



# Open data, FAIR data and RDM: the ugly duckling

Sarah Jones

Digital Curation Centre, Glasgow

[sarah.jones@glasgow.ac.uk](mailto:sarah.jones@glasgow.ac.uk)

Twitter: @sjDCC

Slides available at: <https://doi.org/10.5281/zenodo.1196631>





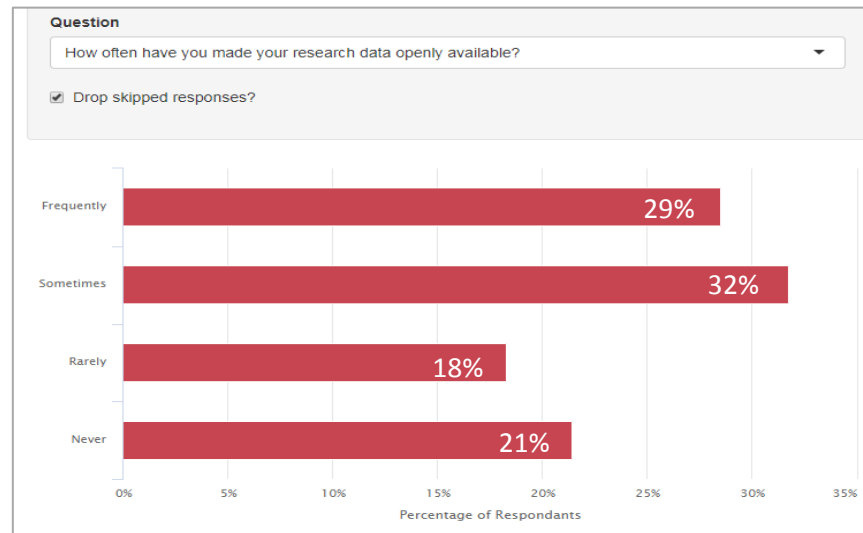
# How many researchers make data open?

79% of researchers have made data openly available

[The State of Open Data 2017](#)

Digital Science

2300 respondents worldwide



only 1 in 10 provides their research data as open data for the public

[Researchers and their data \(2015\)](#)

eInfrastructures Austria

3026 Austrian respondents

64% agree that they are willing to share their data

[Open Data: the researcher perspective \(2017\)](#), Elsevier

1162 respondents worldwide

68% of researchers already share data or expect to do so in future

[Jisc DAF studies \(2016\)](#)

1185 UK respondents

# How do researchers share data?

Over half only allow access on request.

54% share data by using external storage devices or email.

[eInfra Austria: Researchers and their data](#)

Of 13 methods stated, top 4 options for currently sharing data were:

1. Emailing data files (65%)
2. Cloud service e.g. Dropbox, Googledrive (59%)
3. Portable storage (35%)
4. Supplementary data (20%)

Formal repository (public / institutional) c.12%

[Jisc DAF studies](#)

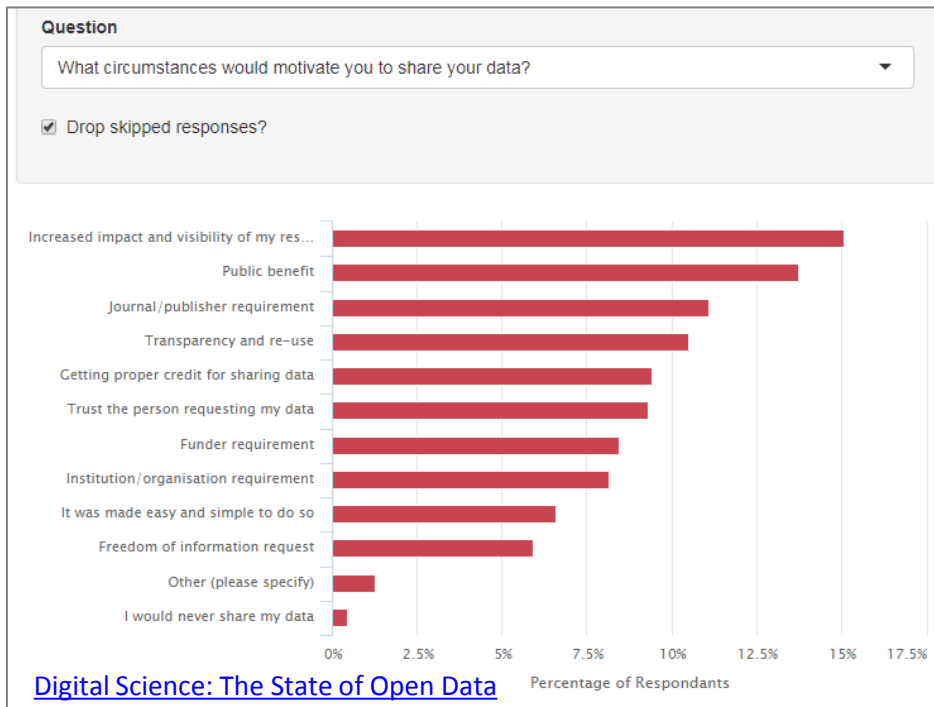
“When asked where they have published data, most commonly respondents had done so as an appendix to an article (just over 30%) with a data repository close behind (just under 30%) and 20% having published in a data journal.”

[Digital Science: The State of Open Data](#)

Less than 15% publish data in a repository.

