



The benefits & practice of openness

Sarah Jones

Digital Curation Centre, University of Glasgow

sarah.jones@glasgow.ac.uk

Twitter: @sjDCC



What does it mean to be open?



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](http://www.rin.ac.uk/our-work/data-management-and-curation/open-science-case-studies)



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>



Open access to publications

- Free, immediate, online access to the results of research
- 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



Open data

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

<http://opendefinition.org>

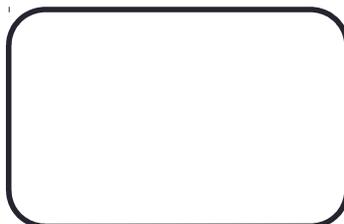
Tim Berners-Lee’s proposal for five star open data - <http://5stardata.info>

- ☐ 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

Openness at every stage



impactstory.org



myexperiment

www.myexperiment.org



DataCite

www.datacite.org



www.taverna.org.uk

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

LabTrove

www.labtrove.org



www.sherpa.ac.uk/rome

[Q](#)



 **OPEN WETWARE**

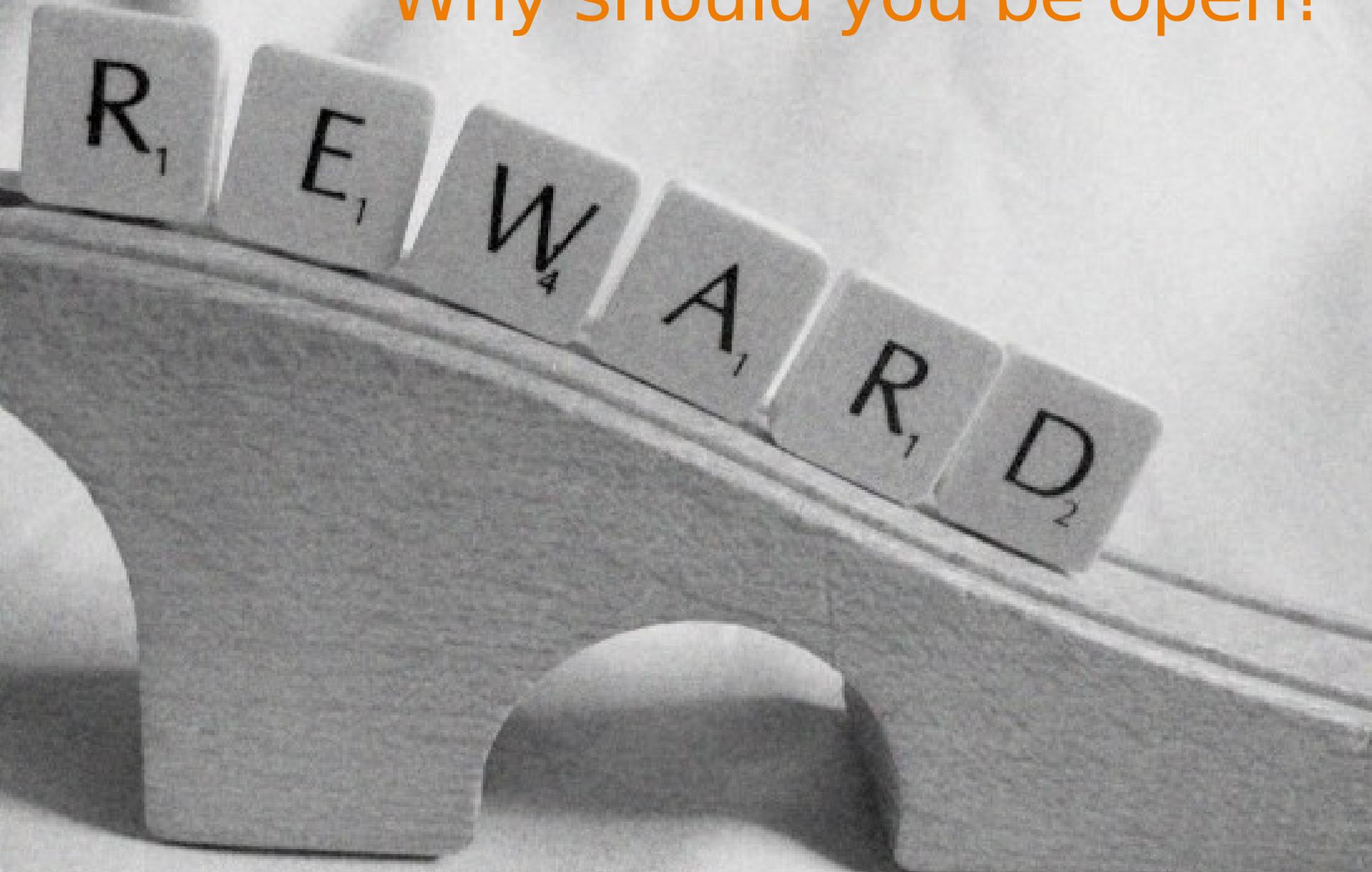
<http://openwetware.org>

GitHub

<https://github.com>



Why should you be open?



It's part of good research practice

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

- Ewan Birney

#OpenData
#OpenScience



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

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



Saving wasted time

OA helps to reduce time spent finding/accessing material:

“If around 60 minutes were characteristic for researchers (the average time spent trying to access the last research article they had difficulty accessing), then in the current environment the time spent dealing with research article access difficulties might be costing around DKK 540 million (EUR 72 million) per year among specialist researchers in Denmark alone.”

Access to research and technical information in Denmark,
Houghton, Swan & Brown (2011)

<http://eprints.ecs.soton.ac.uk/22603>



Cut down on academic fraud

nature International weekly journal of science Login

