Open data, open science

Jennifer Lin, PhD
Senior Product Manager, PLOS

Building an environment for open data is building an environment for open science

open science needs:

- open access
- open data
- open source
- open infrastructure

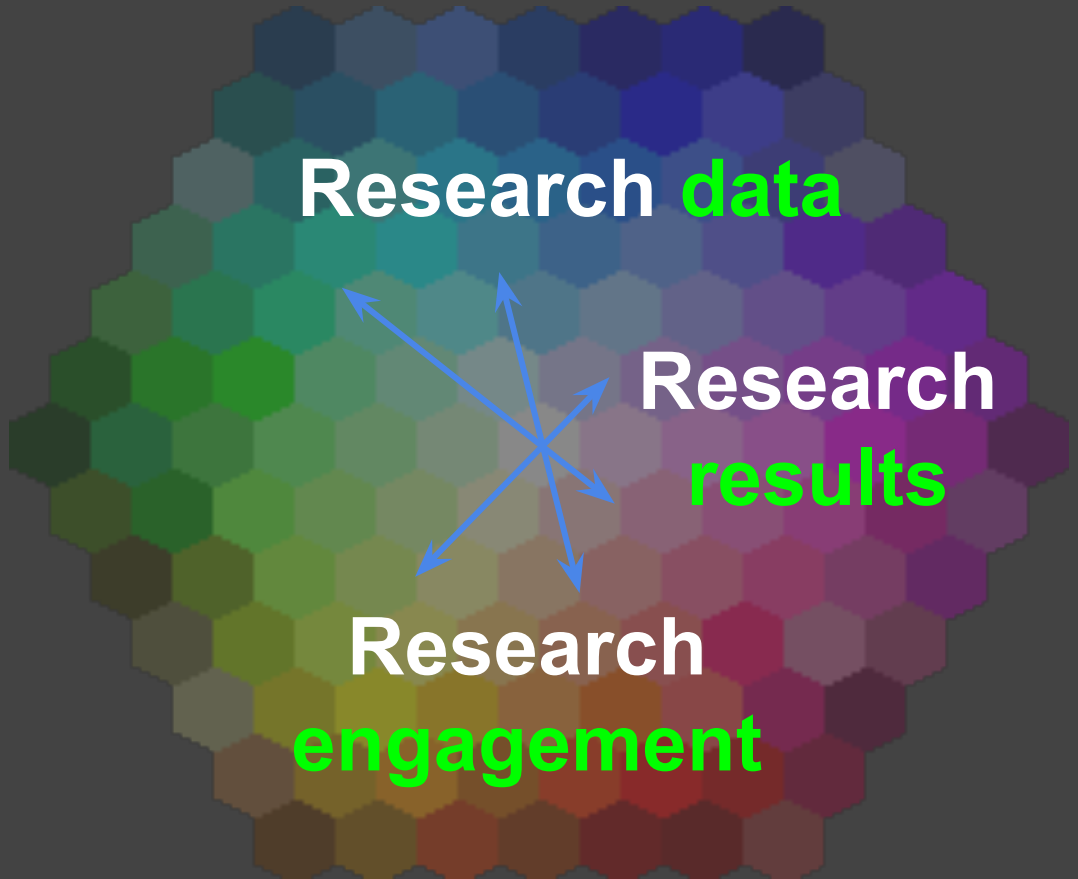
Why Open?

There are deep inefficiencies in the research enterprise:

- doing research
- finding research
- communicating results
- communicating evidence of contribution

An *open* research environment enables a far more *efficient r*esearch environment.

Framework



OPEN ACCESS

PERSPECTIVE

Recommendations for the Role of Publishers in Access to Data

Jennifer Lin, Carly Strasser

Published: October 28, 2014 • DOI: 10.1371/journal.pbio.1001975

16 Saves | 1 Citation | 9,461 Views | 225 Shares

Abstract

As appeals for public access of research data continue t
—alongside funders, institutions, and libraries—are expa
Here we outline eight recommendations and a set of sug
promote and contribute to increasing access to data. Thi
that brought together data stewardship leaders across st
recommendations were subsequently refined by the com
gathered online and in meetings.

