

FAIR Code Sharing in Open Access Repositories

Maria Coteria and Andrew Mckenna-Foster, Figshare

Open Repositories 2023

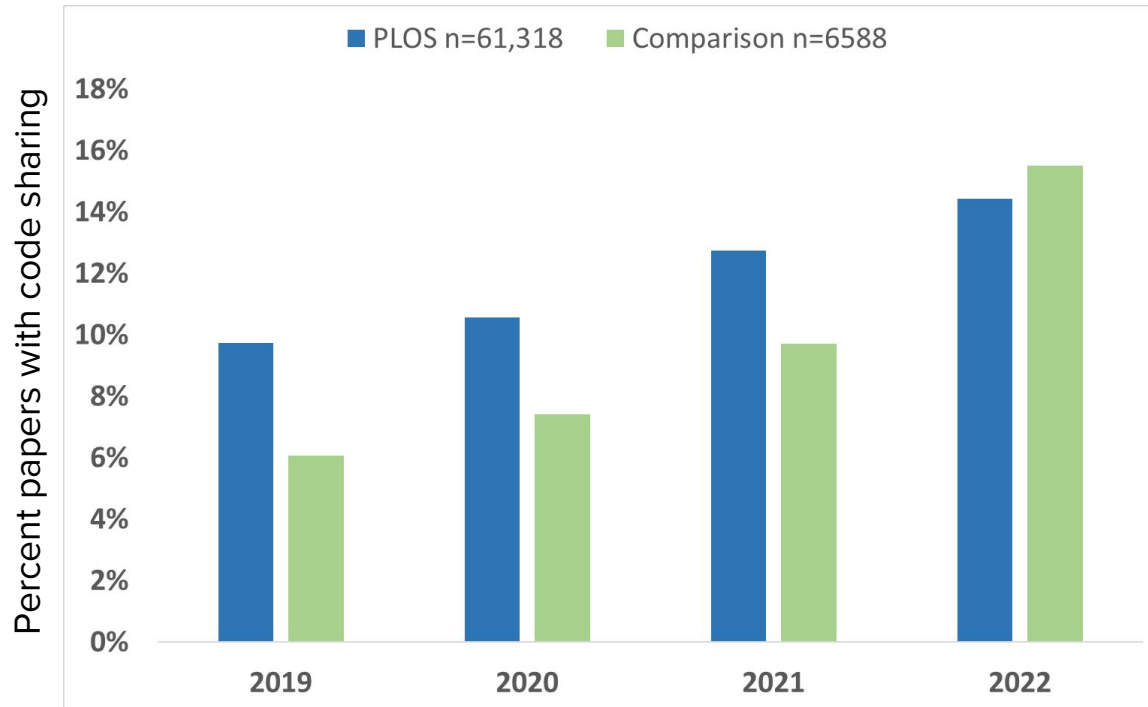
Stellenbosch, South Africa

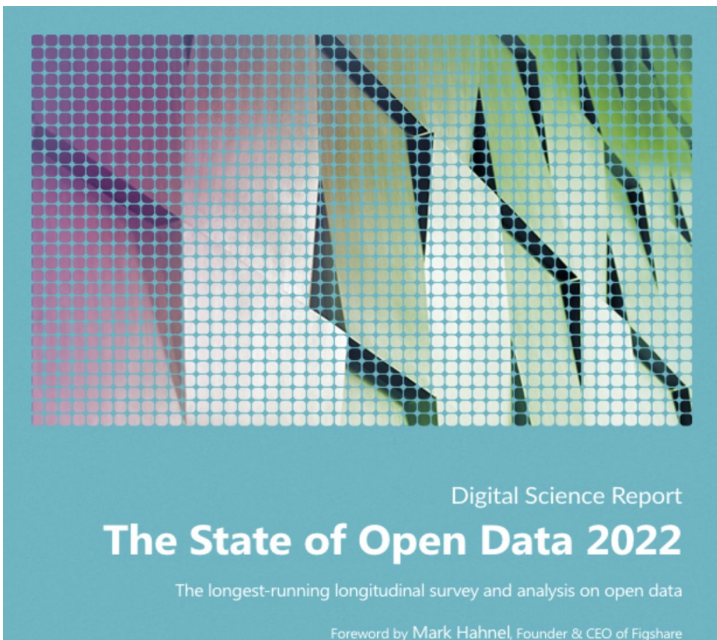
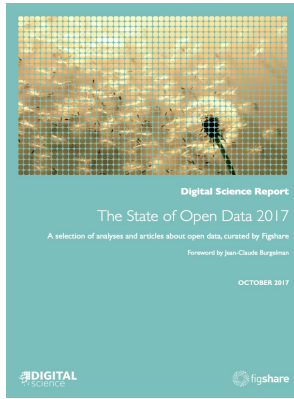
June 14, 2023


PLOS Open Science Indicators Dataset - December 2022

- Start defining and measuring Open Science practices
- Understand the current state of Open Science adoption
- Track progress over time
- Focus on detection of three Open Science practices:
 - preprint hosting
 - data sharing (in particular data shared in data repositories)
 - **code sharing**

PLOS Open Science Indicators Dataset - Shared Code





 **figshare** **SPRINGER NATURE**

The State of Open Data

Has been running for 7 years

Had over 25,000 respondents from 192 countries over that time

Provided us with a sustained look at the state of open data over time

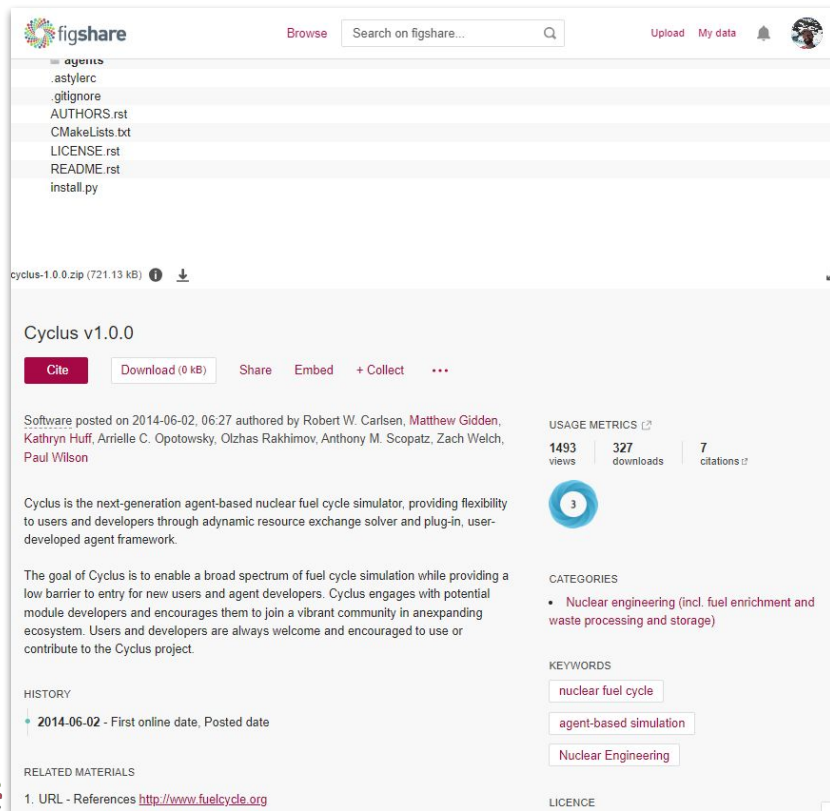
Figshare State of Open Data Report - October 2021

“Think about the last occasion you tried to get access to a dataset produced by another research group...”

