Advancing science with Dataverse: Publication, discovery, citation, and exploration of research data

The Dataverse® Project

HARVARD
Dataverse

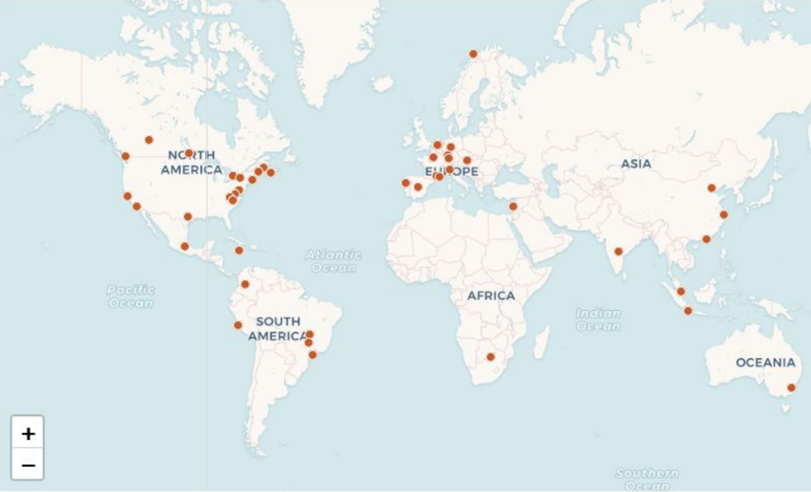
Philip Durbin
@philipdurbin

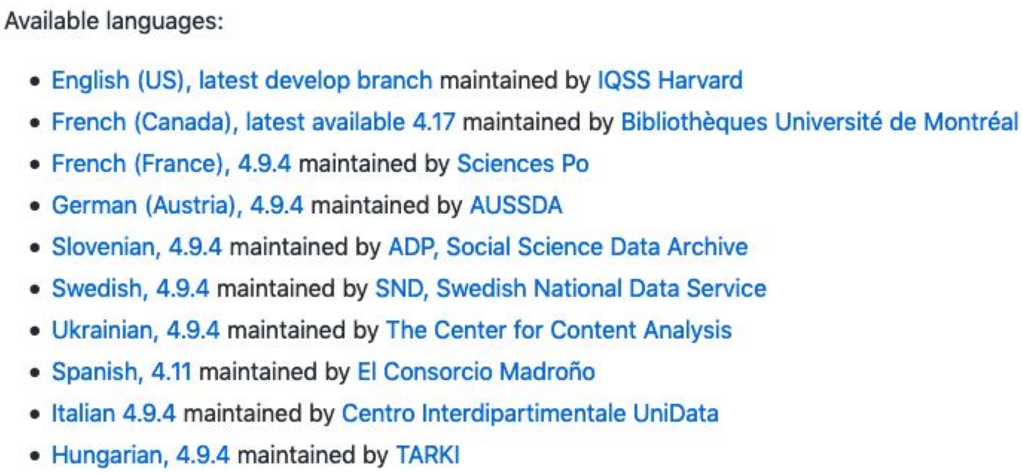
IQSS
The Institute for Quantitative Social Science

What is Dataverse?

research data sharing enthusiasts: scientists, researchers, curators, librarians, etc.

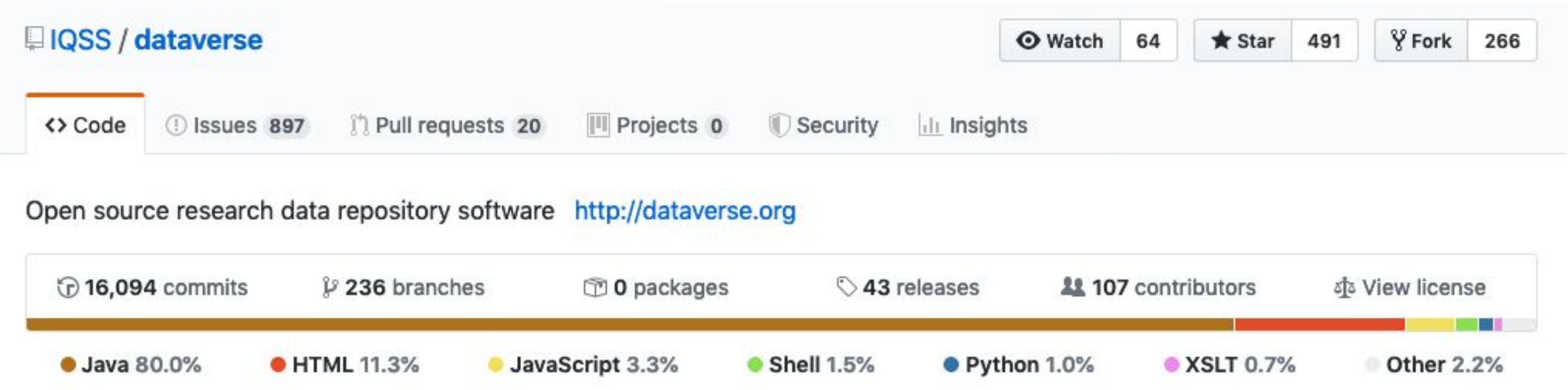
52 Installations





Available languages:

- English (US), latest develop branch maintained by IQSS Harvard
- French (Canada), latest available 4.17 maintained by Bibliothèques Université de Montréal
- French (France), 4.9.4 maintained by Sciences Po
- German (Austria), 4.9.4 maintained by AUSSDA
- Slovenian, 4.9.4 maintained by ADP, Social Science Data Archive
- Swedish, 4.9.4 maintained by SND, Swedish National Data Service
- Ukrainian, 4.9.4 maintained by The Center for Content Analysis
- Spanish, 4.11 maintained by El Consorcio Madroño
- Italian 4.9.4 maintained by Centro Interdipartimentale UniData
- Hungarian, 4.9.4 maintained by TARKI



IQSS / dataverse

Open source research data repository software http://dataverse.org

16,094 commits | 236 branches | 0 packages | 43 releases | 107 contributors | View license

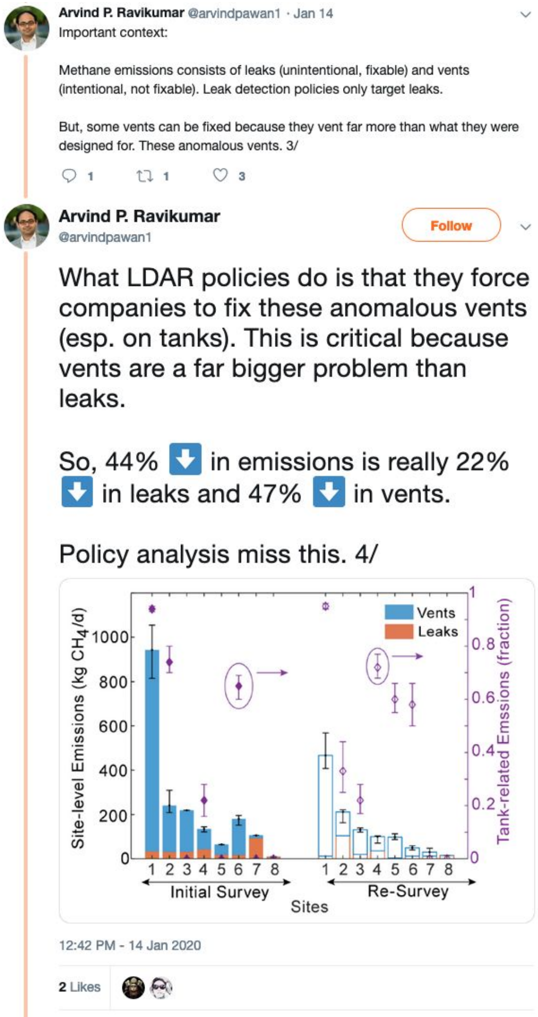
● Java 80.0% ● HTML 11.3% ● JavaScript 3.3% ● Shell 1.5% ● Python 1.0% ● XSLT 0.7% ● Other 2.2%

What is Dataverse for?

Arvind P. Ravikumar

"All my work is built on the premise that climate change is the single biggest existential threat facing humanity."

arvindravikumar.com



Arvind P. Ravikumar @arvindpawan1 · Jan 14

Important context:

Methane emissions consists of leaks (unintentional, fixable) and vents (intentional, not fixable). Leak detection policies only target leaks.

