Basic Dataverse Training for Astronomy

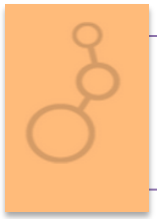
August Muench, CfA 2012

HARVARD LIBRARY

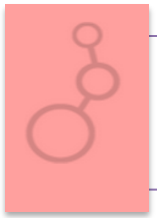
Dataverse is an open-source application for publishing, citing & discovering research data

Dataverse is a collaboration of

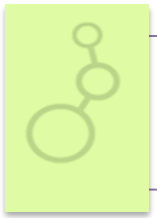
Training Agenda



Examples and Basics of Dataverse



Astronomy – Specific Usages



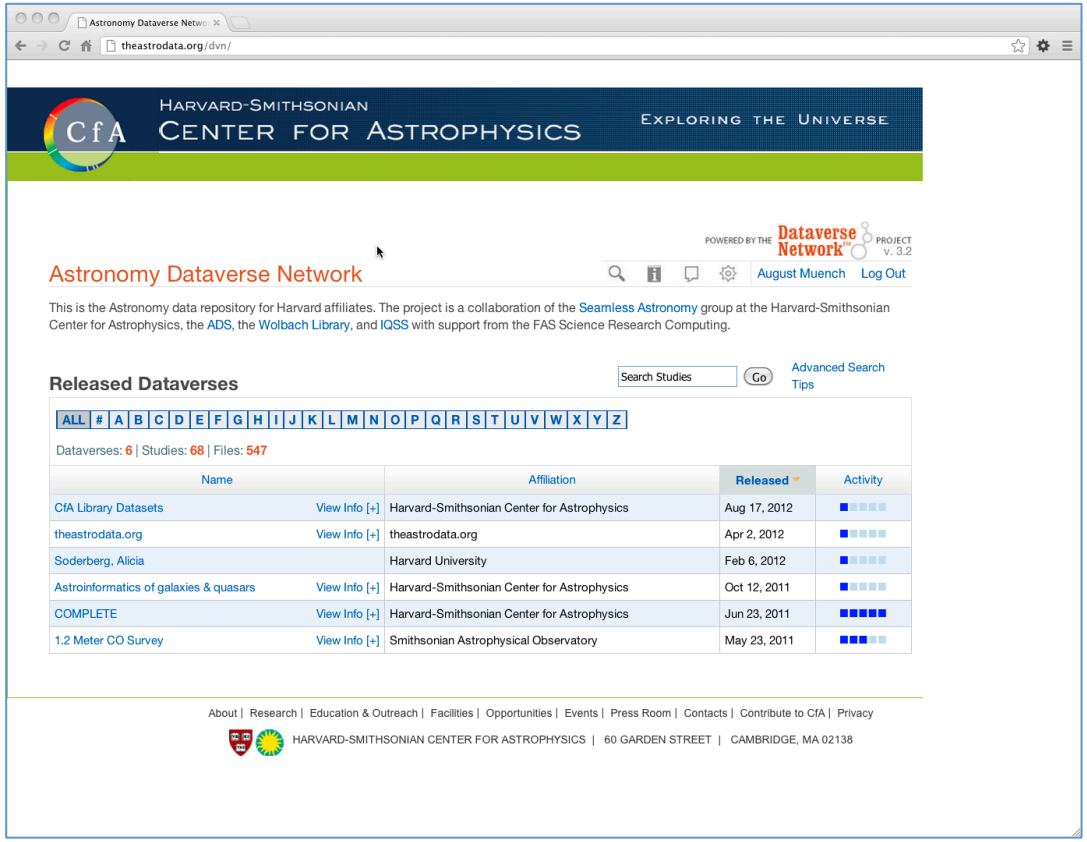
Hands-on Tutorial

Examples & Basics of Dataverse

HARVARD LIBRARY

The Astronomy Dataverse Network:

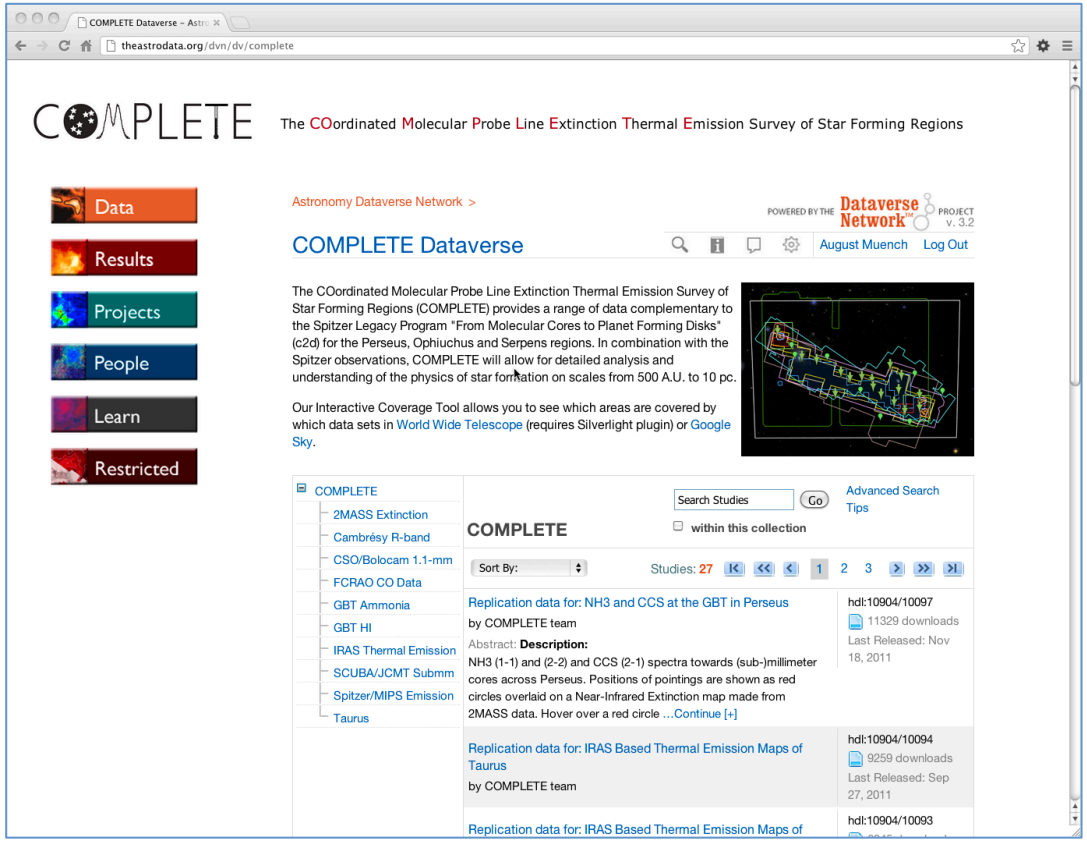
- Is for Harvard faculty, Smithsonian staff, their postdocs and students;
- Connects to:
  - Other Institutional Dataverse Networks;
  - NASA/SAO ADS;
  - The Virtual Observatory *(future)*.



HARVARD LIBRARY

Astronomy Dataverses Include:

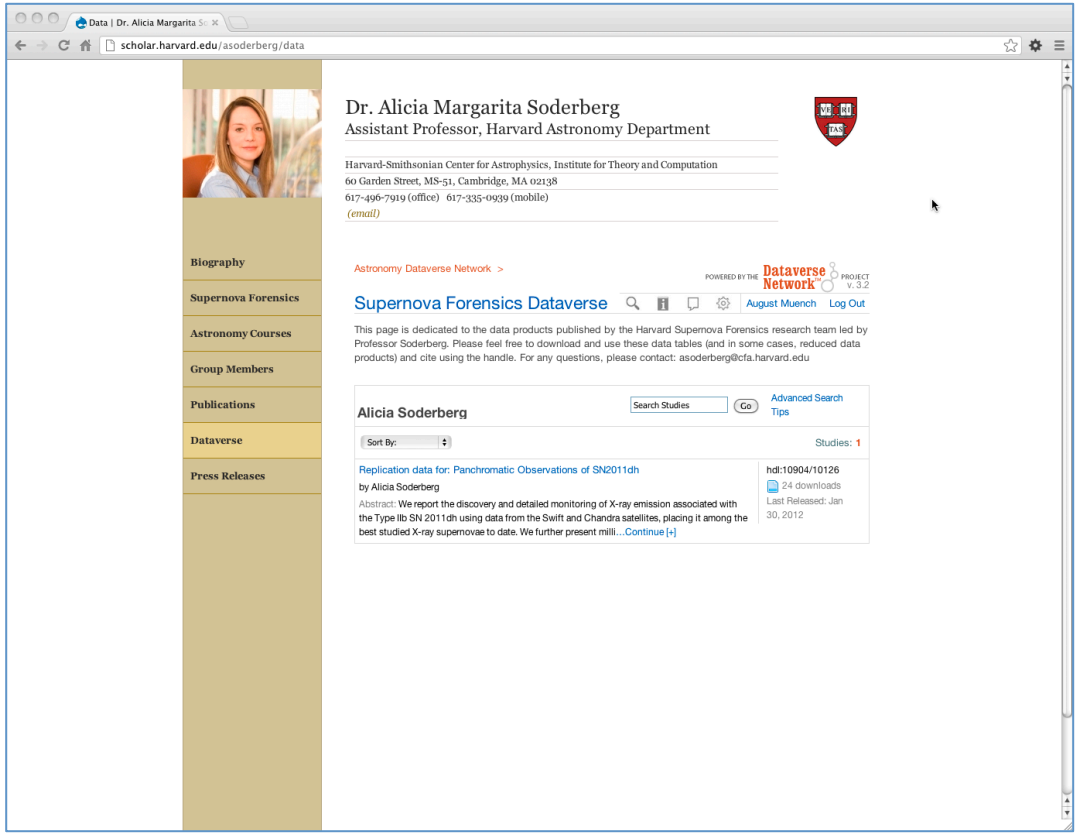
- Projects:
  - **Grant Oriented;**
  - **Collaborations;**



