

How could forest science journals contribute to the development of open data ?

Odile Hologne

▶ **To cite this version:**

Odile Hologne. How could forest science journals contribute to the development of open data ?. Meeting of the editors of Forest Research Journals, Sep 2013, Nancy, France. 25 p. hal-02805741

HAL Id: hal-02805741
https://hal.inrae.fr/hal-02805741

Submitted on 6 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.

How could forest science journals contribute to the development of open data ?

Odile Hologne, Head of the department of scientific information

Agenda

- Data within the publishing ecosystem : a quick overview

- What could be done to promote data sharing ?
  - ✓ From level 0 to level n …

_01

Data within the publishing ecosystem : a quick overiew



Overview : data within the publishing ecosystem

- Journals policies and authors guidelines
- Intellectual property
- Data journals
- Data repositories
- Data and articles citations
- Data citation index

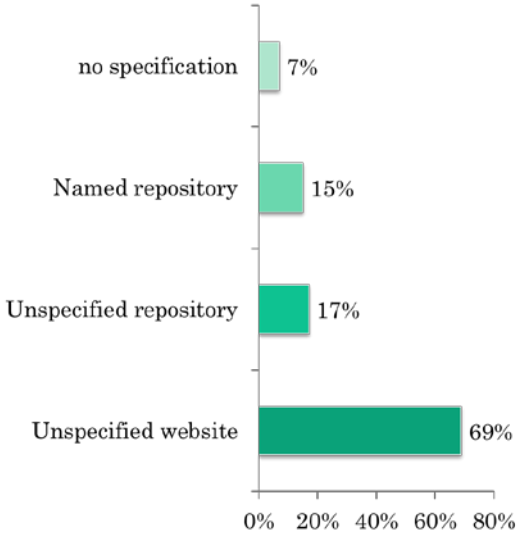


Jord Project

JISC

Centre for Research Communications

- Survey of journals author guidelines



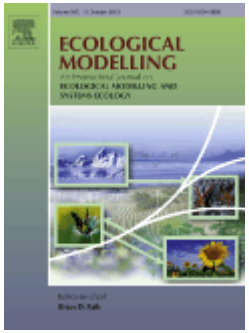
Results of Journal Survey	
Total no. of Journals surveyed	371
Total no. of Journals with data sharing policies	162
Total no. of Journals that make sharing a requirement of publication	31
Total no. of Journals that enforce the policies	27
Total no. of Journals that state consequences for non compliance	7

The JoRD Project was a feasibility study on the possible shape of a central service on journal research data policies, funded by the UK JISC under its Managing Data Research Programme
http://jordproject.wordpress.com/category/data-analysis/

Data within the Author guidelines

Science

Data and materials availability All data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of *Science*.]...[
Science supports the efforts of databases that aggregate published data for the use of the scientific community. Therefore, appropriate data sets (including microarray data, protein or DNA sequences, atomic coordinates or electron microscopy maps for macromolecular structures, and climate data) must be deposited in an approved database, and an accession number or a specific access address must be included in the published paper. We encourage compliance with MIBBI guidelines (Minimum Information for Biological and Biomedical Investigations).



Data at PANGAEA

Electronic archiving of supplementary data enables readers to replicate, verify and build upon the conclusions published in your paper. We recommend that data should be deposited in the data library PANGAEA (http://www.pangaea.de). Data are quality controlled and archived by an editor in standard machine-readable formats and are available via Open Access. After processing, the author receives an identifier (DOI) linking to the supplements for checking. As your data sets will be citable you might want to refer to them in your article. In any case, data supplements and the article will be automatically linked as in the following example: doi:10.1016/0016-7037(95)00105-9. Please use PANGAEA's web interface to submit your data (http://www.pangaea.de/submit/).

