

Publisher perspective – Data Policies and Data Accessibility Statements

Laura Armstrong, Research Services





Research data

What is research data?

 everything that is collected, observed or created for the purposes of analysis that underpins a research output (e.g. publication or creative work).¹

Why publish data?

- research/er integrity, institutional obligation, funder requirement, publisher requirement, increase citation and dissemination, research impact, enable collaboration, encourage innovation and reuse, preservation, use in teaching, public record...

^{1.} University of Auckland Research Hub > <u>Guide to Managing Research Data</u>



Thanks to Rebecca Grant, Research Data Manager at Springer Nature, for permission to reuse slides from the Council of Australian University Librarians (CAUL) webinar, May 2018.

Individual slides taken from Rebecca's presentation are unchanged here and explains why there are two styles of branding present within this presentation.

Due to use of Springer Nature's slides this presentation is CC-BY-ND.

Contact: rebecca.grant@springernature.com

SPRINGER NATURE

Springer Nature is a leading research, educational and professional publisher, providing quality content to our communities through a range of innovative platforms, products and services.

Home to brands including Springer, Nature Research, BioMed Central, Palgrave Macmillan and Scientific American.

As the leading open access publisher, we see the rise of open research in all its manifestations as one of the major forces reshaping the way that researchers communicate and collaborate to advance the pace and quality of discovery.

Our focus is on investing in and creating tools, services or training that help the research community to understand and utilise new ideas and concepts.

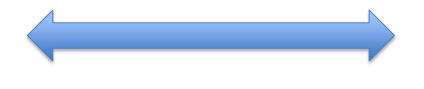


Data Policy

Publishers state (website, journal submission process) to what degree and how they encourage - require authors to share their data at the review - post-publication stage.

Type or strength of policy differs journal, publisher and discipline.

Contact author, embargo



Required at review, correction issued

^{1.} University of Auckland Research Hub > <u>Guide to Managing Research Data</u>

THE CASE FOR RESEARCH DATA PUBLISHING: ENHANCING REPRODUCIBILITY Data availability has been shown to enable reproducibility

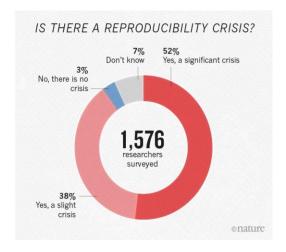
Evidence is mounting on costs & scale of the issue

The Economist

The How to do a nuclear deal with Iran Investment tips from Nobel economists Junk bonds are less than the meaning of Sachin Tendulkar

- Irreproducible biology research costs US \$28 billion per year¹
- Pharma companies report 75%+ failure rates replicating conclusions of peer-reviewed papers ^{2,3}

A recent *Nature* survey⁴ highlights concern in the research community



>50% of researchers couldn't reproduce their own experiments

>70% couldn't reproduce the work of others

There is evidence that data availability increases reproducibility

A study⁵ of eighteen *Nature Genetics* papers found :

- Two could be reproduced fully
- Six were reproduced partially
- Ten could not be reproduced

"The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis."

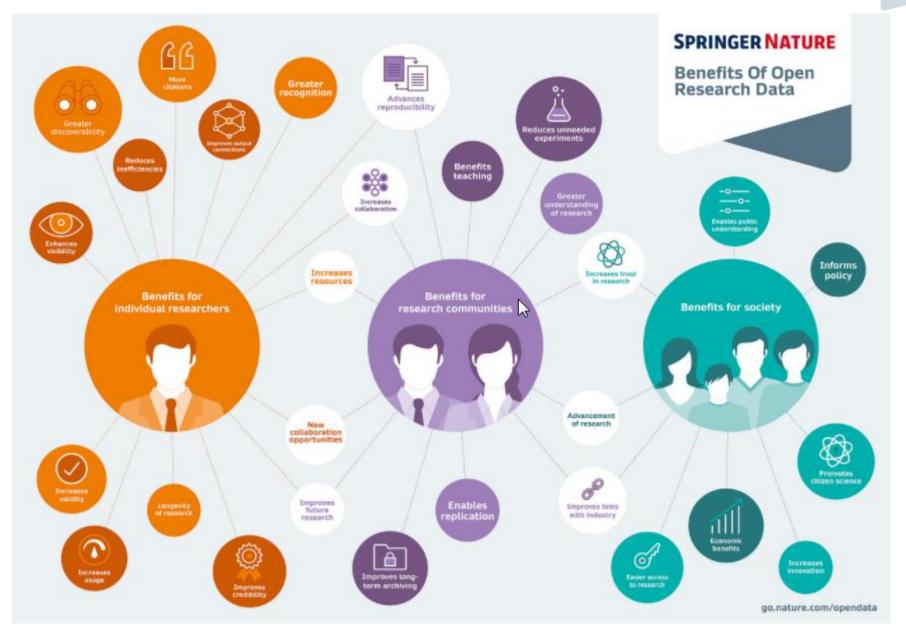
- Nature Genetics **41**, 149–155 (2009)

