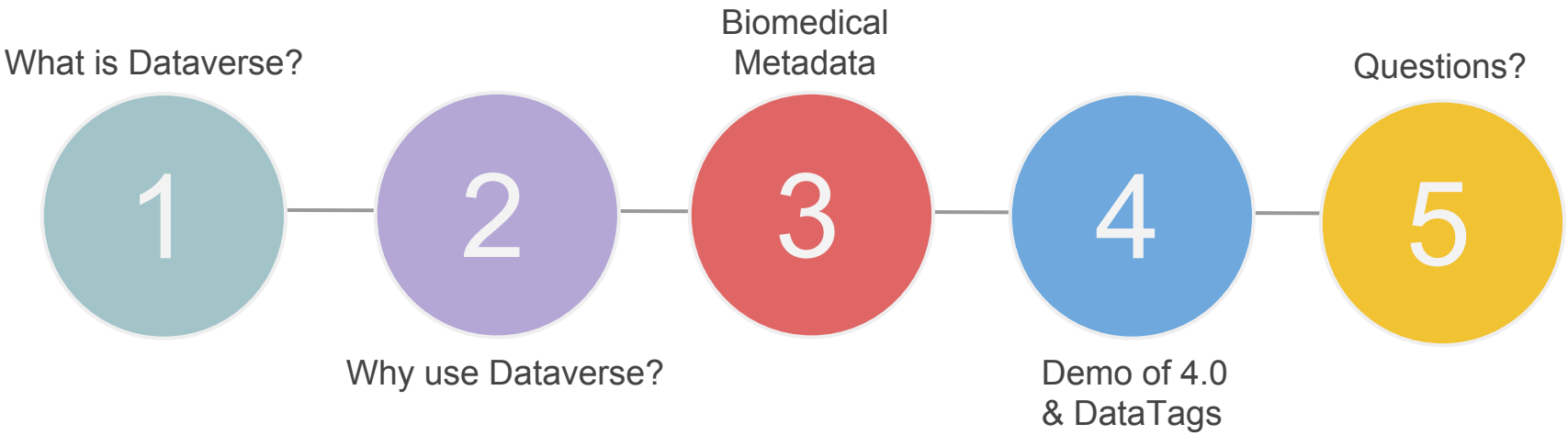
The Dataverse Network™ Project

DATAVERSE 101: For Data Discovery Group

Elizabeth Quigley & Eleni Castro
Countway Library of Medicine - May 8, 2014

Agenda

The Dataverse Network™ Project



What is Dataverse?

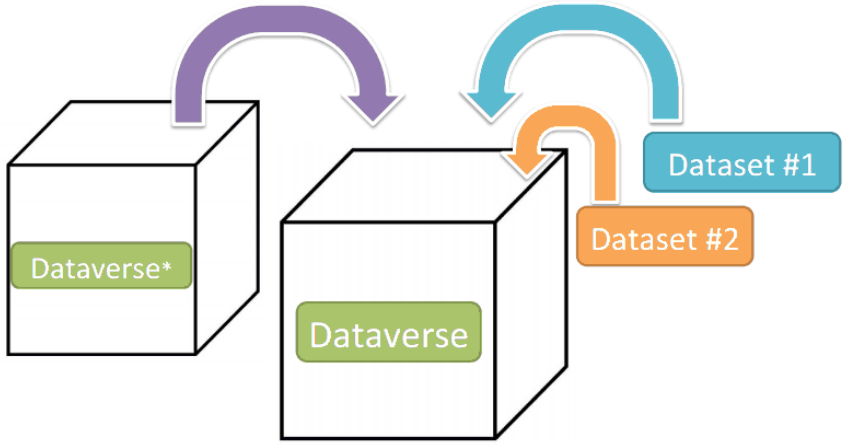
The Dataverse Network™ Project

A repository for sharing, citing, analyzing, and preserving research data.

What is a dataverse?

