SUPPORTING DATA ACCESS and MANAGEMENT with THE DATAVERSE

Mercè Crosas, Ph.D.
Director of Product Development
Institute for Quantitive Social Science, Harvard University
mcrosas@iq.harvard.edu

The Institute *for* Quantitative Social Science at Harvard University

The Dataverse Network® Project

HARVARD UNIVERSITY

WHY

did we build the Dataverse Network?

A PROBLEM

Different needs for Archives, Data Libraries, Researchers, Journals, ...

We need to preserve the data forever

I want to curate the data

We need persistent links

I want recognition for my data

I want to review submissions and control the access

A SOLUTION

Centralized Professional Archiving with **Distributed** Control and Recognition

- Persistent identifiers
- Fixity
- Backups and recovery
- Metadata standards
- Conversion standards
- Preservation standards

+

- Branding and visibility
- Data discovery
- Ease of use
- Scholarly citation
- Control over updates
- Terms of access and use

HOW

does it work?

DATAVERSE NETWORK

A **centralized** software installation and data repository

Dataverse

Dataverse

Dataverse

Dataverse

Individual **distributed** data archive with its own branding

Collections

Study

Study

Study

Metadata

Data Files

RICH DATA SUPPORT

Data Files

Tabular Data:
STATA
SPSS
CSV + control card
Tab delimited + DDI

Social Network Data:
GraphML

+ complementary files
in **any** format

- Convert to preservation format (primary data + variable metadata)
- Calculate Universal Numerical Fingerprint (UNF)
- Define Terms of Use and Permissions
- Download in multiple formats
- Download a subset of the data
- Generate summary statistics
- Apply Zelig (R) statistical methods
- Visualize time series

FORMAL, PERSISTENT DATA CITATIONS

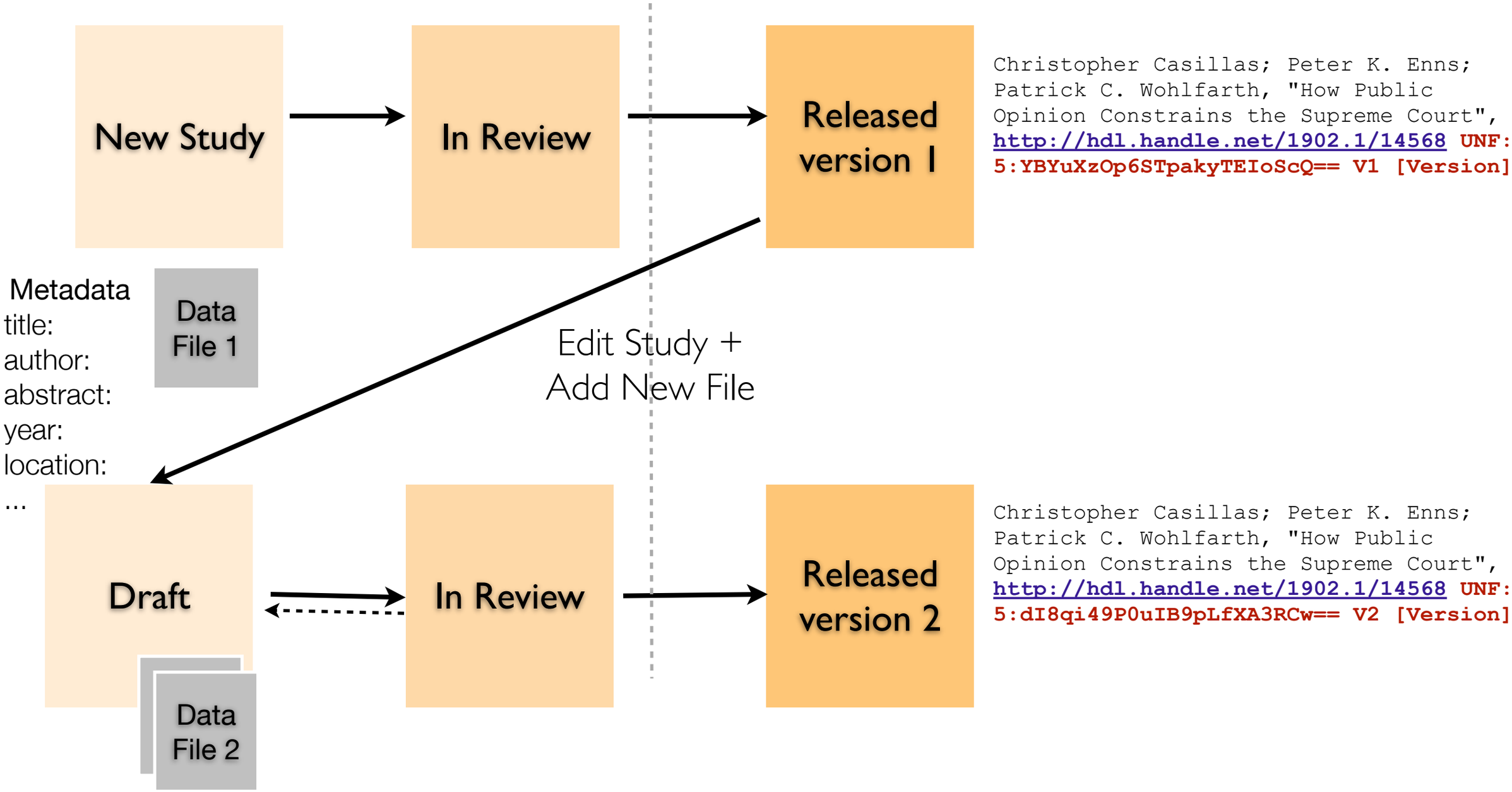
Authors
Year
Title
Persistent URL and ID
UNF
Distributor
Version
+ other optional fields (data subset)

```
Sidney Verba; Joseph DiNunzio; Christina Spaulding, 1985, "Harvard
Sexual Harassment Survey, 1983", http://hdl.handle.net/1902.1/00686
UNF:3:eG6QJsi2VsgIMV49FtGcDA== Murray Research Archive
[Distributor] V1 [Version]
```

