# FROM CLAY TO BRICKS: DATA SHARING, CITATIONS, & TRANSPARENCY

ISMTE

Baltimore | 08.02.2018

Jen Mavzer, Coronis Group

# Data Acceleration Regime

Technology enables production of pattern-based data

Continuous production of standardized data streams

Massive sets of networked/organized data

New data-driven approaches

Kaplan and di Lenardo.
*Front Digit Humanit* (2017)

"Data! data! data!" he cried impatiently.
"I can't make bricks without clay."

Arthur Conan Doyle
*The Adventure of the Copper Beeches* (1892)

Science is built of facts, as a house is with bricks. But a collection of facts is no more a science than a heap of bricks is a house.

Henri Poincaré
*Science and Hypothesis* (1901)
"Hypotheses in Physics"

Post Bulletin (2014)

## Chaos in the Brickyard

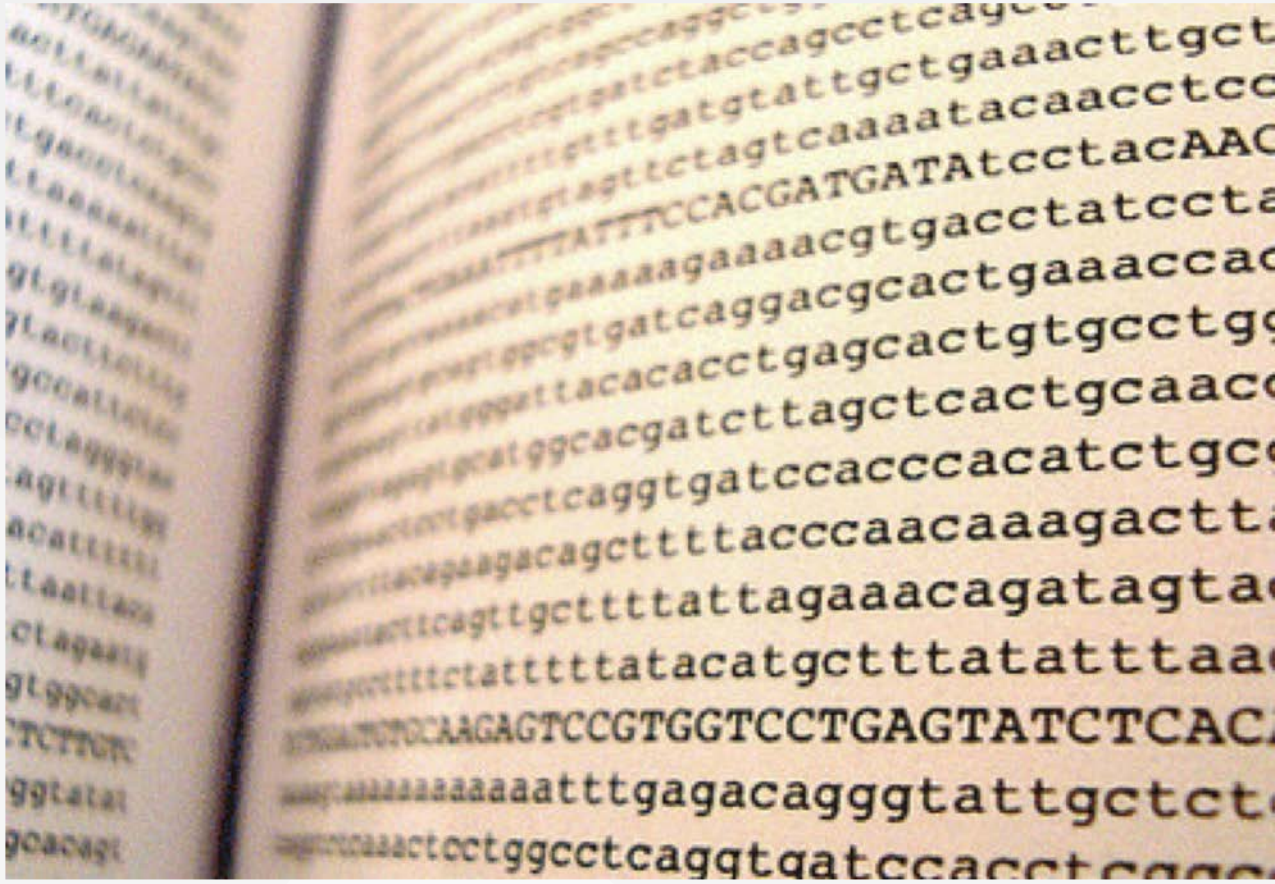
Once upon a time, among the activities and occupations of man there was an activity called scientific research and the performers of this activity were called scientists. In reality, however, these men were builders who constructed edifices, called explanations or laws, by assembling bricks, called facts.

Forscher BK. *Science*. 1963

# Poverty of physical metaphors

"the new gold"
"the new oil"
"the new *crude* oil."
"the new bacon"

neither a commodity, nor finite resource…




Source: Rolf Viervant
http://www.viervant.nl/the-size-of-the-genome/

# Roadmap

**Defining data**
What is it? How is data related to reproducibility and transparency?

**Data sharing**
How is it done? What are the pros and cons? What's the evidence?

**Data policies**
Who is mandating open or shared data? Key principles? Standards?

**Data publishing & citation**
How are others dealing with data? What are best practices?

# DEFINING DATA

"…a scientific publication is not the scholarship itself, it is merely advertising of the scholarship…"

Buckheit J, Donoho DL
*Wavelets and Statistics* (1995)

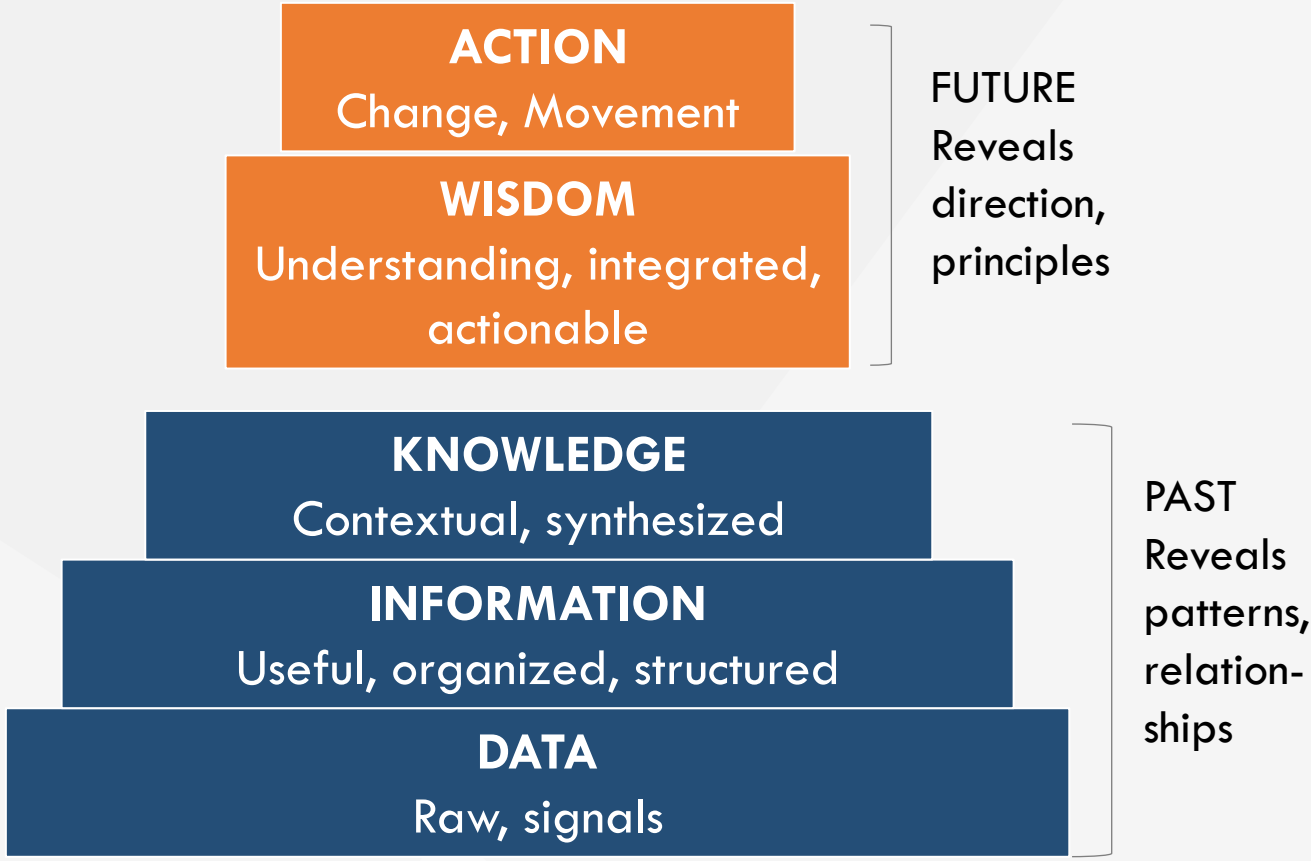
# data

ˈdadə,ˈdādə/
(Latin: *datum* "(thing) given or granted"
cf. Greek: *dedomenon*)

1. Facts and statistics collected together for reference or analysis.

2. The quantities, characters, or symbols on which operations are performed by a computer, being stored and transmitted in the form of electrical signals.

3. Things known or assumed as facts, making the basis of reasoning or calculation.

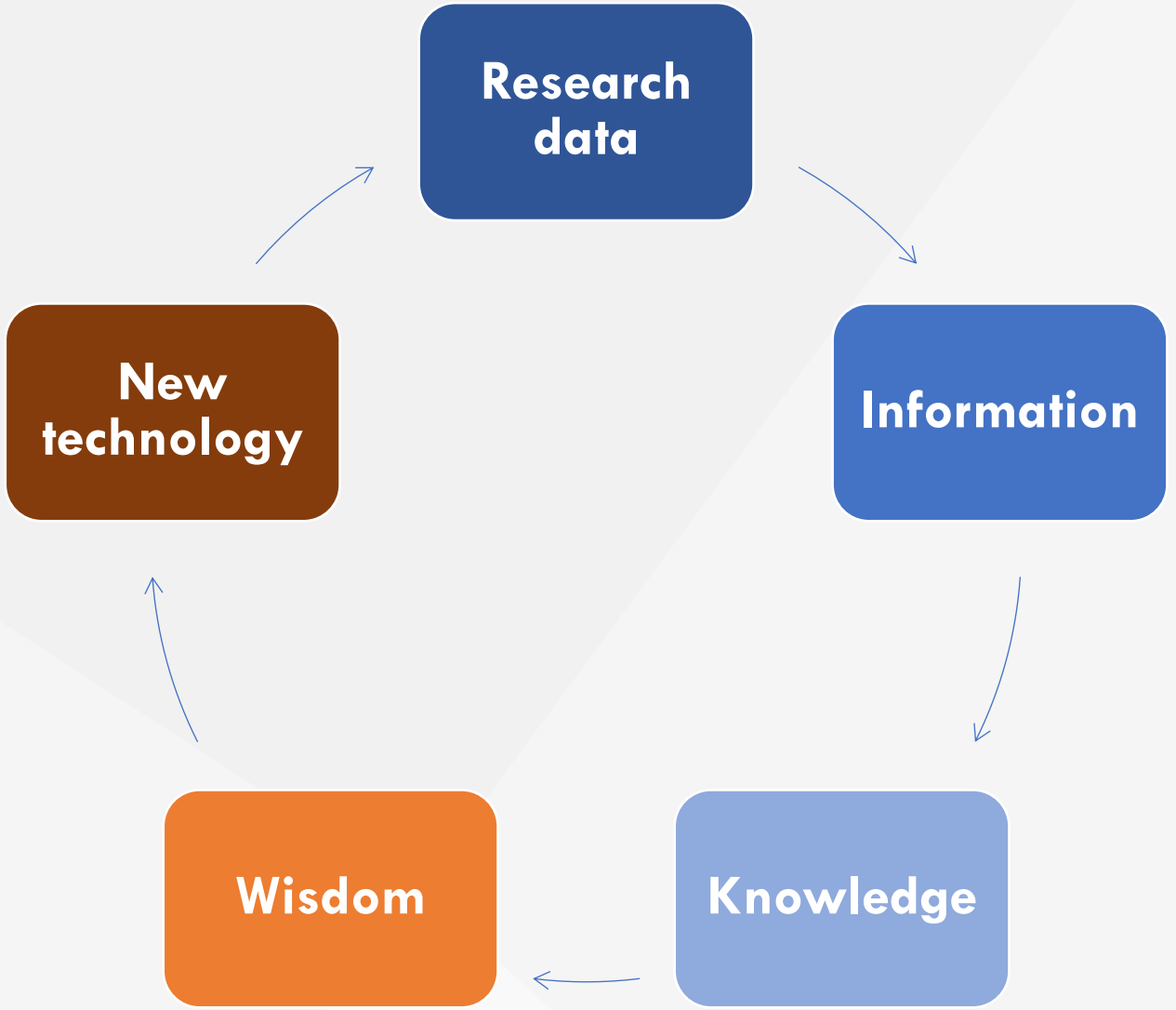



Ackoff R. *J Appl System Analysis* (1989)
Hey J. Intergovernmental Oceanographic Commission (2004)
AGT International., via *IEEE GlobalSpec* (2015)

# Research data

**OECD definition:**

Factual records—particularly those which may be **digital and machine readable**—used as primary sources for scientific research that are necessary to validate research findings.




