Best Practice in Open Science

Iryna Kuchma, EIFL Open Access Programme Manager,
Twitter: @irynakuchma

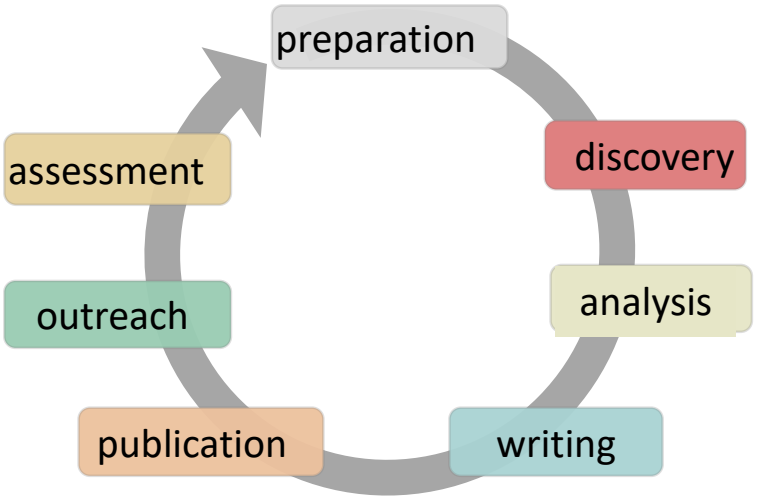
Innovations in scholarly communication

https://101innovations.wordpress.com/

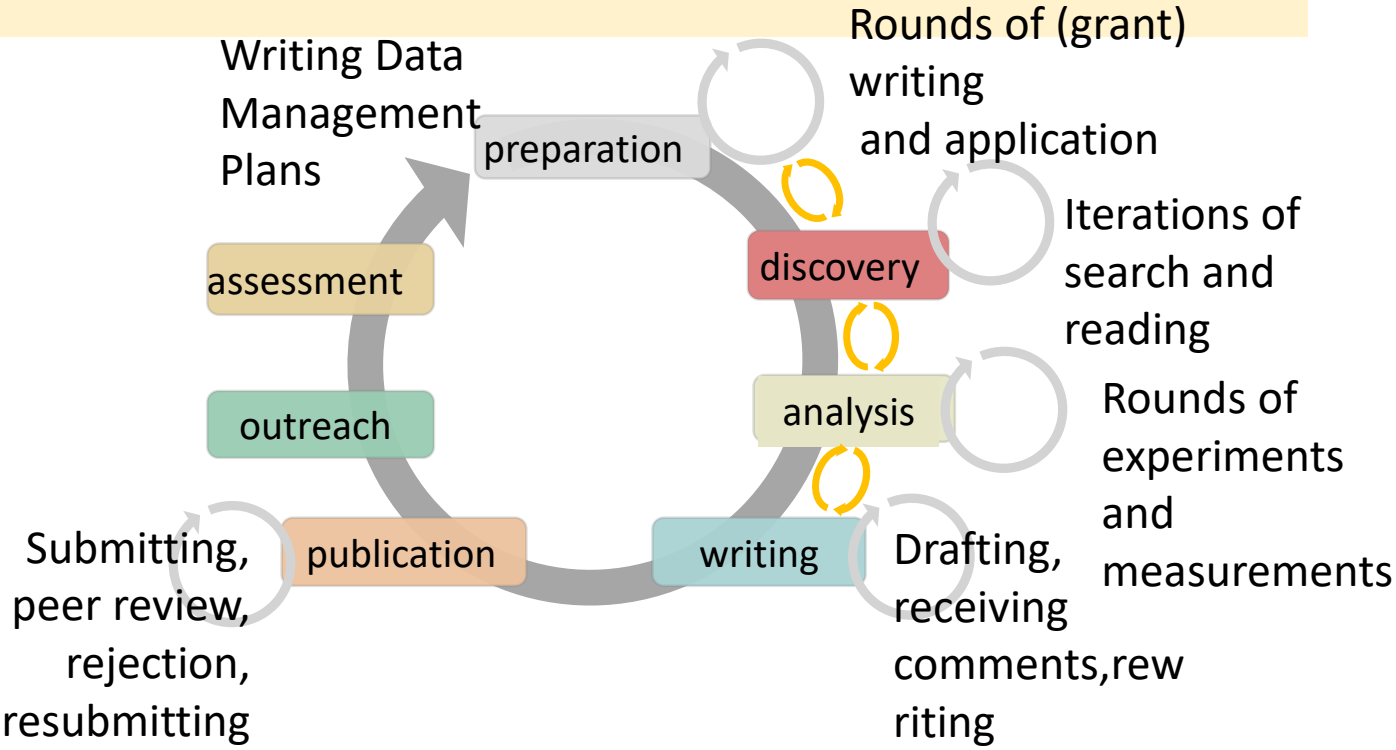
Bianca Kramer & Jeroen Bosman

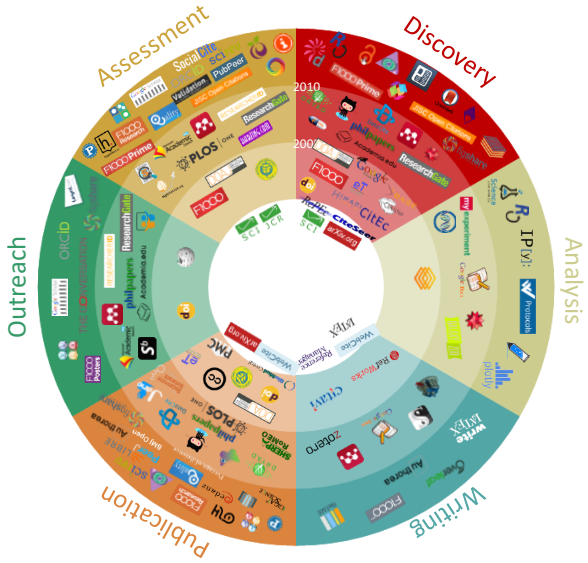
@MsPhelps
@jeroenbosman

A model of the research workflow

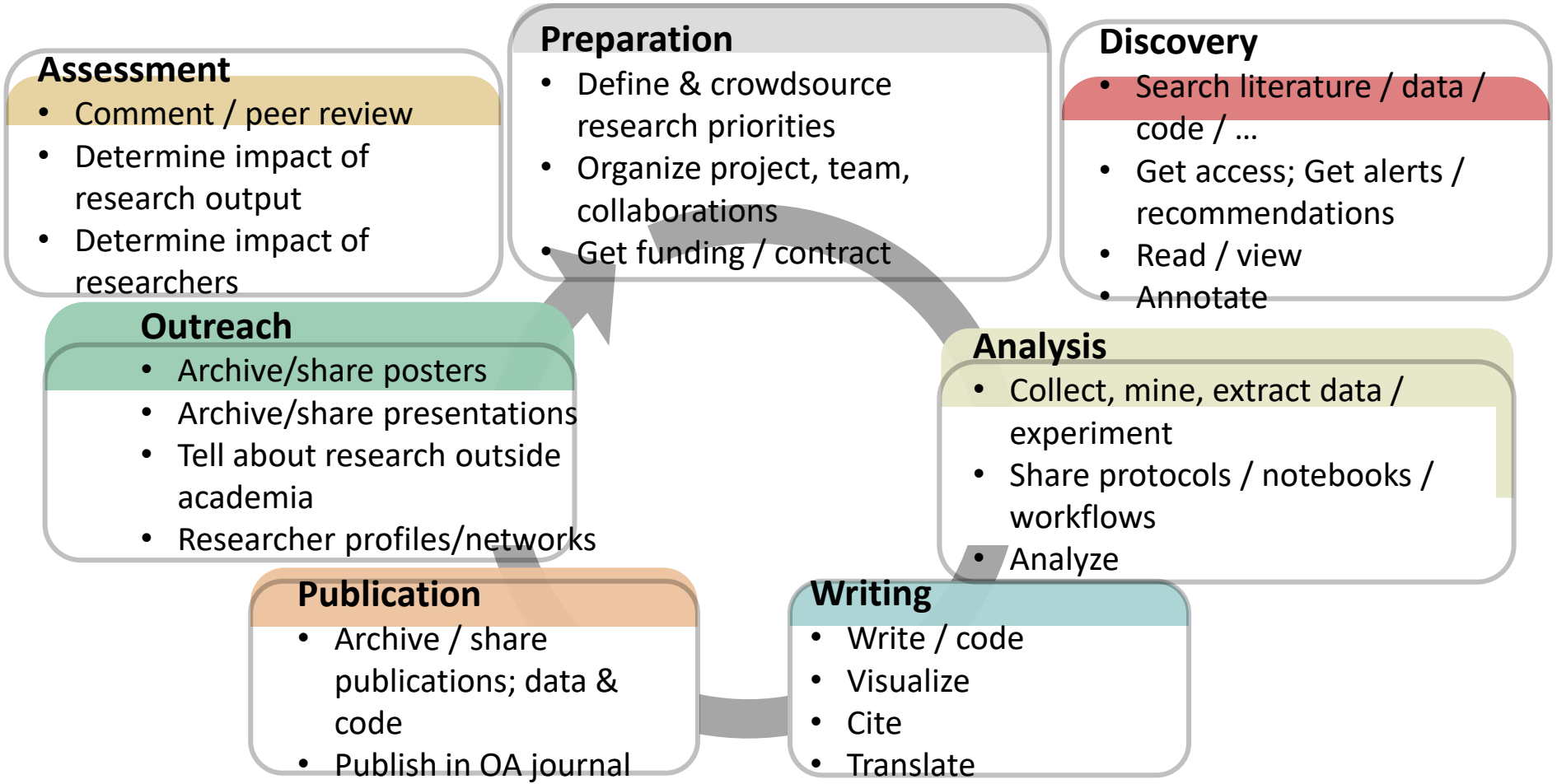


A model of the research workflow




Discovery
Analysis
Writing
Publication
Outreach
Assessment
2010
2007

Opening up the research workflow



Open Science practices

involve public / patients in drafting research proposals

openly share project proposals

share hypothesis before starting research (if possible/relevant)

having open discovery of open access materials

extensively search for existing data before generating your own

use easily attainable software to allow anyone to reproduce your results

Open Science practices

sharing protocols openly, online

store data in the most open format possible

cite OA versions of literature & provide data and code citations

acknowledge contributor roles in a publication

translate research objects in world languages

publish preprints, encourage feedback / open peer review

Open Science practices

publish pre-publication history (version + reviews)

making conflicts of interest transparant

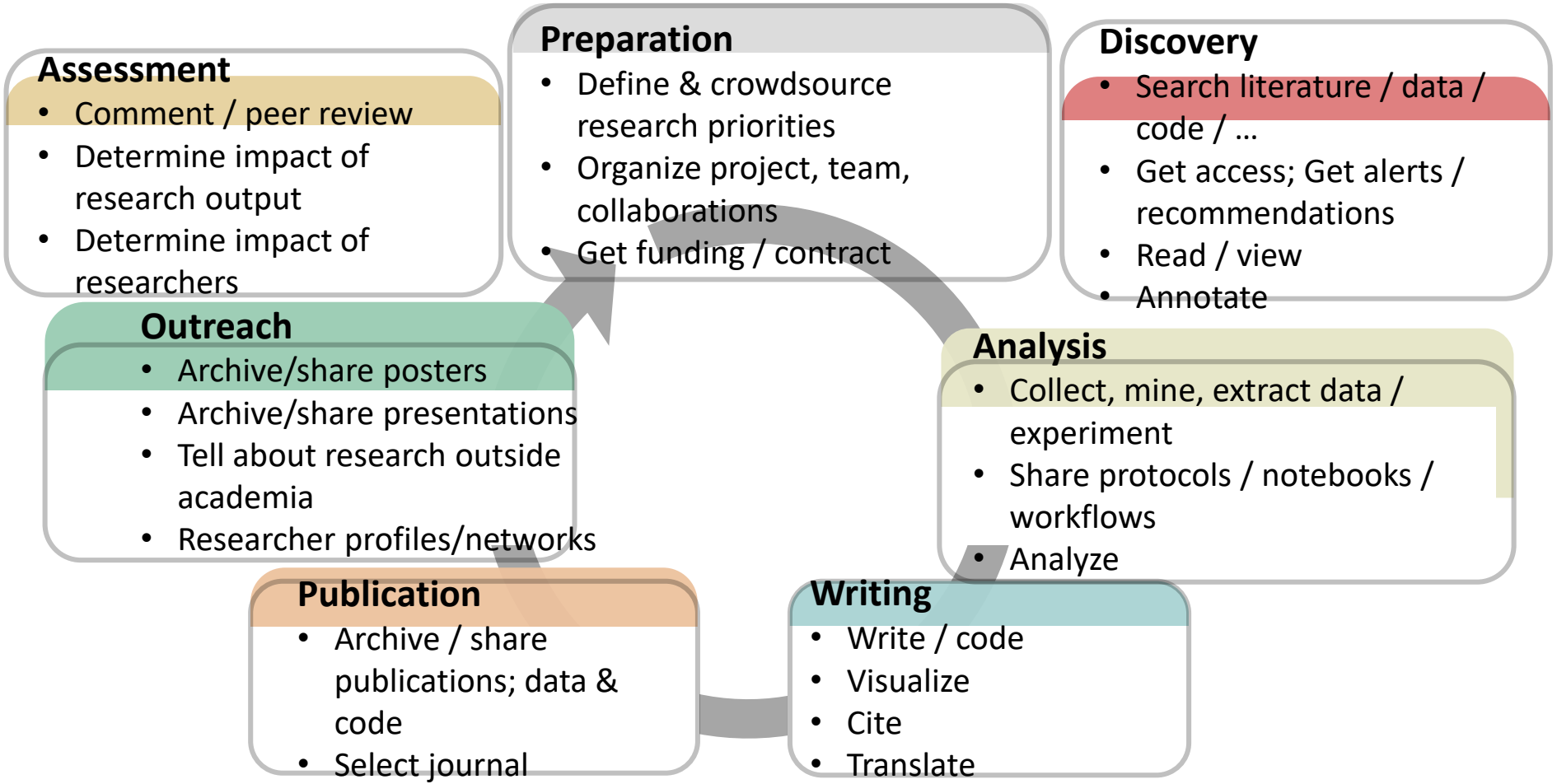
using academic social networks to find and communicate with other researchers

refuse to be part of all male of all white panels

having all types of review openly available

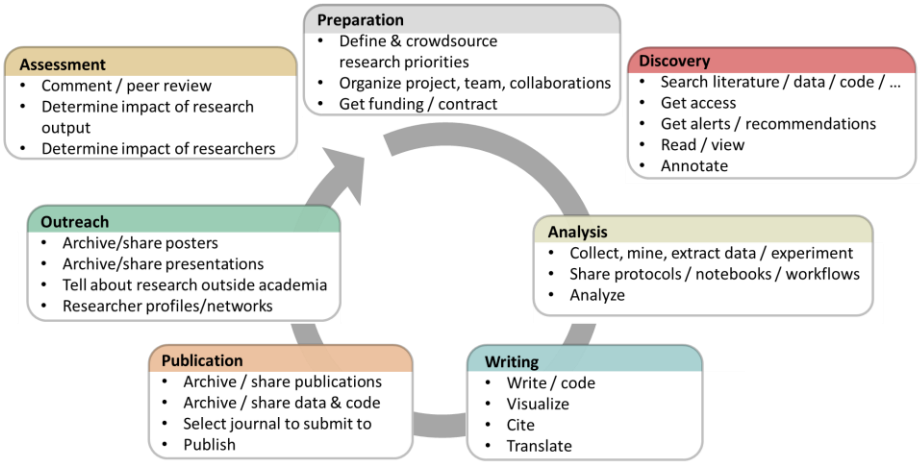
use metrics of commercial /social applications to assess research

Opening up the research workflow

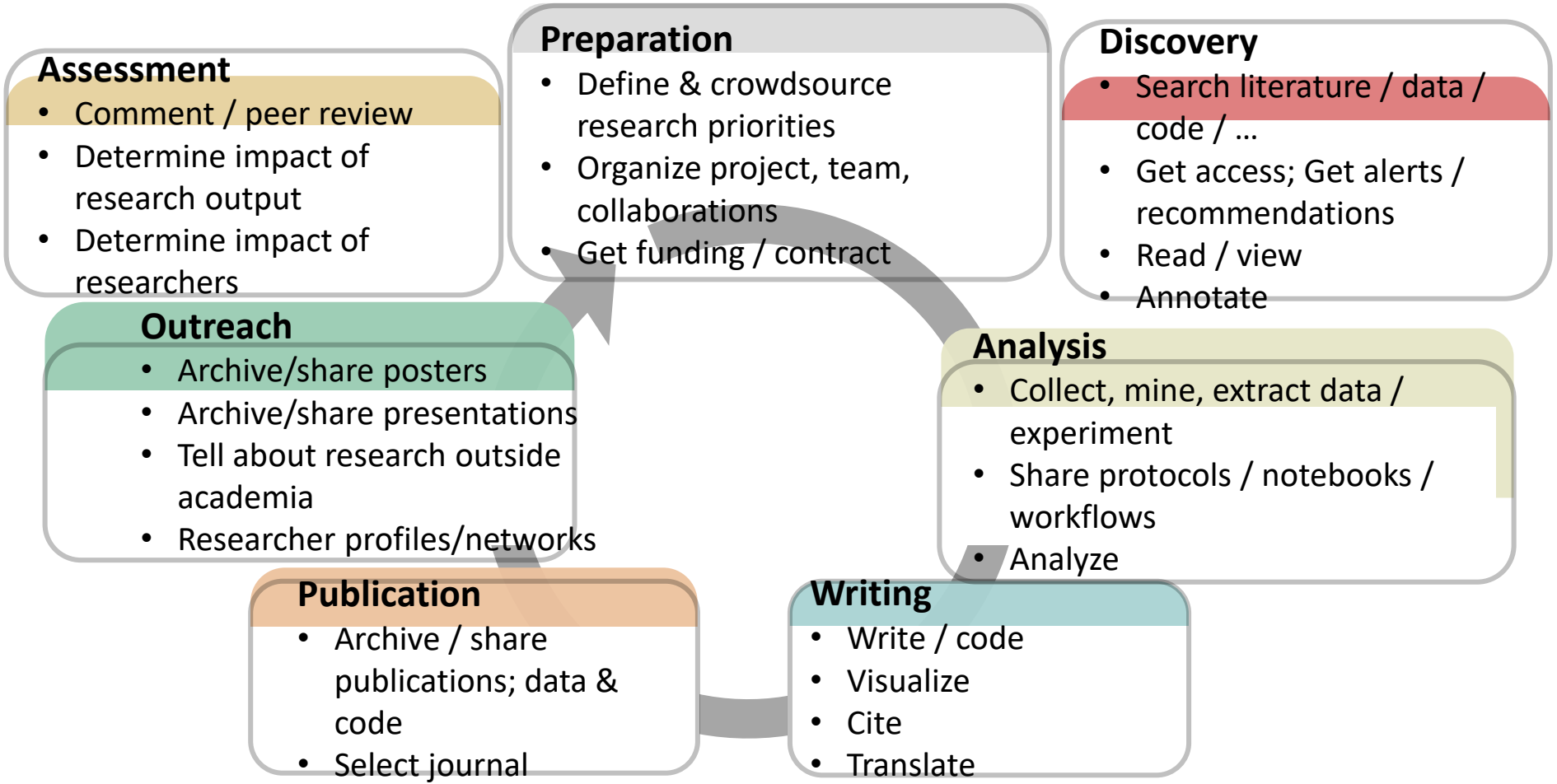


Activity

Tell us about the activities that you are already doing and about one or two that you haven't done yet and would like to explore



Opening up the research workflow

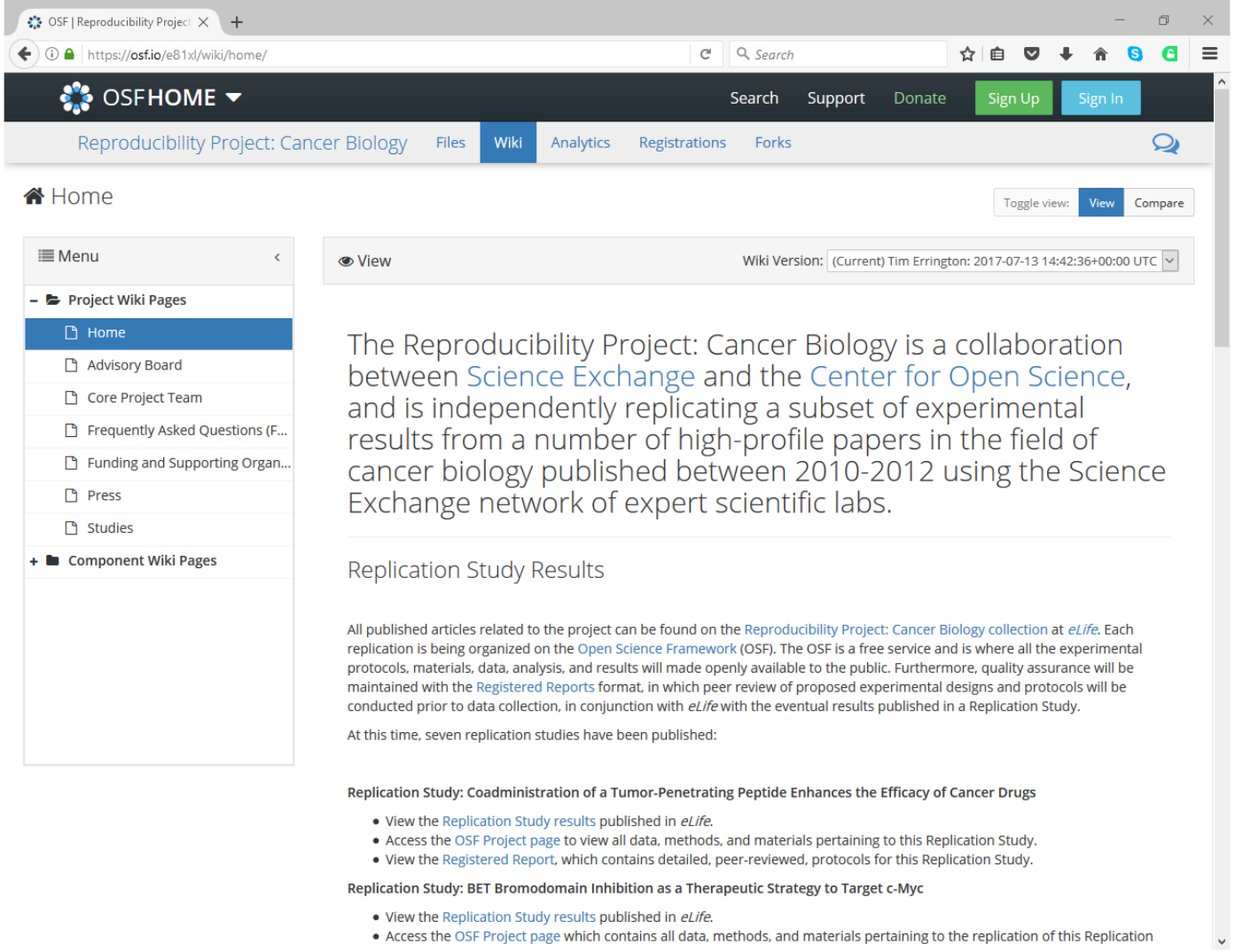


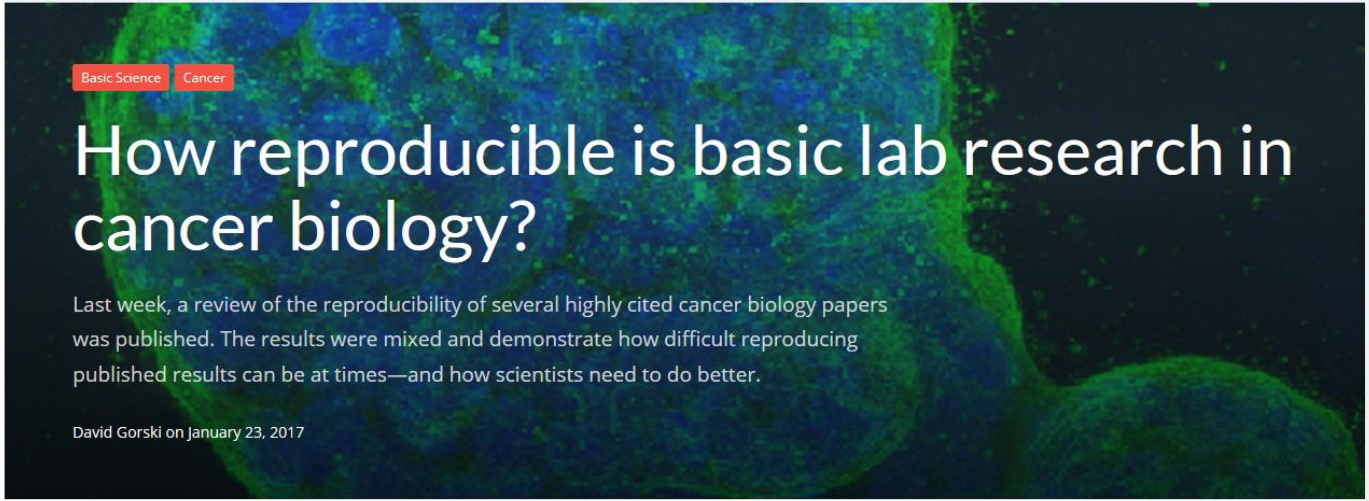
Reproducibility

"Mostly due to current methods capture and data malpractice, approximately 50% of all research data and experiments is considered not reproducible, and the vast majority (likely over 80%) of data never makes it to a trusted and sustainable repository."

