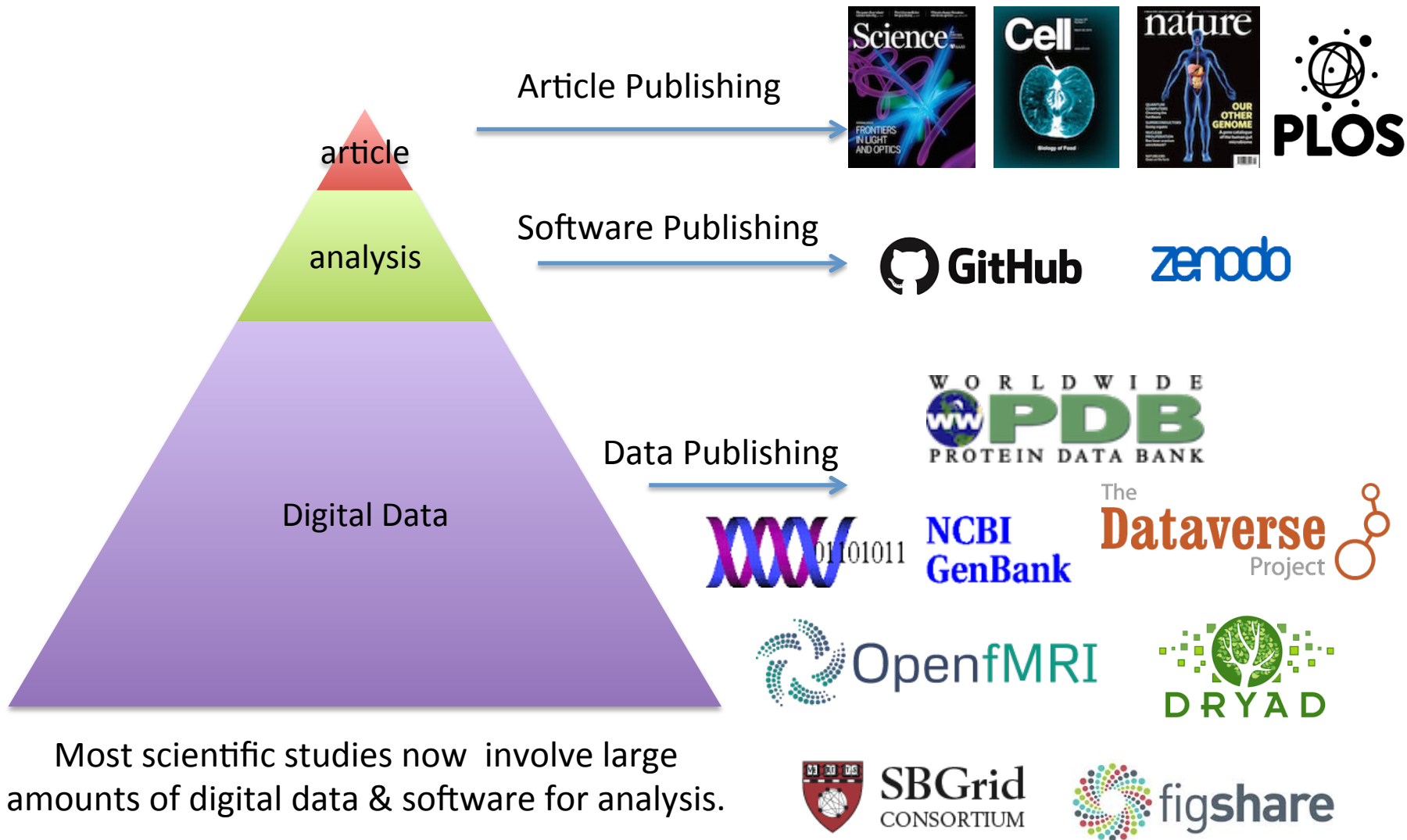


Mercè Crosas, Ph.D.
Chief Data Science and Technology Officer
Institute for Quantitative Social Science
Harvard University
[@mercecrosas](https://twitter.com/mercecrosas)