OECD Principles and Guidelines (2007)

# Open research data

"Data that can be freely used, re-used and redistributed by anyone — subject only, at most, to the requirement to attribute and sharealike."

Entails legal (i.e. licensing) and technical (machine-readability) requirements, i.e.:

- **Availability and Access:** data should be available, easy to access, preferably in modifiable format.
- **Re-use and Redistribution:** provided under licensing terms that permit reuse, redistribution, and recombination
- **Universal Participation:** everyone must be able to use, re-use and redistribute.

Open Knowledge International, Open Data Handbook (2018)

**Open data is like a renewable energy source: it can be reused without diminishing its original value, and reuse creates new value.**

Jean-Claude Burgelman
European Commission

Figshare (2017)

# Data, reproducibility, and transparency

Transparency requires availability of both null and statistically significant results to allow others to accurately assess evidence.

BBC News

## What is reproducibility?

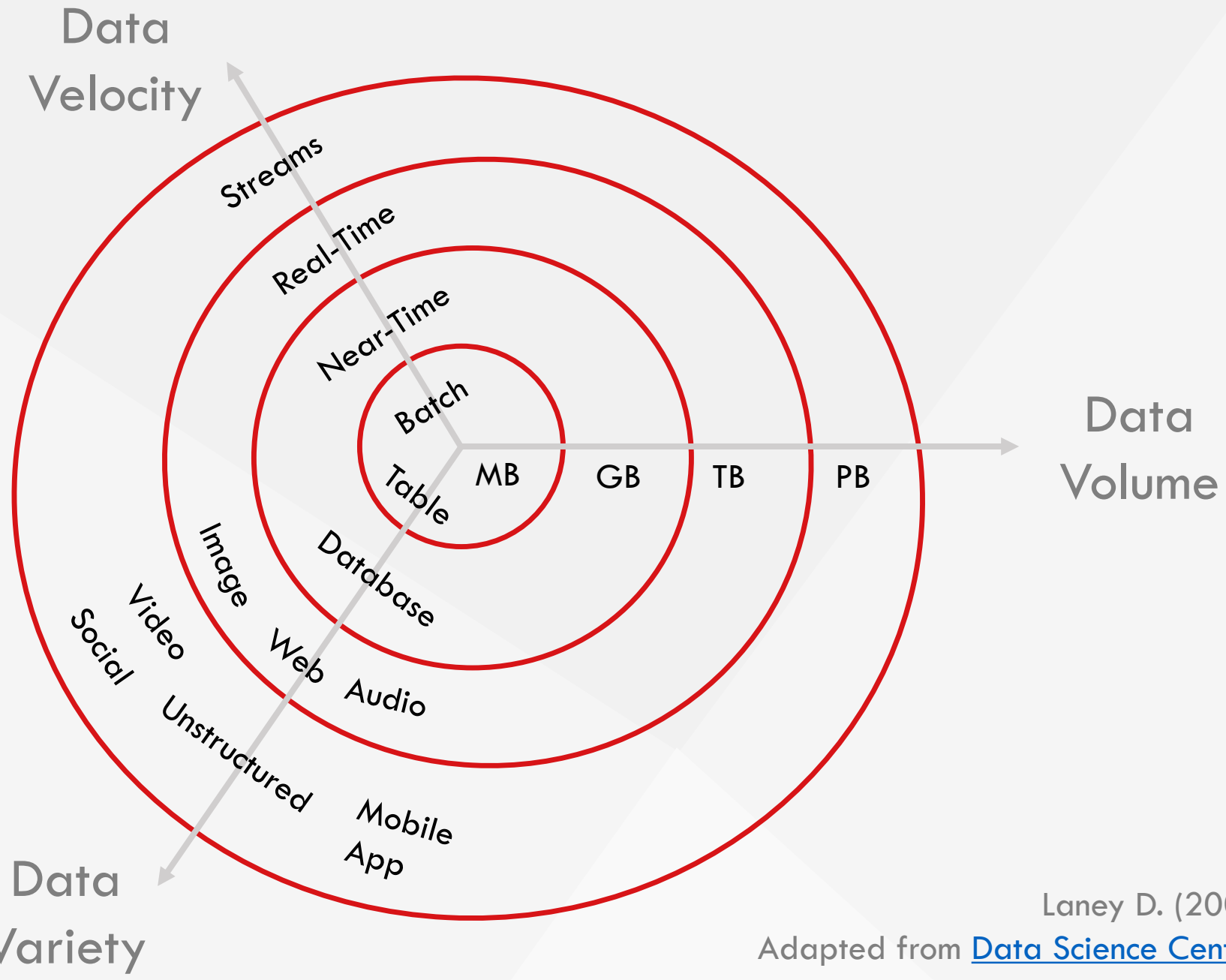
1. **Methods reproducibility**—sufficient detail given to enable a study to be repeated
2. **Results reproducibility** (i.e. replicability, "robustness")—findings can be repeated by others.
3. **Inferential reproducibility**—similar conclusions are drawn about results.

Nosek et al. *Science* (2015)
Goodman et al. *Sci Transl Med* (2016)

# Big data

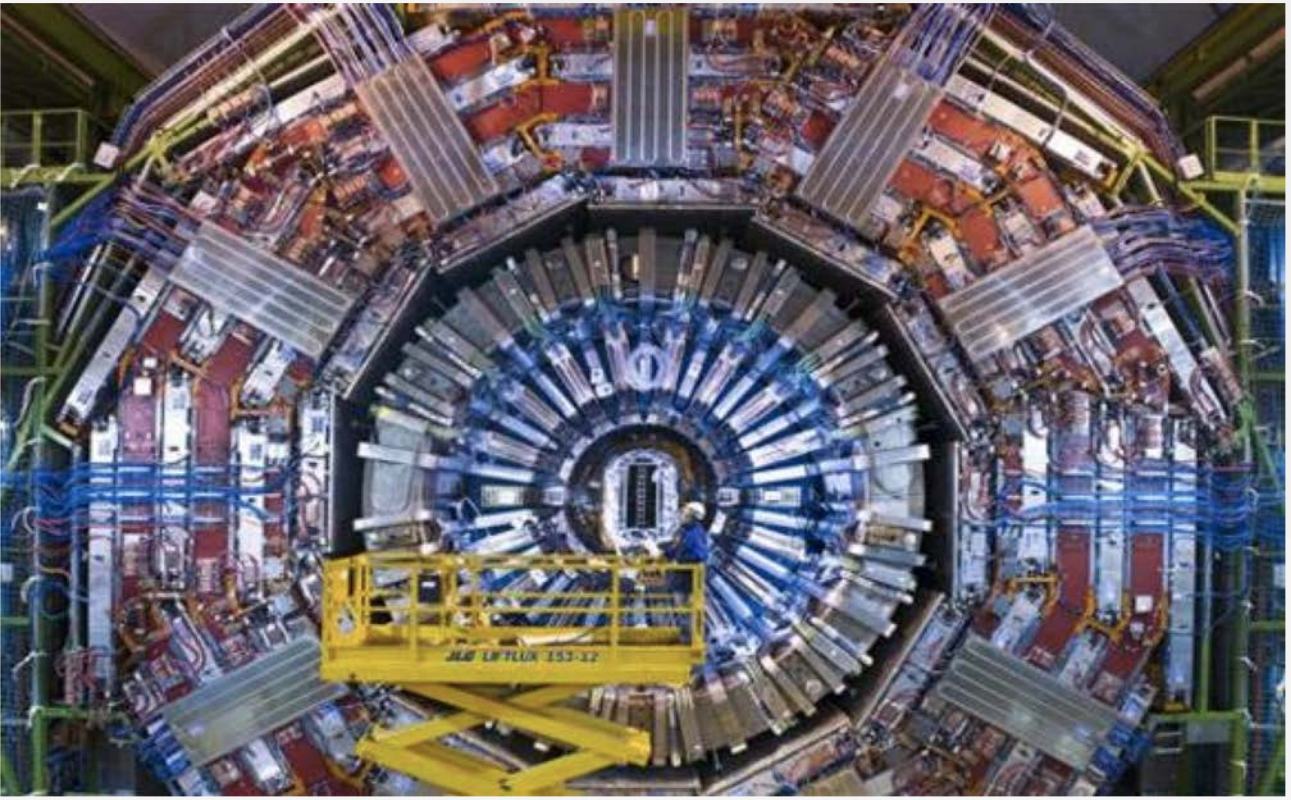
**"Expanding on three fronts at an increasing pace…"**

Fourth V = Veracity (i.e. uncertainty of data)




