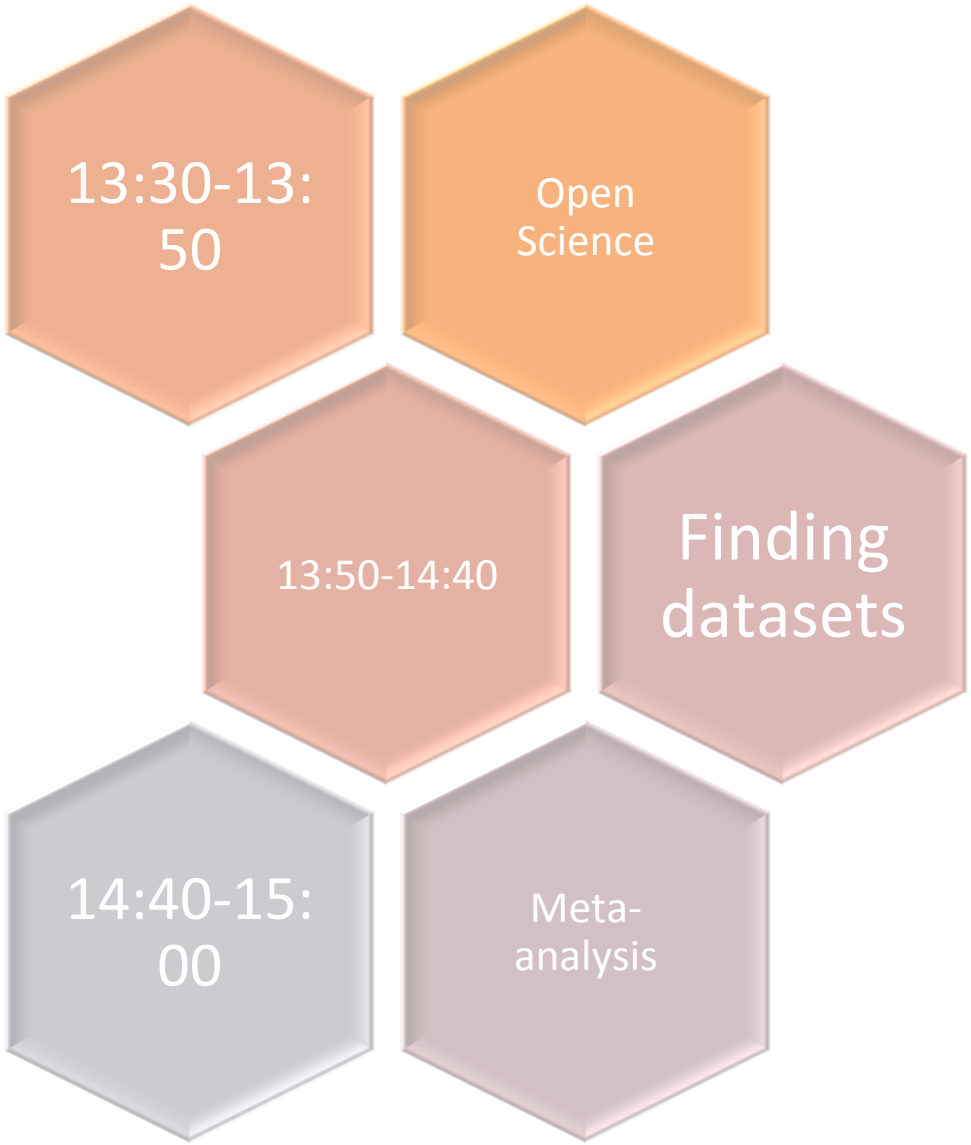
Evolutionary biology and Open Science: practices, challenges and opportunities



