Weaving Together Preservation and Active Research

Sayeed Choudhury, Johns Hopkins University
Rick Johnson, University of Notre Dame
Jeffrey Spies, Center for Open Science
David Wilcox, DuraSpace
OSF

Workflow management, workflow integration
jeff@cos.io  @jeffspies
Incentives for individual success are focused on getting it published, not getting it right.

Nosek, Spies, & Motyl, 2012
Publish Report
Search / Discovery
Develop Idea
Design Study
Collect Data
Store Data
Analyze Data
Write Report
Reproducibility
We could . . . demonstrate that it makes research more efficient, of higher quality, and more accessible.
Better, we could . . . demonstrate that researchers will get published more often.
Even better, we could . . . make it easy.
Best, we could . . . make it automatic.
Preservation must be integrated rather than appended to research workflow.
Simplified scientific collaboration
Powerful end-to-end support for your research.

http://osf.io
Collaboration Documentation Archiving
Replication Studies

Study 3: Gupta et al. 2010, Nature

Contributors: Tim Errington, Elizabeth Iorns, William Gunn, Fraser Elisabeth Tan, Sarah Statt, Joelle Lomax, Nicole Perfido

Date Created: 2013-10-22 02:04 PM | Last Updated: 2015-01-20 06:16 PM

Category: Project

Wiki

This project contains all information pertaining to the replication of key experiments from this paper. It includes the detailed protocols, including reagents and author clarifications. We also include any comments from other contributors, researchers from the Science Exchange network, and further information with the original authors that we have learned since the beginning of the project. When experimental studies begin all data collected will also be deposited here, including data analysis...

Citation

osf.io/4bokd

Components

Coded Paper

Errington, Iorns, Gunn & 2 more

8 contributions

Recent Activity

All times displayed at -0700 UTC offset.

2015-01-20 06:16 PM Tim Errington added Nicole Perfido as contributor(s) to Project 3: Gupta et al. 2010, Nature...
## Revisions

<table>
<thead>
<tr>
<th>Version ID</th>
<th>Date</th>
<th>User</th>
<th>Download</th>
<th>MD5</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2015-08-17 01:05 PM</td>
<td>Sara Bowman</td>
<td>14</td>
<td>66518</td>
</tr>
<tr>
<td>3</td>
<td>2015-08-17 12:49 PM</td>
<td>Sara Bowman</td>
<td>0</td>
<td>53411</td>
</tr>
<tr>
<td>2</td>
<td>2015-08-17 12:32 PM</td>
<td>Sara Bowman</td>
<td>0</td>
<td>d6d969</td>
</tr>
<tr>
<td>1</td>
<td>2015-08-17 12:25 PM</td>
<td>Sara Bowman</td>
<td>0</td>
<td>122f1b</td>
</tr>
</tbody>
</table>
Other Features

- Granular privacy/sharing
- Granular permissions
- Analytics dashboards
- Persistent, citable identifiers
- Persistent content
- **Project snapshotting** (i.e., registration)
- Licensing
- Forking
Connects Services Researchers Use
<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component: Demo Add-Ons</td>
</tr>
<tr>
<td>GitHub: AndrewSallans/demofiles master d2e68a6246</td>
</tr>
<tr>
<td>ExamplePythonNotebook.ipynb</td>
</tr>
<tr>
<td>ExampleImage.jpg</td>
</tr>
<tr>
<td>ExampleCSV.csv</td>
</tr>
<tr>
<td>ExampleImage.png</td>
</tr>
<tr>
<td>ExampleSPSS.sav</td>
</tr>
<tr>
<td>ExamplePython.py</td>
</tr>
<tr>
<td>ExampleSpreadsheet.xlsx</td>
</tr>
<tr>
<td>ExamplePDF.pdf</td>
</tr>
<tr>
<td>ExampleR.r</td>
</tr>
<tr>
<td>ExampleWordDocument.docx</td>
</tr>
<tr>
<td>Amazon Simple Storage Service: osfdemofiles</td>
</tr>
<tr>
<td>FigShare: demofiles:892</td>
</tr>
<tr>
<td>Dropbox: /demofiles</td>
</tr>
<tr>
<td>ExampleImage.jpg</td>
</tr>
<tr>
<td>ExampleImage.png</td>
</tr>
<tr>
<td>ExamplePDF.pdf</td>
</tr>
<tr>
<td>ExamplePython.py</td>
</tr>
<tr>
<td>ExampleSPSS.sav</td>
</tr>
</tbody>
</table>
OSF Application Framework

- Workflow
- Authentication
- Permissions
- File Storage
- File Rendering
- Meta-database
- Persistence
- Integrations
- Search
- SHARE

osf.io

osf.io/preprints

osf.io/registries

journals

grants management

university systems

SHARE
Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency


Added on: August 29, 2016 | Last edited: August 29, 2016

Abstract

Beginning January 2014, Psychological Science gave authors the opportunity to signal open data and materials if they qualified for badges that accompanied published articles. Before badges, less than 3% of Psychological Science articles reported open data. After badges, 23% reported open data, with an accelerating trend; 39% reported open data in the first half of 2015, an increase of more than an order of magnitude from baseline. There was no change over time in the low rates of data sharing among comparison journals. Moreover, reporting openness does not guarantee openness. When badges were earned, reportedly available data were more likely to be actually available, correct, usable, and complete than when badges were not earned. Open materials also increased to a weaker degree, and there was more variability among comparison journals. Badges are simple, effective signals to promote open practices and improve presence of data and...
OSF Application Framework

- Workflow
- Authentication
- Permissions
- File Storage
- File Rendering
- Meta-database
- Persistence
- Integrations
- Search
- SHARE

osf.io

osf.io/preprints

osf.io/registries

journals

grants management

university systems
The Fedora Perspective
Fedora Facts

Managed by DuraSpace (not-for-profit)

Funded by the community

Collaboratively developed by the community

Supported by 2 full-time staff members (not developers)
Fedora...