Laney D. (2001)
Adapted from Data Science Central

# Big questions

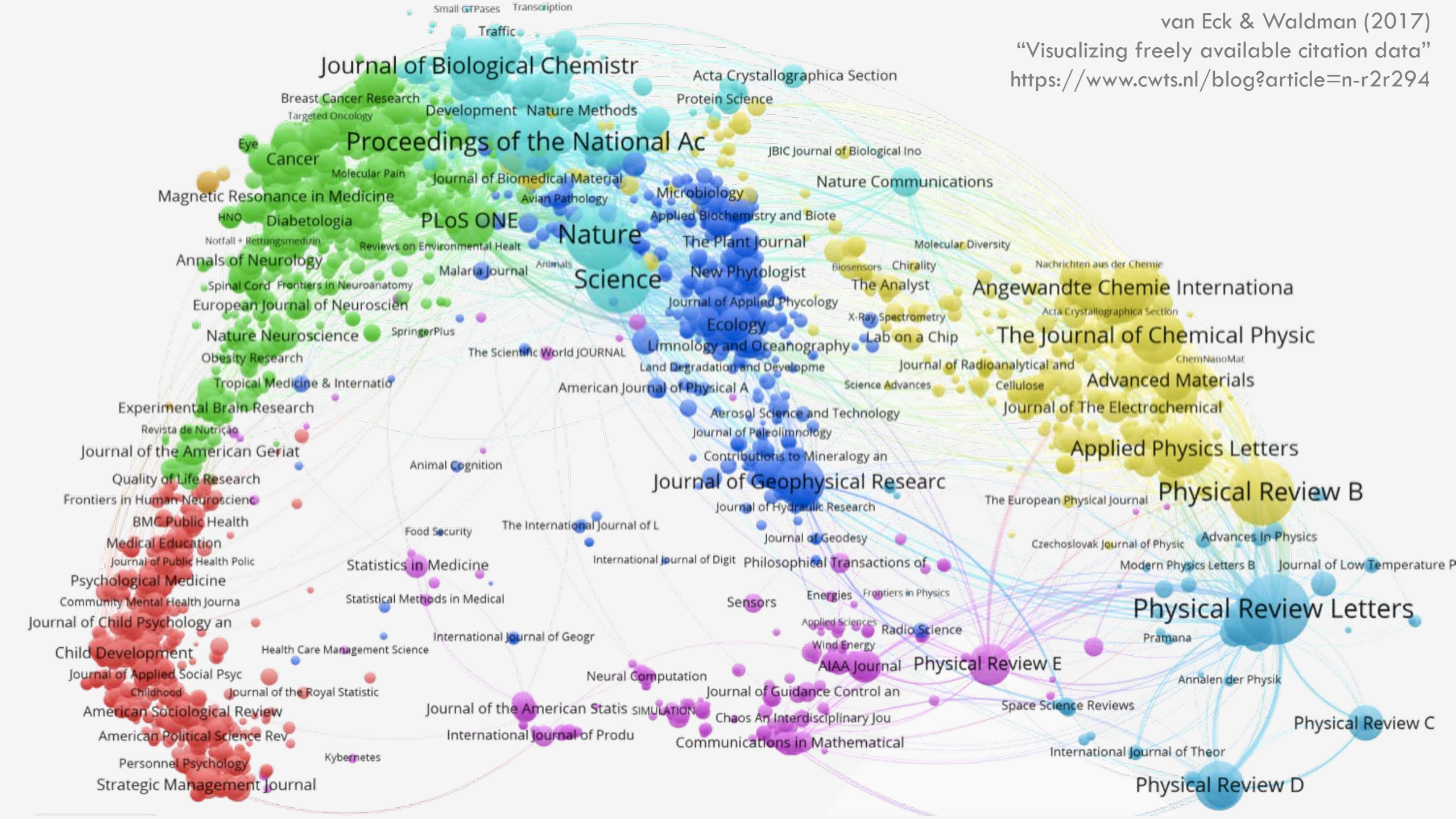


CERN LHC (2017)

- Human Genome Project
- Human Connectome Project
- CERN LHC (and FCC)
- FDA GenomeTrakr
- SETI@home
- USGS Geochemical Landscapes
- ESRI GIS & remote sensing projects
- Digital humanities and social sciences

…lead to production of
**big scholarly data**

Xia F. *IEEE Trans Big Data* (2017)




van Eck & Waldman (2017)
"Visualizing freely available citation data"
https://www.cwts.nl/blog?article=n-r2r294

# DATA SHARING

Stanford University (2013)
Mapping the Republic of Letters

# Data sharing

The practice of making data used in scholarly research available to others.

### What kind?

- Observational
  Experimental
- Derived/compiled
- Simulation
- Reference/canonical

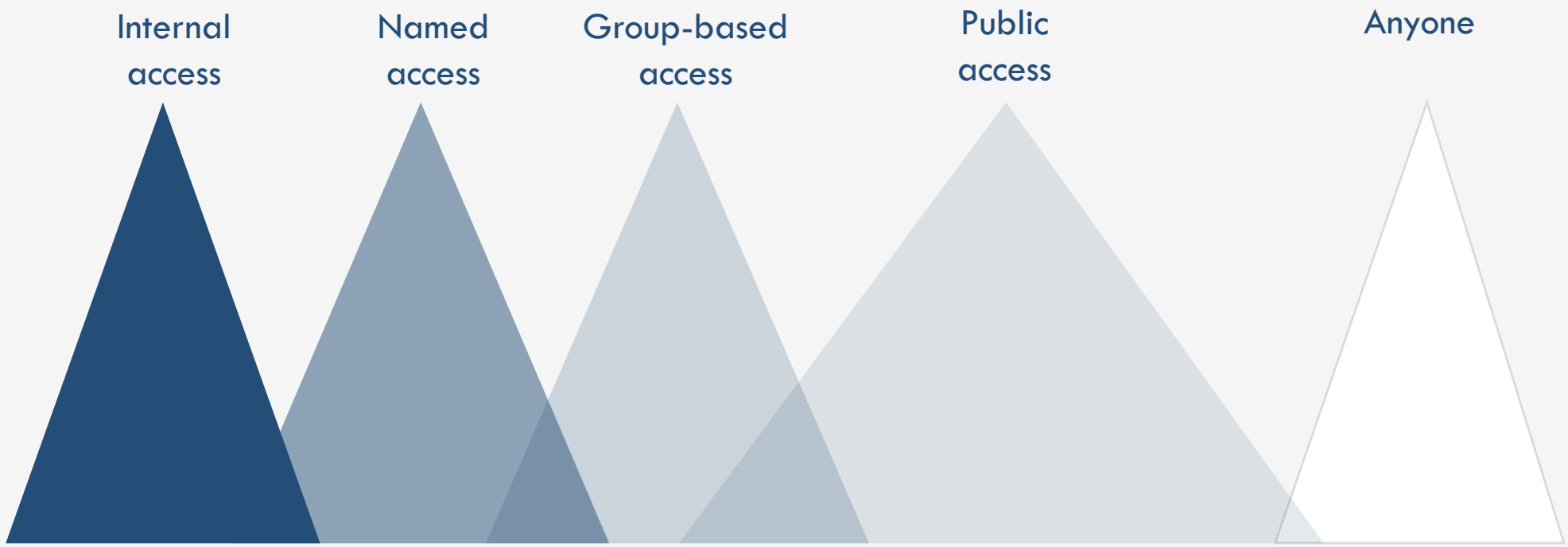
### Which format?

- Text, spreadsheets
- Multimedia
- Models
- Code/software
- Protocols
- Instrument- or discipline-specific

### How?

- Direct request (email, direct contact)
- Supplementary material to an article (not rec'd)
- Personal, institutional, or project webpage
- Institutional repository
- Discipline-specific repository
- General purpose repository

# The data spectrum




Modified from the Open Data Institute
http://theodi.org/data-spectrum

# Data repositories

**Institutional**

- Harvard Dataverse
- Purdue University Research Repository

**Discipline-specific**

- Archaeology Data Service
- GenBank
- PANGAEA

**General purpose**

- Dataverse
- Dryad
- Figshare
- DataHub
- GitHub
- Zenodo

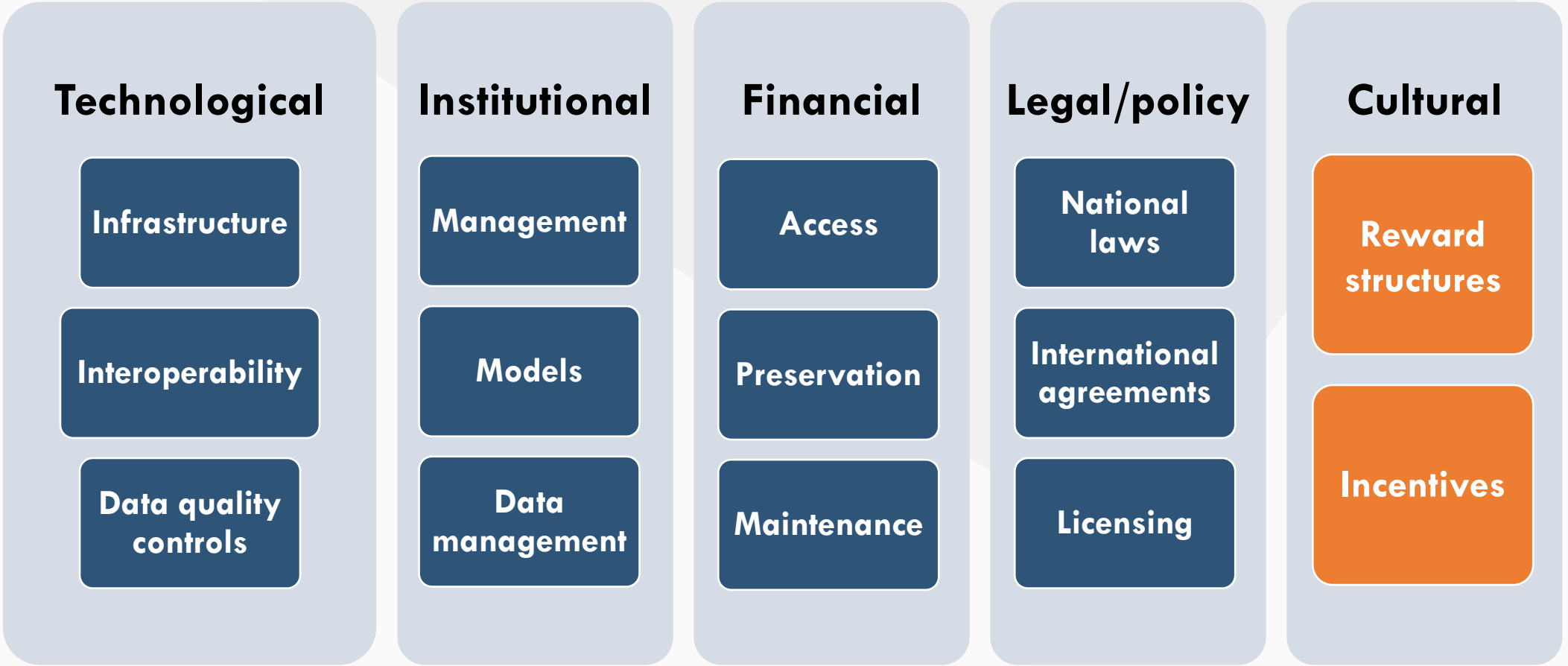
**Directories**

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

FAIRsharing.org
standards, databases, policies

OpenDOAR

# Data sharing issues




Adapted from
OECD Principles and Guidelines for Access to Research Data from Public Funding (2007)

# Benefits of data sharing

## Authors

- Enables exploration of topics not addressed by original researchers
- Encourages diversity of analysis
- Incentivizes the production of higher quality data/analyses
- Supports validation, research transparency, reproducibility and replicability
- Can lead to reuse and discovery, secondary analyses
- Supports the translation of research into practice

## Journals

- Demonstrates commitment to quality
- Discourages fraud
- Data-equipped papers receive increased citations[1]
- Allows journals/publishers to align with funder or government mandates, community norms
- Journals with strong data sharing requirements correlated with higher impact[2]