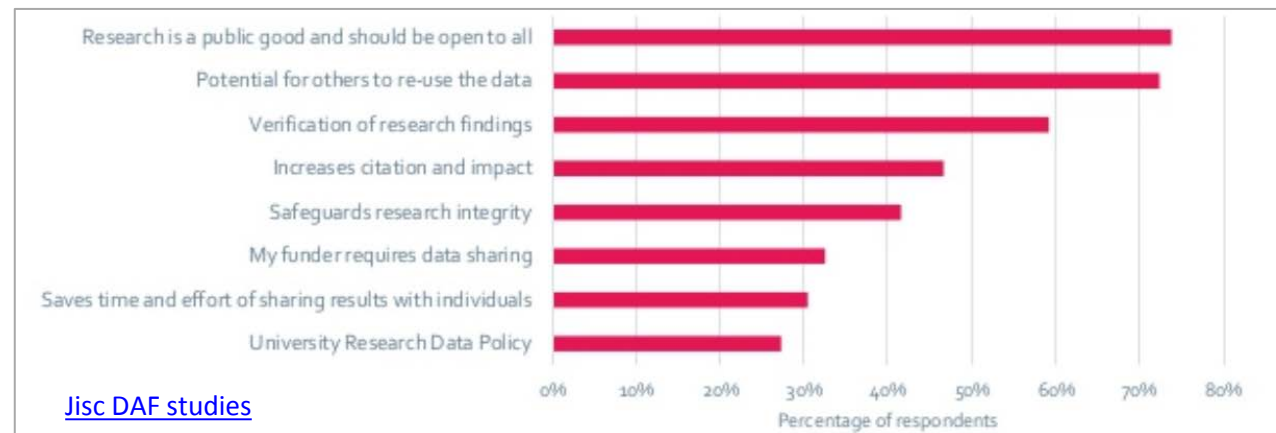
[Elsevier: Open Data - the researcher perspective](#)

# Why do researchers share data?



*“For more than half of the researchers, the most attractive incentives for sharing their data were increased visibility and impact, new cooperation opportunities, recognition in professional circles, as well as their contributions being regarded as scientific output.”*

[Infra Austria: researchers and their data](#)



# Data storage and loss

36% had experienced loss and 83% of this was due to physical storage media

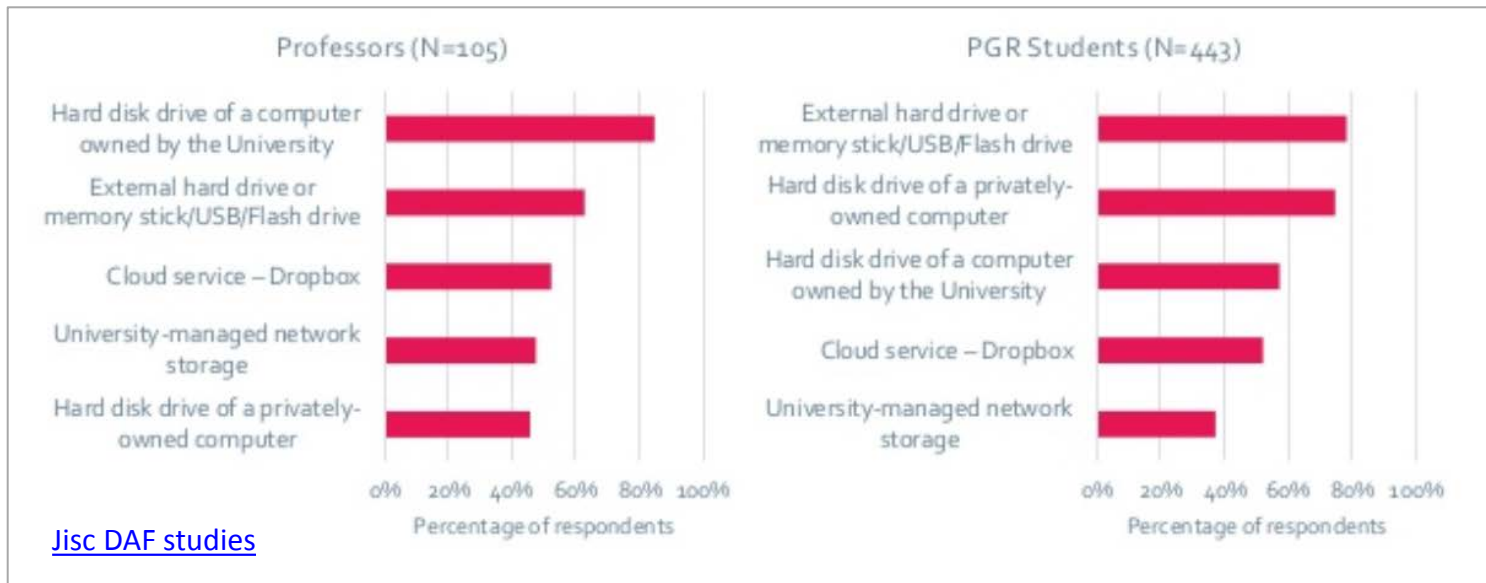
[Digital Science: The State of Open Data](#)

More than one-third had experienced data loss.

Strong preference to store on business/private computer, external hard drive & usb

[Infra Austria: researchers and their data](#)

