

# データ引用の現状とこれまでの進展

The current trends and progress in data citation

JOSS, 27 May 2019

大場郁子

ADVANCING  
**DISCOVERY**

# 現状と今後のトレンドについて

# データ引用とは何か?

- *A reference made to data in the same way as researchers routinely provide a bibliographic reference to journal articles and books*
- When public datasets have Digital Object Identifiers (DOIs), or equivalent identifiers, **it is the same as citing a journal article**
- **For researchers (authors) data citation should be simple and straightforward**
- Include the minimum information recommended by DataCite and follow journal style  
e.g. for *Nature*: authors, title, publisher (repository name), identifier, year

Creator (author)

Title

77. Di Stefano, B., Collombet, S. & Graf, T. Time-resolved gene expression profiling during reprogramming of C/EBP $\alpha$ -pulsed B cells into iPS cells. *figshare* [https://dx.doi.org/10.6084/m9.figshare.939408\\_D1](https://dx.doi.org/10.6084/m9.figshare.939408_D1) (2014).

 [Show context](#)

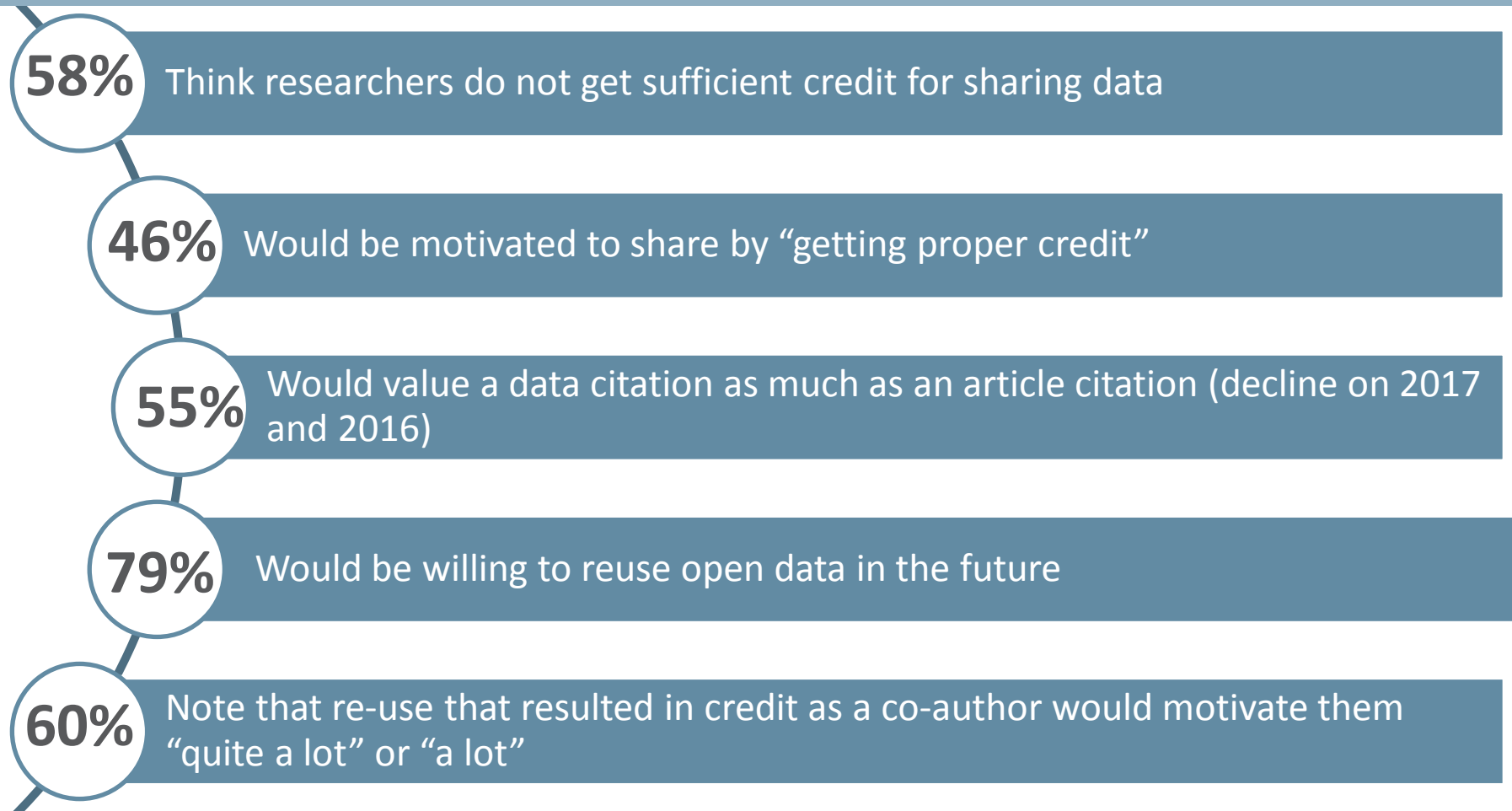
Publisher (repository name)

