



Elizabeth Quigley
Usability Specialist
IQSS @ Harvard University
equigley@iq.harvard.edu

Introduction to Dataverse

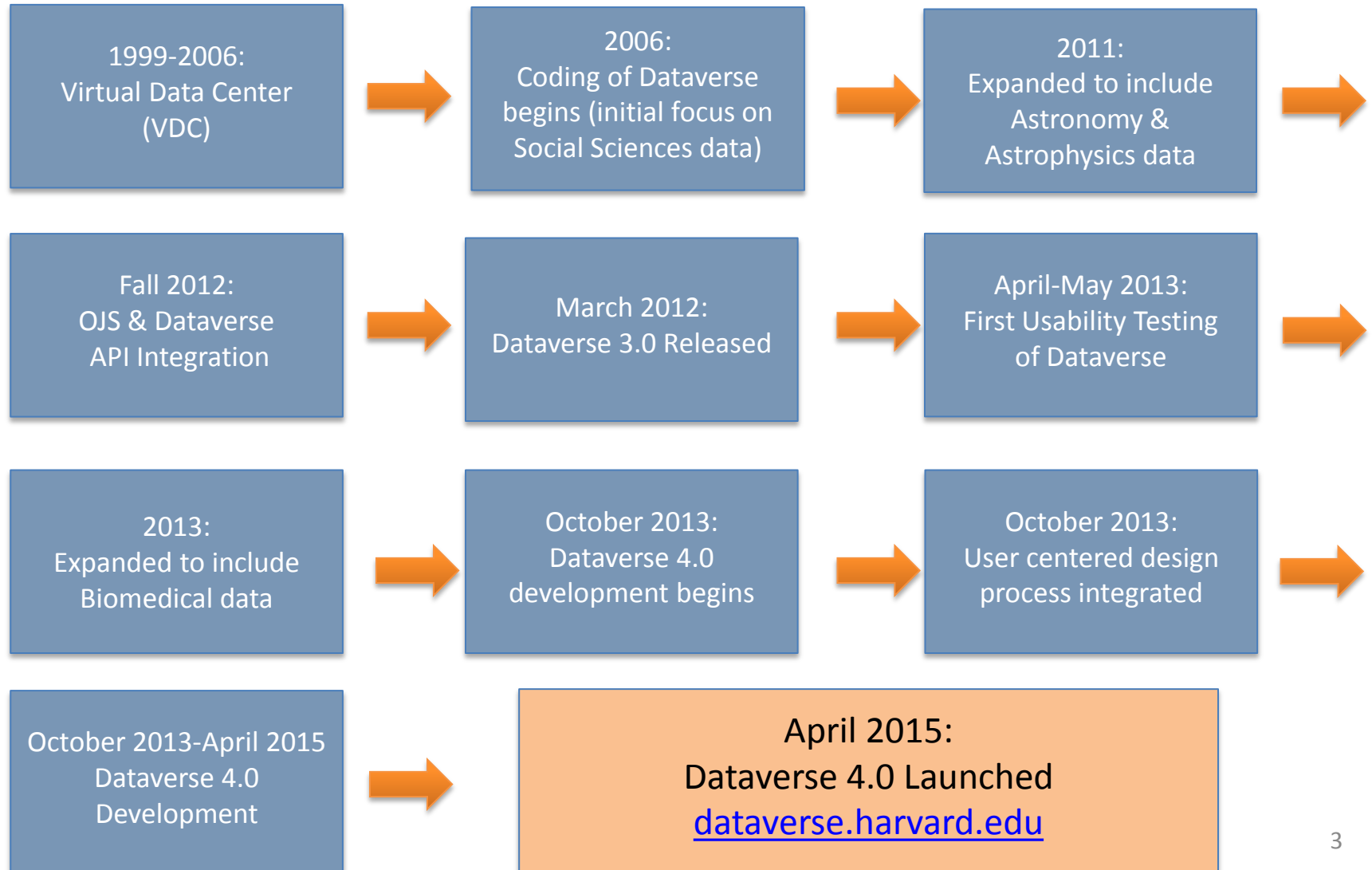
Software framework for publishing, citing and preserving research data (open source on [github](#) for others to install)