1. Leitner et al. *Front Neurosci.* (2016)
2. Piwowar & Chapman (2007)

# Data sharing concerns

## Authors

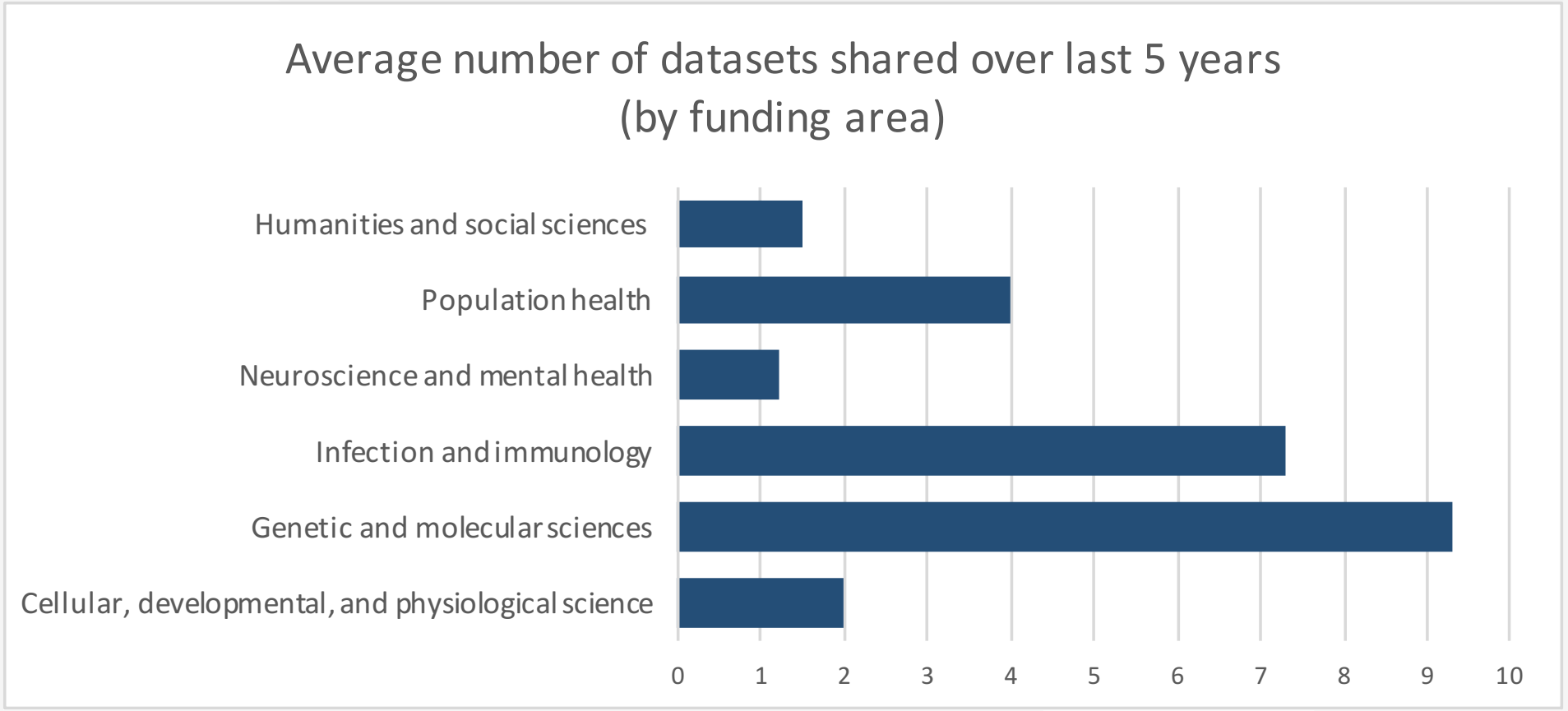
- Lack of acknowledgement or citation, ”what’s in it for me?”
- Inappropriate reuse of the shared data
- Cost and time to prepare data and metadata.
- Impossible to de-identify some data
- Consent forms did not include intention to share data
- Others detecting mistakes in the primary analyses, “looking bad”

## Journals

- Lack of clarity regarding how much data checking is expected (editors/ reviewers)
- Administrative burden
- Sensitive information contained in dataset
- Journal policy at odds with community standards
- Unextractable data files
- Publication priority concerns
- Authors may talk the talk, but full data inaccessible

# Data sharing by discipline

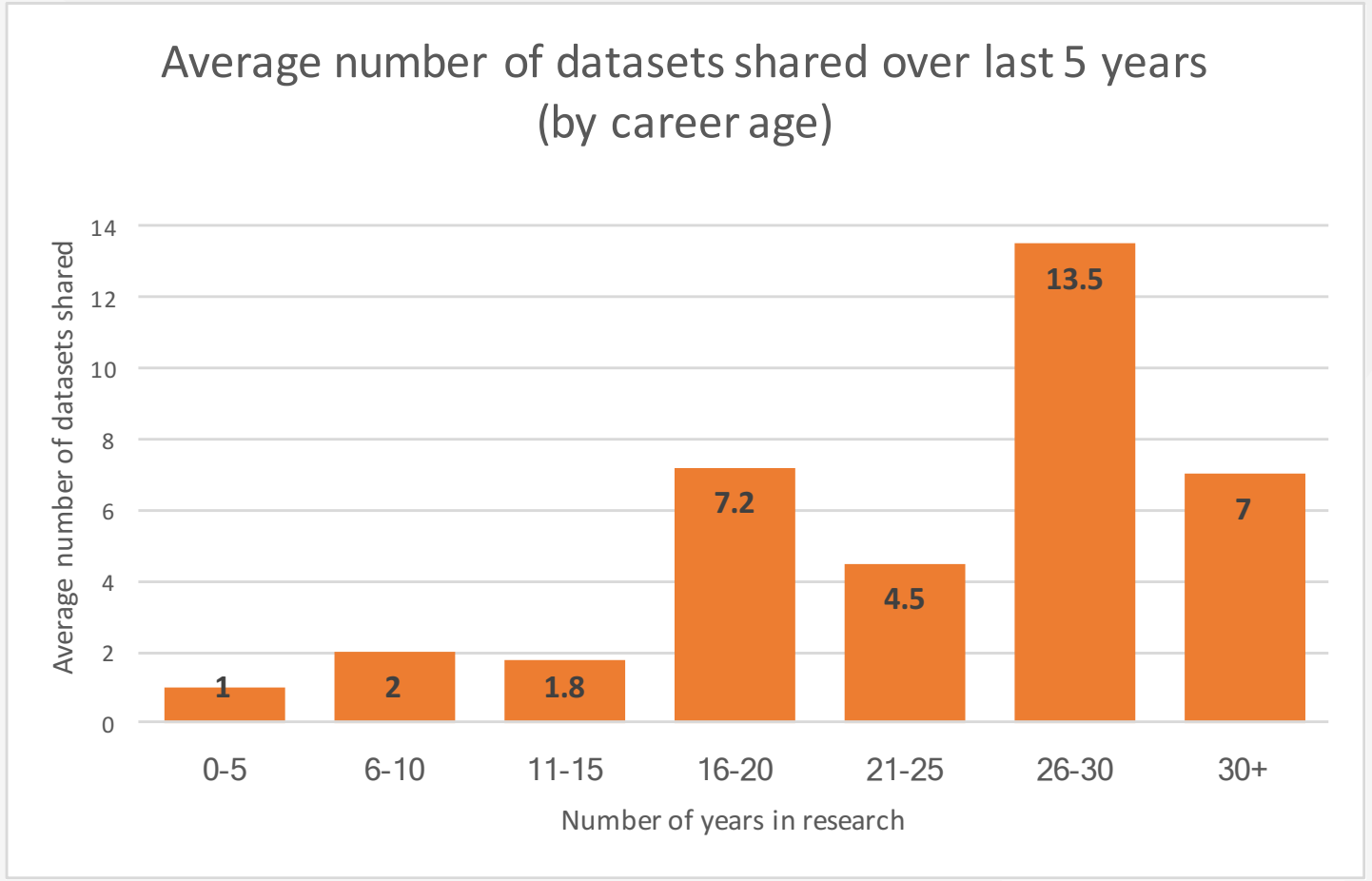
**In this survey study, 95% of respondents generated data and 51% made it available.**




Van den Eynden et al (2016)

# Data sharing by career age

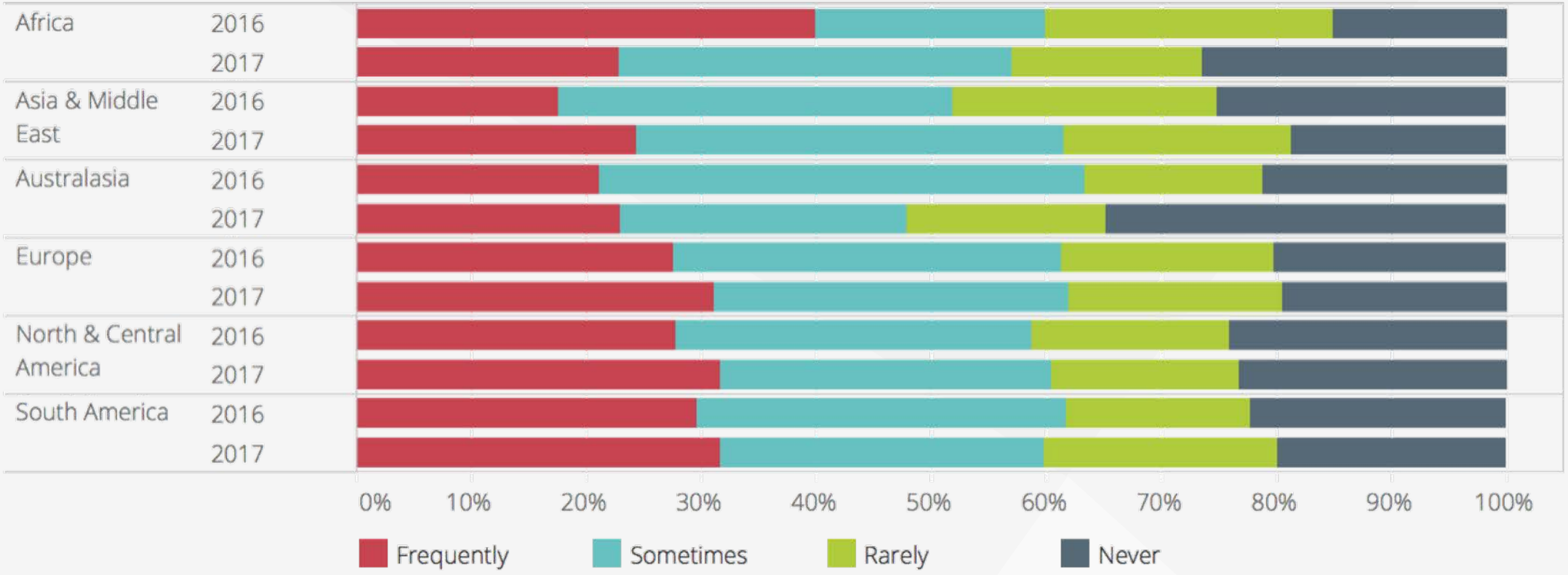
**Datasets were made open in 80% of cases; 19% made data available on request.**




Van den Eynden et al (2016)

# Current state of open data sharing

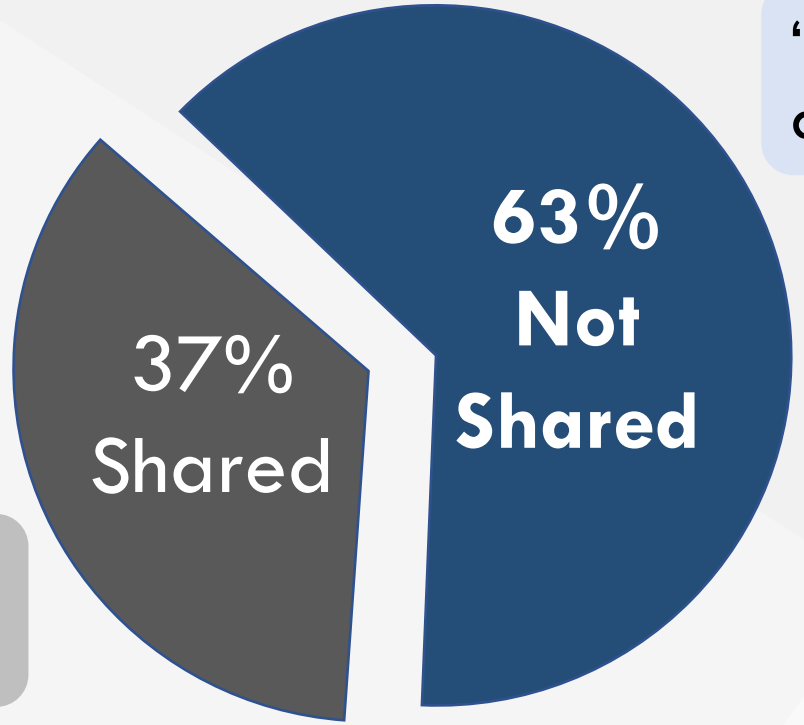
**Overall, 60% routinely share datasets—only 21% have never made data open.**




Figshare, State of Open Data Survey (2017)

# Observational study data sharing

**Review of data sharing in *BMJ* articles describing cohort, case-control, and cross-sectional studies published 2015-2017**




"no additional data available"

"data are not publicly available"

"code [or technical appendix] available" but no data reference provided

"data are available upon request"

reference to data location provided

Adapted from
McDonald L et al. *F1000 Res.* (2017)

# Clinical trial data sharing

**Survey study to assess clinical trial data sharing behavior across 90 trials (2012-2016)**

More than half of respondents had **prepared a data sharing plan**; only 29% were written.