Identifier (DOI)

Publication year

SPRINGER NATURE

研究者はデータを共有することによりクレジットが得られることを求めており、データ引用によりこれが可能となる



## 助成金団体は研究データとデータ引用を重要視している

UK concordat on open research data, endorsed by funding agencies including Wellcome Trust and the UK research councils: ***“The obligation to recognise through citation and acknowledgement the original creators of the data must be respected”***

<https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf>

**National Science Foundation (US)** *“For all new grant applications from 14 January [2013], the US National Science Foundation (NSF) asks a principal investigator to list his or her research “products” rather than “publications” in the biographical sketch section. This means that, according to the NSF, a scientist’s worth is not dependent solely on publications. Data sets, software and other non-traditional research products will count too.”*

<http://www.nature.com/nature/journal/v493/n7431/full/493159a.html>

# ジャーナルのデータポリシーによりデータ引用を推奨

## Policy Types



- Nearly 1,700 (~70%) Springer Nature journals now have a policy
- **Four types of policy, which increase in their stringency from but all promote data citation in instructions to authors**
- All policies and journals promote data citation in Information for authors
- Similar data policy initiatives since introduced at Elsevier<sup>2</sup>, Wiley<sup>3</sup>, Taylor & Francis<sup>4</sup> and BMJ group

1. *Standardising and harmonising research data policy in scholarly publishing*

Iain Hrynaszkiewicz, Aliaksandr Birukou, Mathias Astell, Sowmya Swaminathan, Amye Kenall, Varsha Khodiyar  
International Journal of Digital Curation; doi: <https://doi.org/10.2218/ijdc.v12i1.531>

2. <https://www.elsevier.com/authors/author-services/research-data/data-guidelines>

3. <https://authorservices.wiley.com/author-resources/Journal-Authors/licensing-open-access/open-access/data-sharing.html>

4. <https://authorservices.taylorandfrancis.com/understanding-our-data-sharing-policies/>

# ジャーナルポリシー事例：ネイチャー・リサーチ

# nature research

## Data availability statements and data citations policy: guidance for authors

### Policy summary

---

All manuscripts reporting original research must include a data availability statement. Authors are also encouraged to include formal citations to datasets in article reference lists where deposited datasets are assigned Digital Object Identifiers (DOIs) by a data repository.

This policy builds upon our long-standing [policy on data availability](#), which requires that authors make materials, data, code, and associated protocols promptly available to readers without undue qualifications. The preferred way to share large data sets is via public repositories. For certain types of data, data sharing is mandatory. This policy does not introduce new data sharing mandates but aims to make the circumstances for data availability more transparent to readers.