THE EMBO JOURNAL

Data deposition

Large-scale datasets, sequences, atomic coordinates and computational models should be deposited in one of the relevant public databases prior to submission (provided private access is available at the database) and authors should include accession codes in the Materials & Methods section. The suggested wording for referring to accession identifiers in a manuscript is the following: "The [protein interaction | microarray | mass spectrometry] data from this publication have been submitted to the [name of the database] database [URL] and assigned the identifier [accession | permalink | hashtag]." If necessary, please include in the manuscript the relevant information (username and password) for confidential access by peer-reviewers.

nature genetics

NATURE GENETICS | EDITORIAL

Respectful re-use

Nature Genetics **44**, 1073 (2012) | doi:10.1038/ng.2433
Published online 26 September 2012

The impact of the papers we publish depends increasingly on the data they describe. In insisting on data access for referees and readers, we prioritize scientific integrity above all and place the interests of research participants before impact.

The Nature research journals insist that materials, methods and data be made available and that authors detail any conditions for access where these exist. In our shared guide to authors (http://www.nature.com/authors/policies/availability.html), we state:

The preferred way to share large data sets is via public repositories. Some of these repositories offer authors the option to host data associated with a manuscript confidentially, and provide anonymous

print
email
pdf
download citation
order reprints
rights and permissions

Data sharing and article citation

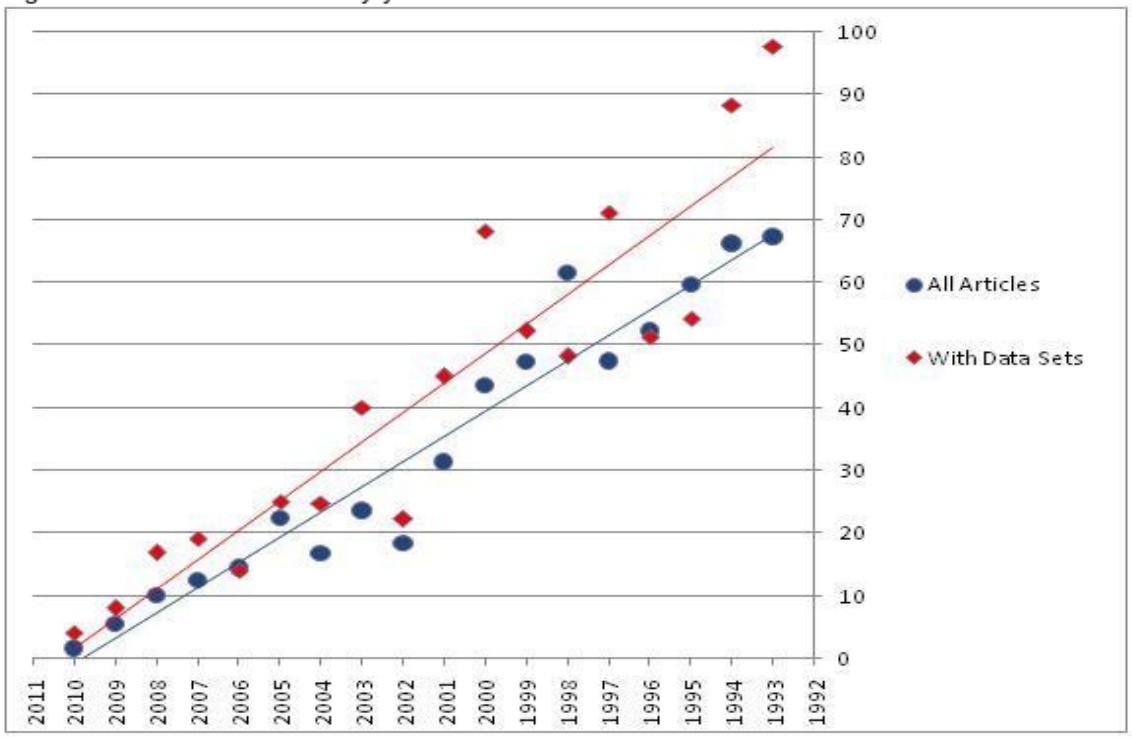
Data Sharing Effect on Article Citation Rate in Paleoceanography

posted Nov 27, 2011 4:11 AM by Michael Diepenbroek [**updated Nov 27, 2011 11:21 PM**]

Jon Sears (AGU) - Abstract for the AGU 2011:

> The validation of scientific results requires reproducible methods and data. Often, however, data sets supporti
> PANGAEA® data library measurably increases the citation rate of articles published between 1993 and 2010 i
> available supporting data sets received 19.94% (8,056) of the aggregate citations (40,409). Publicly available d
> sampled over the 18-year study period (1,331), and the increase is fairly consistent over time (14 of 18 years).