DATA VERSIONING AND MANAGEMENT

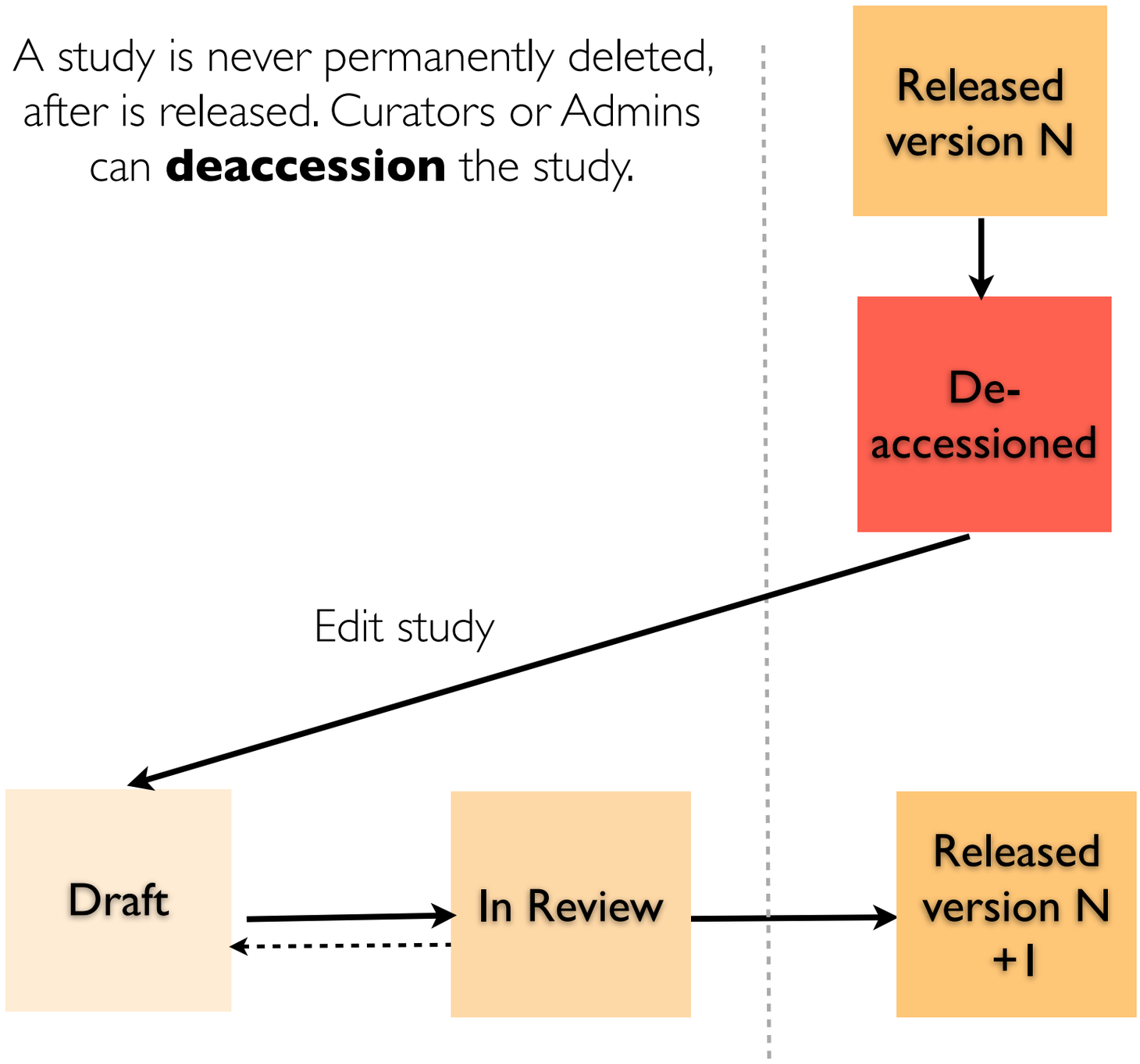
Contributors, curators, admins view

End user view



DATA NEVER DELETED

A study is never permanently deleted, after is released. Curators or Admins can **deaccession** the study.



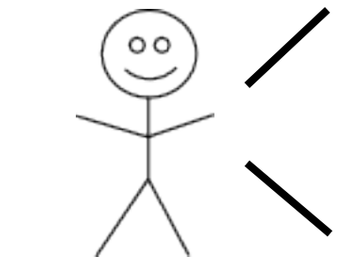
Christopher Casillas; Peter K. Enns; Patrick C. Wohlfarth, "How Public Opinion Constrains the Supreme Court", http://hdl.handle.net/1902.1/14568 UNF: 5:dI8qi49P0uIB9pLfXA3RCw== VN [Version]

http://hdl.handle.net/1902.1/14568

This study is deaccessioned.
[Go to this other study]

WORKFLOWS FOR ALL

Regular Dataverse

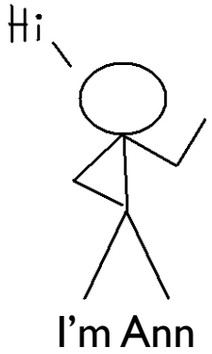


