



UNIVERSITY OF  
OXFORD



UNIVERSITY OF  
CAMBRIDGE



OPEN KNOWLEDGE

# How to open up our research data (and why it matters)



Jenny Molloy | YEAR Annual Conference, Helsinki, May 2015

# You are free to:



Copy, share, adapt, or re-mix;



Photograph, film, or broadcast;



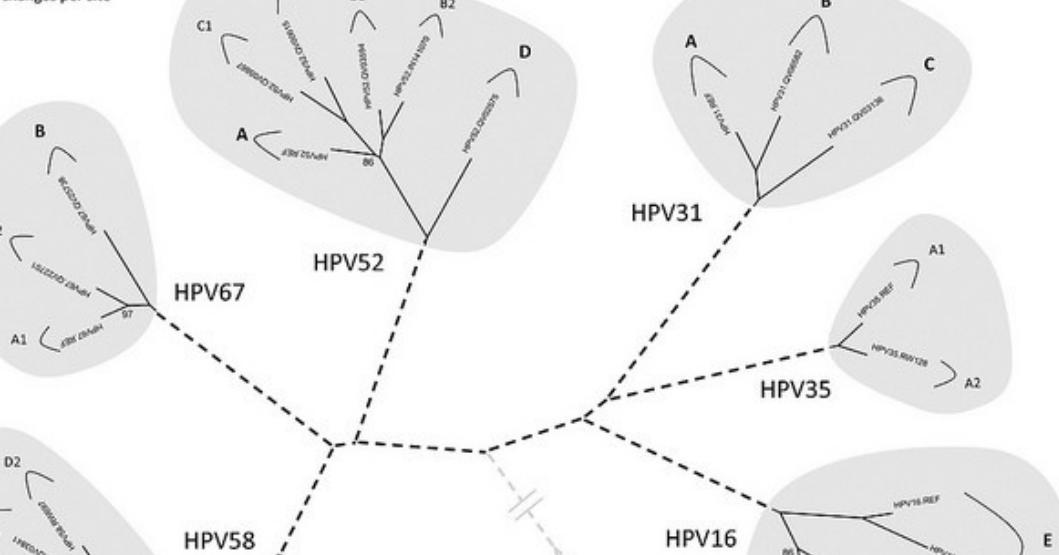
Blog, live-blog, or post video of;

# This presentation. Provided that:



You attribute the work to its author and respect the rights and licenses associated with its components.

What  
does  
“Data”  
mean?