Figure: Plot of mean citations by year to articles with data sets vs. citations to all articles in Paleoceanography betwe



Link to data and effect on citation

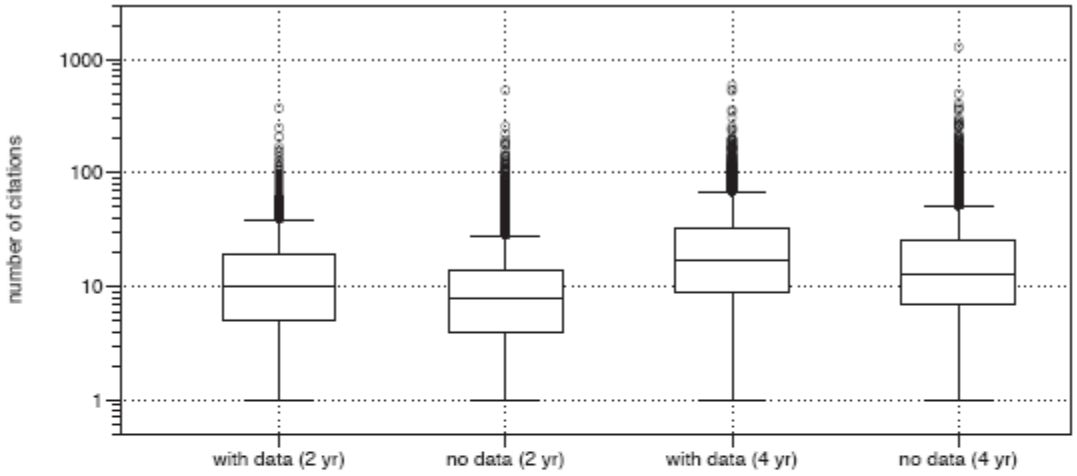


Figure 1. Distribution of citations of articles published in *The Astrophysical Journal* (including *Letters* and *Supplement*), *The Astronomical Journal*, *The Monthly Notices of the R.A.S.* and *Astronomy & Astrophysics* including *Supplement*), during the period 1995 through 2000. The extent of the box corresponds with the interquartile range of the citations and whiskers extend to 1.5 times the interquartile range. The horizontal lines within the boxes correspond with the medians boxes correspond respectively with the citation distributions and without data links 2 years after publication, and 4 years medians are respectively at 10, 8, 17 and 13 citations.

http://arxiv.org/pdf/1111.3618v1.pdf

Linking to Data - Effect on Citation Rates in Astronomy

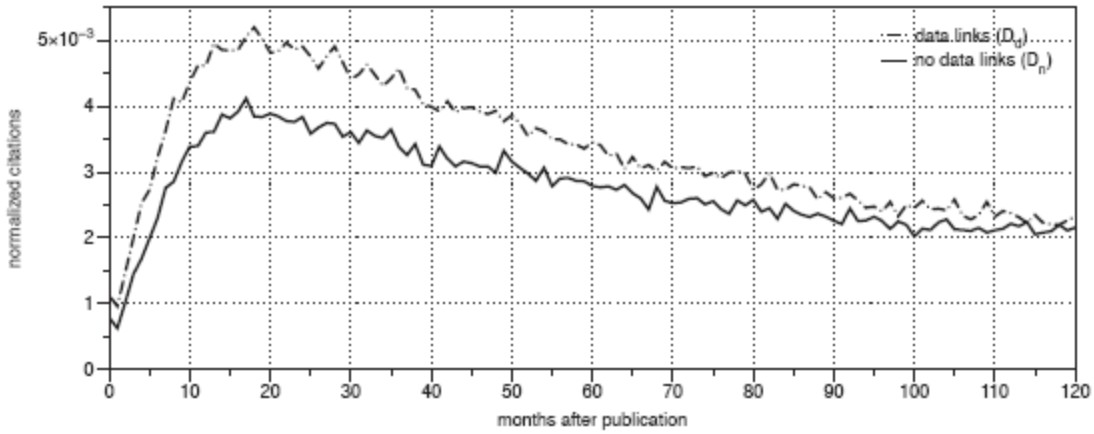


Figure 2. The normalized number of citations for data sets D_d and D_n. The citations have been normalized by the total number of citations.

