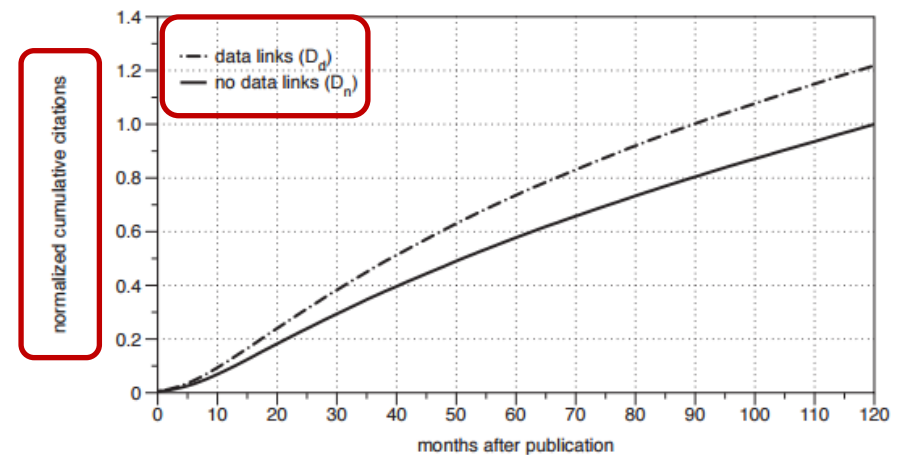# Research data

COPE Seminar, May 2017

Iain Hrynaszkiewicz

SPRINGER NATURE

Benefits of data sharing/publishing – what's the evidence?

- Increased citations & use of papers with data available or linked[1-6]

- Increase reproducibility (quality/robustness) of research[7]

- Increased productivity[8]

- Reducing harm and costs of biased/non-transparent research[9,10]

- Helps researcher career progression[11]
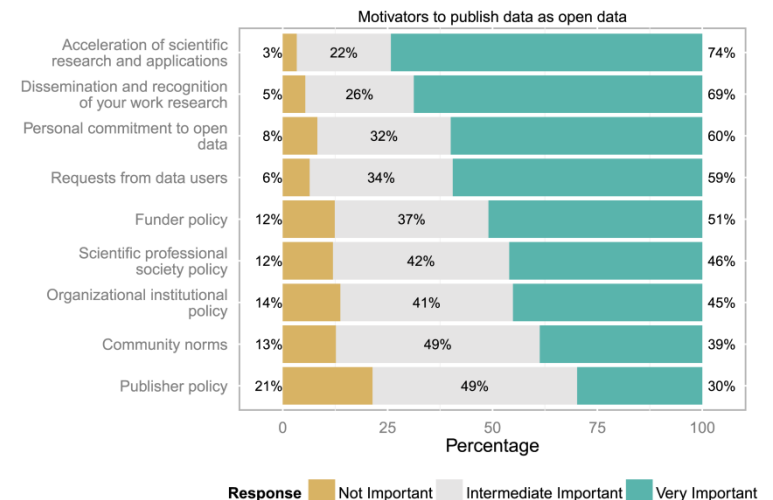
- Better return on investment in research funding[12]

1. Henneken & Accomazzi (2011) https://arxiv.org/abs/1111.3618
2. Belter (2014) http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0092590
3. Piwowar & Vision (2013) https://doi.org/10.7717/peerj.175
4. Piwowar et al (2007)http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308
5. Benevant et al (2016) http://link.springer.com/article/10.1007%2Fs11192-016-1868-7
6. Leitner et al (2016) http://journal.frontiersin.org/article/10.3389/fnins.2016.00419/full
7. Ioannidis et al (2009) https://www.nature.com/ng/journal/v41/n2/full/ng.295.html
8. Pienta et al (2010) https://deepblue.lib.umich.edu/handle/2027.42/78307
9. Eyding et al (2010) http://www.bmj.com/content/341/bmj.c4737
10. Cochrane Database Syst Rev. 2012 DOI: 10.1002/14651858.CD008965.pub3
11. McKiernan et al (2016) http://dx.doi.org/10.7554/eLife.16800
12. Report finds genomics effort has added US$1 trillion to US economy http://www.nature.com/news/economic-return-from-human-genome-project-grows-1.13187

Source: http://arxiv.org/pdf/1111.3618v1.pdf (1)

SPRINGER NATURE

# More than half of researchers[1] share data – why?

- **97%** - to accelerate research and its applications[2]

- **96%** - increased visibility and discovery of their research data[2,3]

- **95%** - increased usability of their research data[3]

- **>90%** - credit mechanism for deposit of data[2,3]

- **88%** - to comply with funder policy[2]

Motivators to publish data as open data

	Not Important	Intermediate Important	Very Important
Acceleration of scientific research and applications	3%	22%	74%
Dissemination and recognition of your work research	5%	26%	69%
Personal commitment to open data	8%	32%	60%
Requests from data users	6%	34%	59%
Funder policy	12%	37%	51%
Scientific professional society policy	12%	42%	46%
Organizational institutional policy	14%	41%	45%
Community norms	13%	49%	39%
Publisher policy	21%	49%	30%

