

# Research Data Management and Open Access to publications and data

Iryna Kuchma, Open Access Programme Manager



Attribution 4.0 International



# Corona Virus Disease

COVID-19

[Subscribe](#)

Research outcomes

Search by title, author, abstract, DOI, orcid...

SEARCH

[Advanced Search](#)

## SUMMARY

PUBLICATIONS

**148,938**

RESEARCH DATA

**8,074**

SOFTWARE

**583**

OTHER RESEARCH

**8,221**

This portal provides access to publications, research data, projects and software that may be relevant to the Corona Virus Disease (COVID-19). The OpenAIRE COVID-19 Gateway aggregates COVID-19 related records, links them and provides a single access point for discovery and navigation. We tag content from the OpenAIRE Research Graph (10,000+ data sources) and additional sources. All COVID-19 related research results are linked to people, organizations and projects, providing a contextualized navigation.

Curated by: [Alessia Bardi](#), [Iryna Kuchma](#), [Evgeny Bobrov](#), [Ivana Truccolo](#), [Elizabete Monteiro](#)

...

[View all 8 curators](#)

Created: 16-Mar-2020 Members: 37

Projects: [130](#) Content Providers: [15](#)

Linked to [7](#) Zenodo Communities

<https://covid-19.openaire.eu>

## Subjects

[COVID19](#), [SARS-CoV](#), [HCoV-19](#), [mesh:C000657245](#), [MERS-CoV](#), [Síndrome Respiratorio Agudo Severo](#), [mesh:COVID-19](#), [COVID2019](#), [COVID-19](#), [SARS-CoV-2](#), [2019 novel coronavirus](#), [severe acute respiratory syndrome coronavirus 2](#), [Orthocoronavirinae](#), [Coronaviridae](#), [mesh:D045169](#), [coronavirus](#), [SARS](#), [coronaviruses](#), [coronavirus disease-19](#), [sars cov 2](#)

[View all](#)



### Filters

X Open Access

Access Mode (1) Clear

Open Access (118,091)

Research Type (4)

- Publications
- Research data
- Software
- Other research products

Year range

e.g. 1800 - e.g. 2030

THIS YEAR | LAST 5 YEARS | LAST 10 YEARS

Funder (26)

- National Institutes of H... (2,785)
- European Commission (1,197)
- Wellcome Trust (830)
- Canadian Institutes of He... (417)
- Research Council UK (377)
- National Science Foundation (349)

Top 100 values are shown in the filters

Search...

Sort by:  
Name

Results per page:

10

Sort by:

Relevance

Download Results

118,091 RESEARCH OUTCOMES, PAGE 1 OF 11,810

1 2 3 4 5 >

Subscribe

Publication . Article . Other Literature Type . 2005

### Angiotensin-converting enzyme 2 protects from severe acute lung failure

OPEN ACCESS ENGLISH

Authors: Imai, Yumiko; Kuba, Keiji; Rao, Shuan; Huan, Yi; Guo, Feng; Guan, Bin; Yang, Peng; Sarao, Renu; Wada, Teiji; Leong-Poi, Howard; ...

DOI: 10.1038/nature03712

Publisher: Nature Publishing Group UK

Project: CIHR

Acute respiratory distress syndrome (ARDS), the most severe form of acute lung injury, is a devastating clinical syndrome with a high mortality rate (30–60%) (refs 1–3). Predisposing factors for ARDS are diverse<sup>1,3</sup> and include sepsis, aspiration, pneumonias and infectio...

Publication . Article . 2020

### Africa Against COVID-19: Can National Solidarity Complements Health System Weaknesses? -the Moroccan Experience-

OPEN ACCESS

Authors: Amine Souadka; Amine Benkabbou; Mohammed Anass Majbar; Hajar Essangri; Laila Amrani; Raouf Mohsine; EL Houcine Akhnif;

Publisher: Medical and Surgical Research Journals Group (MSRJGroup)

<jats:p>NA</jats:p>



### Type (34)

- Article (94,927)
- Other literature type (24,986)
- Preprint (11,021)
- Dataset (2,670)
- Other ORP type (1,923)
- Report (1,365)

Top 100 values are shown in the filters

Sort by:  
Re... ▼

- Article (94,927)
- Other literature type (24,986)
- Preprint (11,021)
- Dataset (2,670)
- Other ORP type (1,923)
- Report (1,365)

- View less

### Language (49)

- English (52,854)
- Spanish; Castilian (1,363)

Publication . Article . 2020

## SARS-CoV-2: Naso-bronchial cytological correlations.



OPEN ACCESS ENGLISH

Authors: Gelardi, Matteo; Notargiacomo, Mario; Trecca, Eleonora M.C.; Cassano, Michele;

DOI: [10.1016/j.amjoto.2020.102549](https://doi.org/10.1016/j.amjoto.2020.102549)

Publisher: Elsevier BV

Publication . Article . Other Literature Type . 2017

## Tuberculosis in the elderly: Why inflammation matters

OPEN ACCESS

Authors: Piergallini, Tucker J; Turner, Joanne;

DOI: [10.1016/j.exger.2017.12.021](https://doi.org/10.1016/j.exger.2017.12.021)

Publisher: Elsevier BV

Abstract Growing old is associated with an increase in the basal inflammatory state of an individual and susceptibility to many diseases, including infectious diseases. Evidence is growing to support the concept that inflammation and disease...

- WHAT IS REQUIRED?
- HOW TO COMPLY?
- WHICH REPOSITORY TO USE?
- WHEN TO PROVIDE ACCESS?
- ARE PUBLICATION COSTS SUPPORTED?
- START PLANNING EARLY ON!
- HOW CAN OPENAIRE HELP?
- VIEW OUR WEBINARS RECORDINGS
- COVID-19 GUIDELINES

## COVID-19 guidelines

### Guidelines for publications

- Make all research publications relevant to the outbreak immediately available, through deposition of a copy of the published, or final, peer-reviewed version, in a repository (through which open access to the deposited copy shall be ensured), at the latest at the time of publication, under a Creative Commons Attribution 4.0 International Public License (CC BY 4.0) or a license with equivalent rights.
- Make research findings available via preprint servers before journal publication, or via platforms that make publications openly accessible before peer-review. Include clear statements regarding the availability of underlying data. Some reliable and currently very relevant preprint archives are [bioRxiv](#) (life sciences), [medRxiv](#) (medical), [PsyArxiv](#) (behavioural sciences), [SocArXiv](#) (social sciences), [ArXiv](#) (o.a. physics, mathematics, computer science) and [Open Science Framework \(OSF\) preprints](#) or [Zenodo](#) (the latter two are multidisciplinary archives).
- Provide information via the repository about any research output or any other tools and instruments needed to re-use and/or validate the conclusions of the scientific publication. This includes for example software, workflows, models, materials etc. If possible, provide access to the tools or instruments themselves.
- Include metadata of deposited publications under a Creative Commons Public Domain Dedication (CC 0 1.0) or equivalent, in line with the FAIR principles (in particular machine actionable) and provide information at least about the following:

## Support

### RESOURCES

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

### HELPDESK

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

### TRAINING

- [Webinars](#)
- [Workshops](#)
- [Community of Practice](#)
- [Content Providers Community Calls](#)

## Coronavirus Disease Research Community - COVID-19

### Recent uploads

Search Coronavirus Disease Research Community - COVID-19



July 15, 2020 (v1)

Journal article

Open Access

View

#### Feministki w sieci

Dominika Urzędowska; Joanna Suchomska;

Artykuł podejmuje kwestie funkcjonowania nowych ruchów społecznych na przykładzie działań podejmowanych w obronie praw reprodukcyjnych i dostępu do edukacji seksualnej w Polsce w latach 2016-2020. Szczególną uwagę zwracamy na rolę nowych mediów w kształtowaniu się, organizacji i

Uploaded on October 19, 2020

October 18, 2020 (v 1.1)

Presentation

Open Access

View

#### applied humanities in a data enabled society

Wandl-Vogt, Eveline;

Invited keynote for the Data and Digital Humanities Conference 2020 (online). Introducing the concept of applied humanities, referring to the ALLEA discussion paper "flourishing in a data enabled society" and giving examples of exploration space at the Austrian Academy of Sciences and the A

Uploaded on October 18, 2020

October 18, 2020 (32)

Dataset

Open Access

View

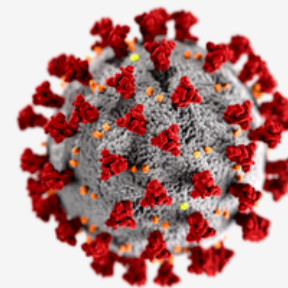
#### A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration

Banda, Juan M.; Tekumalla, Ramya; Wang, Guanyu; Yu, Jingyuan; Liu, Tuo; Ding, Yuning; Artemova, Katya; Tutubalina, Elena; Chowell, Gerardo;

Version 32 of the dataset. In version 30 we added a few additional historical tweets in Russian provided by our coauthors. Due to the relevance of the COVID-19 global pandemic, we are releasing our dataset of tweets acquired from the Twitter

New upload

Community



#### Coronavirus Disease Research Community - COVID-19

This community collects research outputs that may be relevant to the Coronavirus Disease (COVID-19) or the SARS-CoV-2. Scientists are encouraged to upload their outcome in this collection to facilitate sharing and discovery of information. Although Open Access articles and datasets are recommended, also closed and restricted access material are accepted. All types of research outputs can be included in this Community (Publication, Poster, Presentation, Dataset, Image, Video/Audio, Software, Lesson, Other).

Read more

October 18, 2020

Dataset **Open Access**

# A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration

Banda, Juan M.; Tekumalla, Ramya; Wang, Guanyu; Yu, Jingyuan; Liu, Tuo; Ding, Yuning; Artemova, Katya; Tutubalina, Elena; Chowell, Gerardo

**Version 32 of the dataset.** In version 30 we added a few additional historical tweets in Russian provided by our coauthors.

Due to the relevance of the COVID-19 global pandemic, we are releasing our dataset of tweets acquired from the Twitter Stream related to COVID-19 chatter. Since our first release we have received additional data from our new collaborators, allowing this resource to grow to its current size. Dedicated data gathering started from March 11th yielding over 4 million tweets a day. We have added additional data provided by our new collaborators from January 27th to March 27th, to provide extra longitudinal coverage. Version 10 added ~1.5 million tweets in the Russian language collected between January 1st and May 8th, gracefully provided to us by: Katya Artemova (NRU HSE) and Elena Tutubalina (KFU). From version 12 we have included daily hashtags, mentions and emojis and their frequencies the respective zip files. From version 14 we have included the tweet identifiers and their respective language for the clean version of the dataset. Since version 20 we have included language and place location for all tweets.

The data collected from the stream captures all languages, but the higher prevalence are: English, Spanish, and French. We release all tweets and retweets on the full\_dataset.tsv file (762,035,280 unique tweets), and a cleaned version with no retweets on the full\_dataset-clean.tsv file (182,857,703 unique tweets). There are several practical reasons for us to leave the retweets, tracing important tweets and their dissemination is one of them. For NLP tasks we provide the top 1000 frequent terms in frequent\_terms.csv, the top 1000 bigrams in frequent\_bigrams.csv, and the top 1000 trigrams in frequent\_trigrams.csv. Some general statistics per day are included for

37,769

views

30,809

downloads

[See more details...](#)

Indexed in

OpenAIRE

**Publication date:**

October 18, 2020

**DOI:**DOI [10.5281/zenodo.4104587](https://doi.org/10.5281/zenodo.4104587)**Keyword(s):**[social media](#) [twitter](#) [nlp](#) [covid-19](#)  
[covid19](#)**Related identifiers:**Continued by  
<http://www.panacealab.org/covid19/>  
(Other)Supplement to  
<https://arxiv.org/abs/2004.03688>

October 17, 2020

Dataset **Open Access**

# BIP4COVID19: Impact metrics and indicators for coronavirus related publications

Thanasis Vergoulis; Ilias Kanellos; Serafeim Chatzopoulos; Danae Pla Karidi; Theodore Dalamagas

This dataset contains impact metrics and indicators for a set of publications that are related to the [COVID-19 infectious disease](#) and the coronavirus that causes it. It is based on:

1. The [CORD-19 dataset](#) released by the team of [Semantic Scholar](#)<sup>1</sup> and
2. The curated data provided by the [LitCovid hub](#)<sup>2</sup>.

These data have been cleaned and integrated with data from [COVID-19-TweetIDs](#) and from other sources (e.g., PMC). The result was dataset of 189,733 unique articles along with relevant metadata (e.g., the underlying citation network). We utilized this dataset to produce, for each article, the values of the following impact measures:

- **Influence:** Citation-based measure reflecting the total impact of an article. This is based on the PageRank<sup>3</sup> network analysis method. In the context of citation networks, it estimates the importance of each article based on its centrality in the whole network. This measure was calculated using the PaperRanking (<https://github.com/diwis/PaperRanking>) library<sup>4</sup>.
- **Popularity:** Citation-based measure reflecting the current impact of an article. This is based on the RAM<sup>5</sup> citation network analysis method. Methods like PageRank are biased against recently published articles (new articles need time to receive their first citations). RAM alleviates this problem using an approach known as "time-awareness". This is why it is more suitable to capture the current "hype" of an article. This measure was calculated using the PaperRanking (<https://github.com/diwis/PaperRanking>) library<sup>4</sup>.
- **Social Media Attention:** The number of tweets related to this article. Relevant data were collected from the [COVID-19-TweetIDs](#) dataset. In this version, only tweets between Jan 21<sup>st</sup>

22,376

views

5,597

downloads

[See more details...](#)

Indexed in

**Publication date:**

October 17, 2020

**DOI:**

DOI [10.5281/zenodo.4099186](https://doi.org/10.5281/zenodo.4099186)

**Keyword(s):**

[COVID-19](#) [coronavirus](#) [scientometrics](#)  
[bibliometrics](#)

**Related identifiers:**

Cites  
<https://pages.semanticscholar.org/coronavirus-research> (Dataset)  
<https://github.com/diwis/PaperRanking> (Software)





Search



Upload

Communities

Log in

Sign up

April 20, 2020

Dataset **Open Access**

# COVID-19 CT Lung and Infection Segmentation Dataset

[ORCID](#) Ma Jun; [ORCID](#) Ge Cheng; [ORCID](#) Wang Yixin; [ORCID](#) An Xingle; [ORCID](#) Gao Jiantao; [ORCID](#) Yu Ziqi; [ORCID](#) Zhang Mingqing; [ORCID](#) Liu Xin; [ORCID](#) Deng Xueyuan; [ORCID](#) Cao Shucheng; [ORCID](#) Wei Hao; [ORCID](#) Mei Sen; [ORCID](#) Yang Xiaoyu; [ORCID](#) Nie Ziwei; [ORCID](#) Li Chen; Tian Lu; [ORCID](#) Zhu Yuntao; Zhu Qiongjie; Dong Guoqiang; He Jian

This dataset contains 20 labeled COVID-19 CT scans. Left lung, right lung, and infections are labeled by two radiologists and verified by an experienced radiologist.

To promote the studies of annotation-efficient deep learning methods, we set up three segmentation benchmark tasks based on this dataset <https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark>.

In particular, we focus on learning to segment left lung, right lung, and infections using

- pure but limited COVID-19 CT scans;
- existing labeled lung CT dataset from other non-COVID-19 lung diseases;
- heterogeneous datasets include both COVID-19 and non-COVID-19 CT scans.