BMC : Data within the article = CCO

Editorial

Highly accessed | Open Access

Licensing the future: report on BioMed Central's public consultation on open data in peer-reviewed journals

Iain Hrynaszkiewicz[1], **Stefan Busch**[2*] and **Matthew J Cockerill**[2]

* Corresponding author: Stefan Busch stefan.busch@biomedcentral.com ▾ Author Affiliations

1 Faculty of 1000, London, UK

2 BioMed Central Ltd, 236 Gray's Inn Road, London WC1X 8HB, UK

For all author emails, please log on.

BMC Research Notes 2013, **6**:318 doi:10.1186/1756-0500-6-318

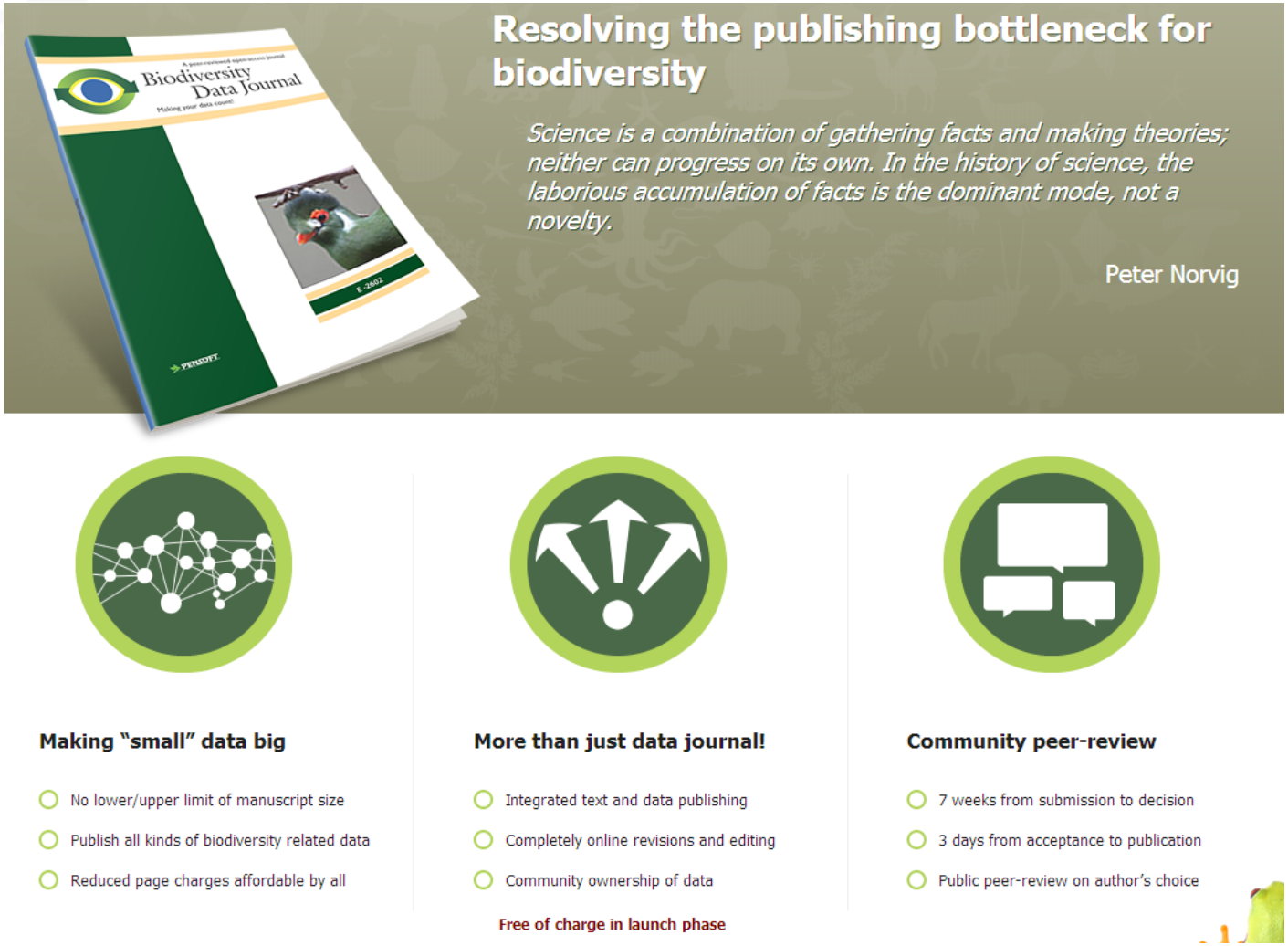
The electronic version of this article is the complete one and can be found online at:http://www.biomedcentral.com/1756-0500/6/318

