Open Science

Sarah Jones
Digital Curation Centre, University of Glasgow
sarah.jones@glasgow.ac.uk
Twitter: @sjDCC

Outline of the session

- Introduction to open science
- Why be open?
- How to make your publications and data open
- Questions and discussion

FOSTER

WHAT IS OPEN SCIENCE?

Some definitions and clarifications

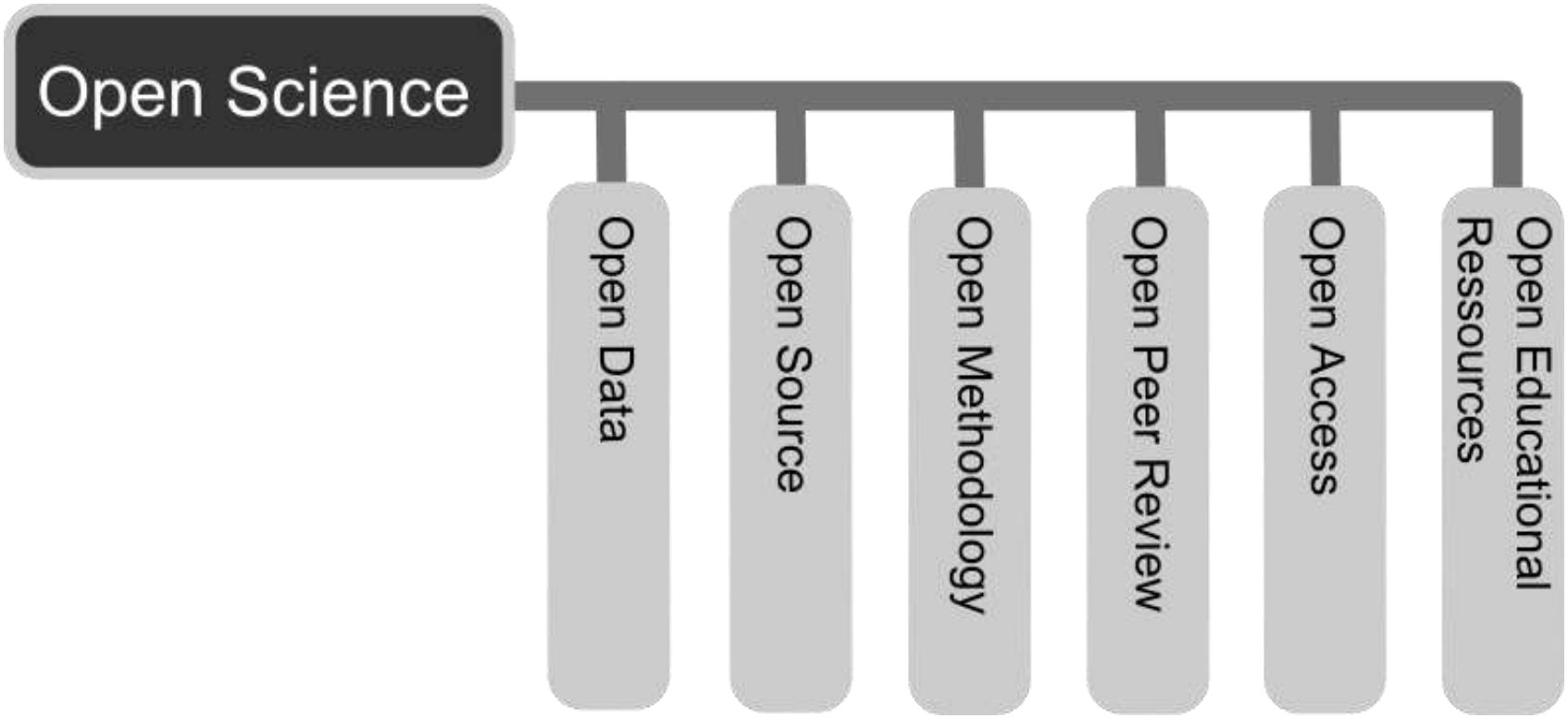
What is open science?