Preview

COVID-19-CT-Seg\_20cases.zip

ReadMe.txt	359 Bytes
coronacases_001.nii.gz	93.2 MB
coronacases_002.nii.gz	70.1 MB
coronacases_003.nii.gz	67.3 MB
coronacases_004.nii.gz	88.9 MB
coronacases_005.nii.gz	96.5 MB
coronacases_006.nii.gz	71.6 MB
coronacases_007.nii.gz	85.0 MB

12,541

views

15,240

downloads

[See more details...](#)

Indexed in

OpenAIRE

**Publication date:**

April 20, 2020

**DOI:**

DOI [10.5281/zenodo.3757476](https://doi.org/10.5281/zenodo.3757476)

**Keyword(s):**

COVID-19 Segmentation Lung Infection  
Computed tomography

**Related identifiers:**

Derived from <https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark> (Dataset)

**Communities:**

[Coronavirus Disease Research](#)



Search



Upload

Communities

Log in

Sign up

October 6, 2020

Dataset **Open Access**

# Novel Coronavirus (COVID-19) Cases in The Netherlands

De Bruin, J

Data collector(s)

Phil, T

Project leader(s)

De Bruin, J

Project member(s)

Voorwaard, R; Hoogerwerf, M

Researcher(s)

Menger, V; Kocken, I

On 27 February 2020, the first case of COVID-19 disease was confirmed in The Netherlands by RIVM (National Institute for Public Health and the Environment). In the weeks after, thousands of people were diagnosed with the infectious disease. Data on COVID-19 case counts are important for research and applications on various topics like epidemiology and statistics.

This dataset contains reported case counts derived from official sources like RIVM (National Institute for Public Health and the Environment), LCPS (National Coordination Center for Patient Distribution), and NICE (National Intensive Care Evaluation). Data from these sources are collected, standardized, and published in various formats on a daily basis.

The README document in this repository provides an overview of the available datasets, their file location(s), and codebooks. Copies of the original data are stored in the folder named 'raw\_data'. Scripts to process the raw data into standardized files can be found in the folder workflows.

9,368

views

2,488

downloads

[See more details...](#)

Indexed in

**Publication date:**

October 6, 2020

**DOI:**DOI [10.5281/zenodo.4068121](https://doi.org/10.5281/zenodo.4068121)**Keyword(s):**

[covid-19](#) [corona](#) [coronavirus](#)  
[netherlands](#) [dataset](#) [timeseries](#)  
[pandemic](#) [rivm](#) [sars-cov-2](#) [nederland](#)

**Communities:**

[Coronavirus Disease Research](#)  
[Community - COVID-19](#)  
[Zenodo](#)

**License (for files):**



Search

Upload Communities

Log in

Sign up

August 13, 2020

Dataset Open Access

# Linked COVID-19 Data: Robert Koch-Institut (RKI)

Florian Thiery

Linked COVID-19 Data derived from

**Robert Koch Institute**

using the COVID-19 Ontology

[10.5281/zenodo.3757828](https://doi.org/10.5281/zenodo.3757828)

developed for the Linked COVID-19 Data Dashboard: <http://covid19data.link>

8,465

views

4,218

downloads

[See more details...](#)

Indexed in

OpenAIRE

This files include data for

- covid19\_rki\*.ttl - single COVID-19 cases per day, collected by the RKI
  - <https://www.arcgis.com/home/item.html?id=dd4580c810204019a7b8eb3e0b329dd6>
  - "confirmed" is calculated by the column "Refdatum"
- covid19\_rki\_ger.ttl - COVID-19 cases per day for Germany, collected by the RKI
- covid19\_rki\_ger\_cum.ttl - COVID-19 cases accumulated per day for Germany, collected by the RKI
- covid19\_rki\_fs.ttl - COVID-19 cases per day for the federal states in Germany, collected by the RKI
- covid19\_rki\_fs\_cum.ttl - COVID-19 cases accumulated per day for the federal states in Germany, collected by the RKI

This RDF files are based on the NPGEO Corona Hub 2020

- [https://opendata.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6\\_0.geojson](https://opendata.arcgis.com/datasets/dd4580c810204019a7b8eb3e0b329dd6_0.geojson)
- <https://npggeo-corona-npggeo-de-hub.arcgis.com/datasets>

### Publication date:

August 13, 2020

### DOI:

DOI [10.5281/zenodo.3983086](https://doi.org/10.5281/zenodo.3983086)

### Keyword(s):

Linked Data COVID-19

Robert-Koch Insitute

### Subject(s):

COVID-19

2019-20 COVID-19 pandemic

2020 coronavirus pandemic in Germany

Linked COVID-19 Data

### Related identifiers:

August 13, 2020

Dataset **Open Access**

# Linked COVID-19 Data: Johns Hopkins University (JHU) and European Centre for Disease Prevention and Control (ECDC)

Florian Thiery

Linked COVID-19 Data derived from

**Johns Hopkins University**

and

**European Centre for Disease Prevention and Control**

using the COVID-19 Ontology

[10.5281/zenodo.3757828](https://doi.org/10.5281/zenodo.3757828)developed for the Linked COVID-19 Data Dashboard: <http://covid19data.link>

This files include data for

- covid19\_jhu.ttl - COVID-19 data collected by the JHU
- covid19\_ecdc.ttl - COVID-19 data collected by the ECDC

This RDF files are based on

- <https://pomber.github.io/covid19/timeseries.json>
- <https://opendata.ecdc.europa.eu/covid19/casedistribution/json/>

7,757

views

808

downloads

[See more details...](#)

Indexed in

**Publication date:**

August 13, 2020

**DOI:**DOI [10.5281/zenodo.3983085](https://doi.org/10.5281/zenodo.3983085)**Keyword(s):**

Linked Data COVID-19

Johns Hopkins University

European Centre for Disease Prevention and Control

**Subject(s):**

COVID-19

2019-20 COVID-19 pandemic

Linked COVID-19 Data

September 17, 2020

Dataset Open Access

# Crowdsourced air traffic data from The OpenSky Network 2020

Xavier Olive; Martin Strohmeier; Jannis Lübbe

## Motivation

The data in this dataset is derived and cleaned from the full OpenSky dataset to illustrate the development of air traffic during the COVID-19 pandemic. It spans all flights seen by the network's more than 2500 members since 1 January 2019. More data will be periodically included in the dataset until the end of the COVID-19 pandemic.

## License

See LICENSE.txt

## Disclaimer

The data provided in the files is provided as is. Despite our best efforts at filtering out potential issues, some information could be erroneous.

- Origin and destination airports are computed online based on the ADS-B trajectories on approach/takeoff: no crosschecking with external sources of data has been conducted. Fields **origin** or **destination** are empty when no airport could be found.
- Aircraft information come from the OpenSky aircraft database. Fields **typecode** and **registration** are empty when the aircraft is not present in the database.

## Description of the dataset

One file per month is provided as a csv file with the following features:

- **callsign**: the identifier of the flight displayed on ATC screens (usually the first three letters are reserved for an airline: AFR for Air France, DLH for Lufthansa, etc.)
- **number**: the commercial number of the flight when available (the matching with the callsign

6,983

views

6,117

downloads

See more details...

Indexed in

OpenAIRE

## Publication date:

September 17, 2020

## DOI:

DOI 10.5281/zenodo.4088202

## Keyword(s):

aviation flightlist ads-b covid19

## Related identifiers:

Documented by  
<https://traffic-viz.github.io/scenarios/covid19.html> (Software)

Supplement to  
10.1109/IPSJ.2014.6846743 (Journal article)

Clinical



Omics



Epidemiology



Social Sciences



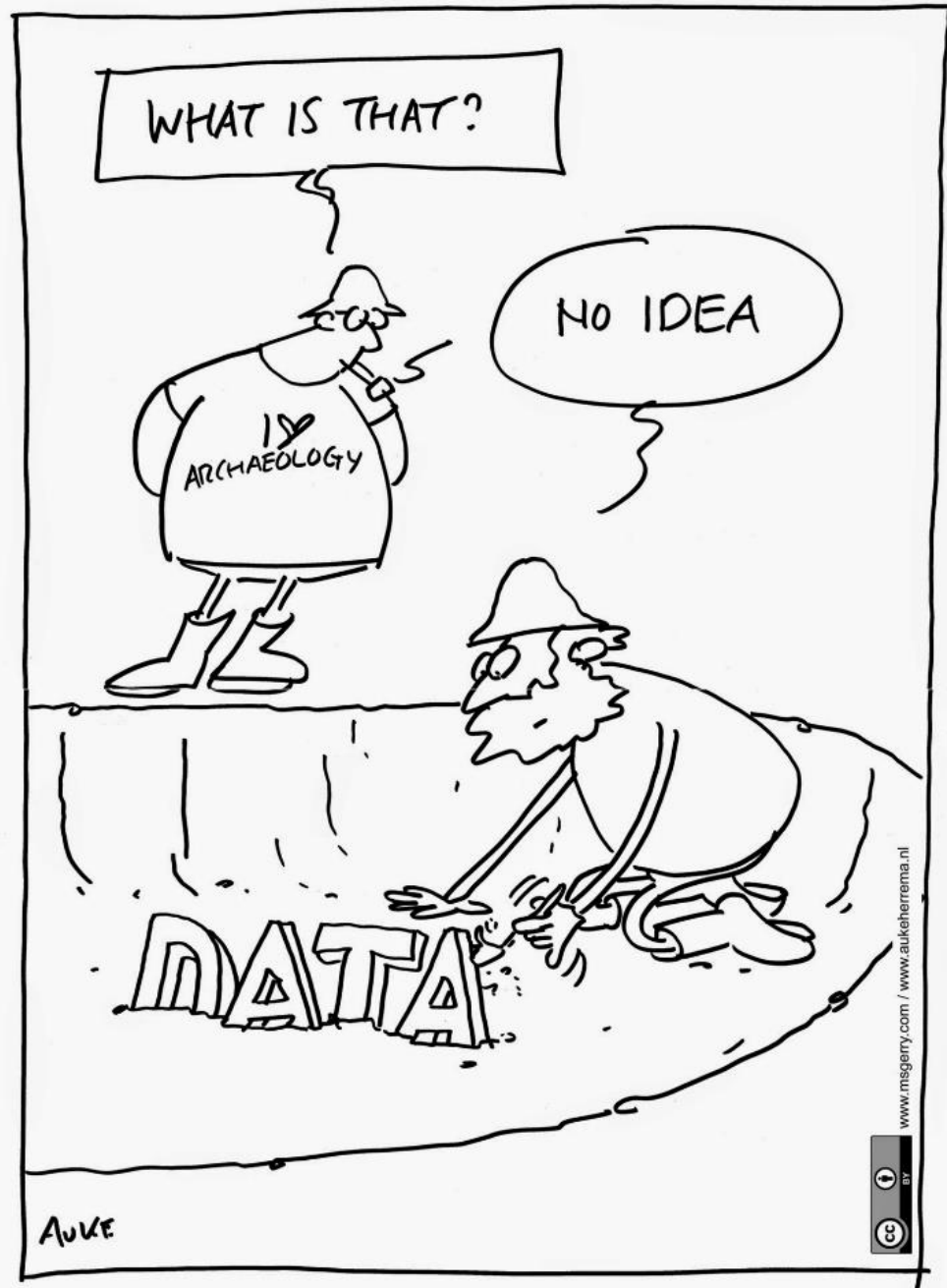
Community Participation for Data Sharing under COVID-19

Indigenous Data under COVID-19

Legal and Ethical Considerations under COVID-19

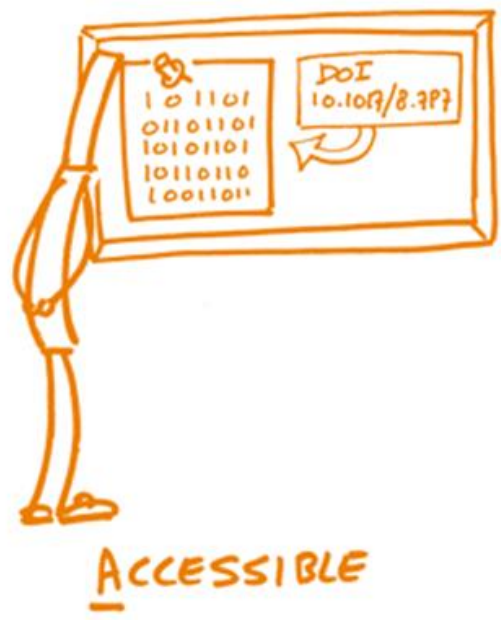
Research Software and Data Sharing under COVID-19

- April 1st through June 30th, 2020
- Weekly update webinar and writing sprint
- Over 600 data professional experts
- 4 Research domains and 4 cross cutting areas



DATA FOR FUTURE GENERATIONS

# FAIR DATA PRINCIPLES







# Levels of openness

**Open data** - the Open Data Institute (ODI) defines Open Data as those that anyone can access, use and share. According to the ODI, open data must be licensed to make clear that anyone can use the data in any way they want, including transforming, combining, and sharing it with others, even for commercial purposes. The ODI provides a great introduction to all aspects of Open Data in their [Open Data Essentials course](#). We highly recommend reviewing these modules.

**Shared data** - similar to Open data, shared data may be made widely accessible but could have some conditions such as non-commercial reuse or reuse with attribution. It is important to note that not all shared data has to be available to anyone. Sometimes shared data is only made available to specific groups such as peers from another university.

**Closed data** - if researchers are dealing with highly sensitive data - such as sensitive personal data or commercially sensitive data - it may not be possible to share the data at all. However, even in such cases a metadata description of the research data should be shared. Sharing of sensitive data can also be supported by making use of safe havens where only authorised users are given controlled access.

# Definition of Open Data

Open Data are online, free of cost, accessible data that can be used, reused and distributed provided that the data source is attributed.

# Tip - use 5 Star Open Data Model to explain FAIR



make your stuff available on the Web (whatever format) under an open license



make it available as structured data (e.g., Excel instead of image scan of a table)



make it available in a non-proprietary open format (e.g., CSV instead of Excel)

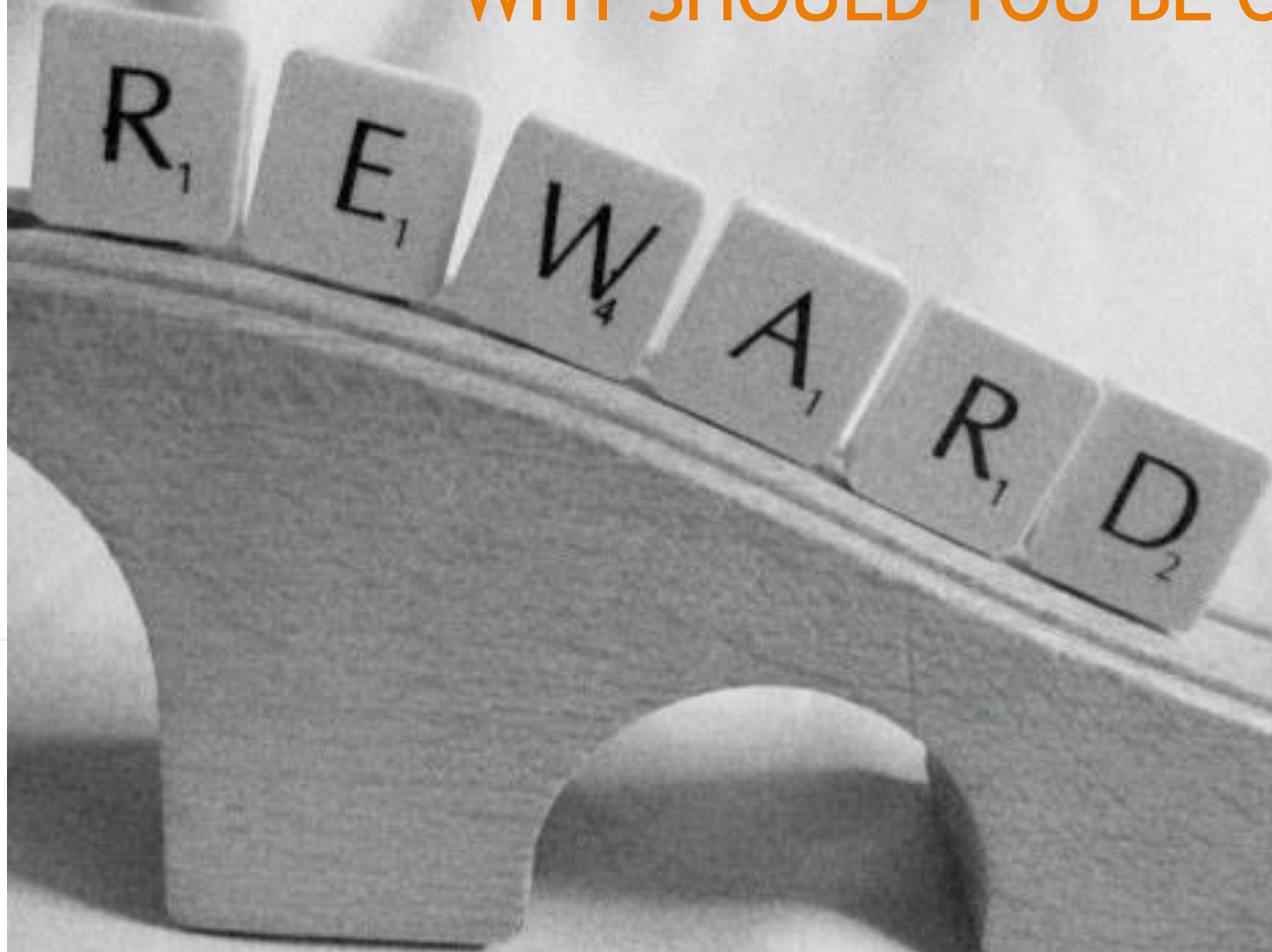


use URIs to denote things, so that people can point at your stuff



link your data to other data to provide context

# WHY SHOULD YOU BE OPEN?



# It's part of good research practice

"It was *\*never\** acceptable to publish papers without making data available."