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
- Default CC0 Waiver for all uploaded datasets

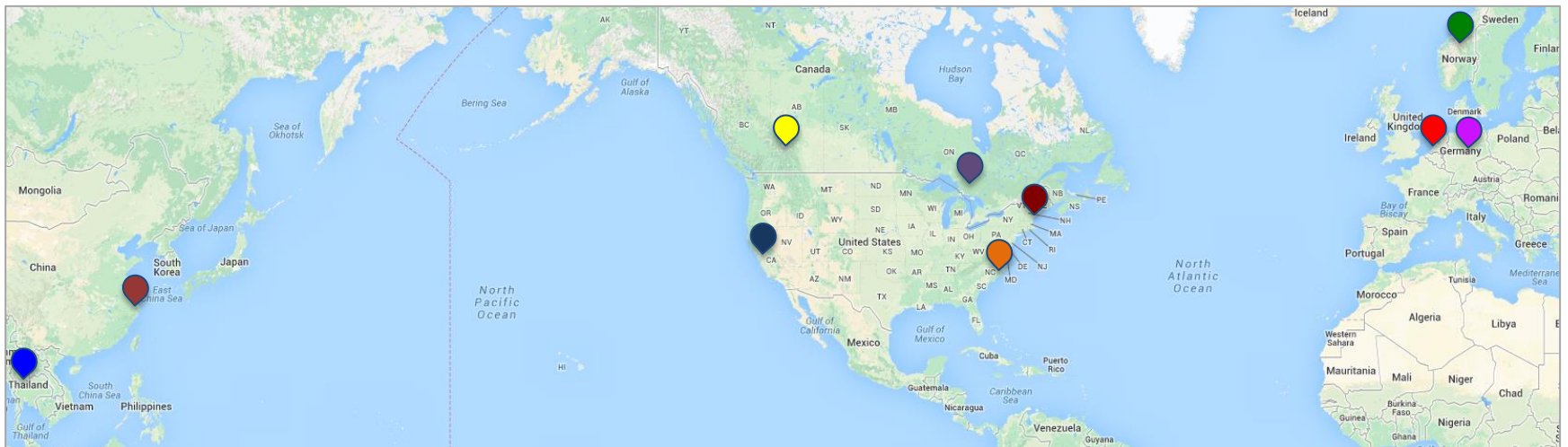
Dataverse Milestones



Who uses Dataverse?

- Researchers
- Librarians
- Data Archivists
- Journals
- Courses
- Institutions and Organizations

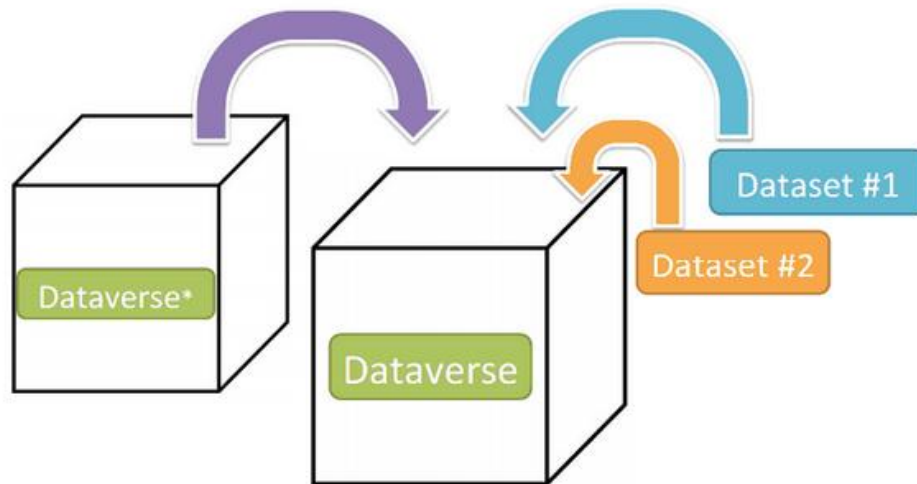
Dataverse Around the World



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?