Patient consent was addressed inconsistently, 48% of respondents had **not addressed data sharing** in patient consent forms.

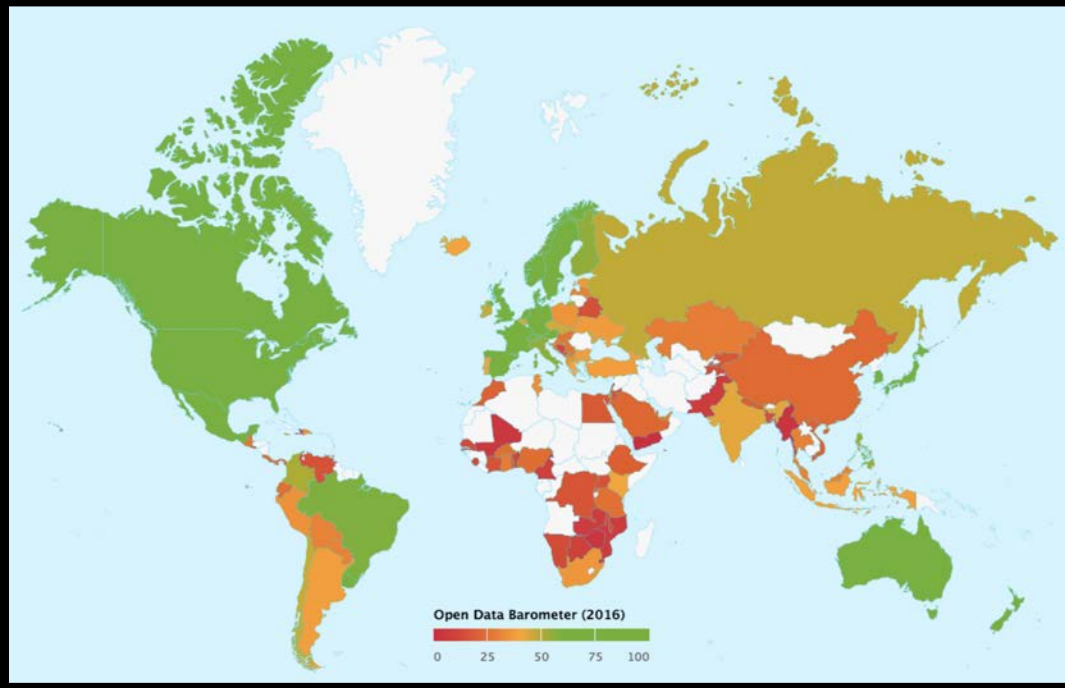
Of those who had granted direct sharing requests, 92% reported that **preparing data for sharing required substantial time** (median 18 hours).

*Table.* Characteristics of Studies and Experiences of Data Sharing*

| Variable | All Studies, *n* (%) | Studies Granting ≥1 Data Sharing Request, % |
|---|---|---|
| **Journal** | | |
| *Annals of Internal Medicine* | 25 (27.8) | 16.0 |
| *PLOS Medicine* | 24 (26.7) | 20.8 |
| *The BMJ* | 41 (45.6) | 39.0 |
| **Funding source** | | |
| Any industry | 14 (15.6) | 50.0 |
| Nonindustry | 76 (84.4) | 23.7 |
| **Data sharing plan** | | |
| Written | 26 (28.9) | 50.0 |
| Discussed | 23 (25.6) | 34.8 |
| None | 41 (45.6) | 9.8 |
| **Consent form** | | |
| PHI shared with safeguards | 4 (4.4) | 25.0 |
| Only anonymized data shared | 27 (30.0) | 37.0 |
| Sharing not addressed | 43 (47.8) | 14.0 |
| Other | 16 (17.8) | 50.0 |
| **Deidentification to national standards** | | |
| By expert | 5 (5.6) | 40.0 |
| Deidentified but not by expert | 48 (53.3) | 29.2 |
| Not deidentified | 26 (28.9) | 15.4 |
| Other | 11 (12.2) | 45.5 |

Tannenbaum S et al. *Ann Int Med.* 2018.

# DATA POLICIES




World Wide Web Foundation.
Open Data Barometer 4th ed.

Wagner, Jonkers. *Nature*. (2017)

# Funders with data sharing policies

- Bill and Melinda Gates Foundation
- Gordon and Betty Moore Foundation
- Howard Hughes Medical Institute
- Laura and John Arnold Foundation
- MacArthur Foundation
- NASA
- National Endowment for the Humanities
- National Institutes of Health
- National Institute of Justice
- National Ocean and Atmospheric Administration
- National Science Foundation
- Simons Foundation
- US Department of Education,
- Institute of Education Sciences
- US Department of Energy
- US Department of Agriculture
- US Department of Transportation
- US Geological Survey
- US Environmental Protection Agency

SHERPA Juliet: Research Funders' Open Access Policies. (2018)
MIT Libraries: Research Funder Requirements. (2017)
US Environmental Protection Agency News Release (2018)

# Global policy timeline

| | | |
|---|---|---|
| **1997** | National Research Council, *Bits of Power: Issues in Global Access to Scientific Data* | |
| **2002** | Budapest Open Access Initiative | |
| **2003** | Bethesda Statement on OA Publishing | Berlin Declaration on Open Access to Knowledge |
| **2007** | OECD Principles and Guidelines for Access to Research Data | |
| **2010** | NSF Data Management Plan Requirements | |
| **2012** | Royal Society, *Science as an Open Enterprise* | SF Declaration on Research Assessment |
| **2013** | OSTP memo, "Increasing Access to the Results of Federally Funded Scientific Research" | G8 Science Ministers Statement<br>G8 Open Data Charter |
| **2014** | Horizon 2020 Open Data Policy Pilot | FORCE11 Joint Declaration of Data Citation Principles |
| **2015** | Leiden Manifesto for research metrics | |
| **2016** | FAIR guiding principles for scientific data management and stewardship | ICJME proposal, "Sharing Clinical Research Data" |
| **2017** | ICJME requirement, "Data Sharing Statements for Clinical Trials" | |

# FAIR Data Principles

Translating the FAIR principles in practice differs depending on discipline, but, in broad strokes, research data ought to be:

## FINDABLE

Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

## ACCCESSIBLE

Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

## INTEROPERABLE

Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

## REUSABLE

Data and collections have a clear usage licenses and provide accurate information on provenance.

Adapted from Wilkinson MD et al. *Sci Data* (2016)
Icons by Flaticon (find, key, cloud by Freepik; reuse by Vitaly Gorbachev)

FORCE11

# Joint Declaration of Data Citation Principles

1. Data citations should be accorded the same **importance** as citations of other research objects.
2. Data citations should facilitate giving scholarly **credit and attribution.**
3. Wherever a claim relies upon data, the data itself should be cited as **evidence.**
4. Data citations require **persistent identifiers** that are machine actionable, globally unique, and widely used by a community.
5. Data citations should facilitate **access** to the data directly, along with metadata
6. Unique identifiers, and metadata describing the data, and its disposition, should **persist.**
7. Data citations should facilitate identification, access, and **verification** of the data that support a claim.
8. Data citations should accommodate difference in field-specific practices, but remain **interoperable.**

FORCE11 Joint Declaration of Data Citation Principles. (2014)

# ICMJE Data Sharing Policy

## Sharing Clinical Trial Data (2016)

- Proposed that authors of all clinical trial manuscripts share **de-identified individual-patient data** underlying results within 6 months of publication
- Suggested new requirements for describe the results of any analysis performed with publicly shared data (credit, cite, attest fair use)

## Sharing Statements for Clinical Trials (2017)

- Requires the inclusion of a **data sharing statement** in all manuscripts in member journals starting July 1, 2018.
- Clinical trials that begin January 1, 2019, must include a **data sharing plan** in the trial's registration.
- Does not require clinical data sharing or secondary analyses

# Journal/publisher policies

## Variable at best. Often, MIA.

- Journal data policies are often vague (do not define data, where/when to deposit, expectations of access) [1]
- Higher-impact journals tend to have **enforceable data policies**[2]
- Journal data policies can be difficult to find, and even more difficult to understand… leads to perceived compliance problems and confusion from peer review to post-publication. [3]
- Most journal policies do not guide authors well on ensuring data are available and reusable. [4]


1. Sturges et al. *J Assoc Inform Sci Tech*. 2015
2. Piwowar & Chapman. *AMIA Annu Symp Proc*. 2008
3. Vasilevsky NA et al. *PeerJ*. 2017
4. Naughton & Kernohan. *Insights*.2016

# Toward policy standardization

## Policy Types

| Type 1 | Type 2 | Type 3 | Type 4 |
| --- | --- | --- | --- |
| Data sharing and data citation is encouraged but not required | Data sharing and evidence of data sharing encouraged | Data sharing encouraged and statements of data availability required | Data sharing, evidence of data sharing, and peer review of data required |

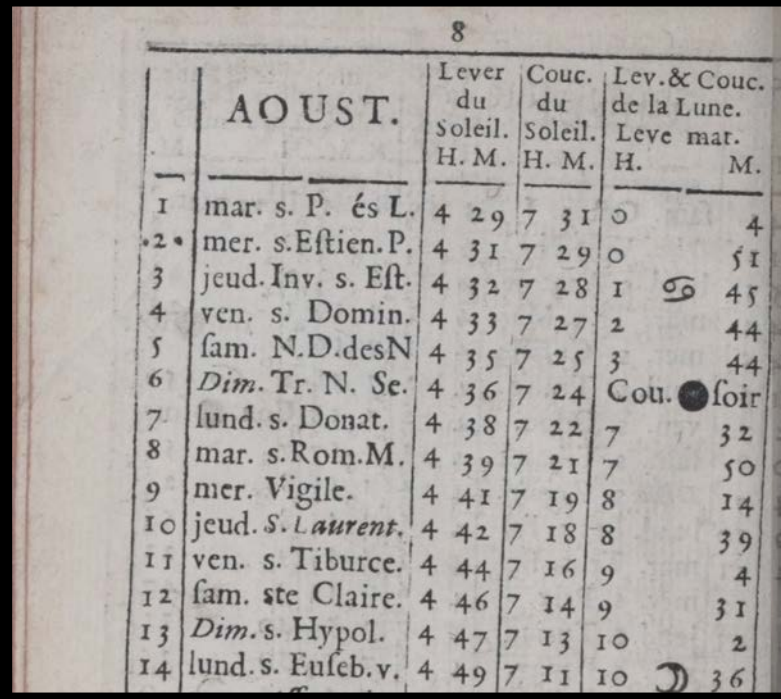
## Process

1. Identify and agree the most relevant policy type for individual journals
2. Implement standard text and processes into relevant journal guides and publishing workflows
3. Provide a consistent and easy-to-follow journal data policy for authors, researchers, and peer reviewers.

