

# Publier un data paper

*Laurence Dedieu*  
*Editrice scientifique*

*Dist (Délégation à l'Information Scientifique et Technique)*



# Publier un data paper

## Data paper

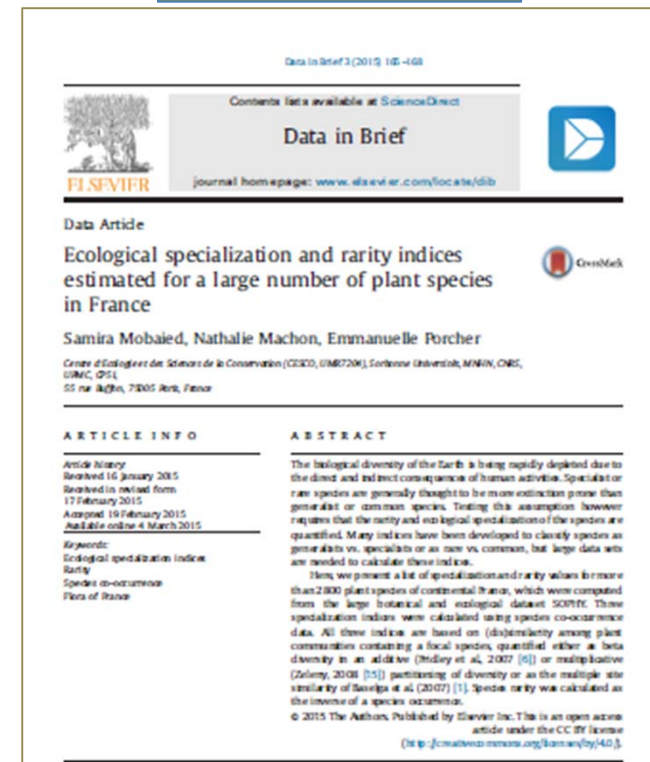
Jeu de données

Réutiliser et  
Valoriser le jeu de données

Mettre à disposition  
d'une  
communauté  
scientifique

Informers la  
communauté  
scientifique

Entrepôt de  
données

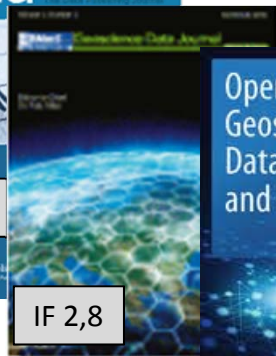


- Décrit un jeu de données
- Montre son potentiel
- Donne accès aux données

# Revues pour publier des jeux de données



IF 6,7



IF 2,8

Geoscience Data Journal



IF 5,3



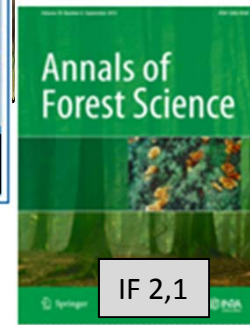
IF 1,3



IF 4,8



IF 6



IF 2,1



IF 3,5



IF 2,3



Water Resources Research



IF 4,4



# Data in Brief



Title, Authors

Abstract

Subject area	<i>Physics, Chemistry, Biology, Earth and Planetary Sciences, Environmental Sciences</i>
Type of data	
How data was acquired	<i>Microscope, survey, SEM, NMR, mass spectroscopy, etc.</i>
Data format	
Experimental factors	<i>Brief description of any pretreatment of samples</i>
Data source location	
Data accessibility	<i>State if data is with this article or in public repository ( name, DOI, URL)</i>

**Value of the data :** 3-5 points to describe :

*why this data is of value to the scientific community.*

*how the data could be potentially valuable to other researchers.*

**Data**

*Briefly describe the data to give the reader context before you describe Materials and Methods*

**Experimental Design, Materials and Methods**

*Complete description of the Experimental design and methods used to acquire the data.*

Acknowledgements

References

Title, Authors

Abstract



## Background & Summary

*overview of the study and the context  
and **potential reuse value**.*

## Data Records

*explain each data record + data formats + data repository*

## Methods

*methods, procedures and processing steps leading to production of the data files,  
including computational analyses*

## Technical Validation

*to justify the technical quality of the dataset  
the reliability of the data.*

## Usage Notes (optional)

*brief instructions to assist with data reuse by other researchers*

Acknowledgements

Author Contributions

References



## DATA PAPER

### Title

---

Author names, affiliations

Abstract, Key words

**Dataset name, data centre, and unique identifier.**

Identifier: xxxxx

Creator: xxxxx

Publisher: xxxxxx

Publication year: xxxxxx

---

### Introduction

Rationale for collecting the dataset.

#### 1. Data production methods

Description of the experimental set-up for the acquisition of the data.

#### 2. Dataset location and format

Description of location, format and accessibility of the dataset.

#### 3. Dataset use and reuse

Description of actual and potential uses for the dataset.

### Supporting information

The following supporting information is available as part of the online article:

Video S1. xxxxxx

Figure S1. xxxxxx

Table S1. xxxxxx

Appendix S1. xxxxxx

### Acknowledgements

The research presented in this paper was funded by xxxxxx. The authors are grateful to xxxx for xxxx.

### References

Authors. [year of publication]. Article title. *Journal Name* [vol]: pp–pp, doi:xxxxxxx.

