

THE EVIL IS IN THE DETAILS

(something more on Open Access,
Data Management and FAIR data)



Elena Giglia
IIT, Genova, 2019
elena.giglia@unito.it
 @egiglia



Open Access

The [Open Content](#) definition defines five rights^[2], also known as the five Rs, that should hold true for a copyrightable work to be truly open:

- **Retain** - the right to make, own, and control copies of the content (eg, download, duplicate, store, and manage)
- **Reuse** - the right to use the content in a wide range of ways (eg, in a class, in a study group, on a website, in a video)
- **Revise** - the right to adapt, adjust, modify, or alter the content itself (eg, translate the content into another language)
- **Remix** - the right to combine the original or revised content with other material to create something new (eg, incorporate the content into a mashup)
- **Redistribute** - the right to share copies of the original content, your revisions, or your remixes with others (eg, give a copy of the content to a friend)

<http://opencontent.org/definition/>



Open Access

Berlin Declaration

1. The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards, will continue to provide the mechanism for enforcement of proper attribution and responsible use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.

The pillars

The background of the slide is a photograph of the Temple of Apollo at the Acropolis in Athens. It shows several tall, fluted Doric columns supporting a heavy entablature. The sky is a clear, vibrant blue. The text is overlaid on this image.

Knowledge is a common

Scholarly communication is
a great conversation

The results of publicly funded research
must be publicly available

Benefits / real impact

THE IMPACT OF OPEN ACCESS

Demonstrating Achievements

Open access for Institutions

Building Reputation	Garnering Funding	Developing Alumni Connections
Strengthening Recruiting	Demonstrating Learning Outcomes	Professionalizing Students
Contributing Stories for Public Relations	Aiding Accreditation	Boosting Operational Efficiency

Reputation Building

Open access for Authors

Amplifying Scholarly Expertise	Preserving Scholarly Legacy	Developing Niche Field
Forging Business & Government Partnerships	Popularizing Research in Mainstream Outlets	Finding Collaborators
Modeling Innovative Teaching	Launching Scholarly Career	Improving Quality

Advancing Knowledge

Open access for Readers

Affecting Public Policy	Advancing Innovation	Linking Global Experts
Building Local Community	Improving Access to Education	Datamining Research
Updating Practitioners	Connecting Cultures	Informing Prospective Applicants

How Open Access works /
green and gold



Green road – self-archiving

A vintage bicycle with a brown frame and black tires is leaning against a light-colored stone wall. The bicycle has a green fender on the front and a rear rack. Large green leaves are growing in front of the bicycle, partially obscuring it. A white mailbox is visible at the bottom right.

**Author self-archives
in an Open Access repository
his/her final version of the paper,
wherever it was published,
according to copyright agreements**

Green road – self-archiving

82% of international publishers allows some kind of self-archiving (Elsevier, Wiley, Springer...); list:



...possible limitations:

- no pdf with the publisher's layout
- embargo period requested

ONLY IF
you have signed a copyright
transfer agreement...
if not, **the author** is the rights
holder

Author's Pre-print:	✓ author can archive pre-print (ie pre-refereeing)
Author's Post-print:	✗ subject to Restrictions below , author can archive post-print (ie final draft post-refereeing)
Restrictions:	• 12 months embargo
Publisher's Version/PDF:	✗ author cannot archive publisher's version/PDF
General Conditions:	• Some journals have separate policies, please check with each journal directly • On author's personal website, institutional repositories, arXiv, AgEcon, PhilPapers, PubMed Central, RePEc or Social • Author's pre-print may not be updated with Publisher's Version/PDF • Author's pre-print must acknowledge acceptance for publication • Non-Commercial • Publisher's version/PDF cannot be used • Publisher source must be acknowledged with citation • Must link to publisher version with set statement (see policy) • If OnlineOpen is available, BBSRC, EPSRC, MRC, NERC and STFC authors, may self-archive after 12 months • If OnlineOpen is available, AHRC and ESRC authors, may self-archive after 24 months

Arsenate toxicity on the apices of *Pisum sativum* L. seedling roots: Effects on mitotic activity, chromatin integrity and microtubules

Stefania Dho, Wanda Camusso, Marco Mucciarelli, Anna Fusconi

Abstract

Arsenic (As) is one of the most toxic plant growth. Despite the growing of this element on meristem activity study, short-term experiments with whether plant growth impairment was studied by evaluating apical fragmentation and microtubule or that arsenate, at the lowest concentrations, whilst the other concentrations, mitotic and labelling index (after bromo-deoxyuridine administration) (through immunofluorescence). Mitotic metaphases increased, as did the mitotic spindles, which closely resemble those induced by colchicine. Chromosome breaks and anaphase/telophase bridges were virtually absent. These data point to a poor clastogenic activity of As and implicate that microtubules are one of the main targets of As.

Keywords

Pea; Arsenic; Apical meristems; Aberrations; Immunofluorescence; TUNEL test

1. Introduction

Arsenic (As) is a toxic element, frequently found in soils and water. A main natural source of As is the erosion of mother rock, even though a consistent part of As environmental pollution comes from human activities (Meharg and Hartley-Whitaker, 2002 and Patra et al., 2004). The As in unpolluted fresh water is usually in the range 1–10 µg/l. According to EPA and WHO, the maximum permissible As concentration in drinking water is 50 µg/l (Mandal and Suzuki, 2002).

Arsenic is a well-established human carcinogen (Qin et al., 2008a) and has been shown to be genotoxic in a variety of *in vitro* studies (Hughes, 2002). In plants, it severely affects growth and development, and its toxicity is strongly dependent on the concentration, exposure time and physiological state of the plant (Singh et al., 2007). However, plants vary in their sensitivity to As, and a wide range of species have been identified in As-contaminated soils (Meharg and Hartley-Whitaker, 2002). Besides, hyperaccumulators such as *Pteris vittata*, which tolerate high internal As content, may also use this As to defence themselves against herbivore attack (Mathews et al., 2009).

Higher plants take up As mainly as arsenate (V), the dominant form of phytoavailable As in aerobic soils. According to Meharg and Hartley-Whitaker (2002), As competes with phosphate for plant phosphate transporters. Upon absorption, most arsenate is rapidly reduced to arsenite (III), due to an arsenate reductase activity (Xu et al., 2007), hence, the arsenate cytoplasmic concentration is generally not high enough to exert toxicity (Meharg and Hartley-Whitaker, 2002). Both As species interfere with various metabolic pathways: arsenate, as an analogous chemical to phosphate, may replace phosphate in the ATP and in various



UNIVERSITÀ DEGLI STUDI DI TORINO

This Accepted Author Manuscript (AAM) is copyrighted and published by Elsevier. It is posted here by agreement between Elsevier and the University of Turin. Changes resulting from the publishing process – such as editing, corrections, structural formatting, and other quality control mechanisms – may not be reflected in this version of the text. The definitive version of the text was subsequently published in ENVIRONMENTAL AND EXPERIMENTAL BOTANY, 69(1), 2010, 10.1016/j.envexpbot.2010.02.010.

You may download, copy and otherwise use the AAM for non-commercial purposes provided that your license is limited by the following restrictions:

- (1) You may use this AAM for non-commercial purposes only under the terms of the CC-BY-NC-ND license.
- (2) The integrity of the work and identification of the author, copyright owner, and publisher must be preserved in any copy.
- (3) You must attribute this AAM in the following format: Creative Commons BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>), 10.1016/j.envexpbot.2010.02.010

The definitive version is available at:
<http://linkinghub.elsevier.com/retrieve/pii/S0098847210000353>



Contents lists available at ScienceDirect

Environmental and Experimental Botany

journal homepage: www.elsevier.com/locate/envexpbot



Arsenate toxicity on the apices of *Pisum sativum* L. seedling roots: Effects on mitotic activity, chromatin integrity and microtubules

Stefania Dho^a, Wanda Camusso^a, Marco Mucciarelli^b, Anna Fusconi^{a,*}

^a Dipartimento di Biologia Vegetale, CEBIOVEM, Viale Mattioli 25, I-10125 Torino, Italy

^b Dipartimento di Morfologia Vegetale, Via Leonardo da Vinci 44, I-10095 Grugliasco (To), Italy

ARTICLE INFO

Article history:

Received 20 July 2009

Received in revised form 9 February 2010

Accepted 14 February 2010

Keywords:

Pea
Arsenic
Apical meristems
Aberrations
Immunofluorescence
TUNEL test

ABSTRACT

Arsenic (As) is one of the most toxic pollutants in the environment, where it severely affects both animal and plant growth. Despite the growing literature data on As effects on plant development, alterations induced by this element on meristem activity of the root have not been explored to any great extent. In the present study, short-term experiments with arsenate have been conducted on *Pisum sativum* L. seedlings to assess whether plant growth impairment is due to DNA/chromosome or mitotic microtubule damages. Root growth was studied by evaluating apical meristem activity and cell elongation. Mitotic aberrations, DNA fragmentation and microtubule organization of the apical cells were also analyzed. The results have shown that arsenate, at the lowest concentration (0.25 µM), slightly increases root growth and some related parameters, whilst the other concentrations have a dose-dependent negative effect on root growth, on the mitotic and labelling index (after bromo-deoxyuridine administration), and on the mitotic arrays of microtubule (through immunofluorescence). The main effects on mitosis occurred for 25 µM As. The percentage of metaphases increased, as did the irregular metaphases and c-mitoses. This was related to alterations in the mitotic spindles, which closely resemble those induced by colchicine. Chromosome breaks and anaphase/telophase bridges were virtually absent, whilst DNA fragmentation only increased from 25 µM arsenate onwards. These data point to a poor clastogenic activity of As and implicate that microtubules are one of the main targets of As.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Arsenic (As) is a toxic element, frequently found in soils and water. A main natural source of As is the erosion of mother rock, even though a consistent part of As environmental pollution comes from human activities (Meharg and Hartley-Whitaker, 2002; Patra et al., 2004). The As in unpolluted fresh water is usually in the range 1–10 µg/l. According to EPA and WHO, the maximum permissible As concentration in drinking water is 50 µg/l (Mandal and Suzuki, 2002).

Arsenic is a well-established human carcinogen (Qin et al., 2008a) and has been shown to be genotoxic in a variety of *in vitro* studies (Hughes, 2002). In plants, it severely affects growth and development, and its toxicity is strongly dependent on the concentration, exposure time and physiological state of the plant (Singh et al., 2007). However, plants vary in their sensitivity to As, and a wide range of species have been identified in As-contaminated soils (Meharg and Hartley-Whitaker, 2002). Besides, hyperaccumulators such as *Pteris vittata*, which tolerate high internal As content,

may also use this As to defence themselves against herbivore attack (Mathews et al., 2009).

Higher plants take up As mainly as arsenate (V), the dominant form of phytoavailable As in aerobic soils. According to Meharg and Hartley-Whitaker (2002), As competes with phosphate for plant phosphate transporters. Upon absorption, most arsenate is rapidly reduced to arsenite (III), due to an arsenate reductase activity (Xu et al., 2007), hence, the arsenate cytoplasmic concentration is generally not high enough to exert toxicity (Meharg and Hartley-Whitaker, 2002). Both As species interfere with various metabolic pathways: arsenate, as an analogous chemical to phosphate, may replace phosphate in the ATP and in various phosphorylation reactions, leading to the disruption of the energy flow in cells. The toxicity of arsenite is mainly ascribed to its reaction with sulphhydryl groups of proteins that interfere with their functions (Meharg and Hartley-Whitaker, 2002; Patra et al., 2004).

Exposure to high concentrations of As induces the production of reactive oxygen species (ROS) (Singh et al., 2007; Wang et al., 2007; Lin et al., 2008; Shri et al., 2009) and the conversion of arsenate to arsenite is regarded as one of the causes of ROS generation (Wang et al., 2007). Oxidative stress induced by As can damage cells, mainly through lipid peroxidation of membranes (Singh et al., 2007) and DNA fragmentation, as has been demonstrated in leaves and roots

* Corresponding author. Tel.: +39 011 6705968; fax: +39 011 6705962.
E-mail address: anna.fusconi@unito.it (A. Fusconi).

AUTHORS DON'T HAVE TO CHANGE THEIR HABITS

- **YOU CAN GO ON PUBLISHING IN THE MOST REPUTATED JOURNALS/SERIES**, as requested by research assessment exercises
- then you make the content free by self-archiving your final version of the work



YOU-CAN-DO-IT!!!!!!

YOU CAN DO OPEN ACCESS
WHATEVER THE CRITERIA OF
RESEARCH EVALUATION

THEY ARE NOT MUTUALLY
EXCLUSIVE!!!

Green road

Why in an Open Access repository?

- provides a **unique and persistent ID** (handle), VS personal page
- long-term **preservation**
- is **no-profit** VS Research Gate / Academia.edu

- THEY CAN CLOSE TOMORROW
- THEY CAN BE BOUGHT TOMORROW

...two of a kind

OFFICE OF SCHOLARLY COMMUNICATION
UNIVERSITY OF CALIFORNIA

[HOME](#) • [FEATURES](#) • A SOCIAL NETWORKING SITE IS NOT AN OPEN ACCESS REPOSITORY

A social networking site is not an open access repository

	Open access repositories	Academia.edu	R
Supports export or harvesting	Yes	No	
Long-term preservation	Yes	No	
Business model	Nonprofit (usually)	Commercial. Sells job posting services, hopes to	

 **LC**
Lenz Caemmerer

<https://goo.gl/RnUszK>

Attorneys
and Notaries

Basel
Attorneys - Notaries:
Dr. Felix Iselin, notary
Dr. Gerd Thoenen, M.
f, notary
y, notary
y, Inheritance Law
H, LL.M.
ten-Kiefer
ten
ert
m, LL.M.
fio - EU-Attorney
ber-Stadt

ResearchGate GmbH

To
Dr
Ho
An
Ja
Inv
Be



ResearchGate vs. publishers

ResearchGate vs. Publishers: The Saga Continues...

Last updated May 8, 2018

May 2018

NEWS

ResearchGate bows to pressure from publishers on copyrighted material



BY REBECCA TRAGER | 15 NOVEMBER 2017

Networking site has moved 1.7 million journal articles from five major publishers so they are no longer accessible to the public [Nov. 15, 2017](#)

<http://osc.universityofcalifornia.edu/2015/12/a-social-networking-site-is-not-an-open-access-repository/>



Björn Brembs

@brembs

Following

In risposta a @ReaderMeter e @mrgunn

Every publisher with an embargo policy does obviously not believe they add any value themselves. So why should we?

Embargo?



1) IF YOU ASK FOR EMBARGO, WHAT KIND OF VALUE YOU ADD AS A PUBLISHER?

2) «WE PREVENT READING TO PROTECT READERS»
[!!!]



Björn Brembs

@brembs

Following

Elsevier stooping to new lows (no, didn't think it was possible either): "we prevent reading to protect readers" One more in a long list of Elsevier's insults to their customers' intelligence: bjoern.brembs.net/2017/10/why-

Segui



mrgunn

@mrgunn

In risposta a @brembs e @ReaderMeter

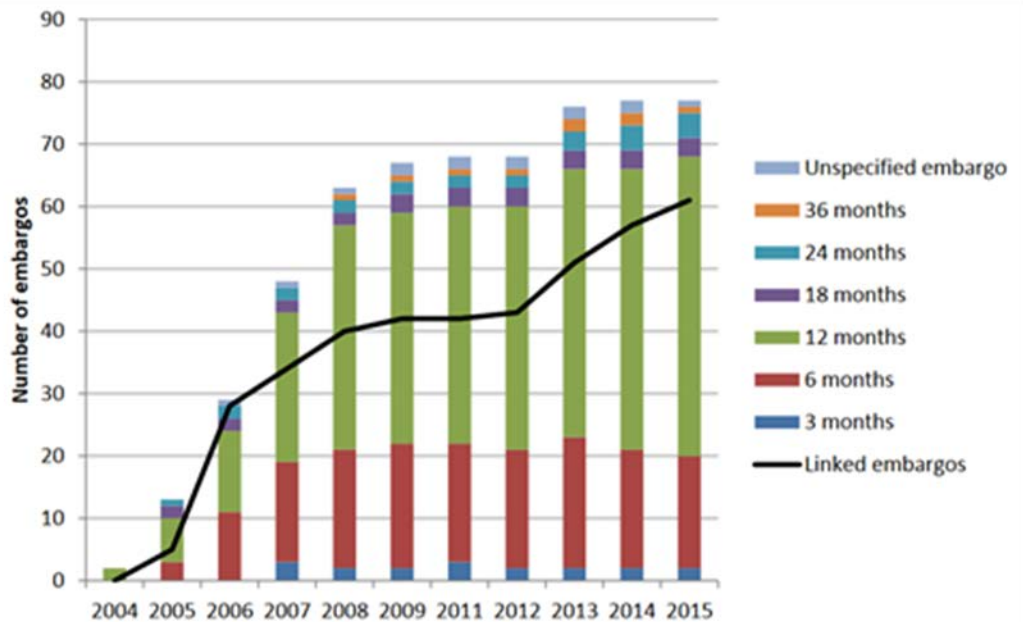
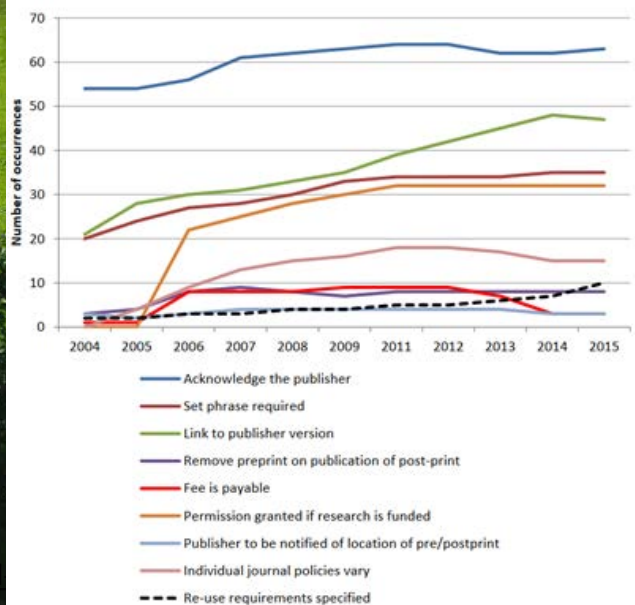
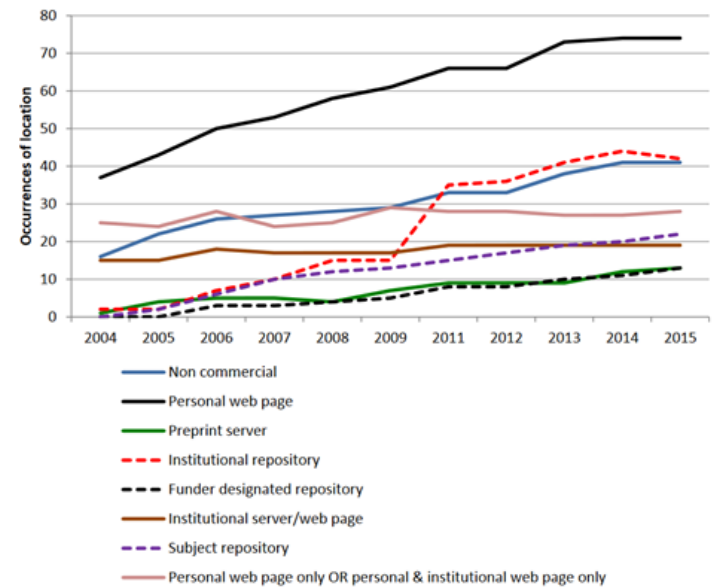
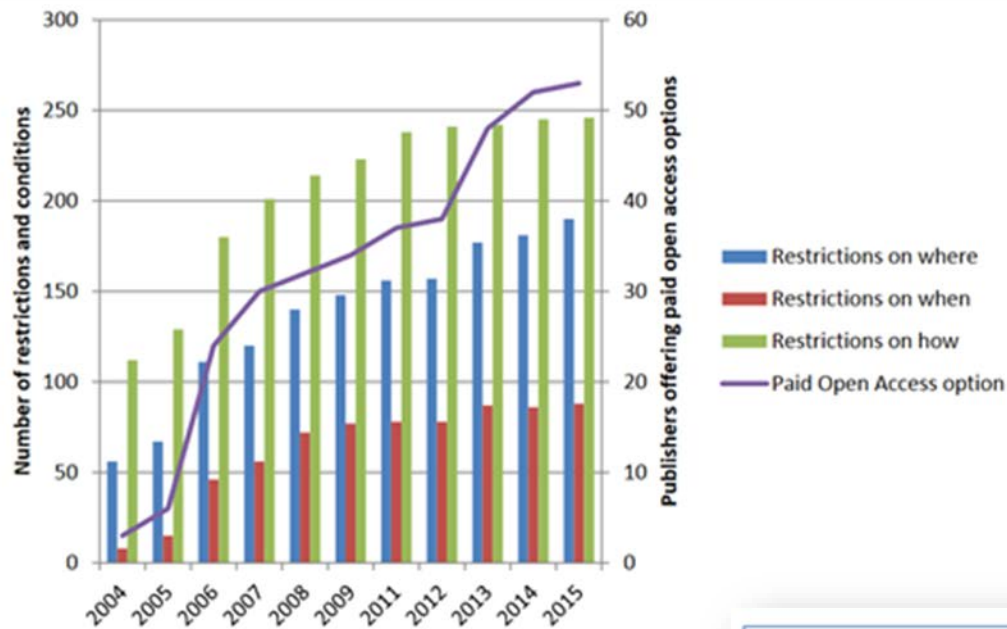
The whole point of the embargo is so that subscribers can get their money's worth. As you know, I wish everything was open access, but I'm just putting out there why embargoes exist.

aderMeter

go is so that subscribers can get their money's worth. As 'as open access, but I'm just putting out there why



A labyrinth?



Gold road Publishing Open Access

- you have to **change your publication venue**
- choose one of **more than 13.000** Open Access journals listed in **DOAJ, Directory of Open Access Journals**
- **26,6%** of the journals require **APCs, Article Processing Charges, [250 to 2900 \$]**
- [bear in mind: even traditional publishers have been (and still are) asking for fees for figure, pages over...]

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

Apply News About <https://doaj.org/API>

Pay attention!

SUBSCRIPTIONS

- paid **every year**
- **increase** every year
- **CLOSE the content** for those who have no subscription

APC

- paid **once and forever**
- **OPEN the content** to all

DON'T MIX

- NATIVE OPEN ACCESS PUBLISHERS [NO REVENUE BUT APCs]
- TRADITIONAL PUBLISHERS OFFERING AN «OPEN OPTION»
[MAIN REVENUE STREAM IS STILL SUBSCRIPTIONS]

Are OA journals different?

PUBLISH **DATA** WITH THE ARTICLE

- **TRANSPARENCY**
- **REPRODUCIBILITY**

PUBLISH **REVIEWS** WITH THE ARTICLE

- **TRANSPARENCY**
- **PIECES OF KNOWLEDGE**

PUBLISH UNDER CREATIVE COMMONS **LICENCES**

- **REUSE**
- **TEXT AND DATA MINING**

PUBLISH **QUICKLY**

(OFTEN) PUBLISH IN **MACHINE-READABLE FORMATS**

- **TEXT AND DATA MINING**

A new model



JOURNALS ARE COMMUNITY OWNED

The Fair Open Access Principles

1. The journal has a transparent ownership structure, and is controlled by and responsive to the scholarly community.
2. Authors of articles in the journal retain copyright.
3. All articles are published open access and an explicit open access licence is used.
4. Submission and publication is not conditional in any way on the payment of a fee from the author or its employing institution, or on membership of an institution or society.
5. Any fees paid on behalf of the journal to publishers are low, transparent, and in proportion to the work carried out.

The mission of the foundation is

- a. to promote and support initiatives concerning (Fair) Open Access publications in the broadest sense;
- b. to acquire resources and financially sustain (Fair) Open Access publications;
- c. to support foundations financially and otherwise in various disciplines (xxxOA's) that pursue the same goals;
- d. to expand the [Open Library of Humanities](#) to other disciplines.
- e. to propagate and promote the principles of Fair Open Access over all disciplines of science.

Quality?

Stop Predatory Journals

About Contribute Hijacked Journals Metrics Publishers

List of Predatory Journals

This is a list of possibly [predatory journals](#). The kernel for this list was extracted from the archive of Beall's list at [web.archive.org](#). It will be updated as new information or suggested edits are submitted by the maintainers of this site.

This list is only for individual journals. See the other list for [publishers](#) potentially engaging in pre practices.

