





**GBIF**

Global Biodiversity  
Information Facility

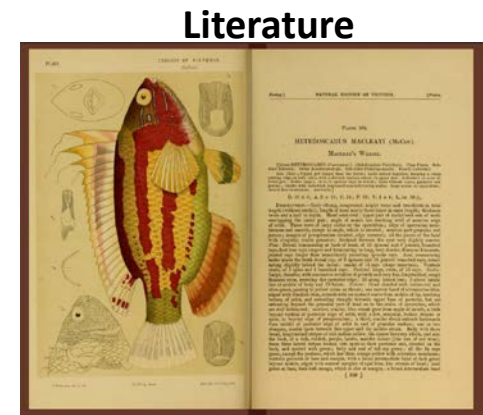
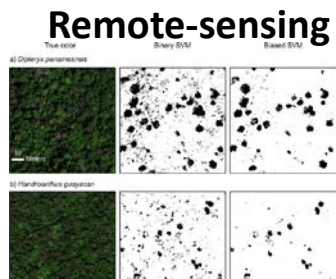
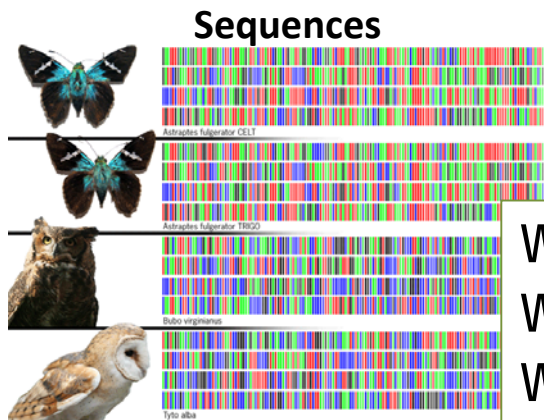
# How to Publish a Data Paper

**Dmitry Schigel**  
GBIF Scientific officer



Photo: *Meripilus giganteus*, 18 Sep 2016  
<https://flic.kr/p/LV1yEu>

# STREAMS OF SPECIES OCCURRENCE DATA



What species?  
 Where was it found?  
 When was it found?  
 What is the evidence?  
 Other information

- Specimen data
- Sampling event information
- Sequences, images, etc.

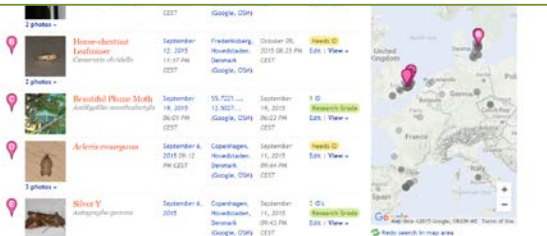


ScientificName: *Imbophorus pallidus*  
 Family: Pterophoridae  
 Locality: Stirling Range  
 Country: Australia  
 State: WA  
 Latitude: -34.3  
 Longitude: 118.0  
 CoordinatePrecision: 10000m  
 CoordinateMethod: Google Earth  
 DateCollected: 1963-09-15  
 BasisOfRecord: Preserved specimen  
 TypeStatus: Paratype

What species?  
 Where was it found?  
 When was it found?  
 What is the evidence?  
 Other information

- Specimen data
- Sampling event information
- Sequences, images, etc.

**Specimens**



**Citizen science**



**Animal tracking**



# Types of data shared through GBIF

what



Specimen



Fossil



Material sample



Literature occurrence



Observation  
Human observation  
Living specimen



Machine observation

**sample-based data**



**Barcoding**  
**Metagenomics**  
**eDNA**

---

## What is a data paper?

A **data paper** is a peer reviewed document describing a dataset, published in a peer reviewed journal. Data papers provide recognition for this effort by means of a scholarly article.

Searchable metadata document describing a dataset or groups of datasets

**! no interpretation, analyses, or hypothesis testing**  
**= data table + data story + journal link**





---

## The data paper concept

- Scholarly publication, established scientific process
- Rigorous data stewardship to create and curate data as first class objects; peer-review.
- Recognize data publication as traditional scientific articles – impact factor, citation, etc.

---

## Three main actors in data paper process

1. **Authors** – researchers, scholars, citizen scientists
2. **Publishers** – journals
3. **Data repositories** – e.g., GBIF, Figshare, Dryad, Zenodo, etc.