But, some vents can be fixed because they vent far more than what they were designed for. These anomalous vents. 3/

Arvind P. Ravikumar
@arvindpawan1

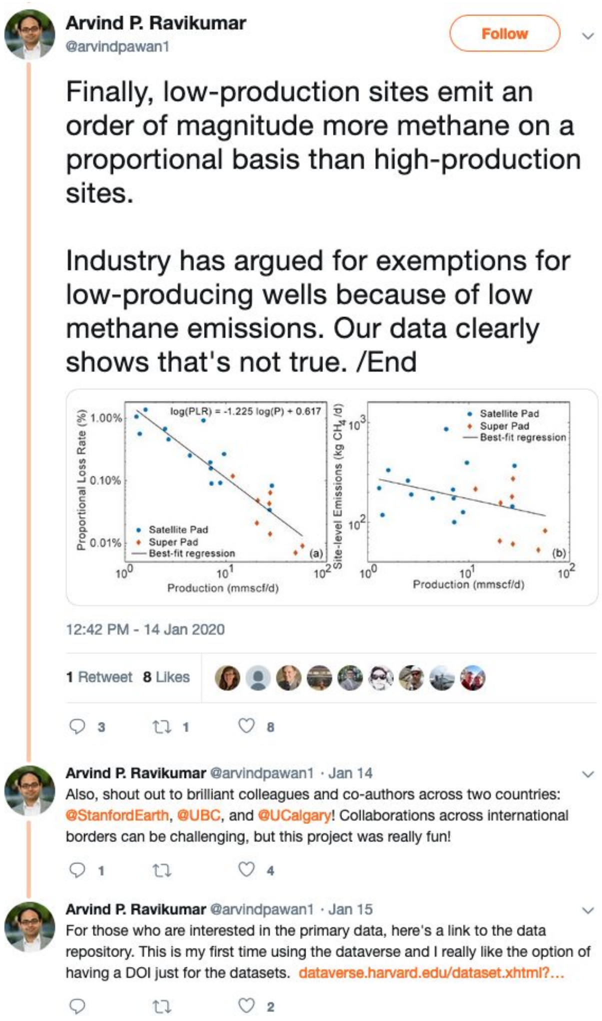
What LDAR policies do is that they force companies to fix these anomalous vents (esp. on tanks). This is critical because vents are a far bigger problem than leaks.

So, 44% ⬇ in emissions is really 22% ⬇ in leaks and 47% ⬇ in vents.

Policy analysis miss this. 4/

12:42 PM - 14 Jan 2020

2 Likes



Arvind P. Ravikumar
@arvindpawan1

Finally, low-production sites emit an order of magnitude more methane on a proportional basis than high-production sites.

Industry has argued for exemptions for low-producing wells because of low methane emissions. Our data clearly shows that's not true. /End

12:42 PM - 14 Jan 2020

1 Retweet 8 Likes

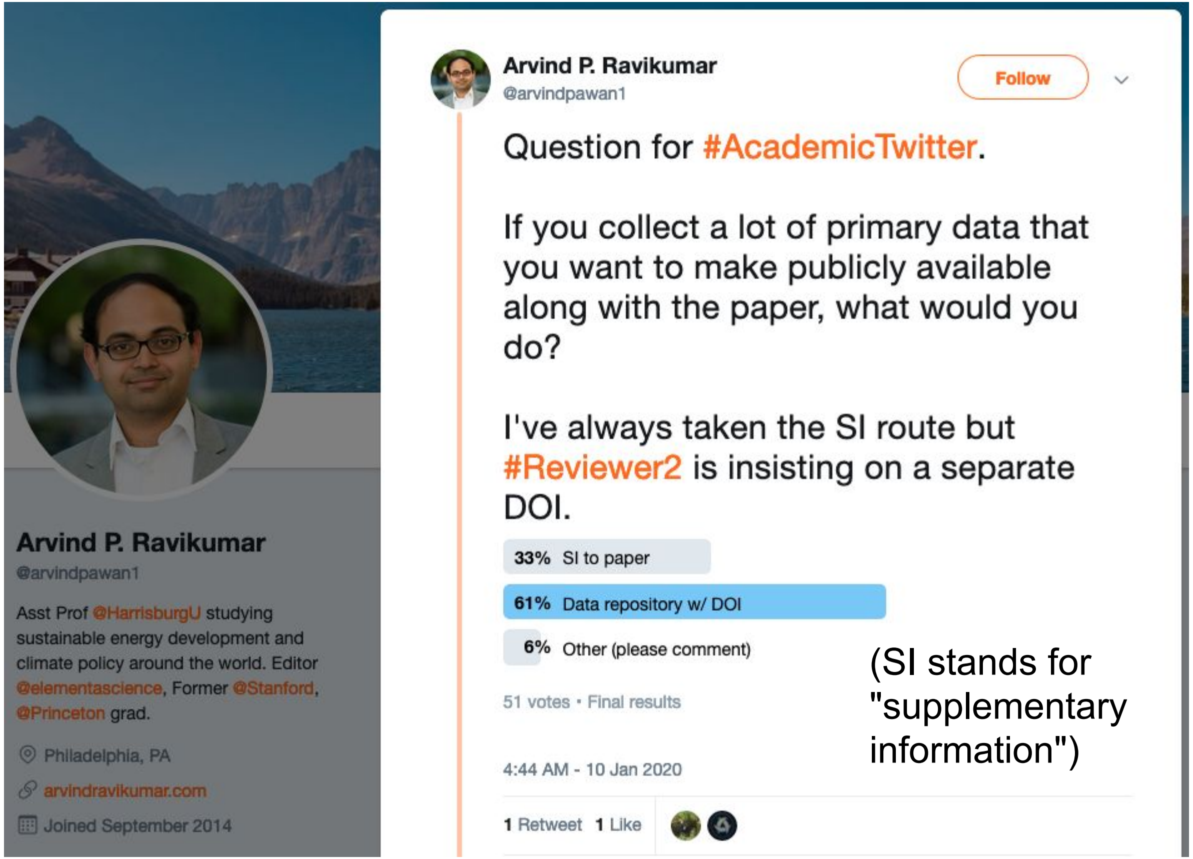
Arvind P. Ravikumar @arvindpawan1 · Jan 14

Also, shout out to brilliant colleagues and co-authors across two countries: @StanfordEarth, @UBC, and @UCalgary! Collaborations across international borders can be challenging, but this project was really fun!

Arvind P. Ravikumar @arvindpawan1 · Jan 15

For those who are interested in the primary data, here's a link to the data repository. This is my first time using the dataverse and I really like the option of having a DOI just for the datasets. dataverse.harvard.edu/dataset.xhtml?...

https://twitter.com/arvindpawan1/status/1217185235298934785



Arvind P. Ravikumar
@arvindpawan1

Asst Prof @HarrisburgU studying sustainable energy development and climate policy around the world. Editor @elementascience, Former @Stanford, @Princeton grad.

Philadelphia, PA
arvindravikumar.com
Joined September 2014

Arvind P. Ravikumar
@arvindpawan1

Question for #AcademicTwitter.

If you collect a lot of primary data that you want to make publicly available along with the paper, what would you do?

I've always taken the SI route but #Reviewer2 is insisting on a separate DOI.

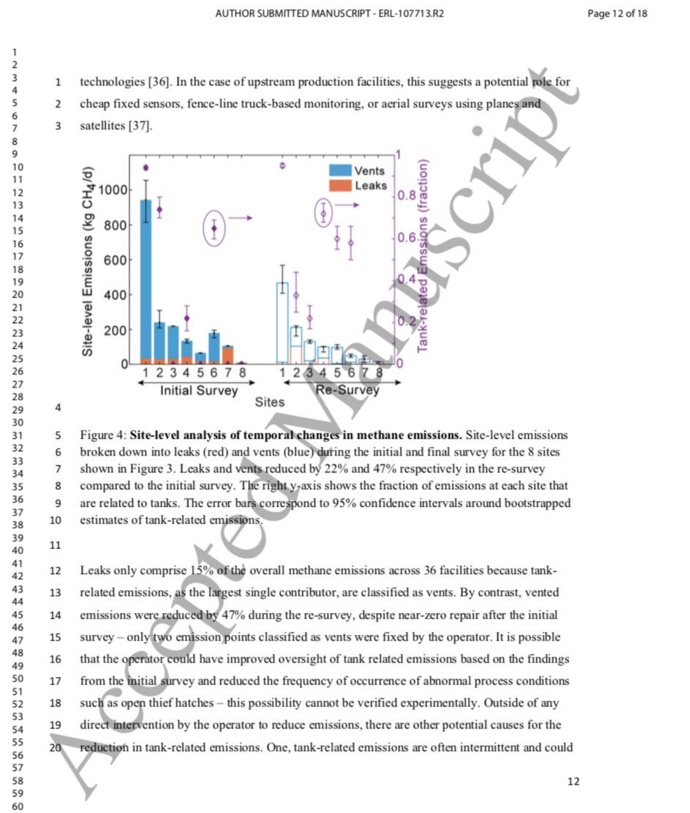
33%	SI to paper
61%	Data repository w/ DOI
6%	Other (please comment)

