

Odum Institute iRODS Policies to Support Preservation NDSLabs Demonstration

Jonathan Crabtree

Reagan Moore

Don Sizemore

Craig Willis



National Science Foundation
Cooperative Agreement:
OCI-0940841

Project Goals

- Design curation/preservation workflow integration
- Connect research environment with archive
- Connect archive with national architecture
- Policy based
- Open source focused
- As “pluggable” as possible

Bringing great tools together

- iRODS
- Dataverse
- Modeshape
- BitCurator
- BagIt
- Preservation Audit Standards
 - ISO 16363
 - Data Seal of Approval
- Databook Architecture
 - Apache Service Mix
 - iRODS Rule Integration
 - Indexing Engine

Integration of the Dataverse

The screenshot shows the AJPS Dataverse website. At the top, the Dataverse logo is on the left, and navigation links (About, Guides, Support, Sign Up, Log In) are on the right. Below the header, the AJPS logo and name are prominently displayed. A sub-header identifies it as the 'American Journal of Political Science (AJPS) Dataverse' from Michigan State University, with the URL 'ajps.org'. A breadcrumb trail shows 'Harvard Dataverse > American Journal of Political Science (AJPS) Dataverse'. On the right, there are icons for email and sharing. A paragraph states the journal's commitment to significant advances in knowledge and understanding of citizenship, governance, and politics. Below this is a search bar with the placeholder 'Search this dataverse...', a 'Find' button, and a link to 'Advanced Search'. To the right of the search bar is a '+ Add Data' button. On the left side, there are filters for 'Dataverses (0)', 'Datasets (188)', and 'Files (1,178)'. Below these are filters for 'Publication Date' (2014 (66), 2013 (64), 2015 (31), 2012 (27)) and 'Author Name' (Hainmueller, Jens (4), Brookman, David (3)). The main content area shows '1 to 10 of 188 Results'. It includes a 'Sort' dropdown and pagination controls (Previous, 1, 2, 3, 4, 5, Next). Two dataset entries are visible: 1. 'Replication data for: The Externalities of Inequality: Fear of Crime and Preferences for Redistribution in Western Europe' by Rueda, David; Stegmueller, Daniel, dated Jun 4, 2015. The abstract discusses redistribution preferences between the rich and the poor. 2. 'Replication data for: Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities' by Coppock, Alexander; Green, Donald P., dated May 29, 2015.

Leveraging iRODS

- Automate preservation actions
- Rule based policy management
- Scalable storage infrastructure
- Enhances secure data sharing possibilities
- Diversifies Dataverse storage options

Preservation Policy Enforcement

- Odum Current Production Efforts
 - iRODS configured to utilize BitCurator
 - Testing sensitive information identification utilizing iRODS/BitCurator/BulkExtractor
 - Selected ISO 16363 policies implemented with iRODS rules
 - Rules published to GitHub
 - Utilizing irsync client/server method to push data from Odum Dataverse to iRODS and policies to archive data in DFC
 - Integrated staging iRODS server to traverse UNC Firewall

Next Steps

- Stress test sensitive information identification and valid file formats
- Add more complex sensitive information formats
- Expand iRODS rules based policy reporting
- Complete testing of iRODS/Modeshape/Dataverse integration
- Enable Dataverse/iRODS metadata lookup

Links

- <https://github.com/DICE-UNC/dfc-dataverse-integration>
- <https://github.com/DICE-UNC/indexing-irods>
- <https://github.com/akio-sone/dvn/tree/Odum-Ext>
- https://github.com/donsizemore/odum/tree/master/irods_rules
- <https://en.wikipedia.org/wiki/BagIt>
- <http://datafed.org>
- <http://www.odum.unc.edu/odum/home2.jsp>
- <http://Dataverse.org>
- <http://nationaldataservice.org>

Thanks

Demonstration of workflow and tools in NDSLabs