Option A - The **admin or curator** adds Mike as a contributor. Mike can now add studies and edit his **own** studies in this Dataverse.

or

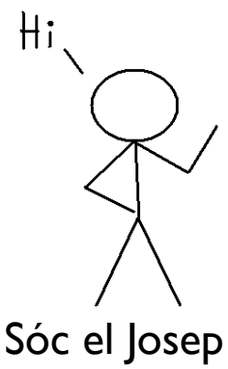
Option B - The **admin or curator** adds Mike as a contributor. Mike can add studies and edit **any** study in this Dataverse.

Open Dataverse



Ann **creates an account** and starts adding and editing her **own** studies in an open Dataverse. The admin or curator reviews the studies before releasing them.

Wiki Dataverse



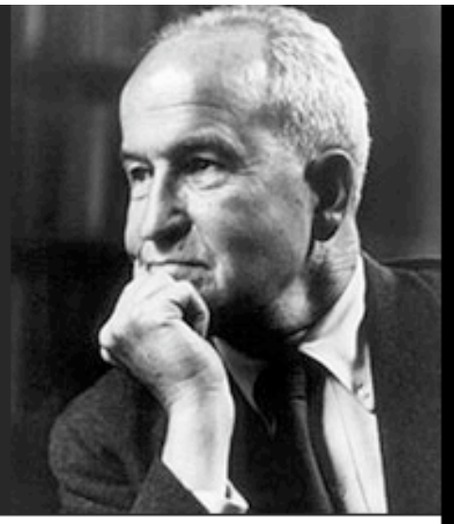
Josep **creates an account** and starts adding new studies and editing **any** study in a wiki Dataverse. The admin or curator reviews the studies before releasing them.

