Open Data Sharing – PLOS ONE's Perspective

Meg Byrne

Senior Editor, PLOS ONE

July 11, 2017

Making data available fosters scientific progress

Data availability allows:

- Validation, replication, reanalysis, new analysis
- Reproducibility
- Increased value of research
- Reduction of the burden on authors
- Easier citation of data

journals.plos.org/plosone/s/data-availability

PLOS


PLOS ONE
PLOS BIOLOGY
PLOS MEDICINE
PLOS COMPUTATIONAL BIOLOGY
PLOS GENETICS
PLOS NEGLECTED TROPICAL DISEASES
PLOS PATHOGENS

PLOS

PLOS Data Policy

PLOS journals require authors to make **all data underlying the findings described in their manuscript fully available without restriction** when at all possible.

When submitting a manuscript online, authors must provide a **Data Availability Statement** describing compliance with PLOS' policy. If the article is accepted for publication, the data availability statement will be published as part of the final article.

Since March 2014

PLOS

Data Availability Statement Published on Each Article

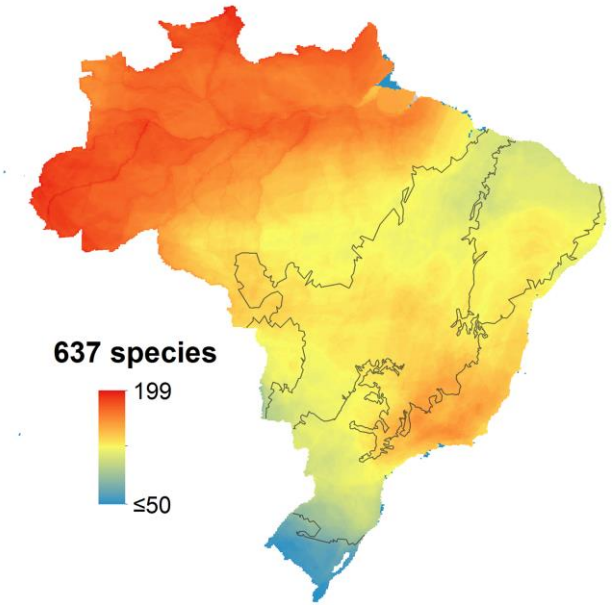
OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Patterns of Vertebrate Diversity and Protection in Brazil

Clinton N. Jenkins, Maria Alice S. Alves, Alexandre Uezu, Mariana M. Vale

Published: December 17, 2015 • https://doi.org/10.1371/journal.pone.0145064



Editor: Adam Stow, Macquarie University, AUSTRALIA

Received: October 31, 2015; **Accepted:** November 29, 2015; **Published:** December 17, 2015

Data Availability: Biodiversity results, including GIS-ready datasets for open-access use, are available online at http://BiodiversityMapping.org and the Dryad Digital Repository: (http://dx.doi.org/10.5061/dryad.6rv61).

Funding: CNJ received support from the Ciência Sem Fronteiras program (A025_2013), MASA received support from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, process 08792/2009-2), and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ, process E-26/102.837868 /2012). MMV received support from CNPq (grant no. 444704/2014-0), MCTI/CNPq/FAPs PELD (Grant No. 34/2012), CNPq PPBio/Rede BioM.A. (Grant No. 477524/2012-2), FAPERJ (grant no. E-26/111.577/2014) and RedeCLIMA Program (grant no. 01.0405.01). 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.

What Data?

The policy applies to **the dataset used to reach the conclusions** drawn in the manuscript with **related metadata and methods, and any additional data required to replicate** the reported study findings in their entirety.

Authors need not submit your entire dataset, or all raw data collected during an investigation, but **they must provide the portion that is relevant to the specific study.**

PLOS

Minimal Dataset

- The **values** behind the means, standard deviations and other measures reported
- The **values** used to build graphs
- The **points** extracted from images for analysis.

Authors are not required to make all images available, but we do require a **sample Western Blot, Immunohistochemistry image, fMRI image, etc.** to be included with the submission files or in a public repository.