"science carried out and communicated in a manner which allows others to contribute, collaborate and add to the research effort, with all kinds of data, results and protocols made freely available at different stages of the research process."

Research Information Network, Open Science case studies
www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies

FOSTER

More than open access publishing



FOSTER

Why open access?

Open Access Explained!
www.youtube.com/watch?v=L5rVH1KGBCY

Open access to publications

- Free, immediate, online access to the results of research
- Free to reuse e.g. to build tools to mine the content
- Two routes to make sure anyone can access your papers
  - Gold route: paying APCs to ensure publishers makes copy open
  - Green route: self-archiving Open Access copy in repository
- Find out what your publisher allows on SHERPA RoMEO
  - www.sherpa.ac.uk/romeo

FOSTER

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

Open methods

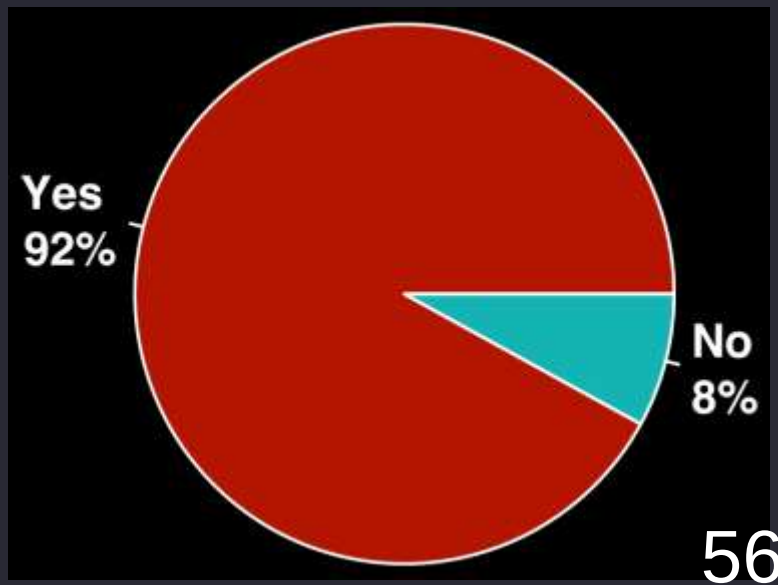
- Documenting and sharing workflows and methods
- Sharing code and tools to allow others to reproduce work
- Using web based tools to facilitate collaboration and interaction from the outside world
- *Open netbook science* – "when there is a URL to a laboratory notebook that is freely available and indexed on common search engines."
  http://drexel-coas-elearning.blogspot.co.uk/2006/09/open-notebook-science.html

FOSTER

Reliance on specialist research software

Slide from Neil Chue-Hong, Software Sustainability Institute

Do you use research software?