^{1.} Freedman, L. P., Cockburn, I. M. & Simcoe, T. S. PLoS Biol. 13, e1002165 (2015) http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1002165

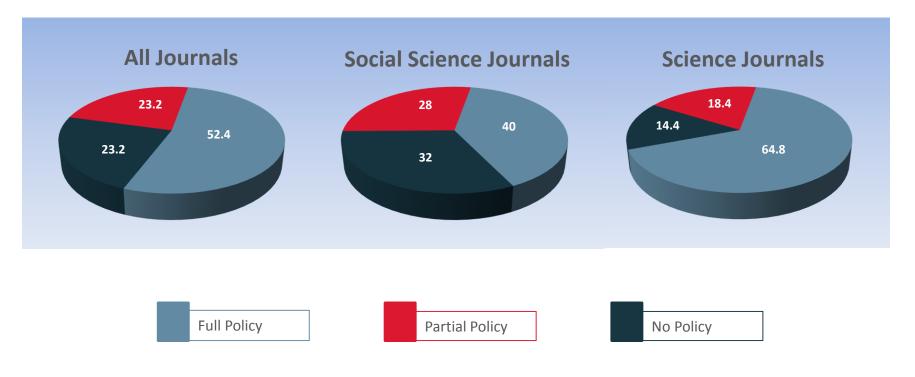
^{2.} Begley, C. G. & Ellis, L. M. Nature 483, 531–533 (2012), 3. Prinz, F., Schlange, T. & Asadullah, K. Nature Rev. Drug Discov. 10, 712 (2011)

^{4.} Baker (2015) http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

^{5.} loannidis et al (2009) https://www.nature.com/ng/journal/v41/n2/full/ng.295.html



Understanding journal data policies is difficult



Data source: Linda Naughton, JISC Journal Research Data Policy Bank project presentation (n = 250)

"The evidence shows that the current research data policy ecosystem is in critical need of standardization and harmonization"

Springer Nature research data policy initiative

Policy Types

Type 1

Data sharing and data citation is encouraged but not required

Type 2

Data sharing and evidence of data sharing encouraged

Type 3

Data sharing encouraged and statements of data availability required

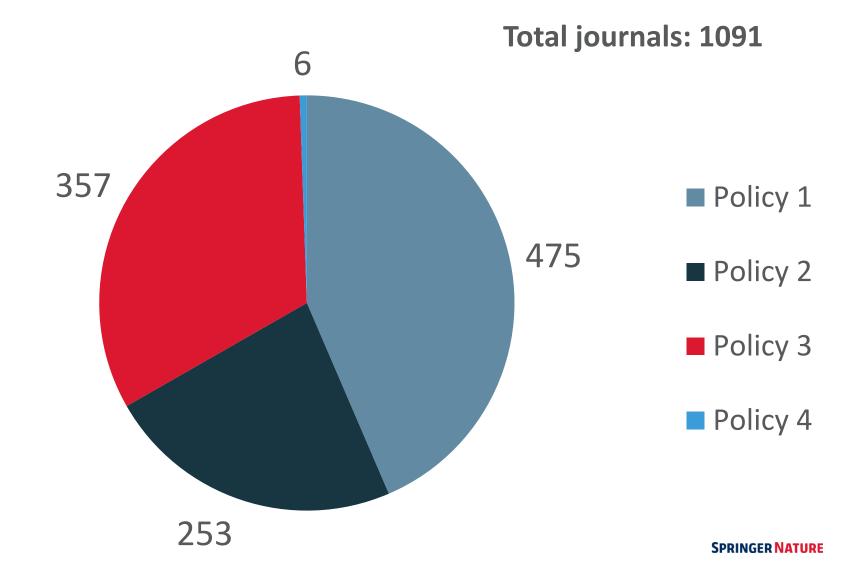
Type 4

Data sharing, evidence of data sharing and peer review of data required

- Over 1,400 (>50%) Springer Nature journals have adopted a standard research data policy as of May 2018
- Practical and pragmatic
- Preference data archiving in repositories over supplementary information
- Support community specific policies, mandates and repositories
- Promote use of data availability statements
- Promote data citation
- Offer support from the publisher via our helpdesk

https://www.springernature.com/gp/authors/research-data-policy/journal-policies/15369670