<https://www.nature.com/documents/nr-data-availability-statements-data-citations.pdf>

# 論文内のデータ引用事例

Paper published in  
**Nature** (<https://doi.org/10.1038/nature23654>)



Letter | Published: 30 August 2017

## Early members of 'living fossil' lineage imply later origin of modern ray-finned fishes

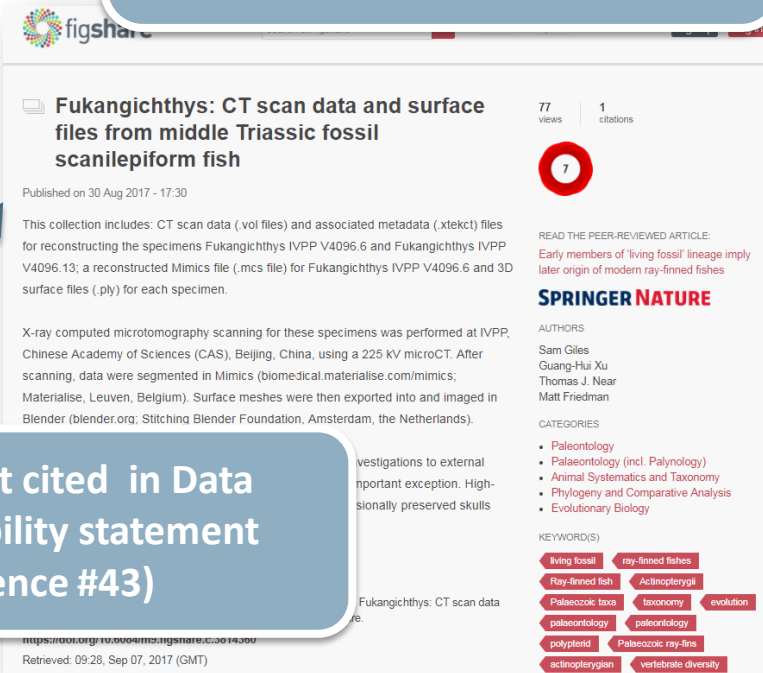
Sam Giles✉, Guang-Hui Xu, Thomas J. Near & Matt Friedman

Nature

### Data availability

The CT data that support the findings of this study, as well as 3D surface files of described material, are available in figshare with the identifier <https://doi.org/10.6084/m9.figshare.c.3814360>. All other data files are included in the Supplementary Information.

Dataset published in the Springer  
**Nature figshare repository**  
(<https://doi.org/10.6084/m9.figshare.c.3814360>)



Dataset cited in Data  
availability statement  
(Reference #43)

Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006)

