

SPARC-ARL Forum: Connecting articles
and data to expand open access to
research

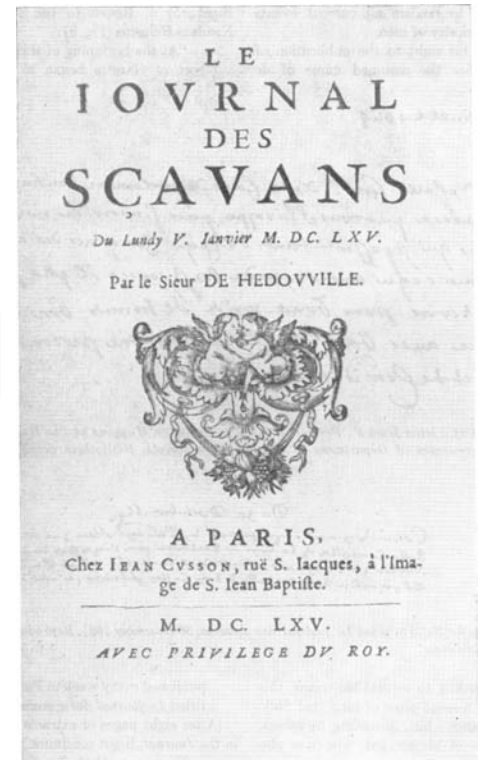
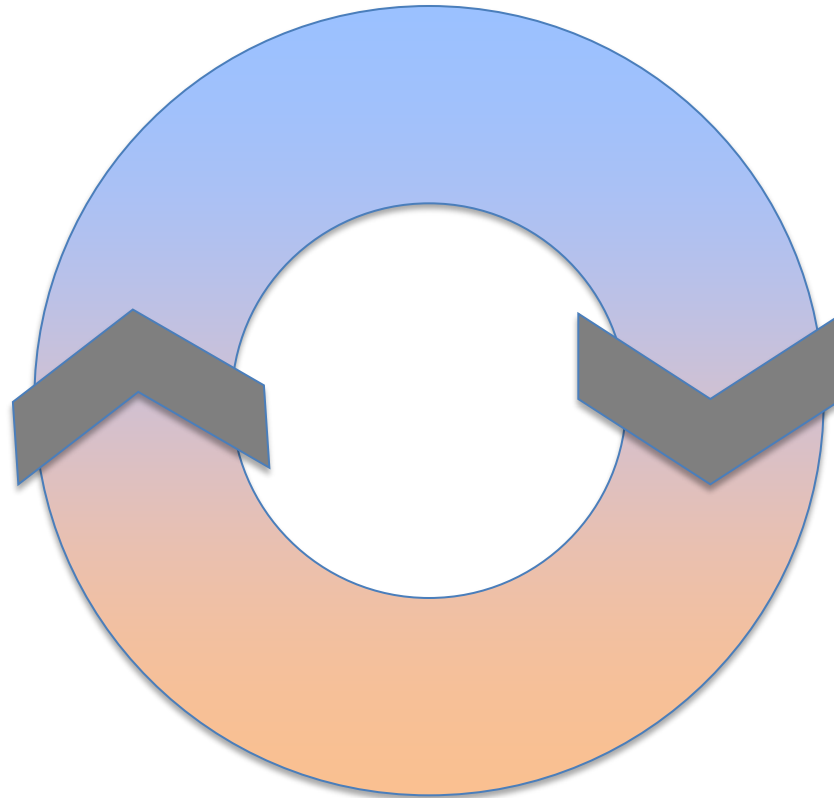
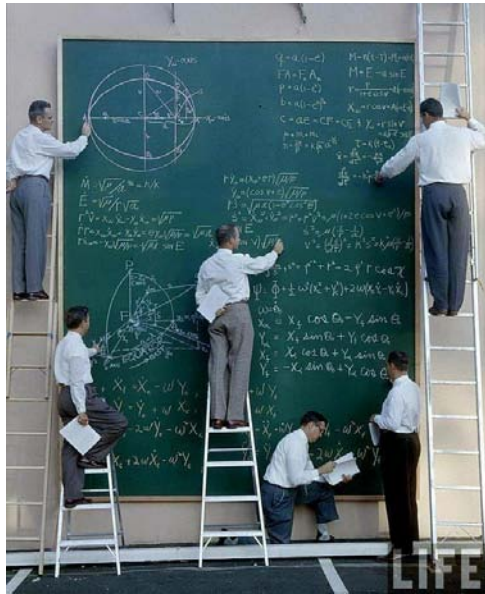
A disciplinary repository perspective

Todd Vision

Associate Director for Informatics, National
Evolutionary Synthesis Center

Department of Biology, University of North
Carolina at Chapel Hill

Traditional scholarly communication



Magnoliids were sister to a monocot + eudicot clade (Fig. 3). Within monocots, the Poaceae (grass family) had 100% support, and within the grasses, the Panicoideae clade had 100% bootstrap support (Fig. 3). In the core eudicot clade, the Caryophyllales (100% support) were sister to the rosids (99% support) and asterids (100% support) (Fig. 3).

There were several differences in the species trees obtained using ML gene trees versus NJ/PP gene trees. For example, the relationships among eucosid lineages differed slightly; however, in both analyses, Malpighiales

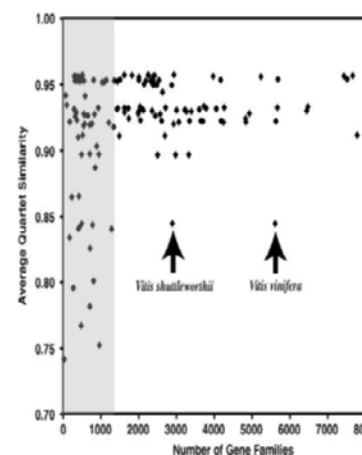


FIGURE 2. Average quartet similarity for each taxon among bootstrap trees. Each point in the graph represents a single taxon. The x-axis shows the number of gene families trees that have data for that taxon. The y-axis shows the average percentage frequency of (four taxon statements) containing the taxon that are identical to two bootstrap trees. The shaded area in the graph contains all taxons that are present in less than 1300 gene families.

In this study, we used GTP to find species trees that minimize the total number of duplications across a collection of nearly 18,896 plant gene trees. The sequencing includes extensive collections of existing data that have rarely been used for plant phylogenetics.

A computational system to select candidate genes for complex human traits

Table 2. Tests using susceptibility genes for complex human traits

Complex trait	OMIM	Review(s) ^a	Gene ^b	Reviews	
				Rank	Total
Age-related macular degeneration	603075	15094132; 15350892	<i>CFH</i>	7263	137
ARMD (second run)	603075	N/A ^c	<i>LOC387715</i>	—	137
			<i>C2</i>	—	—
Alzheimer's disease	104300	15225164	<i>CFB</i>	—	—
Asthma	600807	12810182; 14551038	<i>LOC439999</i>	—	135
Autism	209850	11733747; 12142938	<i>NPSR1</i>	1117	138
Celiac disease	212750	12907013; 12699968; 14592529	<i>EN2</i>	98	136
			<i>MYO9B</i>	234	130
Myocardial infarction	608446	15861005; 16041318	<i>LT44H</i>	122	140
Parkinson's disease	168600	16026116; 16278972	<i>SEM45A</i>	4548	134
Rheumatoid arthritis	180300	15478157; 12915205	<i>PTPN22</i>	333	132
Schizophrenia	181500	15340352; 16033310	<i>FCRL3</i>	3743	132
Type 1 diabetes mellitus	222100	12270944; 11921414	<i>ENTH</i>	10013	146
			<i>SUMO4</i>	12123	142
			<i>PTPN22</i>	165	142
			<i>IL2RA</i>	130	142
Type 2 diabetes mellitus	125853	15662000; 15662001; 15662002; 15662003	<i>CTLA4</i>	78	142
			<i>TCF7L2</i>	2911	139
Totals				725 ^d	138

^aPubMed IDs of review articles used in corpus.

^bFor references see Methods section. HUGO approved gene symbols used to identify genes.

^cNo suitable review corpus available (see Methods section).

^dThe OMIM record is insufficiently detailed and was not used.

^eMedian result.

^fMean result.

or sum

$$\phi_{n,i} = \sum_{j=1}^n s_{ij}$$

of the transformed scores for gene i .

The fourth method, referred to as in4, differs from the other three by considering both the score of a gene within a data source as well as the number of genes returned for that data source. First, a transformed score s_{ij} is obtained.

$$s_{ij} = \frac{r_{ij}}{\sum_{i=1}^n r_{ij}}$$

The transformed gene scores are then summed together to provide a final score for each gene.

$$\phi_{n,i} = \sum_{j=1}^n s_{ij} \frac{r_{ij}}{G}$$

where r_{ij} is the number of genes returned for source j and

$$G = \sum_{j=1}^n r_{ij}$$

2.4 Implementation

The CASAR algorithms were written using Perl version 5.8.1 and Java version 1.4.2. The vector space similarity searches were performed using a modified version of the Perl module

SearchVec
a 2003.02.01
downloaded
Intermedia
under the s

2.5 Selection

To assess if genes were evidence of disorders. *PTPN22* (Pimm et al. 2006), *CFH* (Pimm et al. 2006), *LT44H* (Pimm et al. 2006), *CFB* (Pimm et al. 2006), *LOC439999* (Pimm et al. 2006), *ENTH* (Pimm et al. 2006), *SUMO4* (Pimm et al. 2006), *FCRL3* (Pimm et al. 2006), *IL2RA* (Pimm et al. 2006), *CTLA4* (Pimm et al. 2006), *TCF7L2* (Pimm et al. 2006). Each disorder was assigned a score of 1 or 0 (Table 2). (Wheeler et al. 2006). We removed genes that were not in the OMIM database. Thus, the

1794 Kries et al.

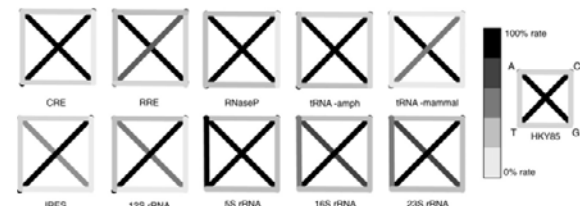


FIG. 3.—Heatmap of nucleotide substitution rates for each alignment. The color scale indicates the rate of substitution for each nucleotide substitution model for each molecule. Within a molecule, rates were scaled to the maximum rate (black). Diagonal lines depict transitions. The edges of the square depict transversions. The 100% model, which was used for the rate ratios reported throughout this article, is shown for comparison on the right.

Substitution Patterns in RRE

We examined 3 possible explanations for the surprising result that $j_{\text{RRE}} > j_{\text{CRE}}$. First, because both the RRE and CRE secondary structures occur within coding regions, we examined the possibility that the difference between j_{RRE} and j_{CRE} is driven by selection on the protein sequence. We recalculated j_{RRE} and j_{CRE} for both molecules using only data from 4-fold degenerate sites in paired and unpaired regions. In CRE, the presence of codons affects the estimates in the predicted direction (4-fold degenerate sites $j_{\text{RRE}} > j_{\text{CRE}}$, all sites $j_{\text{RRE}} > j_{\text{CRE}}$) through the 4-fold sites over- and under-represented in the predicted pattern. We had less power to compare 4-fold degenerate sites at the paired and unpaired sites of RRE because there were too few 4-fold degenerate unpaired sites, and there was insufficient sequence variability at these sites. However, the 4-fold degenerate paired sites did show

a higher j_{RRE} ($j_{\text{RRE}} = 5.761$ with 95% CI [4.79–10.40]) than the paired sites as a whole ($j_{\text{RRE}} = 5.421$ with 95% CI [3.51–5.20]). This suggests that the presence of protein-coding constraints does impede compensatory evolution at paired sites in RNA secondary structures, although it does not explain why j_{RRE} would be "greater" than j_{CRE} .

Second, we examined the possibility that we had used a nonrepresentative sample of RRE sequences. To confirm that the observed substitution patterns in RRE were not specific to the particular set of HIV sequences we examined (which were all derived from subtype B), we estimated j_{RRE} and j_{CRE} from 2 additional RRE alignments of sequences drawn from higher taxonomic levels: sequences from different subtypes (1 sequence each from A, B, C, F, G, H, J, and K) and sequences from different groups (1–2 sequences each from M, N, and O) of HIV. In both these alignments, the results were qualitatively similar to those for subtype B (j_{RRE} was significantly higher than j_{CRE}).

Third, we considered whether the RRE estimates were disproportionately influenced by a portion of the molecule that experiences a type of selection that differs from the molecule as a whole. We systematically removed each stem-loop of RRE and reestimated j_{RRE} and j_{CRE} for the resulting partial structures. The j_{RRE} and j_{CRE} estimates were qualitatively similar for all these partial structures (table 5).

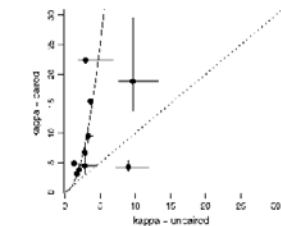


FIG. 4.—Transition-transversion rate ratios ($j_{\text{RRE}}/j_{\text{CRE}}$) for each alignment. The dashed line represents a 1:1 relationship between j_{RRE} and j_{CRE} . The solid line represents the predicted relationship $j_{\text{RRE}}/j_{\text{CRE}} = 1$. Note that the CRE data point is from the analysis of 4-fold degenerate sites in paired and unpaired regions.

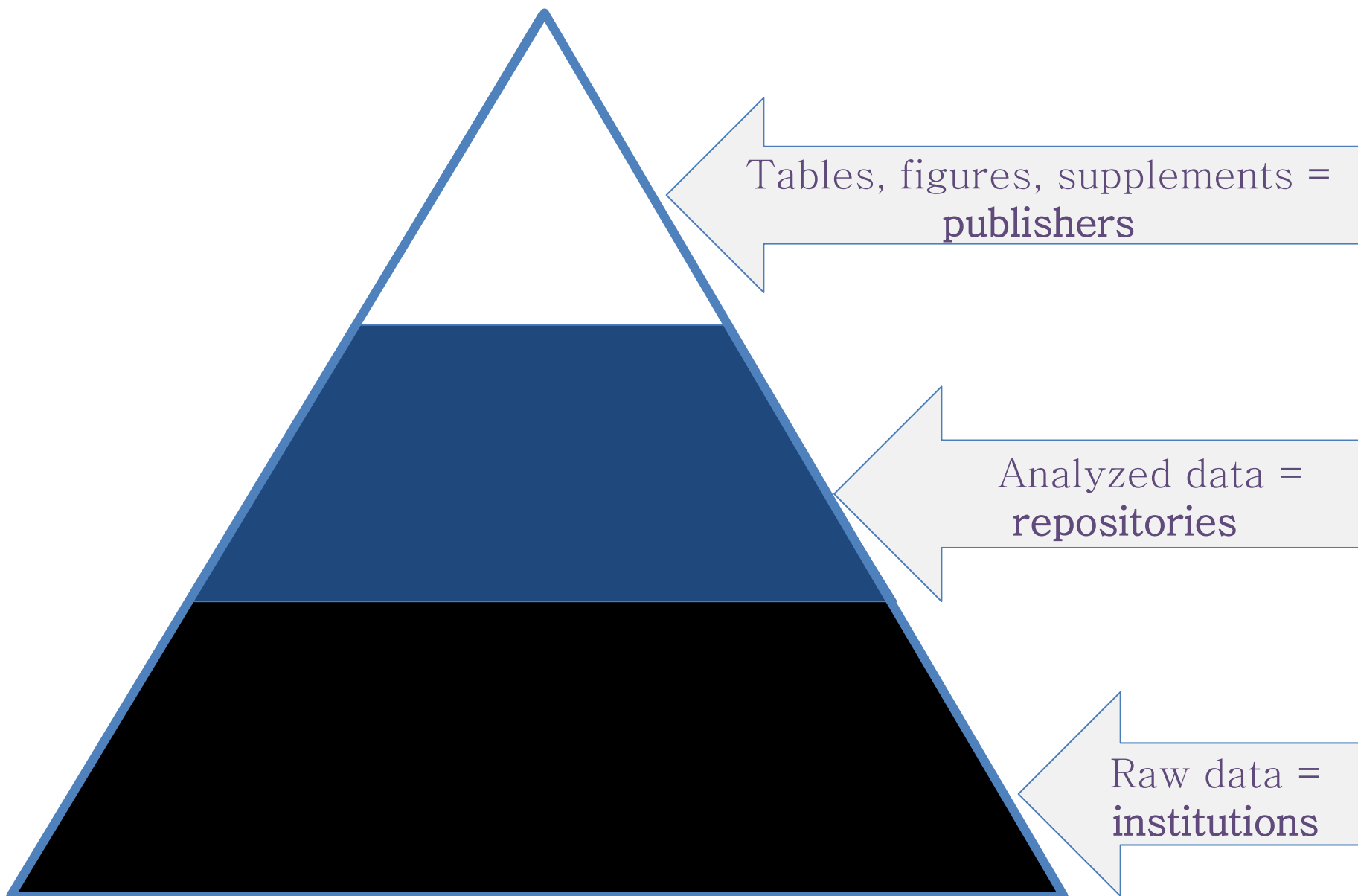
Table 3. Transition-Transversion Rate Ratios ($j_{\text{RRE}}/j_{\text{CRE}}$)

Structure	j_{RRE}	j_{CRE}	$j_{\text{RRE}}/j_{\text{CRE}}$	k
RRE	5.10	4.21	0.01	540.05 ^a
CRE	0.50	15.34	3.00	73.40 ^a
CRE	12.62	22.80	2.48	177.16 ^a
SS tRNA	3.70	4.44	2.82	36.00 ^a
TS tRNA	3.74	3.75	2.02	100.04 ^a
TS tRNA	2.57	3.06	1.71	120.11 ^a
A tRNA	0.04	9.48	3.30	204.73 ^a
M tRNA	11.98	18.78	9.06	122.24 ^a
TS tRNA	3.90	6.09	2.83	131.01 ^a
RNase P	2.98	4.80	1.30	59.21 ^a

^a tRNA value significant at $P < 0.0001$.