## Choix d'une revue

Data in Brief  
ELSEVIER



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

Localisation des  
données

Intégrées dans le  
datapaper

Entrepôt de données  
(éditeur, généralistes,  
thématiques)

# Ecological Research et entrepôt LTER



Ecol Res (2015) 30: 415  
DOI 10.1007/s11284-015-1252-8



## DATA PAPER

Itsarapong Voraphab · Yupa Hanboonsong  
Yousichi Kobori · Hiroaki Ikeda · Takeshi Osawa

### Insect species recorded in sugarcane fields of Khon Kaen Province, Thailand, over three seasons in 2012

Received: 1 August 2014 / Accepted: 15 February 2015 / Published online: 6 March 2015  
© The Ecological Society of Japan 2015

**Abstract** The diversity of insect species in cultivated fields provides a number of ecosystem benefits, including natural pest control and pollination. However, current knowledge of the insect fauna in agricultural landscapes is extremely limited when compared with information available for other ecosystems, such as forests and rivers. Here, we present data on insect diversity, including functional feeding group diversity, in sugarcane fields of Khon Kaen Province, Thailand. Sugarcane is a major Thai crop. We collected insects using light trap systems located in 16 sugarcane field plots. Each light trap was deployed in an individual 0.16 ha plot. Trapping was conducted over three seasons, including both dry and rainy periods, in March, June, and October of 2012. On each trapping day, we collected insects in the evening from 18:00 to 19:30. The traps yielded a total of 98,423 individuals including 143 species belonging to 26 families. We classified collections into five functional feeding groups: herbivores, predators, decomposers, parasitoids, and pollinators. The dominant functional

species were herbivores, followed in rank order by predators; most individuals in the traps were predatory species. We assembled an insect fauna data set that will be useful in the study of southeast Asian agricultural ecosystems and provided the data in the Darwin Core Archive format.

**Keywords** Agricultural ecosystem · Decomposer · Herbivore · Parasitoid · Pollinator · Predator

*Ecological Research Data Paper Archives* [http://db.cger.nies.go.jp/JaLTER/ER\\_DataPapers/](http://db.cger.nies.go.jp/JaLTER/ER_DataPapers/)  
Accession number: ERDP-2015-01

The complete data set for this abstract published in the Data Paper section of the journal is available in electronic format in *Ecological Research Data Paper Archives* at [http://db.cger.nies.go.jp/JaLTER/ER\\_DataPapers/archives/2015/ERDP-2015-01](http://db.cger.nies.go.jp/JaLTER/ER_DataPapers/archives/2015/ERDP-2015-01).

I. Voraphab · Y. Hanboonsong  
Division of Entomology, Faculty of Agriculture, Khon Kaen University, Khon Kaen 40002, Thailand

Y. Kobori  
Japan International Research Center for Agricultural Sciences, Tsukuba, Japan

H. Ikeda · T. Osawa (✉)  
National Institute for Agro-Environmental Sciences, Tsukuba, Japan  
E-mail: [tosawa@affrc.affrc.go.jp](mailto:tosawa@affrc.affrc.go.jp)  
Tel.: +81.78.803.8148  
Fax: +81.78.803.8199

T. Osawa  
Japan Node of the Global Biodiversity Information Facility, Tsukuba, Japan

LTER

International Long-Term Ecological  
Research Network  
(Entrepôt thématique)



## DATA PAPER

### Title

Author names, affiliations

Abstract, Key words

**Dataset : name, data centre, and unique identifier**

Identifier: xxxxx

Creator: xxxxx

Publisher: xxxxxx

Publication year: xxxxxx

### Introduction

Rationale for collecting the dataset.

#### 1. Data production methods

Description of the experimental set-up for the acquisition of the data.

#### 2. Dataset location and format

Description of location, format and accessibility of the dataset.

#### 3. Dataset use and reuse

Description of actual and potential uses for the dataset.

### Acknowledgements

The research presented in this paper was funded by xxxxxx. The authors are grateful to xxxx for xxxx.

### Approved Thematic Data repositories

commonly used by the scientific community it supports, formal data management policy in place, provide a stable URL and unique identifier for the dataset.

[4TU.ResearchData](#)

[British Atmospheric Data Centre \(BADC\)](#)

[British Oceanographic Data Centre \(BODC\)](#)

[CISL Research Data Archive](#)

[CSIRO Data Access Portal](#)

[Deep Carbon Observatory](#)

[earthH<sub>2</sub>Observe](#)

[Environmental Information Data Centre \(EIDC\)](#)

[IEDA:EarthChem](#)

[IEDA:MGDS](#)

National Center for Atmospheric Research (NCAR), USA

[Earth Observing Lab \(EOL\)](#),

[Research Data Archive \(RDA\)](#)

[National Geoscience Data Centre \(NGDC\)](#)

[NERC Earth Observation Data Centre \(NEODC\)](#)

[NOAA National Climatic Data Center \(NCDC\)](#)

[NOAA National Oceanographic Data Center \(NODC\)](#)

[NOAA National Geophysical Data Center \(NGDC\)](#)

[PANGAEA](#)

[Polar Data Centre \(PDC\)](#)

[Figshare](#)

[Zenodo](#)

# Software papers et entrepôts



Journal of  
open research software

Li, X et al 2017 Xanthos – A Global Hydrologic Model. *Journal of Open Research Software*, 5: 21, DOI: <https://doi.org/10.5334/jors.181>

## SOFTWARE METAPAPER

### Xanthos – A Global Hydrologic Model

Xinya Li<sup>1</sup>, Chris R. Vernon<sup>1</sup>, Mohamad I. Hejazi<sup>2</sup>, Robert P. Link<sup>2</sup>, Leyang Feng<sup>2</sup>, Yaling Liu<sup>2</sup> and Lynn T. Rauchenstein<sup>3</sup>

<sup>1</sup> Pacific Northwest National Laboratory, Richland, WA 99352, US

<sup>2</sup> Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, Maryland, US

<sup>3</sup> (Formerly) Pacific Northwest National Laboratory, Richland, WA 99352, US

Corresponding author: Xinya Li ([Xinya.Li@pnnl.gov](mailto:Xinya.Li@pnnl.gov))

Xanthos is an open-source hydrologic model, written in Python, designed to quantify and analyse global water availability. Xanthos simulates historical and future global water availability on a monthly time step at a spatial resolution of 0.5 geographic degrees. Xanthos was designed to be extensible and used by scientists that study global water supply and work with the Global Change Assessment Model (GCAM). Xanthos uses a user-defined configuration file to specify model inputs, outputs and parameters. Xanthos has been tested using actual global data sets and the model is able to provide historical observations and future estimates of renewable freshwater resources in the form of total runoff.