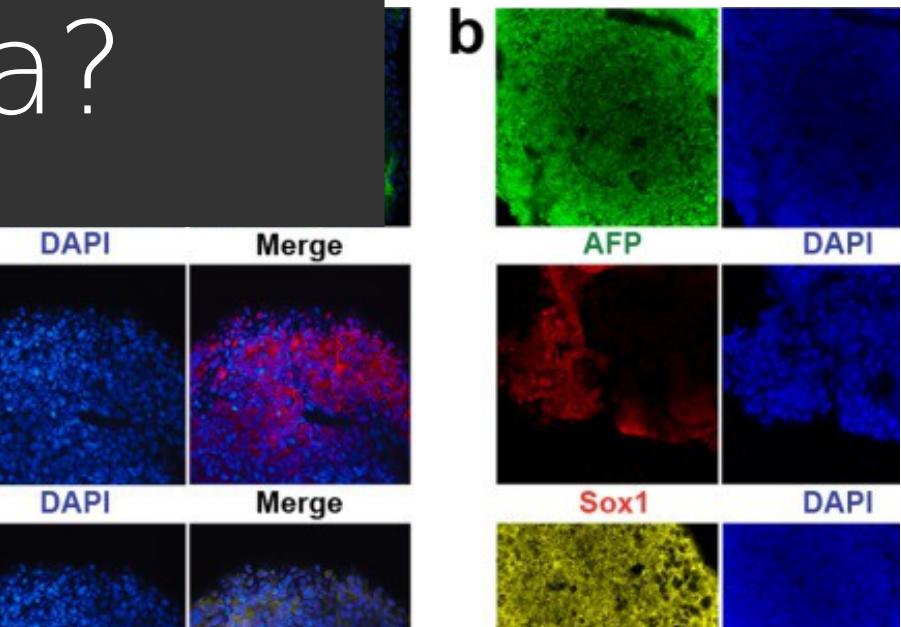
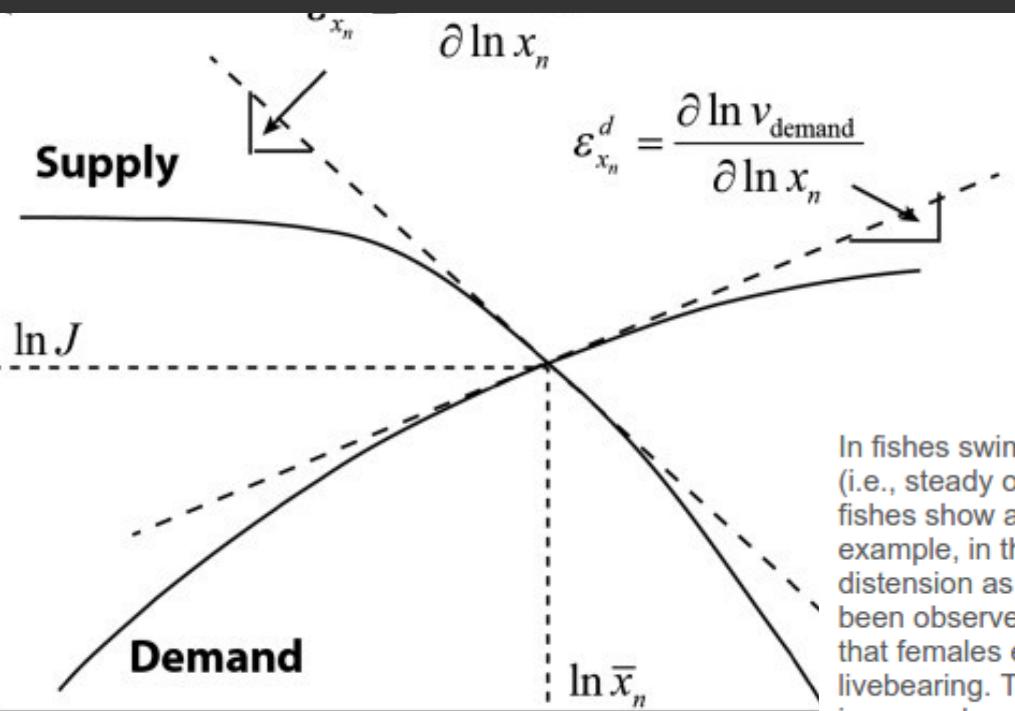


Repetitive element representation within Allometrix mouse microa

Microarray platform	LTR
Murine genome u74a v2 affy_mg_u74a_v2	243 (0.038) 79
Mouse genome 430 2.0 affy_mouse430_v2	2085 (0.330) 93
Mouse genome 430A 2.0 affy_mouse430a_v2	3 86 (0.061) 94
Mouse gene 1.0 ST affy_mogene_1_0_st	1581 (0.250) 233

Numbers of probes corresponding to LTR, LINE, and SINE elements within the Allometrix mouse microarray platforms are shown. Shortened platform names correspond to the names of the platforms within the 'oligo' Bioconductor R package. Numbers in brackets indicate the number of all individual LTR, LINE, or SINE probes.

# What is your data?



In fishes swimming performance is a major determinant of survival probability [16]. Swimming ability (i.e., steady or unsteady swimming measures) is strongly related to body form [16]–[19]. Livebearing fishes show a change in body form and increased overall mass in the latter stages of pregnancy; for example, in the livebearing fish *Brachyrhaphis rhabdophora*, females exhibit increased abdominal distension as pregnancy progresses [20], [21]. Pregnancy-related reduction in escape velocity has been observed in western mosquitofish (*Gambusia affinis*) [22], [23] and these studies have suggested that females experience a viability cost of reproduction as a consequence of the physical burden of livebearing. The argument is that changes in shape lead to reduced swimming ability which leads to increased mortality of pregnant females. However, the magnitude of this viability cost of reproduction is not well understood.

What  
does  
“Open”  
mean?