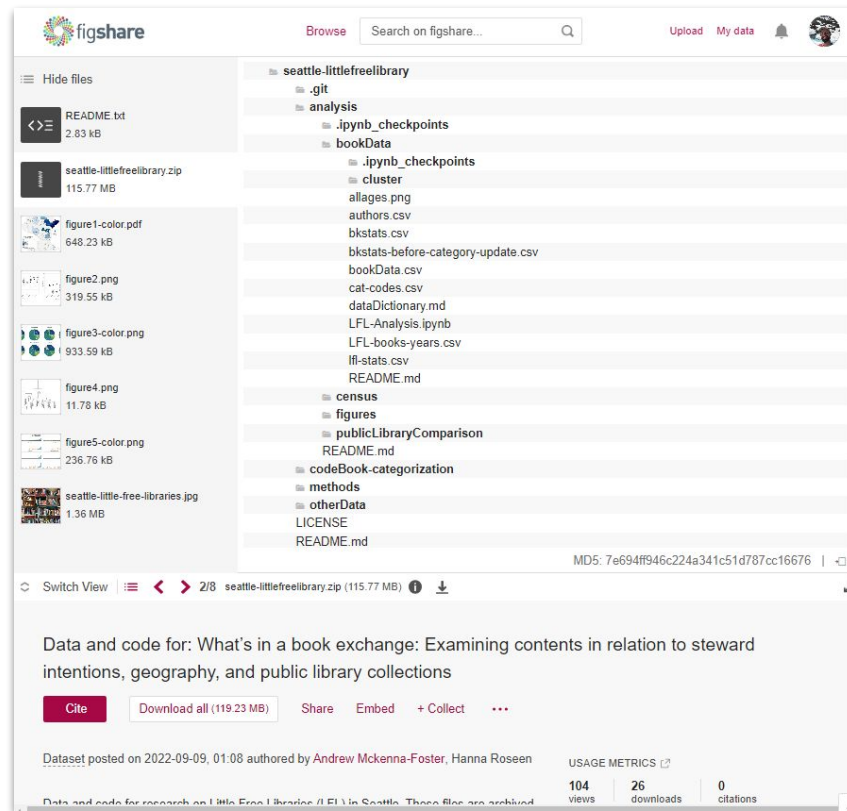
17% of those
who reused data
found **code** included

16% of those
who reused data
found **software
requirements**
included

Software/Code can take many forms



The screenshot shows the Figshare interface for the dataset 'Cyclus v1.0.0'. The top navigation bar includes 'Browse', a search bar, and 'Upload My data'. The left sidebar lists files such as 'agents', '.astylarc', '.gitignore', 'AUTHORS.rst', 'CMakeLists.txt', 'LICENSE.rst', 'README.rst', and 'install.py'. Below the sidebar, the main content area displays 'Cyclus v1.0.0' with a 'Cite' button, a 'Download (0 kB)' button, and options for 'Share', 'Embed', '+ Collect', and a menu icon. The description states: 'Software posted on 2014-06-02, 06:27 authored by Robert W. Carlsen, Matthew Gidden, Kathryn Huff, Arrielle C. Opatowsky, Olzhas Rakhimov, Anthony M. Scopatz, Zach Welch, Paul Wilson'. Usage metrics show 1493 views, 327 downloads, and 7 citations. A circular badge with the number '3' is visible. Categories include 'Nuclear engineering (incl. fuel enrichment and waste processing and storage)'. Keywords are 'nuclear fuel cycle', 'agent-based simulation', and 'Nuclear Engineering'. The history section shows '2014-06-02 - First online date, Posted date'. Related materials include '1. URL - References <http://www.fuelcycle.org>'. The license section is partially visible at the bottom.



The screenshot shows the Figshare interface for the dataset 'seattle-littlefreelibrary'. The top navigation bar includes 'Browse', a search bar, and 'Upload My data'. The left sidebar lists files such as 'README.txt', 'seattle-littlefreelibrary.zip', 'figure1-color.pdf', 'figure2.png', 'figure3-color.png', 'figure4.png', 'figure5-color.png', and 'seattle-little-free-libraries.jpg'. The main content area displays a file tree for 'seattle-littlefreelibrary' with subfolders like '.git', 'analysis', 'bookData', 'cluster', 'census', 'figures', 'publicLibraryComparison', 'codeBook-categorization', 'methods', 'otherData', and 'LICENSE'. The title is 'Data and code for: What's in a book exchange: Examining contents in relation to steward intentions, geography, and public library collections'. It includes a 'Cite' button, a 'Download all (119.23 MB)' button, and options for 'Share', 'Embed', '+ Collect', and a menu icon. The description states: 'Dataset posted on 2022-09-09, 01:08 authored by Andrew McKenna-Foster, Hanna Roseen'. Usage metrics show 104 views, 26 downloads, and 0 citations. The MD5 hash is 'MD5: 7e694ff946c224a341c51d787cc16676'.

Github citations, no DOIs

5 (2023) 200048

Contents lists available at ScienceDirect

Systems and Software Engineering

ELSEVIER journal homepage: www.journals.elsevier.com/systems-and-software-engineering

A benchmark dataset for defect detection and classification of electroluminescence images of PV modules

Lawrence Pratt^{a,b,*}, Jana Mattheus^a, Richard Klei

^a CSIR, Meiring Naude Road, Pretoria, South Africa
^b School of Computer Science and Applied Mathematics, University of the Witwatersrand

ARTICLE INFO

Keywords:
Electroluminescence
EL
PV
Semantic segmentation
Machine learning

ABSTRACT

Electroluminescence (EL) images of crystalline silicon solar cells are invisible to the naked eye. This paper presents a benchmark dataset for defect detection and classification. The dataset consists of 593 cell images with ground truth masks corresponding to the pixel-level labels for each feature and defect. Four deep learning models (U-Net, FCN, PSPNet, and DeepLabV3+) were trained using convolutional