**Keywords:** Water Balance Model; Global Hydrologic Model; Runoff; River Routing; Potential Evapotranspiration (PET); Global Change Assessment Model (GCAM); Python

**Funding statement:** PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.



## Choix d'une revue



**Localisation des données**

**Modalités de diffusion des données**

Intégrées dans le datapaper

Entrepôt de données  
(éditeur, généralistes, thématiques)

Licences de diffusion

Accès sur demande

Accès après embargo

## Revue et modalités de diffusion des données



Données accessibles  
dès publication



Embargo 2 ans possible

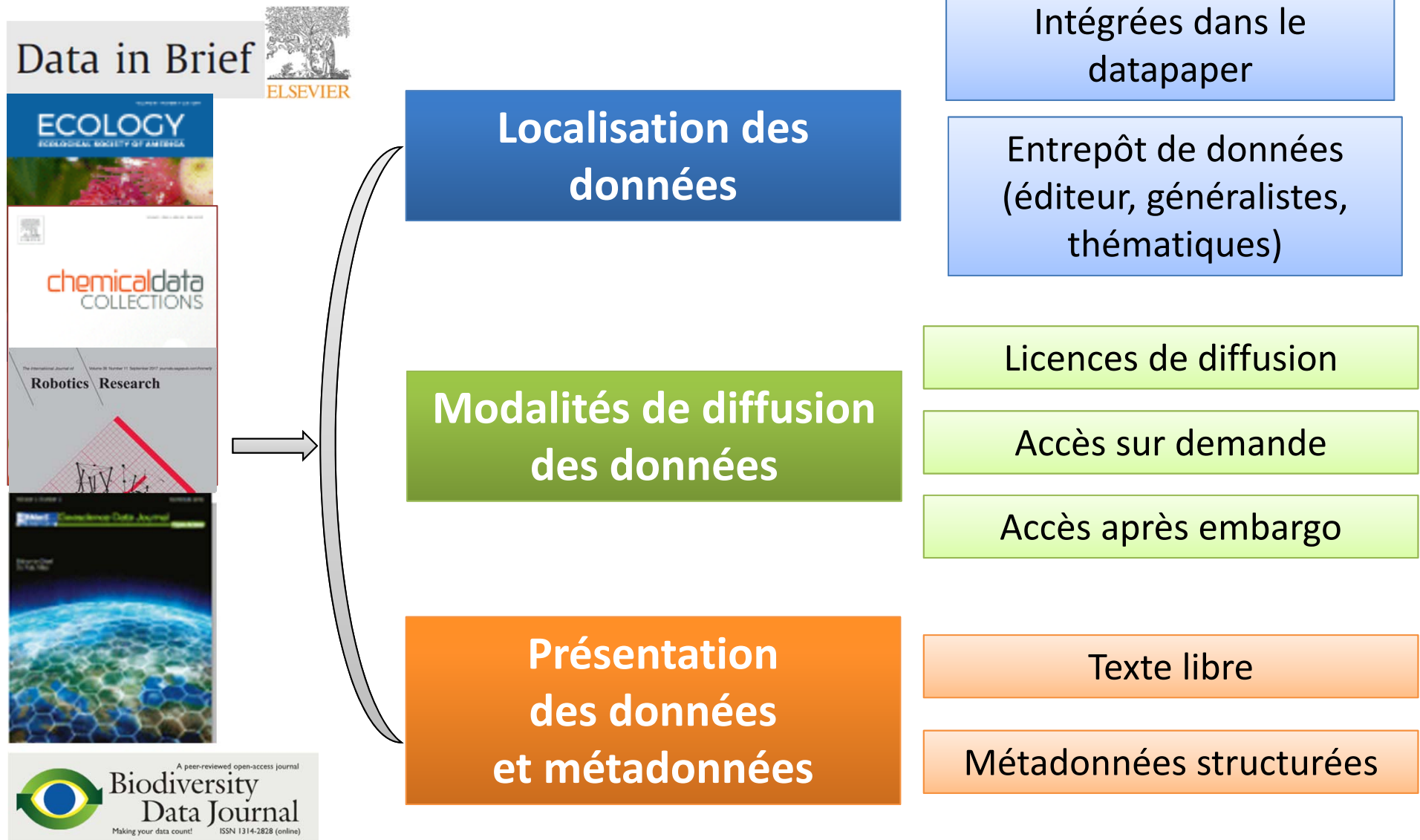


Pas de restriction à une utilisation non commerciale (NC)  
→ licence CC-0 ou CC-BY

Licence CC-BY-NC  
Uniquement utilisation en recherche



## Choix d'une revue





*Ecology*



---

**Data Papers**

---

Ecology, 98(8), 2017, pp. 2224  
© 2017 by the Ecological Society of America

Fifty years of continuous precipitation and stream chemistry data from the Hubbard Brook ecosystem study (1963–2013)

GENE E. LIKENS<sup>1</sup>

*Cary Institute of Ecosystem Studies, 2801 Sharon Turnpike, Millbrook, New York 12545 USA*

officially began on 1 June 1963. This analysis of (at least) weekly stream water 1963–2014 (from 1 June 1963 to 30 May 2014) and precipitation chemistry for Precipitation Experimental Forest are reported

um; chloride; Hubbard Brook Experimental sulfate.