PLOS

Exceptions

- Data cannot be made publicly available for **ethical or legal reasons**, e.g., public availability would compromise patient confidentiality or participant privacy.

- Data deposition **could present some other threat**, such as revealing the locations of fossil deposits, endangered species, or farms/other animal enclosures etc.

- Data are owned by a **third party**.

PLOS

Where?

Data deposition in public repository

- Strongly recommended
- Discipline-specific repositories preferable
- Authors must specify DOIs or accession numbers

Supporting information files

- Can accept up to ~100 MB of data
- Each file has its own DOI and is available in Figshare

In the body of the manuscript

PLOS

Citation: Drake JM, Kaul RB, Alexander L et al. (2015) Ebola Cases and Health System e1002056. doi:10.1371/journal.pbio.10020

Academic Editor: Steven Riley, Imperi

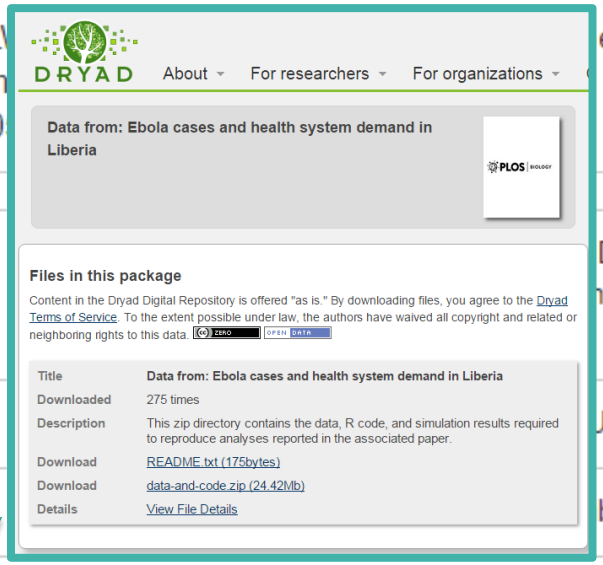
Received: October 31, 2014; **Accepte** 2015

Copyright: © 2015 Drake et al. This is terms of the Creative Commons Attribu distribution, and reproduction in any me are credited

Data Availability: All files are available http://doi.org/10.5061/dryad.17m5q.

Funding: This research was funded by (http://www.nih.gov/). The content is so necessarily reflect the official views of t no role in study design, data collection of the manuscript.

Competing interests: The authors hav



Doebley JF (2014) The Role of *cis* . PLoS Genet 10(11): e1004745.

United States of America

ber 9, 2014; **Published:** November 6, 2014

Copyright: © 2014 Lemmon 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: The authors confirm that all data underlying the findings are fully available without restriction. The raw sequence data has been deposited in NCBI Sequence Read Archive with accessions SRX710894-711341 and the Gene Expression Omnibus (GEO) Series with accession number GSE61810 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61810). Supplemental datasets have been made available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.4kh67.

Funding: This work was supported by the National Science Foundation grants

esign,

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

The data availability statement is openly available, and machine-readable as part of the PLOS search API

Data in Supporting Information

plos.org create account sign in

PLOS ONE | TENTH ANNIVERSARY

Publish About Browse Search

advanced search

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Regulation of Heat Exchange across the Hornbill Beak: Functional Similarities with Toucans?

T. M. F. N. van de Ven, R. O. Martin, T. J. F. Vink, A. E. McKechnie, S. J. Cunningham

Published: May 18, 2016 • https://doi.org/10.1371/journal.pone.0154768

21 Save	4 Citation
4,520 View	113 Share

Data Availability: All relevant data are within the paper and its Supporting Information files.

PLOS

Data in Supporting Information

Regulation of Heat Exchange across the Hornbill Beak: Functional Similarities with Toucans?