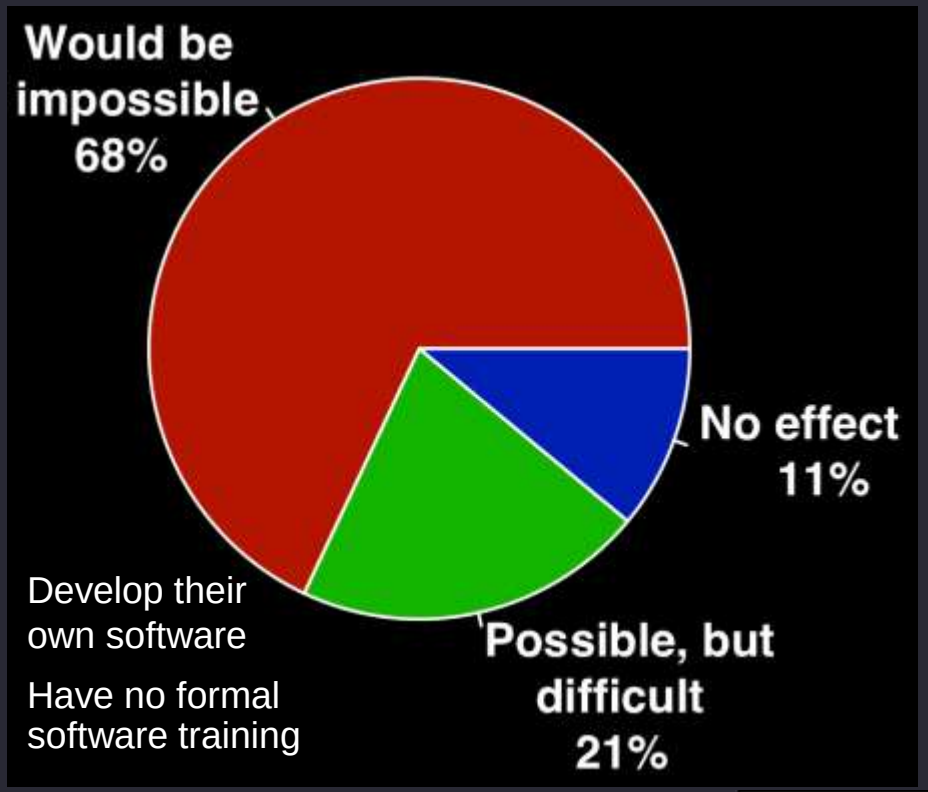
56%

Develop their own software

71%

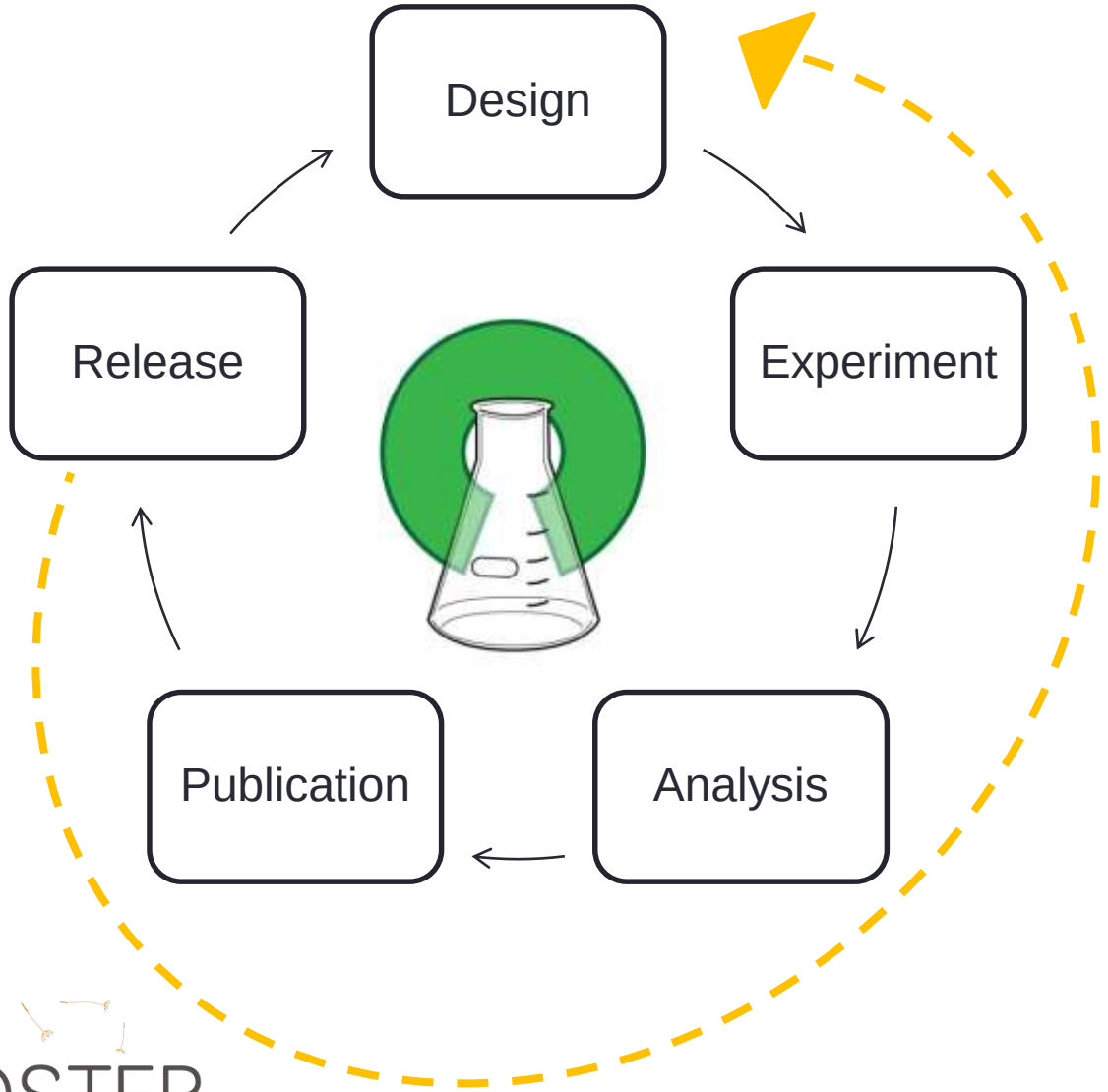
Have no formal software training

What would happen to your research without software



Survey of researchers from 15 UK Russell Group universities conducted by SSI between August - October 2014.
DOI: 10.5281/zenodo.14809

Openness at every stage



Change the typical lifecycle

Publish earlier and release more

Papers + Data + Methods + Code…

Support reproducibility

FOSTER

Degrees of openness

Five star open data

★★★★★

Unable to share
Under embargo

Open — Restricted — Closed