ABCDEFGHIJKLMNOPQRSTUVWXYZ

A

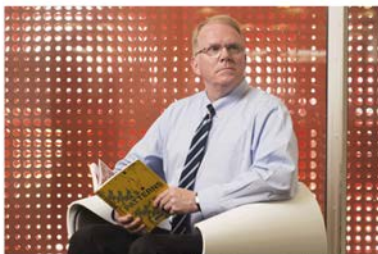
- Academic Exchange Quarterly
- Academic Research Reviews

Scholarly Open Access

PUBLISHING

Why Beall's List Died — and What It Unresolved About Open Access

By Paul Basken | SEPTEMBER 12, 2017 ✓ PREMIUM



Theo Stroomer for The Chronicle

Jeffrey Beall, an academic librarian at the U. of Colorado at Denver, abruptly shuttered a blacklist of journals he deemed untrustworthy nine months ago. But while the project has ended, debates over its merit and impact live on.

Nine months after an academic librarian deleted his careful list shaming more than a thousand scientific journals as unscrupulous, the Beall's List Mystery remains unsolved.

Why, after toiling so hard for five years and creating a resource cherished by scientists wary of exploitative publishers — did the University of Colorado at Denver's Jeffrey Beall abruptly give it all up? Who, or what, forced his hand?

it is NOT junk

a blog about genomes, DNA, evolution, open science, baseball

Beall's Litter

By MICHAEL EISEN | Published: DEC 11, 2017

Jeffrey Beall, a librarian at the University of Colorado at Denver, in science publication circles for his access publishers and curating a list of people seeking to navigate the sea of legitimate, many scammers — that is

Unfortunately, as he has gained a reputation for trying to identify bad open access publishing in general. Then that Beall is a credible contributor with an article he published last year can't really describe it. So I'm just publishing in an open access journal (so do), along with my comments

The Open-Access Movement is

Jeffrey Beall

Cites & Insights

Crawford at Large
Libraries • Policy • Technology • Media

Volume 14, Number 4: April 2014

ISSN 1534-0937

Walt Crawford

Intersections

Ethics and Access 1: The Sad Case of Jeffrey Beall

Open access (OA) is all about ethics, economics and equity, and the three interact in various ways. OA is inherently at the intersection of libraries, media, policy and technology—but that's a different issue.

This is the first of a trio of essays: two related to fairly specific situations, one covering a range of ethical discussions. Depending on how you define

ways a little different, however. He first encountered OA when reviewing a publisher, Bentham Open, for *The Charleston Advisor*. It's a very negative review for what seem to be good reasons, and at the time Beall seemed to be at least potentially positive about OA itself, based on the first sentence of this extract:

The Open Access model is a good one, for it makes research freely available to everyone. However, Bentham Open is exploiting the good will of those who established the Open Access model by twisting it and exploiting it for profit. Just because a journal is Open Access doesn't make it legitimate or high quality.

I don't imagine there are many lower quality folks who would argue with that last sentence, which

citesandinsights.info/civ1414.pdf

<http://www.michaeliseisen.org/blog/?p=1500>

Walt at Random

The library voice of the radical middle.

<http://walt.lishost.org/2016/01/trust-me-the-other-problem-with-87-of-bealls-lists/>

« Gold Open Access Journals 2011-2015: A SPARC Project

Not quite gone: a short catchall post »

“Trust Me”: The Other Problem with 87% of Beall's Lists

Here's the real thing: I could only find any discussion at all in Beall's blog for 230 of the 1,834 journals and publishers in his 2016 lists—and those cases don't include even 2% of the journals in DOAJ.

Now for the shorter version...

As long-time readers will know, I don't much like blacklists. I admit to that prejudice: I don't think blacklists are good ways to solve problems.

And yet, when I first took a hard look at Jeffrey Beall's lists in 2014, I was mostly assessing whether the lists represented as massive a problem as Beall seemed to assert. As you may know, I concluded that they did not.

But there's a deeper problem—one that I believe applies whether you dislike blacklists or mourn the passing of the *Index Librorum Prohibitorum*. To wit, Beall's lists don't meet what I would regard as minimal standards for a blacklist even if you agree with all of his

UC
pator
gives
work
| my
tion,
nical
rong
nder
most
lore

Predatory?

QUALITY IS ABOUT EDITORIAL WORKFLOW NOT BUSINESS MODEL

(Springer 107 retractions for false review,
Elsevier 7 journals withdrawn – they were paid by
pharmaceutical firms)

NOT PEER-REVIEWED

"Preprint" is a venue for early communication or feedback before peer review. Data may be used for research purposes only. Learn more about preprints or browse peer-reviewed articles instead.

Ten myths around open scholarly publishing

Librarian review Science and Medical Education Science Policy

THEY ARE 2%-5%

The screenshot shows the 'Think, Check, Submit' website. At the top, it says 'Think, Check, Submit' and 'From Think, Check, Submit'. Below this is a video player showing three people. To the right of the video player, there is a list of questions: 'Are you submitting your research to a trusted journal?' and 'Is it the right journal for your work?'. Below these questions are three buttons: 'THINK' (red), 'CHECK' (orange), and 'SUBMIT' (green). At the bottom, it says 'Only if you can answer 'yes' to the questions on our check list'.

IF REVIEWS WERE PUBLIC...

AND, IN THE END, IF YOU WERE NOT SO PRESSED TO PUBLISH OR PERISH, WOULDN'T YOU BE MORE CAREFUL???

HOW TO MAKE YOUR RESEARCH OPEN ACCESS

FOR FREE AND LEGALLY

💡 Check
thinkchecksubmit.org
The DOAJ

Do you know a free Open Access journal?



Publish via 'gold' route



Most Open Access journals do not have publishing charges

💡 Check
Your library
Your university
Your funders

Do you have funding for Open Access?



Publish via 'gold' route



⚠️ Note: Some publishers impose embargo periods on post-print publication

💡 Check
sherpa.ac.uk/romeo

Can you publish the post-print?



Publish your post-print



Post-print: manuscript after it has been peer reviewed, but before type-setting by the publisher

💡 Check
sherpa.ac.uk/romeo

Can you publish the pre-print?

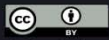


Publish your pre-print



Pre-print: manuscript that has not yet been subject to formal peer review, distributed to receive early feedback on research from peers

Choose a different journal



Jon Tennant and
Lisa Matthias

... in practice

THERE IS ALWAYS A WAY

“Open Access is too expensive.”



The «red road»

Gold road is not the «Open Choice»
offered by traditional commercial
publishers
(Elsevier, Springer, Wiley...)

paying a 3000 \$ APC,
your single article becomes Open Access,
but the journal is still by subscription

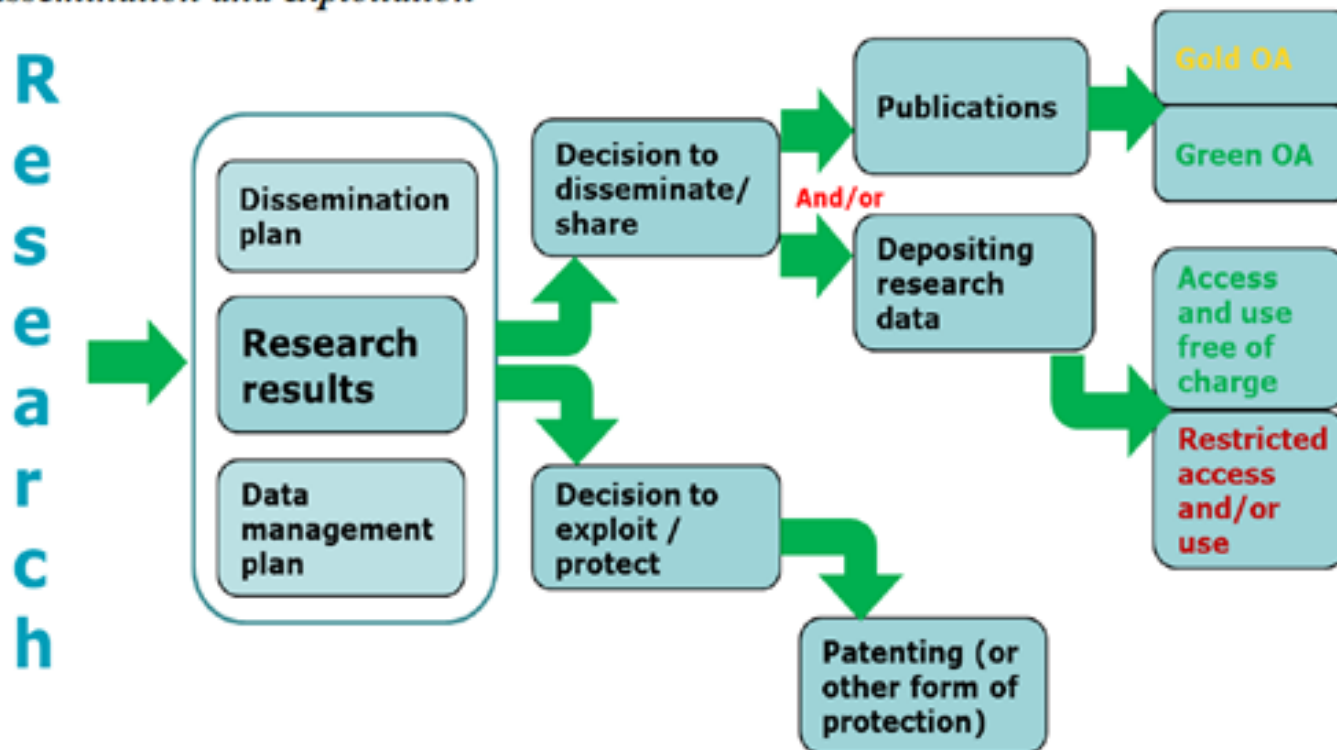
so, you pay twice...

you have to use it only if the funder asks
for a shorter embargo period
[in Horizon2020!]

[TO BE AVOIDED – YOU GET THE SAME RESULT
SELFARCHIVING FOR FREE]

HORIZON2020

Graph: Open access to scientific publication and research data in the wider context of dissemination and exploitation



Horizon2020



TEXT:

1. **DEPOSIT A MACHINE READABLE VERSION (PRESERVATION/TEXT MINING)**
ALWAYS – EVEN WHEN PUBLISHING OA
2. **GIVE ACCESS WITHIN 6/12 MONTHS [IF LONGER EMBARGO, PAY OPEN CHOICE – IT'S ELIGIBLE FOR REIMBOURSEMENT]**
3. **IN METADATA ALWAYS THE GRANT NUMBER**
4. **ASSOCIATE DATA WHENEVER POSSIBLE**

Dissemination costs

H2020 AGA — Annotated Model Grant Agreement: V4.1 – 26.10.2017

General

Annotated Model Grant Agreement 6.2.D3 (pag. 84)



D.3 Costs of other goods and services (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are eligible, if they are:

- (a) purchased specifically for the action and in accordance with Article 10.1.1 or

1. Costs of other goods and services (D.3): Types of costs — Form — Eligibility conditions — Calculation

The budget category applies to all RIA, IA and CSA grants under the General MGA.

The additional costs are covered by **Article 16** and other research infrastructure costs.

Open access — Costs related to open access to peer-reviewed scientific publications and research data are eligible, if the eligibility conditions are fulfilled. With explicit agreement by the Commission/Agency, it can also include fees levied for a membership scheme (if this is a requirement for publishing in open access or if membership is a pre-condition for significantly lower article processing charges).

1.1 What? This category covers the costs of dissemination of the action (or contributed in-kind against payment), including:

- costs for consumables and supplies (*e.g. raw materials etc.*)
- dissemination costs (including costs regarding open access to peer-reviewed scientific publications, *e.g. article processing or equivalent charges*, costs related to open access to research data and related costs, *such as data maintenance or storage and conference fees for presenting project-related research*)
- costs related to intellectual property rights (IPR) (including costs to protect the results or royalties paid for access rights needed to implement the action)

Dissemination costs

Annotated Model Grant
Agreement
Art. 29.2
(pag. 246)

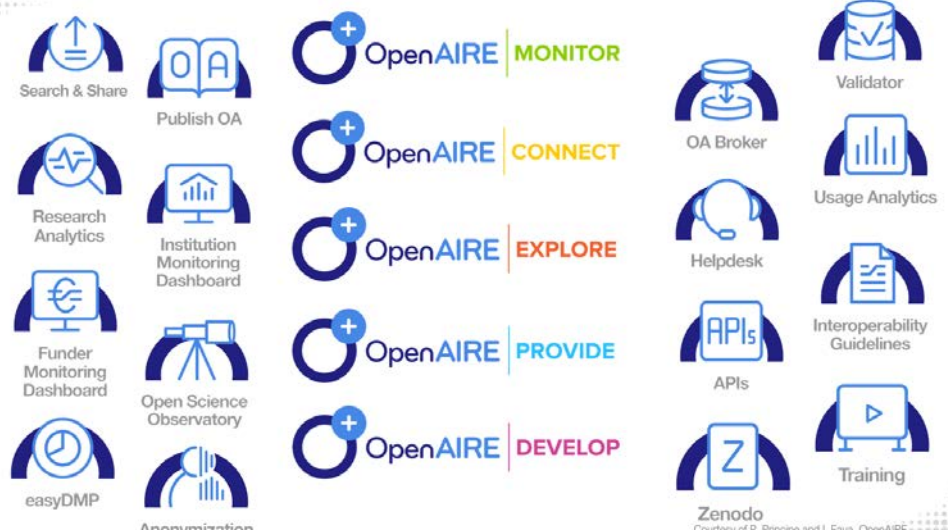


In gold open access, the payment of publication costs (article processing charges) is shifted from readers' subscriptions to (generally one-off) payments by the author. Such author processing costs may be eligible (see [Article 6.2.D.3](#)) — if incurred before the end of the action. Currently, an action for dealing with such costs incurred after the end of the action in FP7 is being piloted and further action in H2020 will be considered based on the outcome.

 Although gold open access already ensures access via the publisher, the beneficiaries must nevertheless also provide open access via a repository (see *below*) in order to ensure the long-term preservation and availability of the publication.

**COSTS MUST INCURR
BEFORE THE END OF
THE PROJECT**

...and you have OpenAIRE



National Open Access Desks (NOADs)

Facts

- Research is global, support is local
- Diversity in culture & maturity of national/local infras
- Not one size fits all in OA and open science



Our pan-European network

→ 34 countries

Key national organizations

→ 4 area coordinators

Moving to OS

Linked to infras around the world

What we support



OS policies



- Harmonization for policy makers
- Training
- Support

Infrastructure



- Interoperability
- Setup
- Connectivity
- Repositories

Open Research Data



- FAIR
- Open data
- Tools
- Legal
- Compliance

OA to publications



- Guides
- Tools/repositories
- Licenses
- Compliance

...let's talk about data

HOW MANY OF YOU WERE IN
TROUBLE IF DROPBOX BECAME
FOR FEE?

HOW MANY OF YOU WERE IN
TROUBLE IF THE PC BROKE
DOWN NOW?

HOW MANY OF YOU WERE IN
TROUBLE IF THE USB STICK GOT
LOST?

HOW MANY OF YOU WERE IN
TROUBLE IF THE FILES ON G-
DRIVE DISAPPEARED?



COPYRIGHT Italia Avv. Simone Aliprandi, Ph.D. - Copyright-Italia.it / Array Law Firm
www.copyright-italia.it - www.aliprandi.org - www.array.eu **ARRAY**

il backup: definizione (meno seria)

DEFINITION OF «BACKUP»:
WHAT YOU HAD TO DO
BEFORE

Why should you take care of your data?



... this is the data steward's nightmare:

- no backup
- no software
- - no data legend

... data are fragile

Scientists losing data at a rapid rate

Decline can mean 80% of data are unavailable after 20 years.

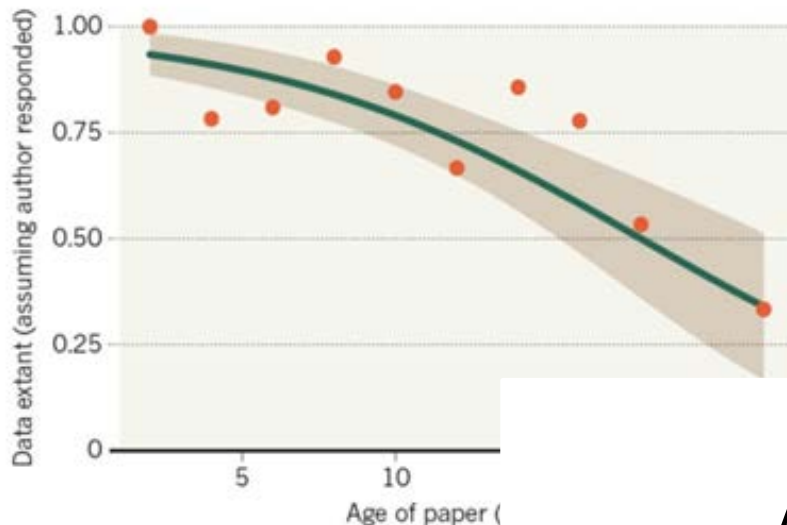
Elizabeth Gibney & Richard Van Noorden

19 December 2013

[Rights & Permissions](#)

MISSING DATA

As research articles age, the odds of their raw data being extant drop dramatically.



80% will be lost
in 20 years

...THAT'S WHY YOU NEED
A DATA MANAGEMENT PLAN.
IT'S NOT JUST AN ADMINISTRATIVE BURDEN



WHERE DO YOU STORE YOUR DATA?

Why should you take care of your data / 2

Jisc

How and why you should manage your research data: a guide for researchers

An introduction to engaging with research data management processes. [JISC Guide](#)

NOT TO LOOSE THEM

SOME KIND OF DATA ARE «UNIQUE EVENTS»
(metereology, sismology...)

WHEN DATA ARE ORGANIZED, YOUR RESEARCH IS MORE EFFICIENT

(IF OPEN) TO BE MORE VISIBLE

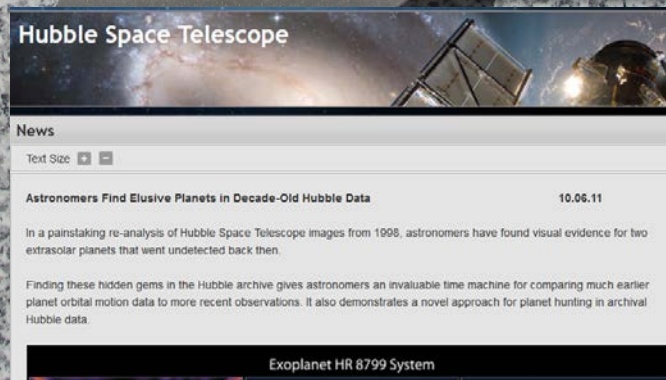
TO ALLOW FOR CHECKS AND VALIDATION

(IF OPEN) TO BOOSTER COLLABORATIONS

TO IMPROVE RESEARCH INTEGRITY

(IF OPEN) TO ALLOW REUSE

TO BE REPRODUCIBLE

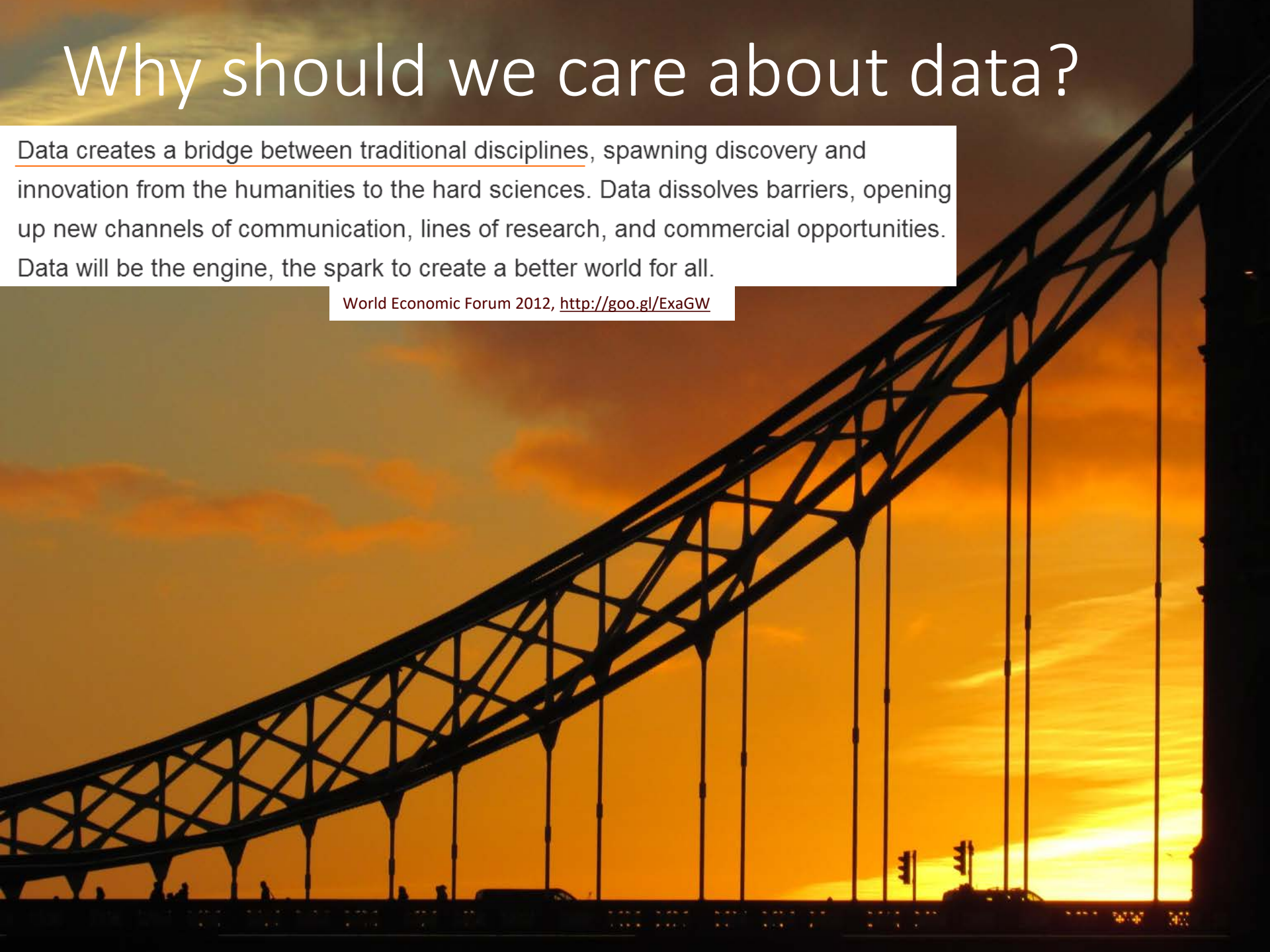


«the coolest thing to do with your data will be thought of by someone else» [R.Pollock]

Why should we care about data?

Data creates a bridge between traditional disciplines, spawning discovery and innovation from the humanities to the hard sciences. Data dissolves barriers, opening up new channels of communication, lines of research, and commercial opportunities. Data will be the engine, the spark to create a better world for all.

World Economic Forum 2012, <http://goo.gl/ExaGW>



Why should we care about data?

The Vienna Declaration on the European Open Science Cloud

Vienna, 23 November 2018

e 2 0
u 1 8
- a t

BECAUSE EOSC IS HERE TO STAY

Vienna, Nov.23, 2018

We, Ministers, delegates and other participants attending the launch event of the European Open Science Cloud (EOSC):

- 1. Recall** the challenges of data driven research in pursuing excellent science as stated in the “EOSC Declaration” signed in Brussels on 10 July 2017.
- 2. Reaffirm** the potential of the European Open Science Cloud to transform the research landscape in Europe. Confirm that the vision of the European Open Science Cloud is that of a research data commons, inclusive of all disciplines and Member States, sustainable in the long-term.
- 3. Recognise** that the implementation of the European Open Science Cloud is a process, not a project, by its nature iterative and based on constant learning and mutual alignment. Highlight the need for continuous dialogue to build trust and consensus among scientists, researchers, funders, users and service providers.
- 4. Highlight** that Europe is well placed to take a global leadership position in the development and application of cloud services for Science. Reaching out over time to and open to the world,
- 5. Recall** that the Council

SEAMLESS ACCESS TO OPEN BY DEFAULT FAIR DATA

9. Call for the European Open Science Cloud to provide all researchers in Europe with seamless access to an open-by-default, efficient and cross-disciplinary environment for storing, accessing, reusing and processing research data supported by FAIR data principles.

10. Note that the 2018 EOSC Summit (held on 11 June 2018) called for acceleration towards making the European Open Science Cloud a reality, hinting at the need to further strengthen the ongoing dialogue across institutions and with stakeholders, for a new governance framework to be launched in Vienna, on 23 November 2018.

[EOSC – what?]