T. M. F. N. van de Ven, R. O. Martin, T. J. F. Vink, A. E. McKechnie, S. J. Cunningham

Published: May 18, 2016 • http://dx.doi.org/10.1371/journal.pone.0154768

S1 Data.XLSX

	1	2	3	4	5	
1	Bird ID	Sex	Bill height (mm)	Bill surface (mm2)	Bill surface (m2)	Body m
2	3	1	30.943	1707.055	0.00341411	223.7
3	4	1	30.427	1818.746	0.003637492	229.5
4	8	1	30.773	1647.224	0.003294448	235.3
5	9	1	30.711	1947.31	0.00389462	237.3
6	10	1	29.249	1696.357	0.003392714	237.4
7	13	1	29.542	1844.555	0.00368911	252.5
8	14	1	34.766	2065.919	0.004131838	270.4
9	15	1	27.996	1608.615	0.00321723	258.9
10	17	1	29.922	1727.775	0.00345555	231.4

Fig1 | Fig4, Table1 | S2, Table 2 | S3, Table 3 | Table 4 | S4

figshare 1 / 6 download

Download this file (96.28 kB)
Download all (7.79 MB)

Numerical data used in preparation of Figs 1 and 4; Tab
Figs.
(XLSX)

PLOS

Data in Supporting Information

Regulation of Heat Exchange across the Hornbill Beak: Functional Similarities with Toucans?

T. M. F. N. van de Ven, R. O. Martin, T. J. F. Vink, A. E. McKechnie, S. J. Cunningham

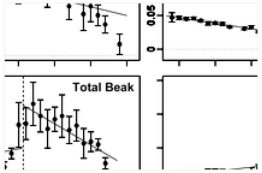
Published: May 18, 2016 • http://dx.doi.org/10.1371/journal.pone.0154768

figshare

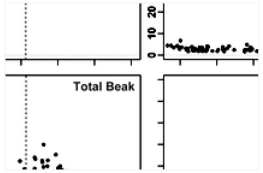
search on figshare

Browse Upload Sign up Log in

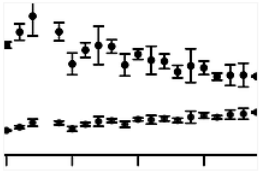
S1 Data.XLSX (96.28 kB)



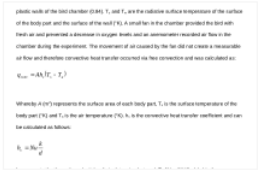
S1 Fig.TIF (1.18 MB)



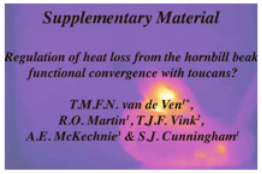
S2 Fig.TIF (1.11 MB)



S3 Fig.TIF (424.14 kB)



S1 Text.DOCX (22.28 kB)



S1 Video.MP4 (4.98 MB)

Download all (7.8 MB) Share Cite Embed + Collect (you need to log in first)

6 files

van de Ven, T. M. F. N.; Martin, R. O.; Vink, T. J. F.; McKechnie, A. E.; Cunningham, S. J. (2016): Regulation of Heat Exchange across the Hornbill Beak: Functional Similarities with Toucans?. figshare.

https://dx.doi.org/10.1371/journal.pone.0154768

Retrieved: 18 27, May 23, 2016 (GMT)

Place your mouse over the citation text to select it

PLOS

Restricted Data

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Elephant Management in North American Zoos: Environmental Enrichment, Feeding, Exercise, and Training

Brian J. Greco, Cheryl L. Meehan, Lance J. Miller, David J. Shepherdson, Kari A. Morfeld, Jeff Andrews, Anne M. Baker, Kathy Carlstead, Joy A. Mench

Published: July 14, 2016 • https://doi.org/10.1371/journal.pone.0152490

Data Availability: For reasons relating to protection of the facilities and animals included in this study, access restrictions apply to the individual-level data underlying the findings. A data set of de-identified, population-level data is available at doi: 10.6084/m9.figshare.3383554.

PLOS

Restricted Data

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Can machine-learning improve cardiovascular risk prediction using routine clinical data?

