

---

# Dataverse: Helping Researchers Publish Their Data Through Automation

— Eleni Castro, Research Coordinator  
IQSS, Harvard University  
IDCC 2016 - Feb 24, 2016 —



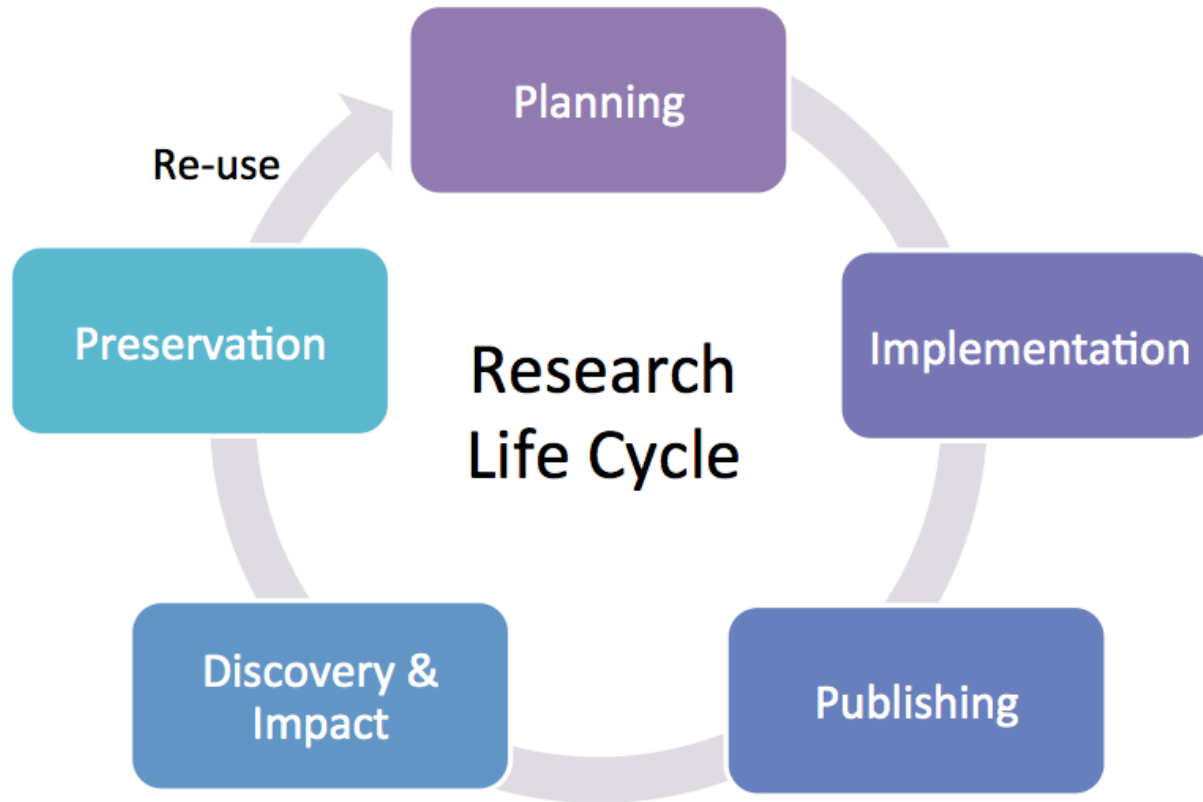
**Helping Researchers Share & Archive Data At Their Point of Need**

# 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

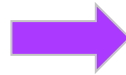
In 2016: adding meta-tags and schema.org metadata for datasets

