

# ICPSR



## Researcher Passport & Virtual Data Enclave

**Maggie Levenstein**  
**ICPSR Director**

**Statistics Canada**  
**August 21, 2019**

# ICPSR

Founded in 1962 by 22 universities, now  
consortium of ~800 institutions world-wide

Focus on social and behavioral science data

Current holdings

- 11,000 studies, quarter million files
- 1500 are *restricted studies*, almost always to protect confidentiality
- Bibliography of Data-related Literature with 80,000 citations

Approximately 60,000 active MyData  
("shopping cart") accounts

Thematic collections of data about addiction  
and HIV, aging, arts and culture, child care and  
early education, criminal justice, demography,  
health and medical care, and minorities



# Differentiated access to restricted data

- Goal: Protect confidentiality
  - Of individuals, groups, organizations
- Implementation can rely on safe data, safe places, safe people
  - But you don't need all three
- Differentiated access leverages
  - Different data
  - Different researchers
  - Different computing environments

# Alternative access modalities

## **Encrypted download**

- ✓ Approved, secure local computing environment
- ✓ Researcher conducts disclosure review
- ✓ Restricted data use agreement with sanctions

## **Virtual Data Enclave (VDE)**

- ✓ Data stays on secure server
- ✓ Access from approved, local computing environment
- ✓ Third party disclosure review
- ✓ Restricted data use agreement with sanctions

## **Physical Data Enclave (PDE)**

- ✓ Access from approved, secure computing environment
- ✓ Third party disclosure review
- ✓ Restricted data use agreement with sanctions

## **Secure online analysis**

- ✓ Data stays on secure server
- ✓ Automatic disclosure review

# Integrating access modalities with researcher credentials

PI SCORE	DATA SECURITY LEVEL				
	LOW	MODERATE	HIGH	HIGHEST	
8+	unrestricted	secure download	secure download	VDE / physical enclave	Platinum
7	unrestricted	secure download	virtual enclave	physical enclave	Gold
6	unrestricted	secure download	virtual enclave	no access	Silver
5	unrestricted	virtual enclave	no access	no access	Bronze
4	unrestricted	virtual enclave	no access	no access	Copper
0-3	unrestricted	no access	no access	no access	Tin

# Virtual Data Enclave

Data stays on server in secure data center

- Behind secure firewall

Computation and analysis conducted on virtual machine

- Behind secure firewall
- Analytical software must be on virtual machine
- No internet access
- No access to local machine