51 votes • Final results

4:44 AM - 10 Jan 2020

1 Retweet 1 Like

(SI stands for "supplementary information")



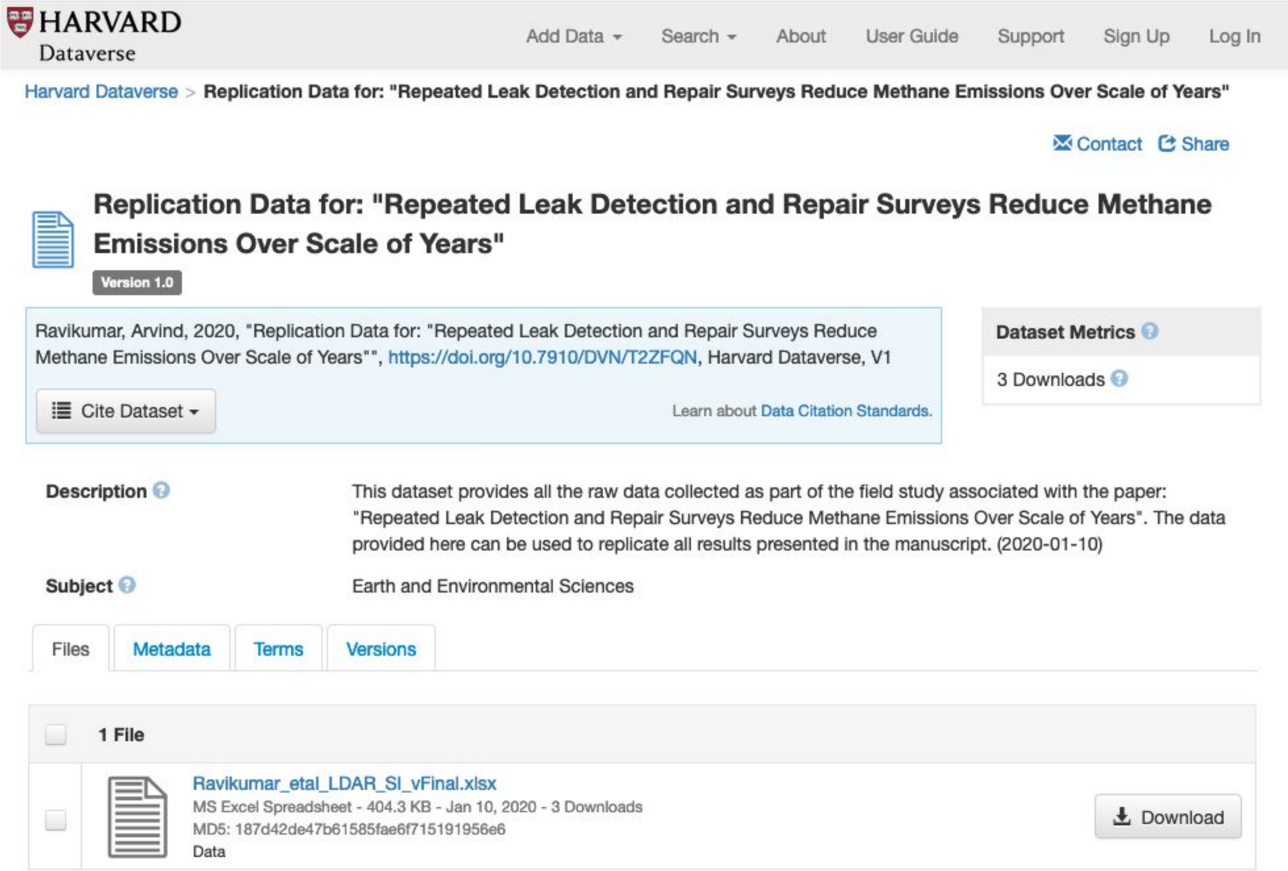
technologies [36]. In the case of upstream production facilities, this suggests a potential role for cheap fixed sensors, fence-line truck-based monitoring, or aerial surveys using planes and satellites [37].

Figure 4: **Site-level analysis of temporal changes in methane emissions.** Site-level emissions broken down into leaks (red) and vents (blue) during the initial and final survey for the 8 sites shown in Figure 3. Leaks and vents reduced by 22% and 47% respectively in the re-survey compared to the initial survey. The right y-axis shows the fraction of emissions at each site that are related to tanks. The error bars correspond to 95% confidence intervals around bootstrapped estimates of tank-related emissions.

Leaks only comprise 15% of the overall methane emissions across 36 facilities because tank-related emissions, as the largest single contributor, are classified as vents. By contrast, vented emissions were reduced by 47% during the re-survey, despite near-zero repair after the initial survey – only two emission points classified as vents were fixed by the operator. It is possible that the operator could have improved oversight of tank related emissions based on the findings from the initial survey and reduced the frequency of occurrence of abnormal process conditions such as open thief hatches – this possibility cannot be verified experimentally. Outside of any direct intervention by the operator to reduce emissions, there are other potential causes for the reduction in tank-related emissions. One, tank-related emissions are often intermittent and could

https://twitter.com/arvindpawan1/status/1215621920080699392

https://doi.org/10.1088/1748-9326/ab6ae1



Harvard Dataverse > Replication Data for: "Repeated Leak Detection and Repair Surveys Reduce Methane Emissions Over Scale of Years"

Replication Data for: "Repeated Leak Detection and Repair Surveys Reduce Methane Emissions Over Scale of Years"

Ravikumar, Arvind, 2020, "Replication Data for: "Repeated Leak Detection and Repair Surveys Reduce Methane Emissions Over Scale of Years"", https://doi.org/10.7910/DVN/T2ZFQN, Harvard Dataverse, V1

Learn about Data Citation Standards.

Dataset Metrics

3 Downloads

Description This dataset provides all the raw data collected as part of the field study associated with the paper: "Repeated Leak Detection and Repair Surveys Reduce Methane Emissions Over Scale of Years". The data provided here can be used to replicate all results presented in the manuscript. (2020-01-10)

Subject Earth and Environmental Sciences

□ 1 File

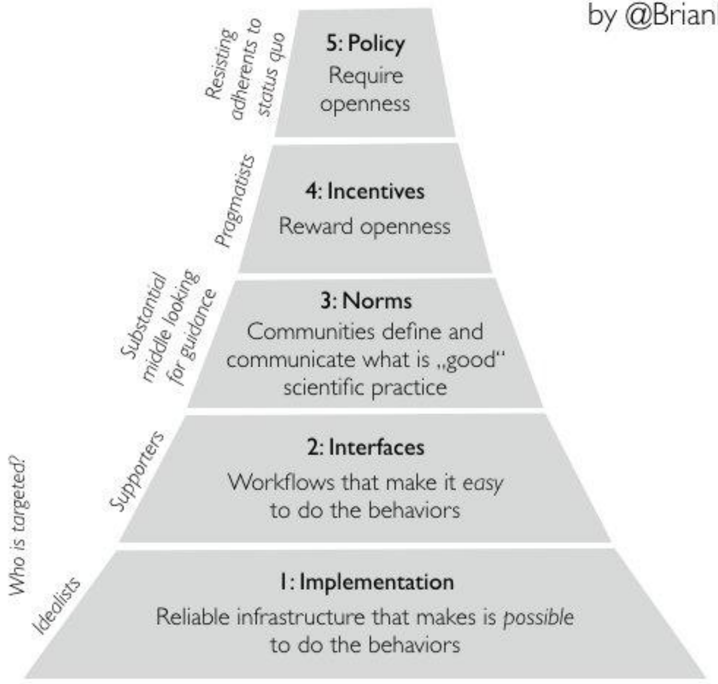
□ Ravikumar_etal_LDAR_SI_vFinal.xlsx
MS Excel Spreadsheet - 404.3 KB - Jan 10, 2020 - 3 Downloads
MD5: 187d42de47b61585fae6f715191956e6
Data

Ravikumar, Arvind, 2020, "Replication Data for: "Repeated Leak Detection and Repair Surveys Reduce Methane Emissions Over Scale of Years"", https://doi.org/10.7910/DVN/T2ZFQN, Harvard Dataverse, V1