"© 2013 < Author > et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated."

Data journals, data articles

A growing list : http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList

A data paper describes a dataset, giving details of its collection, processing, software, file formats etc, without the requirement of novel analyses or ground breaking conclusions. It allows the reader to understand the when, how and why data was collected and what the data-product is



Nature

SCIENTIFIC DATA

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

Helping you publish, discover, and reuse research data



Welcome to *Scientific Data*

Scientific Data is a new open-access, online-only publication for descriptions of scientifically valuable datasets. It introduces a new type of content called the Data Descriptor, which will combine traditional narrative content with curated, structured descriptions of research data, including detailed methods and technical analyses supporting data quality. *Scientific Data* will initially focus on the life, biomedical and environmental science communities, but will be open to content from a wide range of scientific disciplines. Publications will be complementary to both traditional research journals and data repositories, and will be designed to foster data sharing and reuse, and ultimately to accelerate scientific discovery.

Scientific Data will launch in Spring 2014 and be open for submissions in Autumn 2013. Sign up for our

Data repositories

Dryad is an international repository of data underlying peer-reviewed articles in the basic and applied biosciences. Dryad enables scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. Dryad is governed by a consortium of journals that collaboratively promote data archiving and ensure the sustainability of the repository.

The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.
Most of the data are freely available and can be used under the terms of the license mentioned on the data set description. A few password protected data sets are under moratorium from ongoing projects. The description of each data set is always visible and includes the principle investigator (PI) who may be asked for access.
Each dataset can be identified, shared, published and cited by using a Digital Object Identifier (DOI). Data are archived as supplements to publications or as citable data collections. Citations are available through the portal of the German National Library of Science and Technology (GetInfo).

ESA Data Registry s a publicly accessible registry describing scientific data sets on ecology and the environment. The data sets registered here are associated with articles published in the journals of the Ecological Society of America. They are registered here in order to facilitate communication and data sharing by scientists.

THE DATA CITATION INDEX℠

DEFINITIONS:

Data repository: a database or collection comprising data studies, data sets and/or microcitations which stores and provides access to the raw data. Constituent data studies, and sometimes individual data sets, are marked up with metadata providing a context for the available raw data.

Data study: description of studies or experiments held in repositories with the associated data which have been used in the data study. (Includes serial or longitudinal studies over time). Data studies can be a citable object in the literature and may have cited references attached in their metadata, together with information on such aspects as the principal investigators, funding information, subject terms, geographic coverage etc. The level of metadata provided varies between repositories.

Data set: a single or coherent set of data or a data file provided by the repository, as part of a collection, data study or experiment. Data sets may present in a number of file formats and media types: they may be number based files such as spreadsheets, images, video, audio, databases etc. Data sets can be a citable object in the literature and may have cited references attached in their metadata, but more commonly they inherit the metadata of the overall study in which they are used.