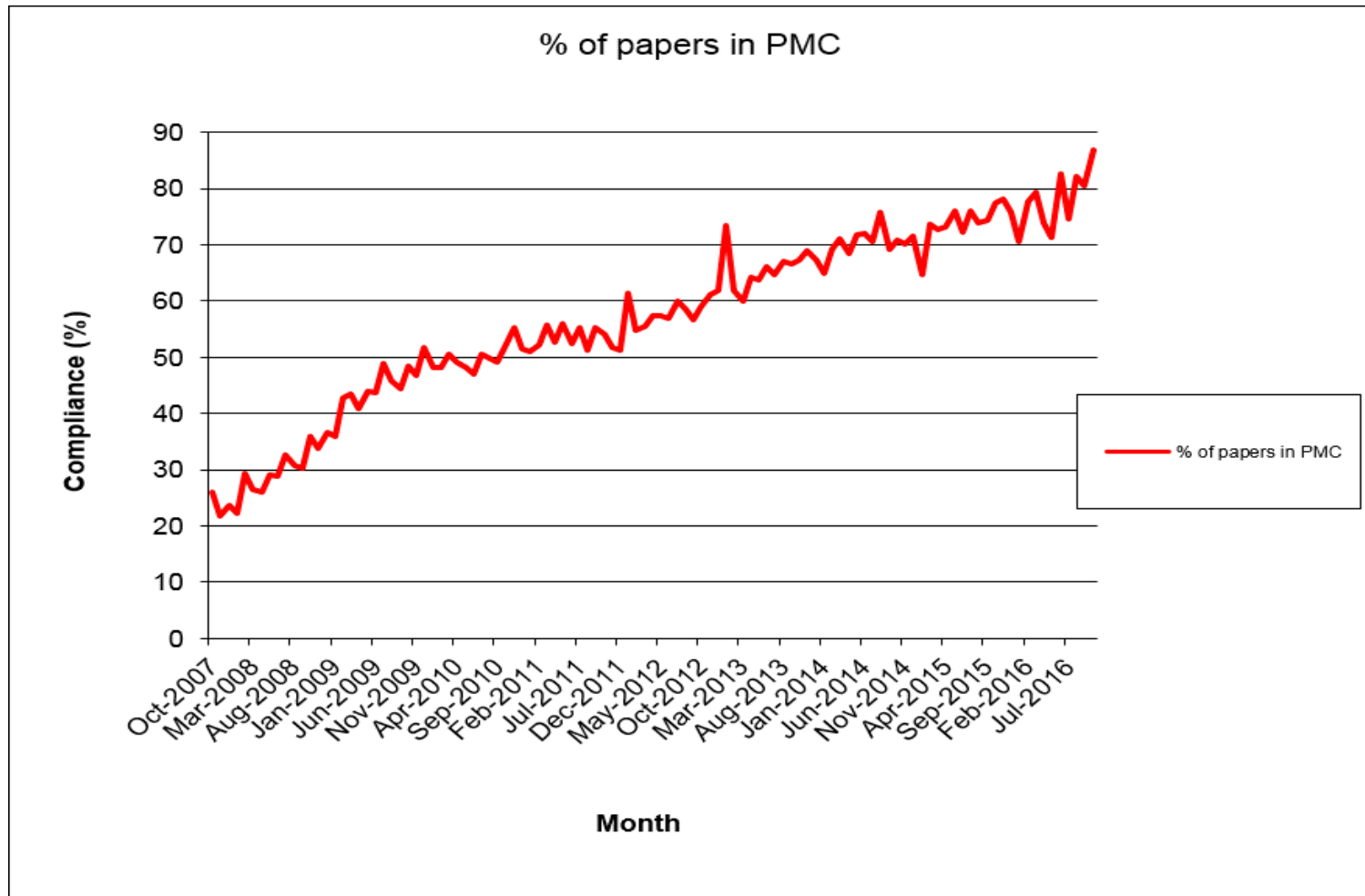
17% of respondents had lost data



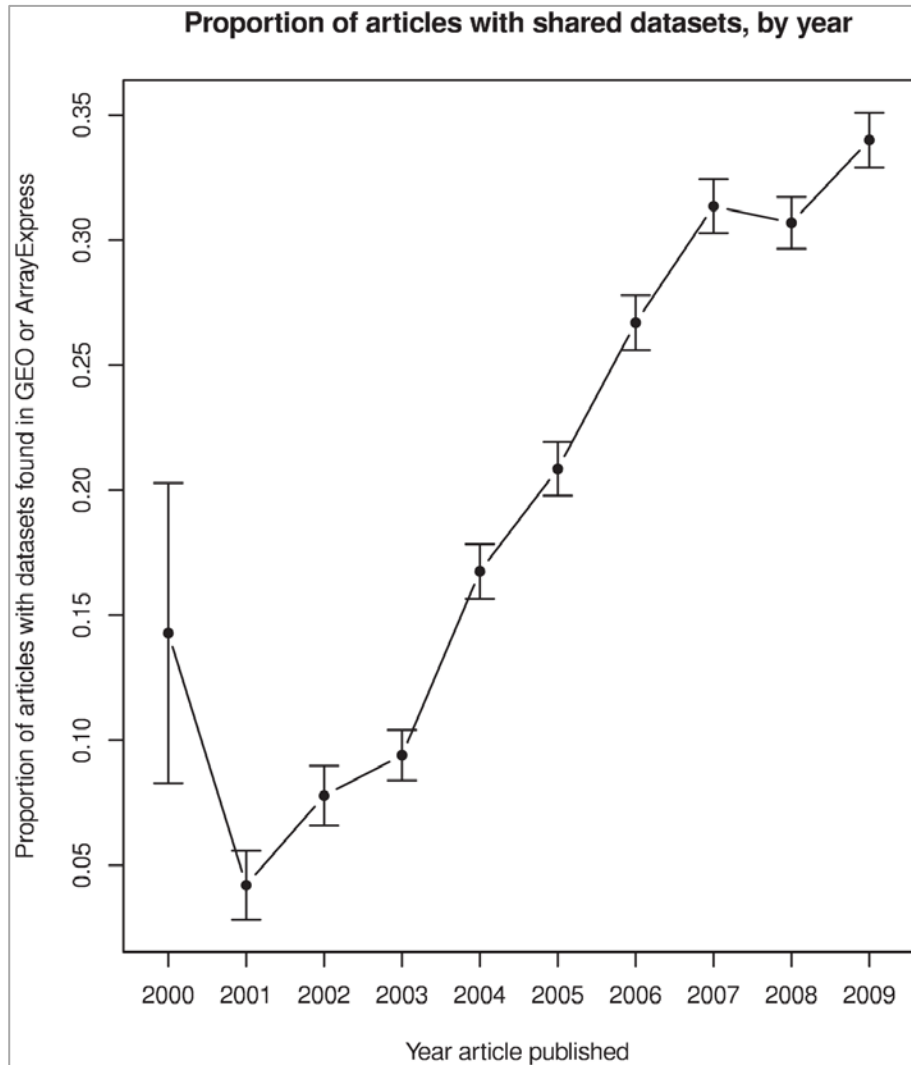
[Jisc DAF studies](#)