# Research Life Cycle Workflow



# 1. Planning Phase

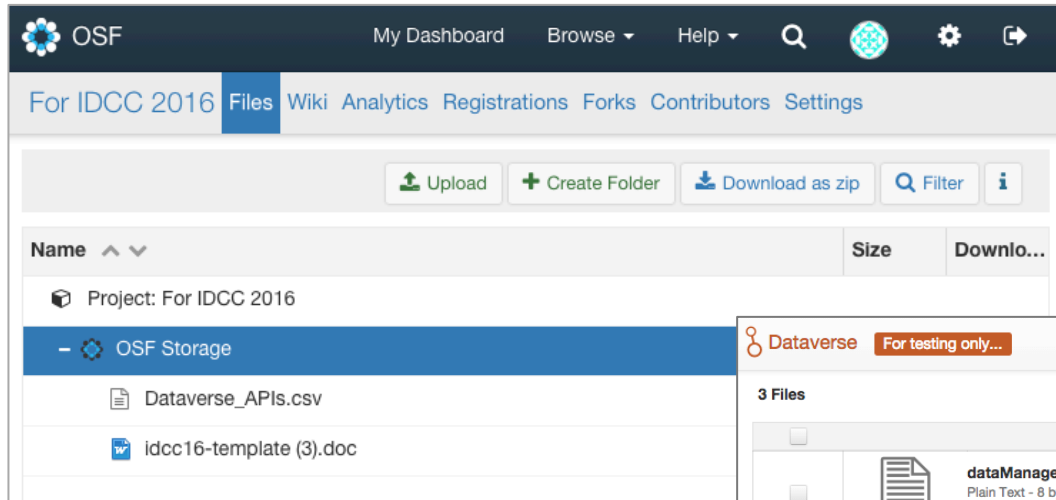
# Future Integration with

A screenshot of the DMPTool "Create New DMP" form. The header shows the DMPTool logo and a welcome message to Eleni Castro. The navigation bar includes links for Home, My Dashboard, DMP Requirements, Public DMPs, News, Help, Contact Us, About, and Log Out. The main content area shows the Harvard University logo and name, along with a "Contact Harvard DMPTool Support" link. Below this is a "DMP Overview" section with a form for creating a new DMP. The form includes fields for "DMP Template" (Alfred P. Sloan Foundation), "DMP Title" (For IDCC 2016), "Proposal Solicitation Number" (IDCC022416), and "Proposal Submission Deadline" (02/24/16). There are also radio buttons for "Visibility" (public, institutional, private) and a text field for "Add Co-owners". At the bottom are "Back" and "Save and Next >>" buttons.A screenshot of the Dataverse dataset view for "For IDCC 2016". The header shows the Dataverse logo, a "For testing only..." banner, and navigation links for About, Guides, Support, and a user profile for Eleni Castro. The main content area shows the dataset title "For IDCC 2016" with "Draft" and "Unpublished" status. Below the title is a citation for "Castro, Eleni, 2016, 'For IDCC 2016', http://dx.doi.org/10.5072/FK2/X4Z3ZF, Demo Dataverse, DRAFT VERSION". A "Download Citation" button is on the right. Below the citation is a description of the dataset: "Description: Test" and "Subject: Arts and Humanities". There are tabs for "Files", "Metadata", "Terms", and "Versions". A search bar is present with the text "Search this dataset...". Below the search bar is a list of files, with one file highlighted: "dataManagementPlan-Sloan.txt", which is a Plain Text file of 8 bytes, dated Feb 20, 2016, with 0 downloads. The MD5 hash is 1b63257ccffef6c29e14add016df707.

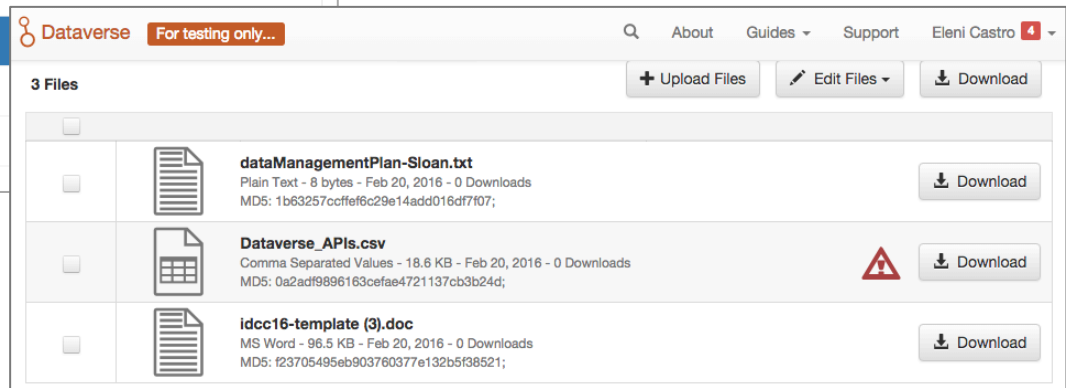
See: <http://blog.dmptool.org/2016/01/22/dmptool-maintenance-and-a-roadmap>