The Dataverse Network™ Project

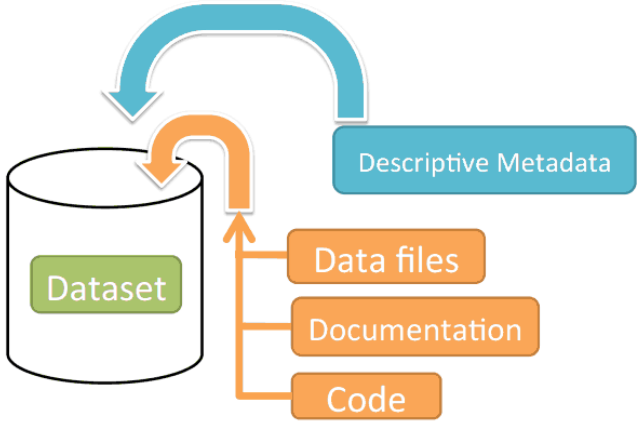
Schematic Diagram of a **Dataverse** in Dataverse 4.0



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.

Why use Dataverse?

The Dataverse Network™ Project

Upload datasets and files:

- Receive a formal data citation for a dataset and its files, with a persistent identifier (DOI).

- Provide as much descriptive metadata as possible:
  - for others to **discover** a dataset and its files
  - for others to **reuse** a dataset for their own research

Why use Dataverse?

The Dataverse Network™ Project

Find and use data:

- Search for dataverses, datasets, and files (incl. variables)
- Browse dataverses, datasets, and files
- Download data
- Subset & Analysis within Dataverse using Zelig statistical analysis software (R)

Why use Dataverse?

The Dataverse Network™ Project

Reasons to share data:

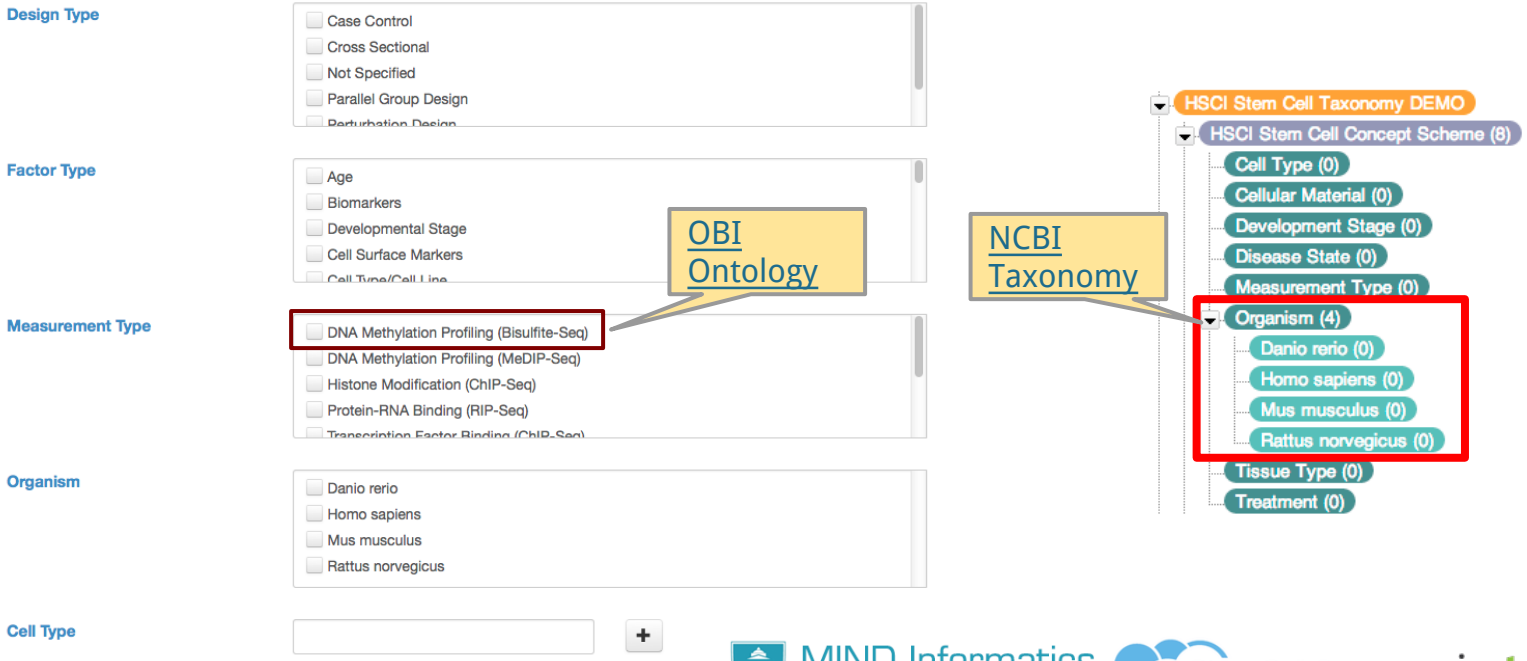
- Fulfill data management plan requirements (e.g., PLOS, NIH, NSF, etc).
- Get recognition and credit via data citations.
- Allow collaborators to contribute to your Dataverse (e.g, Journal, Project).
- Restrict data to your team until ready to publish
- Facilitate discovery and reuse of your data through extensive metadata.
- Enable reproducible research, which contributes to the validation and verification of science.
- ...

