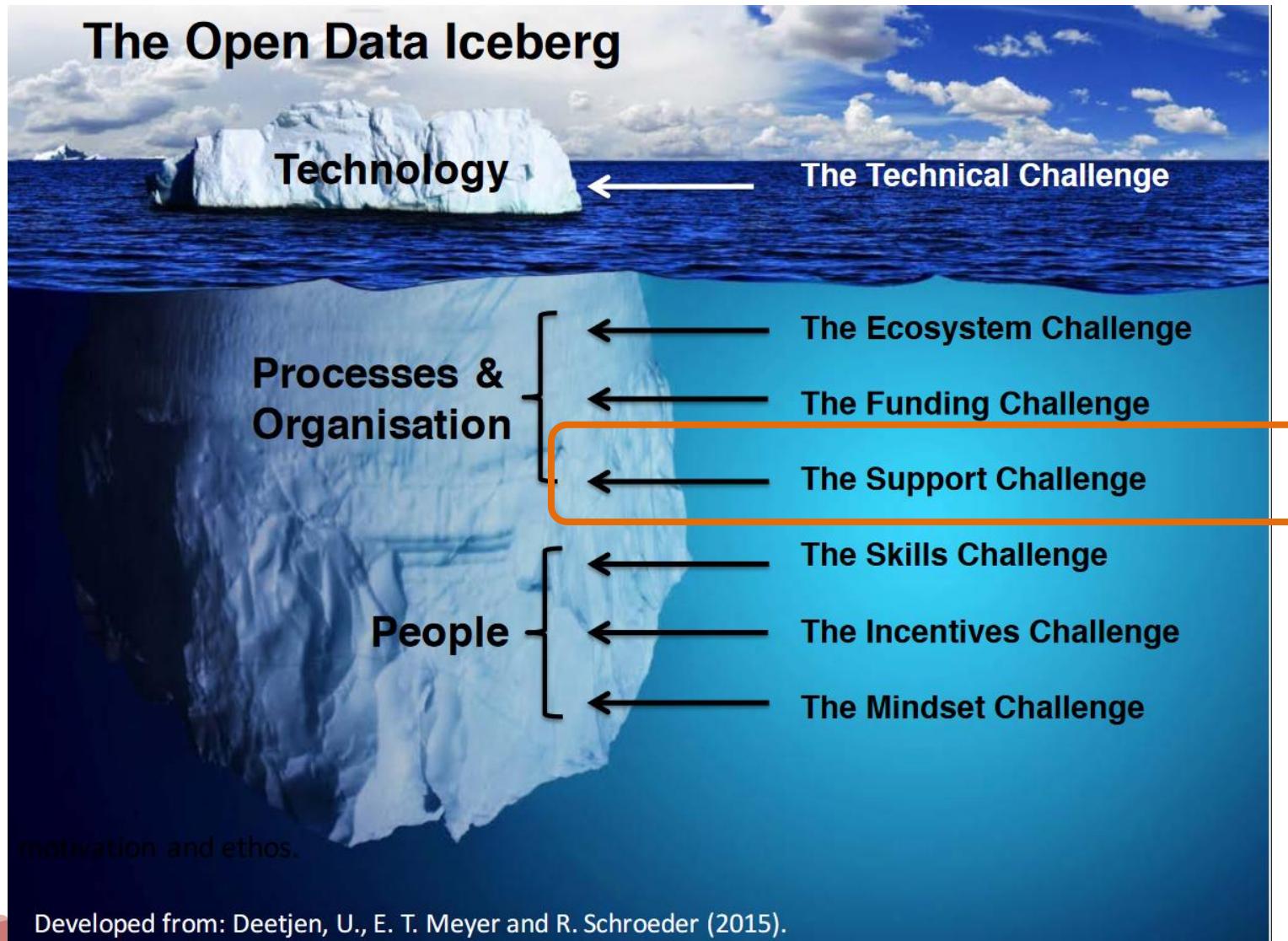
Figure 1. Lifecycle of an automated and integrated journal and data publishing workflow.

- 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 ...

Altman et al., 2015

ist@inra

Importance de l'accompagnement



Un site Inra dédié, pour vous accompagner ...