## 2. Implementation Phase

# OSF Dataverse Add-On to archive data



via SWORD API

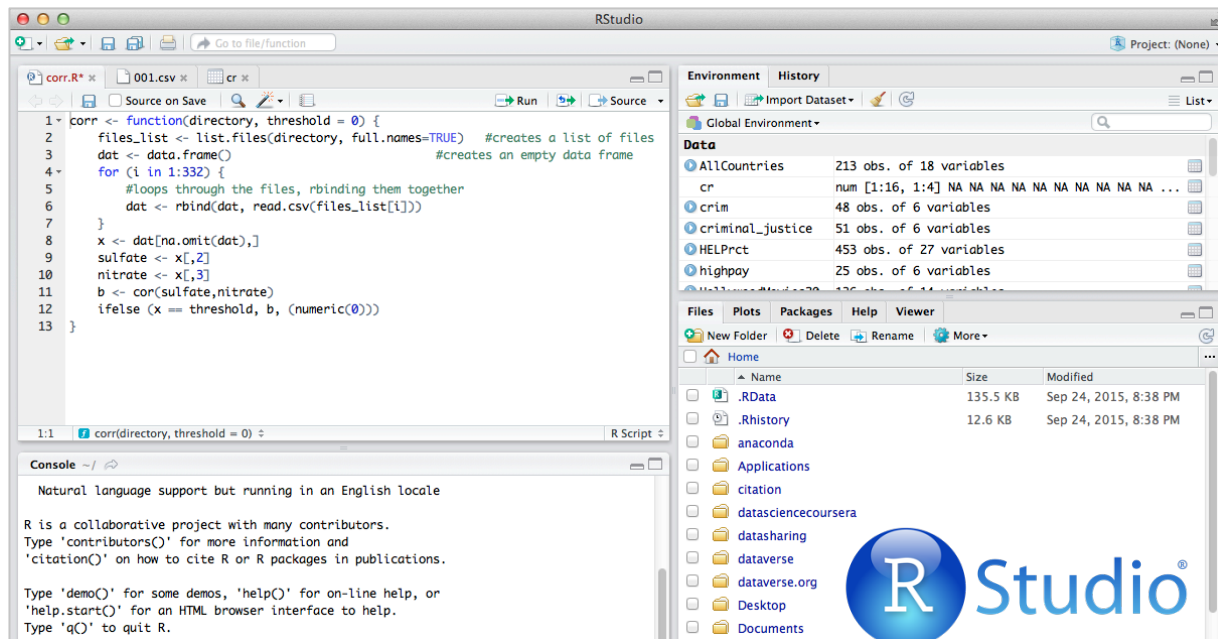


See: <https://osf.io/getting-started/#dataverse>




# R package to deposit data & search Dataverse

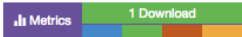
Thomas Leeper's code: <https://github.com/rOpenSci/dvn>





# Data Visualizations from Dataverse...

 Dataverse


Search About Guides - Support Sign Up Log In



## Fairfax County High School Attendance Areas

Government of Fairfax County, Virginia, 2015, "Fairfax County High School Attendance Areas", <http://dx.doi.org/10.7910/DVNE6H DUZ>, Harvard Dataverse, V1

 Download Citation -

If you use these data, please add this citation to your scholarly resources. Learn about [Data Citation Standards](#).


Description	Boundaries for Fairfax County High School Attendance
Subject	Other
Keyword	Fairfax County (Va.)--Maps


Files


Metadata


Terms


Versions


Search this dataset... 


2 Files 






**High\_School\_Attendance\_Areas.cpg**  
Plain Text - 5 bytes - Jul 16, 2015 - 0 Downloads  
MD5: ae3b3df9970b49b6523e608759bc957d;  
**fairfax; high school; boundary**