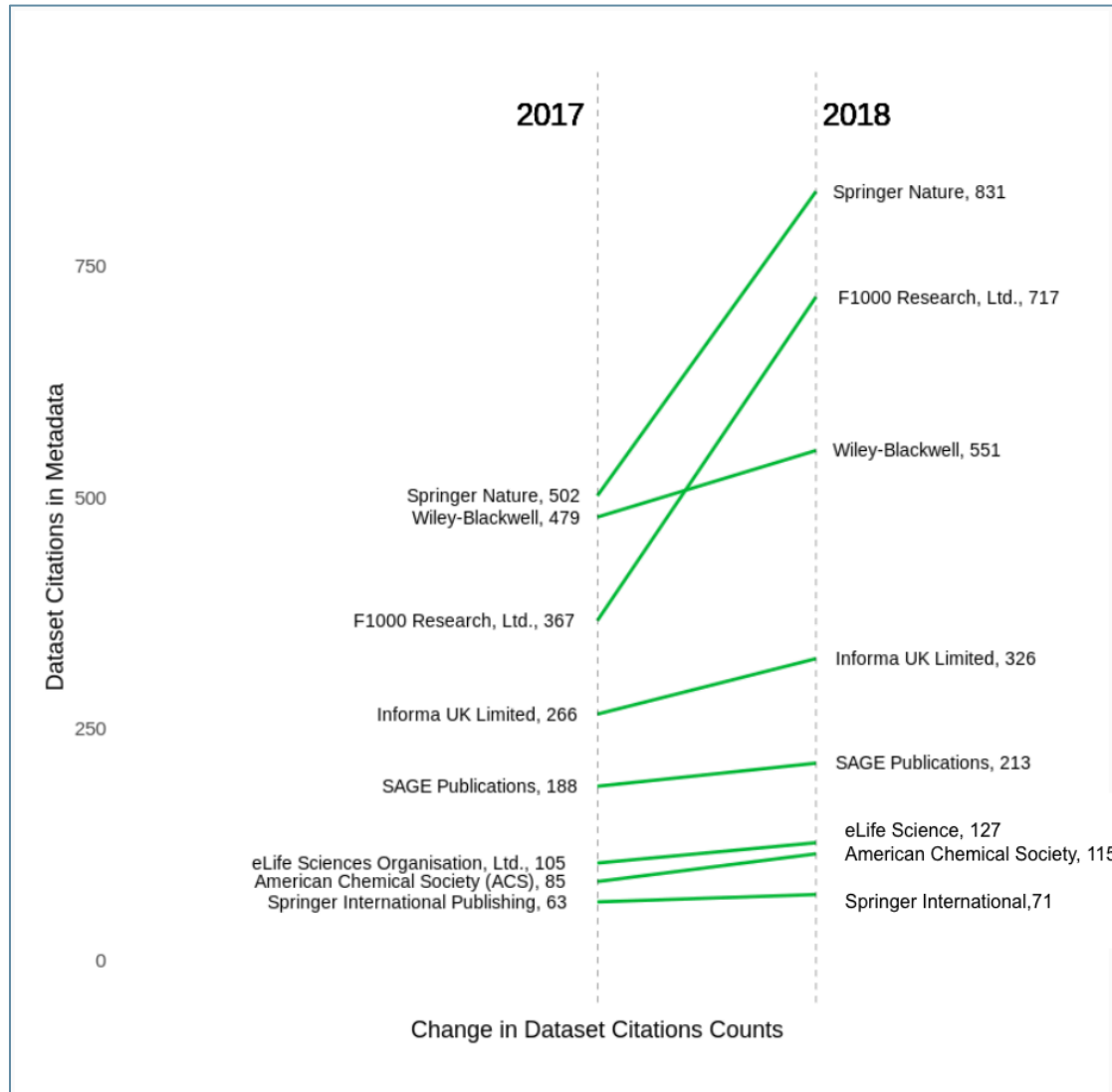
CAS PubMed Article Google Scholar

Giles, S. Xu, G.-H., Near, T. J. & Friedman, M. *Fukangichthys*: CT scan data and surface files from Middle Triassic fossil scanilepiform fish. <https://doi.org/10.6084/m9.figshare.c.3814360> (2017)

Google Scholar



# ジャーナルのポリシーによりデータ引用は増えているのか？



- According to data from DataCite and CrossRef, yes
- But progress has been slow and only a small proportion of papers include data citations in the reference lists
- More commonly and traditionally, data are referenced in the main text rather than reference list

# データ引用の認知度を向上するための取り組み

# 業界・コミュニティ内でジャーナルのデータ・引用ポリシーを標準化するための取り組み


Co-chairs:



Natasha Simons (ANDS), Simon Goudie (Wiley), Azhar Hussain (Jisc), Iain Hrynaskiewicz (Springer Nature)

- A group in the Research Data Alliance has defined a common framework for research data policy for all journals and publishers
- Includes support for data citation as a mandatory feature in all policies

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



[ABOUT RDA](#)
[GET INVOLVED](#)
[GROUPS](#)
[RECOMMENDATIONS & OUTPUTS](#)
[RDA FOR D](#)

## Data policy standardisation and implementation

[Home](#) » [Working And Interest Groups](#) » [IG](#)

IG


### Group details

**Status:** Recognised & Endorsed

**Chair (s):** Iain Hrynaskiewicz, Natasha Simons, Simon Goudie, Azhar Hussain

**Secretariat Liaison:** Kathy Fontaine

**TAB Liaison:** Devika Madalli

 IG Established

[History](#)

---

### Introduction

Increasing the availability of research data for reuse is in part being driven by research data policies and the number of funders and journals and institutions with some form of research data policy is growing. The research data policy landscape of funders, institutions and publishers is however too complex (Ref: <http://insights.uksg.org/articles/10.1629/uksg.284/>) and the implementation and implications of policies for researchers can be unclear. While around half of researchers share data, their primary motivations are often to carry out and publish good research, and to receive renewed funding, rather than making data available. Data policies that support publication of research need to be practical and seen in this context to be effective beyond specialist data communities and publications.