Stephen F. Weng, Jenna Reps, Joe Kai, Jonathan M. Garibaldi, Nadeem Qureshi

Published: April 4, 2017 • https://doi.org/10.1371/journal.pone.0174944

107 Save	0 Citation
30,524 View	540 Share

Data Availability: This dataset contains patient level health records with intellectual property rights held by The Crown copyright, which is subject to UK information governance laws. The authors will make their data available upon specific requests subject to the requestor obtaining ethical and research approvals from the Clinical Practice Research Datalink Independent Scientific Advisory Committee (https://www.cprd.com/intro.asp) at the UK Medicines and Health Products Regulatory Agency.

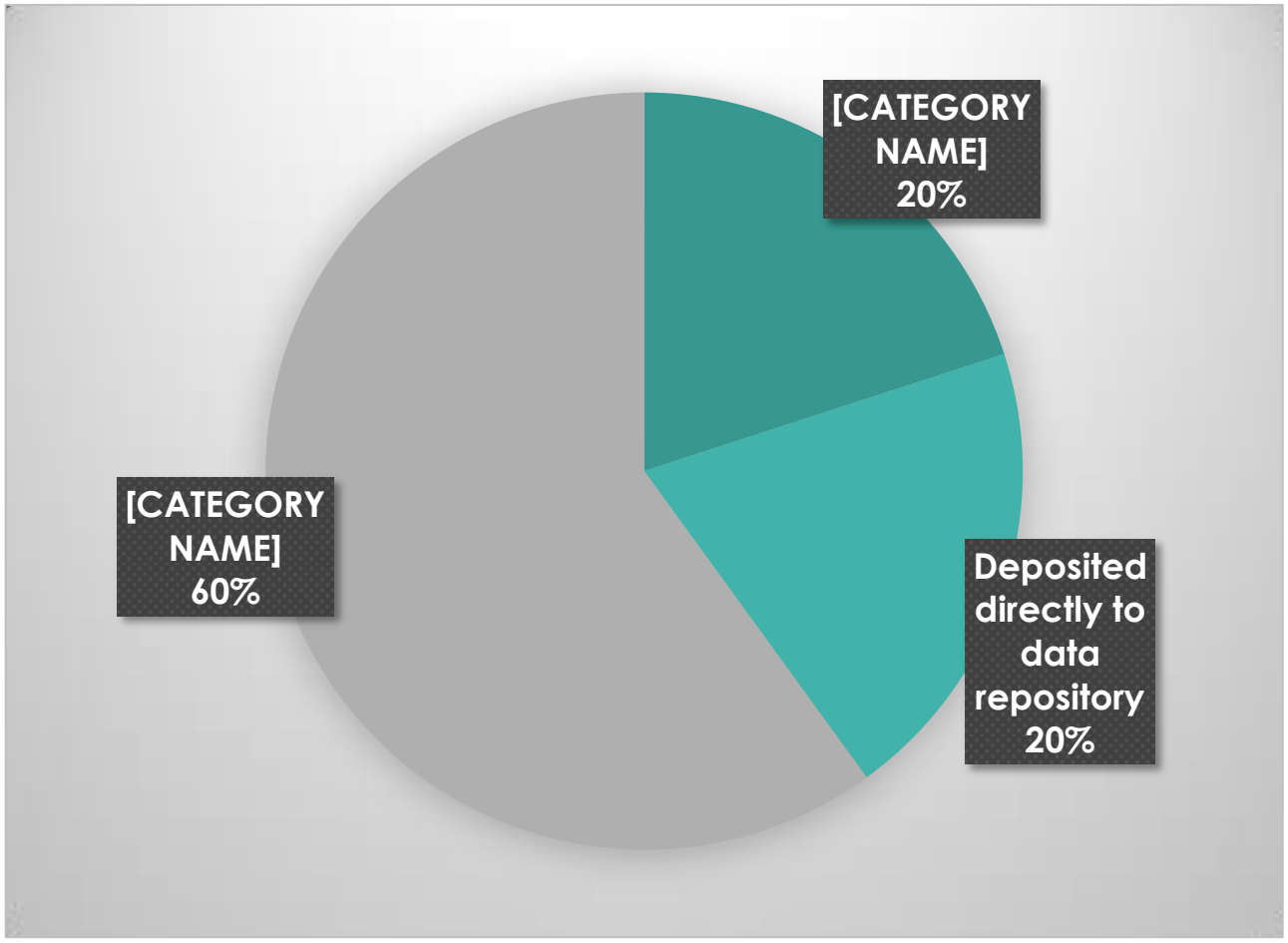
- Authors state
  1) Reasons for restrictions on making data publicly available.
  2) Contact information or instructions for requesting the data.
- Statement considered by Editor, Reviewers, and Journal Staff during review process.