Content that can be freely used, modified and shared by anyone for any purpose

Limits on who can use the data, how or for what purpose

- Charges for use
- Data sharing agreements
- Restrictive licences
- Peer-to-peer exchange
- …

FOSTER

WHY PRACTICE OPEN SCIENCE?

Benefits and drivers

It’s part of good research practice

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

- Ewan Birney

#OpenData
#OpenScience

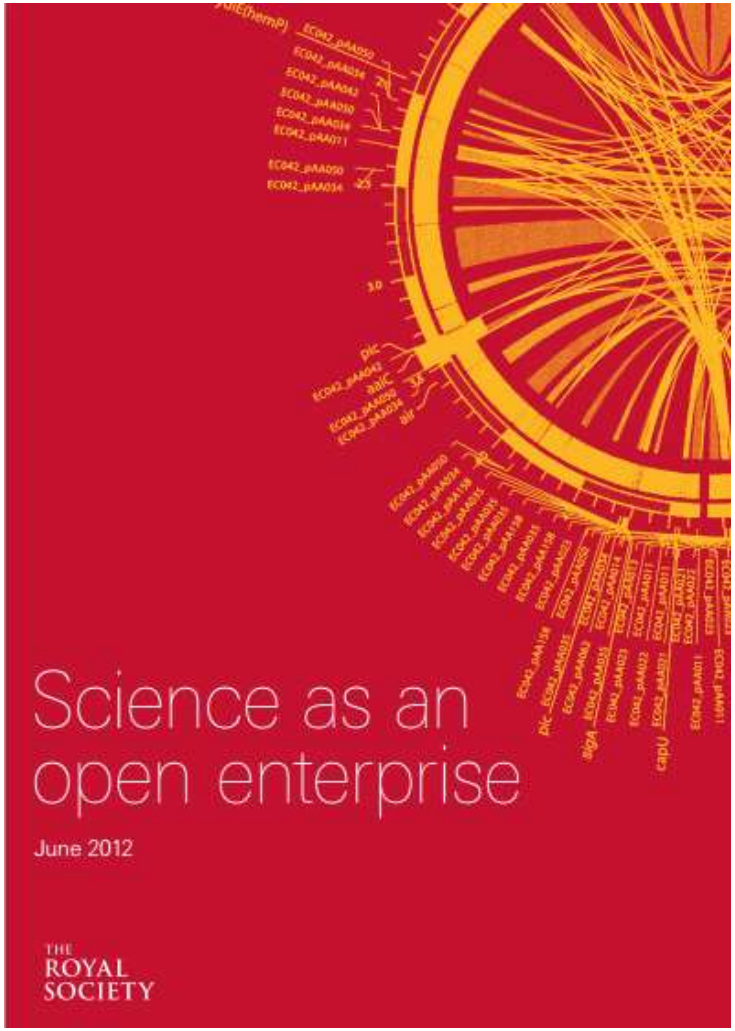
Original image via doi:10.1038/461145a. “Research cannot flourish if data are not preserved and made accessible. Data management should be woven into every course in science.” - *Nature* 461, 145

