Sharing Your Data with Dataverse

Elizabeth Quigley
User Experience Lead
The Dataverse Project (@dataverseorg)
equigley@iq.harvard.edu

What is Dataverse?

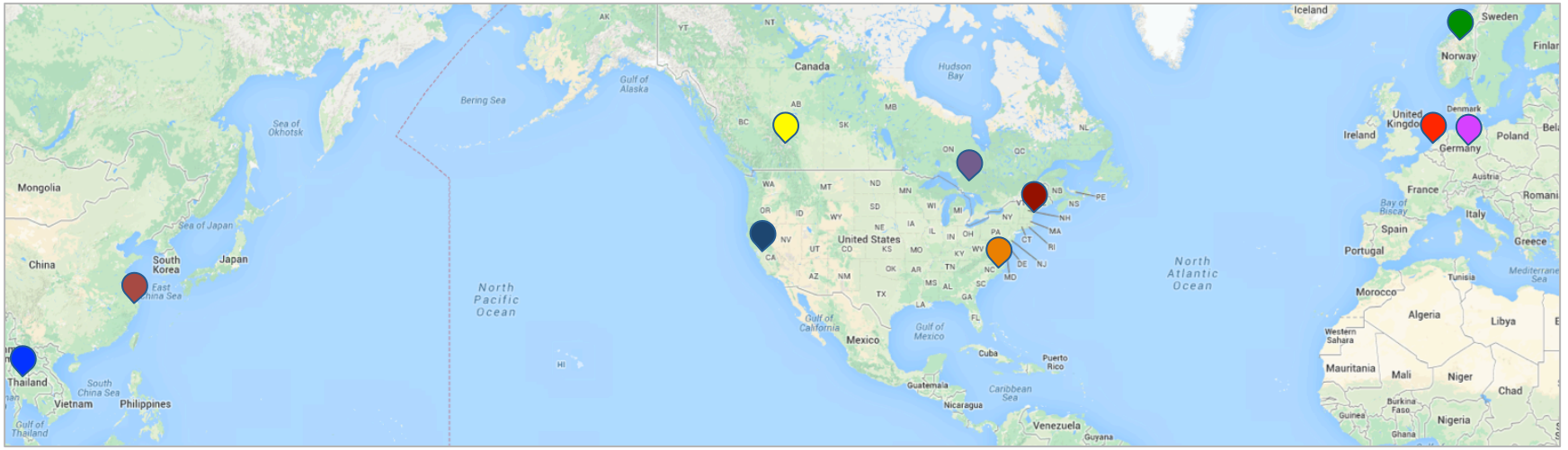
A software framework for publishing, citing and preserving research data (open source on github for others to install)

Developed by the Institute for Quantitative Social Science at Harvard University.

Provides incentives for researchers to share:

- Recognition & credit via data citations
- Control over data & branding
- Fulfill Data Management Plan requirements

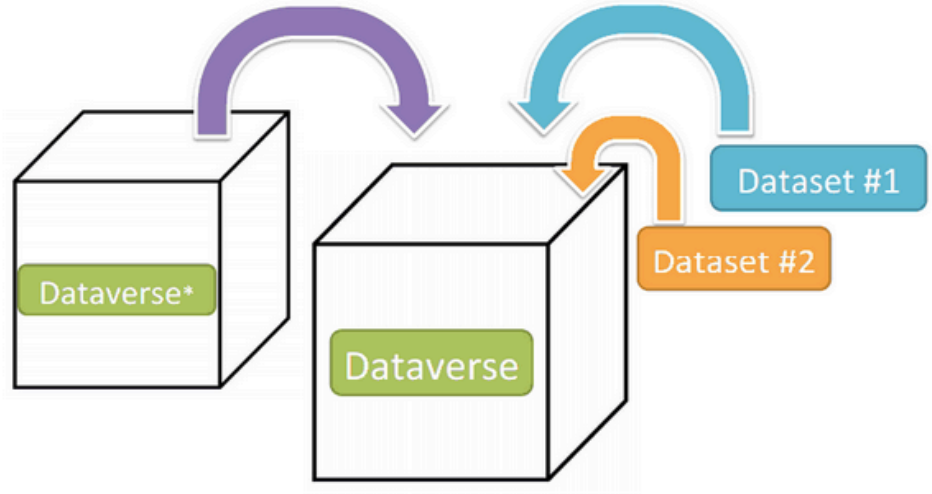
The Dataverse Community



Institutions can setup/host their own Dataverse repository (UNC ODUM, Fudan Univ, Scholars Portal, DANS, etc) and within them can have dataverses for a variety of users (across all research domains): Researchers, Projects, Journals, etc.

What is a Dataverse or Dataset?