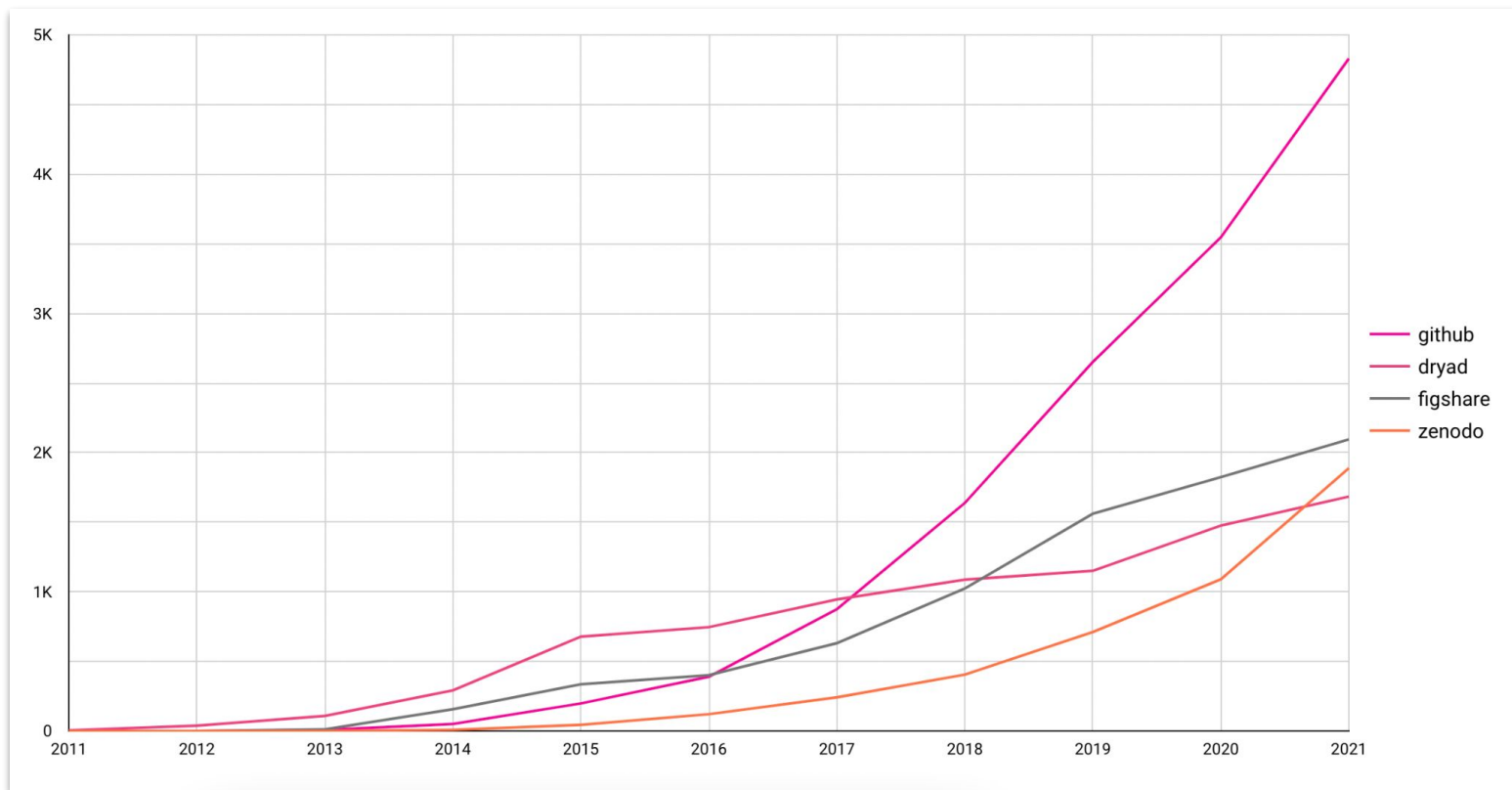
- [43] D. Gupta (2021) Image Segmentation Keras : Implementation of Segnet, FCN, UNet, PSPNet and other models in Keras [Source code], Accessed 23 Jan 2022, <https://github.com/divamgupta/image-segmentation-keras>.
- [44] N. Tomar (2022) Semantic-Segmentation-Architecture [Source code], Accessed 23 Jan 2022, <https://github.com/nikhilroxtomar/Semantic-Segmentation-Architecture>.
- [45] Y.Kamikawa (2020) keras-PSPNet [Source code], Accessed 23 Jan 2022, <https://github.com/ykamikawa/tf-keras-PSPNet>.
- [46] S. Kawakita (2021) example_camvid_multiclassB_quita.ipyn [Source code], Accessed 23 Jan 2022, <https://github.com/shirokawakita/multiclass-segmentation>.

Data availability

The data will be made public on the github repo referenced in the paper.

<https://doi.org/10.1016/j.sasc.2023.200048>

Github's growth in data availability statements is dramatic



Data from dimensions.ai

We know how software *should* be shared
and cited

00100101010
111110011000
001111111000
010001111110
01010101010
11001001001
01110000100
10001001010
01000101000
01010010010
10010010100

Software/Code is part of FAIR Workflows

FAIR Workflow = Using Persistent identifiers and metadata to make every aspect of a research project FAIR, from grant submission to publication.

FAIR Workflows Project



Figure 1. A Diagram of the proposed FAIR Workflows based on a research lifecycle.

Code can be highly reusable and gets cited more than other outputs

Of the top 10 most highly cited code / software records in Figshare:

Total citations: 221

Median: 13

Range: 9 to 86

[a figshare article - Research Data. Is it being cited?](#)

Software citation principles from Smith et al. 2016



Unique identification
Persistence
Accessibility
Specificity

The easiest way to accomplish these is use a repository that provides a PID and can version records...

...which many repositories do!

If you provide it, will researchers use it? (and correctly?)



Under leg chainsaw juggling - Guinness World Records <https://www.youtube.com/watch?v=aH4mQVflaPE>

Lots of shared software out there

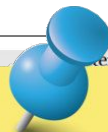
Zenodo: 103,418 records | Figshare: 10,647 records

But using a repository does not guarantee proper sharing or citing

A few examples...

00100101010
111110011000
001111111000
010001111110
11001001001
01110000100
10001001010
01000101000
01010010010
10010010100

Github citations, no DOIs



	citation	code
Unique identification	-	-
Persistence	-	-
Accessibility	-	-
Specificity	-	✓

5 (2023) 200048

ments lists availabl

[43] D. Gupta (2021) Image Segmentation Keras : Implementation of Segnet, FCN, other models in Keras [Source code], Accessed 23 Jan 2022, <https://github.com/divamgupta/image-segmentation-keras>.
semantic-segmentation-architecture [Source code], Accessed 23 Jan 2022, <https://github.com/nikhilroxtomar/Semantic-Segmentation-Architecture>

) keras-PSPNet [Source code], Accessed 23 Jan 2022, <https://github.com/mikawa/tf-keras-PSPNet>.

) example_camvid_multiclassB_quita.ipyn [Source code], Accessed 23 Jan 2022, <https://github.com/shirokawakita/multiclass-segmentation>

, Accessed

public on the github repo referenced in the

line silicon solar cells. The dataset
pixel-level labels for each feature and
were trained using conv-lass

Citing in the text/DAS, software snapshots available

The screenshot shows the journal page for 'Methods in Ecology and Evolution', Volume 6, Issue 11, November 2015. The article title is 'A new method to analyse species abundances in space using generalized dimensions' by Leonardo A. Saravia. The page includes a search bar, navigation menus, and a list of sections. The article text is partially obscured by a text box.