- Ewan Birney

#OpenData  
#OpenScience



Original image via doi:10.1038/461145a. "Research cannot flourish if data are not preserved and made accessible. Data management should be woven into every course in science." - *Nature* 461, 145

# Cut down on academic fraud

**nature** International weekly journal of science Login

[nature news home](#) [news archive](#) [specials](#) [opinion](#) [features](#) [news blog](#) [nature journal](#)

[comments on this story](#)

Published online 1 November 2011 | *Nature* **479**, 15 (2011) | doi:10.1038/479015a  
[Updated](#) online: 1 November 2011  
[Updated](#) online: 8 December 2011

**News**

## Report finds massive fraud at Dutch universities

### Investigation claims dozens of social-psychology papers contain faked data.

Ewen Callaway

When colleagues called the work of Dutch psychologist Diederik Stapel too good to be true, they meant it as a compliment. But a preliminary investigative report ([go.nature.com/tqmp5c](http://go.nature.com/tqmp5c)) released on 31 October gives literal meaning to the phrase, detailing years of data manipulation and blatant fabrication by the prominent Tilburg University researcher.



Dutch psychologist Diederik Stapel.  
*Persbureau van Eindhoven*

"We have some 30 papers in peer-reviewed journals where we are actually sure that they are fake, and there are more to come," says Pim Levelt, chair of the committee that investigated Stapel's work at the university.

Stapel's eye-catching studies on aspects of social behaviour such as power and stereotyping garnered wide press coverage. For example, in a recent *Science* paper (which the investigation has not identified as fraudulent), Stapel reported that untidy environments encouraged discrimination ([Science 332, 251-253; 2011](#)).

#### Related stories

- [Seven days: 9-15 September 2011](#)  
14 September 2011
- [Chaos promotes stereotyping](#)  
07 April 2011

#### Naturejobs

**Tenure-Track Faculty Positions (Assistant / Associate / Full Professor) Yale University, Department of Genetics**  
Yale University School of Medicine

**Assistant Professor**  
Harvard Medical School

- [More science jobs](#)
- [Post a job for free](#)

#### Resources

- [PDF Format](#)
- [Send to a Friend](#)
- [Reprints & Permissions](#)
- [RSS Feeds](#)

#### external links

- [Tilburg University](#)
- [Interim investigation report](#)

#### Stories by subject

- [Brain and behaviour](#)
- [Lab life](#)

#### Stories by keywords

- [Diederik Stapel](#)
- [Tilburg University](#)
- [Academic fraud](#)
- [Retractions](#)
- [Social psychology](#)

#### This article elsewhere

[Blogs linking to this article](#)

- [Add to Digg](#)
- [Add to Facebook](#)
- [Add to Newsvine](#)
- [Add to Del.icio.us](#)
- [Add to Twitter](#)

# Validation of results

“It was a mistake in a spreadsheet that could have been easily overlooked: a few rows left out of an equation to average the values in a column.

The spreadsheet was used to draw the conclusion of an influential 2010 economics paper: that public debt of more than 90% of GDP slows down growth. This conclusion was later cited by the International Monetary Fund and the UK Treasury to justify programmes of austerity that have arguably led to riots, poverty and lost jobs.”

## The error that could subvert George Osborne's austerity programme

The theories on which the chancellor based his cuts policies have been shown to be based on an embarrassing mistake

Charles Arthur and Phillip Inman

The Guardian, Thursday 18 April 2013 21.10 BST



George Osborne says that Ken Rogoff, the man whose economic error has been uncovered, has strongly influenced his thinking. Photograph: Stefan Wermuth/PA

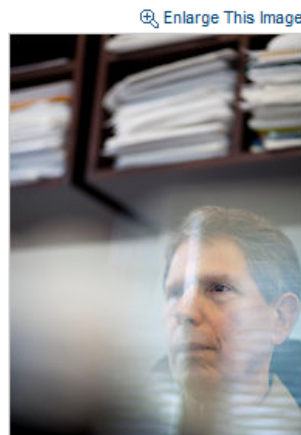
[www.guardian.co.uk/politics/2013/apr/18/uncovered-error-george-osborne-au](http://www.guardian.co.uk/politics/2013/apr/18/uncovered-error-george-osborne-au)

# More scientific breakthroughs

## Sharing of Data Leads to Progress on Alzheimer's

By GINA KOLATA  
Published: August 12, 2010

In 2003, a group of scientists and executives from the [National Institutes of Health](#), the [Food and Drug Administration](#), the drug and medical-imaging industries, universities and nonprofit groups joined in a project that experts say had no precedent: a collaborative effort to find the biological markers that show the progression of [Alzheimer's disease](#) in the human brain.



[Enlarge This Image](#)

Now, the effort is bearing fruit with a wealth of recent scientific papers on the early diagnosis of Alzheimer's using methods like PET scans and tests of spinal fluid. More than 100 studies are under way to test drugs that might slow or stop the disease.

And the collaboration is already serving as a model for similar efforts against [Parkinson's disease](#). A \$40 million project to look for biomarkers for Parkinson's, sponsored by the [Michael J. Fox Foundation](#), plans to enroll 600 study subjects in the United States and Europe.

*“It was unbelievable. Its not science the way most of us have practiced in our careers. But we all realised that we would never get biomarkers unless all of us parked our egos and intellectual property noses outside the door and agreed that all of our data would be public immediately.”*

*Dr John Trojanowski, University of Pennsylvania*

[www.nytimes.com/2010/08/13/health/research/13alzheimer.html?pagewanted=all&\\_r=0](http://www.nytimes.com/2010/08/13/health/research/13alzheimer.html?pagewanted=all&_r=0)



Open data linked to higher cita... X +

THE ROYAL SOCIETY OF CHE... (GB) https://www.chemistr... Search

NEWS RESEARCH OPINION FEATURES CULTURE CAREERS PODCASTS WEBINARS IYPT COLLECTIONS REGISTER

# Open data linked to higher citations for journal articles

NEWS

BY REBECCA TRAGER | 11 JULY 2019

SOURCE: © GETTY IMAGES

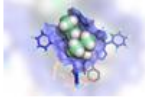



## Studies that provide access to underlying data are cited 25% more often than those that don't


Research papers that make their underlying data openly available are significantly more likely to be cited in future work, according to an analysis led by researchers at the Alan Turing Institute in London that has been [published as a preprint](#). The study, which is currently under peer review, examined nearly 532,000 articles in over 350 open access journals published by Public Library of Science (PLoS) and BioMed Central (BMC) between 1997 and 2018, and found those that linked directly to source data sets received 25% more citations on average.

<https://www.chemistryworld.com/news/open-data-linked-to-higher-citations-for-journal-articles/3010723.article>

LATEST POPULAR

 Molecular amplifier makes purest chiral Möbius aromatic

 Curbing industry's carbon emissions

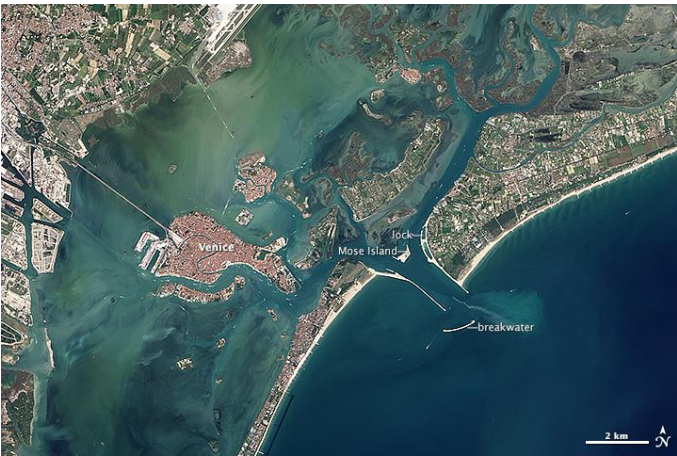
 Science scorecards reveal state of research in G20 nations

# Increased use and economic benefit

The case of NASA Landsat satellite imagery of the Earth's surface:

## Up to 2008

- Sold through the US Geological Survey for US\$600 per scene
- Sales of 19,000 scenes per year
- Annual revenue of \$11.4 million



## Since 2009

- Freely available over the internet
- Google Earth now uses the images
- Transmission of 2,100,000 scenes per year.
- Estimated to have created value for the environmental management industry of \$935 million, with direct benefit of more than \$100 million per year to the US economy
- Has stimulated the development of applications from a large number of companies worldwide

## BE PART OF THE NEW ERA OF OPEN SCIENCE

here's one example of the gains arising from open research data

### Bioinformatics Institute

**€1.3 billion** per year

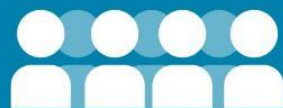
Benefits identified by the European Bioinformatics Institute to users and their funders just by making scientific information freely available to the global life science community...



equivalent to **more than 20 times** the direct operational cost of the Institute

Source: Charles Beagrie Ltd. for EMBL-EBI

## BE PART OF THE NEW ERA OF OPEN SCIENCE



reach more  
people,  
have greater  
impact



avoid  
duplication  
of efforts



preserve data  
for future  
researchers



simplify final  
Horizon 2020  
reporting  
thanks to an  
up-to-date DMP

# re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

Search

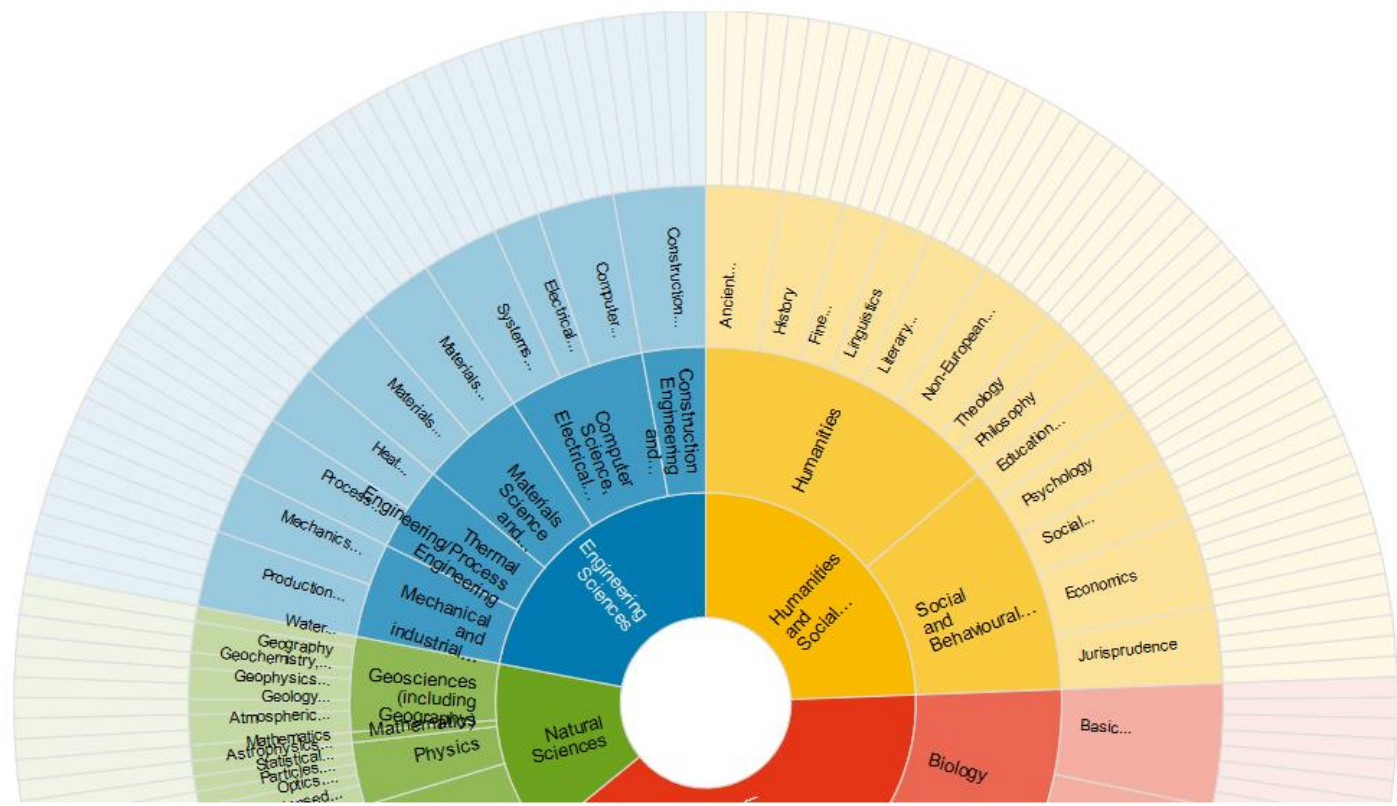


# Browse by subject

Graphical

Text

click to zoom into subjects or to select a bottommost subject in the hierarchy as filter for the re3data search page  
ctrl + click on a top subject to select it as filter



# Browse by subject

Graphical

Text

## A. Humanities and Social Sciences

### a. Humanities

#### I. Ancient Cultures

1. Prehistory
2. Classical Philology
3. Ancient History
4. Classical Archaeology
5. Egyptology and Ancient Near Eastern Studies

#### II. History

1. Medieval History
2. Early Modern History
3. Modern and Current History
4. History of Science

#### III. Fine Arts, Music, Theatre and Media Studies

1. Art History
2. Musicology
3. Theatre and Media Studies

#### IV. Linguistics

1. General and Applied Linguistics
2. Individual Linguistics
3. Typology, Non-European Languages, Historical Linguistics

#### V. Literary Studies

1. Medieval German Literature
2. Modern German Literature
3. European and American Literature
4. General and Comparative Literature and Cultural Studies

# Browse by content type

- [Archived data](#)
- [Audiovisual data](#)
- [Configuration data](#)
- [Databases](#)
- [Images](#)
- [Networkbased data](#)
- [Plain text](#)
- [Raw data](#)
- [Scientific and statistical data formats](#)
- [Software applications](#)
- [Source code](#)
- [Standard office documents](#)
- [Structured graphics](#)
- [Structured text](#)
- [other](#)