REGULAR DATAVERSE

Mercè Crosas, LIBER June 2011

Henry A. Murray Research Archive at Harvard University

HOME POLICIES CONTACT US



All IQSS Dataverses >

Murray Research Archive Original Collection Dataverse

POWERED BY THE Dataverse Network™ PROJECT

Search User Guides Report Issue Log In Create Account

The Henry A. Murray Research Archive is the endowed, permanent repository for quantitative and qualitative research data at the Institute for Quantitative Social Science, and provides physical storage for the entire IQSS Dataverse Network. Our collection comprises over 125 terabytes of data, audio, and video.
More information about the Murray Archive can be found at our Website

- Henry A. Murray Research Archive Ori
  - Diversity samples: Race, Ethnicity,
  - Economic Theory and Demography
  - Education
  - Family. Marriage. Women
  - Health
  - Politics and Government
  - Replications, Extensions and Follow
  - Work

Henry A. Murray Research Archive Original Collection

Search Studies Go Advanced Search Tips

within this collection

Sort By: Studies: 501 1 2 3 4

Attitudes of Participants at the Iowa Women's Political Caucus, 1973

by Barbara Burrell

Abstract: This study was designed to obtain a profile of women who are active in a movement generating social change through advocating political equality. The sample consists of participants at the first state ...

hdl:1902.1/00508
112 downloads
Last Released: Jun 21, 2011

Survey of Child-Care Needs of Lower-Paid Personnel, 1980

by Harvard-Radcliffe Child Care Council

Abstract: The purpose of this study was to identify unmet needs for child care among lower-paid personnel employed at Harvard University. This identification was aimed both at estimating the numbers of lower-paid ...

hdl:1902.1/00566
31 downloads + analyses
Last Released: Jun 21, 2011

Early Head Start Research and Evaluation Project, 1996 - 2001

by Administration for Children and Families

hdl:1902.1/00097
2436 downloads + analyses

OPEN DATAVERSE

Mercè Crosas, LIBER June 2011

The Review of Economics and Statistics

All IQSS Dataverses >

Review of Economics and Statistics Dataverse

POWERED BY THE Dataverse Network™ PROJECT

Search User Guides Report Issue Log In Create Account

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 93, Issue 1
    - Volume 93, Issue 2
    - Volume 93, Issue 3
    - Volume 93, Issue 4

Search Studies Go Advanced Search Tips

within this collection

Review of Economics and Statistics

Sort By: Studies: 33 1 2 3 4

Replication data for: New Eurocoin: Tracking Economic Growth in Real Time
by Filippo Altissimo; Riccardo Cristadoro; Mario Forni; Marco Lippi; Giovanni Veronese
hdl:1902.1/15447
27 downloads
Last Released: Apr 19, 2011

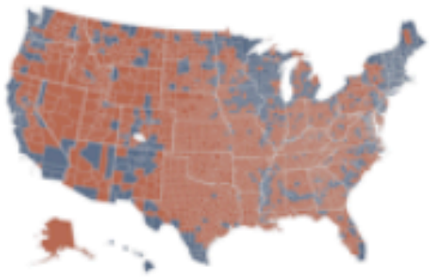
Replication data for: Inequality and Redistribution: Evidence from U.S. Counties and States, 1890-1930
by Rodney Ramcharan
hdl:1902.1/15784
236 downloads
Last Released: Apr 19, 2011

Replication data for: Corruption and Bilateral Trade Flows: Extortion or Evasion?
by Pushan Dutt; Daniel Traca
hdl:1902.1/15394
28 downloads + analyses
Last Released: Mar 31, 2011

Replication data for: Universal Vouchers and Racial and Ethnic Segregation
by Eric Brunner; Jennifer Imazeki; Stephen Ross
hdl:1902.1/15356
36 downloads + analyses
Last Released: Mar 31, 2011

WIKI DATAVERSE

Mercè Crosas, LIBER June 2011



HARVARD ELECTION DATA ARCHIVE

Sharing and Improving Election Data

DATA PEOPLE HOME

Data

All IQSS Dataverses >

Election Data Archive Dataverse

POWERED BY THE Dataverse Network™ PROJECT

Search User Guides Report Issue Log In Create Account

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.