Box 2. Recommendations for Publishers to Increase Access to Data

1. Establish and enforce a mandatory data availability policy.
2. Contribute to establishing community standards for data management and sharing.
3. Contribute to establishing community standards for data preservation in trusted repositories.
4. Provide formal channels to share data.
5. Work with repositories to streamline data submission.
6. Require appropriate citation to all data associated with a publication—both produced and used.
7. Develop and report indicators that will support data as a first-class scholarly output.
8. Incentivize data sharing by promoting the value of data sharing.

Community...

PLOS Data Access Policy Update
No barriers to enter.
All welcome.
All data underlying the published results should be
fully available without restriction with rare exception
and described in the article publication.

Aims

- Establish clarity with respect to authors' obligations
- Highlight author's responsibility to determine and describe a data sharing plan to be published with the article
- Enhance enforcement mechanism
- Ensure transparency so that compliance with the policy is externally visible to readers (and editors & referees during peer review)
- Ensure policy is workable across scientific fields, and take account of special considerations (in relation to privacy of human subjects, and other issues)
- **Not change WHAT data needs to be shared –** rather, focus on **WHERE** it is housed, **WHEN** it is shared, and **HOW** authors provide access for those who want it

***Sine Systemate Chaos?* A Versatile Tool for Earthworm Taxonomy: Non-Destructive Imaging of Freshly Fixed and ...**

Rosa Fernández, Sebastian Kvist, Jennifer Lenihan, Gonzalo Giribet, Alexander Ziegler

PLOS

- **Abstract**
- Introduction
- Materials and Methods
- Results
- Discussion
- Conclusions
- Supporting Information
- Acknowledgments
- Author Contributions
- References

- Reader Comments (0)
- Media Coverage (0)
- Figures

Fixed and Museum Specimens Using Micro-Computed Tomography. PLoS ONE 9(5): e96617. doi:10.1371/journal.pone.0096617

Editor: Jacob Guy Bundy, Imperial College London, United Kingdom

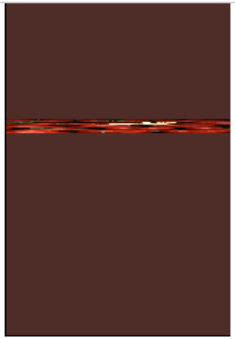
Received: December 11, 2013; **Accepted:** April 9, 2014; **Published:** May 16, 2014

Copyright: © 2014 Fernández 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.

Funding: Funding for RF was provided by the Ramón Areces Foundation, for SK by the Wenner-Gren Foundations and Helge Ax:son Johnson's Foundation, and for AZ by the Deutsche Forschungsgemeinschaft through grants no. ZI-1274/1-1 and ZI-1274/1-2. Research was also supported by internal funds from the Museum of Comparative Zoology and from the Department of Invertebrate Zoology. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Data Availability: The authors confirm that all data underlying the findings are freely available without restriction. Specimen data may be accessed through the Museum of Comparative Zoology at Harvard University's MCZbase (http://mczbase.mcz.harvard.edu/) under the sample IDs http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:24804, http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:24805, http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:95557, http://mczbase.mcz.harvard.edu/guid/MCZ:IZ:95901. MicroCT scans are available through GigaScience's GigaDB (http://gigadb.org/site/index) at http://dx.doi.org/10.5524/100092.