Stores, preserves, and provides access to digital objects

Supports flexible and complex content models for objects

Supports complex semantic relationships between objects inside and outside the repository using RDF

Supports millions of objects, both large and small

Interoperates with other applications and services
Key Priorities

Aligning with standards

Pursuing integrations

Supporting research data management
Fedora + OSF

Integrates archiving and preservation with active research

Leverages Fedora’s native linked data functionality

Supports longer-term goal of open source solutions for research lifecycle
Community Support

Public wiki and issue tracker

Open meetings and mailing lists

Communication, outreach, marketing
Notre Dame
CurateND and Research Sharing
How Can We Address this Question?

“Can I start archiving my work while I am working?”

- ND College of Science Deans and Chairs, Fall 2014

If the answer is no...

- How do we get in front of the research data lifecycle early enough where we are not rushing/begging to get metadata created?
- How do we deal with the large volume and frenetic pace of data creation in computational analysis?
OSF for Institutions at Notre Dame

GENERATE & STORE

amazon web services

box

Dropbox

CRC

CENTER FOR RESEARCH COMPUTING
https://crc.nd.edu

COLLABORATE & PROCESS

ARCHIVE

Curate ND

REUSE

EXECUTE

http://www.nationaldataservice.org

PRESERVE & SHARE
OSF for Institutions at Notre Dame

Data Conservancy Packaging Tool

Amazon Web Services
Box
Dropbox
CRC

CurateND

EXECUTE
ARCHIVE
REUSE

http://www.nationaldataservice.org

GENERICATE & STORE
COLLABORATE & PROCESS
PRESERVE & SHARE

Center for Research Computing
https://crc.nd.edu
Getting Closer to Researchers

- Aligns IDR with existing research workflows
- Researchers choose working storage provider
- Simple, user-driven archival process
- Data flow between computational and preservation environments
- Reuse preserved data in other projects
- Archive data into Fedora 3 or Fedora 4
- Reference data model for preserving research data
rick.johnson@nd.edu

@rick_nd
Data Conservancy
The Data Conservancy (DC) was launched through a grant from the National Science Foundation’s DataNet program, which built upon prior experience with managing data from the Sloan Digital Sky Survey. The grant provided the DC team an opportunity to broaden its data infrastructure development and gain better understanding of the challenges in collecting, preserving and curating different types of research data. Since the DataNet funding, the Data Conservancy has redesigned and refactored its core infrastructure to leverage existing software and technologies and to build deeper connections with both research and technology communities. Most notably, we have embraced approach of data representation by the Linked Data Platform (LDP) by building our data archive with the Fedora 4 repository platform and leading the development of the RMap Services with funding from the Sloan Foundation.

**Packaging Specification**
- Based on popular BagIt specification
- Domain model agnostic
- Adds semantic information about content
- May be used with any RDF-based domain model

**Packaging Tool**
- A JavaFX point-and-click interface
- Produces DC-specification compliant packages
- Supports multiple domain models
- Allows semantic enrichment

**Package Ingest Service**
- Deposits package content into an archive
- Exposes content as linked data
- Fedora 4 is the current reference implementation of a DC archive

**Current DC Components**
- Packaging Specification
- Packaging tools
- Package Ingest Service
- Fedora data archive
- Fedora API-X framework
- RMap Services

**RMap Services**
- Protocol for Linked Data representations
- Developed through the RMap project
- Captures relationships between publication and underlying data
- Distributed Scholarly Compound Object (DISCO) protocol for resource aggregation
- OAI-ORE based
- REST APIs are available

**Fedora API-X**
- Extends core functionalities of a Fedora 4 repository
- Facilitate:
  - Mapping between domain specific data models and Fedora data model
  - Support for commonly used web-service standards
  - Domain specific federated discovery and access
  - Support advance data curation capabilities
RMap examples -- OSF project from SHARE
RMap Examples -- arXiv.org
RMap examples -- UPenn Repository
RMap examples -- Figshare
Resources

OSF

Fedora
- http://fedorarepository.org
- https://wiki.duraspace.org/display/FF

Notre Dame
- http://curate.nd.edu
- http://osf.nd.edu

Johns Hopkins
- http://dataconservancy.org/
- http://rmap-project.info/rmap/
Acknowledgements

COS
● cos.io/about_sponsors
● cos.io/about_team

Fedora
● http://fedorarepository.org/membership

Notre Dame
● Data and Software Preservation for Open Science, http://daspos.org
● Ian Taylor, Center for Research Computing, University of Notre Dame

Johns Hopkins
● NSF Datanet grant, Sloan Foundation grant for RMap
● Karen Hanson, Elliot Metsger, Mark Patton, Hanh Vu (and others on Data Conservancy team)
Questions?