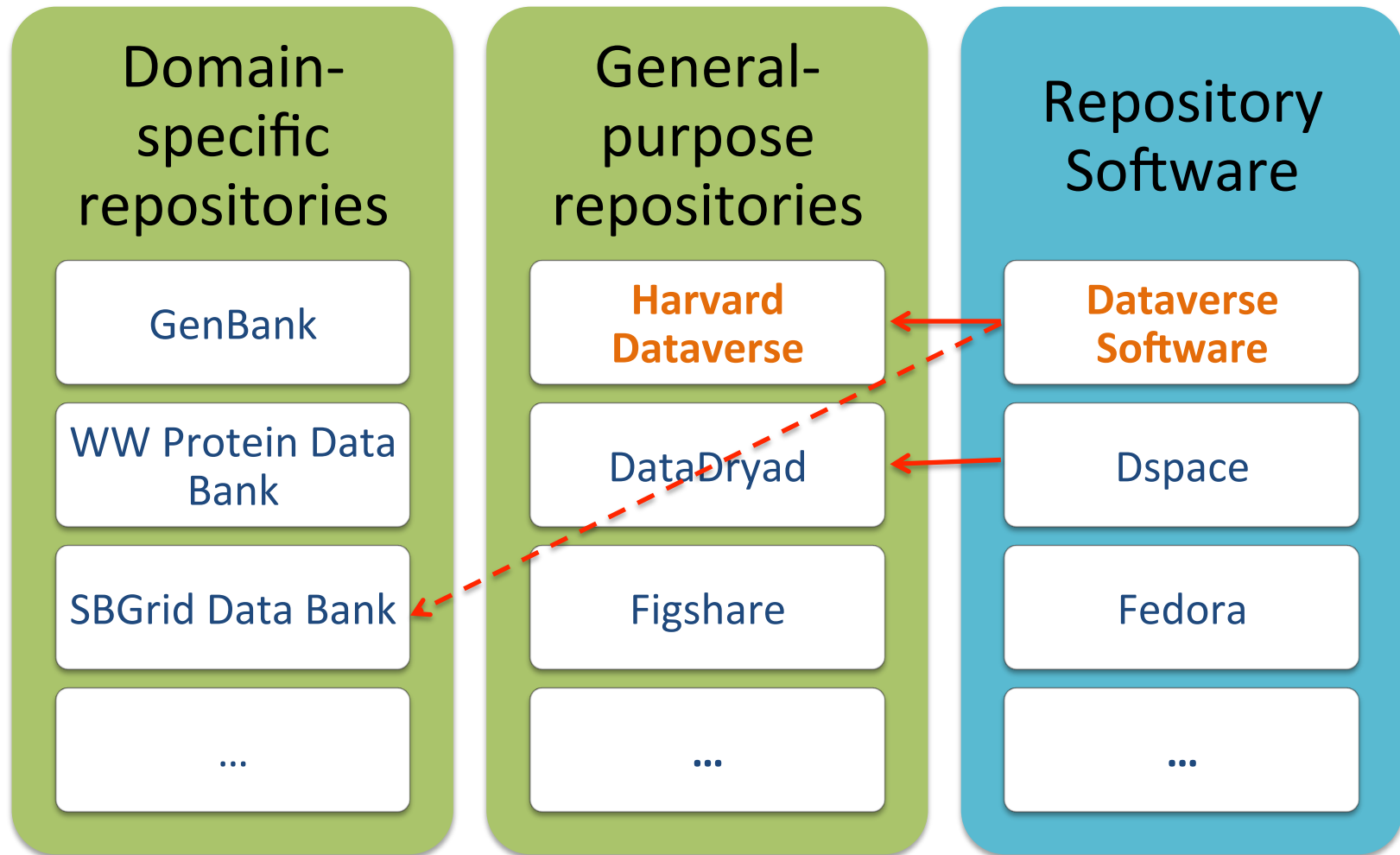
THE DATAVERSE COMMONS

The Future of the Commons, November 18, 2015,

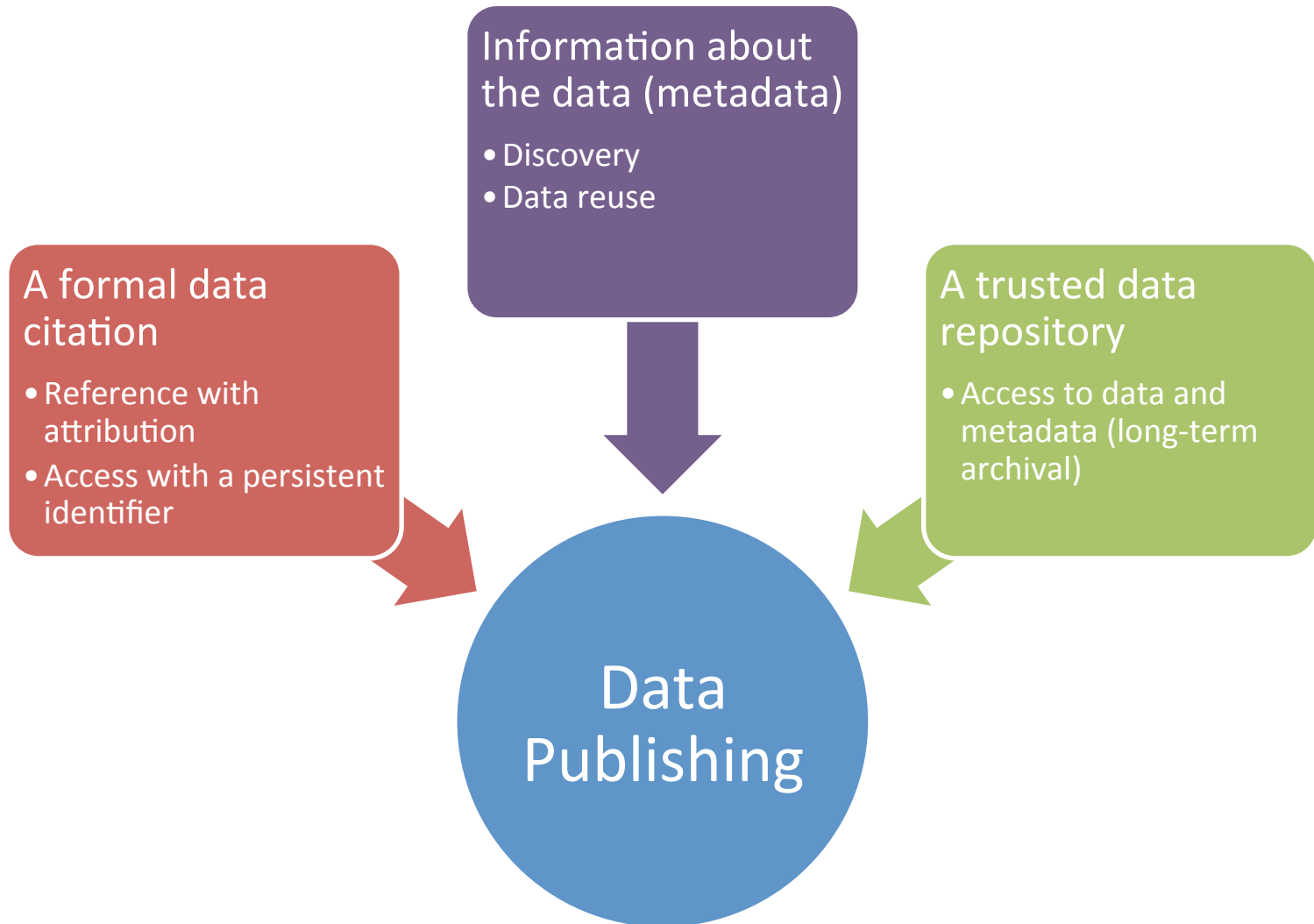
Today's Scholarly Publication Landscape



Data Repositories vs Repository Software



Dataverse follows best practices to support Data Publishing





dataverse.org

Open-source software developed at Harvard's IQSS since 2006
Installed in 12 sites world wide
Serving 100s of universities and organizations

 Dataverse Repositories 





Harvard Dataverse

A collaboration with Harvard Library, Harvard University IT, and IQSS

Metrics 1,417,679 Downloads



Share, publish, and archive your data. Find and cite data across all research fields.

Harvard Dataverse: dataverse.harvard.edu

Started as a community data repository for Social Science

Now open to all research fields and all researchers

More than 1300 dataverses

More than 59,000 datasets

More than 1,400,000 downloads



The Institute for Quantitative Social Science



HARVARD LIBRARY



HARVARD UNIVERSITY

Information Technology

Search

- Data
- Dataset
- File

Dataverse