...sharing what we already create

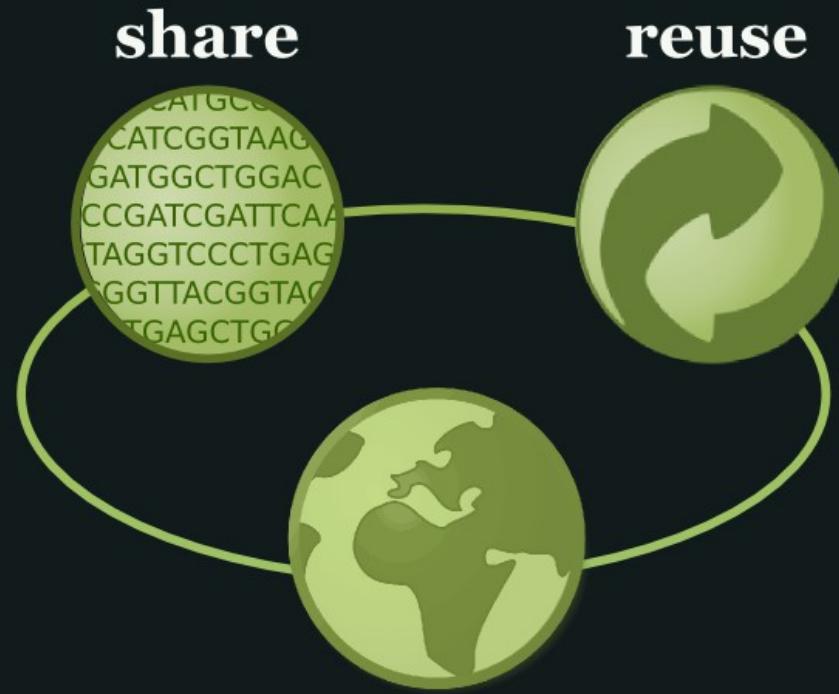
<http://www.flickr.com/photos/yourdon/3088582622>

A piece of data  
or content is  
**OPEN**  
if **anyone** is free to  
use, reuse, and redistribute it

subject **only**, at most, to the  
requirement to attribute  
and/or share-alike. *[opendefinition.org](http://opendefinition.org)*

# Panton Principles

principles for open data in science



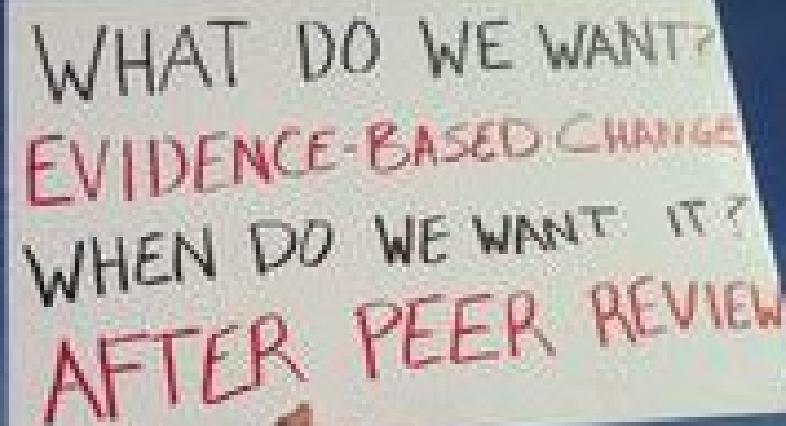
**Science** is based on  
**building on, reusing and**  
**openly criticising** the  
published body of scientific  
knowledge.

For science to  
**effectively function**,  
and for society to **reap**  
**the full benefits** from  
scientific endeavours,

it is **crucial** that  
science data be  
made **open**.