photo: Flickr Commons

Biomedical Metadata

The Dataverse Network™ Project

Initially working with Stem Cell Commons research metadata/ontologies for Proof of Concept.



MGH 1811 MIND Informatics MASSACHUSETTS GENERAL HOSPITAL

Stem Cell Commons

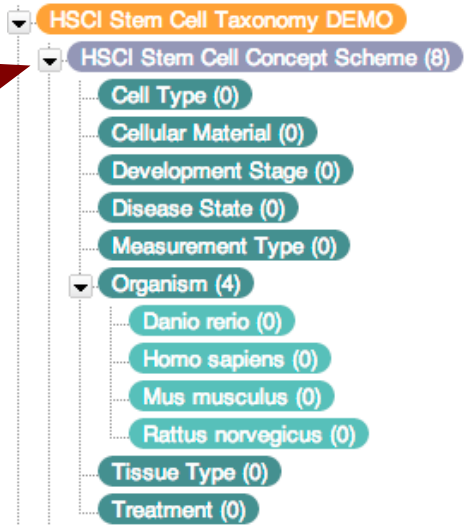
isatab specification

Bio Metadata Challenges

The Dataverse Network Project

How feasible is it to support a one-size-fits-all solution for biomedical metadata?

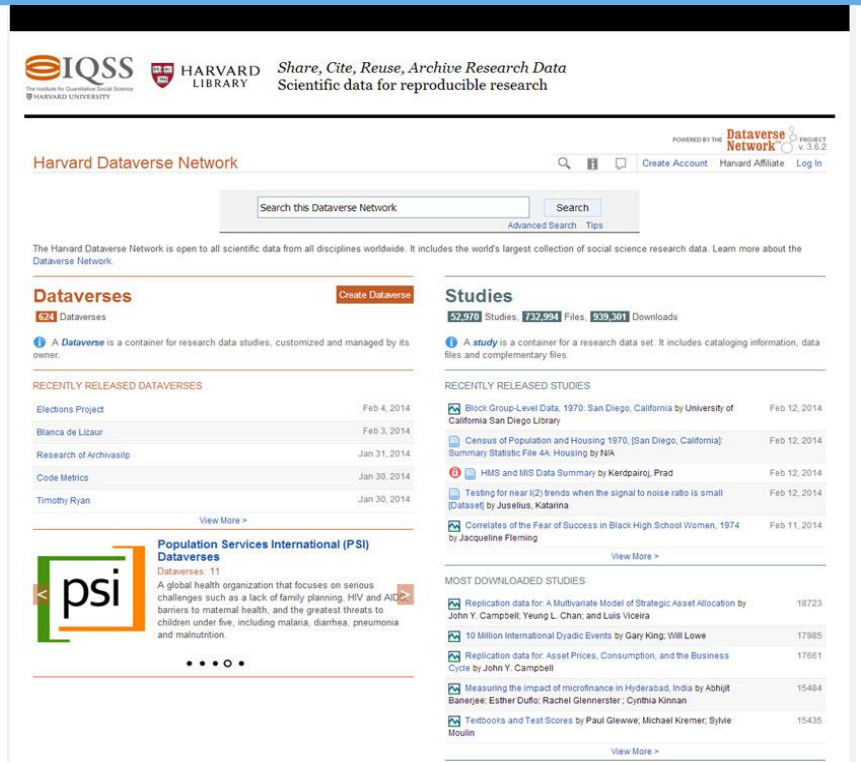
1. Template Approach: work w/ different groups to support their specific needs. (↑Time)
   (e.g., HSCI Stem Cell Commons, FAIRPORT, MIBBI (e.g., fMRI)).