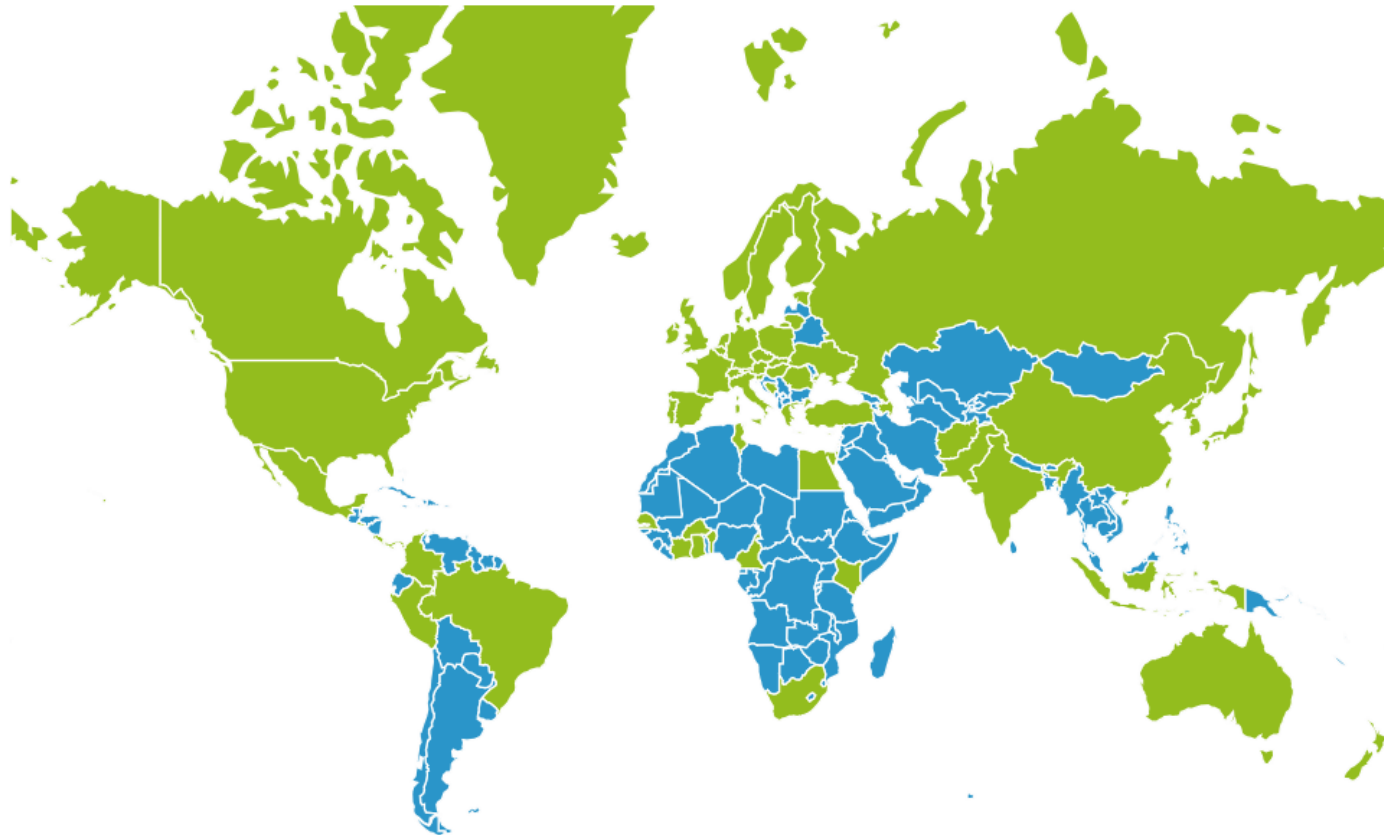




# Browse by country

Graphical

Text






# Browse by country

Graphical


Text

- International
-  Afghanistan
-  Australia
-  Austria
-  Azerbaijan
-  Belgium
-  Benin
-  Burkina Faso
-  Bosnia and Herzegovina
-  Brazil
-  Canada
-  Switzerland
-  China
-  Cote d'Ivoire
-  Cameroon
-  Colombia
-  Costa Rica
-  Cyprus
-  Czech Republic
-  Germany
-  Denmark
-  European Union
-  Egypt
-  Spain
-  Estonia

## Filter

[Reset all](#)Subjects Content Types Countries [Germany \(1\)](#)[Kazakhstan \(1\)](#)Data access Database access Database licenses Data licenses Data upload Enhanced publication Institution responsibility type Institution type Keywords Metadata standards PID systems Provider types Quality management Repository languages Software Repository types 

Search...

 Search[Toggle short help](#)[← Previous](#) **1** [Next →](#)

Sort by ▾

Found 1 result(s)

## Pyramids of the Steppe / Kazakhstan, Land of Seven Rivers



KUSS

Subject(s)

[History](#) [Ancient Cultures](#) [Humanities and Social Sciences](#) [Humanities](#)

Content type(s)

[Images](#) [Structured graphics](#) [Standard office documents](#) [Databases](#)

Country

[Kazakhstan](#) [Germany](#)

Atlas of the early iron age kurgans (burial mounds) of the Saks in the Land of Seven Rivers between the river Ili and the foothills of the Transili-Alatau as well as in the mountain valleys of the northern Tien-Shans.

# Open Data doesn't just happen - data management planning helps!

- What data will be created (format, types, volume...)
- Standards and methodologies to be used, documentation
- How ethics and Intellectual Property will be addressed
- Plans for storage and back-up
- Plans for data sharing and access
- Strategy for long-term preservation





Home

Public DMPs

Public Dataset Desc.

Co-Branding

Support

Send feedback

About Terms Of Service

Glossary User Guide

### What is ARGOS?

ARGOS is an open extensible service that simplifies the management, validation, monitoring and maintenance and of Data Management Plans. It allows actors (researchers, managers, supervisors etc) to create actionable DMPs that may be freely exchanged among infrastructures for carrying out specific aspects of the Data management process in accordance with the intentions and commitment of Data owners.



### Latest Activity

All (5)

Public DMPs (5)

Public Datasets (5)

Sort by: Published

Search

[nature](#) > [scientific data](#) > [comment](#) > [article](#)


[Open Access](#) | [Published: 15 March 2016](#)

## The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons 

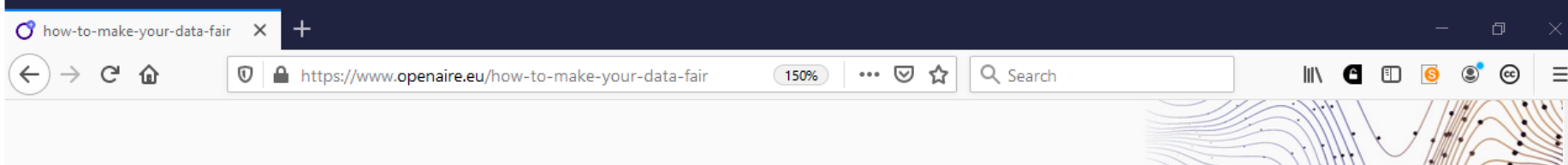
*Scientific Data* **3**, Article number: 160018 (2016) | [Cite this article](#)

**145k** Accesses | **1757** Citations | **1591** Altmetric | [Metrics](#)

 An [Addendum](#) to this article was published on 19 March 2019

### Abstract

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data



INTRODUCTION

WHAT IS FAIR DATA?

FAIR - IN DEPTH

HOW FAIR ARE YOUR DATA?

MORE RESOURCES

TRAINING MATERIALS

## FAIR - in depth

### Findable

Make your data findable by ensuring it:

- Has a persistent identifier
- Has rich metadata
- Is searchable and discoverable online

**Persistent identifiers** (PIDs) are important because they unambiguously identify your data and facilitate data citation. An example of a PID is a Digital Object Identifier (DOI). When depositing your data in a repository, make sure you select a repository that assigns a persistent identifier (for example [Zenodo](#)).

The **metadata** describing your data supports findability, citation and reuse. Rich metadata provides important context for the interpretation of your data and makes it easier for machines to conduct automated analysis. Follow standard metadata schemes, general ones such as [Dublin Core](#), or discipline specific. Consult the [DCC metadata directory](#), the [RDA Metadata Directory](#) and a portal of data standards at [FAIRsharing](#).

<https://www.openaire.eu/how-to-make-your-data-fair>

## Accessible

Make your data accessible by ensuring it:

- Is retrievable online using standardised protocols
- Has restrictions in place if necessary

**Remember that not all data has to be made open.** Data can be restricted and still be FAIR. However, if access is allowed, data should be retrievable without the need for specialised protocols. In addition, even if the full content is not made openly available, the data must be as findable as possible.

*As Open as Possible, As Closed as Necessary*

**Where can I keep my data?** Not necessarily opening it up, but keeping it somewhere safe for the long-term. You should look for a repository that does the following:

1. Stores the data safely
2. Make sure the data is findable
3. Describes the data appropriately (metadata)
4. Adds license information

You can **deposit data** to a general repository (e.g. [Zenodo](#), [Harvard Dataverse](#)) or a subject-specific repository (e.g. [Dryad](#)). Looking for your discipline? Search [www.re3data.org](http://www.re3data.org) for more suitable data repositories. See a demonstration of searching for [research data repositories using the re3data directory](#).



# Interoperable

Make your data interoperable by using:

- Common formats and standards
- Controlled vocabularies

**Interoperable data** means it can be integrated with other data, applications and workflows. Think about not creating data with proprietary software and making it available in open formats. Remember to use community agreed schemas, controlled vocabularies, keywords, thesauri or ontologies where possible.

## Reusable

Make your data reusable by ensuring it:

- o Is well-documented
- o Has clear licence and provenance information

**Create documentation**, e.g. a README file to help ensure that your data can be correctly interpreted and reanalyzed by others. A README plain text file should contain the following information:

- o for each filename, a short description of what data it includes, optionally describing the relationship to the tables, figures, or sections within the accompanying publication;
- o for tabular data: definitions of column headings and row labels; data codes (including missing data); and measurement units;
- o any data processing steps, especially if not described in the publication, that may affect interpretation of results;
- o a description of what associated datasets are stored elsewhere, if applicable;
- o whom to contact with questions.

If text formatting is important for your README, PDF format could also be used.

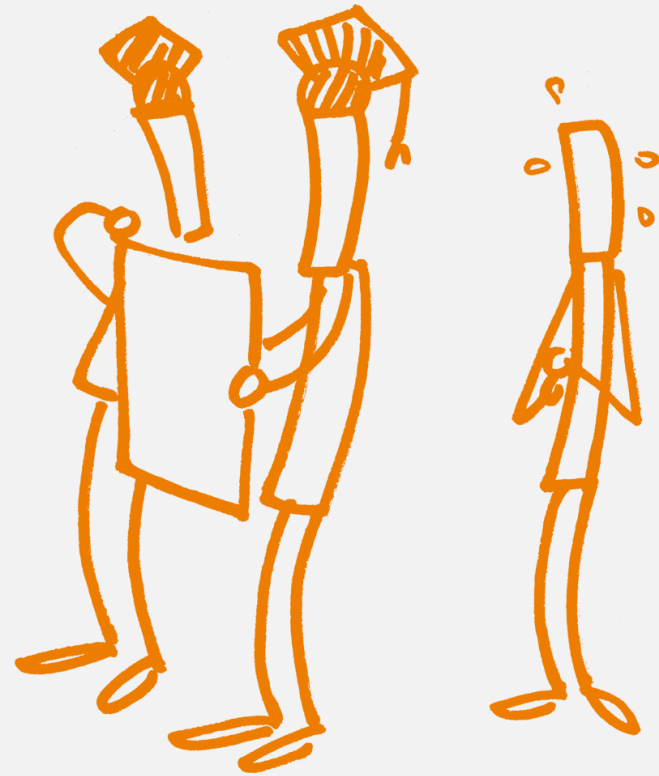
Source: <https://datadryad.org/pages/readme>

Data should have a **clear license** to govern the terms of its reuse. Guidance from the DCC can help you to [understand data licensing](#). This guide outlines the pros and cons of each approach e.g. the limitations of some Creative Commons options. The OA guidelines under Horizon 2020 recommend CC-0 or CC-BY as a straightforward and effective way to make it possible for others to mine, exploit and reproduce the data. See [p11 of this document](#).

**Check out:** EUDAT provides a [wizard](#) to help you choose an

# Misconception

If I share my data early, I'll be scooped!



# Pre-registration timestamps your work

## Register Your Project



Open Science Framework

A registration on OSF creates a frozen, time-stamped version of a project that cannot be edited or deleted. The *original project* can still be edited, while the registered version cannot. You might create a registration to capture a snapshot of your project at certain points in time - such as right before data collection begins, when you submit a manuscript for peer review, or upon completion of a project.

Registrations can be made public immediately or embargoed for up to 4 years. Registrations cannot be deleted, but they can be withdrawn. [Withdrawing a registration](#) removes the content of the registration but leaves behind basic metadata, like registration title, contributors, and a reason for the withdrawal (not required).

# Tips - share preprints too

- Early feedback on methods and initial findings
- Time to correct and mistakes before publishing
- Recognition for your ideas by peers

## Misconception #2:

I have to keep and share everything.



# Deciding which data need to be kept after the project ends

Five steps to follow

- ① **Could** this data be re-used
- ② **Must** it be kept as evidence or for legal reasons
- ③ **Should** it be kept for its potential value
- ④ **Consider costs** – do benefits outweigh cost?
- ⑤ **Evaluate criteria** to decide what to keep

5 steps to decide what data to keep

[www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep](http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep)

# What should be preserved and shared?

- The **data** needed to validate results in scientific publications (minimally!).
- The associated **metadata**: the dataset's creator, title, year of publication, repository, identifier etc.
  - Follow a metadata standard in your line of work, or a generic standard, e.g. Dublin Core or DataCite, and be FAIR.
  - The repository will assign a persistent ID to the dataset: important for discovering and citing the data.



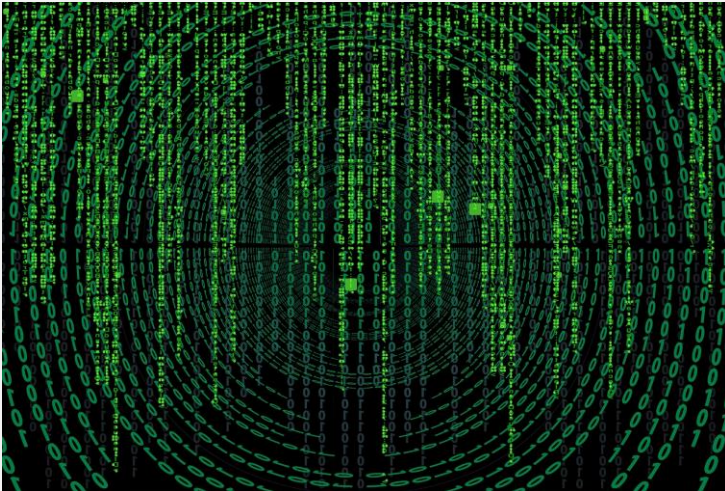
# What should be preserved and shared? (2)

- **Documentation**: code books, lab journals, informed consent forms - domain-dependent, and important for understanding the data and combining them with other data sources.
- **Software**, hardware, tools, syntax queries, machine configurations - domain-dependent, and important for using the data. (Alternative: information about the software etc.)

Basically, everything that is needed to replicate a study should be available. Plus everything that is potentially useful for others.

# Tip - link data to other outputs for context (reuse)

## Open Data



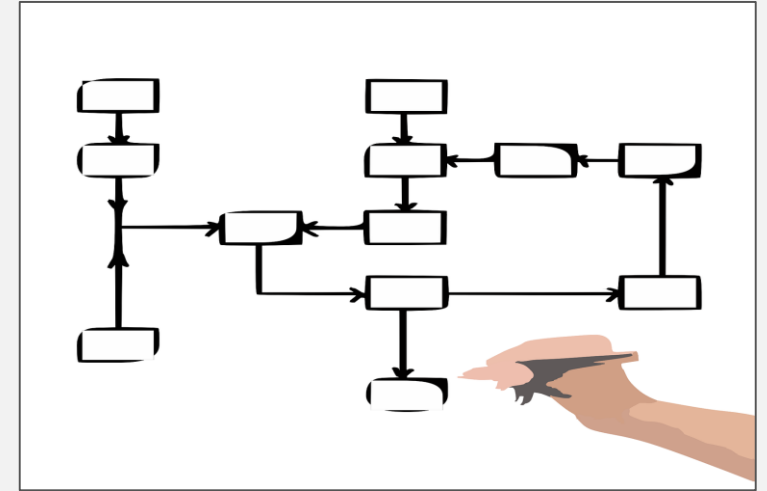
To support validation and facilitate reuse

## Open Code



Software created to analyse and/or visualise the data

## Open Workflows



What steps were taken and in what order?

