

A Data Sharing Story

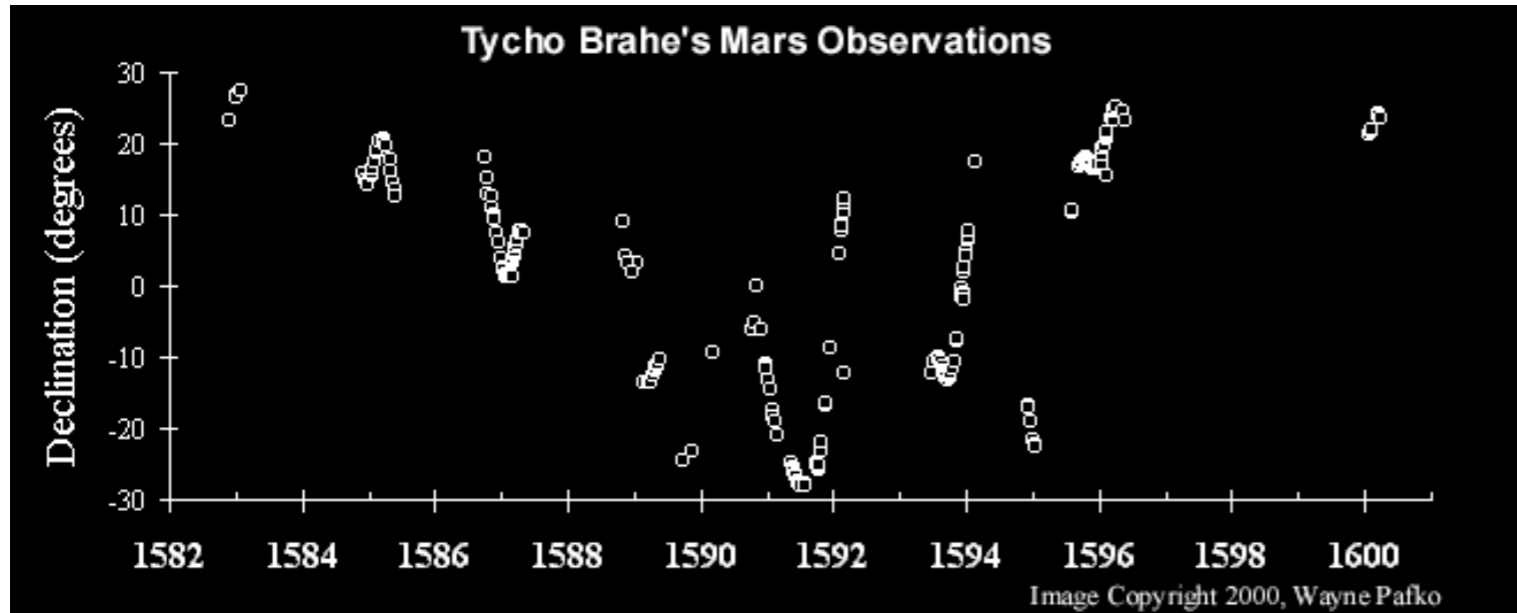
Mercè Crosas, Ph.D.
Director of Product Development
Institute for Quantitative Social Science (IQSS)
Harvard University

How it all
started

Success stories
from social
science

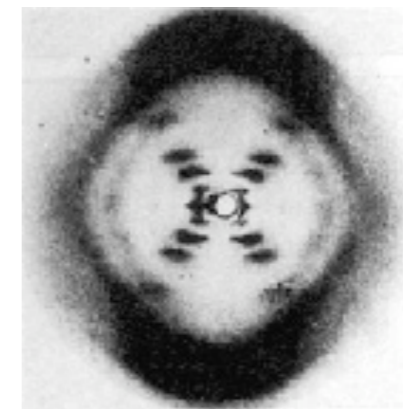
Where we are
heading

Sharing of Data Enables the Maximum Advance of Science



In the Early days of modern science, **Tycho Brahe's data on Mars** allowed Kepler to define the laws of planetary motion.