# Wellcome OA compliance rates



# Sharing of microarray data



Increase from c.5-35% in under a decade

Best-practice guidelines for sharing microarray data are fairly mature

Two centralized databases have emerged

Unusually strong data sharing requirements in some journals

Piwowar, H. (2011) Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. PLOS One <https://doi.org/10.1371/journal.pone.0018657>





# Data policy changes

Emphasis on data sharing more than RDM

Increasingly 'open' and 'FAIR' rhetoric

## 2002 (handbook)

- General issues relating to data
- Management responsibilities for data within NERC
- Planning for the management of data
- Access to, and charges for, NERC's data
- The implications for scientists holding data

## 2010

- Data acquisition
- Data management
- Access and use
- Charging for Access to NERC's Data

## 2016

- Access to data
- NERC's environmental data centres
- Data collection
- Open access to data underpinning research publications

# Forerunners to FAIR



[OECD Principles and Guidelines for Access to Research Data from Public Funding](#) (2007)

- A. Openness
- B. Flexibility
- C. Transparency
- D. Legal conformity
- E. Protection of IP
- F. Formal responsibility
- G. Professionalism
- H. Interoperability
- I. Quality
- J. Security
- K. Efficiency
- L. Accountability
- M. Sustainability



[Science as an Open Enterprise](#) (2012)

notion of *'intelligent openness'* where data are accessible, intelligible, assessable and useable



*"Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards."*

[G8 Science Ministers Statement](#) (2013)

# Good understanding of FAIR, but...

To what extent do the following statements represent your experience of using the H2020 template?

	Agree		Neither agree nor disagree		Disagree	
	%	n	%	n	%	n
I don't understand what FAIR means	10%	17	16%	28	74%	125

*“We understand the basic principle of FAIR, but the terminology is often difficult to grasp immediately. Things could be explained better in plain language”*

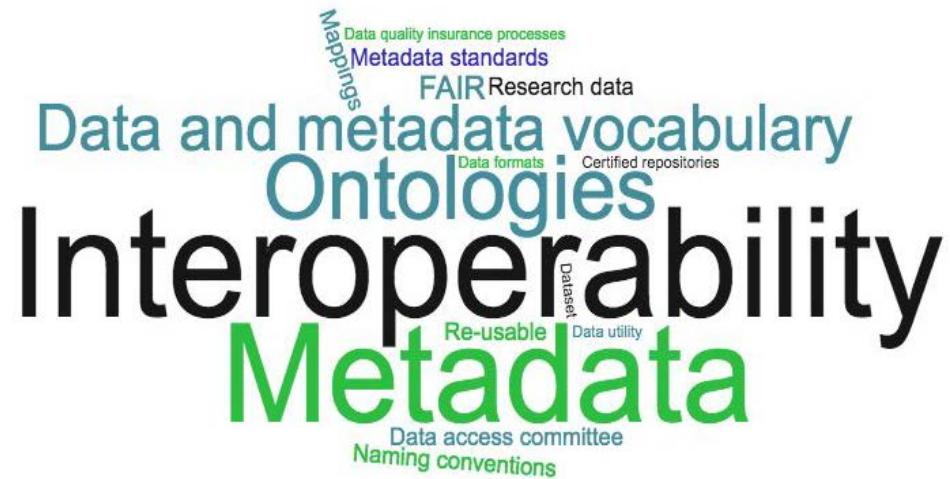
*“The term interoperable is quite confusing sometimes and mixed with re-use.”*

*“I could do with help understanding the section on Making data interoperable as I don't understand a number of the terms and concepts.”*

Table from Q4, comments from Q5

# Language is a barrier

**Respondents mentioned  
40 terms which were  
unclear to them**



*“Researchers are not familiar with the following terms/phrases : Metadata, standards for metadata/data, ontologies, mapping with ontologies, interoperability, ... . All the ICT jargon”*

*“With the help from Swedish National Data Service we could clarify many questions. Without this help we would not be able to finish the DMP.”*

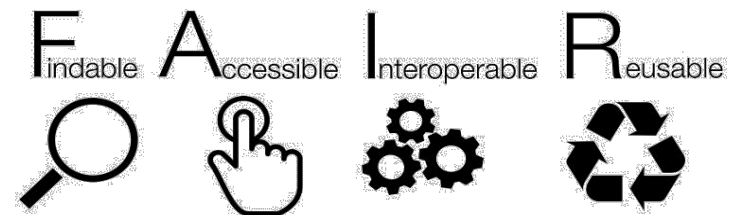
# Conflation of FAIR and open

Making data FAIR ensures it can be found, understood and reused

Data can be shared under restrictions & still be FAIR

Open data is a subset of all the data shared

***"As open as possible, as closed as necessary"***



# Confusion in DMPs

Overly broad definitions of data, including publications, presentations, meeting minutes, dissemination materials, digital photos, project website...

Talk of making data available by gold or green open access

Blurring of methods to store and share data within consortium versus long-term preservation e.g. backup to googledocs, use Dropbox to give public access...