 Researcher
 Research F
 Organizati
 Journal (58)
 Teaching Cou

Publication Date

- 2015 (14,971)
- 2011 (10,075)
- 2007 (9,586)
- 2012 (8,645)
- 2009 (6,251)

More...



Oct 11, 2015 - MIT Libraries Dataverse

Centre for European Policy Studies, 2015, "Lending to Households in Europe (1995-2014): ECRI Statistical Package 2015", <http://dx.doi.org/10.7910/DVN/51SIMV>, Harvard Dataverse, V1

The ECRI Statistical Package on Lending to Households in Europe is a collection of data on lending to non-financial corporations and households, including consumer credit, housing and other loans, in Europe, covering 40 countries: the 28 EU member states, three EU candidate count...

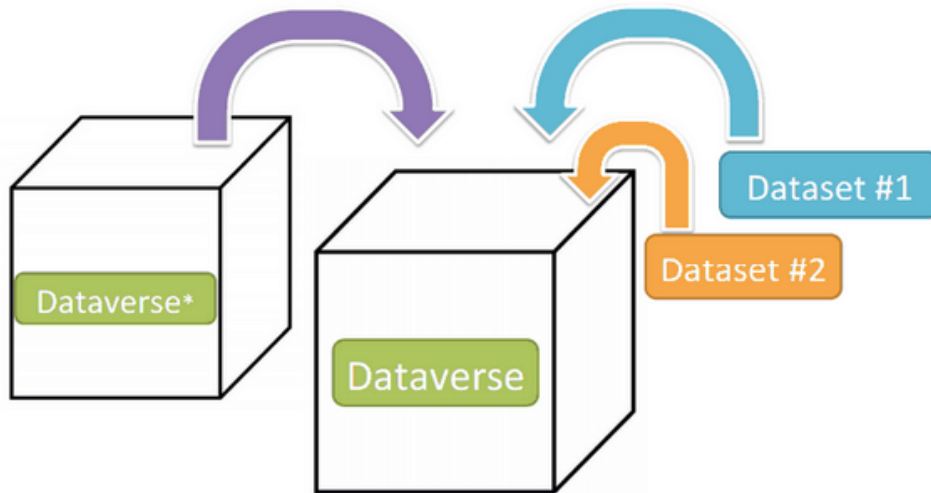


Replication Data for: A rhythm landscape approach to the developmental dynamics of birdsong rhythm

Oct 10, 2015

Dataverses are containers for Datasets

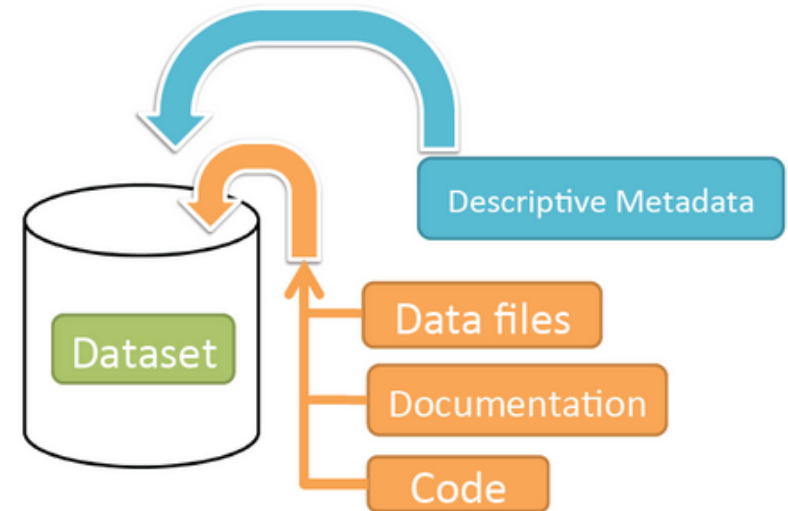
Schematic Diagram of a **Dataverse** in Dataverse 4.0



Container for your **Datasets** and/or **Dataverses***

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.

