

Data Sharing

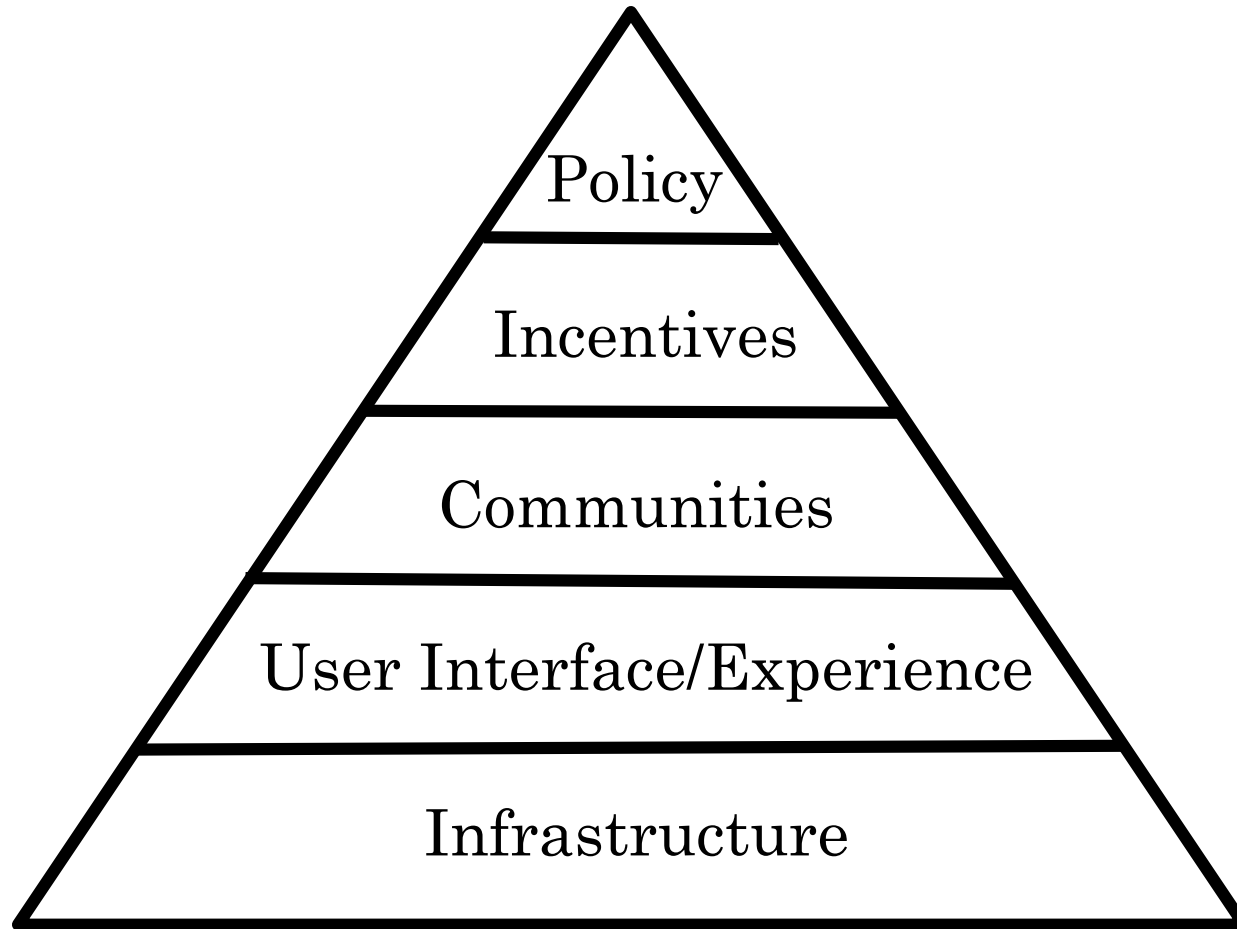
principle > practise > progress

Hollydawn Murray, PhD

Head of Data and Software Publishing, F1000 Research

@ilex_meunier

Changing Research Culture



Make it required

Make it rewarding

Make it normative

Make it easy

Make it possible

2013