Cultural change

How to achieve a cultural change towards open science

Based on a tweet storm by @BrianNosek



5: Policy
Require openness

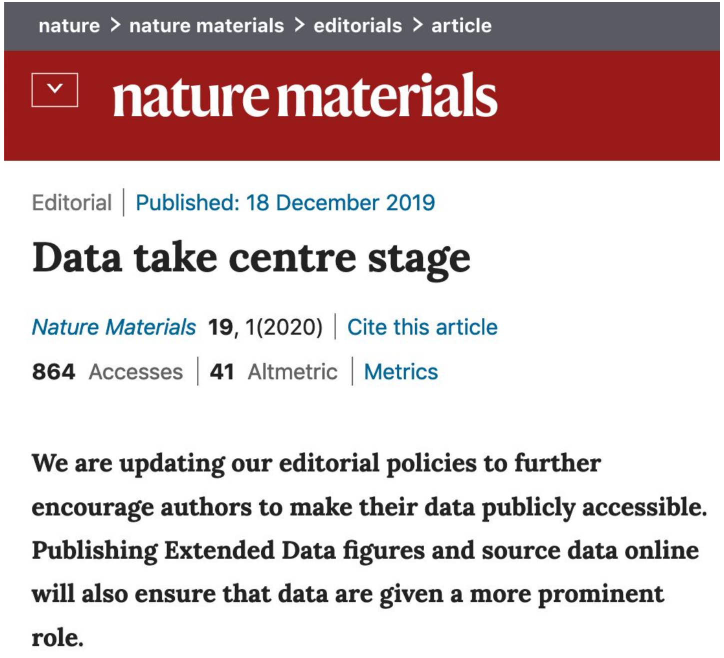
4: Incentives
Reward openness

3: Norms
Communities define and communicate what is „good" scientific practice

2: Interfaces
Workflows that make it *easy* to do the behaviors

1: Implementation
Reliable infrastructure that makes is *possible* to do the behaviors

Who is targeted?
Resisting adherents to status quo
Pragmatists
Substantial middle looking for guidance
Supporters
Idealists



nature > nature materials > editorials > article

nature materials

Editorial | Published: 18 December 2019

Data take centre stage

Nature Materials **19**, 1(2020) | Cite this article

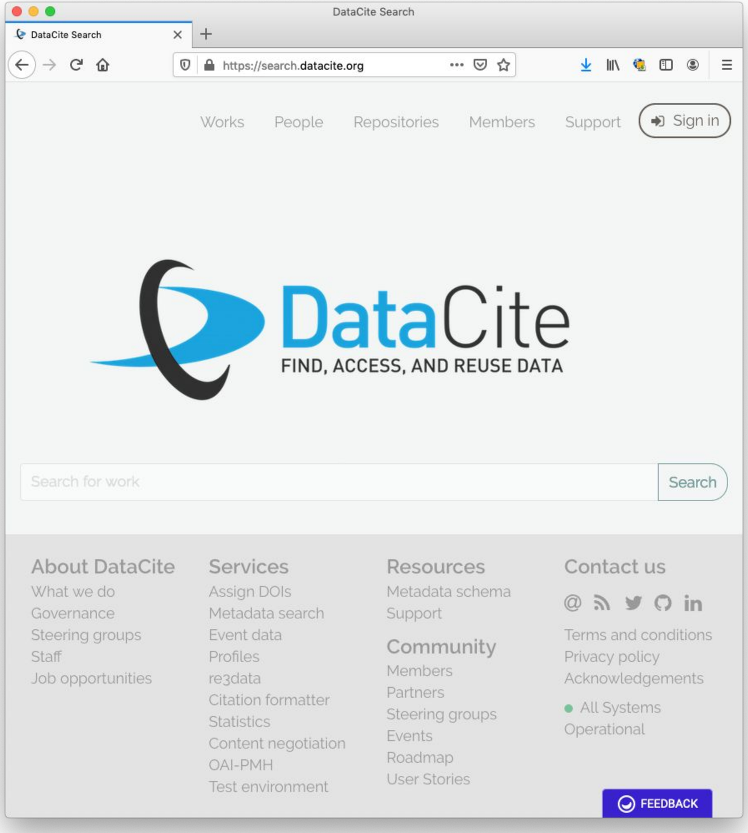
864 Accesses | **41** Altmetric | Metrics

We are updating our editorial policies to further encourage authors to make their data publicly accessible. Publishing Extended Data figures and source data online will also ensure that data are given a more prominent role.

https://www.nature.com/articles/s41563-019-0574-2

https://twitter.com/BrianNosek/status/973506782063677440

Findable



DataCite Search

https://search.datacite.org

Works People Repositories Members Support Sign in

DataCite
FIND, ACCESS, AND REUSE DATA

Search for work | Search

About DataCite
What we do
Governance
Steering groups
Staff
Job opportunities

Services
Assign DOIs
Metadata search
Event data
Profiles
re3data
Citation formatter
Statistics
Content negotiation
OAI-PMH
Test environment

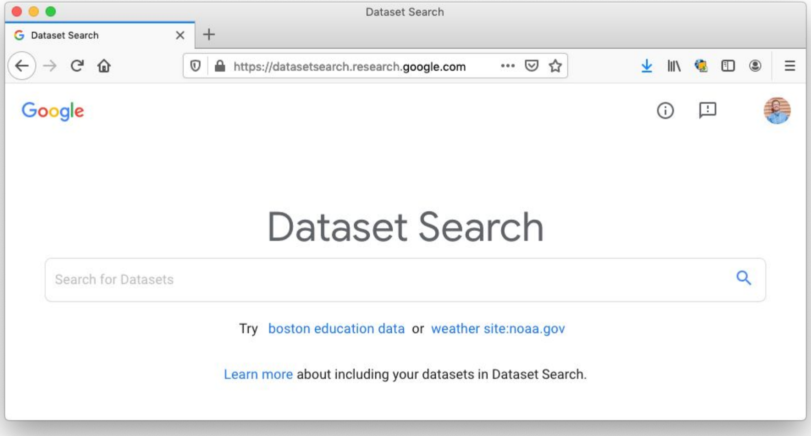
Resources
Metadata schema
Support

Community
Members
Partners
Steering groups
Events
Roadmap
User Stories

Contact us
Terms and conditions
Privacy policy
Acknowledgements
● All Systems Operational

FEEDBACK

https://search.datacite.org



Dataset Search

https://datasetsearch.research.google.com

Google

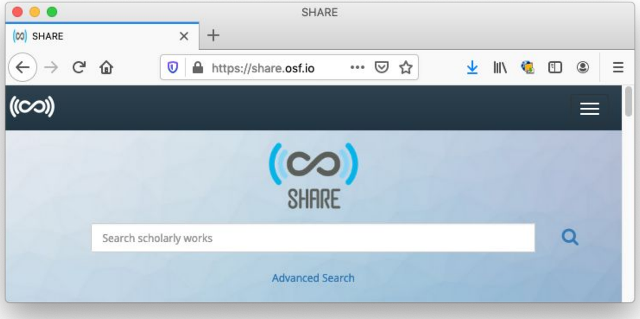
Dataset Search

Search for Datasets

Try boston education data or weather site:noaa.gov

Learn more about including your datasets in Dataset Search.