Each Dataverse can be for a researcher, a research project, a department, a journal, or a larger organization.

Dataverse offers a rich feature set to publish research data

Credit and Visibility

- Standard, persistent data citation
- Branding for each dataverse
- Widgets to embed in your own website

Discovery

- Faceted search for all metadata
- Standard metadata:
 - citation
 - scientific domain
 - file-level

Access Control & Roles

- CCO waiver for public datasets
- Tiered access:
 - terms of use
 - guestbook
 - restricted data
- Publishing workflow
- Multiple roles:
 - contribute
 - curate, review
 - administrate

Data Features

- Versioning
- Conversion of tabular data files to standard format
- Automatic extraction of file metadata (R, STATA, SPSS, XSLX, FITS)

Interoperability through APIs

Journal Systems (Open Journal System, ScholarOne); Open Science Framework Data Analysis (TwoRavens); Spatial Viz (WorldMap); Preservation systems (Archivematica)

Impact on the Social Science research community and on the World

Anti-Slavery Petitions of Massachusetts

Date range: 1775 - 1825 254 of 3694 rows Timeline playback

| Sigs. | Years | Location | Topic |
|-------|-----------|-----------|---|
| 130 | 1794 | Newbury | Captivity in Algiers |
| 98 | 1798 | Nantucket | Better treatment of men on whaling ships |
| 80 | 1787 | Boston | Support for transportation to Africa |
| 80 | 1811 | Mashpee | Rights, liberties, and resources |
| 77 | 1824 | Boston | Measures for securing the right of suffrage |
| 72 | 1807-1808 | Mashpee | Indian guardians |

Antislavery petitions data

HARVARD ELECTION DATA ARCHIVE
Sharing and Improving Election Data
(email)

HOME PEOPLE MAPS PUBLICATIONS

About the Harvard Election Data Archive

This archive contains data on election results, voting behavior, and electoral politics, with particular focus on the United States. The core data for the archive are state, county and district level election returns for all recent state and federal elections in the United States.

To receive e-mail updates about the status of the archive and new data releases, e-mail your name and institutional affiliation to heda@gov.harvard.edu.

The Harvard Election Data Archive is sponsored by the Sloan Foundation, the Dean of the Faculty of Arts and Sciences at Harvard University, and the Institute for Quantitative Social Science.

Thank you to our colleagues Michael McDonald, Nathaniel Persily, and David Bradlee for their help with data collection.

Redistricting Software Redistricting Resources
[Dave's Redistricting](#) [DrawCongress.org \(Columbia Univ\)](#)

New Data: Alaska
 New Data: Hawaii
 New Data: North Dakota
 New Data: South Dakota
 New Data: Washington
 New Data: Wyoming

Election Data Archive

The
Dataverse
Project

HAVERFORD COLLEGE

Project TIER
Teaching Integrity in Empirical Research

Home

About

Protocol

Fellowships & Workshops

Events

Student Work & Testimonials

People

Partners

Publications

Further Reading

Tweets

Brian Nosek @briannosek
New editorial from Psych Science on improving reproducibility: [https://doi.org/10.1126/science.1257521](#)
53 Retweeted by Project TIER

Project TIER @Project_TIER
Beyond the PPI in Empirical Economic Research | AEA Annual Mtg. 1/5/2016 10:15 am, Hilton Union Square, Franciscan C [goo.gl/vqW49v](#)

Project TIER @Project_TIER
Replication and Transparency Workshop Jan 6-7, 2016 [https://doi.org/10.1126/science.1257521](#) via @uchicago

What's New

- Announcing the 2015-2016 TIER Faculty Fellows
- Project TIER establishes Advisory Board.

Project TIER is supported by

Alfred P. Sloan FOUNDATION

Project TIER

Boston Area Research Initiative

ABOUT | NEWS & EVENTS | PROJECTS | RESOURCES | FELLOWSHIPS | PUBLICATIONS | CONNECT

SEARCH [] GO

BARI's Mission

The Boston Area Research Initiative (BARI) seeks to spur original urban research on the cutting edge of social science and public policy. In conducting and interpreting this research, BARI seeks to forge mutually beneficial relationships among the region's scholars, policymakers, practitioners and civic leaders.