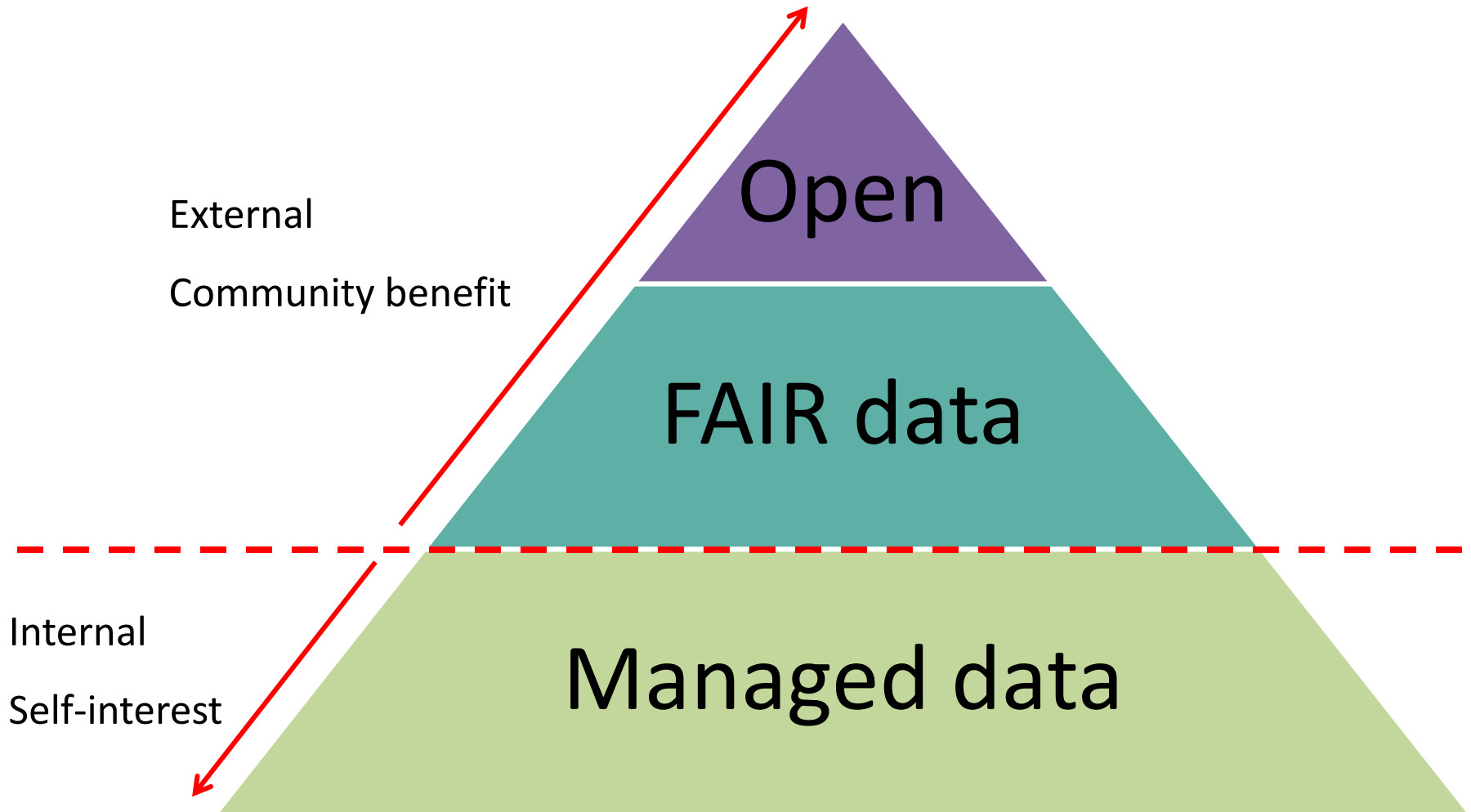
# Be careful what you say...



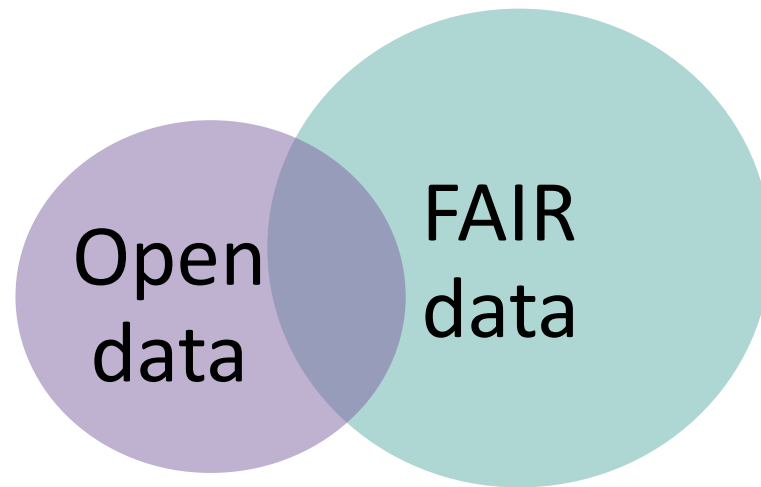




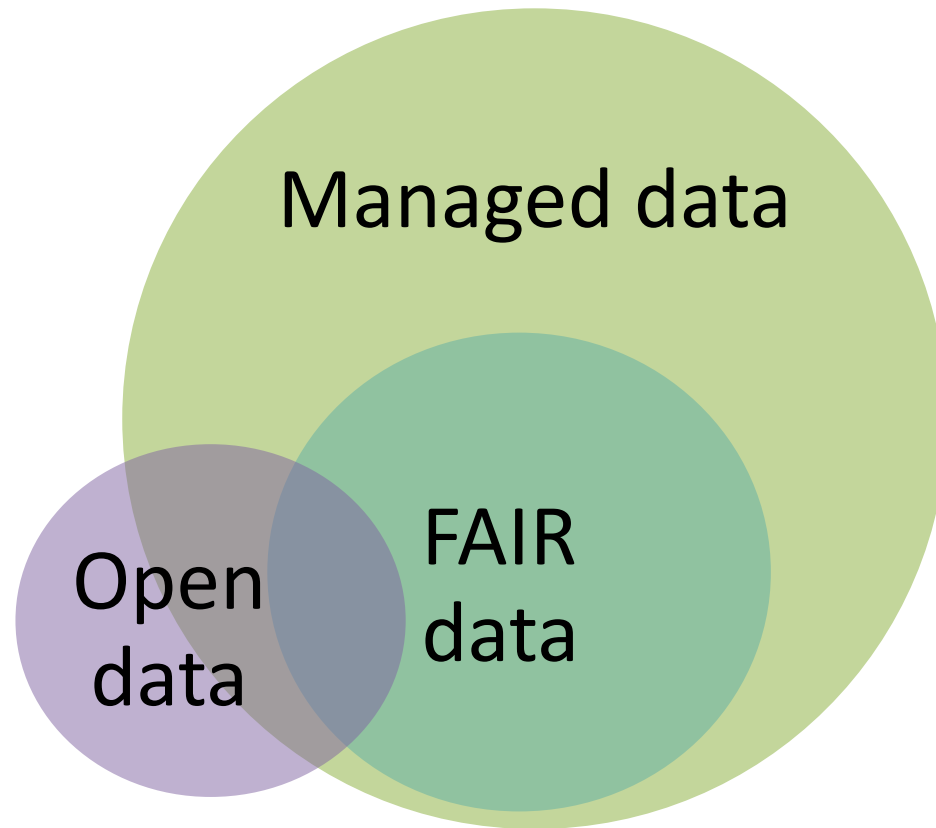
# How do Open, FAIR & RDM intersect?