## Class II. Research origin descriptors

## Site Description

- Site type
- Geography (location, size)
- Habitat
- Geology, Landform
- Climate

## Experimental or sampling design

- Design characteristics
- Variables included
- Species sampled
- Data collection period, frequency

## Research methods

- Field/Laboratory
- Instrumentation

### Class III. Data set status and accessibility

## Status

- Latest update
- Latest archive date
- Metadata status
- Data verification

## Accessibility

- Storage location and medium
- Contact person(s)
- Copyright restriction
- Costs

## Filename

ecy1894-sup-0001-MetadataS1.doc

Word document, 126K

[ecy1894-sup-0002-DataS1.zip](#)

Zip archive, 684K

## Class IV. Data structural descriptors

## Data Set Files

- Data set Identity
- Size
- Format and storage mode

## Critères de choix d'une revue

- **Champ scientifique et lectorat/futurs utilisateurs**
- **Instructions aux auteurs:** modèle, recommandations, contraintes
- **Qualité et échelle de vos jeux de données**
  - ✓ **Global Ecology & Biogeography:** *datasets of **broad** macro-ecological interest.*
  - ✓ **GigaScience:** *“**big-data**” studies from entire spectrum of life and biomedical sciences.*
  - ✓ **Water Resources Research :** ***important** Earth and space science datasets.*
- **Coût de publication :** gratuit; 250 \$ - 2000 \$
- **Facteur d'impact**

## Le datapaper en résumé

- Décrit des jeux de données et les méthodes d'obtention  
*ni analyses, ni résultats, ni interprétation*
- Donne accès à ces jeux de données



Doit montrer le potentiel des données



Réutilisation de ses données

- **Publics** : communautés scientifiques, enseignants, décideurs, secteur privé, créateurs de start-up, grand public, membres d'ONG, journalistes
- **Réutilisations** : dans votre domaine, dans d'autres domaines
- **Perspectives d'application, développement ?** : potentiel stratégique, commercial, ...

## Critères d'évaluation (peer-reviewing)



Les datapapers sont évalués

- **Potentiel / valeur de réutilisation** des données
- **Valeur scientifique et originalité** des données
- **Rigueur et adéquation** de la méthode
- **Description suffisante** pour permettre à d'autres chercheurs de comprendre, d'interpréter et de réutiliser les données
- **Accès aux données**

## Publier des jeux de données



A condition d'en avoir le droit, et donc de respecter:

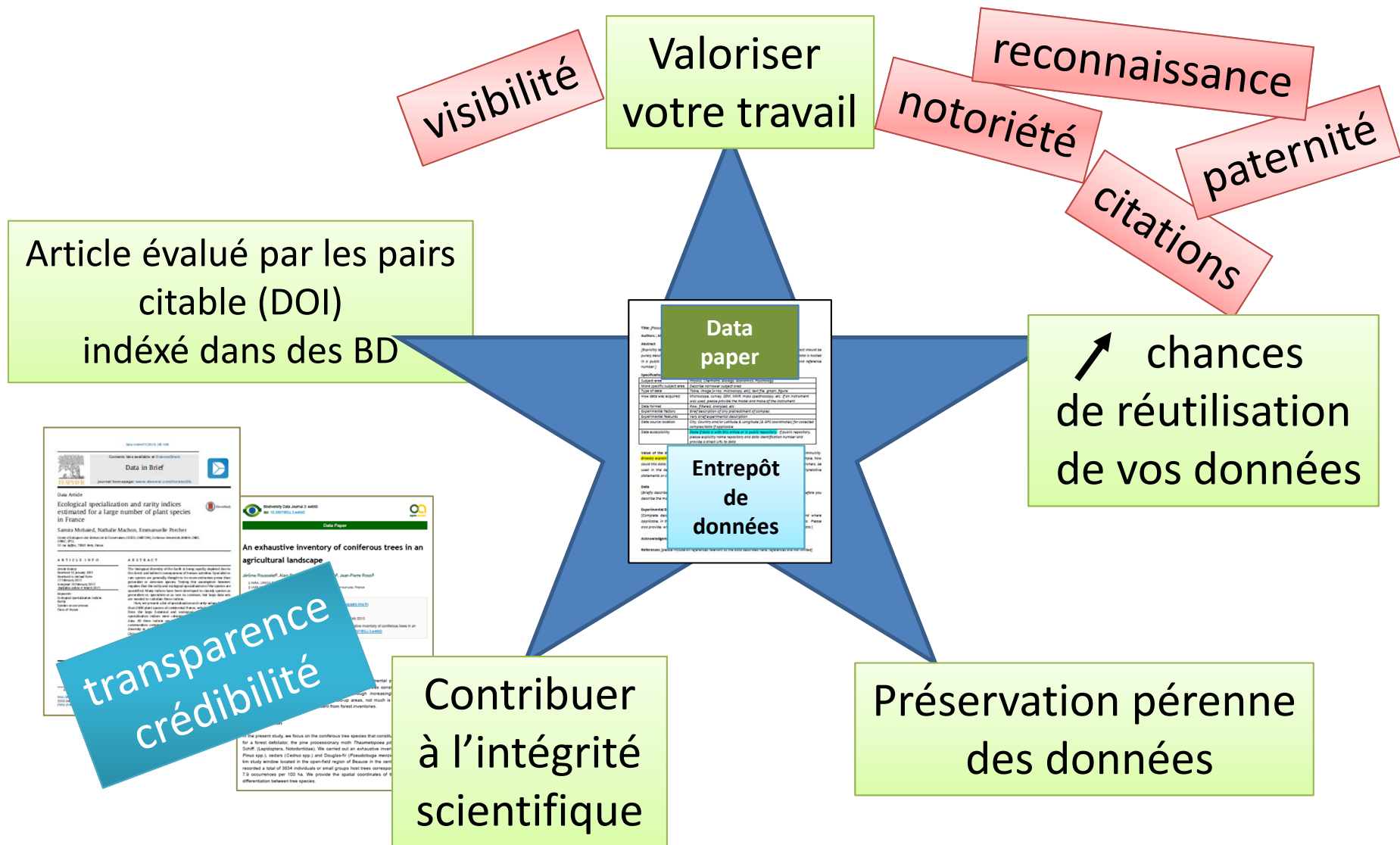
- Droits de propriété intellectuelle ou industrielle
- Obligations institutionnelles, liées au bailleur,
- Contraintes de partenariats
- Cadres réglementaire ou éthique: *règles d'Accès et Partage des Avantages*
- Protection de données personnelles (anonymisation), cliniques, sensibles, etc.
- Accord de tous les contributeurs