What happens to
the data behind
the million plus
research articles
published each
year?





Data journals

(GIGA)ⁿ
SCIENCE

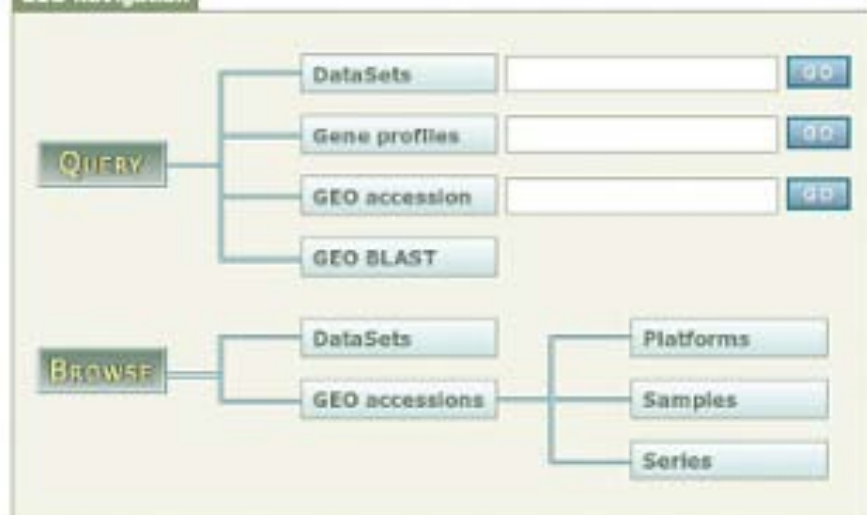
nature publishing group 



SCIENTIFIC DATA



Gene Expression Omnibus: a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles. [More information »](#)

GEO navigation

Submitter login

[» New account](#)
[» Recover password](#)

Site contents
Public data

Platforms	11,154
Samples	882,875
Series	36,158
DataSets	3,101

Documentation

[Overview](#) | [FAQ](#) | [Find](#)
[Submission guide](#)
[Linking & citing](#)
[Journal citations](#)
[Construct a Query](#)
[Programmatic access](#)
[DataSet clusters](#)
[GEO announce list](#)
[Data disclaimer](#)
[GEO staff](#)

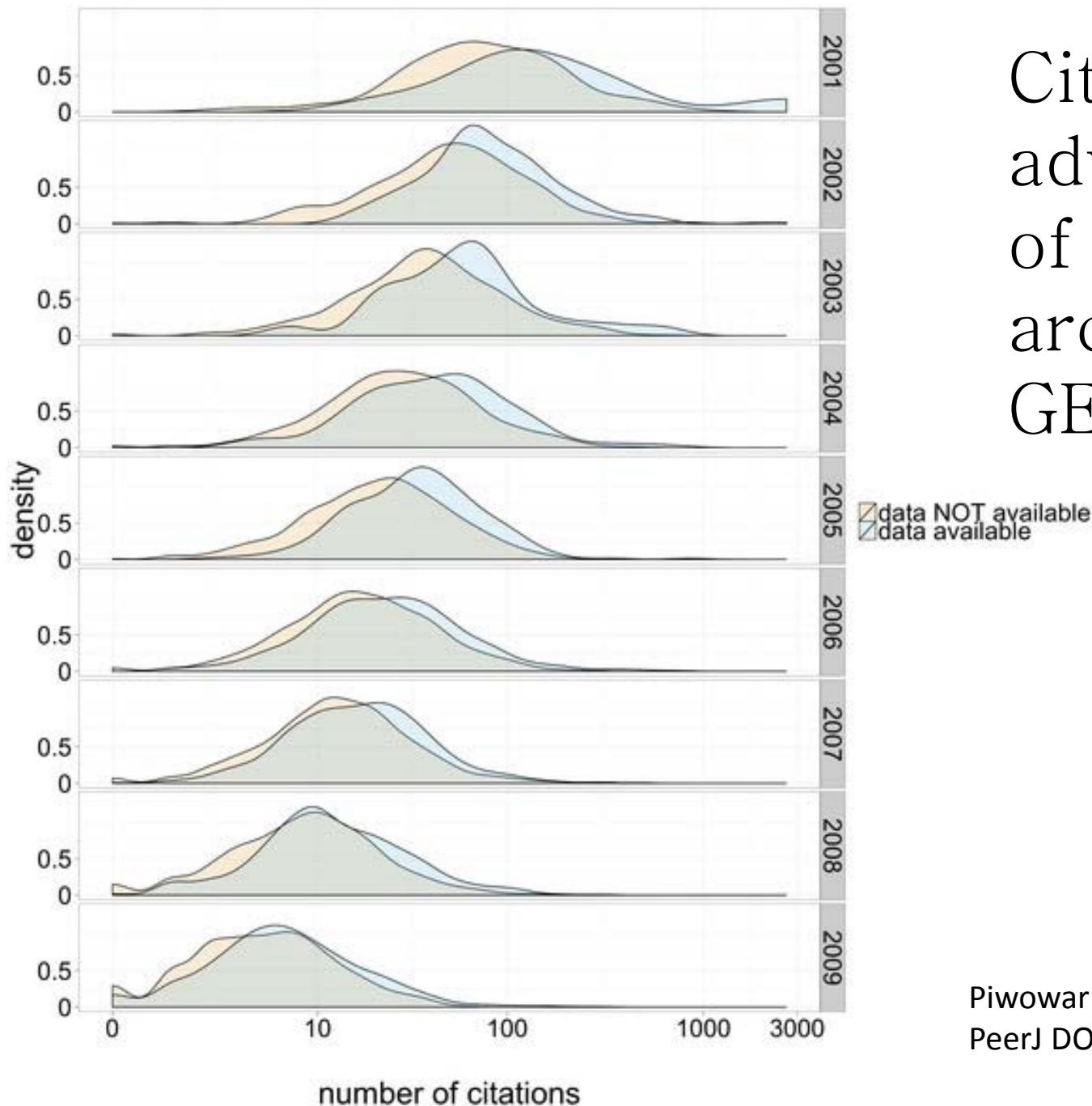
Query & Browse

[Repository browser](#)
[GEO2R](#)
[FTP site](#)
[GEO Profiles](#)
[GEO DataSets](#)

Submit

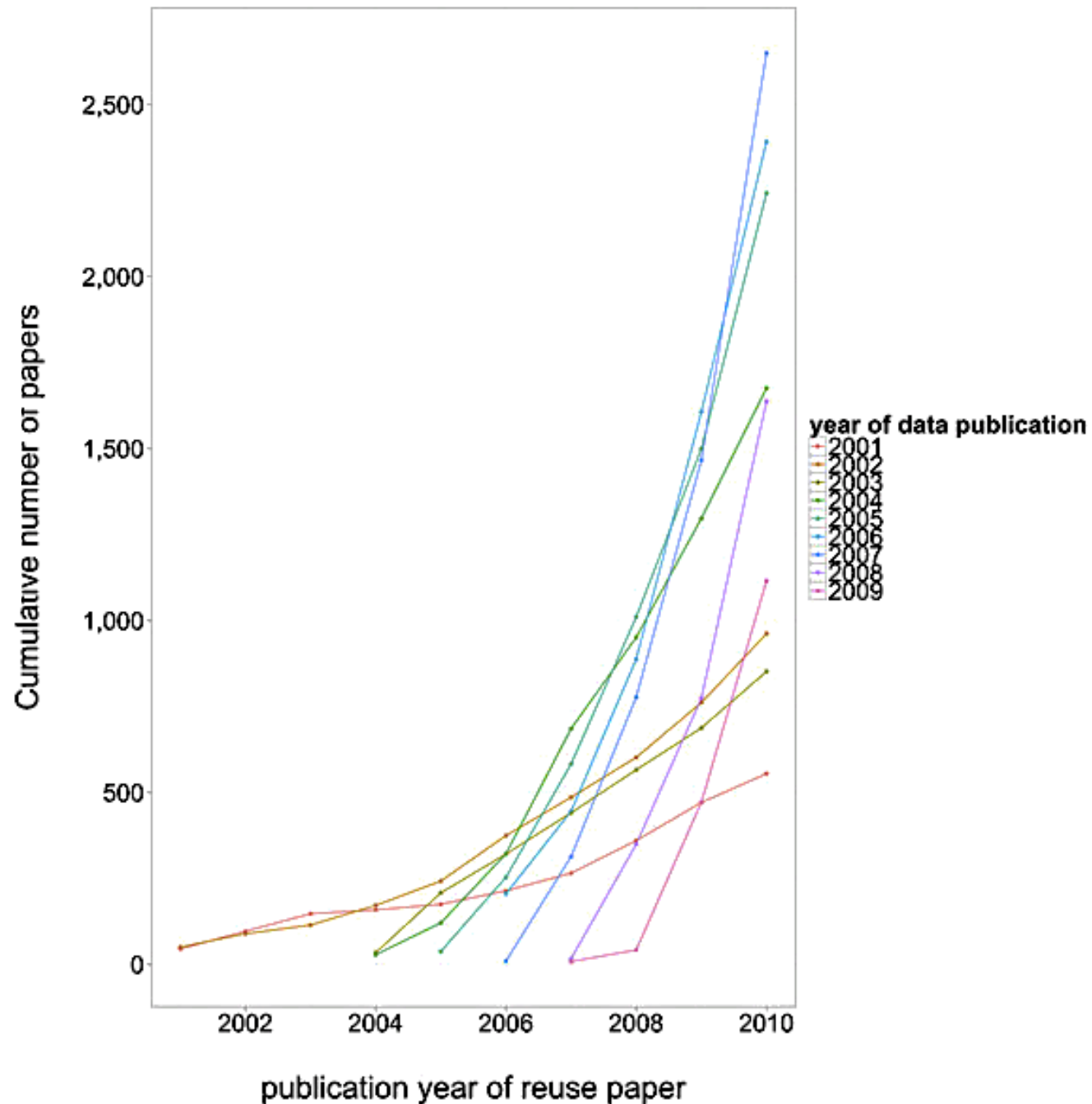
[New account](#)

Citation advantage of having archived GEO data

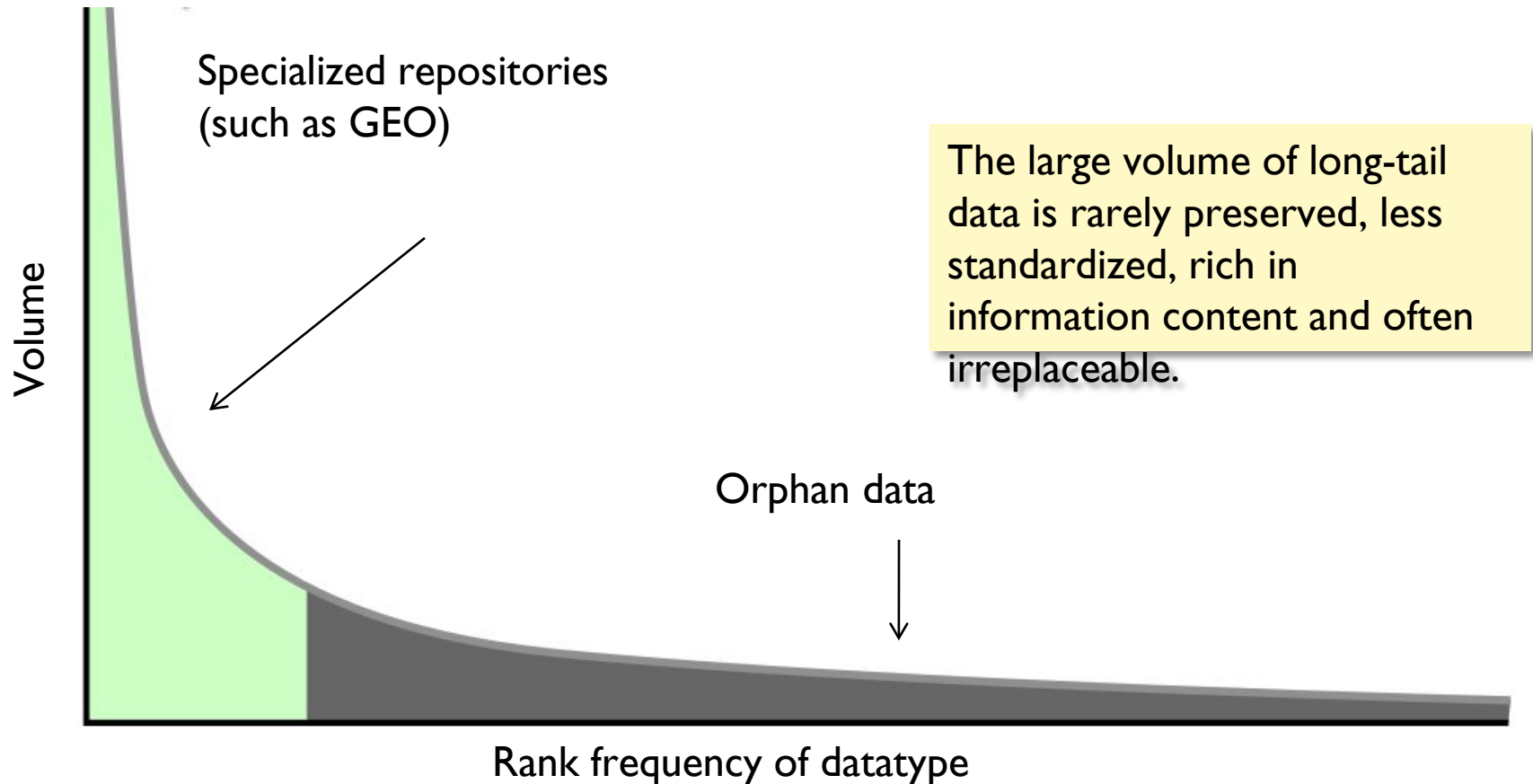


Piwowar and Vision (2013),
PeerJ DOI:10.7717/peerj.175

1.5-fold
reuse of
GEO data
after 5
years, and
the rate of
reuse is
growing

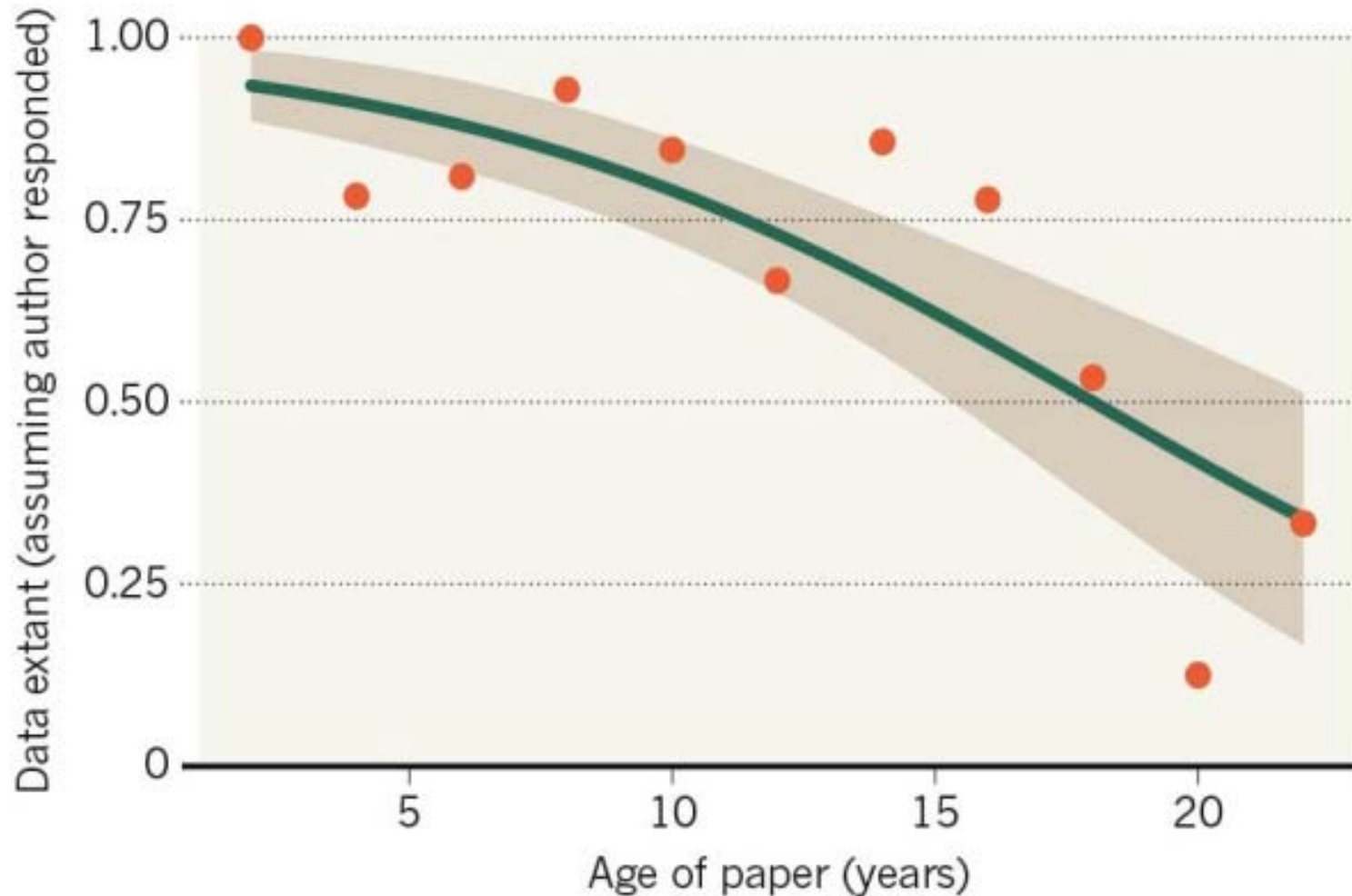


The long tail of orphan data



After Heidorn (2008) <http://hdl.handle.net/2142/9127>

80% of ecology data irretrievable after 20 years



Brussels Declaration on STM Publishing

“Raw research data should be made freely available to all researchers. Publishers encourage the public posting of the raw data outputs of research. Sets or sub-sets of data that are submitted with a paper to a journal should wherever possible be made freely accessible to other scholars”