# Automatically generate a data paper: example

Get data Share Tools Inside GBIF

OCCURRENCE DATASET | REGISTERED JULY 18, 2019

## 2019 ForBio and UiO International Dead Wood Course. Biodiversity assesment

Published by [VORONEZHSKIY STATE NATURE BIOSPHERE RESERVE NAMED AFTER V. PERSHOV](#)

Ludmila Kalina • Artyom Leostin • Dabao Lu • Inger Skrede • Ilya Viner • Nina Filippova • Maria Shumskaya • Dmitry Schigel • Elena Starodubtseva

DATASET METRICS ACTIVITY DOWNLOAD HOME PAGE

198 OCCURRENCES

International ForBio and UiO Dead Wood Course 2019 has hold in the Voronezhsky State Nature Biosphere Reserve on 24-28 June 2019, bringing together 12 instructors and 24 students from Russia, Norway, Finland, Lithuania, Latvia, Switzerland, and the US. The intensive five-day course program was comprised by lectures, workshops and excursions. More about the course program: <http://udgprim.com/teaching/iwo2019/>, course summary: <http://zapovednik.ru/ru/en/news-and-events/2019-07-04-forbio>

Metadata last modified: July 22, 2019  
Data last changed: July 18, 2019  
Hosted by: Institute of Mathematical Problems of Biology, Russian Academy of Sciences

198 Occurrences 100% With taxon match 100% With co

198 GEOREFERENCED RECORDS

1. GBIF

arpha

Feedback Tips and tricks Dr. Siro Masinde

Start a manuscript

2. Arpha

Article structure will be defined automatically, once you have chosen article type and journal. Once a manuscript is created, its type cannot be altered. Open a new manuscript, if you want to change article type.

Journals

- Biodiversity Data Journal
- Research Ideas and Outcomes
- One Ecosystem
- BioDiscovery

Article type

Research ideas	Grant proposals	Brief research outcomes	Early research outcomes
<ul style="list-style-type: none"><li>Data Management Plan (Biosciences)</li><li>Data Management Plan (Generic)</li><li>Data Management Plan (NSF Generic)</li><li>PhD Project Plan</li><li>PhD Project Plan (Free Text)</li><li>PostDoc Project Plan</li><li>PostDoc Project Plan (Free Text)</li><li>Research Idea</li><li>Small Grant Proposal</li><li>Small Grant Proposal (Free Text)</li><li>Software Management Plan</li></ul>	<ul style="list-style-type: none"><li>DfG Grant Proposal</li><li>FP7 Grant Proposal</li><li>Grant Proposal</li><li>Grant Proposal (Free Text)</li><li>H2020 Grant Proposal</li><li>NH Grant Proposal</li><li>NSF Grant Proposal</li></ul>	<ul style="list-style-type: none"><li>Conference Abstract</li><li>Correspondence</li><li>Ecosystem Inventory</li><li>Ecosystem Service Mapping</li><li>Ecosystem Service Models</li><li>Monitoring Schema</li><li>Research Poster</li><li>Research Presentation</li><li>Single-media Publication</li></ul>	<ul style="list-style-type: none"><li>Case Study</li><li>Case Study (Free Text)</li><li>Data Paper (Biosciences)</li><li>Data Paper (Generic)</li><li>Forum Paper (Free Text)</li><li>Methods</li><li>Methods (Free Text)</li><li>Opinion Article</li><li>Opinion Article (Free Text)</li><li>Project Report</li><li>Project Report (Free Text)</li><li>Questionnaire</li><li>Software Description</li></ul>

3. BDJ

PENSOFT

About Pensoft Books E-Books Journals News & Blog Contact Register | Login

## Biodiversity Data Journal

Making your data count! ISSN 1314-2324 online

Start a manuscript

Articles About

Filter

Search

by Taxon

- Animalia
- Archaena
- Bacteria
- Chromista
- Fungi
- Plantae
- Protozoa

by Subject

- Agriculture & Forestry
- Aquatic biology
- Biosciences
- Biodiversity & Conservation
- Biogeography
- Bioinformatics

Articles

1 2 3 4

Section [Data Papers]

10 Collections - Digitising the British and Irish Butterflies in the Natural History Museum, London

Gordon Paterson doi: 10.3897/BDJ.4.49559

13/04-2016 Unique: 22 Total: 28 Reprints: € 2,10

10 AXIOM: Amphipod crustaceans from insular *Posidonia oceanica* seagrass meadows

Lois Michel, Nicolas Stavers, André Neughebaert, Gilles Lepoint doi: 10.3897/BDJ.4.41019