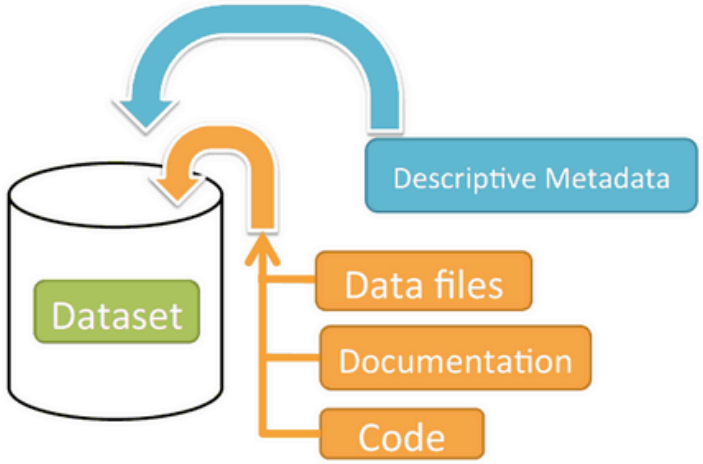
Schematic Diagram of a **Dataverse** in Dataverse 4.0



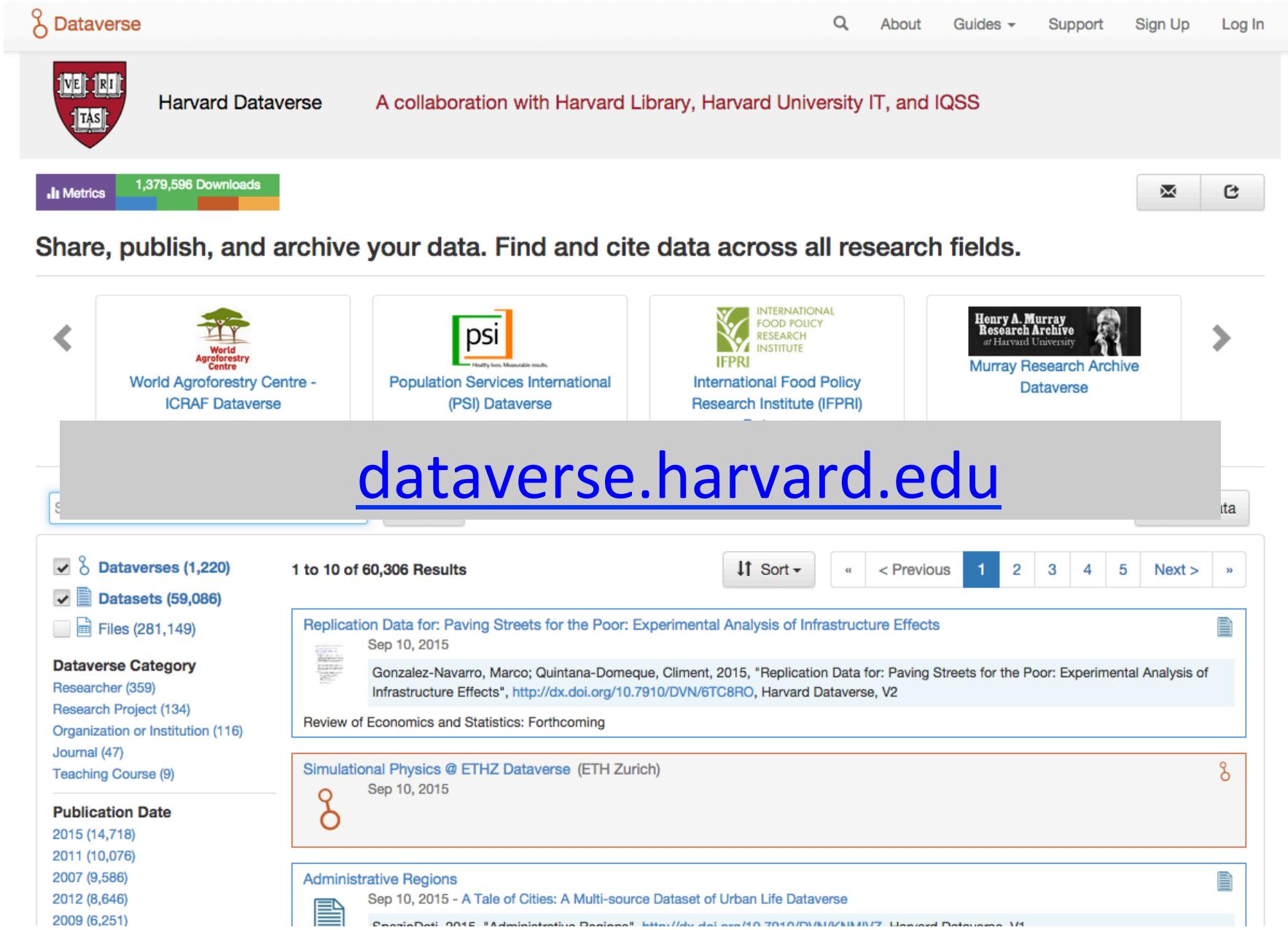
Container for your **Datasets** and/or **Dataverses***