SPRINGER NATURE

Hrynaszkiewicz et al. *IJDC*. (2017)

# DATA PUBLISHING +CITATIONS



8

| | AOUST. | Lever du Soleil. H. M. | Couc. du Soleil. H. M. | Lev. & Couc. de la Lune. Leve mat. H. M. |
|---|---|---|---|---|
| 1 | mar. s. P. és L. | 4 29 | 7 31 | 0 4 |
| 2 | mer. s. Eſtien. P. | 4 31 | 7 29 | 0 51 |
| 3 | jeud. Inv. s. Eſt. | 4 32 | 7 28 | 1 ♋ 45 |
| 4 | ven. s. Domin. | 4 33 | 7 27 | 2 44 |
| 5 | ſam. N.D.desN | 4 35 | 7 25 | 3 44 |
| 6 | *Dim.* Tr. N. Se. | 4 36 | 7 24 | Cou. ● ſoir |
| 7 | lund. s. Donat. | 4 38 | 7 22 | 7 32 |
| 8 | mar. s. Rom. M. | 4 39 | 7 21 | 7 50 |
| 9 | mer. Vigile. | 4 41 | 7 19 | 8 14 |
| 10 | jeud. *S. Laurent.* | 4 42 | 7 18 | 8 39 |
| 11 | ven. s. Tiburce. | 4 44 | 7 16 | 9 4 |
| 12 | ſam. ste Claire. | 4 46 | 7 14 | 9 31 |
| 13 | *Dim.* s. Hypol. | 4 47 | 7 13 | 10 2 |
| 14 | lund. s. Euſeb. v. | 4 49 | 7 11 | 10 ☽ 36 |

*Connaissance des temps*, 1679

# Publishing data

## Data availability statements

**Purpose:** provide a statement and description where data underlying results may be found.

Publishers/journals and may also provide templates with preferred language.

**Reproducible Research Statement:** *Study protocol:* Available from Dr. Gross (e-mail, cary.gross@yale.edu). *Statistical code:* Not applicable. *Data set:* Deidentified data set available from Dr. Gross (e-mail, cary.gross@yale.edu).

doi:10.7326/M18-0723

*Ann Int Med* (2018)
doi:10.7326/M18-0723

---

Data deposition: The data that support the findings of this study can be accessed on GitHub at https://github.com/jmlbr/body-machine_interface_drone.

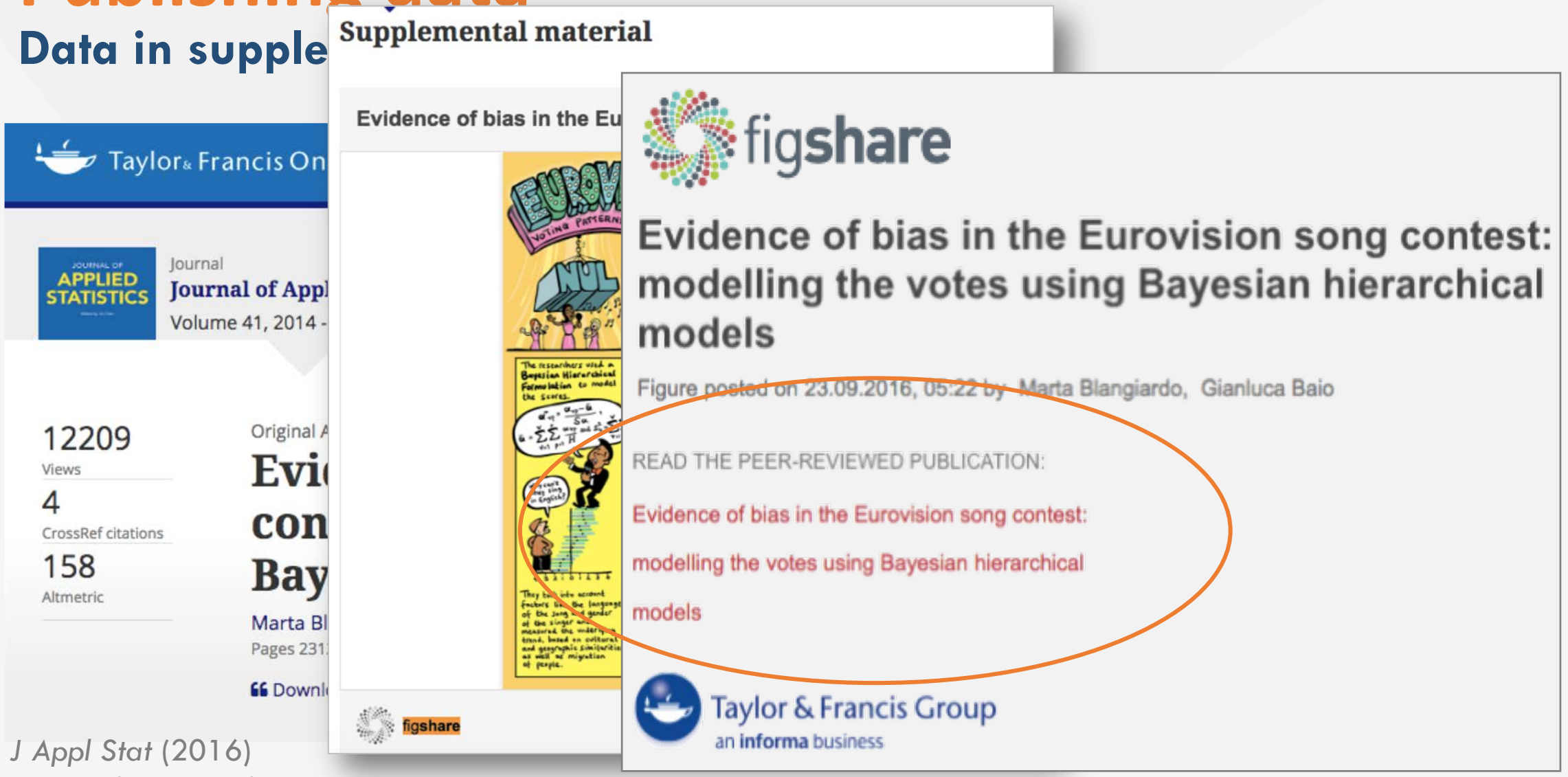
---

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1718648115/-/DCSupplemental.

*PNAS* (2018)
doi.org/10.1073/pnas.1718648115

# Publishing data

## Data in supple



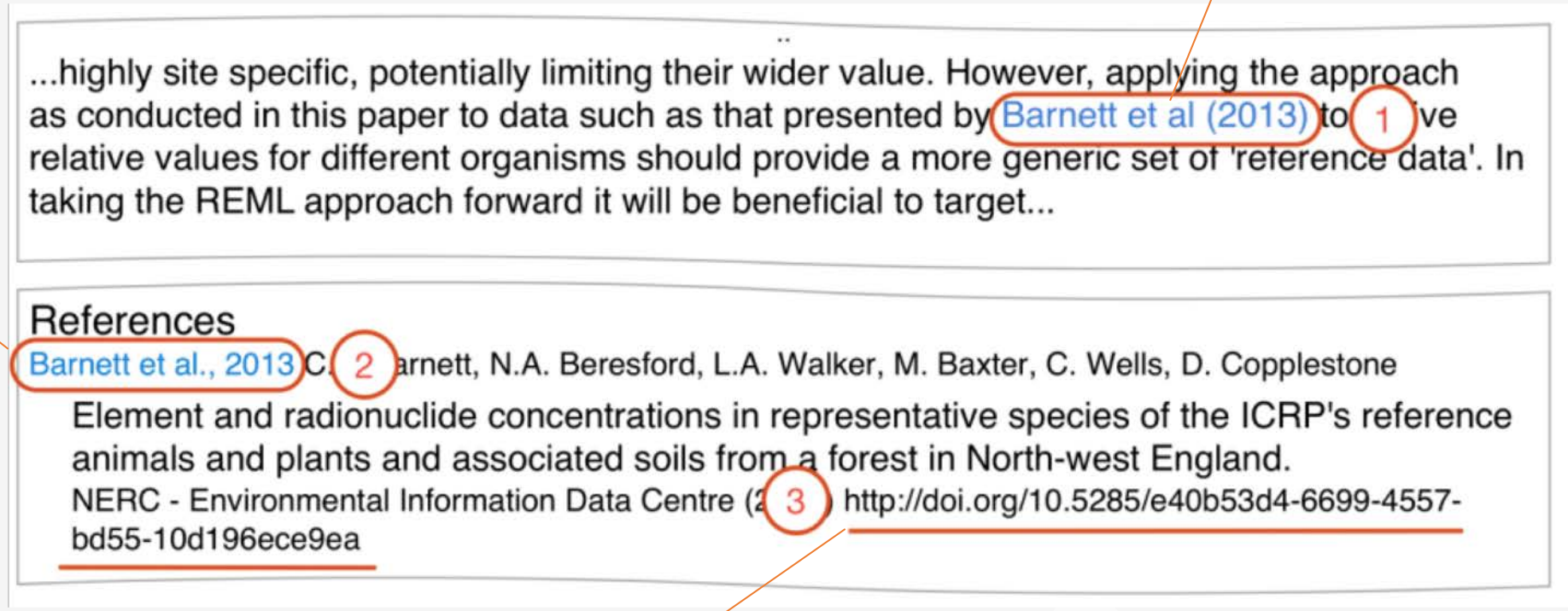

*J Appl Stat* (2016)
doi.org/10.1080/02664763.2014.909792

# Publishing data

## Data cited in reference list

1. Citation in text points to reference list

2. Citation format in reference list contains author, title, repository, year



...highly site specific, potentially limiting their wider value. However, applying the approach as conducted in this paper to data such as that presented by Barnett et al (2013) to (1) ve relative values for different organisms should provide a more generic set of 'reference data'. In taking the REML approach forward it will be beneficial to target...

References

Barnett et al., 2013 C (2) arnett, N.A. Beresford, L.A. Walker, M. Baxter, C. Wells, D. Copplestone

Element and radionuclide concentrations in representative species of the ICRP's reference animals and plants and associated soils from a forest in North-west England.

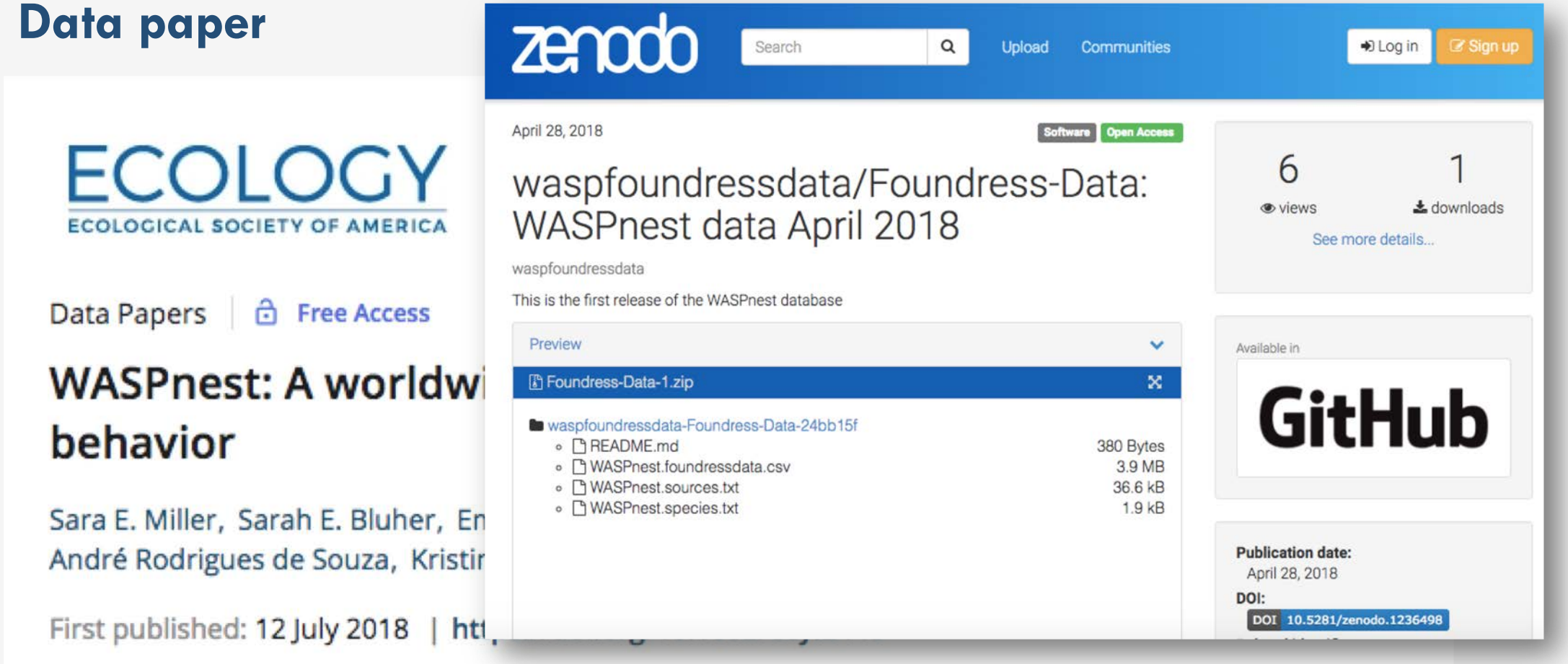
NERC - Environmental Information Data Centre (2 (3) http://doi.org/10.5285/e40b53d4-6699-4557-bd55-10d196ece9ea