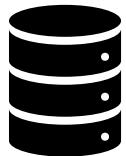


Thrift Shop

Macklemore & Ryan Lewis featuring Wanz



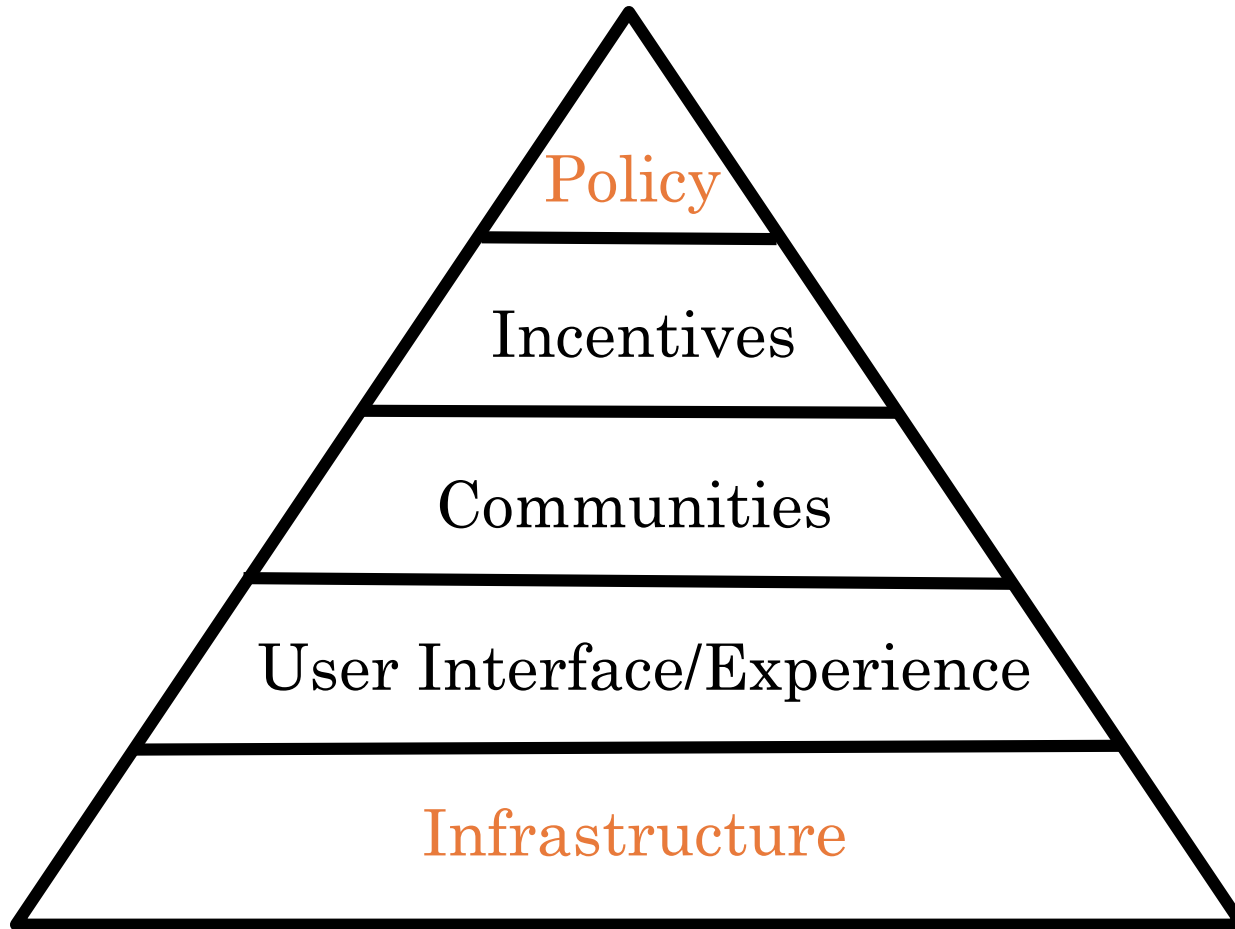
James E. Rothman, Randy W. Schekman and Thomas C. Südhof
for their discoveries of machinery regulating vesicle traffic, a major
transport system in our cells