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

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

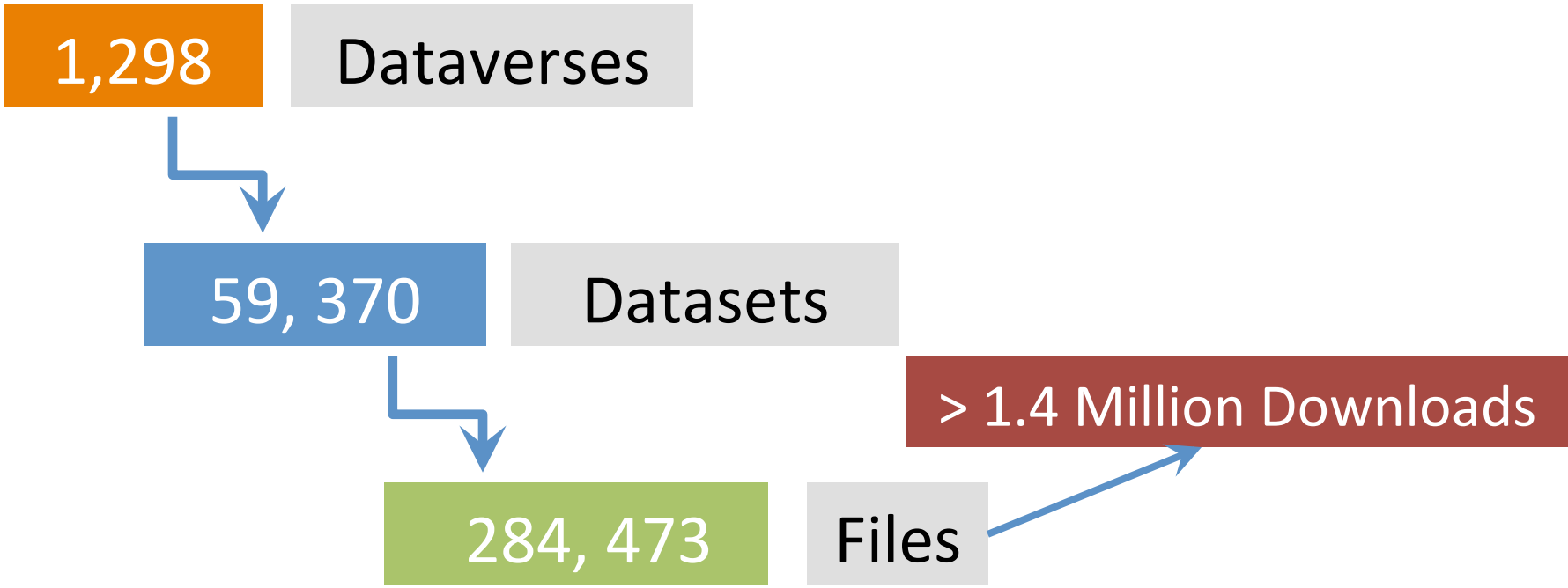


Container for your data, documentation, and code.


Dataverse
About
Guides
Support
Sign Up
Log In
Harvard Dataverse
A collaboration with Harvard Library, Harvard University IT, and IQSS
Metrics
1,379,596 Downloads
Share, publish, and archive your data. Find and cite data across all research fields.
World Agroforestry Centre - ICRAF Dataverse
Population Services International (PSI) Dataverse
International Food Policy Research Institute (IFPRI)
Murray Research Archive Dataverse
dataverse.harvard.edu
Dataverses (1,220)
Datasets (59,086)
Files (281,149)
Dataverse Category
Researcher (359)
Research Project (134)
Organization or Institution (116)
Journal (47)
Teaching Course (9)
Publication Date
2015 (14,718)
2011 (10,076)
2007 (9,586)
2012 (8,646)
2009 (6,251)
1 to 10 of 60,306 Results
Sort
< Previous
1
2
3
4
5
Next >
Replication Data for: Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects
Sep 10, 2015
Gonzalez-Navarro, Marco; Quintana-Domeque, Climent, 2015, "Replication Data for: Paving Streets for the Poor: Experimental Analysis of Infrastructure Effects", http://dx.doi.org/10.7910/DVN/6TC8RO, Harvard Dataverse, V2
Review of Economics and Statistics: Forthcoming
Simulational Physics @ ETHZ Dataverse (ETH Zurich)
Sep 10, 2015
Administrative Regions
Sep 10, 2015 - A Tale of Cities: A Multi-source Dataset of Urban Life Dataverse

Harvard Dataverse

Open to all repository instance at Harvard currently has:



*number from October 25, 2015

DATAVERSE BEST PRACTICES

Dataverse Best Practices (1)

- Standard Metadata Schemas
  - DDI & OAI DC
  - New in 4.0:
    - DataCite 3.1
    - ISA-Tab (biomedical)
    - VO Resource (astronomy)
    - DC Terms
  - Metadata can be exported in JSON & XML

Dataverse Best Practices (2)