Source: ***Realising the European Open Science Cloud, EC DG Research & Innovation 2016***
http://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf#view=fit&pagemode=none

FOSTER


OSF | Reproducibility Project
https://osf.io/e81xl/wiki/home/
Search
OSFHOME
Search
Support
Donate
Sign Up
Sign In
Reproducibility Project: Cancer Biology
Files
Wiki
Analytics
Registrations
Forks
Home
Toggle view:
View
Compare
Menu
View
Wiki Version:
(Current) Tim Errington: 2017-07-13 14:42:36+00:00 UTC
Project Wiki Pages
Home
Advisory Board
Core Project Team
Frequently Asked Questions (F...
Funding and Supporting Organ...
Press
Studies
Component Wiki Pages
The Reproducibility Project: Cancer Biology is a collaboration between Science Exchange and the Center for Open Science, and is independently replicating a subset of experimental results from a number of high-profile papers in the field of cancer biology published between 2010-2012 using the Science Exchange network of expert scientific labs.
Replication Study Results
All published articles related to the project can be found on the Reproducibility Project: Cancer Biology collection at eLife. Each replication is being organized on the Open Science Framework (OSF). The OSF is a free service and is where all the experimental protocols, materials, data, analysis, and results will made openly available to the public. Furthermore, quality assurance will be maintained with the Registered Reports format, in which peer review of proposed experimental designs and protocols will be conducted prior to data collection, in conjunction with eLife with the eventual results published in a Replication Study.
At this time, seven replication studies have been published:
Replication Study: Coadministration of a Tumor-Penetrating Peptide Enhances the Efficacy of Cancer Drugs
View the Replication Study results published in eLife.
Access the OSF Project page to view all data, methods, and materials pertaining to this Replication Study.
View the Registered Report, which contains detailed, peer-reviewed, protocols for this Replication Study.
Replication Study: BET Bromodomain Inhibition as a Therapeutic Strategy to Target c-Myc
View the Replication Study results published in eLife.
Access the OSF Project page which contains all data, methods, and materials pertaining to the replication of this Replication



Basic Science Cancer

How reproducible is basic lab research in cancer biology?

Last week, a review of the reproducibility of several highly cited cancer biology papers was published. The results were mixed and demonstrate how difficult reproducing published results can be at times—and how scientists need to do better.

David Gorski on January 23, 2017

When deciding what to write about this week, I had thought about expounding on, for instance, my concerns regarding vaccine policy given the new administration, but I think I've done enough of that for the moment at my not-so-super-secret other blog. Besides, there will be plenty of time and many opportunities to return to my concerns in that area

Reproducibility

Method Reproducibility

the provision of enough detail about study procedures and data so the same procedures could, in theory or in actuality, be exactly repeated.

Result Reproducibility (aka replicability)

obtaining the same results from the conduct of an independent study whose procedures are as closely matched to the original experiment as possible

What does research reproducibility mean? Steven N. Goodman, Daniele Fanelli, John P. A. Ioannidis Science Translational Medicine 8 (341), 341ps12.
[doi: 10.1126/scitranslmed.aaf5027]
http://stm.sciencemag.org/content/scitransmed/8/341/341ps12.full.pdf

Reproducibility in the research workflow

assessment

- check (statistical) methods /reporting
- welcome replication studies

publication

- use executable/forkable publications
- use IDs for preregistrations, data, methods, materials, contributors

preparation

- pre-register (can be embargoed)

experimenting / analysis

- share protocols, scripts
- use materials ids (RRIDs)
- use open hardware
- document steps, file management
- share data

Preregistration Challenge

https://cos.io/prereg/

Search

CENTER FOR OPEN SCIENCE

About COS | Our Products | Our Services | Our Communities | Blog | Contact Us | Donate to COS

Help support open science today. Donate Now

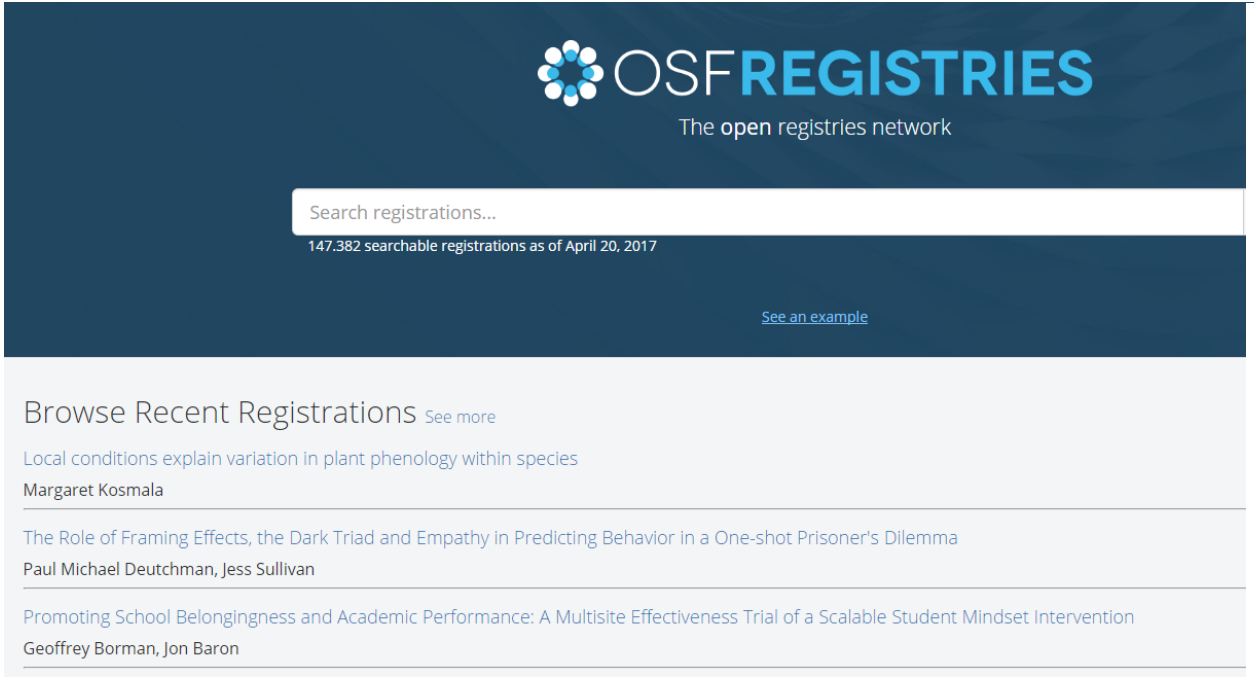
Preregistration makes your science better.

PREREGISTRATION CHALLENGE

If you have a project that is entering the planning or data collection phase, we'd like you to try out a preregistration. Through our **$1 Million Preregistration Challenge**, we're giving away $1,000 to 1,000 researchers who preregister their projects before they publish them. It's straightforward to complete and will really enhance your research output.

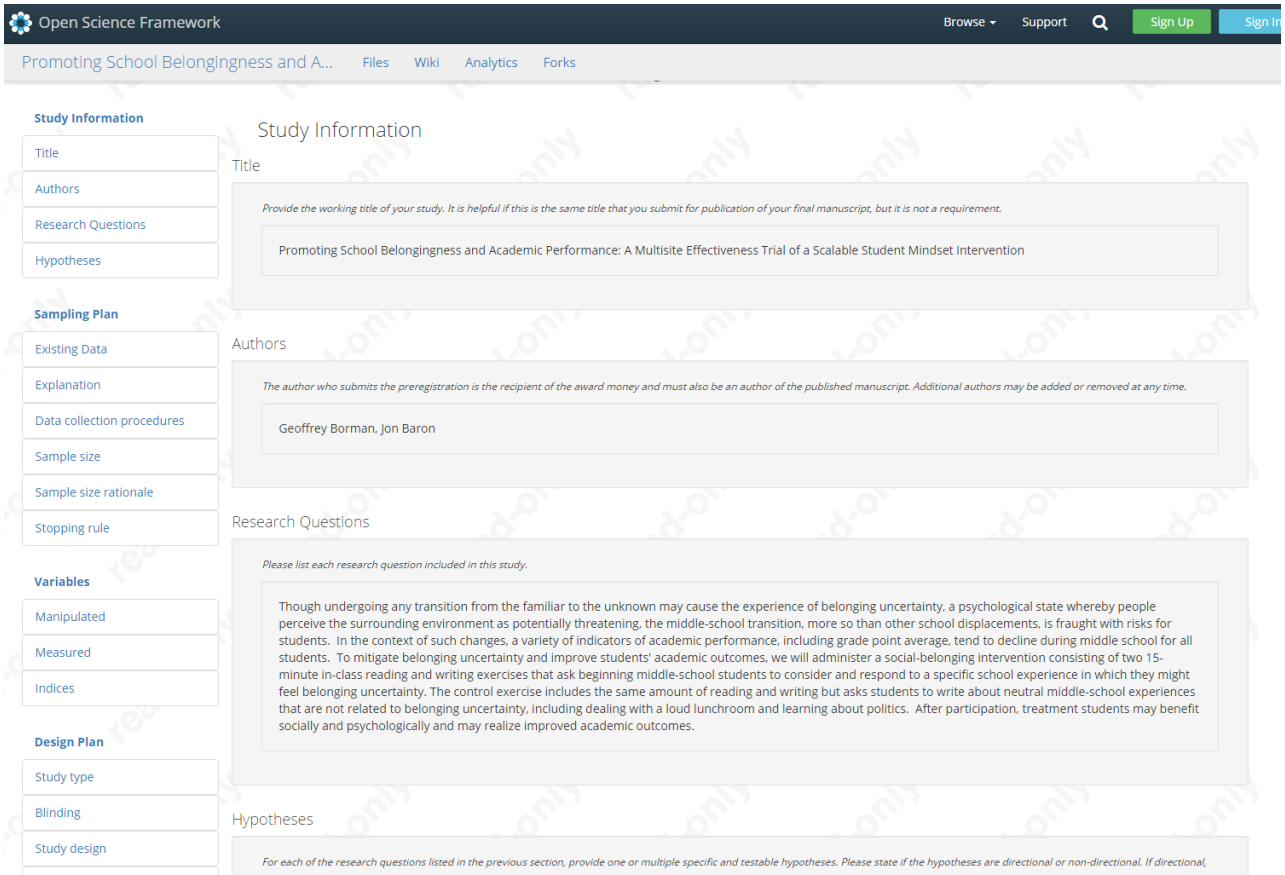
Get Started Now

Pre-registering, e.g. at OSF or AsPredicted

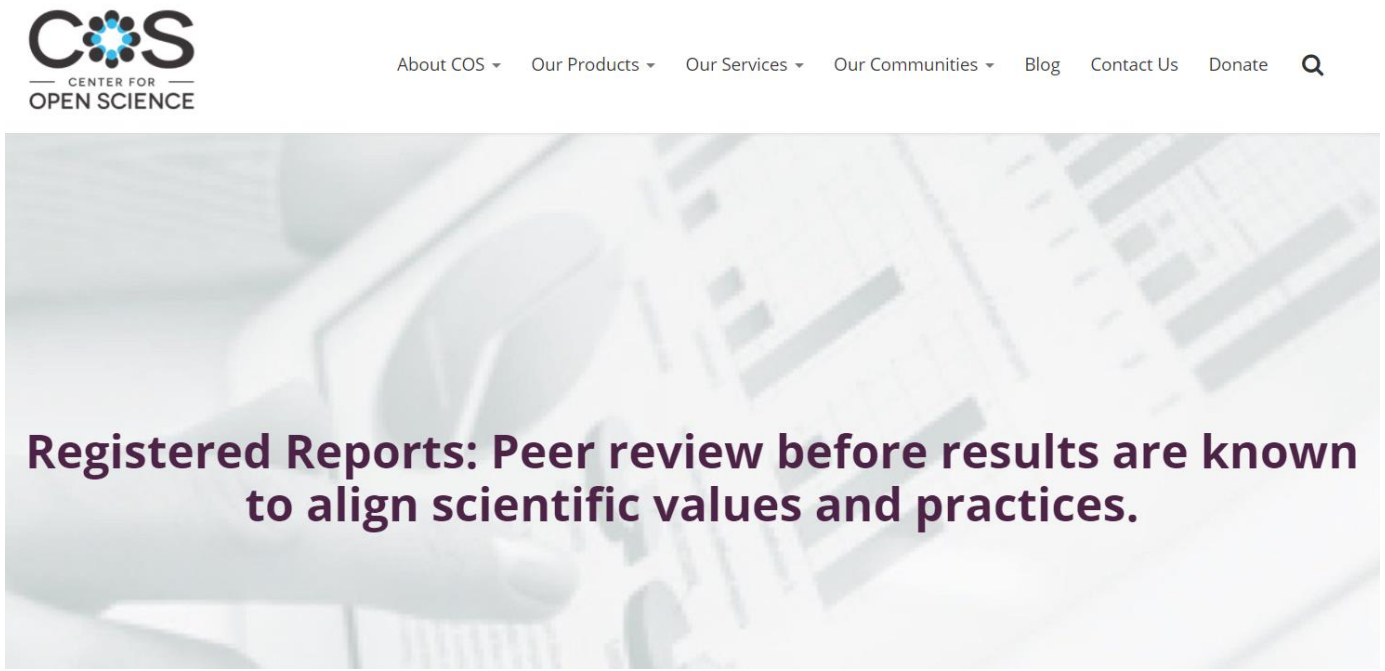


Make it easy to verify your hypothesis and analysis plans. Prevent p-hacking

Pre-registering, e.g. at OSF or AsPredicted

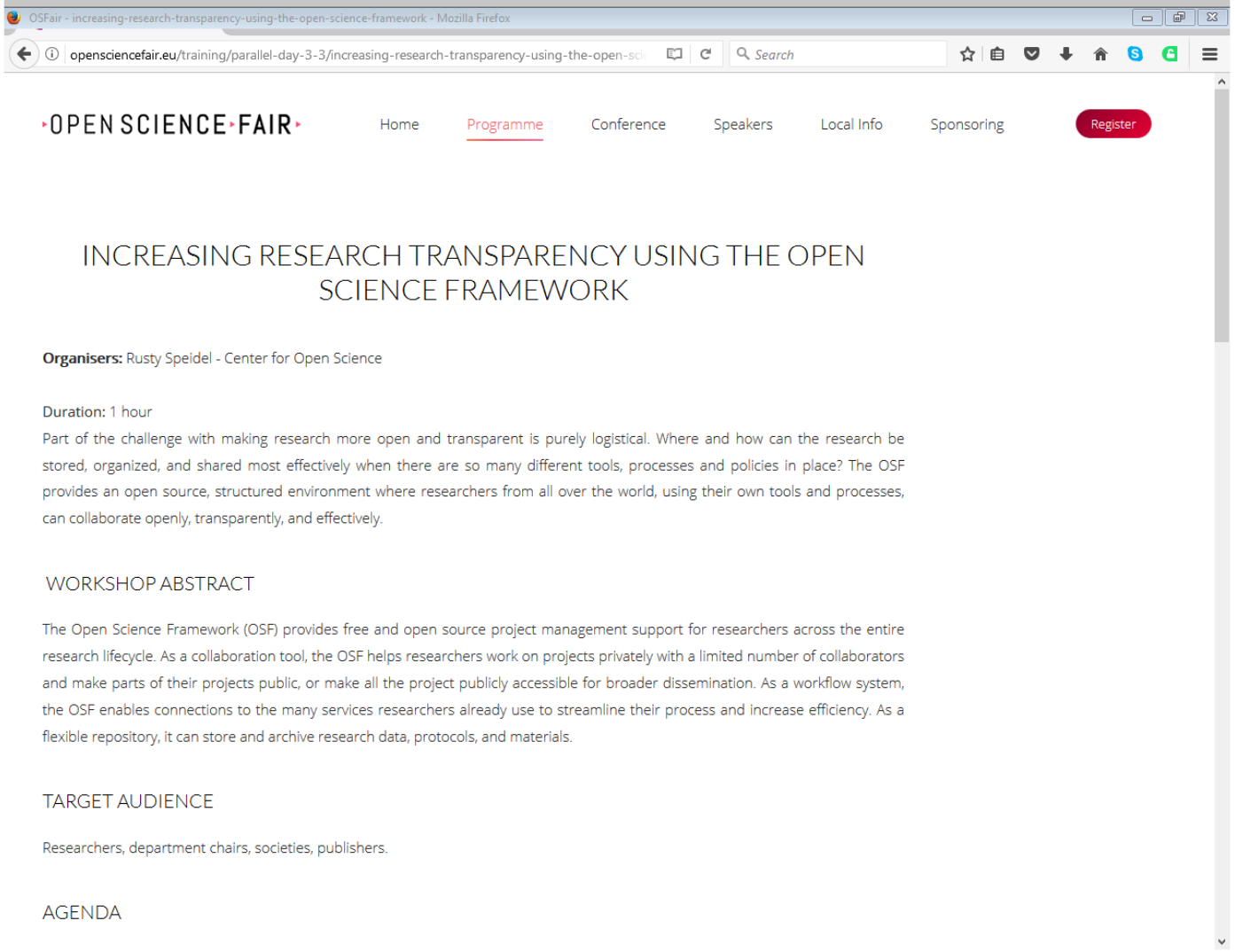


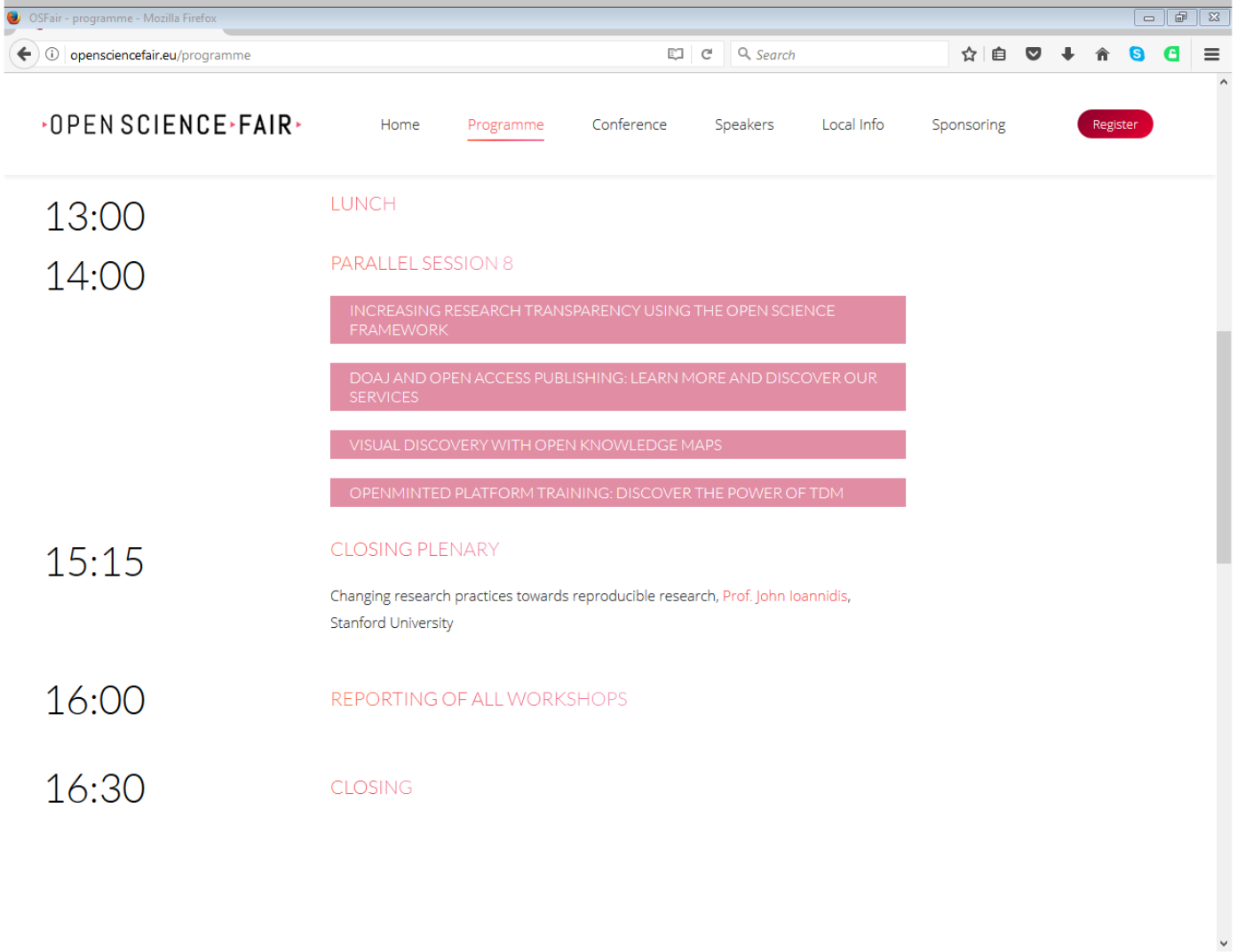
Pre-registering, e.g. at OSF or AsPredicted



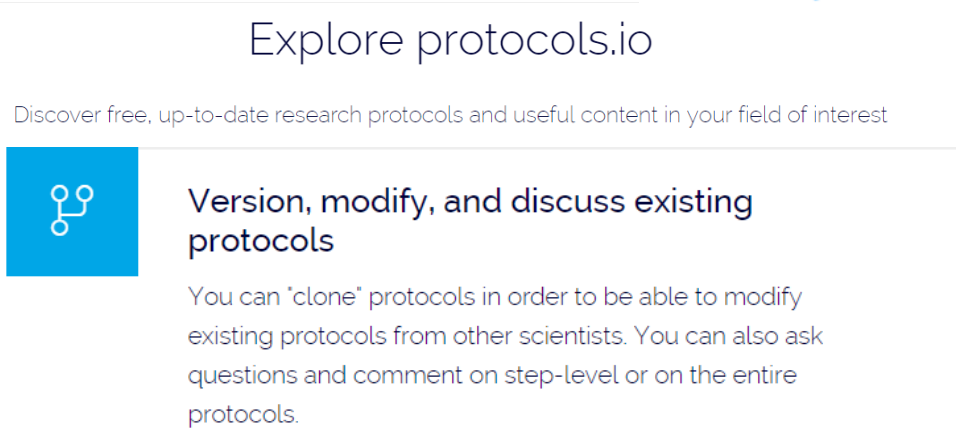
Aspredicted.org

Source: https://aspredicted.org/

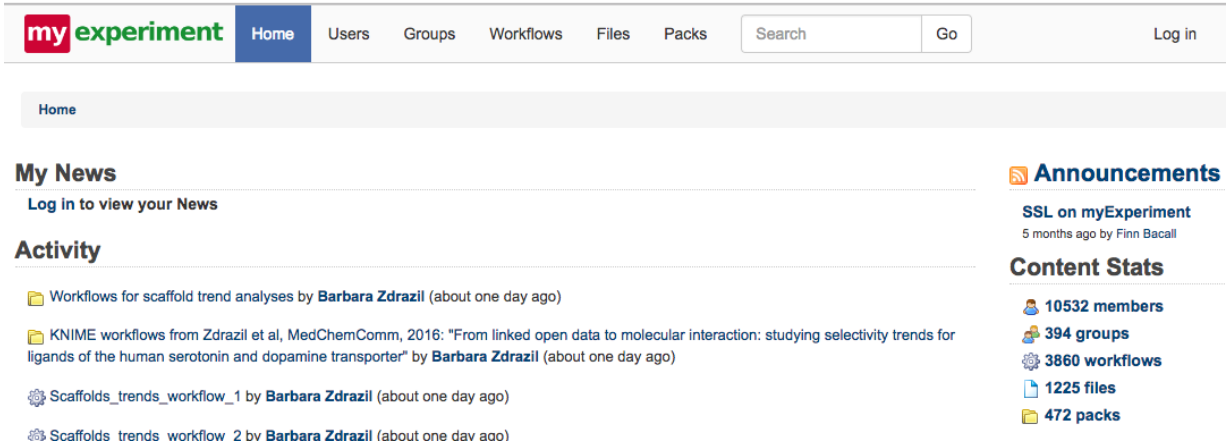

OSFair - increasing-research-transparency-using-the-open-science-framework - Mozilla Firefox
opensciencefair.eu/training/parallel-day-3-3/increasing-research-transparency-using-the-open-sci
Search
OPEN SCIENCE FAIR
Home
Programme
Conference
Speakers
Local Info
Sponsoring
Register
INCREASING RESEARCH TRANSPARENCY USING THE OPEN SCIENCE FRAMEWORK
Organisers: Rusty Speidel - Center for Open Science
Duration: 1 hour
Part of the challenge with making research more open and transparent is purely logistical. Where and how can the research be stored, organized, and shared most effectively when there are so many different tools, processes and policies in place? The OSF provides an open source, structured environment where researchers from all over the world, using their own tools and processes, can collaborate openly, transparently, and effectively.
WORKSHOP ABSTRACT
The Open Science Framework (OSF) provides free and open source project management support for researchers across the entire research lifecycle. As a collaboration tool, the OSF helps researchers work on projects privately with a limited number of collaborators and make parts of their projects public, or make all the project publicly accessible for broader dissemination. As a workflow system, the OSF enables connections to the many services researchers already use to streamline their process and increase efficiency. As a flexible repository, it can store and archive research data, protocols, and materials.
TARGET AUDIENCE
Researchers, department chairs, societies, publishers.
AGENDA


OSFair - programme - Mozilla Firefox
opensciencefair.eu/programme
Search
OPEN SCIENCE FAIR
Home
Programme
Conference
Speakers
Local Info
Sponsoring
Register
13:00
LUNCH
14:00
PARALLEL SESSION 8
INCREASING RESEARCH TRANSPARENCY USING THE OPEN SCIENCE FRAMEWORK
DOAJ AND OPEN ACCESS PUBLISHING: LEARN MORE AND DISCOVER OUR SERVICES
VISUAL DISCOVERY WITH OPEN KNOWLEDGE MAPS
OPENMINTED PLATFORM TRAINING: DISCOVER THE POWER OF TDM
15:15
CLOSING PLENARY
Changing research practices towards reproducible research, Prof. John Ioannidis, Stanford University
16:00
REPORTING OF ALL WORKSHOPS
16:30
CLOSING

Sharing methods and materials, e.g. at Protocols.io or RRID



MyExperiment – research workflow



Source: https://www.myexperiment.org/home

sharing notebooks e.g. at ONSNetwork or OSF

Computing – Oly BGI GBS Reproducibility; fail?

OK, so things have improved since the last attempt at getting this BGI script to run and demultiplex the raw data.

I played around with the index.lst file format (based on the error I received last time, it seemed like a good possibility that the file formatting was incorrect) and actually got the script to run to completion! Granted, it took over 16hrs (!!), but it completed!

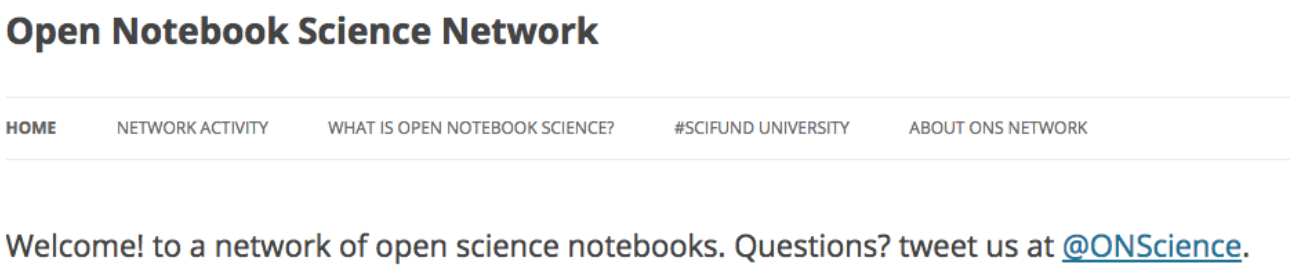
See the Jupyter notebook link below.