05/04-2016 Unique: 127 Total: 147 Reprints: € 5,50

10 Terrestrial arthropods of Steel Creek, Buffalo National River, Arkansas. IV. Asilidae and other Diptera

Michael Skvaris, Jeffrey Barnes, Danielle Fisher, Ashay Dowling doi: 10.3897/BDJ.4.49977

07/04-2016 Unique: 130 Total: 196 Reprints: € 4,80

10 Species composition, community and population dynamics of two gallery forests from the Brazilian Cerrado domain

Markus Gastauer, Roosevelt Almeida, Angela Masaki, Eco Diniz, Luis Moreira, João Mera-Neto





Intergovernmental open data infrastructure

Established in 2001, OECD recommendation

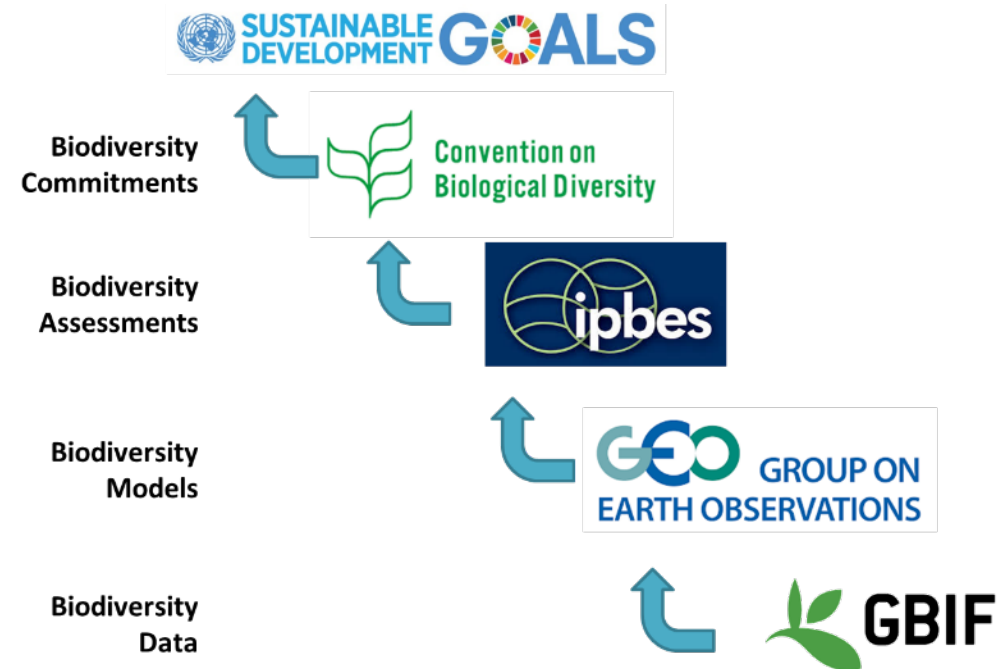
Network for free and open access  
to biodiversity data from all countries

Voluntary memorandum of understanding

Funds: governments of the participant countries

96 participants: 58 countries and 38 organizations

<https://www.gbif.org/the-gbif-network>






GBIF | Global Biodiversity Information Facility

# Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

WHAT IS GBIF? ABOUT GBIF RUSSIAN FEDERATION



Occurrence records  
1,335,543,292

Datasets  
45,635

Publishing institutions  
1,439

Peer-reviewed papers using data  
3,758



News

Belarus extends GBIF's European membership map eastward  
16 July 2019



Data use

Minimizing biodiversity loss in the Brazilian Cerrado  
2 July 2019



Data use

Science Review 2019  
Stay up to date on the latest research investigations enhanced and supported by free and open access to biodiversity data.



News

Programme seeks Biodiversity Open Data Ambassadors to expand best practices  
10 July 2019



News

BIFA funds nine Asian data mobilization projects  
26 June 2019



Taxonomy

*Atlantodesmus sierwaldae* sp. nov.  
Cladistic analysis and description of a new species of the Brazilian genus *Atlantodesmus* Hoffman, 2000 (Diplopoda: Polydesmida: Chelodesmidae)



Tool

The GBIF network  
Dozens of countries and organizations working together to make species data findable, accessible, interoperable and reusable.