Research Data Alliance

the first plenary meeting and official launch held in Gothenburg, Sweden

2013



Make it required

Make it rewarding

Make it normative

Make it easy

Make it possible

2013

Policy




- Mandatory data sharing, via a repository with an identifier under CC0 license
- Exceptions for ethical, security, and data protection reasons
- List of approved repositories
- Guidance on preparing data
- No “data available upon request” option

Infrastructure

- F1000Research staff broker between authors and generalist repositories
- Figshare and Dataverse visualisation widgets in article
- Data Notes

Data File 2: raw transcript read counts			571 views	0 shares	126 downloads
33	Atp5b	NM_016774	NM_016774		
34	Rpl13a	NM_009438;BC065799	NM_009438		
35	Ldha	NM_010699;AK171373;BC094019	NM_010699		
36	Rps6	NM_009096	NM_009096		
37	Hnrpa3	BC085474;NM_053263;BC116251;NM_146130;NM_198090	NM_053263;NM_1		
38	Rpl7	NM_011291	NM_011291		
39	Rps3a	NM_016959	NM_016959		
40	Eef1g	NM_026007	NM_026007		
41	Pol	AK149078			
42	Trim28	NM_011588	NM_011588		
43	Rpl7	NM_011291	NM_011291		

Sheet1 Sheet2 Sheet3

figshare  Share  Cite  Download

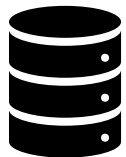
2014



Happy
Pharrell Williams