Policy adoption across journals





SPRINGER NATURE

Q Search

EN Y





Submitting to Research Data Support

Who can use Research Data Support

Pricing for Research Data Support

Benefits of Research Data Support

Help & FAQs

Data policy types

Data availability statements

Data policy FAQs

Research Data Helpdesk

Recommended Repositories

Journal policies & services

Find Your Journal's Data Policy and Services

Below is the list of Springer Nature journals that have adopted a standardised research data policy, including information on which data policy type they have implemented and if they currently provide Springer Nature Research Data Support through the manuscript submission system.

If your journal does not currently provide Research Data Support during manuscript submission, you can still access the service through this form.

Journal name	Policy type	Research Data Support
3D Printing in Medicine	3	Available at publication
AAPS Open	2	Available at publication
Academic Questions	1	Available at publication
Accreditation and Quality Assurance	1	Available at publication

https://www.springernature.com/qp/authors/research-data-policy/springer-nature-journals-data-policy-type/12327134

Type 4



SCIENTIFIC DATA

Search | Submit | Login

Policies

Editorial & Publishing Policies

For Referees

Data Policies

Recommended Data Repositories

Data Policies

Data Descriptors, Scientific Data's primary article type, describe scientifically valuable datasets. These datasets must be made available to editors and referees at the time of submission, and must be shared with the scientific community as a condition of publication. Here, we provide information on the types of data that should be archived, guidance for authors on selecting a suitable repository for their data, and how to archive sensitive data.

Scientific Data's data policies are compatible with the standardised research data policies set out by Springer Nature.

Please read on for our data deposition policies, and please contact us if you would like additional advice on how best to meet these requirements for your own data.

Research Data Alliance Interest Group

Co-chairs:









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

Primary objective is to help define common frameworks for research data policy – starting with journals and publishers and latterly funding agencies



ABOUT RDA GET INVOLVED GROUPS RECOMMENDATIONS & RDA FOR DISC

Data policy standardisation and implementation

Home » Working And Interest Groups » Inte

IG

Group details

Status: Recognised & Endorsed

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

Secretariat Liaison: Kathy Fontaine
TAB Liaison: Devika Madalli

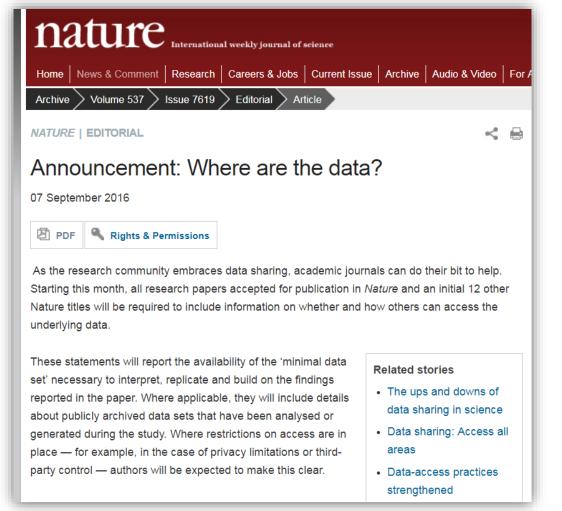
⊘ IG Established

—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



Two phases of research were undertaken to analyse the impact of this new policy.

The aim of this research was to assess the ways by which researchers chose to make their data available, and to measure the additional time required by editors and production staff to add data availability statements to manuscripts (which gives an indication of cost to the publisher).

http://www.nature.com/news/announ cement-where-are-the-data-1.20541

What is a data availability statement (DAS)?

A statement about where data supporting the results reported in the article can be found:

- The datasets generated during and/or analysed during the current study are available in the [NAME] repository, [PERSISTENT WEB LINK TO DATASETS].
- The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.
- All data generated or analysed during this study are included in this published article (and its supplementary information files).