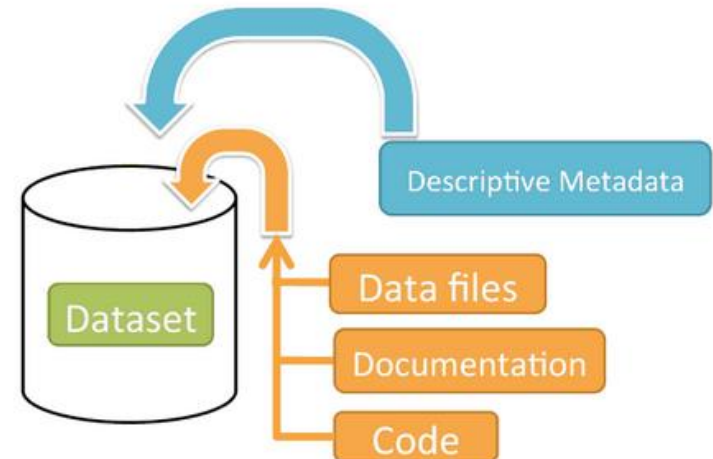
Schematic Diagram of a **Dataverse** in Dataverse 4.0



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.

Harvard Dataverse

- Dataverse installation run at Harvard University

The screenshot displays the Harvard Dataverse website. At the top, the 'Dataverse' logo is on the left, and navigation links for 'About', 'Support', 'Contact', 'Sign Up', and 'Log In' are on the right. Below the navigation bar, the Harvard University crest is shown next to the text 'Harvard Dataverse' and a subtitle: 'A collaboration with Harvard Library, Harvard University IT, and IQSS'. A 'Metrics' bar indicates '1,232,230 Downloads'. Below this, four partner logos are featured: World Agroforestry Centre, Population Services International (PSI), International Food Policy Research Institute (IFPRI), and the Henry A. Murray Research Archive. A large grey banner with the URL dataverse.harvard.edu is overlaid across the middle. Below the banner, a search bar is visible. The main content area shows search results for '1 to 10 of 58,761 results'. On the left, filters for 'Dataverses (991)', 'Datasets (57,770)', and 'Files (268,009)' are listed, along with 'Dataverse Category' and 'Affiliation'. The first search result is 'Raw IAT Data - FINAL APRIL2015' by Bernard Groen, dated April 13, 2015. The result includes a document icon, a citation snippet, and a description: 'This is the final IAT data set for healthcare and social care participants. Doctoral study at Durham University experiment 5/5'. A second result, 'Complex Integration: Status Inequalities As a Hindrance to Successful Integration', is partially visible below.

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

1,232,230 Downloads

World Agroforestry Centre - Population Services International International Food Policy Murray Research Archive Original Collection Dataverse

dataverse.harvard.edu

Search this dataverse... Find Advanced Search Add Data

☒ **Dataverses (991)**
☒ **Datasets (57,770)**
☐ **Files (268,009)**

Dataverse Category
Organization or Institution (22)
Journal (13)
Researcher (3)
Teaching Course (1)

Affiliation
Harvard (4,335)