1. When publishing data make an  
**explicit** and **robust** statement of  
your wishes.

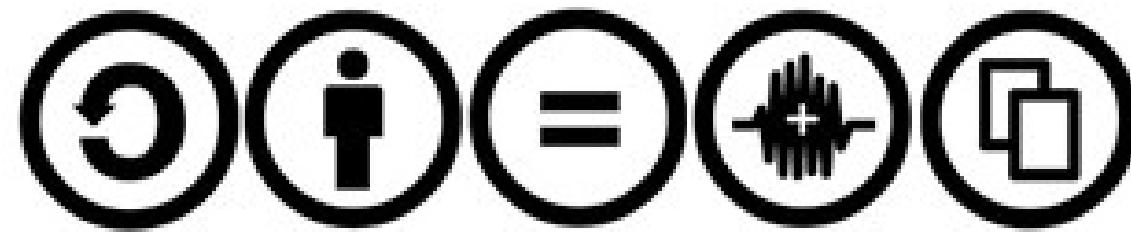
*pantonprinciples.org*



WHAT DO WE WANT?  
EVIDENCE-BASED CHANGE  
WHEN DO WE WANT IT?  
AFTER PEER REVIEW

2. Use a **recognized** waiver or license  
that is **appropriate** for data.

*pantonprinciples.org*



**creative  
commons**



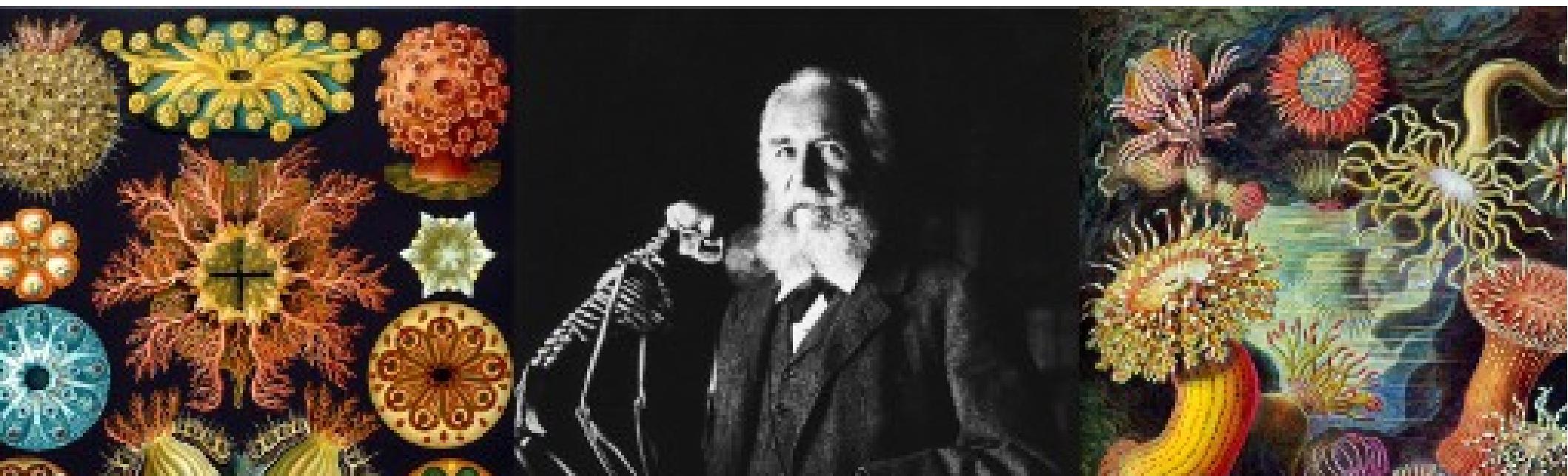
3. If you want your data to be **effectively** used and added to by others it should be open as defined by the Open Knowledge/ Data Definition – in particular **non-commercial** and other **restrictive clauses** should **not** be used.

*pantonprinciples.org*



4. **Explicit** dedication of data underlying published science into the **public domain** via PDDL or CCZero is **strongly recommended** and ensures compliance with both the Science Commons Protocol for Implementing Open Access Data and the Open Knowledge/Data Definition.

*pantonprinciples.org*



EASY STREET

So - do we  
actually  
**share?**

# Empirical Study of Data Sharing by Authors Publishing in PLoS Journals

Caroline J. Savage, Andrew J. Vickers 

**bioethics**

Clinical  
Chemistry

## Data Sharing in Medical Research: An Empirical Investigation

Daniel D. Reidpath<sup>1</sup>, Pascale A. Allotey<sup>2</sup>

Issue

Article first published online: 16 DEC 2002

DOI: 10.1111/1467-8519.00220



Bioethics  
Volume 15, Is  
125–134, Ap

## Data Submission and Quality in Microarray-Based MicroRNA Profiling

Kenneth W. Witwer<sup>1,\*</sup>

10,992

10

93

36

VIEWS

CITATIONS

ACADEMIC  
BOOKMARKS

SOC  
SHAI

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

## Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results

Jelte M. Wicherts , Marjan Bakker, Dylan Molenaar

## When Data Sharing Gets Close to 100%: What Human Paleogenetics Can Teach the Open Science Movement

Paolo Anagnostou , Marco Capocasa, Nicola Milia, Emanuele Sanna, Cinzia Battaggia, Daniela Luzi, Giovanni Destro Bisol 

Some of  
the time....