https://datasetsearch.research.google.com



SHARE

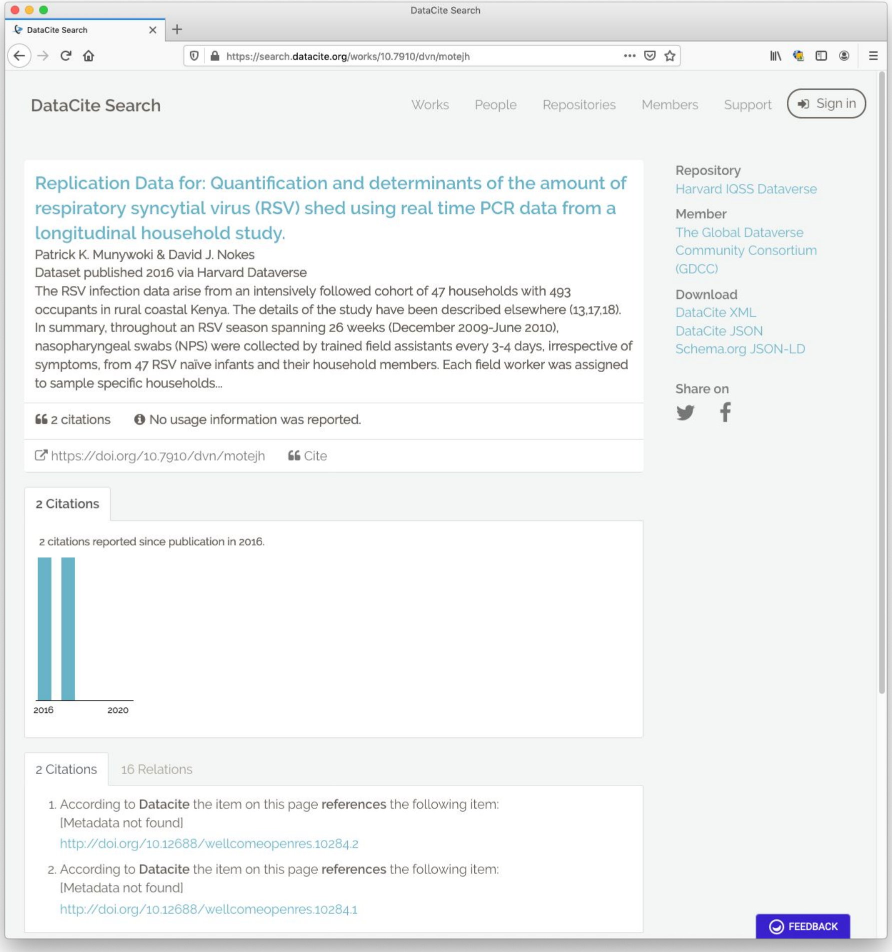
https://share.osf.io

SHARE

Search scholarly works

Advanced Search

https://share.osf.io



DataCite Search

Works People Repositories Members Support Sign in

Replication Data for: Quantification and determinants of the amount of respiratory syncytial virus (RSV) shed using real time PCR data from a longitudinal household study.

Patrick K. Munywoki & David J. Nokes

Dataset published 2016 via Harvard Dataverse

The RSV infection data arise from an intensively followed cohort of 47 households with 493 occupants in rural coastal Kenya. The details of the study have been described elsewhere (13,17,18). In summary, throughout an RSV season spanning 26 weeks (December 2009-June 2010), nasopharyngeal swabs (NPS) were collected by trained field assistants every 3-4 days, irrespective of symptoms, from 47 RSV naïve infants and their household members. Each field worker was assigned to sample specific households...

2 citations No usage information was reported.

https://doi.org/10.7910/dvn/motejh Cite

Repository

Harvard IQSS Dataverse

Member

The Global Dataverse Community Consortium (GDCC)

Download

DataCite XML

DataCite JSON

Schema.org JSON-LD

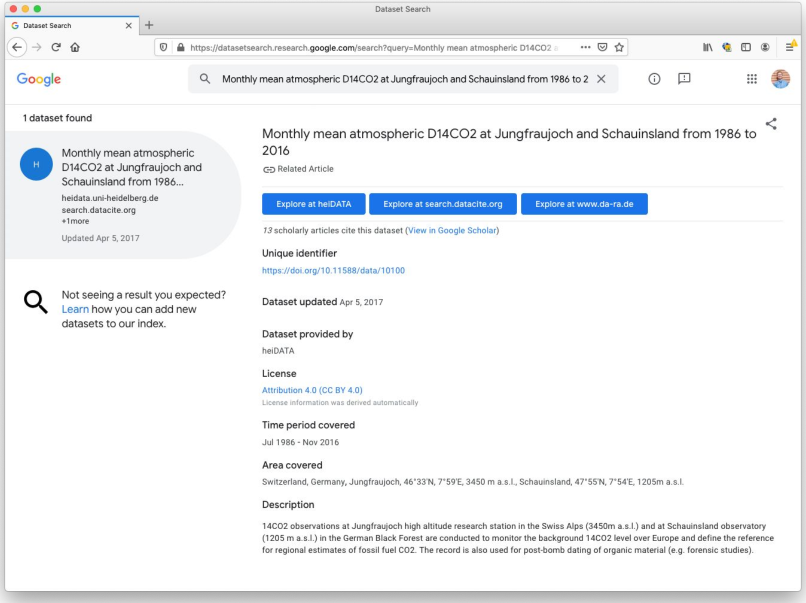
Share on

2 Citations

2 citations reported since publication in 2016.

2 Citations 16 Relations

1. According to **Datacite** the item on this page **references** the following item:
   [Metadata not found]
   http://doi.org/10.12688/wellcomeopenres.10284.2
2. According to **Datacite** the item on this page **references** the following item:
   [Metadata not found]
   http://doi.org/10.12688/wellcomeopenres.10284.1



Dataset Search
Dataset Search
https://datasetsearch.research.google.com/search?query=Monthly mean atmospheric D14CO2 a
Google
Monthly mean atmospheric D14CO2 at Jungfraujoch and Schauinsland from 1986 to 2

1 dataset found

Monthly mean atmospheric D14CO2 at Jungfraujoch and Schauinsland from 1986...
heidata.uni-heidelberg.de
search.datacite.org
+1more
Updated Apr 5, 2017

Not seeing a result you expected? Learn how you can add new datasets to our index.

Monthly mean atmospheric D14CO2 at Jungfraujoch and Schauinsland from 1986 to 2016

Related Article

Explore at heiDATA | Explore at search.datacite.org | Explore at www.da-ra.de

13 scholarly articles cite this dataset (View in Google Scholar)

Unique identifier

https://doi.org/10.11588/data/10100

Dataset updated Apr 5, 2017

Dataset provided by

heiDATA

License

Attribution 4.0 (CC BY 4.0)

License information was derived automatically

Time period covered

Jul 1986 - Nov 2016

Area covered

Switzerland, Germany, Jungfraujoch, 46°33'N, 7°59'E, 3450 m a.s.l., Schauinsland, 47°55'N, 7°54'E, 1205m a.s.l.

Description

14CO2 observations at Jungfraujoch high altitude research station in the Swiss Alps (3450m a.s.l.) and at Schauinsland observatory (1205 m a.s.l.) in the German Black Forest are conducted to monitor the background 14CO2 level over Europe and define the reference for regional estimates of fossil fuel CO2. The record is also used for post-bomb dating of organic material (e.g. forensic studies).

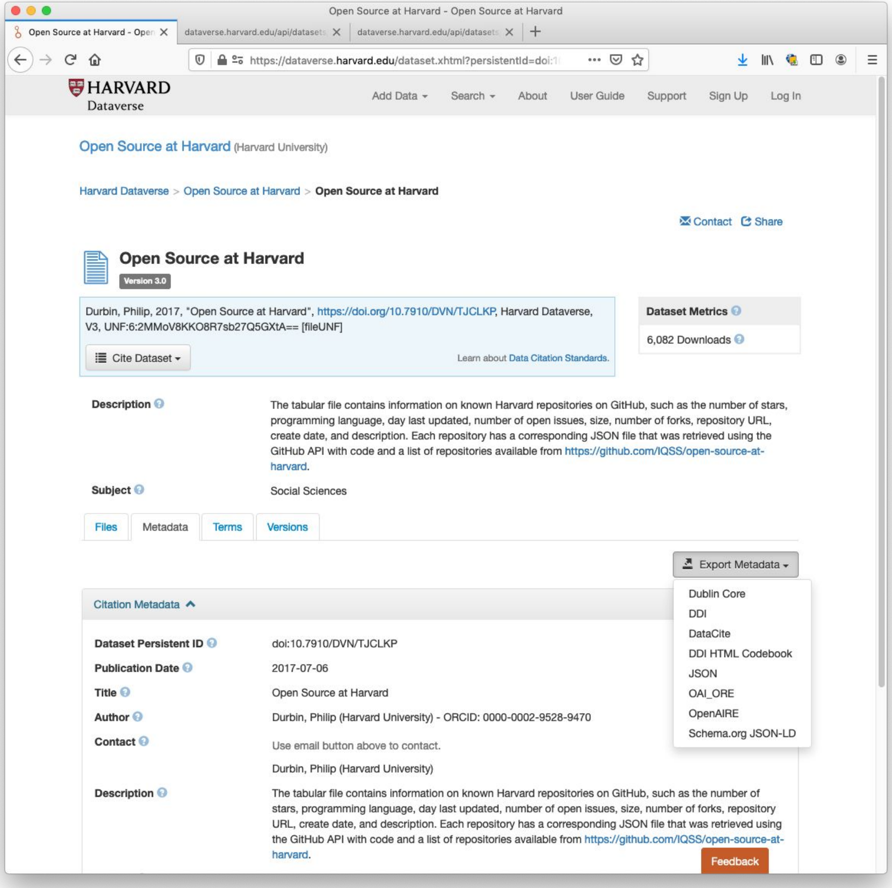
https://datasetsearch.research.google.com

DataCite

FIND, ACCESS, AND REUSE DATA

https://search.datacite.org

Accessible



Open Source at Harvard (Harvard University)

Harvard Dataverse > Open Source at Harvard > Open Source at Harvard

Open Source at Harvard

Durbin, Philip, 2017, "Open Source at Harvard", https://doi.org/10.7910/DVN/TJCLKP, Harvard Dataverse, V3, UNF:6:2MMoV8KKO8R7sb27Q5GXtA== [fileUNF]

Learn about Data Citation Standards.