1 to 10 of 58,761 results

Sort < > < Previous 1 2 3 4 5 Next >

Raw IAT Data - FINAL APRIL2015
Apr 13, 2015 - Bernard Groen Dataverse

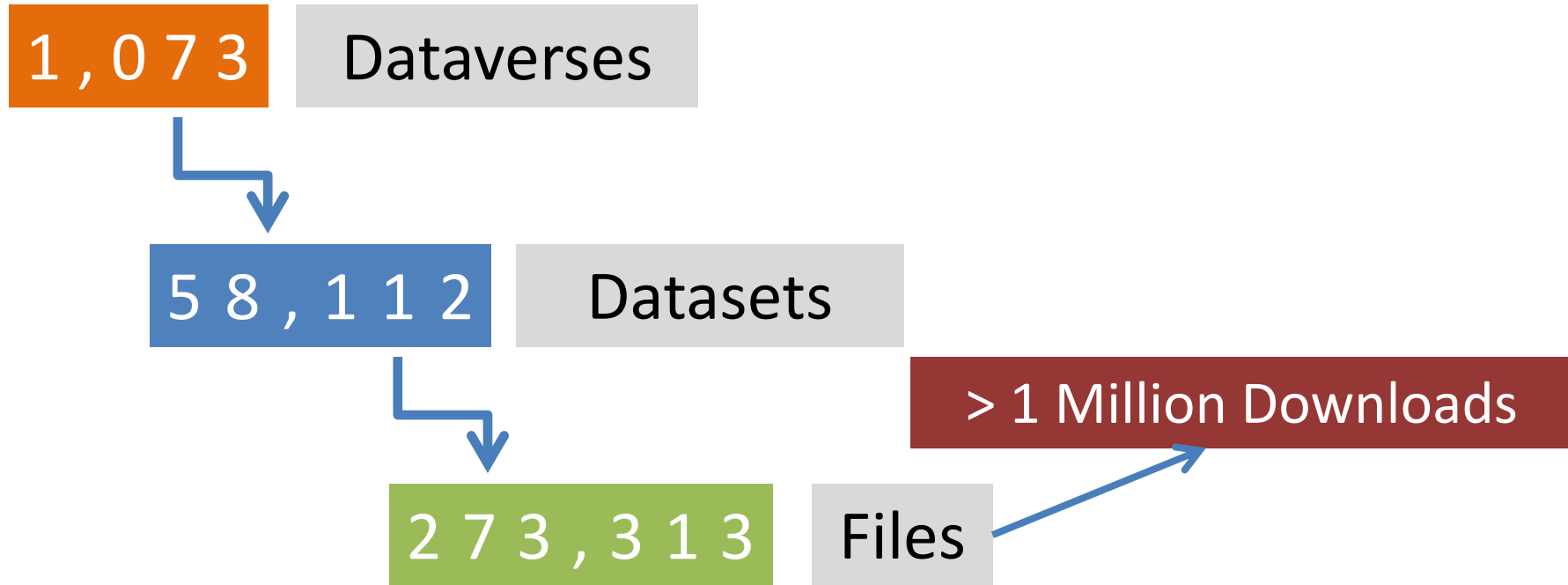
Groen, Bernard, 2015, "Raw IAT Data - FINAL APRIL2015", <http://dx.doi.org/10.7910/DVN/C4Z9F0>, Harvard Dataverse, V1 [UNF:6.lgCG3iYILcB54ag3QYxLwg==]

This is the final IAT data set for healthcare and social care participants. Doctoral study at Durham University experiment 5/5

Complex Integration: Status Inequalities As a Hindrance to Successful Integration
Apr 13, 2015 - Bernard Groen Dataverse

Harvard Dataverse

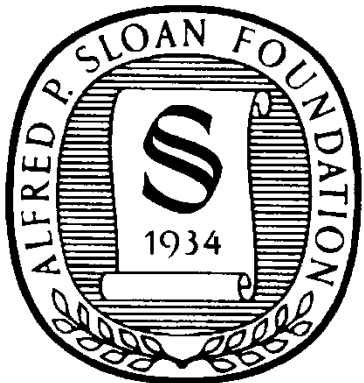
Open to all repository instance at Harvard currently has:



*number from June 2, 2015

Dataverse + OSF

- OSF Add-on uses Dataverse SWORD API for Data Deposit
 - [SWORD API guide](#)
 - projects.iq.harvard.edu/ojs-dvn/home
- June 9th: Common Models and APIs for Data Publishing and Citation in Cambridge, MA



PKP
PUBLIC
KNOWLEDGE
PROJECT

Dataverse + OSF

CenterForOpenScience / osf.io

Watch 27

Star 106

Fork

[Production & Staging] Dataverse: Deleting a Published version of a file is allowed, but not really. #2541

New Issue

Closed nicipfeiffer opened this issue on Apr 21 · 5 comments



nicipfeiffer commented on Apr 21

Collaborator

Steps

1. Go to a project with dataverse configured
2. Click to view a published file that is also in draft version
3. Click on the delete button the files detail page
4. Confirm that you want to delete the file
5. Return to the Files tab
6. Look for the file in the grid under Published files

Expected

Completing the steps to delete the published version of a file should not be allowed since those are read-only. Returning to the file tab and seeing the file remain when the steps for deleting were completed with no errors.

Actual

By completing the steps with no errors, the user is led to believe that they in fact, deleted a published file. However, they only deleted the draft version of the file.

Labels

5 - pending review

bug: production

Milestone

No milestone

Assignee

No one assigned

Notifications

Subscribe

You're not receiving notifications from this thread.

3 participants



Ibanner commented on Apr 21

Collaborator

@nicipfeiffer After the steps above, is the Published file still visible? But not any longer the Draft file?



nicipfeiffer commented on Apr 21

Collaborator

@Ibanner Yes, the published file is still visible. The draft file is now deleted. The inconsistency, in my opinion, is that the user is completing the deleting action from the published file. I think this is misleading.



Ibanner commented on Apr 22

Collaborator

@rriebz Did your resolving commit here get merged into Production? Thanks.



rriebz commented on Apr 22

Collaborator

I'm bundling the commit with a few other changes, which includes a better distinction between draft/published files and some related permissions fixes. Still a WIP.

Dataverse + OSF

- Researcher working in OSF decides to use the Dataverse add on
- Work in OSF with ability to publish datasets in Dataverse without having to leave OSF
- Increases data discoverability for a researcher's dataset




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

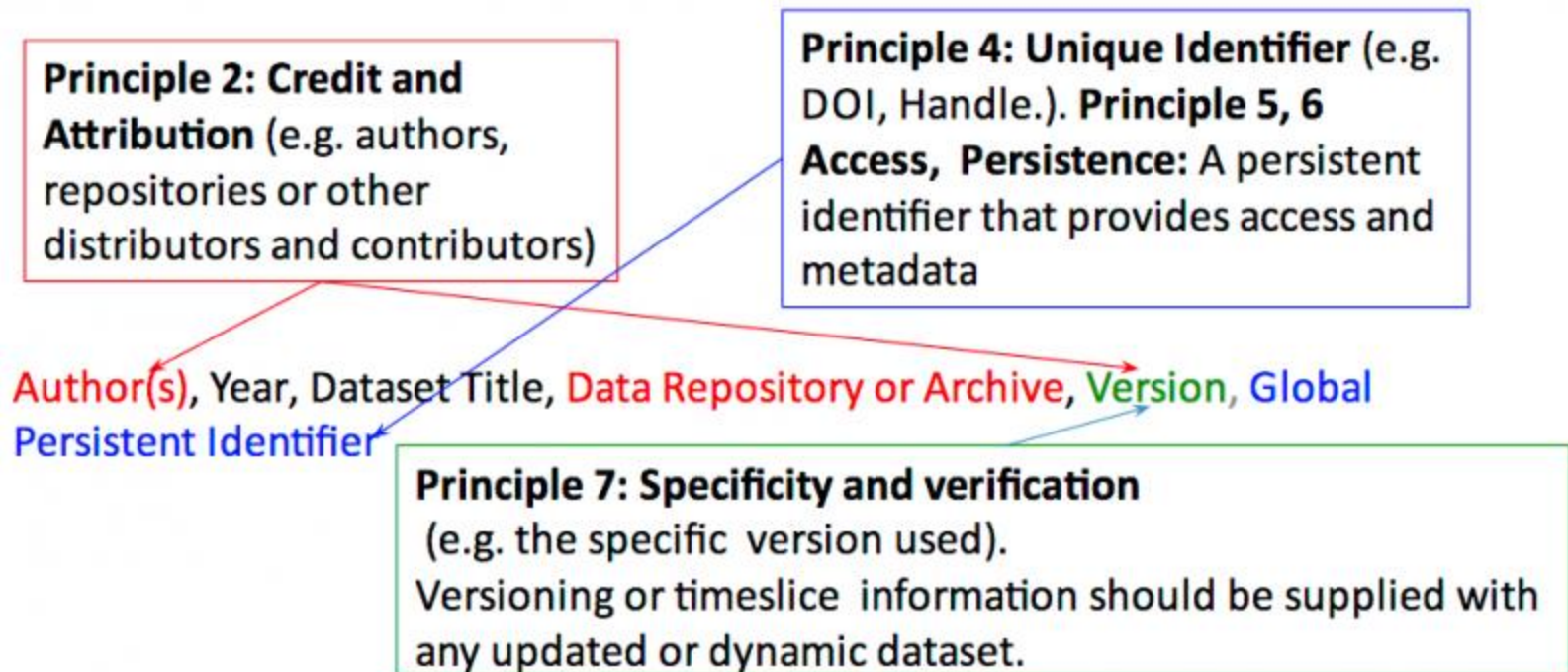
Dataverse Best Practices (2)