In groups of 3, you have 10 mins to list  
3 reasons why...

YOU WOULD share your data  
YOU might NOT share

Terrorists and/or criminals will use it	It's too complicated	People may misinterpret the data	It's not very interesting
I don't mind, but someone else might	We might want to use it in a(nother) paper	People will contact me to ask about stuff	Data Protection/ National Security
It's too big	People will see that my data is embarrassingly bad	It's not useful to anyone else	It's not a priority and I'm busy
I don't know how	I'm not sure I own the data	Someone might steal/plagiarise it and not credit me	My funder doesn't require it

Terrorists and/or criminals will use it	It's too complicated	People may misinterpret the data	It's not very interesting
I don't mind, but someone else might	We might want to use it in a(nother) paper	People will contact me to ask about stuff	Data Protection/ National Security
It's too big	People will see that my data is embarrassingly bad	It's not useful to anyone else	It's not a priority and I'm busy
I don't know how	I'm not sure I own the data	Someone might steal/plagiarise it and not credit me	My funder doesn't require it



# 5 steps to Sharing

# 1

Get  
motivated:  
start  
early



# CASH REWARD

for returning my lost backpack



505Adventure.com

- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is **VERY** important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: (██████████)

Thank you!!

Think you don't need a research data management plan?

Learn from those who have gone before you and think again!





Talk about data sharing to your colleagues, bosses, collaborators, funders, ethics committees, data providers, research subjects, librarians, departments, other students, friends, pets...

2

Stay on top  
of your data





## Pick a face:

What was that unnamed column?

Which version of the data did I end up feeding into R?

Which of my three lab books did the protocol end up in?

To share, first YOU have to document and understand.

# Find the best practises in your research community

e.g. [biostandards.org](https://www.biostandards.org)

**Policies Registry**



A catalogue of 21 data preservation, management and sharing policies from international funding agencies and regulators.

Call for collaborators

If you have catalogues you would like us to link to, contact us.

**Standards Registry**



A catalogue of reporting standards and organizations that develop these. We have 71 reporting guidelines, 339 terminology artefacts, and 177 models and formats.

Partly compiled by linking to

BioPortal mibbi equator

**Databases Registry**



A catalogue of 676 databases, described according to the BioDILcore guidelines, along with the standards used within them.

Partly compiled with the support of

OXFORD UNIVERSITY PRESS re3data.org

There are many online courses and guides for multiple disciplines, plus all those people to whom you've already been talking!