- Metadata is always public once a dataset is published

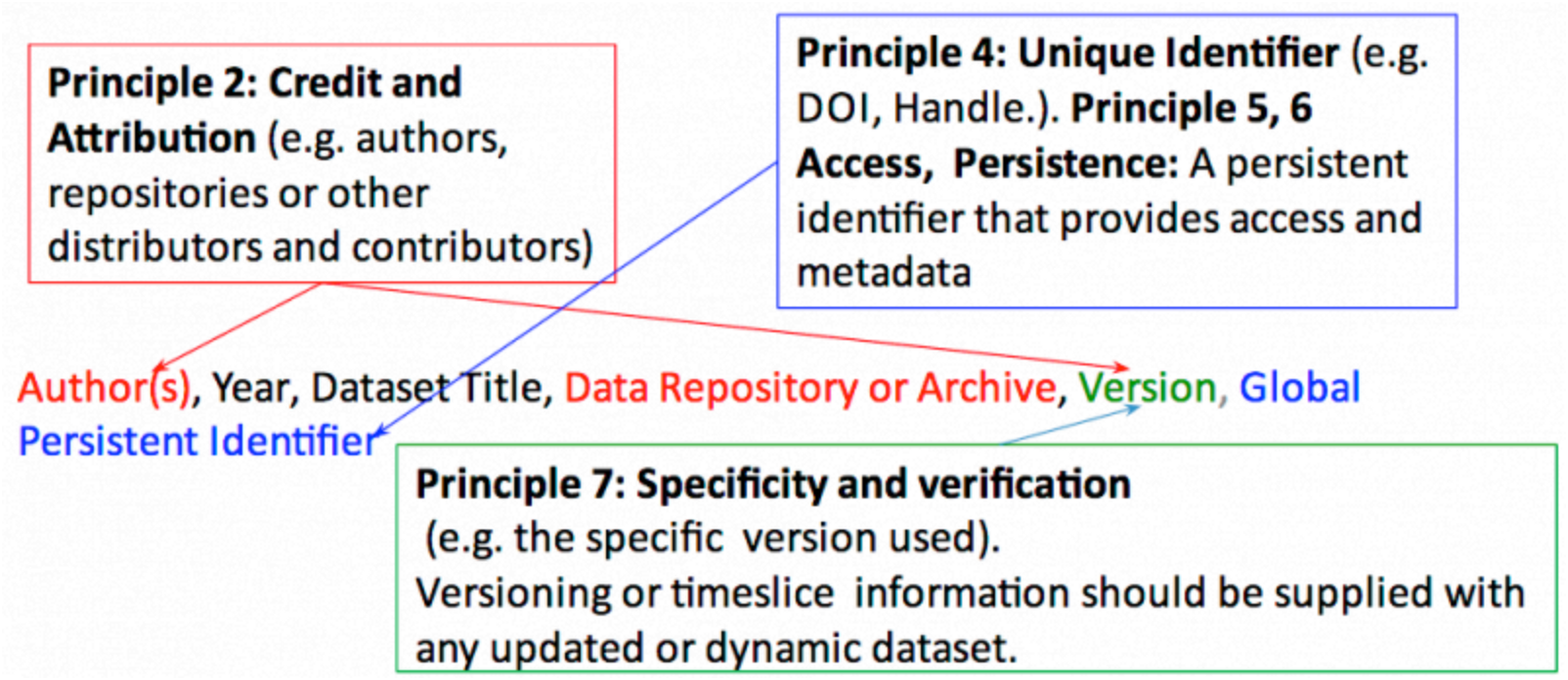
- By default, datasets receive CC0 Waiver 

- Even though default is CCO and we encourage open/public data, when needed, data files in a dataset can be made restricted, or terms of use can be added

Dataverse Best Practices (3)

- Formal Data Citation
  - Originally based off Altman + King 2007
  - Endorse + comply w/ 2014 Joint Declaration of Data Citation Principles (FORCE11)
    - Lead by Merce Crosas, Director of Data Science @ IQSS
  - Versioning and File Fixity
- Persistent IDs: DOI (DataCite/EZID)
  - Resolve to a dataset landing page, not directly to the data files

Data Citation Example



Dataverse Best Practices (4)

- Preservation format conversion for tabular data (extract column/variable metadata)
- File Fixity:
  - UNF (Altman, 2008) for tabular data
  - MD5 checksums for other files

Dataverse Best Practices (5)

- Data-PASS: (ICPSR, ODUM, NARA, ROPER,…)
  - Member of Data-PASS
- OAI-PMH: Harvesting metadata (DC, DDI)
  - From other Dataverse installations
  - From other OAI-DC compliant repositories
- If necessary: Deaccession a Dataset

DISCOVERING + PUBLISHING DATA WITH DATAVERSE

Benefits for Researchers Using Dataverse

- Add your own custom branding and increase your datasets' visibility by embedding them on your website with our widgets.
- Safe and long-term data storage in preservation format.
- Share your data with everyone, or only specific individuals you approve.
- No need to translate data when statistical software formats change.
- Domain specific metadata: making it easy for others to find your data and associated scholarship.
- Allow users to download your data in any format and run many advanced statistical methods online.

HARVARD SHARED INSTITUTIONAL DATA

1 to 5 of 5 Results

Sort

Harvard Faculty Finder

Sep 11, 2015

Waldo, Jim, 2015, "Harvard Faculty Finder", http://dx.doi.org/10.7910/DVN/PLMNRW, Harvard Dataverse, V1

The Harvard Faculty Finder creates creates an institution-wide view of the breadth and depth of Harvard faculty and scholarship, and it helps students, faculty, administrators, and the general public locate Harvard faculty according to research and teaching expertise. More inform...

Library Cloud

Sep 11, 2015

Weinberger, David, 2015, "Library Cloud", http://dx.doi.org/10.7910/DVN/WXAKPE, Harvard Dataverse, V1

Catalog metadata++ about Harvard Library's collection + VIA images + Archival items. Total of about 17M items. Provides an API, which is documented at https://wiki.harvard.edu/confluence/display/LibraryStaffDoc/Library+Cloud . A overview is attached as the data file. Entry point...