BARI's Strategy

- Facilitate original, policy-relevant research that also addresses scholarly questions
- Create two-way "data pipelines" and tools for analysis and mapping
- Offer scholars and students opportunities to work

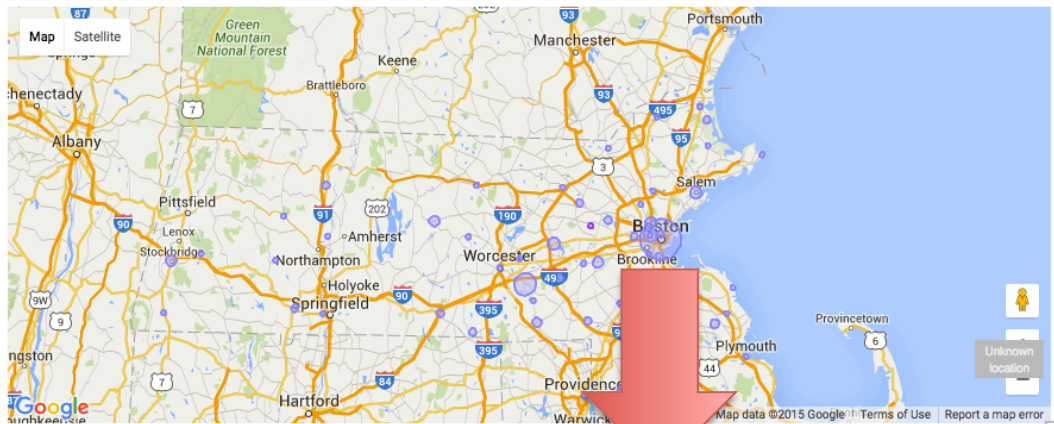
Mapping Calls for Service

The timing and location of a year's worth of service calls to city are shown in a new BARI video. > READ MORE

NEWS & EVENTS

BARI Cosponsors Series on Transforming Boston: From Basket Case to Innovation Hub. Politics, planning, and development in Boston from the end of World War II to the present will be the focus of a four-part series of events

Boston Area Research Initiative



| Sigs. | Years | L |
|-------|-----------|---|
| 130 | 1794 | N |
| 98 | 1798 | N |
| 80 | 1787 | E |
| 80 | 1811 | E |
| 77 | 1824 | E |
| 72 | 1807-1808 | M |
| 63 | 1814 | C |
| 60 | 1782 | E |
| 59 | 1811-1812 | C |
| 57 | 1794 | H |
| 54 | 1815 | C |

Boston 1775-1825 :: petition details

Support for transportation to Africa
 Creation date: 1787-01-04
 Activity dates: 1787-01-04 to 1787-01-04
 Location: Boston
 Total 80 signatures
 Selected signatories: Samuel Stevens, Brister Splanzer, Lanchester Hill, Pring Spear, Prince Rand, Prince Gray, Samuel Beene
 House Unpassed Legislation 1787, Docket 2358, SC1/series 230, Petition of Samuel Stevens
[View details in Dataserve](#) | [View scanned petition](#)

Measures for securing the right of suffrage
 Creation date: (unknown)
 Activity dates: 1824-05-29 to 1824-05-31
 Location: Boston
 Total 77 signatures
 Selected signatories: Jonathan Simonds, Ebenezer Clough, William Barry, Charles Lincoln Jr., Russell Jarvis, William Shed
 Senate Unpassed Legislation 1825, Docket 8102, SC1/series 231, Petition of Jonathan Simonds
[View details in Dataserve](#) | [View scanned petition](#)

Military service
 Creation date: 1782-09-20
 Activity dates: 1782-09-20 to 1782-09-20
 Location: Boston
 Total 36 signatures
 Selected signatories: John Fogg, Daniel Ginnings, Cesar Prescott, Cesar Blake, Solomon Whitney, Peter Keter, David Downing
 Massachusetts Archives Collection. v.237-Revolution Resolves, 1782. SC1/series 45X, Petition of John Fogg
[View details in Dataserve](#) | [View scanned petition](#)

Inspecting 60 of 254 petitions in 1775-1825 out of 3694 petitions total.

Close

Antislavery Petition Data

Daniel Carpenter, Garth Griffin (Harvard University)

3,500 antislavery and antisegregation petitions sent to Massachusetts from 1600s to 1870

Antislavery Petitions Massachusetts Dataserve (Harvard University)

Harvard Dataserve > Antislavery Petitions, Massachusetts Archives, Boston

House Unpassed Legislation 1787, Docket 2358, SC1/series 230, Petition of Samuel Stevens

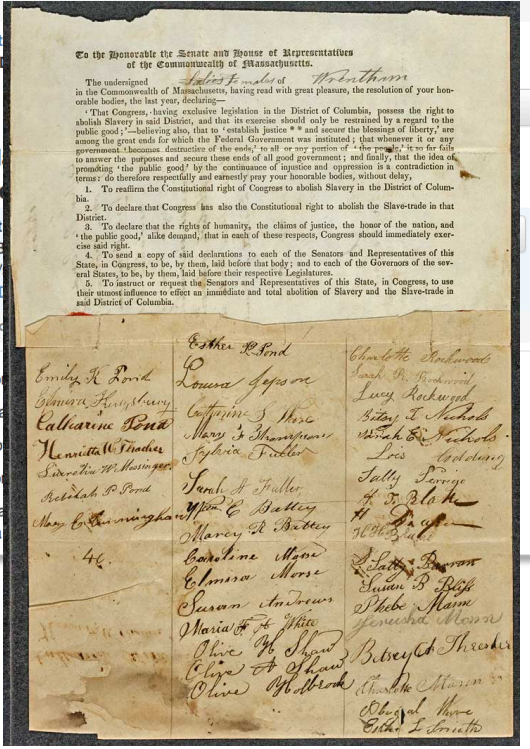
Metrics Shares Coming soon...