Results:

Well, although the script finished and kicked out all the demultiplexed FASTQ files, the contents of the FASTQ files don't match (the read counts differ between these results and the BGI files) the original set of demultiplexed files. I'm not entirely sure if this is to be expected or not, since the script allows for a single nucleotide mismatch when demultiplexing. Is it possible that the mismatch could be interpreted slightly differently each time this is run? I'm not certain.

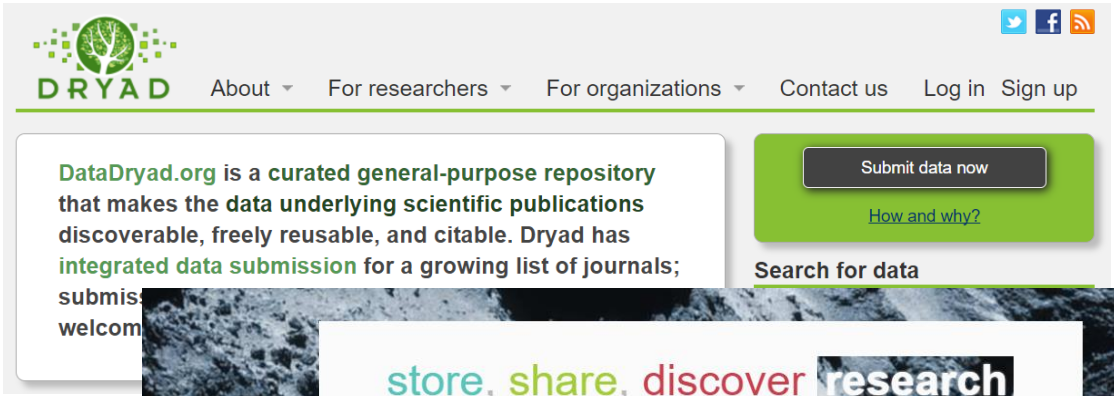
Get feedback from peers, help form your thoughts, feel less alone while doing the analyses. Spot mistakes early on.

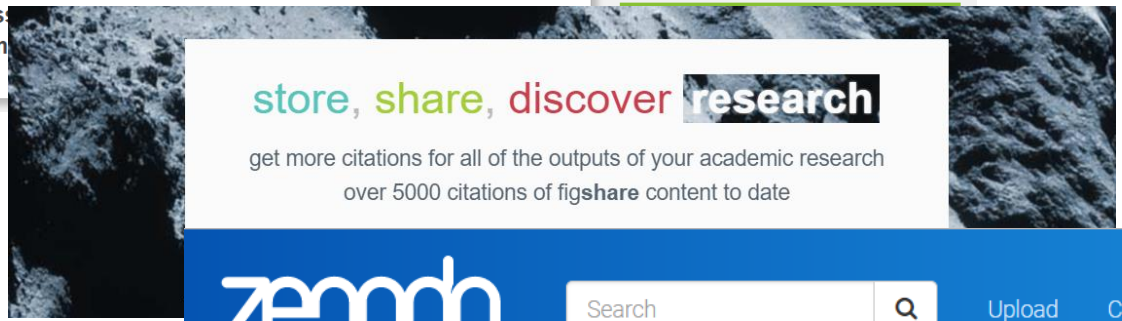
Open Notebook Science Network



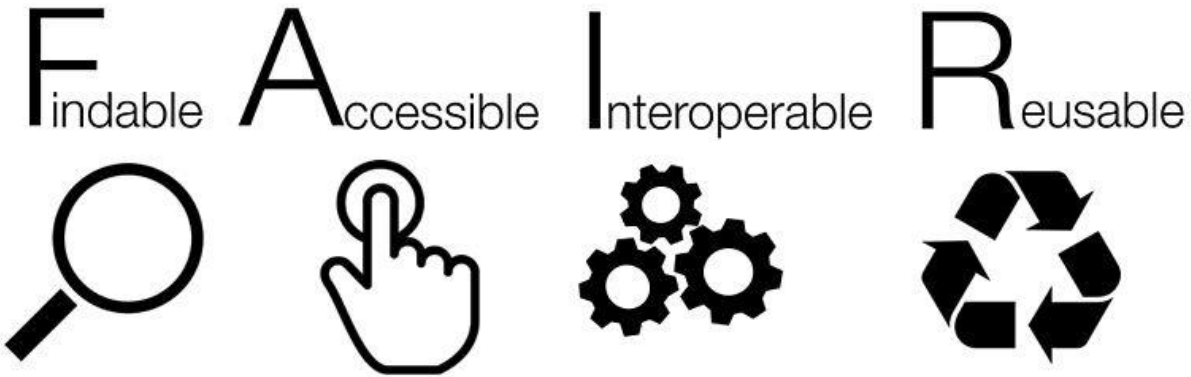
Source: http://onsnetwork.org/

Sharing data, e.g. at Dryad, Figshare or Zenodo

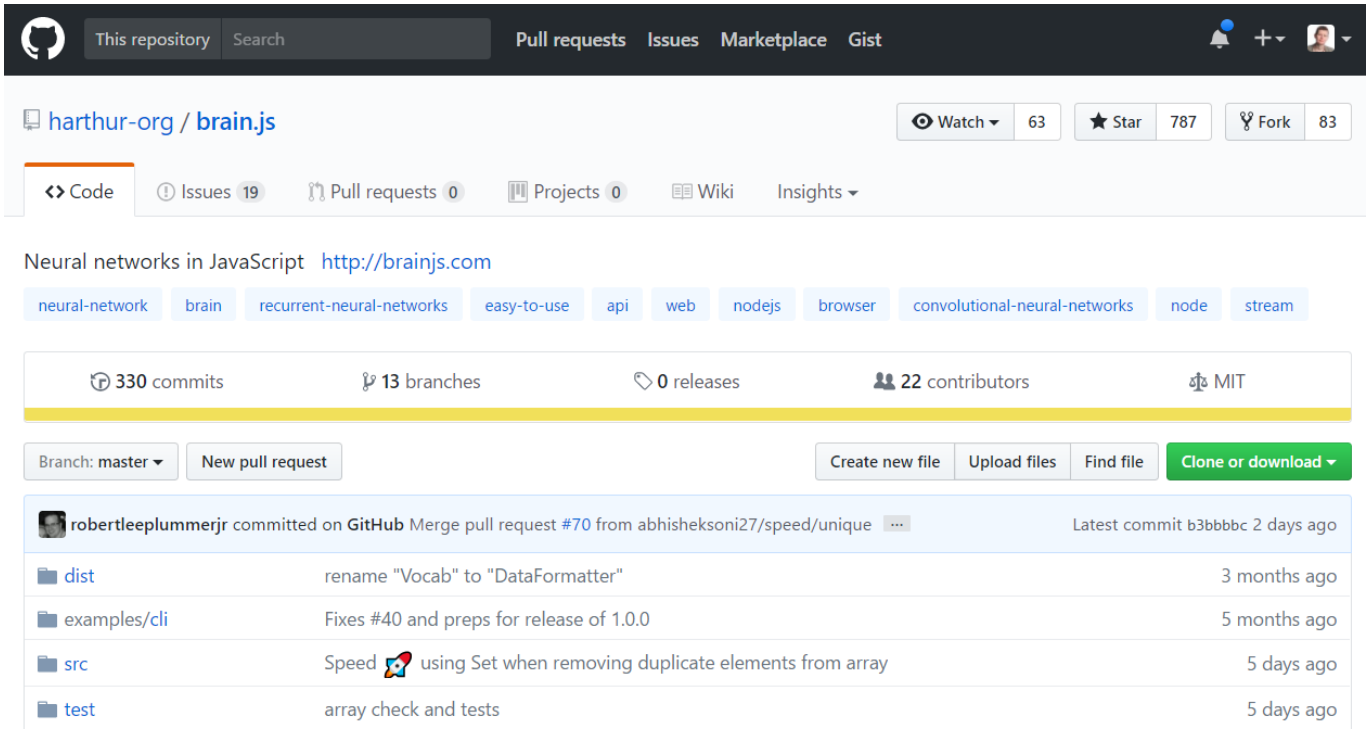




Sharing research data,
e.g. at Dryad, Figshare or Zenodo



Sharing code e.g. at GitHub with GNU OR MIT license



Get people to check, contribute to and use and build on your code

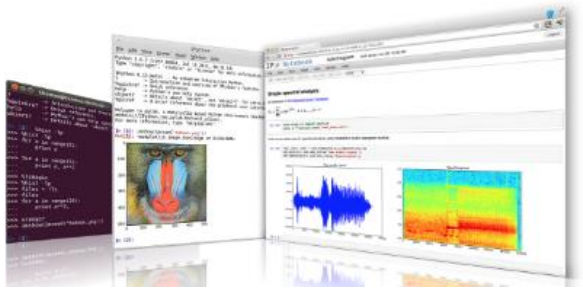
Ipython - notebook

IP[y]: IPython
Interactive Computing

Install · Documentation · Project · Jupyter · News · Cite · Donate · Books

IPython provides a rich architecture for interactive computing with:

- A powerful interactive shell.
- A kernel for Jupyter.
- Support for interactive data visualization and use of GUI toolkits.
- Flexible, embeddable interpreters to load into your own projects.
- Easy to use, high performance tools for parallel computing.



Source: http://ipython.org/notebook.html

Peer Reviewers' Openness Initiative

Supporting the spread of open research practices

Search ...

RECENT POSTS

PRO Initiative media

RECENT COMMENTS

Richard Morey on Making your materials public

Richard Morey on Making your materials public

Matt Superdock on Making your materials public

Tomasz Witkowski on Signatories

Graeme Armstrong on Join the Initiative

ARCHIVES

November 2015

CATEGORIES

Uncategorized

META

Log in

Entries RSS

Comments RSS

WordPress.org

THE PRO INITIATIVE for open science

We believe that openness and transparency are core values of science. For a long time, technological obstacles existed preventing transparency from being the norm. With the advent of the internet, however, these obstacles have largely disappeared. The promise of open research can finally be realized, but this will require a cultural change in science. The power to create that change lies in the peer-review process.

We suggest that beginning January 1, 2017, **reviewers make open practices a pre-condition for more comprehensive review.** This is already in reviewers' power; to drive the change, all that is needed is for reviewers to collectively agree that the time for change has come.

Read the paper about the PRO initiative

- Read the Initiative
- Read our guidelines for reviewers
- PRO-related media
- **See the signatories**

What do researchers use ?

Analyze

What tools/sites do you use to analyze data / texts etc.?

Choose as many as you like

A R

B SPSS

C Matlab

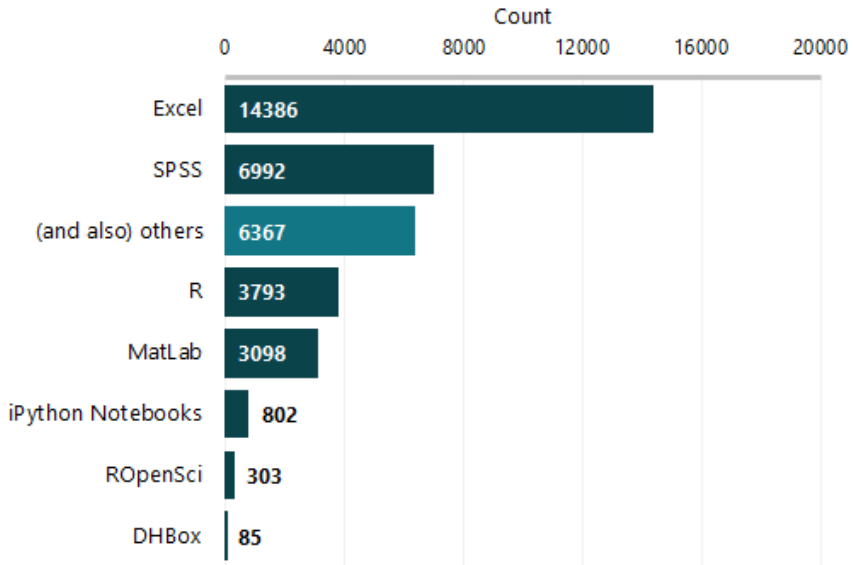
D Excel

E iPython Notebook

F ROpenSci

G DHbox

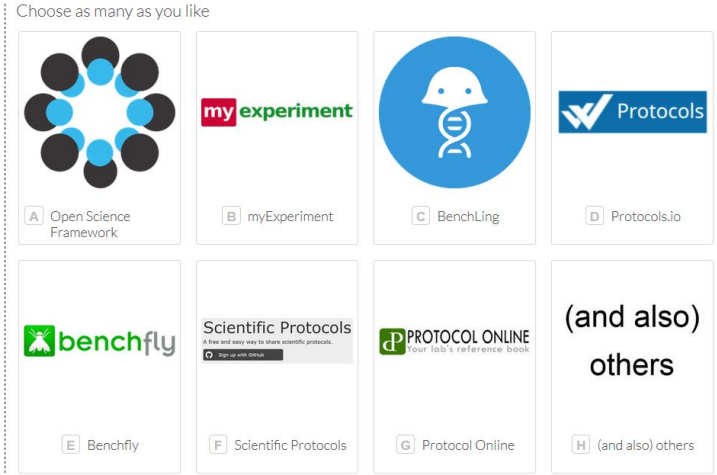
H (and also) others

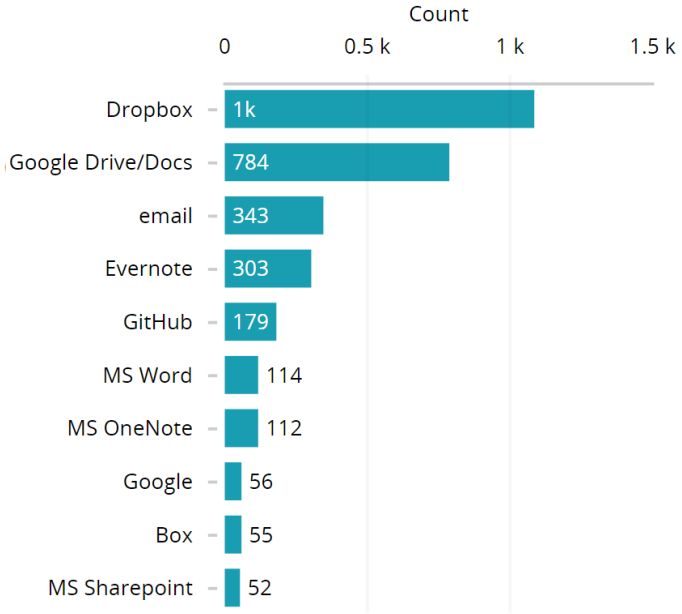


Share notebooks / protocols

What tools/sites do you use to share notebooks / protocols / workflows?

Choose as many as you like

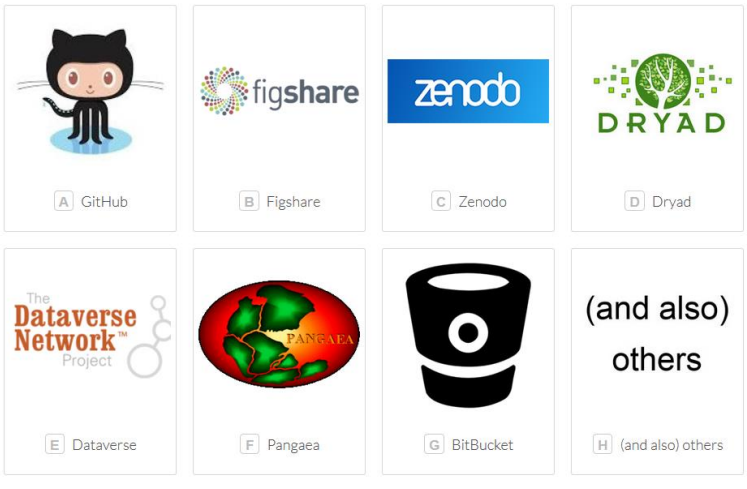


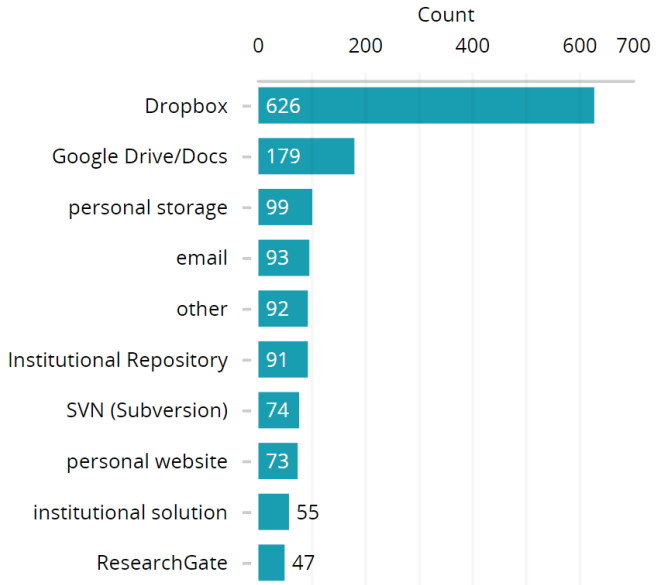


Archive / share data & code

What tools/sites do you use to archive/share data & code?

Choose as many as you like





Open Peer Review

PEER REVIEW AT THE CROSSROADS

Organisers: Edit Görögh, Birgit Schmidt, Tony Ross-Hellauer - Ugottingen
Duration: 1.5 hours
The workshop will present and discuss tools and methods related to open and alternative peer review.

WORKSHOP ABSTRACT

The workshop builds on the results of the OpenUp landscape scan and the OpenAIRE report on open peer review. The workshop has multiple purposes including (1) assessing existing and evolving methods and functions of alternative peer review mechanisms, (2) breaking down peer review into the basic processes to identify the benefits and challenges, and (3) identifying questions and issues that need further investigation.

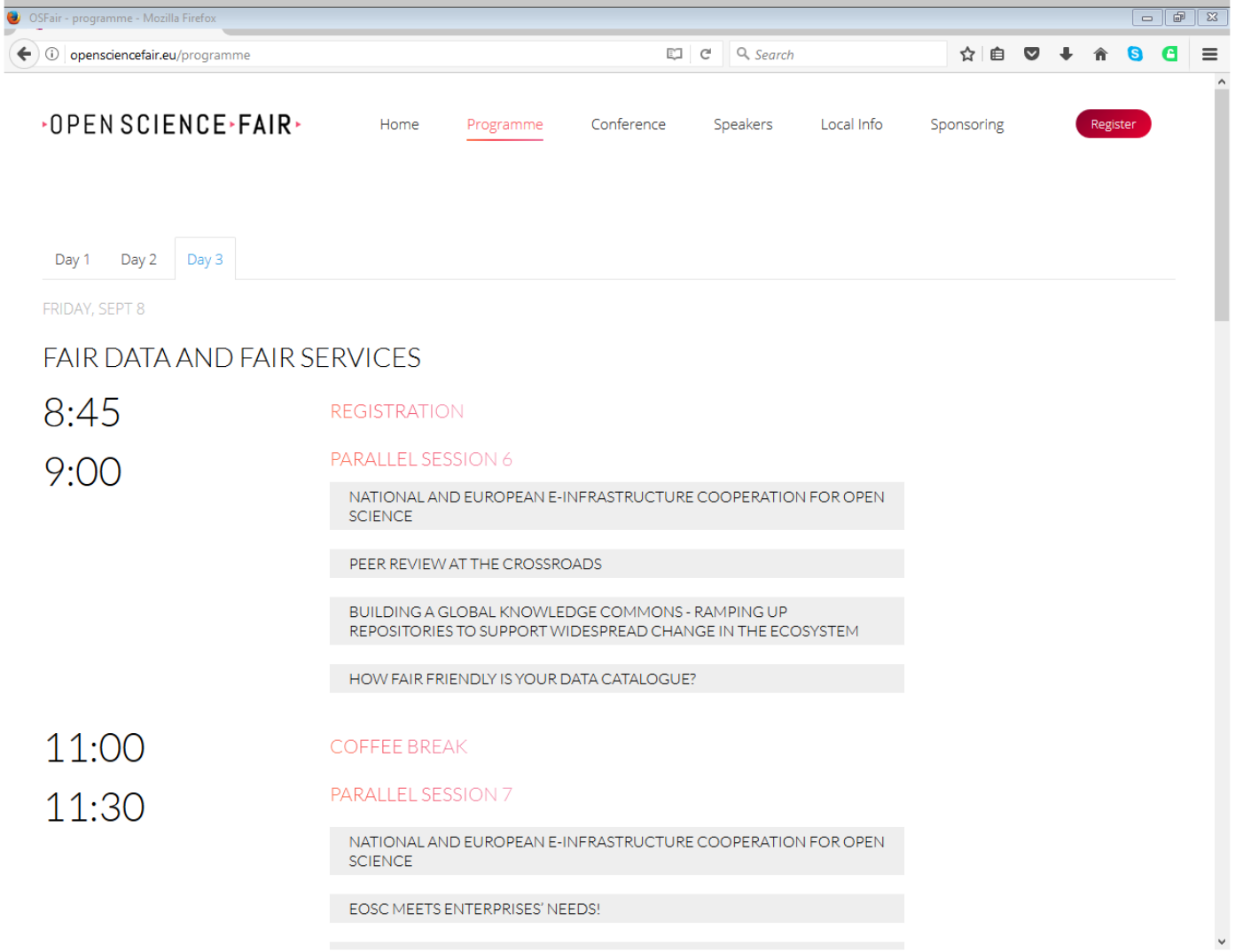
Group discussions will also touch upon issues such as the sustainability, long-term availability of alternative review tools, and their uptake by researchers, and the incorporation of these methods into institutional, national, funders' and publishers' policies.

OpenUP and OpenAIRE are dedicated to engage with different (disciplinary, inter-disciplinary) research communities from the social sciences, life sciences, energy, arts and humanities to identify the requirements from the emerging trends as posed by Open Science and e-infrastructural interconnected environments. Both projects aim at developing a sustainable framework that is relevant for and responsive to the Open Science needs.

External speakers from the research, publisher and infra communities will ensure a broad range of perspectives.

KEYWORDS

Peer review, open/alternative peer review tools and methods, Open Science, Framework


OSFair - programme - Mozilla Firefox
opensciencefair.eu/programme
Search
OPEN SCIENCE FAIR
Home
Programme
Conference
Speakers
Local Info
Sponsoring
Register
Day 1
Day 2
Day 3
FRIDAY, SEPT 8
FAIR DATA AND FAIR SERVICES
8:45
REGISTRATION
9:00
PARALLEL SESSION 6
NATIONAL AND EUROPEAN E-INFRASTRUCTURE COOPERATION FOR OPEN SCIENCE
PEER REVIEW AT THE CROSSROADS
BUILDING A GLOBAL KNOWLEDGE COMMONS - RAMPING UP REPOSITORIES TO SUPPORT WIDESPREAD CHANGE IN THE ECOSYSTEM
HOW FAIR FRIENDLY IS YOUR DATA CATALOGUE?
11:00
COFFEE BREAK
11:30
PARALLEL SESSION 7
NATIONAL AND EUROPEAN E-INFRASTRUCTURE COOPERATION FOR OPEN SCIENCE
EOSC MEETS ENTERPRISES' NEEDS!