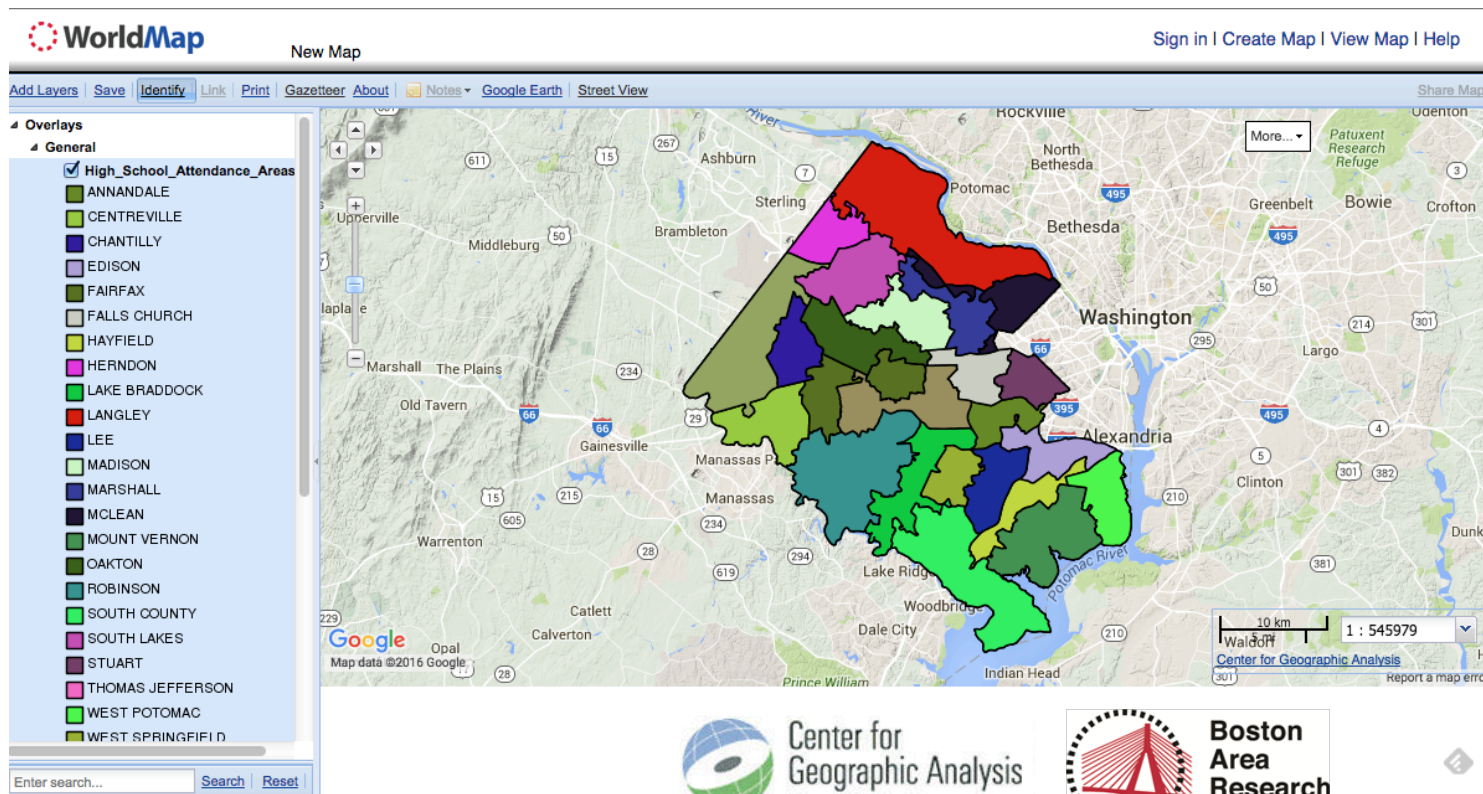


**High\_School\_Attendance\_Areas.zip**  
Shapefile as ZIP Archive - 1.5 MB - Jul 16, 2015 - 1 Download  
MD5: 82182337589405f6ce6d9af3d3cb95bd;  
**fairfax; high school; boundary**