Guidance

Establishing a national biodiversity information facility in Chile  
Experiences of setting up and running a node in Latin America

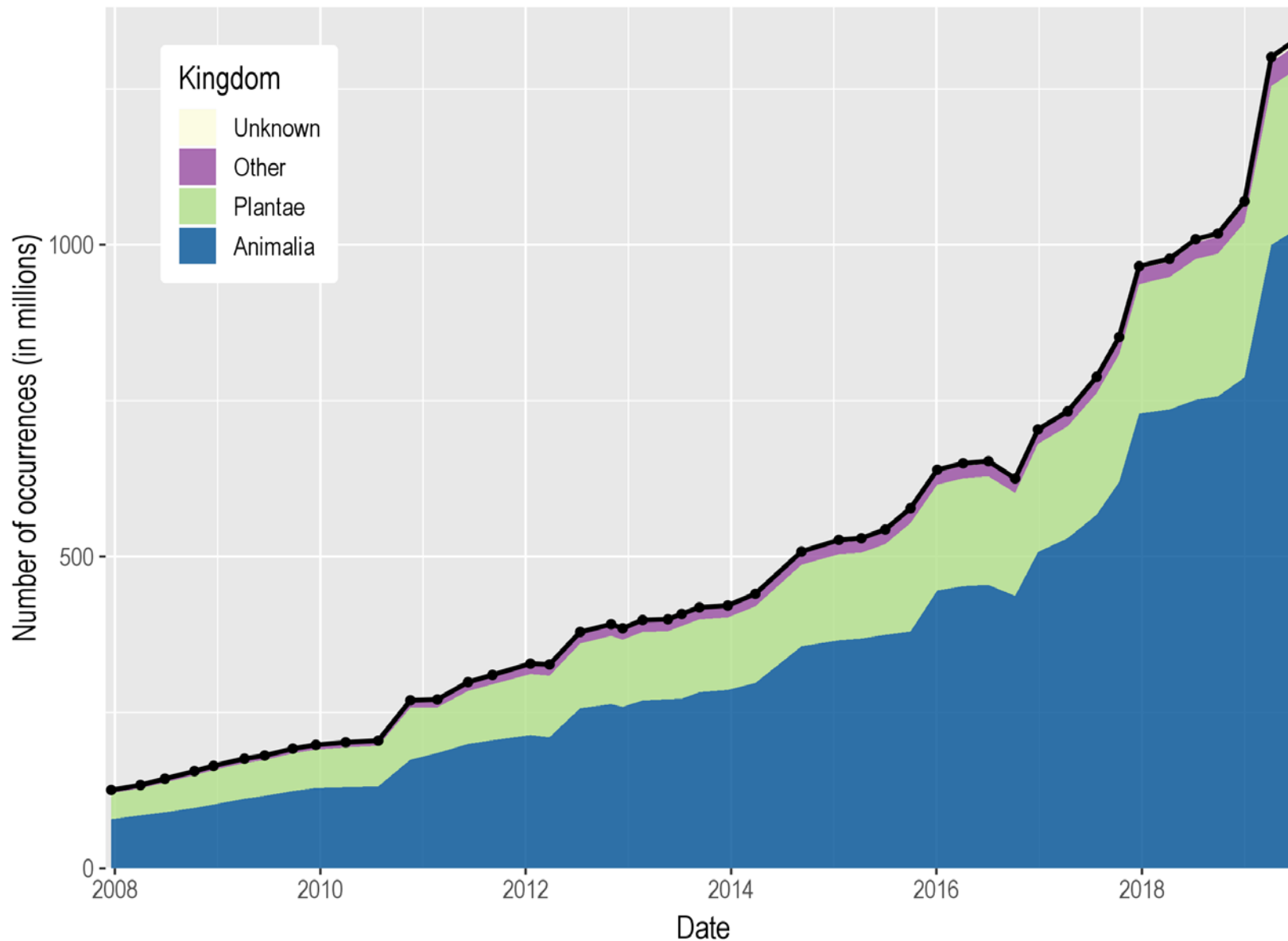


Global Biodiversity  
Information Facility

Data published  
through GBIF.org

data availability

Species occurrence records accessible through GBIF over time



[www.gbif.org/analytics/global](http://www.gbif.org/analytics/global)





# BY THE NUMBERS

7 Aug 2019

*Species occurrence records*

**1,335,646,137**

*Datasets*

**45,632**

*Country  
Participants*

**58**

*Organizational  
Participants*

**38**

*Publishers*

**1,439**

# Data citation: tracking and display



data access and use

...ure, how it's discovered and linked to GBIF-mediated data.