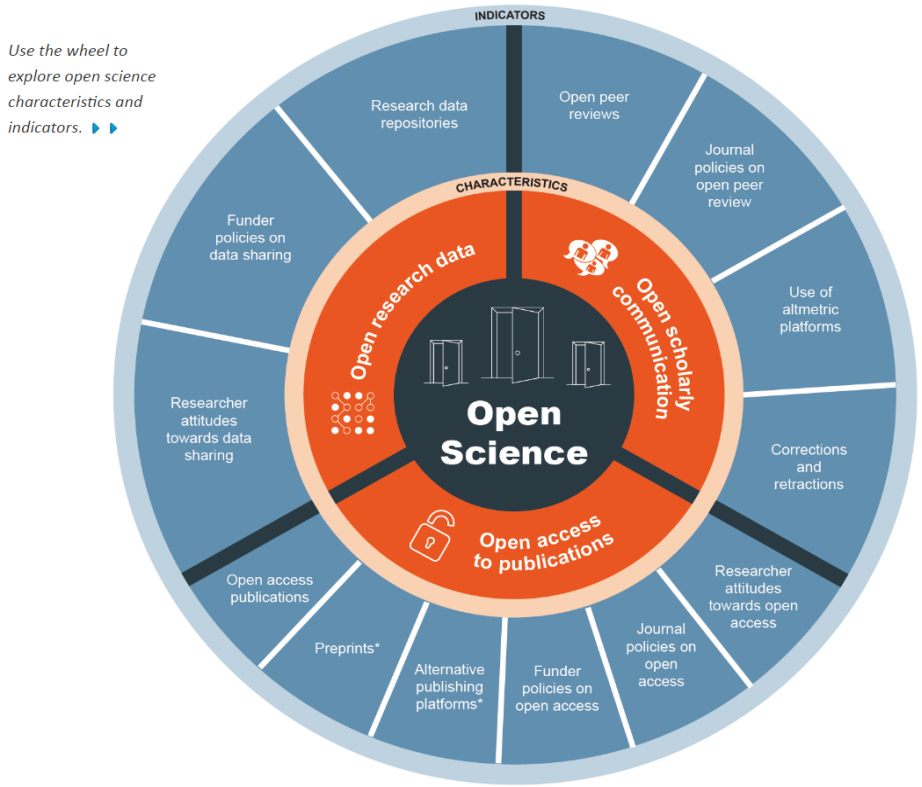
Disciplinary variety and Open Science

	ARTS & HUMANITIES	SOCIAL SCIENCE	LIFE/HEALTH	PHYSICAL SCIENCES
Research types	often exploratory research	often confirmatory research	often confirmatory research	often confirmatory research?
Data	often texual data	also qualitative data, sometimes sensitive data	sensitive patient data / big datasets	big datasets
Publ. Types	books, chapters, articles	mostly articles and chapters	mostly articles, (syst.) reviews	preprints, conf papers, articles
Collaboration	typically 1	typically 1-4	typically 3-10	typically 3-many
Languages	native languge & some English	English, some native languages	English	English
Funding	small scale funding	small & medium scale funding	large scale funding	large scale funding
Review	double blind	double + single blind	single blind	single blind

Research characteristics and Open Science options/issues

Characteristic	Open Science options/issues
Research types	Preregistration different for exploratory research
Data	Costs of archiving large datasets / conderations of anonimity/sensitiveness / Patentable code/outcomes
Publ. Types	Limited OA Book options / Book publishers small and many / Preprints
Collaboration	Reaching agreements with co-authors
Languages	Not all languages always accepted / Non-native English researcherds at disadvantage
Funding	Large projects have funding but may ‘dictate’ way of archiving/publishing/communicating
Review	Closed and blind variants of peer review are deeply rooted

Open Science monitor (European Union)

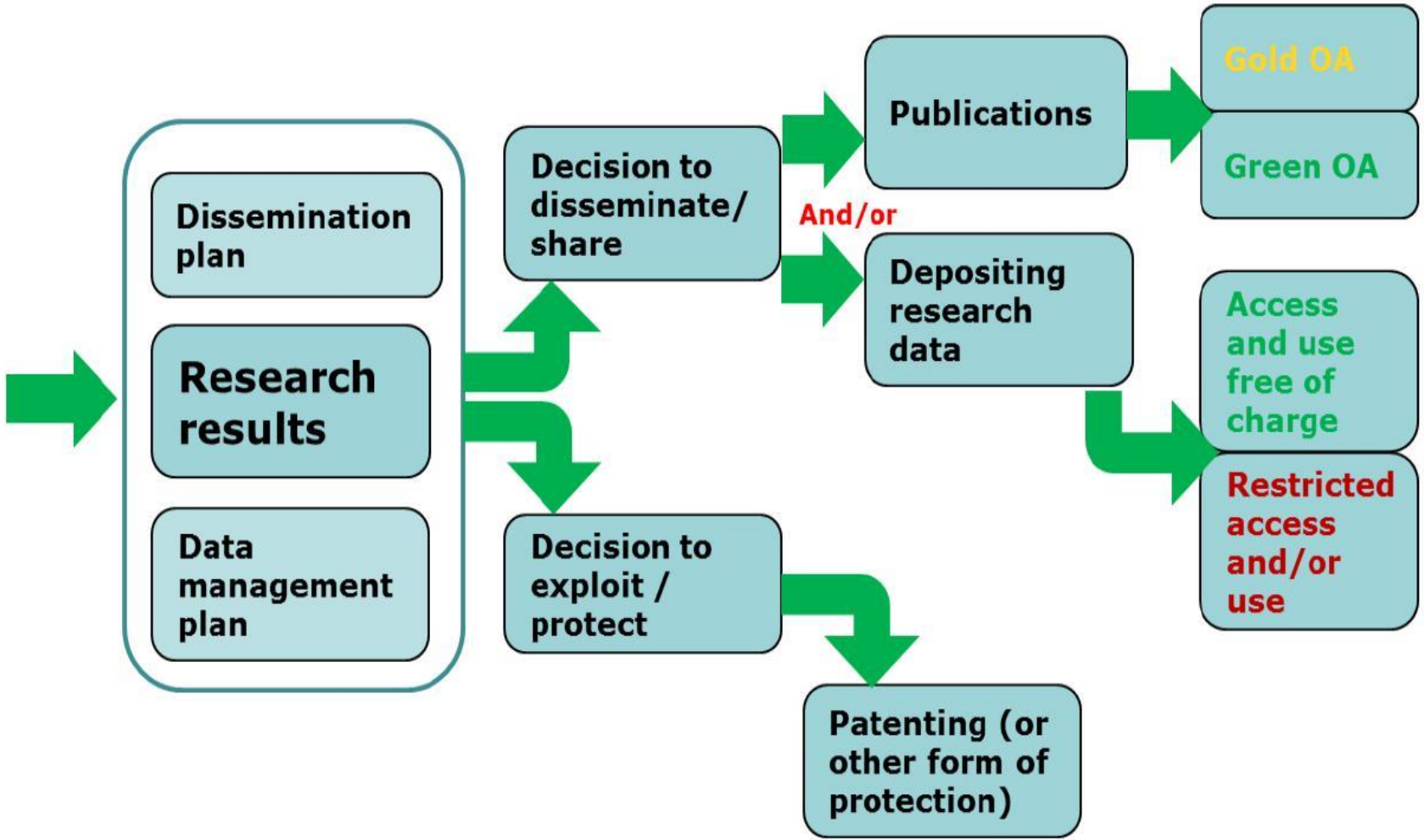


http://ec.europa.eu/research/openscience/index.cfm?pg=home&section=monitor

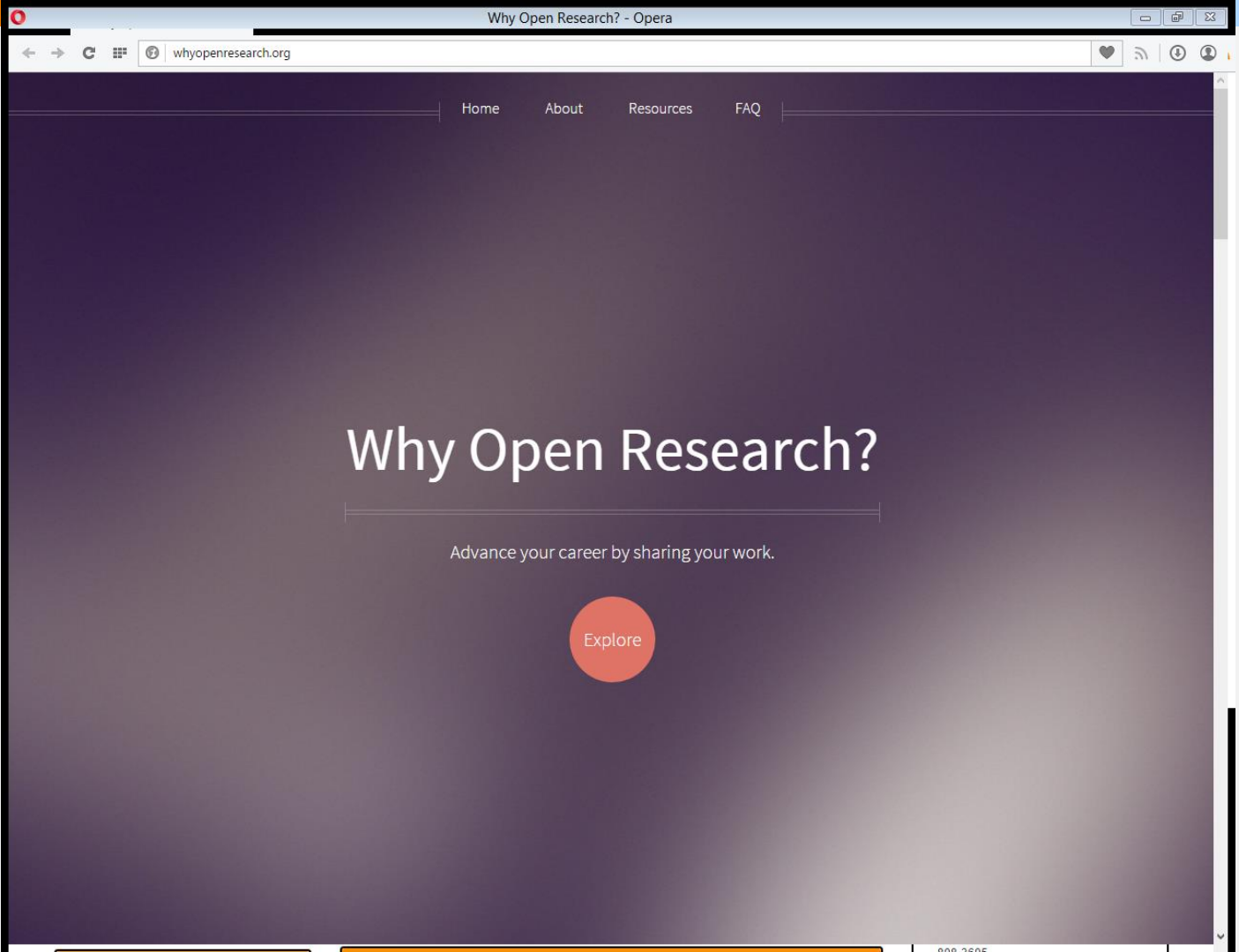
With thanks to Bianca Kramer & Jeroen Bosman, Utrecht University Library

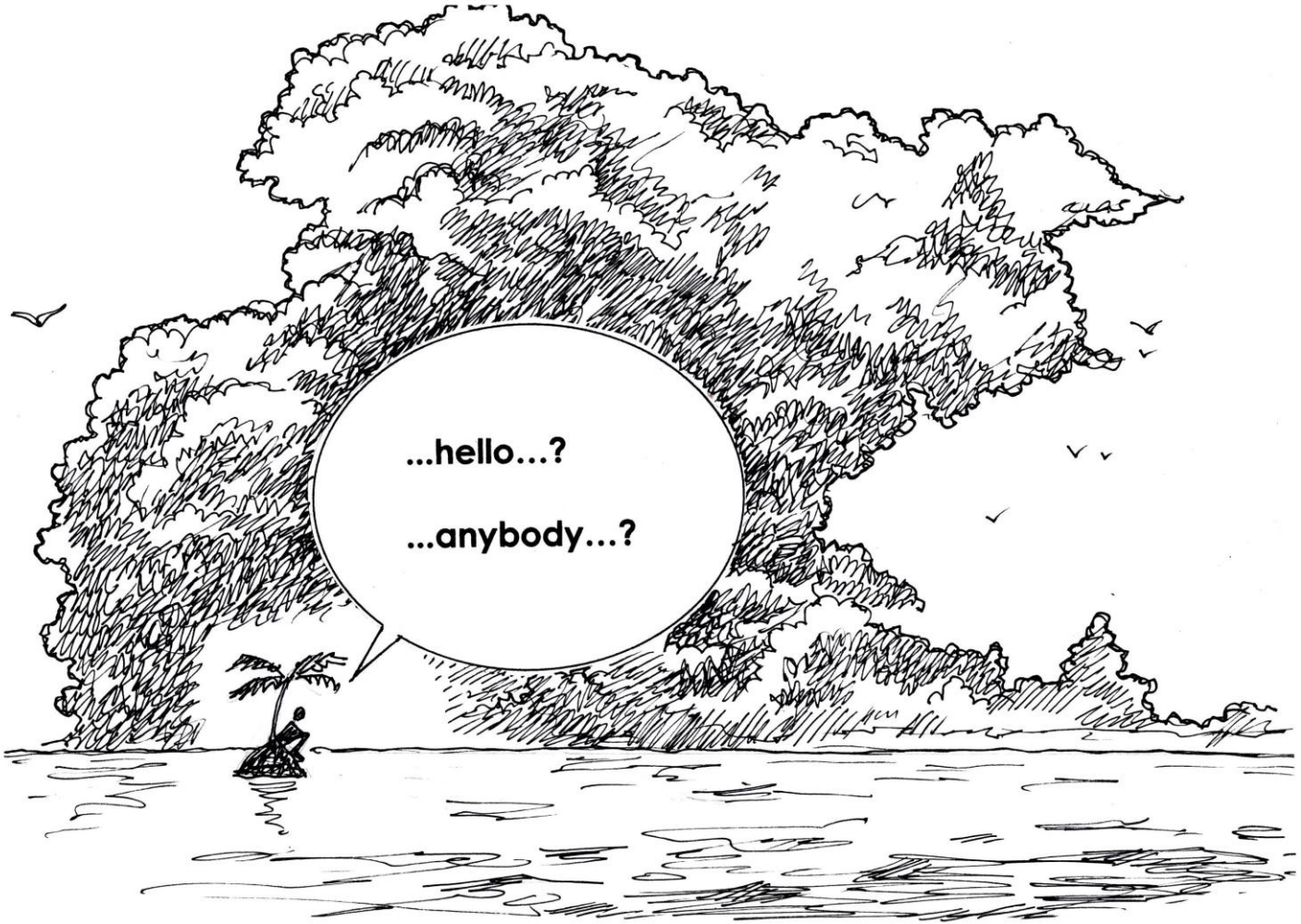
for re-using their slides presented at Open Access, Open Data, Open Science EIFL Train-the-trainer programme

@MsPhelps
@jeroenbosman


Research
Dissemination plan
Research results
Data management plan
Decision to disseminate/ share
Decision to exploit / protect
Publications
And/or
Depositing research data
Gold OA
Green OA
Access and use free of charge
Restricted access and/or use
Patenting (or other form of protection)

Open Access to publications


Why Open Research? - Opera
whyopenresearch.org
Home
About
Resources
FAQ
Why Open Research?
Advance your career by sharing your work.
Explore


...hello...?
...anybody...?

OA publishing: Costs

Find a no-cost open access (OA) journal - about 65% of fully OA journals do not charge for publishing an article

OA publishing: Costs (2)

Notable examples of OA journals that do not currently charge fees include:

ROYAL SOCIETY
OPEN SCIENCE

OA publishing: Costs (3)

It's important to note that researchers in *any* country can request a fee waiver if unable to pay

Examples of publisher fee waiver policies (non-exhaustive list)

BioMed Central's open access waiver fund

Hindawi's waiver policy

PLOS's
Global Participation Initiative

Take back control

“Know your rights. Keep your rights. You should decide how your work is used.”

“Many subscription publishers require authors to sign a copyright transfer agreement. Sign this and you no longer own your work, the publisher does. The publisher decides who can read, share, and reuse the content. Do you think you should have to ask the publisher for permission to reuse your own work? No? Then take back control.”

Negotiate to keep your rights

“You can negotiate the terms of your publishing agreement. Want to retain your copyright? Want the right to post a free copy in an open repository? Want to reuse the content? Simply ask. Submit an author addendum that describes the rights you want to retain. SPARC provides a template addendum [pdf]. More info is available here.”

Negotiate to keep your rights (2)

The Scholar's Copyright Addendum Engine can help you generate a customized author addendum.

Not all publishers will accept author addenda, but some will. And it never hurts to ask…

Consider publishing in an OA journal

“Instead of negotiating with a subscription publisher, you can go with an OA publisher and keep all your rights.

OA publishers do not require a copyright transfer agreement; authors retain copyright.”

Consider publishing in an OA journal (2)

“Articles are most often distributed under a Creative Commons Attribution (CC BY) license, which allows anyone to read, share, and reuse the content provided they attribute the original source. Creative Commons also has other licenses, depending on the types of reuse rights you as the copyright holder want to grant users of your work. Don't know which license to choose? This simple license selector can help.”


"My research behind the pay wall is of no use"
- Prateek Mahalwar, PhD Candidate at the Max Planck Institute for Developmental Biology
open access week


"Access to information is a human right, but it is often treated as a privilege. This has to change—and it will take all of us to make it happen."
- Erin McKiernan, physiologist, neuroscientist, and open access advocate
open access week

How to find a suitable OA journal?

1. Browse the Directory of Open Access Journals

The DOAJ is the foremost trusted listing of over 9,914 registered OA journals. Searches by subject, article processing charges, journal license, publisher, country of publisher, and full text language, type of peer review (blind, double-blind, open, etc.) are available.

FOSTER

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

How to find a suitable OA journal? (2)

2. Use the CoFactor Journal Selector Tool

The Cofactor Journal Selector Tool is not exclusive to OA journals, but allows authors to filter by several options, including whether the journal is fully open, or has an open publishing option (hybrid journal).

FOSTER

How to find a suitable OA journal? (3)

3. Talk to your mentors, librarians, and colleagues who focus on scholarly communication issues in your field

Scholars from many disciplines are seeking to promote the development and growth of rigorous OA publishing options in their fields.

FOSTER

Erin McKiernan

Erin McKiernan is a researcher in experimental and computational neuroscience, and an advocate for open access, open data, and open science. You can follow her on Twitter @emckiernan13

August 2014

Early career researchers / University research: if you believe in openness, stand up for it

Publishing openly provides greater exposure, boosts prospects and can lead to more citations, says **Erin McKiernan**

22 Aug 2014 34

Make a list of OA publication options in your particular field. Chances are you will be surprised by the range of possibilities

Erin McKiernan

Discuss access issues with your collaborators up front, before the research is done and the articles written.

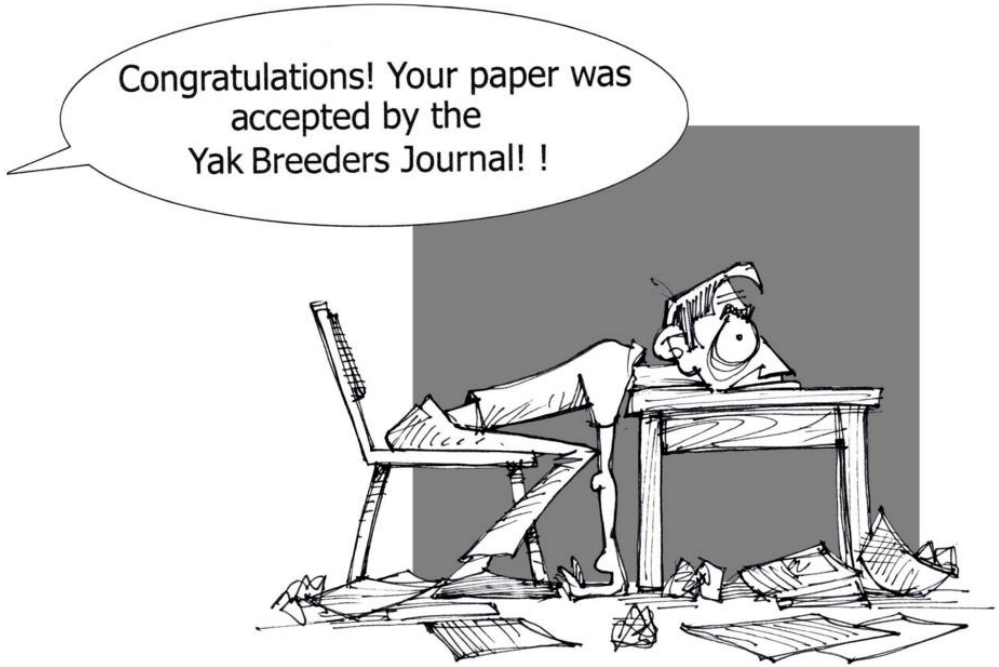
Erin McKiernan

If for some reason you do publish a closed-access article, remember that you can self-archive a copy of your article in a disciplinary or institutional or shared repository.

Erin McKiernan

Publish where you want

Publish in open journals, or in the journal of your choice and archive a free copy.



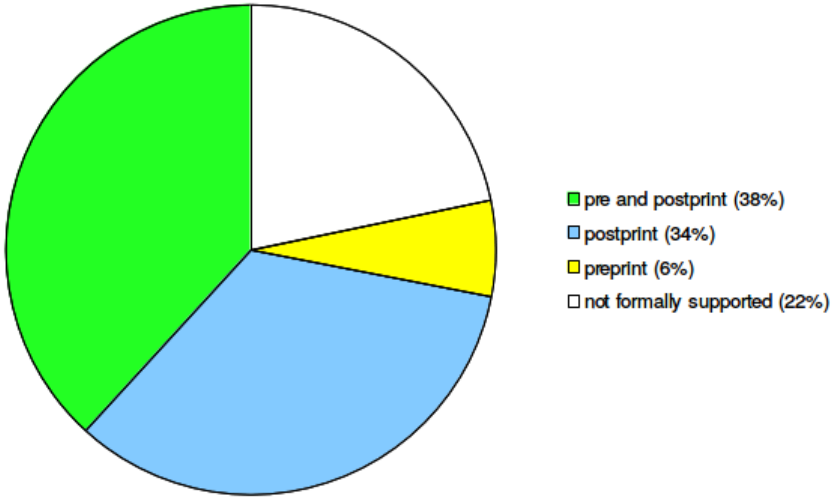
Publish where you want

“You don’t have to sacrifice quality or academic freedom to publish openly.

There are many high-quality OA journals. But if you can't find what you deem a suitable OA venue, it's important to remember that open publishing is not restricted to these journals. You can publish in any journal you like and make a copy of your manuscript available (self-archive) in an OA repository.”

FOSTER

~78% of publishers allow authors to openly archive a version of their published manuscript: Breakdown of archiving policies from over 2,100 publishers. Source: Data from SHERPA/RoMEO. Accessed October 2015 and plotted by E.C. McKiernan (CC BY)



SHERPA/RoMEO

. . . opening access to research

Publisher copyright policies & self-archiving

Search

◉ Journal titles or ISSNs ○ Publisher names

◉ Exact title ○ starts with ○ contains ○ ISSN

Advanced Search | Search | Reset

Use this site to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement.

Special RoMEO Pages

- RoMEO Statistics
- Application Programmers' Interface (API)
- Publisher Categories in RoMEO
- Definitions and Terms

Additions and Updates

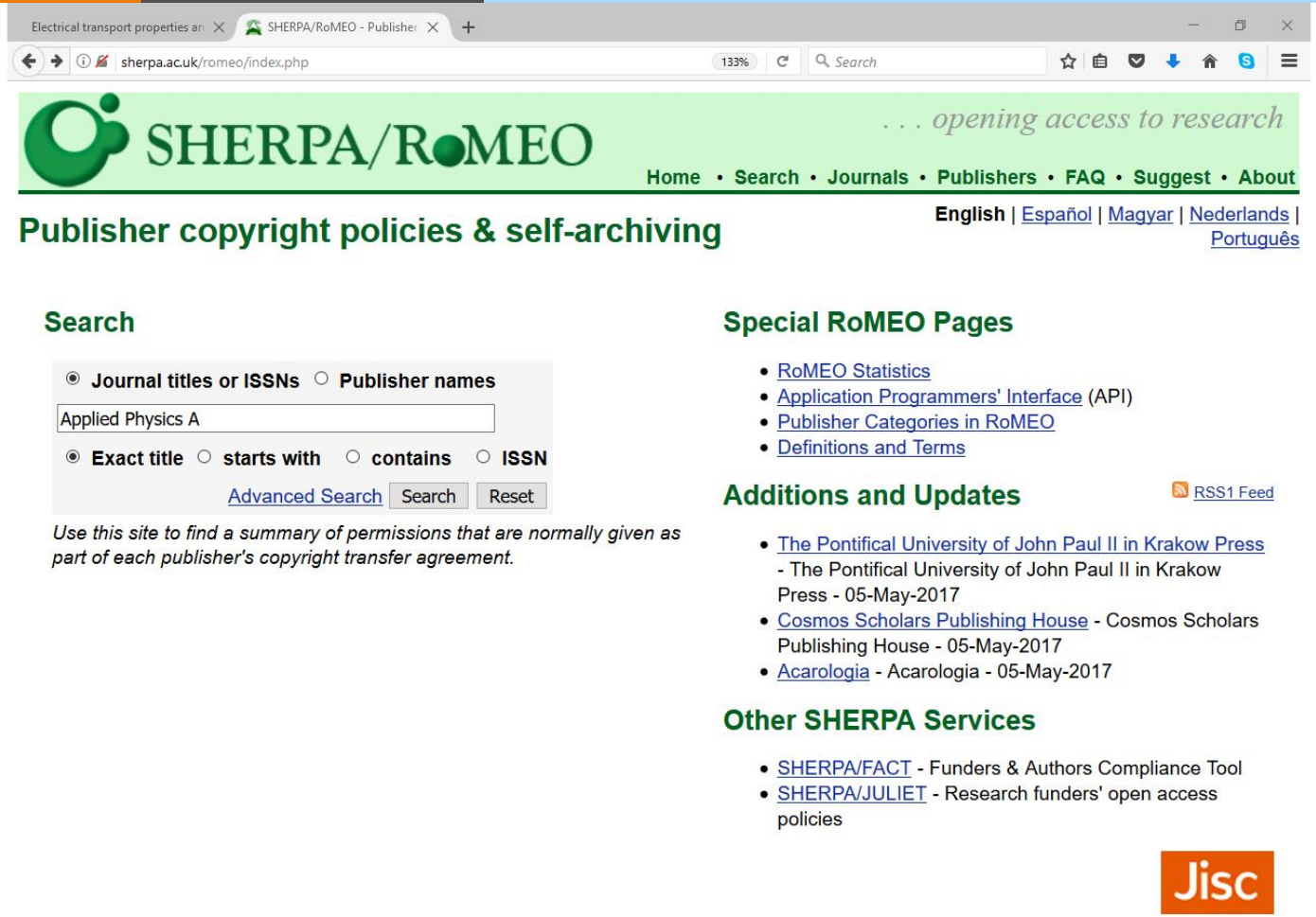
RSS1 Feed

- Bergen Open Access Publishing - Bergen Open Access Publishing - 18-Aug-2017
- Classical Association of the Middle West and South - Classical Association of the Middle West and South - 18-Aug-2017
- International Medical Society - International Medical Society - 18-Aug-2017

Other SHERPA Services

- SHERPA/FACT - Funders & Authors Compliance Tool
- SHERPA/JULIET - Research funders' open access policies

Jisc


Electrical transport properties an
SHERPA/RoMEO - Publisher
sherpa.ac.uk/romeo/index.php
133%
Search
SHERPA/RoMEO
. . . opening access to research
Home • Search • Journals • Publishers • FAQ • Suggest • About
Publisher copyright policies & self-archiving
English | Español | Magyar | Nederlands | Português
Search
Journal titles or ISSNs
Publisher names
Applied Physics A
Exact title
starts with
contains
ISSN
Advanced Search
Search
Reset
Use this site to find a summary of permissions that are normally given as part of each publisher's copyright transfer agreement.
Special RoMEO Pages
RoMEO Statistics
Application Programmers' Interface (API)
Publisher Categories in RoMEO
Definitions and Terms
Additions and Updates
RSS1 Feed
The Pontifical University of John Paul II in Krakow Press - The Pontifical University of John Paul II in Krakow Press - 05-May-2017
Cosmos Scholars Publishing House - Cosmos Scholars Publishing House - 05-May-2017
Acarologia - Acarologia - 05-May-2017
Other SHERPA Services
SHERPA/FACT - Funders & Authors Compliance Tool
SHERPA/JULIET - Research funders' open access policies
Jisc
This work is licensed under CC BY-NC-ND. About using our content
Contact us

SHERPA/RoMEO . . . opening access to research

Search - Publisher copyright policies & self-archiving

One journal found when searched for: **applied physics a**

Journal:	**Applied Physics A** (ISSN: 0947-8396, ESSN: 1432-0630)
RoMEO:	This is a RoMEO **green** journal
Paid OA:	A paid open access option is **available** for this journal.
Author's Pre-print:	✓ author **can** archive pre-print (ie pre-refereeing)
Author's Post-print:	✓ author **can** archive post-print (ie final draft post-refereeing)
Publisher's Version/PDF:	✗ author **cannot** archive publisher's version/PDF
General Conditions:	• Author's pre-print on pre-print servers such as arXiv.org • Author's post-print on author's personal website immediately • Author's post-print on any open access repository after 12 months after publication • Publisher's version/PDF cannot be used • Published source must be acknowledged • Must link to publisher version • Set phrase to accompany link to published version (see policy) • Articles in some journals can be made Open Access on payment of additional charge
Mandated OA:	(*Awaiting information*)
Paid Open Access:	Open Choice
Notes:	• Publisher last reviewed on 26/07/2016
Copyright:	Self-archiving policy - Authors Rights - Funder Compliance
Updated:	16-May-2014 - **Suggest an update for this record**
Link to this page:	http://sherpa.ac.uk/romeo/issn/0947-8396/
Published by:	**Springer Verlag (Germany)** - *Green Policies in RoMEO*

This summary is for the journal's ***default*** policies, and changes or exceptions can often be negotiated by authors.