HARVARD LIBRARY

Astronomy Dataverses Include:

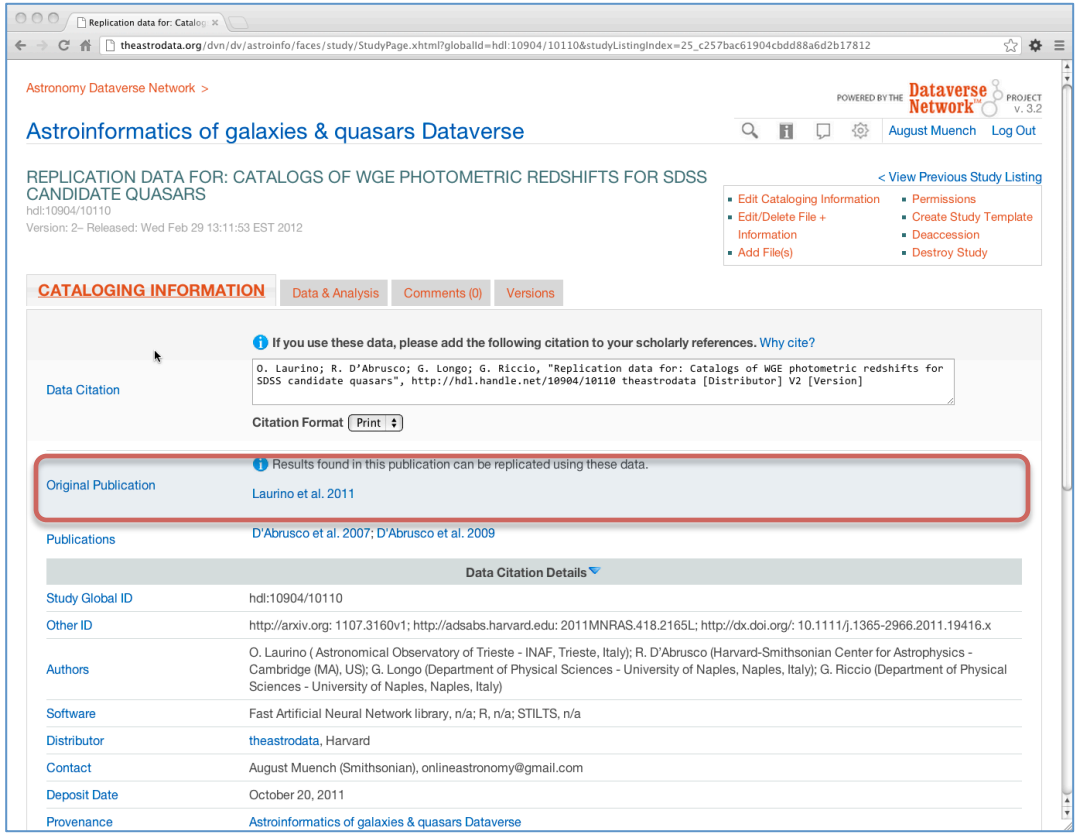
- Projects:
  - Grant Oriented;
  - Collaborations;
  - **Individuals.**



HARVARD LIBRARY

Astronomy Dataverses Include:

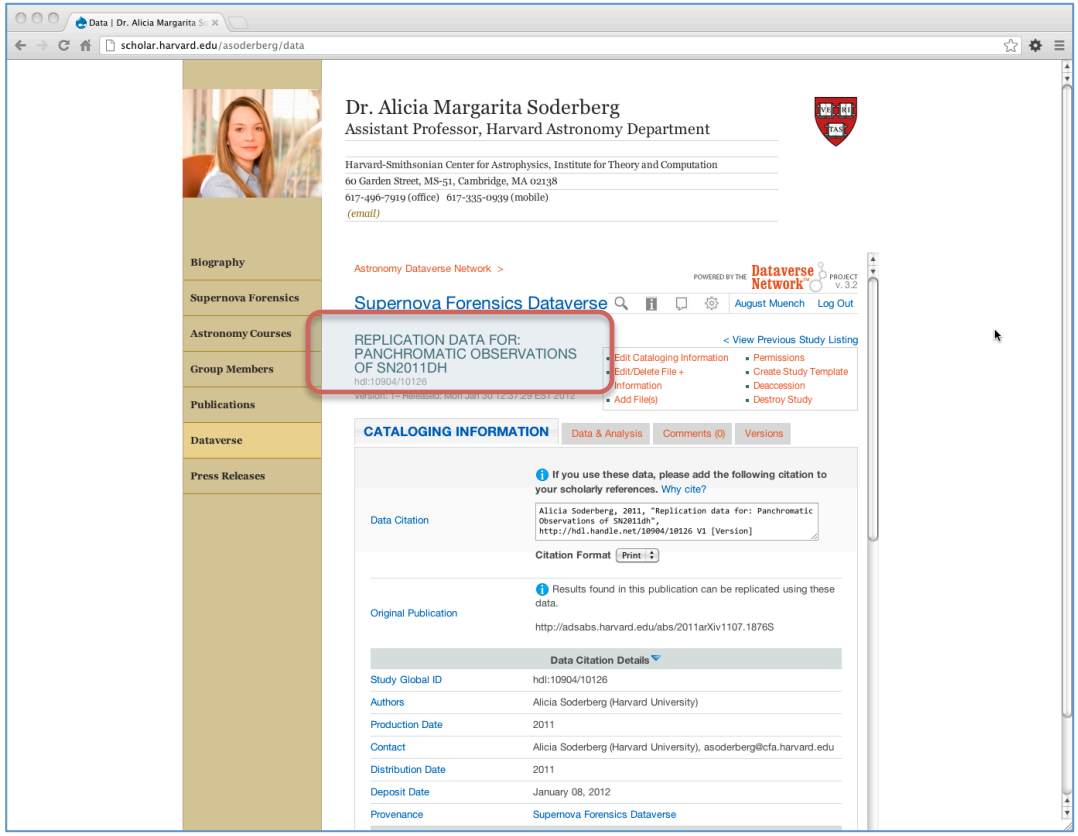
- Projects:
  - Grant Oriented;
  - Collaborations;
  - Individuals
- **Data for:**
  - **Papers**



HARVARD LIBRARY

Astronomy Dataverses Include:

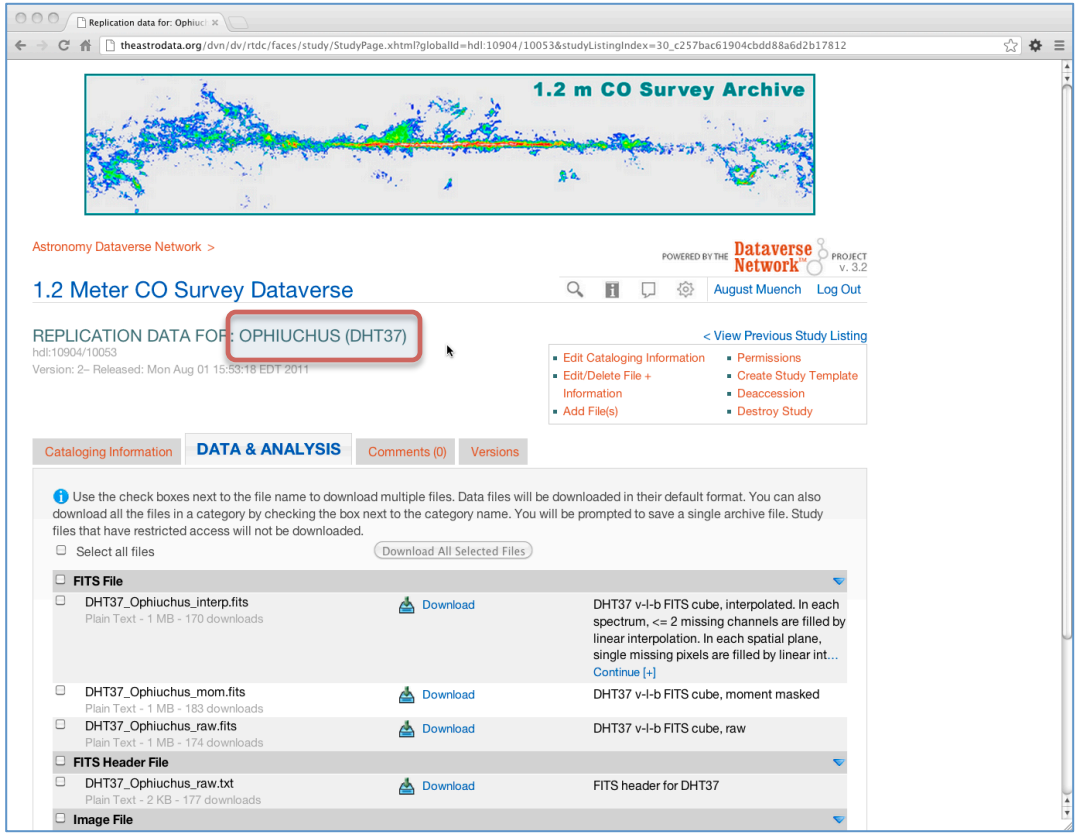
- Projects:
  - Grant Oriented;
  - Collaborations;
  - Individuals.
- **Data for:**
  - Papers
  - **Astronomical Objects**



HARVARD LIBRARY

Astronomy Dataverses Include:

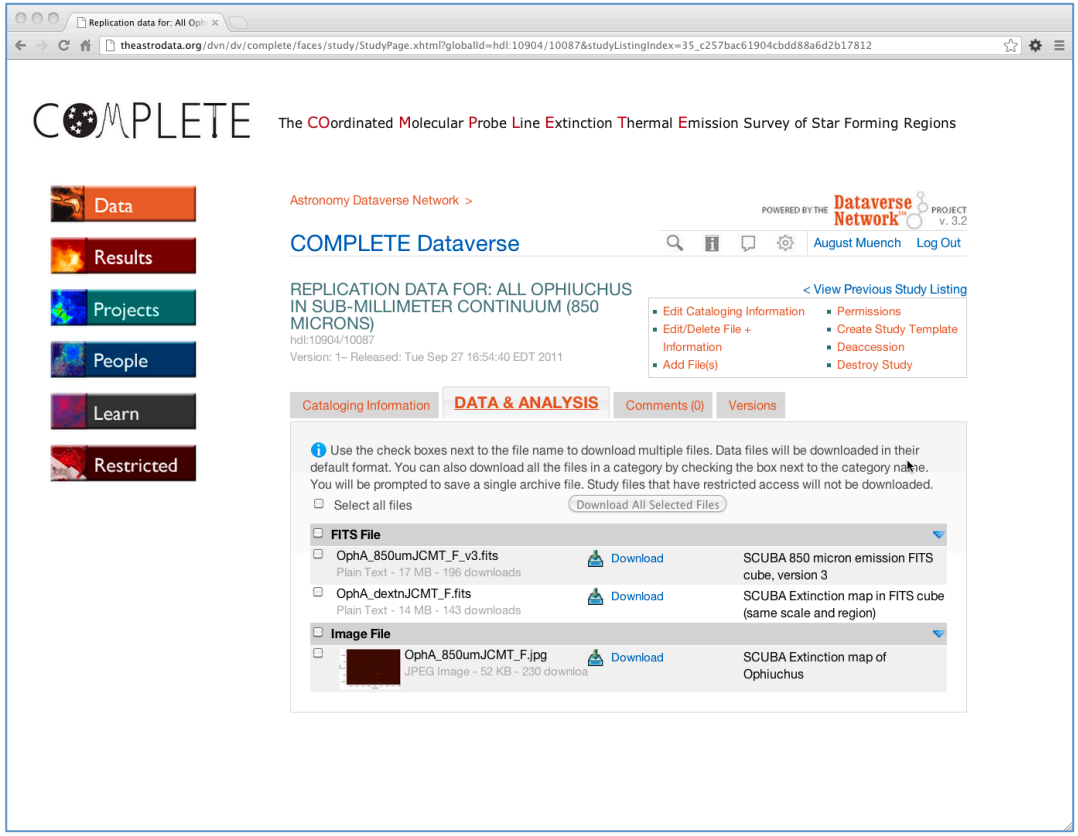
- Projects:
  - Grant Oriented;
  - Collaborations;
  - Individuals
- **Data for:**
  - Papers
  - Astronomical Objects
  - **Observations**



HARVARD LIBRARY

Astronomy Dataverses Include:

- Projects:
  - Grant Oriented;
  - Collaborations;
  - Individuals.
- **Data for:**
  - Papers
  - Astronomical Objects
  - Observations
  - **Aggregations of datasets by wavelength, time, etc.**



HARVARD LIBRARY

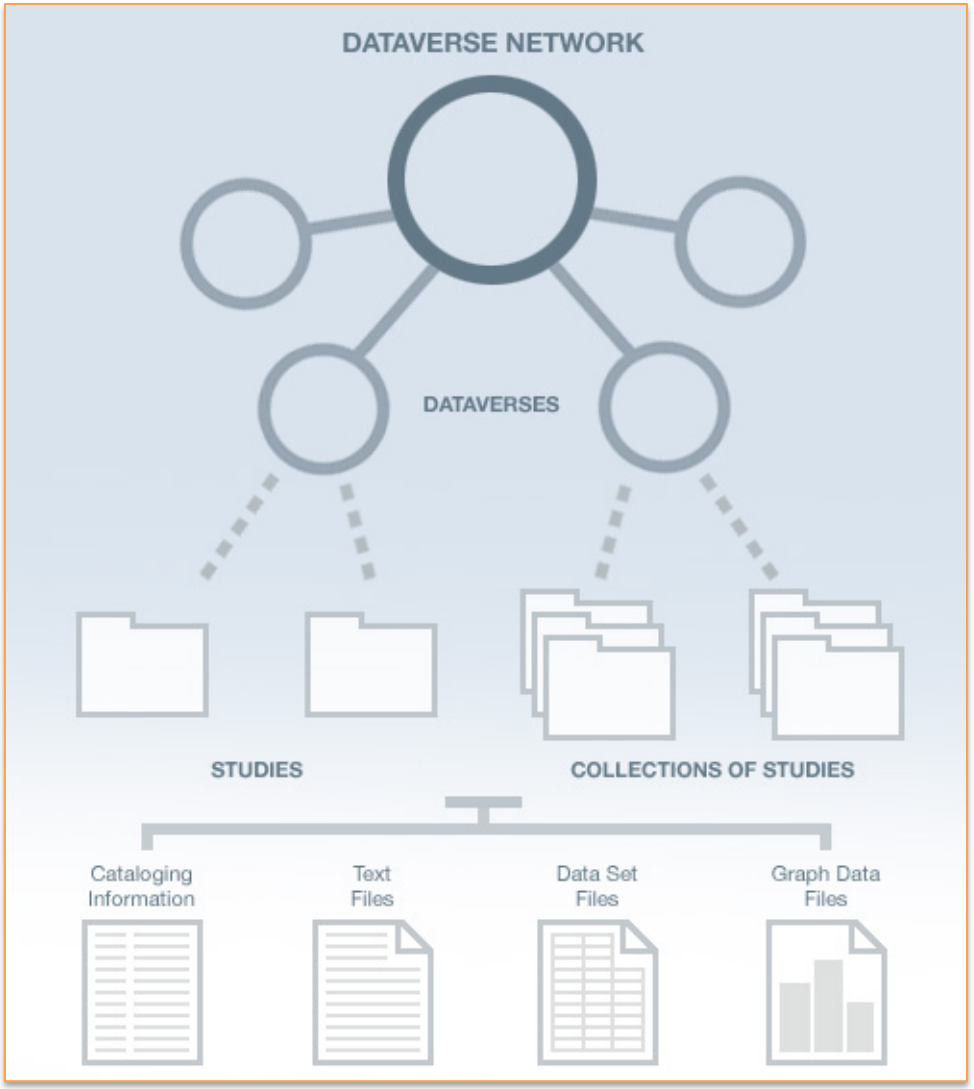
Researchers use the Dataverse Network to:

- Archive data;
- Supplement publications;
- Fulfill data management requirements of funding agencies;
- Increase recognition;
- Enhance impact;
- Gain access control.