- Metadata is always public once a dataset is published
- By default, datasets receive CC0 Waiver 
- Even though default is CC0 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

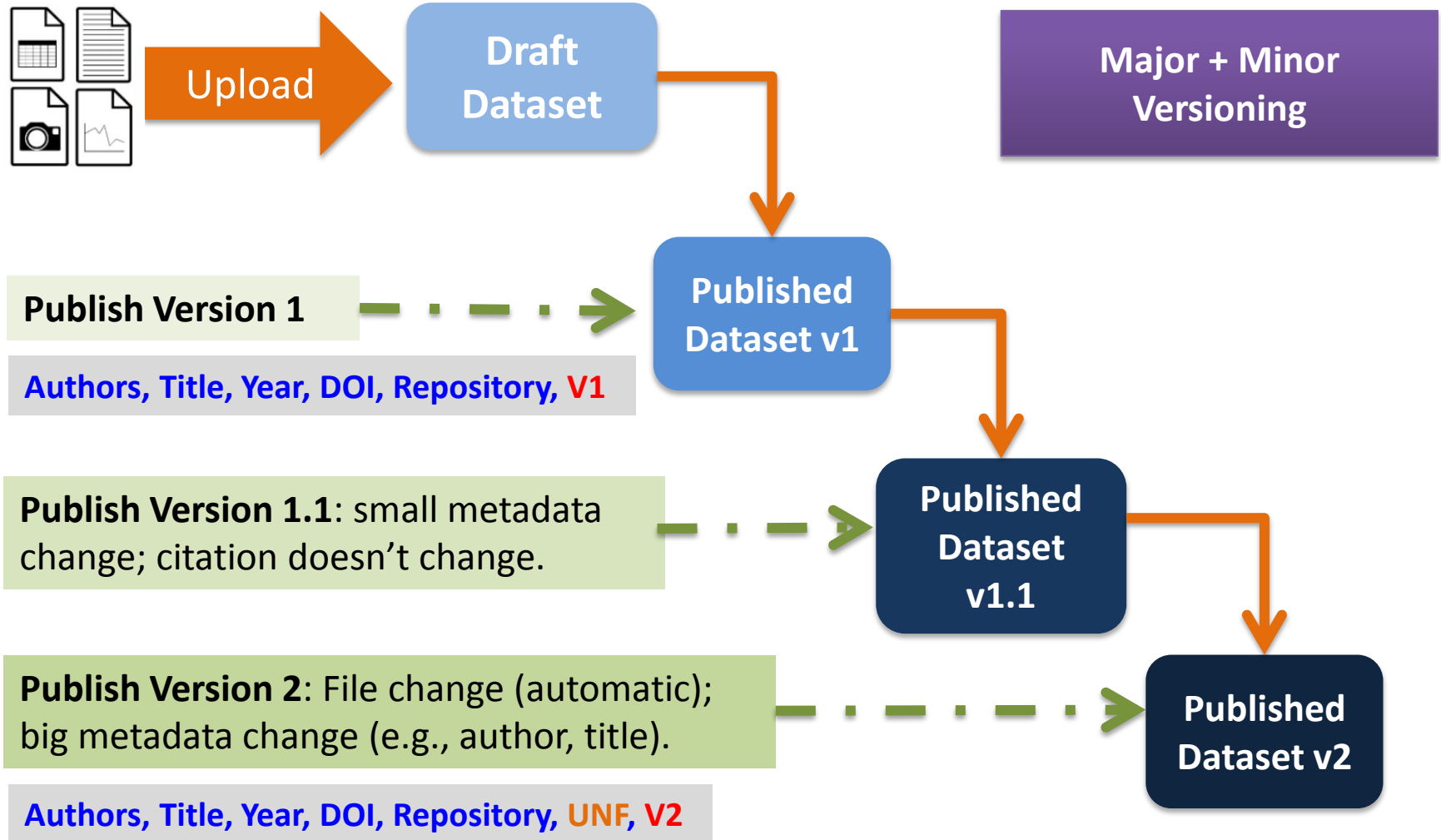
Dataverse Best Practices (5)