Preprint and postprint versions

Preprints are all the versions of an academic article or other publication **before it has been submitted for peer review**, while the **postprint** is the form of the article **after all the peer review changes are in place**.

FOSTER

Repositories – institutional & disciplinary

OpenDOAR

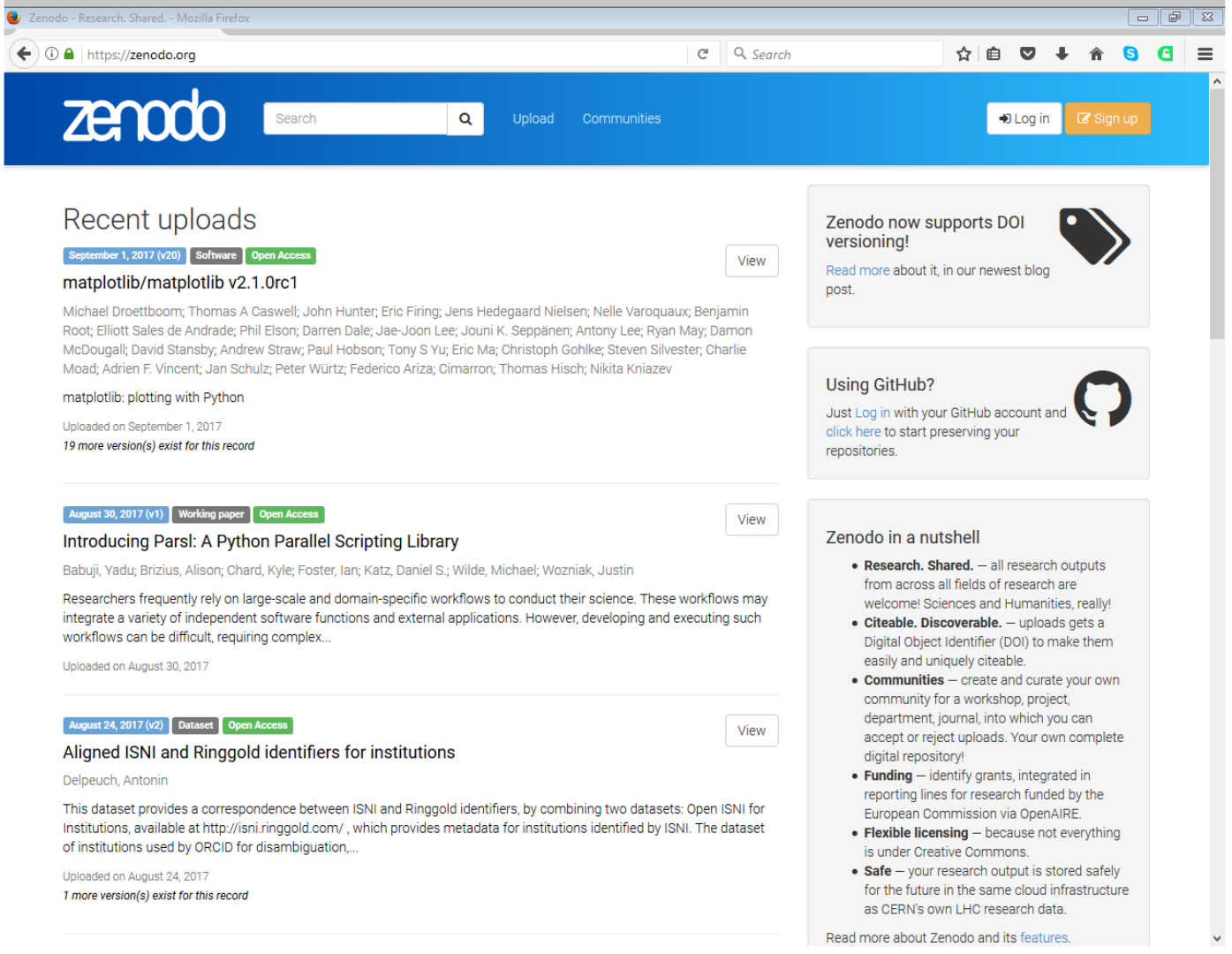
The Directory of Open Access Repositories

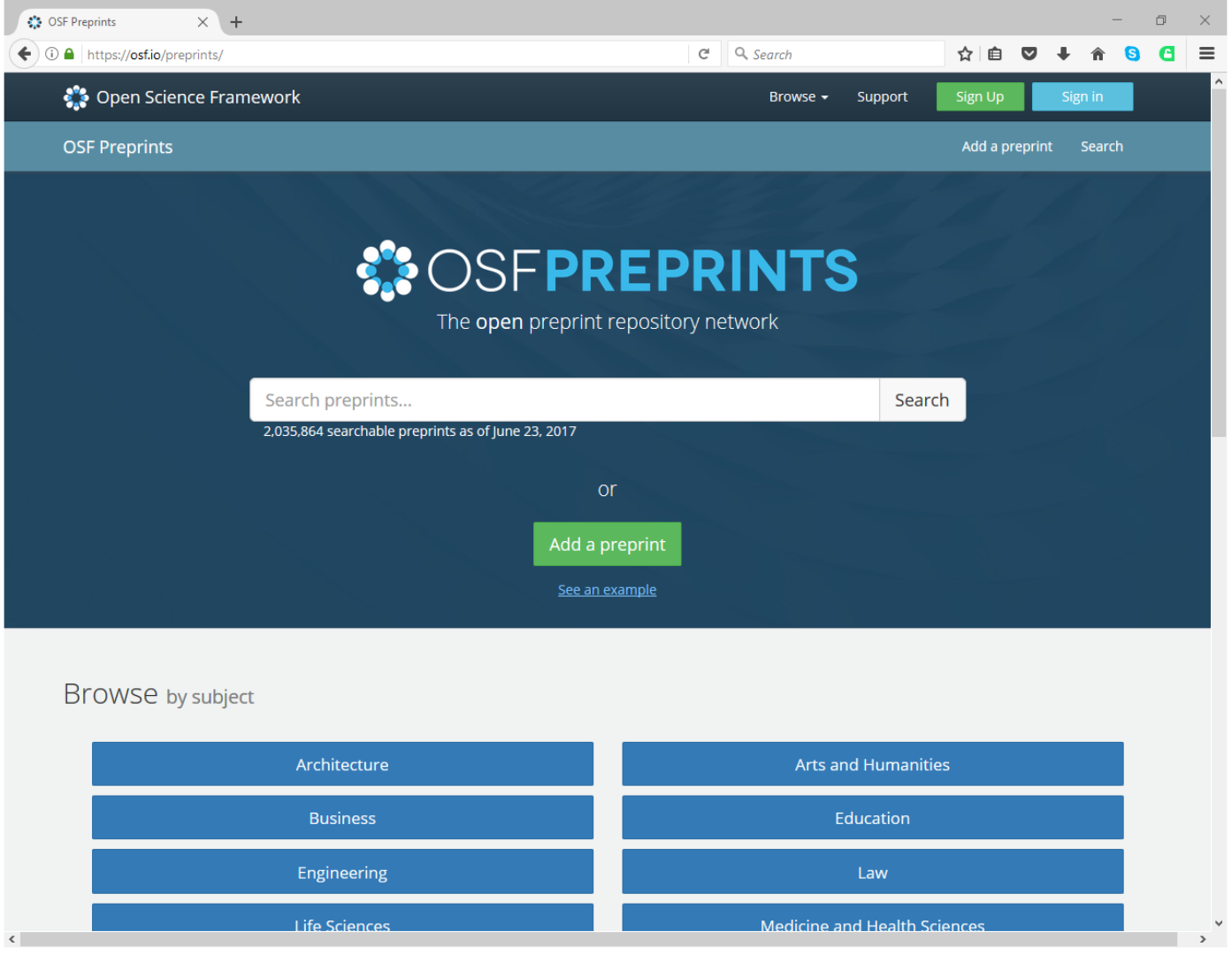
http://opendoar.org/

http://oad.simmons.edu

Registry of Open Access Repositories

http://roar.eprints.org/


Zenodo - Research. Shared. - Mozilla Firefox
https://zenodo.org
Search
zenodo
Search
Upload
Communities
Log in
Sign up
Recent uploads
September 1, 2017 (v20)
Software
Open Access
View
matplotlib/matplotlib v2.1.0rc1
Michael Droettboom; Thomas A Caswell; John Hunter; Eric Firing; Jens Hedegaard Nielsen; Nelle Varoquaux; Benjamin Root; Elliott Sales de Andrade; Phil Elson; Darren Dale; Jae-Joon Lee; Jouni K. Seppänen; Antony Lee; Ryan May; Damon McDougall; David Stansby; Andrew Straw; Paul Hobson; Tony S Yu; Eric Ma; Christoph Gohlke; Steven Silvester; Charlie Moad; Adrien F. Vincent; Jan Schulz; Peter Würtz; Federico Ariza; Cimarron; Thomas Hisch; Nikita Kniazev
matplotlib: plotting with Python
Uploaded on September 1, 2017
19 more version(s) exist for this record
August 30, 2017 (v1)
Working paper
Open Access
View
Introducing Parsl: A Python Parallel Scripting Library
Babuji, Yadu; Brizius, Alison; Chard, Kyle; Foster, Ian; Katz, Daniel S.; Wilde, Michael; Wozniak, Justin
Researchers frequently rely on large-scale and domain-specific workflows to conduct their science. These workflows may integrate a variety of independent software functions and external applications. However, developing and executing such workflows can be difficult, requiring complex...
Uploaded on August 30, 2017
August 24, 2017 (v2)
Dataset
Open Access
View
Aligned ISNI and Ringgold identifiers for institutions
Delpeuch, Antonin
This dataset provides a correspondence between ISNI and Ringgold identifiers, by combining two datasets: Open ISNI for Institutions, available at http://isni.ringgold.com/ , which provides metadata for institutions identified by ISNI. The dataset of institutions used by ORCID for disambiguation,...
Uploaded on August 24, 2017
1 more version(s) exist for this record
Zenodo now supports DOI versioning!
Read more about it, in our newest blog post.
Using GitHub?
Just Log in with your GitHub account and click here to start preserving your repositories.
Zenodo in a nutshell
Research. Shared. — all research outputs from across all fields of research are welcome! Sciences and Humanities, really!
Citeable. Discoverable. — uploads gets a Digital Object Identifier (DOI) to make them easily and uniquely citeable.
Communities — create and curate your own community for a workshop, project, department, journal, into which you can accept or reject uploads. Your own complete digital repository!
Funding — identify grants, integrated in reporting lines for research funded by the European Commission via OpenAIRE.
Flexible licensing — because not everything is under Creative Commons.
Safe — your research output is stored safely for the future in the same cloud infrastructure as CERN's own LHC research data.
Read more about Zenodo and its features.


OSF Preprints
https://osf.io/preprints/
Search
Open Science Framework
Browse
Support
Sign Up
Sign in
OSF Preprints
Add a preprint
Search
OSFPREPRINTS
The open preprint repository network
Search preprints...
Search
2,035,864 searchable preprints as of June 23, 2017
or
Add a preprint
See an example
Browse by subject
Architecture
Arts and Humanities
Business
Education
Engineering
Law
Life Sciences
Medicine and Health Sciences


» A social networking site is not an open access repository Office of Scholarly Communication - Opera
osc.universityofcalifornia.edu/2015/12/a-social-networking-site-is-not-an-open-access-repository/
Home
UC Open Access Policies
Scholarly Publishing
Publishing Tools
News and Archives
About OSC
Campus Resources
OFFICE OF SCHOLARLY COMMUNICATION
UNIVERSITY OF CALIFORNIA
Comments? Feedback?
HOME » FEATURES » A SOCIAL NETWORKING SITE IS NOT AN OPEN ACCESS REPOSITORY
A social networking site is not an open access repository
Home
UC Open Access Policies
Deposit Your Work
Get a Waiver/Embargo
Implementation Plan
Policy FAQ
Policy History
Publishers Contacted
Read the Policies
Presidential OA Policy
Systemwide Senate
UCSF Senate
OA Policy Contacts
Get the Word Out!
Scholarly Publishing
Copyright & Licensing
"What's the difference between ResearchGate, Academia.edu, and the institutional repository?"
"I put my papers in ResearchGate, is that enough for the open access policy?
These and similar questions have been been common at open access events over the past couple of years. Authors want to better understand the differences between these platforms and when they should use one, the other, or some combination.
First, a brief primer on what each service has to offer:
ResearchGate and Academia.edu
ResearchGate and Academia.edu are social networking platforms whose primary aim is to connect researchers with common interests. Users create profiles on these services, and are then encouraged to list their publications and other scholarly activities, upload copies of manuscripts they've authored, and build connections with scholars they work or co-author with. Essentially
UC OPEN ACCESS POLICIES
Learn more
DEPOSIT WAIVER FAQ
UC PUBLISHING TOOLS
eScholarship University of California
dash

	Open access repositories	Academia.edu	ResearchGate
Supports export or harvesting	Yes	No	No
Long-term preservation	Yes	No	No
Business model	Nonprofit (usually)	Commercial. Sells job posting services, hopes to sell data	Commercial. Sells ads, job posting services
Sends you lots of emails (by default)	No	Yes	Yes
Wants your address book	No	Yes	Yes
Fulfills requirements of UC's OA policies	Yes	No	No

Get more funding

whyopenresearch.org/funding

Search

Get more funding

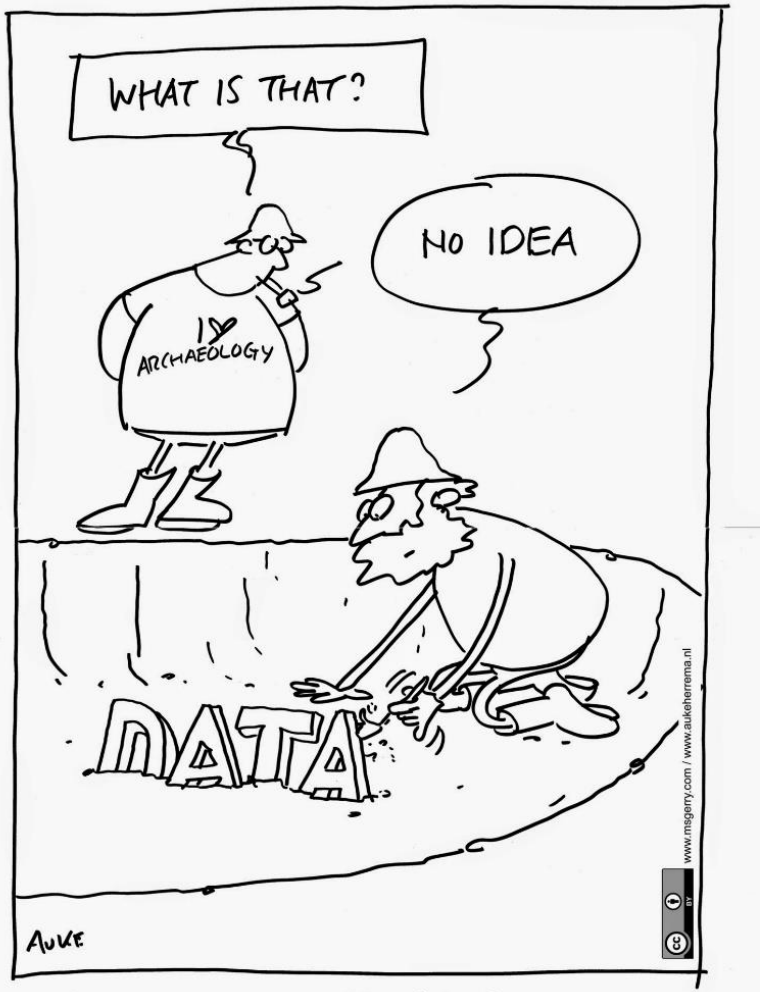
Meet funder requirements and qualify for special scholarships and grants.

Get that promotion

Open research is becoming increasingly recognized in promotion and tenure.

With thanks to Erin C. McKiernan, Twitter: @emckiernan13, and John McKiernan for re-using the content and visuals from the Why Open Research website http://whyopenresearch.org

Open Access to Research Data



DATA FOR FUTURE GENERATIONS

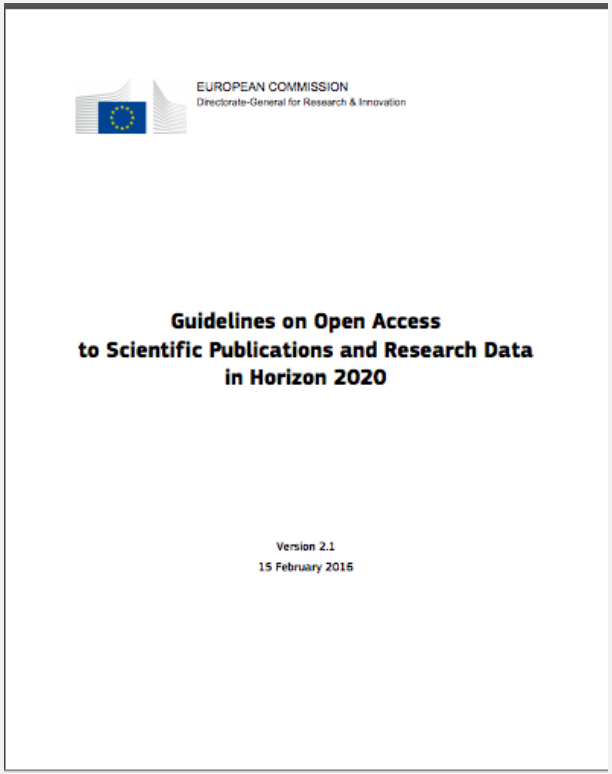
”If we wait 5 years for (Arctic) data to be released, the Arctic is going to be a very different place”

Parsons, Arctic Research Scientist

Source: http://www.nature.com/nature/journal/v461/n7261/index.html

FOSTER

Open Science is now a requirement


EUROPEAN COMMISSION
Directorate-General for Research & Innovation

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020

Version 2.1
15 February 2016

Research results:

"each beneficiary must ensure open access to all peer-reviewed scientific publications" (page 4)

Research data:

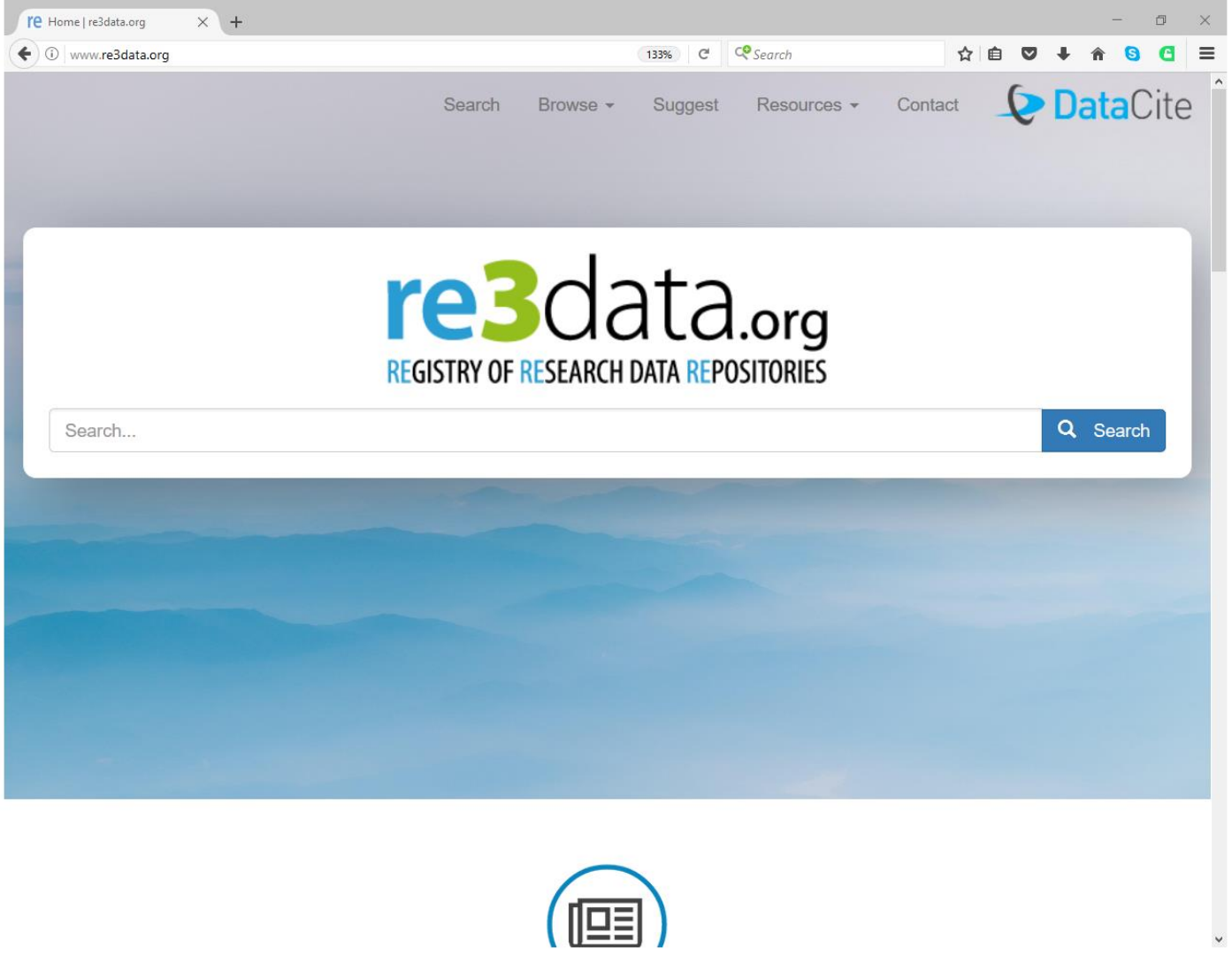
"A new feature of Horizon 2020 is the Open Research Data Pilot (ORD Pilot), designed to improve and maximise access to and reuse of research data generated by projects… The Pilot on Open Research Data will be monitored throughout Horizon 2020 with a view to further developing Commission policy on open research." *(page 7)*

FOSTER

Report URL:
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Funders recognise it

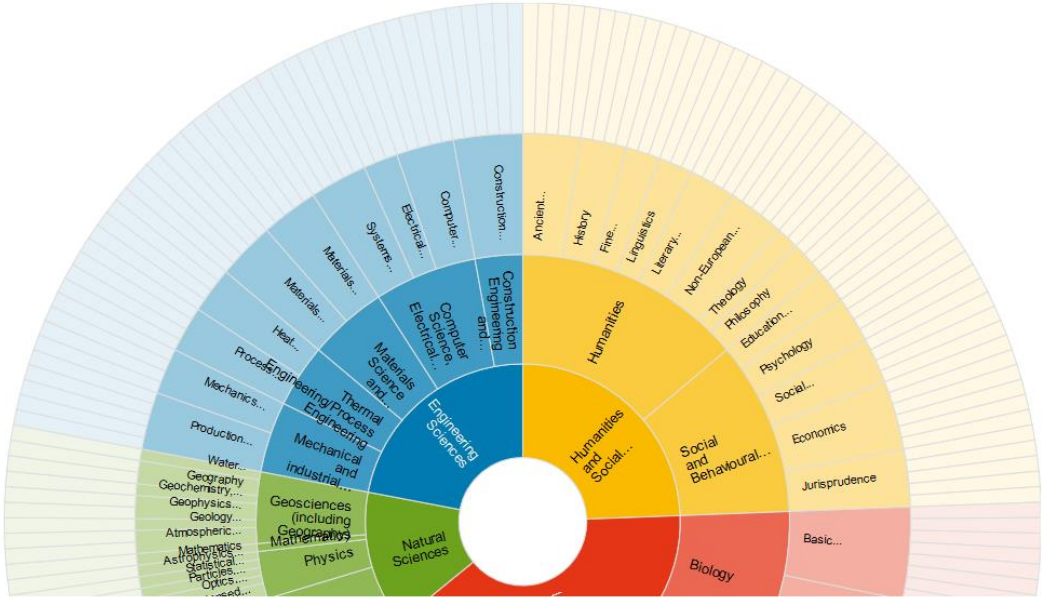
FOSTER


Home | re3data.org
www.re3data.org
133%
Search
Search
Browse
Suggest
Resources
Contact
DataCite
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 hierachy 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 | re3data

www.re3data.org/browse/by-country/

133%

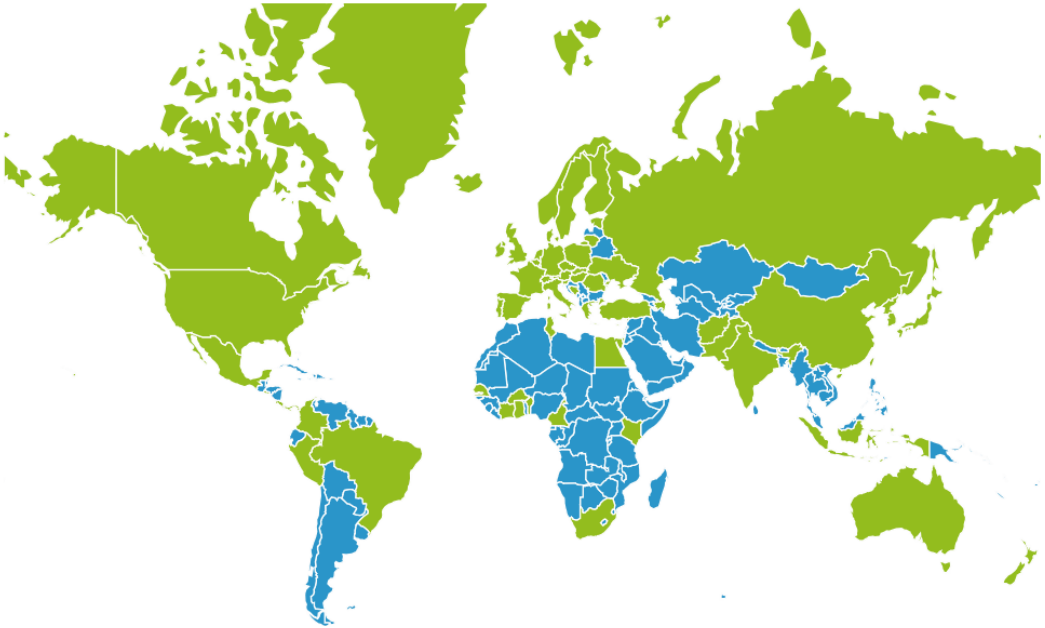
Search

re3data.org

Search Browse Suggest Resources Contact DataCite

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

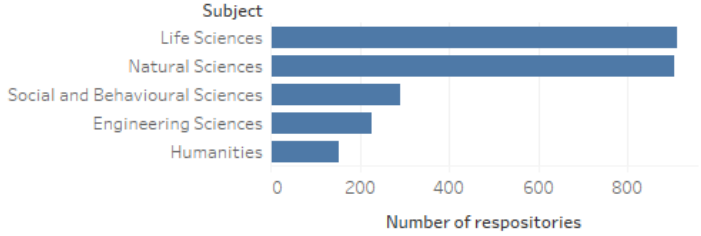
Number of data repositories

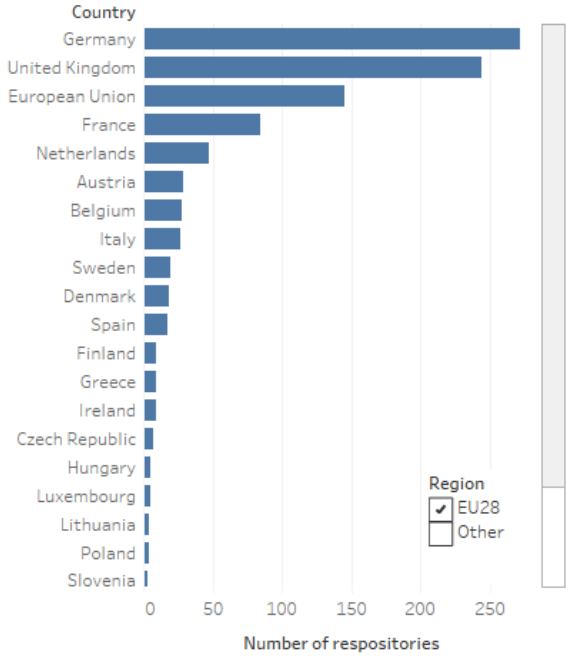
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

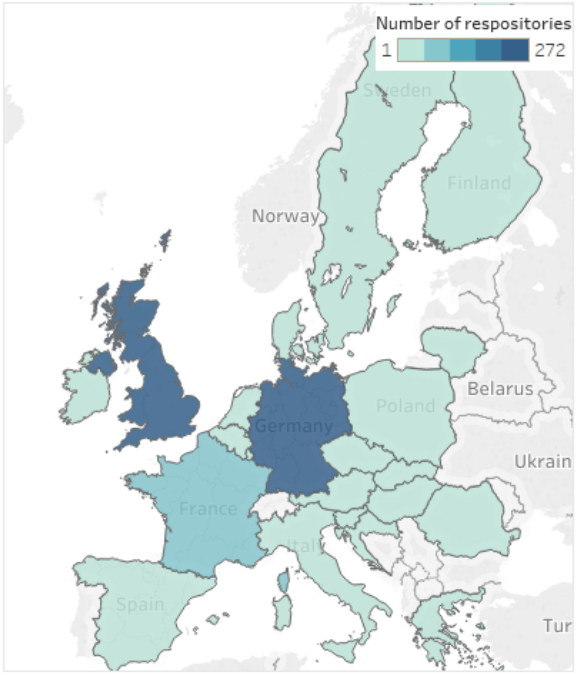
This visualisation shows the number of data repositories recorded in re3data.org. Repositories can cover multiple subjects, and be run from multiple countries.