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

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

News

Report finds massive fraud at Dutch universities

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

Ewen Callaway

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



Dutch psychologist Diederik Stapel.
Persbureau van Eindhoven

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

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

Related stories

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

Naturejobs

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

Assistant Professor
Harvard Medical School

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

Resources

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

external links

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

Stories by subject

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

Stories by keywords

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

This article elsewhere

[Blogs linking to this article](#)

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

Validation of results

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

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

The error that could subvert George Osborne's austerity programme

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

Charles Arthur and Phillip Inman

The Guardian, Thursday 18 April 2013 21.10 BST



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

Acceleration of the research process

“As more papers are deposited and more scientists use the repository, the time between an article being deposited and being cited has been shrinking dramatically, year upon year.

This is important for research uptake and progress, because it means that in this area of research, where articles are made available at – or frequently before – publication, the research cycle is accelerating.”

Open Access: Why should we have it? Alma Swan
www.keyperspectives.co.uk

More scientific breakthroughs

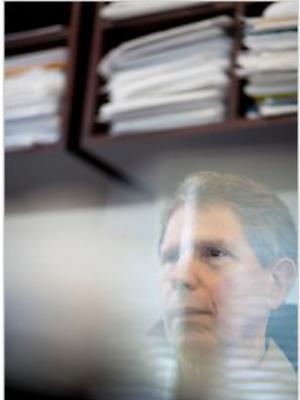
Sharing of Data Leads to Progress on Alzheimer's