In the 20th century, **Rosalind Franklin's X-Ray data of DNA** allowed Watson and Crick to discover the double helix model of the structure of DNA.



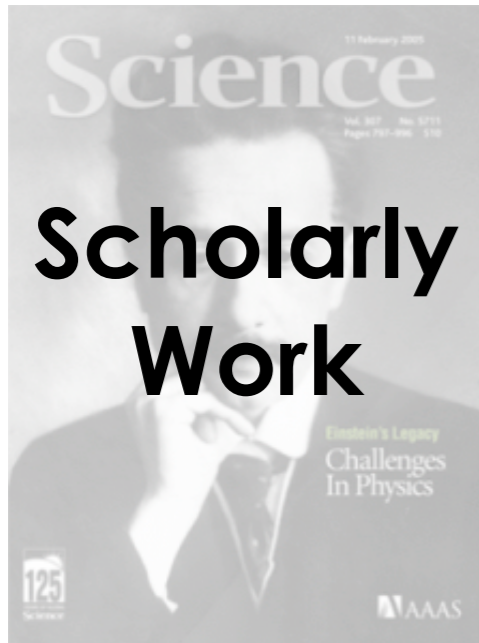
(Unfortunately, in these cases the data were “stolen” rather than “shared” - we can do better!)

Data Sharing in the Digital Age - the Intellectual Beginning

"The replication standard holds that sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author."

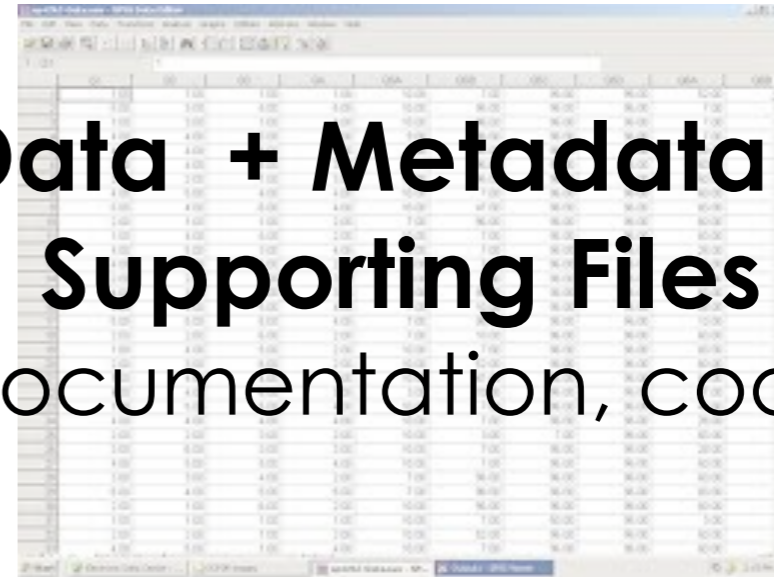
- ▶ King, Gary. 1995 "Replication, Replication"
- ▶ Altman, Micah, King, Gary. 2007 "A Proposed Standard for the Scholarly Citation of Quantitative Data"

A Basic Principle



+

**Data + Metadata +
Supporting Files**
(documentation, code)



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00	6.00
7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00	7.00
8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00	8.00
9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00	10.00
11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00	11.00
12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00	12.00
13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00	13.00
14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00
15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00	15.00
16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00	16.00
17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00	17.00
18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00	18.00
19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00	19.00
20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00	20.00

Formal Data Citation:

Authors, Year, Title, **Persistent Identifier (handle)**, **Universal Numerical Fingerprint (UNF)**, Distributor, Version, [+ Optional Fields]

=

A third party can replicate and reuse,
thus validate and advance science

What You Need to Make it Work

A repository for research data that takes care of **long term preservation and good archival practices**, while the **researcher keeps control of and gets recognition for his data**