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: 20 1 2

New Jersey Data Files

by Stephen Ansolabehere; Jonathan Rodden

Abstract: New Jersey Shapefile and Election Results

hdl:1902.1/16222

0 downloads + analyses

Last Released: Jun 22, 2011

New Hampshire Data Files

by Stephen Ansolabehere; Jonathan Rodden

Abstract: New Hampshire Shapefile and Election Results

hdl:1902.1/16219

2 downloads + analyses

Last Released: Jun 21, 2011

DATAVERSE

BEYOND

SOCIAL SCIENCE

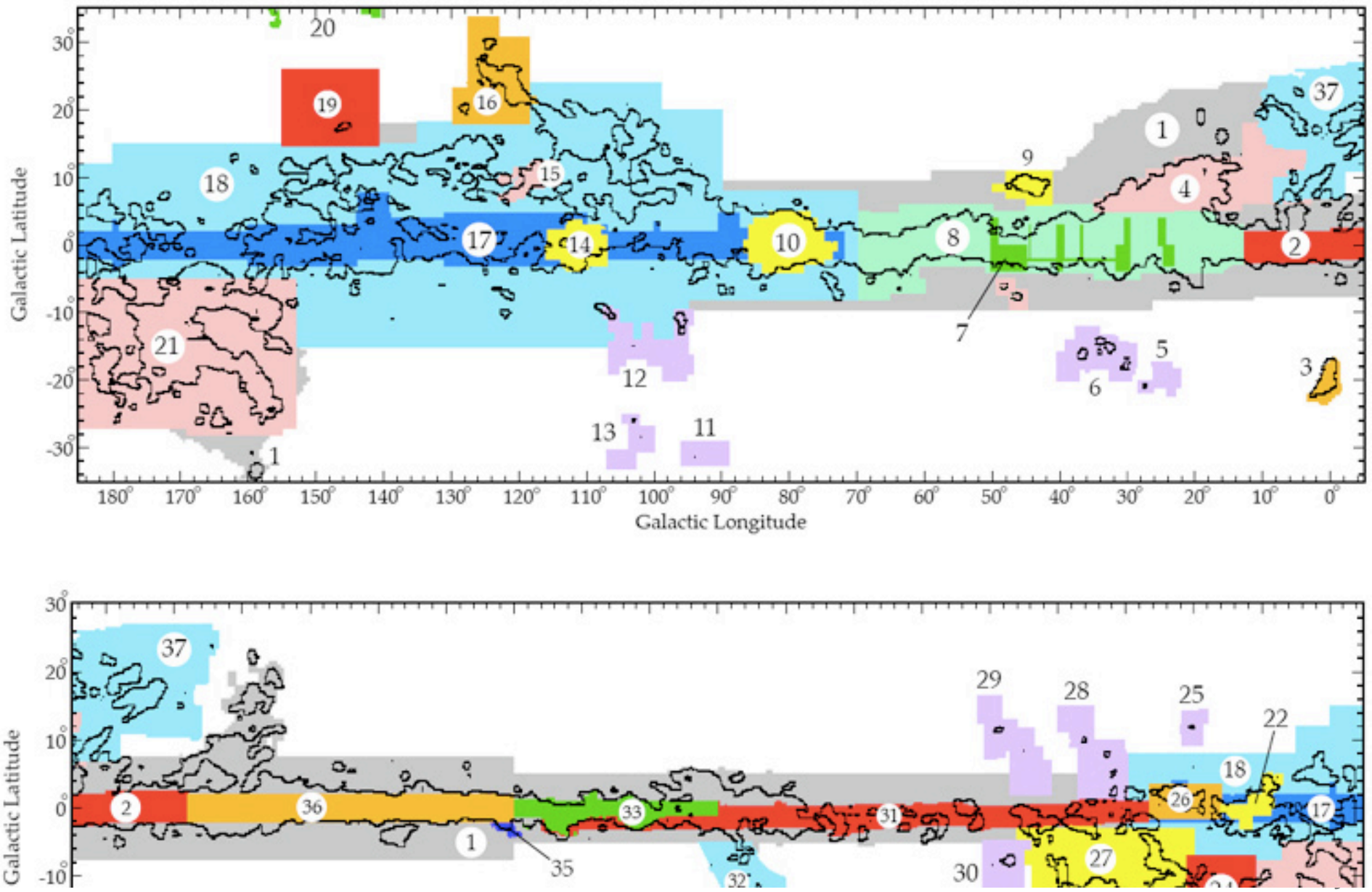
Mercè Crosas, LIBER June 2011

Radio Telescope Data Center

Smithsonian Astrophysical Observatory

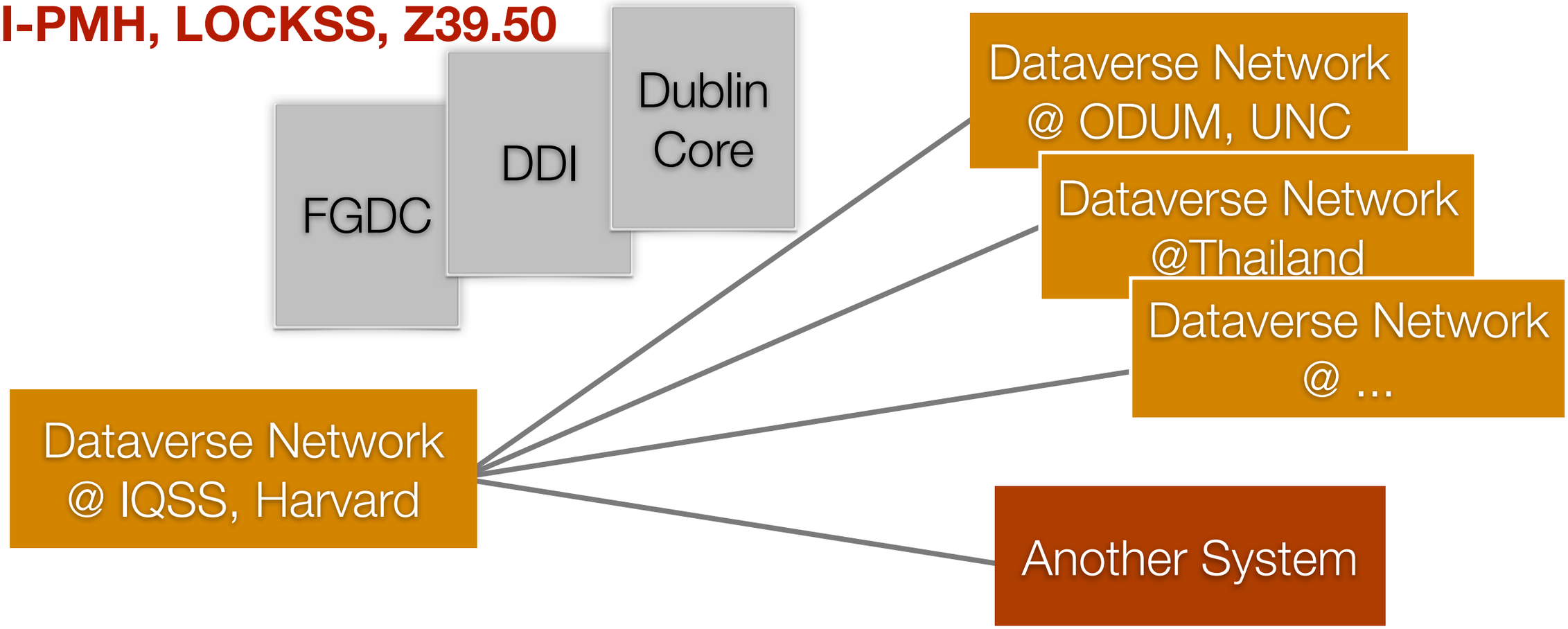
- RTDC Home Page
- R&G Division
- Submillimeter Array
  - Searching Archives
  - Inspecting Data
  - Processing Data
- 1.2 Meter Telescopes
  - CO Survey Archive
- R&G Computers
- Software Packages
- Protecting Your Data

The whole-Galaxy CO survey presented in Dame et al. (2001) is a composite of 37 separate surveys that are described and numbered in Table 1 of the paper. The data from most of these surveys can be accessed by clicking on the survey number in the map below, which is Figure 1 from the paper. Larger composites of these individual surveys are available from the link below.



SUPPORTING STANDARDS

OAI-PMH, LOCKSS, Z39.50



- Study and variable metadata are exported into **XML** (Dublin Core, Data Descriptive Initiative - DDI, ...) and MARC.
- **OAI-PMH** for harvesting metadata.
- **LOCKSS** for data duplication in multiple locations.
- **Z39.50** for distributed search.
- **E-Z Proxy** to authenticate for data access.

THANK YOU

GRÀCIES

http://thedata.org

Should Libraries play a role promoting good practices and standards for data sharing and management?

Should research institutions and funding agencies require data sharing? and should their Libraries be the place for data deposit?