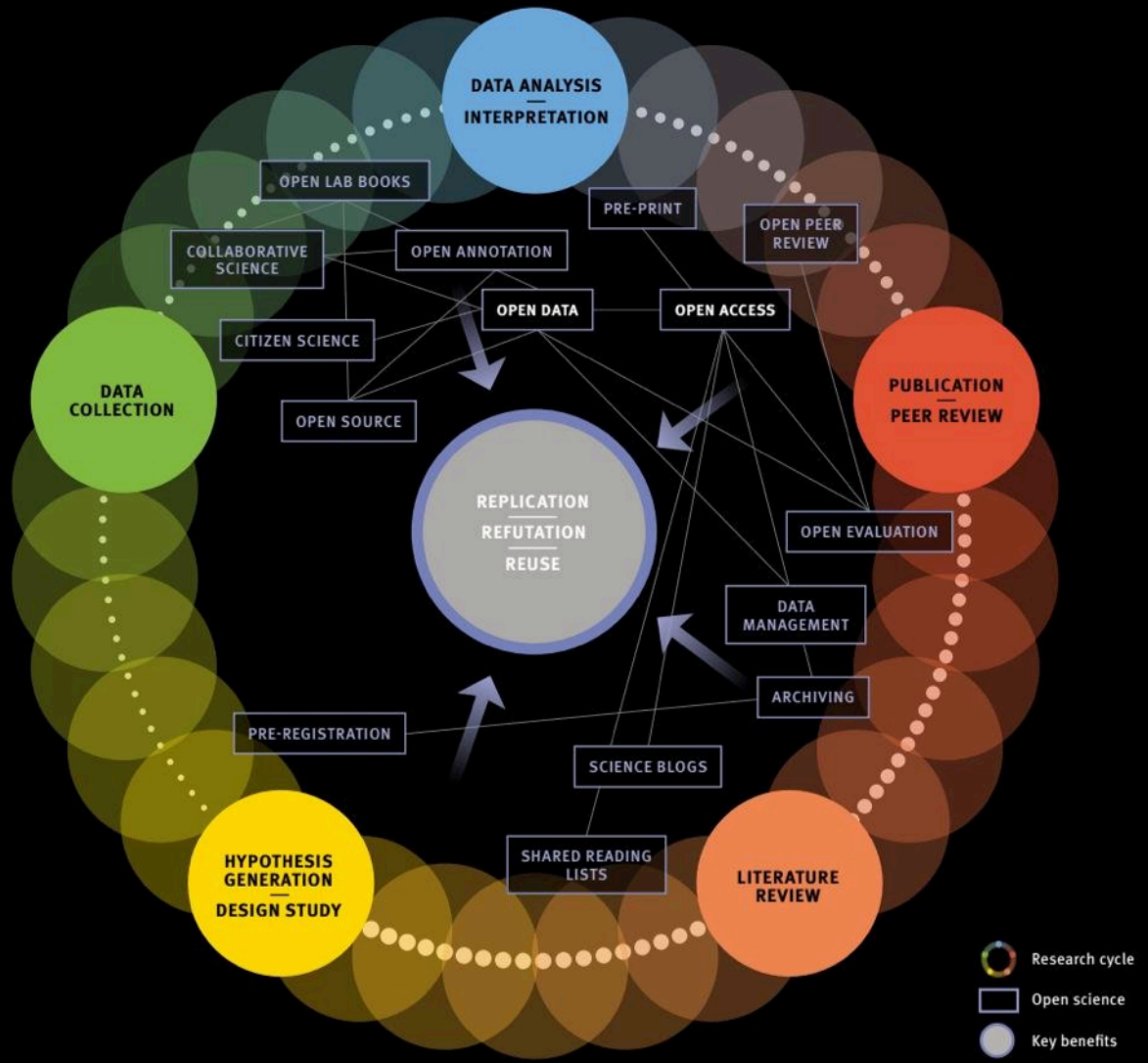
By Antica Culina
Netherlands Institute of Ecology
@antica_c
A.Culina@nioo.knaw.nl

By 2020, we want all European researchers to be able to deposit, access and analyse European scientific data through a European Open Science Cloud..

Speech by Commissioner Carlos Moedas in Amsterdam, NL: "Open science: share and succeed", 4 April 2016