THE EUROPEAN OPEN SCIENCE CLOUD? SOME NUANCES AND DEFINITIONS

Imagine a federated, globally accessible environment where researchers, innovators, companies and citizens can publish, find and re-use each other's data and tools for research, innovation and educational purposes. Imagine that this all operates under well-defined and trusted conditions, supported by a sustainable and just value for money model. This is the environment that must be fostered in Europe and beyond to ensure that European research and innovation contributes in full to knowledge creation, meet global challenges and fuel economic prosperity in Europe. This we believe encapsulates the concept of the European Open Science Cloud (EOSC), and indeed such a federated European endeavour might be expressed as the European contribution to an Internet of FAIR Data and services.

The European Open Science Cloud is a supporting environment for Open Science and not an 'open Cloud' for science.

The EOSC aims to accelerate the transition to more effective Open Science and Open Innovation in a Digital Single Market by removing the technical, legislative and human barriers to the re-use of research data and tools, and by supporting access to services, systems and the flow of data across disciplinary, social and geographical borders. The term European Open Science Cloud requires some reflection to dispel incorrect associations and clarify boundaries; in fact the term 'cloud' is a metaphor to help convey the idea of seamlessness and a commons.

[EOSC is not only based on data, but also on data stewardship]



The number of people with these skills needed to effectively operate the EOSC is, we estimate, likely exceeding half a million within a decade. As we further argue below, we believe that the implementation of the EOSC needs to include instruments to help train, retain and recognise this expertise, in order to support the 1.7 million scientists and over 70 million people working in innovation⁹. The success of the EOSC depends upon it.

Open Working

An Experiment in Open Working from 4TU.Centre for Research Data & TU Delft Research Data Services (Note! This is a test)

[HOME](#) [ABOUT OPEN WORKING AT TU DELFT](#) [DRAFT DATA MANAGEMENT PLAN CATALOGUE](#) [DATA STEWARDSHIP](#) [CONTACT](#)

FEBRUARY 23, 2018

We are hiring (again!) – Data Steward position at TU Delft

WE ARE HIRING



e-Infrastructure Reflection Group

◀ We need 500.000 respected data stewards to operate the European Open Science Cloud

04/05/16 09:08

At the e-IRG workshop in Amsterdam, we had the opportunity to talk to Barend Mons who is chairing the High Level Expert Group on the European Open Science Cloud, an advisory group to the European Commission. To be successful, the European Science Cloud needs a lot of experts to operate it, Barend Mons told us. Data stewards that have a lot of knowledge about managing and maintaining data. Experts who are well respected with a solid career path. Barend Mons also discussed several other findings of the Expert group, whose



Why should we care about data

DATA AB INITIO

MAKING RESEARCH DATA RIGHT, FROM THE START

K.Birney, 2015

and think, "surely I've covered this one my blog?"
up when I wrote [December's Exit Strategy post](#).

and, as you don't want to be stuck with
a non-representative set of data, it's a good format.



Il debito pubblico deprime la crescita? Il clamoroso errore di Carmen Reinhart e Kenneth Rogoff

2013

Publicato da keynesblog il 18 aprile 2013 in consigliati, Economia, ibt, Teoria economica



Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Thomas Herndon*

Michael Ash

Robert Pollin

April 15, 2013

Herndon, 2013

We replicate Reinhart and Rogoff (2010a and 2010b) and find that **coding errors**, **selective exclusion** of available data, and **unconventional weighting** of summary statistics **lead to serious errors** that inaccurately represent the relationship between public debt

and GDP growth among 20 advanced economies in the post-war period. Our finding is that when properly calculated, the average real GDP growth rate for countries carrying a public-debt-to-GDP ratio of over 90 percent is actually 2.2 percent, not -0.1 percent as published in Reinhart and Rogoff. That is, contrary to RR, average GDP growth at public debt/GDP ratios over 90 percent is not dramatically different than when debt/GDP ratios are lower.

We also show how the relationship between public debt and GDP growth varies significantly by time period and country. Overall, the evidence we review contradicts Reinhart and Rogoff's claim to have identified an important stylized fact, that public debt loads greater than 90 percent of GDP consistently reduce GDP growth.

Why should we care about data

1995 WaveLab and Reproducible Research

Jonathan B. Buckheit and David L. Donoho

Stanford University, Stanford CA 94305, US

*An article about computational science in a scientific publication is **not** the scholarship itself, it is merely **advertising** of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

A PAPER WITHOUT
DATA OR SOFTWARE IS
MERELY ADVERTISING



No data?

Is withholding your data simply bad science, or should it fall under scientific misconduct?



A recent study sent data requests to 200 stated 'data available upon request'. Most community think about those withholdin misconduct? **Nicole Janz** argues that if professional standards in research, and Classifying data secrecy as misconduct



Alastair Dunning

@alastairdunning

Following

To me, data are like footnotes. I might not always read them, but I get suspicious if they are not there.

Traduci dalla lingua originale: inglese

12:49 - 27 feb 2018

<https://twitter.com/alastairdunning/status/968453078218395648>

2 Retweet 8 Mi piace



Jon Tennant

@Protohedgehog

Following

My first talk of the year! Message is going to be that the opposite of 'open science' isn't 'closed science' - it's bad science.

**Gold Standard
Research Integrity**

**Questionable Research
Practices**

**Scientific
Misconduct**



Data secrecy

Open data
Open code
Pre-registration
Version control

P-hacking
Sloppy statistics
Peer review abuse
Inappropriate research design
Not answering to replicators
Lying about authorships

Fabrication
Falsification
Plagiarism

Glossary



5 WAYS TO THINK OF DATA :

- THE WAY DATA ARE COLLECTED
- THEIR FORM
- THEIR FORMAT
- THEIR SIZE/VOLUME
- THE WORKFLOW PHASE THEY ARE IN

THEY COULD
REQUIRE A
DIFFERENT
APPROACH/
TREATMENT

▣ The way the data is collected.

- ▣ By experimenting, simulations, observations, derived data, reference data.

▣ The data forms.

- ▣ For example text documents, spreadsheets, lab journals, logs, questionnaires, software code, transcripts, code books, audio and video recordings, photos, samples, slides, artefacts, models, scripts, databases, metadata, etc.

▣ The formats for electronic storage of the research data.

▣ The size (volume) of the data files.

▣ The *research lifecycle* phase the data is in.

3 pillars

 **DCC** because good research needs good data

Home Digital curation About us News Events Resources Training Projects Com

Home > Resources

<http://www.dcc.ac.uk/resources>

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
- Publications and presentations
- Roles
- Curation journals
- Informatics research
- External resources
- Online Store

Resources for digital curators

With just a few clicks, you can access any of the data curation resources collected by the DCC since we opened our doors in 2004.

Our comprehensive and easily accessible digital library is completely free to use and aims to provide you with everything you need to evaluate and implement those digital curation techniques most suited to your particular research project.

What's more, we regularly review and update our library, adding the latest resources developed for data curators.

Select from the links below and left to access high-level digital curation briefing papers; legal watch, standards watch and technology watch papers; case studies and interviews; and instalments from our detailed Curation Reference Manual.

International Journal of Digital Curation

Our highly regarded digital journal, published twice a year, features general articles and peer reviewed papers and serves as an invaluable channel for the

 **ands**
AUSTRALIAN NATIONAL DATA SERVICE

About ANDS | Events | Contact Us

<https://www.ands.org.au/>

About us News and Events Partners and Communities Working with data Online Services Guides and resources

Findable Accessible Interoperable Reusable

Is your data FAIR?

Find out how the FAIR principles can help you maximise the value of data

 **DANS**

Netherlands Contact Search this website

<https://dans.knaw.nl/en>

HOME DEPOSIT SEARCH TRAINING AND CONSULTANCY PROJECTS ABOUT DANS NEWS AND EVENTS

Welcome at DANS: the Netherlands institute for permanent access to digital research resources.

What can we do for you?

ARCHIVING

Deposit your datasets in DataverseNL or EASY or send research data for NARCIS.

DEPOSIT

REUSE

Find datasets, publications, researchers, projects and institutions via NARCIS and EASY.

SEARCH

TRAINING & CONSULTANCY

Let DANS advise you on data management and certification of digital archives.

ADVICE FROM DANS

SPOTLIGHT
DRYAD and DANS partner for long-term preservation research data

 **DRYAD**

Dryad and DANS announce a new collaboration to ensure long-term preservation and accessibility to curated scientific data. Over 50,000 researchers who have already deposited research data with Dryad can count on continuous open access to their data packages with an extra layer of security and recoverability as a result of this partnership.

NEWS
CoreTrustSeal certification launched

The Data Seal of Approval (DSA) and ICSU World Data System (WDS) announce the launch of a new certification organization: CoreTrustSeal.

DANS Retweeted
Marta Terepek
Nice demo by @pkdoom
@DANSKNANW - tool to help historians decide which @re3data repository to use to archive their datasets: ddrs-dev.danah.eu #dccc18

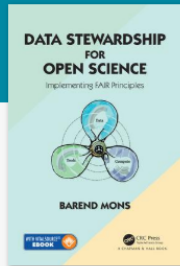
Feb 20, 2018

... and a master



Taylor & Francis Group
an informa business

<https://www.taylorfrancis.com/books/9781498753180>



Data Stewardship for Open Science Implementing FAIR Principles

the worst way imaginable to communicate the outcome of the scientific process. If science has become indeed data driven and *data is the oil of the 21st century*, we better put data centre stage and publish data as first-class research objects, obviously with supplementary narrative where needed, steward them throughout their life cycle, and make them available in easily reusable format.

Yet another recent study claimed that only about 12% of NIH funded data finds its way to a trusted and findable repository. Philip Bourne, when associate director for data science at the U.S.A. National Institutes of Health coined the term dark data for the 88% that is lost in amateur repositories or on laptops. When we combine the results of the general reproducibility related papers and the findability studies,

GET ACCESS

PREVIEW PDF



Monsense and more... @barendmons · 2 h

Finally! Tomorrow the book goes to the printer: Data Stewardship for Open Science: Implementing FAIR Principles

Traduci dalla lingua originale: inglese



Data Stewardship for Open Science: Implementing ...
Data Stewardship for Open Science: Implementing FAIR Principles has been written with the intention of making scientists, funders, and innovators in all disciplines an...
crcpress.com



In conclusion to this paragraph, my statement in 2005: Text-mining? Why bury it first and then mine it again? [Mons, 2005] is still frighteningly relevant.

A good data steward publishes data with a supplementary article(Data(+)).

...fast track

Au Loup Garou Gourmand
La Maison des
100 Bières Bretonnes

escience
vidensportal <https://vidensportal.deic.dk/RDMelearn>

eScience Få styr på data Supercomputing Træningskurser Om os Podcasts

Fjern + Få styr på data + eLearning course about the importance of good research data management (RDM)

eLearning course about the importance of good research data management (RDM)

Within the framework of the Danish National Forum for Data Management, the Danish Universities have developed the eLearning course "Research Data Management".

90% of the world's data was created within the last two years

Take the course

Module 1: Introduction



Reference: Viachos, E., Larsen, A.V., Zürcher, S., Hansen, A.F. (2019). 'Introduction'. In: Holmstrand, K.F., den Boer, S.P.A., Viachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), *Research Data Management* (eLearning course). doi: 10.11581/dtu.00000048

Module 2: FAIR principles



Reference: Martínez-Lavanchy, P.M., Hüser, F.J., Buss, M.C.H., Andersen, J.J., Begtrup, J.W. (2019). 'FAIR Principles'. In: Holmstrand, K.F., den Boer, S.P.A., Viachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), *Research Data Management* (eLearning course). doi: 10.11581/dtu.00000049

Module 3: Data Management Plans



Reference: den Boer, S.P.A., Buss, M.C.H., Hüser, F.J., Smed, U. (2019). 'Data Management Plans'. In: Holmstrand, K.F., den Boer, S.P.A., Viachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), *Research Data Management* (eLearning course). doi: 10.11581/dtu.00000050

Club
Passion
PANTHER
TOUR FINISTERE
19-21 AOUT 2013

Cost



[DM costing tool](#)

[What's Zingtree?](#)

Data Management Costing Tool

Data Management costing tool

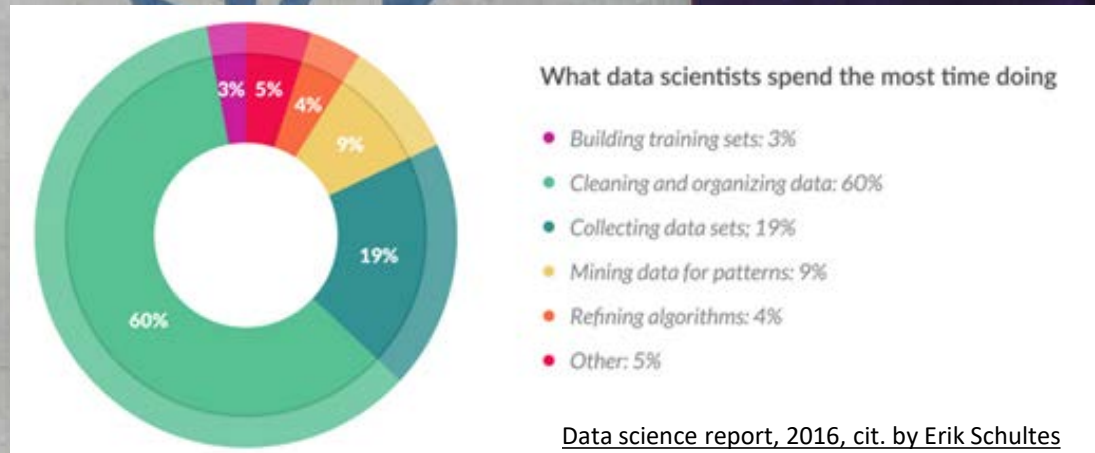


Welcome to the Data Management Costing Tool. This is for TU Delft researchers and staff to help determine costs and staffing requirements in project proposals. Let's start with some questions about your project which will help us estimate the data management needs of your project.



**THERE ARE COSTS. IT'S TRUE.
BUT HOW MUCH COULD IT COST **NOT TO** CURATE AND
PRESERVE DATA?**

Costs



Cost of not having FAIR research data

Cost-Benefit analysis for FAIR research data

Following this approach, we found that the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year. In addition, we also listed a number of consequences from not having FAIR which could not be reliably estimated, such as an impact on research quality, economic turnover, or machine readability of research data. By drawing a rough parallel with the European open data economy, we concluded that these unquantified elements could account for another €16bn annually on top of what we estimated. These results relied on a combination of desk research, interviews with the subject matter experts and our most conservative assumptions.

10,2 bn

16 bn

26,2 bn

Cost of not having FAIR data, 2018



...one step behind...

[the foundation]

Information Guide: Introduction to Ownership of Rights in Research Data. CREATE, University of Glasgow, 2018

Burrow, S., Margoni, T. and McCutcheon, V. (2018) Information Guide: Introduction to Ownership of Rights in Research Data. CREATE, University of Glasgow, 2018. Documentation. University of Glasgow. <http://eprints.gla.ac.uk/171314/>



Guides for Researchers

How do I know if my research data is protected?

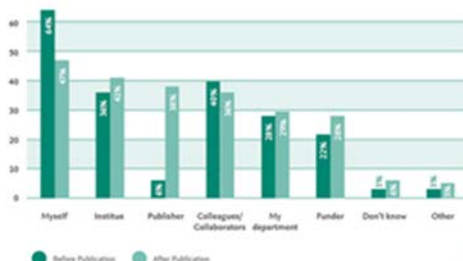
Learn more about what is research data and their protection by intellectual property rights

OpenAIRE

RESEARCH DATA IS NOT «MINE»
NO COPYRIGHT AS NO CREATIVITY
ON DATA PER SE

This time though it happened. What it was: 64% of researchers believe they own the data they generated for their research.

Figure 3. Research data ownership before and after publication (%; n=1162)



The result comes from a **solid piece of academic research** based on equally solid (open) data. The study and the report 'Open Data - the Researcher Perspective' were done by **CWTS / Leiden** and **Elsevier**. Credit giving, check.

Of course, the study reports other equally surprising results



Wainer Lusoli

@w_lusoli

Following

repeat with me: **#researchdata** is NOT mine. I was paid to get it, I'll get a **#nobel** 4 it, but it's NOT mine [linkedin.com/pulse/repeat-m ...](https://www.linkedin.com/pulse/repeat-m...) **#opendata**

Traduci dalla lingua originale: inglese



Repeat with me: research data is not mine

Seldom do I see something that truly shakes me at work. You know, work is work, I am no neurosurgeon, no médecin sans frontières nor am I a social

[linkedin.com](https://www.linkedin.com)

11:18 - 12 apr 2017

14 Retweet 18 Mi piace



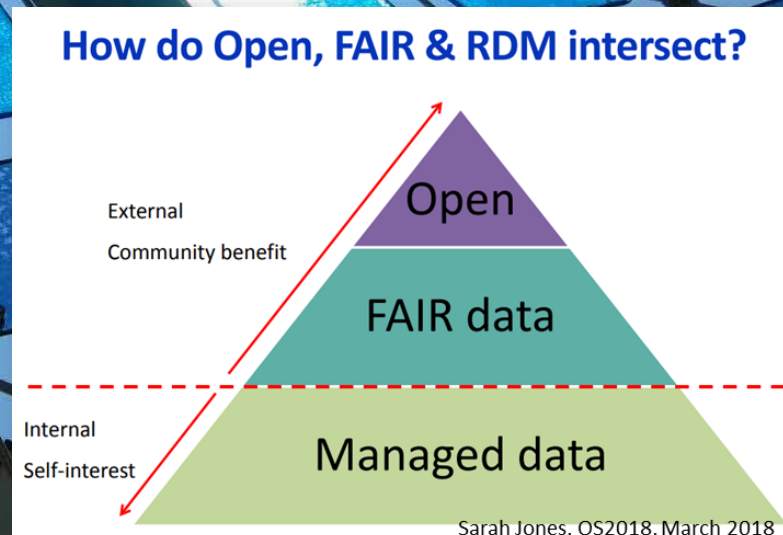
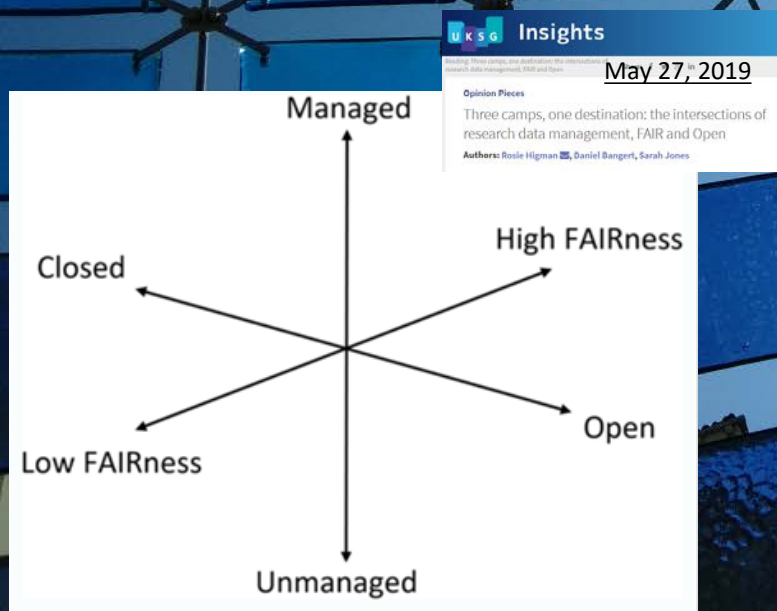
Lusoli, Apr.2017

[the 3 steps]

Open

FAIR

Managed



1. Data should be managed

Data management is an active process by which digital resources remain discoverable, accessible and intelligible over the longer term, a process that invests data and datasets with the potential to accrue value as assets enjoying far wider use than their creators may have anticipated. In the world of research, such a value-adding process is a significant contributor to the much desired achievement of impact.

ORGANIZATION
(file naming,
versioning...)

METADATA

BACKUP AND
STORAGE

LONG TERM
PRESERVATION

LEGAL ASPECTS

2. Data should be FAIR

TO BE FINDABLE:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

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
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

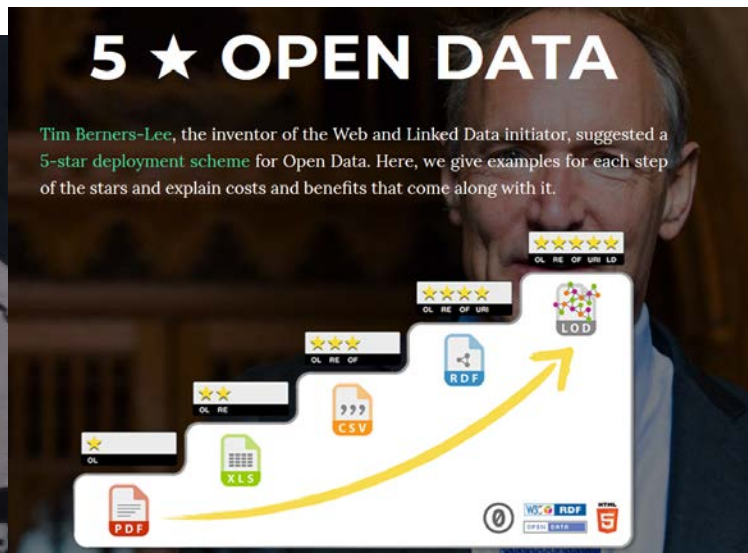
- R1. meta(data) have a plurality of accurate and relevant attributes.
- R1.1. (meta)data are released with a clear and accessible data usage license.
<https://www.force11.org/group/fairgroup/fairprinciples>
- R1.2. (meta)data are associated with their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

open
data
is about
MORE
THAN
DISCLOSURE
it must be
Fair

- Findable
- Accessible
- Interoperable
- Reusable

«ACCESIBLE»
DOES NOT MEAN «OPEN».
DATA CAN BE CLOSED,
PROVIDED YOU – AND
MACHINES - KNOW WHERE
TO FIND THEM AND AT WHAT
ACCESS CONDITIONS

3. Data COULD be Open



- ★ make your stuff available on the Web (whatever format) under an open license¹
- ★★ make it available as structured data (e.g., Excel instead of image scan of a table)²
- ★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)³
- ★★★★ use URIs to denote things, so that people can point at your stuff⁴
- ★★★★★ link your data to other data to provide context⁵

MANAGING DATA



|D|C|C

because good research needs good data

... from the beginning

Research Data Management: Get it right from the beginning

May 2018



Good RDM = Higher quality, efficiency and value for your research



Add a "version management" tab to your spreadsheet.

Now, let me expand on this idea.

Start by adding an extra "version management" tab to a new spreadsheet. In this sheet, carefully write down a version name (name of the file, typically) in the first column, in the second column the date, and in a third column an explanation of all changes you made to the sheet. Carefully fill out this sheet every single time you move something around, or tinker with the sheet.

If you're a starting PhD student, start doing this the very next time you build a new sheet. Thank me later.

If you already have multiheaded monstrous sheets: start by managing them in this way, and take a few extra hours to redefine the logic behind what you did earlier. Your dissertation writing self will thank you.



Expert guide

Data Management Expert Guide

- 1. Plan >
- 2. Organise & Document >
- 3. Process >
- 4. Store >
- 5. Protect >
- 6. Archive & Publish >
- 7. Discover >



Plan

In this introductory tour, you will become aware of what data management and a data management plan (DMP) are and why they are important. General concepts such as social science data and FAIR data will be explained. Based on our recommendations and good practice examples, you will be able to start writing your DMP.

Organise & Document

If you are looking for good practices in designing an appropriate data file structure, naming, documenting and organising your data files within suitable folder structures, this chapter is for you.

Process

Store

To be able to plan a storage and backup strategy, you will learn about different storage and backup solutions and their advantages and disadvantages. Also, measures to protect your data from unauthorised access with strong passwords and encryption will be explained.

Protect

This chapter highlights your legal and ethical obligations and shows how a combination of gaining consent, anonymising data, gaining clarity over who owns the copyright to your data and controlling access can enable the ethical and legal sharing of data.

Archive & Publish

When you arrive at this chapter you will have learnt to differentiate between currently available data publication services. You will also find a number of stepping stones on how to promote your data.

Discover

How can you discover and reuse existing or previously collected datasets?

Some [practical] support

AT THE END OF EACH STEP,
THERE IS A SECTION «ADAPT
YOUR DMP» ACCORDING TO
WHAT YOU HAVE JUST
LEARNT



⊕ Versioning

⊖ Interoperability

In order to be able to link your work to other research, it might be useful to build on established terminologies as well as commonly uses coding and soft- and hardware wherever this is possible.

- Which *software and hardware* will you use? How does this relate to other research?

If applicable:

- Will established *terminologies/ontologies* (i.e. structured controlled vocabularies) be used in the project? If not, how does yours relate to established ones?
- Which *coding* is used (if any)? How does this relate to other research?

