Tiered Access to Research Data for Secondary Analysis

John E Marcotte, Sarah Rush & Kelly Ogden-Schuette
DSDR
ICPSR, Institute for Social Research
University of Michigan
Tiered Access to Research Data

• Richness of research data for secondary analysis...
  ...Enabled researchers to analyze complex topics
  ...Increased disclosure risk

• Two levels of either *Public-access* or *Restricted-access* are no longer sufficient

• Seven Tiers

• Ten Controls that cover both technology and human behavior
Researchers

• Researchers must qualify for access
  Typical requirements: PI eligible, Doctoral degree, Grantee

• Requirements of the research data determine the appropriate tier of access

• All access is through the tier or a more restrictive tier
Data Repositories

• Some data repositories have established different levels of access that have some overlap with the seven tiers.

• Levels usually focus on technology such as encryption and 2-factor authentication

• Our unique contribution is how we define our tiers in terms of both human and technical controls to prevent the release of disclosive information.
Ten Controls

1. Application
   Do researchers have to apply for access?

2. Approval
   Does access have to be approved?

3. Agreement
   Who can agree to “Terms of Use” or sign “Data Use Agreement”?

4. Period of Access
   How long can a researcher have access?

5. Research location
   Where can researchers access the data?
Ten Controls

6. **Encryption**
   Is encryption at rest and in transit required?

7. **Internet**
   Is a connection (inbound and outbound) to the Internet allowed?

8. **Output**
   Does output have to be vetted by external personal?

9. **Proctor**
   Is a proctor present while research accesses the data?

10. **View of Data**
    Are researchers allowed to view the data?
0-Unrestricted

- Public-access research-data (also called “public-use” and “open-data”)
- Researcher can download
- *Example study:* Baby’s First Years (public)
- *Implementation:* Website and bandwidth to handle download demand.
- *Weakness:* Data might still have hidden risks.
- *Impediment to research:* Data may not contain sufficient information for analysis

☐ Application  ☐ Approval  ☐ Agreement  ☐ Period  ☐ Location
☐ Encryption  ☐ No Internet  ☐ Output  ☐ Proctor  ☐ No View
• Researcher must provide valid email and research purpose before download

• Example study: National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2018 [Public Use] requests registration from all data analyzers

• Implementation: Registration system to collect information. Website and bandwidth to handle download demand.

• Weakness: Researchers could provide inaccurate information.

• Impediment to research: Researcher must provide information to access data.

☑ Application  □ Approval  □ Agreement  □ Period  □ Location
□ Encryption  □ No Internet  □ Output  □ Proctor  □ No View
2-Approved

- Researcher must provide be approved before download
- May have additional requirements including Agreement, Period, Location and Encryption.

*Example studies:* Some Panel Study of Income Dynamics (PSID) and Health and Retirement Study (HRS) data fall into this category.

*Implementation:* Application system with encrypted download.

*Weakness:* Researchers could leak data inadvertently.

*Impediment to research:* Researchers must apply for access to research-data

- Application
- Approval
- Agreement
- Period
- Location

- Encryption
- No Internet
- Output
- Proctor
- No View
3-Local

- Researcher receives data with approved security plan
- Researcher analyzes data at local university or organization
- Requires Data Use Agreement signed by Researcher and Institutional Representative
- First tier of “restricted-access”

**Example study:** Study of Early Child Care and Youth Development (SECCYD)

**Implementation:** Standalone (non-networked computer) in a locked private office. Some organizations may have an acceptable server set up.

**Weakness:** The research-data with re-identification and harm risks are not under the control of the repository. Unauthorized access is possible.

**Impediment to research:** Difficult to collaborate with a research team. Universities and organizations may be reluctant to permit a non-networked computer.

- Application
- Approval
- Agreement
- Period
- Location
- Encryption
- No Internet
- Output
- Proctor
- No View
4-Remote

• Researcher comes to data electronically with approved security plan
• Researcher analyzes data on “virtual” enclave system; output only released after review
• Requires Data Use Agreement signed by Researcher and Institutional Representative
• Second tier of “restricted-access”

• Example study: L.A. FANS (restricted data)
• Implementation: Terminal Server or Virtual Desktop Infrastructure (VDI) that prevents files from being copied off the server or VDI.
• Weakness: Researchers could still transcribe information from the screen.
• Impediment to research: Researchers must wait for the release of results. Available software may be limited.

☑ Application  ☑ Approval  ☑ Agreement  ☑ Period  ☑ Location
☑ Encryption  ☑ No Internet  ☑ Output  □ Proctor  □ No View
5-Coldroom

- Researcher comes to data in person with pre-approved materials
- Researcher analyzes data in “coldroom”; output only released after review
- Requires Data Use Agreement signed by Researcher and Institutional Representative
- Proctor is present and inspects materials when leaving coldroom
- Third tier of “restricted-access”

Example study: Videos; Data with direct identifiers
Implementation: Locked room with proctor
Weakness: Researchers could still look up an individual record.
Impediment to research: Accessing the data requires travel to the cold room and an appointment.