Consulter vos juristes, ...



# Les avantages du datapaper

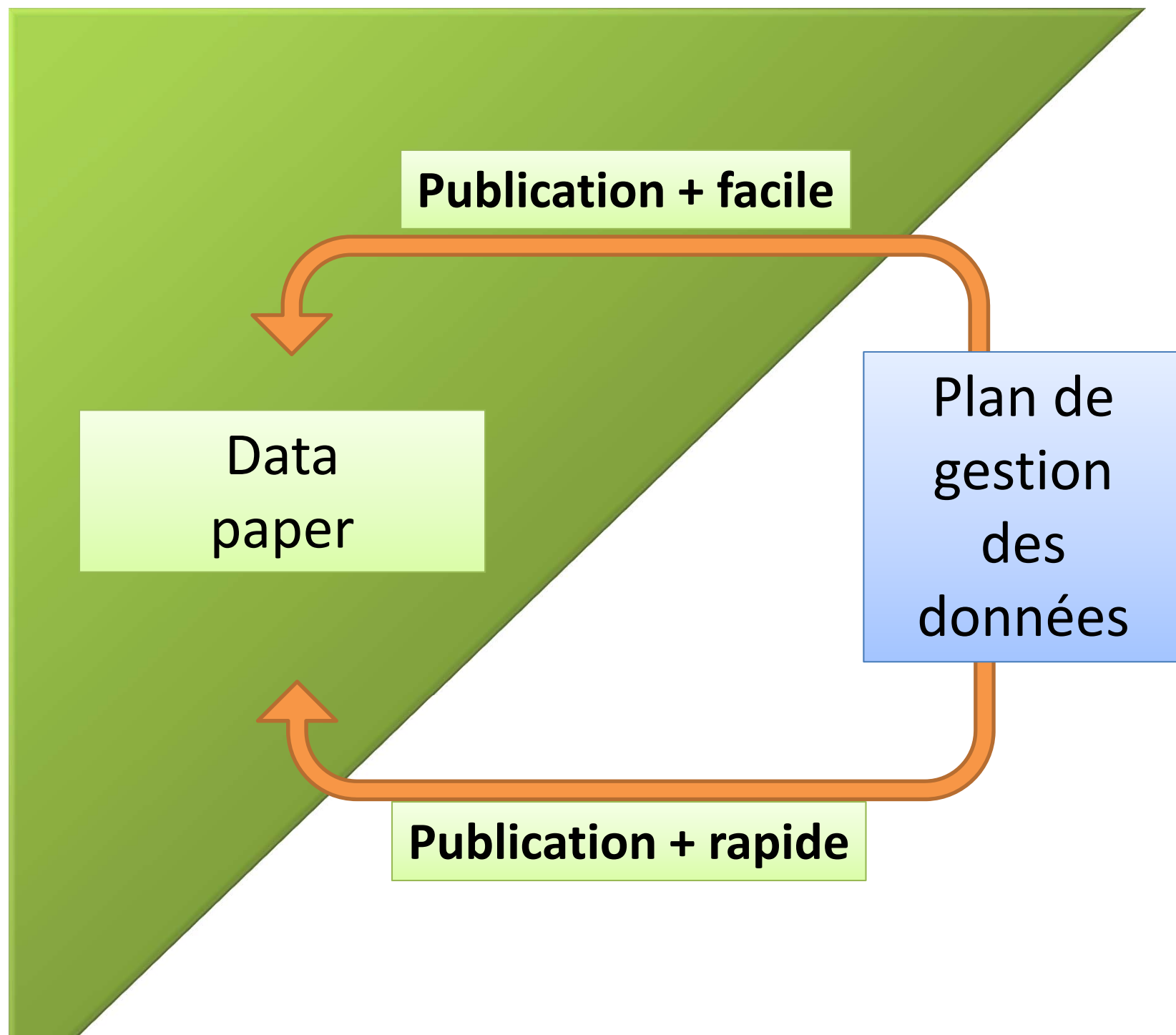


**Publication + facile**

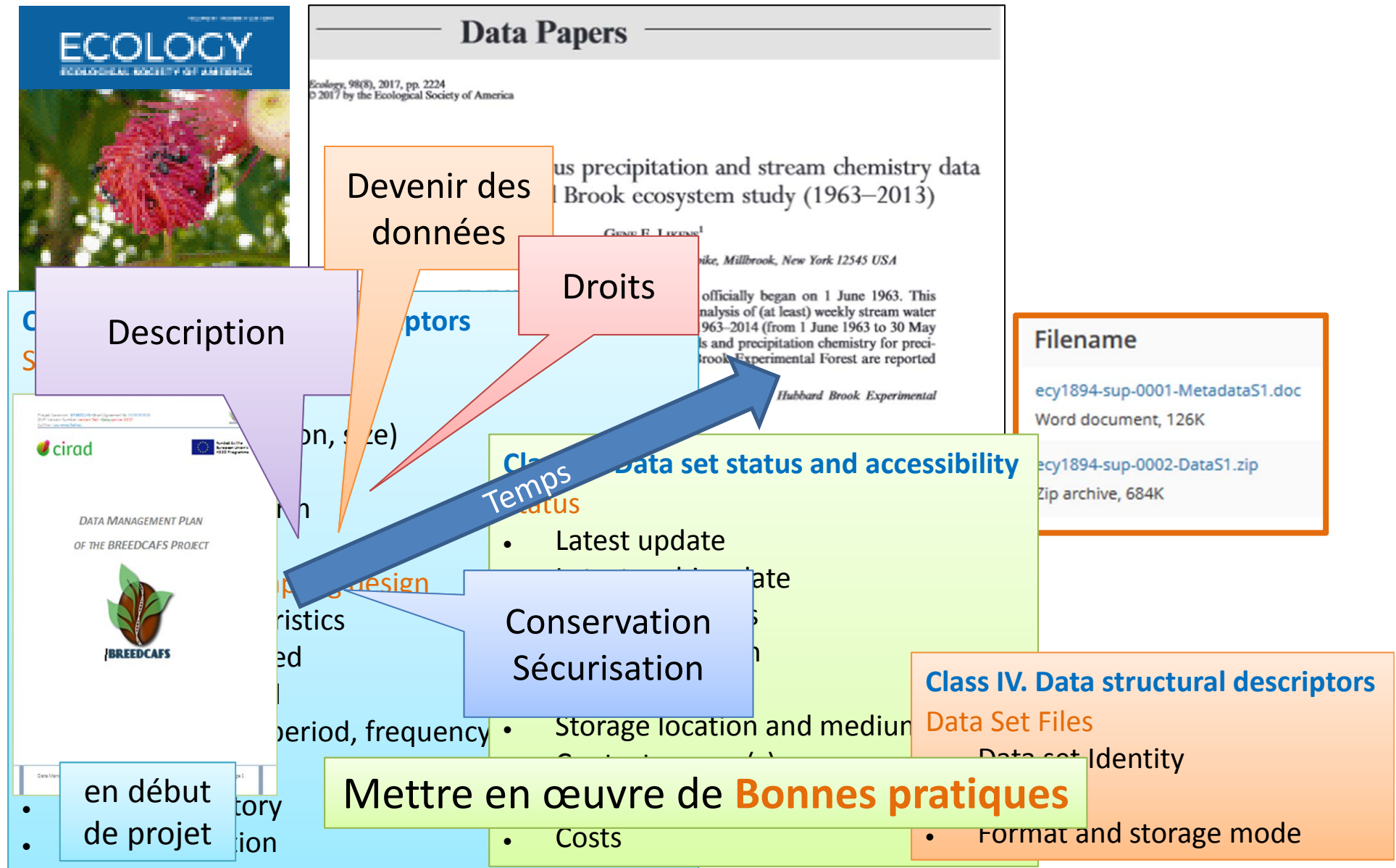
Data  
paper

Plan de  
gestion  
des  
données

**Publication + rapide**



# Plan de gestion des données



## Le PGD répond à un constat

Minimiser les  
risques  
de perte de données

Éviter les  
duplications  
d'expériences

Réduire les coûts

Pourquoi ?

Assurer la compréhension et réutilisation

Mise en œuvre de **Bonnes pratiques**

Partage  
des données



# Le PGD concerne tous producteurs de données

Doctorant

Scientifique:  
*Retraite  
changement d'activités*

Travail  
en équipe

Projet en  
partenariat

Facilite l'échange de données,  
la comparaison/combinaison de données  
Besoin de normes/métadonnées communes

Partage  
des données

Pour qui ?

Assurer la compréhension et réutilisation

Mise en œuvre de **Bonnes pratiques**





## Le PGD garantit des bonnes pratiques

### Modèle de PGD

#### Data set description

*Type, format,...*

*Données pré-existantes ?*

#### Standards and metadata

#### Archiving / preservation

*Volume, coût, ...*

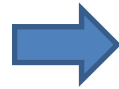
#### Data sharing

*Quelles données,*

*Comment,*

*Où ?*

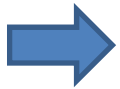
*Droits, Éthique,...*



Pose de bonnes questions  
dès le début du projet

- Comment décrire au mieux les données pour assurer leur compréhension ?
- Mode de stockage / sécurisation ?
- Quelles données conserver ? Quel financement ?
- Quelles données partager ? Intérêt pour qui ?
- Avons-nous les droits ? *Données pre-existantes ?*
- Aspects éthiques / réglementaires ?  
*données sensibles, données personnelles, ayant un impact sur homme/environnement, relevant du secret professionnel, ...*
- Responsabilités ?  
*gestion des données, stockage, ...*

## Le PGD renseigne les modalités de partage



Partage fortement recommandé  
par les institutions, bailleurs, revues, ...