Required by many journals/publishers e.g. PLOS, Royal Society, BMJ, Hindawi and funders including RCUK

A common disadvantage is the need for power during heating and reaction time.

Conclusion

Filter paper is already a well-known field support method as it allows the collection, transfer, and storage of samples without a cold chamber. Considering the designed-in field-friendliness of this PURE–LAMP combination method, use of this system with dried blood samples will allow maximum exploitation of this characteristic while still performing accurately.

Cost-effectiveness is beyond the scope of this paper, but technically this system is predicted to be extendable to the field, while it is effective for malaria diagnosis in travellers.

Additional file

Additional file 1. List of samples analysed: Combination of PURE-DNA extraction and LAMP-DNA amplification methods for accurate malaria diagnosis on dried blood spots. PURE-LAMP: Procedure for ultra rapid extraction–loop-mediated isothermal amplification; Pan: *Plasmodium* genus; Pf: *Plasmodium falciparum*; Pv: *P. vivax*; Pm: *P. malariae*; Po: *P. ovale*; Poc: *P. o. curtisi*; Pow: *P. o. wallikeri*; RBC: red blood cells; T1: band of the rapid test containing the antibody for detection of a *Plasmodium falciparum*-specific HRP2 antigen; T2: band containing the antibody for detection of a *Plasmodium* common aldolase antigen. +: positive; -: negative; \pm : positive (pale band). *DNA extracted from 200 μL of fresh blood or 100 μL of frozen RBC concentrate, nested PCR method 1 [19]. **DNA extracted from 200 μL of fresh blood or 100 μL of frozen RBC concentrate, nested PCR method 2 [20]. ***DNA extracted from 3 dried blood spots of φ 3 mm, nested PCR method 2 [20].

Availability of data

The dataset generated during this study is included in this published article and its Additional file.

Ethics approval and content to participate

The study protocol was carefully reviewed and approved by the NCGM Ethics Committee (No. 1111). All subjects submitted a signed consent form.

Funding

This study was partly supported by an AMED grant of the SATREPS project for the "Development of innovative research technique in genetic epidemiology of malaria and other parasitic diseases in the Lao PDR for containing their expanding endemicity".

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 30 May 2018 Accepted: 15 October 2018

Published online: 22 October 2018

References

- WHO. World malaria report 2017. Geneva: World Health Organization; 2017.
- Gilles HM, Warrell DA. Bruce-Chwatt's essential malariology. 3rd ed. London: Edward Arnold; 1993.
- Hänscheid T. Diagnosis of malaria: a review of alternatives to conventional microscopy. Clin Lab Haematol. 1999;21:235–45.
- Abdul-Ghani R, Al-Mekhlafi AM, Karanis P. Loop-mediated isothermal amplification (LAMP) for malarial parasites of humans: would it come to clinical reality as a point-of-care test? Acta Trop. 2012;122:233–40.
- Wongsrichanalai C, Barcus MJ, Muth S, Sutamihardja A, Wernsdorfer WH. A review of malaria diagnostic tools: microscopy and rapid diagnostic test (RDT). Am J Trop Med Hyg. 2007;77(Suppl 6):119–27.
- McMorrow ML, Aidoo M, Kachur SP. Malaria rapid diagnostic tests in elimination settings—can they find the last parasite? Clin Microbiol Infect. 2011;17:1624–31.
- Moody A. Rapid diagnostic tests for malaria parasites. Clin Microbiol Rev. 2002;15:66–78.

Type 4: Malaria Journal 2018 17:373 https://doi.org/10.1186/s12936-018-2527-7

7. Harrischeid 1, Globusch Mr. How useruns rich in die diagnosis d

The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health resea...

Daniel John Exeter, Jinfeng Zhao, Sue Crengle, Arier Lee, Michael Browne





Abstract

Introduction

Data and methods

Results

Geographical variations in multiple deprivation in New Zealand

Discussion

Supporting information

Acknowledgments

Author Contributions

References

Reader Comments (0) Media Coverage (0) Figures Citation: Exeter DJ, Zhao J, Crengle S, Lee A, Browne M (2017) The New Zealand Indices of Multiple Deprivation (IMD): A new suite of indicators for social and health research in Aotearoa, New Zealand. PLoS ONE 12(8): e0181260. https://doi.org/10.1371/journal.pone.0181260