Response   Not Important   Intermediate Important   Very Important

1. Admin, Wiley (2016): Wiley Data Sharing Survey. figshare. 10.6084/m9.figshare.3468368.v2 (**n= 2250**)
2. Schmidt et al. (2016). PLoS ONE 11(1): e0146695. doi:10.1371/journal.pone.0146695 (**n=1248**) (& image credit, CC BY)
3. Nature Publishing Group (2014): Data publication survey - raw data. figshare. http://dx.doi.org/10.6084/ m9.figshare.1234052 (**n=387**)

SPRINGER NATURE

What challenges do researchers face?

- **64%** unsure about open licensing of research data[1]

- **56%** do not use a metadata standard[2]

- **54%** would like more guidance complying with funder policies[1]

- **54%** do not have enough time to make data available[1]

- **45%** unaware of a repository for some of their data[3]

- **39%** uncertain about meeting costs of making data open[1,2]
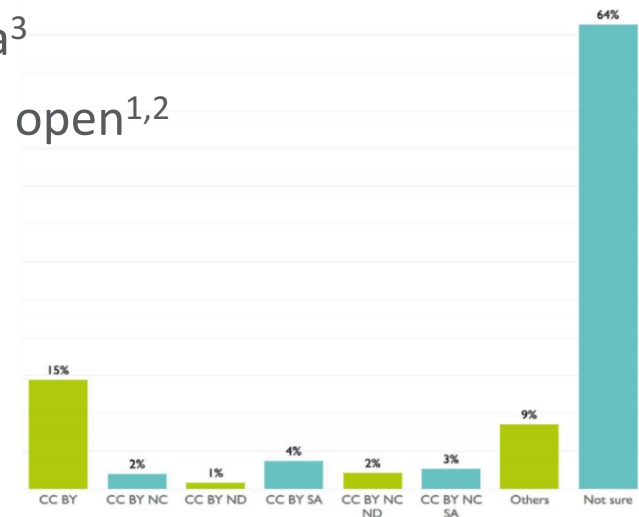
1. Treadway et al. (2016). figshare. https://dx.doi.org/10.6084/m9.figshare.4036398.v1 (**n= 2061**) (& image credit CC BY)
2. Tenopir et al. (2011). PLoS ONE 6(6): e21101. doi:10.1371/journal.pone.0021101 (**n=1315**)
3. Nature Publishing Group (2014). figshare. http://dx.doi.org/10.6084/ m9.figshare.1234052 (**n=387**)

SPRINGER NATURE

What are publishers doing about it?

- **Content types** e.g. data articles and journals

- **Credit and incentives** e.g. data citation and data articles

- **Encouraging reuse** e.g. open licenses

- **Data quality** e.g. data peer review, community standards and repositories

- **Data discoverability** e.g. repository partnerships, linking, integration with submission systems and research data metadata

- **Raising awareness** e.g. editorials, outreach

- **Guidance** e.g. information for authors

- **Policy – *and its implementation***

- Lin J, Strasser C (2014) Recommendations for the Role of Publishers in Access to Data. PLoS Biol 12(10): e1001975. doi:10.1371/journal.pbio.1001975
- Hrynaszkiewicz I, Li P, Edmunds SC. Open science and the role of publishers in reproducible research. In: Stodden V, Leisch F, Peng, RD, editors. Implementing Reproducible Research. CRC Press; 2014. Public (https://osf.io/35s9d/)

SPRINGER NATURE

Journal data policy landscape

- Data sharing encouraged (some Springer journals)

- Data sharing implied as a condition of submission/publication with mandates for specific data types (Nature pre-2016, some Springer journals)

- Mandated data availability statements in every paper and mandates for specific data types (Royal Society, BioMed Central, Palgrave Communications, Nature 2016 –)

- Mandated data sharing for all, with exceptions, with statement in paper (e.g. PLOS)

- Mandated data sharing for all with statement and link to data (e.g. *Molecular Ecology*, *American Economics Review*)

- Mandated open data and data citation as a condition of submission (*Scientific Data, GigaScience, F1000Research*)
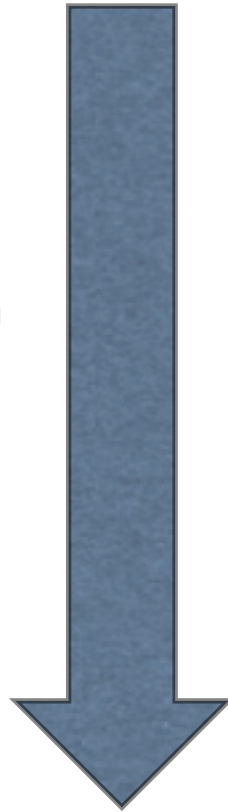
STRONGER