Signatories include Elsevier, Nature Publishing Group, Springer, Oxford U Press, Wiley-Blackwell. 46 publishers and 13 trade organizations as of 2012

http://www.stmassoc.org/public_affairs_brussels_declaration.php

Joint Data Archiving Policy (JDAP)

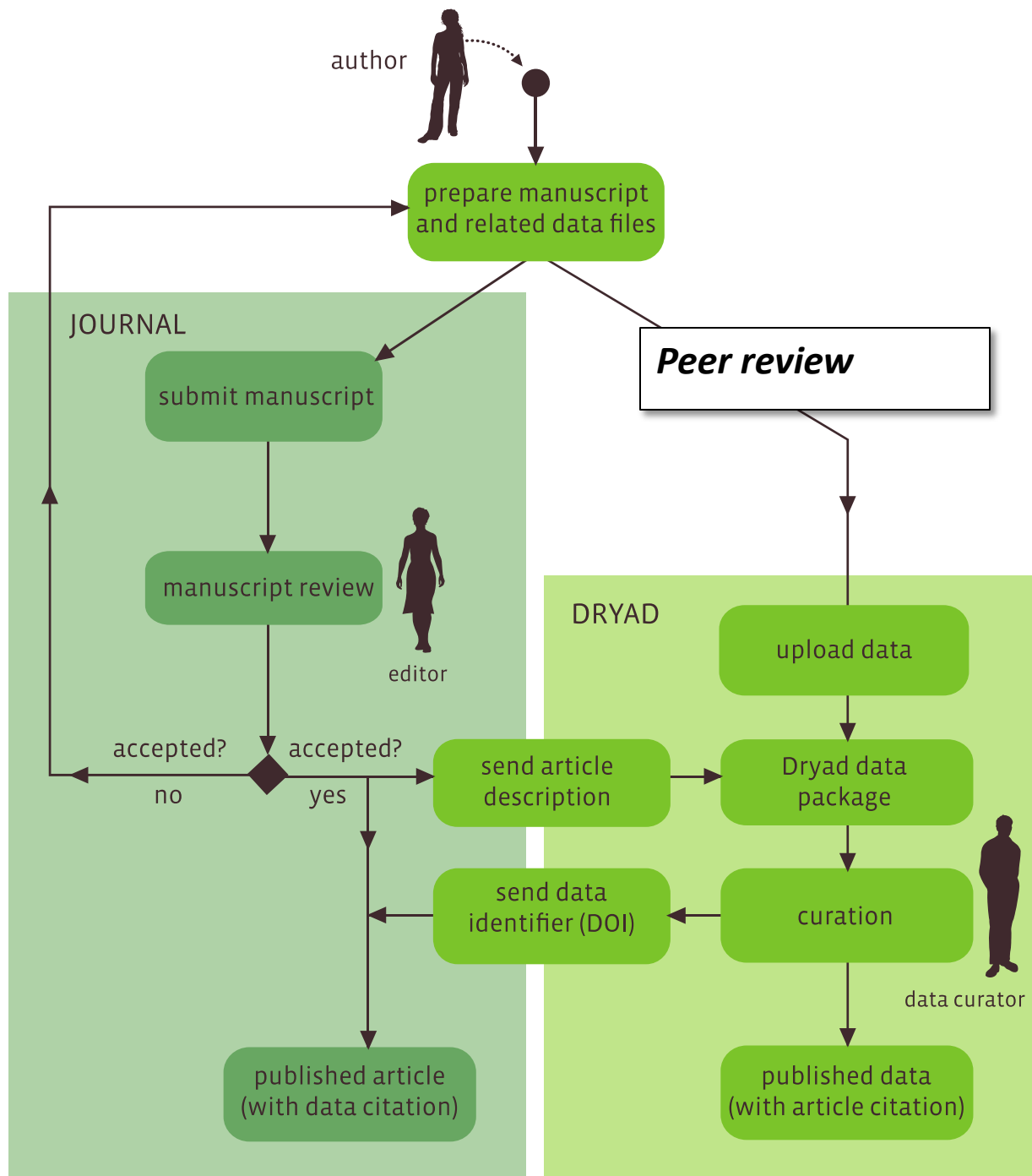
Data are important products of the scientific enterprise, and they should be **preserved** and **usable** for decades in the future.

As a condition for publication, data supporting the results in the article should be deposited in an **appropriate public archive**.

Authors may elect to **embargo** access to the data for a period up to a year after publication.

Exceptions may be granted at the discretion of the editor, especially for sensitive information.

<http://datadryad.org/pages/jdap>




[Submit data now](#)
[How and why?](#)

Search for data

[Go](#)
[Advanced search](#)

Browse for data

[Recently published](#)
[Popular](#)
[By Author](#)
[By Journal](#)

Recently Published Data

Phadke S, Zufall R, Paixao T, Pham S, Pham T (2013) Data from: Genetic background alters dominance relationships between mat alleles in the ciliate *Tetrahymena thermophila*. *Journal of Heredity* [doi:10.5061/dryad.qv731](https://doi.org/10.5061/dryad.qv731)

Chiu C, Jost L, Chao A (2013) Data from: Phylogenetic beta diversity, similarity, and differentiation measures based on Hill numbers. *Ecological Monographs* [doi:10.5061/dryad.vr4k2](https://doi.org/10.5061/dryad.vr4k2)

Salgado-Salazar C, Rossmann AY, Chaverri P (2013) Data from: Not as ubiquitous as we thought: taxonomic crypts, hidden diversity and cryptic speciation in the cosmopolitan fungus *Theloneotria discophora* (Nectriaceae, Hypocreales, Ascomycota). *PLoS ONE* [doi:10.5061/dryad.q3s66](https://doi.org/10.5061/dryad.q3s66)

Morgan EMJ, Green BS, Murphy NP, Strugnelli JM (2013) Data from: Investigation of genetic structure between deep and shallow populations of the southern rock lobster, *Jasus edwardsii* in Tasmania, Australia. *PLoS ONE* [doi:10.5061/dryad.656gf](https://doi.org/10.5061/dryad.656gf)

Robinson BW (2013) Data from: Evolution of growth by genetic accommodation in Icelandic freshwater stickleback. *Proceedings of the Royal Society B*

Be part of Dryad



Publishers, societies, universities, libraries, funders, and other stakeholder organizations are invited to become [members](#). Tap into an active knowledge-sharing

network, receive discounts on submission fees, and help shape Dryad's future.

[Submission integration](#) is a free service that allows publishers to coordinate manuscript and data submissions. It makes submitting data easy for researchers; makes linking articles and data easy for journals; and enables confidential review of data prior to publication.



Submission fees support the cost of keeping Dryad's content free to use. Flexible [pricing plans](#) provide volume discounts.

How to enable the different stakeholders (funders, publishers, societies, libraries, etc.) to all contribute to the governance and sustainability of an open access resource?

Dryad Members (30 and growing)

American Association for the Advancement of Science * American Society of Naturalists * The American Genetic Association * Botanical Society of America * British Ecological Society * BMJ Publishing Group, Ltd. * The Biological Journal of the Linnean Society * BioMed Central * Ecology Letters * Ecological Society of America * Elementa: Science of the Anthropocene * European Society for Evolutionary Biology * Evolutionary Applications * The Genetics Society * German National Library of Medicine * HighWire * Molecular Ecology * Molecular Ecology Resources * Molecular Phylogenetics and Evolution * Oikos * Oxford University Press * The Paleontological Society * Pensoft Publishers * PLOS * The Royal Society * Society for Molecular Biology and Evolution * Society for the Study of Evolution * Society of Systematic Biologists * United States Fish and Wildlife Service * Wiley-Blackwell

Data Publication Charges

Payment Plan	Member	Non-member	Minimum purchase
1. Voucher Plan	\$65 per data package	\$70 per data package	25 vouchers
2. Deferred Payment Plan	\$70 per data package	\$75 per data package	1 yr contract
3. Subscription Plan	annual fee based on \$25 per published research article	annual fee based on \$30 per published research article	2 yr contract
For individuals: Pay on submission	NA	\$80 per data package, payable by the submitter	1 data package

- *Waivers*
 - for authors from countries classified by the World Bank as low-income or lower-middle-income economies.
- *Surcharges*
 - for non-integrated submissions.
 - for data packages greater than 10 GB.

Integrated Journals and Costs for Submitters

Show all amounts in USD\$

All journals with either [integrated data submission](#) or sponsored [Data Publishing Charges](#) are listed below. Note that for large data packages, submitters will be asked to pay an additional \$15 for the first GB beyond 10GB and \$10 for each GB thereafter.

Journal	Submission				Payment
	Integrated	Submit: R: Prior to A: On Acceptance	Data embargo allowed	Metadata hidden until publication	
<i>If your journal is not listed below...</i>	<i>N</i>	<i>A</i>	<i>Y</i>	<i>Y</i>	<i>\$90</i>
American Naturalist, The	Y	A	Y	N	American Society of Naturalists
Biodiversity Data Journal	N	A	Y	Y	Pensoft
Biological Journal of the Linnean Society	Y	A	Y	N	\$80
Biology Letters	Y	A	Y	Y	The Royal Society
BioRisk	Y	R	Y	N	Pensoft
BMJ Open	Y	R	N	N	\$80
Comparative Cytogenetics	Y	R	Y	N	Pensoft
Ecological Monographs	Y	A	Y	Y	Ecological Society of America

[Submit data now](#)

[How and why?](#)

Search for data

Enter keyword, DOI, etc.

[Advanced search](#)

[Go](#)

Be part of Dryad

Learn more about:

[Membership](#)

[Submission integration](#)

[Pricing plans](#)



CHONE
CANADIAN HEALTHY OCEANS NETWORK

news

Animal Without a Brain Can "Sneeze," Surprising Study Shows

Jan 15, 2014

CHONE's Dr. Verena Tunnicliffe to receive The Murray A. Newman Award

Dec 20, 2013

Arctic Sea Ice Loss Recorded in Algae

Nov 29, 2013

[More news...](#)

you are here: [home](#)

Welcome to the Canadian Healthy Oceans Network website!

The Canadian Healthy Oceans Network, or **CHONE** (pronounced Ko-nee), is an NSERC strategic network focused on biodiversity science for the sustainability of Canada's three oceans. CHONE, a marine worm that occurs in all of Canada's three oceans is the equivalent of the canary in the coal mine: it indicates a healthy benthic habitat.

The network includes about 150 researchers from 14 universities across Canada, the federal Department of Fisheries and Oceans, and seven other government laboratories, to carry out thirty-five collaborative research projects in three interconnected themes.



CHONE's research in Canada's three oceans: the Pacific, the Arctic and the Atlantic, showing Discovery Corridors in white. These locations are focal areas for biodiversity research.

Data conclaves



Home	About TRY	Data Portal	Contact	Registration
Datasets	Explore Data	Download Data	Request Data	Contribute Data

Data Portal

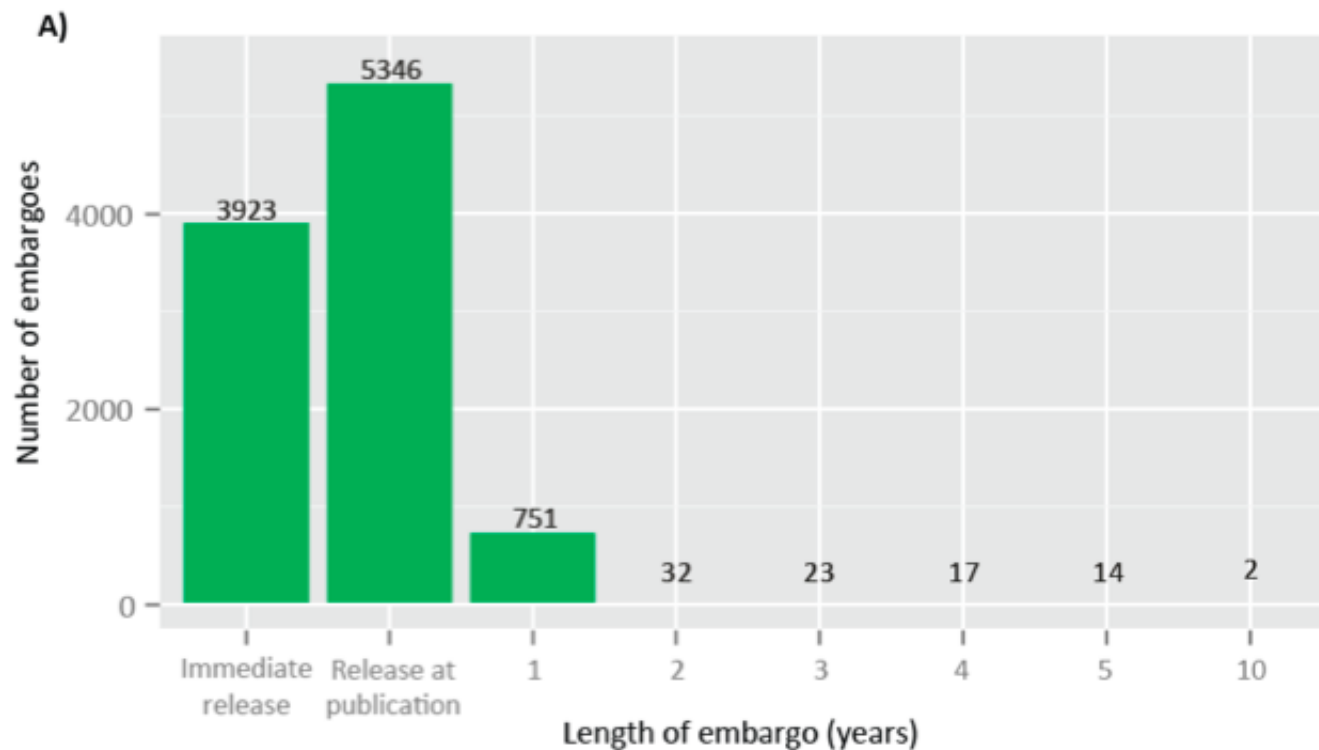
The TRY database contains public and non-public data. The public data are free to be downloaded; only a registration is needed for web security reasons. However, the majority of data in the TRY database is non-public. These data are available upon request.

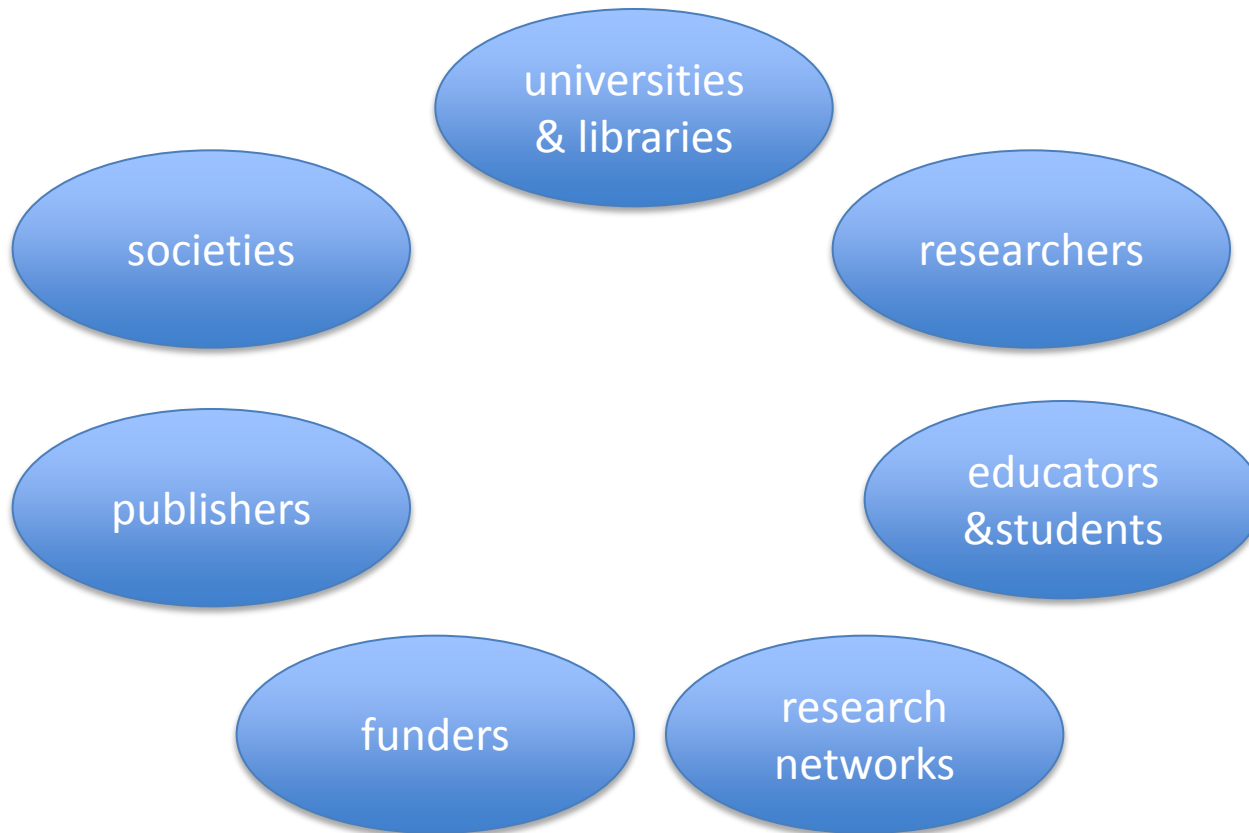
The TRY Data Portal enables you to search the TRY database ([Data Explorer](#)) and to [download](#) the **free data**. The TRY Data Portal also supports the submission of a proposal to [request non-public data](#) and provides information how to [contribute data](#).

Datasets	Information about all datasets in TRY with links, tables, and maps.
Explore Data	Explore the contents of the TRY database (trait data are not available this way).
Download Data	Direct download of public files (mainly supplemental material).
Request Data	A wizard to submit a proposal to TRY to receive trait data .
Contribute Data	Learn how to contribute data to TRY.



