What is Dataverse?

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

Provides incentives for researchers to share:
- Recognition & credit via data citations
- Control over data & branding
- Fulfill journal data availability and funder requirements.

Harvard Dataverse (open to all; repository instance at Harvard) currently has:

- **1,442** Dataverses
- **60,194** Datasets
- **300,284** Files

> 1.5 Million Downloads

What we have done/are doing.

- Add provenance as a new kind of file metadata.
- Provenance schema interoperable with w3 PROV standard.
- Integrate provenance schema in Dataverse.
- Represent provenance as computation that can produce the artifact.
- Represent provenance data as a graph, explicitly revealing the relationship between inputs and outputs (using DOIs).
- Extend Dataverse versioning to match provenance changes.

A new kind of metadata

Provenance Browser

http://projects.iq.harvard.edu/dataprovenance