✔ Application ✔ Approval ✔ Agreement ✔ Period ✔ Location ✔ Encryption ✔ No Internet ✔ Output ✔ Proctor ☐ No View
6-Batch

• Researcher only receives summary results
• Researcher analyzes data by submitting batch jobs; output only released after review
• Requires Data Use Agreement signed by Researcher and Institutional Representative
• Fourth tier of “restricted-access”

Example study: Data with high sensitivity and high re-identification risks

Implementation: Batch system. A server with synthetic data and the software available in the batch system for testing programs will enable the system to run smoothly.

LISSY at the Cross-national Data Center in Luxembourg is an implementation of this tier. The retired ANDRE system at the National Center for Health Statistics (NCHS) was another example.

Impediment to research: Without access to the data, analysis is cumbersome and requires much more time.

Although 6-Batch is more restrictive than 5-Coldroom, the tier does not require travel to a specific location.
## Access Tier by Security Control

<table>
<thead>
<tr>
<th>Tier</th>
<th>Description</th>
<th>Application</th>
<th>Approval</th>
<th>Agreement</th>
<th>Period of Access</th>
<th>Research Location</th>
<th>Encryption</th>
<th>Internet</th>
<th>Output</th>
<th>Proctor</th>
<th>View Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>researcher may download</td>
<td>none</td>
<td>none</td>
<td>Researcher</td>
<td>No limit</td>
<td>public or private</td>
<td>not required</td>
<td>allowed</td>
<td>not vetted</td>
<td>not monitored</td>
<td>allowed</td>
</tr>
<tr>
<td>1-Registered</td>
<td>researcher must provide additional info such as research purpose before download</td>
<td>submit information</td>
<td>none</td>
<td>Researcher</td>
<td>No limit</td>
<td>public or private</td>
<td>not required</td>
<td>allowed</td>
<td>not vetted</td>
<td>not monitored</td>
<td>allowed</td>
</tr>
<tr>
<td>2-Approved</td>
<td>researcher must be approved before download</td>
<td>must apply</td>
<td>approved</td>
<td>Researcher &amp; Advisor</td>
<td>Limited</td>
<td>private</td>
<td>at rest in transit</td>
<td>allowed</td>
<td>not vetted</td>
<td>not monitored</td>
<td>allowed</td>
</tr>
<tr>
<td>3-Local</td>
<td>researcher receives data with approved security plan</td>
<td>must apply</td>
<td>approved</td>
<td>Researcher &amp; Institution</td>
<td>Specified period</td>
<td>private</td>
<td>at rest, real-time in transit</td>
<td>blocked</td>
<td>self-vetted</td>
<td>not monitored</td>
<td>allowed</td>
</tr>
<tr>
<td>4-Remote</td>
<td>researcher comes to data electronically with approved security plan</td>
<td>must apply</td>
<td>approved</td>
<td>Researcher &amp; Institution</td>
<td>Specified period</td>
<td>private</td>
<td>at rest in transit</td>
<td>blocked except session</td>
<td>externally vetted</td>
<td>not monitored</td>
<td>allowed</td>
</tr>
<tr>
<td>5-Coldroom</td>
<td>researchers cannot access the data</td>
<td>must apply</td>
<td>approved</td>
<td>Researcher &amp; Institution</td>
<td>Specified period</td>
<td>private</td>
<td>at rest in transit</td>
<td>blocked</td>
<td>externally vetted</td>
<td>watched during access</td>
<td>not allowed</td>
</tr>
<tr>
<td>6-Batch</td>
<td>researchers can only access summary results</td>
<td>must apply</td>
<td>approved</td>
<td>Researcher &amp; Institution</td>
<td>Specified period</td>
<td>private</td>
<td>at rest in transit</td>
<td>only batch submissions</td>
<td>externally vetted</td>
<td>monitored batch jobs</td>
<td>not allowed</td>
</tr>
</tbody>
</table>
# Public-access vs. Restricted-access

<table>
<thead>
<tr>
<th></th>
<th>Public-access</th>
<th>Restricted-access</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>• Research Only</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No attempt to identify respondents</td>
<td></td>
</tr>
<tr>
<td><strong>Request Data</strong></td>
<td>No application</td>
<td>Application</td>
</tr>
<tr>
<td><strong>Understanding</strong></td>
<td>Terms of Use</td>
<td>Data Use Agreement</td>
</tr>
<tr>
<td><strong>IRB</strong></td>
<td>Exempt</td>
<td>Possible Review</td>
</tr>
<tr>
<td><strong>Disclosure Risk</strong></td>
<td>Data: Very Low</td>
<td>Results: Very Low</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>No security requirements</td>
<td>Security Plan</td>
</tr>
<tr>
<td><strong>Period of Access</strong></td>
<td>Unlimited</td>
<td>End-date</td>
</tr>
<tr>
<td><strong>Access Tier</strong></td>
<td>0-Unrestricted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 3-Local</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 4-Remote</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 5-Coldroom</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 6-Batch</td>
<td></td>
</tr>
</tbody>
</table>
Connect with DSDR

User Support
(734) 763-8937
DSDR@icpsr.umich.edu

Subscribe to DSDR News for the latest updates!

Facebook: @DSDRdata