3. Persistent identifier present

bioRxiv (2017)
dx.doi.org/10.1101/100784

# Publishing data

## Data paper




*Ecology* (2018)
doi.org/10.1002/ecy.2448

# Publishing data

## Data journal

Data Records

et Biology
der Creative Commons
ense. EMODnet Biology
ous sources and processes them
menting common standards
Ocean Biogeographic
lobal Biodiversity Information
re. EMODnet Biology hosts also
and documents that would
he datasets described in this

nd Experimental Geophysics
r Environmental Protection
://dx.doi.org/10.14284/284

/doi.org/10.14284/170 (2017)

nd Experimental Geophysics
r Environmental Protection
://dx.doi.org/10.14284/290

nd Experimental Geophysics
r Environmental Protection

Data Citations

Erratum: Fish and fishery historical data since the 19th century in the Adriatic Sea, Mediterranean

Tomaso Fortibuoni, Simone Libralato, Enrico Arneri, Otello Giovanardi, Cosimo Solidoro & Saša Raicevich

*Scientific Data* **5**, Article number: 180144 (2018) | Download Citation

The original article was published on 12 September 2017

*Scientific Data* 4:170104 doi: 10.1038/sdata.2017.104 (2017), Published 12 September 2017; Updated 24 July 2018

In the HTML version of this Data Descriptor, Data Citation 8 incorrectly listed the repository as VLIZ instead of EMODnet Bathymetry.

*Sci Data* (2017)
doi.org/10.1038/sdata.2017.104

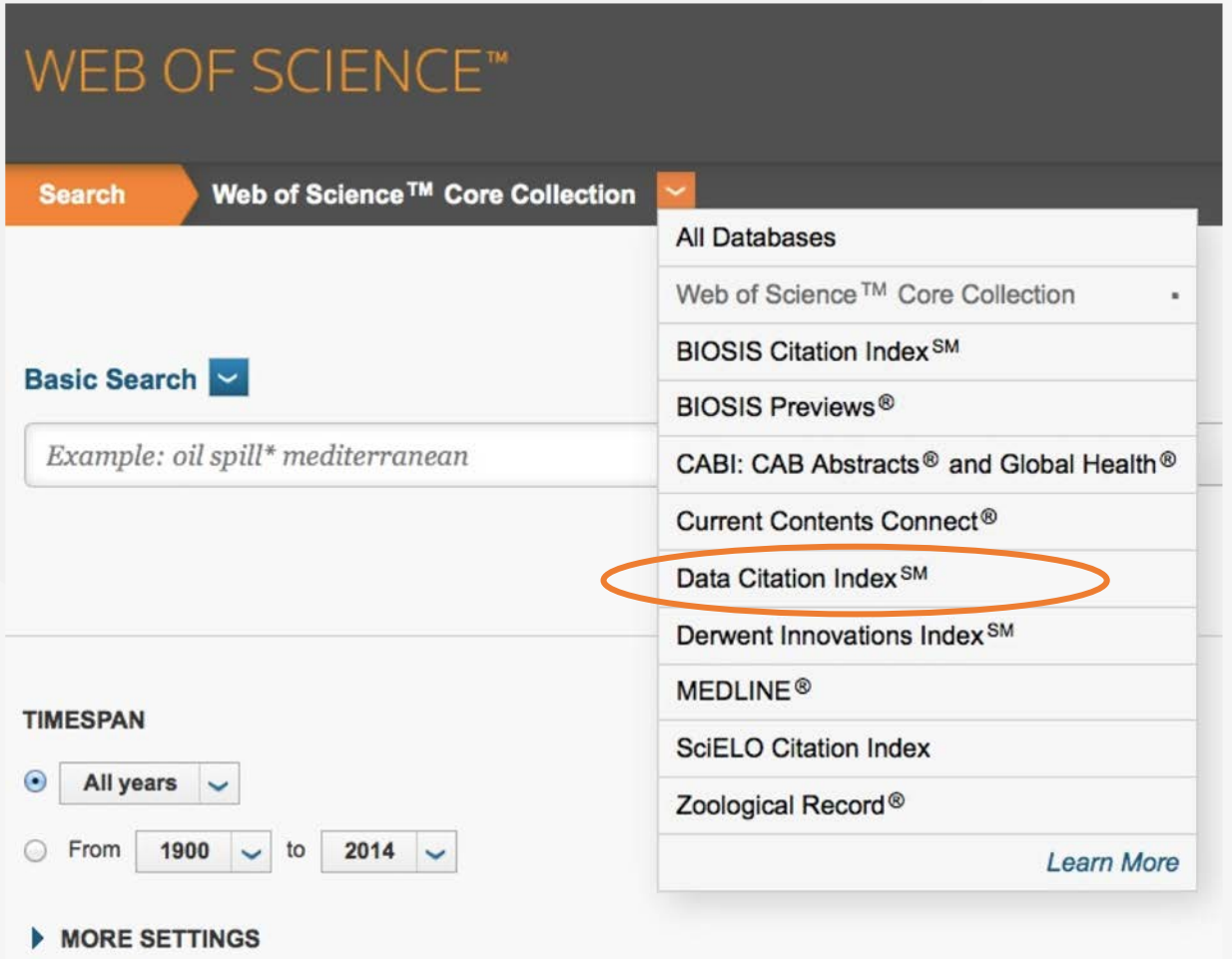
# Tracking citations

## Clarivate Data Citation Index

Indexes data from more than 300 global repositories, sorts records into document types:

- **Data Study:** description of the study or experiment with its associated data.
- **Data Set:** A single set of data or a data file provided by the repository as part of a collection, data study, or experiment.
- **Software:** A computer program or package in source code or compiled form, which can be installed on another machine and used to support & analyze research.

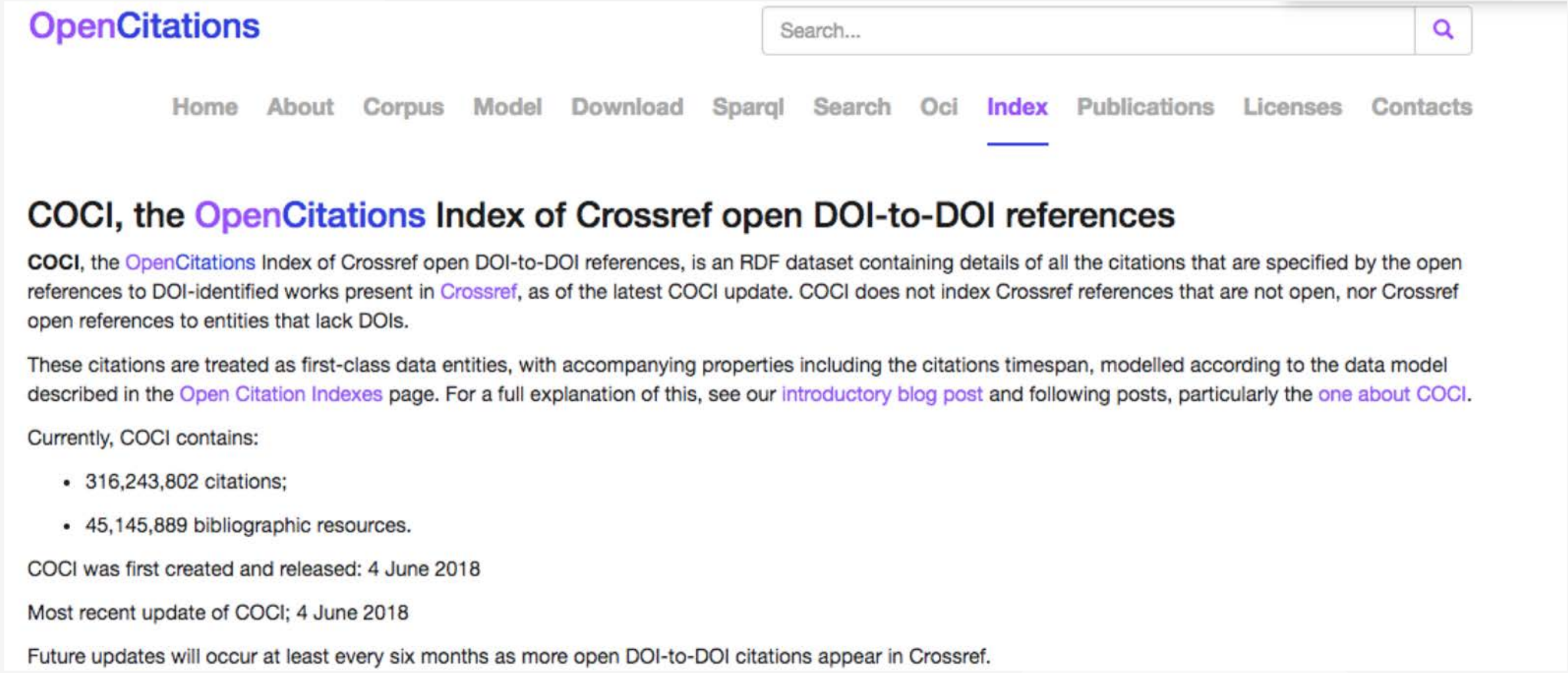
Citation data to data linked with associated paper(s).




https://clarivate.com/products/web-of-science/web-science-form/data-citation-index/