# Consider who else has a say about sharing data

- Collaborators
- Research participants
- Commercial partners
- Data repository
- Publishers
- Institutions, funders



# How to make data open?



<https://okfn.org>

1. Choose your dataset(s)
  - What can you open? You may need to revisit this step if you encounter problems later.
2. Apply an open license
  - Determine what IP exists. Apply a suitable licence e.g. CC-BY
3. Make the data available
  - Provide the data in a suitable format. Use repositories.
4. Make it discoverable
  - Post on the web, register in catalogues...

# Research data lifecycle

**RE-USING DATA:** follow-up research, new research, undertake research reviews, scrutinising findings, teaching & learning

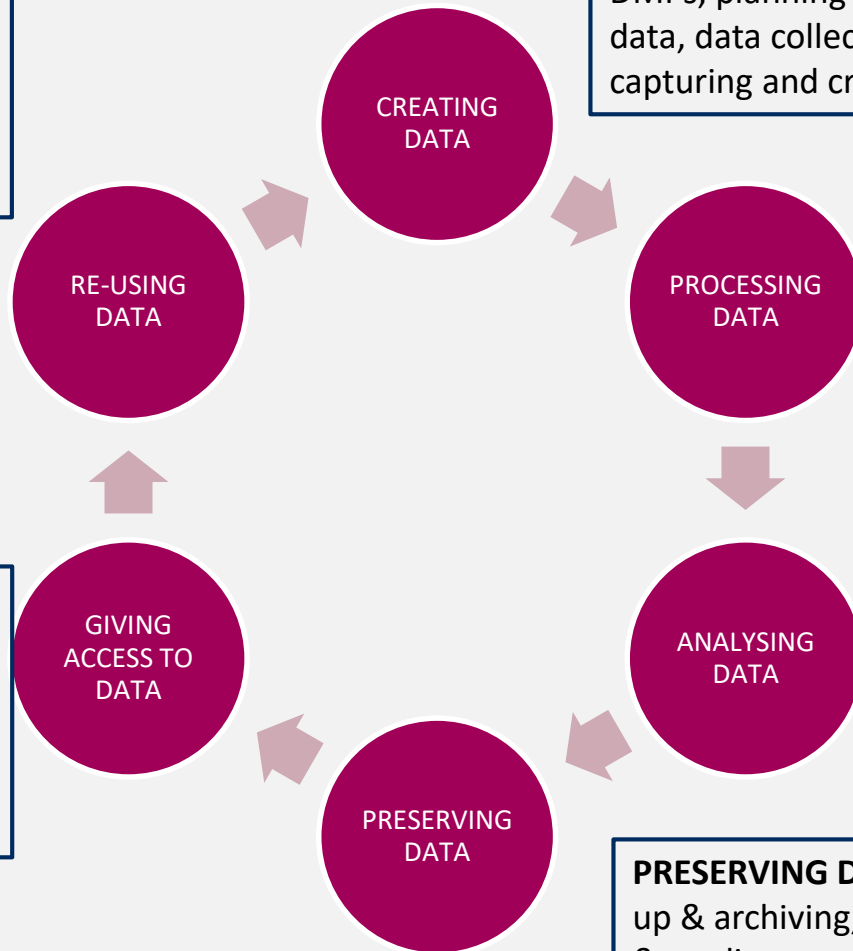
**CREATING DATA:** designing research, DMPs, planning consent, locate existing data, data collection and management, capturing and creating metadata

**PROCESSING DATA:** entering, transcribing, checking, validating and cleaning data, anonymising data, describing data, manage and store data

**ANALYSING DATA:** interpreting, & deriving data, producing outputs, authoring publications, preparing for sharing

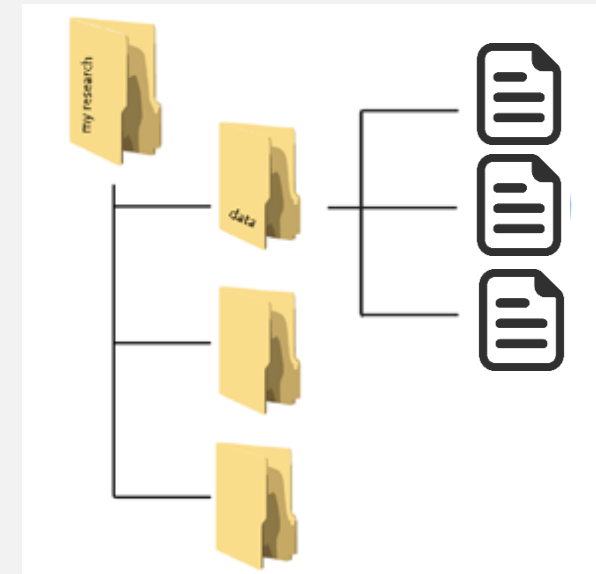
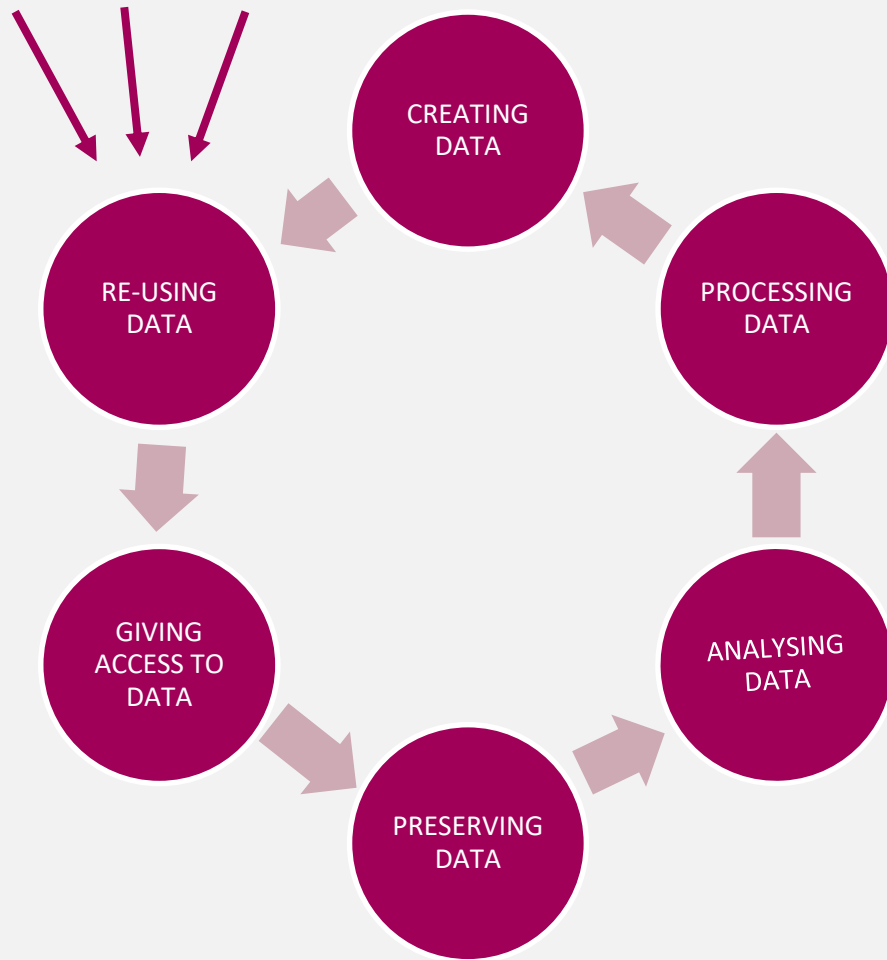
**PRESERVING DATA:** data storage, back-up & archiving, migrating to best format & medium, creating metadata and documentation

**ACCESS TO DATA:** distributing data, sharing data, controlling access, establishing copyright, promoting data



# Planning trick 1: think backwards

What data organisation would a re-user like?



# Data organisation

## Meaningful file names

Below are tips on meaningful and consistent file names. Read more in '[Choosing a file name](#)'.<sup>(2)</sup>

- ❑ Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD or yymmdd).
- ❑ Do not use strange characters like ?\!@\*%{[<> in the file name.
- ❑ Use traceable file names, such as Project\_Instrument\_locatie\_YYYYMMDD.ext.
- ❑ Make sure to only use each file once in the folder structure. If you store a file in more than one place, several versions of the same file can unwillingly be created.
- ❑ See also [version management](#).

It is good practice to note the file naming and its meaning in a readme.txt.

Even if a researcher is well underway with his project consistent file naming is still an option by using a [bulk file rename utility](#).<sup>(3)</sup> It is important, however, to check if this bulk renamer delivers on its promises.



white\_data\_20140708.csv



blue\_data\_20140708.docx



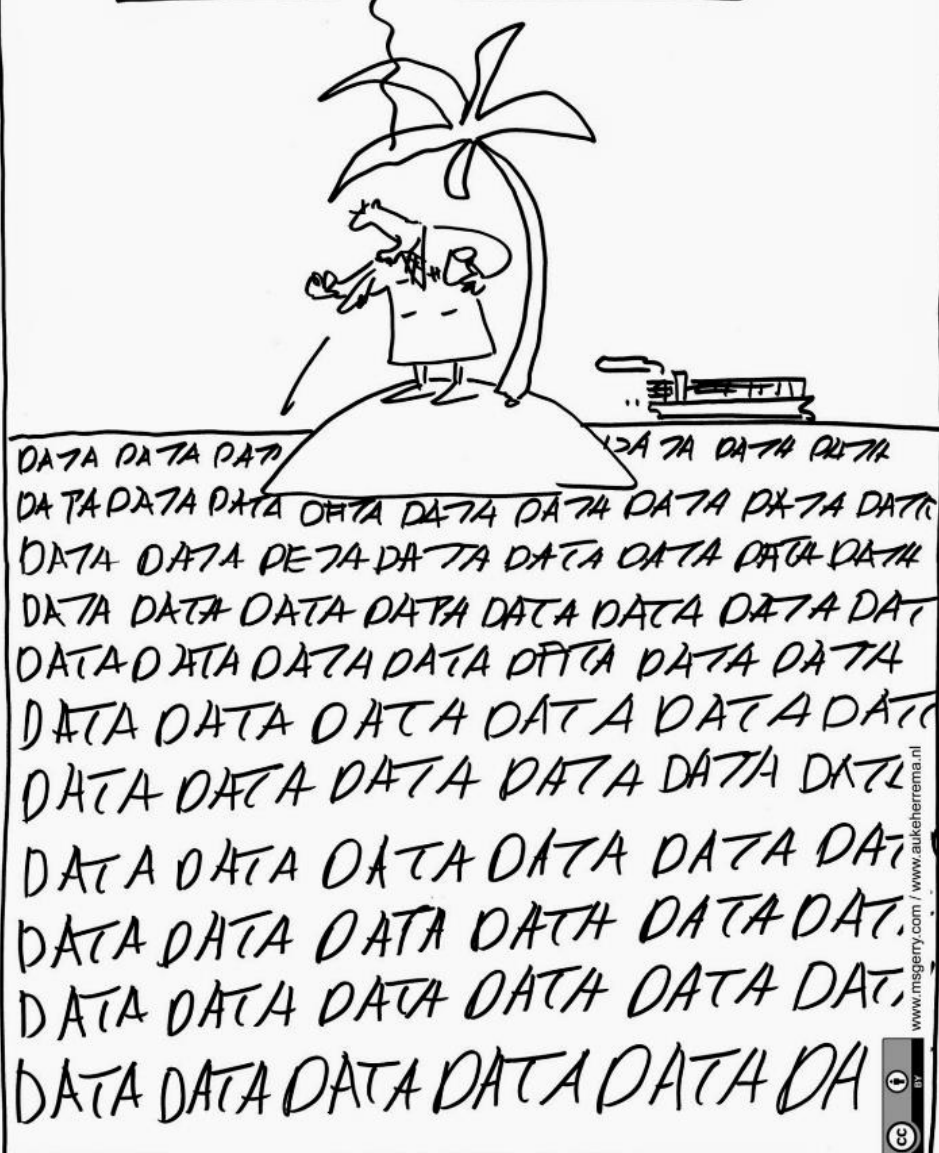
red\_data\_20140708.R



red\_data\_20140708\_v02.R

*File naming and version management*

NO USE, CAN NOT DRINK IT

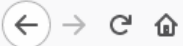


DATA OCEAN





REUSABLE DATA



## 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

github

@twitter

linkedin

facebook

# Bioinformatics

## Standards [Add](#)

### [MIBBI \(Minimum Information for Biological and Biomedical Investigations\)](#) [Edit](#)

A common portal to a group of nearly 40 checklists of Minimum Information for various biological disciplines. The MIBBI Foundry is developing a cross-analysis of these guidelines to create an interoperable, extensible community of standards.

The concept was realized initially through the joint efforts of the Proteomics Standards Initiative, the Genomic Standards Consortium and the MGED RSBI Working Groups. The latest project to register with MIBBI is the MIABie guidelines for reporting biofilm research, as of January 2012.

### [PDBx/mmCIF \(Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework\)](#) [Edit](#)

Protein Data Bank archive (PDB) is the single worldwide archival repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, managed by the Worldwide PDB (wwPDB). The PDB Exchange Dictionary (PDBx) is used by the wwPDB to define data content for deposition, annotation and archiving of PDB entries. PDBx incorporates the community standard metadata representation, the Macromolecular Crystallographic Information Framework (mmCIF), originally developed under the auspices of the International Union of Crystallography (IUCr). PDBx has been extended by the wwPDB to include descriptions of other experimental methods that produce 3D macromolecular structure models such as Nuclear Magnetic Resonance Spectroscopy, 3D Electron Microscopy and Tomography.

## Extensions [Add](#)

## Tools [Add](#)

### [PDBx/mmCIF Software Resources](#) [Edit](#)

Parsing, validation, and visualization tools and libraries supporting PDBx/mmCIF, the data standard used by the Worldwide Protein Data Bank.