> 70,000

papers published with a data availability statement

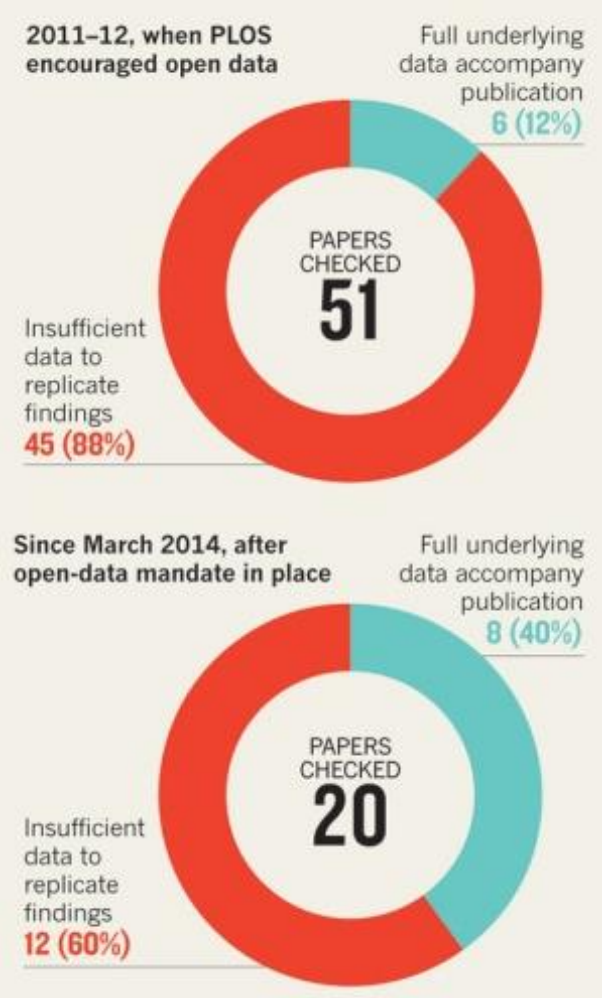
PLOS

Where are the data (PLOS ONE)?



PLOS

Impact of Data Policy



An increase in data sharing from 12% to 40%

Not seeing full compliance but we are seeing a significant improvement

Recent analysis saw an increase to 67% (Tim Vines, personal communication)

Source: 'Confusion over publisher's pioneering open-data rules' *Nature* **515,** 478 (27 November 2014) doi:10.1038/515478a