### ➤ Ne concerne pas toutes les données



*« As open as possible, as closed as necessary »*

Protection des données contractuelles / exploitation

Protection des données sensibles, personnelles, de santé,...

Document évolutif → à renseigner au fur et à mesure du projet

H2020: versions 6, 18, 36 mois

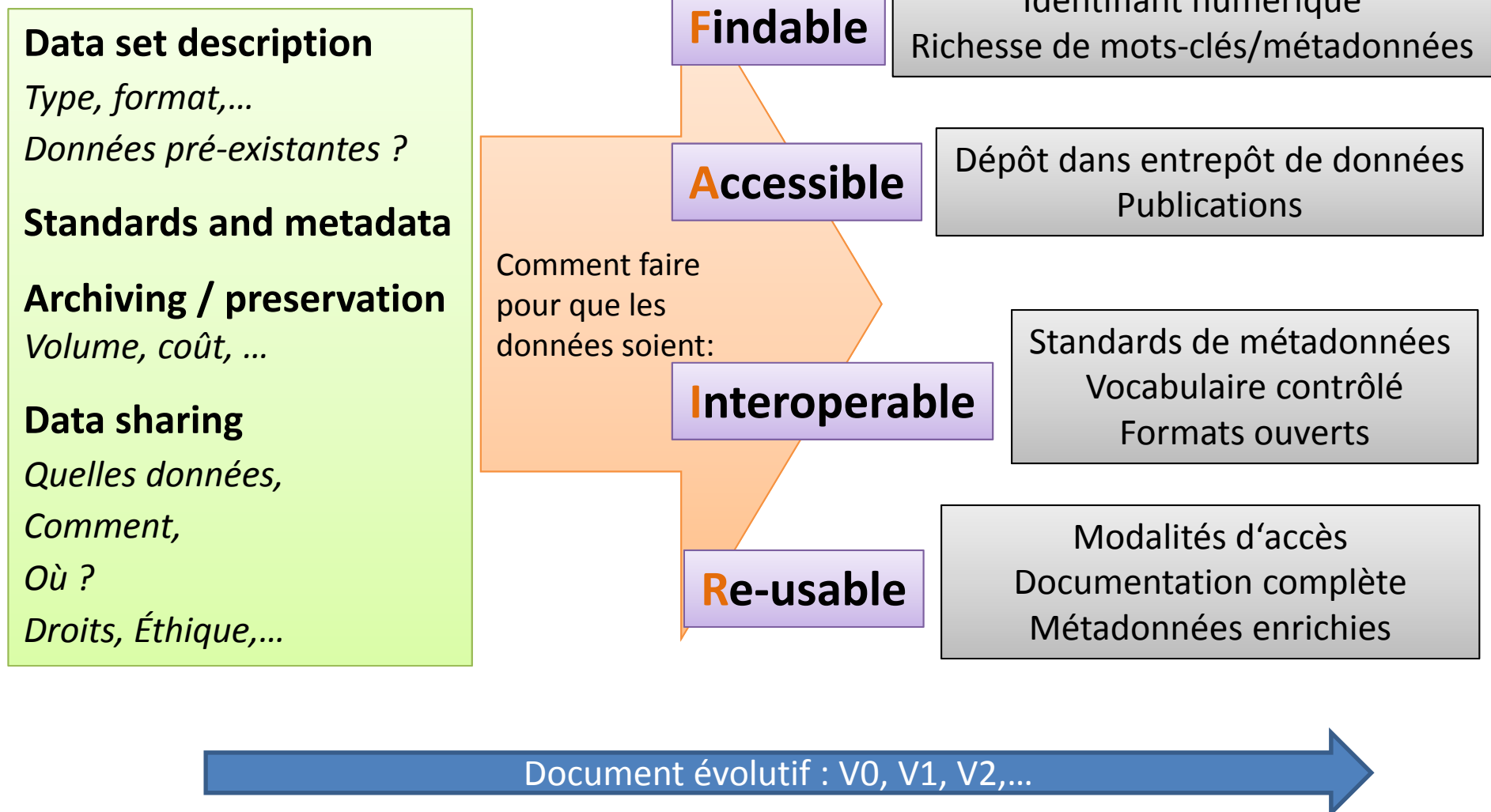
### ➤ Choix des modalités d'accès à vos données