Fig. S1 (A) Embargo selections of Dryad data authors for the 10,108 files in Dryad (datadryad.org) deposited from inception to September 20 2013. Data include only journals for which the authors had the option of selecting an embargo. **(B)** Long-term embargoes (> 1 year) by journal that granted them. Data were obtained from [42].







What can and should academic libraries be doing to develop, maintain, host, govern and sustain *disciplinary* research data repositories to ensure that they remain open for the long term?

To learn more



- <http://datadryad.org>
- <http://blog.datadryad.org>
- <http://datadryad.org/wiki>
- <http://code.google.com/p/dryad>
- dryad-users@nescent.org
-  @datadryad
-  Dryad

Data in traditional journals

Function	Role of journal
Registration	✓
Certification	✓ for methods
Dissemination	✓ limited
Archiving	✓ limited

My Collection

1 products (expand all)

add products

refresh metrics

download ▾

Tweet 0

dataset

Data from: Towards a worldwide wood economics spectrum

(2009) Zanne, Lopez-Gonzalez, Coomes et al. *Dryad Digital Repository*

highly cited

highly viewed

highly discussed

viewed by scholars



discussed by public



cited by scholars



What do these numbers mean?

[Sign In](#)[Marked List \(0\)](#)[EndNote](#)[ResearcherID](#)[Citation Alerts](#)[Journal Alerts](#)[Saved Searches](#)[Log Out](#)[Help](#)[All Databases](#)[Select a Database](#)[Data Citation Index](#)[Additional Resources](#)[Search](#)[Cited Reference Search](#)[Advanced Search](#)[Search History](#)Data Citation IndexSM[<< Back to results list](#)

Record 1 of 1

Record from Data Citation IndexSM[Add to Marked List \(0\)](#)

Send to:

[my.endnote.com](#)Boron in **Antarctic** granulite-facies rocks: under what conditions is boron retained in the middle crust?From Repository: [Australian Antarctic Data Centre](#)Author(s): [Carson, Christopher](#); [Grew, Edward](#)Source: Australian Antarctic Data Centre Source URL: [http://gcmd.nasa.gov/KeywordSearch/Metadata.do?](http://gcmd.nasa.gov/KeywordSearch/Metadata.do?Portal=amd_a&MetadataView=Full&MetadataType=0&KeywordPath=8&OrigMetadataNode=AADC&EntryId=ASAC_2350)

2012) Published Year: 2009

2012) Published Year: 2009

Cited References: 0

Abstract: This metadata record describes data collected as part of ASAC project 2350 - Boron in Antarctic granulite-facies rocks: under what conditions is boron retained in the middle crust? As a direct result of the field mapping during this project (and previous fieldwork by myself and others) 'we' have produced a 1:25000 map of the geology of the Larsemann Hills. This was collaboration between the AAD and Geoscience Australia (with considerable assistance by Phil O'Brien and Henk Broilma) and published by GA earlier in 2007. The map is referenced below. Additionally, several papers are linked to this record, plus copies of the field report and two documents which details the photos taken, and the locations of the field sites. Extended abstract The Larsemann Hills region is dominated by two major lithological associations, a Palaeoproterozoic felsic/mafic orthogneiss complex (Sostrene Orthogneiss) which occurs as basement to a sequence of pelitic, psammitic and felsic paragneiss (supergroup = Brattstrand Paragneiss) and felsic intrusives. The depositional age of the Brattstrand Paragneiss sequences are controversial but isotopic data suggest derivation from the basement Sostrene Orthogneiss. Current geochronology indicates that the region experienced medium to low pressure granulite-facies metamorphism during the Early Palaeozoic (~500 Ma). Although the paragneiss sequences record no evidence of earlier metamorphism, relicts of a previous metamorphic event at ~1000 Ma are preserved in the Sostrene Orthogneiss. Within the Larsemann Hills region, the Early Palaeozoic event is characterised by peak metamorphism of ~7 kbar at ~800-850 degrees C, with the post-peak evolution characterised by decompression, with some cooling, to 4 kbars at 750 degrees C, then to 2-3 kbar at 600-650 degrees C during final stages of orogenesis, with exhumation largely driven by crustal extension. Tectonic models generally argue for a continental-continental collisional scenario, with thermal input derived from a thinned mantle lithosphere. Structural evolution The various high-grade structural frameworks proposed by different workers have been distilled by Fitzsimons (1997) into three major events Da, Db and Dc which broadly correlates D1, D2 and D3 proposed by Stuwe et al. (1989), Thost et al. (1994), Carson et al. (1995b) and D1, D2 and D3-D6 of Dirks and Hand (1995) and D3, D4 and D5 of Fitzsimons and Harley (1991). Within the Larsemann Hills, the dominant outcrop structures are attributed to Db (using the nomenclature of Fitzsimons, 1997). Db can be sub-divided into low and high strain zones, low strain zones preserve complex multiple fold generations that fold lithological layering (Da) and high-strain zones which transpose Da into a new planar gneissosity, Db. Similarly, Dc high-strain zones overprint and locally transpose Db structures, which are completely replaced by a new gneissic layering. Sc, and mineral lineation, Lc, in the northern and southern regions of the Larsemann Hills. Much of the Larsemann Hills is, therefore, a window of Dc low-strain in which Db structures are preserved, although these are reorientated by large, relatively open, upright Dc low-strain folds. Fold hinges and mineral extension lineations preserved on anisoleic surfaces within both domains are co-linear and have a characteristic orientation: easterly to southerly plunging for Db and consistently

This dataset has been cited 58 times

Times Cited: 58

[Create Citation Alert](#)

This article has been cited 58 times in Web of Knowledge.

WADOSKI ER. [not available] EOS T AGU S, 2009.

WADOSKI ER. [not available] GEOL SOC AM ABSTR PR, 2009.

GREW ES. [not available] GRAN GRAN 2009 JUL 1, 2009.

[\[view all 58 citing articles \]](#)*View the articles that have cited this dataset*

Cited References: 0

Additional information

[How to cite this Resource](#)*See how to cite this dataset*

Suggest a correction

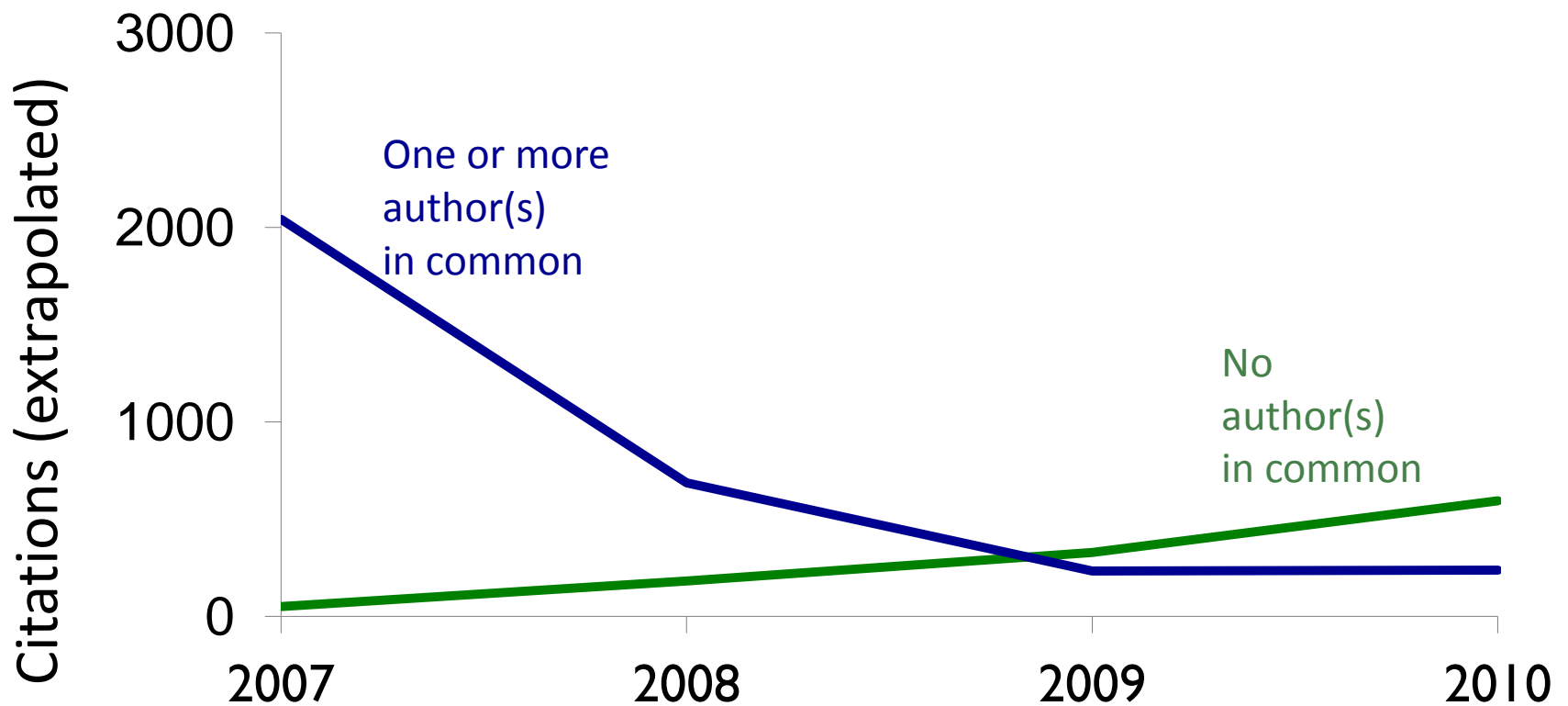
If you would like to improve the quality of the data in this record, please [suggest a correction](#).

TABLE III^a.*Measurements of Twenty-eight Adult and Young Females which Perished.*

	TOTAL LENGTH.	ALAR EXTENT.	WEIGHT.	LENGTH OF BEAK AND HEAD.	LENGTH OF HUMERUS.	LENGTH OF FEMUR.	LENGTH OF TIBIO- TARSUS.	WIDTH OF SKULL.	LENGTH OF KEEL OF STERNUM.
37 ♀	155	240	26.3	31.4	.709	.710	1.125	.614	.815
38 ♀	156	240	25.8	31.5	.715	.678	1.127	.597	.812
39 ♀	160	242	26.	32.0	.740	.732	1.157	.597	.854
40 ♀	152 ¹	232 ³	23.23	30.3	.6762	.683	1.048	.590	.780
41 ♀	160	250	26.5	31.7	.741	.731	1.187	.615	.886
42 ♀	155	237	24.2	31.	.727	.723	1.118	.610	.787
43 ♀	157	245	26.9	32.2	.766	.751	1.2272	.620	.841
44 ♀	165 ³	245	27.7	32.12	.7801	.7573	1.195	.633	.895
45 ♀	153 ²	231 ²	23.9	30.1	.6803	.6623	1.0423	.592	.781
46 ♀	162	239	26.1	30.3	.709	.685	1.092	.587	.911
47 ♀	162	243	24.6	31.6	.741	.729	1.162	.605	.840
48 ♀	159	245	23.6	31.8	.727	.700	1.129	.610	.855
49 ♀	159	247	26.	30.9	.711	.666	1.098	.580	.7492
50 ♀	155	243	25.	30.9	.730	.711	1.127	.598	.839
51 ♀	162	252	24.8	31.9	.752	.738	1.180	.615	.875
52 ♀	152 ¹	230 ¹	22.82	30.4	.682	.664	1.0423	.5511	.7341
53 ♀	159	242	24.8	30.8	.717	.667	1.090	.575	.809
54 ♀	155	238	24.6	31.2	.706	.702	1.102	.588	.7583
55 ♀	163	249	30.52	33.41	.767	.7671	1.2073	.6401	.896
56 ♀	163	242	24.8	31.	.713	.713	1.128	.607	.813
57 ♀	156	237	23.9	31.7	.718	.716	1.090	.611	.800
58 ♀	159	238	24.7	31.5	.726	.701	1.145	.600	.800
59 ♀	161	245	26.0	32.1	.751	.704	1.142	.607	.819
60 ♀	155	235	22.61	30.7	.695	.692	1.119	.584	.771
61 ♀	162	247	26.1	31.9	.751	.735	1.157	.618	.802
62 ♀	153 ²	237	24.8	30.6	.732	.718	1.172	.594	.802
63 ♀	162	245	26.2	32.5	.728	.731	1.102	.614	.832
64 ♀	164	248	26.1	32.3	.739	.707	1.159	.592	.823
Average . .	158	241	25.3	31.4	.726	.709	1.131	.601	.820
General average for 64 birds . . .	160	245	25.8	31.5	.728	.709	1.128	.601	.834

Bumpus HC (1898) The Elimination of the Unfit as Illustrated by the Introduced Sparrow, *Passer domesticus*. *Biological Lectures from the Marine Biological Laboratory*: 209-226.



Proprietary interest in one's own GEO data is short-lived



Data from: Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring



Files in this package

To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.  

Title	Sequences after initial processing in QIIME (seqs.fna, library 1)
Downloaded	61 times
Description	Raw sequence data from the Roche 454 GS FLX sequencer, region 1 (split_library_output_1). These data are the output of the command: split_libraries.py -m 454_Map.txt -f 1.TCA.454Reads.fna -q 1.TCA.454Reads.qual -o split_library_output_1/ -l 100 -L 700 -H 9 -M 2 -b 10
Download	README.txt (3.755Kb) View File Details
Download	seqs.fna (20.96Mb) View File Details

Title	Sequences after initial processing in QIIME (seqs.fna, library 2)
Downloaded	33 times

Journals integrated with Dryad

American Naturalist • [Biological Journal of the Linnean Society](#) • Biology Letters • [BioRisk](#) • BMJ Open • [Comparative Cytogenetics](#) • Ecological Monographs • [Ecology and Evolution](#) • Ecology Letters • [Elementa: Science of the Anthropocene](#) • eLife • [Evolution](#) • Evolutionary Applications • [Functional Ecology](#) • gms German Medical Science • [GMS Medizinische Informatik, Biometrie und Epidemiologie](#) • Heredity • [International Journal of Myriapodology](#) • Journal of Animal Ecology • [Journal of Applied Ecology](#) • Journal of Ecology • [Journal of Evolutionary Biology](#) • Journal of Fish and Wildlife Management • [Journal of Heredity](#) • Journal of Hymenoptera Research • [Journal of Open Public Health Data](#) • Journal of Paleontology • [Methods in Ecology and Evolution](#) • Molecular Ecology • [Molecular Ecology Resources](#) • MycoKeys • [Nature Conservation](#) • NeoBiota • [Palaeontology](#) • Paleobiology • [PhytoKeys](#) • PLOS Biology • [PLOS Genetics](#) • Subterranean Biology • [Systematic Biology](#) • ZooKeys

The fallibility of published results

- More than half of published papers contain statistical errors.
- 5–10% of papers contain errors that change the conclusions.

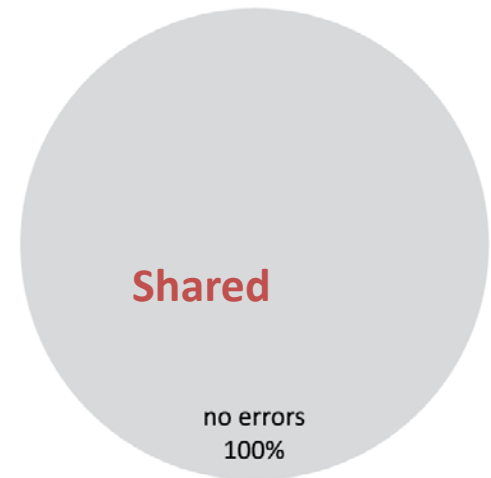
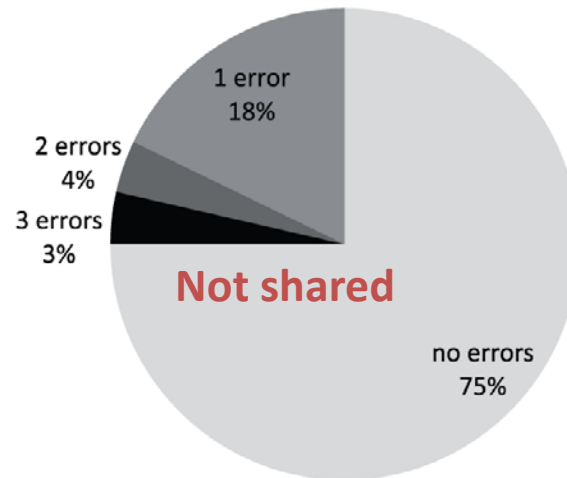
Gore et al. (1977) Br. Med. Journal 85–87

Hurlbert and White (1993) Bull. Marine Sci. 53, 128–153

McGuigan (1995) Br. J. Psychiatry 167, 683–688

It's the dodgy stuff that doesn't get shared

**Reporting errors
concerned with $p < .05$:**

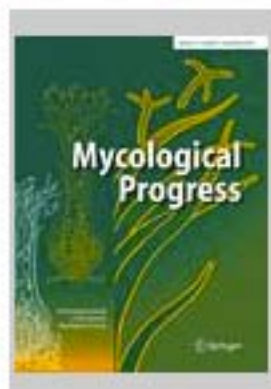


When using this data, please cite the original article:

Chave J, Coomes D, Jansen S, Lewis SL, Swenson NG, Zanne AE (2009) Towards a worldwide wood economics spectrum. *Ecology Letters* 12: 351–366. doi:10.1111/j.1461-0248.2009.01285.x

Additionally, please cite the Dryad data package:

Zanne AE, Lopez-Gonzalez G, Coomes DA, Ilic J, Jansen S, Lewis SL, Miller RB, Swenson NG, Wiemann MC, Chave J (2009) Data from: Towards a worldwide wood economics spectrum. Dryad Digital Repository. doi:10.5061/dryad.234



ORIGINAL ARTICLE

Lepidostroma vilgalysii, a new basidiolichen from the New World

Brendan P. Hodkinson, Jessie K. Uehling and Matthew E. Smith

Materials and Methods

The analyzed alignment file was formatted for and deposited in TreeBASE (<http://www.treebase.org/>; study accession URL: <http://purl.org/phylo/treebase/phyloids/study/TB2:S12029>; submission 12029) and analysis files were made available through the Dryad data repository (<http://datadryad.org/>; data package URL: <http://dx.doi.org/10.5061/dryad.j1g5dh23>; data package DOI: 10.5061/dryad.j1g5dh23; Hodkinson et al. 2011).