**Leptodactylus mystaceus (Spix, 1824): Distribution extension for the Brazilian Cerrado (Anura: Leptodactylidae)** [↔](#) **Literature**

Andreani, T. Borges, R. Santos, L. (2017) Herpetology Notes  
Fifty-seven species of anurans belonging to the genus *Leptodactylus* are recorded and widely distributed throughout Brazil (Sá et al., 2014; SBH, 2016). Among these species, *Leptodactylus mystaceus* (Spix, 1824) spreads throughout the whole Northern region and some areas of northeastern and central...

Journal Article  
Data used in study [DOI 10.15468/dl.s1xtf6](#)

**DO ENVIRONMENTAL FACTORS AFFECT THE TAXONOMIC RELIABILITY OF LEAF CUTICULAR MICROMORPHOLOGICAL CHARACTERS? A CASE STUDY IN PODOCARPACEAE** [↔](#) **Literature**

Clugston, J. Jeffree, C. Ahrends, A. Mill, R. (2017) Edinburgh Journal of Botany  
Leaf cuticle micromorphology has been cited as an important set of taxonomic characters in gymnosperms, but previous studies have largely been based on small sample sizes. The premise of this study was to understand whether external factors affect cuticular micromorphology of Podocarpaceae. Two exam...

Journal Article  
Data used in study [DOI 10.15468/6e8nje](#) [DOI 10.15468/bkzv11](#) [DOI 10.15468/dgbpla](#) [DOI 10.15468/dl.s1xtf6](#) [DOI 10.15468/hja69f](#) [DOI 10.15468/hnhrq3](#) [DOI 10.15468/i9bj5r](#) [DOI 10.15468/ib5ypt](#) [DOI 10.15468/ly60bx](#) [DOI 10.15468/mug7kr](#) [DOI 10.15468/nc6rxy](#) [DOI 10.15468/ucmdjy](#) [DOI 10.15468/vtfbe3](#) [DOI 10.15468/x5ucvh](#) [DOI 10.15468/xy0eoi](#) [DOI 10.15468/yo3mmu](#) [DOI 10.15468/ypoair](#) [DOI 10.15468/z7tjps7](#)

**One for each ocean: revision of the *Bursa granularis* (Röding, 1798) species complex (Gastropoda: Tonnoidea: Bursidae)** [↔](#) **Literature**

Sanders, M. Merle, D. Bouchet, P. Castelin, M. Beu, A. Samadi, G. (2017) Journal of Molluscan Studies  
*Bursa granularis* (Röding, 1798) is a tonnoidean gastropod that is regarded as broadly distributed throughout the Indo-Pacific and tropical western Atlantic. Because of its variable shell, it has received no less than thirteen names, now all synonymized under the name *B. granularis*. We sequenced a fra...

Journal Article  
Data used in study [DOI 10.15468/dl.ma8dle](#)

## NMNH Extant Specimen Records

Published by [National Museum of Natural History, Smithsonian Institution](#)  
Thomas Orrell

OCCURRENCE DATASET | REGISTERED OCTOBER 13, 2016

DATASET METRICS ACTIVITY DOWNLOAD HOME PAGE

7,571,188 OCCURRENCES 218 CITATIONS

Public records of accessioned specimens and observations curated by the National Museum of Natural History, Smithsonian Institution. These data are from the Departments of Botany, Entomology, Invertebrate Zoology, Vertebrate Zoology (Amphibians & Reptiles, Birds, Fishes, and Mammals) and include more than 70 primary type specimen records.

National Museum of Natural History  
Metadata last modified April 16, 2019  
Data last changed April 16, 2019  
Hosted by National Museum of Natural History, Smithsonian Institution  
License: CC BY-NC-ND  
How to cite [DOI 10.15468/hnhrq3](#)

7,571,188 Occurrences 99.7% With taxon match 28% With coordinates 79% With year

183 GEOREFERENCED RECORDS

Updated 11 hours ago © OpenStreetMap contributors, © OpenMapTiles, GBIF.

year 1653 - 2019 EXPLORE

103 OCCURRENCES WITH IMAGES

SEE GALLERY

# Diversity of biodiversity data journals



## Journals that publish biodiversity-related data papers

Please notify us at [communication@gbif.org](mailto:communication@gbif.org) with corrections or updates.

