Science requires community access to data
Free, open source software for archiving, sharing, referencing, extracting, and analyzing research data, dataverse.org
Partnership with the Harvard Library and Harvard University IT to provide a community data repository across sciences and humanities, dataverse.harvard.edu:
- 1000 published dataverses (by hundreds of universities)
- 57,791 published datasets (by thousands of scholars)
- 1,235,658 downloads

Mercè Crosas, Director of Data Science, IQSS, Harvard University, @mercecrosas
Publishing Datasets in Dataverse, even when Data are Confidential

What we support now:
• Datasets published in Dataverse get a data citation, with a DOI and attribution
• Metadata describing the datasets are always public, but files in dataset can be public or restricted
• Datasets are given a CC0 waiver as default
• BUT, CC0 can be changed if special terms of use apply to the data

What we are building:
• A solution to publish, query and analyze confidential data:
  • DataTags
  • Privacy-preserving tools
• As part of a collaboration with Harvard SEAS, Berkman center, IQSS (including the Data Privacy Lab), and MIT
<table>
<thead>
<tr>
<th>Tag Type</th>
<th>Description</th>
<th>Storage &amp; Transit</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td><strong>Non-confidential</strong> information, stored and shared freely.</td>
<td>Clear</td>
<td>Open</td>
</tr>
<tr>
<td>Green</td>
<td><strong>Not harmful</strong> personal information, shared with some access control.</td>
<td>Clear</td>
<td>Email, OAuth verified registration</td>
</tr>
<tr>
<td>Yellow</td>
<td><strong>Potentially harmful</strong> personal information, shared with loosely verified and/or approved recipients.</td>
<td>Encrypted</td>
<td>Password, Registered, Approval, click-through DUA</td>
</tr>
<tr>
<td>Orange</td>
<td><strong>Sensitive</strong> personal information, shared with verified and/or approved recipients under agreement.</td>
<td>Encrypted</td>
<td>Password, Registered, Approval, signed DUA</td>
</tr>
<tr>
<td>Red</td>
<td><strong>Very sensitive</strong> personal information, shared with strong verification of approved recipients under signed agreement.</td>
<td>Encrypted</td>
<td>Two-factor Auth, Registered, Approval, signed DUA</td>
</tr>
<tr>
<td>Crimson</td>
<td><strong>Maximum sensitive</strong>, explicit permission for each transaction, strong verification of approved recipients under signed agreement.</td>
<td>Double Encrypted</td>
<td>Two-factor Auth, Registered, Approval, signed DUA</td>
</tr>
</tbody>
</table>

Dataset

Surveys

Handling Access Control

DUAs, Legal Policies

DataTags

Direct Access

Criminal Penalties

Privacy Preserving Access

None

Minimal

Shame

Civil Penalties

Differential Privacy

Max Control

The Dataverse Project