Share  
like  
you  
want  
to

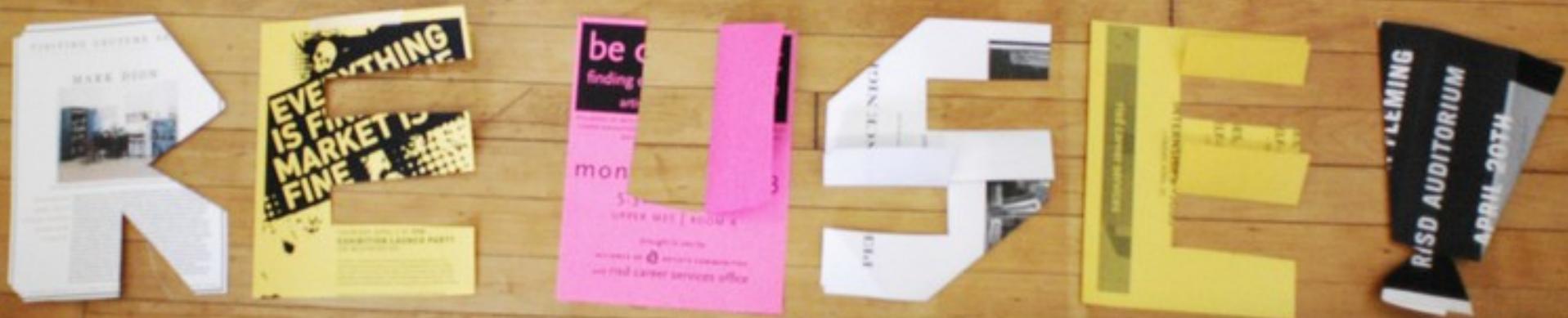
3





# Discoverability

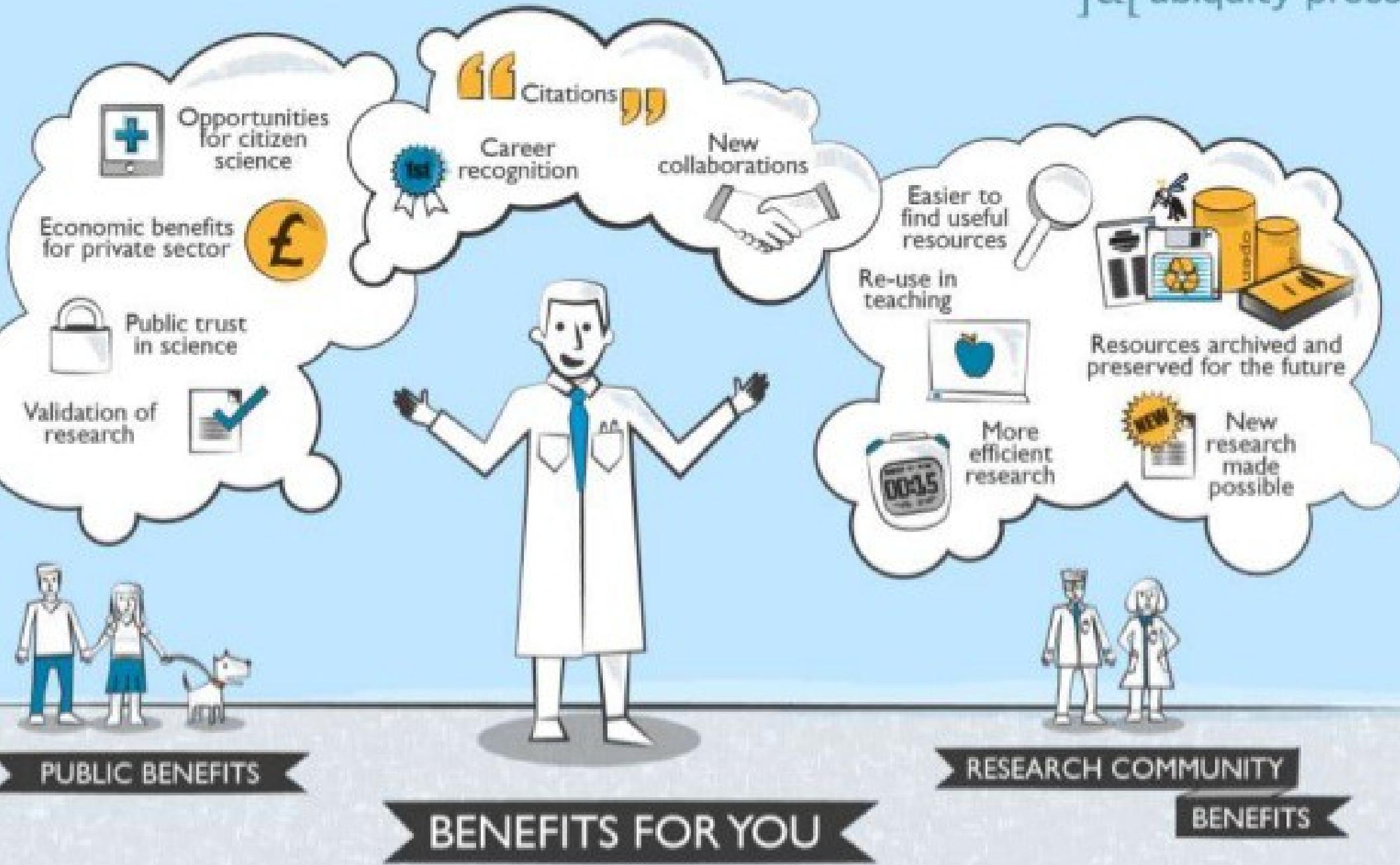
# Get your data ready for



A close-up photograph of several green smiley face stickers. One yellow smiley face is positioned in the center foreground, slightly overlapping the others. All the smiley faces have black outlines and dark blue eyes. The background is dark.