Dataset Metrics

6,082 Downloads

Description The tabular file contains information on known Harvard repositories on GitHub, such as the number of stars, programming language, day last updated, number of open issues, size, number of forks, repository URL, create date, and description. Each repository has a corresponding JSON file that was retrieved using the GitHub API with code and a list of repositories available from https://github.com/IQSS/open-source-at-harvard.

Subject Social Sciences

Dataset Persistent ID doi:10.7910/DVN/TJCLKP

Publication Date 2017-07-06

Title Open Source at Harvard

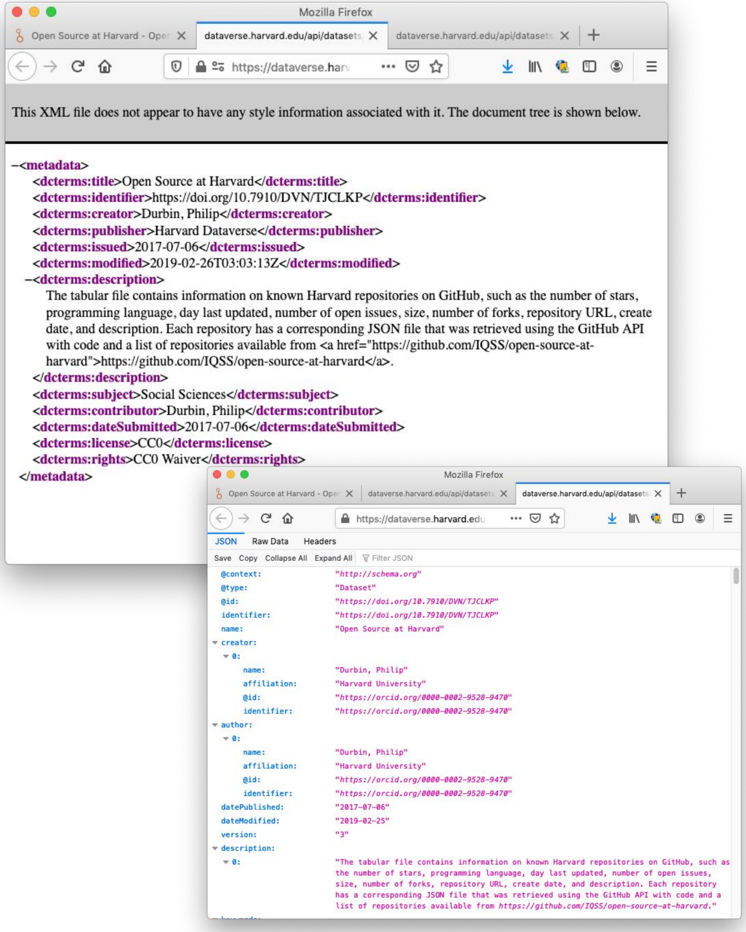
Author Durbin, Philip (Harvard University) - ORCID: 0000-0002-9528-9470

Contact Use email button above to contact.

Durbin, Philip (Harvard University)

Description The tabular file contains information on known Harvard repositories on GitHub, such as the number of stars, programming language, day last updated, number of open issues, size, number of forks, repository URL, create date, and description. Each repository has a corresponding JSON file that was retrieved using the GitHub API with code and a list of repositories available from https://github.com/IQSS/open-source-at-harvard.

https://doi.org/10.7910/DVN/TJCLKP

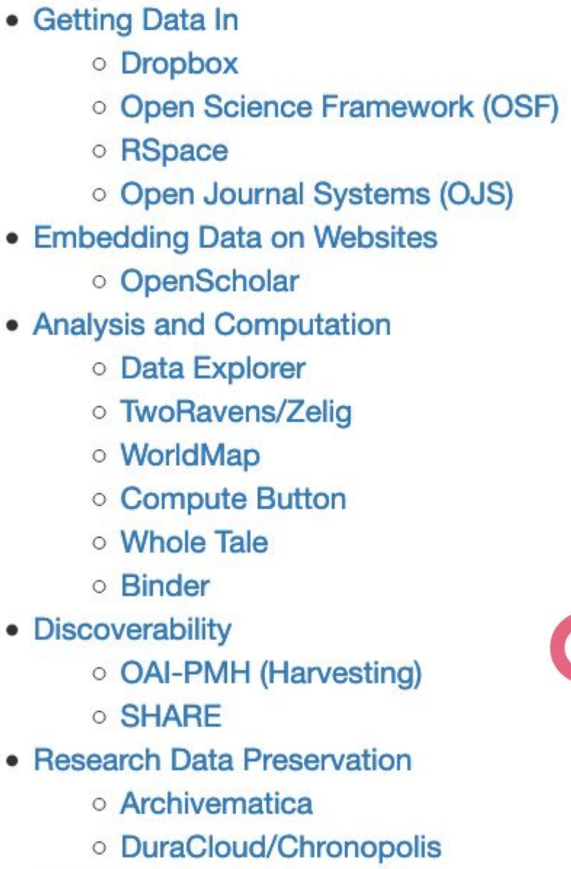

Mozilla Firefox
Open Source at Harvard - Open
dataverse.harvard.edu/api/datasets
dataverse.harvard.edu/api/datasets
https://dataverse.harv
Mozilla Firefox
https://dataverse.harvard.edu
JSON
Raw Data
Headers
Save
Copy
Collapse All
Expand All
Filter JSON

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
-<metadata>
   <dcterms:title>Open Source at Harvard</dcterms:title>
   <dcterms:identifier>https://doi.org/10.7910/DVN/TJCLKP</dcterms:identifier>
   <dcterms:creator>Durbin, Philip</dcterms:creator>
   <dcterms:publisher>Harvard Dataverse</dcterms:publisher>
   <dcterms:issued>2017-07-06</dcterms:issued>
   <dcterms:modified>2019-02-26T03:03:13Z</dcterms:modified>
 -<dcterms:description>
     The tabular file contains information on known Harvard repositories on GitHub, such as the number of stars,
     programming language, day last updated, number of open issues, size, number of forks, repository URL, create
     date, and description. Each repository has a corresponding JSON file that was retrieved using the GitHub API
     with code and a list of repositories available from <a href="https://github.com/IQSS/open-source-at-
     harvard">https://github.com/IQSS/open-source-at-harvard</a>.
   </dcterms:description>
   <dcterms:subject>Social Sciences</dcterms:subject>
   <dcterms:contributor>Durbin, Philip</dcterms:contributor>
   <dcterms:dateSubmitted>2017-07-06</dcterms:dateSubmitted>
   <dcterms:license>CC0</dcterms:license>
   <dcterms:rights>CC0 Waiver</dcterms:rights>
 </metadata>
```