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.



FOSTER

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

Some benefits of openness

- You can access relevant literature – not behind pay walls
- Ensures research is transparent and reproducible
- Increased visibility, usage and impact of your work
- New collaborations and research partnerships
- Ensure long-term access to your outputs
- Help increase the efficiency of research

FOSTER

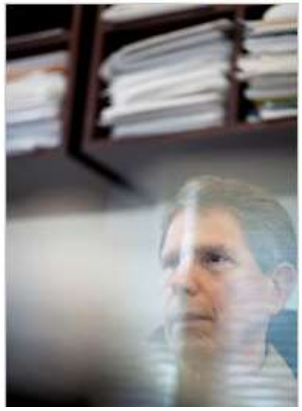
More scientific breakthroughs

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

Get 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

Funder imperatives...

"The European Commission's vision is that information already paid for by the public purse should not be paid for again each time it is accessed or used, and that it should benefit European companies and citizens to the full."

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

FOSTER

Open Science in Horizon 2020

Peer-reviewed publications

- Mandated to deposit machine-readable electronic copy of paper in repository by the date of publication
- Ensure OA via green/gold routes
- Embargo of 6 months (STEM) or 12 months (HSS) allowed
- Bibliographic metadata must be made openly available
- Aim to deposit research data

Research data

- Pilot for projects in named areas. Other can participate voluntarily.
- Applies to research data underlying publications, plus any other data as decided by project.
- Participants must:
  - Write a DMP as a project deliverable
  - Deposit data in a repository
  - Make it possible for others to access, mine, exploit and reuse the data
  - Share information on the tools needed

"As open as possible, as closed as necessary "

HOW TO PRACTICE OPEN SCIENCE?

Making your publications and data open

Routes to open access publication

Researcher **decides where to publish**

Check SHERPA RoMEO to see what OA and self-archiving options are available www.sherpa.ac.uk/romeo

Publish in an **open access journal**

Pay Article Processing Charge (APC) - *if required*

Immediate open access (via publisher)

GOLD OA ROUTE

GREEN OA ROUTE

Publish in a **subscription-based journal**

Search for a repository http://opendoar.org

Self-archive in a repository, based on publisher policy.

Immediate or delayed open access, depending on publisher's policy.

IF OPTION EXISTS
e.g. a **'hybrid' journal**
(*a subscription-based journal that has a paid open access option*)

Pay Article Processing Charge (APC)

Immediate open access (via publisher)

FOSTER

Sherpah RoMEO

Search again?