Archived Tweets

29 May 2014

Stéphane Boyer @Zelpapang
CT scanning #earthworm, a non-invasive method to describe worms internal morphology. #PLOSONE http://t.co/9mVNG6fHN3 http://t.co/21yWMTok5F

24 May 2014

G. Chevroux Maguarie

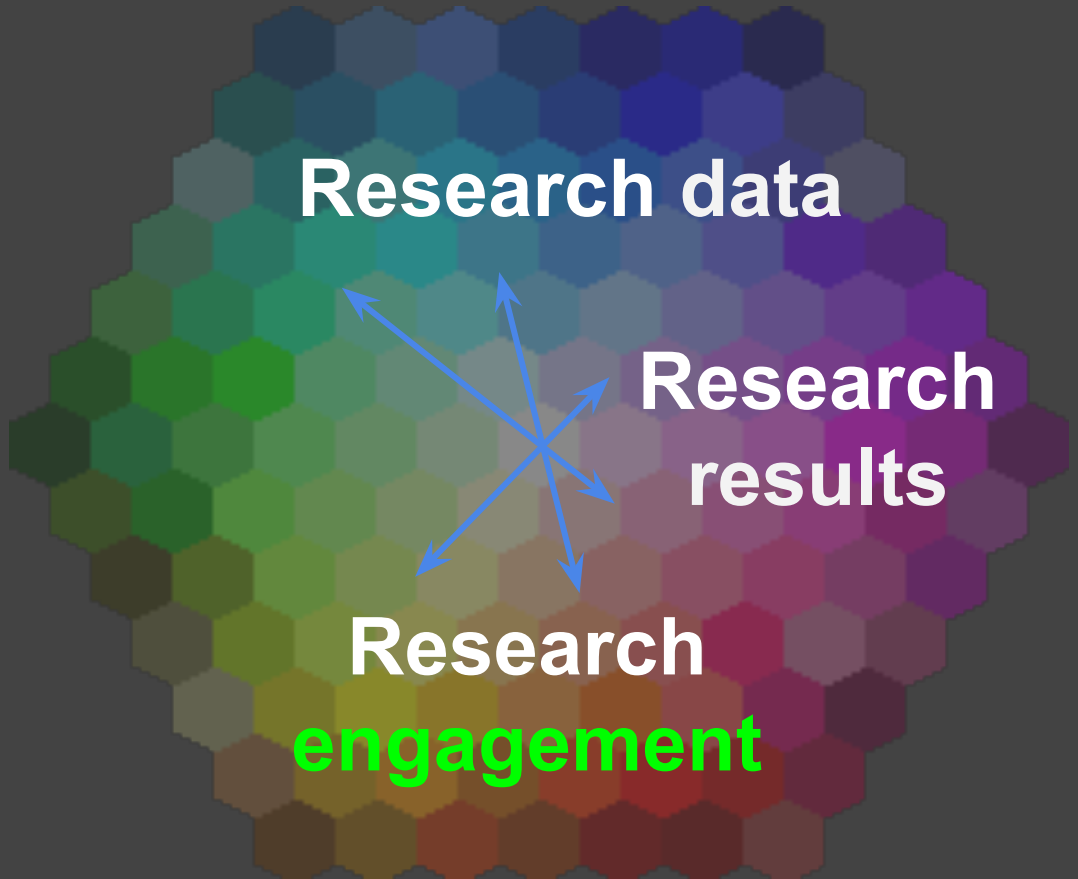
Data on Data @PLOS

(April 1, 2014 - early Feb 2015)

# of papers with Data Availability Statements	~11k
Fully shared	
% data deposited in repositories	~11%
% all data included in manuscript/SI files	~57%
Restrictions	
% data held by 3rd parties	~2%
% data containing sensitive information (available upon request)	~2%

Most widely used repositories	● figshare ● Dryad ● NCBI Genbank ● NCBI GEO ● NCBI SRA ● Dataverse ● EBI ● Github ● Array Express ● RCSB Protein Databank
Persistent identifiers used	18% articles ● DOI ● GEO identifier ● Genbank identifier ● PDB identifier
Increase in rate of deposits since policy	yes
Increase in SI files since policy	yes

Open metrics, Open science



Scholars access, cite, share, mention, recommend research

Usage

PLOS Journals
(HTML, PDF, XML)