References

- Diego, C.
- Hodkinson BP, Lendemer JC, Esslinger TL (2010) *Parmelia barrenoae*, a macrolichen new to North America and Africa. *North American Fungi* 5(3):1–5
- Hodkinson BP, Uehling JK, Smith ME (2011) Data from: *Lepidostroma vilgalysii*, a new basidiolichen from the New World. Dryad Digital Repository. doi:10.5061/dryad.j1g5dh23
- Honegger R (1996) Mycobionts. In: Nash TH III (ed) *The biology of lichens*. Cambridge University Press, Cambridge, pp 25–36
- Katoh K, Misawa K, Kuma K-I, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066

Dryad Governance

- Independent nonprofit with a board elected by its membership
 - Members may include any stakeholder organization: journal, society, publisher, funder, research institution
 - Board need not be drawn from membership organizations
- The membership provides an invaluable forum
 - Coordinating data policies
 - Raising awareness of research, best practices, etc.
 - Helping shape Dryad's business model

Contrast of Dryad with (typical) SOM

	Dryad	SOM
Discoverable: indexed and exposed to both web and bibliographic search engines	✓	✗
Identifiable: DataCite DOIs within articles serve as permanent, resolvable identifiers	✓	✗ *
Permanent: processes in place to promote preservation (incl. format migration)	✓	✓ / ✗ **
Curated: quality control by both automated processes and human inspection	✓	✗ *
Ease of deposit: streamlined deposit, allowance for large and complex datasets	✓	✓ / ✗ **
Formatted for reuse: support for non-PDF file formats	✓	✓ / ✗ **
Updatable: new versions of data files can be added, metadata can be enhanced	✓	✗
Support for embargoes: can delay release of data in accordance with journal policy	✓	✗
Free reuse: no paywall, clear terms of reuse (all data released under CC Zero)	✓	✓ / ✗ **
Economy of scale: cost efficiency from shared infrastructure	✓	✓ / ✗ **

* A few publisher SOM sites are exceptions to the general rule

** Practices differ among publishers, see Smit (2011), doi:10.1045/january2011-smit

What is the return on investment?

- A rigorous framework is lacking, but...
- Marginal cost of data archiving is small
 - Dryad's DPC is ~5% of a PLOS ONE APC
 - Less than 0.5% of the costs per article for a research funder (\$25–33K)
- Does publication of the data have 1/20 the value of the article?

Vision (2011) Open data and social contract of scientific publishing. *BioScience*, 60(5):330-330
Piwowar H, Vision TJ, Whitlock MC (2011) Data archiving is a good investment. *Nature* 473:285

Revenue model

- Deposit fees
 - Pay upfront for both ingest and long-term preservation
 - At 10K data packages/yr, ~ €40 per deposit
 - Non-operational costs covered by membership fees, grants for R&D, etc
- Payment schemes
 - Journal subscription
 - Per-deposit fees, covered by journal (pre or post-submission)
 - Author-pays (last resort!)
- Neutral w.r.t. the business model of journals/publishers

Beagrie N, Eakin-Richards L, Vision TJ (2009) Business models and cost estimation: Dryad repository case study. iPRES 2010

Costs

- Moderate economies of scale are required
 - At 10K packages/yr, <\$50/deposit, depending on curation
- What are the costs for SOM?
 - Journal of Clinical Investigation: \$300 flat fee
 - Ecological Archives: \$250 <10Mb, more fees beyond that
 - FASEB: \$100 per file

Beagrie N, Eakin-Richards L, Vision TJ (2009) Business models and cost estimation: Dryad repository case study. iPRES 2010

Arrival and Diversification of Caviomorph Rodents and Platyrrhine Primates in South America

CÉLINE PO

¹Laboratoire de Paléontologie, Place E. Bataillon
²Department of Biology

Abstract.—Platyrrhine period of isolation from investigated the timing sequences from three calibration constraints relaxed molecular clock along the branch lead respectively). The time dates of origin (upper platyrrhine primates (or 20.1 ± 3.3) Mya, a Considering both the primates from Africa during the Middle or arrival of primates and glaciations, and these in the Oligocene, and Caviomorphia; colonization



Access
To read this story in full
nature.com > Journal home

Letter

Nature 442, 684–687 (1 May 2006)

Minimal ProtoH

D. Chourroux¹, F. Delsuc², Renfer¹, M. F. Jensen¹, B.

1. Sars International Centre Thormøhlensgt. 55, 5008 B
2. Laboratoire de Paléontologie l'Evolution, UMR5554-CNRS, Montpellier Cedex 05, France
3. American Hospital of Paris sur-Seine Cedex, France
4. Children's Hospital and USA
5. Department of Biological Sciences 1, Irvine, California

Correspondence to: D. C requests for materials sh (Email: Daniel.chourroux@genbank.org) (Email: Ulrich.technau@genbank.org) GenBank with accession

Bilateral animals have patterning the main Comparison of Hox an that both clusters orig cluster named ProtoH least the precursors of

The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils?

Emmanuel J. P. Douzery^{*†}

^{*}Departments of Paleontology, Phys la Recherche Scientifique), University of Reading, Whiteknights Mixte de Recherche 7622, Centre Na and [†]Canadian Institute for Advanced

Edited by Wen-Hsiung Li, University

The use of nucleotide and amino understanding of the timing of Molecular estimates of divergence al and are generally much m fossil record. The limited num and pervasive variations in ev sources of such discrepancies. amino acid sequences of 129 determine the divergence time animals, fungi, plants, and v variations in their evolutionar teinity of the fossil record, we clock simultaneously calibrated We show that, according to 95 otic kingdoms diversified 950– mals diverged from choanofla bated age of the split between occurred 642–761 Mya. The o robust with respect to prior calibrations. Interestingly, these much more recent than those o global molecular clock, yet bila ~100 million years more ancie



- Article Views
- Abstract
- Full Text
- Full Text (PDF)
- Figures Only
- Supporting Online Material
- Podcast Interview
- VERSION HISTORY
- 330/6009/1381 (most recent)
- science.1194167v1
- Article Tools
- Save to My Folders
- Download Citation
- Alert Me When Article is Cited

OPEN ACCESS Freely available online

PLOS PATHOGENS

A New Malaria Agent in African Hominids

Benjamin Ollomo^{1*}, Patrick Durand^{2*}, Franck Prugnolle^{2*}, Emmanuel Douzery³, Céline Arnathau², Dieudonné Nkoghe¹, Eric Leroy^{1,4}, François Renaud^{2*}

¹Unité des Maladies Virales Émergentes, Centre International de Recherches Médicales de Franceville, Franceville, Gabon, ²Laboratoire Génétique et Evolution des Maladies Infectieuses, UMR 2724 CNRS-IRD, IRD Montpellier, Montpellier, France, ³Laboratoire de Paléontologie, Phylogénie & Paléobiologie, Institut des Sciences de l'Evolution (UMR 5554 CNRS), Université Montpellier II, Place E. Bataillon, Montpellier, France, ⁴Unité Émergence des Pathologies Virales, UMR 190 IRD-Université de la Méditerranée, CIRMF, Franceville, Gabon

Abstract

Plasmodium falciparum is the major human malaria agent responsible for 200 to 300 million infections and one to three million deaths annually, mainly among African infants. The origin and evolution of this pathogen within the human lineage is still unresolved. A single species, *P. reichenowi*, which infects chimpanzees, is known to be a close sister lineage of *P. falciparum*. Here we report the discovery of a new *Plasmodium* species infecting Hominids. This new species has been isolated in two chimpanzees (*Pan troglodytes*) kept as pets by villagers in Gabon (Africa). Analysis of its complete mitochondrial genome (5529 nucleotides including Cyt b, Cox I and Cox II genes) reveals an older divergence of this lineage from the clade that includes *P. falciparum* and *P. reichenowi* (~21 ± 9 Myrs ago using Bayesian methods and considering that the divergence between *P. falciparum* and *P. reichenowi* occurred 4 to 7 million years ago as generally considered in the literature). This time frame would be congruent with the radiation of hominoids, suggesting that this *Plasmodium* lineage might have been present in early hominoids and that they may both have experienced a simultaneous diversification. Investigation of the nuclear genome of this new species will further the understanding of the genetic adaptations of *P. falciparum* to humans. The risk of transfer and emergence of this new species in humans must be now seriously considered given that it was found in two chimpanzees living in contact with humans and its close relatedness to the most virulent agent of malaria.

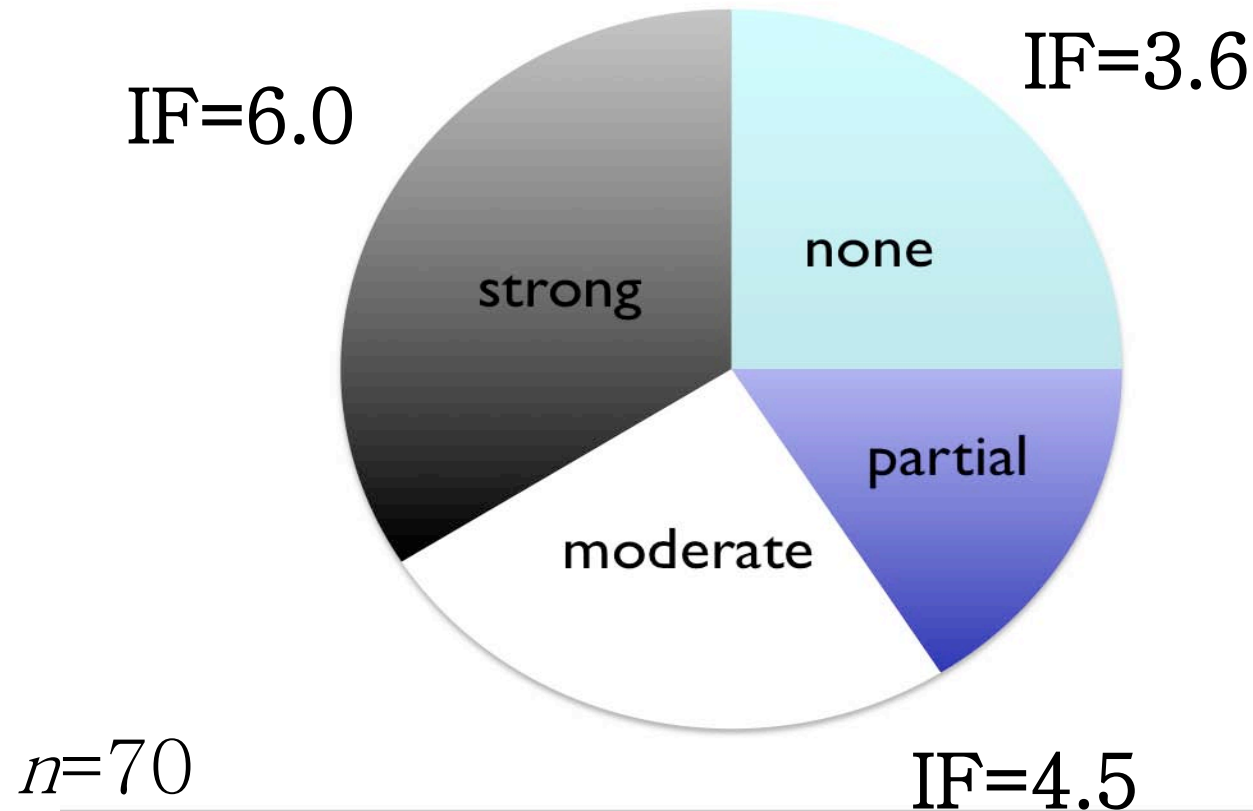
Citation: Ollomo B, Durand P, Prugnolle F, Douzery E, Arnathau C, et al. (2009) A New Malaria Agent in African Hominids. PLoS Pathog 5(5): e1000446. doi:10.1371/journal.ppat.1000446

20 papers from Delsuc and Douzery going back to 2002

What do the data look like?

- Size
 - Mean number of files: 2.4
 - Mean storage requirement: 9.7 Mb
- Format types
 - 60% plain text (unrecognized content standard)
 - 11% Excel
 - 5% PDF
 - 3% Word

High impact factor journals have stronger data archiving policies





Plan

Collect

Excel:mac 2008

Assure

Describe

Preserve



Analyze

Integrate

Discover

DataONE
Powered by MERCURY

Semantic Tools
for Data Management

MATLAB
The Language of Technical Computing



PLOS Data Policy

Data Access for the Open Access Literature

Margaret Winker, MD

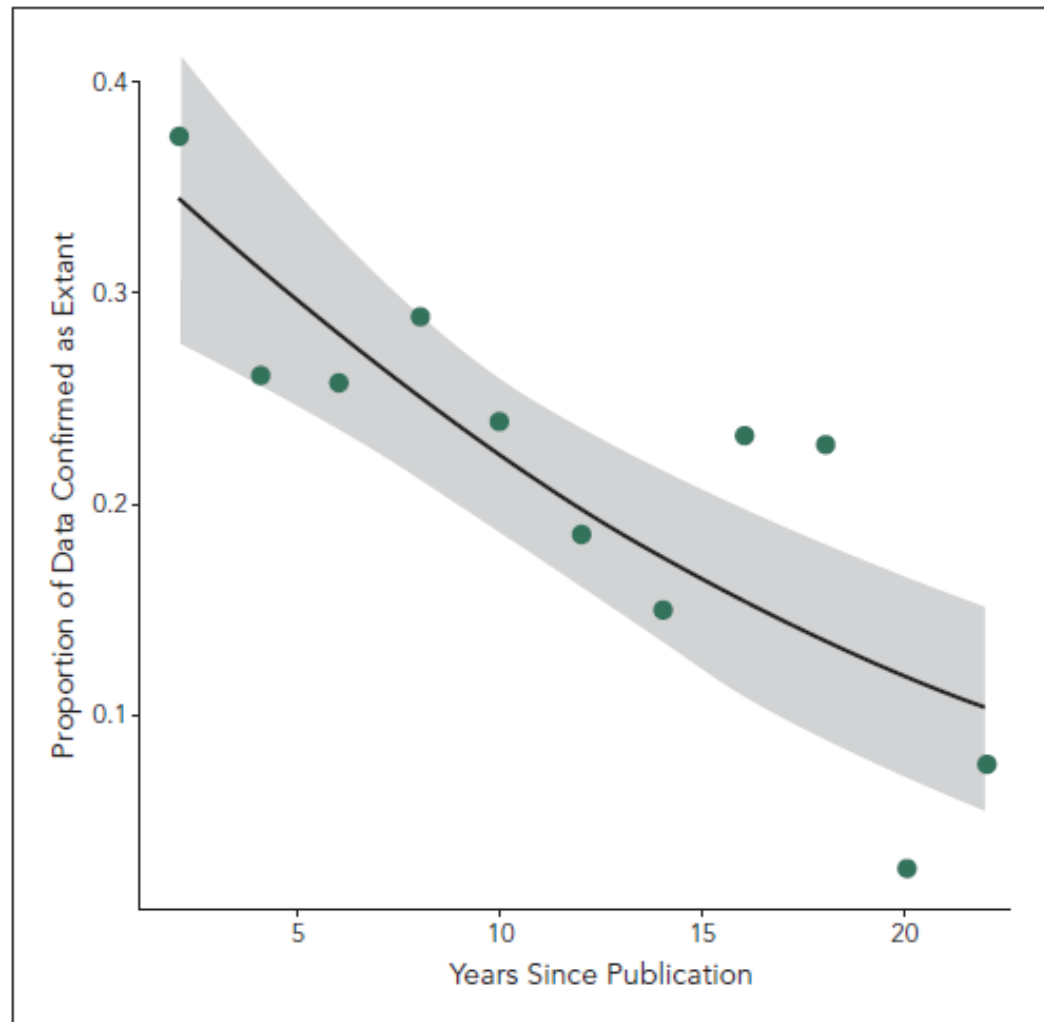
Senior Research Editor, *PLOS Medicine*

January 25, 2014

Data availability allows

- ✓ Replication
- ✓ Validation
- ✓ New analysis
- ✓ Interpretation
- ✓ Inclusion in meta-analyses
- ✓ Facilitation of reproducibility of research
- ✓ Better 'bang for the buck' out of research investment

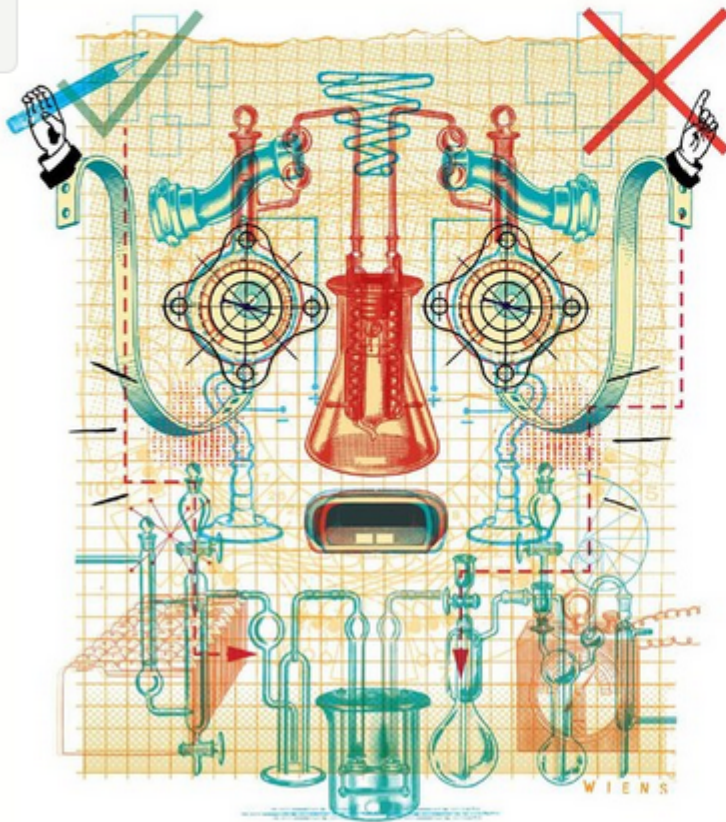
Figure 5. Proportion of Papers With Data Available, 1991-2011



From: How Does the Availability of Research Data Change With Time Since Publication? Timothy H. Vines and colleagues, Abstract (podium), Peer Review Congress, 2013

New Truths That Only One Can See

JAN. 20, 2014



Since 1955, [The Journal of Irreproducible Results](#) has offered “spoofs, parodies, whimsies, burlesques, lampoons and satires” about life in the laboratory. Among its greatest hits: “Acoustic Oscillations in Jell-O, With and Without Fruit, Subjected to Varying Levels of Stress” and “Utilizing Infinite Loops to Compute an Approximate Value of Infinity.” The good-natured jibes are a backhanded celebration of science. What really goes on in the lab is, by implication, of a loftier, more serious nature.