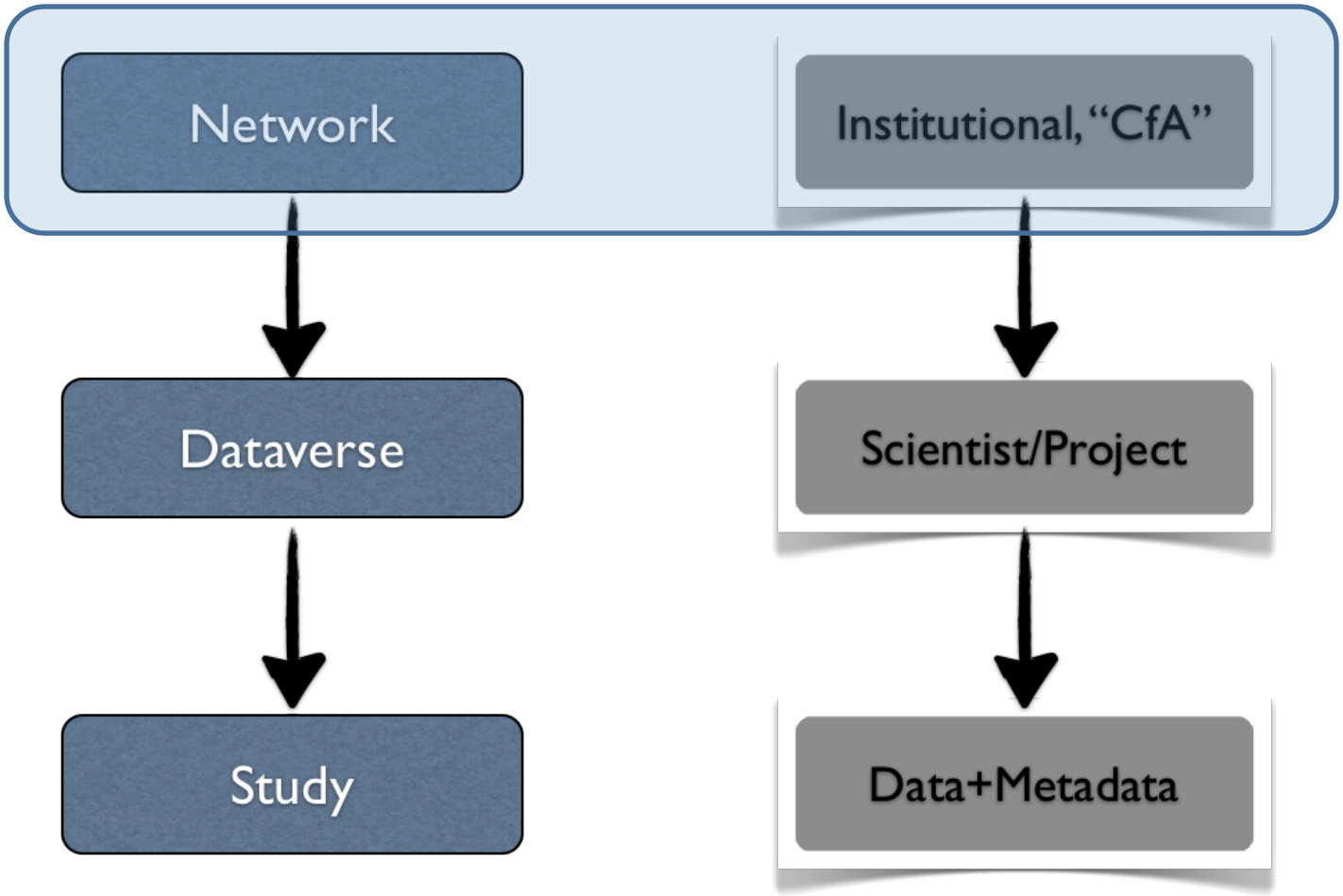
Defining Terms

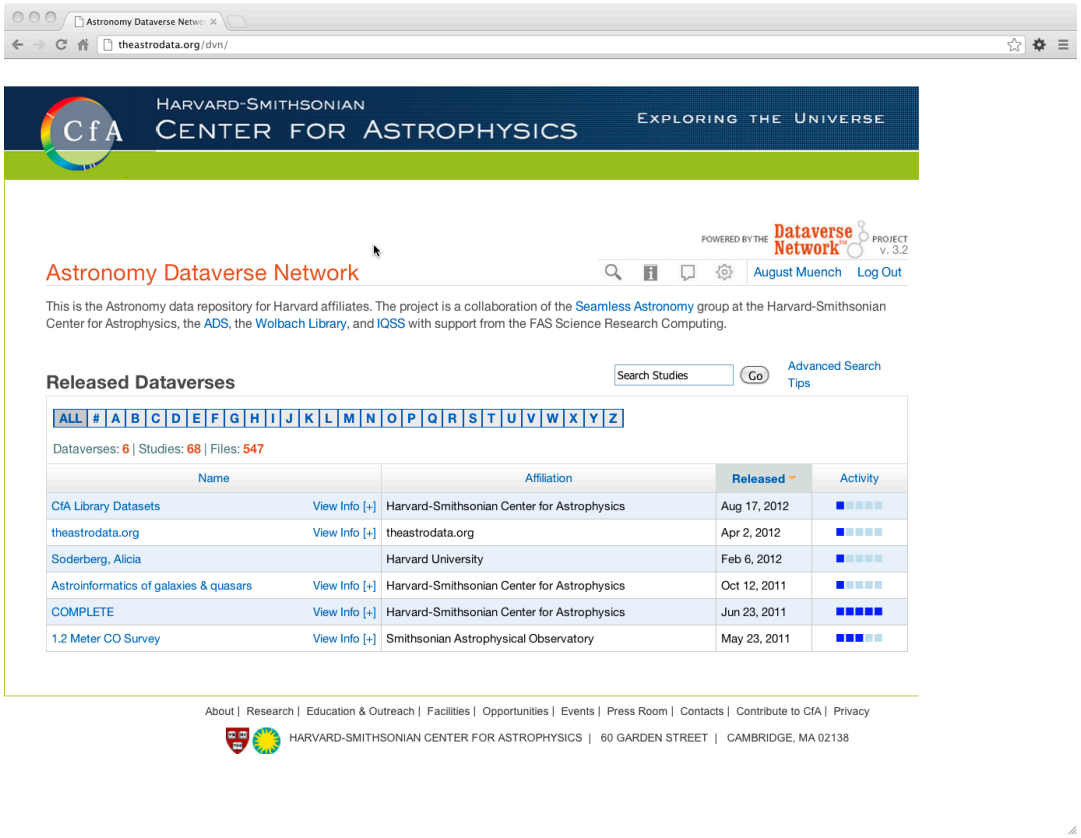
Dataverse Network:
Web-based software for data management that can be used as a data repository to store, preserve, cite and publish data.

- **Dataverse**: a *container*, created by a user (or project or institution) on the Dataverse Network web site, *to store all of their research data.*
  - **Study**: a specific cataloged object within a user's or institution's Dataverse to which related data files are attached. These can be files from *published research*, from research that is *in progress*, or just a *set of data.*
    - **Files**: These are the *research data and supplemental information* files produced during research that need to be preserved.


DATAVERSE NETWORK
DATAVERSES
STUDIES
COLLECTIONS OF STUDIES
Cataloging Information
Text Files
Data Set Files
Graph Data Files

Dataverse Network Diagram

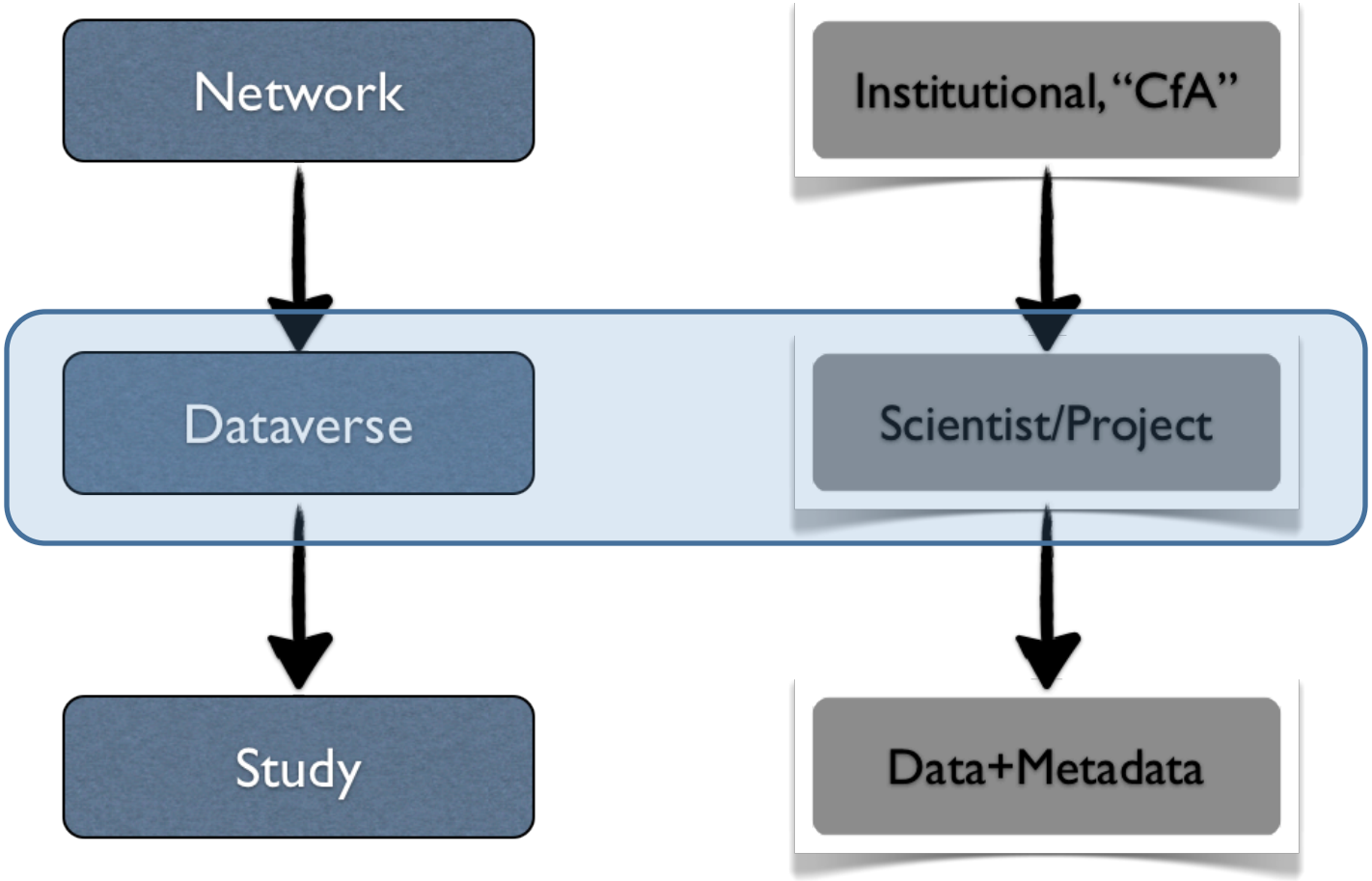


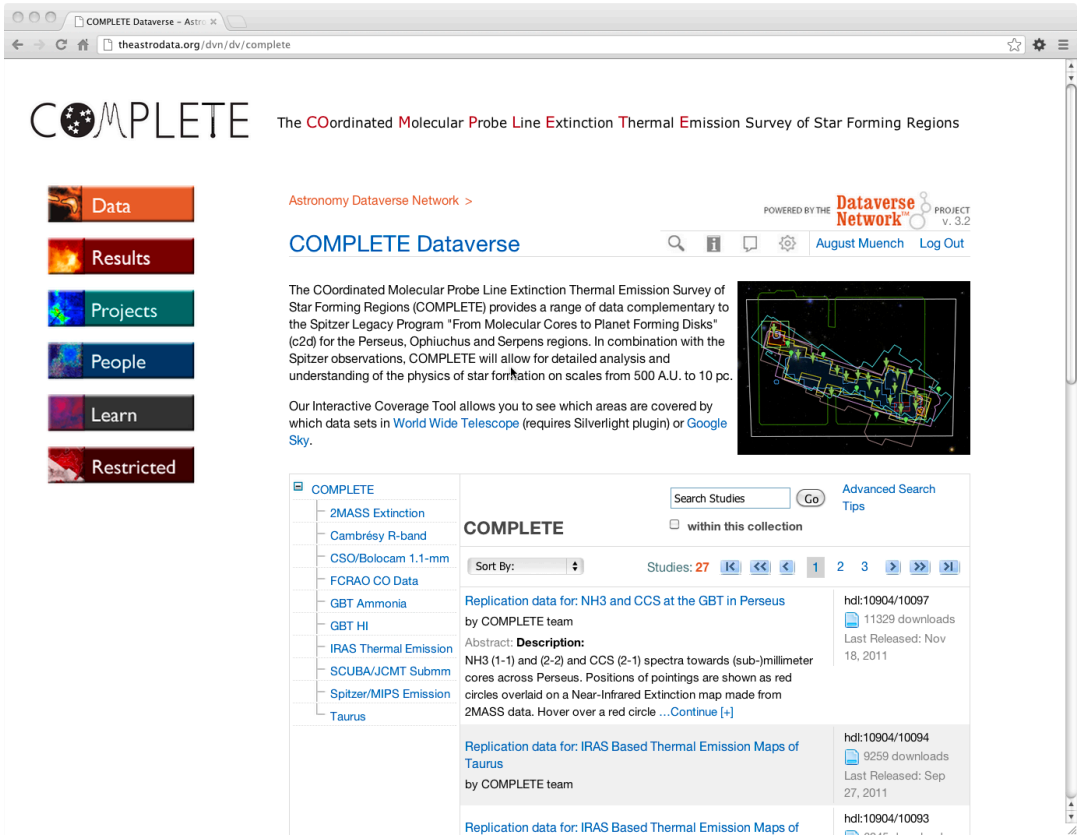


The Astronomy Dataverse Network

The Astronomy Dataverse *network* is the administration layer, or an umbrella that lets us manage all the projects from the Harvard-Smithsonian Center for Astrophysics.

