Dataverse: Helping Researchers Publish Their Data Through Automation

Eleni Castro, Research Coordinator
IQSS, Harvard University
IDCC 2016 - Feb 24, 2016
Helping Researchers Share & Archive Data At Their Point of Need
Our Quest For Interoperability and Automation

- **OAI-PMH** for harvesting metadata from Dataverse
- **SWORD API**: depositing metadata + data from a SWORD client into Dataverse
- **Search API**: searching dataverses, datasets and files within Dataverse
- **Data Access API**: downloading files from datasets found in Dataverse
- **Native API**: for performing GUI and super-user functionality programmatically via REST

In 2016: adding meta-tags and schema.org metadata for datasets

Research Life Cycle Workflow

- Planning
- Implementation
- Publishing
- Discovery & Impact
- Preservation

1. Planning Phase
Future Integration with DMP Tool

See: http://blog.dmptool.org/2016/01/22/dmptool-maintenance-and-a-roadmap
2. Implementation Phase
OSF Dataverse Add-On to archive data

via SWORD API

See: https://osf.io/getting-started/#dataverse
R package to deposit data & search Dataverse

Thomas Leeper’s code: https://github.com/rOpenSci/dvn
Data Visualizations from Dataverse...
Data Visualizations from Dataverse via WorldMap

http://worldmap.harvard.edu
Data Visualizations and Analysis with ClioInfra

via Data Access API + Native API

https://www.clio-infra.eu/
3. Publishing Phase
Integrate Journal and Data Publishing Workflows

Paper: http://journal.code4lib.org/articles/10989
Future: Integrate data quality review + verification

http://ajps.org/2015/03/26/the-ajps-replication-policy-innovations-and-revisions/
1. Allow users to authenticate using their ORCID ID.
2. Automatically insert ORCID ID into Dataset and search ORCID ID to insert for co-authors.
3. Add to and update ORCID records (Subject to permissions granted by iD holders).

See: Requiring ORCID in Publication Workflows: Open Letter
4. Discovery & Impact Phase
SHARE is building a free, open, data set about research and scholarly activities across their life cycle.

Explore SHARE’s database

Become a SHARE Notify Provider

http://www.share-research.org/projects/share-notify/
Send Dataset Metadata to DataCite

DataCite Metadata 3.0

Coming soon in Dataverse
Future: Measure Dataset Impact with Altmetrics

Example from Univ of Southampton

Example from Univ of Zurich

See Repository Badges documentation:
5. Preservation Phase
Scholars Portal Dataverse Integration With Archivematica

1.1 Ingest script issues API call for new or updated study
1.2 Dataverse returns json file for study
1.3 Ingest script issues API call for data files listed in json file
1.4 Dataverse returns all data files listed in json file
1.5 If bundled files are retrieved, ingest script unpackages the bundle(s)
1.6 Ingest script extracts metadata from json file
1.7 Ingest script generates METS file describing contents of transfer
1.8 Ingest script initiates ingest into Archivematica
1.9 Archivematica performs pre-configured processing micro-services
1.10 Archivematica generates AIP METS file incorporating content from transfer METS file
1.11 Archivematica packages AIP and places it in archival storage

Image source & read more: https://wiki.archivematica.org/Dataverse
Helping Future Researchers Re-Use Data
Thank You!
Questions?

ecastro@fas.harvard.edu