Journal	Publisher	Open Access (license)	APC estimate	Journal Impact Factor (2017)	Scopus CiteScore (2017)
<a href="#">Arxius de Miscel·lània Zoològica</a>	Nat Hist Museum of Barcelona	CC BY	?	-	0.31
<a href="#">Biodiversity Data Journal</a>	Pensoft	CC BY	?	-	0.96
<a href="#">Biogeographia</a>	eScholarship	CC BY	?	-	-
<a href="#">BioInvasions Records</a>	REABIC Journals	CC BY	EUR 300	1.189	1.35
<a href="#">BioRisk</a>	Pensoft	Yes	EUR 300	-	6.00
<a href="#">Biota Colombiana</a>	Humboldt Institute, Colombia	CC BY-NC-ND	?	-	-
<a href="#">BMC Ecology</a>	Biomed Central	CC BY	GBP 1370	2.315	2.35

At least 30 journals

Impact Factor 0–8.7

Scopus CiteScore 0–8.1

APC 0–2014 USD



# Benefits of openness



Global Biodiversity  
Information Facility

- Increases the efficiency of research
- Promotes scholarly rigor and quality of research
- Enables tracking of data use and data citation through DOIs
- Expands the spectrum of academic products through data papers
- Enhances visibility and scope for engagement
- Enables researchers to ask new research questions
- Enhances collaboration and community-building
- Increases the economic and social impact of research
- International conventions and requirements from funding agencies

OPEN ACCESS Freely available online PLOS ONE

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

Heather A. Piwowar\*, Roger S. Day, Douglas B. Fridman

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

**Background:** Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings:** We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ( $p=0.006$ ) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. **Significance:** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar HA, Day RS, Fridman DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. doi:10.1371/journal.pone.0030038

**INTRODUCTION**

Sharing information facilitates science. Publicly sharing detailed research data—sample attributes, clinical factors, patient outcomes, DNA sequences, raw mRNA microarray measurements—with other researchers allows these valuable resources to contribute far beyond their original analysis[1]. In addition to being used to confirm original results, raw data can be used to explore related or new hypotheses, particularly when combined with other publicly available data sets. Raw data is indispensable when investigating and developing study methods, analysis techniques, and software implementations. The larger scientific community also benefits: sharing data encourages multiple perspectives, helps to identify errors, discourages fraud, is useful for training new researchers, and increases efficient use of funding and patient population resources by avoiding duplicate data collection.

Believing that that these benefits outweigh the costs of sharing research data, many initiatives actively encourage investigators to make their data available. Some journals, including the *PLoS* family, require the submission of detailed biomedical data to publicly available databases as a condition of publication[2–4]. Since 2003, the NIH has required a data sharing plan for all large funding grants. The growing open-access publishing movement will perhaps increase peer pressure to share data.

However, while the general research community benefits from shared data, much of the burden for sharing the data falls to the study investigator. Are there benefits for the investigators themselves?

A currency of value to many investigators is the number of times their publications are cited. Although limited as a proxy for the scientific contribution of a paper[5], citation counts are often used in research funding and promotion decisions and have even been assigned a salary-increase dollar value[6]. Boosting citation rate is thus a potentially important motivator for publication authors.

In this study, we explored the relationship between the citation rate of a publication and whether its data was made publicly available. Using cancer microarray clinical trials, we addressed the following questions: Do trials which share their microarray data receive more citations? Is this true even within lower profile trials? What other data-sharing variables are associated with an increased citation rate? While this study is not able to investigate causation, quantifying associations is a valuable first step in understanding these relationships. Clinical microarray data provides a useful environment for the investigation: despite being valuable for reuse and extremely costly to collect, it is not yet universally shared.

**RESULTS**

We studied the citations of 85 cancer microarray clinical trials published between January 1999 and April 2003, as identified in a systematic review by Nizami and Ioannidis[7] and listed in Supplementary Text S1. We found 41 of the 85 clinical trials (48%) made their microarray data publicly available on the internet. Most data sets were located on lab websites (28), with a few found on publisher websites (8), or within public databases (9) in the Stanford Microarray Database (SMD)[8], 6 in Gene Expression Omnibus (GEO)[9], 2 in Array Express[10], 2 in the NCI Gene Expression Data Portal (GDDP)(gddp.nci.nih.gov); some datasets in more than one location). The internet locations of the datasets are listed in Supplementary Text S2. The majority of datasets were made available concurrently with the trial publication, as illustrated within the WayBackMachine internet archives (www.archive.org/web/web.php) for 25 of the datasets and mention of supplementary data within the trial publication itself for 10 of the remaining 16 datasets. As seen in Table 1, trials published in high impact journals, prior to 2001, or with US authors were more likely to share their data.