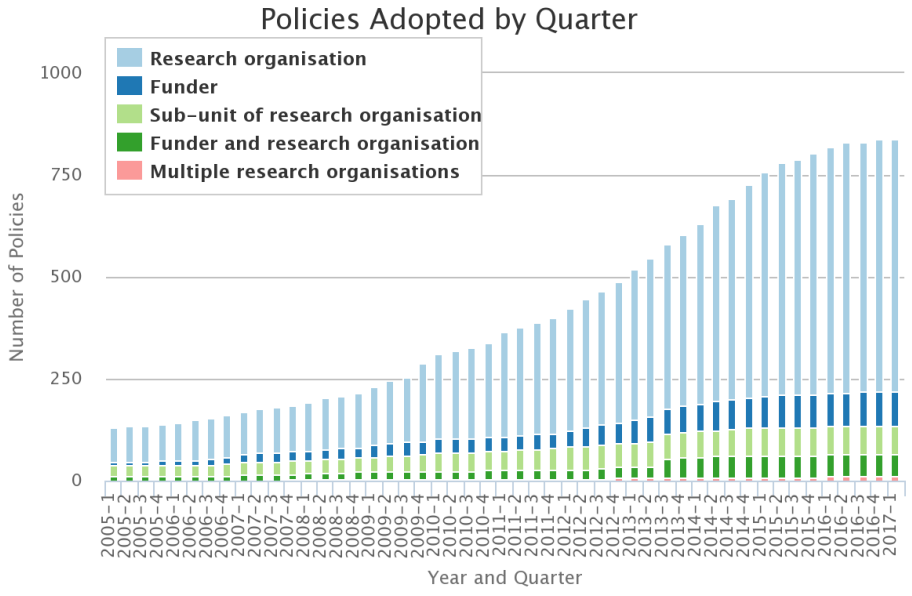
The elements of open science

This grassroots movement has created a plethora of new concepts. Here's an overview.



ROARMAP

Registry of Open Access Repository Mandates and Policies



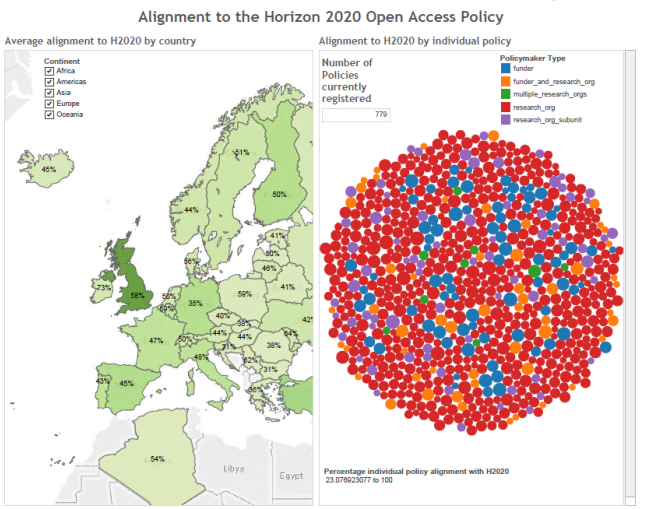


Table 1 A list, with brief explanations, of each of the eight ex
parency and Openness Promotion (TOP) guidelines (https://c

TOP Guideline
1. Citation standards (citation of data sets, etc.)
2. Data transparency (data archiving)
3. Analytic methods (code) transparency (code archiving)
4. Research materials transparency (materials archiving)
5. Design and analysis transparency (reporting of details of methods and results)
6. Pre-registration of studies (registering study prior to initiation)
7. Pre-registration of analysis plans (registering analysis plan prior to study initiation)
8. Replication (a study designed to replicate a previously published study)

Conservation Biology

Editorial

Promoting transparency in conservation science

ECOLOGY LETTERS

Ecology Letters, (2016) **19**: 726–728 doi: 10.1111/ele.12610

EDITORIAL **Promoting transparency in evolutionary biology and ecology**

T. H. Parker,[1]* S. Nakagawa,[2] J. Gurevitch,[3] and IIEE (Improving Inference in Evolutionary Biology and Ecology) workshop participants[†]

[1]*Department of Biology Whitman College Walla Walla, USA,* [2]*School of Biological Earth and Environmental Sciences University of New South Wales Sydney, Australia and* [3]*Department of Ecology and Evolution Stony Brook University USA*