When $\rho = 1$, there is no replacement of species and the model is completely neutral. A more thorough description of the model is given in Appendix S1 (Supporting Information), and its C++ source code is available at <https://github.com/lasaravia/neutral> and figshare <http://dx.doi.org/10.6084/m9.figshare.969692>.

Data accessibility
All data and R scripts used in this manuscript are available at figshare <http://dx.doi.org/10.6084/m9.figshare.1276105>.

<https://doi.org/10.1111/2041-210X.12417>

Citing in the text/DAS, software snapshots available

BRITISH ECOLOGICAL SOCIETY

Search

JOURNALS TOPICS

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Research Article | Free Access

A new method to analyse species abundances in space using generalized dimensions

Leonardo A. Saravia

First published: 19 August 2015 | <https://doi.org/10.1111/2041-210X.12100>

SECTIONS

Summary

1. Species–area relationships (SAR) and species-abundance distributions (SAD) are among the most studied patterns in ecology, due to their application to both theoretical and conservation issues. One problem with SAR is that different theories can generate the same prediction, so they cannot be used to detect different mechanisms. A solution is to search for more sensitive patterns, from SAR to the whole SAD. A generalized dimension (D_q) to study the scaling of SAD, but to date, there has been no equivalent way to express SAD is the rank-abundance curve (RAC). We introduce a new way to study SAD scaling using a spatial version of RAD: the species-rank surface (SRS), which can be analysed using D_q . Thus, there is an old D_q based on SAR (D_q^{SAR}), and a new one based on SRS (D_q^{SRS}). I perform spatial simulations to examine the relationship of D_q with SAD, spatial patterns and number of species. Finally, I compare the power of both D_q SAD, SAR exponent

when $\rho = 1$, then
thorough descri
C++ source code
<http://dx.doi.org/>

Data access
All data and R scri
<http://dx.doi.org/>

figshare

Browse Search on figshare... Upload My data

1/1

Neutr... zip (49.32 kB) ips... exe (348.78 kB)

ARCHIVE EXE

Switch View | 2 files

Neutral: a stochastic spatial multispecies model

Cite Download all (469.96 kB) Share Embed + Collect

Version 4 Software posted on 2017-01-23, 08:55 authored by Leonardo Saravia

C++ code for stochastic spatial neutral/hierarchical models: this is the release v1.0.1, including calculation of multifractal spectra and species clusters sizes and windows compatibility.

For the latest version of this code, please see <https://github.com/lSaravia/neutral>

HISTORY

- 2017-01-23 - First online date, Posted date

USAGE METRICS

895 views	109 downloads	10 citations
-----------	---------------	--------------

CATEGORIES

- Bioinformatics and computational biology not

Citing in the text/DAS, software snapshots available

BRITISH ECOLOGICAL SOCIETY

Search

JOURNALS TOPICS

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

Research Article | Free Access

A new method to analyse species abundances in space using generalized dimensions

Leonardo A. Saravia

First published: 19 August 2015 | <https://doi.org/10.1111/2041-210X>

SECTIONS

Summary

1. Species–area relationships (SAR) and species-abundance distributions (SAD) are among the most studied patterns in ecology, due to their application to both theoretical and conservation issues. One problem with these patterns is that different theories can generate the same prediction, so they cannot be used to detect different mechanisms.
2. A solution is to search for more sensitive patterns, from SAR to the whole SAD. A generalized dimension (D_q) to study the scaling of SAD, but to date, there has been no method to detect different mechanisms.
3. An equivalent way to express SAD is the rank-abundance distribution (RAD). We introduce a new way to study SAD scaling using a spatial version of RAD: the species-rank surface (SRS), which can be analysed using D_q . Thus, there is an old D_q based on SAR (D_q^{SAR}), and a new one based on SRS (D_q^{SRS}). I perform spatial simulations to examine the relationship of D_q with SAD, spatial patterns and number of species. Finally, I compare the power of both D_q . SAD, SAR exponent

when $p = 1$, then
thorough descri
C++ source code
<http://dx.doi.org/>

Data access
All data and R scri
<http://dx.doi.org/>

figshare

Browse Search on figshare... Upload My data

1/1

Neutr... zip (49...

Switch View | 2

Neutral: a s

Cite

Version 4 v So

C++ code for sto
of multifractal sp

For the latest ver

HISTORY

2017-01-23 - First online date, Posted date

CATEGORIES

Bioinformatics and computational biology not

	citation	code
Unique identification	✓	✓
Persistence	✓	✓
Accessibility	✓	✓
Specificity	-	✓

Citation in paper, data and software in different repositories

The image shows a screenshot of an eLife research article page. The page header includes the eLife logo, navigation links (Home, Magazine, Community, About), a search bar, an alerts bell icon, and a 'SUBMIT YOUR RESEARCH' button. A blue banner at the top of the article content area reads 'Accepted manuscript, PDF only. Full online version available upon publication'. The article title is 'Tradeoffs explain scaling, sex differences and seasonal oscillations in the remaining weapons of snapping shrimp (*Alpheidae*)'. The authors are Jason P Dinh and S N Patek, from Duke University, United States. The article was published on May 9, 2023, with the DOI <https://doi.org/10.7554/eLife.84589>. A callout box titled 'Data availability' is overlaid on the right side of the page, stating: 'Data, metadata, and code are available on Dryad: <https://doi.org/10.5061/dryad.qz612jmkf>'. Below the callout box, there are 'Share' and 'Comment' buttons, and a view count of '138 views'. At the bottom right, it says 'Accepted Manuscript published May 9, 2023 (This version)'. The page footer shows 'Article' and 'Abstract' tabs.

eLife

Home Magazine Community About

Search Alerts

SUBMIT YOUR RESEARCH

Accepted manuscript, PDF only. Full online version available upon publication

Research Article

Ecology, Evolutionary Biology

Tradeoffs explain scaling, sex differences and seasonal oscillations in the remaining weapons of snapping shrimp (*Alpheidae*)

Jason P Dinh, S N Patek

Duke University, United States

May 9, 2023 · <https://doi.org/10.7554/eLife.84589>

Data availability

Data, metadata, and code are available on Dryad:
<https://doi.org/10.5061/dryad.qz612jmkf>

Share Comment

138 views

<https://doi.org/10.7554/eLife.84589>

Accepted Manuscript published
May 9, 2023 (This version)

Article Abstract

Citation in paper, data and software in different repositories

The screenshot shows a DRYAD research article page. The article title is "Tradeoffs and benefits explain scaling, sex differences, and seasonal oscillations in the remarkable weapons of snapping shrimp (Alpheus spp.)". The authors are Jason P. Dinh and S. N. Patek. The article is published in eLife. On the right side, there are sections for "Data files" and "Related works". The "Related works" section contains two entries: "Article" with DOI <https://doi.org/10.7554/eLife.84589> and "Software" with DOI <https://doi.org/10.5281/zenodo.7316000>. A green box highlights the software DOI entry.