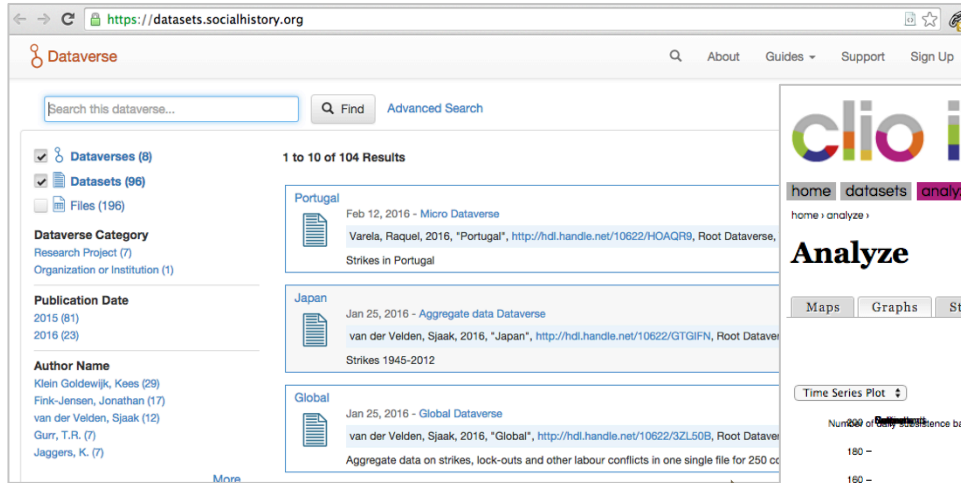
 Explore 

# Data Visualizations from Dataverse via WorldMap



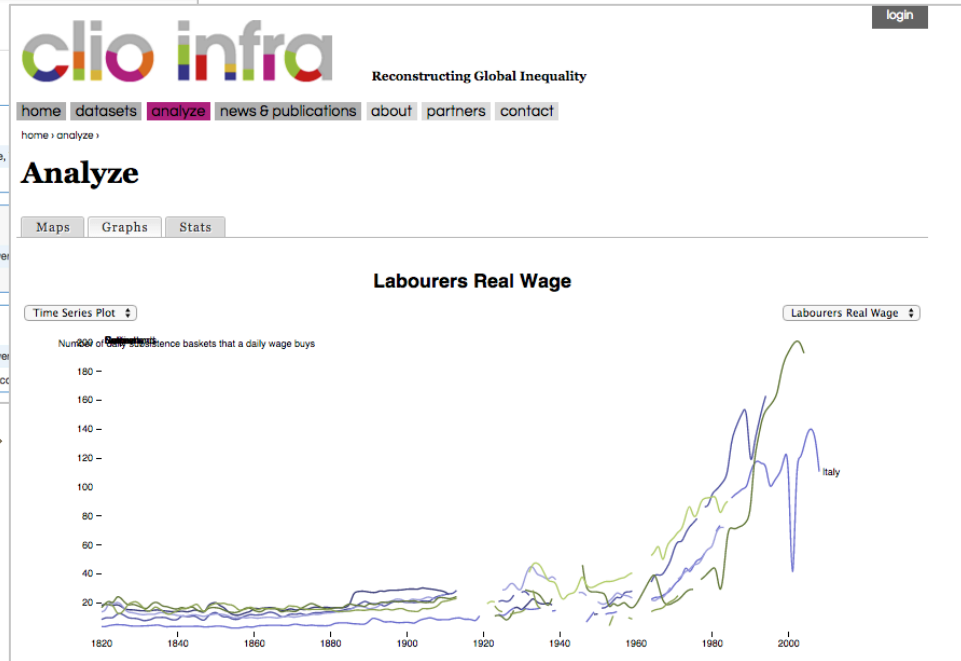
<http://worldmap.harvard.edu>

# Data Visualizations and Analysis with ClioInfra



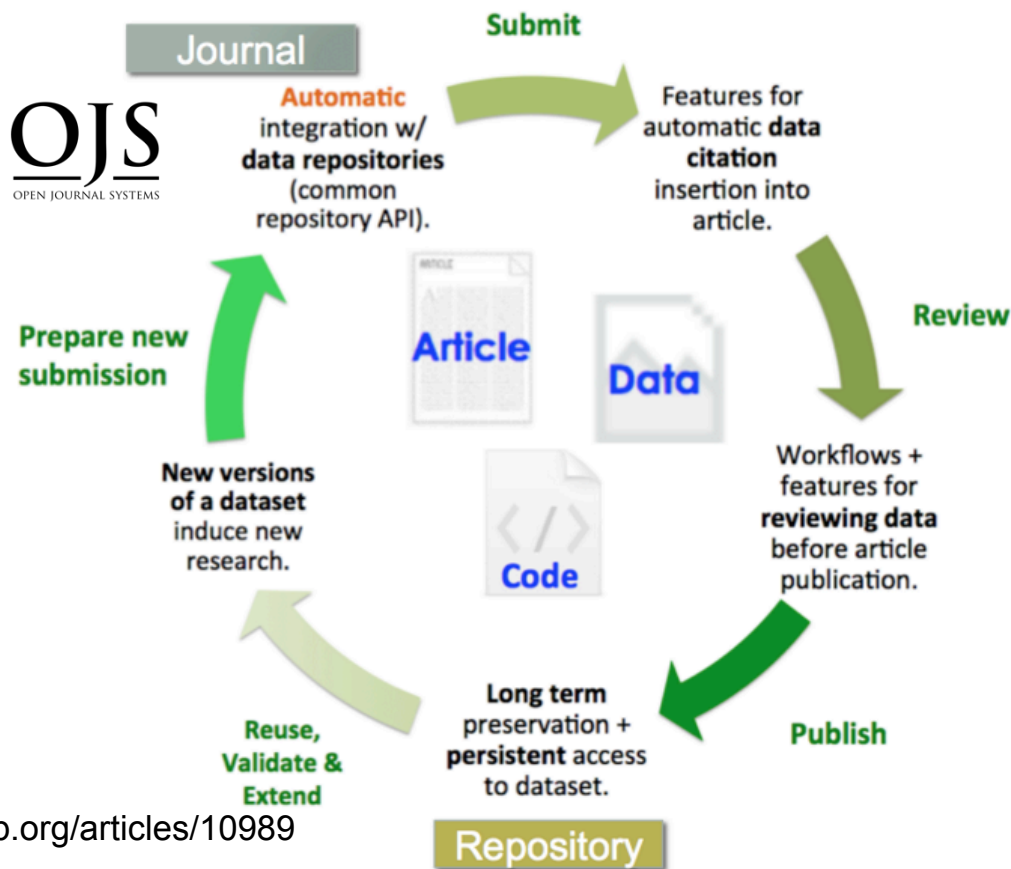
The screenshot shows the Dataverse website at <https://datasets.socialhistory.org>. The interface includes a search bar, navigation links (About, Guides, Support, Sign Up), and a sidebar with filters for Datasets (8), Files (196), and Datasets (96). The main content area displays search results for 'Portugal', 'Japan', and 'Global' datasets, each with a brief description and a link to the dataset.