## Use Cases [Add](#)

[BioModels Database](#) [Edit](#)

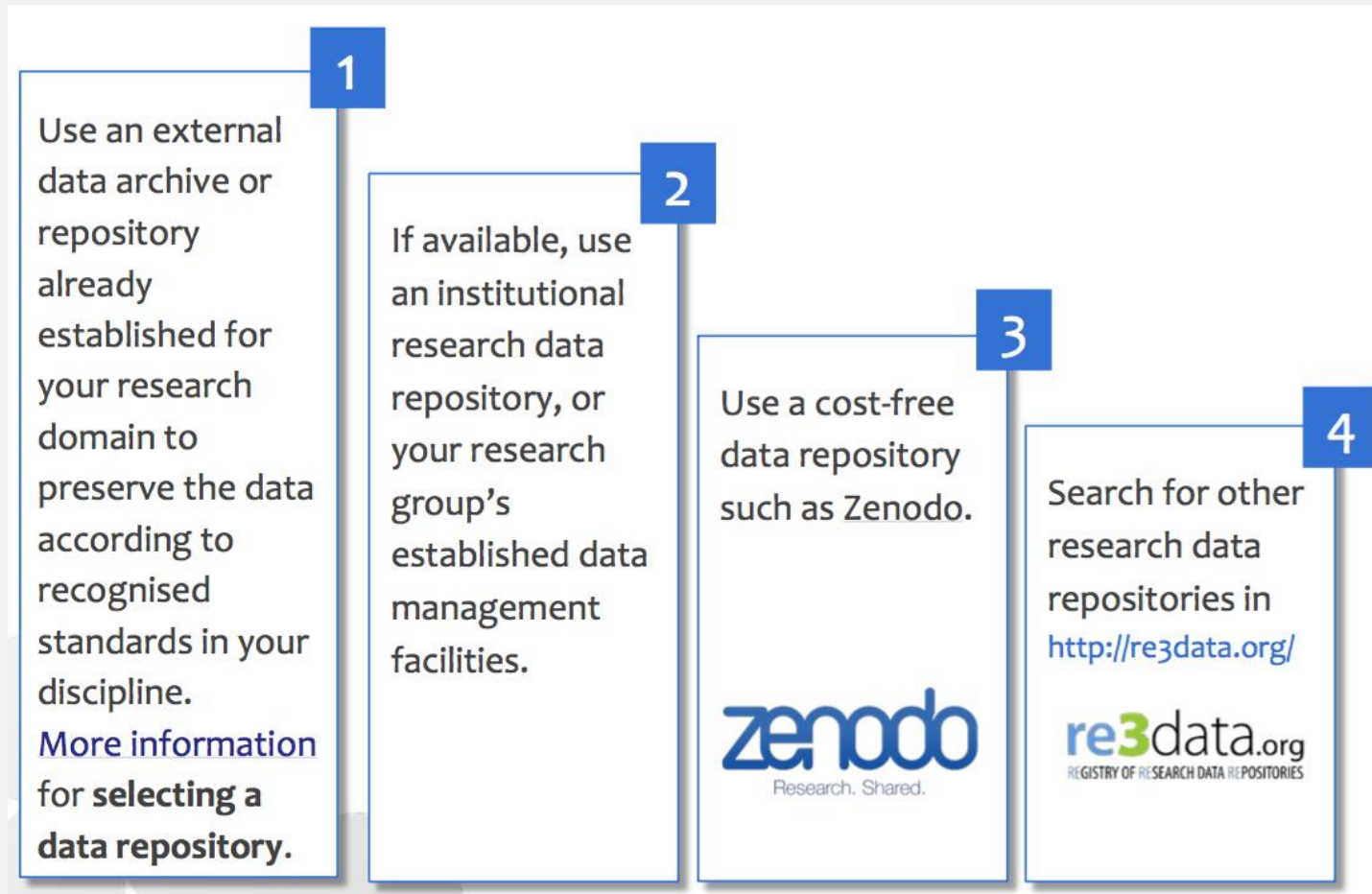
# Archiving, repositories, ehm?

Select a data repository that will preserve your data, metadata and possibly tools in the long term.

It is advisable to **contact the repository of your choice when writing the first version of your DMP.**

Repositories may offer guidelines for sustainable data formats and metadata standards, as well as support for dealing with sensitive data and licensing.

# Where to find a repository?



Zenodo: <http://www.zenodo.org>

Re3data.org: <http://www.re3data.org>

# How to select a repository?

Main criteria for choosing a data repository:

Certification as a ‘Trustworthy Digital Repository’, with an explicit ambition to keep the data available in the long term.

- Three common certification standards for TDRs:



Data Seal of Approval: <http://datasealofapproval.org/en>

nestor seal: [http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel\\_node.html](http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html)

ISO 16363: <http://www.iso16363.org>

# How to select a repository? (2)

- Matches your particular data needs: e.g. formats accepted; mixture of Open and Restricted Access.
- Provides guidance on how to cite the data that has been deposited.
- Gives your submitted dataset a persistent and globally unique identifier: for sustainable citations - both for data and publications - and to link back to particular researchers and grants.

## (All) Research. Shared.

### — your one stop research shop!

All research outputs from across all fields of science are welcome! Zenodo accept any file format as well as both positive and negative results. However, we do promote peer-reviewed openly accessible research, and we do curate your upload before putting it on the front-page.



[www.zenodo.org](http://www.zenodo.org)

## Citeable. Discoverable.

### — be found!

Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable. Zenodo further supports harvesting of all content via the OAI-PMH protocol.

## Community Collections

### — create your own repository

Zenodo allows you to create your own collection and accept or reject all uploads to it. Creating a space for your next workshop or project have never been easier. Plus, everything is citeable and discoverable.

## Safe

### — more than just a drop box!

Your research output is stored safely for the future in same cloud infrastructure as research data from CERN's [Large Hadron Collider](#) using a CERN's battle-tested repository software [INVENIO](#) used by some of the world's largest repositories such as [INSPIRE HEP](#) and [CERN Document Server](#).

## Reporting

### — tell your funding agency!

Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](#). Just upload your research on Zenodo and we will take care of the reporting for you. We plan to extend with futher funding agencies in the future so stay tuned!

## Flexible Licensing

### — not everything is under Creative Commons

Zenodo encourage you to share your research as openly as possible to maximize use and re-use of your research results. However, we also acknowledge that one size does not fit all, and therefore allow for uploading under a multitude of different licenses and access levels\*.

\* You are responsible for respecting applicable copyright and license conditions for the files you upload.

 [New Upload](#)

Drafts 0

Published 0

Sort

Most recent

asc.

# Get started!

Make your first upload - all research outputs from across all fields of research are welcome.

[New Upload](#)[About](#)[About](#)  
[Contact](#)  
[Policies](#)[Resources](#)[Features](#)  
[FAQ](#)[Developers](#)[REST API](#)  
[OAI-PMH](#)[Contribute](#)[GitHub](#)  
[Donate](#)[Funded by](#)



Delete

Save

Publish

## New upload

**Instructions:** (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Files

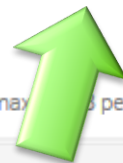
Choose files

Start upload

Drag and drop files here

— or —

Choose files

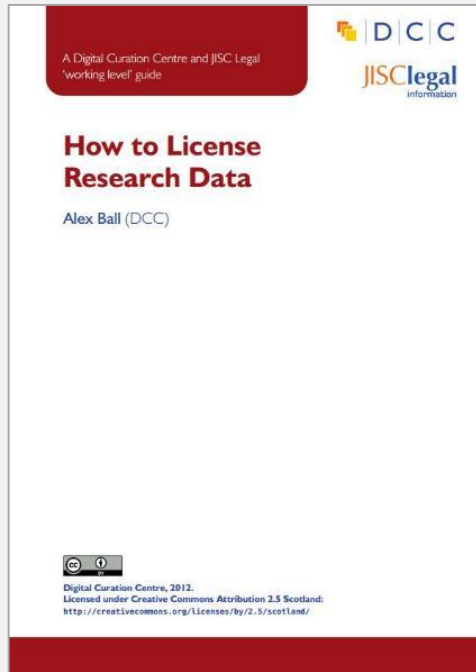


(minimum 1 file required, max  per dataset - [contact us](#) for larger datasets)

Upload type

required

# Licensing research data



Horizon 2020 Open Access guidelines point to:



This DCC guide outlines the pros and cons of each approach and gives practical advice on how to implement your licence

## CREATIVE COMMONS LIMITATIONS



NC Non-Commercial  
*What counts as commercial?*



ND No Derivatives  
*Severely restricts use*

**These clauses are not open licenses**

[www.dcc.ac.uk/resources/how-guides/license-research-data](http://www.dcc.ac.uk/resources/how-guides/license-research-data)

# EUDAT licensing tool

Answer questions to determine which licence(s) are appropriate to use

Do you own copyright and similar rights in your dataset and all its constitutive parts?

Do you allow others to make commercial use of you data?

**Creative Commons Attribution (CC-BY)**  
This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

**Public Domain Dedication (CC Zero)**  
CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

<http://ufal.github.io/public-license-selector>

NON PECUNIAE INVESTIGATIONIS CURATORE

SED VITAE FACIMUS PROGRAMMAS DATORUM PROCURATIONIS

(Not for the research funder, but for life we make data management plans)


- Make your research easier
- Stop yourself drowning in irrelevant stuff
- Save data for later
- Avoid accusations of fraud or bad science
- Write a data paper
- Share your data for re-use
- Get credit for it

RDM Handbook | Support

https://www.openaire.eu/rdm-handbook

EXPLORE PROVIDE CONNECT MONITOR DEVELOP

SERVICES SUPPORT OPEN SCIENCE IN EUROPE ABOUT

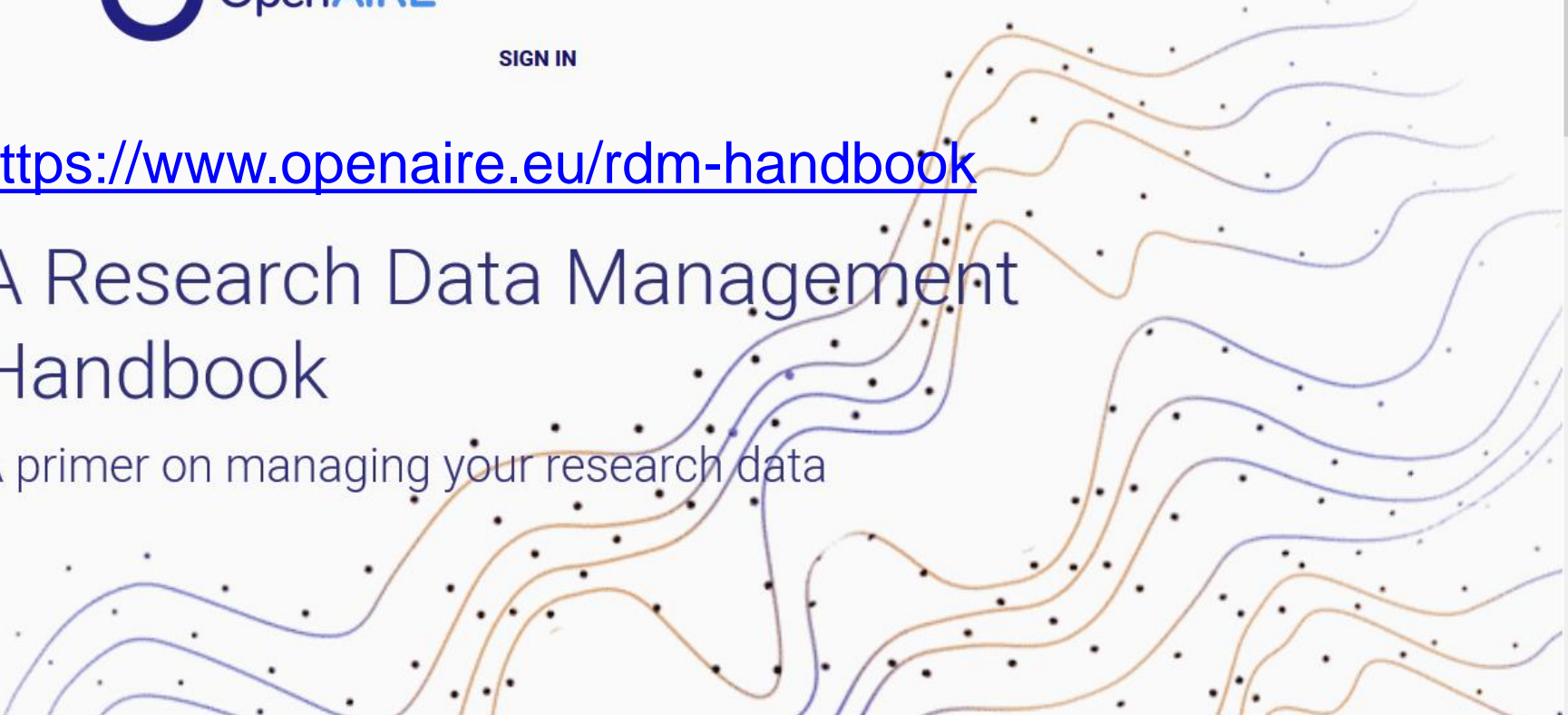


SIGN IN

<https://www.openaire.eu/rdm-handbook>

# A Research Data Management Handbook

A primer on managing your research data



OPEN DATA

HOW TO MAKE DATA OPEN?

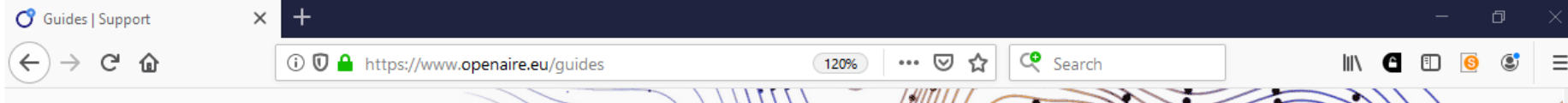
WHY MANAGE DATA?

## Open Data

“Open data and content can be freely used,

Support

RESOURCES



## <https://www.openaire.eu/guides>

### Researchers

[How to comply with H2020 mandates - publications](#)

[How to comply with H2020 mandates - research data](#)

[Research Data Management costs in H2020 projects](#)

[How to make your data FAIR](#)

[How to find a trustworthy repository for your data](#)

[How to create a Data Management Plan for H2020 projects](#)

[Data formats for preservation](#)

[How do I know if my research data is protected](#)

[How do I license my research data](#)

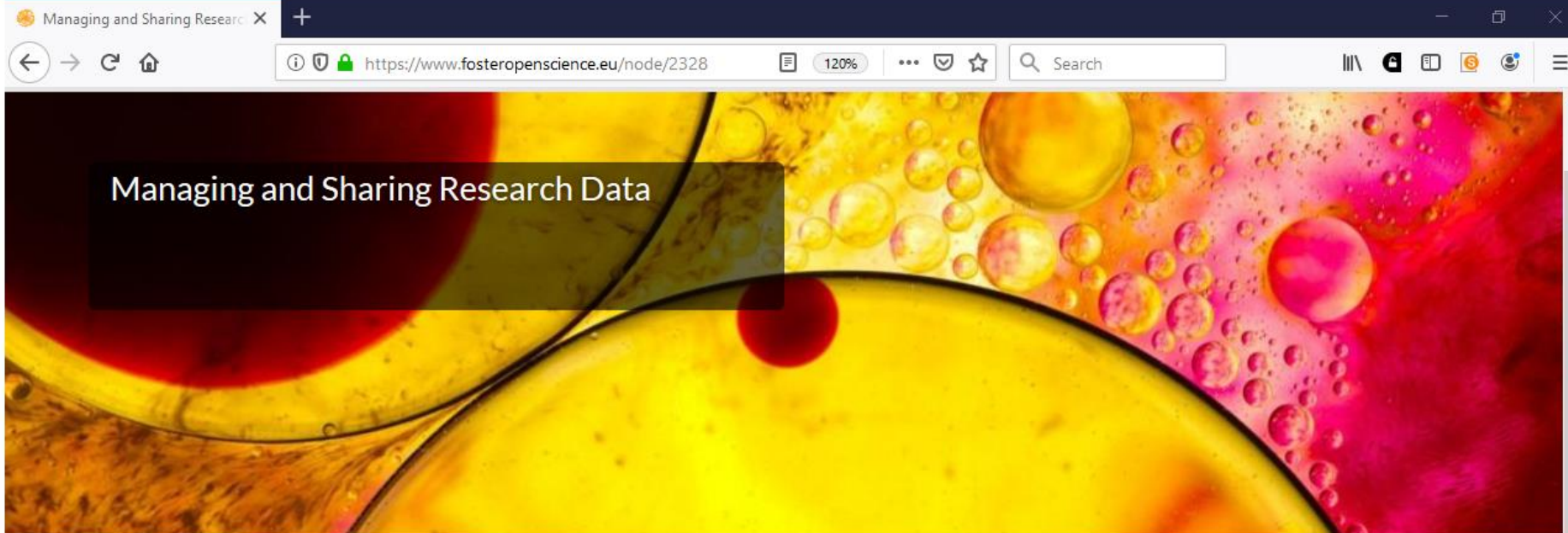
[Can I reuse someone else's research data](#)

[How to deal with non-digital data](#)

[How to deal with sensitive data](#)

[Raw data, backup and versioning](#)

[How can identifiers improve the dissemination of your research outputs?](#)



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:

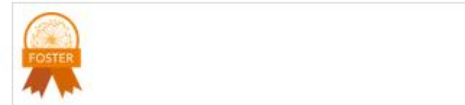
- understand which data you can make open and which need to be protected
- know how to go about writing a data management plan
- understand the FAIR principles
- be able to select which data to keep and find an appropriate repository for them
- learn tips on how to get maximum impact from your research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

For a citable version or to use this course offline, please refer to the print version which is available from [Zenodo](#).