DRYAD | Explore data | Search | Who we are | What we do | Join us | Help | Login

Research Article
Ecology, Evolutionary Biology

Tradeoffs explain scaling and seasonal oscillations in the remarkable weapons of snapping shrimp

Jason P. Dinh, S. N. Patek

Duke University, United States

May 9, 2023 · <https://doi.org/10.7554/eLife.84589>

Tradeoffs and benefits explain scaling, sex differences, and seasonal oscillations in the remarkable weapons of snapping shrimp (Alpheus spp.)

Dinh, Jason, Duke University, <https://orcid.org/0000-0001-6471-9047>
Patek, S. N., Duke University, <https://orcid.org/0000-0001-9738-882X>
jasonpdinh@gmail.com, snp2@duke.edu

Publication date: November 18, 2022
Publisher: Dryad
<https://doi.org/10.5061/dryad.qz612jmkf>

Data files

Download dataset

> November 18, 2022

Related works

Article
<https://doi.org/10.7554/eLife.84589>

Software
<https://doi.org/10.5281/zenodo.7316000>

Citation

Dinh, Jason; Patek, S. N. (2022). Tradeoffs and benefits explain scaling, sex differences, and seasonal oscillations in the remarkable weapons of snapping shrimp (Alpheus spp.). *eLife*, 12, 84589. <https://doi.org/10.7554/eLife.84589>

Citation in paper, data and software in different repositories



Home

Research Article
Ecology, Evolutionary Biology

Tradeoffs explain and seasonal weapons of s

Jason P Dinh, S N Patek
Duke University, United States

May 9, 2023 · <https://doi.org/10.7554/eLife.84589>

Article



Tradeoffs and benefits explain scaling, seasonal oscillations in the remarkable weapons of snapping shrimp (*Alpheus* spp.)


Dinh, Jason, Duke University, <https://orcid.org/0000-0001-9323-2100>
Patek, S. N., Duke University, <https://orcid.org/0000-0001-9323-2100>
jasonpdinh@gmail.com, snp2@duke.edu

Publication date: November 18, 2022
Publisher: Dryad
<https://doi.org/10.5061/dryad.qz612jmkf>

Citation

Dinh, Jason; Patek, S. N. (2022). Tradeoffs and benefits explain scaling, seasonal oscillations in the remarkable weapons of snapping shrimp (*Alpheus* spp.). *eLife*, 12, 84589. <https://doi.org/10.7554/eLife.84589>

Abstract




Search [] [] Upload Communities Log in Sign up

November 18, 2023


Tradeoffs and benefits explain scaling, sex differences, and seasonal oscillations in the remarkable weapons of snapping shrimp (*Alpheus* spp.)

14 views 7 downloads

Part of



Indexed in



Publication date: November 18, 2023
DOI: [DOI: 10.5281/zenodo.7316900](https://doi.org/10.5281/zenodo.7316900)
Keyword(s): snapping shrimp, Alpheus, animal weapons, handicap principle, reproduction
Related identifiers: Cited by 10.7554/eLife.84589

Funding provided by: National Science Foundation
Crossref Funder Registry ID: <http://dx.doi.org/10.13039/100000001>
Award Number: IOS 2019323
Funding provided by: Duke University
Crossref Funder Registry ID: <http://dx.doi.org/10.13039/100006510>
Award Number: Biology Department: Grant-in-Aid of Research

Name	Size	Download
WeaponTradeoffsandBenefits_Code_Final.Rmd	44.7 kB	Download

md5:5867417682b3ff4e7549435f4c7276cd7

Citation in paper, data and software in different repositories

eLife Home

DRYAD

Tradeoffs and benefits explain seasonal oscillations in the snapping shrimp (*Alpheus* spp)

Dinh, Jason, Duke University, <https://orcid.org>
 Patek, S. N., Duke University, <https://orcid.org>
 jasonpdinh@gmail.com, snp2@duke.edu

Publication date: November 18, 2022
 Publisher: Dryad
<https://doi.org/10.5061/dryad.qz612jmkf>

Citation

Dinh, Jason; Patek, S. N. (2022). Tradeoffs and benefit

zenodo Search Upload Community Log in Sign up


	citation	code
Unique identification	✓	✓
Persistence	✓	✓
Accessibility	✓	✓
Specificity	-	✓

Citing in the references...

Duffy et al. *Malar J* (2016) 15:258
DOI 10.1186/s12936-016-1296-4

Malaria Journal

RESEARCH Open Access



Differences in PfEMP1s recognized by antibodies from patients with uncomplicated or severe malaria

Michael F. Duffy^{1*}, Rintis Noviyanti², Takafumi Tsuboi³, Zhi-Ping Feng^{4,5}, Leily Triana⁶, Eizo Takashima³, Fransisca Sumardy¹, Daniel A. Lampah⁶, Louise Turner⁷, Thomas Peter Siba⁹, Stephen J. Rogerson¹, Thor G. Theander⁷, Jutta Marfurt¹⁰, Ric N. Price¹¹, Graham V. Brown¹² and Anthony T. Papenfuss^{4,5,13,14}

Abstract

Background: *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) variants and mediate pathogenic cytoadhesion and antigenic variation in malaria. PfEMP1s can be divided into three principal groups (A, B and C) and they contain conserved arrangements of functional domains called domain cassettes. Despite their tremendous diversity there is compelling evidence that a restricted subset of PfEMP1s is expressed in severe disease. In this study antibodies from patients with severe and uncomplicated malaria were compared for differences in reactivity with a range of PfEMP1s to determine whether antibodies to particular PfEMP1 domains were associated with severe or uncomplicated malaria.

Methods: Parts of expressed *var* genes in a severe malaria patient were identified by RNAseq and several of these partial PfEMP1 domains were expressed together with others from laboratory isolates. Antibodies from Papuan patients to these parts of multiple PfEMP1 proteins were measured.

Results: Patients with uncomplicated malaria were more likely to have antibodies that recognized PfEMP1 of Group C type and recognized a broader repertoire of group A and B PfEMP1s than patients with severe malaria.

Conclusion: These data suggest that exposure to a broad range of group A and B PfEMP1s is associated with protection from severe disease in Papua, Indonesia.

Keywords: Severe malaria, *var* genes, PfEMP1

Background molecules are encoded by the *var* multigene family [1–3].