via Data Access API  
+ Native API

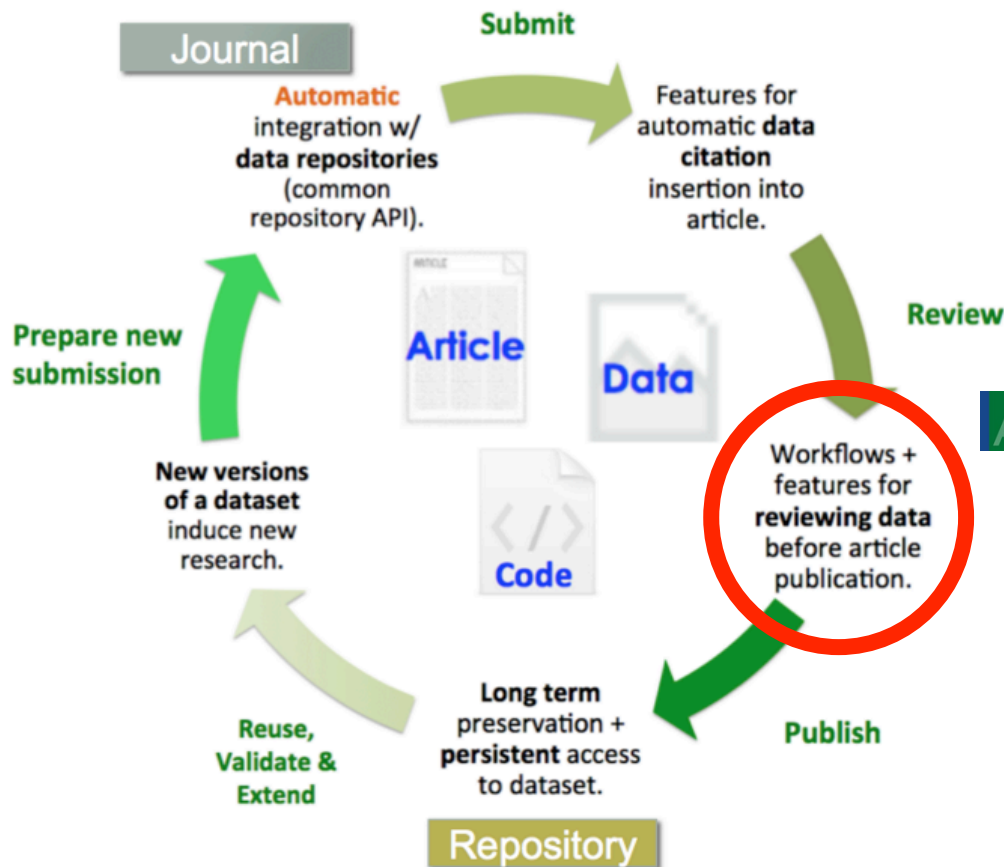


## 3. Publishing Phase

# Integrate Journal and Data Publishing Workflows



# Future: Integrate data quality review + verification



AJPS AMERICAN JOURNAL  
of POLITICAL SCIENCE



# Future: Dataverse / ORCID Integration



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

# Expand Dataset Discovery via SHARE Notify



▼ About SHARE

SHARE Notify

News

Knowledge Base

Contact

SHARE is building a free, open, data set about research and scholarly activities across their life cycle.

**Explore SHARE's  
database**

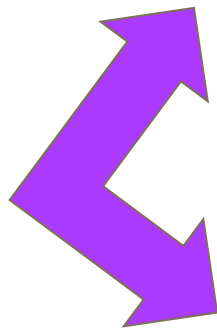
Become a SHARE Notify  
Provider

<http://www.share-research.org/projects/share-notify/>

# Send Dataset Metadata to DataCite

```
<?xml version="1.0" encoding="UTF-8"?>

<resource xmlns="http://datacite.org/schema/kernel-3"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://datacite.org/schema/kernel-3
  http://schema.datacite.org/meta/kernel-3/metadata.xsd">
  <identifier identifierType="DOI">10.5072/DVN/ANBT9E</identifier>
  <creators>
    <creator>
      <creatorName>Admin, Dataverse</creatorName>
      <nameIdentifier schemeURI="http://orcid.org/"
nameIdentifierScheme="ORCID">0001-0002-0003-0078</nameIdentifier>