Adapt your DMP: part 6

This is the sixth 'Adapt your DMP' section in this tour guide. To adapt your DMP, consider the following elements and corresponding questions:

⊖ Deposit your data

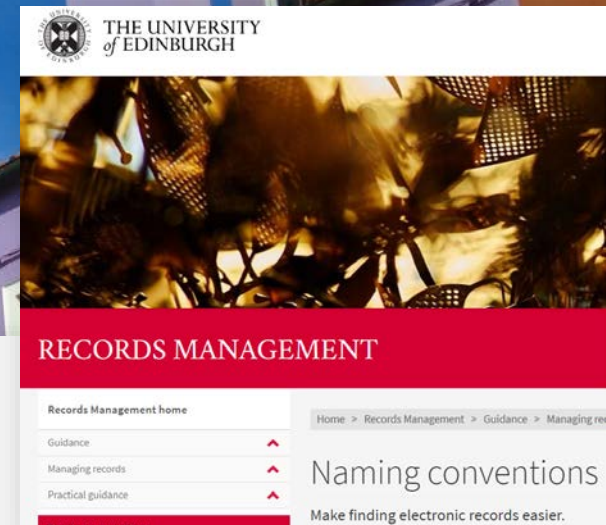
- Will the data you produce and/or used in the project be useable by third parties, in particular after the end of the project?
- Which data and associated metadata, documentation and code will be deposited?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included?
- Is it possible to include the relevant software (e.g. in open source code)?
- What data quality assurance processes will you apply?

Data management ABC – File naming

File naming conventions

The conventions comprise the following 13 rules. Follow the links for examples and explanations of the rules.

1. Keep file names short, but meaningful
2. Avoid unnecessary repetition and redundancy in file names and file paths.
3. Use capital letters to delimit words, not spaces or underscores
4. When including a number in a file name always give it as a two-digit number, i.e. 01-99, unless it is a year or another number with more than two digits.
5. If using a date in the file name always state the date 'back to front', and use four digit years, two digit months and two digit days: YYYYMMDD or YYYYMM or YYYY or YYYY-YYYY.
6. When including a personal name in a file name give the family name first followed by the initials.
7. Avoid using common words such as 'draft' or 'letter' at the start of file names, unless doing so will make it easier to retrieve the record.
8. Order the elements in a file name in the most appropriate way to retrieve the record.
9. The file names of records relating to recurring events should include the date and a description of the event, except where the inclusion of any of either of these elements would be incompatible with rule 2.
10. The file names of correspondence should include the name of the correspondent, an indication of the subject, the date of the correspondence and whether it is incoming or outgoing correspondence, except where the inclusion of any of these elements would be incompatible with rule 2.
11. The file name of an email attachment should include the name of the correspondent, an indication of the subject, the date of the correspondence, 'attach', and an indication of the number of attachments sent with the covering email, except where the inclusion of any of these elements would be incompatible with rule 2.
12. The version number of a record should be indicated in its file name by the inclusion of 'V' followed by the version number and, where applicable, 'Draft'.
13. Avoid using non-alphanumeric characters in file names.



Data management ABC – File naming / 2



Data versioning

Data versioning



What do we mean by the term 'data versioning'?

A version is “a particular form of something differing in certain respects from an earlier form or other forms of the same type of thing □”. In the research environment, we often think of versions as they pertain to resources such as manuscripts, software or data. We may regard a new version to be created when there is a change in the structure, contents, or condition of the resource.

In the case of research data, a new version of a dataset may be created when an existing dataset is reprocessed, corrected or appended with additional data. Versioning is one means by which to track changes associated with 'dynamic' data that is not static over time.

Why is data versioning important?

Increasingly, researchers are required to cite and identify to support research reproducibility and trustworthiness: accurately indicate exactly which version of a dataset particularly challenging where the data to be cited are accessed via a web service.

Numbering system 1

Data versioning follows a similar path to software versioning, usually applying a two-part numbering rule: Major.Minor (e.g. V2.1). Major data revision indicates a change in the formation and/or content of the dataset that may bring changes in scope, context or intended use. For example, a major revision may increase or decrease the statistical power of a collection, require change of data access interfaces, or enable or disable answering of more or less research questions. A Major revision may incorporate:

- substantial new data items added to /deleted from a collection
- data values changed because temporal and/or spatial baseline changes
- additional data attributes introduced
- changes in a data generation model
- format of data items a changed
- major changes in upstream datasets.

Minor revisions often involve quality improvement over existing data items. These changes may not affect the scope or intended use of initial collection. A Minor revision may include:

- renaming of data attribute
- correction of errors in existing data
- re-running a data generation model with adjustment of some parameters
- minor changes in upstream datasets.

Unlike the software domain, the data community doesn't yet have a standard numbering system. Three representative data version numbering patterns in use include:

Numbering system 1

Numbering system 2

Numbering system 3

What tools are available for data versioning?

There is no one-size-fit-all solution for data versioning and tracking changes. Data come in different forms and are managed by different tools and methods. In principle, data managers should take advantage of data management tools that support versioning and track changes.

Example approaches include:

Git (and Github) for Data □ (with size <10Mb or 100k rows) which allows:

- effective distributed collaboration – you can take my dataset, make changes, and share those back with me (and different people can do this at once)
- provenance tracking (i.e. what changes came from where)
- sharing of updates and synchronizing datasets in a simple, effective, way.

Data versioning at ArcGIS □

- Users of ArcGIS can create a geodatabase version, derived from an existing version. When you create a version, you specify its name, an optional description, and the level of access other users have to the version. As the owner of the version, you can change these properties or delete a version at any time.

Data management ABC – Versioning

University of Leicester

Good Practice and Guidance – Document Version Control Chart (Draft)

1. Create Document/File

- Save the document according to file naming guidance/good practice.

2. Document Identification

- Identify on the document e.g. in header or footer, the author, filename, page number and date the document is created/revised.

3. Version Control Table

- Versions and changes documented with Version Control Table where significant/formal/project based.

4. Version Number

- Current version number identified on the first page and where appropriate, incorporated into the header or footer of the document.
- Version number is included as part of the file name.

5. First Draft Version

- Named as version "0-1" (no full stops in electronic file names).
- Subsequent draft versions 0-2, 0-3, 0-4 ...

6. First Final/Approved Version

- When document is final/approved it becomes version 1-0.

7. Changes to Final Version

- Changed/revised final version becomes x-1.
- Subsequent drafts to Final version become e.g. 1-1, 1-2, 1-3 etc.

8. Further Final/Approved Documents

- Version number increased by "1-0" e.g. 1-0, 2-0, 3-0 etc.
- e.g. Amendments to Final 1-0 are 1-1, 1-2, 1-3 and as approved becomes 2-0.

Data management ABC – Versioning



Version control

Version control can be done through:

- Uniquely identifying different versions of files using a systematic naming convention, such as using version numbers or dates (date format should be YYYY-MM-DD, see '[File naming](#)');
 - Record the date within the file, for example, 20010911_Video_Twintowers;
 - Process the version numbering into the file name, for example, HealthTest-00-02 or HealthTest_v2;
 - **Don't** use ambiguous descriptions for the version you are working on. Who will know whether MyThesisFinal.doc, MyThesisLastOne.doc or another file is really the final version?
- Using version control facilities within the software you use;
- Using versioning software like [Subversion](#) (2017);
- Using file-sharing services with incorporated version control (but remember that using commercial cloud services as the Google cloud platform, Dropbox or iCloud comes with specific rules set by the provider of these services. Private companies have their own terms of use which applies for example to copyrights);
- Designing and using a version control table. In all cases, a file history table should be included within a file. In this file, you can keep track of versions and details of the changes which were made. Click on the tab to have a look at [an example which was taken from the UK Data Service](#) (2017c).

Data Management



Home Articles About

About Pensoft

Books

E-Books

Blog

Jo

Project Report

Research Ideas and Outcomes 4: e26439
<https://doi.org/10.3897/rio.4.e26439>

May 9, 2018

Reviewable

v1

Support Your Data: A Research Data Management Guide for Researchers

▼ John A Borghi, Stephen Abrams, Daniella Lowenberg, Stephanie Simms, John Chodacki

Abstract ▲

Researchers are faced with rapidly evolving expectations about how they should manage and share their data, code, and other research materials. To help them meet these expectations and generally manage and share their data more effectively, we are developing a suite of tools which we are currently referring to as "Support Your Data". These tools, which include a rubric designed to enable researchers to self-assess their current data management practices and a series of short guides which provide actionable information about how to advance practices as necessary or desired, are intended to be easily customizable to meet the needs of a researchers working in a variety of institutional and disciplinary contexts.

Suppl. material 5: Draft Guide - Preparing [doi](#)

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

Brief description: A draft guide that corresponds with the "Getting your data ready for analysis" row of the RDM rubric. Suggested points of customization are highlighted in yellow (discipline-specific) and red (institution-specific).

Filename: Draft Guide - Preparing.odt

[Download file](#) (59.52 kb)

Suppl. material 6: Draft Guide - Analyzing [doi](#)

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

Brief description: A draft guide that corresponds with the "Analyzing your data and handling the outputs" row of the RDM rubric. Suggested points of customization are highlighted in yellow (discipline-specific) and red (institution-specific).

Filename: Draft Guide - Analyzing.odt

[Download file](#) (51.82 kb)

Suppl. material 7: Draft Guide - Sharing [doi](#)

Authors: John Borghi

Data type: OpenDocument Text (.odt) file

Brief description: A draft guide that corresponds with the "Sharing and publishing your data" row of the RDM rubric. Suggested points of customization are highlighted in yellow (discipline-specific) and red (institution-specific).

	Ad Hoc	One-Time	Active and Informative	Optimized for Re-Use
Planning your project	When it comes to my data, I have a "way of doing things" but no standard or documented plans.	I create some formal plans about how I will manage my data at the start of a project, but I generally don't refer back to them.	I develop detailed plans about how I will manage my data that I actively revisit and revise over the course of a project.	I have created plans for managing my data that are designed to stream its future use by myself others.
Organizing your data	I don't follow a consistent approach for keeping my data organized, so it often takes time to find things.	I have an approach for organizing my data, but I only put it into action after my project is complete.	I have an approach for organizing my data that I implement prospectively, but it not necessarily standardized.	I organize my data so th others can navigate, understand, and use it without me being present.
Saving and backing up your data	I decide what data is important while I am working on it and typically save it in a single location.	I know what data needs to be saved and I back it up after I'm done working on it to reduce the risk of loss.	I have a system for regularly saving important data while I am working on it. I have multiple backups.	I save my data in a manner and location designed maximize opportunities for re-use by myself and others.
Getting your data ready for analysis	I don't have a standardized or well documented process for preparing my data for analysis.	I have thought about how I will need to prepare my data, but I handle each case in a different manner.	My process for preparing data is standardized and well documented.	I prepare my data in such a way as to facilitate use by both myself and others in the future.
Analyzing your data and handling the outputs	I often have to redo my analyses or examine their products to determine what procedures or parameters were applied.	After I finish my analysis, I document the specific parameters, procedures, and protocols applied.	I regularly document the specifics of both my analysis workflow and decision making process while I am analyzing my data.	I have ensured that the specifics of my analysis workflow and decision making process can be understood and put into action by others.

Data management ABC – Data entry

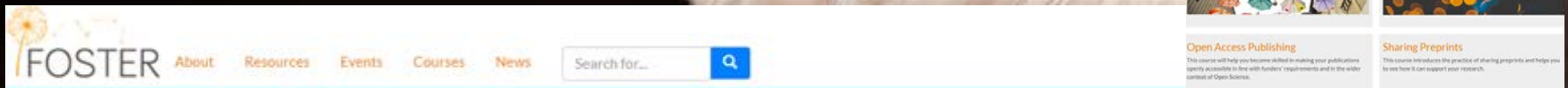
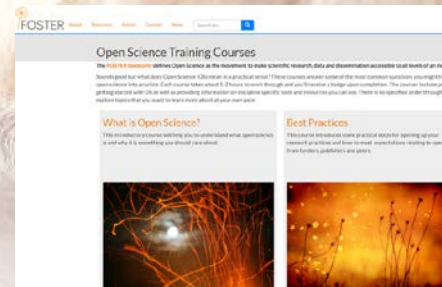


Data Management Expert Guide

- 1. Plan >
- 2. Organise & Document >
- 3. Process >
 - Data entry and integrity**
 - Quantitative coding
 - Qualitative coding
 - Weights of survey data
 - File formats and data conversion
 - Data authenticity
 - Wrap up: Data quality
 - Adapt your DMP: part 3
 - Sources and further reading
- 4. Store >
- 5. Protect >
- 6. Archive & Publish >

- ⊕ Check the completeness of records
- ⊕ Reduce burden at manual data entry
- ⊕ Minimise the number of steps
- ⊕ Conduct data entry twice
- ⊕ Perform in-depth checks for selected records
- ⊕ Perform logical and consistency checks
- ⊕ Automate checks whenever possible

Learn to manage



Managing and Sharing Research Data

Data-driven research is becoming increasingly common in a wide range of academic disciplines, from Archaeology to Zoology, and spanning Arts and Science subject areas alike. To support good research, we need to ensure that researchers have access to good data. Upon completing this course, you will:

- understand which data you can make open and which need to be protected
- know how to go about writing a data management plan
- understand the FAIR principles
- be able to select which data to keep and find an appropriate repository for them
- learn tips on how to get maximum impact from your research data

[Start the Free Course](#)

<https://www.fosteropenscience.eu/node/2328>



Full details

Level of knowledge: Introductory: no previous knowledge is required

Topics



Learn to protect

What are personal data?

Click the plus sign to expand the text box

- + What are personal data?
- + Protecting personal data
- + Legal requirements - EU General Data Protection Regulation (GDPR)
- + Legal requirements - GDPR research exemptions



Data Protection and Ethics

This course covers data protection in particular and ethics more generally. It will help you understand the basic principles of data protection and introduces techniques for implementing data protection in your research processes. Upon completing this course, you will know:

- what personal data are and how you can protect them
- what to consider when developing consent forms
- how to store your data securely
- how to anonymise your data

[Start the Free Course](#)



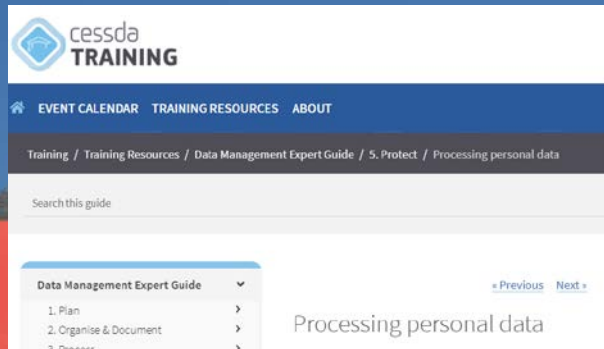
Full details

Level of knowledge: Introductory: no previous knowledge is required

Topics



[personal data]



I. Process lawfully, fair and transparent

The participant is informed of what will be done with the data and data processing should be done accordingly.

II. Keep to the original purpose

Data should be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.

III. Minimise data size

Personal data that are collected should be adequate, relevant and limited to what is necessary.

IV. Uphold accuracy

Personal data should be accurate and, where necessary kept up to date. Every reasonable step must be taken to ensure that personal data that are inaccurate are erased or rectified without delay.

V. Remove data which are not used

Personal data should be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.

VI. Ensure data integrity and confidentiality

Personal data are processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

[applicable laws]



Science Europe 2018

Privacy

- **Personal Data Protection Acts** are present in all European countries and concern general laws regulating the protection of personal data. They are based on European Directive 95/46/EC.⁸ This Directive will be replaced in the near future by the General Data Protection Regulation (GDPR),¹⁰ which all EU Member States will have to implement in their national legislation by May 2018.
- **Obligations to Report Data Leakage Acts** are additions to the Personal Data Protection Acts. They deal with the publication of personal data and contain sanctions in the form of penalties.
- **Medical Treatment Agreement Acts** regulate the use and preservation of personal (patient) data in and for medical research.
- **Scientific Medical Research with Humans Acts** regulate scientific research in the medical field, in particular how to handle personal health-related data. These make ethical reviews compulsory for all medical research projects.

Intellectual Property Rights

- **Copyright Acts** regulate the rights of the creator of a work. One distinguishes between exploitation rights and personal intellectual rights ('moral rights').
- The **Database Rights Act** recognises the investments made in creating and/or compiling a database. It is based on European Directive 96/9/EC.¹¹
- **Related Rights Acts** or **Neighbouring Rights Acts** mostly refer to the rights of performers, phonogram producers, and broadcasting organisations.
- **Patent Acts** are for the protection of patents. Publication of research results (including data) is restricted during the application stage of a patent.

Public data

- **Public Records Acts** (Public Archives Acts) oblige all public administration offices and services to preserve their documents and transfer these, after appraisal and selection, to public archives.
- **Public Sector Information Acts** (concerning re-usability of public data) are based on European Directive 2013/37/EU¹² that focuses on the economic aspects of the re-use of public information. It encourages Member States to make as much of this information as possible available for re-use. This also covers content held by museums, libraries, and archives, but does not apply

- **Freedom of information Acts** regulate and enable citizen access to documents held by public authorities or companies carrying out work for a public authority. They do not specifically deal with access to research data.
- **Heritage Acts** are relevant for archaeological research data in so far as that they regulate ownership of documentation (data) from archaeological excavations.
- **Statistical Information Acts** regulate the competencies of the statistics authorities in data gathering as well in access to data.
- **Land Registry Acts** (cadastral information) regulate the competencies of the national land registries and access to their data, with special provisions concerning personal data contained in their various databases.

Codes of Conduct/Ethical Issues

- **Codes of Conduct**, where these exist on a national level or in an institution, should be taken into account in DMPs. They contain the general principles of good academic teaching and research.
- **Codes of Practice** for the use of personal data in scientific and scholarly research are based on the Personal Data Protection Acts¹³ and prescribe how to handle personal data in research practice.
- **Codes of Conduct for Medical Research** regulate how researchers should handle medical personal data. They may be based on Medical Treatment Agreement Acts.

[Data and GDPR]

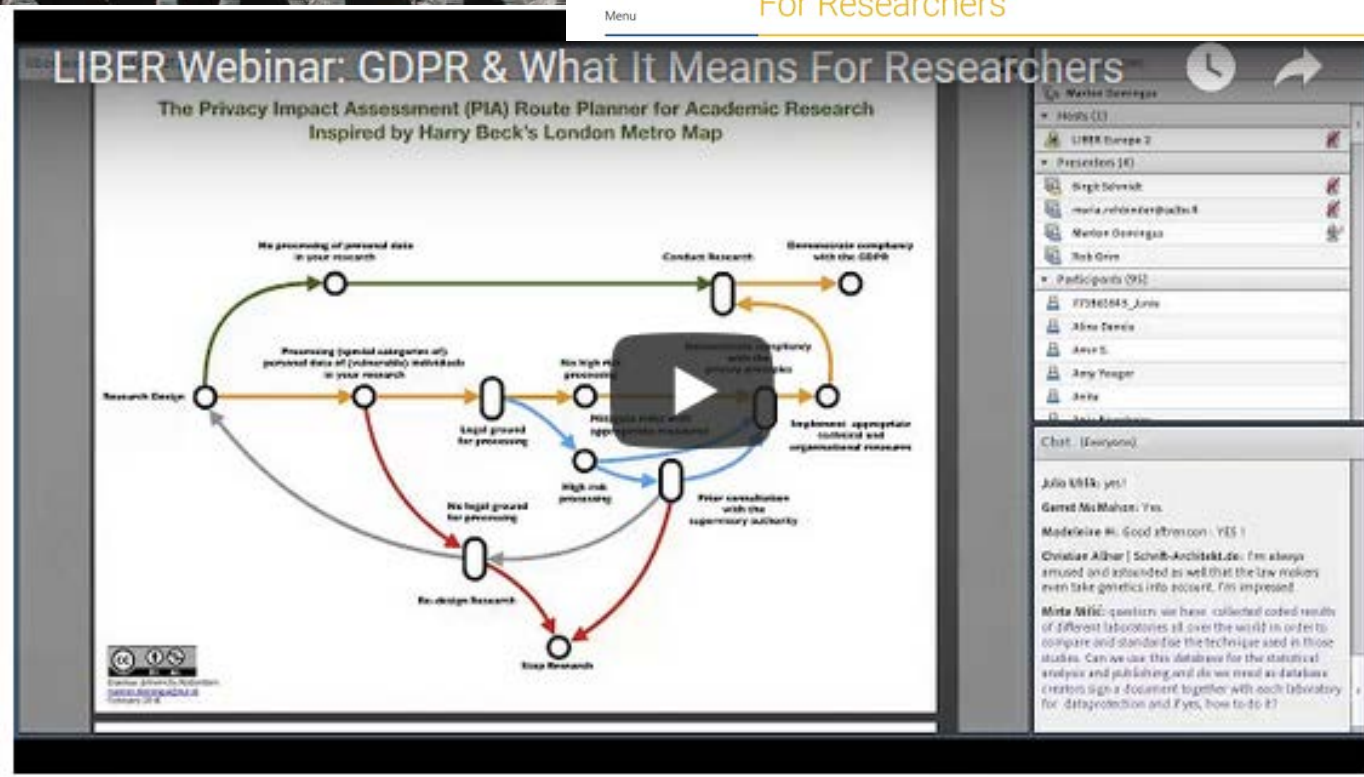
LIBER Ligue des Bibliothèques Européennes de Recherche
Association of European Research Libraries

About Us Network Strategy **May 22 2018** Login 🔍 ☰

Scholarly Communication Digital Skills & Services Research Infrastructure Events

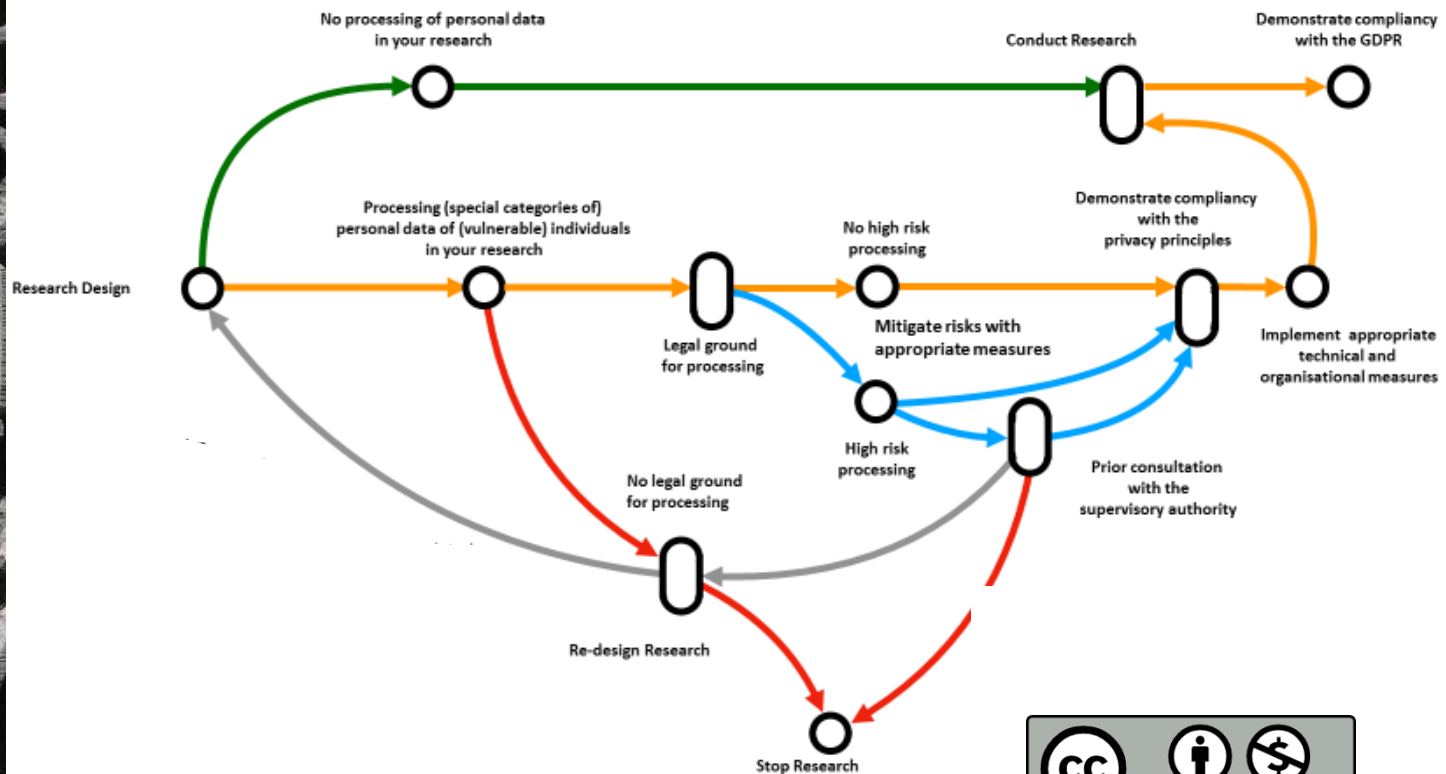
Menu

Webinar Video: GDPR & What It Means For Researchers



[Data and GDPR]

The Privacy Impact Assessment (PIA) Route Planner for Academic Research Inspired by Harry Beck's London Metro Map



Erasmus University Rotterdam
marlon.domingus@eur.nl
February 2018

The Logic of a Privacy Impact Assessment (PIA) for Academic Research

Q1. Do you process (special categories of) personal data of (vulnerable) individuals in your research?