Dataverse Network Diagram



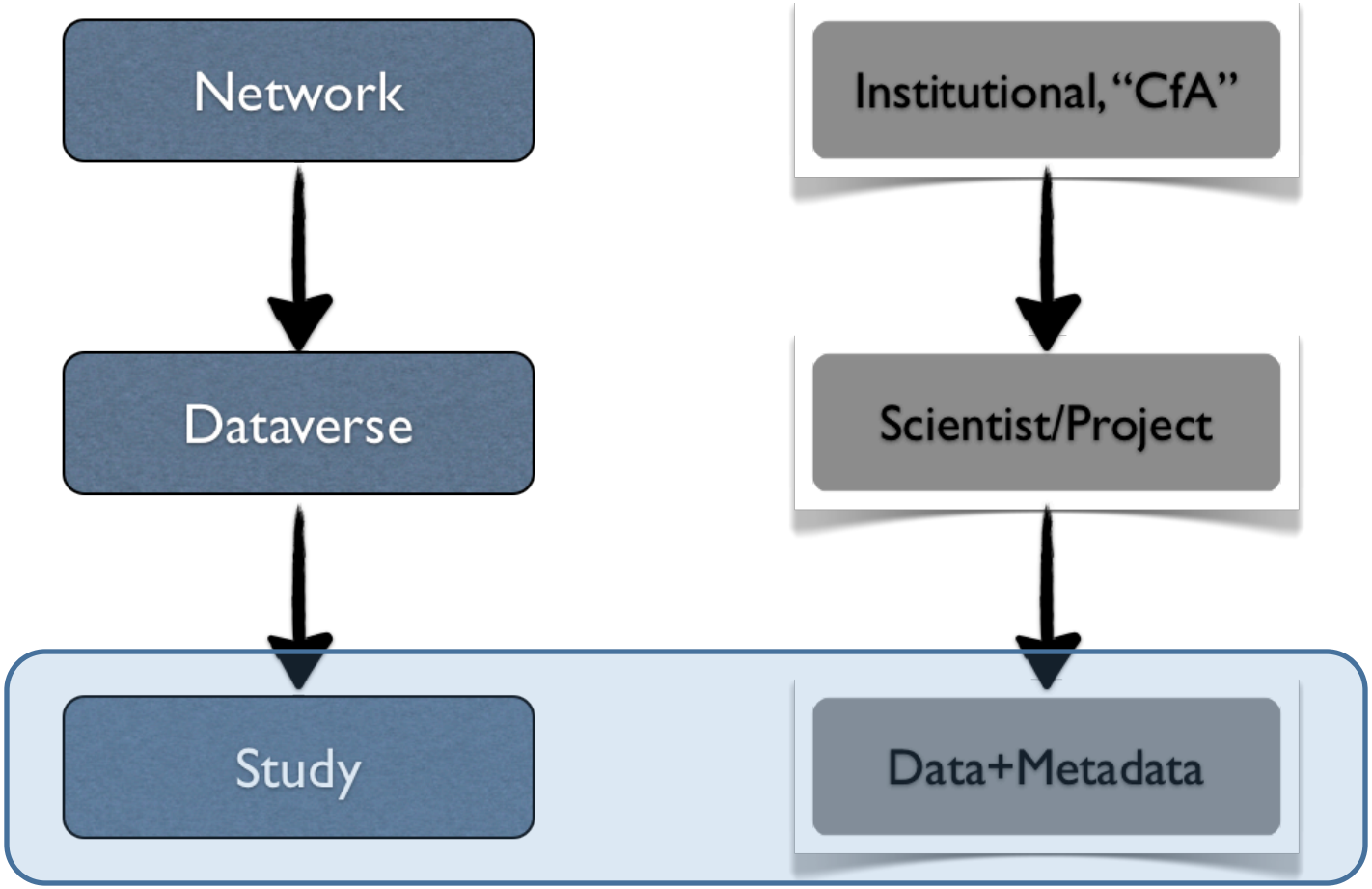


The "COMPLETE" Dataverse

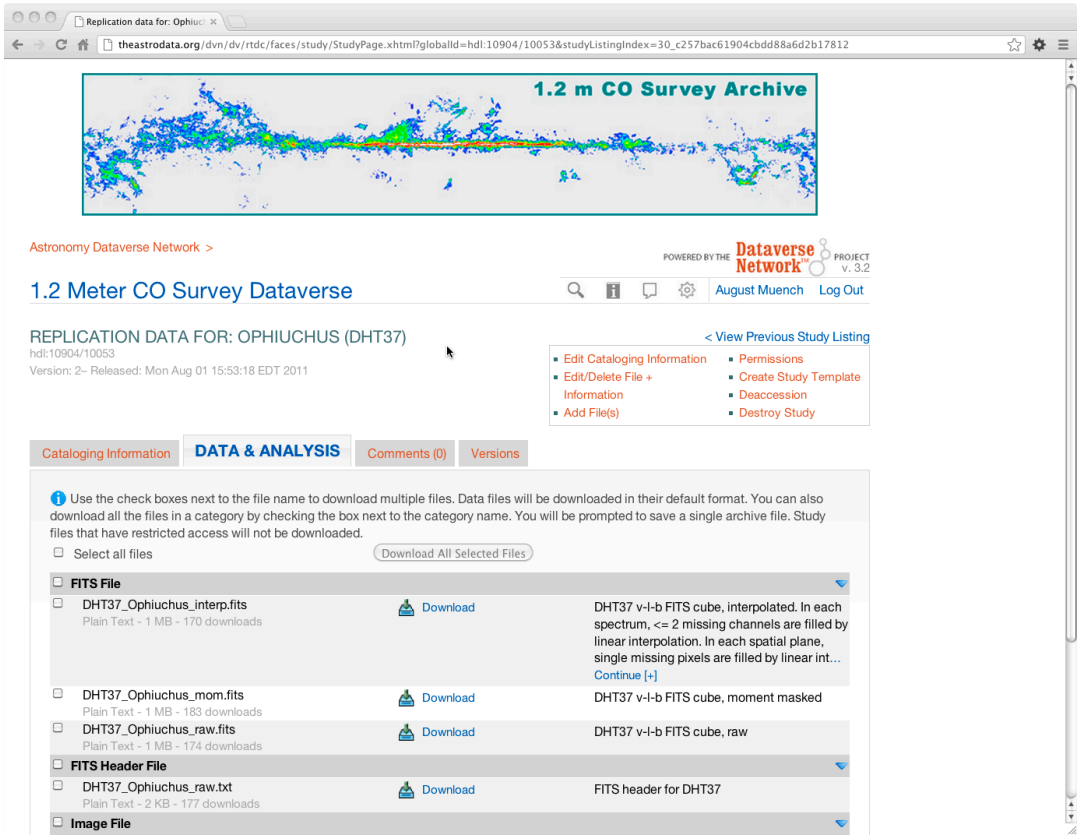
The COMPLETE project created a single *Dataverse* for all the datasets and subprojects related to their work.

HARVARD LIBRARY

Dataverse Network Diagram



HARVARD LIBRARY



The "OPHIUCHUS (DHT37)" Study

The 1.2m CO Survey broke the sky up into individual observing runs with a particular instrument and survey pattern. All the "DHT37" run data are contained in this *study*.