Haga el curso en español

Start the Free Course

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



## Full details

**Level of knowledge:** Introductory: no previous knowledge is required

## Topics





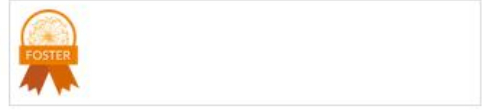
## Data Protection and Ethics

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

[Haga el curso en español](#)

[Start the Free Course](#)



### Full details

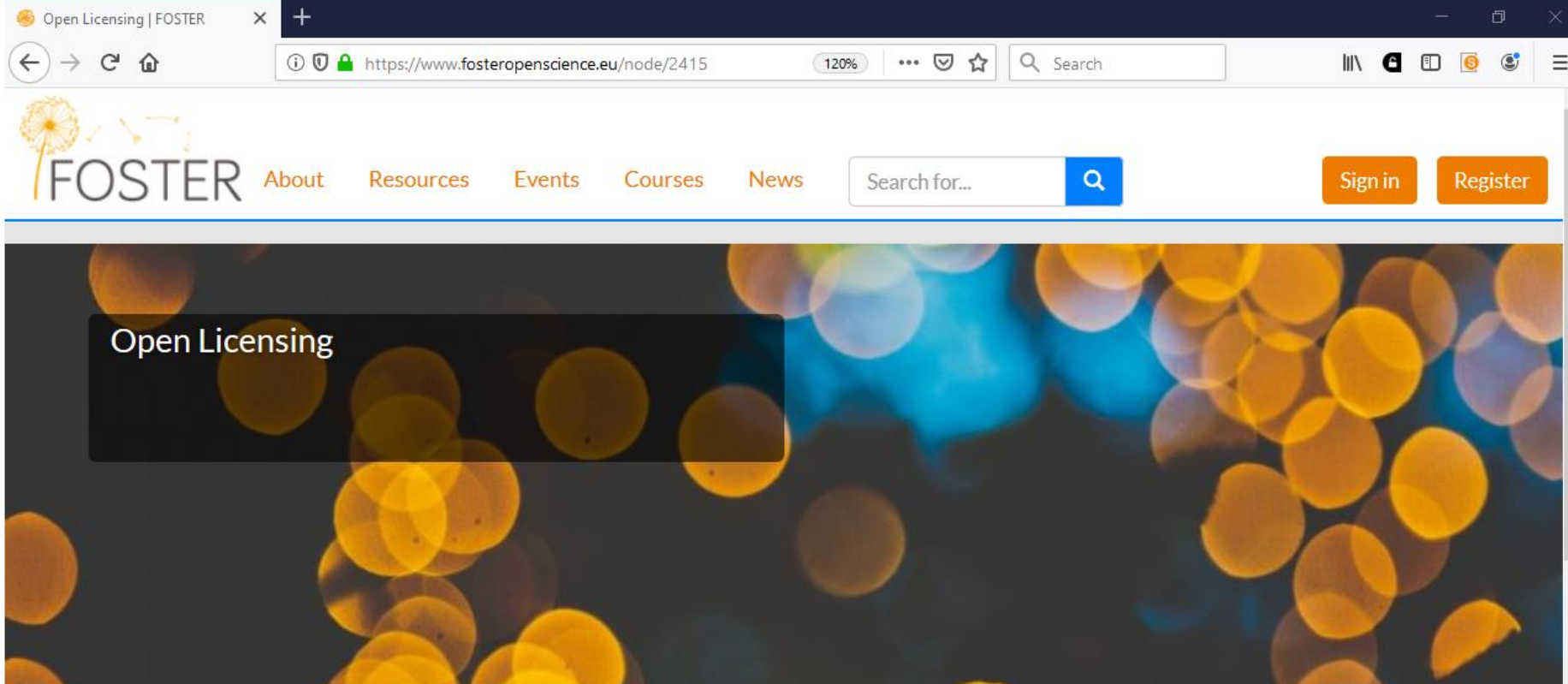
**Level of knowledge:** Introductory: no previous knowledge is required

### Topics



<https://www.fosteropenscience.eu/node/2330>





Licensing your research outputs is an important part of practicing Open Science. In this course, you will:

- know what licenses are, how they work, and how to apply them
- understand how different types of licenses can affect research output reuse
- know how to select the appropriate license for your research

Haga el curso en español

Start the Free Course



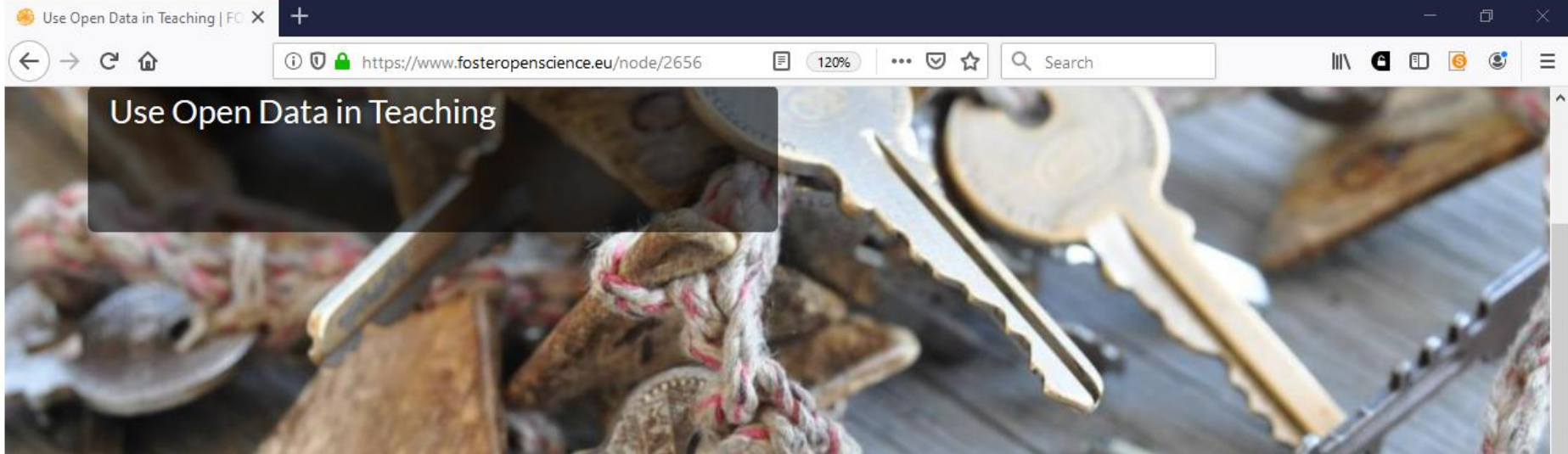
## Full details

**Level of knowledge:** Introductory: no previous knowledge is required

## Topics



<https://www.fosteropenscience.eu/node/2415>



In higher education and research, the topics of open science and research data management have gained interest. This module stimulates the use of open research data in teaching, thus furthering the open science movement.

With the help of the teacher community, the module offers inspiring resources, such as good practices and examples of lesson plans and learning activities. It also provides practical information on how to use open data in teaching and indirectly encourages the acquisition of research data management literacy among students.

This module can help you:

- Integrate open data in your teaching
- Stimulate open science
- Stimulate reuse

This course was developed as part of the [Use \(Open Research\) Data in Teaching project \(UDIT\)](#) which was initiated in 2017. The objective of the project is to encourage and help teachers in higher education use open research data in their teaching. The motivation of the project is the importance we see in active learning and in open science, and the increasing focus dedicated to these two concepts, both on the institutional, national, and international level. The partner institutions in the project are Radboud University, UiT The Arctic University of Norway, and University of Amsterdam/Amsterdam University of Applied Sciences.

<https://www.fosteropenscience.eu/node/2656>

## Full details

### Topics



### Audience

Librarians and Repository managers

PHD Students

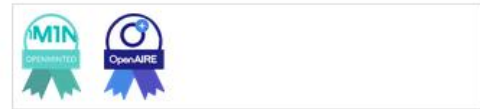


## Assessing the FAIRness of Data

In this course you'll learn how to go about assessing the FAIRness of research data using freely accessible tools and resources. This course will:

- introduce you to the key terms and explain what they mean in a practical sense
- tell you how data management planning can help to make data FAIR from the very start of research projects
- show you how you can use freely available tools to help assess the FAIRness of data

[Start the Free Course](#)



Full details

Reuse this course

If you want to use this course in your LMS.

[Download SCORM](#)

<https://www.fosteropenscience.eu/node/2644>

# Sherpa Juliet

About Search Statistics Contact Admin

## Research Funders' Open Access Policies

Sherpa Juliet is a searchable database and single focal point of up-to-date information concerning funders' policies and their requirements on open access, publication and data archiving.

Search for a funder policy  Search

**SERVICES**

### Open access services from Jisc

Services to support open access

**SERVICE**

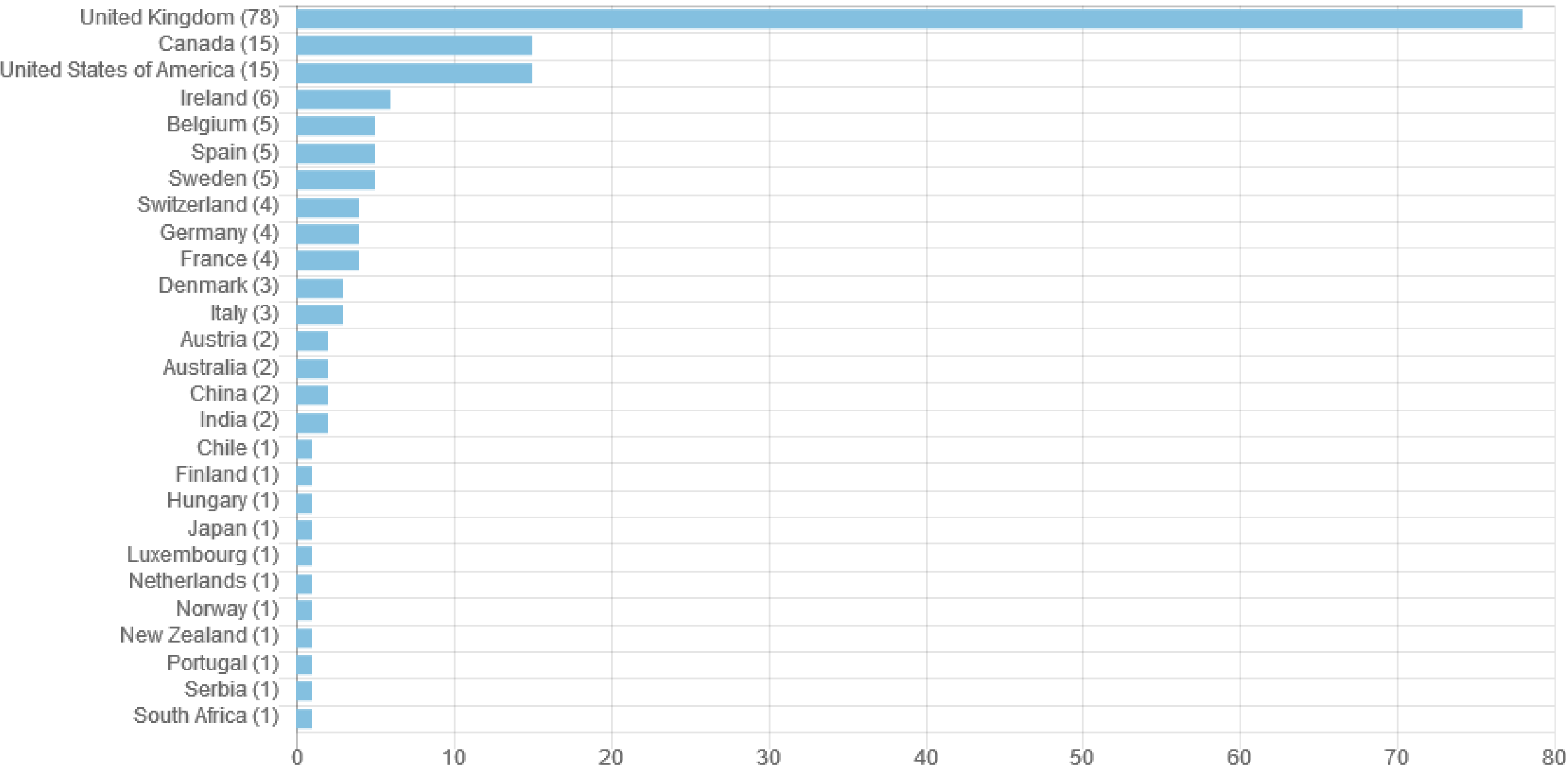
### Sherpa Services

Helping authors and institutions make informed and confident decisions in

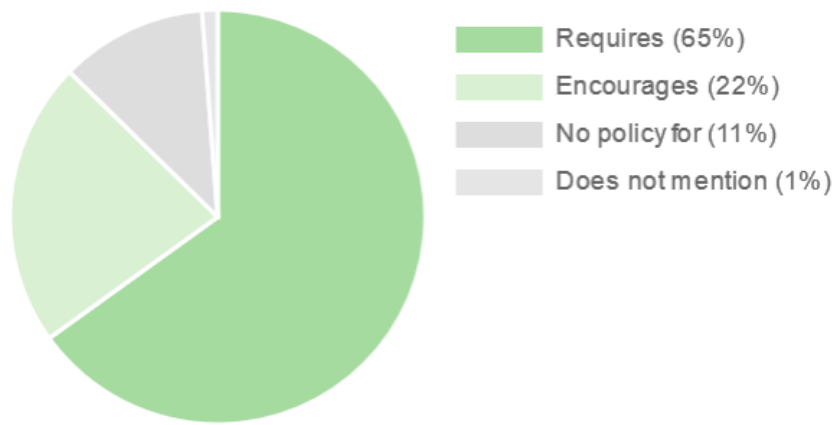
**GUIDE**

### Managing open access costs

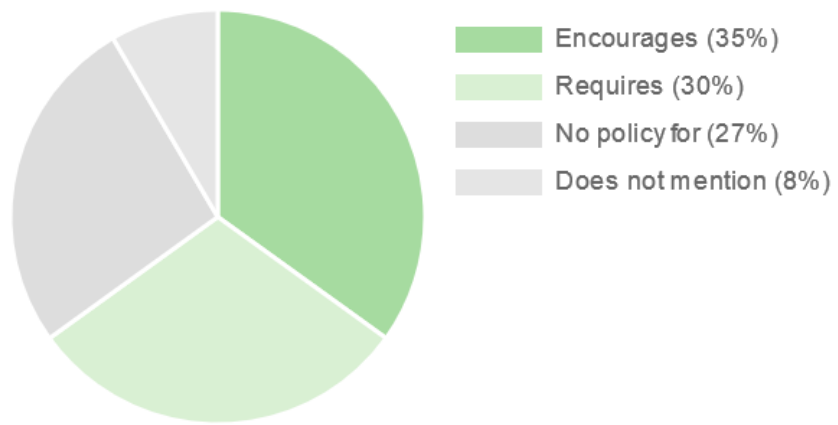
A guide from Jisc



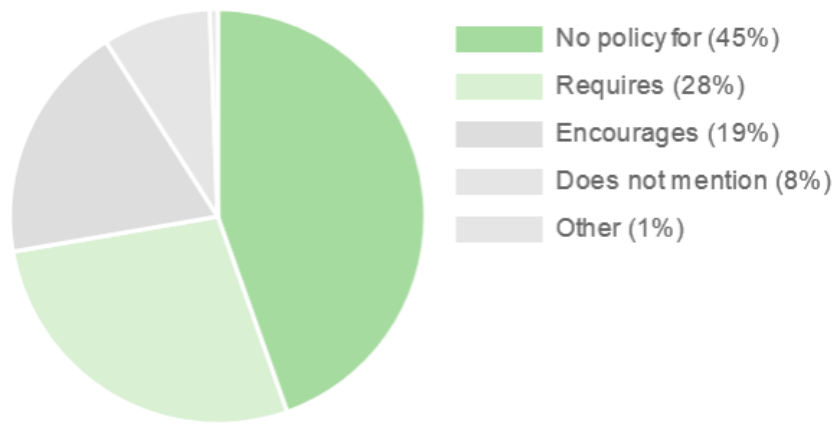
### Open Access Archiving Policies



### Open Access Publishing Policies



### Data Archiving Policies





Leabharlann UCD  
UCD Library

Search Library Supports

- Explore Collections
- Use the Library
- Guides & Help
- Cultural Heritage
- Contact Us / About Us
- A-Z Index

## <https://libguides.ucd.ie/openaccess/openaccessstory>

Library / LibGuides / Researcher Guides / Open Access for Research Impact / My Open Access Story

### Open Access for Research Impact: My Open Access Story

Search this Guide Search

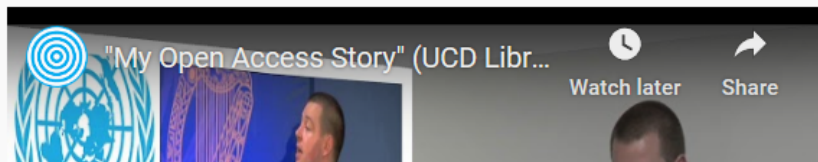
Discover the benefits of increasing the visibility and impact of your research outputs.