- read mapping by seed-and-vote. *Nucleic Acids Res.* 2013;41:e108.
51. Crusoe M, Edverson G, Fish J, Howe Adina, McDonald E, et al. The khmer software package: enabling efficient sequence analysis. *Figshare.* 2014. doi:10.6084/m9.figshare.979190.
52. Schulz MH, Zehring DR, Vingron M, Birney E. Oases: robust de novo RNA-

...but the record contains a PDF

Duffy et al. *Malar J* (2016) 15:258
DOI 10.1186/s12936-016-1296-4

Malaria Journal

RESEARCH Open Access

Differences in PfEMP1s recognized by antibodies from patients with uncomplicated or severe malaria

Michael F. Duffy^{1*}, Rintis Noviyanti², Takafumi Tsuboi³, Zhi-Ping Feng^{4,5}, Leily Triani⁶, Eizo Takashima³, Fransisca Sumardy¹, Daniel A. Lampah⁶, Louise Turner⁷, Thomas Peter Siba⁹, Stephen J. Rogerson¹, Thor G. Theander⁷, Jutta Marfurt¹⁰, Ric N. Price¹¹, Graham V. Brown¹² and Anthony T. Pappenfuss^{4,5,13,14}

Abstract
Background: *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) variants and mediate pathogenic cytoadhesion and antigenic variation in malaria. PfEMP1s can be divided into three principal groups (A, B and C) and they contain conserved arrangements of functional domains called domain cassettes. Despite their tremendous diversity there is compelling evidence that a restricted subset of PfEMP1s is expressed in severe disease. In this study antibodies from patients with severe and uncomplicated malaria were compared for differences in reactivity with a range of PfEMP1s to determine whether antibodies to particular PfEMP1 domains were associated with severe or uncomplicated malaria.
Methods: Parts of expressed *var* genes in a severe malaria patient were identified by RNAseq and several of these partial PfEMP1 domains were expressed together with others from laboratory isolates. Antibodies from Papuan patients to these parts of multiple PfEMP1 proteins were measured.
Results: Patients with uncomplicated malaria were more likely to have antibodies that recognized PfEMP1 of Group C type and recognized a broader repertoire of group A and B PfEMP1s than patients with severe malaria.
Conclusion: These data suggest that exposure to a broad range of group A and B PfEMP1s is associated with protection from severe disease in Papua, Indonesia.
Keywords: Severe malaria, *var* genes, PfEMP1

Background molecules are encoded by the *var* multigene family [1–3].

figshare

Browse Search on figshare... Upload My data

1

The khmer software package: enabling efficient sequence analysis

Michael R. Crusoe, Greg Edverson, Jordan Fish, Adina Howe, Eric McDonald, Joshua Nahum, Kaben Nanlohy, Humberto Ortiz-Zuazaga, Jason Pell, Jared Simpson, Camille Scott, Ramakrishnan Rajaram Srinivasan, Qingpeng Zhang, and C. Titus Brown*

* E-mail: Corresponding ctb@msu.edu

Abstract

The khmer package is a freely available software library for working efficiently with fixed length DNA words, or k-mers. khmer provides implementations of a probabilistic k-mer counting data structure, a compressible De Bruijn graph representation, De Bruijn graph partitioning, and digital normalization. khmer is implemented in C++ and Python, and is freely available under the BSD license at <http://github.com/ged-lab/khmer/>.

khmer-software.pdf (40.02 kB)

The khmer software package: enabling efficient sequence analysis

Cite Download (0 kB) Share Embed + Collect ...

Version 3 Dataset posted on 2014-04-01, 05:02 authored by C. Titus Brown, Michael R. Crusoe, Greg Edverson, Jordan Fish, Adina Howe, Eric McDonald, Joshua Nahum, Kaben Nanlohy, Humberto Ortiz-Zuazaga, Jason Pell, Jared Simpson, Camille Scott, Ramakrishnan Rajaram Srinivasan, Qingpeng Zhang

USAGE METRICS

5736 views	1937 downloads	12 citations
------------	----------------	--------------

A citation handle for the khmer software, v1.0

HISTORY

- 2014-04-01 - First online date. Posted date

RELATED MATERIALS

1. URL - References <http://github.com/ged-lab/khmer/>

CATEGORIES

- Bioinformatics and computational biology not elsewhere classified

KEYWORDS

bioinformatics Bioinformatics

...but the citation is for a PDF

Duffy et al. *Malar J* (2016) 15:258
DOI 10.1186/s12936-016-1296-4

Malaria Journal

RESEARCH Open Access

Differences in PfEMP1s recognized by antibodies from patients with uncomplicated or severe malaria

Michael F. Duffy^{1*}, Rintis Noviyanti², Takafumi Tsuboi³, Zhi-Ping Feng^{4,5}, Leily Trianty², Boni F. Sebayang², Eizo Takashima³, Fransisca Sumardy¹, Daniel A. Lampah⁶, Louise Turner⁷, Thomas Lavstsen⁷, Freya J. I. Fowkes⁸, Peter Siba⁹, Stephen J. Rogerson¹, Thor G. Theander⁷, Jutta Marfurt¹⁰, Ric N. Price^{10,11}, Alexander M. Anstey¹⁰, Graham V. Brown¹² and Anthony T. Pappenfuss^{4,5,13,14}

Abstract
Background: *Plasmodium falciparum* erythrocyte membrane protein 1 (PfEMP1) variants and mediate pathogenic cytoadhesion and antigenic variation in malaria. PfEMP1s are categorized into three principal groups (A, B and C) and they contain conserved arrangements of functional domains called domain cassettes. Despite their tremendous diversity there is compelling evidence that a restricted subset of PfEMP1s is expressed in severe disease. In this study antibodies from patients with severe and uncomplicated malaria were compared for differences in reactivity with a range of PfEMP1s to determine whether antibodies to particular PfEMP1 domains were associated with severe or uncomplicated malaria.
Methods: Parts of expressed *var* genes in a severe malaria patient were identified by RNAseq and several of these partial PfEMP1 domains were expressed together with others from laboratory isolates. Antibodies from Papuan patients to these parts of multiple PfEMP1 proteins were measured.
Results: Patients with uncomplicated malaria were more likely to have antibodies that recognized PfEMP1 of Group C type and recognized a broader repertoire of group A and B PfEMP1s than patients with severe malaria.
Conclusion: These data suggest that exposure to a broad range of group A and B PfEMP1s is associated with protection from severe disease in Papua, Indonesia.
Keywords: Severe malaria, *var* genes, PfEMP1

Background molecules are encoded by the *var* multigene family [1–3].

	citation	code
Unique identification	✓	✓
Persistence	✓	✓
Accessibility	✓	-
Specificity	-	-

RELATED MATERIALS

1. URL - References <http://github.com/qed-lab/khmer/>

• Bioinformatics and computational biology not elsewhere classified

KEYWORDS

bioinformatics Bioinformatics

Repository record is only a pointer record, no files



Browse

Search on figshare...



Log in Sign up

File(s) not publicly available

Reason: I am trying to have one single repo only, But figshare should still link to the file.

PIVlab - Time-Resolved Digital Particle Image Velocimetry Tool for MATLAB

Cite

Share

+ Collect

Version 19 Online Resource posted on 2021-01-28, 13:40 authored by [William Thielicke](#), Eize J.