4

Make the  
most of it



5



[citation needed]

Play fair

# 5 steps to Sharing

## 1. Get motivated and start early:

Have a plan

Manage expectations

## 2. Keep on top of your data

Will you understand it in four years?

What are the best practices in your field?

## 3. Share it like you really want to

Make it discoverable and reuseable

Metadata, metadata, metadata

## 4. Make the most of it

Write data papers, report it to funders

Check for citations

## 5. Play fair

Set an example: cite data properly.

Where applicable, contact data creators – collaborate!

Any  
questions?



This presentation and all original content is licensed under CC-BY 3.0 <http://creativecommons.org/licenses/by/3.0/>, Credited images below remain under the license used by the creator, including derivative images such as crops and recolourised versions produced for this presentation.

## Image Credits:

All OKF images under CC-BY

Sharing Slide: Photo: Ed Yourdon under CC-BY-NC-SA. Slide by Cameron Neylon

Open by Todd Huffman on Flickr, licensed under CC-BY 2.0 <https://www.flickr.com/photos/oddwick/2087333395>

Licencias Creative Commons by Arturo Quirantes, licensed under CC-BY-NC-SA 2.0

<http://elprofedefisica.naukas.com/2011/04/18/defendiendo-las-licencias-creative-commons-un-caso-personal/>

Restaurant's "Sorry we're Closed" sign by Nick Papakyriazis on Flickr, licensed under CC-BY-NC-SA 2.0

<https://www.flickr.com/photos/mlrs193/6015396482>

Easy Street by HA! Designs – Artbyheather on Flickr, licensed under CC-BY-NC-SA

<https://www.flickr.com/photos/hadesigns/2688188166>

BINGO! By Marlon E on Flickr, licensed under CC-BY-SA 2.0

<https://www.flickr.com/photos/sjsharktank/4630354008>

Researcher data sharing image: Ainsley Seago in Roche DG, Lanfear R, Binning SA, Haff TM, Schwanz LE, et al. (2014) Troubleshooting Public Data Archiving: Suggestions to Increase Participation. PLoS Biol 12(1): e1001779. doi:10.1371/journal.pbio.1001779

What's your motivation? By opensource.com on Flickr, licensed under CC-BY-SA 2.0

<https://www.flickr.com/photos/opensourceway/4504724163>

Backpack photo by Peter Murray-Rust, licensed under CC-BY 2.0

<http://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan/>

Discussion at OKCon Berlin by Open Knowledge Foundation on Flickr, licensed under CC-BY 2.0

<https://www.flickr.com/photos/okfde/9504259162>

Social Media Information Overload by Mark Smiciklas on Flickr, licensed under CC-BY-NC 2.0

<https://www.flickr.com/photos/intersectionconsulting/7537238368>

Stress by Bernard Goldbach on Flickr, licensed under CC-BY 2.0

<https://www.flickr.com/photos/topgold/6273248505/>

Jumpin' Giminy by Linda Tanner on Flickr, licensed under CC-BY-NC-ND 2.0

<https://www.flickr.com/photos/goingslo/8684346961/in/photostream/>

Reuse! By marcos ojeda on Flickr, licensed under CC-BY-NC-ND 2.0

<https://www.flickr.com/photos/subliminal/3486562413/>

Smile! By seanbjak on Flickr, licensed under CC-BY-ND 2.0

<https://www.flickr.com/photos/sean-b/245744537/>

Data sharing infogrpahic by Ubiquity Press

<http://www.slideshare.net/brianhole/data-citation-a-critical-role-for-publishers>

"Citation needed" by [www.futureatlas.com](http://www.futureatlas.com) on Flickr, licensed under CC-BY 2.0

<https://www.flickr.com/photos/87913776@N00/5129607997>