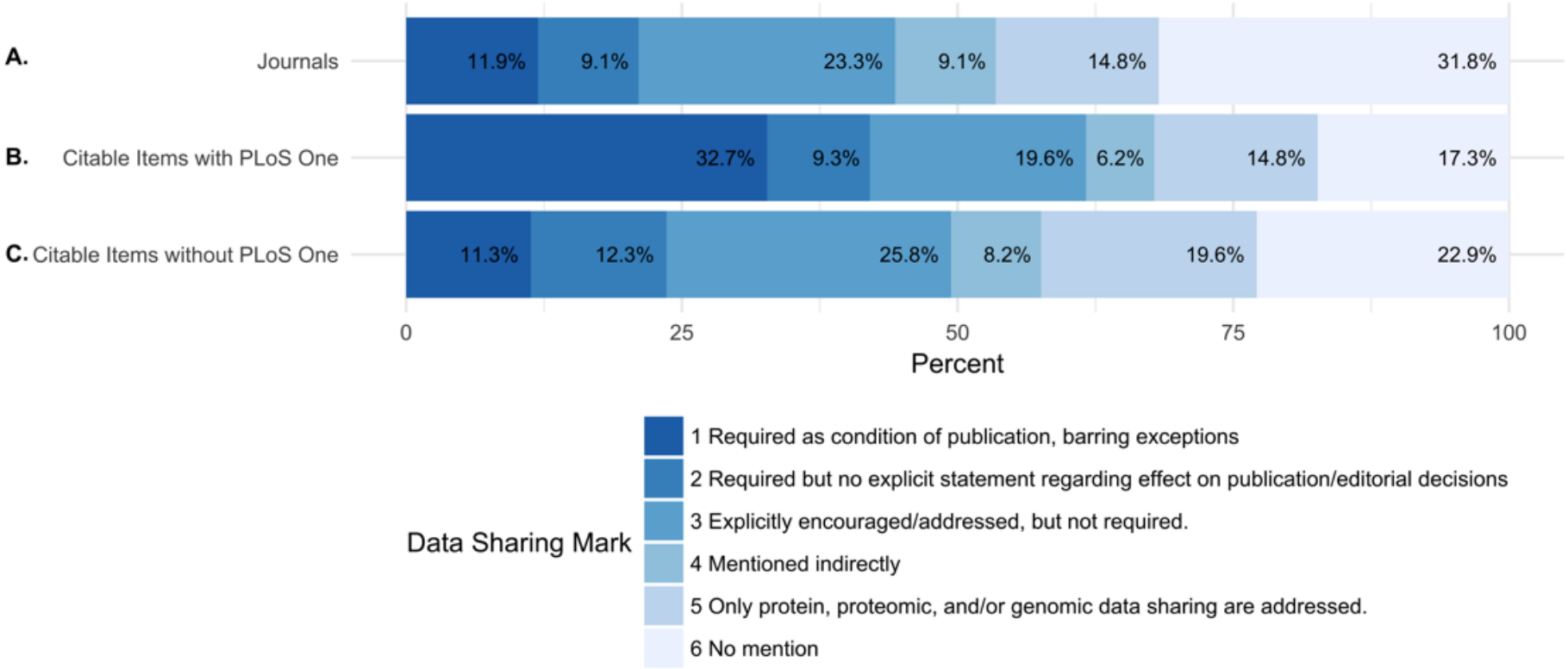
PLOS

PLOS ONE data checks

- **At submission**
  - Ask authors for initial data available statement
  - Check for unacceptable restrictions
- **During review**
  - Academic Editors and Reviewers assess underlying data
  - Send additional information authors
- **At accept**
  - Check all data availability statements
  - Check clinical datasets for potentially identifying information
- **Post-publication**
  - Follow up with authors as needed

PLOS

PLOS ONE significantly increases citable biomedical research items with open data

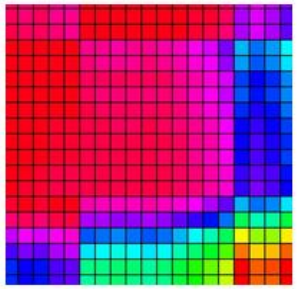


From Vasilevsky NA, Minnier J, Haendel MA, Champieux RE. (2017) Reproducible and reusable research: are journal data sharing policies meeting the mark?
PeerJ 5:e3208 https://doi.org/10.7717/peerj.3208

PLOS

PLOS ONE Datasets Collection: Noteworthy Examples Across Disciplines

Social Networks



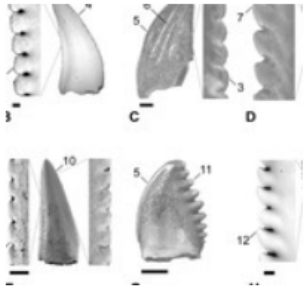
Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter

Peter Sheridan Dodds, Kameron Decker Harris, Isabel M. Kloumann, Catherine A. Bliss, Christopher M. Danforth

PLOS ONE: 07 Dec 2011

Sune Lehmann: "This paper has made an MTurk generated list of word-valences openly available to the research community. As a first-cut sentiment analysis method, this dataset is invaluable and I've downloaded it at least dozens of times to use for both teaching & research."