Focus on one or more countries or subjects by selecting them on the barcharts. The CTRL key can be used to select multiple bars.

To reset, use the reset button in the bar below.







Open data

"Open data and content can be freely used, modified and shared by anyone for any purpose"

http://opendefinition.org

Tim Berners-Lee's proposal for five star open data - http://5stardata.info

- ★ make your stuff available on the Web (whatever format) under an open licence
- ★★ make it available as structured data (e.g. Excel instead of a scan of a table)
- ★★★ use non-proprietary formats (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

FOSTER

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…

FOSTER

WHY SHOULD YOU BE OPEN?

Image by wonderwebby CC-BY-NC-SA www.flickr.com/photos/wonderwebby/2723279491

PUBLICATIONS AND DATA

It’s part of good research practice

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

- Ewan Birney

#OpenData
#OpenScience

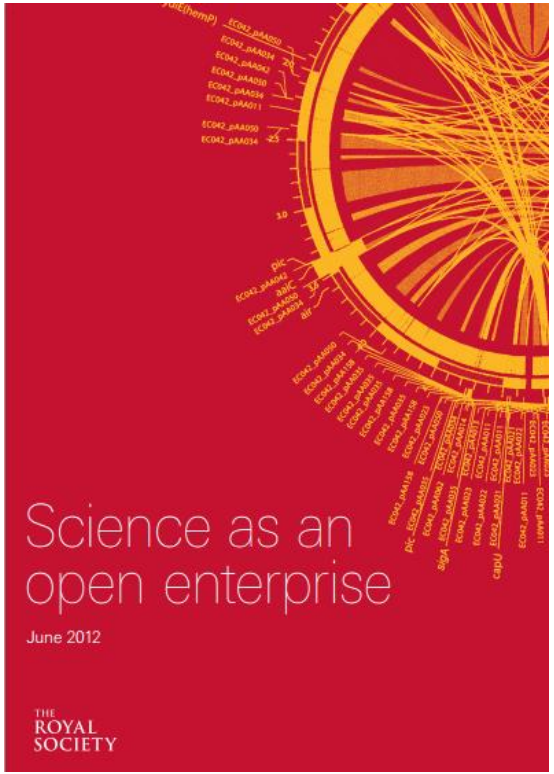
Original image via doi:10.1038/461145a. “Research cannot flourish if data are not preserved and made accessible. Data management should be woven into every course in science.” - *Nature* 461, 145

FOSTER

Science as an open enterprise

"Much of the remarkable growth of scientific understanding in recent centuries is due to open practices; open communication and deliberation sit at the heart of scientific practice."

Royal Society report calls for 'intelligent openness' whereby data are accessible, intelligible, assessable and usable.



https://royalsociety.org/policy/projects/science-public-enterprise/Report

FOSTER

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

Stories by subject
- Brain and behaviour
- Lab life

Stories by keywords
- Deiderik 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

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) 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 Eijndhoven

"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

FOSTER

www.nature.com/news/2011/111101/full/479015a.html

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-austerity

FOSTER

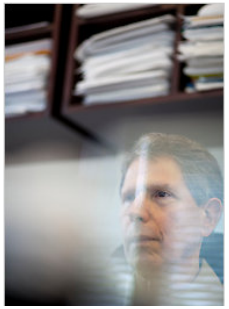
More scientific breakthroughs

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

FOSTER

A citation advantage

A study that analysed the citation counts of 10,555 papers on gene expression studies that created microarray data, showed:

"studies that made data available in a public repository received 9% more citations than similar studies for which the data was not made available"

Data reuse and the open data citation advantage, Piwowar, H. & Vision, T. https://peerj.com/articles/175

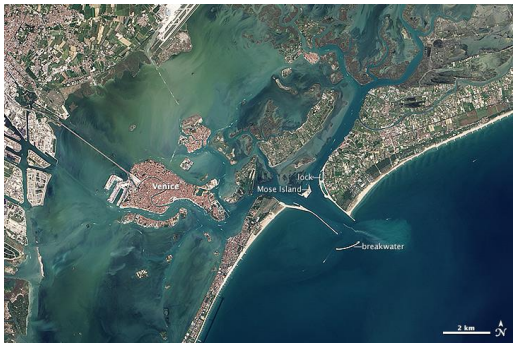
FOSTER

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

http://earthobservatory.nasa.gov/IOTD/view.php?id=83394&src=ve

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

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

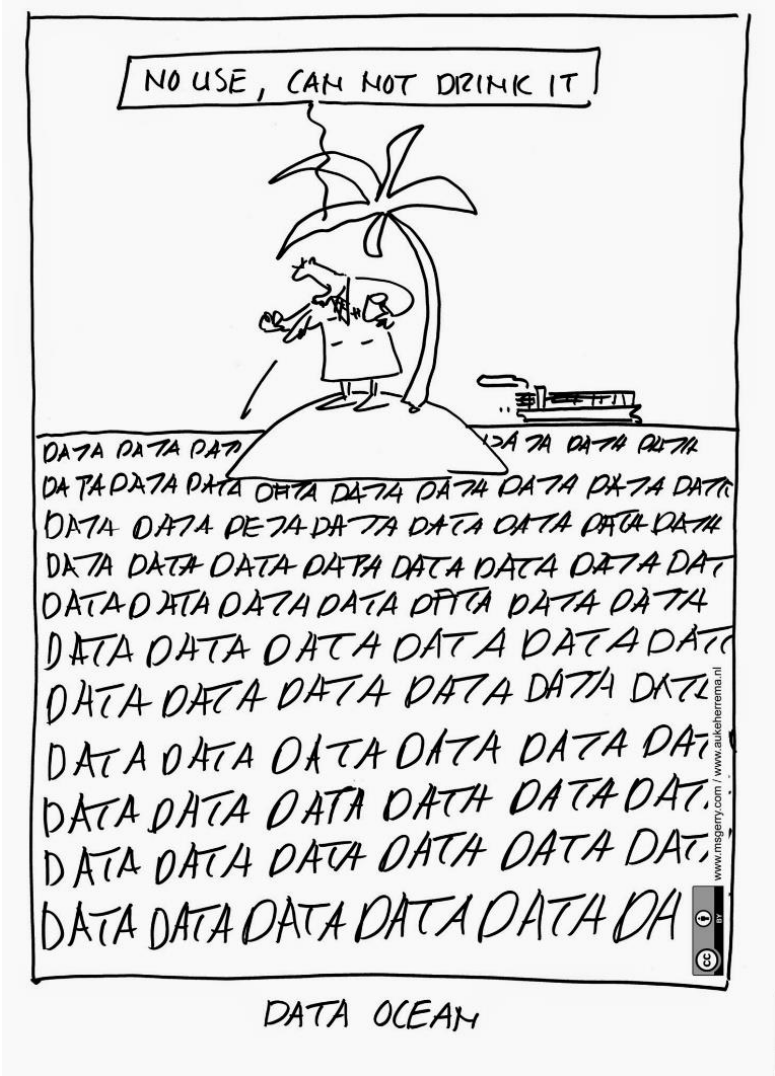


Image courtesy of http://aukeherrema.nl CC-BY

WHAT IS A DMP & WHY WRITE ONE?

Data Management Plans (DMP)

A DMP is a brief plan to define:

- how the data will be created
- how it will be documented
- who will be able to access it
- where it will be stored
- who will back it up
- whether (and how) it will be shared & preserved

DMPs are often submitted as part of grant applications, but are useful whenever researchers are creating data.

European Commission

OPEN RESEARCH DATA IN HORIZON 2020

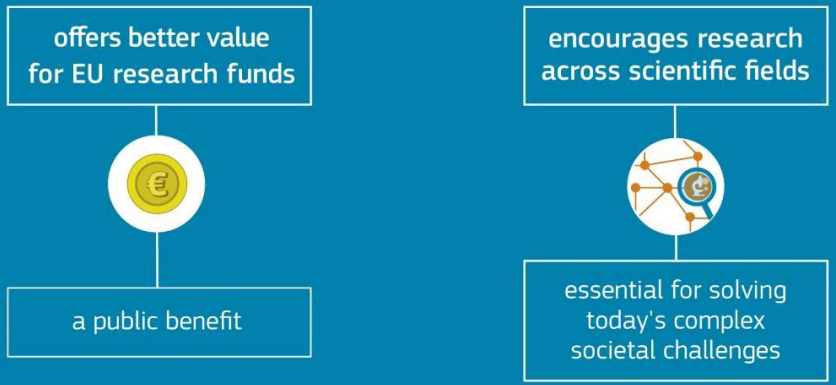
Jean-François Dechamp & Daniel Spichtinger

European Commission
Directorate-General for Research & Innovation

Research and Innovation

CHALLENGE

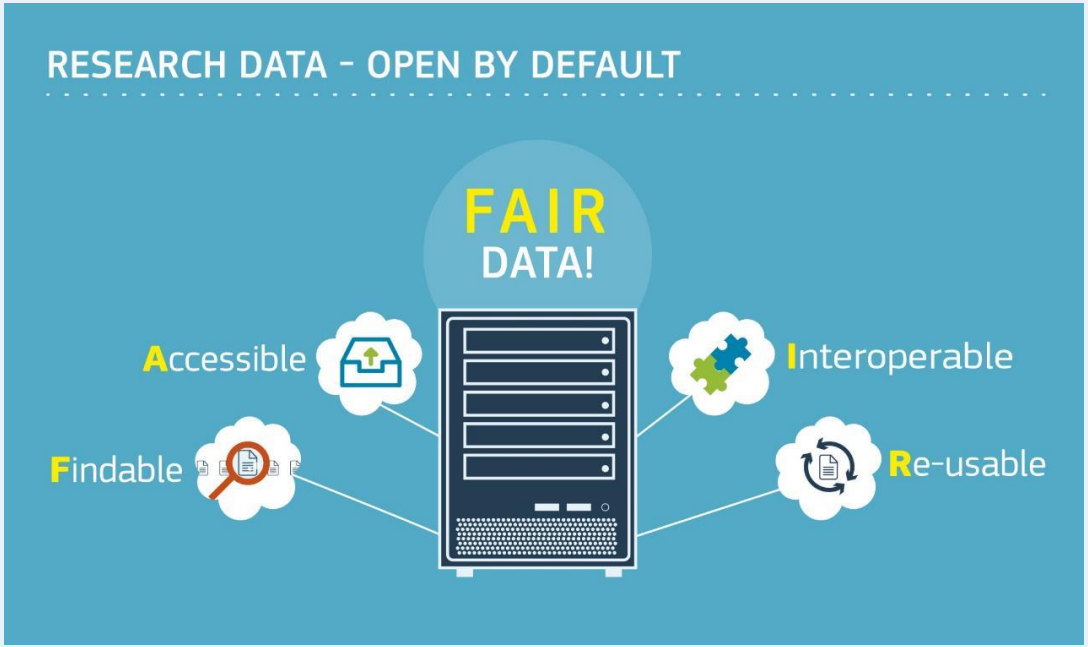
Wider access to scientific facts and knowledge helps researchers, innovators and the public find and re-use data, and check research results:



SOLUTION

Horizon 2020 already mandates open access to **all scientific publications**

From 2017,
research data is **open by default,**
with possibilities to **opt out**


RESEARCH DATA - OPEN BY DEFAULT
FAIR
DATA!
Accessible
Interoperable
Findable
Re-usable

Making data FAIR

- **Findable** – Assign persistent IDs, provide rich metadata, register in a searchable resource,...
- **Accessible** - Retrievable by their ID using a standard protocol, metadata remain accessible even if data aren't...
- **Interoperable** - Use formal, broadly applicable languages, use standard vocabularies, qualified references...
- **Reusable** - Rich, accurate metadata, clear licences, provenance, use of community standards

www.force11.org/group/fairgroup/fairprinciples

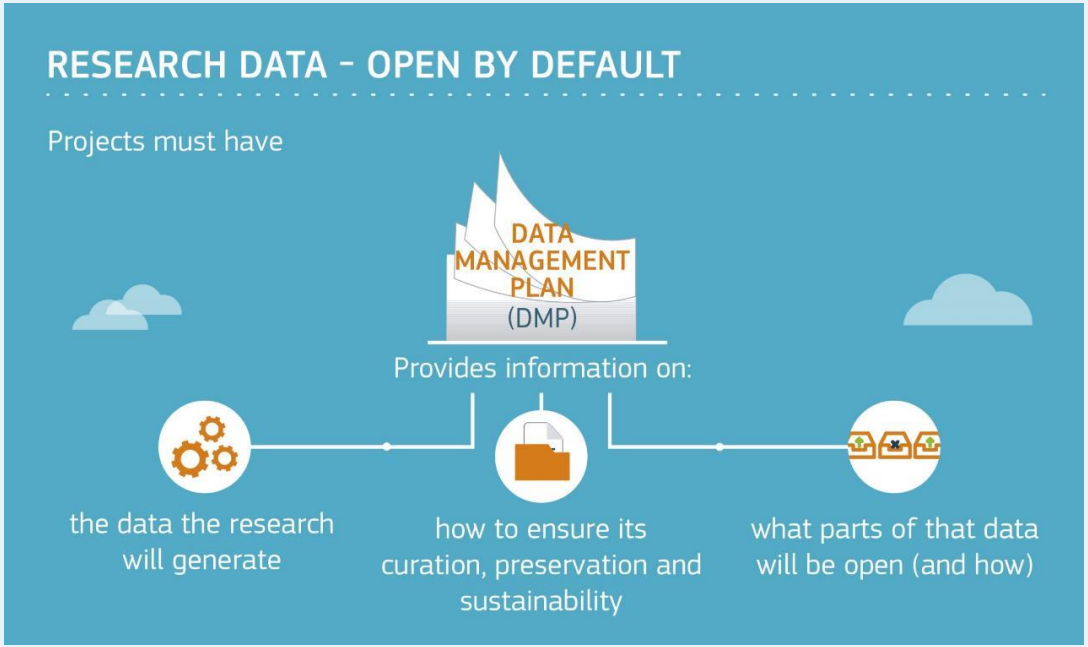
RESEARCH DATA - OPEN BY DEFAULT

Horizon 2020 grantees are required

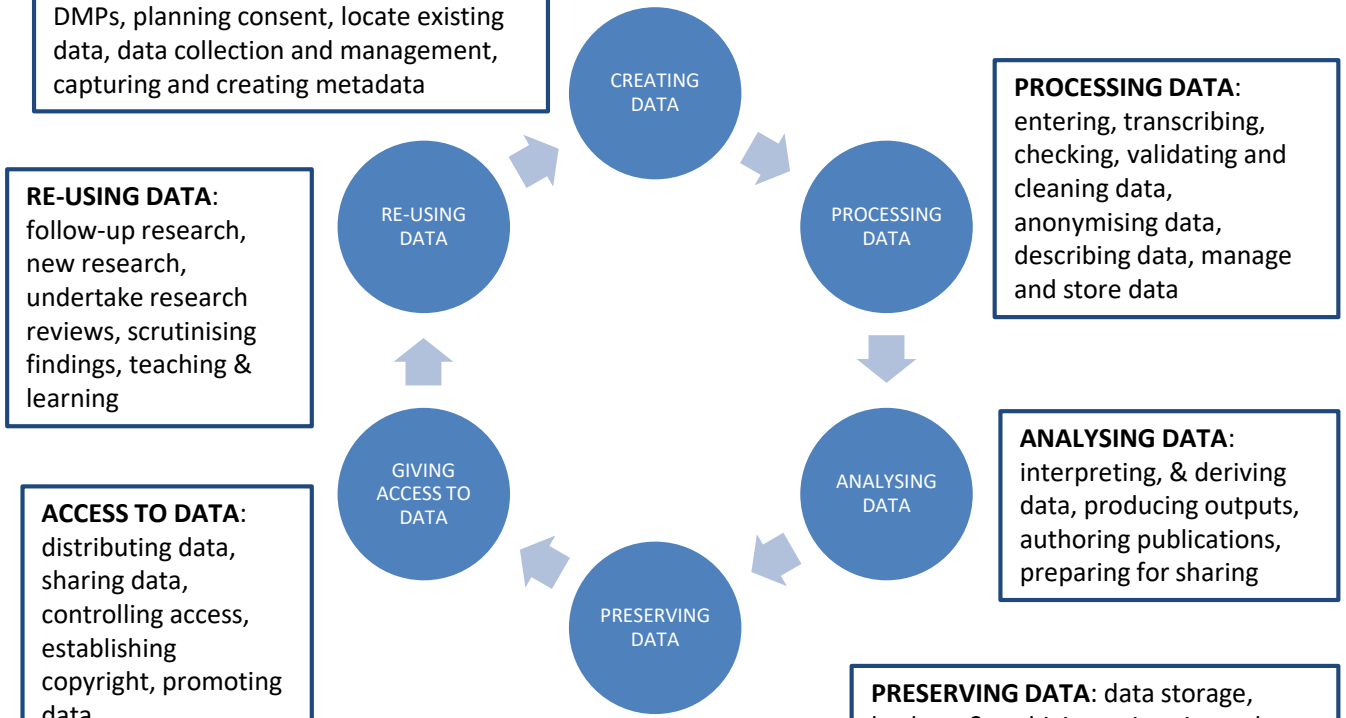
take measures to ensure open access to the **data underlying their scientific publications**

provide open access to **any other research data of their choice**

Horizon 2020 grantees are **encouraged to also share datasets** beyond publication


RESEARCH DATA - OPEN BY DEFAULT
Projects must have
DATA MANAGEMENT PLAN (DMP)
Provides information on:
the data the research will generate
how to ensure its curation, preservation and sustainability
what parts of that data will be open (and how)

Research data lifecycle



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

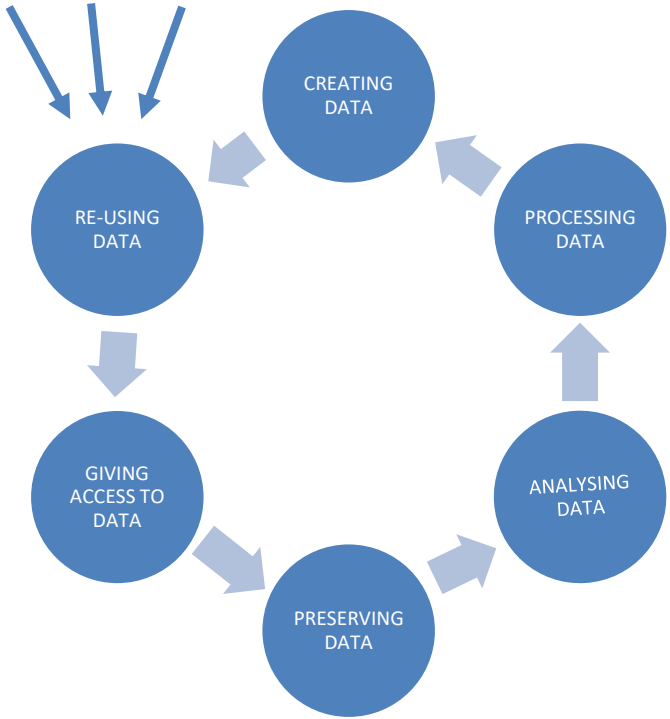
ACCESS TO DATA: distributing data, sharing data, controlling access, establishing copyright, promoting data

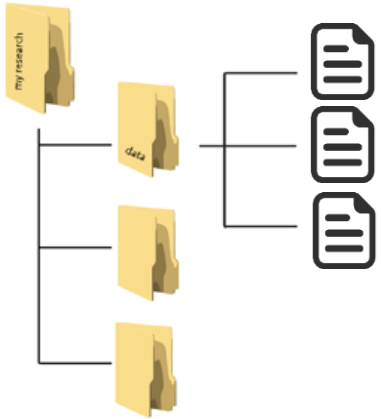
RE-USING DATA: follow-up research, new research, undertake research reviews, scrutinising findings, teaching & learning

Ref: UK Data Archive: http://www.data-archive.ac.uk/create-manage/life-cycle

Planning trick 1: think backwards

What data organisation would a re-user like?





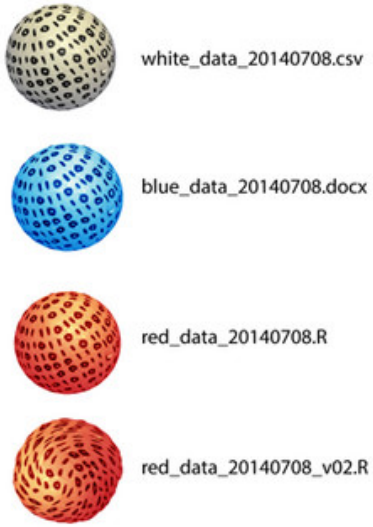
Data organisation

Meaningful file names

Below are tips on meaningful and consistent file names. Read more in '*Choosing a file name*'. (2)

- Make sure to use consistent file names. When you use a date in the file name, choose a notation (for instance, YYYYMMDD of 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.