```
  @context:          "http://schema.org"
  @type:             "Dataset"
  @id:               "https://doi.org/10.7910/DVN/TJCLKP"
  identifier:        "https://doi.org/10.7910/DVN/TJCLKP"
  name:              "Open Source at Harvard"
▼ creator:
  ▼ 0:
      name:          "Durbin, Philip"
      affiliation:   "Harvard University"
      @id:           "https://orcid.org/0000-0002-9528-9470"
      identifier:    "https://orcid.org/0000-0002-9528-9470"
▼ author:
  ▼ 0:
      name:          "Durbin, Philip"
      affiliation:   "Harvard University"
      @id:           "https://orcid.org/0000-0002-9528-9470"
      identifier:    "https://orcid.org/0000-0002-9528-9470"
  datePublished:     "2017-07-06"
  dateModified:      "2019-02-25"
  version:           "3"
▼ description:
  ▼ 0:               "The tabular file contains information on known Harvard repositories on GitHub, such as
                     the number of stars, programming language, day last updated, number of open issues,
                     size, number of forks, repository URL, create date, and description. Each repository
                     has a corresponding JSON file that was retrieved using the GitHub API with code and a
                     list of repositories available from https://github.com/IQSS/open-source-at-harvard."
```

Interoperable



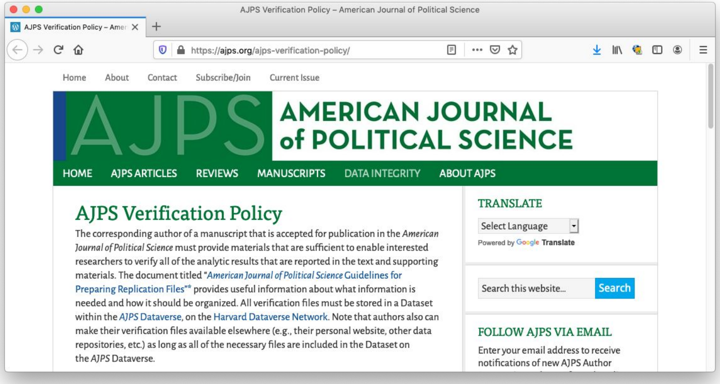
- Getting Data In
  - Dropbox
  - Open Science Framework (OSF)
  - RSpace
  - Open Journal Systems (OJS)
- Embedding Data on Websites
  - OpenScholar
- Analysis and Computation
  - Data Explorer
  - TwoRavens/Zelig
  - WorldMap
  - Compute Button
  - Whole Tale
  - Binder
- Discoverability
  - OAI-PMH (Harvesting)
  - SHARE
- Research Data Preservation
  - Archivematica
  - DuraCloud/Chronopolis

http://guides.dataverse.org/en/4.19/admin/integrations.html

Reusable

THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

https://xkcd.com/1838/


AJPS Verification Policy – American Journal of Political Science
AJPS Verification Policy – Amer
https://ajps.org/ajps-verification-policy/
Home
About
Contact
Subscribe/Join
Current Issue
AJPS
AMERICAN JOURNAL of POLITICAL SCIENCE
HOME
AJPS ARTICLES
REVIEWS
MANUSCRIPTS
DATA INTEGRITY
ABOUT AJPS

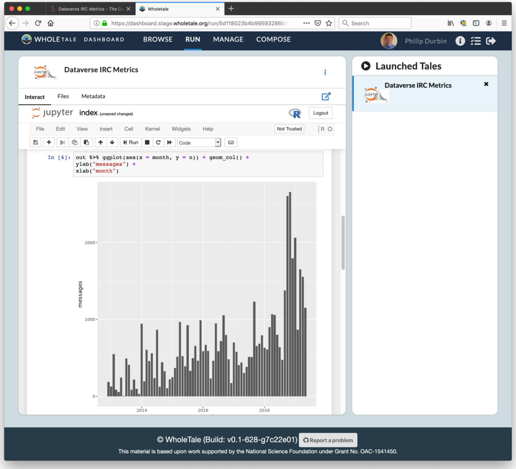
AJPS Verification Policy

The corresponding author of a manuscript that is accepted for publication in the *American Journal of Political Science* must provide materials that are sufficient to enable interested researchers to verify all of the analytic results that are reported in the text and supporting materials. The document titled "*American Journal of Political Science* Guidelines for Preparing Replication Files" provides useful information about what information is needed and how it should be organized. All verification files must be stored in a Dataset within the *AJPS* Dataverse, on the Harvard Dataverse Network. Note that authors also can make their verification files available elsewhere (e.g., their personal website, other data repositories, etc.) as long as all of the necessary files are included in the Dataset on the *AJPS* Dataverse.

TRANSLATE
Select Language
Powered by Google Translate
Search this website...
Search
FOLLOW AJPS VIA EMAIL
Enter your email address to receive notifications of new AJPS Author

https://ajps.org/ajps-verification-policy/

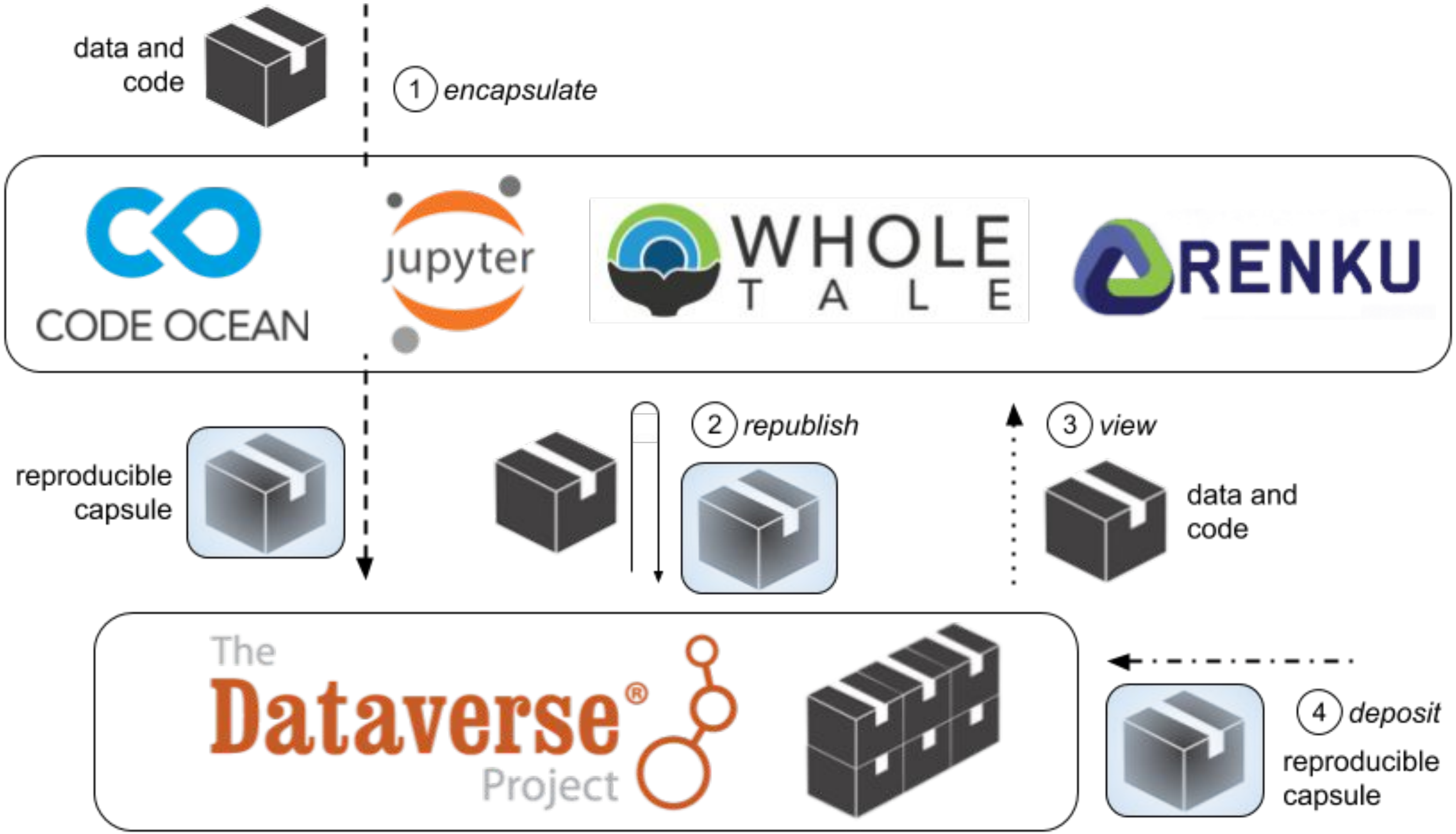
jupyter