Researcher



Centralized
Data Repository

- ✓ Deposits data and enters metadata
- ✓ Gets data citation
- ✓ Displays data on own web site
- ✓ Manages access control
- ✓ Updates new versions

- ✓ Backups and replication of data in different locations (LOCKSS)
- ✓ Re-format for long term accessibility
- ✓ Extract Metadata from data sets
- ✓ Metadata standards (DDI, Dublin Core)
- ✓ Inter-operability (OAI, APIs)

CDD-Hewlett Data Center
A Center Committed To The Provision Of Objective, Verifiable Data For Socio-Economic Development Of Ghana

ABOUT US PEOPLE ANNOUNCEMENTS DATA CALENDAR CONTACT

Data

IQSS Dataverse Network >
Ghana Center for Democratic Development Dataverse

An independent, non-profit, non-government research think tank dedicated to the promotion of democracy, good governance and economic liberalization of Ghana and Africa as a whole.

Search Studies Go Advanced Search Tips

Sort By: Studies: 23

Ghana: Local Public Finance and Decentralization, 1994-2004
by Tewodaj Mogues
Abstract: The purpose of this data compilation effort was to use the

hdi:1902.1/17846
79 downloads + analyses
Last Released: Mar 19, 2012

April 2012

Sidney Verba
Carl H. Pforzheimer University Professor
Department of Government, Harvard University (email)

BIO PUBLICATIONS DATA

Data

IQSS Dataverse Network >
Sidney Verba Dataverse

Search Studies Go Advanced Search Tips

Sort By: Studies: 11

Political Participation in America, 1967
by Verba, Sidney; Nie, Norman H.
Abstract: The focus of this study is individual political participation, both at the local and national levels. Respondents were asked to describe the most important recent changes and problems within their communities, how ...

hdi:1902.2/7015
1 download/analysis
Last Released: Feb 9, 2011

Public Opinion and the War in Vietnam Study, 1966
by Verba, Sidney; Brody, Richard et al.
Abstract: This study, conducted by the National Opinion Research Center,

hdi:1902.2/7295
0 downloads + analyses
Last Released: Feb 9, 2011

James M. Snyder, Jr.
Professor of Government
1737 Cambridge St. Cambridge, MA 02138, 617-496-1089, jsnyder@gov.harvard.edu (email)

Data

IQSS Dataverse Network >
James Snyder Dataverse

Search Studies Go Advanced Search Tips

Sort By: Studies

Replication data for: Electoral Selection with Parties and Primaries
by James M. Snyder, Jr.; Michael M. Ting
114 downloads + analysis
Last Released: Apr 11, 2011

State Legislative Election Returns, 1982-2003
by Carsey, Thomas M.; Berry, William D.; Nanni, Richard G.; Powell, Lynda W.; Snyder, James M.
Abstract: This data collection contains information on state legislative election returns from 1987 through 2003. Each state's returns are organized by individual candidate who ran for state legislative ...

hdi:1902.1/15624
10 downloads
Last Released: Feb 8, 2011

An Empirical Investigation of the Dynamics of PAC Contributors
by Romer, Thomas; Snyder, James M. Jr.
Abstract: These data and/or computer programs are part of ICPSR's Publication-Related Archive and are distributed exactly as they arrived from the data depositor. ICPSR has not checked or processed this material. Users ...

hdi:1902.2/1092
0 downloads
Last Released: Feb 8, 2011

Dataverse Network
(centralized repository)

Metadata Data Files

University of California at San Diego

Christopher J. Fariss's Home Page

IQSS Dataverse Network >
Christopher J. Fariss Dataverse

Please visit my website for more information.