YES

NO
Proceed - no measures required for safeguarding privacy.

"Personal Data" (GDPR*, Article 4):

Any information relating to an identified or identifiable natural person: a name, an identification number, location data, an online identifier, one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

"Special Categories of Personal Data (Sensitive Data)" (GDPR, Article 9):

Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

Action

Records of processing activities (GDPR*, Article 30):

The university shall maintain a digital record of the processing activities in your research to demonstrate compliancy to the GDPR. This register contains:

1. The name and contact details of the researcher, the research partners and service providers;
2. The purposes of the processing;
3. A description of the categories of data subjects and of the categories of personal data;
4. The categories of recipients to whom the personal data have been or will be disclosed;

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Online available at: <http://eur-lex.europa.eu/legal-content/EN/TEXT/PDF/?uri=CELEX:32016R0679&from=EN>

** Article 29 Data Protection Working Party: Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Adopted on 4 April 2017. As last Revised and Adopted on 4 October 2017. Online available at: https://ec.europa.eu/newsroom/document.cfm?doc_id=47711

Q2. What is the legal ground for this processing?

Lawfulness of Processing (GDPR*, Article 6, 89):

1. The individuals participating in your research have freely given their explicit consent for one or more specific purposes.
2. Your research contributes to a legitimate interest, yet results in no high risks for the individuals participating in the research.
3. Your research has a scientific, historical or statistical purpose, yet results in no high risks for the individuals participating in the research.

Action

Data protection by design and by default (GDPR*, Article 25):

Implement appropriate technical and organisational measures:

1. **Individual participating in your research (data subject).** Is the participant well informed, aware of possible risks for her/him and aware of the purpose of the research?
2. **Data.** Is the data de-identified and encrypted?
3. **Access Management.** How is access managed and controlled for the PI / team (expanded) / public?
4. **Software / Platform.** Are the *Terms of Service* for used software / platform checked (where is the data and who has access and has which usage rights)?
5. **Devices.** Are devices used safe? Encrypted drive, encrypted communication, strong password / two factor authentication.
6. **Partners.** Are the research partners / service partners trusted and are appropriate legal agreements made, with regards to roles, rights and responsibilities?
7. **Safe and secure collaboration.** Is the ((cross border) communication to, in and from the) collaboration platform end to end encrypted, are roles and permissions defined and implemented, is logging and monitoring implemented?

8. **Risk definition and mitigation.** Are risks defined and mitigated? Is a risk audit procedure started?

YES

NO
Stop research or redefine research.

Q3. Is this processing a high risk processing?

Criteria for high risk processing (WP29 - DPIA Guideline**):

1. Evaluation or scoring
2. Automated-decision making with legal or similar significant effect
3. Systematic monitoring
4. Sensitive data or data of a highly personal nature
5. Data processed on a large scale
6. Matching or combining datasets
7. Data concerning vulnerable data subjects
8. Innovative use or applying new technological or organisational solutions
9. When the processing itself prevents data subjects from exercising a right or using a service or a contract

YES

NO
Proceed - measures required for safe-guarding privacy.

Action

Prior consultation (GDPR*, Article 36):

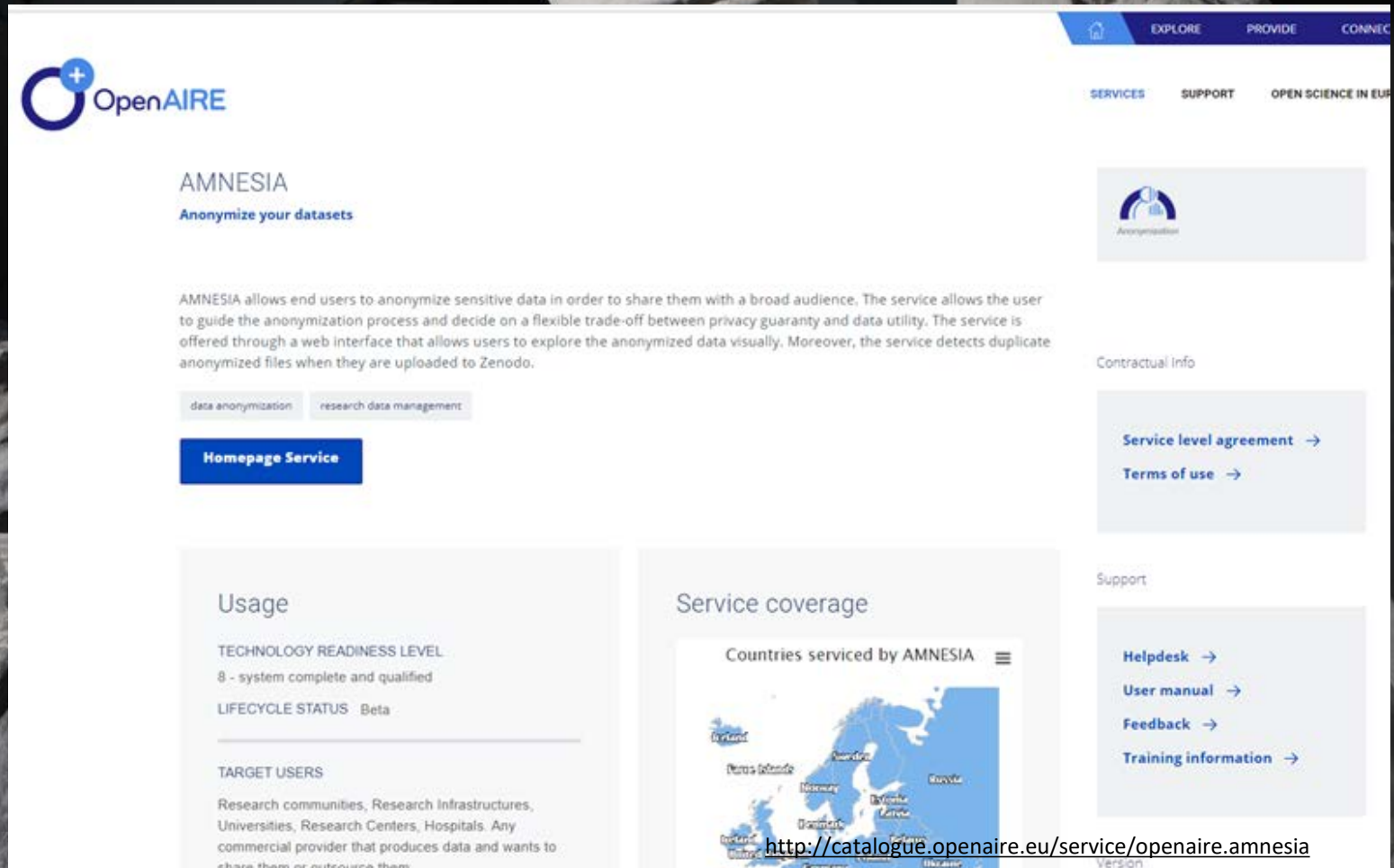
1. The Data Protection Officer shall, on behalf of the researcher, consult the supervisory authority, prior to the processing (the research) when the processing would result in a high risk *in the absence of measures* to mitigate the risk.

Action

Principles relating to processing of personal data (GDPR*, Article 5):

Demonstrate compliancy with the principles: lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, confidentiality and accountability.

[anonymizing data]



OpenAIRE

AMNESIA
Anonymize your datasets

AMNESIA allows end users to anonymize sensitive data in order to share them with a broad audience. The service allows the user to guide the anonymization process and decide on a flexible trade-off between privacy guaranty and data utility. The service is offered through a web interface that allows users to explore the anonymized data visually. Moreover, the service detects duplicate anonymized files when they are uploaded to Zenodo.

data anonymization research data management

Homepage Service

Usage

TECHNOLOGY READINESS LEVEL
8 - system complete and qualified

LIFECYCLE STATUS Beta

TARGET USERS

Research communities, Research Infrastructures, Universities, Research Centers, Hospitals. Any commercial provider that produces data and wants to share them or outsource them.

Service coverage

Countries serviced by AMNESIA



<http://catalogue.openaire.eu/service/openaire.amnesia>

Contractual Info

[Service level agreement →](#)
[Terms of use →](#)

Support

[Helpdesk →](#)
[User manual →](#)
[Feedback →](#)
[Training information →](#)

FAIRIFYING DATA



FAIR data

F indable

discoverable with machine readable metadata, identifiable and locatable by means of a standard identification mechanism

A ccessible

available and obtainable to both human and machine

I nteroperable

both syntactically parseable and semantically understandable, allowing data exchange and reuse among scientific disciplines, researchers, institutions, organisations and countries

R eusable

sufficiently described and shared with the least restrictive licences, allowing the widest reuse possible across scientific disciplines and borders, and the least cumbersome integration with other data sources

FAIR in a nuts

FAIR data training
Findable
Accessible
Interoperable
Reusable
FAIR for Developers
FAIR data self-assessment tool

f t in e s +SHARE

F1. (meta)data are assigned a globally unique and et

There are many resources created by the ARDC on the topic of metadata

- Metadata guide
- Data versioning

The ARDC has information on persistent identifiers on three different levels

- Persistent identifiers: awareness level
- Persistent identifiers: working level
- Persistent identifiers: expert level

It is also a provider of services for minting persistent identifiers of many d
of the data being identified):

- Digital Object Identifier (DOI) System for research data
- Handle minting Service (Identify My Data)
- International Geo Sample Numbers (IGSN)

Complementary to the assignment of persistent identifiers is their proper



FAIR data

Article type: Research Article

Authors: Mons, Barend^{a, b, c, *} | Neylon, Cameron^d | Velterop, Jan^e | Dumontier, Michel^f | da Silva Santos, Luiz Olavo Bonino^{b, g} | Wilkinson, Mark D.^h

4. ...and what FAIR is not

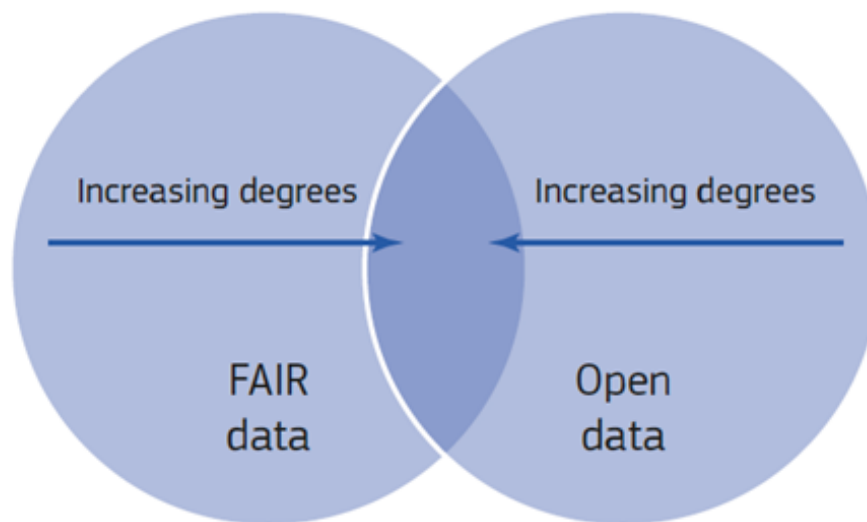
FAIR is not a standard: The FAIR guiding principles are sometimes incorrectly referred to as a 'standard', even though the original publication explicitly states they are not [25]. The guiding principles allow many different approaches to rendering data and services Findable, Accessible, Interoperable, to serve the ultimate goal: the reuse of valuable research objects. Standards are prescriptive, while guidelines are permissive. We suggest that a variety of valuable standards can and should be developed, each of which is guided by the FAIR Principles. FAIR simply describes the qualities or behaviours required of data resources to achieve – possibly incrementally – their optimal discovery and scholarly reuse.

FAIR is not equal to RDF, Linked Data, or the Semantic Web The reference article in Scientific Data [25] emphasises the machine-actionability of data and metadata. This implies (in fact, requires) that resources that wish to maximally fulfil the FAIR guidelines must utilise a widely-accepted machine-readable framework for data and knowledge

FAIR is not just about humans being able to find, access, reformat and finally reuse data: The official press release following the publication of the FAIR Principles states the authors' position clearly: "The recognition that computers must be capable of accessing a data publication autonomously, unaided by their human operators, is core to the FAIR Principles. Computers are now an inseparable companion in every research endeavour". In recent surveys, the time reportedly spent by PhD students and other researchers in projects dealing with discovering and reusing multiple data sources – so called 'data munging' – has been pegged at 80% [19]. Were these colleagues and their machine-assistants only having to deal with FAIR data and services, this wasted time would be reduced to a fraction of what it is today. The avoidance of time-wasting would be a first return on investment in good data stewardship. To serve this potentially enormous cost reduction, FAIR compliant (meta)data and services should be actionable by machines without human supervision whenever and wherever possible.

FAIR is not equal to Open: The 'A' in FAIR stands for 'Accessible under well defined conditions'. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles, although inspired by Open Science, explicitly and

FAIR / Open



Turning FAIR into reality, 2018

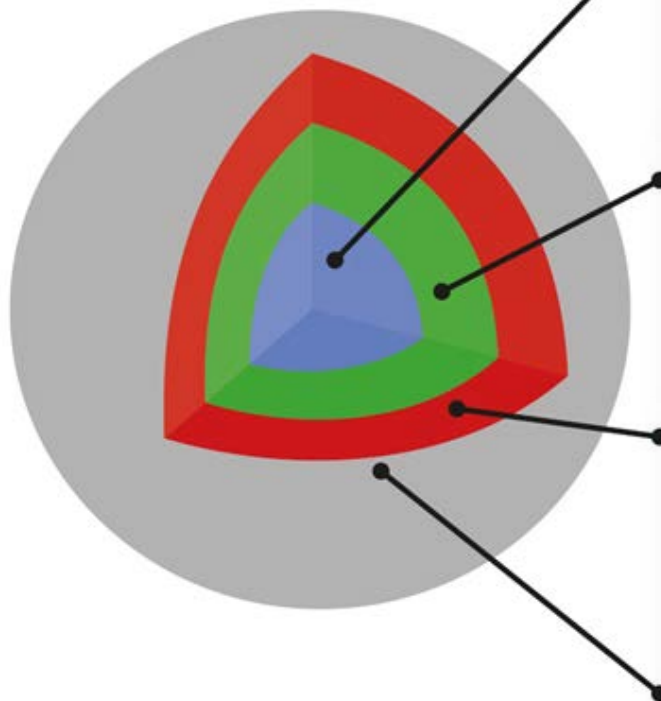
Figure 4. The relationship between FAIR and Open

Data can be FAIR or Open, both or neither. The greatest benefits come when data are both FAIR and Open, as the lack of restrictions supports the widest possible reuse, and reuse at scale. To maximise the benefits of making FAIR data a reality, and in the context of Open Science initiatives, the FAIR principles should be implemented in combination with a policy requirement that research data should be Open by default – that is, Open unless there is a good reason for restricting access or reuse. In recent European Commission formulations, the maxim ‘as open as possible, as closed as necessary’ has been introduced, which is a helpful articulation of the principles

Rec. 17: Align and harmonise FAIR and Open data policy

Policies should be aligned and consolidated to ensure that publicly-funded research data are made FAIR and Open, except for legitimate restrictions. The maxim ‘as Open as possible, as closed as necessary’ should be applied proportionately with genuine best efforts to share.

FAIR data - the ideal FAIR object



DIGITAL OBJECT

Data, code and other research outputs

At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Digital Objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and support citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code used to process and analyse the data.

METADATA

Contextual documentation

In order for Digital Objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.

FAIR: technology VS domain

FAIR Principles

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

Technology



Domain-relevant content

Findable:

F1 (meta)data are assigned a globally unique and persistent identifier;

F2 data are described with rich metadata;

F3 metadata clearly and explicitly include the identifier of the data it describes;

F4 (meta)data are registered or indexed in a searchable resource;

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;

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;

Reusable:

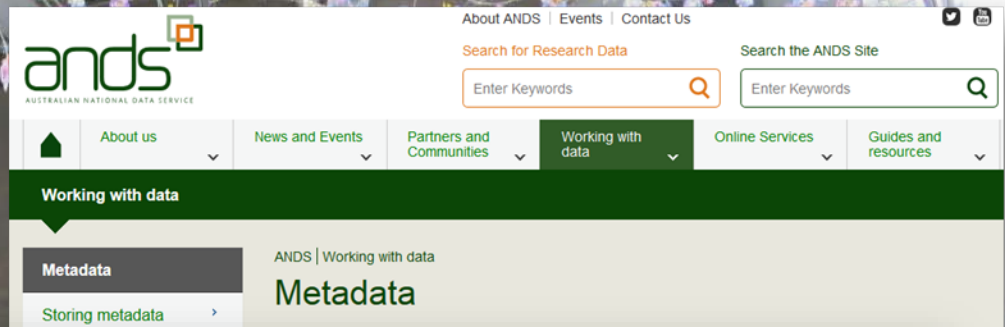
R1 meta(data) are richly described with a plurality of accurate and relevant attributes;

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;

F = Findable. Metadata



Types of metadata	Goal	Example
Descriptive metadata	<i>The minimal metadata, required to find a digital object.</i> <i>If there are additional contextual metadata, a user will have a better idea on how to use the data</i>	<i>Author, title, abstract, date</i> <i>Contextual metadata are for example location, time, data collection method (tools)</i>
Structural metadata	<i>These link the individual objects of a unity</i>	<i>Links to related digital objects, (e.g. the article written based on the linked research data)</i>
Technical metadata	<i>Information on the technical aspects of the data set</i>	<i>Data format, hardware/software used, calibration, version, authentication, encryption, metadata standard</i>
Administrative metadata	<i>Metadata focusing on user rights and management of digital objects</i>	<i>License, possible reasons for an embargo, waivers</i>

F = Findable. Metadata standards

Metadata

RDA | Metadata Directory

Edit this page

View the standards

View the extensions

View the tools

View the use cases

Browse by subject areas

Contribute

Add standards

Add extensions

Add tools

Add use cases

Arts and Humanities [Edit](#)

- [Archaeology](#) [Edit](#)
- [Creative art and design](#) [Edit](#)
- [Heritage Studies](#) [Edit](#)
- [Historical and Philosophical Studies](#) [Edit](#)
- [History by Area](#) [Edit](#)
- [History](#) [Edit](#)
- [Law](#) [Edit](#)
- [Music](#) [Edit](#)

Engineering [Edit](#)

- [Architecture](#) [Edit](#)
- [Building Conservation](#) [Edit](#)

Life Sciences [Edit](#)

- [Agricultural Economics](#) [Edit](#)
- [Agricultural Science](#) [Edit](#)
- [Animal pathology](#) [Edit](#)
- [Animal physiology](#) [Edit](#)
- [Biochemistry](#) [Edit](#)
- [Biodiversity](#) [Edit](#)
- [Bioengineering](#) [Edit](#)
- [Biogeography](#) [Edit](#)
- [Bioinformatics](#) [Edit](#)

Physical Sciences & Mathematics [Edit](#)

- [Astronomy](#) [Edit](#)
- [Astrophysics](#) [Edit](#)
- [Cartography](#) [Edit](#)
- [Chemistry](#) [Edit](#)
- [Climatology](#) [Edit](#)
- [Crystallography](#) [Edit](#)
- [Environmental Science](#) [Edit](#)
- [Geology](#) [Edit](#)
- [Geoscience](#) [Edit](#)
- [Glaciology](#) [Edit](#)
- [Hydrogeology](#) [Edit](#)
- [Hydrography](#) [Edit](#)
- [Hydrology](#) [Edit](#)
- [Marine Science](#) [Edit](#)
- [Maritime Geography](#) [Edit](#)
- [Materials Science](#) [Edit](#)
- [Meteorology](#) [Edit](#)
- [Minerology](#) [Edit](#)
- [Nuclear and Particle Physics](#) [Edit](#)
- [Oceanography](#) [Edit](#)
- [Palaeontology](#) [Edit](#)
- [Physics](#) [Edit](#)
- [Planetary science](#) [Edit](#)
- [Remote Sensing](#) [Edit](#)
- [Soil Science](#) [Edit](#)
- [Solar physics](#) [Edit](#)

Social and Behavioral Sciences [Edit](#)

- [Anthropology](#) [Edit](#)
- [Demography](#) [Edit](#)
- [Economics](#) [Edit](#)
- [Geography](#) [Edit](#)
- [Health Policy](#) [Edit](#)
- [Human and Social Geography](#) [Edit](#)
- [Planning \(Urban, Rural and Regional\)](#) [Edit](#)
- [Politics](#) [Edit](#)
- [Sociology](#) [Edit](#)

General Research Data [Edit](#)

- [Multi-disciplinary](#) [Edit](#)

F = findable. Metadata tools

What CEDAR does

<https://metadatacenter.org/>

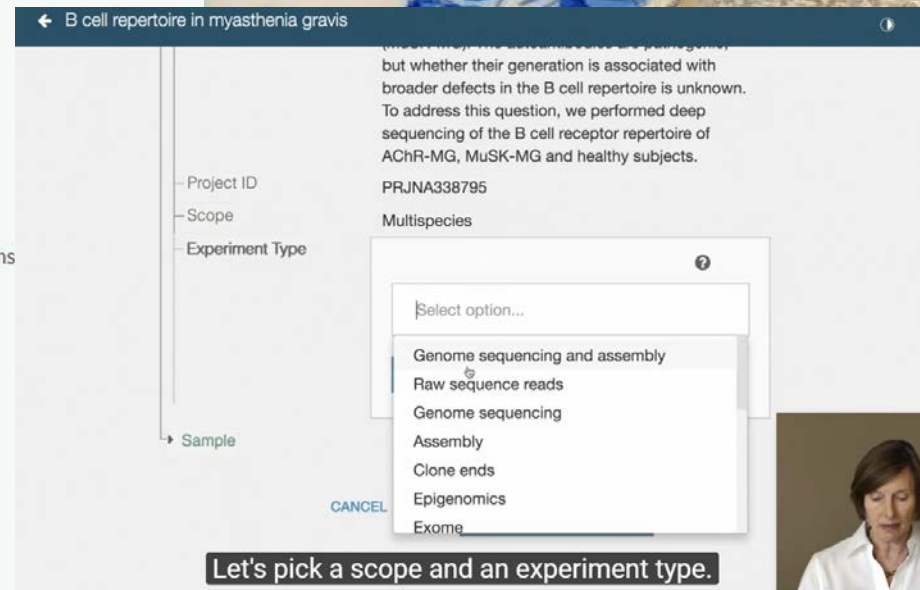
The CEDAR Workbench, as we refer to the suite of CEDAR tools, makes it easy to collect and use metadata. Eventually our tools will create a metadata record to its eventual processing, and even enhancement, by users and analysts. But for now, CEDAR tools help users collect, store, and download the information that users have provided.

What can CEDAR do for me already?

As of its production release, in February 2017, CEDAR addresses these scenarios:

- create user-friendly, shareable forms for collecting metadata, with features like
 - nested and repeatable elements and fields
 - reusable elements
 - control over tool tips, field titles, and field descriptions
- share your forms and metadata
 - provide a link to your metadata editors, so they can enter metadata responses based on your forms
 - share your forms and other content with individuals or a group
 - create and manage groups to make permissions simpler
- associate your questions (fields) and possible answers (values) with controlled terms
 - select any term or collection of terms from the NCBO BioPortal semantic repository
 - combine different terms from different controlled vocabularies into a single set of options
 - create your own terms, or term lists ('value sets') that can be re-used
- view responses meeting your (simple) search criteria, in several forms
 - CEDAR Metadata Editor's metadata view
 - an in-line JSON-LD format, used by CEDAR for all its metadata instances
 - download of JSON-LD files via the [CEDAR REST API](#), for offline integration with your workflow
- use the Workbench Desktop interface to manage your content
 - use My Workspace to see your items, or Shared with Me to see other items you can access
 - select an item and control-click or use the 3-dot menu in the upper right to share it, copy it, delete it, or get info on it
- enable intelligent metadata suggestions in your template by using a field's Suggestions tab
 - CEDAR keeps track of metadata entered for that field
 - users will see a drop down list of the most popular metadata entries, and can select from them
- remotely access CEDAR content and capabilities using the [CEDAR REST API](#)

With these capabilities, you can capture simple or rich metadata for your project, build a repository of project metadata, or design particular needs. Advanced users can even submit metadata entries through CEDAR's REST API.



Let's pick a scope and an experiment type.