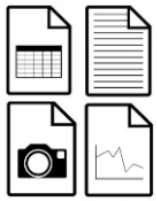
Harvard Library Bibliographic Dataset

Sep 11, 2015

Waldo, Jim, 2015, "Harvard Library Bibliographic Dataset", http://dx.doi.org/10.7910/DVN/LZDQYN, Harvard Dataverse, V1

This dataset contains over 12 million bibliographic records for materials held by the Harvard Library, including books, journals, electronic resources, manuscripts, archival materials, scores, audio, video and other materials. The metadata has been created, acquired and modified...

Rigorous Data Publishing Workflows



Upload

Draft Dataset

Major + Minor Versioning

Publish Version 1

Published Dataset v1

Authors, Title, Year, DOI, Repository, V1

Publish Version 1.1: small metadata change; citation doesn't change.

Published Dataset v1.1

Publish Version 2: File change (automatic); big metadata change (e.g., author, title).

Published Dataset v2

Authors, Title, Year, DOI, Repository, UNF, V2

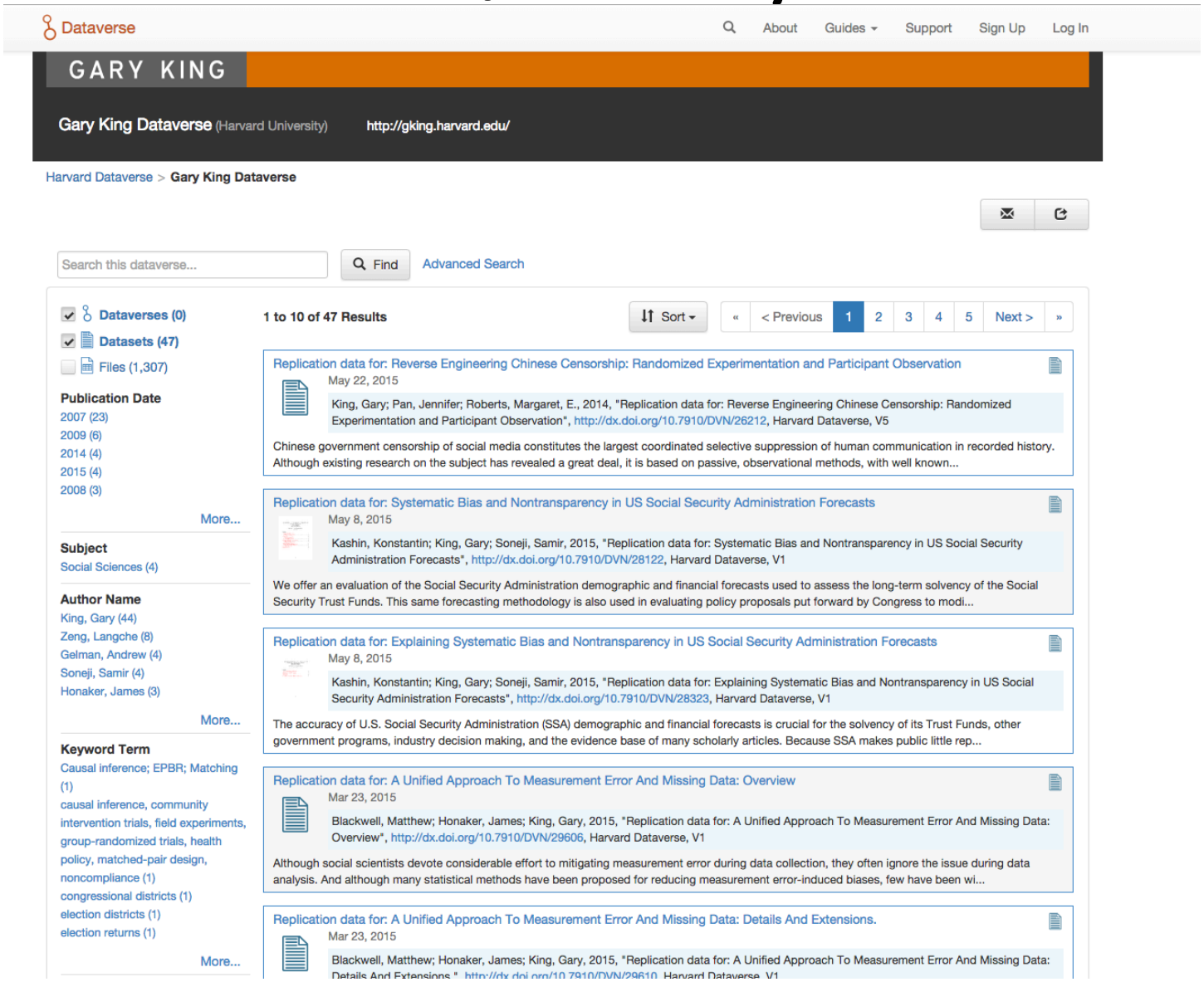
The Dataverse Project

LIVE DEMO TIME!