Editor: Isil Ergin, Ege Universitesi, TURKEY

Received: November 29, 2016; Accepted: June 22, 2017; Published: August 3, 2017

Copyright: © 2017 Exeter et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: All research outputs related to this publication including the Data Zone concordance tables and the Data Zone-level ranks of The Index of Multiple Deprivation (IMD) and its Domains are freely available for download from https://figshare.com/articles/Data_Zones_for_New_Zealand/5207884 and https://figshare.com/articles/The_Index_of_Multiple_Deprivation_IMD_/5207887 respectively, or from the IMD project's website http://www.fmhs.auckland.ac.nz/imd. 2013 Census Meshblock files used as the building blocks to develop Data Zones can be obtained from http://www.statistics.govt.nz/browse_for_stats/Maps_and_geography /Geographic-areas/digital-boundary-files.aspx#censusbased. Access to the raw data used to develop the indicators used in the development of the IMD are available from the Integrated Data Infrastructure (IDI), which is maintained by Statistics New Zealand, subject to ethical approval, for researchers who meet the criteria for access to confidential data. Information regarding access to the IDI is available from http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx or by contacting

access2microdata@stats.govt.nz<mailto:access2microdata@stats.govt.nz>. Some data used in the Access Domain of the IMD were obtained from http://koordinates.com/layer/152-nz-petrol-stations/ and https://koordinates.com/layer/189-nz-supermarkets/. Further information about data availability can be obtained from the senior author, Dr Daniel Exeter (d.exeter@auckland.ac.nz<mailto:d.exeter@auckland.ac.nz>) and from http://www.fmhs.auckland.ac.nz/imd.

New Zealand

Socioeconomic asp...

Housing

Health services rese...

Crime

Employment

Social research

C



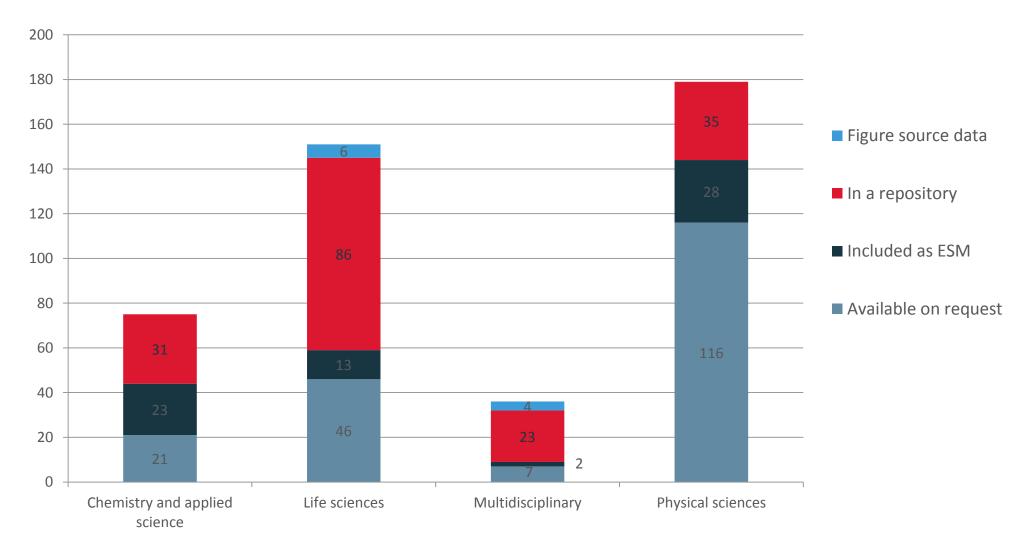


4 Aug 2017

Daniel Exeter @djexeter @djexeter just published #PLOSONE The New Zealand Indices of Multiple Deprivation (IMD)... https://t.co/3MUfKcXkLI #imd #datazones #biqdata



How are different disciplines sharing their data?