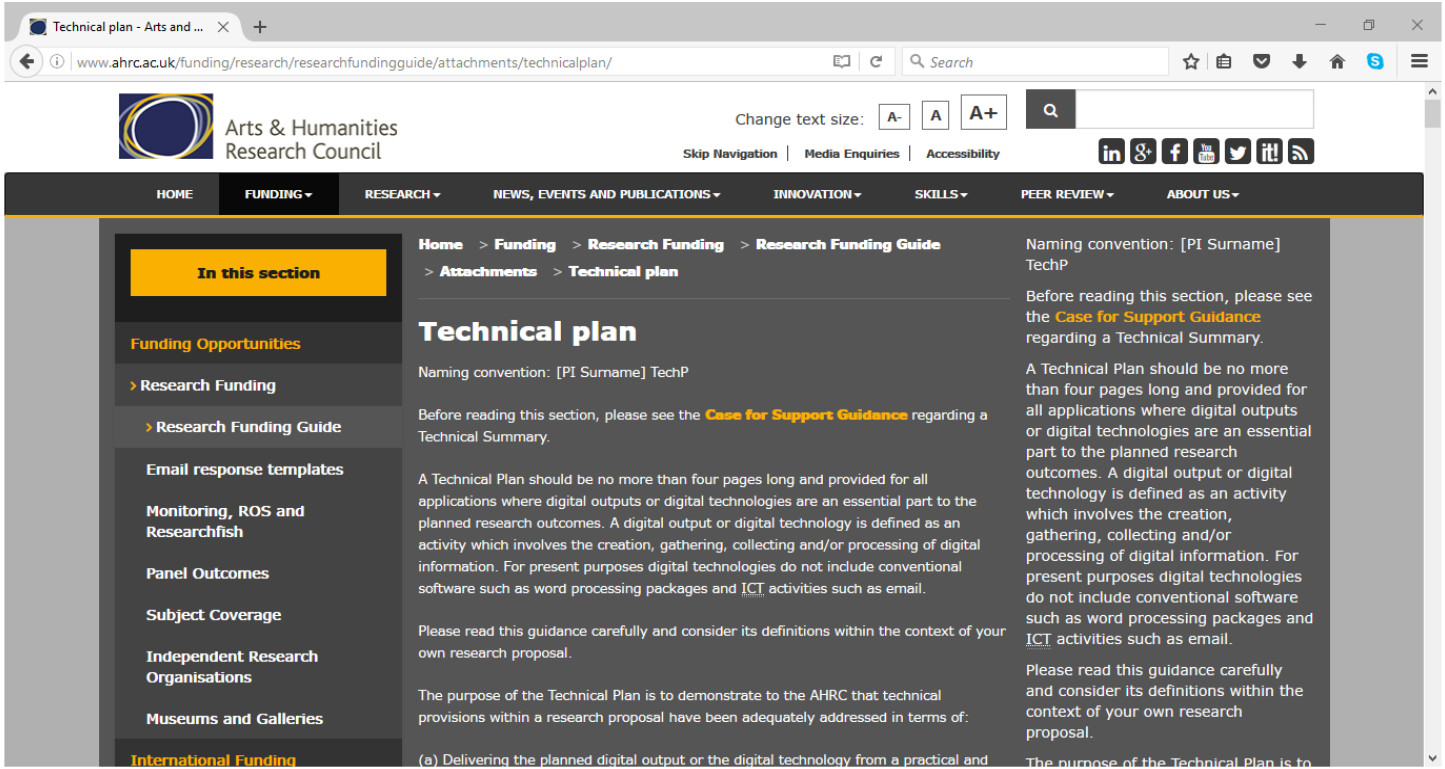


File naming and version management

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

http://datasupport.researchdata.nl/en/start-de-cursus/iii-onderzoeksfase/organising-data

Some other funders that require DMPs


Technical plan - Arts and ...
www.ahrc.ac.uk/funding/research/researchfundingguide/attachments/technicalplan/
Search
Arts & Humanities Research Council
Change text size: A- A A+
Skip Navigation | Media Enquiries | Accessibility
HOME
FUNDING
RESEARCH
NEWS, EVENTS AND PUBLICATIONS
INNOVATION
SKILLS
PEER REVIEW
ABOUT US
In this section
Funding Opportunities
Research Funding
Research Funding Guide
Email response templates
Monitoring, ROS and Researchfish
Panel Outcomes
Subject Coverage
Independent Research Organisations
Museums and Galleries
International Funding
Home > Funding > Research Funding > Research Funding Guide > Attachments > Technical plan
Technical plan
Naming convention: [PI Surname] TechP
Before reading this section, please see the Case for Support Guidance regarding a Technical Summary.
A Technical Plan should be no more than four pages long and provided for all applications where digital outputs or digital technologies are an essential part to the planned research outcomes. A digital output or digital technology is defined as an activity which involves the creation, gathering, collecting and/or processing of digital information. For present purposes digital technologies do not include conventional software such as word processing packages and ICT activities such as email.
Please read this guidance carefully and consider its definitions within the context of your own research proposal.
The purpose of the Technical Plan is to demonstrate to the AHRC that technical provisions within a research proposal have been adequately addressed in terms of:
(a) Delivering the planned digital output or the digital technology from a practical and
Naming convention: [PI Surname] TechP
Before reading this section, please see the Case for Support Guidance regarding a Technical Summary.
A Technical Plan should be no more than four pages long and provided for all applications where digital outputs or digital technologies are an essential part to the planned research outcomes. A digital output or digital technology is defined as an activity which involves the creation, gathering, collecting and/or processing of digital information. For present purposes digital technologies do not include conventional software such as word processing packages and ICT activities such as email.
Please read this guidance carefully and consider its definitions within the context of your own research proposal.
The purpose of the Technical Plan is to

Why manage data?

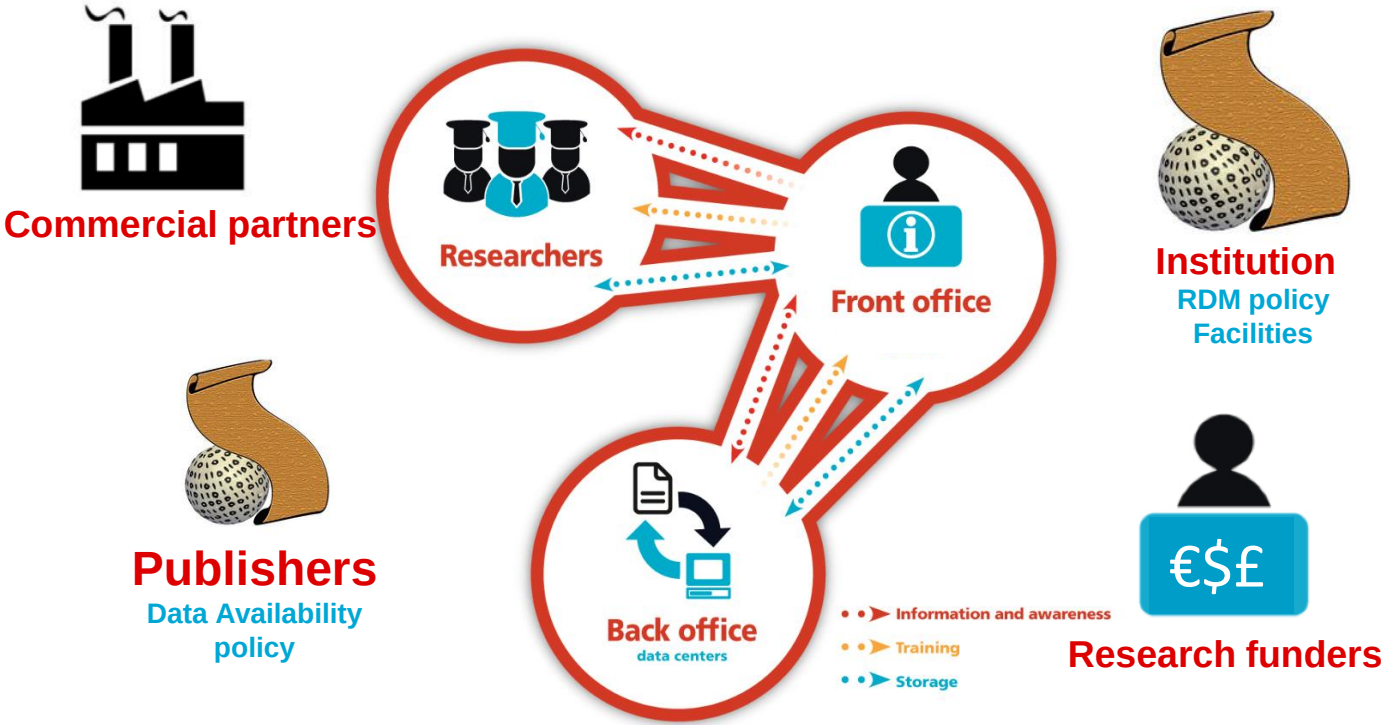
NON PECUNIAE INVESTIGATIONIS CURATORE

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

Planning trick 2: include RDM stakeholders



https://www.openaire.eu/briefpaper-rdm-infonoads

Responsibilities in RDM

- The principal investigator – ultimately responsible for the data and for data management
- Researchers, research assistants and/or data managers – involved in day-to-day data management
- The institution's management – draft and enforce data policies; raise data awareness
- The institution's research office consisting of library, IT and legal services – provide external data, tools, secure storage and access; expertise on rights management and ethics, data citation, metadata, access and licenses, funder requirements; raise data awareness
- Research funders – encourage good data practices; invest in data infrastructure; raise data awareness
- Project partners in academic and other research institutions as well as commercial partners
- Academic publishers – impose requirements on the availability of data underlying submitted and/or published papers; provide identifiers to cite papers and link to related data
- Research data repositories – preserve data long term; provide persistent identifiers and data discovery service

https://www.openaire.eu/briefpaper-rdm-infonoads

A DMP is about ‘keeping’ data

- Storing data < > archiving data
- Archived data < > findable data
- Findable < > accessible
- Accessible < > understandable
- Understandable < > usable

- A USB stick is not safe
- A persistent ID is essential but no guarantee for usability
- Data in a proprietary format is not sustainable

How to deal with data and context?

- Versioning, back-up, storage and archiving
  - During the project and in the long term
- Ethics, consent forms, legal access
- Security and technical access
- Usage licences

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.

Research Data Alliance (RDA) http://rd-alliance.github.io/metadata-directory/standards/
FAIR Guiding Principles for scientific data management & stewardship http://www.nature.com/articles/sdata201618
How to select and appraise research data:www.dcc.ac.uk/resources/how-guides/appraise-select-research-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.)

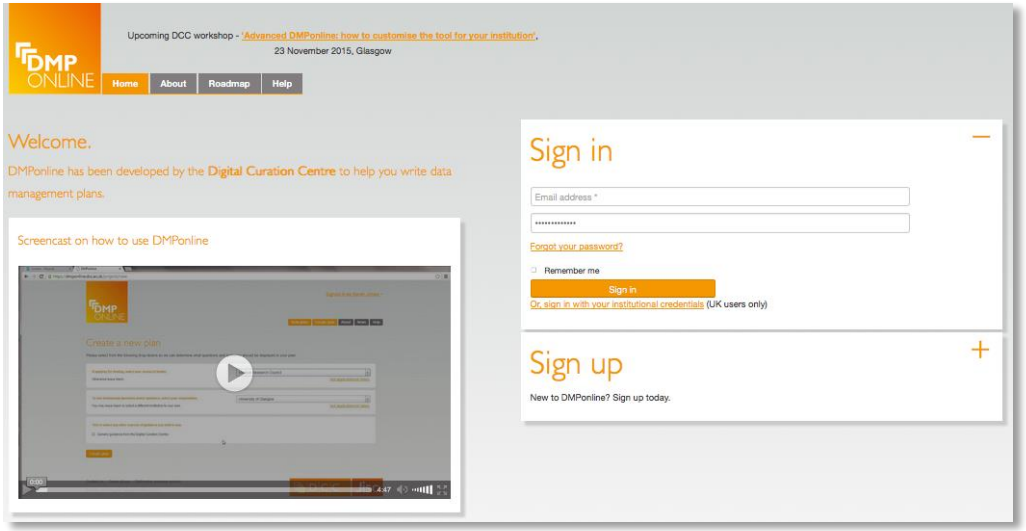
Basically, everything that is needed to replicate a study should be available. Plus everything that is potentially useful for others.

Research Data Alliance (RDA) http://rd-alliance.github.io/metadata-directory/standards/
FAIR Guiding Principles for scientific data management & stewardship http://www.nature.com/articles/sdata201618
How to select and appraise research data:www.dcc.ac.uk/resources/how-guides/appraise-select-research-data

DMPonline

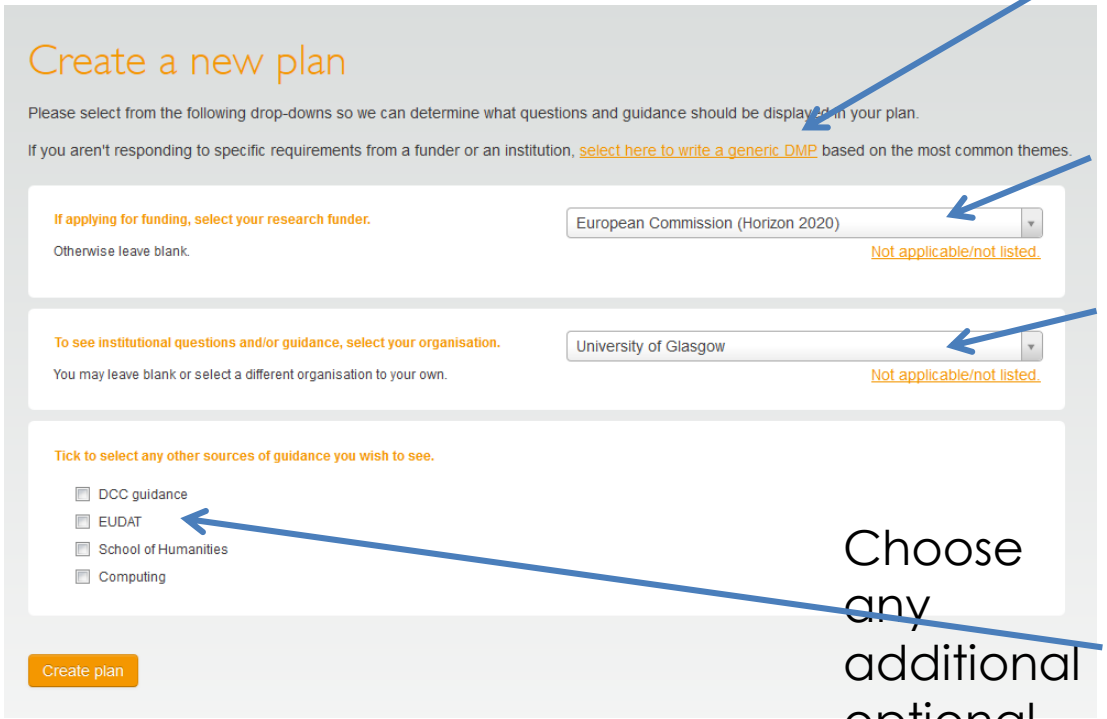
A web-based tool to help researchers write DMPs

Includes a template for Horizon 2020, guidance from EUDAT and OpenAIRE



https://dmponline.dcc.ac.uk

How the tool works



Click to write a generic DMP

Or choose your funder to get their specific template

Pick your uni to add local guidance and to get their template if no funder applies

Choose any additional optional guidance

RESEARCH DATA - OPEN BY DEFAULT

Data management costs are fully eligible for funding

No repository imposed: deposit data where you want

Data description examples

The final dataset will include self-reported demographic and behavioural data from interviews with the subjects and laboratory data from urine specimens provided.

From NIH data sharing statements

Every two days, we will subsample E. affinis populations growing under our treatment conditions. We will use a microscope to identify the life stage and sex of the subsampled individuals. We will document the information first in a laboratory notebook and then copy the data into an Excel spreadsheet. The Excel spreadsheet will be saved as a comma separated value (.csv) file.

From DataOne – E. affinis DMP example

Metadata examples

Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively.

From ICPSR Framework for Creating a DMP

We will first document our metadata by taking careful notes in the laboratory notebook that refer to specific data files and describe all columns, units, abbreviations, and missing value identifiers. These notes will be transcribed into a .txt document that will be stored with the data file. After all of the data are collected, we will then use EML (Ecological Metadata Language) to digitize our metadata. EML is one of the accepted formats used in ecology, and works well for the types of data we will be producing. We will create these metadata using Morpho software, available through KNB. The metadata will fully describe the data files and the context of the measurements.

From DataOne – E. affinis DMP example

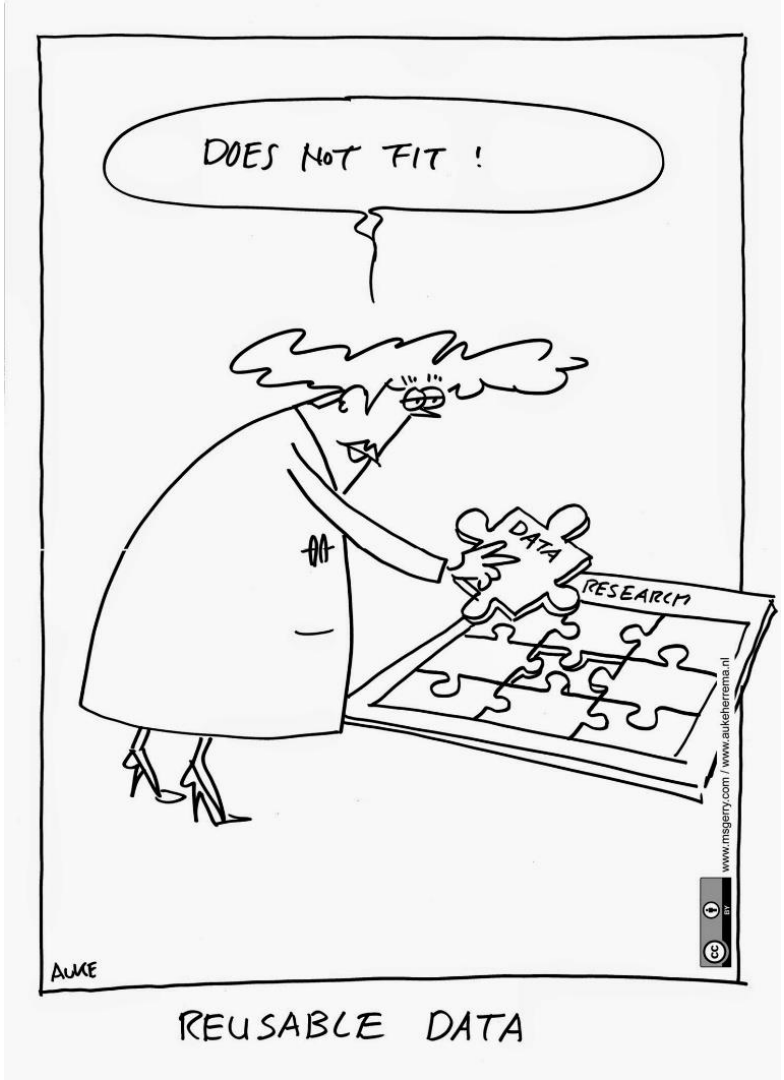
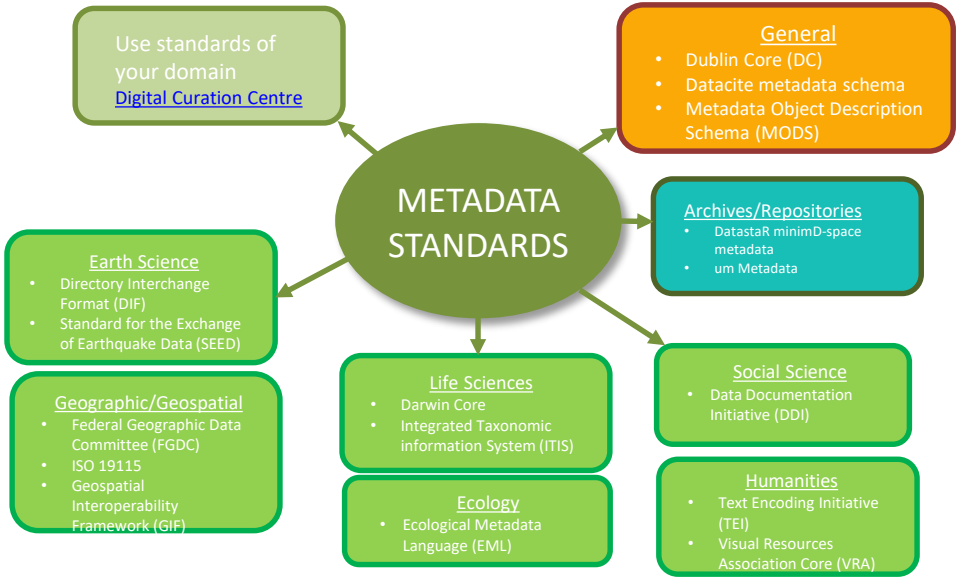


Image courtesy of http://aukeherrema.nl CC-BY



OpenAIRE

European Commission

Metadata standards

Use relevant standards for interoperability

Search by Discipline

Biology

Earth Science

General Research Data

Physical Science

Social Science & Humanities

RDA RESEARCH DATA ALLIANCE

www.dcc.ac.uk/resources/metadata-standards

FOSTER

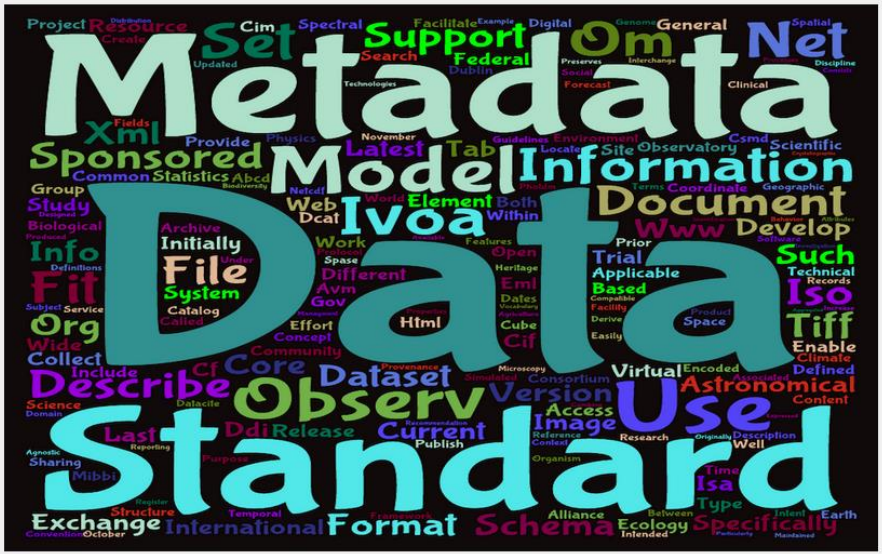
Metadata

RDA | Metadata Directory

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



Metadata Standards Directory Working Group

The RDA Metadata Standards Directory Working Group is supported by individuals and organizations involved in the development, implementation, and use of metadata for scientific data. The overriding goal is to develop a collaborative, open directory of metadata standards applicable to scientific data can help address infrastructure challenges.

The RDA Metadata Standards Directory is maintained by Sean Chen, Kate Anne Alderete, and Alex Ball.
The theme is maintained by Dustin Allen.
This page was generated by GitHub Pages.

http://rd-alliance.github.io/metadata-directory

Data sharing examples

The videos will be made available via the bristol.ac.uk website (both as streaming media and downloads) HD and SD versions will be provided to accommodate those with lower bandwidth. Videos will also be made available via Vimeo, a platform that is already well used by research students at Bristol. Appropriate metadata will also be provided to the existing Vimeo standard.

All video will also be available for download and re-editing by third parties. To facilitate this Creative Commons licenses will be assigned to each item. In order to ensure this usage is possible, the required permissions will be gathered from participants (using a suitable release form) before recording commences.

From University of Bristol Kitchen Cosmology DMP

We will make the data and associated documentation available to users under a **data-sharing agreement** that provides for: (1) a commitment to using the data only for research purposes and not to identify any individual participant; (2) a commitment to securing the data using appropriate computer technology; and (3) a commitment to destroying or returning the data after analyses are completed.

From NIH data sharing statements

Examples restrictions

Because the STDs being studied are reportable diseases, we will be collecting identifying information. Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, we will make the data and associated documentation available to users only under a data-sharing agreement.

From NIH data sharing statements

Examples restrictions (2)

1. Share data privately within 1 year.
   Data will be held in Private Repository, but metadata will be public
2. Release data to public within 2 years.
   Encouraged after one year to release data for public access.
3. Request, in writing, data privacy up to 4 years.
   Extensions beyond 3 years will only be granted for compelling cases.
4. Consult with creators of private CZO datasets prior to use.
   Pis required to seek consent before using private data they can access

From Boulder Creek Critical Zone Observatory DMP

Archiving examples

The investigators will work with staff at the UKDA to determine what to archive and how long the deposited data should be retained. Future long-term use of the data will be ensured by placing a copy of the data into the repository.

From ICPSR Framework for Creating a DMP

Data will be provided in file formats considered appropriate for long-term access, as recommended by the UK Data Service. For example, SPSS Portal forat and tab-delimited text for qualitative tabular data and RTF and PDF/A for interview transcripts. Appropriate documentation necessary to understand the data will also be provided. Anonymised data will be held for a minimum of 10 years following project completion, in compliance with LSHTM's Records Retention and Disposal Schedule. Biological samples (output 3) will be deposited with the UK BioBank for future use.

From Writing a Wellcome Trust Data Management and Sharing Plan

Sharing data: what is meant?

With collaborators while research is active

Dropbox

Data are mutable

(Open) data sharing

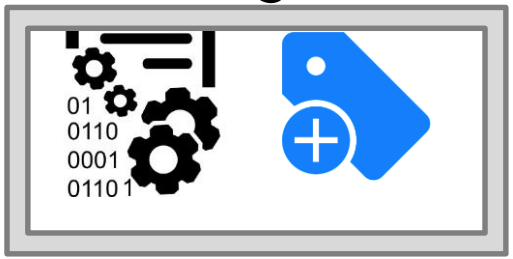
Data are stable, searchable, citable, clearly licensed

Storing data: what is meant?

Storing and backing up files while research is active

Likely to be on a networked filestore or hard drive

Archiving or preserving data in the long-term



Likely to be deposited in a digital repository

afeguarded and preserved

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?