      <affiliation>Dataverse.org</affiliation>
    </creator>
    <creator>
      <creatorName>Junior, Joey Jojo</creatorName>
      <nameIdentifier schemeURI="http://isni.org/isni/"
nameIdentifierScheme="ISNI">0001-0002-0003-3716</nameIdentifier>
      <affiliation>KMS</affiliation>
    </creator>
    <creator>
      <creatorName>Moony, Padfoot</creatorName>
      <nameIdentifier schemeURI="http://id.loc.gov/authorities/names/"
nameIdentifierScheme="LCNA">jseljtf1wERJFGLKS</nameIdentifier>
      <affiliation>Gryffindoor</affiliation>
    </creator>
  </creators>
  <titles>
    <title>Test DataCite Metadata - after merge</title>
  </titles>
  <publisher>Root Dataverse</publisher>
  <publicationYear>2016</publicationYear>
  <resourceType resourceTypeGeneral="Dataset"/>
  <descriptions>
    <description descriptionType="Abstract">Added description to DataCite
xml</description>
```



Long-term identifiers made easy

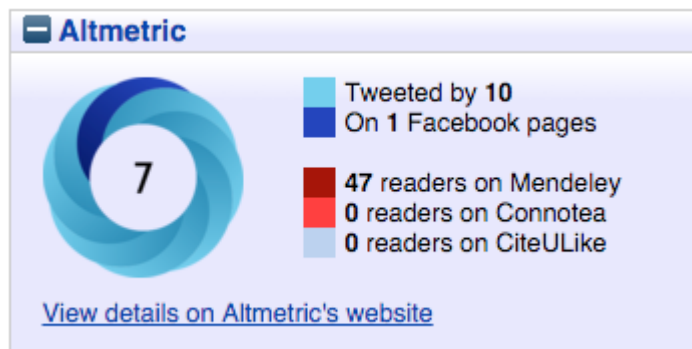
A SERVICE OF THE UNIVERSITY OF CALIFORNIA CURATION CENTER — UC3



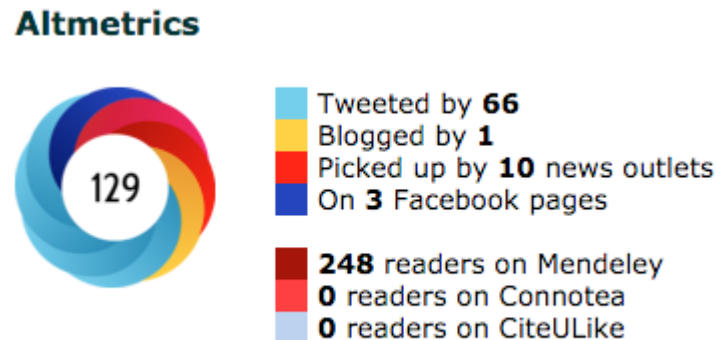
## Metadata Store

Coming soon in Dataverse

# Future: Measure Dataset Impact with Altmetrics



[Example from Univ of Southampton](#)