The impact on authors and editors of introducing Data Availability Statements at Nature journals Rebecca Grant, Iain Hrynaszkiewicz; doi: https://doi.org/10.1101/264929

Find the best repository for your data

Biological sciences

Chemistry & chemical biology

Earth & environmental sciences

Generalist repositories

https://tinyurl.com/y8hodaon

Health sciences

Physics, astrophysics & astronomy

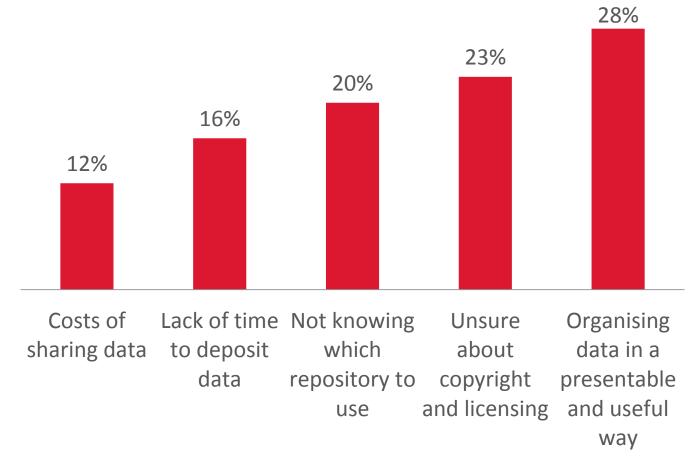
Social science

Other repositories



What problems do authors have in sharing datasets?



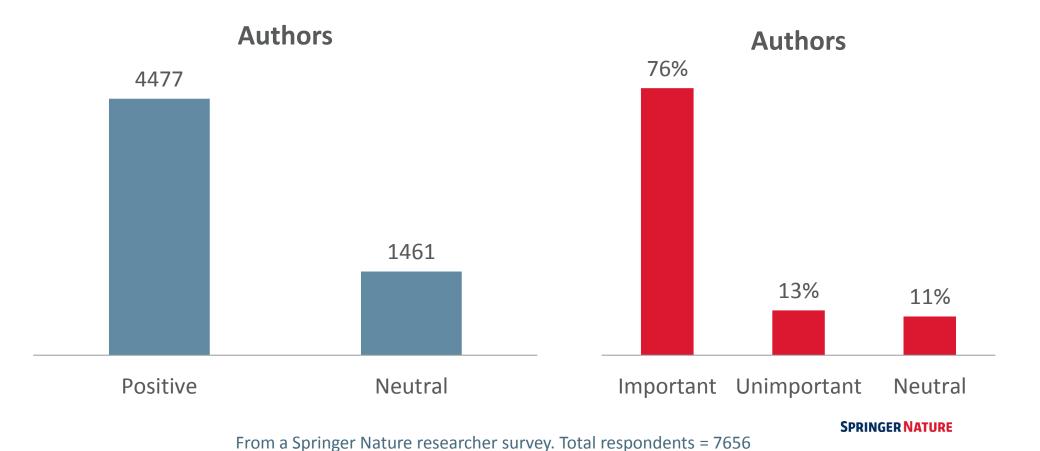


From a Springer Nature researcher survey. Total respondents: 7719

SPRINGER NATURE

How interested are authors in a service that helps them to deposit their data in a repository?

How important is data discoverability to authors?



Submitting a dataset: the researcher uploads the file through our secure figshare submission portal

Add files to your Springer Nature submission		
Springer Nature Research Data Support will help you prepare your research data (including code, text, raw and processed data, video, audio and images) for sharing and reuse. Have a question? email researchdata@springernature.com	ı	
The cost of Research Data Support is €300 per accepted submission, which includes up to 50 GB of permanent online storage. View more information on the services. By submitting your files to Research Data Support you are accepting responsibility for payment. For guidance on completing the submission process, view our tutorial page.		
Upload your file(s). If you have multiple files, add all your files at once		
drag and drop your file(s) here or Browse Researchers with their n	s _l	

Researchers provide us with their name and contact information, and choose an appropriate licence for their dataset

Example output of Research Data Support

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



NATURE | LETTER

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

Sam Giles, Guang-Hui Xu, Thomas J. Near & Matt Friedman Affiliations | Contributions | Corresponding author

Nature (2017) | doi:10.1038/nature23654 Received 05 December 2016 | Accepted 18 July 2017 | Published onl

🖺 PDF 👲 Citation 🔍 Rights & permissions 📈 Article metrics

Modern ray-finned fishes (Actinopterygii) comprise half of extant vi are widely thought to have originated before or near the end of the (around 385 million years ago) 1, 2, 3, 4. Polypterids (bichirs and rope earliest-diverging lineage of living actinopterygians, with almost all interpreted as more closely related to other extant actinopterygians