House Unpassed Legislation 1787, Docket 2358, SC1/series 230, Petition of Samuel Stevens

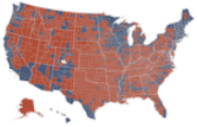
Digital Archive of Massachusetts Archives, Boston, Massachusetts Archives, Boston, Massachusetts Archives, Boston, Massachusetts Archives, Boston

Legislation 1787, Docket 2358, SC1/series 230, Petition of Samuel Stevens", <http://dx.doi.org/10.7910/2012-10-15>

If you use these data, please add this citation:



Authors: Daniel Carpenter, National Endowment for the Humanities, Institutional Development Initiative at Harvard University, and Harvard University Library.



[HOME](#) /

About the Harvard Election Data Archive

This archive contains data on election results, voting behavior, and electoral politics, with particular focus on the United States. The core data for the archive are state, county and district level election returns for all recent state and federal elections in the United States.

To receive e-mail updates about the status of the archive and new data releases, e-mail your name and institutional affiliation to heda@gov.harvard.edu.

The Harvard Election Data Archive is sponsored by the Sloan Foundation, Faculty of Arts and Sciences at Harvard University, and the Institute for Social Science.

Latest Updates

- [Precinct-Level Election Data posted.](#) (1/23/14)

[Redistricting Software](#)

[Redistricting Resources](#)

[Dave's Redistricting](#)

[DrawCongress.org](#) (Columbia Law School)

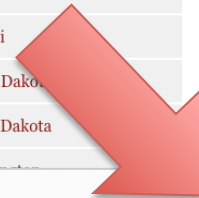
New Data: Alaska

New Data: Hawaii

New Data: North Dakota

New Data: South Dakota

New Data: Washington



Election Data Archive

Steve Ansholabehere (Harvard), Jonathan Rodden (Stanford)

A collaborative archive to share election results, voting behavior, and electoral politics.

The screenshot shows the Dataverse interface for the Harvard Election Data Archive. At the top, there is a navigation bar with links for 'About', 'Guides', 'Support', 'Sign Up', and 'Log In'. Below this is the archive's header with the logo and name. The main content area shows the path 'Harvard Dataverse > Election Data Archive Dataverse > Alaska Data Files'. A 'Metrics' bar indicates 1,505 Downloads. The title 'Alaska Data Files' is followed by a citation: 'Stephen Ansholabehere; Jonathan Rodden, 2011, "Alaska Data Files", http://hdl.handle.net/1902.1/17216, Harvard Dataverse, V1'. A 'Description' section states 'Alaska Shapefile and Election Results'. Below this are tabs for 'Files', 'Metadata', 'Terms', and 'Versions'. A search box is present with the text 'Search this dataset...'. At the bottom, a list of 8 files is shown, including 'AK_08_merge_final.tab' (466 Downloads), 'AK_final.dbf' (173 Downloads), and 'AK_final.sbn' (85 Downloads). Each file entry includes a file icon, name, details, and download options.

Alaska electoral data:
1,500 data downloads

Project TIER

Teaching Integrity in Empirical Research

- Home
- About
- Protocol
- Fellowships & Workshops
- Events
- Student Work & Testimonials
- People
- Partners
- Publications
- Further Reading

PROJECT TIER

Tweets

- Brian Nosek** @BrianNosek
New editorial from Psych Science on improving reproducibility:
ps.sagepub.com/content/early/...
Retweeted by Project TIER
- Project TIER** @Project_TIER
Beyond the PDF in Empirical Economic Research | AEA Annual Mtg.
1/5/2016 10:15 am, Hilton Union Square, Franciscan C goo.gl/wgv6Bv
- Project TIER** @Project_TIER
Replication and Transparency Workshop Jan 6-7, 2016
bitss.org/2015/10/23/rep... via @ucbits

What's New

- Announcing the 2015-2016 TIER Faculty Fellows
- Project TIER establishes Advisory Board.

Questions

We welcome questions and comments on any aspect of our work. Please feel free to email us at TIER@Haverford.edu



Project TIER

Richard Ball, Norm Medeiros (Haverford College)

Teaching empirical research with reproducibility in mind to future scholars

Provides a protocol to document all steps in data management and analysis:

- Data Files
- Metadata Files
- Computing Command Files
- Readme File

The screenshot shows the Dataverse website interface. At the top, there's a navigation bar with 'Dataverse' and search options. Below that, the Project TIER logo is displayed. The main content area features a search bar and a list of results. Two results are visible: 'Colgate Dataverse' (Colgate University) dated Jul 8, 2015, and 'Haverford Dataverse' (Haverford College) dated Jun 17, 2015. On the left side, there are filters for 'Dataverses (3)', 'Datasets (10)', and 'Files (126)'. There are also filters for 'Dataverse Category' (Teaching Course (2), Organization or Institution (1)) and 'Publication Date' (2015 (9), 2013 (3)).