- More information: https://www.openaire.eu/opendatapilot-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
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. www.openaire.eu/opendatapilot-repository

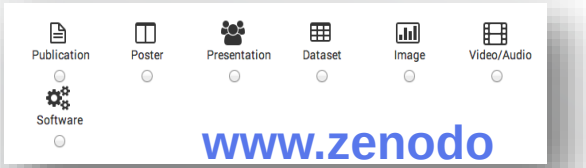
Zenodo (OpenAIRE/CERN repository)

www.zenodo.org

Zenodo Repository

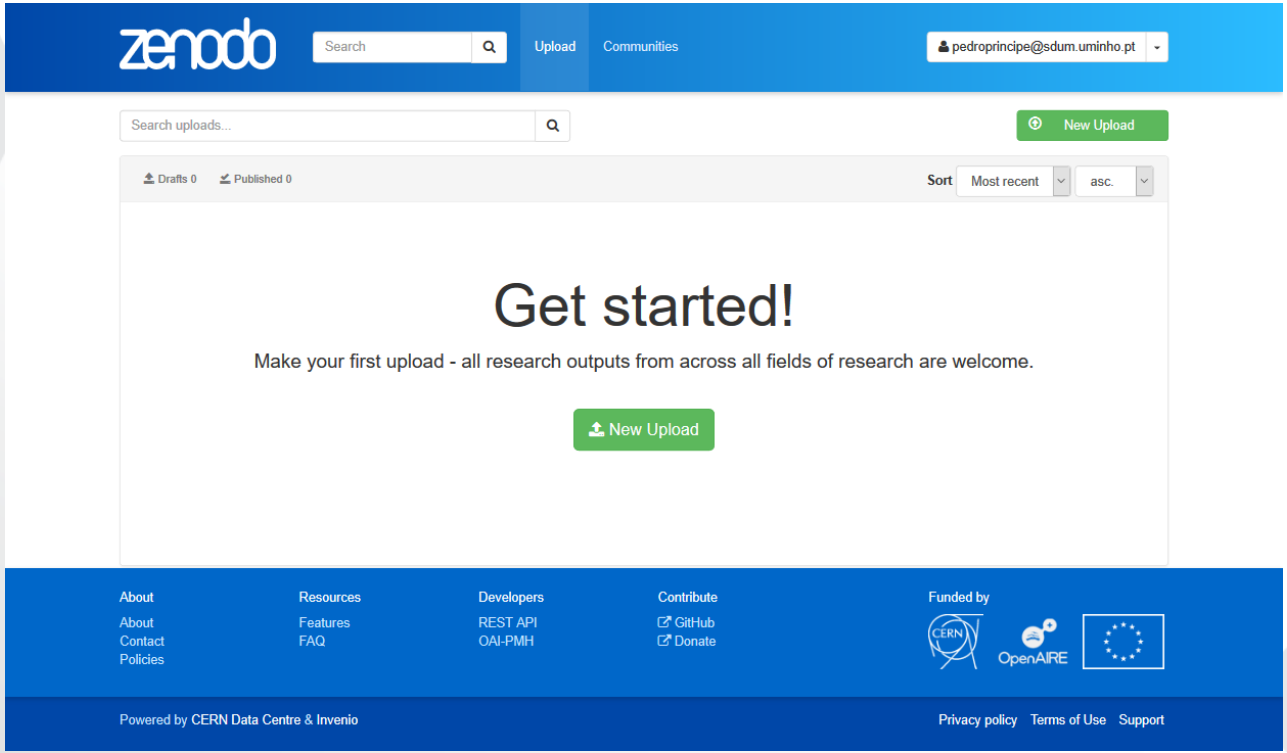
“Catch-all” repository: OpenAIRE-CERN joint effort

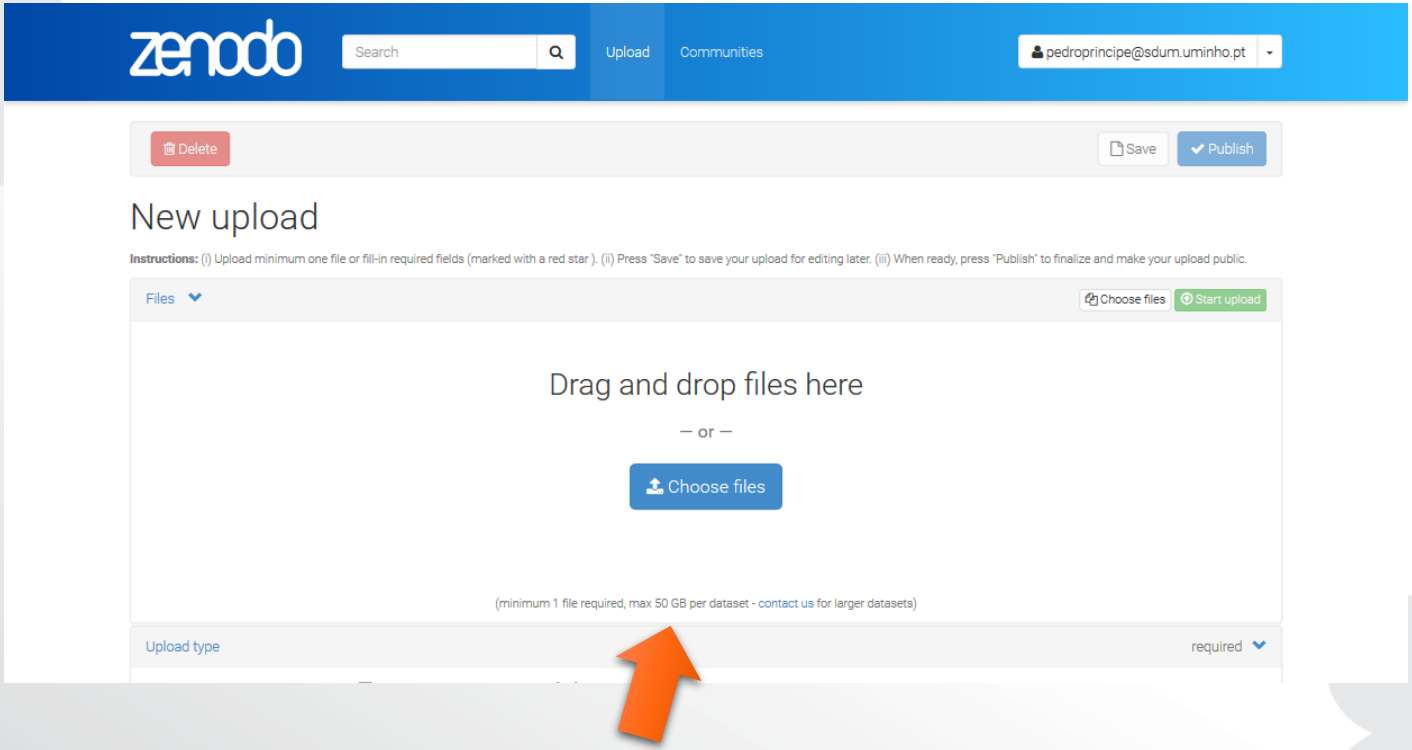
- **Multiple data types**
  - **Publications**
  - **Long tail of research data**
- **Citable data (DOI)**
- **Links to funding, pubs, data, software**



www.zenodo.org

H2020: Option to gather, preserve and share
project’s scientific output

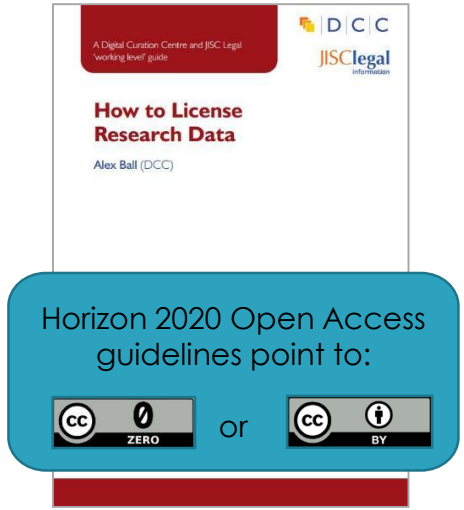

zenodo
Search
Upload
Communities
pedroprincipe@sdum.uminho.pt
Search uploads...
New Upload
Drafts 0
Published 0
Sort
Most recent
asc.
Get started!
Make your first upload - all research outputs from across all fields of research are welcome.
New Upload
About
About
Contact
Policies
Resources
Features
FAQ
Developers
REST API
OAI-PMH
Contribute
GitHub
Donate
Funded by
CERN
OpenAIRE
Powered by CERN Data Centre & Invenio
Privacy policy
Terms of Use
Support


zenodo
Search
Upload
Communities
pedroprincipe@sdum.uminho.pt
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 50 GB per dataset - contact us for larger datasets)
Upload type
required

Guidelines on DMPs

- How to develop a DMP
  www.dcc.ac.uk/resources/how-guides/develop-data-plan
- RDM brochure and template
  https://dans.knaw.nl/en/about/organisation-and-policy/information-material?set_language=en
- OpenAIRE guidelines
  www.openaire.eu/opendatapilot-dmp
- ICPSR framework for a DMP
  www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/framework.html

Licensing research data



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?

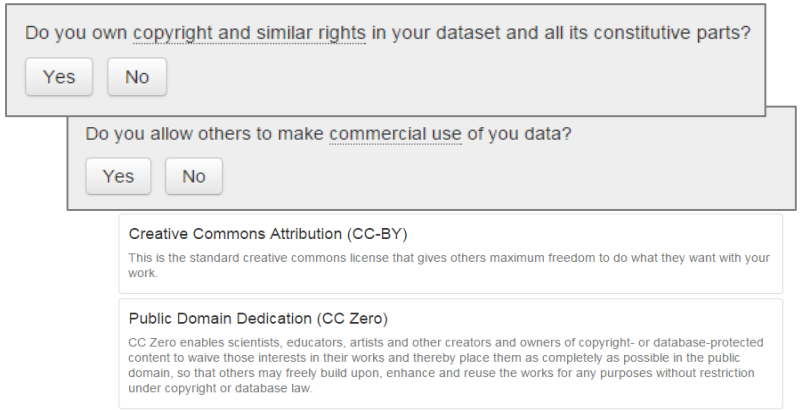
ND No Derivatives
Severely restricts use

These clauses are not open licenses

www.dcc.ac.uk/resources/how-guides/license-research-data

EUDAT licensing tool

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



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

Other resources

Where to keep research data http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data

Five steps to decide what data to keep http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep

Re3data http://www.re3data.org/

Figshare https://figshare.com/

Genbank https://www.ncbi.nlm.nih.gov/genbank/

How to write a lay summary http://www.dcc.ac.uk/resources/how-guides/write-lay-summary

Lay summaries https://www.bhf.org.uk/research/information-for-researchers/how-to-apply/lay-summaries

With thanks to

Marjan Grootveld:
marjan.grootveld@dans.knaw.nl
Sarah Jones: sarah.jones@glasgow.ac.uk

Acknowledgements:
Thanks to DANS and DCC for reuse of slide

www.eudat.eu www.openaire.eu

Research ethics and Data Protection

with thanks to GESIS

Research ethics: useful links

- National advisory board on research ethics (Helsinki, 2009): Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review: http://www.tenk.fi/sites/tenk.fi/files/ethicalprinciples.pdf
- RatSWD (German Data Forum): Principles and Review Procedures of Research Ethics in the Social and Economic Sciences: https://www.ratswd.de/dl/RatSWD_Output9.5_Summary_Research_Ethics.pdf
- Ethics Assessment in Different Fields: Humanities by Rok Benčin, Jelica Šumič Riha, Rado Riha, Scientific Research Centre of the Slovenian Academy of Sciences and Arts (ZRC SAZU): http://satoriproject.eu/media/2.e-Humanities.pdf

Research ethics

“The ethics of data focuses on ethical problems posed by the collection and analysis of large datasets and on issues ranging from the use of big data in biomedical research and social sciences, to profiling, advertising and data philanthropy as well as open data.”

Research ethics (2)

“Key issues concern possible re-identification of individuals through data-mining, -linking, -merging and re-using of large datasets, as well as risks for so-called ‘group privacy’, when the identification of types of individuals, independently of the de-identification of each of them, may lead to serious ethical problems, from group discrimination (e.g. ageism, ethnicism, sexism) to group-targeted forms of violence.”

Research ethics (3)

“Trust and transparency are also crucial topics in the ethics of data, in connection with an acknowledged lack of public awareness of the benefits, opportunities, risks and challenges associated with data science. For example, transparency is often advocated as one of the measures that may foster trust. However, it is unclear what information should be made transparent and to whom information should be disclosed.”

Research ethics (4)

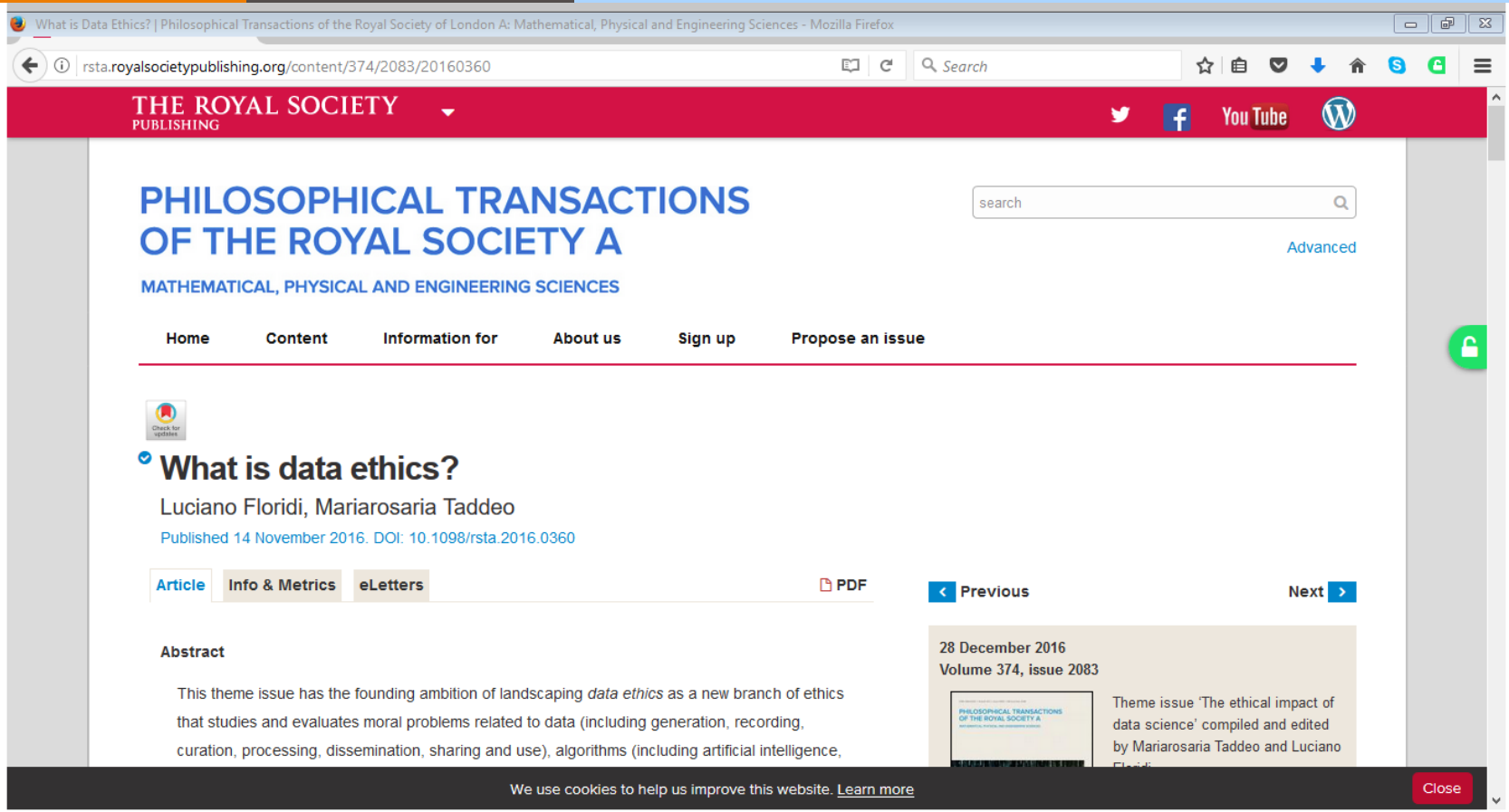
“The ethics of algorithms addresses issues posed by the increasing complexity and autonomy of algorithms broadly understood (e.g. including artificial intelligence and artificial agents such as Internet bots), especially in the case of machine learning applications. In this case, some crucial challenges include moral responsibility and accountability of both designers and data scientists with respect to unforeseen and undesired consequences as well as missed opportunities.”

Research ethics (5)

“Unsurprisingly, the ethical design and auditing of algorithms' requirements and the assessment of potential, undesirable outcomes (e.g. discrimination or the promotion of antisocial content) is attracting increasing research.”

Research ethics (6)

“Finally, the ethics of practices (including professional ethics and deontology) addresses the pressing questions concerning the responsibilities and liabilities of people and organizations in charge of data processes, strategies and policies, including data scientists, with the goal to define an ethical framework to shape professional codes about responsible innovation, development and usage, which may ensure ethical practices fostering both the progress of data science and the protection of the rights of individuals and groups. **Three issues are central in this line of analysis: consent, user privacy and secondary use.**"



http://rsta.royalsocietypublishing.org/content/374/2083/20160360

Some more useful links

- Guide to Research Ethics - Research with human participants (faculty of humanities): http://www.humanities.uct.ac.za/sites/default/files/image_tool/images/2/HumFaculty%20Ethics%20Guidebook%20August%20l%202016%281%29.pdf
- CESSDA User Guide on Research Data Management Data Consent and Ethics: https://cessda.net/content/download/245/2411/file/CESSDA%20User%20Guide%20for%20data%20management_8_Data%20consent%20and%20ethics.pdf

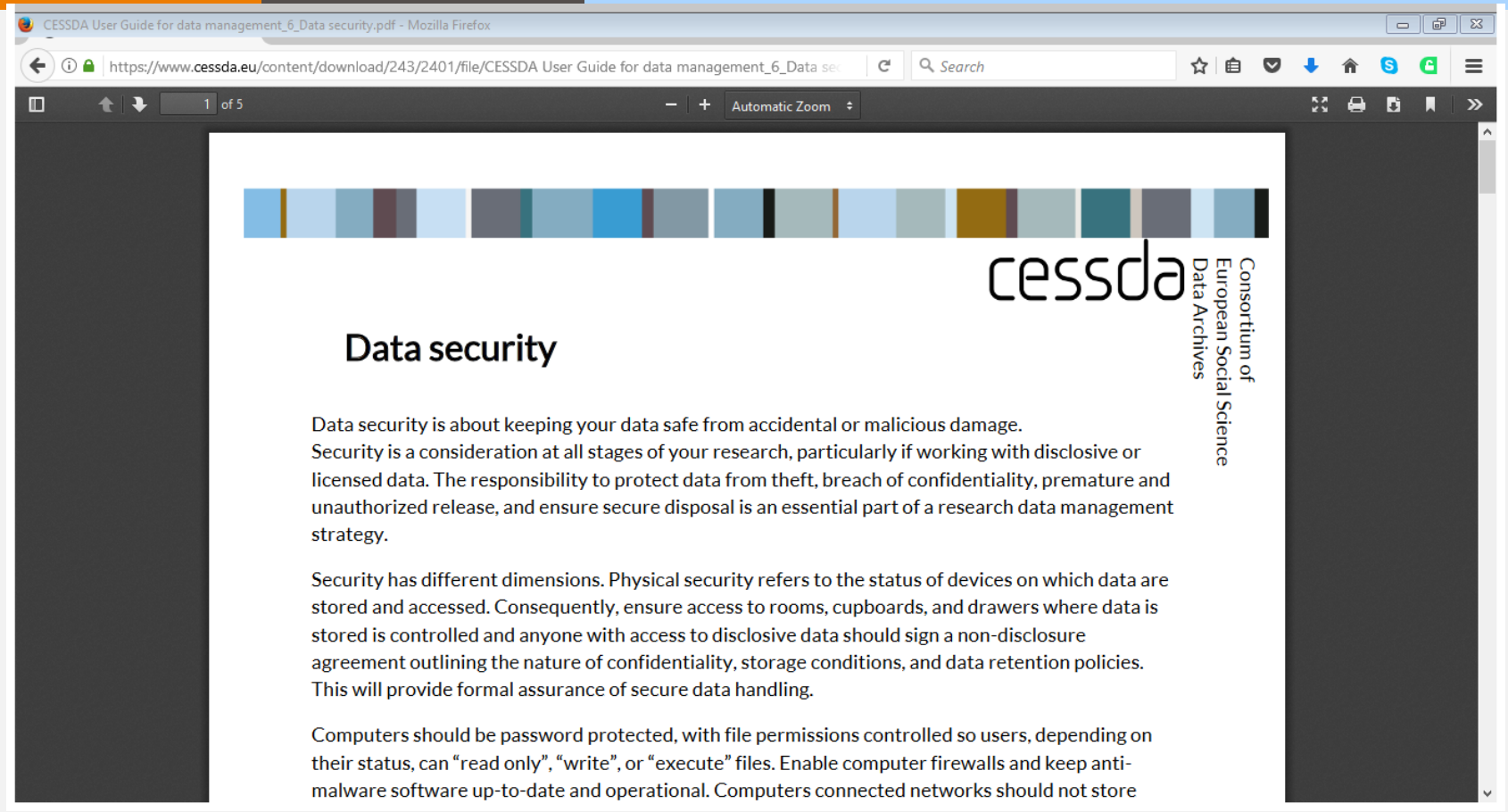
https://www.fosteropenscience.eu/content/research-ethics-and-legal-compliance-informed-consent-and-data-licensing

Data Protection

- Data protection is especially difficult in qualitative data (interviews, videos)
- Informed consent, participation is voluntary, aim and scope of survey and (re-)use of data must be transparent
- Participants in surveys can withdraw consent at any point in time, also after survey was completed, but only until data is completely anonymized

Data Protection (2)

Complete anonymisation is often very difficult or impossible to achieve for micro data, data can be shared when it is “factually anonymous” (at least by German court ruling, not quite sure about other countries)

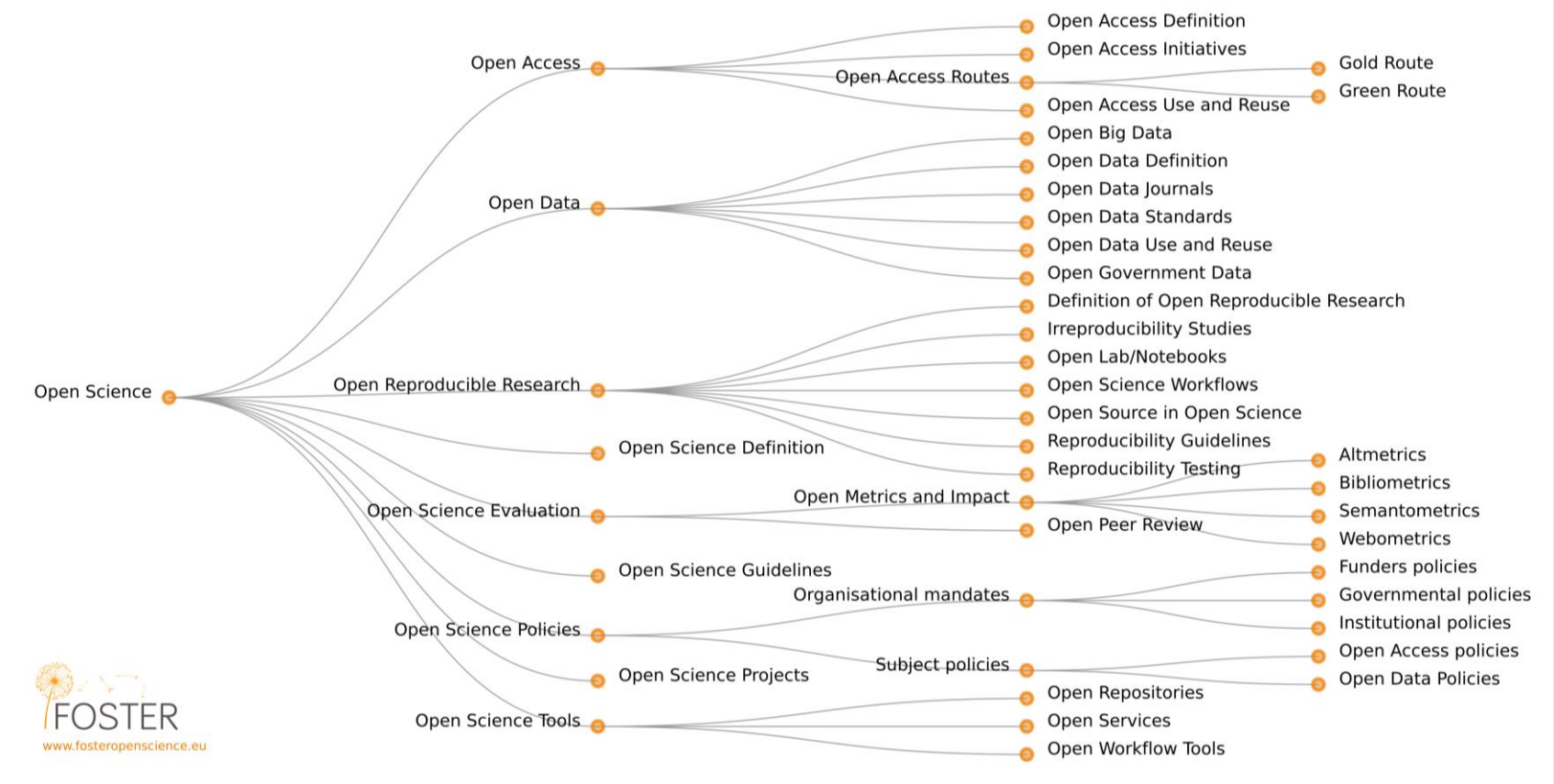


https://cessda.net/content/download/243/2401/file/CESSDA%20User%20Guide%20for%20data%20management_6_Data%20security.pdf

Data protection: tips

- Collected survey data and personal data (such as addresses, telephone number etc. needed for field work) must be stored separately
- Sensitive information (such as on race, ethnicity, health, religion, political views and engagement, sexual orientation) need even stronger protection
- Re-identification of survey participants is strictly forbidden for both primary researchers and secondary users

Open Science taxonomy



Paper available at http://oro.open.ac.uk/44719/. Image available at http://oro.open.ac.uk/47806/

FOSTER – an Open Science portal

COURSE: Introduction to Open Science

Intended audience Researchers and Students Level: Introductory: no previous knowledge is required

The following course is a general introduction to the various components and philosophies of Open Science, that can directly enrich each step of the scholarly lifecycle (Open Notebook Science, OpenData, Open Research Software, Open Access). The overall objective of the course is to provide an introduction to why Open Science is essential to rigorous, reproducible and transparent research, as well as to future research evaluation criteria focused on societal impact.

> Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods.

The learning outcomes of this course are:

- Understand the relevance of OS in relation to research integrity, reproducibility and impact;
- Identify suitable tools to help you embrace OS at each stage of the research lifecycle;
- Understand the potential of OS in supporting innovation and economic growth

FOSTER

Course: Introduction to Open Science

- Online material with re-usable/open licenses
- Videos, readings, quizzes, certificate
- Self-paced for the time being
- Forum where learners can post questions

FOSTER

Direct link to the course https://www.fosteropenscience.eu/content/open-science-scientific-research

FOSTER Objectives & project activities 2017-2019

- Strengthening Open Science training capacity in ERA
- Focusing on practical implementation of Open Science & ‘training the trainers’
- Training resources: new topics RDM & Open Data + intermediate & advanced level, and discipline specific
- Involving disciplines:
  - Humanities
  - Social sciences
  - Life sciences

DARIAH-EU
cessda
elixir

New Open Science training resources:
toolkit & training handbook

New functionalities on portal:
badging & gaming

More e-learning & face to face
trainings & training calendar

Initiate bootcamp & network for
Open Science trainers

Save the date:
Open Science Trainer Bootcamp 18-20 April

Become an Open Science trainer in our 3 day programme in Barcelona!

For more information check the news on our website.

www.fosteropenscience.eu/news

Book sprint: Open Science Training Handbook, 12-15 February, Germany

Share your experience and help to write a book!

Open Science trainers and educators will collaboratively author a training handbook.

Check the news on our website and apply now.

www.fosteropenscience.eu/news

Thank you!
Questions?

iryna.kuchma@eifl.net