1. Title: **ESTerases and alpha/beta Hydrolase Enzymes and Relatives.**
Editor(s): Hotelier, Thierry; Renault, Ludovic; Cousin, Xavier; et al.
Source: ESTerases and alpha/beta Hydrolase Enzymes and Relatives
Source URL: **http://bioweb.ensam.inra.fr/ESTHER/general?what=index**
Document Type: Repository Times Cited: **2** (from All Databases)
[⊞ View abstract]

1. Title: **Enzymatic Activity and Protein Interactions in Alpha/Beta Hydrolase Fold Proteins: Moonlighting Versus Promiscuity**
Author(s): Marchot, Pascale; Chatonnet, Arnaud
Source: PROTEIN AND PEPTIDE LETTERS Volume: **19** Issue: **2** Pages: **132-143** Published: **FEB 2012**
Times Cited: **3** (from All Databases)
S·F·X [⊞ View abstract]

2. Title: **ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins**
Author(s): Hotelier, T; Renault, L; Cousin, X; et al.
Source: NUCLEIC ACIDS RESEARCH Volume: **32** Special Issue: **SI** Pages: **D145-D147** DOI: **10.1093/nar/gkh141** Published: **JAN 1 2004**
Times Cited: **79** (from All Databases)
S·F·X Full Text [⊞ View abstract]

Building a Culture of Data Citation

CREATE

Australian researcher creates a research dataset and a publication related to the dataset



Dataset is stored in a publicly accessible repository



1

Researcher uses ANDS services to mint a Digital Object Identifier (**doi**) for the dataset





2

Researcher future funding and promotion influenced by dataset citation metrics



REWARD

7

3



doi is used in data citation

4

USE



Research community use the **doi** to access the dataset and carry out related research





Research community generate new publications using the **doi** to reference the dataset

5

Citation metrics services (eg Scopus, Web of Knowledge) accumulate citation references to the dataset and publication





MEASURE

6

Funding and research groups review publication and dataset citation metrics

ands.org.au

SCIENCE & IMPACT

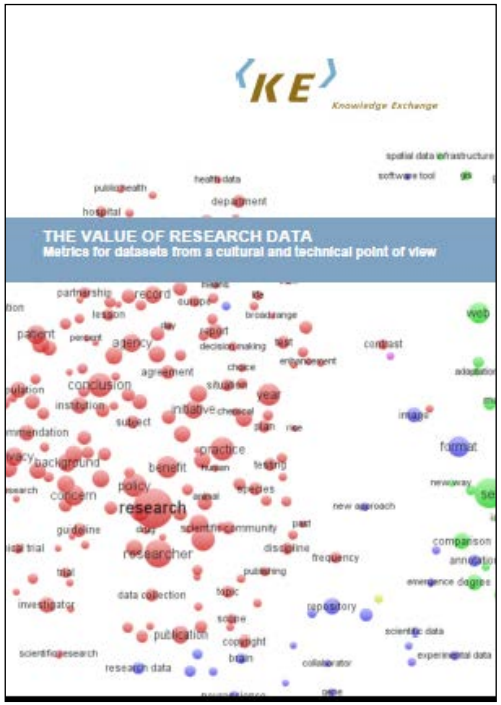
ands

AUSTRALIAN NATIONAL DATA SERVICE

Research data assessment

- The value of research data –Knowledge Exchange report, april 2013

http://www.knowledge-exchange.info/datametrics



_02

What you could do ….