- 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




PUBLISHING WITH DATAVERSE

Rigorous Data Publishing Workflows



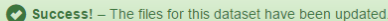
Publishing a Dataset

 **Dataverse** For testing only...





Search About Support Contact Elizabeth Quigley 2

TCDL Demo Dataverse (Harvard University) Unpublished


Demo Dataverse > TCDL Demo Dataverse > **TCDL Demo Dataset**



Metrics 0 Downloads

TCDL Demo Dataset Draft Unpublished

Quigley, Elizabeth, 2015, "TCDL Demo Dataset", <http://dx.doi.org/10.5072/FK2/C7SKKC>, Demo Dataverse, DRAFT VERSION
[UNF:6:/KWizLRBtSgASy2TgzB8xw==] 
If you use these data, please add this citation to your scholarly resources. [Why Cite?](#)



Description










Dataset created for demo purposes for Texas Conference on Digital Libraries.

Subject

Arts and Humanities

Files Metadata Terms Versions

 Download  Upload + Edit Files

	AK_2002.tab Tabular Data - 35.4 KB - Apr 21, 2015 - 0 Downloads Original File MD5: 4f82c6d25c7fbc90e7c16abfa92de1b; 18 Variables, 517 Observations - UNF:6:/KWizLRBtSgASy2TgzB8xw==	 Explore  Download
	HowToDeposit4-01.png PNG Image - 90.5 KB - Apr 21, 2015 - 0 Downloads MD5: 96104c15decdb3429cc586889e3cafb1;	 Download
	IQSSLogo-2014-Large (2).png PNG Image - 108.5 KB - Apr 21, 2015 - 0 Downloads MD5: 3db97461ea101a8200fa6fccdd01808;	 Download
	l1.avge.fits FITS - 7.7 MB - Apr 21, 2015 - 0 Downloads MD5: 811f4d104ce96d34e66b552e8c84af48; This is a FITS file with 1 (primary) HDU. The following recognized metadata keys have been found in the FITS file: INSTRUME; NAXIS0; NAXIS1; TELESCOP; DATE-OBS; CRVAL2; NAXIS; OBJECT; CD1_1; CRVAL1; EXPTIME;	 Download