7, 8, 9, 10. By contrast, the earliest material assigned to the polypterid lineage is mid-Cretaceous in age (around 100 million years old)¹¹, implying a quarter-of-a-billion-year palaeontological gap. Here we show that scanilepiforms, a widely distributed radiation from the Triassic period (around 252-201 million years ago), are stem polypterids. Importantly, these fossils break the long polypterid branch and expose many supposedly primitive features of extant polypterids as reversals. This shifts numerous Palaeozoic rayfins to the actinopterygian stem, reducing the minimum age for the crown lineage by roughly 45 million years. Recalibration of molecular clocks to exclude phylogenetically reassigned Palaeozoic taxa results in estimates that the actinopterygian crown lineage is about 20-40 million years younger than was indicated by previous molecular analyses 1, 2, $^{3,\,4}$. These new dates are broadly consistent with our revised palaeontological timescale and coincident with an interval of conspicuous morphological and taxonomic diversification among ray-fins centred on the Devonian–Carboniferous boundary 12, 13, 14. A shifting timescale, combined with ambiguity in the relationships of late Palaeozoic actinopterygians, highlights this part of the fossil record as a major frontier in understanding the evolutionary assembly of modern vertebrate diversity.

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 43 with the identifier

https://doi.org/10.6084/m9.figshare.c.3814360. All other data files are included in the Supplementary Information.

References

Main · Methods · References · Acknowledgements · Author information · Extended data figures and tables · Supplementary information · Comments

1. Hurley, I. A. et al. A new time-scale for ray-finned fish evolution. Proc. R. Soc. Lond. B 274, 489-498 (2007)

> Data availability statement included with the paper

Dataset published in the Springer **Nature figshare repository** (https://doi.org/10.6084/m9.figshare.c.3 814360)

Fukangichthys: CT scan data and surface files from middle Triassic fossil scanilepiform fish

Published on 30 Aug 2017 - 17:30

fia**sha**

This collection includes: CT scan data (vol files) and associated metadata (xtekct) files for reconstructing the specimens Fukangichthys IVPP V4096.6 and Fukangichthys IVPP V4096.13: a reconstructed Mimics file (.mcs file) for Fukangichthys IVPP V4096.6 and 3D

my of Sciences (CAS), Beijing, China, using a 225 kV microCT, Afte vere segmented in Mimics (biomedical.materialise.com/mimics; ven Belgium). Surface meshes were then exported into and imaged in r.org; Stitching Blender Foundation, Amsterdam, the Netherlands).

orm fossils are heavily compressed, limiting investigations to external liddle Triassic Fukangichthys represents an important exception. Highcomputed tomography (µCT) of three-dimensionally preserved skulls

uang-Hui; Near, Thomas J.; Friedman, Matt (2017): Fukangichthys: CT scan data om middle Triassic fossil scanilepiform fish, figshare.

0.6084/m9.figshare.c.3814360

Sep 07 2017 (GMT)