demo.dataverse.org

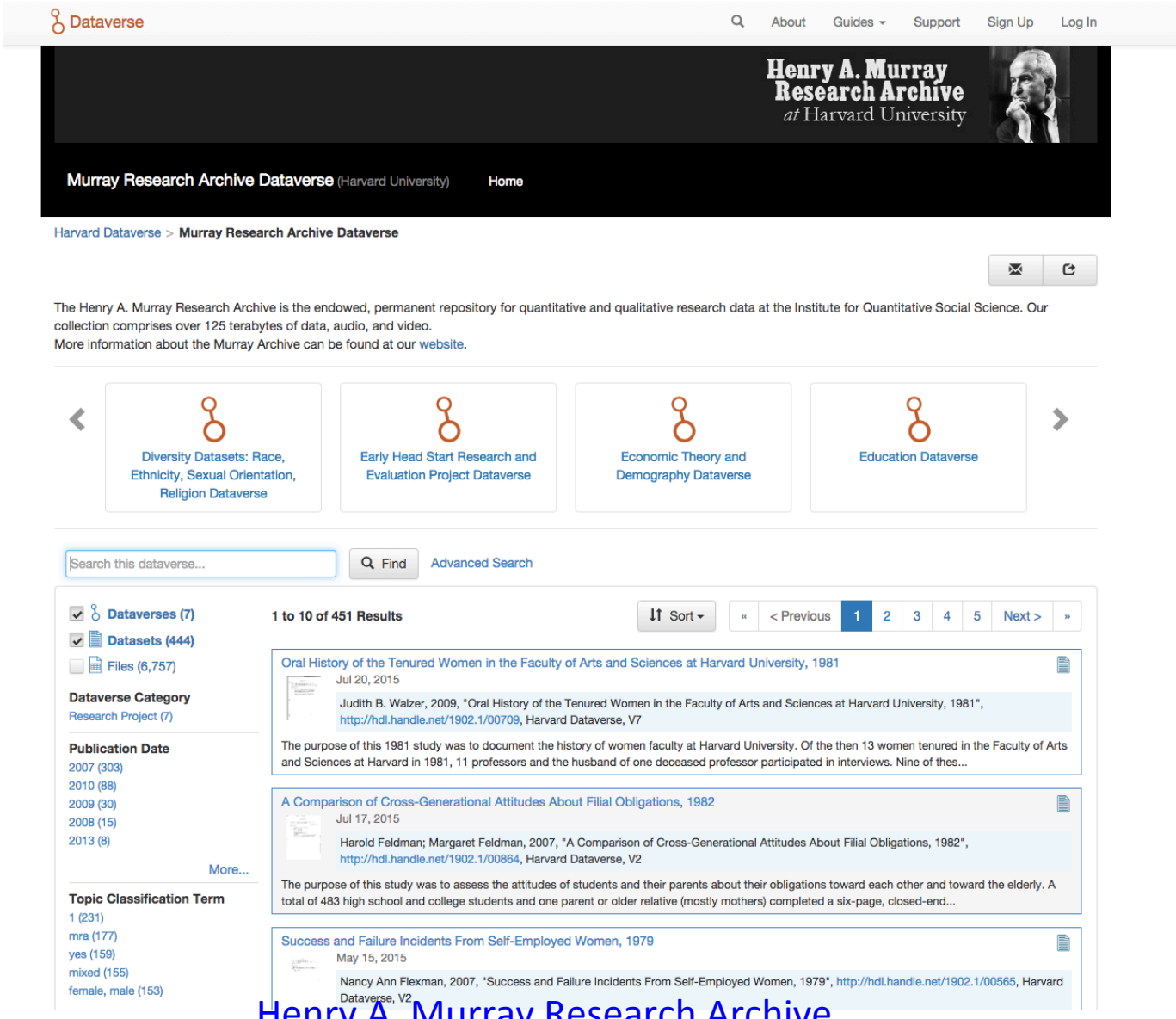
ORGANIZING AN INSTITUTION'S DATA

Researcher/Faculty Member



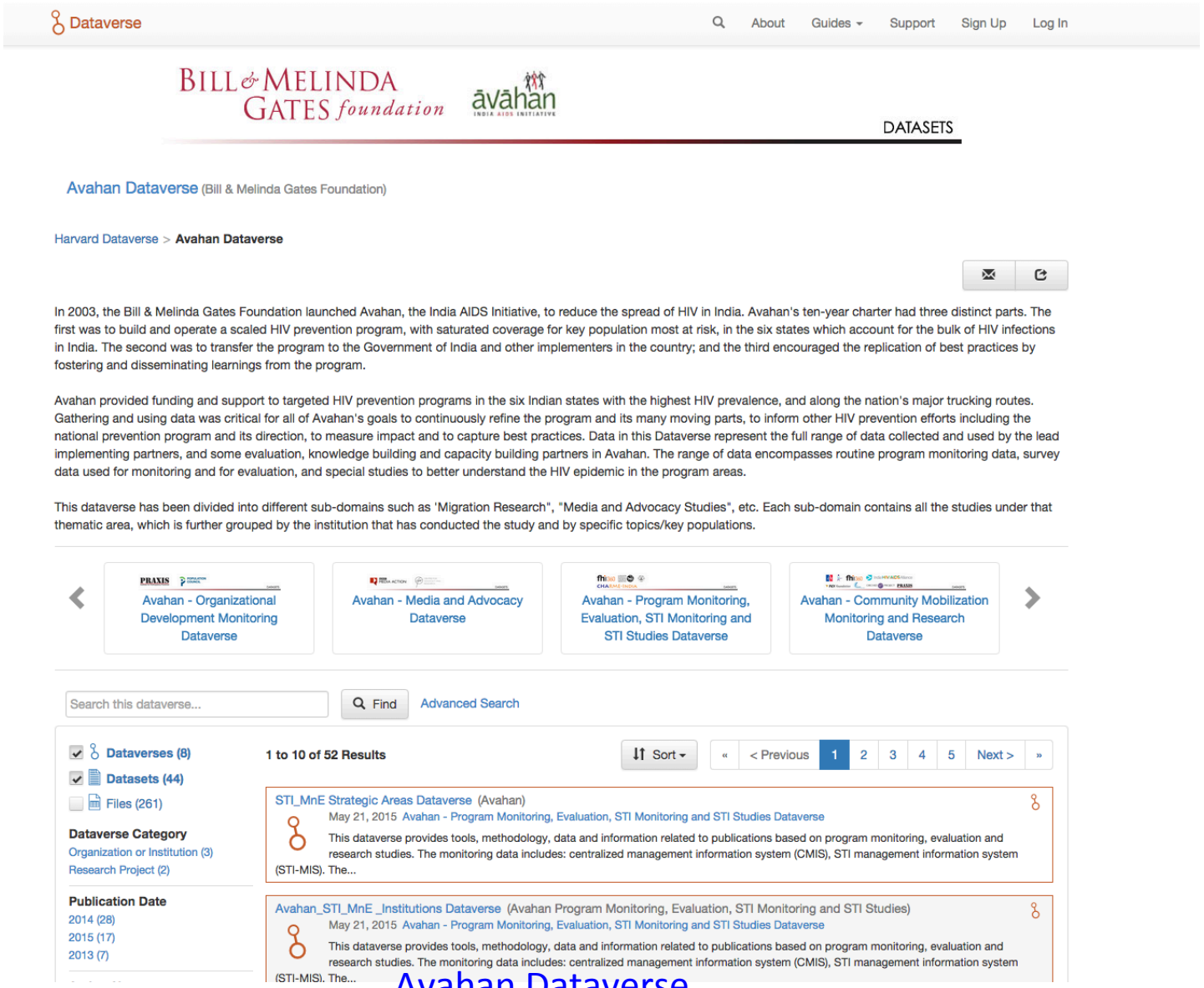