### Use cases and user scenarios

# 大規模な機械可読データ引用を実現するための出版社間 コラボレーション

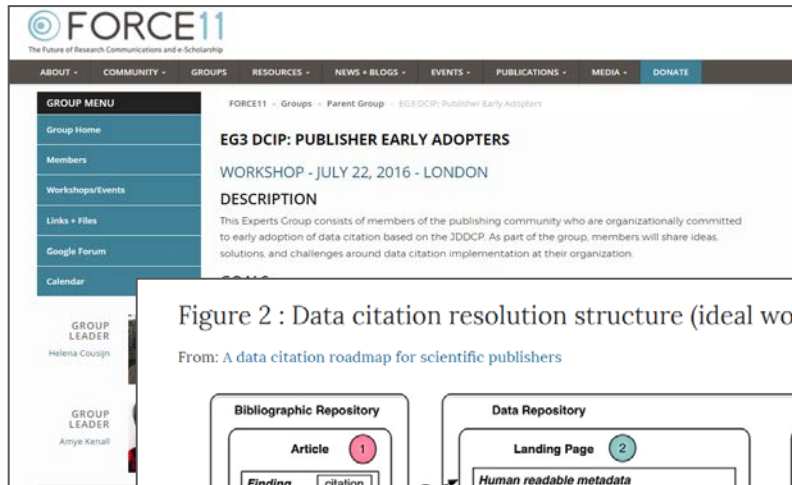
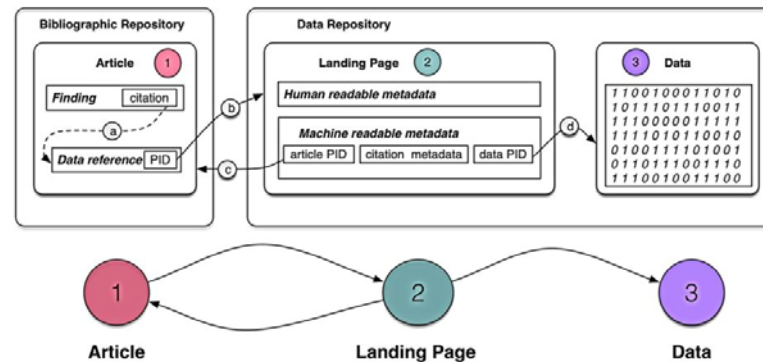


Figure 2 : Data citation resolution structure (ideal workflow).

From: A data citation roadmap for scientific publishers



Articles (1) link to datasets in appropriate repositories, on which their conclusions are based, through citation to a dataset (a), whose unique persistent identifier (PID) resolves (b) to a landing page (2) in a well-supported data repository. The data landing page contains human- and machine-readable metadata, to support search and to resolve (c) back to the citing article, and (d) a link to the data itself (3).

- Working group of publishers defined a Publishers' Data Citation Roadmap to aid in implementation of data citation in articles and books
- Document covers policy, author guidelines and **content production and technical issues**
- Many publishers collaborating with infrastructure providers (repositories, DataCite, CrossRef) and policy makers to take a common approach
- An aim is to keep the process simple for researchers (authors) while delivering benefits of better measurement of credit and reuse of data by improving content (XML) structure

**A Data Citation Roadmap for Scientific Publishers.** Helena Cousijn, Amye Kenall, Emma Ganley, Melissa Harrison, David Kernohan, Fiona Murphy, Patrick Polischuk, Maryann Martone, Timothy Clark  
*Scientific Data* **volume5**, Article number: 180259 (2018)

<https://www.force11.org/group/dcip/eg3publisherearlyadopters>

# 出版社によるデータ引用導入のタイムライン

Table 2: Estimated data citation implementation timelines for eight academic publishers.

