PROMOTING TRANSPARENCY IN SOCIAL SCIENCE RESEARCH
SESSION: OPEN DATA AND PROTECTION OF RESEARCH SUBJECTS

DECEMBER 12-13, 2013, UC BERKELEY

Mercè Crosas, Ph.D.
Twitter: @mercecrosas
Institute for Quantitative Social Science
Harvard University
1. AN INTEGRATED PUBLISHING WORKFLOW FOR OPEN DATA

2. DATATAGS: SHARING EVEN SENSITIVE DATA WITH CONFIDENCE
AN INTEGRATED PUBLISHING WORKFLOW FOR OPEN DATA

Enabling data deposit to a public repository as part of article’s submission to a journal

http://projects.iq.harvard.edu/ojs-dvn
Journal’s Publishing System

Article Submission Workflow

Plugin for Data Deposit
1) Enter citation metadata
2) Select data Files

Get data citation

Insert Data Citation to References section

Open Data Repository

Data Sharing Workflow

Deposit metadata + data files

Dataset (metadata + data) in review, not released

Release data when article is published

Dataset (metadata + data) released
**Pilot Project:** Integrate PKP’s Open Journal System (OJS) with Dataverse

- **OJS Journal**
- **Harvard Dataverse Network**

**OJS plugin for:** Data + metadata + supporting files, sent via SWORD API to the Dataverse
Published Article + Link to Data

Moving toward a recovery-oriented approach in the Swedish mental health system

Carolina Klockmo

Data

1. Replication data for: Moving toward a recovery-oriented approach in the Swedish mental health system

Type:

Data Citation:

Data Set
Carolina Klockmo, 2012, "Moving toward a recovery-oriented approach in the Swedish mental health system", hdl:99999/00000

Full Record Available In:
COACTION DATaverse NETWORK

Data Tools:

INDEXING METADATA
ANALYZE
DOWNLOAD
Data in the Dataverse + Link to Article

Replication data for: Moving toward a recovery-oriented approach in the Swedish mental health system.

Klockmo, Carolina, 2012-12-13, "Replication data for: Moving toward a recovery-oriented approach in the Swedish mental health system", http://hdl.handle.net/TEST/10000 V1 [Version]

Vulnerable Groups & Inclusion. http://dx.doi.org/10.3402/vgi.v3i0.18879
ID: DOI:10.3402/vgi.v3i0.18879
Link
Progress to-date

**Phase I:** Completed testing with 50+ journals from 6 Publishers.

**Phase II:** Extend testing to >100 journals (2014)
DATATAGS: SHARING EVEN SENSITIVE DATA WITH CONFIDENCE

A standardized framework for sharing data when data cannot be 100% open

Mercè Crosas, Latanya Sweeney, Michael Bar-Sinai

http://privacytools.seas.harvard.edu/
<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>Clear &amp; 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>

Privacy Laws

• 2187 privacy related laws exist in the United States, which can be grouped in ~ 30 types:

Explicit consent, medicals records, student records, arrest and conviction records, bank and financial records, cable television, computer crime, credit reporting, criminal justice, electronic surveillance, employment records, government information on individuals, identity theft, insurance records (including use of Genetic information), library records, mailing lists, special medical records (including HIV testing), non-electronic surveillance, polygraphing in employment, privacy statutes/state constitutions, privileged communications, social security numbers, tax records, telephone services, testing in employment, tracking technologies, voter records.

Acknowledgement: Robert Smith, Latanya Sweeney
Phase I Implementation, Education and Outreach

- DataTags.org web application, initially supporting HIPAA regulations:
  - Explicit Consent
  - Medical Records
- Integrate with Secure Dataverse
- Alpha, Beta testing with hundreds of Social Scientists
- YouTube videos on how to use it

On a parallel track:
- HIPAA certification
- Pilot at Harvard Medical School and School of Public Health
DATA TAGS DEMO

http://datatags.org
Mercè Crosas, Latanya Sweeney, Michael Bar-Sinai