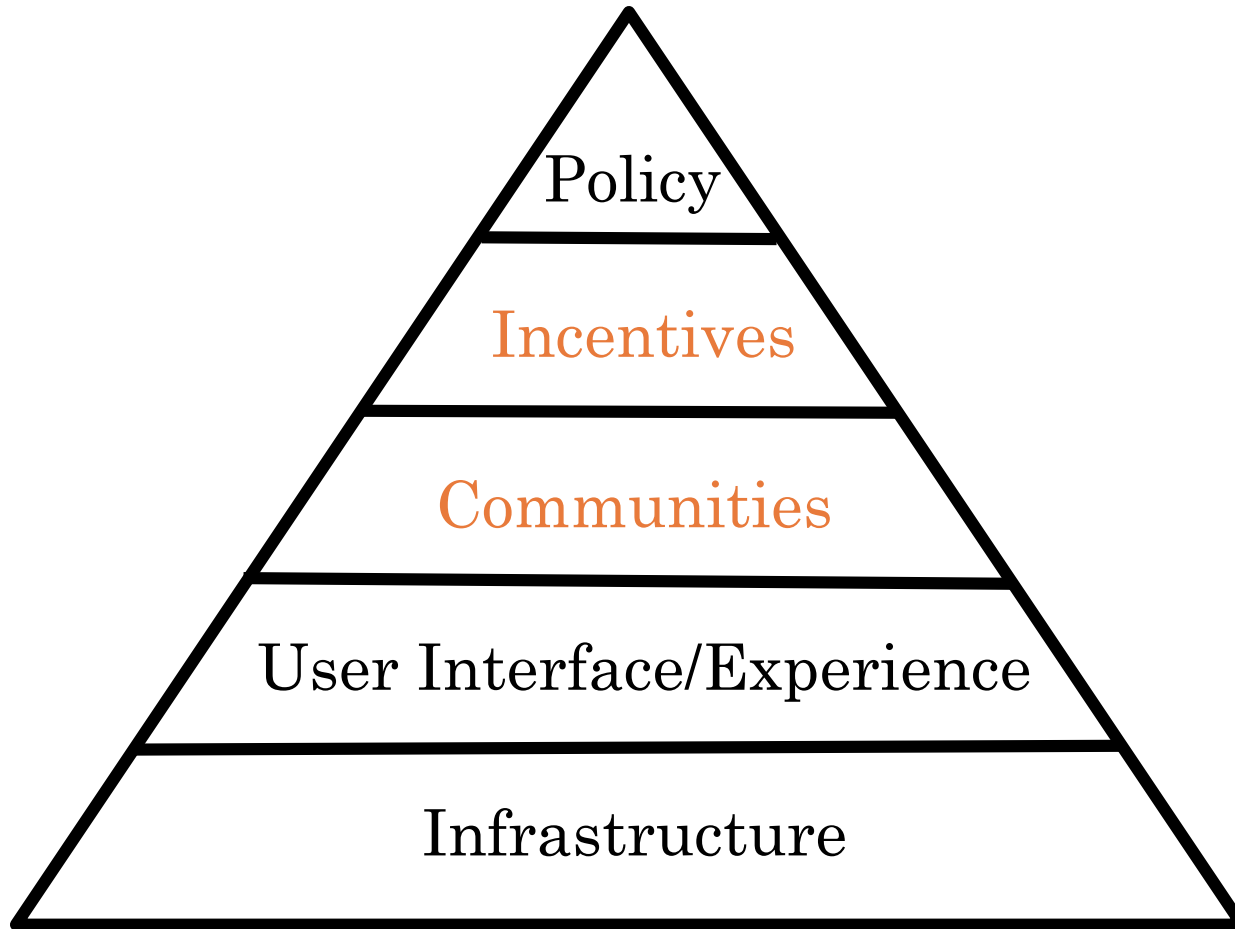


John O'Keefe, May-Britt Moser and Edvard I. Moser
for their discoveries of cells that constitute a positioning system in the brain



FORCE11 Data Citation Principles
like the citation of other sources, data citation is good research practice and is part of the scholarly ecosystem supporting data reuse

2014



Make it required

Make it rewarding

Make it normative

Make it easy

Make it possible

2014

- Adopted FORCE11 Data Citation Principles
- Data availability statements, where data was presented
- Use of ‘own brand’ repository

Data availability

F1000Research: Dataset 1. Tooth size-arch length discrepancies and teeth eruption in children with and without dental crowding, [10.5256/f1000research.3196.d27729](https://doi.org/10.5256/f1000research.3196.d27729)²⁴

Data availability

F1000Research: Dataset 1. *Drosophila* lifespan and fecundity data set, [10.5256/f1000research.3975.d28311](https://doi.org/10.5256/f1000research.3975.d28311)⁴⁷

2015



Uptown Funk!

Mark Ronson Featuring Bruno Mars

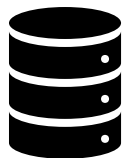


William C. Campbell and Satoshi Ōmura

for their discoveries concerning a novel therapy against infections caused by roundworm parasites