2. For broader support allow free text (↓Quality) instead of just controlled vocabularies (e.g., OBO)

Transition from 3.6 to 4.0

The Dataverse Network™ Project



Dataverse v3.6

Transition from 3.6 to 4.0

The Dataverse Network™ Project

What's changing?

- User interface and simpler workflows
- Support for more metadata
- Account page redeveloped with new functionality
- New interactive data exploration and analysis tool

SolaFide

Let's check out Dataverse 4.0 Beta…..

dataverse-demo.iq.harvard.edu

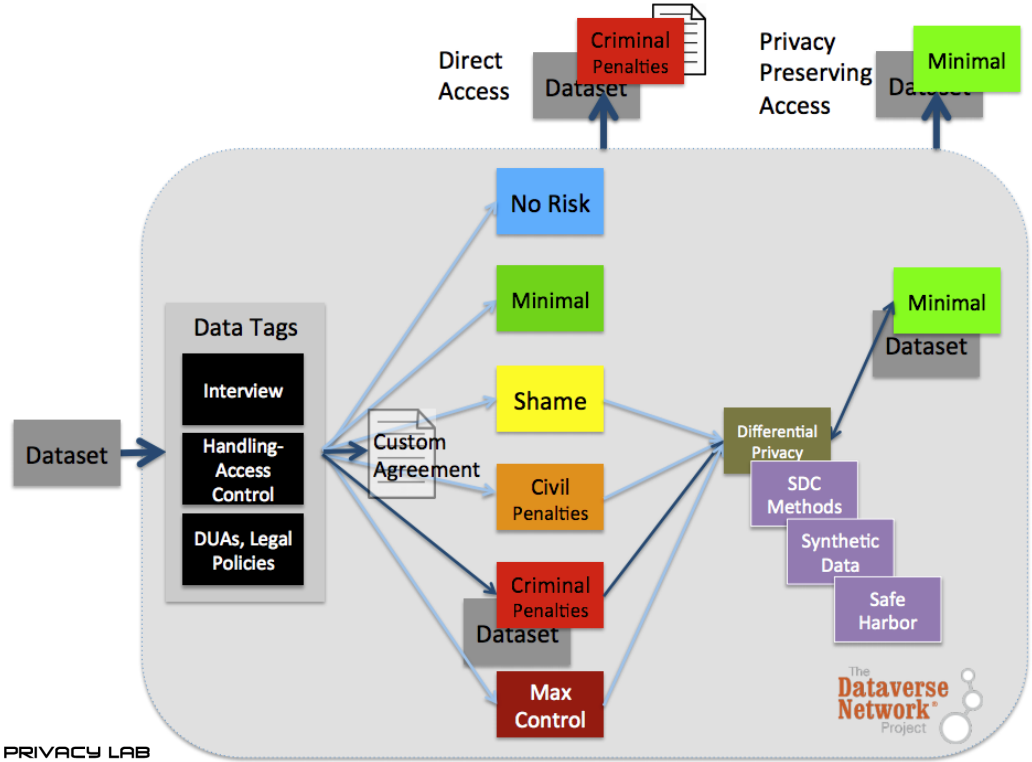
Sharing Privacy Sensitive Data

The Dataverse Network Project

Coming soon:

Tools for sharing privacy sensitive research data.

→ DataTags (demo)



NSF

Berkman
The Berkman Center for Internet & Society at Harvard University

DATA PRIVACY LAB

Questions?

The Dataverse Network™ Project

To learn more about Dataverse and the Data Science Team at IQSS, visit datascience.iq.harvard.edu

Elizabeth Quigley: equigley@iq.harvard.edu & Eleni Castro: ecastro@fas.harvard.edu