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

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,

...r, H. & Vision, T. <https://peerj.com/articles/175>



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

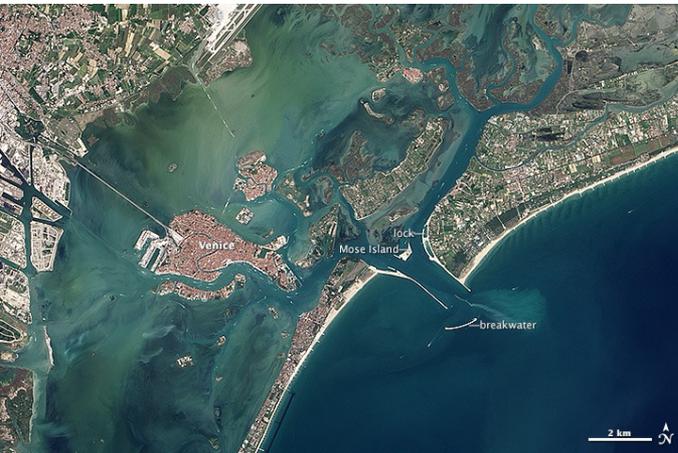
Freely available over the internet

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=v>

But there are also opportunity costs



THE OPPORTUNITY COST OF MY #OPENSOURCE WAS 35 HOURS + \$690

By Emilio Bruna
<http://brunalab.org/blog/2014/09/04/the-opportunity-cost-of-my-opensource-was-35-hours-690>

For his most recent paper:

1. Double checking the main dataset and reformatting to submit to Dryad: **5 hours**
2. Creating complementary file and preparing metadata: **3 hours**
3. Submission of these two files and the metadata to Dryad: **45 minutes**
4. Preparing a map of the locations: **1 hour**
5. Submission of map to Figshare: **15 minutes**
6. Cleaning up and documenting the code, uploading it to GitHub: **25 hours**
7. Cost of archiving in Dryad: **US\$90**
8. Page Charges: **\$600**

So what needs to change?

Conclusions from Emilio Bruna:

- Develop a better system of incentives from the community for archiving data and code
- Teach our students how to do this NOW - it's much easier if you develop good habits early
- Minimise the actual and opportunity costs

**We need to stop telling people “You should”
and get better at telling people “Here’s how”**



openness

connection

How to practice open science

Conducting science in the open: UsefulChem

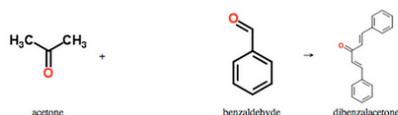


- Wiki Home
- Recent Changes
- Pages and Files
- Members

- All Experiments
- UC blog
- ONSchallenge
- UC on ChemSpider
- Mailing List
- Libraries
- References
- Experiment Format
- Extra Credit
- Paper03 Draft
- Isolated Compounds
- Alicia's Masters Thesis
- CombiUgi Project
- Open Web Drug Dev.
- To Do List
- Ugi Chemicals
- Ugi NMR Analysis
- Tim's Paper
- SLchemPaper
- ChemGamePaper
- DoUgi
- bots

Stemmeter

EXP284



Researcher

Matthew McBride

Objective

To synthesize [dibenzalacetone](#) from acetone and benzaldehyde using NaOH as the catalytic base in a 1:1 water/ethanol solvent using a modification of the procedure described in [Hawbecker78](#). The Hawbecker procedure describes using 6.0 equivalents of benzaldehyde, but this procedure uses 2.9 equivalents.

Procedure

(1E,4E)-1,5-Diphenyl-1,4-pentadien-3-one: A solution of NaOH (0.902 g, 22.55 mmol, 1.6 eq) in distilled water (40 mL) at room temperature was added to a stirred solution of benzaldehyde (4.361 g, 41.13 mmol, 2.9 eq) in ethanol (40 mL). Acetone (0.833 g, 14.36 mmol, 1.0 eq) was added to the reaction mixture. The reaction mixture was then stirred at room temperature for 30 minutes. The precipitated product was recovered by suction filtration, washed 1x with 2-3mL of 1:1 ethanol/water, dried over suction for 25 minutes, and recovered (3.044 g, 90.6%) as yellow crystals (mp 103-104°C).