Restricted data use agreement

- Institutional and researcher signature
- Explicit sanctions

# Virtual Data Enclave

Third party review of all output to avoid disclosure

- Before output sent to researcher outside VDE environment

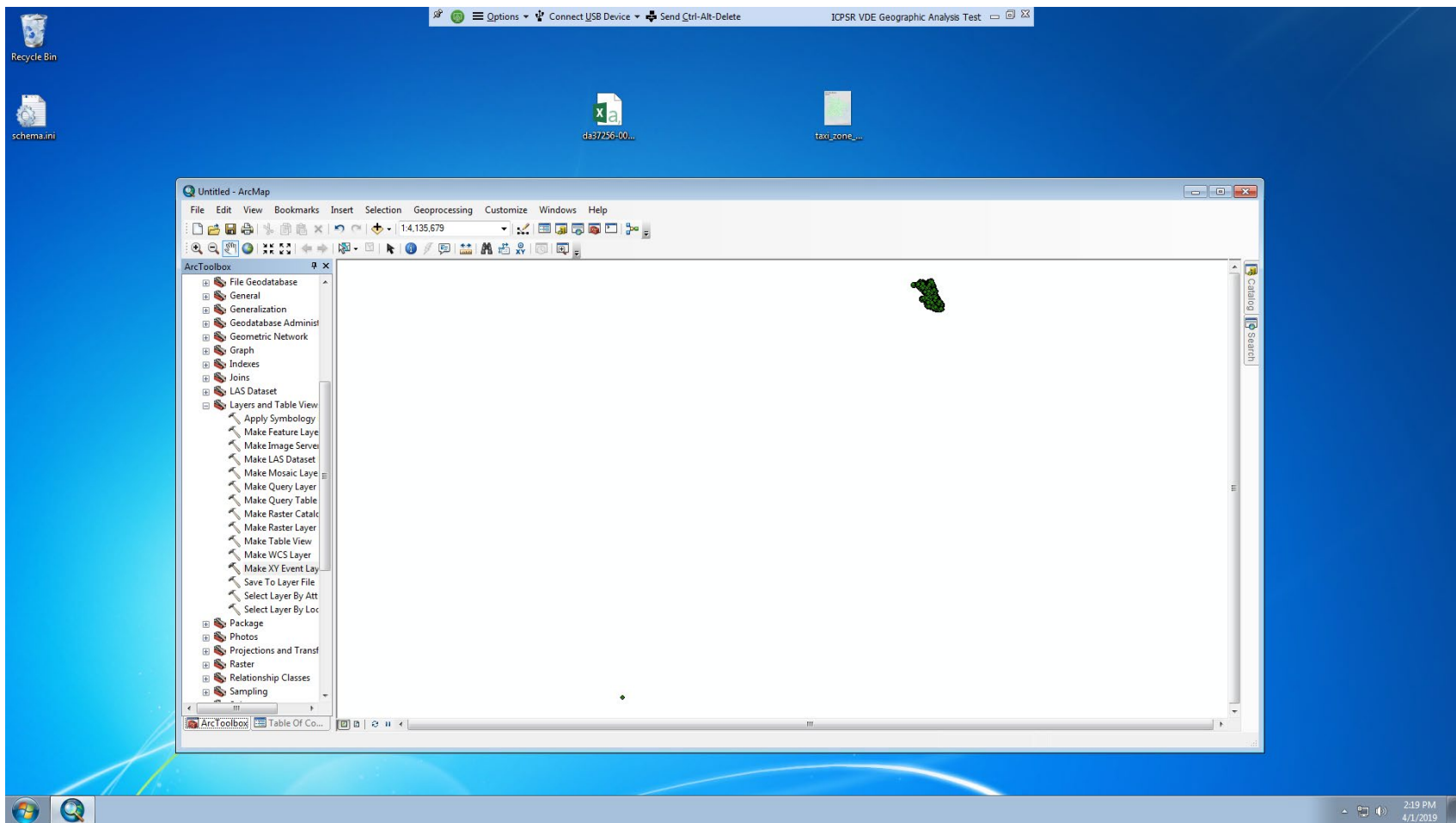
Requires [training](#) and quiz

- Establish culture of confidentiality

Research goal: automate disclosure

Mask output, not underlying data

# ICPSR VDE windows environment





# FSRDC VDE Pilot

## Michigan Federal Statistical Research Data Center

- Longstanding physical enclave collaborating with US Census Bureau
- Will use VDI technology to allow researchers with approved projects to connect to FSRDC servers at US Census Bureau from approved locations outside a physical RDC
  - Relatively low risk datasets from US Census Bureau
  - Other statistical agencies invited to participate
- Will launch this fall

# Researcher Passport

## Researcher Passport: Improving Data Access and Confidentiality Protection

- ICPSR's Strategy for a Community-normed System of Digital Identities of Access
- <https://deepblue.lib.umich.edu/handle/2027.42/143808>
- Identifies inconsistent language and policies that impede access

## Passports for safe people

- Verified identities, institutional affiliation
- Training
- Experience (good and bad)

## Visas to control access

- Permission to “enter” (access) specific data specifying
  - Passport holder
    - Project, Place, Period

# Researcher Passport and Radius

https://statsnap.icpsr.umich.edu/statsnap/

Bookmarks Imported From Fire... Issues Other bookmark

Contact Us Johanna Bleckman (logout)

## StatSnap

An online tool designed to explore, subset, and analyze data in a snap!

**BETA**

Welcome to StatSnap, ICPSR's new online exploration and analysis tool! This tool is still in development and this Beta version is limited to frequencies and cross-tabulations. If you need to analyze data with a complex sample design using weights, please return to the study homepage and choose the "Freqs/Crosstabs (Legacy)" option. If you need to do more sophisticated analyses select the "Full Analysis Capabilities (SDA)" option. [We'd love to hear your feedback](#)

### Welcome to ICPSR's Online Statistical Analysis System

[Browse Studies and Datasets.](#)

#### Your History

#### Popular Datasets to Use with StatSnap

- [Carolina Abecedarian Project \(ABC\) and the Carolina Approach to Responsive Education \(CARE\), Age 21 Follow Up Study, 1993 - 2003](#)
- [21st Century Americanism: Nationally Representative Survey of the United States Population, 2004](#)
- [WHO Study on Global AGEing and Adult Health \(SAGE\): Wave 0, 2002-2004](#)
- [Annual Parole Survey, 2007](#)
- [National Survey on Drug Use and Health, 2012](#)

# Researcher Passport

[Sign In](#)

[My Passport](#) | [About](#) | [FAQs](#)

**RESEARCHER  
PASSPORT**  
by ICPSR

**BETA**

[Learn about the development phases and submit feedback.](#)

Improving Data Access and Confidentiality Protection

[Apply now](#)

What is a Researcher Passport?

It's a digital identity, or profile, that captures and verifies the information that data repositories need to know in order to share their data with you. It can then be provided to participating repositories to expedite your access to their data.

- ✔ Complete your profile
- ✔ Submit your application
- ✔ Share your passport as you apply for data access

[Watch our one-minute video](#)

**Establishing shared  
understanding across  
repositories of what it  
means to be a trusted  
researcher**

# Researcher passport: next steps

- Digital identifiers associated with ICPSR MyData accounts
  - Open badges
  - Integrated into VDE application process
- Repositories issue visas to access restricted data
  - Visa controls access to secure download or VDE
    - Visa issued after signed DUA
- Establishing content of training requirements
- Governance system to provide due process for allegations of misconduct



# LINKAGE LIBRARY

Maintaining datasets to support the data linkage community



### What is Linkage Library



**Comment**  
Comment on projects



**Projects**  
Some have data, Some have code, Some have both



**GitHub**  
Interface with GitHub

The logo for Linkage Library features a blue arc at the top with a yellow dot on the left and a blue dot on the right. Below the arc, the words "LINKAGE LIBRARY" are written in a bold, blue, sans-serif font.

# LINKAGE LIBRARY

Encourage researchers to share linked (or linkable) data and linkage strategies

- Algorithms, code, documents

Compare approaches across projects, datasets, disciplines

- Improve linkage practices
- Improve transparency
- Commenting – to improve linkage & build community

[LinkageLibrary.org](http://LinkageLibrary.org)



This is DataJeff. He loves data.