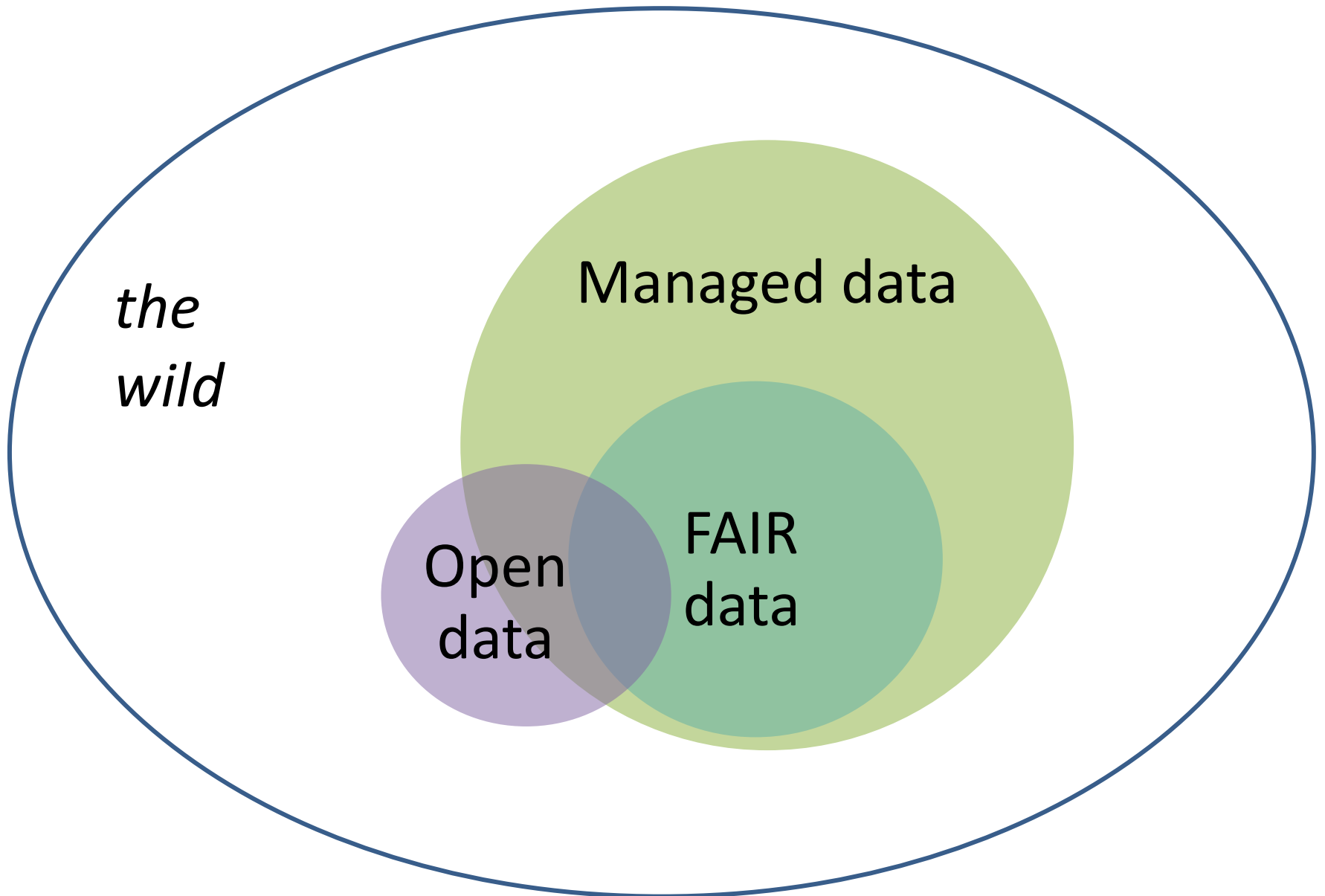
# Open data and FAIR data



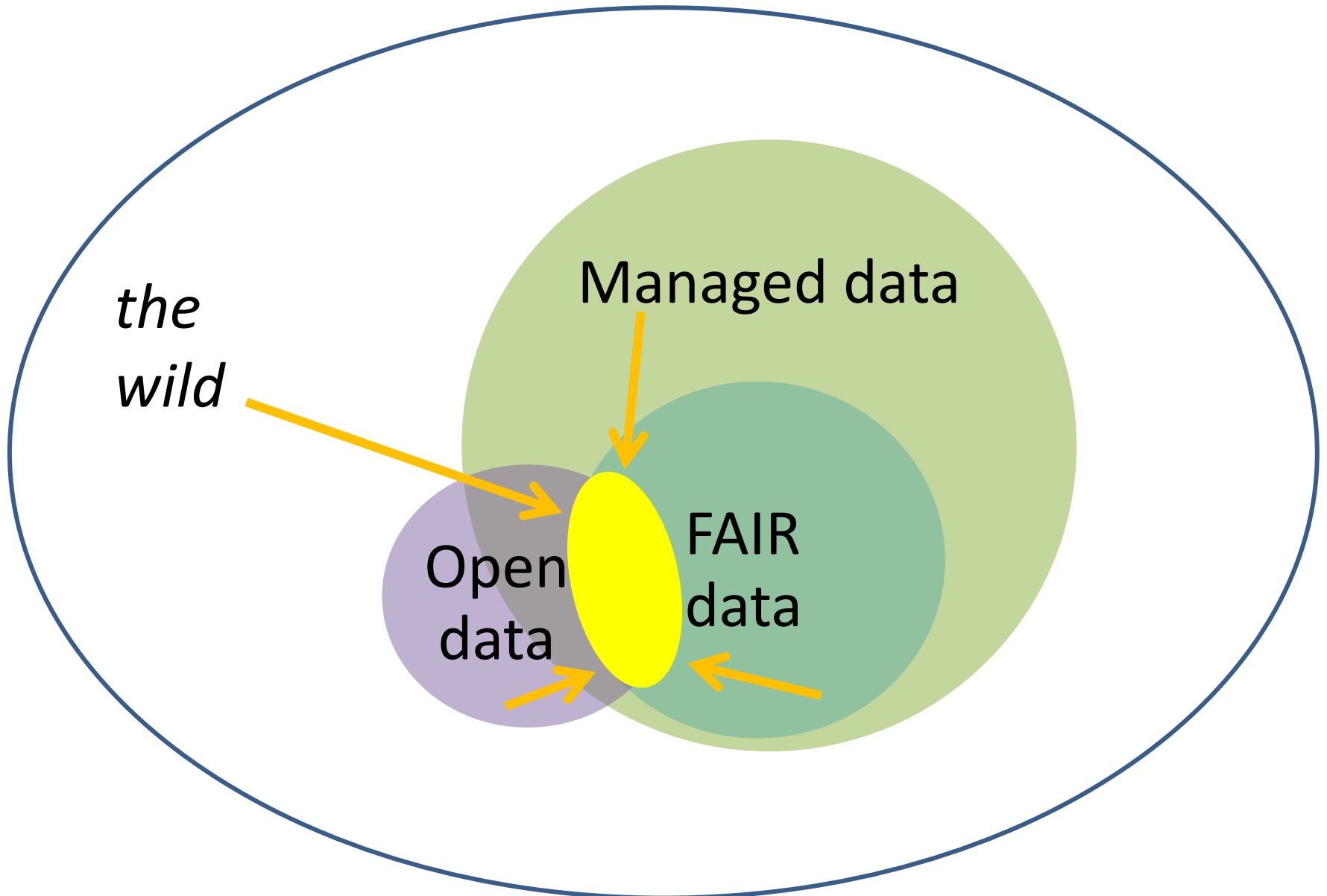