F = Findable. Persistent identifiers

DataCite About us ▾ Services ▾ Resources ▾ Community ▾ Become a member

WELCOME TO DATACITE

the leading global provider of DOIs for research data.

ORCID Connecting Research and Researchers

FOR RESEARCHERS FOR ORGANIZATIONS ABOUT HELP SIGN IN

SIGN IN REGISTER FOR AN ORCID ID LEARN MORE

6,055,250 ORCID IDs and counting. See more learn more

We need your feedback! Please tell us about your understanding and perceptions of ORCID and your experience of using your ID by completing our [community survey](#). Thank you!

DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. [Find out more](#)

- 1 REGISTER** Get your unique ORCID identifier [Register now!](#)
Registration takes 30 seconds.
- 2 ADD YOUR INFO** Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn).

<https://orcid.org/>

LATEST NEWS
Tue, 26 Feb 2019
Construyendo una Infraestructura para Apoyar a los Investigadores - Una introducción

re3data.org
Find an appropriate repository to access and deposit research data with re3data.org

Generate your references automatically with our easy-to-use citation formatting tool.
<https://www.datacite.org/>

- ASSIGN A DOI
DIGITAL OBJECT IDENTIFIER
- USE ORCID

ORCID – ID. You are unique

CREATE AND MANTAIN YOUR
ORCID-ID.
IT MAKES THE WHEELS TURNING



F = Findable. Persistent identifiers

Here are some identifier schemes:

- **ARK (Archival Resource Key)** – a URL with extra features allowing you to ask for descriptive and archival metadata and to recognize certain kinds of relationships between identifiers. ARKs are used by memory organizations such as libraries, archives, and museums. They are resolved at "http://www.nt2.net". Resolution depends on HTTP redirection and can be managed through an API or a user interface.
- **DOI (Digital Object Identifier)** – an identifier that becomes actionable when embedded in a URL. DOIs are very popular in academic journal publishing. They are resolved at "http://dx.doi.org". Resolution depends on HTTP redirection and the Handle identifier protocol, and can be managed through an API or a user interface.
- **Handle** – an identifier that becomes actionable when embedded in a URL. Handles are resolved at "http://www.handle.net/". Resolution depends on HTTP redirection and the Handle protocol, and can be managed through an API or a user interface.
- **InChI (IUPAC International Chemical Identifier)** – a non-actionable identifier for chemical substances that can be used in printed and electronic data sources, thus enabling easier linking of diverse data compilations.
- **LSID (Life Sciences Identifier)** – a kind of URN that identifies a biologically significant resources, including species names, concepts, occurrences, and genes or proteins, or data objects that encode information about them. Like other URNs, it becomes actionable when embedded in a URL.
- **NCBI (National Center for Biotechnology Information) ACCESSION**
• a non-actionable number in use by NCBI.
- **PURL (Persistent Uniform Resource Locator)** – a URL that is always redirected through a hostname (often purl.org). Resolution depends on HTTP redirection and can be managed through an API or a user interface.
- **URL (Uniform Resource Locator)** – the typical "address" of web content. It is a kind of URI (Uniform Resource Identifier) that begins with "http://" and consists of a string of characters used to identify or name a resource on the Internet. Such identification enables interaction with representations of the resource over a network, typically the World Wide Web, using the HTTP protocol. Well-managed URL redirection can make URLs as persistent as any identifier. Resolution depends on HTTP redirection and can be managed through an API or a user interface.
- **URN (Uniform Resource Name)** – an identifier that becomes actionable when embedded in a URL. Resolution depends on HTTP redirection and the DDDS protocol, and can be managed through an API or a user interface. A browser plug-in can save you from typing a hostname in front of it. https://dmptool.org/general_guidance#persistent-identifiers

A = Accessible

ACCESSIBLE≠OPEN
«ACCESS» CAN ALSO BE
RESTRICTED OR EMBARGOED

- **Open access**

Data that can be accessed by any user whether they are registered or not. Data in this category shouldn't contain personal information (unless consent is given (see '[Informed consent](#)').

- **Access for registered users (safeguarded)**

Data that is accessible only to users who have registered with the archive. This data contains no direct identifiers but there may be a risk of disclosure through the linking of indirect identifiers.

- **Restricted access**

Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information.

Restricted access requires long-term commitment of the researcher or person responsible for the data to handle the upcoming the permission requests.

- **Embargo**

Besides offering the opportunity for restricted access 'for eternity' most data repositories allow you to place a temporary embargo on your data. During the embargo period, only the description of the dataset is published. The data themselves will become available in open access after a certain period of time.

HARVARD
Dataverse

We've changed hosting! Contact us if you see any issues.

Metrics 3,074,518 Downloads

Share, archive, and get credit for your data. Find and cite data across all research fields

Search this dataverse... Find Advanced Search

☒ Datasets (2,537)
 ☒ Datasets (75,771)
 ☐ Files (372,751)

Dataverse Category
Research Project (774)
Researcher (722)
Organization or Institution (738)

1 to 10 of 78,308 Results

Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multi-component, community-based intervention in Zimbabwe
Feb 27, 2018 - Replication Studies Dataverse

Yu, Fang; Hein, Nicholas; Bagenda, Danstan, 2018, "Preventing HIV and HSV-2 through improving knowledge and attitudes: a replication study of a multi-component, community-based intervention in Zimbabwe", doi:10.7910/DATASERVSE.V1

DRYAD About For researchers For organizations Contact us Log in Sign up

Publishers: Simplify data submission.
Strengthen links between articles and data.
For free.

Integrate your journal with Dryad today »

Submit data now
How and why?

Search for data
Enter keyword, author, title, DOI, etc. Go
Advanced search

Latest from @datadryad

Tweets by @datadryad

Dryad @datadryad
Excited to announce submission integration with JAMIA Open, a new gold OA title from @AMIAinformatics! academic.oup.com/jamiaopen #opendata

Browse for data
Recently published Popular

Recently published data

Young RJ (2017) Data from: The tonic immobility test: do wild and domesticated mantella frogs (Mantella aurantiaca) have the same response? PLOS ONE 12(10):1-10. doi:10.5061/dryad.v172f

zenodo Search Upload Communities

All versions Found 359191 results.

Access Right
☐ Open (338942)
☐ Closed (19395)
☐ Restricted (328)
☐ Embargoed (306)

File Type
☐ Png (128118)
☐ Pdf (117520)
☐ Jpg (49381)
☐ Zip (30738)

February 27, 2018 (v1.0.1) Software Open Access
air-community/air-standards: Early revision of AIRR definitions
Ahmad Syed, Christian Busse, Uri Laserson, Scott Christley, Jason Vander
An early revision of the AIRR definitions with corresponding reference library
Updated on February 27, 2018

February 27, 2018 (v0.9.7) Software Open Access
gpertea/fgtrim: fgtrim release v0.9.7
Geo Perlea,
filtering and trimming next generation sequencing reads
Updated on February 27, 2018

Open Science Framework
A scholarly commons to connect the entire research cycle

figshare Search on figshare Browse Upload

store, share, discover research

get more citations for all of the outputs of your academic research over 5000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

General depositories for research data

The following depositories are of interest to researchers in all domains:

- Zenodo (not-for-profit, hosted by CERN): <https://zenodo.org>:
- Dryad (not-for-profit membership organisation): <http://www.datadryad.org>
- Figshare (free service provided by private company): <https://figshare.com>
- Open Science Framework (not-for-profit, developed and maintained by the Center for Open Science¹): <https://osf.io>
- Harvard Dataverse (not-for-profit, hosted by the Institute for Quantitative Social Studies IQSS at Harvard University): <https://dataverse.harvard.edu>

A = Accessible. Data repositories

Checklist: is it the right repository for your data?

The checklist that follows addresses the five key questions posed in this guide:

1. is the repository reputable?
2. will it take the data you want to deposit?
3. will it be safe in legal terms?
4. will the repository sustain the data value?
5. will it support analysis and track data usage?

3. Will the data be safe in legal terms?

For this criterion we first consider the basic legal terms and conditions to check. Here only one capability level is given, as a repository either will or will not comply. Then we consider licensing, disclosure risk and access control, where repository may offer different levels of capability you can match to your needs. Note that Re3data provides relevant details for the repositories it lists, under 'terms'.

Legal terms and conditions
Level 1
Personal data or data which may identify individuals when linked to other data should not be stored outside the European Economic Area, unless in a legal jurisdiction that ensures personal data is adequately protected ☐
By agreeing to the terms and conditions the depositor will not be breaching other Data Protection principles, or the terms of any confidentiality agreement with data subjects or owners (e.g. consent form, consortium agreement) ☐

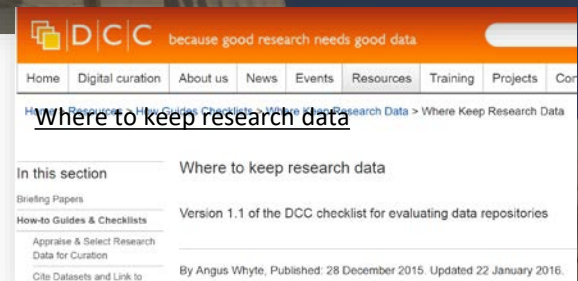
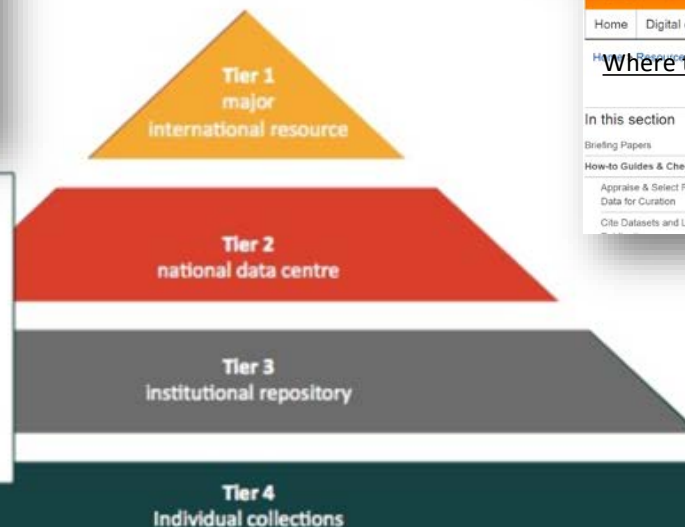
Findable, accessible and interoperable

Level 1	Level 2	Level 3
Metadata publishing: Data collections are catalogued in a repository according to funder expectations so that they are discoverable by title, creator, and date of deposition ☐	Repository publishes other pertinent information as metadata fields to enhance cross-disciplinary discovery ☐	Metadata is catalogued to enhance reuse according to sector-leading standards, or to fulfil domain-specific purposes ☐
Stable identifiers: Enables a DOI or other open standard identifier to be assigned to a landing page for each ingested dataset/ collection ☐	Supports assignment of related persistent IDs per dataset/ collection ☐	Supports assignment of multiple persistent IDs at different levels of granularity within dataset/ collection ☐
Discovery metadata: Provides Datacite mandatory metadata and exposes it according to open	Provides metadata elements to enable broader discovery (e.g. geo-spatial) to reflect best practice	Exposes discovery metadata as Linked Open Data to optimise automatic

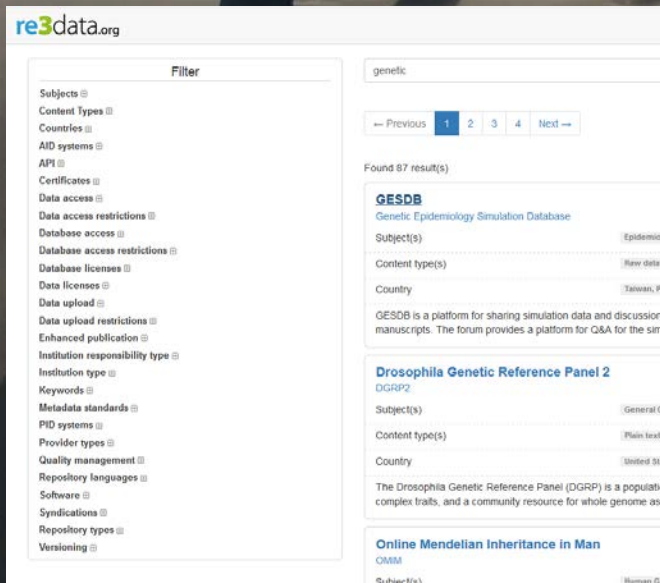
Value of data increases up the tiers: from individual to community to social value.

Each higher tier brings greater responsibility and demands for access.

As infrastructure increases so must the attention given to standards, sustainability and provenance.



A = Accessible. Looking for a data repository?



2,000 Data Repositories and Science Europe's Framework for Discipline-specific Research Data Management

By offering detailed information on more than 2,000 research data repositories, re3data has become the most comprehensive source of reference for research data infrastructures globally. Through the development and advocacy of a framework for discipline...

[Read more](#)

Three new DOI Fabrica features to simplify account management

Last month we launched DOI Fabrica, the modernized version of the DataCite Metadata Store (MDS) web frontend. It is the one place for DataCite providers and their clients to create, find, connect and track every single DOI from their organization...

[Read more](#)

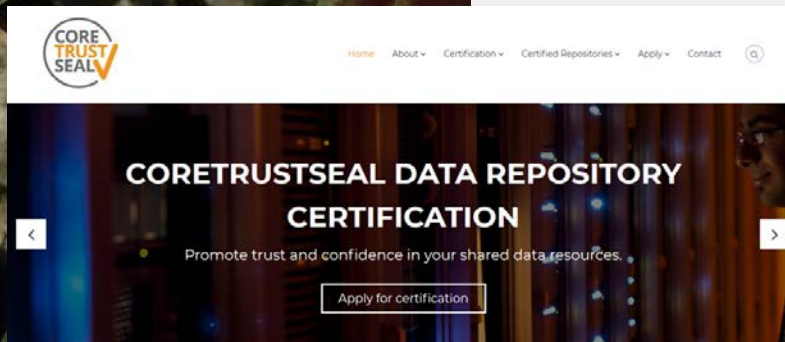
One step closer towards instant DOI search results

Art Art? You might be wondering, what this pink and green picture illustrates? A few months ago we couldn't show you this picture; the data that we used to created it, did not exist. And the answer to what this illustrates – this is simply a distorted...

[Read more](#)

<https://www.re3data.org/>

[trusted repositories]



CRITERIA FOR THE SELECTION OF TRUSTWORTHY REPOSITORIES



Trustworthy repositories should meet the following minimum criteria:

- ☐ **1. Provision of Persistent and Unique Identifiers (PIDs)**
 - a. Allow data discovery and identification
 - b. Enable searching, citing, and retrieval of data
 - c. Provide support for data versioning
- ☐ **2. Metadata**
 - a. Enable finding of data
 - b. Enable referencing to related relevant information, such as other data and publications
 - c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
 - d. Use metadata standards that are broadly accepted (by the scientific community)
 - e. Ensure that metadata are machine-retrievable
- ☐ **3. Data access and usage licences**
 - a. Enable access to data under well-specified conditions
 - b. Ensure data authenticity and integrity
 - c. Enable retrieval of data
 - d. Provide information about licensing and permissions (in ideally machine-readable form)
 - e. Ensure confidentiality and respect rights of data subjects and creators
- ☐ **4. Preservation**
 - a. Ensure persistence of metadata and data
 - b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)

A = Accessible. Data journals

Data Journals

Hier entsteht eine Liste von Data Journals, die vorwiegend Data Papers

- Atomic Data and Nuclear Data Tables [\(Elsevier\)](#)
- Biodiversity Data Journal [\(Pensoft Publishers\)](#)
- Biomedical Data Journal [\(Procon Ltd.\)](#)
- BMC Research Notes [\(Biomed Central\)](#)
- Chemical Data Collections [\(Elsevier\)](#)
- Data [\(MDPI\)](#)
- Data in Brief [\(Elsevier\)](#)
- Dataset Papers in Science [\(Hindawi Publishing Corporation\)](#)
- Earth System Science Data - ESSD [\(Copernicus Publications\)](#)
- Ecological Archives [\(Ecological Society of America - ESA\)](#)
- European Data Watch [\(European Data Watch\)](#)
- F1000Research [\(F1000 Research\)](#)
- Genomics Data [\(Elsevier\)](#)
- Geoscience Data Journal [\(Wiley\)](#)
- GigaScience [\(BioMed Central\)](#)
- Internet Archaeology [\(Internet Archaeology\)](#)
- Journal of Open Psychology Data (JOPD) [\(Ubiquity Press\)](#)
- Journal of Chemical & Engineering Data [\(ACS Publications\)](#)
- Journal of Physical and Chemical Data [\(AIP Publishing\)](#)
- Nuclear Data Sheets [\(Elsevier\)](#)
- Open Archaeology Data [\(Ubiquity Press\)](#)
- Open Data Journal for Agricultural Research [\(diverse\)](#)
- Open Health Data [\(Ubiquity Press\)](#)
- Open Journal of Bioresources [\(Ubiquity Press\)](#)
- Open Network Biology [\(BioMed Central\)](#)
- Research Data Journal for the Humanities and Social Sciences [\(Brill\)](#)
- Scientific Data [\(Nature Publishing Group\)](#)

Dataset Description

Object Name

- *walkers* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for records made by individual walkers during stage-one fieldwalking.
- *counts* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for potsherds counted during stage-one fieldwalking.
- *pottery* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main pottery database, assembled various artefact specialists.
- *petrography* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for those sherds sampled for thin section petrography.
- *lithics* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main lithics database.
- *other* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all non-ceramic and non-lithic finds.
- *structs* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all standing remains, except for terraces.
- *coast* – a vector polygon dataset (.shp and associated files) with the shape of Antikythera's coastline.
- *geology* – a vector polygon dataset (.shp and associated files) with the main bedrock units on Antikythera.
- *tracts* – a vector polygon dataset (.shp and associated files) with the main stage-one survey units.
- *grids* – a vector polygon dataset (.shp and associated files) with the main stage-two survey units.
- *terraces* – vector line dataset (.shp and associated files) with all observable agricultural terraces (i.e. the location

I don't need

- *other* – primarily Andrew Bevan (UCL), with further assistance from James Conolly (Trent)
- *geology* – a combination of fieldwork by Ruth Siddall (UCL) and remote sensing by Andrew Bevan (UCL)

Repository Location

UK Archaeology Data Service Collection 1115 (doi: 10.5284/1012484)

Publication Date

05/02/2012

Language

English (a Greek language summary of the project methods and results can be found at www.ucl.ac.uk/asp/ or www.tuarc.trentu.ca/asp/).

License

Creative Commons CC-BY 3.0

Reuse Potential

Due to their unusual coverage of an entire landscape, these datasets would provided a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (<http://dx.doi.org/10.5284/1001000> to the fact that the ASP data locations, dates and structures and terraces

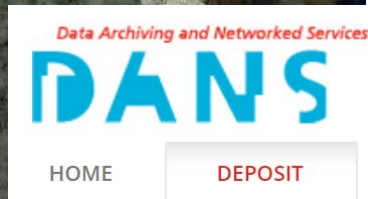
Data journals

Panayiota Polydoratou

Alexander Technological Educational Institute of Thessaloniki

European Commission Workshop
Alternative Open Access Publishing Models: Exploring New Territories in
Communication
Brussels, 12 October 2015

A = Accessible. Formats



Type	• Preferred format(s)	• Non-preferred format(s)
Text documents	• PDF/A (.pdf)	• ODT (.odt) • MS Word (.doc, .docx) • RTF (.rtf) • PDF (.pdf)
Plain text	• Unicode text (.txt)	• Non-Unicode text (.txt)
Markup language	• XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es	• SGML (.sgml)
Spreadsheets	• ODS (.ods) • CSV (.csv)	• MS Excel (.xls, .xlsx) • PDF/A (.pdf) • OOXML (.docx, .docm)
Databases	• SQL (.sql) • SIARD (.siard) • DB tables (.csv)	• MS Access (.mdb, .accdb) (v. 2000 or later) • dBase (.dbf) • HDF5 (.hdf5, .he5, .h5)
Statistical data	• SPSS Portable (.por) • SPSS (.sav) • STATA (.dta) • DDI (.xml) • data (.csv) + setup (.txt)	• SAS (.7dat; .sd2; .tpt) • R (* under examination)
Raster images	• JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png) • JPEG 2000 (.jp2)	• DICOM (.dcm) (by mutual agreement)

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) delimited text and command ("setup") file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings, variable names	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tiff) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages

Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti
Image data	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)

Video data	MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	AVCHD video (.avchd)
Documentation and scripts	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHTML 1.0

UK Data Service			
	About us	Get data	Use data
Home > Manage data > Format your data > Recommended formats			
Recommended formats			

A = Accessible. Preservation

LONG TERM OR SHORT TERM?

Checksum Checker

Software for Digital Preservation

Download version 3.0.1, released 25 March 2014 AEST

Checksum Checker is free and open source software developed by the National Archives of Australia. Checksum Checker is a piece of software that is used to monitor the contents of a digital archive for data loss or corruption.

Checksum Checker is a component of the Digital Preservation Software Platform (DPSP).

Features

As part of the Digital Preservation Recorder (DPR) workflow, checksums are generated for each Archival Information Package (AIP). Checksum Checker generates a new checksum for each AIP and compares it against the stored checksum. If the checksums do not match, then the AIP is flagged as being corrupt.

Checksum Checker incorporates the following features:

- Checksum Checker functions as a service.
- Checksum Checker sends automated emails to a nominated administrator email address, coinciding with certain events (such as the start of a checking run or when an error is encountered).

Checksum Checker is released under the GPLv3, and is available for download. <http://checksumchecker.sourceforge.net/>



Storage Solutions	Advantages	Disadvantages	Suitable for
Personal Computer & Laptop	<i>Always available</i> <i>Portable</i>	<i>Drive may fail</i> <i>Laptop may be stolen</i>	<i>Temporary storage</i>
Networked drives File servers managed by your university, research group or facilities like a NAS-server	<i>Regularly backed up</i> <i>Stored securely in a single place</i>	<i>Costs</i>	<i>Master copy of your data</i> <i>(if enough storage space is provided ..)</i>
External storage devices USB flash drive, DVD/CD, external hard drive	<i>Low cost</i> <i>Portability</i>	<i>Easily damaged or lost</i>	<i>Temporary storage</i>
Cloud services	<i>Automatic synchronization between folders and files</i> <i>Easy to access and use</i>	<i>It's not sure whether data security is taken care of</i> <i>You don't have direct influence on how often backups take place and by whom</i>	<i>Data sharing</i>

1

2

3

4

5

6

Organize and document research data. Make digital versions of paper data documentation in a PDF/A format (suitable for long-term storage).

A = Accessible. Preservation

Portable devices

Cloud storage

Local storage

Networked drive



Laptops, tablets, external hard-drives, flash drives and Compact Discs

Advantages

- Allow easy transport of data and files without transmitting them over the Internet. This can be especially helpful when working in the field.
- Low-cost solution.

Disadvantages/Risks

- Easily lost, damaged, or stolen and may, therefore, offer an unnecessary security risk.
- Not robust for long-term storage or master copies of your data and files.
- Possible quality control issues due to version confusion.

Precautions for (sensitive) personal data

Use encrypted password

Advantages

- Automatic backups.
- Often automatic version control.

Disadvantages/Risks

- Not all cloud services are secure. May not be suitable for sensitive data containing personal information about EU citizens.
- Insufficient control over where the data is stored and how often it is backed up.
- Free services by commercial providers (e.g. Google Drive, Dropbox) may claim rights to use content you manage and share them for their own purposes.
- Data can be lost if your account is suspended or accidentally deleted, or if the provider goes out of business.

Precautions for (sensitive) personal data

- Encrypt all (sensitive) personal data before uploading it to the cloud. This is particularly important to avoid conflict with European data protection regulations if you do not know in which countries servers used for storage and backup are located (see '[Security](#)' for more information on encryption; also see '[Protecting data](#)').

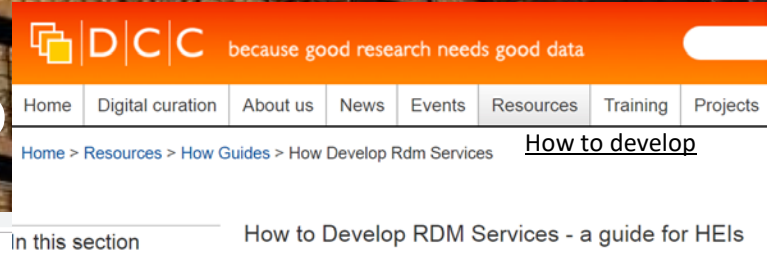
Recommendations

- Do: use cloud services for granting shared, remote and easy access to data and other files to all involved in the project.
- Do: Read the terms of service. Especially focus on rights to use content given to the service provider.
- Do: Opt for European, national, or institutional cloud services which store data in Europe if possible.
 - [B2drop](#) (EUDat, n.d.) is an example of a European cloud storage solution.
 - [SWITCHdrive](#) (SWITCH, 2017) is a Swiss solution.
 - [DataVerseNL](#) (Data Archiving and Networked Services, 2017) is an example of a service for Dutch researchers that allows the storage and sharing of data both during and after the research period.
- Don't: make this your only storage and backup solution.
- Don't: use for unencrypted (sensitive) personal data.