[View details on Altmetric's website](#)

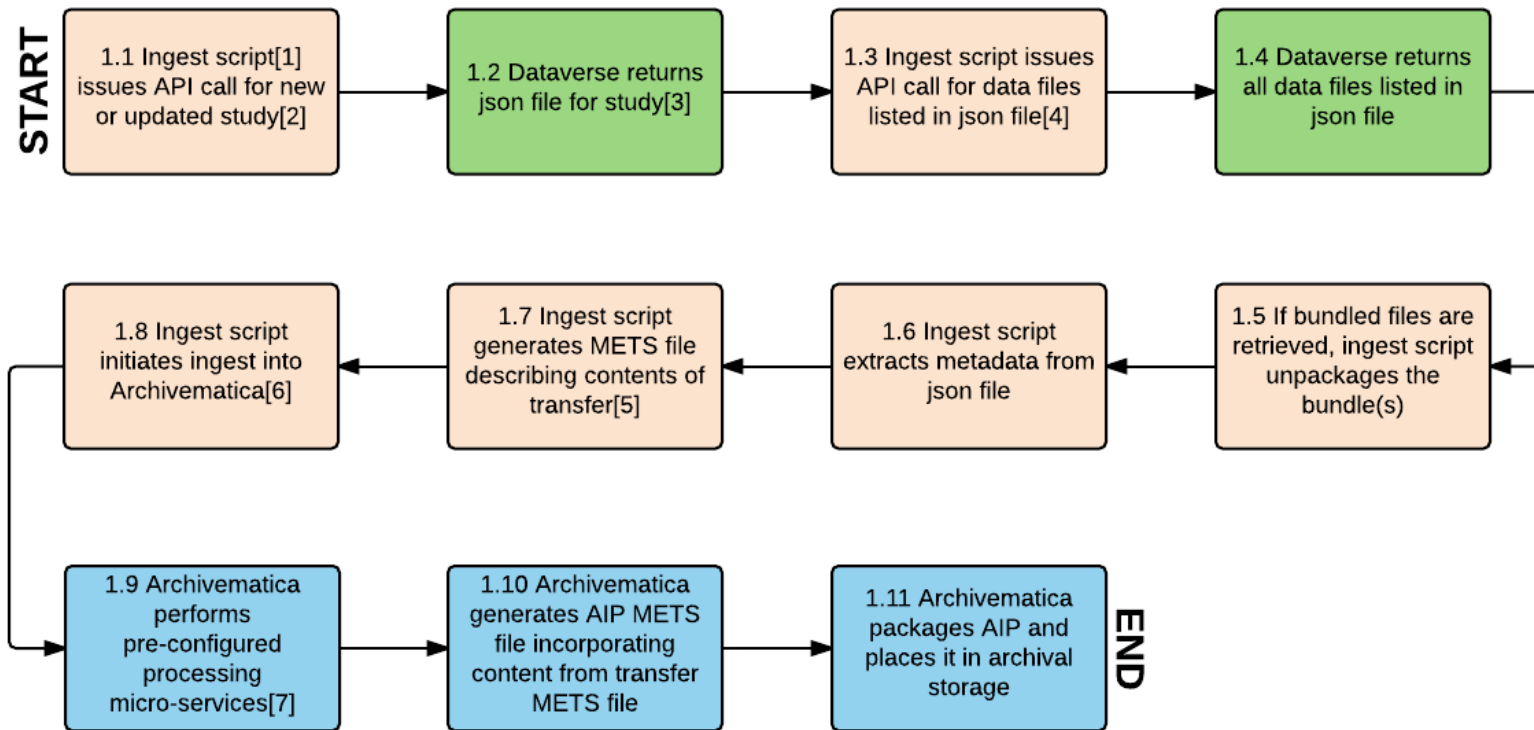
[Example from Univ of Zurich](#)

See Repository Badges documentation:

<https://www.altmetric.com/products/free-tools/institutional-repository-badges/>

## 5. Preservation Phase

# Scholars Portal Dataverse Integration With Archivematica



# Helping Future Researchers Re-Use Data

Thank You!  
Questions?