[Gary King Dataverse]

Research Project/Archive



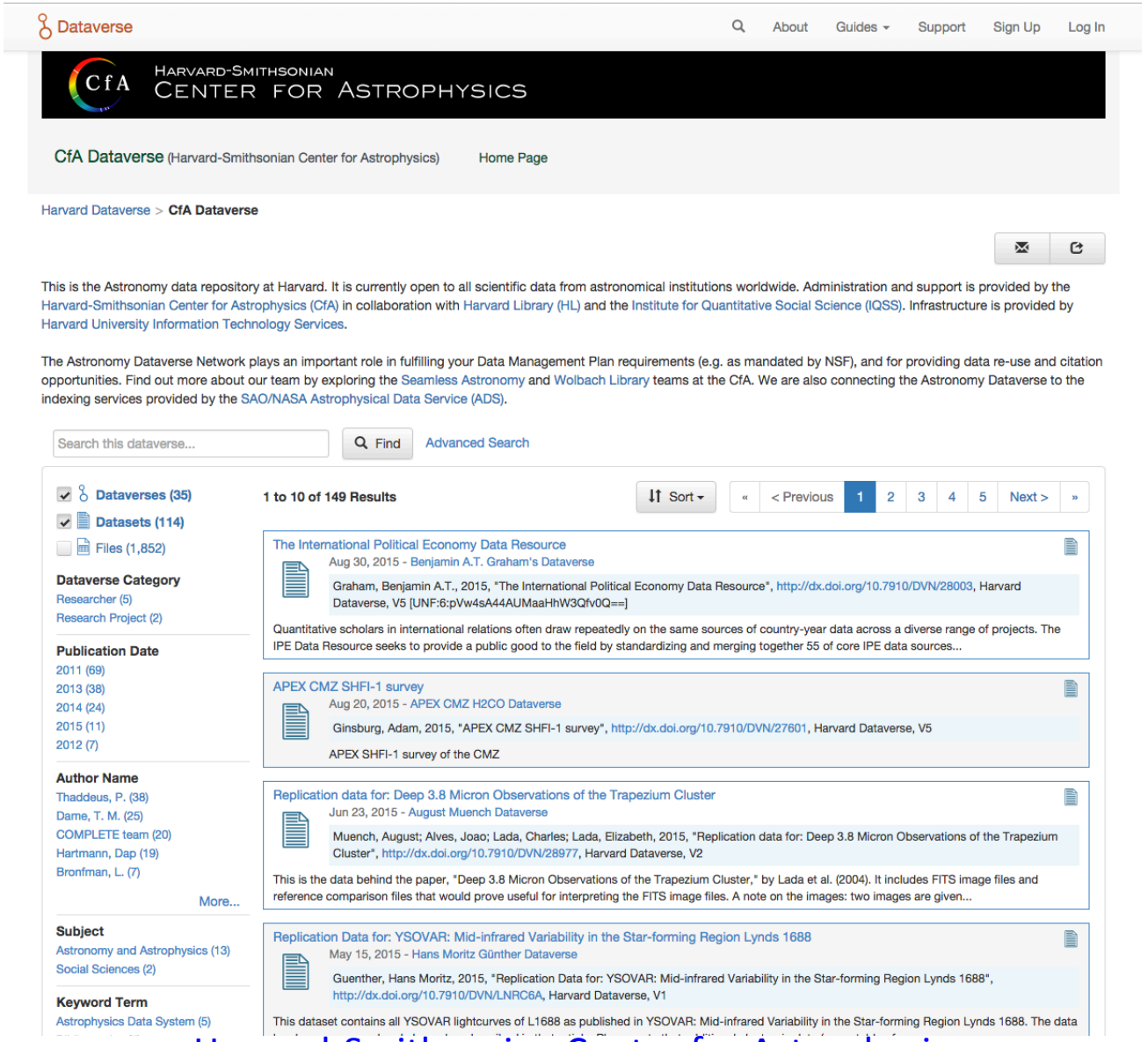
[Henry A. Murray Research Archive]