Results

[Reaction Preparation Sheet](#)

Characterization of Product 284A

Amount: 3.044 g

Appearance: Bright yellow crystals

Melting Point: 103-104°C (lit mp 110-112°C [Hawbecker78](#))

[HNMR Spectrum](#) of UCEXP284_A

¹H NMR (500 MHz, CDCl₃) δ 7.72 (d, J = 16.3 Hz, 2H), δ 7.58 (m, 4H), δ 7.38 (m, 6H), δ 7.06 (d, J = 15.5 Hz, 2H)

Second [HNMR Spectrum](#) of UCEXP284_A2

Purity: Very pure sample of trans-dibenzalacetone by NMR

Discussion

This procedure produced pure trans-dibenzalacetone without having to recrystallize. The two characteristic doublets of trans-dibenzalacetone had coupling constants of 16.25Hz (located at 7.7ppm) and 15.5Hz (located at 7.1ppm). These coupling constants of near 15Hz indicates that the trans product was recovered.

The limiting reagent was the acetone (0.833 g, 14.36 mmol) and the percent yield of this reaction was 90.6%. An excess of benzaldehyde (2.87 equivalent) was used and the benzaldehyde was dissolved into the ethanol prior to being introduced into the reaction mixture. This was a major modification from the Hawbecker78 synthesis procedure, which used 6.0 equivalents of benzaldehyde and only listed yields of 40-60%. Additional modifications include that no scratching was needed to form crystals in the reaction mixture. The yellow crystals formed in the mixture immediately as they were formed. A second crop of crystals did not need to be recovered from the filtrate, but were recovered from the first filtration. The most important modification is that the product was recovered from the reaction mixture pure and did not need to be recrystallized. The Hawbecker78 procedure recrystallized the crystals from ethanol using water to crash the crystals out of solution. Product 284A was not recrystallized and was determined to be pure by HNMR.

The measured melting point of the product (103-104°C) is less than other [recorded melting point values](#). The melting point has been recorded to be around 110°C and the melting point measured in this experiment was 6-7°C less.

Conclusion

3.044 g of pure trans-dibenzalacetone was produced in a yield of 90.6%

Log

2012-07-17

14:28 0.902 g of NaOH pellets were placed in a 250 mL [Erlenmeyer flask](#) (#1).

14:30 Added 40 mL of distilled H₂O to the flask and placed on stir plate with stir bar to dissolve the pellets.

14:33 Added 40 mL of ethanol to a separate flask (#2)

Edit 0 24 ...



<http://usefulchem.wikispaces.com/EXP284>

Collaboration & sharing: MyExperiment

Version 7 (latest) (of 7) View version: **7 (latest)**

Version created on: 02/09/11 @ 11:43:00 by: Paul Fisher | Revision comment

Last edited on: 02/09/11 @ 11:44:57 by: Paul Fisher

Title: Pathways and Gene annotations for QTL region
Type: Taverna 2

Preview

(Click on the image to get the full size)

[Download Scalable Diagram \(SVG\)](#)

Workflow Type
Taverna 2

Original Uploader

Paul Fisher

License
All versions of this Workflow are licensed under:

Credits (1)
(People/Groups)
Paul Fisher

Attributions (0)
(Workflows/Files)
None

Tags (21)
 Original Uploader tags
adad | annotation | chromosome | data-driven | disease | ensembl | entrez | **gene** | genes | genotype | **kegg** | mouse | nbic onworkflows | **pathway** | pathway-driven | pathways | phenotype | qtl | shim | subworkflow | uniprot
[Log in to add Tags](#)

Shared with Groups (0)
None

Ratings (10)
Current: **4.6 / 5**
(10 ratings)
[Log in to rate and see breakdown of ratings](#)

Attributed By (7)
(Workflows/Files)
 The impact of workflow tools on data-centric research
Item doesn't exist anymore
 Pathways and Gene annotations for QTL region
 microRNA to KEGG Pathways and Abstracts
 Pathways and Gene annotations for QTL region
 KEGG Gene IDs to KEGG Pathways
 Pathways and Gene annotations for Arabidopsis affy data