PubMed Central
(HTML, PDF)

figshare

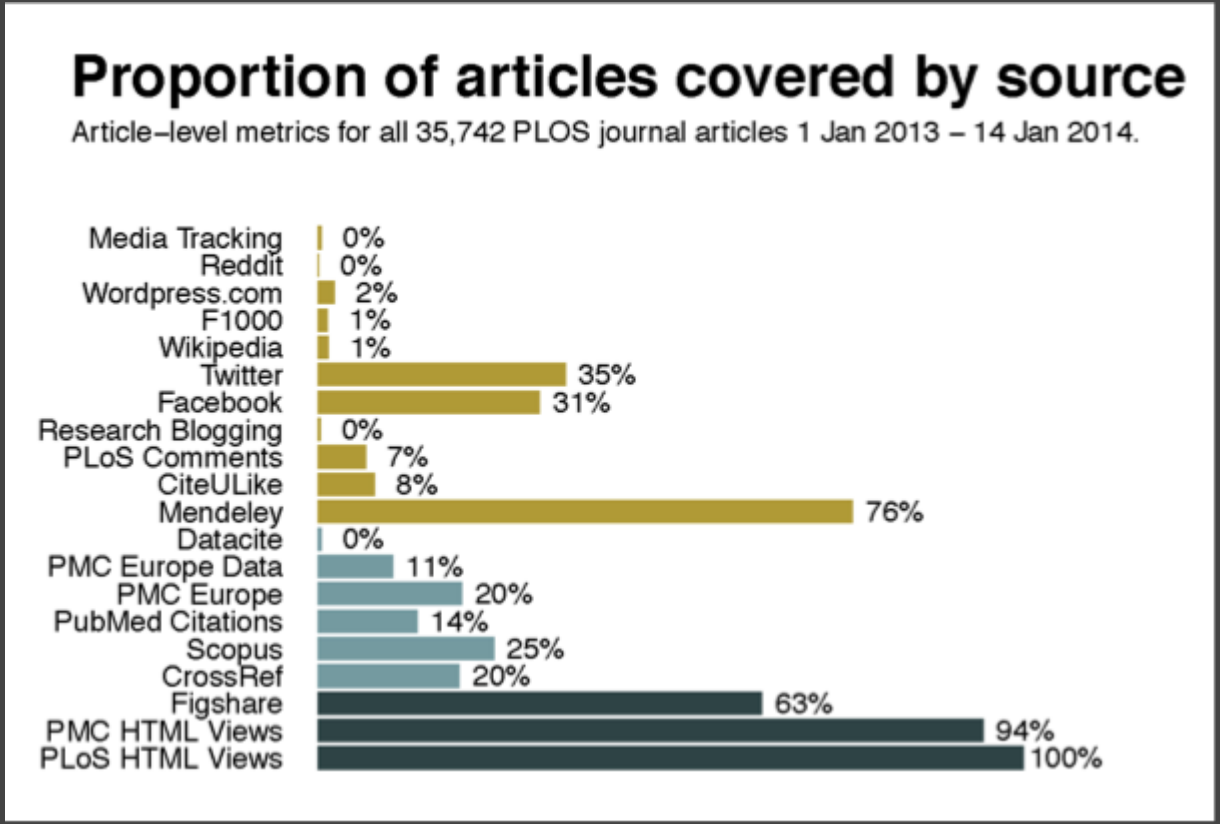
Citations

CrossRef
Scopus
Web of Science
PubMed Central
PMC Europe
PMC Europe
database Citations
DataCite

Altmetrics

PLOS Comments
F1000Prime
Mendeley
CiteULike
ResearchBlogging
ScienceSeeker
Facebook
Twitter
Wikipedia
Reddit