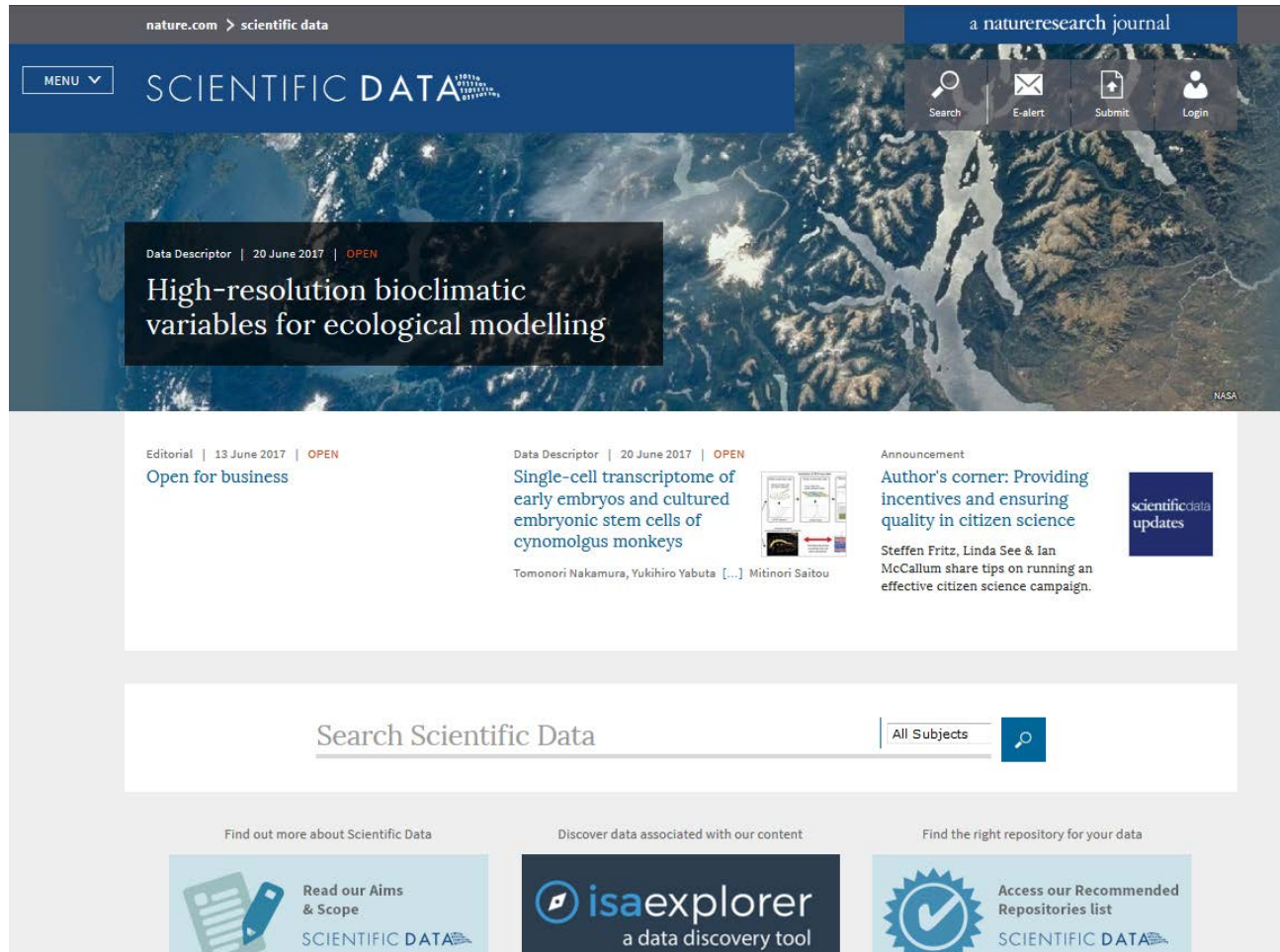
From: [A data citation roadmap for scientific publishers](#)

Publisher	Planning	Implementation	Planned go-live date
eLife	Q1–Q3 2017	Q1–Q3 2018	Live
Elsevier	Q2–Q3 2016	Q4 2016	Live
EMBO Press	Q1–Q2 2017	Q3 2017–Q1 2018	Live
Frontiers	Q1–Q4 2017	Q1– Q2 2018	Live
PLOS	Q1–Q4 2017	2018	Q1/Q2 2019
SpringerNature	2016–2017	Q2 2017–Q4 2018	Q4 2018
Taylor & Francis	Q1–Q2 2017	Q4 2017 continuing through 2018	Live
Wiley	Q1–Q2 2017	Q4 2017 continuing through 2018	2018

NB. The data citation rollout at the given publishers will be in line with their respective data policies and will apply to all journals whose content is based on a dataset or which references datasets.