Transparency and reproducibility in evolutionary research

Ruth G. Shaw,[1] Allen J. Moore, Mohamed Noor, and Michael G. Ritchie

[1]*E-mail: shawx016@umn.edu*

Received June 6, 2016
Accepted June 6, 2016

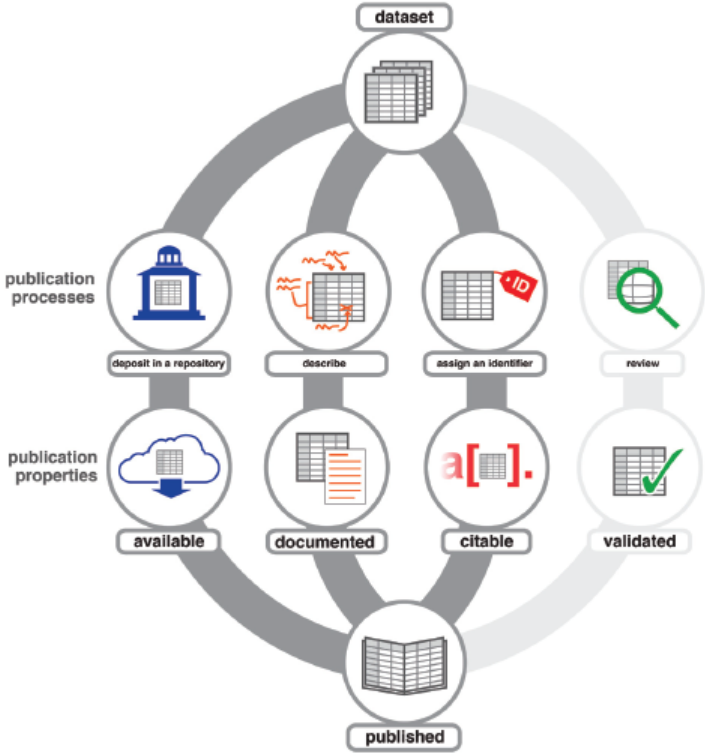
Promoting Transparency in Evolutionary Biology and Ecology

Author(s): James F. Smith , T. H. Parker , S. Nakagawa , J. Gurevitch , and TTEE (Tools for Transparency in Ecology and Evolution) Working Group

Source: Systematic Botany, 41(3):495-497.

Open data = a piece of data that anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike. Equivalent to FAIR data.

FAIR data = to enable data to be found and used, data should ideally adhere to the Findable Accessible Interoperable and Reusable (FAIR) principles. FAIR data are equivalent to *Open Data.*



Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

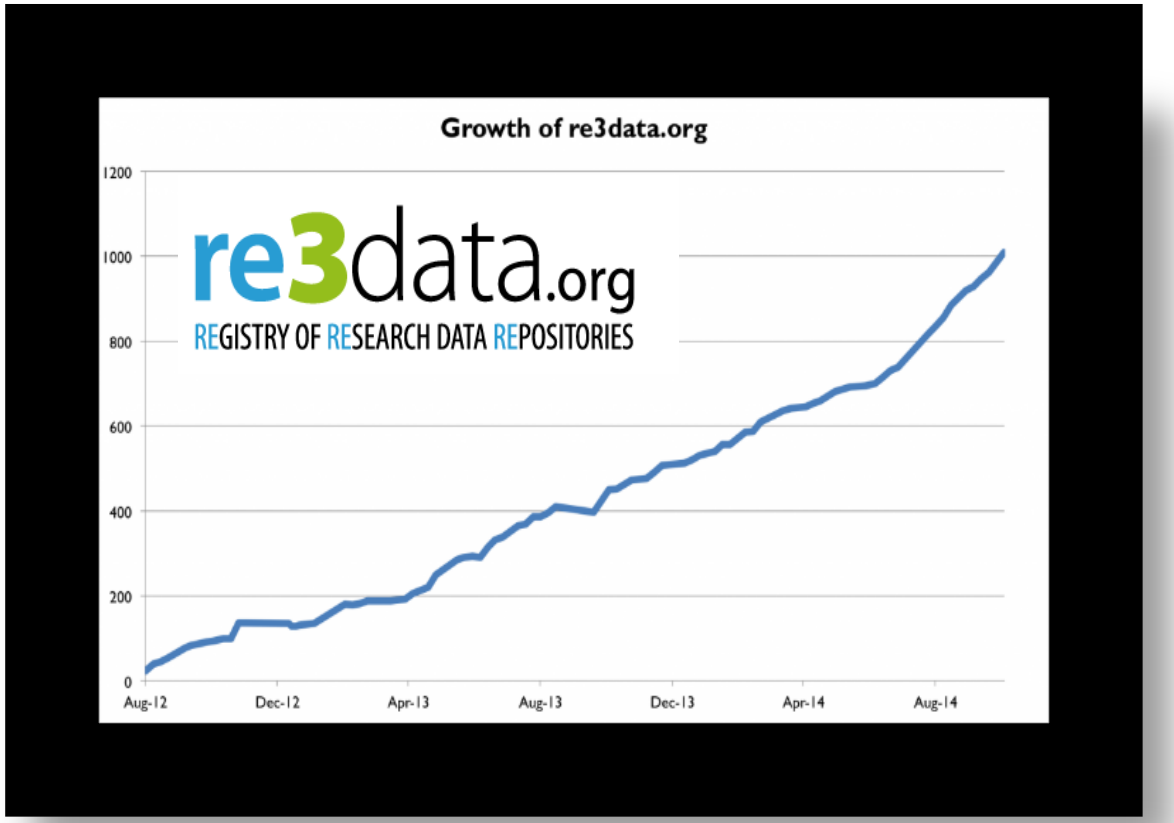
To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