# Open data, FAIR data & RDM

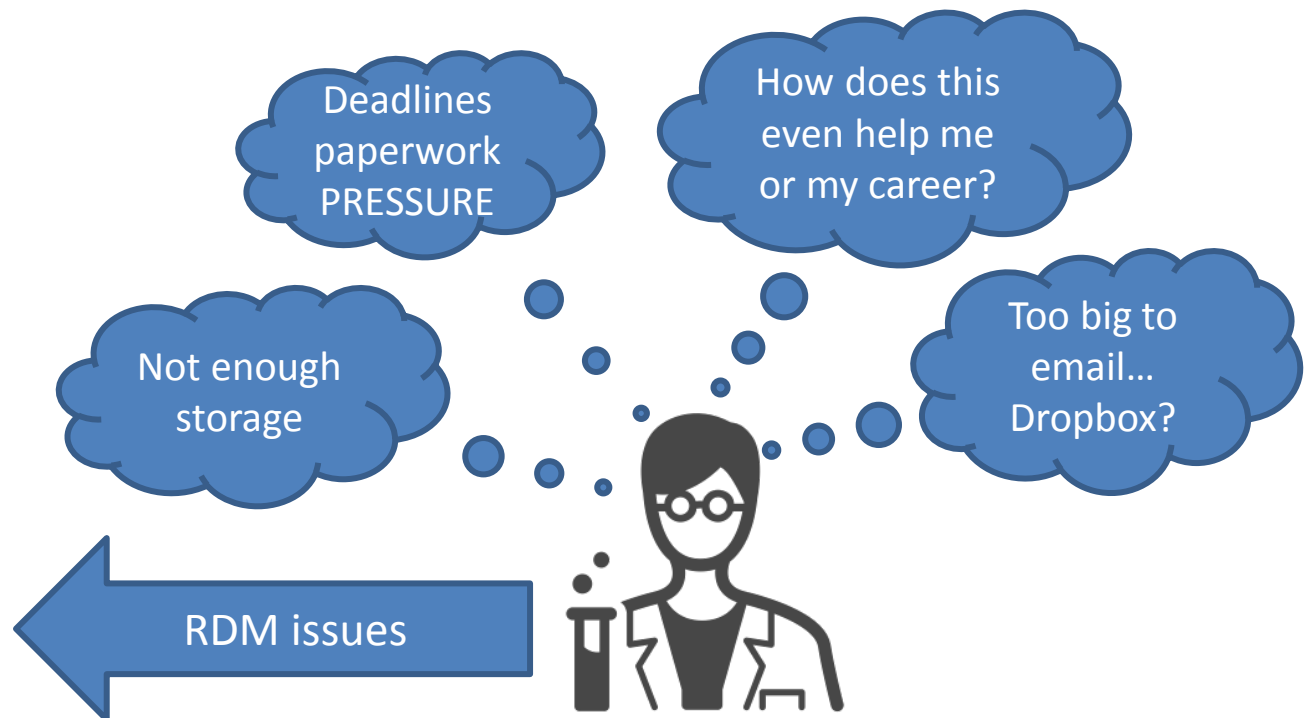
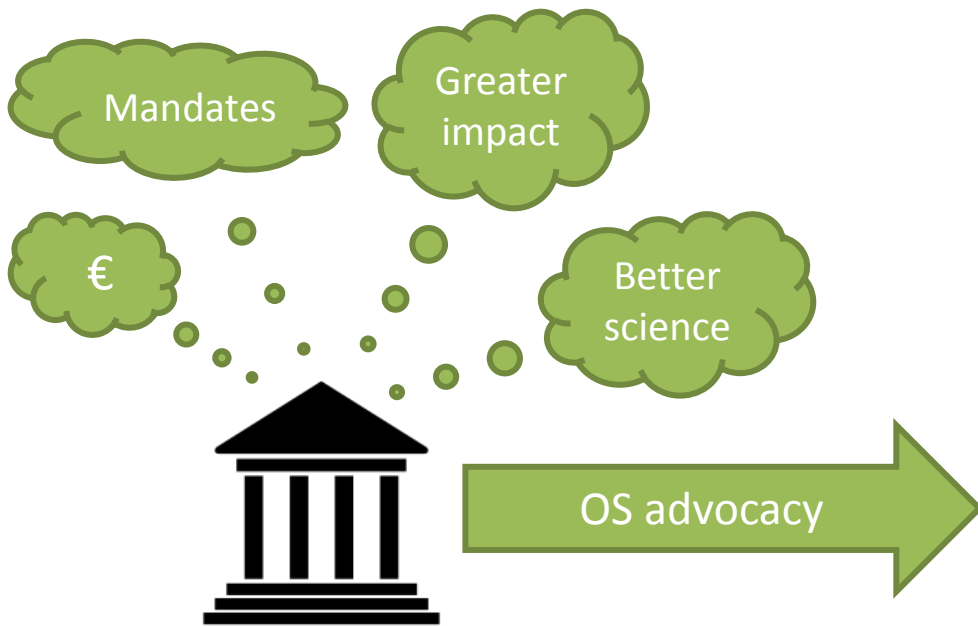