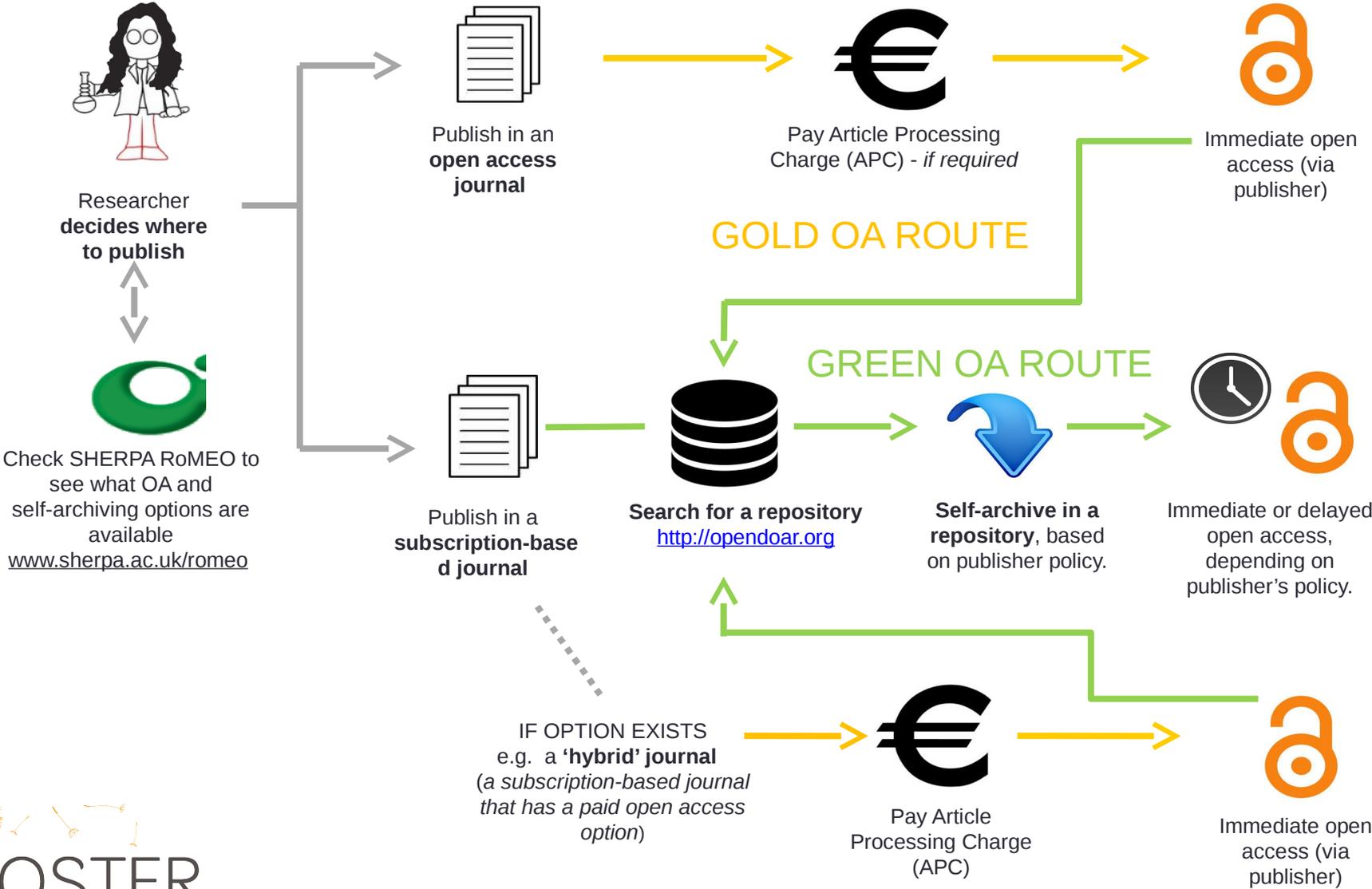
Favoured By (11)
 Katy Wolstencroft
 David Withers
 Taverna
 Xiaoliang
 Kawther
 AbuJarour
 Ali Rezaee
 Delistyle777
 Gamble
 Wotan
 Stian Solland-Reyes

Statistics
17944 viewings
6275 downloads
[\[see breakdown \]](#)

[More](#)



Open access publication



Sherpah RoMEO

Search again?