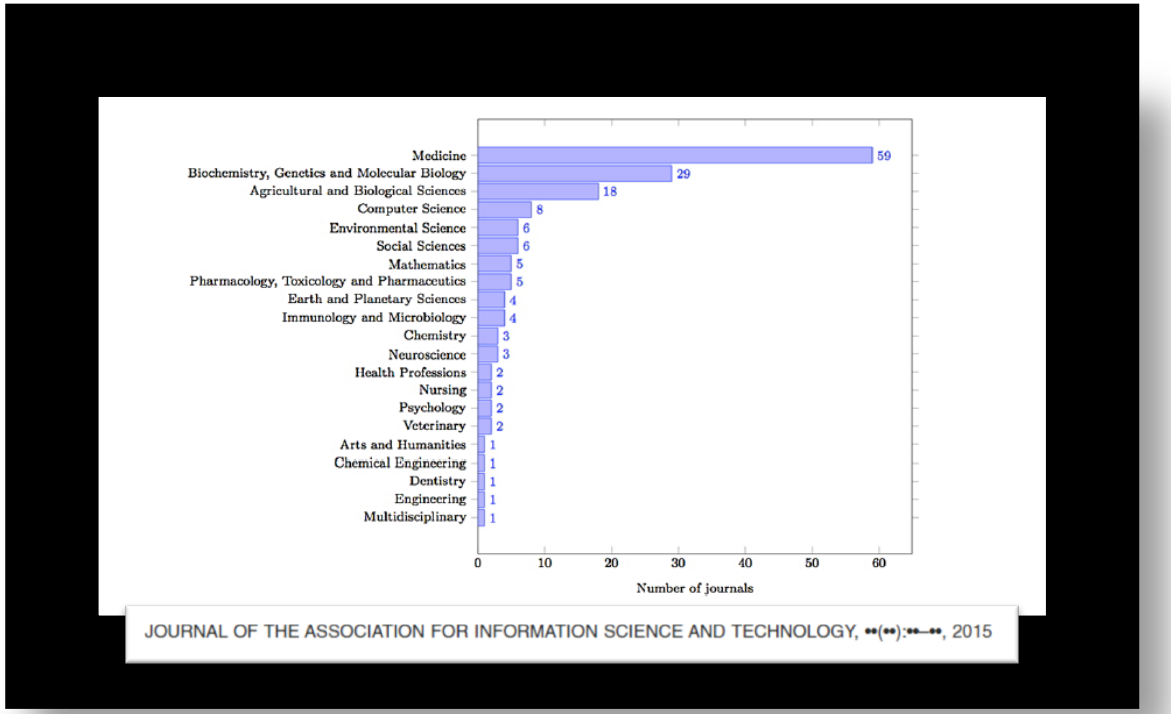
To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Figure 1. **To be published, datasets are typically deposited in a repository to make them available, documented to support reproduction and reuse, and assigned an identifier to facilitate citation.** Some, but not all, publishers review datasets to validate them.