# All research data



# Increasing that which is FAIR & open





# Data engagement programmes

Data conversations at  
Lancaster University

Provide a forum for  
researchers to speak about  
their data

Engage the non-converted

Use peers to spread RDM /  
OS message

[www.lancaster.ac.uk/library/  
rdm/data-conversations](http://www.lancaster.ac.uk/library/rdm/data-conversations)



Data champions at Cambridge  
& data stewards at TU Delft

Local support & help

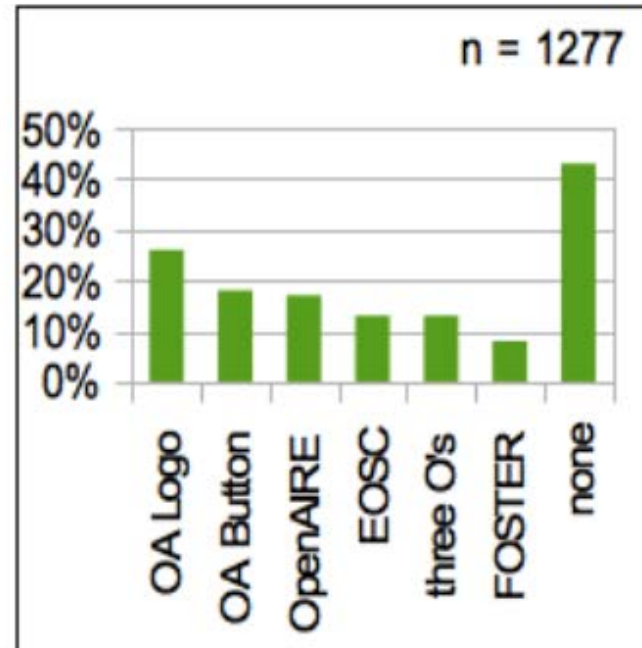
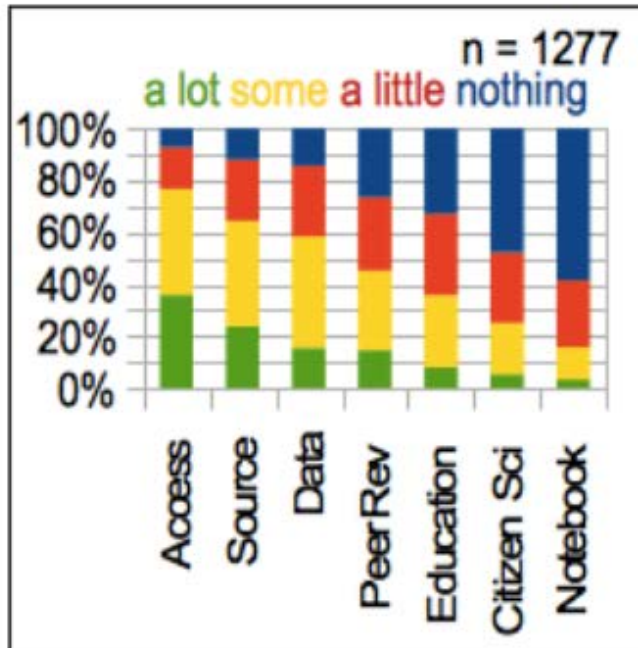
Researcher-focused & led

Explicitly recognise value and  
role of curation

[https://osc.cam.ac.uk/engaging  
-researchers-good-data-  
management](https://osc.cam.ac.uk/engaging-researchers-good-data-management)



# Awareness of OS & initiatives



European Commission (OSPP) Open Science Policy Platform. (2017) *Providing researchers with the skills and competencies they need to practise Open Science*. Report of the Working Group on Education and Skills under Open Science, doi: 10.2777/121253

# ORD pilot & FAIR data

## ORD Pilot

- Introduction of an Open Research Data pilot in 2014
- Expansion from 7 to 9 work areas in 2016
- ‘Open data by default’ since 2017. Need to actively opt out.

## FAIR data

- New ‘FAIR Data Management guidelines’ in July 2016
- Increasing emphasis on data management as well as sharing
- Mantra of “As open as possible, as closed as necessary”

## ?

- *Formal policy in FP9...*
- *Even greater emphasis on research data management...*
- *Mandatory DMP, even in cases of opt out...*

[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

# Manage data so it can flourish



# Thanks for listening

DCC resources on Data Management

[www.dcc.ac.uk/resources](http://www.dcc.ac.uk/resources)

Follow us on twitter:

@digitalcuration and #ukdcc



D|C|C

because good research needs good data