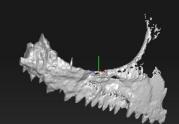
READ THE PEER-REVIEWED ARTICLE

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

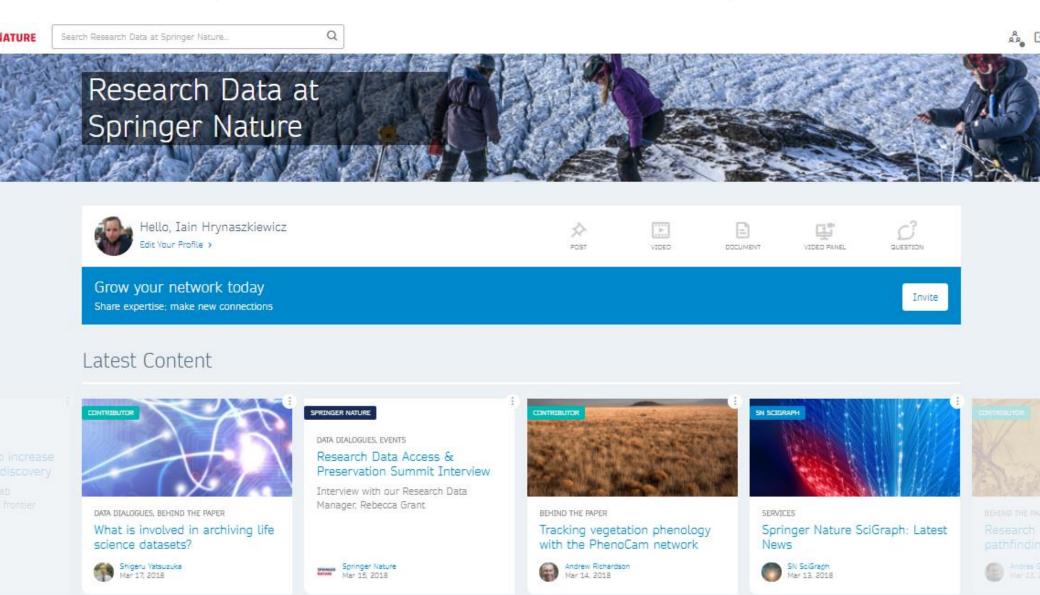
SPRINGER NATURE

Sam Giles Guang-Hui Xu Thomas J. Nea

- Paleontology
- · Palaeontology (incl. Palynology)
- · Animal Systematics and Taxonomy . Phylogeny and Comparative Analysis
- Evolutionary Biology



We're all part of the research data community





New

- Open Data badges
- Data sharing at submission

BMC Microbiology



Read more: Springer Nature launches Open data badges (pilot).

https://researchdata.springernature.com/users/182454-rebecca-pearce/posts/39872-springer-nature-launches-open-data-badges





Springer Nature Open data badge

VERIFY BADGE



Springer Nature Open data badges

Awarded on Sep 26, 2018

Because this article provides data in a publicly available open repository, it has been awarded an Open data badge.

Criteria

An Open data badge is awarded when the following criteria are met:

- 1. A data availability statement is included with the manuscript, stating how the data can be accessed.
- The dataset (or part of the dataset) is deposited in a public repository.
- A DOI, Accession Number, or other appropriate persistent object identifier is supplied for the dataset.
- The dataset provided is relevant to the related paper.

This paper has been assessed by the Springer Nature Research Data Support group.

Evidence

The whole genome shotgun datasets generated in this study have been deposited at DDBJ/ENA/GenBank under the accessions QAEG00000000-QAEN0000000. The versions described in this paper are versions QAEG01000000-QAEN01000000.

Citation: Lienemann et al. BMC Microbiology (2018) 18:102 https://doi.org /10.1186/s12866-018-1250-4

v1.1

BAKED IMAGE

v2.0

BAKED IMAGE



EDITORIAL OPEN

Reproducibility: let's get it right from the start

From September 12th 2018, Nature Communications will be setting a higher standard of data reporting for papers under peer review. We believe that sharing raw data at an early stage with editors and reviewers is the best way to build confidence in the reproducibility of your findings. Learn here how to ensure that your paper makes the grade.

> hether you prefer to call it a crisis, a challenge or a revolution, the growing awareness of reproducibility as an important issue in science is surely a cause for optimism. In biomedical research in particular, journals now compete with each other to demonstrate the strength of their reproducibility credentials, and play an increasingly proactive role in setting standards for how data and methods should be reported in their pages. Many journals and funding bodies have developed checklists and guidelines with the aim of enquiring that require the reporting common to be com-

this can occur at a relatively late stage in review, resulting in a waste of time for reviewers and authors alike. Ensuring that reporting is transparent, right from the start of the peer review process, would allow our reviewers to scrutinize the level of support for the findings with greater confidence and at a point where any problems can more easily be resolved.

For this reason, Nature Communications is now joining the other Nature journals in requiring authors to complete all relevant checklists at the point their paper is sent for peer review. In the life sciences, we will

About For Authors

For Reviewers

Blog

Publications

Submit Manuscript

Preprints

PREPRINTS REASONS TO PREPRINT

GET STARTED

PREPRINT FAQ



OCTOBER 22 - 28

Join us for a live-streamed, interactive preprint journal club event.

Your scientific input is valuable! Make it heard and influence the development of a research article by participating in this Open Access Week event.

Interact with fellow researchers as you contribute your feedback, review and enhance a manuscript that is still undergoing formal peer review at a journal.

Hosted by PLOS and PREreview

REGISTER NOW

PLOS Preprints https://www.plos.org/preprints

Read more: Lin, J. Preprints growth rate ten times higher than journal articles, Crossref Blog. 31st May 2018 https://www.crossref.org/blog/preprints-growth-rate-ten-times-higher-than-journal-articles



Questions?

Thank you

Libraries and Learning Services