It has been jarring to learn in recent

NY Times, January 20, 2014

Table 1: Reproducibility of research findings

Preclinical research generates many secondary publications, even when results cannot be reproduced.

From

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Nature **483**, 531–533 (29 March 2012) | doi:10.1038/483531a

[◀ back to article](#)

Table 1: Reproducibility of research findings

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles [*]	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

Quantifying Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome

Daniel Garijo, Sarah Kinnings, Li Xie, Lei Xie, Yinliang Zhang, Philip E. Bourne , Yolanda Gil 

Published: November 27, 2013 • DOI: 10.1371/journal.pone.0080278

Article

About the Authors

Metrics

Comments

Related Content



Abstract

Introduction

Methods and Analysis

Discussion

Conclusions

Supporting Information

Author Contributions

References

Reader Comments (0)

Figures

Abstract

How easy is it to reproduce the results found in a typical computational biology paper? Either through experience or intuition the reader will already know that the answer is with difficulty or not at all. In this paper we attempt to quantify this difficulty by reproducing a previously published paper for different classes of users (ranging from users with little expertise to domain experts) and suggest ways in which the situation might be improved. Quantification is achieved by estimating the time required to reproduce each of the steps in the method described in the original paper and make them part of an explicit workflow that reproduces the original results. Reproducing the method took several months of effort, and required using new versions and new software that posed challenges to reconstructing and validating the results. The quantification leads to “reproducibility maps” that reveal that novice researchers would only be able to reproduce a few of the steps in the method, and that only expert researchers with advance knowledge of the domain would be able to reproduce the method in its entirety. The workflow itself is published as an online resource together with supporting software and data. The paper concludes with a brief discussion of the complexities of requiring reproducibility in terms of cost versus benefit, and a desiderata with our observations and guidelines for improving reproducibility. This has implications not only in reproducing the work of others from published papers, but reproducing work from one's own laboratory.

Do some
research



Integrated collection
of methods, results,
data, metadata

Write a narrative
description that is
inextricably linked to
the data and methods

Store all data in
accessible,
usable format,
link to publication

Facilitate re-use
& replication –
by people or
machines

Reality

- Patient privacy concerns
- Unextractable data
- Everything “extra” in one file
- Third party licenses
- Proprietary data
- No metadata



Sequencing of Seven Haloarchaeal Genomes Reveals Patterns of Genomic Flux

Erin A. Lynch, Morgan G. I. Langille, Aaron Darling, Elizabeth G. Wilbanks, Caitlin Haltiner, Katie S. Y. Shao, Michael O. Starr, Clotilde Telling, Timothy T. Harkins, Robert A. Edwards, Jonathan A. Eisen , Marc T. Facciotti 

Published: July 24, 2012 • DOI: 10.1371/journal.pone.0041389

Article	About the Authors	Metrics	Comments	Related Content

Download PDF
Print **S**

Reader Comments (3)

[Post a new comment](#) on this article

0
RESPONSES

03 Dec 2012
00:56 GMT
MOST RECENT

Additional data sets
Posted by [Eisen](#) on 03 Dec 2012 at 00:56 GMT

Subject Areas

- Archaeal biology
- Bacterial genomics
- CRISPRs
- Genome evolution
- Genomic database
- Genomic library c
- Halobacterium
- Halobacterium

Current PLOS Data Access Policy

- PLOS is committed to ensuring the availability of data and materials that underpin any published articles
- Authors should make all relevant data immediately available without restrictions (exc patient confidentiality)
- Authors must comply with current best practices in their discipline

Failure to comply

- May affect decision to publish
- If after publication researchers cannot obtain data, PLOS may publish a correction, contact authors' institutions and funders, or in extreme cases withdraw publication

Aims of New PLOS Data Policy

- Ensuring access to the underlying data should be an **intrinsic** part of the scientific publishing process
- To ensure that **all steps**, from authoring to publication, capture **data and its associated metadata well** and then present them in optimal **human and machine-readable formats** to all readers and users of PLOS-published research



Background and rationale

- **Update** PLOS-wide data sharing policy (at <http://www.plosone.org/static/policies#sharing>)
- Establish **clarity** with respect to authors' obligations
- New policy highlights author's responsibility to **determine and describe a data sharing plan**
- New policy contains enhanced **enforcement** mechanism
- Therefore ensures **transparency**, i.e. compliance with policy is externally visible to readers (and to Academic Editors/referees in peer review)
- Aim to ensure policy is **workable across scientific fields**, and takes account of special considerations for privacy (in relation to human-subject research, and other issues)

Key features of the new policy (1)

- “PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.”
- “When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS’s policy. The data availability statement will be published with the article if accepted.”

Definition of “All Data”

- 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
 - Core descriptive data, methods, and study results should be included within the main paper, regardless of data deposition (PMC issue)
 - No “data not shown”
 - Authors who have datasets too large for sharing via repositories or uploaded files should contact the relevant journal for advice

Key features of the new policy (1)

- “PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.”
- “When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS’s policy. The data availability statement will be published with the article if accepted.”

Key features of the new policy (2)

Acceptable data-sharing methods:

- Data deposition (*strongly recommended; must include DOIs or accession numbers*)
- Data in supporting information files (*in file format from which data can be efficiently extracted*)

Choosing a data repository

- Authors must comply with field-specific standards for preparation and recording of data and to select repositories appropriate to their field, eg
 - microarray data in ArrayExpress or GEO
 - gene sequences in GenBank, EMBL or DDBJ
 - deposition of ecological data in Dryad
- Should meet accepted criteria as trustworthy, eg, Centre for Research Libraries or Data Seal of Approval
- If licensing policy is stated, at least CC-BY

Key features of the new policy (2)

Acceptable data-sharing methods:

- Data deposition (*strongly recommended; must include DOIs or accession numbers*)
- Data in supporting information files (*in file format from which data can be efficiently extracted*)

Key features of the new policy (3)

The following two methods may be acceptable alternatives, subject to case-by-case evaluation:

- Data made available to all interested researchers upon request (*only where deposition is not ethical or legal, eg, studies involving human participants; NOT from the authors themselves but from an ethics or oversight committee*)
- Data available from third party (*only where authors did not generate the primary dataset themselves but it is available*)

Data available on request

- Identify the group to which requests should be submitted (eg, a named data access committee or named ethics committee).
 - *Note that it is not acceptable for the authors to be the sole named individuals responsible for ensuring data access*
- The reasons for restrictions on public data deposition must also be specified.

Data with privacy concerns

Guidance on sharing data with privacy issues for human participants:

- Data handled so as to not compromise participants' privacy
- Follow established guidance and applicable local laws; national resources such as NIH
- Authors should indicate, as part of the ethics statement, the ways in which the study participants' privacy was preserved. If license agreements apply, note the process necessary for other researchers to obtain a license.

Key features of the new policy (3)

The following two methods may be acceptable alternatives, subject to case-by-case evaluation:

- Data made available to all interested researchers upon request (*only where deposition is not ethical or legal, e.g. studies involving human participants; NOT from the authors themselves but from an ethics or oversight committee*)
- Data available from third party (*only where authors did not generate the primary dataset themselves but it is available*)

Key features of the new policy (4)

Specification of restrictions unacceptable to PLOS

- Author refusal to share because of personal reasons such as patents or future publications
- Analysis rests solely on proprietary data: If proprietary data are used, the manuscript must include an analysis of public data that validates the conclusions so others can reproduce the analysis and build on the findings

Key features of the new policy (5)

Failure to comply

- May affect decision to consider or to publish the manuscript
- If after publication researchers cannot obtain data, PLOS may publish a correction, contact authors' institutions and funders, or in extreme cases withdraw publication

Questions and comments

- What to do with massive datasets?
- What if the researcher plans to publish additional studies using the data?
- What if competitors take advantage?
- In cases of “data available on request,” what if no data access committee exists and the IRB is not willing to take on the responsibility?

Implementation: submission

Data Availability

Do the authors confirm that all data underlying the findings described in this manuscript are fully available, without restriction or cost, to other researchers who wish to build upon or replicate this work? (See PLOS Data Policy for Details.)

- ☐ Yes
- ☐ No

Please indicate which data underlying the findings of this manuscript are unsuitable for deposit and describe the reasons why.

* typeset

[Instructions](#)

Answer

Required:

Character Count: 0

Limit
10000
characters

Data Location

Where are your datasets? (select all that apply)

[Instructions](#)

Answer

Required:

- ☐ Public repository: DOIs/accession numbers are available now
- ☐ Public repository: DOIs/accession numbers will be available after acceptance
- ☐ Supporting Information files
- ☐ Available upon request (only allowable in limited cases)
- ☐ Data are held by someone other than the authors
- ☐ All data are included within the manuscript

Data Location Description

Please describe where your data may be found. Include URLs, accession numbers and appropriate contact information if the data is available from the authors or a third party.

* typeset

[Instructions](#)

Answer

Required:

Character Count: 0

Limit
10000
characters

Implementation: article

- Abstract
- ▶ Author Summary
- Introduction
- Results
- Discussion
- Materials and Methods
- Supporting Information
- Acknowledgments
- Author Contributions
- References

Reader Comments (2)

Figures

Funding: JBH is supported by the Pennsylvania Commonwealth Health Research Formula Funds and the National Institute of Neurological Disease and Stroke (1R01NS054794) and the National Institute of Mental Health (P50 MH074924-01, awarded to Joseph S. Takahashi, Northwestern University). SP is supported by the Pew Scholars Program in Biomedical Science and the Whitehall Foundation. 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 Statement: The authors confirm that all data underlying the findings described in this manuscript is fully available, without restriction or cost, to other researchers who wish to build upon or replicate this work. All .cel files are available from GEO (liver accession = GSE11923, NIH3T3 accession = GSE11922, U2OS accession = GSE13949) and microarray data are available in a web-based interface at <http://bioinf.itmat.upenn.edu/circa/mouse/>.

Introduction

Circadian rhythms are daily, 24-hour (h) oscillations in physiology and behavior such as food consumption, blood pressure, metabolism, body temperature, and locomotor activity [1],[2]. These rhythms are thought to give an adaptive advantage by allowing an organism to anticipate changes in the environment and regulate physiology accordingly. Moreover, disruptions of circadian rhythms contribute to numerous pathologies including metabolic and cardiovascular disorders, cancer, and aging [3]–[5]. A molecular and cellular clock composed of transcriptional feedback loops generates these oscillations [6]. The central loci of the mammalian clock are

Questions and comments

- FAQ to come
- Authors who have questions or difficulties with the policy, or readers who have difficulty accessing data, are encouraged to contact the relevant journal office or data@plos.org



Open Issues

- How much data, metadata, and explanation are necessary for replication?
- How to cite data
- How to give academic credit for data reuse
- Funding for costs of data sharing
- Who determines who can access private data?
- What file formats should be acceptable?
- Future preservation of data in obsolete file formats
- How much peer review of data is sufficient?

Thank you

- To the PLOS Data Policy Team, particularly
 - Theo Bloom, *PLOS Biology* Editorial Director
 - Emma Veitch, Senior Editor, *PLOS ONE*
 - Emma Ganley, Senior Editor, *PLOS Biology*

**Look for the PLOS Data Policy on
March 1, 2014**



mwinker@plos.org

Libraries, Data, and Publication

Paul J. Bracke

Associate Dean for Research and Assessment
Purdue University Libraries

25 January 2014

SPARC-ACRL Forum on Emerging Issues in Scholarly Communication

PURDUE
UNIVERSITY
LIBRARIES

Overview

- Data Services and Libraries
- PURR
- Data Citation, Identifiers and Article Linking
- What can you do?

Data Services and Libraries

Why data for the Purdue Libraries?

- Directive to find ways of making the Libraries relevant to the campus research enterprise
- Common theme: Researchers need help with data management
- Opportunity to work with researchers directly AND build rich collections of primary source material

Research Data Strategy

- Build relationships and collaborations with campus faculty
 - Conduct research into data behaviors
- Build relationships with OVPR and CIO
- Collaborate on technology projects
 - Small-scale prototypes
 - HUBzero collaborations

Why data for Purdue?

National Science Foundation (NSF) – 2011

“Plans for data management and sharing of the products of research. Proposals must include a supplementary document of no more than two pages labeled ‘Data Management Plan’. This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.”

White House Office of Science and Technology Policy (OSTP) – 2013

*“The Office of Science and Technology Policy (OSTP) hereby directs each Federal agency with over \$100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government. Further, each agency plan for **both scientific publications and digital scientific data...**”*

PURR

Purdue University Research Repository (PURR)

The screenshot shows the PURR website homepage. At the top is a yellow banner with the text "Purdue University Research Repository (PURR)". Below this is a dark navigation bar with the PURDUE UNIVERSITY logo, the text "Purdue University Research Repository", and a "PURR" logo. On the right of the navigation bar are links for "Login", "Register", and "Report a bug". Below the navigation bar is a search bar and a menu with links: "Home", "Publications", "Projects", "Get Started", "Policies", and "Contact Us". The main content area has a section titled "What is PURR?" with a description: "The Purdue University Research Repository (PURR) provides an online, collaborative working space and data-sharing platform to support the data management needs of Purdue researchers and their collaborators." Below this is a "Learn More" button. To the right is a video player showing a PURR introductory video. Below the video player are three sections: "Start Your Research Project" with a "Create a Data Management Plan" link, "Featured Dataset" with a link to "Linking Pressure and Saturation through Interfacial Areas in Porous Media", and "Do you have a question?" with a "Chat with Us" button.

PURDUE UNIVERSITY

Purdue University Research Repository

PURR

Login Register Report a bug

Home Publications Projects Get Started Policies Contact Us

Search

What is PURR?

The [Purdue University Research Repository \(PURR\)](http://purr.purdue.edu) provides an online, collaborative working space and data-sharing platform to support the data management needs of Purdue researchers and their collaborators.

[Learn More](#)

Start Your Research Project

Create a Data Management Plan
Learn about the detailed requirements for your data management plan (DMP). Funding agency requirements are very specific and our DMP resources can help you to clear up any confusion. [Get Started](#)

Linking Research Data to Your Project

Featured Dataset

Linking Pressure and Saturation through Interfacial Areas in Porous Media
By V. Niasar, S. Hassanizadeh, L. Pyrak-Nolte, C. Berentsen

Do you have a question?

Chat with Us
Your Question/Message

PURR: <http://purr.purdue.edu>

Introductory Video: <http://youtu.be/Yw0IJj7FqA8>

What is PURR?

- PURR provides resources for data management planning
- PURR is a web-based platform for sharing data and collaborating on research
- PURR provides a platform for publishing datasets with DataCite DOIs
- PURR provides a platform for long-term archiving of data sets

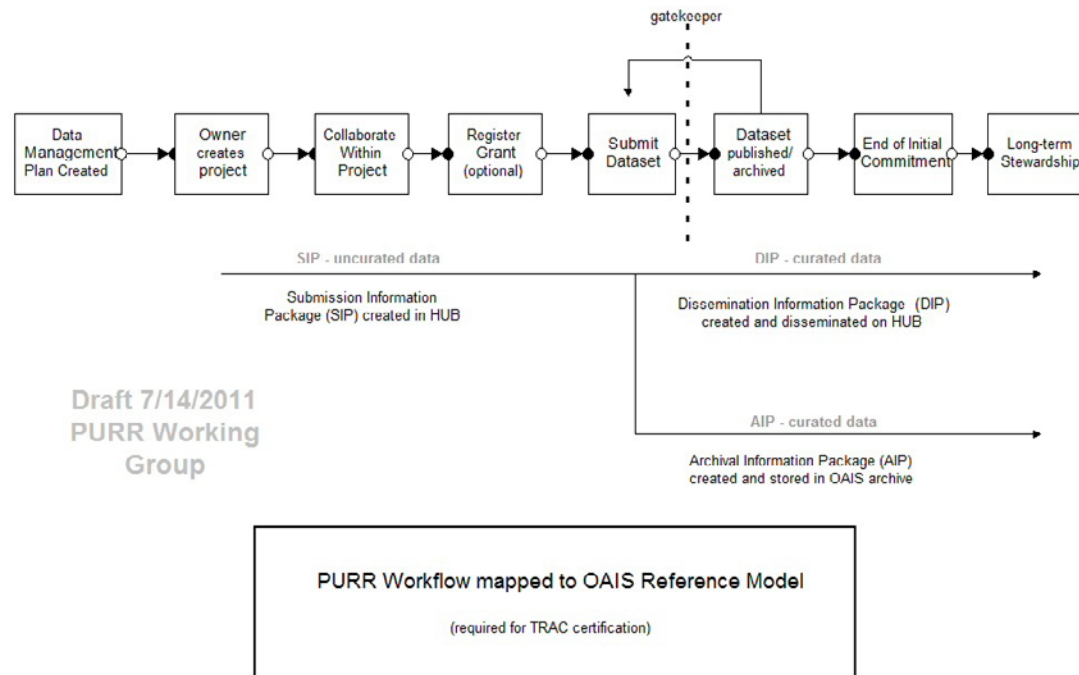
Who can use PURR?