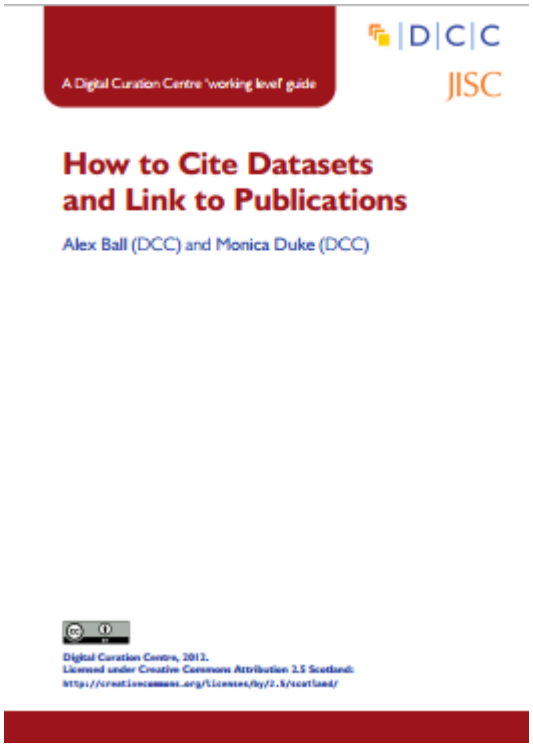
From level 0 of complexity to 3 ….

Level 0 : your author guideline (AG)

- Request for data for the peer reviewing process – research quality
- CC0 for data inside articles
- data as supplementary material

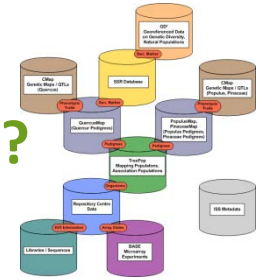
Level 1 : Deposit and cite (AG)

- Give advise to cite data http://www.dcc.ac.uk/resources/how-guides/cite-datasets
- Say that data must be in a public repository



Level 2 : Repository

Evoltree ?



- Your own « Forest data » repository ?
  - ✓ yes : to be a strongest scientific domain
  - ✓ standard development (metadata, data …)
  - ✓ No specific development use sthg like Dspace
- Partnership with an existing one ?
  - ✓ Dryad http://datadryad.org/
  - ✓ KNB http://knb.ecoinformatics.org/m/
  - ✓ ?

Level 3 : help the development of information interoperability

- Develop domain ontologies
- Georeference your information
- Use or develop standards for data interoperability

Level 4 : Forest data journal ?

- to make your scientific domain visible
- to have more reusable data
- to help the developement of standards and data exchange
- to improve the research quality
- to promote a data sharing culture

Summary : What you should do !

- Add an item « Data » in your author guidelines, not only sup. mat.
- Identify public repositories
- Think about your scientific strategy
  - ✓ forest data repository
  - ✓ forest data journal

- And think « open » : open access, open data and open science

A wider sight

Knowledge exchange Report

http://www.knowledge-exchange.info/datametrics

Table 6: Recommendations for the different stakeholders

Stakeholders	Recommendations
Funders	* Demand and reward data sharing activities * Consider data metrics in assessments * Inform policy about the importance and benefits of data sharing * Promote open access of data
Research infrastructures	* Promote policies of data sharing * Promote arguments and incentives in favour of data sharing * Provide options and alternatives to the different types of data sharing activities * Professionalize staff and standardize data sharing activities (collection, curation, dissemination)
Scientists	* Include data sharing as good scientific and scholarly practice * Promote data citation as the formal way of acknowledging data sharing * Perform more research on benefits and possibilities of data sharing * Define codes of conducts for disciplines considering appropriate regulations, i.e. embargo periods, anonymisation etc.
Data centres	* Inform the scientific community about data activities and services * Contribute to reduce the dispersion of data repositories * Develop robust solutions for the preservation and standardisation of the data storage and citations * Develop tools for tracking the users of the repositories
Publishers	* Promote data sharing in their publications and journals * Inform authors about other data sharing stakeholders (e.g. repositories, data centres) * Support open access to data
Libraries	* Promote data publications and data citations * Coach scholars and research managers in their data publication and citation activities * Inform authors about other data sharing stakeholders (e.g. funders, repositories, data centres) * Develop tools to find data repositories * Develop and test appropriate metrics
Publication databases	* Collect and measure data publications and data citations * Facilitate the analysis and metrics of data publications and data citations





Merci pour votre attention…

odile.hologne@versailles.inra.fr

http://prodinra.inra.fr/au/20625T

@Holo_08