Journal titles or ISSNs Publisher names

nature

Exact title starts with contains ISSN

Advanced Search Search Reset

)836, ESSN: 1476-4687)

RoMEO:	This is a RoMEO **yellow** journal
Paid OA:	This journal is not in the list for the paid open access option.
Author's Pre-print:	✓ author **can** archive pre-print (ie pre-refereeing)
Author's Post-print:	**subject to Restrictions below** author **can** archive post-print (ie final draft post-refereeing)
Restrictions:	• 6 months embargo
Publisher's Version/PDF:	✗ author **cannot** archive publisher's version/PDF
General Conditions:	• Authors retain copyright • Published source must be acknowledged and DOI cited • Must link to publisher version • Publisher's version/PDF cannot be used • On author's personal website and institutional repository • If funding agency rules apply, authors may post authors version to their relevant funding body's archive, 6 months
Mandated OA:	Compliance data is available for 28 funders
Paid Open Access:	Open Access Hybrid Model - Selected Titles Only
Copyright:	Pre-publication policy - License to Publish - Manuscript Deposition Service
Updated:	06-Mar-2013 - Suggest an update for this record
Link to this page:	http://www.sherpa.ac.uk/romeo/issn/0028-0836/
Published by:	Nature Publishing Group - *Yellow Policies in RoMEO*

FOSTER

Deposit in your local repository!

- Speak to the library and deposit in your IR
- Consider other relevant repositories for your field too e.g. Arxiv - http://arxiv.org
- Deposit in Zenodo (catch-all repository) http://zenodo.org
- Check OpenDOAR for examples - http://www.opendoar.org

FOSTER

OpenAIRE

Open Access Infrastructure for research in Europe

- aggregates data on OA publications
- mines & enriches it content by linking thing together
- provides services & APIs e.g. to generate publication lists

www.openaire.eu

http://vimeo.com/108790101

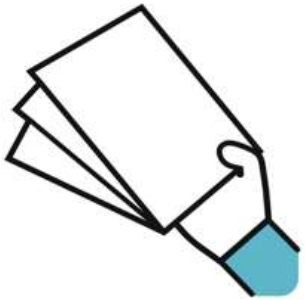
FOSTER

Open access button

Push Button.

The next time you're asked to pay to access academic research. Push the Open Access Button on your phone or on the web.

Get Research.



The Open Access Button will search the web for version of the paper that you can access immediately. If that doesn't work, the Button will email the author and look for more information about the paper.

Make Progress.

If you get your research, you can make progress with your work. If you don't get your research, your story will be used to help change the publishing system so it doesn't happen again.

The Open Access Button helps you get the research you want right now (without paying for it), and adds papers you still need to your wishlist.

FOSTER

https://openaccessbutton.org

How to make data open?

https://okfn.org

1. Choose your dataset(s)
   - What can you may 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

Licensing research data openly

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

Horizon 2020 Open Access guidelines point to:

CC 0 ZERO or CC BY

CREATIVE COMMONS LIMITATIONS

NC Non-Commercial
What counts as commercial?