Designed for:

- Purdue University faculty, staff, and graduate student researchers; their collaborators
- Current and future consumers of their data

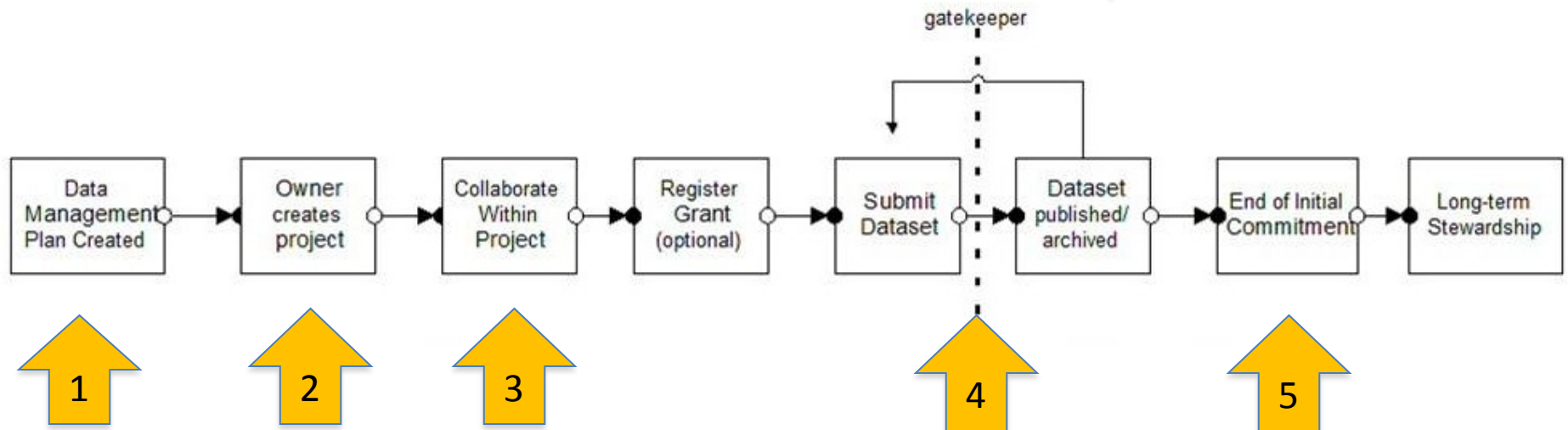
PURR Overview

- Technology platform is HUBzero: <http://hubzero.org>
 - Project and Publication
- Design inspired by the OAIS Reference Model

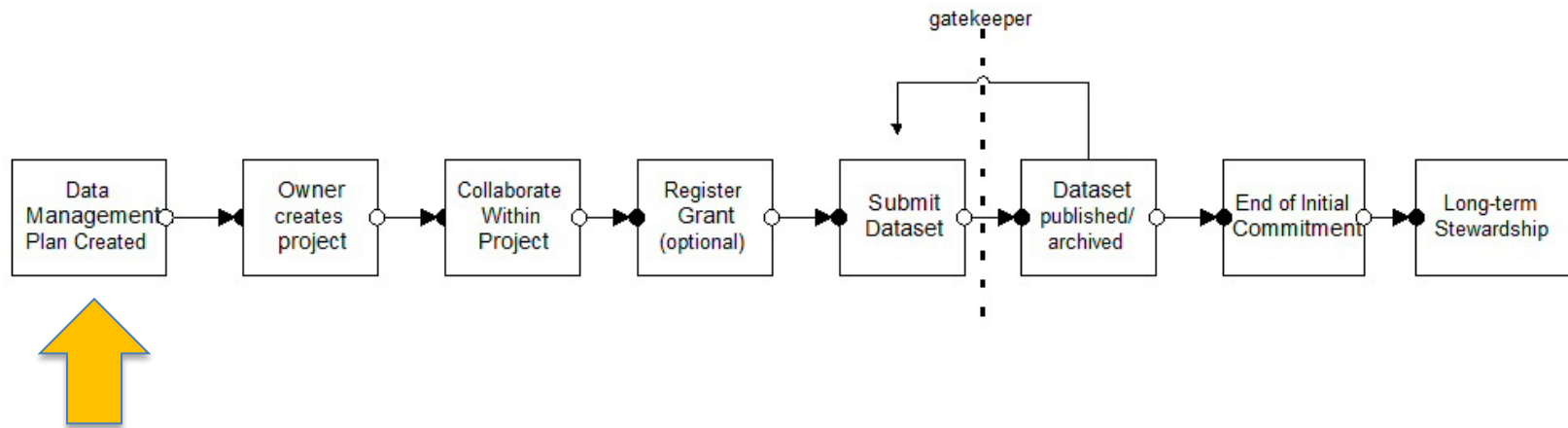


Purdue University Research Repository (PURR)

5 Opportunities

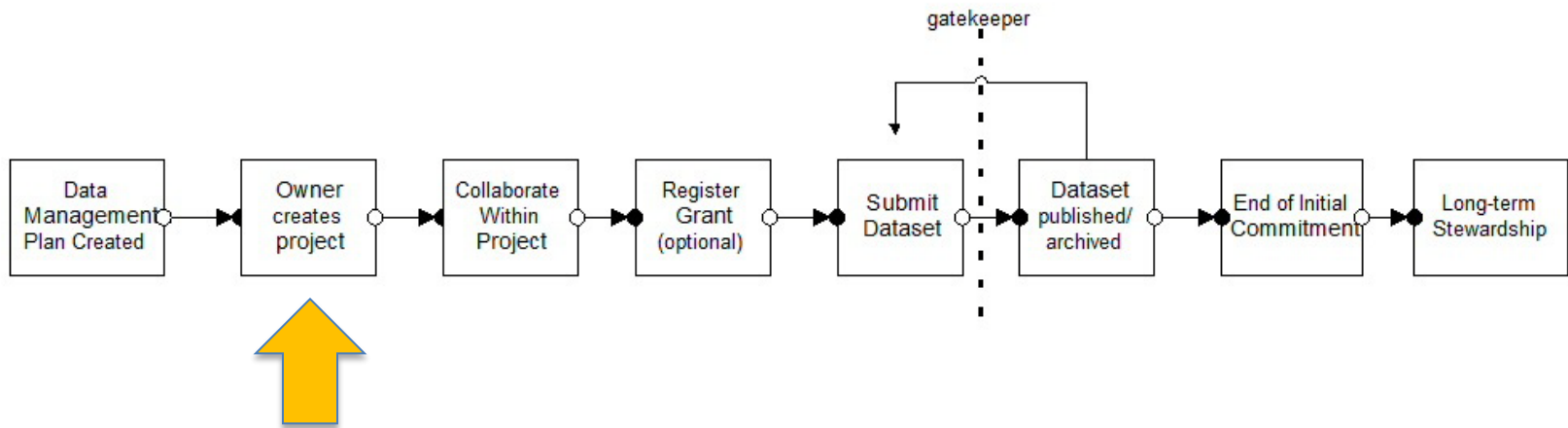


PURR Workflow Diagram



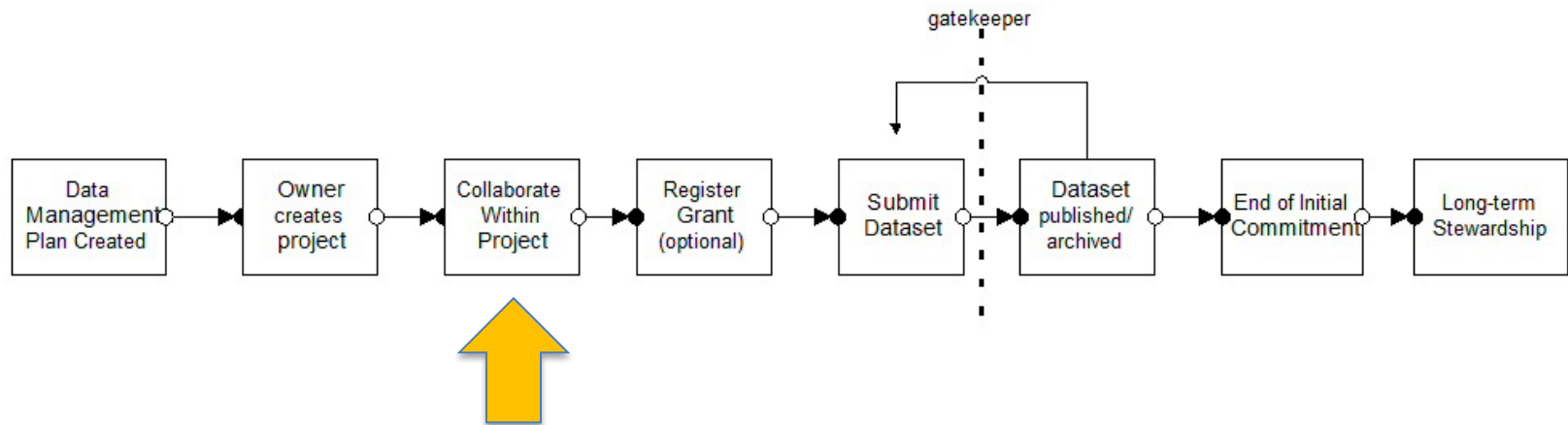
Librarians consult on data management plans in their subject areas.

Creating opportunities for librarians to interact with researchers about data...



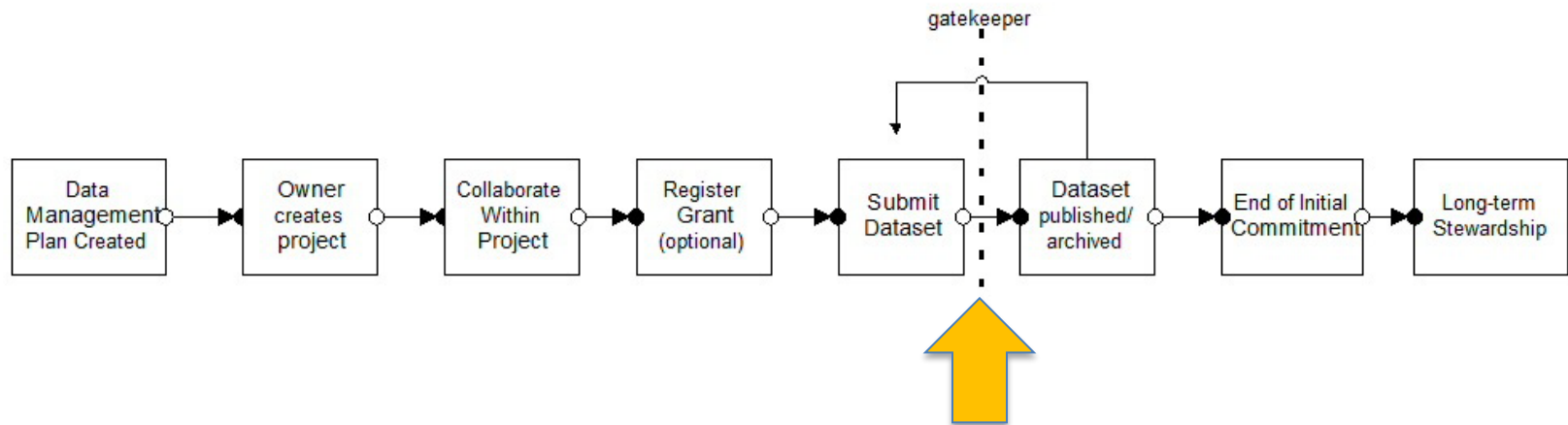
Librarian is notified by e-mail when a new project is created or a grant is awarded, based on department affiliation of Purdue project owner.

Creating opportunities for librarians to interact with researchers about data...



Librarian may consult or collaborate on project if needed.

Creating opportunities for librarians to interact with researchers about data...



Librarians review and post submitted datasets.

Creating opportunities for librarians to interact with researchers about data...

Data Publication

Graph of Flickr Photo-Sharing Social Network Crawled in May 2006

By David F Gleich

Purdue University

Crawl of the Flickr photo-sharing social network from May 2006 returning a graph with 820,878 nodes and 9,837,214 edges. Dataset is distributed as a SMAT file with README file with code to read file in Python and MATLAB.

Listed in Datasets

Download (SMA)

Additional materials available (2)

Version 1.1 - published on Feb 22, 2012
doi:10.4231/D39P2W550 - cite this

0 review(s) (Review this)
0 questions (Ask a question)
0 citation(s)
131 user(s) | 220 download(s)

Share: [Facebook](#) [Twitter](#) [LinkedIn](#) ...

This is a live publication with the page and content publicly available. [Manage this publication.](#)

from project [Networks and Matrix Computations](#)

About Supporting Docs Versions Reviews Questions Citations

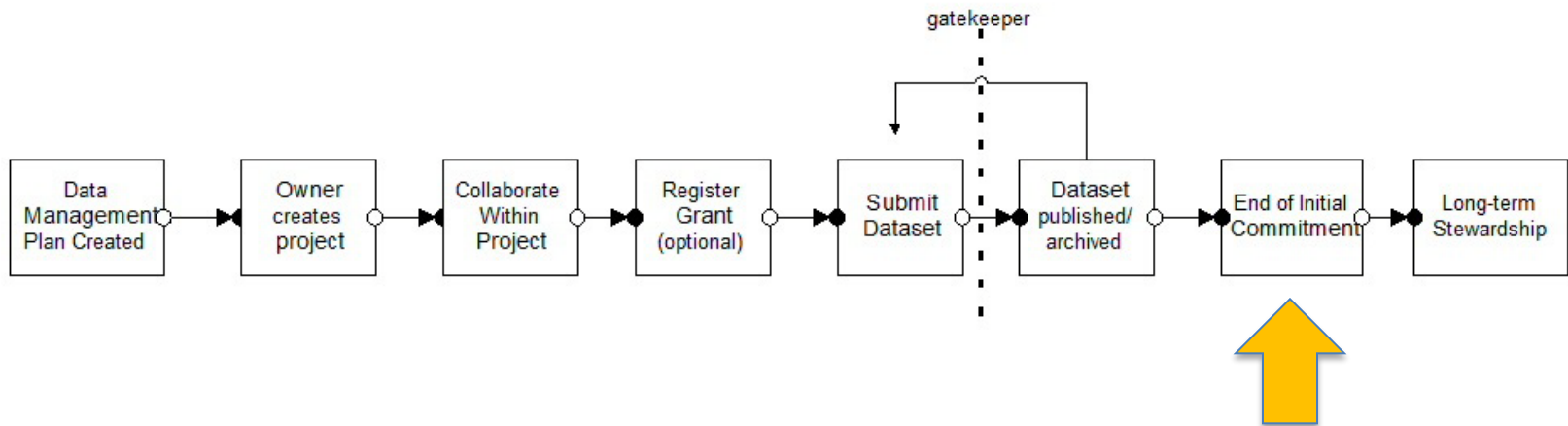


Abstract Flickr is a popular online-community for sharing photos, with millions of users. This graph is representative of its social network, in which the node set V represents users, and the edge set E is such that (u, v) is in E if and only if a user u has added user v as his/her contact. We start with a crawl extracted from Flickr in May 2006. This crawl began with a single user and continued until the total personalized PageRank on the set of uncrawled nodes was less than 0.0001. The result of the crawl was a graph with 820,878 nodes and 9,837,214 edges.

Cite this work Researchers should cite this work as follows:

David F Gleich, (2012), "Graph of Flickr Photo-Sharing Social Network Crawled in May 2006": (DOI: 10.4231/D39P2W550)

[BibTex](#) [EndNote](#)







At the end of initial commitment (10 years), archived and published datasets are remanded to the Libraries' collection. A librarian working with the digital archivist selects (or not) the dataset for the collection.

Creating opportunities for librarians to interact with researchers about data...

PURR Pricing

Storage space for PURR projects currently covered by central resources

<i>Option:</i>	<i>For what:</i>	<i>Space Available:</i>	<i>For how long:</i>	<i>For who:</i>
 Default/trial projects	Just trying things out, or don't need much space	10 GB	3 years	Any Purdue faculty, staff, and graduate student
 Default data publications	Great for small publications	1 GB	10 years	Any Purdue faculty, staff, and graduate student
 Supported projects	Funded projects with PIs from Purdue	100 GB	Life of grant	Purdue faculty with a verifiable grant or account number
 Supported data publications	Publishing work done on a funded project	10 GB	10 years	Purdue faculty with a verifiable grant or account number

Estimated cost of additional space *

<i>Option:</i>	<i>Space Available:</i>	<i>For how long:</i>	<i>For who:</i>	<i>Cost:</i>
Extra project space	As needed	Per year	Any Purdue faculty, staff, and graduate student	\$2.10 per GB per year **
Extra publication space	As needed	10 years	Any Purdue faculty, staff, and graduate student	\$14.30 per GB

<https://purr.purdue.edu/about/pricing>

PURR Collaboration

PURR is a collaboration between:

- Purdue Libraries
- Office of the Vice President for Research
- Information Technology at Purdue

PURR Team

- **Executive Committee:** Dean of Libraries, Vice President for Research, Chief Information Officer
- **Steering Committee:** 2 from libraries, 2 from IT, 2 from research office and sponsored programs, 3 domain faculty researchers
- **Personnel:** Project Director (.50), Technologists (3.85), HUBzero Liaison (.35), Metadata Specialist (.20), Digital Archivist (.25), Digital Data Repository Specialist (1.0)

PURR by the Numbers

At the end of 2013:

- Included in 911 data management plans (DMPs) with grant proposals
- 77 grants awarded
- 266 active research projects

DMP analysis (n=111 most recent NSF proposals from our university)

- 49% PURR
- 29% Local computer or server
- 14% Disciplinary repository (e.g., ICPSR, Protein Data Bank, nanoHUB, NEES)
- 8% No data or not applicable

Data Citation, Identifiers and Article Linking

Data Citation

“The scientist is willing (even eager) to make his data publicly available for a variety of potential uses on the condition that he receives credit, through a citation, when the data are used. He is very interested in employing DOIs for his data to enable their persistence so that the data may be cited.”[1]

“When the data are submitted to the institutional repository the scientist wants a “how-to-cite” note attached to the record so that users will properly cite the dataset. Citations or attribution for use of the data is a high priority. The scientist noted that the ability to connect her datasets with others and the ability to link the data with publications and other metadata is a high priority.”[2]

“The data collected by the scientist is very tightly tied to the scientist’s publications. Experimental context is complex, and may not be easily captured other than by linking publications to the data. “[3]

[1] Carlson, Jake R. (2009) "Traffic Flow - Purdue University," Data Curation Profiles Directory: Vol. 1, Article 4.