Stamhuis

USAGE METRICS

22293
views

3119
downloads

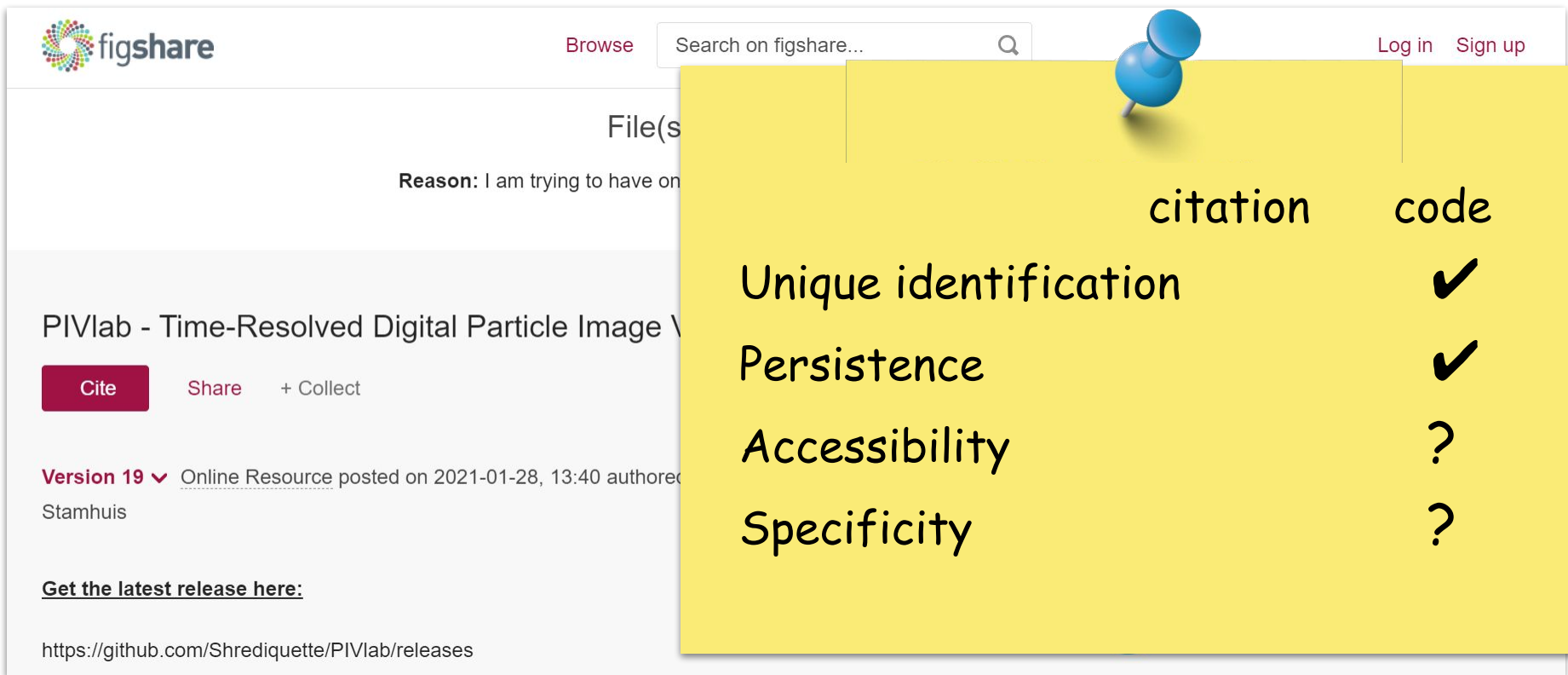
86
citations

Get the latest release here:

<https://github.com/Shrediquette/PIVlab/releases>



Repository record is only a pointer record, no files



figshare

Browse Search on figshare... Q Log in Sign up

File(s)

Reason: I am trying to have on

PIVlab - Time-Resolved Digital Particle Image V

Cite Share + Collect

Version 19 Online Resource posted on 2021-01-28, 13:40 authored by Stamhuis

Get the latest release here:

<https://github.com/Shrediquette/PIVlab/releases>

	citation	code
Unique identification		✓
Persistence		✓
Accessibility		?
Specificity		?

Shared properly, citation confusing

ARTICLE

Measuring Library Broadband Networks to Address Knowledge Gaps and Data Caps

Chris Ritzo, Colin Rhinesmith, and Jie Jiang

ABSTRACT

In this paper, we present findings from a three-year research project funded by the US Institute of Museum and Library Services that examined how advanced broadband measurement capabilities can support the infrastructure and services needed to respond to the digital demands of public library users across the US. Previous studies have identified the ongoing broadband challenges of public libraries while also highlighting the increasing digital expectations of their patrons. However, few large-scale research efforts have collected automated, longitudinal measurement data on library broadband speeds and quality of service at a local, granular level inside public libraries over time including when buildings are closed. This research addresses the following research question: How can public libraries develop a better understanding of the broadband services they receive? In response, we developed a quantitative measurement system that was both developed and deployed across the US. Findings from our analysis of the data can confirm when the library's internet connection is not. When measurements are not consistent with advertised speeds, we analyze the differences and correlate this with additional local factors. These findings support measurements conducted by the library to enable local control and monitoring of this vital service and support critique and interrogation of the differences between internet measurement platforms. In addition, we learned that speed tests are useful for examining these trends but are only a small part of assessing an internet connection and how well it can be used for specific purposes. These findings have implications for state library agencies and federal policymakers interested in having access to data on observed versus advertised speeds and quality of service of public library broadband connections nationwide.

<https://doi.org/10.6017/ital.v41i3.13775>

INTRODUCTION

The COVID-19 pandemic exposed the severity of the digital divide in the United States. During this time, lack of access to computers and the internet has been highlighted among individuals and families with limited monthly incomes in tribal, rural, and urban communities where broadband is neither available nor affordable. Decades of research has shown that this digital divide is further deepened along racial and ethnic lines. Wealthier, white, and more educated individuals

Data Accessibility

The datasets supporting this article have been uploaded to the Harvard Dataverse, located here:

<https://dataverse.harvard.edu/dataverse/mlbn>

Shared properly, citation confusing

ARTICLE

Measuring Library Broadband Networks to Address Knowledge Gaps and Data Caps

Chris Ritzo, Colin Rhinesmith, and Jie Jiang

ABSTRACT

In this paper, we present findings from a three-year research project funded by the US Institute of Museum and Library Services that examined how advanced broadband measurement capabilities can support the infrastructure and services needed to respond to the digital demands of public library users across the US. Previous studies have identified the ongoing broadband challenges public libraries while also highlighting the increasing digital expectations of their patrons. However, few large-scale research efforts have collected automated, longitudinal measurement data on broadband speeds and quality of service at a local, granular level inside public libraries over time, including when buildings are closed. This research addresses the following research question: How can public libraries develop a better understanding of the broadband network they receive? In response, we developed a quantitative measurement system that was both developed for use across the US. Findings from our analysis of the data can confirm when the library's internet connection is not. When measurements are not consistent with differences and correlate this with additional local measurements conducted by the library enable local control and monitoring of this vital service support critique and interrogation of the differences between internet measurement platforms. In addition, we learned that speed tests are useful for examining these trends but are only a small part of assessing an internet connection and how well it can be used for specific purposes. These findings have implications for state library agencies and federal policymakers interested in having accurate data on observed versus advertised speeds and quality of service of public library broadband connections nationwide.

