DATAVERSE COMMUNITY MEETING

10 Years Sharing Data with Dataverse

#dataverse2017
Once there was the VDC
And then came the Dataverse Network
Now we have the Dataverse
RESEARCHERS ARE SHARING AND USING DATA

Harvard Dataverse

- > 70,000 datasets
- 200 datasets/month
- > 340,000 files
- 4,000 files/month
- > 2.5 M downloads
- 60,000 downloads/month
When we started, there were very few journals with data policies, no data requirements from funders.
Weak data sharing and strong data sharing vs. disciplines

Now, Journals across disciplines start supporting data policies

PRIVATE RESEARCH FUNDERS

- Bill and Melinda Gates Foundation Information Sharing Approach
- Sloan Foundation Data Sharing Policy
- Wellcome Trust Data Sharing Policy
- Arnold Foundation
- Moore Foundation
- Robert Wood Johnson Foundation
- HHMI Policy on the Sharing of Publication-Related Materials, Data and Software

PUBLIC RESEARCH FUNDERS

- Department of Agriculture
- Department of Commerce
- Department of Defense
- Department of Education
- Department of Energy
- Department of Health and Human Services
  - Agency for Healthcare Research and Quality (AHRQ)
  - Assistant Secretary for Preparedness and Response (ASPR)
  - Center for Disease Control and Prevention (CDC)
  - Food and Drug Administration (FDA)
  - National Institutes of Health (NIH)
- Department of Homeland Security
- Department of Housing and Urban Development
- Department of Interior
- Department of Labor
- Department of Transportation
- Department of Veterans Affairs
- Environmental Protection Agency (EPA)

And Funders require data sharing
WE ARE EXPERIENCING A CULTURAL CHANGE
WE ARE EXPERIENCING A CULTURAL CHANGE

WE ARE THE CULTURAL CHANGE!
<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1995</td>
<td>King</td>
<td>Replication</td>
</tr>
<tr>
<td>2001</td>
<td>Altman et al</td>
<td>A Digital Library for the Dissemination and Replication of Quantitative Social Science</td>
</tr>
<tr>
<td>2007</td>
<td>Altman and King</td>
<td>A Proposed for the Scholarly Citation of Quantitative Data</td>
</tr>
<tr>
<td>2007</td>
<td>King</td>
<td>An Introduction to the Dataverse Network as an Infrastructure for Data Sharing</td>
</tr>
<tr>
<td>2012</td>
<td>Crosas</td>
<td>The Dataverse Network: an open source application for sharing, discovering, and preserving research data</td>
</tr>
<tr>
<td>2013</td>
<td>Crosas</td>
<td>A Data Sharing Story</td>
</tr>
<tr>
<td>2013</td>
<td>Altman and Crosas</td>
<td>The Evolution to Data Citation: from principles to implementation</td>
</tr>
<tr>
<td>2014</td>
<td>Pepe et al</td>
<td>How Do Astronomers Share Data?</td>
</tr>
<tr>
<td>2014</td>
<td>Goodman et al</td>
<td>Ten Simple Rules for the Care and Feeding of Scientific Data</td>
</tr>
<tr>
<td>2015</td>
<td>Crosas, Honaker, King, Sweeney</td>
<td>Automating Open Science for Big Data</td>
</tr>
<tr>
<td>2015</td>
<td>Castro et al</td>
<td>Achieving Human and Machine Accessibility of Cited Data</td>
</tr>
<tr>
<td>2015</td>
<td>Sweeney, Crosas, Bar-Sinai</td>
<td>Sharing Sensitive Data with Confidence: The DataTags System</td>
</tr>
<tr>
<td>2016</td>
<td>Meyer et al</td>
<td>Data Publication with the Structural Biology Data Grid Supports Live Analysis</td>
</tr>
</tbody>
</table>

The Dataverse project and team leading many aspects of data sharing
AN ACTIVE TEAM AND COMMUNITY

METRICS FROM LAST YEAR, JUNE 2016 TO JUNE 2017
22 COMMUNITY CALLS

190 ATTENDEES
25 ORGANIZATIONS/UNIVERSITIES
10 COUNTRIES
Community

975 GOOGLE GROUP MESSAGES
7,114 IRC MESSAGES
245 UNIQUE USERS
12 SPRINTS
(STARTED IN JANUARY 2017)
220 STANDUP MEETINGS
52,000 SLACK MESSAGES
334 PULL REQUESTS
8,335 GITHUB COMMITS
1,153 SUPPORT TICKETS
A VISION: DATAVERSE AS A KEY PART OF THE FULL RESEARCH DATA LIFECYCLE
TOWARDS A DATA-CENTRIC RESEARCH LIFECYCLE
FROM DATA COLLECTION, TO COMPUTING AND SHARING

Data Collection
- Lab
- E-Notebooks
- Instruments
- Surveys
- ...

Run data & code
Explore & Visualize data
Journals & Funders
Data Citation
Work with Sensitive Data

Assign DUA & metadata
Cloud Computing and Storage

Track Provenance

The Dataverse® Project