HARVARD LIBRARY

Astronomy Specific Usage of the Dataverse Network

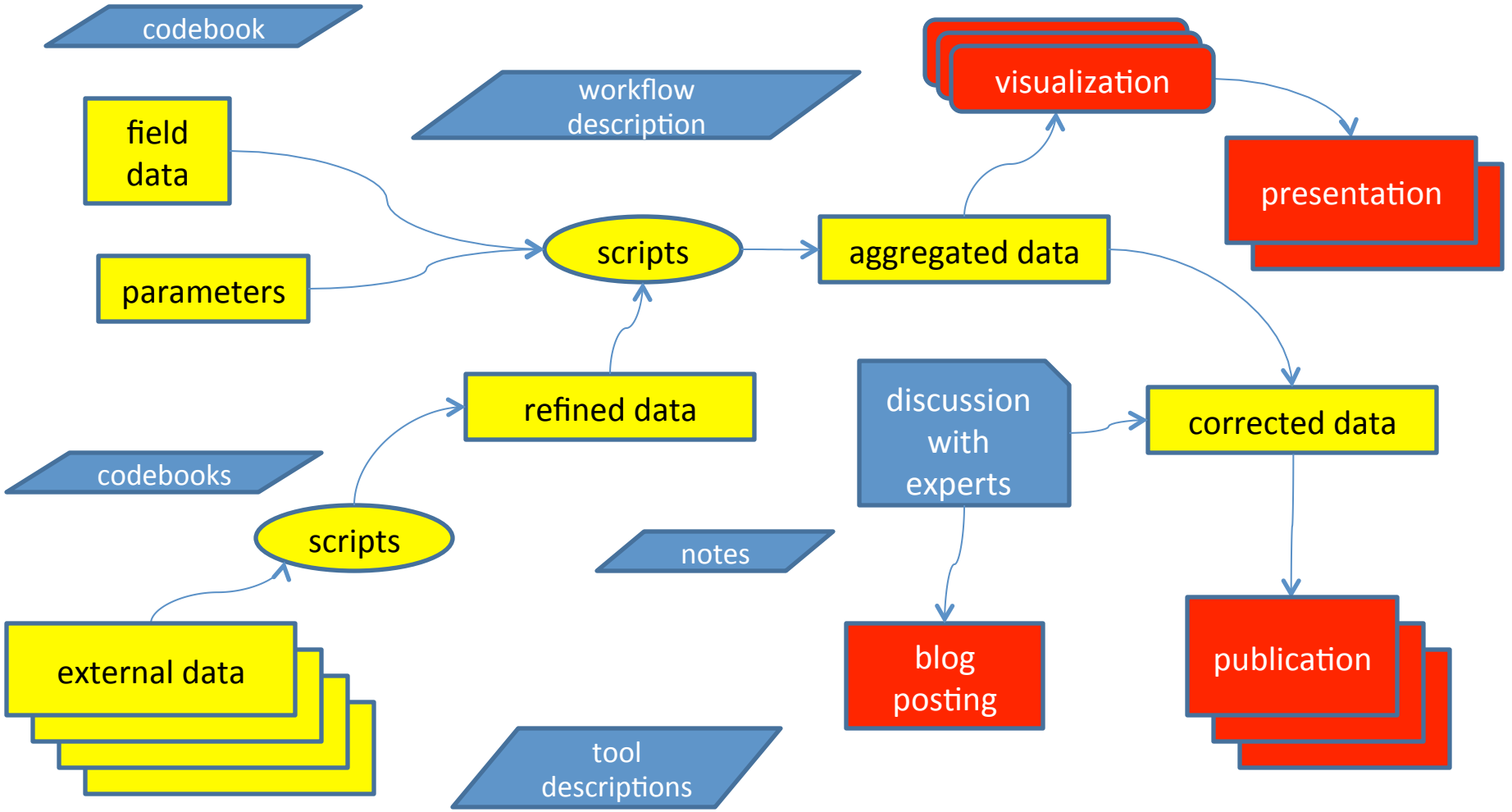
Advice (tips & recommendations) for dealing with file formats, study creation, & terms of use

What Makes a Study?

- Some kind of Homogenous "Packaging" of the Data

- Examples of Astronomy Packages:
  - Papers
  - Objects
  - Observations
  - Wavelength
  - Intersection of any of these.

What do you package into a Study?



HARVARD LIBRARY

Astronomy File Formats

- Try to choose a simple set of file formats:
  - FITS for Images, Spectra, Data cubes, Tables.
  - *USE FITS AS OFTEN AS POSSIBLE*
- Plain Text Files
  - Use File Extensions (e.g., .txt, .md, .rst, .tex)
- Tables
  - If not FITS then:
    - "Structured" Tables => CSV, Tab Delimited, IPAC Table.
    - Don't use whatever random plain text table you happen to have lying around.

Astronomy Metadata

- Use the Example Templates
  - Examples include:
    - "Basic"
    - "Publication Data"
  - You can copy these to your own Dataverse
    - See *Advanced Topics*
- Go back and edit your cataloging data!
- Embed metadata in FITS, Tables.

Astronomy Terms of Use

- Follow Domain Behaviors
  - NASA: Public Domain
  - Institutional Observatory:
    - Proprietary periods
    - Citation Standards

- Additional Terms for Reuse:
  - Citation => *Do you have a specific set of related papers that you want cited when these data are reused?*
  - Sharing => *Do you have expectations about how the data will be reused by others?*

Hands-on Tutorial: Create a Dataverse

Follow steps in this section to create your own Dataverse

Dataverse URLs

- Experiment in the "demo" version of the Dataverse Network:

  http://dvn-demo.iq.harvard.edu/dvn

- Create a real Dataverse in the "production" version of the Dataverse Network:

  http://dvn.iq.harvard.edu (Social Sciences)

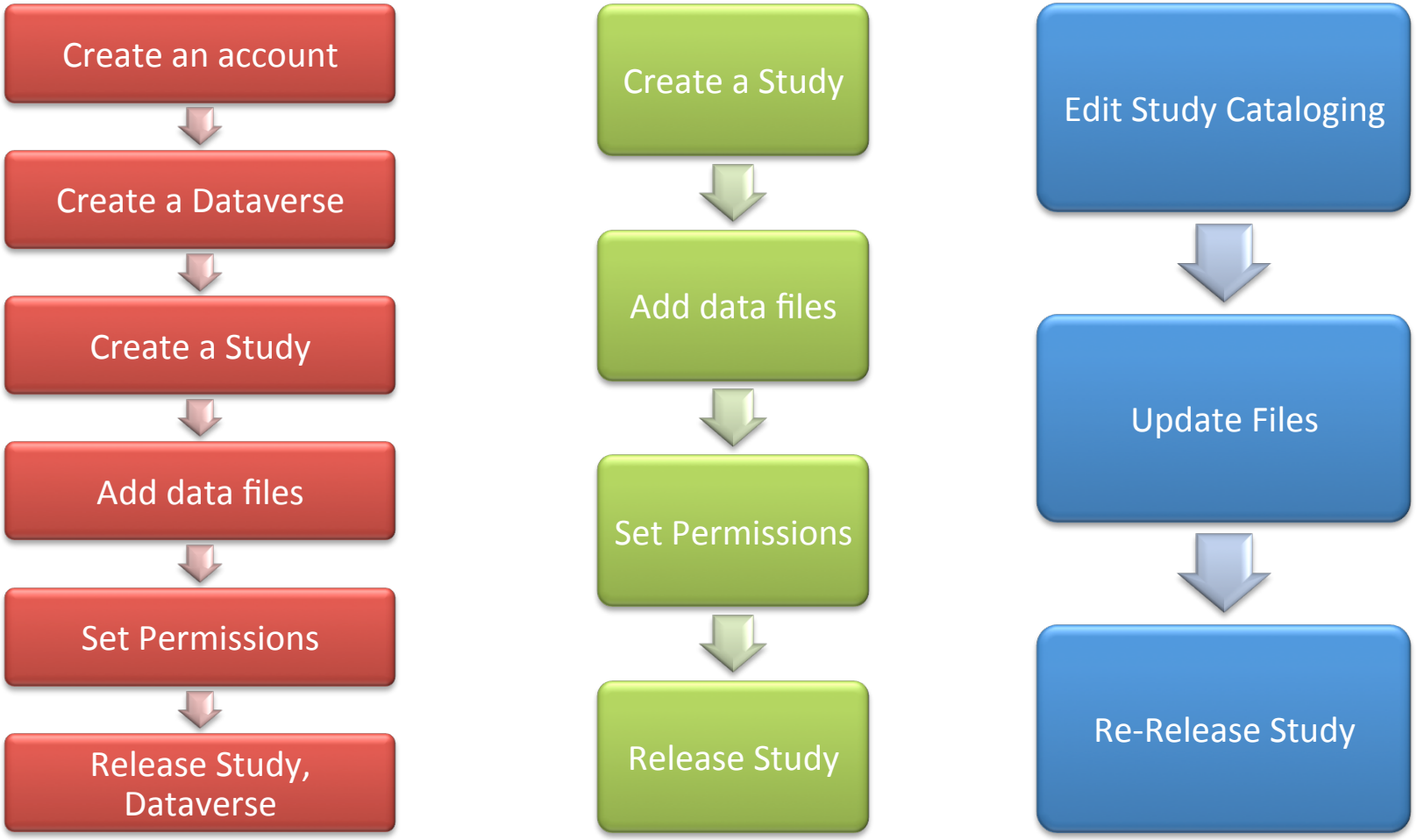
  http://theastrodata.org (Astrophysics)

For More Information

- Digital Scholarship @Harvard
  http://ds.hul.harvard.edu/ds
  - Schedule of Events
  - Handouts
  - Slides
  - Other resources

- Dataverse Support
  dvn_support@help.hmdc.harvard.edu (*registration, bugs, technical help*)
  theastrodata@cfa.harvard.edu (*astronomy user community list*)
    - *To join email majordomo@cfa.harvard.edu with body text as "subscribe theastrodata"*

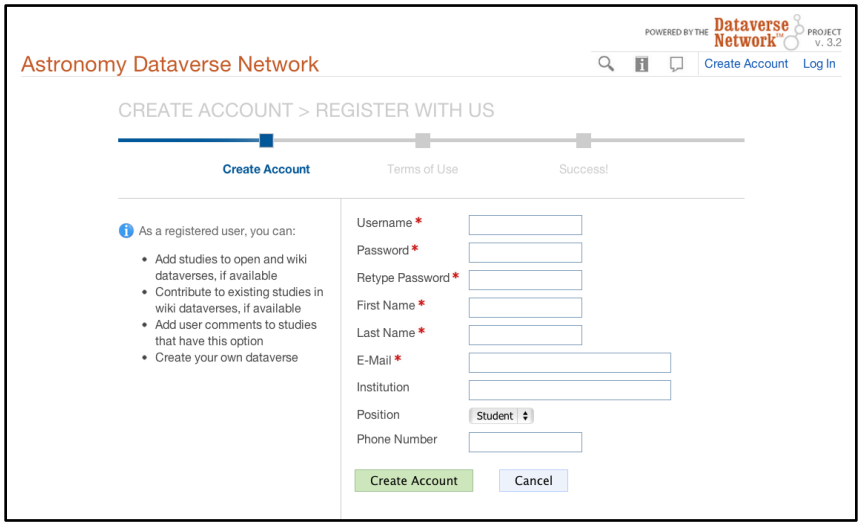
Dataverse Workflow(s)