<https://doi.org/10.6017/ital.v41i3.13775>

INTRODUCTION

The COVID-19 pandemic exposed the severity of the digital divide in the United States. During this time, lack of access to computers and the internet has been highlighted among individuals in families with limited monthly incomes in tribal, rural, and urban communities where broadband is neither available nor affordable. Decades of research has shown that this digital divide is fueled by deepened class, racial, and ethnic lines. Wealthier, white, and more educated individuals

Data Accessibility
The datasets supporting this research are available at <https://dataverse.harvard.edu>

HARVARD Dataverse Add Data Search About User Guide Support Sign Up Log In

Measuring Library Broadband Networks for the National Digital Platform, is a research grant from by the Institute of Museum and Library Services (IMLS) National Leadership Grant for Libraries program (award #L.G-71-18-010-18). The research is led by Dr. Colin Rhinesmith, Associate Professor and Director of the Community Informatics Lab at the Simmons University School of Library and Information Science, along with Measurement Lab (M-Lab) and Internet2 to examine how advanced broadband measurement capabilities can support the infrastructure and services needed to respond to the digital demands of public library users across the U.S.

Search this dataverse... Advanced Search

1 to 7 of 7 Results

- Measuring Library Broadband Networks Training Manual - Final**
May 17, 2021
Ritzo, Chris, Block, Carson, Rhinesmith, Colin, 2021, "Measuring Library Broadband Networks Training Manual - Final", <https://doi.org/10.7910/DVN/8XXXZQ>, Harvard Dataverse, V1
This dataset contains: The Training Manual with complete documentation about the Measuring Library Broadband Networks project, including an overview of the project, measurement computer setup instructions, software running on the MBLN measurement computers, the data visualization...
- Year 2 MLBN Measurements**
May 17, 2021
Ritzo, Chris, 2021, "Year 2 MLBN Measurements", <https://doi.org/10.7910/DVN/0XZF01>, Harvard Dataverse, V1, UNF:6:RD1ubiH4dE8N4k76MnOwMq== [fileUNF]
This dataset contains: Measurements conducted by devices in the MLBN program between 2020-02-01 and 2021-01-31 from 30 participating libraries. Data provenance document describing data collection instruments, data preparation and cleaning methods, generation of relevant metadata...
- MLBN Traceroutes and Scripts**
May 17, 2021
Ritzo, Chris, 2021, "MLBN Traceroutes and Scripts", <https://doi.org/10.7910/DVN/HXTY6Z>, Harvard Dataverse, V1, UNF:6:OFTXUPH5o56d88Ue9l9Uqvw== [fileUNF]
This dataset contains Bash scripts used to collect traceroutes, conducted from the egress measurement device at each participating MLBN library. Traceroutes were conducted to each Ookla and M-Lab server during the course of the MLBN program.
- Murakami Measurement Software**
May 17, 2021
Ritzo, Chris, 2021, "Murakami Measurement Software", <https://doi.org/10.7910/DVN/FAGP6S>

Feedback

Shared properly, not cited usefully

ARTICLE

Measuring Library Broadband Knowledge Gaps and Data

Chris Ritzo, Colin Rhinesmith, and Jie Jiang

ABSTRACT

In this paper, we present findings from a three-year Museum and Library Services that examined how can support the infrastructure and services needed by library users across the US. Previous studies have public libraries while also highlighting the increased few large-scale research efforts have collected and broadband speeds and quality of service at a local including when buildings are closed. This research the following research question: How can public develop a better understanding of the broadband receive? In response, quantitative measurement measurement system that was both developed for across the US. Findings from our analysis of the can confirm when the library's internet connection not. When measurements are not consistent with differences and correlate this with additional local measurements conducted by the library enable to support critique and interrogation of the different addition, we learned that speed tests are useful for of assessing an internet connection and how well have implications for state library agencies and for data on observed versus advertised speeds and quality connections nationwide.

<https://doi.org/10.7910/DVN/FAGPSS>

INTRODUCTION

The COVID-19 pandemic exposed the severity of time, lack of access to computers and the internet families with limited monthly incomes in tribal, neither available nor affordable. Decades of research deepened class, racial, and ethnic lines. Working

HARVARD

Dataverse

Description

This dataset contains the code used on each MLBN measurement device, copied from GitHub: <https://github.com/m-lab/murakami/>

Subject

Computer and Information Science

Keyword

broadband, libraries

License/Data Use Agreement

 CC0 1.0

Files Metadata Terms Versions

Citation Metadata

Persistent Identifier [doi:10.7910](#)

Publication Date 2021-05-

Title Murakami

Author Ritzo, Ch

Point of Contact Use ema

Ritzo, Ch

Description This data

<https://git>

Subject Compute

Keyword broadband

libraries

Funding Information Institute of Museum and Library Services: LG-71-18-0110-18



	citation	code
Unique identification	-	✓
Persistence	-	✓
Accessibility	-	✓
Specificity	-	✓



So, how representative are these examples?

Metadata harvested from Figshare:

Of 4,972 software records published 2013-2023

- 99% have a doi (= unique identifier)
- 100% have repository metadata (= persistence)
- 94% likely have a code file (= accessibility)
 - 19% of total records link to Github
- 100% *can* be versioned and 22% have at least two versions (= specificity)

file_type	count	proportion
m	2012	0.149870
zip	1944	0.144804
R	1103	0.082160
xlsx	827	0.061601
py	755	0.056238
txt	584	0.043501
sigml	328	0.024432
csv	307	0.022868
ipynb	256	0.019069
gz	235	0.017505
pdf	215	0.016015

Conclusions

- Persistent identifiers and versioning are essential for code sharing
- Code in repositories is overall FAIR, citations are the difficulty
 - If your repository...
 - Has a GitHub integration - encourage people to use it
 - Offers flexibility with adding files - consider requiring software files for software records (rather than allowing pdf or no files)
- Provide software citation guidance

References

Digital Science; Simons, Natasha; Goodey, Greg; Hardeman, Megan; Clare, Connie; Gonzales, Sara; et al. (2021). The State of Open Data 2021. Digital Science. Report.

<https://doi.org/10.6084/m9.figshare.17061347.v1>

[Figshare code repository setup & implementation: GitHub, GitLab, and Bitbucket - a help article for using figshare](#)

Public Library of Science (2022). PLOS Open Science Indicators. Public Library of Science. Dataset.

<https://doi.org/10.6084/m9.figshare.21687686.v2>

Smith AM, DS Katz, KE Niemeyer, and FORCE11 Software Citation Working Group (2016) Software citation principles. *PeerJ Computer Science* 2:e86 <https://peerj.com/articles/cs-86/>

Xiaoli Chen, Helena Cousijn, & Kelly Stathis. (2022). Implementing FAIR Workflows D1.1 Workflows Specification. Zenodo. <https://doi.org/10.5281/zenodo.7382642>

Thank you

Maria Cotera <https://orcid.org/0000-0002-1375-0738> maria@figshare.com

Andrew Mckenna-Foster <https://orcid.org/0000-0001-7641-6588> andrew@figshare.com