# Tracking citations

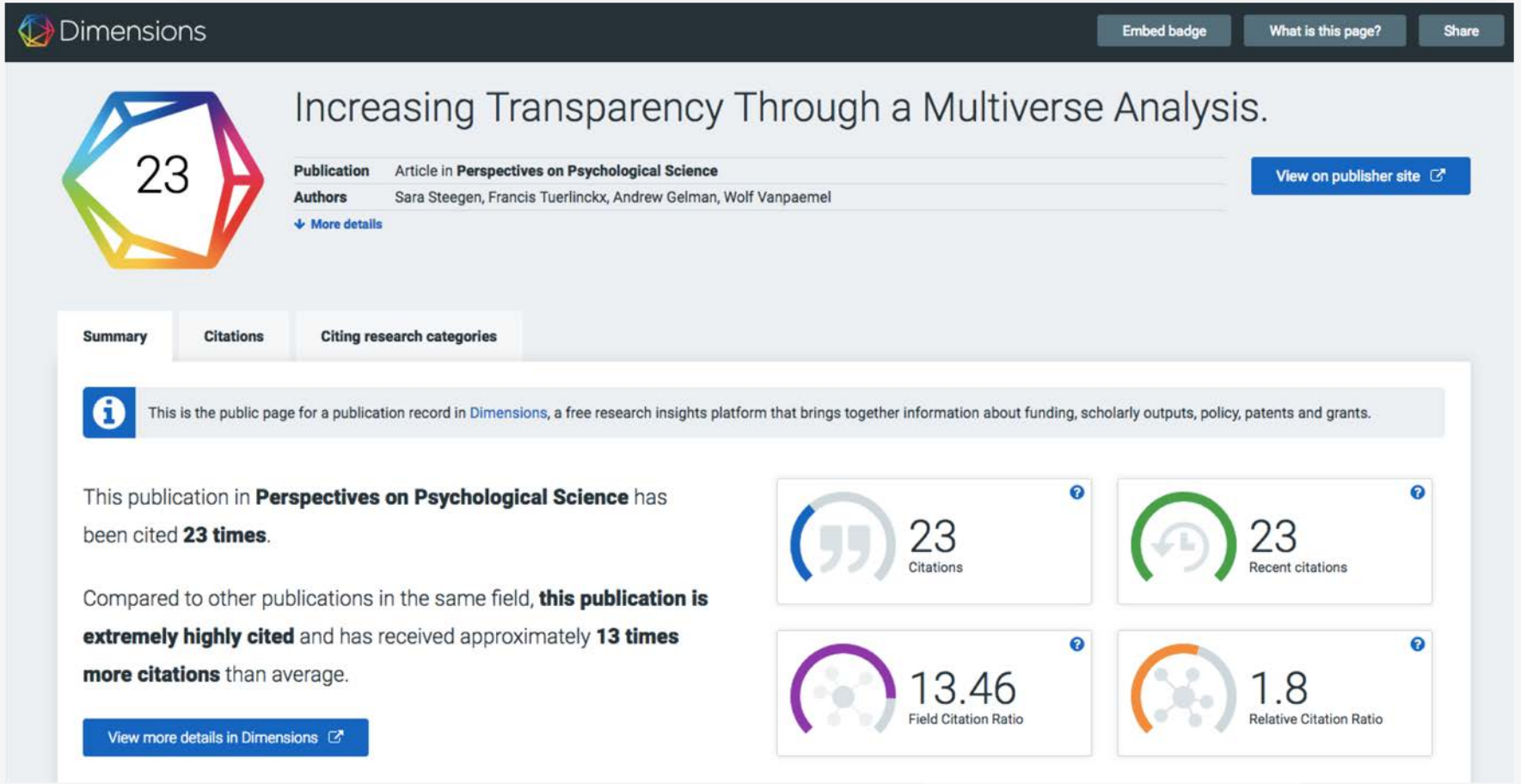
## COCI (OpenCitations Index of Crossref Open DOI-toDOI references)




http://opencitations.net/index/coci

# Tracking citations

## Dimensions platform




https://www.dimensions.ai/

# INCORPORATING DATA

# Considerations in publishing data

## Before submission

- Data required?
- How to format?
- Peer review expected?
- Recommended repositories
- Licensing terms
- Guidelines

## Submission & Review

- Datasets cited and fully referenced?
- Data availability structured and complete?
- Reviewers or editors expected to check?
  - Review forms updated?
  - Does review process require links to blinded datasets?

## Production

- DOIs or other persistent identifiers present?
- Data availability statement converted properly (from ms to typeset page)?
- XML tags in place?

## Publication

- Data citations and availability statements in article?
- Tags present to deliver metadata to CrossRef, PubMed?

# TOP Guidelines for journals

## Eight standard areas

| | |
|---|---|
| Citation Standards | Design and Analytic Transparency |
| Data Transparency | Preregistration of Studies |
| Analytic Methods Transparency | Preregistration of Analysis Plans |
| Research Materials Transparency | Replication |

## Three levels

***E.g. Data Transparency***

**Level 0** – Journal encourages data sharing or says nothing.

**Level 1** – Article states whether data are available and where.

**Level 2** – Data must be posted to a trusted repository; exceptions noted

**Level 3** – Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.

OSF Open Science Framework
Transparency and Openness Promotion Guidelines

Nosek et al. *Science*. (2015)

# Checklists for editors

| TOP Level 2 | **Citation**<br>(Y or N/A requried for each) |
|---|---|
| Y, N, or N/A | Are all data, program code and other methods appropriately cited within the text and listed in the reference section? |
| Y, N, or N/A | Are all data, code, and methods citations given unique, persistent IDs, (e.g. DOI)? |

| TOP Level 2 | **Data, Analytical Methods, Code, and Research Materials Transparency**<br>(Y or N/A required for each) |
|---|---|
| Y, N, or N/A | **If reusing data** available from public repositories, does author provide program code, scripts for statistical packages, and other documentation sufficient to allow an informed researcher to precisely reproduce all published results? |
| | **If using original data**, do the authors... |
| Y, N, or N/A | ...make the data available at a trusted digital repository? (Note: If all data required to reproduce the reported analyses appears in the article text, tables, and figures then it |

# On the horizon…

**Draft policy recommendations from the Research Data Alliance Policy Standardization and Implementation Interest Group to address:**


**Open to anyone – get involved!**

- Definition of research data
- Exceptions to data policies
- Embargoes on data release
- Data sharing via supplemental material
- Data repositories
- Data citation
- Data licensing
- Researcher/author support
- Data availability statements
- Mandatory data sharing/citation
- Peer review of data
- Data management plans

https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation

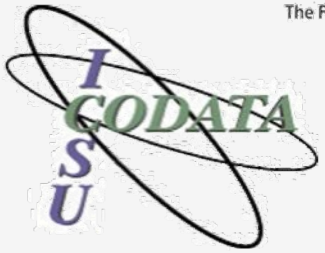
# Best practices

- Update information for authors with clear statement of **journal expectations**
- Specify a policy for **citation of data** (datasets used and generated by study called out in reference list) – **if you ask for it, check it.**
- Detail how to **format data citations** (e.g. author(s), title, year, version, data repository, persistent identifiers) and locate these in the main reference list
- Ask authors for **data availability statements**; incorporate check into workflow
- Establish relationships with and provide guidance on **suitable repositories** (general, institutional, and subjectspecific) and how to find one
- Consider implications of **licensing terms** used to archive data
- Build **author support workflows** (FAQs, specific contact)

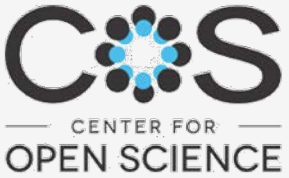
Mayo et al. *IJDC*. (2016)
Austin et al. *Int J Digit Libr* (2017)
Cousijn et al. *bioRxiv*. (2017)

# Resources

- Committee on Data for Science and Technology
- Center for Open Science
- DataCite
- The Dataverse Project
- FORCE11
- Make Data Count
- Open Data Institute
- Open Knowledge International
- Research Data Alliance
  - Policy Standardization and Implementation IG
  - Publishing Data IG










# Reading


Ackoff R. From data to wisdom. *Journal of Applied Systems Analysis*. 1989; 16: 3-9.

Austin CC, et al. Key components of data publishing: using current best practices to develop a reference model for data publishing. *Int J Digit Libr*. 2017; 18(2): 77-92.

Buboltz A. Pages from the first human genome. Broad Institute. October 18, 2010.

Buckheit J, Donoho DL. Wavelab and reproducible research. In: *Wavelets and Statistics*. New York, NY: Springer. 1995:55–81.

Budapest Open Access Initiative. 2002.

Cousijn et al. A data citation roadmap for scientific publishers. bioRxiv. January 19, 2017.

Connoissance des temps. 1679

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014

Fenner M, et al. Code of practice for research data usage metrics release 1. *PeerJ Preprints*. 2018; 6:e26505v1

Figshare (Astell M, Penny D, Treadway J, Fane B.) State of Open Data survey 2017. October 2017.

Forscher BK. Chaos in the brickyard. *Science*. 1963 Oct 18;142(3590):339.

G8 Open Data Charter and Technical Annex. June 18, 2013.

G8 Science Ministers Statement. June 13, 2013.

Goodman SN, et al. What does research reproducibility mean? *Sci Transl Med*. 2016 Jun 1;8(341):341ps12.

Hey J. The data, information, knowledge, wisdom chain: the metaphorical link. Intergovernmental Oceanographic Commission (UNESCO). 2004.

Hrynaszkiewicz I et al. Standardising and harmonising research data policy in scholary publishing. *IJDC*. 2017;12(1): 65-71.

ICJME. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. June 6, 2017.

ICMJE. Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors. January 26, 2016.

Kaplan F, di Lenardo I. Big data of the past. *Front Digit Humanit*. May 29, 2017.

Laney D. 3D data management: Controlling data volume, variety and velocity. Meta Group. February 6, 2001.

Hicks et al. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*. 2015 Apr 23;520(7548):429-31.

McDonald L, et al. A review of data sharing statements in observational studies published in the BMJ: A cross-sectional study. Version 2. *F1000Res*. 2017 Sep 19 [revised 2017 Nov 6];6:1708.

Mannion P. Optimal analysis algorithms are IoT's big opportunity. *IEEE GlobalSpec*. 2015.

Mayo et al. The location of the citation: changing practices in how publications cite original data in the Dryad Digital Repository. *Int J Digital Curation*. 2016;11(1).

MIT Libraries: Research Funder Requirements. August 15, 2017

National Academies of Sciences, Engineering, and Medicine. Big Data and Analytics for Infectious Disease Research, Operations, and Policy: Proceedings of a Workshop. Forum on Microbial Threats. Washington (DC): National Academies Press (US); 2016 Dec 8.

Naughton L., Kernohan D. Making sense of journal research data policies. *Insights*. 2016; 29(1), 84–89.

Nosek et al. Promoting an open research culture. *Science*. 2015 June 26; 348(6242): 1422–1425.

OECD Principles and Guidelines for Access to Research Data from Public Funding. 2007.

O'Mareen et al. Patient consent to publication and data sharing in industry and NIH-funded clinical trials. *Trials*. 2018; 19:269.

Open Knowledge International. Open Data Handbook. July 15, 2018.

OSTP Memorandum on Increasing Access to the Results of Federally Funded Scientific Research. February 22, 2013.

Piwowar H, Chapman W. Identifying Data Sharing in Biomedical Literature. *AMIA Annu Symp Proc*. 2008; 596–600.

San Francisco Declaration on Research Assessment. 2012.

Science International. Open Data in.a Big World: An International Accord. 2015.

SHERPA Juliet. Research Funders' Open Access Policies. July 15, 2018.

Stanford University, Humanities+Design at CESTA. Mapping the Republic of Letters. 2013.

Sturges et al. Research Data Sharing: Developing a Stakeholder-Driven Model for Journal Policies. *J Assoc Inform Sci Tech.* 2015; 66(12):2445-55.

Tannenbaum S. Early Experiences With Journal Data Sharing Policies: A Survey of Published Clinical Trial Investigators. *Ann Int Med.* July 17, 2018.

U.S. Environmental Protection Agency, News Releases, "EPA Administrator Pruitt proposes rule to strengthen science used In EPA regulations." April 28, 2018.

Van Eck NJ, Waltman L. Visualizing freely available citation data using VOSviewer. Centre for Science and Technology Studies (CWTS), Leiden University. October 23, 2017.

Van den Eyden et al. Survey of Wellcome researchers and their attitudes to open research. 2016.

Vasilevsky NA, et al. Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ. 2017; 5*:e3208.

Wagner C, Jonkers K. Open countries have strong science. *Nature.* 4 October 2017;550:32-33.

Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.

World Wide Web Foundation. Open Data Barometer 4th ed. 2017.

Xia F et al. Big Scholarly Data: A Survey. *IEEE Transactions on Big Data.* 2017;3(1):18-35.

Jen Mavzer, Coronis Group
E: consult@coronisgroup.com |
jmavzer@gmail.com
ID: 0000-0002-0212-0773

# THANK YOU!