Tu Youyou

for her discoveries concerning a novel therapy against Malaria

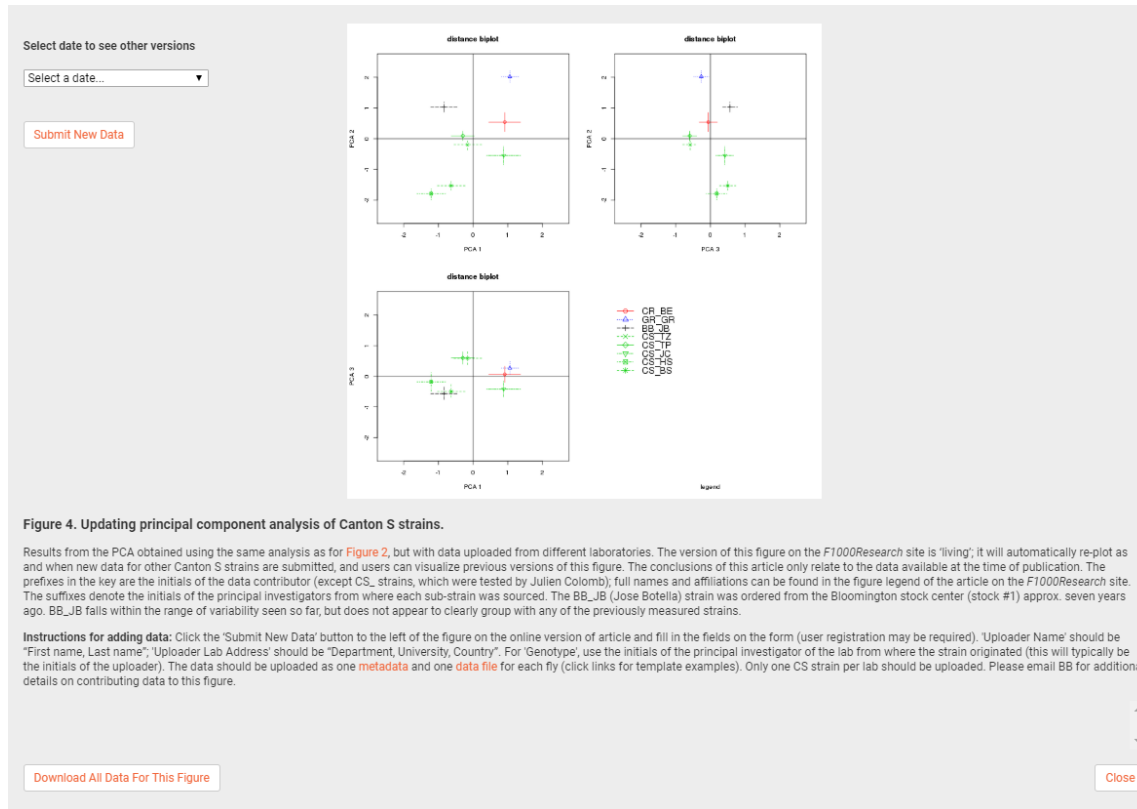


“The Year of Open Data Mandates”

an increase in policies referencing data sharing including ESPRC, NSF, NIH

2015

- Living figure (a PCA that automatically updates when data added)
- No in-house dedicated ‘data roles’



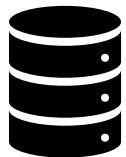
2016



Love Yourself
Justin Bieber



Yoshinori Ohsumi
for his discoveries of mechanisms for autophagy

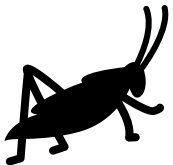


The FAIR Guiding Principles for scientific data management
the first formal publication of the FAIR Principles

Statement on data sharing in public health emergencies
call for all research data gathered during public health emergencies, to be made available as rapidly and openly as possible

2016

- The metadata of each the dataset and article DOI submissions references the related article or dataset, in order to tie authors, their datasets, and articles together



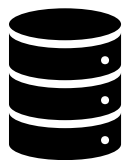
2017



Shape Of You
Ed Sheeran



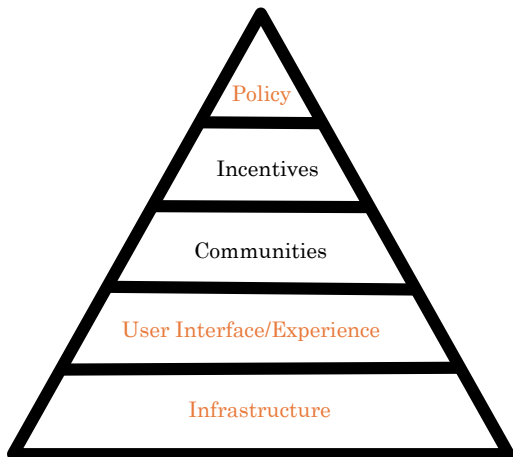
Jeffrey C. Hall, Michael Rosbash and Michael W. Young
for their discoveries of molecular mechanisms controlling the circadian rhythm



ICMJE set out minimum data sharing requirements
specifically for clinical trials, manuscripts must include a data sharing statement as of 1 July 2018 and a data sharing plan for studies beginning after 1 January 2019

2017

- Building data back into objectives
- Mandatory structured peer review questions
- Mandatory data availability statements in ALL articles
- 3rd party data exception added
- Code Ocean widgets and iFrames



Make it required

Make it rewarding

Make it normative

Make it easy

Make it possible

CODE OCEAN BETA F1000Research Env Run

Output - Run 7442850 Results

```

Run environment setup complete, running code...

[1] "Performing analysis of FPKM matrix"
Read 26 records
Performing PCA on log-transformed FPKM matrix
Importance of components:

```

	PC1	PC2	PC3	PC4	PC5
Standard deviation	54.2999	43.5061	39.3992	35.81511	31.77
Proportion of Variance	0.2012	0.1292	0.1059	0.08755	0.06
Cumulative Proportion	0.2012	0.3304	0.4364	0.52391	0.59

	PC7	PC8	PC9	PC10
Standard deviation	24.77243	22.33937	21.69784	21.1730
Proportion of Variance	0.04188	0.03406	0.03213	0.0306
Cumulative Proportion	0.68037	0.71443	0.74656	0.7772

	PC13	PC14	PC15	PC16
Standard deviation	17.41969	17.00162	16.05240	15.65210
Proportion of Variance	0.02071	0.01973	0.01759	0.01672
Cumulative Proportion	0.84751	0.86724	0.88483	0.90155

	PC19	PC20	PC21	PC22
Standard deviation	13.08324	12.96681	12.29599	12.07214
Proportion of Variance	0.01168	0.01148	0.01032	0.00995
Cumulative Proportion	0.94340	0.95487	0.96519	0.97514

	PC25	PC26
Standard deviation	10.47129	2.442e-14
Proportion of Variance	0.00748	0.000e+00
Cumulative Proportion	1.00000	1.000e+00

```

[1] "Testing for significance of Spearman correlation between
Spearman's rank correlation rho

data: plotData$PC1[humanIdx] and plotData$PC1[mouseIdx]
S = 134, p-value = 0.02374
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.6318681

```

Results

Search

Run Time: 0h 00m 36s | Sep 20, 2018 | 12:27

Running...

Output

Run Time: 0h 01m 45s | Sep 20, 2018 | 12:26

Output 14.53 KB

- batchCo... 49.3 KB
- batchC... 50.52 KB
- batchC... 57.82 KB
- batchC... 50.48 KB
- batchC... 57.24 KB
- batchC... 58.31 KB
- batchC... 49.52 KB
- batchC... 48.57 KB
- batchC... 55.91 KB
- batchC... 50.66 KB
- batchC... 59.88 KB

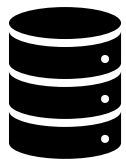
2018



God's Plan
Drake



James P. Allison and Tasuku Honjo
for their discovery of cancer therapy by inhibition of negative immune regulation