HARVARD LIBRARY

Create an Account

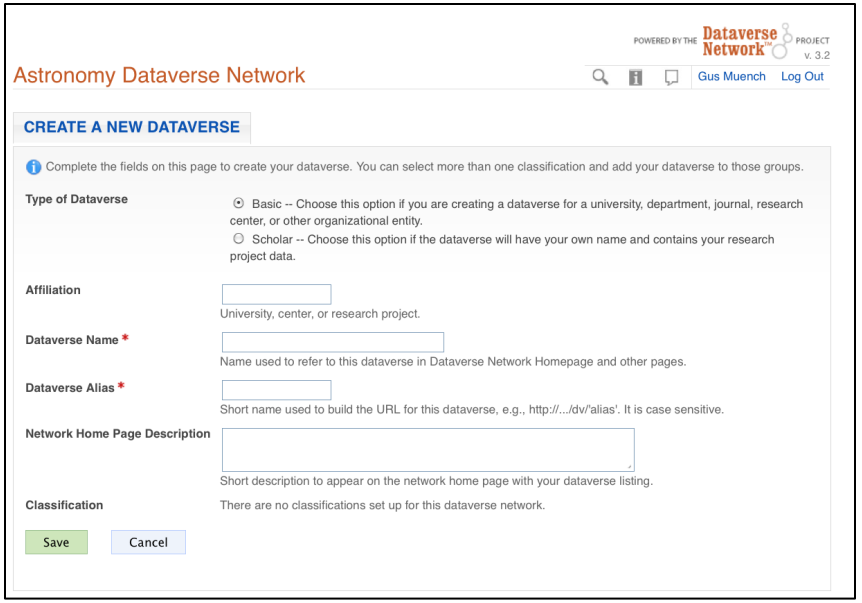
1. Go to the Dataverse Network
   - Demo: http://dvn-demo.iq.harvard.edu/dvn
   - Real: http://theastrodata.org
2. Select "**Create Account**"
3. Complete account form.
4. Log in.



HARVARD LIBRARY

Create Dataverse

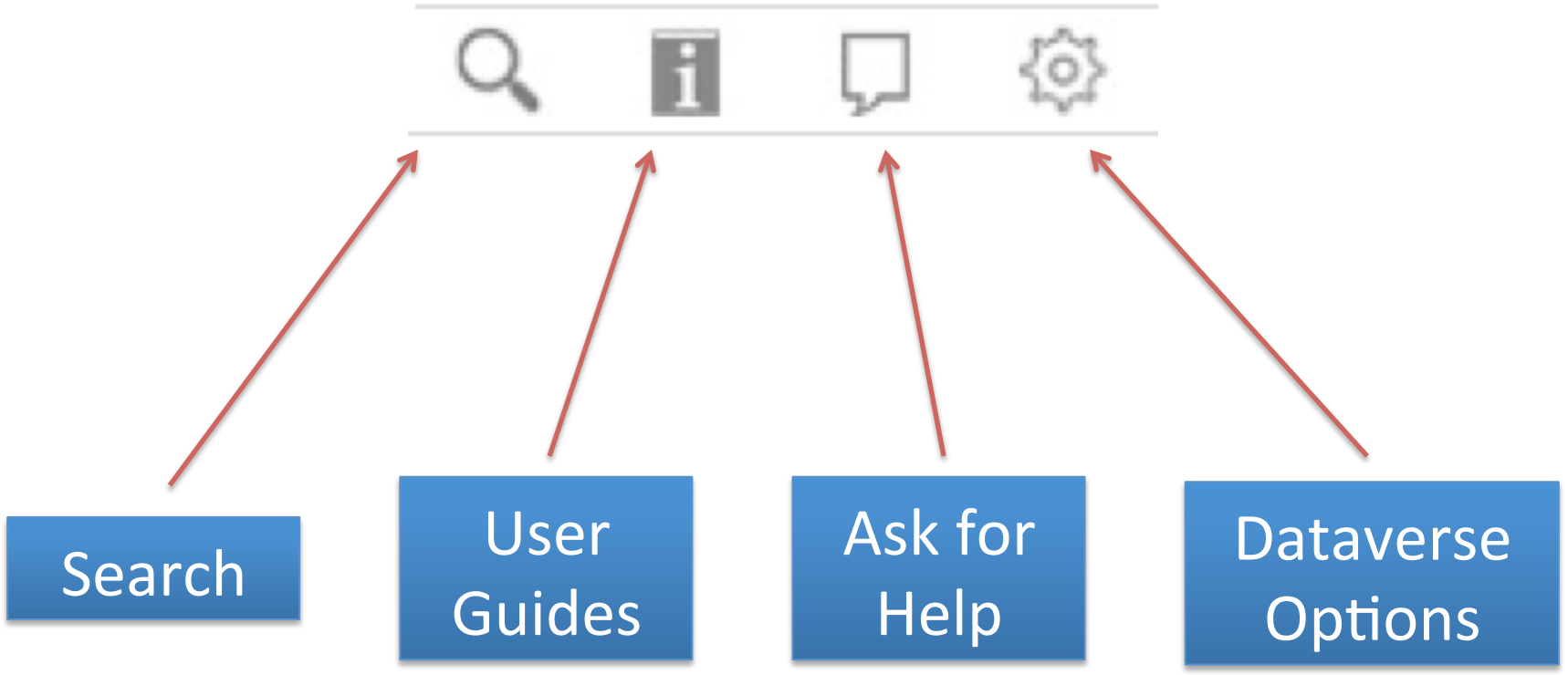
1. Go to the Dataverse Network
   - Demo: http://dvn-demo.iq.harvard.edu/dvn
   - Real: http://theastrodata.org
2. Select "**Create a Dataverse**" (*)
3. Choose Dataverse "*Type*"
   - "*scholar*" for branding as individual;
   - "*basic*" if institution or project.
4. Complete form; **Save**.
5. URL to your Dataverse is emailed to you.
   - Bookmark this URL.
     It is your Dataverse homepage



(*) If the "Create a Dataverse" button is not present then go to:
http://theastrodata.org/dvn/faces/site/AddSitePage.xhtml

HARVARD LIBRARY

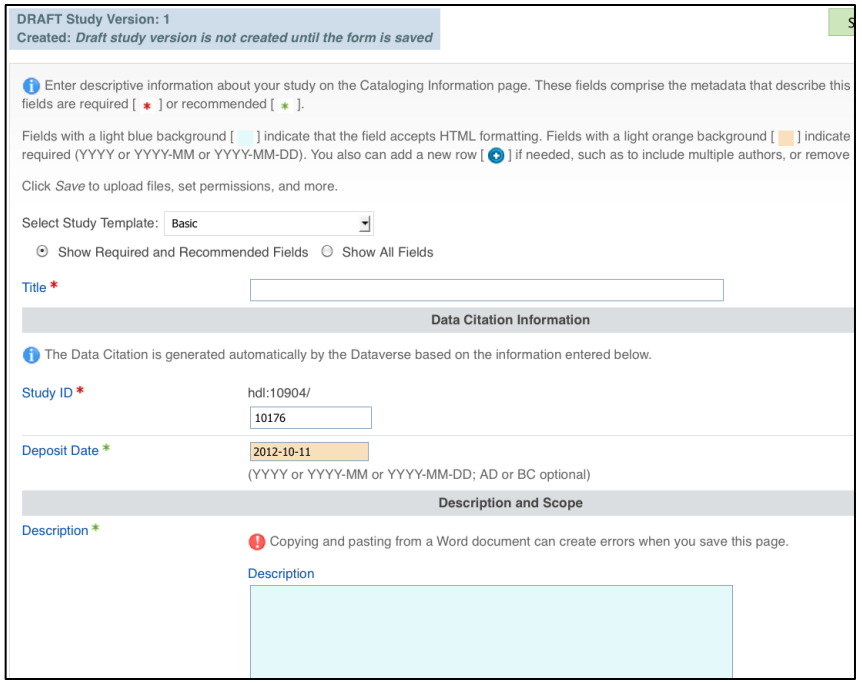
Key to Dataverse Options Menu



HARVARD LIBRARY

Create a Study

1. Go to your Dataverse homepage and click the Dataverse **Options Icon**.
2. Click "**Create Study**".
3. Enter a *title*.
4. A *study ID* is automatically created.
5. Click **Save**.

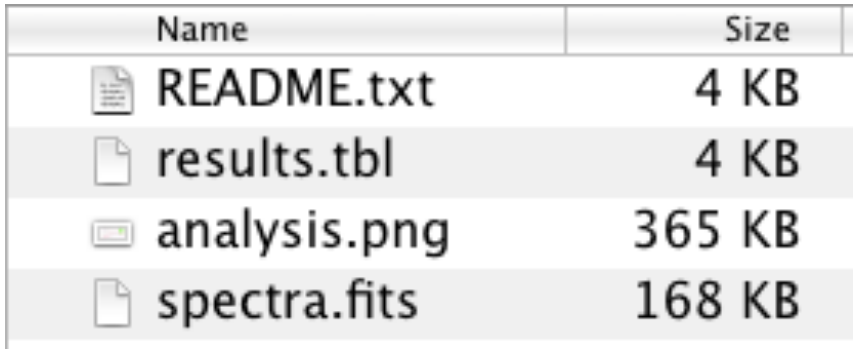


Note: *Title* is the minimum requirement, but additional catalog information should be provided!