licence ouverte / accès restreint / sur enregistrement

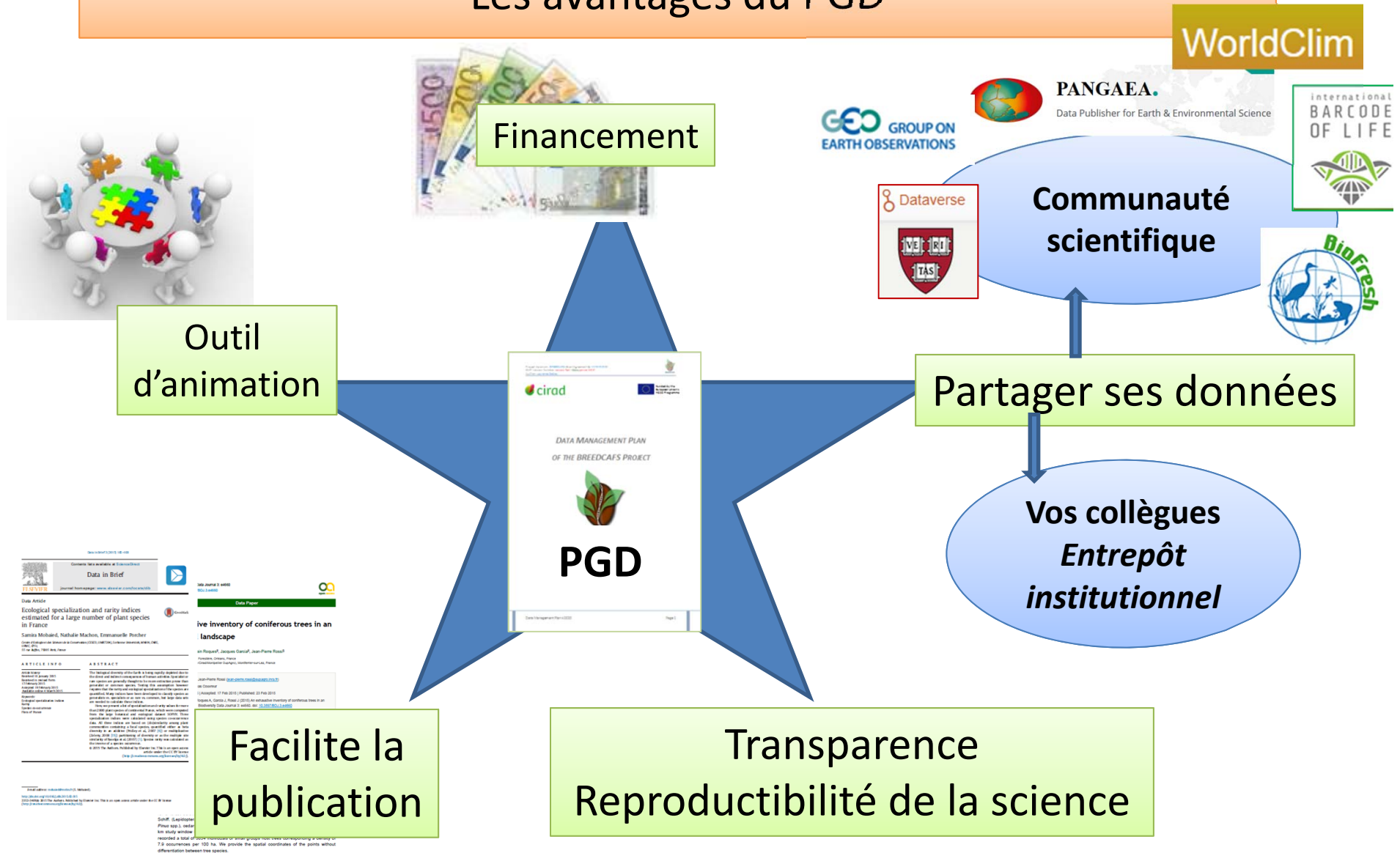
période d'embargo

limiter réutilisation (non commerciale, selon projet)

## Le PGD selon les principes FAIR



# Les avantages du PGD



# Le PGD peut être publié

Pensoft



Research Ideas and Outcomes  
The Open Science Journal

<https://riojournal.com/>

Data Management Plan: Empowering Indigenous Peoples and Knowledge Systems Related to Climate Change and Intellectual Property Rights

Cath Traynor

Data Management Plan doi: 10.3897/rio.3.e15111

25-07-2017 Unique: 296 | Total: 476 Reprint: € 2,30

See collection

HTML XML PDF

Data Management Plan: Opening access to economic data to prevent tobacco related diseases in Africa

Lynn Woolfrey

Data Management Plan doi: 10.3897/rio.3.e14837

24-07-2017 Unique: 242 | Total: 392 Reprint: € 2,60

See collection

HTML XML PDF

Data Management Plan: HarassMap

Reem Wael

Data Management Plan doi: 10.3897/rio.3.e15133

18-07-2017 Unique: 239 | Total: 374 Reprint: € 2,60

See collection

HTML XML PDF

Data Management Plan: IDRC Data Sharing Pilot Project

Cameron Neylon

Data Management Plan doi: 10.3897/rio.3.e14672

27-06-2017 Unique: 320 | Total: 537 Reprint: € 2,60

See collection



HTML XML PDF

Data Management Plan: Brazil's Virtual Herbarium

Dora Ann Lange Canhos

Data Management Plan doi: 10.3897/rio.3.e14675

27-06-2017 Unique: 360 | Total: 590 Reprint: € 2,60

See collection

HTML XML PDF



Deliverable 1.1

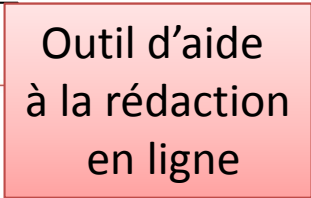
Data Management Plan

DISSEMINATION LEVEL		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	





## Outils



## Research Data Alliance



re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES

<http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>



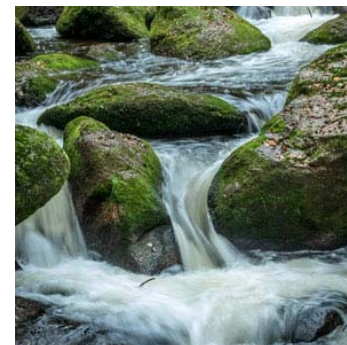
# Merci de votre attention

Laurence Dedieu : [laurence.dedieu@cirad.fr](mailto:laurence.dedieu@cirad.fr)

Pour plus d'information :



<https://coop-ist.cirad.fr/>



La liste de revues publiant des datapapers est accessible à partir de la page:

<https://coop-ist.cirad.fr/aide-a-la-publication/rediger/data-paper/1-qu-est-ce-qu-un-data-paper>