*Meanwhile, researchers are encouraged to cite in their reference lists any publicly available datasets they generate or analyse in their studies*

# データ引用の最先端事例 : *Scientific Data*



Part of the Nature Research Group, *Scientific Data* is an open-access journal for descriptions of scientifically valuable datasets.


- Broad scope covering physical, life and quantitative social sciences
- Data-focused peer-review process
- Supports community data repositories
- Integrated submission of data to general repositories
- Data deposition and data citation are mandatory for submission


# データ引用のベストプラクティス事例 : *Scientific Data*

SCIENTIFIC DATA

Data Descriptor | OPEN | Published: 26 March 2019

## Field-recorded data on habitat, density, growth and movement of *Nephrops norvegicus*

Anne Marie Power , Julian Merder, Patricia Browne, Jan A. Freund, Liam Fullbrook, Conor Graham, Robert J. Kennedy, Jack P. J. O'Carroll, Alina M. Wiczorek & Mark P. Johnson

Scientific Data **6**, Article number: 7 (2019) | Download Citation 

- Data citations are both human readable (pictured) and machine readable
- Publisher XML includes a specific reference item type/tag to indicate which references are for datasets, rather than papers/articles/chapters/websites
- Enables downstream delivery and consumption of data citations and links, and enhancement to the reader experience with data links, for example
- Helps users, such as institutions, infrastructure providers, researchers understand and measure data reuse

### Data Records

Detailed explanation of dataset “**N. norvegicus sampling**”<sup>21</sup> is given in Table 1.

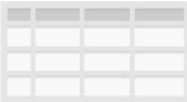


Table 1 Data on the size, growth, distance moved and catch information for *N. norvegicus*

Full size table >>

The first dataset (**N. norvegicus sampling**<sup>21</sup>) consists of:

