The Dataverse Project: An Open-Source Data Repository & Data Sharing Community

Elizabeth Quigley

The Software

Metadata

• Use standard-compliant metadata so metadata can be mapped easily to standard metadata schemas
• Supported metadata schemas:
  • DDI Lite
  • DDI 2.5 Codebook
  • Datacite 3.1
  • DCMI Metadata Terms
  • Dublin Core
  • VOResource Schema (Astronomy & Astrophysics)
  • ISA-Tab Specification (Life Sciences)
  • OBI Ontology (Life Sciences)
  • NCBI Taxonomy for Organisms (Life Sciences)

Preservation & Access

• CC0 waiver as the default
• Ability to add custom Terms of Use & Terms of Access
• Restricted Files with ability for anyone or no one to request access
• Tabular files reformatted into .tab format with variable level XML metadata (based on DDI)
• DOI persistent identifiers from DataCite and EZID
• Datasets cannot be unpublished, only deaccessioned
• Automatic version control for all published datasets

Future Work & Collaborations

• Expand metadata support to: Darwin Core, Biocaddie, schema.org, and Protein Data Bank
• Evaluating a JSON-Id schema for representing all dataset and file-level metadata
• Support for managing large file systems and sensitive data
• OAI-PMH for harvesting dataset metadata
• Integrations with:

The Dataverse Project

The Data Sharing Community

Dataverse repositories around the world

• Over 17 institutions using Dataverse software
  as their own data repository or to support a consortium of universities
• Over 65,000 published datasets
• More than 1,700,000 file downloads

Getting involved with the Dataverse Community:

• Biweekly Community Calls for updates on software development and working groups
• Annual Dataverse Community Meeting
  • July 11-12th
  • Privacy Workshop on July 13th
  • Learn more at projects.iq.harvard.edu/dcm2016
• Active Dataverse Community Google Group for asking questions about the software and providing feedback

Code: github.com/IQSS/dataverse

@dataverseorg | http://dataverse.org | Code: github.com/IQSS/dataverse