The cohort of 85 trials was cited an aggregate of 6239 times in 2004–2005 by 3133 distinct articles (median of 1.0 cohort citation per article, range 1–23). The 48% of trials which shared their data received a total of 5334 citations (85% of aggregate), distributed as shown in Figure 1.

**Academic Editor:** John Ioannidis, University of Ioannina School of Medicine, Greece

Received December 13, 2006; Accepted February 26, 2007; Published March 21, 2007

Copyright: © 2007 Piwowar 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: HAP was supported by NLM Training Grant Number 5T32LM007093-19. The NIH had no role in study design, data collection and analysis, writing the paper, or the decision to submit it for publication. The publication contains any solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

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

\* To whom correspondence should be addressed. E-mail: hpiwowar@cmhi.edu

March 2007 | Issue 3 | e308

---

# Potential benefits of publishing data

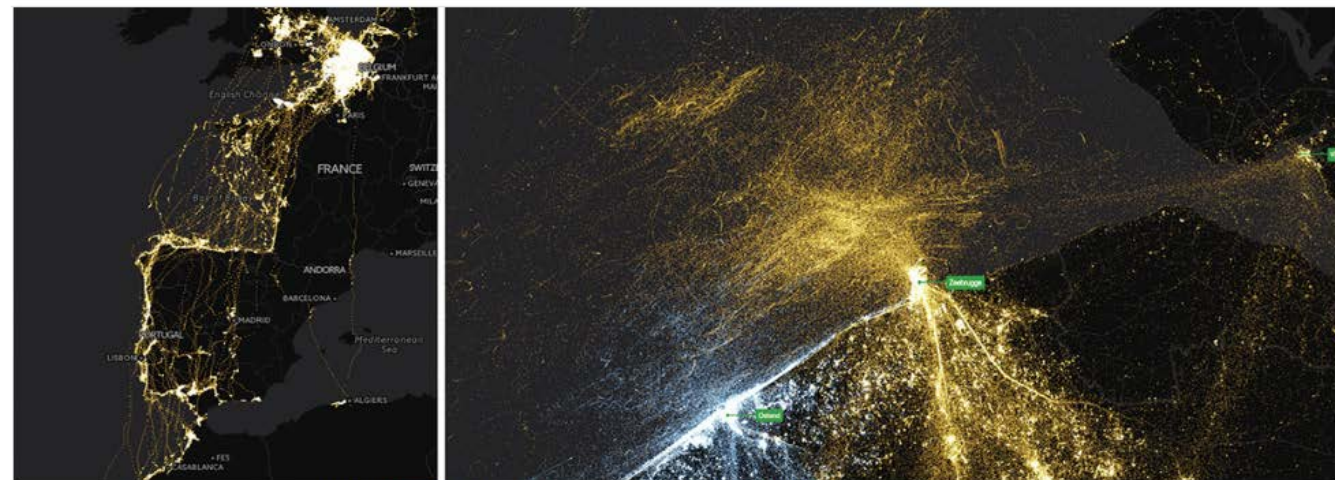
- Enable validity and reproducibility of research
- Increase the visibility and use of research
- Research outputs to answer new questions
- Increase cost-efficiency and data reuse
- Access to new communities - public, policymakers, industry professionals, etc.

# GBIF.org: Data papers

- What is it?
- Webinar video
- Why publish?
- Data paper tools
- Journals with links, costs & metrics
- [helpdesk@gbif.org](mailto:helpdesk@gbif.org)

## Data papers

*Getting scholarly recognition for your datasets*



Map excerpt from data paper by Stienen et al. in Zookeys describing a bird tracking dataset.

### Data paper webinar

A recorded webinar introducing data papers and providing helpful guidelines and tips for writing them is available from [here](#).

### What is a data paper?

A data paper is a peer reviewed document describing a dataset, published in a peer reviewed journal. It takes effort to prepare, curate and describe data. Data papers provide recognition for this effort by means of a scholarly article.

### Data papers explained





**THANK YOU**

[dschigel@gbif.org](mailto:dschigel@gbif.org)

[www.gbif.org](http://www.gbif.org)