Diverse research engagement



Data sharing incentives- Make Data Count

Make Data Count Project Partners

California Digital Library, PLOS, DataONE

NSF Grant Record

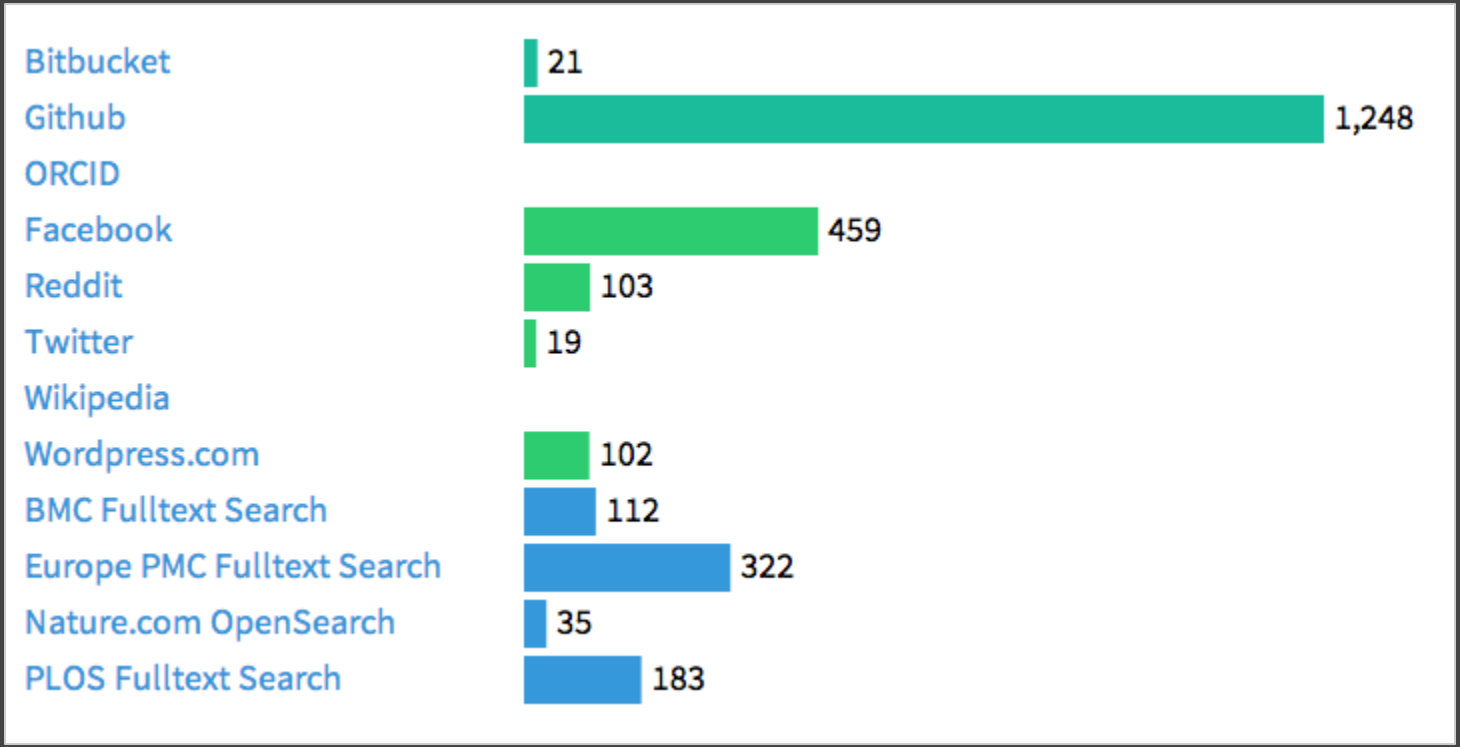
Grant No. 1448821 proposal (eScholarship repository)

Project page articlemetrics.github.io/MDC/

Prototype dlm.plos.org

Software github.com/articlemetrics/lagotto

Open software, open science



Number of software repositories (out of 1,404) with at least one event. Data: software.lagotto.io

“Open” needs highways

Scholarly infrastructure to support:

- publisher and data repository bi-directional linking of literature - data
- integrated data repository and article submission systems
- researcher discovery of data
- funder tracking of grant activity
- ...

Pitch in & help build with us

Researchers & librarians:

- publish in OA journal
- make data available in a public repository
- contribute open peer reviews and provide public commentary to published literature
- demand metrics that capture the reach and reuse of their publications from publishers

Thank you

Open benefits the entire research enterprise, most importantly, **researchers**.

Jennifer Lin (jlin@plos.org)