CESSDA Guide

Different needs, different tools.
During: you need also to share with the team.

What to preserve?



The screenshot shows the top of the DCC (Digital Curation Centre) website. The header is orange with the DCC logo and the tagline "because good research needs good data". Below the header is a navigation menu with links: Home, Digital curation, About us, News, Events, Resources, Training, and Projects. A breadcrumb trail reads "Home > Resources > How Guides > How Develop Rdm Services". The page title is "How to develop" and the main heading is "How to Develop RDM Services - a guide for HEIs".

Establishing criteria for selection decisions

In this section

How to Develop RDM Services - a guide for HEIs

You should establish criteria to guide selection decisions. The DCC's How to Select and Appraise Research Data for Curation[56] proposes seven criteria as outlined below:

1. **Relevance to mission:** the resource content fits any priorities stated in the institution's mission, or funding body policy including any legal requirement to retain the data beyond its immediate use.
2. **Scientific or historical value:** is the data scientifically, socially, or culturally significant? Assessing this involves inferring anticipated future use, from evidence of current research and educational value.
3. **Uniqueness:** the extent to which the resource is the only or most complete source of the information that can be derived from it, and whether it is at risk of loss if not accepted, or may be preserved elsewhere.
4. **Potential for redistribution:** the reliability, integrity, and usability of the data files may be determined; these are received in formats that meet designated technical criteria; and Intellectual Property or human subjects issues are addressed.
5. **Non-replicability:** it would not be feasible to replicate the data/resource or doing so would not be financially viable.
6. **Economic case:** costs may be estimated for managing and preserving the resource, and are justifiable when assessed against evidence of potential future benefits; funding has been secured where appropriate.
7. **Full documentation:** the information necessary to facilitate future discovery, access, and reuse is comprehensive and correct; including metadata on the resource's provenance and the context of its creation

I = Interoperable. Standards



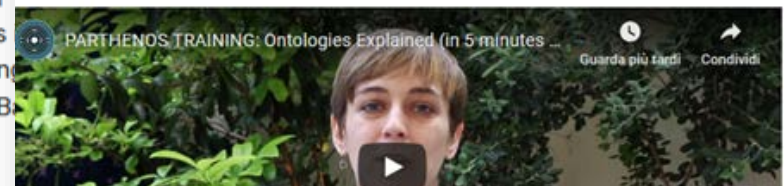
Even perfect metadata may not allow data to become interoperable if a different standard is used. A "standard" refers to a system that structures what types of information are captured in a collection. In our .mp3 library system, a standard is expressed in the header categories such as 'name,' 'time,' 'artist,' and 'album' are listed, with every entry having this filled in. Standards are used to ensure that metadata is as useful as possible for organising a collection, ensuring that common questions (how many songs are there on the album "Big B") can be easily and accurately answered.

How Many Standards Are There and Who Decides Which One To Use?

Different standards have arisen in different kinds of cultural heritage institution: the most common standards in museums are different from those in archives, and those common in libraries are different again.



In addition to metadata and standardised metadata schemas, research infrastructures can also use other forms of "knowledge representation system" to enhance the researcher's experience of the interoperable data they present. When we talk about 'Knowledge Representation Systems' in research infrastructures, we usually mean a specific category of hierarchical systems of terms known more commonly as an 'ontology'. Before the digital age, philosophers referred to an ontology as "the study of the kinds of things that exist". Ontologies are similar to taxonomies, another knowledge organisation framework you probably remember from early lessons in biology.



What is Metadata?

What are Standards?

What Are Knowledge Representation Systems and 'Ontologies'?

Sustainability

Methods and Tools

Networks

I = Interoperable. Standards

Standardization Survival Kit

A collection of research use case scenarios illustrating best practices in Digital Humanities and Heritage research



Browse scenarios



Add a new scenario



About the SSK

Increase efficiency, interoperability and sustainability by using standards

Incorporating standards in all the steps of your research process will make it last longer, easier to update, improve and share. Standards are non legally binding documents produced by an organisation ensuring :

<http://ssk.huma-num.fr/#/>

I = Interoperable

To speed up discovery and uncover new insights, research data should be easily combined with other datasets by humans as well as computer systems.

INTEROPERABLE

10 Establish well documented machine-actionable APIs

Well documented and machine-actionable APIs - a set of subroutine definitions, protocols, and tools for building application software - allow for automatic indexing, retrieval and combining of (meta)data from different data repositories.



Document APIs well and make it possible to deliver the schema of the (meta)data model. Consider showing examples of how to successfully mine data from different endpoints and combine them into new data sets usable for new research.

11 Use open well-defined vocabularies

The description of metadata elements should follow community guidelines that use open, well defined and well known vocabularies. Such vocabularies describe the exact meaning of the concepts and qualities that the data represent.



Use vocabularies relevant to your field, and enrich and structure your research output accordingly from the start of your research project.



Give examples of vocabularies the research community may use, based on research domain specifics.

12 Document metadata models

Clearly documenting metadata models helps developers to compare and make mappings between metadata.



Publish the metadata models in use in your research infrastructure. Document technical specifications and define classes (groups of things that have common properties) and properties (elements that express the attributes of a metadata section as well as the relationships between different parts of the metadata). For metadata mapping purposes, list the mandatory and recommended properties.

13 Prescribe and use interoperable data standards

Using a data standard backed up by a strong community, increases the possibility to share, reuse and combine data collections.



Check with the repository where you want to deposit your data what data standards they use. Structure your data collection in this format from the start of your research project.



Clearly specify which data standard your institution uses, pool a community around them and maintain them especially with a perspective on interoperability. Good examples are CMDI (language studies) and the SIKB0102 Standard (archaeology).

14 Establish processes to enhance data quality

To boost (meta)data quality and, therefore, interoperability, establish (automatic) processes that clean up, derive and enrich (meta)data.



Establish procedures to minimise the risk of mistakes in collecting data. E.g. choose a date from a calendar instead of filling it in by hand.



Invest in tools to help clean up (meta)data and to convert data into standardised and interoperable data formats. Combine efforts to develop workflows and software solutions for such automatic processes, e.g. by using machine learning tools.

15 Prescribe and use future-proof file formats

All data files held in a data repository should be in an open, international, standardised file format to ensure long-term interoperability in terms of usability, accessibility and sustainability.

GUIDELINES
to FAIRify data
management
and make data
reusable

PARTHENOS

2019

I= Interoperable – ac



RightField

Rightfield is an open-source tool for adding ontology term selection to Excel spreadsheets. Rightfield is used by a 'Template Creator' to create semantically aware Excel spreadsheet templates. The Excel templates are then reused by Scientists to collect and annotate their data; without any need to understand, or even be aware of, Rightfield or the ontologies used. Rightfield embedded templates are used within the [Samples](#) framework of the [SEEK](#).

MORE INFORMATION

<https://fair-dom.org/platform/rightfield/>

fic research datasets, models or simulations, processes and
nation about the people and organisations involved. The
i) based on the ISA-Tools format. When paired with our
ection through to publication. Norwegian users benefit from
PI simplifies upload and download of files.

ata sharing within groups and consortia. In addition,

<https://fair-dom.org/platform/>

Metadata	Values (examples)	Notes
Asset Title		The name of the data file
Uploader		The person submitting the asset to SEEK
Uploader SEEK ID		If you add your own SEEK ID, this will help us link this asset with your profile
Project	Project	The project that the asset belongs to
ASSAY		
Assay SEEK ID		If referring to an existing Assay, you can link to it via the Assay SEEK ID.
Assay Title		The title of an existing assay
Assay_type	ExperimentalAssayType	The assay_type describes the type of experiment you are performing
Technology_type	amplification by-product_formation catabolic_response cell_growth_optimisation cell_size Comparative_genomic_hybridization comparative_genomics continuousEnzymatic	Describes the type of instruments and/or equipment used for the experiment
Description		A brief, human readable description.
Experimentalist		The names of the people who carried out the experiments. These can either be SEEK members or external scientists
Date		The start date for the experiment if different from the upload date
SOP		Links to SOPs and protocols used to carry out the experiment. If they are already in SEEK, you can refer to them by their SEEK ID
Publication (optional)		If this data appears in a publication, you can link it directly, or via the assay or study. If it is already registered in SEEK, you can use the PubMed ID or DOI as a reference.
Experimental_conditions		
Item	ExperimentalConditions	The name of the experimental condition you are fixing in your experiment (e.g. temperature, concentration, pH etc). If there is more than 1, please list them in columns across the spreadsheet
Compound (if concentration)		The compound name is only required if the item is concentration.
Unit		The SI units of the experimental conditions measurements.
Start_value (optional)		This field is used for recording changes throughout the experiment to measure different conditions (e.g. pH or dilutions)
End_value (optional)		This field is used for recording changes throughout the experiment to measure different conditions (e.g. pH or dilutions)
Comments		Additional information that would be useful for people reading this data file

[FAIRsharing. To be Interoperable]

FAIRsharing.org
standards, databases, policies

Search all of FAIRsharing

Standards

Databases

Policies

Collections

Add/Claim Content

Stats

Log In or Register

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

HOW CAN WE HELP?

FAIRsharing.org
standards, databases, policies

Search all of FAIRsharing

Standards

Databases

Policies

Collections

Add/Claim Content

Stats

Log In or Register

Showing records 1 - 50 of 1294.

View as Table

View as Grid

Sort by

Name

Recommended Records

Recommended

Associated Publication?

No Publication

Has Publication

Claimed?

No Maintainer

Has Maintainer

Record Status

Uncertain

Deprecated

In development

Ready

Standard Type

Terminology Artifact 726

Model/Format 267

Reporting Guideline 154

Metric 14

Identifier Schema 11

Show More

Domains

Chemical Entity 128

Report 109

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	Is Collection/Recommendation	Status
ABA Adult Mouse Brain	ABA	ABA	Standard	None	None	None	None	None	None	None	R
Access to Biological Collection Data	ABCD	ABCD	Standard	Biotechnology Life Sciences	None	None	GBIF Atlas of Living Australia IPT - GBIF Australia Repository GBIF Spain IPT - GBIF Spain Repository GBIF Canada IPT - GBIF Canada Repository GBIF Colombia IPT - GBIF Colombia Repository Plus 1 more...	ABCD EFG ABCDORNA	None	World Biogeography Information Network	R
Access to Biological Collection Data Extension for Geoscientists	ABCD EFG	ABCD EFG	Standard	Biotechnology Life Sciences	None	None	GeoCase Data Portal	XML ABCD	None	None	R
Access to Biological Collection Data DNA extension	ABCDORNA	ABCDORNA	Standard	Biotechnology Life Sciences	DNA Sequences Data Significant Mutations Proteins Chromosomes Data Proteomes (Data Repository) Plus 1 more...	None	GenBank	ABCD	None	World Biogeography Information Network	Dev
ACE format	ACE format	ACE format	Standard	Life Sciences	DNA Sequences Data Cloning Transcriptome Data Sequencing	None	None	None	None	None	R
Adalab meta ontology	ADALAB.META	ADALAB.META	Standard	None	None	None	None	None	None	None	R
Adalab ontology	ADALAB	ADALAB	Standard	None	None	None	None	None	None	None	R
Achene Drug Reaction Markup Language	EU-ADR ML	EU-ADR ML	Standard	None	Phytochemical Health Record Achene Reaction Chemicals Drug	None	None	None	None	None	U

Natural Sciences

1120

Life Sciences

1087

Biology

454

Addgene

Addgene

Database

Life Sciences

Biotechnology

Chromosomes Data

None

None

GenBank

Sequence

Format

Springer

Nature

Phonics

Availability

FASTA

Data Citation Implementation

FAIR Standards Starter Pack

<https://fairsharing.org>

R = Reusable. Documentation



Project-level documentation



Project-level documentation explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc. In the accordion the questions which your project-level documentation should answer are stated in more detail:

detail:

- ⊕ 1. For what purpose was data created
- ⊕ 2. What does the dataset contain
- ⊕ 3. How was data collected
- ⊕ 4. Who collected the data and when
- ⊕ 5. How was the data processed
- ⊕ 6. What possible manipulations were done to the data
- ⊕ 7. What were the quality assurance procedures
- ⊕ 8. How can data be accessed

Data-level documentation

Data-level or object-level documentation provides information at the level of individual objects such as pictures or interview transcripts or variables in a database. You can embed data-level information in data files. For example, in interviews, it is best to write down the contextual and descriptive information about each interview at the beginning of each file. And for quantitative data variable and value names can be embedded within the data file itself.



⊖ Quantitative data

Variable-level annotation should be embedded within a data file itself. If you need to compile an extensive variable level documentation that can be created by using a structured metadata format.



Data-level documentation for quantitative data

For quantitative data document the following:

- **Information about the data file**
Data type, file type and format, size, data processing scripts.
- **Information about the variables in the file**
The names, labels and descriptions of variables, their values, a description of derived variables, if available, for variables with restrictions etc. The quantitative

R = Reusable. Documentation

≡ ✓ protocols.io

Make your science more reproducible
protocols.io is the #1 open access repository for science methods

Editing Fixation of yeast cells for RNA-FISH

DESCRIPTION
GUIDELINES & WARNINGS
MATERIALS
STEPS

- INOCULATION AND GROWTH
1. Around 10am, start a cell culture in a 50ml tube...
 2. Grow for 8-10 hours in a shaker at 30 °C.
 3. Measure OD in the evening and dilute into 250ml...

- FIXATION
4. Transfer to 50ml falcon tubes.
 5. Add 5ml of Formaldehyde, invert a few times, set...

new step
paste from text
paste from buffer

NEW SECTION (optional)

Fixation
Add 5ml of Formaldehyde, invert a few times, set...

DURATION / TIMER
EXPECTED RESULT
EXTERNAL LINK

NOTES
Optional: Transfer to gentle rocking oven recommended per Anne Dodson, Marc I...

ROBERT BLANK DELETE

RStudio

<https://www.rstudio.com/>

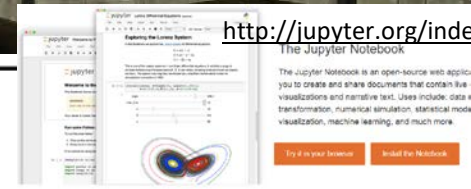
RStudio

Open source and enterprise-ready professional software for R

What is an Open Notebook?

Open Notebooks are documents that contain equations, visualisations, narrative text and live code that can be executed independently and interactively, with output visible immediately beneath the input.

They bring together analysis descriptions and results, which can be executed to perform the data analysis in real time.

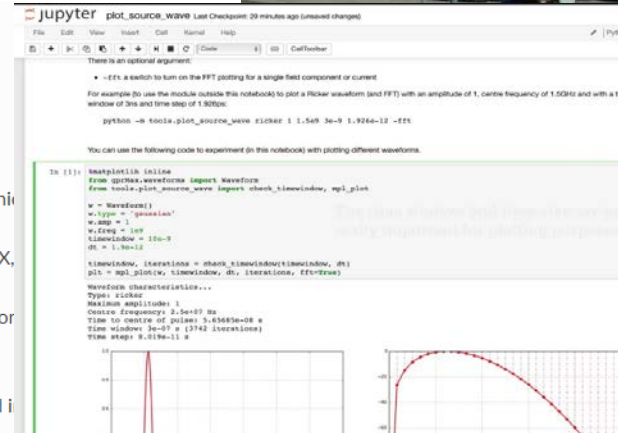


<http://jupyter.org/index>
The Jupyter Notebook

Notebook web application

The notebook web application enables users to:

- Edit code in the browser, with automatic syntax highlighting, indentation, and tab completion/introspection.
- Run code from the browser, with the results of computations attached to the code which generated them.
- See the results of computations with rich media representations, such as HTML, LaTeX, SVG, PDF, etc.
- Create and use interactive JavaScript widgets, which bind interactive user interface controls and visualizations to reactive kernel side computations.
- Author narrative text using the Markdown markup language.
- Include mathematical equations using LaTeX syntax in Markdown, which are rendered in the browser by MathJax.



...why not?

- protocols.io to deposit your methods
- Open Lab Notebook to track anything you do [time consuming the first time, then...]

R= Reusable. Licenses

Copyright: protects the STRUCTURE, selection or arrangement of their contents" (Art. 3) NOT THE DATA

Sui generis database right: protects the «substantial effort» in OBTAINING data [NOT «CREATING»]... the right owner often is the institution

Database=a collection of independent works, data or other materials arranged in a systematic or methodical way (Art.1)



KEEP CALM
AND
REMEMBER:
RAW DATA ARE NOT
PROTECTED BY
COPYRIGHT

DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL
of 11 March 1996

on the legal protection of databases

COUNCIL OF THE EUROPEAN UNION,
European Community, and in particular Article 37 (2), 66 and 100a thereof,

Simone Aliprandi

2014

la QUALI DIRITTI SUI DATI?

raw data

non creative
database

creative
database

copyright

sui generis

NO copyright

sui generis

sui generis+
copyright

[we are not playing together]

Obstacles to the trans-European archiving and sharing of research data

Making research data as openly available as possible is a widely recognised goal. For researchers working on an interdisciplinary project involving several countries, it can be difficult to fully comprehend in which ways open access to research data can be legally obtained. European national laws still diverge.

[A report from Knowledge Exchange](#) (Knowledge Exchange, 2011) concludes that it will remain difficult to predict when particular files of research data are protected because of:

- **Diversity in copyright protection**

Even though most research data will fail to meet the criteria for copyright protection because they are not likely to be considered as “works” (they mainly concern facts), the lack of harmonisation of the criteria for copyright protection in Europe is tricky. E.g., whereas Germany, Denmark and the Netherlands have a relatively similar (higher) originality standard, the UK has a very low standard (skill, judgment and labour) making it

- **Diversity in copyright owner**

If protection applies, the right holder's consent is required for sharing the data. However, the designation of the copyright owner is also different in different jurisdictions. Although in many cases the maker of the work will be considered to be the author and therefore the right holder, only Dutch and UK law designate the employer as the right holder if the work was made in the course of employment.

[CESSDA guide](#)

⊕ Czech Republic

⊕ Finland

⊕ Sweden

⊕ Switzerland

⊕ UK

[Clear rules with the project partner]

Don't even
think of park-
ing here! :)



Safe to be open 2016

Safe to be open

Study on the protection
of research data and
recommendations
for access and usage

Edited by Lucie Guibault
and Andreas Wiebe

- ... SET CLEAR RULES **FROM THE BEGINNING**
- WHO IS THE RIGHT HOLDER (if)
 - WHO HAS THE RESPONSIBILITY OF PRESERVE

R = Reusable. Licenses



How to License Research Data

This guide will help you decide how to apply a licence to your research data, and which licence would be most suitable. It should provide you with an awareness of why licensing data is important, the impact licences have on future research, and the potential pitfalls to avoid. It concentrates on the UK context, though some aspects apply internationally; it does not, however, provide legal advice. The guide should interest both the principal investigators and researchers responsible for the data, and those who provide access to them through a data centre, repository or archive.

<http://www.dcc.ac.uk/resources/how-guides/license-research-data>

R = Reusable. Licences

How to License Research Data

This guide will help you decide how to apply a licence to your research data, and which licence would be most suitable. It should provide you with an awareness of why licensing data is important, the impact licences have on future research, and the potential pitfalls to avoid. It concentrates on the UK context, though some aspects apply internationally; it does not, however, provide legal advice. The guide should interest both the principal investigators and researchers responsible for the data, and those who provide access to them through a data centre, repository or archive.

Creative Commons at a glance

Good for

- very simple, factual datasets
- data to be used automatically

Watch out for

- versions: use v. 4 or later
- attribution stacking
- the NC condition: only use with dual licensing
- the SA condition as it reduces interoperability
- the ND condition as it severely restricts reuse

ODC-By at a glance

Good for

- most databases and datasets
- data to be used automatically
- data to be used for generating non-data products

Watch out for

- attribution stacking

ODC-ODbL at a glance

Good for

- most databases and datasets
- data to be used automatically
- data to be used for generating non-data products

Watch out for

- attribution stacking
- the copyleft condition as it reduces interoperability
- the DRM clause as it may put off some reusers

Public domain at a glance

Good for

- most databases and datasets
- data to be used by anyone or any tool
- data to be used for any purpose

Watch out for

- lack of control over how database is reused
- lack of protection against unfair competition

Creative Commons



FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE V0.1

This information guide contains questions and responses to common concerns surrounding open science and the implications of licensing data under Creative Commons licences. It is intended to aid researchers, teachers, librarians, administrators and many others using and encountering Creative Commons licences in their work.

<https://doi.org/10.5281/zenodo.840651>

What is Open Science?

Open Science is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the **CC0 Public Domain Dedication**, which is first and foremost a waiver, but **can act as a licence** when a waiver is not possible.

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is not protected as a matter of law. Such data can include facts, names, numbers – things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered 'non-original' and thus ineligible for copyright, and it might additionally be excluded

from other forms of protection (like the [EU sui generis database right](#), also known as the 'SGDR', for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a [Public Domain Mark](#).

PUBLIC DOMAIN MARK LOGO



When in doubt, consider which use may be appropriate according to the chart below:

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free



'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

What is Open Science?

Open Science is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the **CC0 Public Domain Dedication**, which is first and foremost a waiver, but **can act as a licence** when a waiver is not possible.

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is **not** protected as a matter of law. Such data can include facts, names, numbers – things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered 'non-original' and thus ineligible for copyright, and it might additionally be excluded

from other forms of protection (like the [EU sui generis database right](#), also known as the 'SGDR', for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a [Public Domain Mark](#).

PUBLIC DOMAIN MARK LOGO



When in doubt, consider which use may be appropriate according to the chart below:

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free



'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

But I would like attribution when others use my dataset. In that case, shouldn't I use a CC BY licence?

We recommend that you avoid using a CC BY licence. Here's why:

While attribution is a genuine, recognisable concern, not only might using a CC BY licence be legally unenforceable when no underlying copyright or SGDR protects the work, but it may also communicate the wrong message to the world. A better solution is to use CC0 and [simply ask for credit](#) (rather than require attribution), and provide a citation for the dataset that others can copy and paste with ease. Such requests are consistent with scholarly norms for citing source materials.

Legally speaking, datasets that are **not** subject to copyright or related rights (and are thus in the public domain) cannot be the object of a copyright licence. Despite this, agreements based in contract law may be enforceable. Creative Commons licences, however, are copyright licences. Therefore, where the conditions for a copyright or related right are not triggered, copyright licences, such as the CC BY licence, [are unenforceable](#).

In some cases, however, rights may exist (like the *sui generis* database right previously mentioned), and permission for others to use your dataset may be legally required. These rights are meant to protect the maker's investment, rather than originality. As such, database rights do not include the moral right of attribution. So by using a CC BY licence, you signal to users that you restrict access to your dataset beyond the protections provided by the law. We are not saying that this cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

Commons Open

USE A CC0

- THEN ASK FOR CREDIT
- PROVIDE A CITATION TO C&P
- BEAR IN MIND IT'S BAD SCIENCE NOT TO CITE THE SOURCE
- CC0 DOES NOT MEAN ACADEMIC UNPOLITENESS

cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

I'm uncomfortable with others using my research for commercial purposes. Should I use a non-commercial licence for my dataset?

We recommend you avoid using a non-commercial licence. Here's why:

For legal purposes, drawing a line between what is and is not 'commercial' can be tricky; it's not as black and white as you might think. For example, if you release a dataset under a non-commercial licence, it would clearly prohibit an organisation

It sounds like you're really pushing for the use of CC0 for open science datasets.

Exactly. Data is only open if anyone is free to use, reuse, and distribute it. This means it must be made available for both commercial and non-commercial purposes under non-discriminatory conditions that allow for it to be modified.

When data is made available for all reuse, others can create new knowledge from combining it. This leads to the enrichment of open datasets and further dissemination of knowledge. Accordingly, CC0 is ideal for open science as it both protects and promotes the unrestricted circulation of data

And remember, it's bad science not to cite the source of data you use. To help others cite your data [include a citation](#) that users can copy and paste to give you credit for your hard work.

I'm uncomfortable permitting use of my research for any and all purposes. Should I use a 'No Derivatives' (ND) licence for my dataset?

We recommend you avoid using a 'No Derivatives' licence. Here's why:

Similar to how a non-commercial licence might restrict meaningful reuse of your dataset, a ND licence can have the same effect: it may prevent someone from recombining and reusing your data for new research. For data to be truly Open Access, it must permit these important types of reuse.