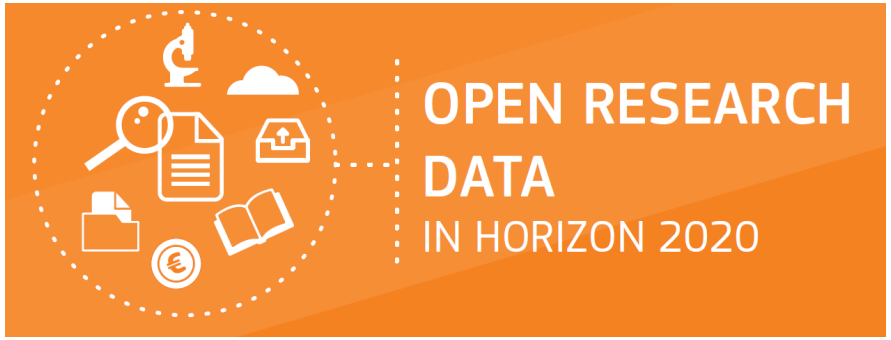
Data journals





nature research

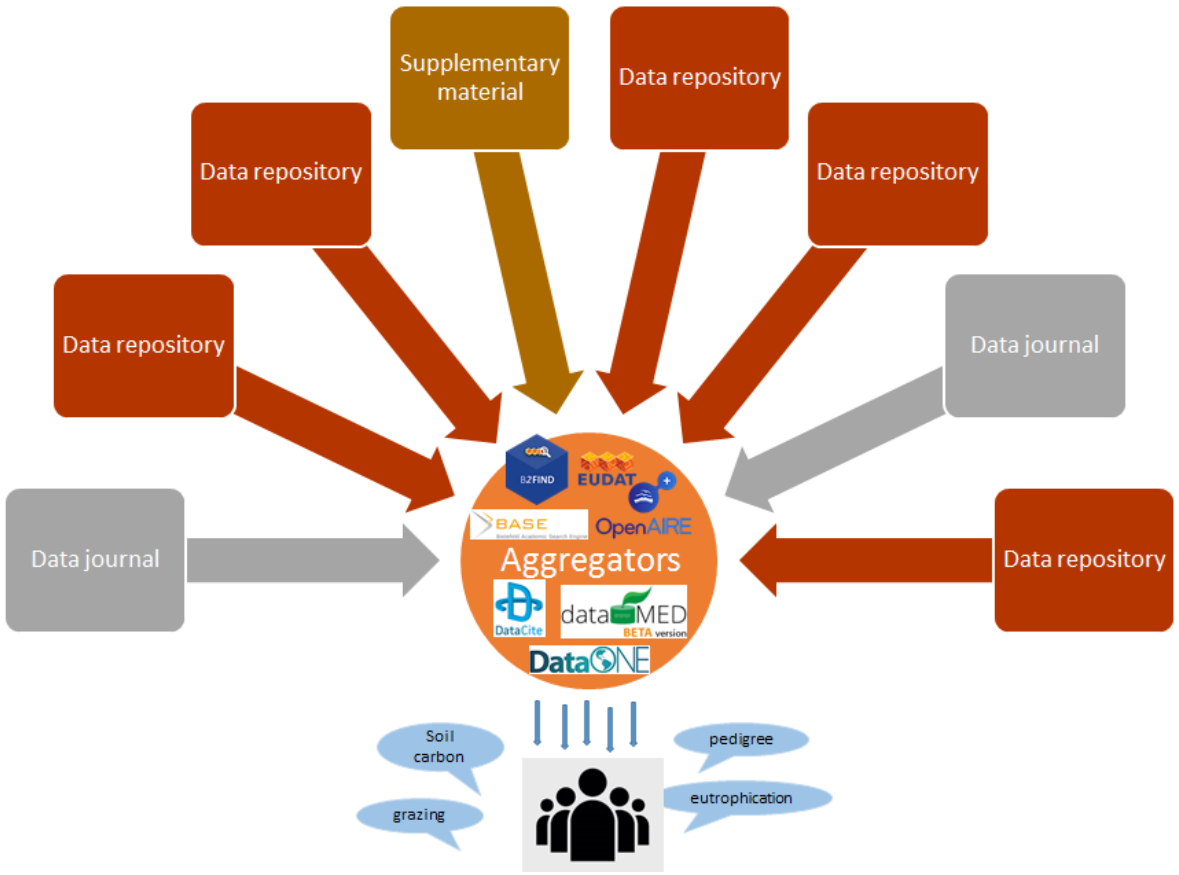
Data availability statements and data citations policy: guidance for authors



Where to find datasets?

Data are usually in:

- Data repository
- Data journal
- Supplement of a publication
- Personal website



- **Aggregators of data repositories** harvest through or host metadata from a set of data repositories.

- **Registries of data sources** are directories of data sources that are intended to provide an organized, up-to-date, and searchable collection of data sources.