Search Studies Go Advanced Search Tips

Sort By: Studies: 3

Replication data for: The Strategic Substitution of United States Foreign Aid
by Christopher J. Fariss
Abstract: I present a foreign policy decision-making theory that accounts for why US food aid is used strategically when other more powerful economic aid tools are at the disposal of policy makers. I focus my ...

hdi:1902.1/14300
56 downloads + analyses
Last Released: May 2, 2011

Replication data for: The Path to Asylum in the US and the Determinants for Who gets in and Why
by Andy J. Rottman; Christopher J. Fariss; Steven C. Poe
Abstract: The migration of political asylum seekers into the United States has long been a salient political topic; however, social scientists have yet to examine this process in its entirety, and in the context ...

hdi:1902.1/12201
107 downloads + analyses
Last Released: May 2, 2011

Henry A. Murray Research Archive
at Harvard University
HOME POLICIES CONTACT US

IQSS Dataverse Network >
Murray Research Archive Original Collection Dataverse

Search Studies Go Advanced Search Tips

Sort By: Studies: 440

Terman Life Cycle Study of Children with High Ability, 1922-1986
by Louis M. Terman; Robert R. Sears; Lee Cronbach; Pauline S. Sears; Albert Hastorf
Abstract: This study began by comparing a group of children with high intelligence quotients with groups of children typical of the general population, to discover similarities and differences. Research was continued ...

hdi:1902.1/00882
188 downloads + analyses
Last Released: Mar 21, 2012

IQSS Dataverse Network >
NSD Norway Dataverse

NSD is one of the largest archives for research data of its kind and provides data to researchers and students in Norway and abroad. Additionally, NSD is a resource centre, which assists researchers with regard to data gathering, data analysis, and issues of methodology, privacy and research ethics.

Search Studies Go Advanced Search Tips

Sort By: Studies: 1041

School Election 1989, Election Results
by Norsk samfunnsvitenskapelig datatjeneste AS
Abstract: SKOLEVALGSRISULTATER 1989: Den 29. og 30. august 1989 ble det gjennomført prøvevalg ved de videregående skolene i forkant av stortingsvalget. Norsk samfunnsvitenskapelig datatjeneste (NSD) stod som ...

hdi:1902.5/NSD0800-3
0 downloads
Last Released: Feb 27, 2012

School Election 1989, Survey among Pupils
by Norsk samfunnsvitenskapelig datatjeneste AS
Abstract: SKOLEVALGSUNDERSØKELSEN 1989 Fra 1989 er det gjennomført skolevalg ved de videregående skolene i forbindelse med stortings- kommune- og fylkestingsvalg. Den foreliggende datafilen er fra en spørreundersøkelse ...

hdi:1902.5/NSD0800-2
0 downloads
Last Released: Feb 27, 2012

School Election 1989, National Study
by Norsk samfunnsvitenskapelig datatjeneste AS
Abstract: "Skolevalget 1989, landsomfattende meningsmåling" er en nasjonal meningsmåling (telefonintervju) gjennomført gjennomført 28 - 30 august 1989 av Norsk Gallup Institutt AS på oppdrag fra Norsk samfunnsvitenskapelig ...

hdi:1902.5/NSD0800-1
0 downloads
Last Released: Feb 27, 2012

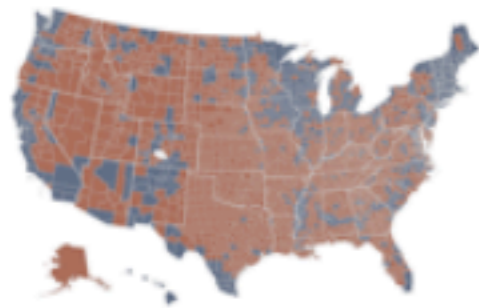
School Election 1991, Election Result
by Norsk samfunnsvitenskapelig datatjeneste AS
Abstract: SKOLEVALGSRISULTATER 1991 Den 28. og 29. august 1991 ble det gjennomført prøvevalg ved de videregående skolene i forkant av kommune/fylkestingsvalget. Datasettet er laget på grunnlag av de resultatene ...

hdi:1902.5/NSD0801-3
0 downloads
Last Released: Feb 27, 2012

How it all
started

Success stories
from social
science

Where we are
heading



HARVARD ELECTION DATA ARCHIVE

Sharing and Improving Election Data



HOME DATA PEOPLE MAPS PUBLICATIONS

A project to share and improve election data:
In less than one year, 43 states, ~ 5000 downloads.