DOI: <http://dx.doi.org/10.5703/1288284315016>

[2] Rutter, Sara. (2011) “Botany/Plant Taxonomy – University of Hawaii,” Data Curation Profiles Directory: Vol. 3, Article 7. DOI:

<http://dx.doi.org/10.5703/1288284315000>

[3] Wright, Sarah J. (2012) “Environmental Science/Herbivory – Cornell University.” Data Curation Profiles Directory: Vol. 4, Article 3.

DOI: <http://dx.doi.org/10.5703/1288284315002>

Data Citation

- Attribution and data citation are :
 - Essential for linking data to publications and other scholarly artifacts
 - Essential for providing incentive and credit
 - Essential for data becoming a first-class citizen of the scholarly world
- Challenging:
 - Creating citable versions of data is not always a simple task
 - Culture of citing data has been uneven and slow in its development

DataCite and Identifiers

An International Organization dedicated to:

- Establishing easier access to scientific research data
- Increasing acceptance of research data as legitimate, citable contributions to the scientific record
- Supporting data archiving that will permit results to be verified and re-purposed for future study

<http://www.datacite.org>



Helping you to find,
access, and reuse data

PURDUE
UNIVERSITY
LIBRARIES


Data Citation and Library Publishing Services

- DOIs provide credibility – established brand for faculty
- Exploring emerging publishing models
 - Open Access
 - Connecting Textual and non-Textual Resources
 - Publishing Data (Data Papers, etc.)

JTRP Technical Reports Data Project

- Pilot between Purdue Press/Library Publishing and an Academic Research Center (JTRP) to:
 - Develop a unified workflow for producing published technical reports and data
 - Link technical reports to their underlying data
 - Link data to technical reports

Linking Data and Reports



e-Pubs

Home About FAQ My Account

Enter search terms:

in this series

Advanced Search

Notify me via email or RSS

Links for Authors


Submit Research
Style Guidelines
Policies and Help Documentation

Links

Purdue Libraries
Purdue University Press Journals
Joint Transportation Research Program
Purdue Road School


Browse

Collections
Disciplines
Authors


Access. Knowledge. Success.

Home > JTRPROGRAM > JTRP > 1521

JTRP TECHNICAL REPORTS



Effects of Realistic Heat Straightening Repair on the Properties and Serviceability of Damaged Steel Beam Bridges

Amit H. Varma, *Purdue University*

Youngmoo Sohn, *Purdue University*

SHARE

1

Recommended Citation


Varma, A. H., and Y. Sohn. *Effects of Realistic Heat Straightening Repair on the Properties and Serviceability of Damaged Steel Beam Bridges*. Publication FHWA/IN/JTRP-2013/03. Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2013. doi: 10.5703/1288284315184.

DOI
10.5703/1288284315184

Comments

Supplementary videos for SPR-3105:

Amit Varma, Youngmoo Sohn, (2013), "S-1 Span Damage - Supplementary Materials for the Report: Effects of Realistic Heat Straightening Repair on the Properties and Serviceability of Damaged Steel Beam Bridges": (DOI: [10.4231/D3RJ48V2G](https://doi.org/10.4231/D3RJ48V2G))



Purdue University Research Repository

Login Register Report a bug

Home Publications Projects Get Started Policies Contact Us

You are here: Publications > Datasets > S-1 Span Damage > Supplementary Materials for the ... > About

S-1 Span Damage - Supplementary Materials for the Report: Effects of Realistic Heat Straightening Repair on the Properties and Serviceability of Damaged Steel Beam Bridges

Version 1.0 - published on Oct 15, 2013
doi:10.4231/D3RJ48V2G - cite this

By Amit H. Varma*, Youngmoo Sohn*
Purdue University

Supplementary Materials for the Report: Effects of Realistic Heat Straightening Repair on the Properties and Serviceability of Damaged Steel Bridges

Listed in Datasets

About Supporting Docs Versions Reviews Questions Citations

Cite this work: Researchers should cite this work as follows:

Amit H. Varma, Youngmoo Sohn, (2013), "S-1 Span Damage - Supplementary Materials for the Report: Effects of Realistic Heat Straightening Repair on the Properties and Serviceability of Damaged Steel Beam Bridges": (DOI: 10.4231/D3RJ48V2G)

[BibTex](#) [EndNote](#)

About Supporting Docs Versions Reviews Questions Citations

Citations Non-affiliated (0) | Affiliated (1)

Affiliated authors

Varma, A. H. and Y. Sohn, "Effects of Realistic Heat Straightening Repair on the Properties and Serviceability of Damaged Steel Beam Bridges," Publication FHWA/IN/JTRP-2013/03, Joint Transportation Research Program, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana, 2013. (DOI: [10.5703/1288284315184](https://doi.org/10.5703/1288284315184))

[BibTex](#) [EndNote](#) [Electronic paper](#)

Purdue Press Supplemental Materials

Purdue University Press

[Home](#)
[About the Press](#)
[For Authors](#)
[For Booksellers](#)
[Rights & Permissions](#)
[Our Books](#)
[Our Journals](#)
[Projects](#)
[Contact](#)

Browse by Signature Area

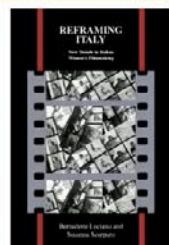
- > Agriculture
- > Business and Leadership
- > Education
- > Global Languages and Literatures
- > Health & Human Sciences
- > History
- > Jewish Studies
- > Library and Information Sciences
- > Philosophy
- > Public Policy
- > Purdue and Indiana
- > Science
- > Technology and Engineering
- > Veterinary Studies
- Uncategorized

Additional Links:

[Support the Press](#)

[Home](#)
[Books](#)
[Titles](#)
[Reframing Italy: New Trends in Italian Women's Filmmaking](#)
[Paperback](#)

Reframing Italy: New Trends in Italian Women's Filmmaking (Paperback)



Bernadette Luciano (Author) Susanna Scarparo (Author)

format: Paperback

publisher: Purdue University Press

pub. date: 11/15/2013

page count: 300pp

series: Purdue Studies in Romance Literatures

in series: 59

subject(s): Classical Studies, Drama, Performing Arts, Global Languages and Literatures

language: English

dimensions: 6.00" x 9.00"

ISBN 10: 1557536554

ISBN 13: 9781557536556

status: Available

[Facebook](#)
[Twitter](#)
[LinkedIn](#)
[YouTube](#)
[Google+](#)

Hide Additional Formats

ePDF

ePub

List Price: \$45.00

Use code 'Purdue Press' at checkout to receive 10% off when placing your order through this website.

Add to Cart

View Cart

[Find at Barnes & Noble.com](#)
[Find at Amazon.com](#)

Book Description

In recent years, Italian cinema has experienced a quiet revolution: the proliferation of films by women. But their thought-provoking work has not yet received the attention it deserves. *Reframing Italy* fills this gap. The book introduces readers to films and documentaries by recognized women directors such as Cristina Comencini, Wilma Labate, Alina Marazzi, Antonietta De Lillo, Marina Spada, and Francesca Comencini, as well as to filmmakers whose work has so far been undeservedly ignored.

An Interview with Alina Marazzi: Supplementary Content to Reframing Italy: New Trends in Italian Women's Filmmaking

By Susanna Scarparo¹, Bernadette Luciano²

¹ Monash University ² University of Auckland

This video interview with Alina Marazzi, a leading female Italian film director, accompanies the 2013 book *Reframing Italy: New Trends in Italian Women's Filmmaking*, by Bernadette Luciano and Susanna Scarparo.

Listed in Datasets

View Publication

Version 2.0 - published on Jun 14, 2013
doi:10.4231/D3F47GT8N - cite this

Licensed under Attribution-NonCommercial 3.0 Unported

[About](#)
[Supporting Docs](#)
[Versions](#)
[Reviews](#)
[Questions](#)
[Citations](#)



Abstract This video interview with Alina Marazzi, a leading female Italian film director, accompanies the 2013 book *Reframing Italy: New Trends in Italian Women's Filmmaking*, by Bernadette Luciano and Susanna Scarparo, and is published in the series *Purdue Studies in Romance Literatures* by Purdue University Press.

“In recent years, Italian cinema has experienced a quiet revolution: the proliferation of films by women. But their thought-provoking work has not yet received the attention it deserves. *Reframing Italy* fills this gap. The book introduces readers to films and documentaries by recognized women directors such as Cristina Comencini, Wilma Labate, Alina Marazzi, Antonietta De Lillo, Marina Spada, and Francesca Comencini, as well as to filmmakers whose work has so far been undeservedly ignored.

“Through a thematically based analysis supported by case studies, Luciano and Scarparo argue that Italian women filmmakers, while not overtly feminist, are producing work that increasingly foregrounds female subjectivity from a variety of social, political, and cultural positions. This book,

PURDUE
 UNIVERSITY
 LIBRARIES

What Can You Do?

1. Talk to Researchers

Data Curation Profiles

- An interview instrument that provides a guide for discussing data with researchers
- Analysis of profiles:
 - Gives insight into faculty needs and attitudes related to data sharing
 - Help assess information needs related to data collections
 - Gives insight into differences between data in various disciplines
 - Help identify possible data services
 - Create a starting point for curating a data set for archiving and preservation


Toolkit: <http://www.datacurationprofiles.org>

Directory: <http://docs.lib.purdue.edu/dcp>

Data Curation
Profiles Toolkit



2. Understand the Landscape




Databib

[Find Repositories](#) | [Submit](#) | [Connect](#) | [About](#)

[Login/Register](#)

Featured Repository



greengenes

603 data repositories
total in Databib.

Recently Added

- National Data Archive on Child Abuse and Neglect
- USGS National Water Information System (NWIS)
- CRYSTMET
- ORGIDS - Open Rotterdam Glaucoma Imaging Data Sets
- Clean Energy Project Database

[Follow @DataBib](#)

Subjects

- Agriculture (11)
- Area, Ethnic, and Gender Studies (9)

DataBib is a searchable catalog / registry / directory / bibliography of **research data repositories**.

Search [Find](#) [Advanced Search](#)

Browse [[Subjects](#) | [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) | [All](#)]

3

3TU.Datacentrum
A multidisciplinary data repository for a consortium of universities in the Netherlands housing over...

A

Access to Archival Databases (AAD)
The AAD is a database through the U.S. National Archives and Records Administration that allows user...

ACEpepDB: Peptide Database
ACEpepDB is a database ran by the Central Food Technological Research Institute. It contains records...

Addgene Plasmid Database
Addgene is a non-profit organization dedicated to making it easier for scientists to share plasmids....

Adult Blood Lead Epidemiology and Surveillance (ABLES) Interactive Database
ABLES provides data on lead exposure of adults in the United States. The data comes from laboratory...

Advanced Cooperative Arctic Data and Information Service (ACADIS)
The Advanced Cooperative Arctic Data and Information Service (ACADIS) program includes data manageme...

Advanced Spaceborne Thermal Emission and Reflection Radiometer [ASTER]
The Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) is an imaging instrument ...

Africa Centre for Health and Population Studies
The Africa Centre offers longitudinal datasets from a rural demographic in KwaZulu-Natal, South Afri...

Agency for Healthcare Research and Quality (AHRQ)
These datasets are made available by the United States Department of Health and Human Services, Agen...

Agri-Environmental Research Data Repository

3. Think About Instruction

Data Citation Services

Libraries » LibGuides » Citing Data

Admin Sign In

Citing Data

Instructions on citing your use of research datasets

Last Updated: May 25, 2011 URL: <http://guides.lib.purdue.edu/datacitation>

[Print Guide](#)

[RSS Updates](#)

[Email Alerts](#)

[SHARE](#) [f](#) [t](#) [e](#) ...

[Home](#) [Print Page](#)

Search:

[This Guide](#) [Search](#)

EXAMPLES OF DATA CITATIONS

Always check your syllabus or author guidelines to see if they contain directions for citing data. Some data distributors will suggest citations that you may use. Most common style guides (e.g., the Chicago Manual of Style) do not give specific instructions for citing data; however, here are three examples from those that do:

Publication Manual of the American Psychological Association (APA), 6th Edition

Pew Hispanic Center. (2004). Changing channels and crisscrossing cultures: A survey of Latinos on the news media [Data file and code book]. Retrieved from <http://pewhispanic.org/datasets/>

Style Manual for Political Science, Revised 2006, APSA

Purdue University. 2007. Controversial Facilities in Japan, 1955-1995 [computer file] (Study #4725). ICPSR04725-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research

HOW DO I CITE DATA?

When you're writing a research paper, it is necessary to cite your use of sources, typically as footnotes at the bottom of the page or in a bibliography at the end of the paper. It is crucial to provide references for your reader to better understand the context of your research and to give credit for people's work that you've used. As research becomes more data-intensive, it is important to cite your use of datasets in addition to traditional publications such as journal articles, books, and conference proceedings.

Digital datasets come in a wide variety of formats. Some examples include:

- spreadsheets
- interview transcripts
- sensor and instrument readings
- high resolution images
- gene sequences
- software source code
- video recordings

** The emerging best practice is to cite data just as you would cite a research article. **

Most traditional forms of documents are not capable of representing these kinds of data, and so datasets can be published separately in data repositories and other web sites. Whether you produced the data yourself or you're using someone else's data in your research, it is important to maintain a linkage between your paper and its supporting datasets by citing them. Not only does this give credit to the person who created the data, but it enables others to reproduce your research and verify your results. In some cases, sharing a dataset may have more scholarly impact than publishing a book or journal article.

There are many challenges in citing data. In most disciplines, there are no clear instructions on how to cite data. In fact, most of the major style guides (APA, MLA, the Chicago Manual of Style) do not directly address the issue of data citation. Data is not recognized as a format in

SUBJECT GUIDE



Michael Witt



Contact Info

Stewart Center Room G50
765-494-8703


[Send Email](#)

Links:

[Website / Blog](#)
[Profile & Guides](#)

[Subjects:](#)

Data Information Literacy

**DII L**
data information literacy

Data Information Literacy > Home

[▶ Home](#)

[▶ News](#)

[▶ About](#)

[▶ Partners](#)

[▶ Materials](#)

[▶ Publications](#)

[▶ Symposium](#)



[▶ Links](#)

Home

Welcome


Welcome to the working space for the [Institute of Museum and Library Services \(IMLS\)](#) funded research project investigating data information literacy (DIL) needs of e-scientists. With the continued evolution of technology driven research or "e-science" impacting the skills necessary for effective data management and curation, a curriculum designed to effectively prepare the next generation of scientists for the dynamic nature of research becomes necessary. This project, consisting of research teams from [Purdue University](#), the [University of Minnesota](#), the [University of Oregon](#), and [Cornell University](#), aims to develop and implement a DIL curriculum in conjunction with university faculty to address these needs.

This website will serve both as a source to obtain continued updates and information on the development of the DIL curriculum as well as a collaborative workspace within which researchers working on the project can meet to discuss their progress. Navigation of the website is available along the left-hand menu. The search box is also available via the menu at the top of the page if you are looking for something specific. Please continue to check back for [updates](#) on the site, where announcements and news will be posted as they become available.





social media feeds

All Inquiries

INSTITUTE of
Museum and Library
SERVICES

PURDUE
UNIVERSITY
LIBRARIES

OREGON



Thank You!



pbracke@purdue.edu
@pjbracke

Future of Research Communication (FORCE11) manifesto

Manifesto from Dagstuhl Perspectives Workshop 11331

Improving The Future of Research Communications and e-Scholarship

Edited by

Philip E. Bourne¹, Timothy W. Clark², Robert Dale³,
Anita de Waard⁴, Ivan Herman⁵, Eduard H. Hovy⁶, and
David Shotton⁷

1 University of California at San Diego, US, pbourne@ucsd.edu

2 Harvard Medical School, US, tim_clark@harvard.edu

3 Macquarie University, AU, Robert.Dale@mq.edu.au

4 Elsevier Laboratories – Jericho, US, A.dewaard@elsevier.com

5 Centrum voor Wiskunde en Informatica – Amsterdam, NL, ivan@w3.org

6 University of Southern California – Marina del Rey, US, hovy@isi.edu

7 University of Oxford, GB, david.shotton@zoo.ox.ac.uk

Abstract

The dissemination of knowledge derived from research and scholarship has a fundamental impact on the ways in which society develops and progresses, and at the same time it feeds back to improve subsequent research and scholarship. Here, as in so many other areas of human activity, the internet is changing the way things work; two decades of emergent and increasingly pervasive information technology have demonstrated the potential for far more effective scholarly communication. But the use of this technology remains limited.

Problems	Recommendations
Formats and Technologies	
Existing formats needlessly limit, inhibit and undermine effective knowledge transfer	Rethink the unit and form of the scholarly publication
Improved knowledge dissemination mechanisms produce information overload	Develop tools and technologies that better support the scholarly lifecycle
Claims are hard to verify and results are hard to reuse	Add data, software, and workflows into the publication as first-class research objects
Business Models and Attribution of Credit	
There is a tension between commercial publishing and the provision of unfettered access to scholarly information	Derive new financially sustainable models of open access
Traditional business models of publishing are being threatened	Derive new business models for science publishers and libraries
Current academic assessment models don't adequately measure the merit of scholars and their work over the full breadth of their research outputs	Derive new methods and metrics for evaluating quality and impact that extend beyond traditional print outputs to embrace the new technologies