[webinar]

Italian Open Science Webinar Series 2019



Italian Open Science Webinar Series 2019

[More Information about the 2019 webinar series.](#)

[data management plan](#) | [OA to research data](#) | [open science](#)



Aspetti legali nella gestione dei dati della ricerca

Thomas Margoni
University of Glasgow - CREATE
OpenAIRE project

Support

RESOURCES

[Open Science Primers](#)
[Guides](#)
[Factsheets](#)
[Use cases](#)

HELPDESK

[FAQs](#)
[Ask a Question](#)

TRAINING

[Webinars](#)
[Workshops](#)
[Community of Practice](#)

OPENING DATA



Why Open Data?



Wilma van Wezenbeek

@wvanwezenbeek

Following

#osc2018 @sjDCC I really like what Sarah said just now "There is more risk in losing your data than sharing your data #openscience"

Traduci il Tweet

11:14 - 13 mar 2018

10 Retweet 10 Mi piace



<https://twitter.com/wvanwezenbeek/status/973502457115537408>



Oct. 2017

Digital Science Report

The State of Open Data 2017

of analyses and articles about open data, curated by Figshare

Foreword by Jean-Claude Burgelman

OCTOBER 2017

"Open data is like a renewable energy source: it can be reused without diminishing its original value, and reuse creates new value."



Sharing data: good for science, good for you

https://www.youtube.com/watch?v=HUp_OAa7n&feature=youtu.be

People will contact me to ask about stuff

YES, I KNOW. FRANKENSTEIN WAS THE DOCTOR, NOT THE MONSTER. FROM FLICKR BY CHOP SHOP GARAGE.

Christopher and Alex (C&A) say: "This is usually an objection of people who feel overworked and that [data sharing] isn't part of their job..." I would add to this that science is all about learning from each other – if a researcher is opposed to the idea of discussing their datasets, collaborating with others, and generally being a good science citizen, then they should be outed by their community as a poor participant.

People will misinterpret the data

C&A suggest this: "Document how it should be interpreted. Be prepared to help and correct such people; those that misinterpret it by accident will be grateful for the help." From the UK Data Archive: "Producing good documentation and providing contextual information for your research project should enable other researchers to correctly use and understand your data."

It's worth mentioning, however, a second point C&A make: "Publishing may actually be useful to counter willful misrepresentation (e.g. of data acquired through Freedom of Information legislation), as one can quickly point to the real data on the web to refute the wrong interpretation."

My data is not very interesting

C&A: "Let others judge how interesting or useful it is — even niche datasets have people that care about them." I'd also add that it's impossible to decide whether a dataset has value to future research. Consider the many datasets collected before "climate change" was a research topic which have now become invaluable to documenting and understanding the phenomenon. From the UK Data Archive:

I might want to use it in a research paper

Anyone who's discussed data sharing with a researcher is familiar with this excuse. The operative word here is *might*. How many papers have we all considered writing, only to have them shift to the back burner due to other obligations? That said, this is a real concern.

C&A suggest the embargo route: "One option is to have an automatic or optional embargo; require people to archive their data at the time of creation but it becomes public after X months. You could even give the option to renew the embargo so only things that are no longer cared about become published, but nothing is lost and eventually everything can become open." Researchers like to have a say in the use of their datasets, but I would caution to have any restrictions default to sharing. That is, after X months the data are automatically made open by the repository.

I would also add that, as the original collector of the data, you are at a huge advantage compared to others that might want to use your dataset. You have knowledge about your system, the conditions during collection, the nuances of your methods, et cetera that could never be fully described in the best metadata.

I'm not sure I own the data

My data is too complicated.

C&A: "Don't be too smug. If it turns out it's not that complicated, it could harm your professional [standing]." I would add that if it's too complicated to share, then it's too complicated to reproduce, which means it's arguably not real scientific progress. This can be solved by more documentation.

My data is embarrassingly bad

C&A: "Many eyes will help you improve your data (e.g. spot inaccuracies)... people will accept your data for what it is." I agree. All researchers have been on the back end of making the sausage. We know it's not pretty most of the time, and we can accept that. Plus it helps you strive will be at managing and organizing data during your next collection phase.

It's not a priority and I'm busy

Good news! Funders are *making* it your priority! New sharing mandates in the OSTP memorandum state that any research conducted with federal funds must be accessible. You can expect these sharing mandates to drift down to you, the researcher, in the very near future (6-12 months).

tor, the wing

hen owner.

arly has owner

... «as open as possible»...

Commission européenne
Europese Commissie



Carlos Moedas ✓

@Moedas

Segui

2/4 "Open as possible, as closed as necessary" is the new principle for all [#data](#) from publicly funded [#research](#) in Europe [#openaccess](#)

RETWEET
76

MI PIACE
32



IT IS A STRUCTURED WAY
TO THINK OF YOUR DATA

CLEAR RULES, LESS
MISTAKES FROM THE
BEGINNING

IT'S A «LIVING DOCUMENT»,
IT GROWS WITH THE PROJECT

Top tip - keep it short and specific! FOSTER toolkit

This very short extract from a presentation by Peter Dukes, Medical Research Council (MRC) back in 2012 provides really useful advice on writing a DMP from the funding body perspective. While it is an example from the Life Sciences, the advice applies to all disciplines. The quality of the video isn't great, unfortunately, but the advice provided definitely is!



The screenshot shows a video player interface. At the top, there's a title bar with 'Advice on writing Data Management Plans' and a 'Research Data' button. Below the title bar, the main content area displays 'Improved Data Management Plans' and a section titled '4. Keep it simple'. This section contains a list of bullet points: 'Informative: two audiences', 'Specific: e.g. name standards', 'Concise: < 1/4 to 3 pages', and 'Don't forget: your achievements & innovation'. A play button is visible in the center of the video player.

- Informative: two audiences
- Specific: e.g. name standards
- Concise: < 1/4 to 3 pages
- Don't forget: your achievements & innovation

WHERE SHOULD YOU PUT ALL THESE
INFORMATION? IN A DMP, THE PERFECT TOOL

DMP - advantages



⊖ Benefit 3. Clarifies needed budget

Data management is not free. You do not want to find yourself running out of funding before the end of the project because you have ignored or underestimated the cost of structured, detailed, and safe data management. Therefore, an important aspect of a DMP is its use in calculating how much money will be required for managing your research data during your research project.

A DMP can be useful in the process of applying for funding. Grant applications should not only include time and resources for collecting, analysing, and publishing on data in their budget, time and resources for careful documentation as well as server space, backup solutions, and documentation software need to be included as well. A DMP is also useful once funding is granted to plan and manage your expenses. Many research funders require a DMP as part of the application and decision-making process. The arguments for making data available are several, the most popular being that the data produced by public funds should be used to the greatest extent possible and available to the public. Unless there are legal, ethical or commercial barriers, data should also be openly available so that research results can be verified, replicated and reused.

Examples of Data Management cost assessments are given by the [University of Utrecht](#) (n.d.) and the Dutch Landelijk Coördinatiepunt Research Data Management ([LCRDM](#), n.d.) inspired by the '[Data management costing tool](#)' by UK Data Service, 2013.

CESSDA Guide

IT'S CRUCIAL
TO BUDGET THE COSTS

DMP Core Requirements

CORE REQUIREMENTS FOR DATA MANAGEMENT PLANS



When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

- ☐ **1. Data description and collection or re-use of existing data**
 - a. How will new data be collected or produced and/or how will existing data be re-used?
 - b. What data (for example the kinds, formats, and volumes) will be collected or produced?

- ☐ **2. Documentation and data quality**
 - a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
 - b. What data quality control measures will be used?

- ☐ **3. Storage and backup during the research process**
 - a. How will data and metadata be stored and backed up during the research process?
 - b. How will data security and protection of sensitive data be taken care of during the research?

- ☐ **4. Legal and ethical requirements, codes of conduct**
 - a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
 - b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
 - c. How will possible ethical issues be taken into account, and codes of conduct followed?

- ☐ **5. Data sharing and long-term preservation**
 - a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
 - b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
 - c. What methods or software tools will be needed to access and use the data?
 - d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

- ☐ **6. Data management responsibilities and resources**
 - a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
 - b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?



CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS



DMP Core Requirements

Translating the Core Requirements into a DMP template

The following example of a data management plan template is based on the core requirements for DMPs.⁶ These core requirements should be considered as a minimum standard, leaving the flexibility to formulate additional guidelines according to the needs of specific domains or to national or local legislation.

The template presented below refers to the 15 questions covering six core requirements for good data management. Additional guidance and explanations are provided to help researchers fill out such a template and to assure that all relevant aspects of research data management are covered. The below table is an example of how the core requirements can be transformed into a DMP template. It will be up to the individual organisations and disciplines to develop templates that fit their needs.

GENERAL INFORMATION

- Administrative information**
- Provide information such as name of applicant, project number, funding programme, version of DMP.

1 DATA DESCRIPTION AND COLLECTION OR RE-USE OF EXISTING DATA

1 a

How will new data be collected or produced and/or how will existing data be re-used?

- Explain which methodologies or software will be used if new data are collected or produced.
- State any constraints on re-use of existing data if there are any.
- Explain how data provenance will be documented.
- Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.

2 DOCUMENTATION AND DATA QUALITY

2 a

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- Indicate which metadata will be provided to help others identify and discover the data.
- Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
- Use community metadata standards where these are in place.
- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
- Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.
- Consider how this information will be captured and where it will be recorded for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks.

2 b

What data quality control measures will be used?

- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.



DMP questions



Overview

Title of the project

Date of this plan

Description of the project

- What is the nature of the project?
- What is the research question?
- What is the project time line?

Origin of Data

- What kind of data will be used during the project?
- If you are reusing existing data: What is the scope, volume and format? How are different data sources integrated?
- If you are collecting new data can you clarify why this is necessary?

Principal researchers

- Who are the main researchers involved?
- What are their contact details?

Collaborating researchers (if applicable)

- What are their contact details and their roles in the project?

Funder (if applicable)

- If funding is granted, what is the reference number of the funding granted?

Data producer

- Which organisation has the administrative responsibility for the data?

Project data contact

- Who can be contacted about the project after it has finished?

Data owner(s)

- Which organisation(s) own(s) the data?
- If several organisations are involved, which organisation owns what data?

Roles

- Who is responsible for updating the DMP and making sure that it's followed?
- Do project participants have any specific roles?
- What is the project time line?

Costs

- Are there costs you need to consider to buy specific software or hardware?
- Are there costs you need to consider for storage and backup?
- Are potential expenses for (preparing the data for) archiving covered?

Organising and documenting your data

Data collection

- How will the data be collected?
- Is specific software or hardware or staff required?
- Who will be responsible for the data collection?
- During which period will the data be collected?
- Where will the data be collected?

Data organisation

- How will you organise your data?
- Will the data be organised in simple files or more complex databases?
- How will the data quality during the project be ensured?
- If data consists of many different file types (e.g. videos, text, photos), is it possible to structure the data in a logical way?

Data type and size

- What type(s) of data will be collected?
- What is the scope, quantity and format of the material?
- After the project: What is the total amount of data collected (in MB/GB)?

File format

- In what format will your data be?
- Does the format change from the original to the processed/final data?
- Will your (final) data be available in an open format?

Folder structure and names

- How will you structure and name your folders?

File structure and names

- How will you structure and name your files?

Documentation

- What documentation will be created during the different phases of the project?
- How will the documentation be structured?

Metadata

- What metadata will be provided with the collected/ generated/ reused data?
- How will metadata for each object be created?
- Is there any program that can be used to document the data?
- Can metadata be added directly into the files or will the metadata be produced in another program or document?

Metadata standard (if applicable)

- What metadata standard(s) will you use?

DMP questions

Processing your data

PROCESS



Versioning

- What is your strategy concerning versioning your data files (and scripts) during the project?
- Will you create and/or follow a convention for versioning your data?
- Who will be responsible for securing that a "Masterfile" will be maintained, documented and versioned according to the project guidelines?
- How can different versions of a data file be distinguished?

Interoperability

- Will you make use of established software and hardware? If not, how does the software and hardware you use relate to other research?

If applicable:

- Will you make use of established terminologies/ontologies (i.e. structured controlled vocabularies) in the project? If not, how do your terminologies relate to established ones?
- Which coding is used (if any)? Will you build on established coding schemes? If not, how does your coding relate to other research?

Storing your data

STORE



Storage

- How and where will the data be stored during the project?
- For how long will the data be stored?

Backup

- How, where and at what intervals will the data be backed-up?
- How will data be recovered in the case of a data loss incident?

Security

- How will sensitive data be protected? (if applicable)
- How will data access be managed?

cessda

Consortium of European
Social Sciences Data Archives

PROTECT



Protecting your data

Ethical review (if applicable)

- Does your project require approval by a local ethics committee?

Informed consent (if applicable)

- Do you require informed consent for your project?
- If so, how will permission be obtained?
- How are consent files organised and stored?

(sensitive) Personal data /confidential information (if applicable)

- How will access to (sensitive) personal data during the project be controlled?
- How will collaborators be granted access to the data in a secure way?
- If the research project is going to have data that includes confidential information or information that requires informed consent, is there a requirement to notify a privacy officer?
- Is there any confidential information within the material that requires special treatment and/or limits the access to it during/after the project?
- How will the material be protected during/after the project?
- How will permissions and restrictions be enforced?

Intellectual property rights (IPR)/Copyrights

- Are there IPR or copyright issues to consider?
- Will permission be needed to collect/reuse the data?
- Will these rights be transferred to another organisation for data distribution and archiving?

Agreements (if applicable)

- What are the agreements with other stakeholders?

Restrictions (if applicable)

- Are there any other restrictions that need to be considered?

DMP question

Archiving and publishing your data



Archiving

- How and where will the data be stored after the project's completion?
- Will you archive your data in a trusted data repository?
- Will your data receive a persistent identifier?

Data formats

- What formats will you provide your data in for archiving (and sharing)?
- Will specific software be required to process your data? Can this software be deposited with the data?

Access (if applicable)

- Will your data be available (Open Access)?
- Will all data or only parts of it be published?
- What licenses do you need for your data?
- How should your data be cited when reused?
- Will there be an embargo period for (all or some of) the data?
- Are there other agreements or restrictions (see above) that need to be considered?
- Are there any legal/ethical restrictions that prevents the publication of all the material?
- Will these restrictions mean that action must be taken before the material can be made available?
- Is there a risk of delayed publication/making data available (all or parts of)?
If so what might be needed to do to avoid this?

DISCOVER



Discovering data

Identification of needs

- Do you plan to use existing data for your research?
- What is the purpose for which you need the data?
- What do you want to learn from the data?
- What type of data do you need?

Search for data

- Do you know where the data may be located?
- How do you plan to search for the data?

Evaluation of data quality

- What is the minimal required quality of the data (in terms of origin, contents, scope, size, methods, etc.)?
- How do you plan to evaluate data quality (evaluation of metadata, tests, analysis, comparisons)?

Gaining access to data

- What are the (expected) terms and conditions for data access and use?
- What is the (expected) process for gaining access to the data?
- What is the (expected) time-span of the process for gaining access to the data?
- What are the (expected) costs for data access and use?

Basic Information.

- State the purpose of the data collection/generation.
- Explain the relation to the objectives of the project
- Consider what data will be collected or created as part of the study (RAW data).
- Consider what data will be produced by processing the RAW data (Secondary, processed data).
- Specify if existing data is being re-used (if any)
- Specify the origin of the data
- Specify the types and formats you plan to use for the data generated/collected (raw, processed, published).
- Consider what data will be published as the result of your study (Published data).

Volume and Life Cycle of the Data.

If you are using FAIRDOM, we will look after data that will be retained and potentially exchanged by your projects. It will help with local storage for temporarily-held local data prior to processing.

For RAW data, please consider the following:

- How much RAW data you think will be produced (Estimates, per month, year, full project duration)?
- Will all of the RAW data be kept for the duration of the study or will the RAW data be deleted once it is processed?
- For large scale RAW data (images, sequence) have you planned the local storage capacity necessary for processing?
- Do you require help to organise a suitable local management system for RAW data?
- Do you have policies that govern the management and usage of RAW data?
- How long will RAW data be kept? Will there be a long-term archive?

For Secondary and Published data, please consider the following:

- What data processing is foreseen in the project?
- How much processed data will be produced, and stored (can you make estimates per month, year, full project)?
- How much of this data will be published? (Estimates per month, year, full project)?
- Does your institution, or the project funders, have policies governing the access and usage of processed data?

Additional for personally sensitive data (e.g medical data)

- When looking at the data flow through the project, define what data is:
 - aggregated (typically safe to share, if names cannot be recovered)
 - anonymized (name cannot be recovered from the data)
 - pseudonymized (name can be recovered by some)
 - non-anonymized (name linked to data)
- Determine which organisational boundaries have to be traversed by which data.
- Make sure with your "local" data protection officer and ethics commission that the data can be shared with your partners along the flow described with the anonymisation levels as described. Why local? Some laws change across surprising boundaries. E.g. in Germany Universities and other public organisations are subject to another data protection law than enterprises. Why seek advice? In some cases you may be required to be able to recover the name-data-relation, e.g. to enable study participants to "leave" a study.

Data Management Checklist

<https://fair-dom.org/knowledgehub/data-management-checklist/>

Data Management Checklist

Making data findable (documentation and metadata management)

- What documentation and metadata will accompany the data (assist its discoverability)? (Details on methodology, definitions, procedures, SOPs, vocabularies, units, dependencies, etc)
- What information is needed for the data to be read and interpreted in the future?
- What naming conventions will be used?
- How will you approach versioning your data?
- How will you capture / create this documentation and metadata?
- How do you ensure the completeness of the captured data?

Making Data Accessible

- Specify which data will be made openly available taking into consideration
 - What ethics and legal compliance issues do you have if any? Do you need consent for data preservation and sharing? Do you have to protect certain data? Is any data sensitive?
 - Do you think you might have Intellectual Property Rights issues? Have you considered ownership of the data, licensing, restrictions on use?
 - Do you think you will need to embargo any data?
- How will you make the data available? (consider the platforms you will use: databases, repositories, etc)
- What methods or software tools are needed to access the data? You should list where the software can be obtained. You should also document how to use the software to access the data. The documentation should be as complete as possible, including examples. If you distribute your system, include the access software and its documentation as part of any distribution.
- If there are any restrictions on accessibility, how will you provide access?

Making Data Interoperable

- What standards (metadata vocabularies, formats, checklists) or methodologies will you use?
- How do you address data and model quality? What validation steps do you foresee?
- Will you use standardised vocabulary for all data types to allow inter-disciplinary interoperability?
- Where you can not use standardised vocabulary for all types of data, can you map to more commonly used ontologies?

Making data Re-usable

- How will you licence your data to permit the widest re-use possible?
- When will the data be made available for re-use? Does this include an embargo period? (if yes, please detail why)
- Which data will be available for re-use during/after the project? For data that is not re-usable, please detail why
- What are your data quality assurance processes?
- How long do you expect your data to remain re-usable?

TO BE CUSTOMIZED

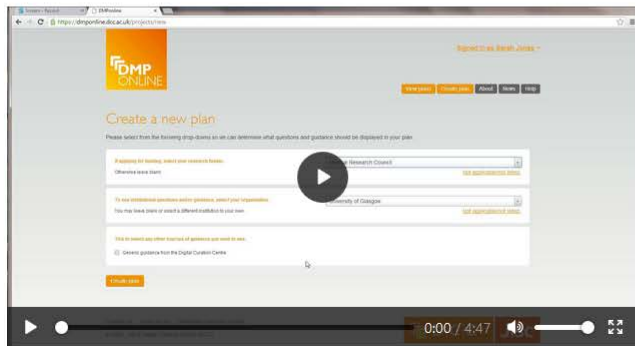
DMP online

[Home](#)[About](#)[Future plans](#)[Help](#)[Change language](#)

Welcome,

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It has been jointly developed by the Digital Curation Centre (DCC) and the University of California Curation Center (UC3).

Screencast on how to use DMPonline



<https://dmponline.dcc.ac.uk/>

Sign in

Veteran tapes

[Project Details](#)[Plan overview](#)[Write Plan](#)[Share](#)[Download](#)[expand all](#) | [collapse all](#)[13/13 answered](#)

Data Collection (2 / 2)

What data will you collect or create?

B *I*

The "Veteran tape " project will collect and generate different types of datasets:

Type of data	Volume	Format	Storage format
Video recordings	600 x 1Gb	.mkv	.mkv
Transcriptions	600 x 1500Kb	MS Word	.txt
Structured interview text	1 x 500Kb	MS word	.txt

For the video recordings the selected format is .mkv; the same .mkv format will be used for the long-term preservation .

Transcriptions will be written in MS Word and then stored as .txt files.

We checked the format compatibility against EASY File format
<https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats>

As the total volume of data is greater than 50Gb, DANS requires a fee for the storage. We are currently in touch with EASY to determine the costs of archiving.

[Save](#)[Guidance](#)[Comments \(1\)](#)[DCC](#)[DCC guidance](#)

Guidance

Questions to consider:

- What type, format and volume of data?
- Do your chosen formats and software enable sharing and long-term access to the data?
- Are there any existing data that you can reuse?

Guidance:

Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.

2. Data Management Plan – general definition

Data Management Plans (DMPs) are a *key element* of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

A DMP is required for all projects participating in the extended ORD pilot, unless they opt out of the ORD pilot. However, projects that opt submit a DMP on a voluntary basis.

DELIVERABLE
DUE IN M6



H2020 Programme

Guidelines on
FAIR Data Management in Horizon 2020

Guide 2016

Version 3.0
26 July 2016

4. Research data management plans during the project life cycle

Once a project has had its funding approved and has started, you **must submit a first version of your DMP** (as a deliverable) within the first 6 months of the project. The Commission provides a DMP template in annex, the use of which is recommended but voluntary.

The DMP needs to be **updated** over the course of the project whenever significant changes arise, such as (but not limited to):

- new data
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)
- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

The DMP should be updated as a minimum in time with the periodic evaluation/assessment of the project. If there are no other periodic reviews foreseen within the grant agreement, then such an update needs to be made in time for the final review at the latest. Furthermore, the consortium can define a timetable for review in the DMP itself.

Periodic reporting

For general information on periodic reporting please check the following sections of the online manual

- How to fill in [reporting tables for publications, deliverables](#)
- [Process for continuous reporting](#) in the grant management system.

Open access to data

GRANT
AGREEMENT
ART. 29.3
pag. 248



3. Open access to research data (Extended Open Research Data Pilot)

What?

Beneficiaries of actions that participate in the Open Research Data Pilot must give **open, free-of-charge access** to the end-user to **digital research data** generated during the action (⚠️ **new in Horizon 2020**).

⚠️ As of the Work Programme 2017, the Open Research Data pilot has been extended to all thematic areas of Horizon 2020 (except ERC PoC actions, SME instrument Ph1 actions, ERA-NET Cofund actions that do not produce data, EJP Cofund actions, and prizes).

Participation is therefore now in principle **the default**. However, actions may **opt out** at any stage — both before signing the GA and afterwards (through an amendment; *see Article 55*) —, if:

- participation is incompatible with the obligation to protect results (*see Article 27*)
- participation is incompatible with the security obligations (*see Article 37*)
- participation is incompatible with rules on protection of personal data
- participation would mean that the project's main aim might not be achieved
- the project will not generate/collect any research data or
- there are other legitimate reasons not to take part.

- PILOT PROJECT EXTENDED TO ALL DISCIPLINES
- OPT OUT CLAUSES

PRINCIPLE: «AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY»

Costs – data curation



Costs related to the implementation of the Open Research Data pilot (e.g. costs for providing open access, related research data management costs, data curation and data storage costs) may be eligible (see [Article 6.2.D.3](#)).

- ELIGIBLE COSTS:
- DATA CURATION
 - DATA STORAGE
 - DATA MANAGEMENT

Open data – H2020

How?

Open access to digital research data involves 3 steps:

Procedure for open access (research data):

Step 1 — **Deposit** the digital research data, preferably in a research data repository.

Step 2 — Provide **open access** by taking measures to enable users to access, mine, exploit, reproduce and disseminate the data free of charge (*e.g. for databases: by attaching an appropriate creative commons licence (CC-BY or CC0 tool) to the data; if the access/use is not subject to any rights: by indicating that no licence is needed*).

Open access must not be given immediately; for data needed to validate the results presented in scientific publications, as soon as possible; for other data, beneficiaries are free to specify embargo periods for their data in the data management plan (as appropriate in their scientific area).

Step 3 — Provide **information**, via the repository, about **tools and instruments** for validating the results.

Where possible, the beneficiaries should provide those tools and instruments (*e.g. specialised software or software code, algorithms, analysis protocols, etc.*).



- 1) DEPOSIT IN A DATA REPOSITORY
- 2) OPEN ACCESS TO DATA ENABLING TDM, EXPLOIT
- 3) POSSIBLE EMBARGO
- 4) INFORMATION TO VALIDATE

ONE DAY OR
DAY ONE
you decide.

THANK YOU!