ND No Derivatives
Severely restricts use

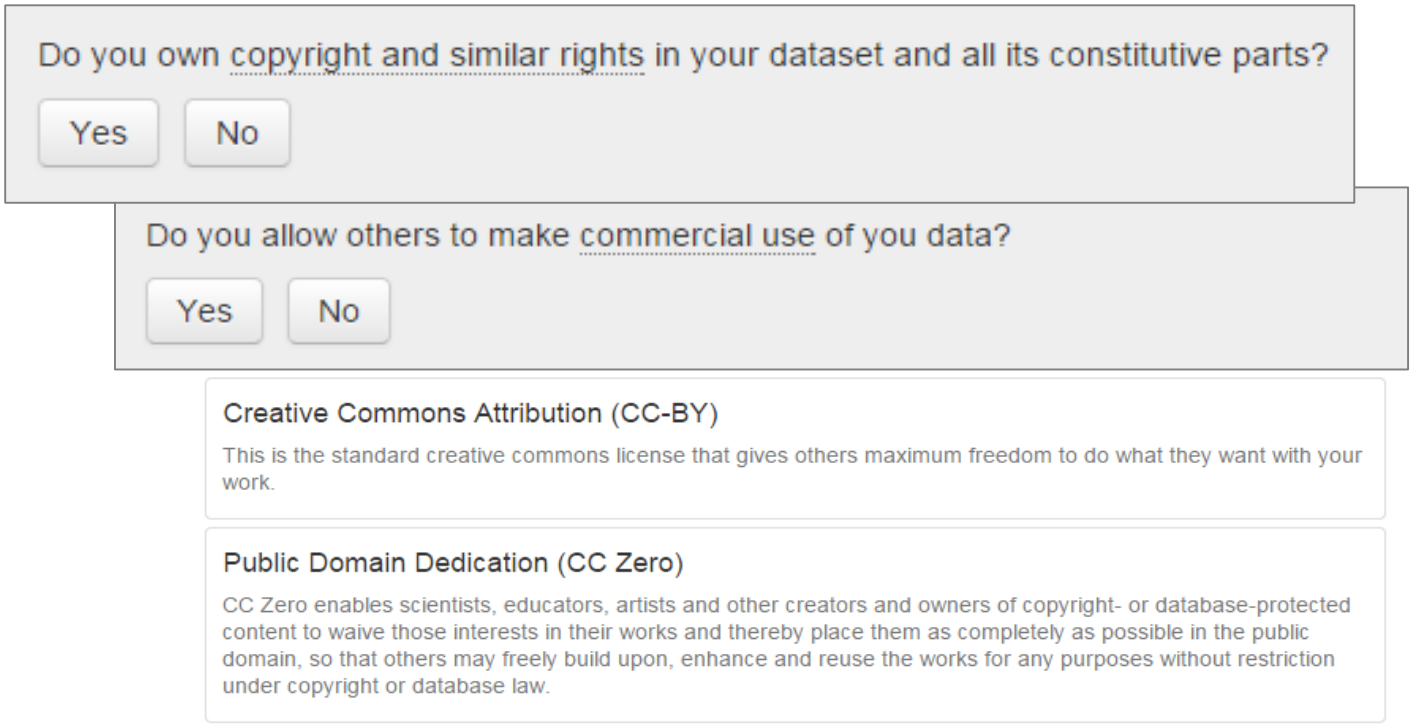
These clauses are not open licenses

FOSTER

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

EUDAT licensing tool

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



FOSTER

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

Metadata standards to use

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

Choosing appropriate file formats

If you want your data to be re-used and sustainable in the long-term, you typically want to opt for open, non-proprietary formats.

Type	Recommended	Avoid for data sharing
Tabular data	CSV, TSV, SPSS portable	Excel
Text	Plain text, HTML, RTF PDF/A only if layout matters	Word
Media	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	Quicktime H264
Images	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF	RDBMS

FOSTER

Further examples:
www.data-archive.ac.uk/create-manage/format/formats-table

Data repositories

- Does your publisher or funder suggest a repository?
- Are there data centres or community databases for your discipline?
- Does your university offer support for long-term preservation?

http://service.re3data.org/search

Zenodo

- OpenAIRE-CERN joint effort
- Multidisciplinary repository
- Multiple data types
  - Publications
  - Long tail of research data
- Citable data (DOI)
- Links funding, publications, data & software

www.zenodo.org

FOSTER

Plan for openness from the outset

Many decisions taken early on in the project will affect whether the data can be made openly available

- Think about where you want to publish and include APCs in grant applications if needed
- Ensure consent agreements also include permission to archive and share data for reuse by others
- Seek permissions for more than just the primary project purpose if signing licences to reuse third-party data. Derivative data may not be able to be shared if it includes somebody else's IP
- Explore the potential for openness when drafting agreements with commercial partners

FOSTER

Thanks – any questions

- DCC resources on Research Data Management
  www.dcc.ac.uk/resources

- FOSTER materials on Open Science
  www.fosteropenscience.eu

Follow us on Twitter:
@fosterscience
#fosteropenscience

FOSTER