The screenshot shows the INRA Data Management and Sharing website. The header features the INRA logo and the tagline "SCIENCE & IMPACT". The main navigation menu includes "Gérer", "Partager / Publier", "Les enjeux pour l'Inra", "Réutiliser", and "Documents de référence". The "Partager / Publier" section is currently active, displaying a banner with the text "gestion et partage des données scientifiques" and a grid of scientific images. Below the banner, the breadcrumb navigation shows "Accueil > Partager / Publier > Valoriser ses données > Publier un Data Paper". The main content area is titled "Publier un Data Paper" with the subtitle "Data as the subject of a paper". It includes a section titled "Qu'est ce qu'un Data Paper ?" with a detailed description and a diagram illustrating the relationship between various data components. A sidebar titled "Boîte à outils" lists various tools and resources, and another sidebar titled "On vous signale" lists recommended links.

Publier un Data Paper
"Data as the subject of a paper"

Qu'est ce qu'un Data Paper ?

A la différence d'un article scientifique classique qui exploite - analyse et interprète les données scientifiques, un Data Paper décrit finement un/des jeu(x) de données de façon à en faciliter la compréhension, et l'éventuelle réutilisation.

Le(s) jeu(s) de données est(sont) déposé(s) dans un (ou plusieurs) entrepôt(s) et associé(s) à un identifiant (souvent le DOI).

Le Data Paper décrit les données

il comprend les éléments descriptifs associés (métadonnées - metadata), et toutes les informations techniques (méthodes, formules, applications logicielles...) utiles à la compréhension de l'obtention des données et à leur réutilisation par d'autres scientifiques

Diagram illustrating the relationships between Data Journal, Data Paper, Dataset, and Data Repository:

```
graph TD; Journal[Journal] -- "publie" --> Paper[Paper]; DataJournal[Data Journal] -- "Est un" --> Journal; DataJournal -- "publie" --> DataPaper[Data Paper]; DataPaper -- "Est un" --> Paper; DataPaper -- "comprend" --> Metadonnees[Métadonnées]; DataPaper -- "cite" --> Dataset[Dataset]; Dataset -- "publie" --> Entrepot[Entrepôt]; Metadonnees <--> Identifiant[Identifiant]; Metadonnees <--> Contenu[Contenu];
```

ist@inra

<http://www.inra.fr/datapartage>

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/>.
- Austin, C.C., Bloom, T., Dallmeier-Tiessen, S. et al. (2016) Key components of data publishing : Using current best practices to develop a reference model for data publishing. *International Journal on Digital Libraries*. doi:10.1007/s00799-016-0178-2
- 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/RDAWDPublishingDataIGWebinar3Workflows.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://docsslide.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. *Journal of the Association for Information Science and Technology*, 66:1747-1762. <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
- Dallmeier-Tiessen, S. (2015). Data Publishing models. <http://fr.slideshare.net/datascienceiqss/dallmeier-tiessen-datapublishingmodelssddataverse>
- 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)
- Drachen, T.M. et al., (2016). Sharing data increases citations. LIBER Quarterly. 26(2), pp.67–82. DOI: <http://doi.org/10.18352/lq.10149>
- 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>
- Hrynaszkiewicz, 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/johkanri.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
- 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>
- Kratz, J., & Strasser, C. (2014). Data publication consensus and controversies. *F1000Res*, 3, 94. <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>
- 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>
- 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
- 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/bgt0op6c7pxf7cxmhen48fe8/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
- Shaklee, P., Cousijn, H. (2015). Can data be peer-reviewed ? <https://www.elsevier.com/connect/can-data-be-peer-reviewed>
- Tenant, J. (2015) Open Sesame! The Risks and Rewards of Open Data for Researchers. <http://www.digitalscience.com/blog/guest/open-sesame-the-risks-and-rewards-of-open-data-for-researchers/>

S'informer en IST ? Des services à découvrir

infodocexpress

1 à 2 heures pour découvrir ou redécouvrir une base, un service, une technique, une nouveauté en IST, ...

2 formules au choix :

- Dans votre unité ou votre centre
- En ligne, depuis votre poste de travail, « classes virtuelles en IST »

En savoir plus

<https://intranet6.inra.fr/ist/Se-former-a-l-IST/infodoc-express>



Question/Réponse
en IST

Posez votre question...
Une équipe de documentalistes
vous répond

- Réponse sous 3 jours (sauf si appel à un expert)
- Anonymat du demandeur
- Réponses publiées, classées et archivées

En savoir plus

<http://ist.blogs.inra.fr/questionreponses/>