Dataverse About Guides Support Sign Up Log In

Boston Area Research Initiative

Boston Area Research Initiative Dataverse (Boston Area Research Initiative) Home Page

Harvard Dataverse > Boston Area Research Initiative Dataverse > Bicycle Collisions in Boston, MA (2009-2012)

145 Downloads

Bicycle Collisions in Boston, MA (2009-2012)

Lopez, Dahlia; Cheevers, Maria; Stidman, Pete; O'Brien, Daniel, 2014, "Bicycle Collisions in Boston, MA (2009-2012)", <http://dx.doi.org/10.7910/DVN/24713>, Harvard Dataverse, V2

[Download Citation](#)

If you use these data, please add this citation to your scholarly resources. Learn about [Data Citation Standards](#).

Description

Bike collisions in Boston (2009 - 2012). The data are derived from Boston Police Department (BPD) incident narrative reports, as organized and compiled by Dahlia Lopez (Harvard Injury Control Research Center) and partners at BPD, BARI and the Boston Cyclists Union. Visitors can also explore the data through an interactive map hosted by [BostonMap](#).

Files Metadata Terms Versions

Search this dataset... Find

3 Files

Bike Collision Codebook.pdf
Adobe PDF - 884.4 KB - Jun 8, 2014 - 45 Downloads
MDS: 3267f885759464416716f102752cbeb;
[Documentation](#)

Bike Collisions.zip
ZIP Archive - 256.9 KB - Jun 8, 2014 - 42 Downloads
MDS: 4164ea35da5b8816908bc57ba8d72ef6;
[Shapefiles](#)

Boston Area Research Initiative
Daniel O'Brien (Northeastern), Robert Sampson, Christopher Winship (Harvard)

Scholars, policymakers, practitioners and civic leaders collaborating on social science research and public policy

boston.com

HOME NEWS WEATHER SPORTS BUSINESS ENTERTAINMENT LIFESTYLE TRAVEL

Cyclists Beware: This New Map Shows "Bike Crash Hotspots" Across Boston

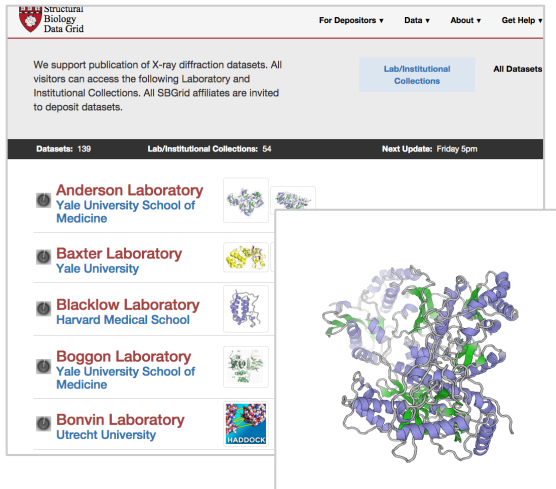
Share 1.2 k Tweet Pin Comment 88 Email

The interactive map plots all of the bicycle-related accidents across Boston.

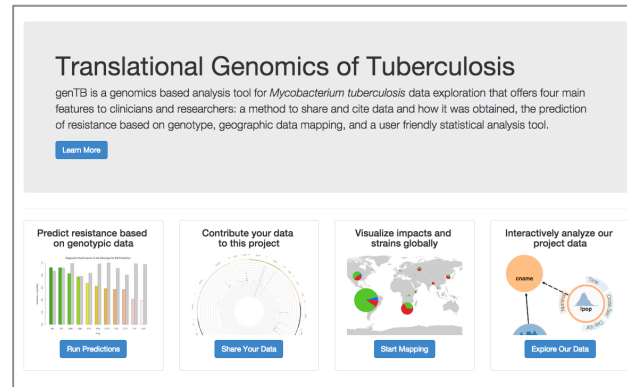
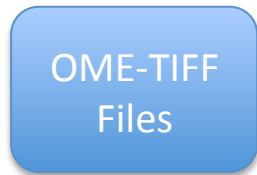
BOSTON AREA RESEARCH INITIATIVE

- Dataset of Bicycle Collisions in Boston (in collaboration with Boston Police, Harvard School of Public Health, and Cyclists Union)
- Data visualization with WorldMap

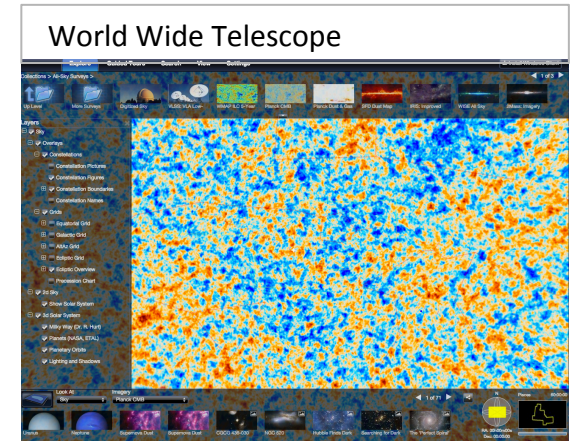
Future impact on other research communities: Biomedical and Astronomy



Structural Biology Data



Tuberculosis Genomics Data

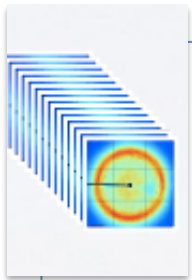


Astronomy Data



- Data archival
- Conversion to standard formats
- Extraction of file-level metadata