Journal titles or ISSNs Publisher names

nature

Exact title starts with contains ISSN

[Advanced Search](#)

1836, 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:	<input checked="" type="checkbox"/> author can archive pre-print (ie pre-refereeing)
Author's Post-print:	<input type="checkbox"/> subject to Restrictions below: author can archive post-print (ie final draft post-refereeing)
Restrictions:	<ul style="list-style-type: none">6 months embargo
Publisher's Version/PDF:	<input checked="" type="checkbox"/> author cannot archive publisher's version/PDF
General Conditions:	<ul style="list-style-type: none">Authors retain copyrightPublished source must be acknowledged and DOI citedMust link to publisher versionPublisher's version/PDF cannot be usedOn author's personal website and institutional repositoryIf 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

Deposit in your local repository!

- Speak to the library and deposit in Lirias - <https://lirias.kuleuven.be>
- Consider other relevant repositories for your field too e.g. Arxiv - <http://arxiv.org>
- Check OpenDOAR for examples - <http://www.opendoar.org>

OpenAIRE

Open Access Infrastructure for research in Europe

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

www.openaire.eu



<http://vimeo.com/108790101>

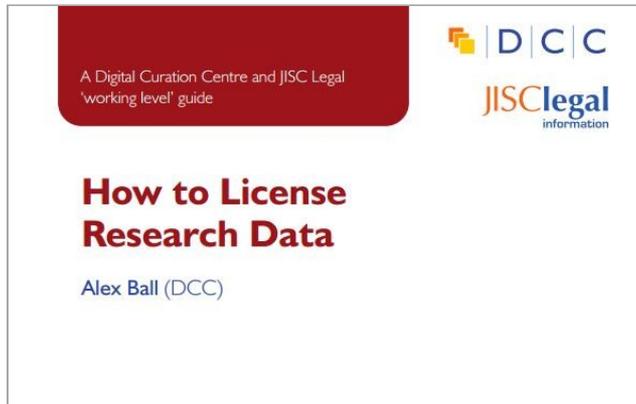
How to make data open?



<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...

Licensing research data



Outlines pros and cons of each approach and gives practical advice on how to implement your licence

Horizon 2020 Open Access guidelines point to:



or



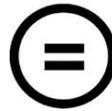
CREATIVE COMMONS LIMITATIONS



NC Non-Commercial
What counts as commercial?



SA Share Alike
Reduces interoperability



ND No Derivatives
Severely restricts use

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

Potential repositories

The screenshot shows the re3data.org website. At the top left is the logo "re3data.org" with the tagline "REGISTRY OF RESEARCH DATA REPOSITORIES". To the right of the logo is the URL "http://service.re3data.org/search". Below the logo is a navigation bar with links: Home, Search, Browse, Suggest, FAQ, About, Schema, Contact, Imprint. The main content area is titled "Search for Repositories (677 Reviewed Repositories)". There is a search input field. Below it are filters for "Subject", "Content Type", and "Country". There are also checkboxes for "Certificates", "Open Access", "Persistent Identifier", and "Include Repositories not yet reviewed by".

Zenodo

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

The screenshot shows the Databib website. At the top left is the logo "Databib" with the tagline "Find Repositories | Submit | Connect | About". To the right of the logo is the URL "http://databib.org". Below the logo is a navigation bar with links: Find Repositories, Submit, Connect, About. The main content area is titled "Featured Repository" and shows a search input field with a "Find" button. Below the search field is a link "Browse [Subjects | ABCDEFGHIJKLMNOPQRSTUVWXYZ]".

Metadata standards

Use relevant standards for interoperability

Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities



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

Thanks - any questions?

Follow us on Twitter:

[@fosterscience](https://twitter.com/fosterscience) and [#fosteropenscience](https://twitter.com/#fosteropenscience)

FOSTER training events and materials:

www.fosteropenscience.eu/events