Organization or Institution



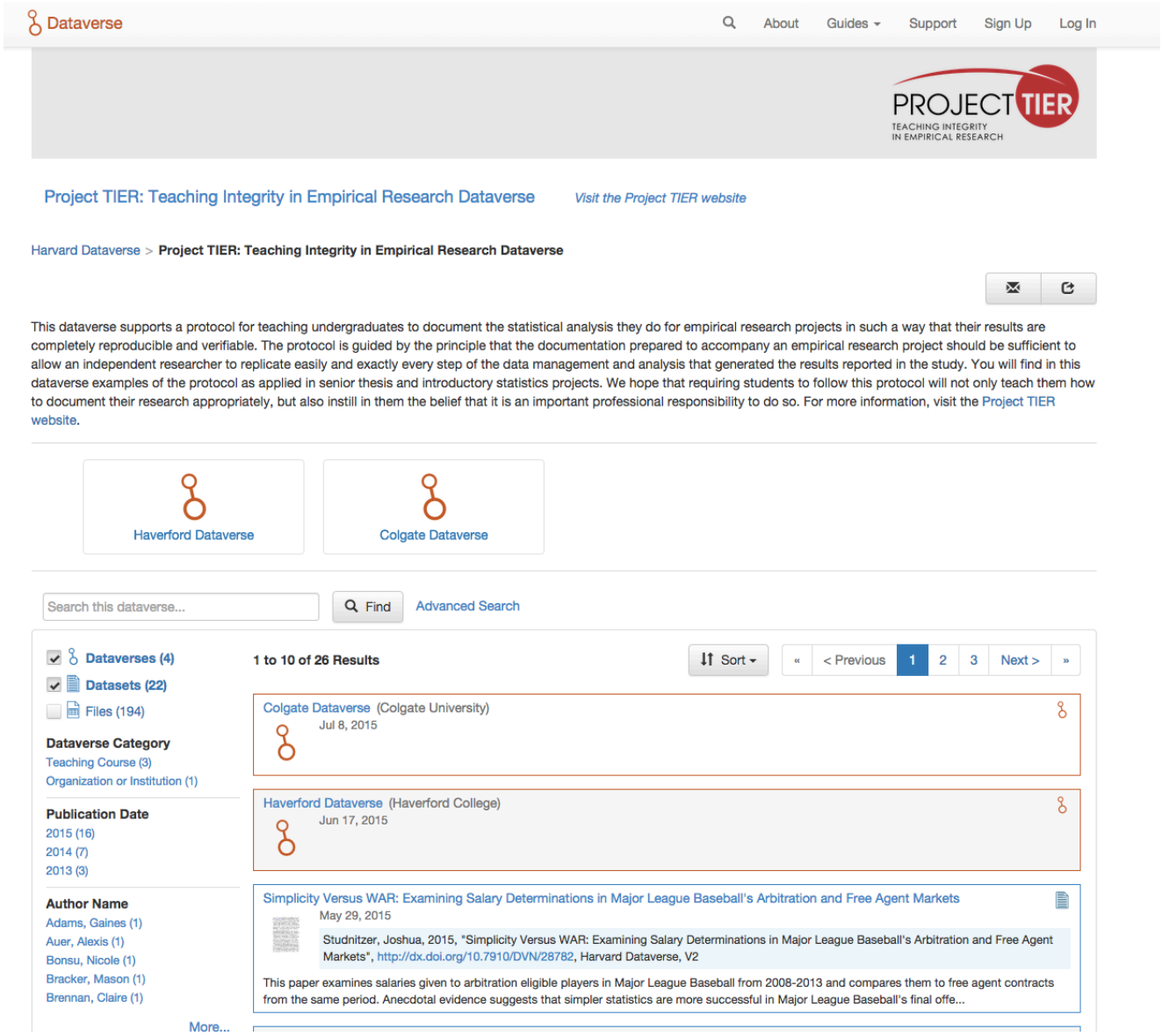
[Avahan Dataverse]

Department



[Harvard-Smithsonian Center for Astrophysics]

Courses



[Project TIER]

Thank you!

Any questions?

Contact: equigley@iq.harvard.edu

Learn more: dataverse.org
Try out Dataverse: demo.dataverse.org

 @dataverseorg

References
Altman M. A Fingerprint Method for Verification of Scientific Data. In A Fingerprint Method for Verification of Scientific Data. Springer-Verlag; 2008.
Altman M, King G. A Proposed Standard for the Scholarly Citation of Quantitative Data. D-Lib Magazine [Internet]. 2007;13(3/4).