Data Science at The Institute for Quantitative and Social Science | Dataverse Project on  | Code available at 
Copyright © 2015, The President & Fellows of Harvard College | [Privacy Policy](#)

Powered by  **Dataverse** Project v. 4.0 build 1042

Publishing Dataverse + Dataset

The screenshot displays the Dataverse web interface. At the top, the header includes the Dataverse logo, a 'For testing only...' badge, and navigation links for search, about, support, contact, and a user profile for Elizabeth Quigley. The main content area shows the breadcrumb path: Demo Dataverse > TCDL Demo Dataverse > TCDL Demo Dataset. A green success message states: 'Success! – The files for this dataset have been updated.' Below this, a modal dialog titled 'Publish Dataset' is open. The dialog contains two messages: a warning that the dataset cannot be published until the parent dataverse is published, and an information message stating that once published, the dataset must remain published. Two buttons, 'Yes, Publish Both' and 'Cancel', are at the bottom of the dialog. The background interface shows the dataset's metrics (0 downloads), description, subject (Arts and Humanities), and tabs for files, metadata, terms, and versions.

Dataverse For testing only...

TCDL Demo Dataverse (Harvard University) Unpublished

Demo Dataverse > TCDL Demo Dataverse > TCDL Demo Dataset

Success! – The files for this dataset have been updated.

Metrics 0 Downloads

TCDL Demo Dataset

Quigley, Elizabeth, 2015, "TCDL Demo Dataset" [UNF:6/KWizLRBtSgA5Y2...]
If you use these data, please add the following citation:

Description Dataset created for demo purposes for Texas Conference on Digital Libraries.

Subject Arts and Humanities

Files Metadata Terms Versions

Download Download Citation

Download Upload + Edit Files

Publish Dataset

⚠ This dataset cannot be published until TCDL Demo Dataverse is published. Would you like to publish both right now?

ℹ Once you publish this dataset it must remain published.

Yes, Publish Both Cancel

Publishing a Dataset

The screenshot shows the Dataverse web interface. At the top, the header includes the Dataverse logo, a "For testing only..." badge, and navigation links for search, about, support, contact, and a user profile for Elizabeth Quigley. The main content area displays the breadcrumb "TCDL Demo Dataverse (Harvard University)" and a success message: "Success! – This dataset has been created." Below this, a metrics bar shows "0 Downloads". The dataset title "TCDL Demo Dataset 2" is followed by a "Draft" status tag. A modal dialog titled "Publish Dataset" is open, displaying a warning icon and the text: "Are you sure you want to publish this dataset? Once you do so it must remain published." The dialog has "Continue" and "Cancel" buttons. In the background, the dataset page shows a description, subject "Arts and Humanities", tabs for "Files", "Metadata", "Terms", and "Versions", and a button to "Upload + Edit Files". A message at the bottom states "There are no files in this dataset."

Dataverse For testing only...

TCDL Demo Dataverse (Harvard University)

Demo Dataverse > TCDL Demo Dataverse > TCDL Demo Dataset 2

Success! – This dataset has been created.

0 Downloads

TCDL Demo Dataset 2 Draft

Quigley, Elizabeth, 2015, "TCDL Demo Dataset 2"
If you use these data, please add this citation to your work.

Description

Subject Arts and Humanities

Files Metadata Terms Versions

There are no files in this dataset.

Publish Dataset

Are you sure you want to publish this dataset? Once you do so it must remain published.

Continue Cancel

Download Citation

Upload + Edit Files

Want to learn more?

Friday at 1:15pm in Blegen Hall 155

Dataverse, a repository framework for all
Panel Discussion on how Dataverse is used
at different universities and institutions

Thank you!

Contact: equigley@iq.harvard.edu

Learn more: 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).