```
In [6]: out %>% ggplot(aes(x = month, y = n)) + geom_col() +
        ylab("messages") +
        xlab("month")
```

© WholeTale (Build: v0.1-628-g7c22e01) Report a problem

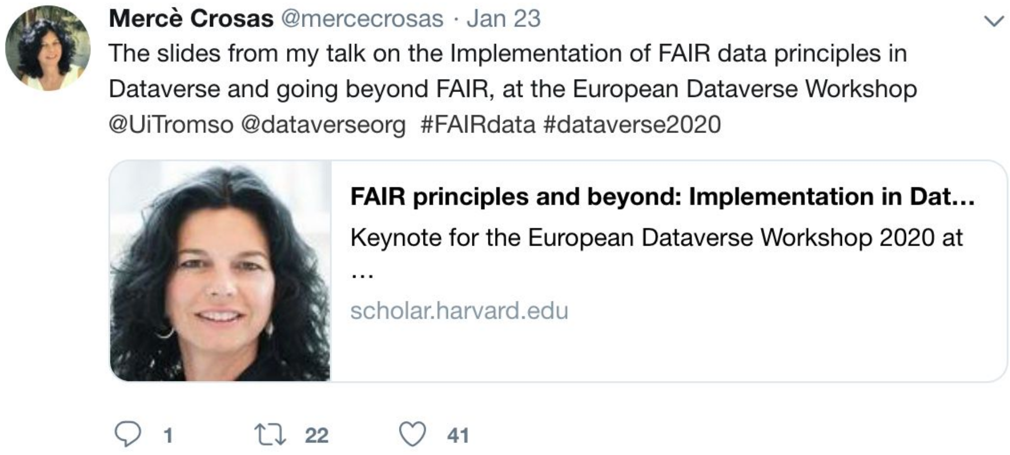
This material is based upon work supported by the National Science Foundation under Grant No. OAC-1541450.



Trisovic, Crosas, et al, 2020, working paper

FAIR Data Principles

- Findable
- Accessible
- Interoperable
- Reusable



Mercè Crosas @mercecrosas · Jan 23

The slides from my talk on the Implementation of FAIR data principles in Dataverse and going beyond FAIR, at the European Dataverse Workshop @UiTromso @dataverseorg #FAIRdata #dataverse2020

FAIR principles and beyond: Implementation in Dat…

Keynote for the European Dataverse Workshop 2020 at …

scholar.harvard.edu

https://twitter.com/mercecrosas/status/1220344995628175360

www.nature.com/scientificdata

SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects[2]—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

There are numerous and diverse stakeholders who stand to benefit from overcoming these obstacles: researchers wanting to share, get credit, and reuse each other's data and interpretations; professional data publishers offering their services; software and tool-builders providing data analysis and processing services such as reusable workflows; funding agencies (private and public) increasingly

Correspondence and requests for materials should be addressed to B.M. (email: barend.mons@dtls.nl).

#A full list of authors and their affiliations appears at the end of the paper.

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18 1

https://dx.doi.org/10.1038/sdata.2016.18

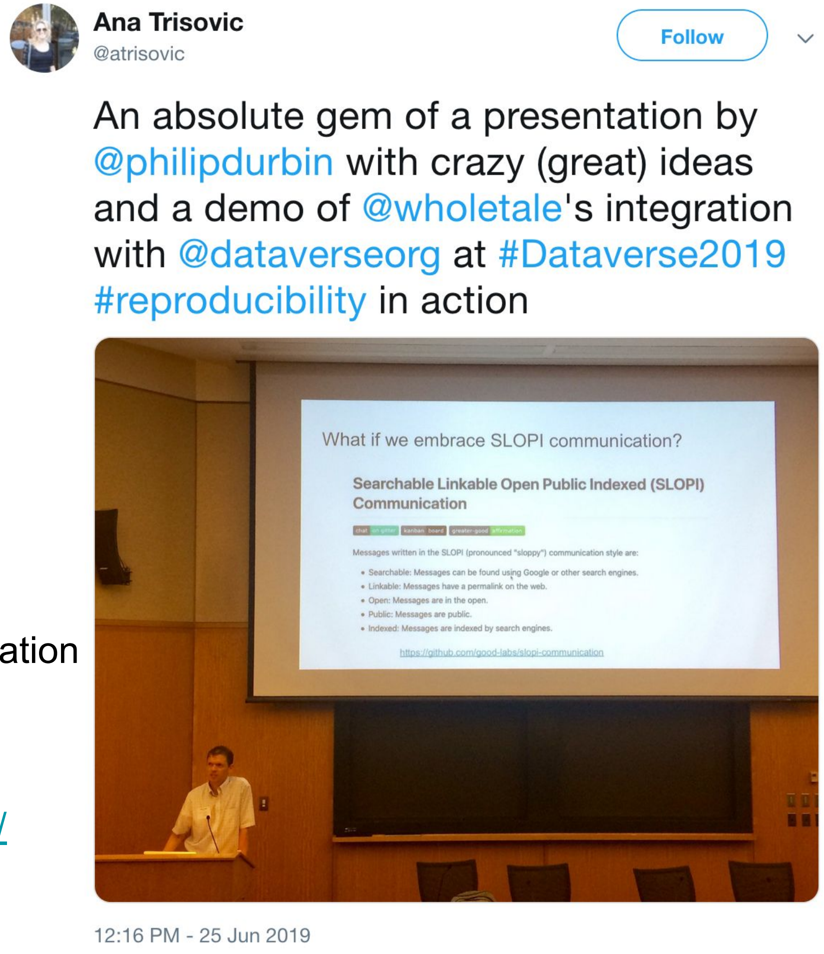
Bonus content

SLOPI

https://github.com/good-labs/slopi-communication

Searchable Linkable Open Public Indexed (SLOPI) Communication
or
Why open source projects should avoid Slack

http://blog.greptilian.com/2020/01/25/slopi-communication/



Ana Trisovic
@atrisovic

An absolute gem of a presentation by @philipdurbin with crazy (great) ideas and a demo of @wholetale's integration with @dataverseorg at #Dataverse2019 #reproducibility in action

What if we embrace SLOPI communication?

Searchable Linkable Open Public Indexed (SLOPI) Communication

Messages written in the SLOPI (pronounced "sloppy") communication style are:

- Searchable: Messages can be found using Google or other search engines.
- Linkable: Messages have a permalink on the web.
- Open: Messages are in the open.
- Public: Messages are public.
- Indexed: Messages are indexed by search engines.

https://github.com/good-labs/slopi-communication

12:16 PM - 25 Jun 2019

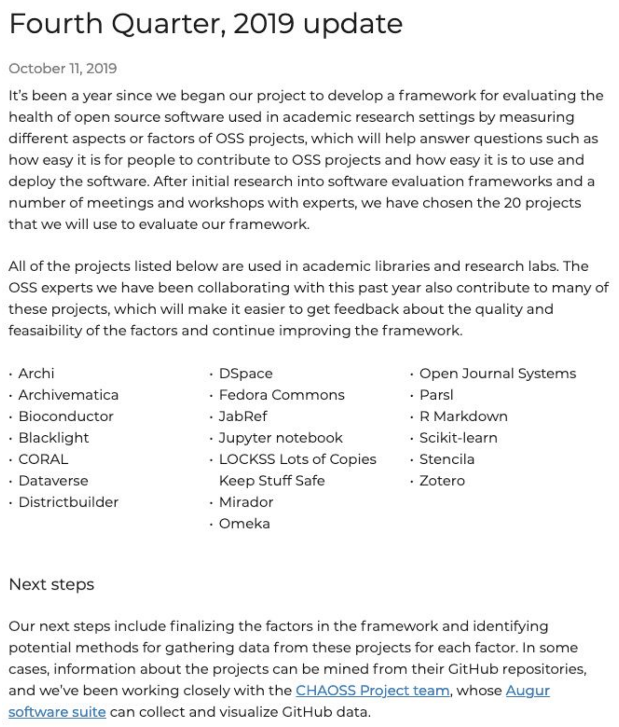
The Open Source Software Health Index Project

CHAOSS

https://chaoss.community

augur

https://github.com/chaoss/augur



Fourth Quarter, 2019 update

October 11, 2019

It's been a year since we began our project to develop a framework for evaluating the health of open source software used in academic research settings by measuring different aspects or factors of OSS projects, which will help answer questions such as how easy it is for people to contribute to OSS projects and how easy it is to use and deploy the software. After initial research into software evaluation frameworks and a number of meetings and workshops with experts, we have chosen the 20 projects that we will use to evaluate our framework.

All of the projects listed below are used in academic libraries and research labs. The OSS experts we have been collaborating with this past year also contribute to many of these projects, which will make it easier to get feedback about the quality and feasaibility of the factors and continue improving the framework.

- Archi
- Archivematica
- Bioconductor
- Blacklight
- CORAL
- Dataverse
- Districtbuilder
- DSpace
- Fedora Commons
- JabRef
- Jupyter notebook
- LOCKSS Lots of Copies Keep Stuff Safe
- Mirador
- Omeka
- Open Journal Systems
- Parsl
- R Markdown
- Scikit-learn
- Stencila
- Zotero

Next steps

Our next steps include finalizing the factors in the framework and identifying potential methods for gathering data from these projects for each factor. In some cases, information about the projects can be mined from their GitHub repositories, and we've been working closely with the CHAOSS Project team, whose Augur software suite can collect and visualize GitHub data.

https://projects.iq.harvard.edu/osshealthindex/blog/fourthquarterupdate

http://blog.greptilian.com/2020/01/26/open-source-health-project/

Thank you!

My website: http://greptilian.com

My blog: http://blog.greptilian.com

@pdurbin on GitHub.

@philipdurbin on Twitter

philip_durbin@harvard.edu

My institute at Harvard: https://www.iq.harvard.edu

For a quick conversation: https://chat.dataverse.org

The Dataverse Project

Created at IQSS at Harvard University