- Vines, T. H. et al. Mandated data archiving greatly improves access to research data. FASEB J. fj.12–218164– (2013). doi:10.1096/fj.12-218164
- Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA (2011) Public Availability of Published Research Data in High-Impact Journals. PLoS ONE 6(9): e24357. doi:10.1371/journal.pone.0024357

SPRINGER NATURE

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 journal

2. Implement standardised 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

Research data questions and issues for publication ethics

- What is a reasonable request/restriction?
- What data are we talking about?
- Corrections e.g. non-compliance with funder data policy
- Versioning and dataset updates
- Data authorship / co-authorship
- Data peer review / data access / reviewer anonymity
- Publishing sensitive data (advertently or inadvertently)
- Retrospective data publication requires correction/update to original article?

Median editor time to add DAS by journal (minutes)
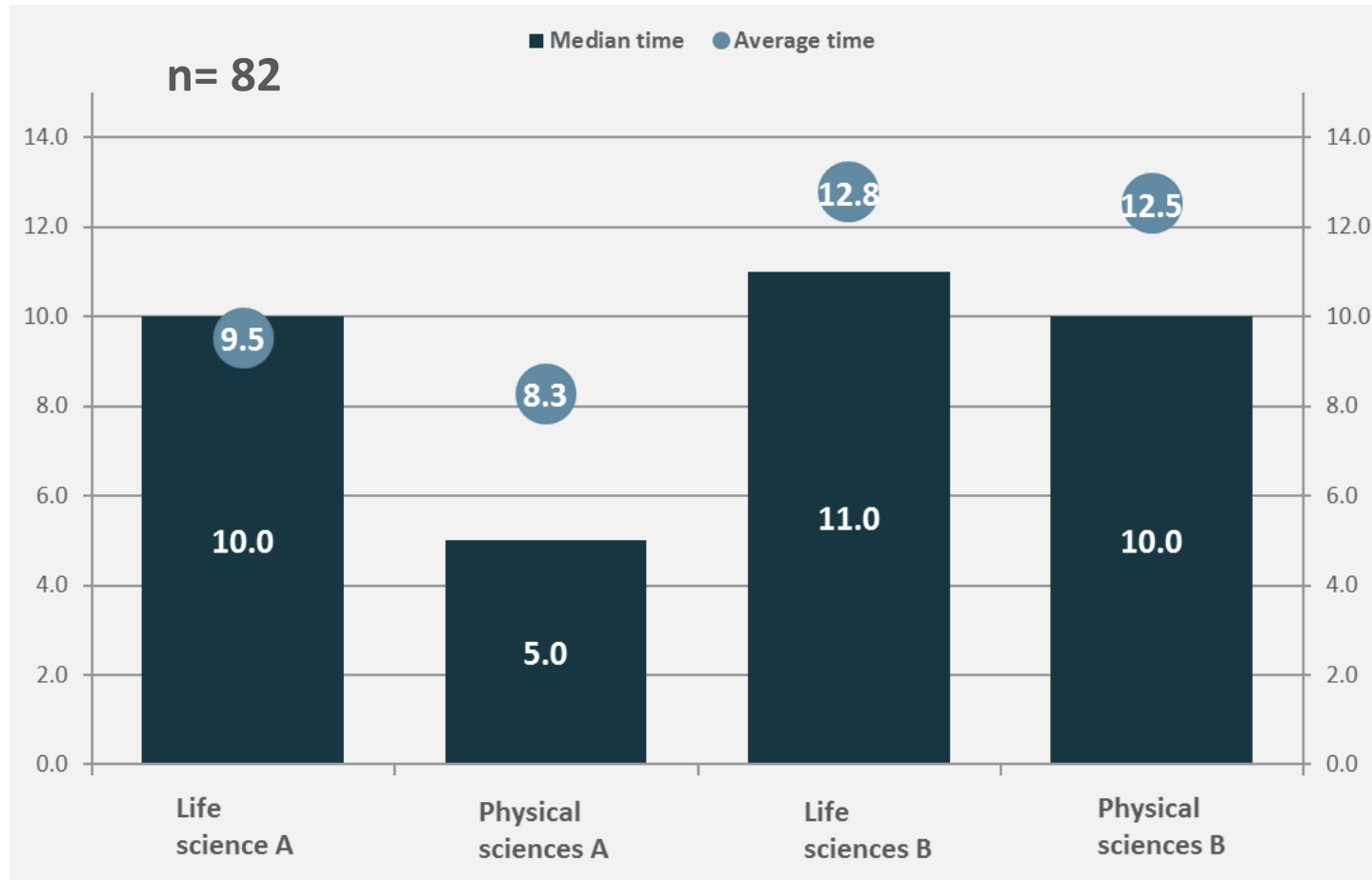
Questions?

iain.hrynaszkiewicz@nature.com

SPRINGER NATURE