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

1999-2006: Virtual Data Center (VDC)

2006: Coding of Dataverse begins (initial focus on Social Sciences data)

2011: Expanded to include Astronomy & Astrophysics data

Fall 2012: OJS & Dataverse API Integration

March 2012: Dataverse 3.0 Released

April-May 2013: First Usability Testing of Dataverse

2013: Expanded to include Biomedical data

October 2013: Dataverse 4.0 development begins

October 2013: User centered design process integrated

October 2013-April 2015 Dataverse 4.0 Development

April 2015: Dataverse 4.0 Launched dataverse.harvard.edu
Who uses Dataverse?

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

Open to all repository instance at Harvard currently has:

1,073 Dataverses
58,112 Datasets
273,313 Files
> 1 Million Downloads

*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
Dataverse + OSF

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

Closed necopter opened this issue on Apr 21 · 5 comments

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 readonly. 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.

@banner commented on Apr 21

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

@banner commented on Apr 21

@necopter 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.

@banner commented on Apr 22

@tiebe Did you resolving commit here get merged into Production? Thanks.

@tiebe commented on Apr 22

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

Principle 2: Credit and Attribution (e.g. authors, repositories or other distributors and contributors)

Principle 4: Unique Identifier (e.g. DOI, Handle.). Principle 5, 6
Access, Persistence: A persistent identifier that provides access and metadata

Author(s), Year, Dataset Title, Data Repository or Archive, Version, Global Persistent Identifier

Principle 7: Specificity and verification
(e.g. the specific version used).
Versioning or timeslice information should be supplied with any updated or dynamic dataset.
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
• 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

Upload

Draft Dataset

Major + Minor Versioning

Published Dataset v1

Published Dataset v1.1

Published Dataset v2

Publish Version 1

Authors, Title, Year, DOI, Repository, V1

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

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

Authors, Title, Year, DOI, Repository, UNF, V2
Publishing a Dataset
Publishing Dataverse + Dataset

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

- Yes, Publish Both
- Cancel
Publishing a Dataset
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