

The Dataverse Project



- Science requires community access to data
- Free, open source software for archiving, sharing, referencing, extracting, and analyzing research data, dataverse.org
- Partnership with the Harvard Library and Harvard University IT to provide a community data repository across sciences and humanities, dataverse.harvard.edu:
 - 1000 published dataverses (by hundreds of universities)
 - 57,791 published datasets (by thousands of scholars)
 - 1,235,658 downloads

Publishing Datasets in Dataverse, even when Data are Confidential

What we support now:

- Datasets published in Dataverse get a data citation, with a DOI and attribution
- Metadata describing the datasets are always public, but files in dataset can be public or restricted
- Datasets are given a CC0 waiver as default
- BUT, CC0 can be changed if special terms of use apply to the data

What we are building:

- A solution to publish, query and analyze confidential data:
 - DataTags
 - Privacy-preserving tools
- As part of a collaboration with Harvard SEAS, Berkman center, IQSS (including the Data Privacy Lab), and MIT

Tag Type	Description	Storage & Transit	Access
Blue	Non-confidential information, stored and shared freely.	Clear	Open
Green	Not harmful personal information, shared with some access control.	Clear	Email, OAuth verified registration
Yellow	Potentially harmful personal information, shared with loosely verified and/or approved recipients.	Encrypted	Password, Registered , Approval click-through DUA
Orange	Sensitive personal information, shared with verified and/or approved recipients under agreement.	Encrypted	Password, Registered, Approval, signed DUA
Red	Very sensitive personal information, shared with strong verification of approved recipients under signed agreement.	Encrypted	Two-factor Auth, Registered, Approval, signed DUA
Crimson	Maximum sensitive , explicit permission for each transaction, strong verification of approved recipients under signed agreement.	Double Encrypted	Two-factor Auth, Registered, Approval, signed DUA