Election Data Archive Dataverse

The Harvard Election Data Archive is a space to share and improve election data. You can create an account and either correct the cataloging information for the studies in this dataverse or upload new data files. Thank you to our colleagues Michael McDonald and David Bradlee for their help with data collection.

The Harvard Election Data Archive by Stephen Ansolabehere and Jonathan Rodden is licensed under a [Creative Commons Attribution 3.0 Unported License](#).

Network™ v. 3.0

[Create Account](#)
[Log In](#)

WIKI DATAVERSE

Create an account to add your own study or contribute to existing studies in this dataverse. Already have an account? [Log In](#)

Election Data Archive

Search Studies

Go

[Advanced Search](#)
[Tips](#)

Sort By:

Studies: 43

Alaska Data Files

by Stephen Ansolabehere; Jonathan Rodden
Abstract: Alaska Shapefile and Election Results

hdl:1902.1/17216

92 downloads + analyses
Last Released: Dec 3, 2011

Wyoming Data Files

by Stephen Ansolabehere; Jonathan Rodden
Abstract: Wyoming Shapefile and Election Results

hdl:1902.1/17215

78 downloads + analyses
Last Released: Dec 3, 2011



James M. Snyder, Jr. Professor of Government

1737 Cambridge St. Cambridge, MA 02138, 617-496-1089, jsnyder@gov.harvard.edu (*email*)



A Data Management Plan for NSF

Bio

Publications

Working Papers

Book Chapters

Dataverse

Classes

IQSS Dataverse Network >

James Snyder Dataverse

POWERED BY THE **Dataverse Network™** PI

James Snyder

[Advanced Search Tips](#)

Sort By:

Studies

[Replication data for: Electoral Selection with Parties and Primaries](#)

by James M. Snyder, Jr.; Michael M. Ting

hdl:1902.1/15624

114 downloads + analyses

Last Released: Apr 11, 2011

[State Legislative Election Returns, 1967-2003](#)

by Carsey, Thomas M.; Berry, William D.; Niemi, Richard G.; Powell, Lynda W.; Snyder, James M.

hdl:1902.2/21480

10 downloads

Last Released: Feb 8, 2011

Abstract: This data collection contains information on state legislative election returns from 1967 through 2003. Each observation in the data refers to an individual candidate who ran for state legislative ...

[An Empirical Investigation of the Dynamics of PAC Contributors](#)

by Romer, Thomas; Snyder, James M. Jr.

hdl:1902.2/1092

0 downloads

Last Released: Feb 8, 2011

Abstract: These data and/or computer programs are part of ICPSR's Publication-Related Archive and are distributed exactly as they arrived from the data depositor. ICPSR has not checked or processed this material. Users ...

Review of Economics and Statistics Dataverse

This page contains replication data and code for the articles published in the Review of Economics and Statistics. Please notice that most articles in volumes previous to volume 92 were not subject to the journal's data availability policy.

OPEN DATAVERSE

Create an account to add your own study to this dataverse. Already have an account? [Log In](#)

- Review of Economics and Statistics
 - Previous Volumes
 - Volume 92
 - Volume 92, Issue 1
 - Volume 92, Issue 2
 - Volume 92, Issue 3
 - Volume 92, Issue 4
 - Volume 93
 - Volume 94

Review of Economics and Statistics

Search Studies

Go


[Advanced Search Tips](#)

within this collection


Sort By: 