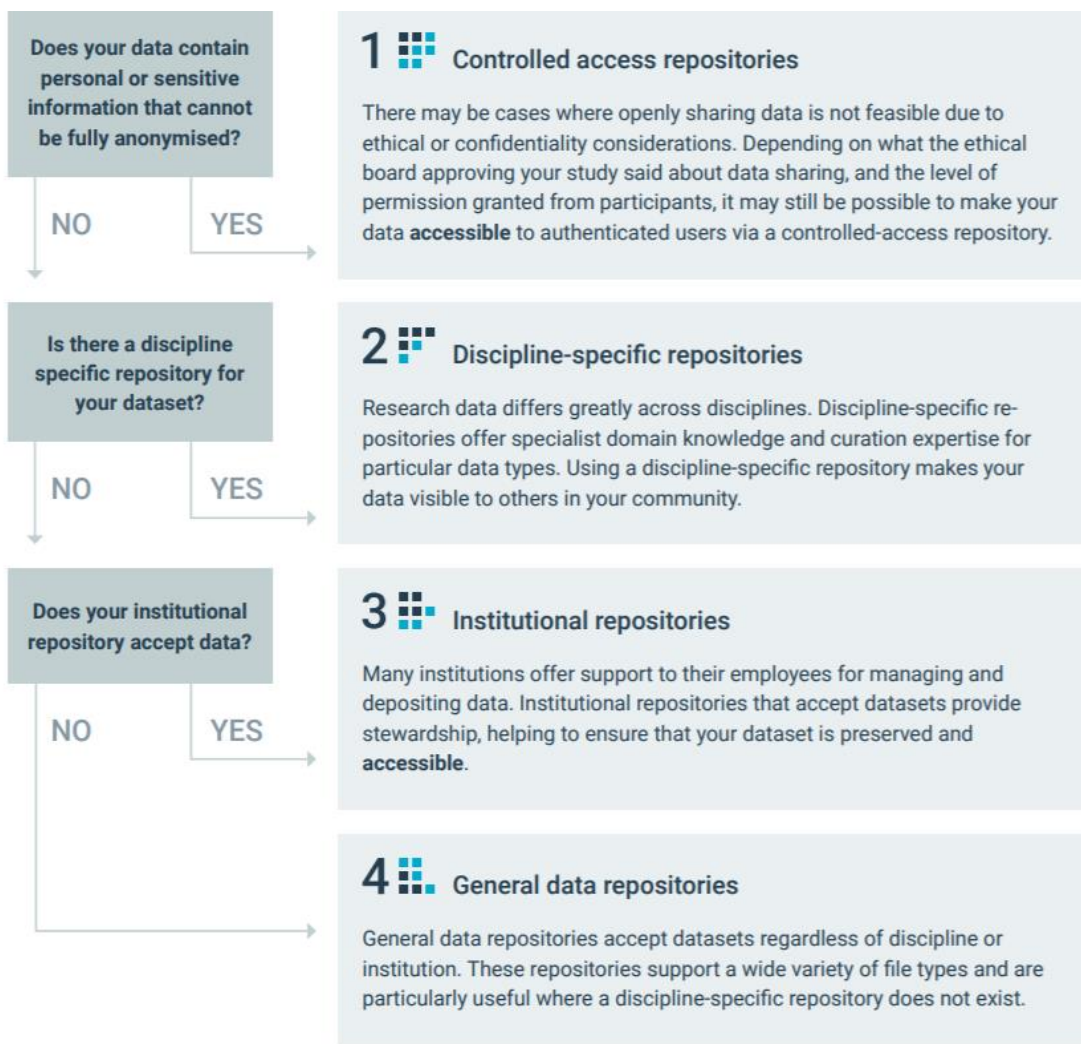


EOSC
officially launched by the Austrian Presidency of the EU

Google Dataset Search
beta release, indexing datasets marked up with schema.org

2018

- Updated guidance to reflect FAIR principles, and be more prescriptive
- Removal of supplementary material
- Additional training materials developed in response to our experiences and barriers to data sharing identified in Figshare's State of Open Data Report
- Springer Nature data services pilot on Wellcome Open Research



2019

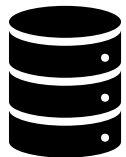


Old Town Road

Lil Nas X Featuring Billy Ray Cyrus



William G. Kaelin Jr, Sir Peter J. Ratcliffe and Gregg L. Semenza
for their discoveries of how cells sense and adapt to oxygen availability

