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 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
Data Availability Statement Published on Each 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

Copyright: © 2015 Jenkins 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: 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.6v61).

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-26/102.837/2012). MMV received support from CNPq (grant no. 444704/2014-0), MCTI/CNPq/FAPs PELD (Grant No. 34/2012), CNPq PPFBio/Rede BioMA (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.
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.
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**.
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
The data availability statement is openly available, and machine-readable as part of the PLOS search API.
Data Availability: All relevant data are within the paper and its Supporting Information files.
# 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](http://dx.doi.org/10.1371/journal.pone.0154768)

## S1 Data: XLSX

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bird ID</td>
<td>Sex</td>
<td>Bill height (mm)</td>
<td>Bill surface (mm²)</td>
<td>Bill surface (m²)</td>
<td>Body mass</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>30.843</td>
<td>1707.055</td>
<td>0.00341411</td>
<td>223.7</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>30.427</td>
<td>1818.746</td>
<td>0.003637492</td>
<td>229.5</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1</td>
<td>30.773</td>
<td>1647.224</td>
<td>0.003294448</td>
<td>235.3</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>1</td>
<td>30.711</td>
<td>1947.31</td>
<td>0.00389462</td>
<td>237.3</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>1</td>
<td>29.249</td>
<td>1696.357</td>
<td>0.003392214</td>
<td>237.4</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>1</td>
<td>29.642</td>
<td>1844.555</td>
<td>0.00368911</td>
<td>252.5</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>1</td>
<td>34.766</td>
<td>2065.919</td>
<td>0.00411838</td>
<td>270.4</td>
</tr>
<tr>
<td>9</td>
<td>15</td>
<td>1</td>
<td>27.996</td>
<td>1608.615</td>
<td>0.00321723</td>
<td>258.9</td>
</tr>
</tbody>
</table>

**Numerical data used in preparation of Figs 1 and 4; Table 3.**

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

Data in Supporting Information
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.
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
Where are the data (PLOS ONE)?

- [CATEGORY NAME] 20%
- Deposited directly to data repository 20%
- [CATEGORY NAME] 60%
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 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 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
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.”

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.”

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

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”


"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**

<table>
<thead>
<tr>
<th>Biochemistry</th>
<th>Neuroscience</th>
<th>Social Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biomedical Sciences</td>
<td>Omics</td>
<td>Structural Databases</td>
</tr>
<tr>
<td>Marine Sciences</td>
<td>Physical Sciences</td>
<td>Taxonomic &amp; Species Diversity</td>
</tr>
<tr>
<td>Model Organisms</td>
<td>Sequencing</td>
<td>Unstructured and/or Large Data</td>
</tr>
</tbody>
</table>

**Cross-disciplinary repositories**

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

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

http://www.data.cam.ac.uk/
Funders can make a significant impact.

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

*Complements work that SPARC did with the Health Research Alliance

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?
Many groups thinking about these questions