Studies: 71    1 2 3 4   

Replication data for: [Can Observers Predict Trustworthiness?](#)
by Michele Belot; V. Bhaskar; Jeroen van de Ven

hdl:1902.1/17866
 4 downloads
Last Released: Mar 20, 2012

Replication data for: [Clash of the Titans: Does Internet Use Reduce Television Viewing?](#)
by Stan J. Liebowitz; Alejandro Zentner

hdl:1902.1/17799
 26 downloads + analyses
Last Released: Mar 19, 2012

Replication data for: [Strategic Citation](#)

hdl:1902.1/17797

Replication Data: when authors submit a paper, they deposit the data in the journal's Dataverse

How it all
started

Success stories
from social
science

Where we are
heading

Reinforcing the Link between Scholarly Work and Research Data

Alfred Sloan Foundation grant

The screenshot shows the homepage of the International Journal of Economics and Finance (IJEF). The header includes the journal's logo and title. Below the header is a navigation menu with links for Home, About, Log In, Register, Search, Current, Archives, Announcements, Thesis Abstracts, Recruitment, Submissions, and Editorial Team. A user login section is visible on the right. The main content area features a search bar, a 'Browse' section with links for 'By Issue', 'By Author', 'By Title', and 'Other Journals', and a 'Current Issue' section with RSS feeds for 3.0, 2.0, and 1.0 versions. A large image of the journal cover is displayed at the bottom.

Open Journal System

→
Seamless
integration
between the
two systems

The screenshot shows the Chris Anderson Dataverse page. The header includes the 'IQSS Dataverse Network' logo and the 'Dataverse Network PROJECT v. 3.0' logo. The main content area displays a list of studies by Chris Anderson. Each study entry includes the title, author, abstract, and download/analysis statistics. The studies listed are: 'Replication data for: Norway: Monthly party support, 1964-92', 'Replication data for: Netherlands: Monthly party support, 1978-92', 'Replication data for: Germany: Monthly party support, 1960-89', 'Replication data for: France: Monthly Prime Minister and Presidential Popularity, 1960-92', and 'Replication data for: EU: Public Support for European Integration, 1973-93'. A search bar and 'Advanced Search Tips' link are also visible.

The Dataverse Network

Working with the Harvard Library to offer Dataverse as a Data Management Plan

- ▶ Provide services on research data management:
 - ▶ Data preservation and curation
 - ▶ Using the Dataverse
- ▶ Centralized hosting for Dataverse Networks in multiple domains



The AstroData Project

Astronomy Dataverse Network

POWERED BY THE **Dataverse Network**™ PROJECT v. 2.2.5

Search [User Guide](#) [Feedback](#) [Log In](#) [Create Account](#)

This is the Astronomy Dataverse Network, a Harvard-affiliated project of the [Astrophysics Astronomy](#) group at the Harvard-Smithsonian Center for Astrophysics, the [ADS](#), the [Wolbach Library](#), and [IQSS](#) with support from the FAS Science Research Computing.

Started as a pilot project at the Harvard-Smithsonian Center for Astrophysics, and it was launched on Monday!

Show All Dataverses

Observational

Panchromatic

All Dataverses

Search Studies

Go

Advanced Search

Tips

ALL A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Does not substitute observatory-specific repositories

Encourages Dataverse expansion:

Support Metadata for multiple domains (v 3.0)

Deposit larger data files

Explore and visualize large Images

Extract metadata from FITS files

Name	Affiliation	Released	Activity
Astrobase.org	harvard.edu	Apr 2, 2012	■ □ □ □ □
Soderberg, Alicia	Harvard University	Feb 6, 2012	■ □ □ □ □
Astroinformatics: galaxies & quasars	Harvard-Smithsonian Center for Astrophysics	Oct 12, 2011	■ □ □ □ □
MPEFT	Harvard-Smithsonian Center for Astrophysics	Jun 23, 2011	■ ■ ■ ■ ■
1.2 Meter CO Survey	Smithsonian Astrophysical Observatory	May 23, 2011	■ ■ ■ □ □



Thanks

The Dataverse Network Project:
thedata.org

Social Science: dvn.iq.harvard.edu

Astronomy: theastrodata.org

mcrosas@iq.harvard.edu