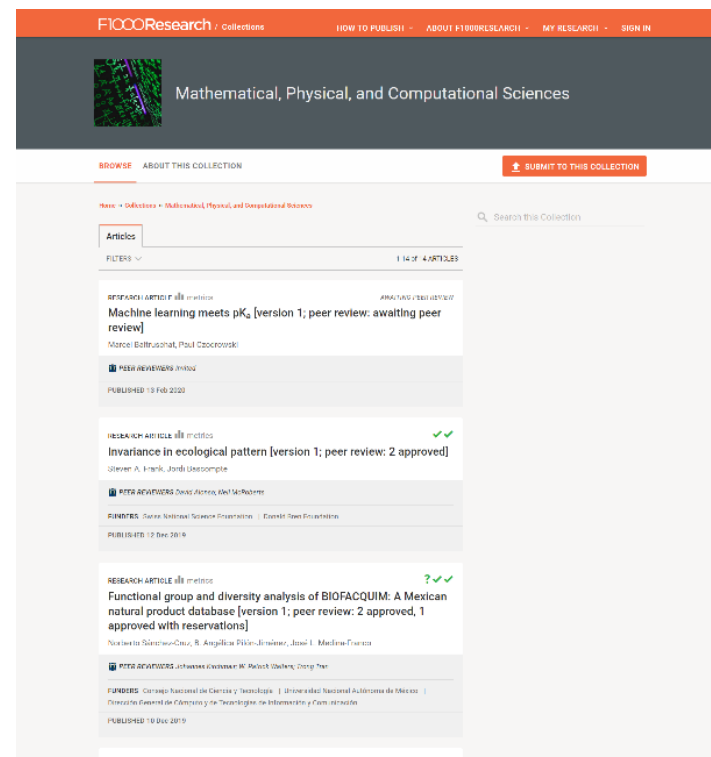
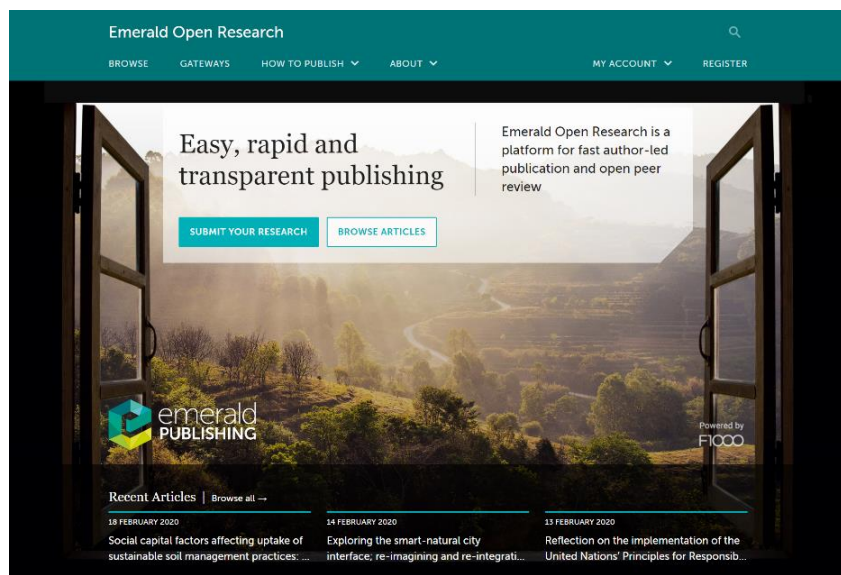


Cost-benefit analysis for FAIR research data

estimates cost of not having FAIR research data at 10.2 billion

2019

- Editorial Data Lead
- Expanded scope to humanities and social sciences and mathematical, physical and computational sciences



Now



Reaching consensus

- FAIRsharing repository criteria
- Data policy framework



Scaling up

- DataSeer Pilot
- Discipline-specific policies



Building a database for data

- Submission form > DA statement
- Record and track



Researcher support

- Formalising
- RDM, specifically DMPs

Lessons learned



BE FLEXIBLE



**CREATE
CAPACITY**



**RUBBER AND
GLUE**

Beyond data

Preprints

Preregistration

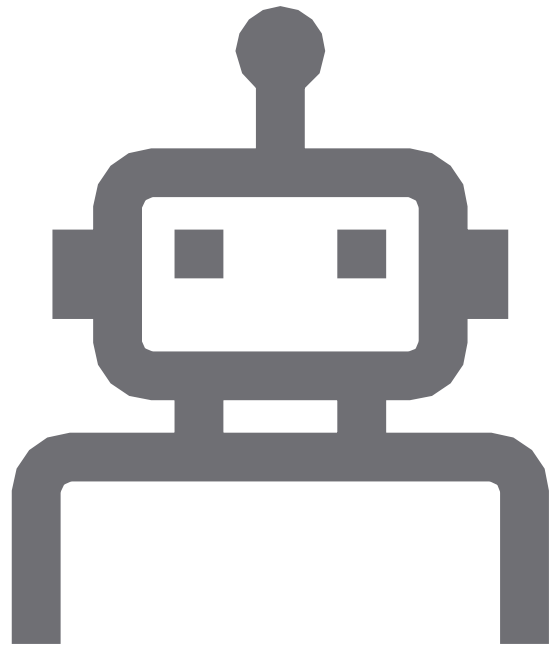
Registered Reports

Open peer review

Replication studies, null results

Detailed methods, analyses, and design

Research material sharing



Don't forget
about software