Current Collaborations: Addressing the Next Challenges in Data Sharing

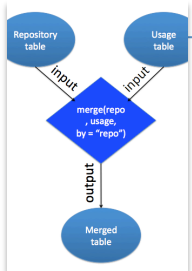


Structural Biology Grid Data Repository (Sliz, HMS, Crosas, IQSS)

THE LEONA M. AND HARRY B.
HELMSLEY
CHARITABLE TRUST



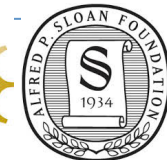
Social Science Big Data (King, Crosas, IQSS, CGA)



Data Provenance (Seltzer, SEAS, Crosas, King, IQSS)



Privacy Tools to share sensitive data (SEAS, Berkman Center, Privacy Lab, IQSS, MIT)

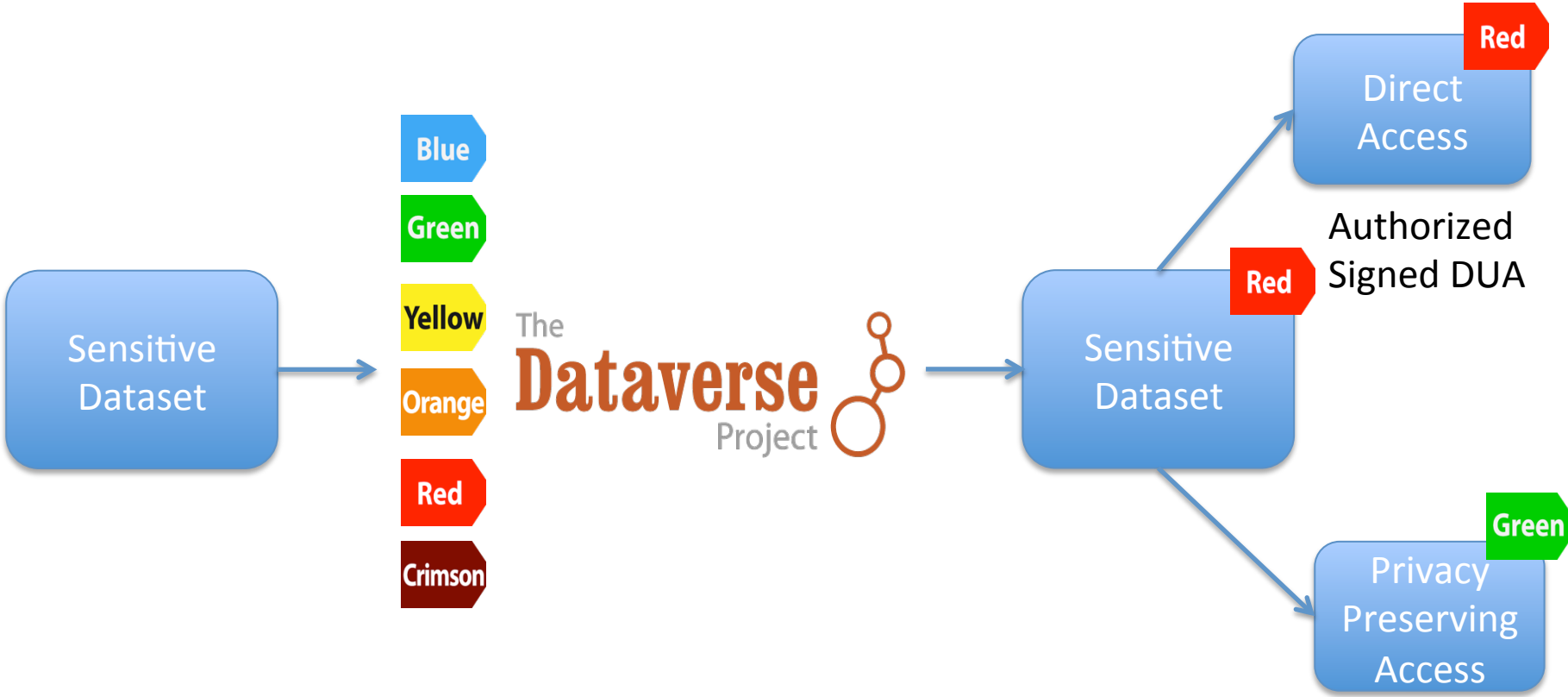


Sharing Sensitive Data with Confidence: DataTags System

| Tag Type | Description | Security Features | Access Credentials |
|----------|-------------------------|--|--|
| Blue | Public | Clear storage, Clear transmit | Open |
| Green | Controlled public | Clear storage, Clear transmit | Email- or OAuth Verified Registration |
| Yellow | Accountable | Clear storage, Encrypted transmit | Password, Registered, Approval, Click-through DUA |
| Orange | More accountable | Encrypted storage, Encrypted transmit | Password, Registered, Approval, Signed DUA |
| Red | Fully accountable | Encrypted storage, Encrypted transmit | Two-factor authentication, Approval, Signed DUA |
| Crimson | Maximally restricted | Multi-encrypted storage, Encrypted transmit | Two-factor authentication, Approval, Signed DUA |

DataTag: A set of security features and access requirements for file handling

Data Sharing Workflow for Sensitive Data



<http://datatags.org>
<http://privacytools.seas.harvard.edu>

@mercecrosas

mcrosas@iq.harvard.edu

<http://scholar.harvard.edu/mercecrosas>

THANKS