HARVARD LIBRARY

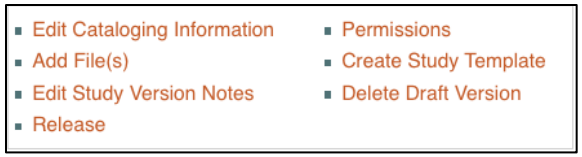
Example Dataset

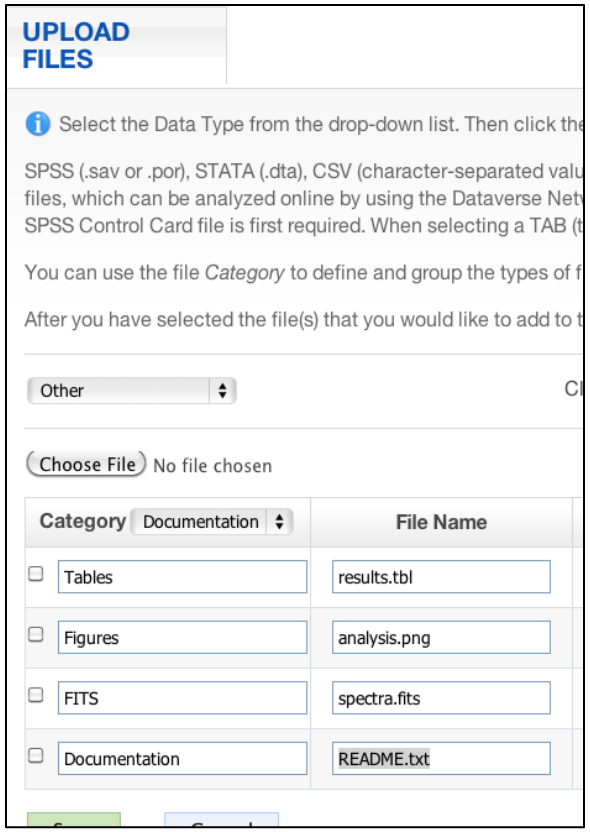
- You can download an example dataset from
  - http://bit.ly/astrodvnex01
- It contains:
  - README.txt
  - spectra.fits
  - analysis.png
  - results.tbl



Add Data & Documentation Files (1)

1. View the study you will add data files to.
2. Click the "**Add File(s)**" link in the Action box on the right.
3. Select the "**Choose a Data Type**":
   - Use "*Other*" for all Astronomy data.
4. Click "**Browse**" & upload a file from your local system.
5. Enter a file "*Category*" for grouping related files within a study:
   - e.g., Documentation, FITS, Image, Tables.
6. Enter a "*Description*" for each file (optional).
7. Click "**Save**" to retain uploaded file(s).



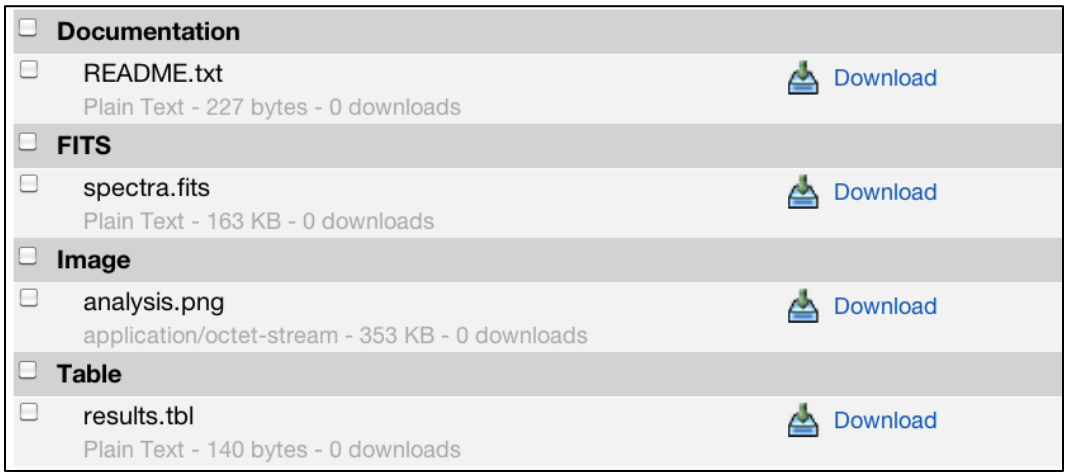


HARVARD LIBRARY

Add Data & Documentation Files (2)

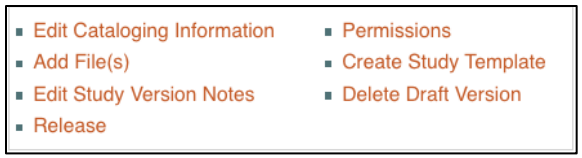
Repeat for all files in dataset:

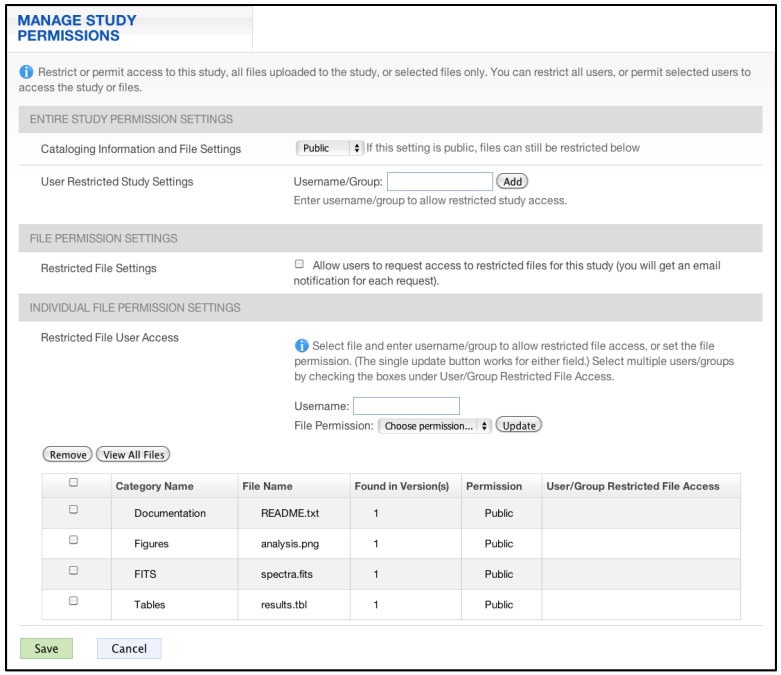
- Make sure to use the correct "*Data type*".
- Click "**Save**" when finished.
- Categories, descriptions, and additional files can also be added later.



Study Permissions

1. View the active study and click the "**Permissions**" Option in the Action box.
2. Two levels of permission are available:
   - *Entire Study*
   - *Individual File*
3. Allow users to request access to restricted files.



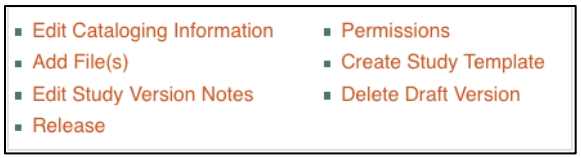


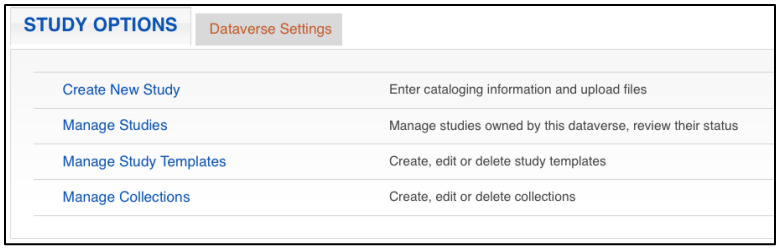
HARVARD LIBRARY

Release The Study

- If viewing the study to be released, click the "**Release**" link in the Action box on the right side..

- Another way to release a study is from the "*Manage Study*" page:
  - Go to your Dataverse home page.
  - Click on the Dataverse **Options Icon**.
  - Click "**Manage Studies**".
  - Click on the "**Release**" link on study's row entry (the far right of the table).



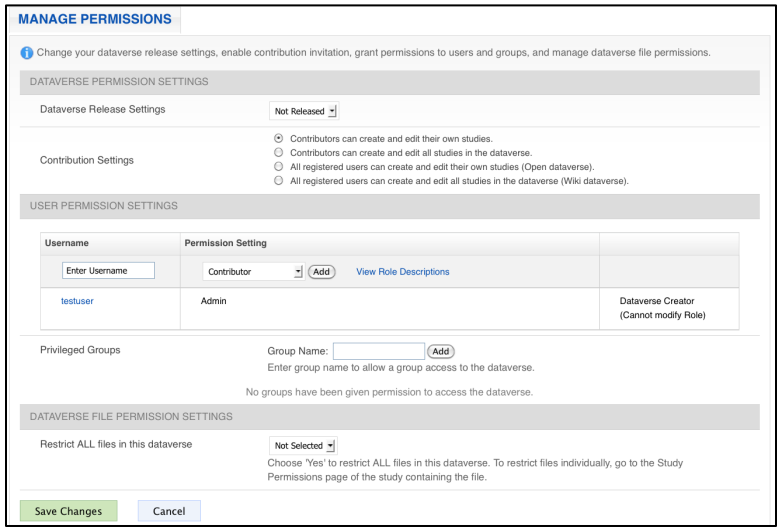


HARVARD LIBRARY

Release the Dataverse

Releasing a Dataverse is as simple as changing the status to released:

1. Navigate to the Dataverse homepage.
2. Click on the "Dataverse options" gear **icon.**
3. Click on the "**Dataverse Settings**" tab.
4. Click on "**Manage Permissions**".
5. At the top of the page, under "*Dataverse Permissions*", change the status to "*Released*".
6. Click "**Save**".



HARVARD LIBRARY

Released, published, & searchable

- You Dataverse is now released and appears on the network homepage for browsing by others.
- Studies in the Dataverse are now searchable from the network homepage. By this definition the data is now “published.”
- Any subsequent editing, including adding files, will create a new draft “version” with those changes until that too is released.
- Updates increment the study version number.
- “Versioning” is an advanced topic to be discussed later.