Paleontology



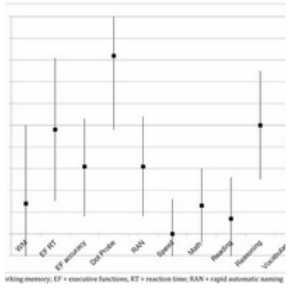
Multivariate Analyses of Small Theropod Dinosaur Teeth and Implications for Paleoecological Turnover through Time

Derek W. Larson, Philip J. Currie

PLOS ONE: 23 Jan 2013

Andrew Farke: "This paper assembles a massive dataset of measurements for over 1,000 teeth of small carnivorous dinosaurs, which has been really useful to help track changes in dinosaur diversity and distribution prior to the big extinction at the end of the Mesozoic."

Applied Psychology



Closing the Achievement Gap through Modification of Neurocognitive and Neuroendocrine Function: Results from a C...

Clancy Blair, C. Cybele Raver

PLOS ONE: 12 Nov 2014

In this cluster randomized controlled trial, Blair and Raver found that teaching self-regulation can help lower achievement gaps in kindergarten, particularly in high-poverty schools. The individual-level data from over 750 children in 29 schools are available in Dryad.

Guidance for researchers

Data policy FAQs
http://journals.plos.org/plosone/s/data-availability#loc-faqs-for-data-policy

Preparing clinical data for publication
http://journals.plos.org/plosone/s/data-availability#loc-clinical-data

"Ten Simple Rules for the Care and Feeding of Scientific Data"

"Ten Simple Rules for Creating a Good Data Management Plan"

"Sharing Research Data and Intellectual Property Law: A Primer"

"Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data"

Recommended repositories

http://journals.plos.org/plosone/s/data-availability#loc-recommended-repositories

Discipline-specific repositories

Biochemistry	Neuroscience	Social Sciences
Biomedical Sciences	Omics	Structural Databases
Marine Sciences	Physical Sciences	Taxonomic & Species Diversity
Model Organisms	Sequencing	Unstructured and/or Large Data

Cross-disciplinary repositories

- Dryad Digital Repository
- figshare
- Harvard Dataverse Network
- Open Science Framework
- Zenodo

PLOS

Institutional repositories adhering to best practices

Continued discussion

"Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results"

"Sharing Detailed Research Data Is Associated with Increased Citation Rate"

"Ethical Challenges of Big Data in Public Health"

"Can Data Sharing Become the Path of Least Resistance?"

"Making Progress Toward Open Data: Reflections on Data Sharing at PLOS ONE"

Institutional support for researchers: Examples

http://www.data.cam.ac.uk/

dash | UCSF DataShare

http://guides.ucsf.edu/datamgmt/share
https://datashare.ucsf.edu/stash/

Resource for funders

POPULAR RESOURCES

Implementing an Open Data Policy: A SPARC Primer for Research Funders

Open Data

*Complements work that SPARC did with the Health Research Alliance

https://sparcopen.org/our-work/implementing-an-open-data-policy/

Funders can make a significant impact.

By the time authors submit articles to a journal, it is often too late to make data shareable.

PLOS

Many questions remain

- How long should researchers store data?
- How much data are needed to replicate a study?
- How should materials sharing differ?
- How do we handle software/code?
- Do we need better/more aligned consenting for patient studies?
- What are best practices for data access committees?
- How can we preserve obsolete formats?
- How should data be cited and authors credited?

PLOS

Many groups thinking about these questions