



Data Sharing for
Demographic Research

DSDR.ICPSR.UMICH.EDU

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.

Seven Tiers

0-Unrestricted

1-Registered

2-Approved

3-Local

4-Remote

5-Coldroom

6-Batch

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

1-Registered

- 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.

☑Application ☑Approval ☑Agreement ☑Period ☑Location
☑Encryption ☑No Internet ☑Output ☑Proctor ☑No View

Access Tier by Security Control

	Tier	Description	Application	Approval	Agreement	Period of Access	Research Location	Encryption	Internet	Output	Proctor	View Data
Public	0-Unrestricted	researcher may download	none	none	Researcher	No limit	public or private	not required	allowed	not vetted	not monitored	allowed
	1-Registered	researcher must provide additional info such as research purpose before download	submit information	none	Researcher	No limit	public or private	not required	allowed	not vetted	not monitored	allowed
	2-Approved	researcher must be approved before download	must apply	approved	Researcher & Advisor	Limited	private	at rest in transit	allowed	not vetted	not monitored	allowed
Restricted	3- Local	researcher receives data with approved security plan	must apply	approved	Researcher & Institution	Specified period	private	at rest, real-time in transit	blocked	self-vetted	not monitored	allowed
	4-Remote	researcher comes to data electronically with approved security plan	must apply	approved	Researcher & Institution	Specified period	private	at rest in transit	blocked except session	externally vetted	not monitored	allowed
	5-Coldroom	researcher comes to data in person with pre-approved materials	must apply	approved	Researcher & Institution	Specified Period	private	at rest in transit	blocked	externally vetted	watched during access	allowed
	6-Batch	researchers cannot access the data researchers can only access summary results	must apply	approved	Researcher & Institution	Specified period	private	at rest	only batch submissions	externally vetted	monitored batch jobs	not allowed

Public-access vs. Restricted-access

	Public-access	Restricted-access
Purpose	<ul style="list-style-type: none"> • Research Only • No attempt to identify respondents 	
Request Data	No application	Application
Understanding	Terms of Use	Data Use Agreement
IRB	Exempt	Possible Review
Disclosure Risk	Data: Very Low	Results: Very Low
Security	No security requirements	Security Plan
Period of Access	Unlimited	End-date
Access Tier	0-Unrestricted	<ul style="list-style-type: none"> • 3-Local • 4-Remote • 5-Coldroom • 6-Batch

Connect with DSDR

User Support

(734) 763-8937

DSDR@icpsr.umich.edu

Subscribe to [DSDR News](#) for the latest updates!



 @DSDRdata