- **Data sources with links to datasets** offer the possibility to reach datasets via links from other scholarly objects, such as literature (e.g. scientific articles, theses, reports)

- **Virtual Research Environments**: provide web user tools for scientists to collaborate or process/manipulate data.

What data I need?

Are there community specific primary data sources? → Identify using ROAR (ROAR http://roar.eprints.org/cgi/search/advanced) or Re3Data.org http://www.re3data.org/

- **Yes**: Search the data within the community specific data sources
- **No**: Use *data discovery sources* to search for data across many primary data sources https://ckan-ecoevo.d4science.org/

I have obtained the datasets

Am I allowed do use data (legal considerations)?

1) *Copyright and database rights attached*
- *CC-BY licence = share and adapt the data, attribution (e.g. citation) required.*
- *CC0 = data in the public domain (no requirements attached)*
- *no permission has been granted = ask permission of reuse from the data owner*

2) Research data protected by terms of use = comply with these

Avoid misinterpretation and biases:

- contact the authors, think about potential spatial, temporal, taxonomic etc biases

Credit the authors

- many datasets have DOI and are accompanied with the information on how to cite them

Name	What type of content	URL (for the search part of the platform)
Aggregators of data repositories		
DataCite	datasets	https://search.datacite.org/
WorldWideScience	datasets, literature, multimedia	http://worldwidescience.org/index.html
BASE	datasets, literature, multimedia, software, other	https://www.base-search.net/
Share	datasets, literature, multimedia, projects, other	https://share.osf.io/discover
Dataone and One Mercury*	datasets	https://search.dataone.org/#data/page/0 https://cn.dataone.org/onemercury/
Science research	datasets, literature, software, multimedia, other	http://scienceresearch.com/scienceresearch/advancedsearch.html
Research Data Australia	datasets	https://researchdata.ands.org.au/
B2Find	datasets, literature, other	http://b2find.eudat.eu/
DataHub	datasets	https://datahub.io/dataset
Dliservice portal	datasets, linked with publications	https://dliservice.research-infrastructures.eu/index.html#/
DataMed	datasets	https://datamed.org/index.php
UK Research data discovery service	datasets	http://ckan.data.alpha.jisc.ac.uk/dataset
ZanRan	datasets	http://www.zanran.com/q/
DataSearch	datasets	https://datasearch.elsevier.com/#/

Data sources with links to datasets		
Europe PMC	Literature (links to datasets)	http://europepmc.org/
OpenAIRE	datasets, literature, software, services	https://www.openaire.eu/search/
BioStudies	descriptions of studies, links to their data	http://www.ebi.ac.uk/biostudies/
GoOA	oa journals + additional files which include tables and supplementary materials, so one can search for data	http://gooa.las.ac.cn/external/about-us.jsp

Registries of data sources		
ROAR	repositories and datasets	http://roar.eprints.org/content.html
OpenDOAR	datasets, literature, software, multimedia	http://www.opendoar.org/search.php
Virtual Research Environments		
D4Science Integrated Data Catalogue	databases, datasets, repositories	https://www.d4science.org/integrated-data-catalogue
Marine LifeWatch	databases, repositories, methods	http://marine.lifewatch.eu

Some other links

https://101innovations.wordpress.com/
http://innoscholcomm.silk.co/
https://docs.google.com/spreadsheets/d/1KUMSeq_Pzp4KveZ7pb5rddcssk1XBTiLHniD0d3nDqo/edit#gid=0

THE ROYAL SOCIETY biology letters

A survey of publication bias within evolutionary ecology

Negative results are disappearing from most disciplines and countries

Daniele Fanelli

"Positive" Results Increase Down the Hierarchy of the Sciences

Daniele Fanelli*

How Many Scientists Fabric
Systematic Review and Met

Daniele Fanelli*

182 | NATURE | VOL 526 | 8 OCTOBER 2015

FOOLING OURSELVES

HUMANS ARE REMARKABLY GOOD AT SELF-DECEPTION.
BUT GROWING CONCERN ABOUT REPRODUCIBILITY IS DRIVING MANY
RESEARCHERS TO SEEK WAYS TO FIGHT THEIR OWN WORST INSTINCTS.

PLoS one

Dissemination biases in ecology: effect sizes matter more than quality

E. Kathryn Barto and Matthias C. Rillig