- What is Open Access (OA)?
- Green v. Gold
- OA Publishing Agreements
- OA & the Humanities
- Funders' Requirements
- Plan S
- OA Requirements in Horizon 2020
- Research Repository UCD
- Finding OA Materials
- Using OA Books to Support Teaching & Learning
- My Open Access Story**
- OA Survey
- Open Scholarship/Science
- OA News & Developments
- Further Help & Resources

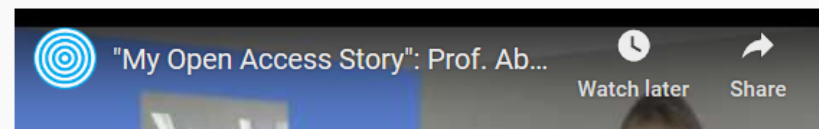
#### My Open Access Story

"My Open Access Story" was held during Open Access Week 2017. At the event UCD researchers spoke about their experience with open access and the benefits of making their research outputs openly accessible in UCD's open access repository, [Research Repository UCD](#).

#### My Open Access Story Compilation



#### Prof. Abbey Hyde, UCD School of Nursing, Midwifery and Health Systems



### My Open Access Story Compilation



### Prof. Abbey Hyde, UCD School of Nursing, Midwifery and Health Systems



### Dr. Muireann Ni Raghallaigh, UCD School of Social Policy, Social Work and Social Justice



### Dr. Liam Thornton, UCD Sutherland School of Law



### Prof. David Brayden, UCD School of Veterinary Medicine



### Assoc. Prof. Tina Hickey, UCD School of Psychology







- “Open access” is the practice of granting free web access to research articles, papers, book chapters, etc.
- “Green” open access means that uploading to an institutional repository is free for authors. It is sometimes referred to as “self-archiving”. Research Repository UCD is UCD’s “green” open access repository:  
[http://researchrepository.ucd.ie/@ucd\\_oa](http://researchrepository.ucd.ie/@ucd_oa)

- “Gold” open access is a channel provided by some publishers whereby authors pay a fee (known as an “article processing charge” or APC) to be published in their open access or hybrid access journal. The APC is typically in the range of €500 - €5,000.

### Uploading to Research Repository UCD

- As part of the process of updating your author profile in the Research Management System (RMS) you can upload your article, paper, chapter etc. See the short video here:

### Citation Advantage

#### RESEARCH IMPACT OF PAYWALLED VERSUS OPEN ACCESS PAPERS

This research presents data from the 1science oaIndx on the average of relative citations (ARC) for 3.3 million papers published from 2007 to 2009 and indexed in the Web of Science (WoS). These data show a decidedly large citation advantage for open access (OA) papers, despite them suffering from a lag in availability compared to paywalled papers.

#### FINDINGS:

- Publishing in paywalled journals without green archiving is never an effective impact maximization strategy
- In total, and for all these fields, publishing in paywalled journals with no additional green archiving always yields below average citedness (the average being 1.0).
- Publishing in paywalled journals is the least impactful strategy overall, and the least impactful in 16 out of 22 fields.
- On average, open access papers produce a 50% higher research impact than strictly paywalled papers.
- In all these fields, fostering open access (without distinguishing between gold and green) is always a better research impact maximization strategy than relying on strictly paywalled papers.
- Having a green copy of a paper is the most impactful research communication strategy overall and the best strategy in 19 fields out of 22.
- Green is nearly always more effective than relying strictly on gold (20 out of 22 fields).
- Gold is the best strategy in biology and biomedical research and very close to green in clinical medicine (likely a reflection of the NIH and Wellcome Trust OA mandates).
- Gold has the least impact in six fields.

The full article is available from the 1Science site (<http://www.1science.com/oaumbr.html>):

- [1Science](#)

### Publishers

Most publishers are very happy to co-exist with green access repositories and do not place embargoes in the way of immediate access. More information on this is available in our outline of the policies of the most common publishers of UCD publications:

- [Publishers' Copyright & Embargo Policies](#) of the most common publishers of UCD publications

### More Information

- For more information contact Joseph Greene, Repository Librarian:  
[joseph.greene@ucd.ie](mailto:joseph.greene@ucd.ie)
- Click on the link below for a PDF version of this Start Guide:
- [Research Repository UCD - Brief Outline](#)  
A short guide to the key benefits and mechanics of uploading to Research Repository UCD.
- [Research Repository UCD \(flyer\)](#)  
Research Repository UCD Publicity Flyer




Library Services for Researchers  
**RESEARCH REPOSITORY UCD**  
Think Local - Reach Global



- Why Deposit?**
- Deposit in the repository and increase the visibility of your work...**
- Showcase your research**
- Become discoverable through Google**
- Increase your citation count**
- Promote your School or Research Centre**
- Disseminate UCD's scholarly research**
- Enhance UCD's reputation**
- Fulfil your funder's requirements**

- Fulfil your obligations to meet your funder's requirements that your research be made available in an open access repository.
- Increase the impact of your research by being universally accessible – there is evidence to suggest that making research open access significantly increases the chances of it being cited.\* All items in the Repository are automatically harvested to search engines such as Google, Google Scholar and also RIAN, the National Portal for Open Access Publications.
- Use the repository to showcase not only your own research but also the research of your School or Research Centre – the Library can assist in creating RSS feeds from the Repository, embedded directly into your School, Institute or Centre website.

## RESEARCH REPOSITORY UCD - FOR ALL RESEARCHERS

- Free global viewing of your research output 
- Increased & timely research visibility and improved impact
- 398,965 downloads last year from Research Repository UCD
- Submit via the UCD Research Management System
- UCD PhD and Research Masters theses now live
- Easy research impact statistics & altmetrics
- Fulfill funders' Open Access requirements
- Build research and industry connections

from the cover page of the Repository version.

"Research Repository" and "Institutional Repository" are often used interchangeably.

### What is the definition of a Repository and how does it differ from a publisher's website?

#### **Repositories**

A repository is an online, open access archive. Repositories can be institutional (e.g. Research Repository UCD), subject - based e.g. ArXiv; PubMed; or centralised e.g. Zenodo

#### **Publisher's Website**

Publishers do not operate according to the principles of open access; nor do they curate their content. Content on a publisher's website may or may not be available into the future; indeed, a publisher's website may or may not be available into the future.

### Do other Universities have research repositories?

Yes, most do. All the universities in Ireland have a repository and all contribute to RIAN - Ireland's National Research Portal


See the following for more information:

- [Research Repository UCD](#)
- [Open Access for Research Impact](#)
- [RIAN - Ireland's National Research Portal](#)
- [ArXiv](#)

Open access to over 900,000 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

### Joseph Greene

[Email Me](#)

**Contact:**  
353 (0)1 7167398  


### Related Guides

- [Open Access for Research Impact](#) by [Michelle Dalton](#)  
Last Updated Jul 23, 2020  
2696 views this year

<https://libguides.ucd.ie/RRU/intro>



## https://libguides.ucd.ie/RRU/Impact

Library / LibGuides / Researcher Guides / Research Repository UCD / Visibility and Impact

### Research Repository UCD: Visibility and Impact

Our guide to Research Repository UCD

- Research Repository UCD - Start Here
- Introduction
- Eligibility
- Types of Material Accepted by Research Repository UCD
- Submitting Materials
- Submit The Correct Version
- Post Submission and Library Services
- Copyright and Licences
- Embargoes
- Third-Party Services e.g. ResearchGate
- Access and Usage
- Visibility and Impact**
- Publishing
- UCD Download Statistics by College
- Testimonials from UCD Researchers

#### Increased Citations, Impact, Visibility, Downloads

Uploading to Research Repository UCD means that your research becomes universally accessible on the Internet and searchable through Google and Google Scholar.

The School of Civil Engineering started uploading items to the Repository in 2009. They now have a total of 587 items there, with download rates increasing year on year.

Research indicates that making your publications open access can significantly increase your chances of being cited; see for example a review of the research in this area by Alma Swan:

- Swan, A. (2010) *The Open Access citation advantage: studies and results to date*. Technical Report, School of Electronics & Computer Science, University of Southampton.

#### RESEARCH IMPACT OF PAYWALLED VERSUS OPEN ACCESS PAPERS

#### Create More Interest

### Functional impact of global rare copy number variation in autism spectrum disorders

Pinto, Dalila; Pagnamenta, Alistair T.; Klei, Lambertus; Regan, Regina; Conroy, Judith; Casey, Jillian; Green, Andrew; Segurado, Ricardo; Shah, Naisha; Ennis, Sean; et al.

Permanent link: <http://hdl.handle.net/10197/4381>

Date: 2010-07-15

#### Recommended citation:

Pinto, Dalila, Pagnamenta, Alistair T., Klei, Lambertus, Regan, Regina, Conroy, Judith, Casey, Jillian, Green, Andrew, Segurado, Ricardo, Shah, Naisha, Ennis, Sean, et al. : Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 15 (466) 2010-07-15, pp.368-372.

[Download this paper](#)

#### Impact and interest

This item's downloads: 77

[See more details](#)



Picked up by 4 news outlets  
Blogged by 13  
Tweeted by 14

- Swam, A. (2010) The Open Access citation advantage: studies and results to date. Technical Report, School of Electronics & Computer Science, University of Southampton.

### RESEARCH IMPACT OF PAYWALLED VERSUS OPEN ACCESS PAPERS

This research presents data from the 1science oalndx on the average of relative citations (ARC) for 3.3 million papers published from 2007 to 2009 and indexed in the Web of Science (WoS). These data show a decidedly large citation advantage for open access (OA) papers, despite them suffering from a lag in availability compared to paywalled papers.

#### FINDINGS:

- Publishing in paywalled journals without green archiving is never an effective impact maximization strategy
- In total, and for all these fields, publishing in paywalled journals with no additional green archiving always yields below average citedness (the average being 1.0).
- Publishing in paywalled journals is the least impactful strategy overall, and the least impactful in 16 out of 22 fields.
- On average, open access papers produce a 50% higher research impact than strictly paywalled papers.
- In all these fields, fostering open access (without distinguishing between gold and green) is always a better research impact maximization strategy than relying on strictly paywalled papers.
- Having a green copy of a paper is the most impactful research communication strategy overall and the best strategy in 19 fields out of 22.
- Green is nearly always more effective than relying strictly on gold (20 out of 22 fields).
- Gold is the best strategy in biology and biomedical research and very close to green in clinical medicine (likely a reflection of the NIH and Wellcome Trust OA mandates).
- Gold has the least impact in six fields.

The full article: Research Impact of Paywalled versus Open Access Papers, is available at <https://www.1science.com/1numbr/>

Pinto, Dalila, Pagnamenta, Alistair T., Klei, Lambertus, Regan, Regina, Conroy, Judith, Casey, Jillian, Green, Andrew, Segurado, Ricardo, Shah, Naisha, Ennis, Sean, et al. : Functional impact of global rare copy number variation in autism spectrum disorders. Nature, 15 (466) 2010-07-15, pp.368-372.

[Download this paper](#)

#### Impact and interest

This item's downloads: 77  
[See more details](#)



- Picked up by 4 news outlets
- Blogged by 13
- Tweeted by 14
- On 3 Facebook pages
- Referenced in 2 Wikipedia pages
- Highlighted by 1 platforms
- 653 readers on Mendeley
- 17 readers on CiteULike

Citations in Scopus®: 800  
[See more details](#)

### Prof. Aidan Moran





Access and Usage

Visibility and Impact

Publishing

UCD Download Statistics by College

Testimonials from UCD Researchers

### Economics

"Given the time lag from creation to publication in academic journals, it is critical that my research is disseminated in a timely manner in order to have a real impact...placing my working papers in the UCD Repository ensures that it is freely available to both academics and non-academics, thus ensuring optimal impact and increasing the visibility of my research"

*Dr. Orla Doyle*  
*UCD School of Economics*

### Social Policy, Social Work and Social Justice

"Placing material in the repository is useful on a number of levels. In the first instance it demonstrates the volume of work we produce and highlights our areas of interest. Also, it provides access for anyone, anywhere to the material I generate. It provides academics like myself with a platform to share material that would generally be inaccessible otherwise for reasons of cost, copyright or general curtailment. Of particular interest for me is that it makes my material readily available to the public...."

*Dr Valerie O'Brien*  
*School of Social Policy, Social Work and Social Justice*

### Physical Sciences

"As most of my research is publicly funded, it is imperative that my results are publicly available. Open access removes barriers to access and encourages innovation and discovery, ultimately advancing knowledge. Open access has the added benefit of encouraging me to write papers with a wider audience in mind, hopefully improving the clarity and impact of our collaborative work. Research Repository UCD in particular provides increased visibility to my work and tools to track and manage these resources. Their team helps navigate all of the copyright issues, making it easy to add publications to the repository."

*Dr. Brian Rodriguez*  
*Senior Lecturer, School of Physics*

### Civil Engineering

"The repository has been central in boosting the impact of our research...it is a main driver in growing the citation statistics of our researchers"

*Eugene O'Brien*  
*Professor of Civil Engineering*

"... has made such an improvement in our dissemination efforts"

*Professor Debra Laefer*  
*School of Civil Engineering*

# With thanks to

Joy Davidson, University of Glasgow

Marjan Grootveld, DANS

Sarah Jones, DCC

Acknowledgements:

Jonathan Rans, DCC

Thanks to DANS and DCC for reuse of slide

<http://openaccessweek.org>

 OPEN ACCESS WEEK 2020

**Open with Purpose:**

Taking Action to Build Structural  
Equity and Inclusion

OCTOBER 19-25

Международная  
Неделя Открытого Доступа 2020

**ОТКРЫТОСТЬ ДЛЯ  
СПРАВЕДЛИВОСТИ И  
ИНТЕГРАЦИИ**

**Среда, 21 октября 11.00 - 12.00**  
"Преимущества открытого доступа и открытой науки"  
Ирина Кучма, Международный консорциум EIFL  
"Открытые архивы Кыргызстана"  
Насымкуль Мамутова, Кыргызский государственный технический  
университет им. И.Раззакова  
**Онлайн-платформа: Zoom**  
Номер встречи: 980 4898 1763  
Код доступа: 508992

**Четверг, 22 октября 14.00 - 15.00**  
"Ограничения и исключения в авторском праве КР"  
Брыс Турсуналиева, Кыргызпатент  
"Правовое регулирование открытых лицензий в КР"  
Айсунуу Чубарова, Юридическая фирма "Арте"  
**Онлайн-платформа: Zoom**  
Номер встречи: 999 3595 9172  
Код доступа: 782880

**Международная Неделя  
Открытого Доступа 2020**

**Открытость для справедливости и  
интеграции**



**Thank you!**  
**Questions?**

[iryana.kuchma@eifl.net](mailto:iryana.kuchma@eifl.net)

**eifl** KNOWLEDGE  
WITHOUT  
BOUNDARIES

[www.eifl.net](http://www.eifl.net)