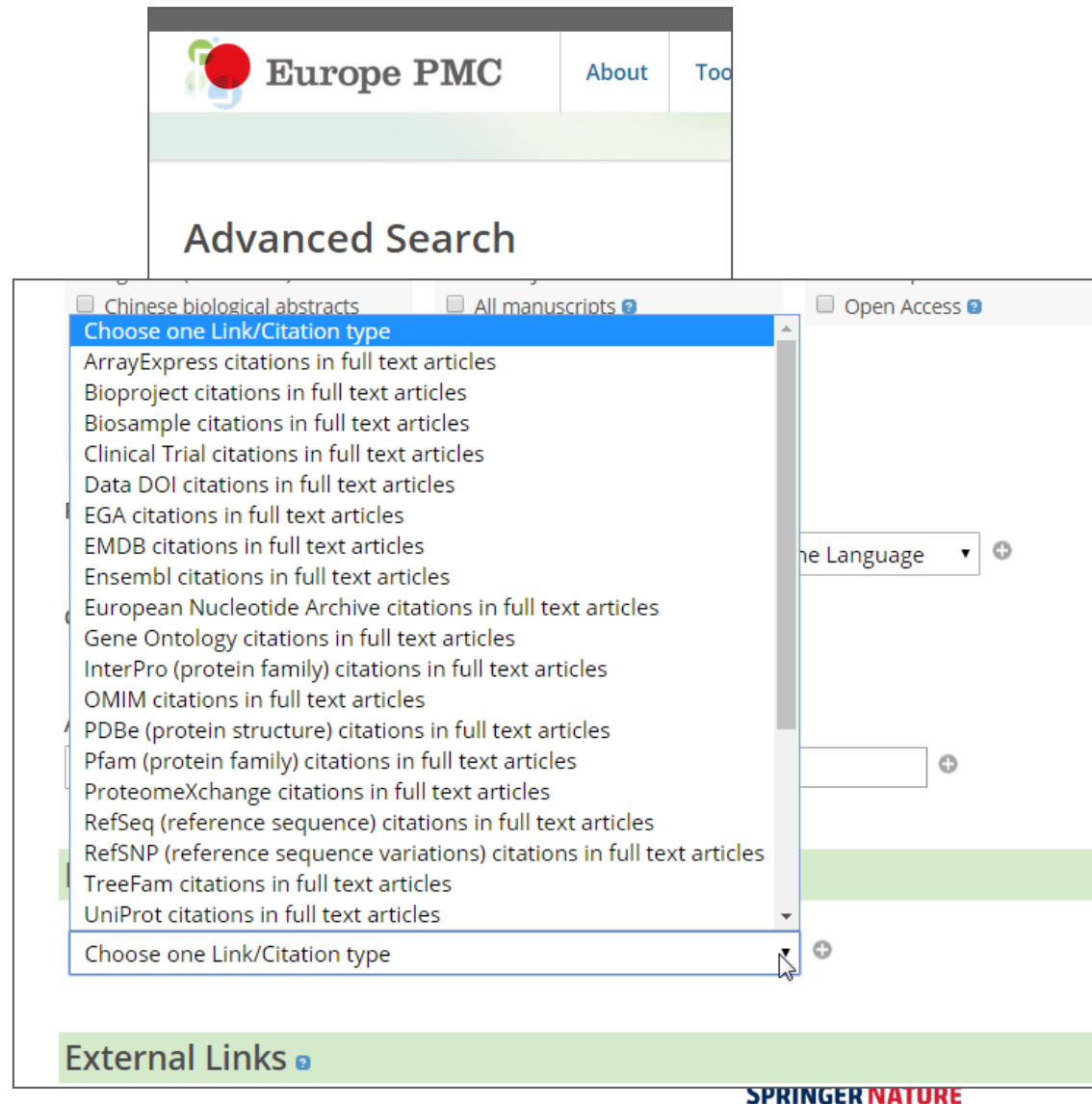
Sections | Figures | **References**

21. Power, A. M. et al. Field-recorded data on habitat, density, growth and movement of *Nephrops norvegicus*. figshare. <https://doi.org/10.6084/m9.figshare.c.4249163> (2019).
22. Lincoln, F. C. Calculating waterfowl abundance on the basis of banding returns. *U.S. Department of Agriculture Circ.* **118**, 1–4 (1930).
23. Petersen, C. G. J. The yearly immigration of young plaice into the Limfjord from the German Sea. *Report of the Danish Biological Station for 1895* **6**, 5–48 (1896).
24. QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org> (2018).
25. BioSonics Inc. *BioSonics*, <http://www.biosonicsinc.com> (2018).
26. Eleftheriou, A. & McIntyre, A. *Methods for the Study of Marine Benthos*. third ed (Blackwell Scientific Publications, Oxford, 2005).



# データー論文間リンクを向上するために、出版社と基盤提供者が協力

- Implementaiton of the Scholix guidelines enables connections between articles and datasets across publishers and repositories.
- Article-data links should be exchanged in a dynamic and standardized way, enhancing discoverability of all contributors' content.
- Information about these links is stored in open hubs (e.g. Crossref, DataCite) so other systems can extract relevant connections.
- Can be used to enhance functionality of literature databases e.g. EU PubMed Central
- <http://www.scholix.org/guidelines>





# ご清聴ありがとうございました。

大場郁子

[ikuko.oba@nature.com](mailto:ikuko.oba@nature.com)

シュプリンガー・ネイチャーにおける研究データサポートやその他の取り組みについては、以下の窓口までお問い合わせください。

Email: [researchdata@springernature.com](mailto:researchdata@springernature.com)

Website: <http://go.nature.com/ResearchDataServices>

Slide acknowledgements:

Rebecca Grant, Varsha Khodiyar, Iain Hrynaszkiewicz

The story behind the image



## Chien Shiung Wu (1912–1997)

Chien Shiung Wu was a Chinese American experimental physicist best known for conducting The Wu experiment that bears her name. This experiment showed that the conservation of parity was violated by a weak interaction and it was possible to distinguish between a mirrored variation of the world and the mirror image of the current world. This discovery earned Wu the Wolf Prize in Physics in 1978.

This presentation is licensed as CC-BY-ND

**SPRINGER NATURE**