

# ICPSR

SHARING DATA TO ADVANCE SCIENCE



## **ICPSR's Restricted-use Data Management and Virtual Data Enclave**

**Data-PASS Alliance Workshop on Data Under Constraint:  
Tools and Strategies for Facilitating Transparency**

**August 30, 2017**

**Justin Noble, Acquisitions Manager  
Inter-university Consortium for Political and Social Research (ICPSR)**

# ICPSR

- Domain Repository for Social & Behavioral Sciences
- International consortium of 760 academic institutions and research organizations
- 55 years of distributing social and behavioral science research data, including government data
- Largest social science data archive in world



9000+ studies



4,700,000+ variables



70,000+ publications



# Restricted-Use Holdings at ICPSR

- Over 1,350 restricted-use datasets available for analysis
- Approximately 20 datasets accessible only on site in physical enclave
- Over 2,000 active restricted-use data agreements
  - Across 500+ institutions
  - Representing 150+ studies
  - Approximately 50 new agreements executed per month

# Methods of Accessing Restricted-Use Data

- ICPSR offers three methods of restricted-use data access:
  - Physical Enclave
  - Secure Dissemination
  - Virtual Data Enclave (VDE)

Principal Investigator(s): Rosay, Andre, University of Alaska-Anchorage; Henry, Tara, Forensic Nurse Services, Inc.

## Summary:

This project examined the characteristics of sexual assault victimizations in Alaska, as observed and recorded by sexual assault nurse examiners in Anchorage, Kodiak, Bethel, Soldotna, Nome, Fairbanks, Homer, and Kotzebue. The sample utilized for this study included all sexual assault nurse examinations conducted in Anchorage from 1996 to 2004, in Bethel and Fairbanks in 2005 and 2006, and in Homer, Kodiak, Kotzebue, Nome, and Soldotna in 2005. A total of 1,699 examinations were collected. More ... ([more info](#))

## Access Notes

- **One or more files in this study are not available for download due to special restrictions**; consult the [restrictions note](#) to learn more. You can [apply online for access to the data](#). A login is required to apply for access.

## Dataset(s)

### Dataset

Documentation: [Codebook.pdf](#) [Questionnaire.pdf](#) [User guide.pdf](#)  
Download: **No downloadable data files available.**

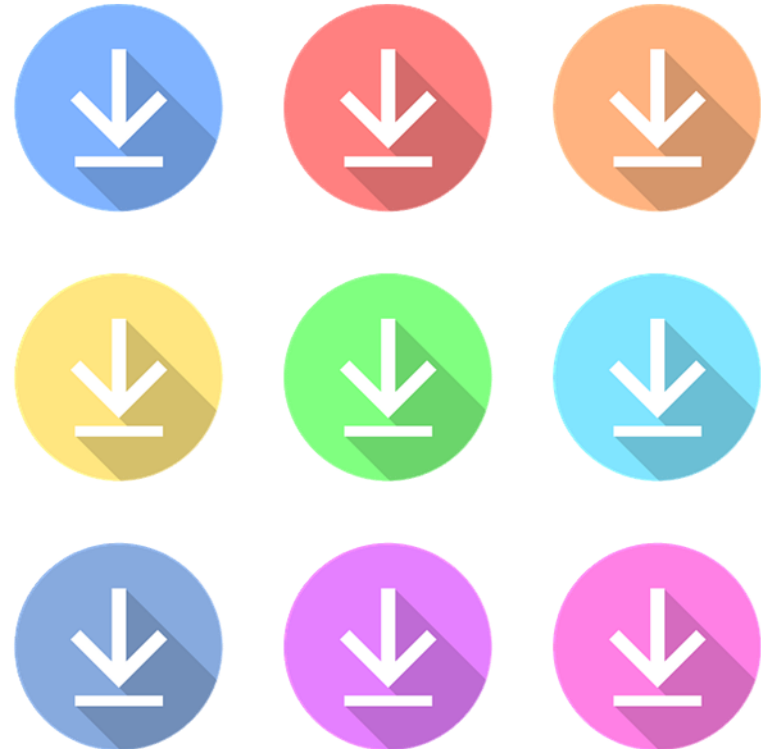
# Physical Enclave

- Researcher must be present in person at ICPSR to analyze data
- All results reviewed for disclosure before release to researcher
- Allows greatest control over data
- Extremely high barrier to access



# Secure Dissemination

- Data are encrypted and delivered via download
- On-line proposal request system
- When request is approved, investigator receives temporary link and password to download files
- Data must be destroyed at the completion of project
- Researcher commits to protect confidentiality of data, but method relies on their trustworthiness







# Virtual Data Enclave (VDE)

- ICPSR hosts Virtual Data Enclave (VDE) providing local access to restricted-use data in secure environment
- VDE is virtual machine launched from researcher's desktop that operates on remote server – at ICPSR
- Data remain on ICPSR file servers and are accessed and analyzed by researcher virtually
- Virtual machine restricts users from printing, emailing, copying, or otherwise moving files outside of the secure environment, either accidentally or intentionally

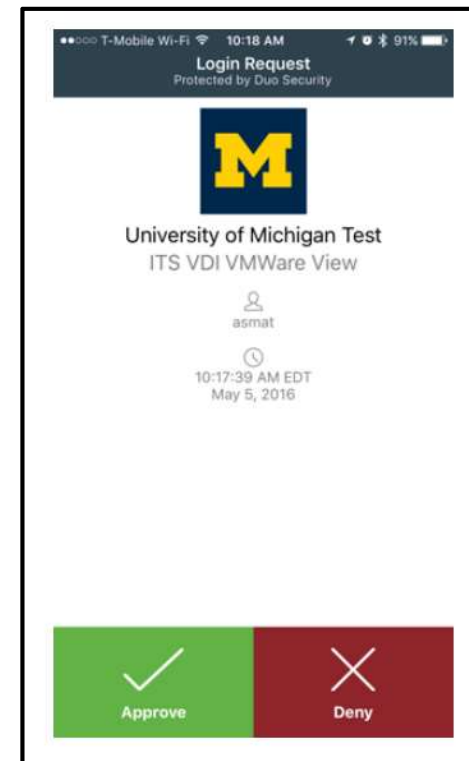
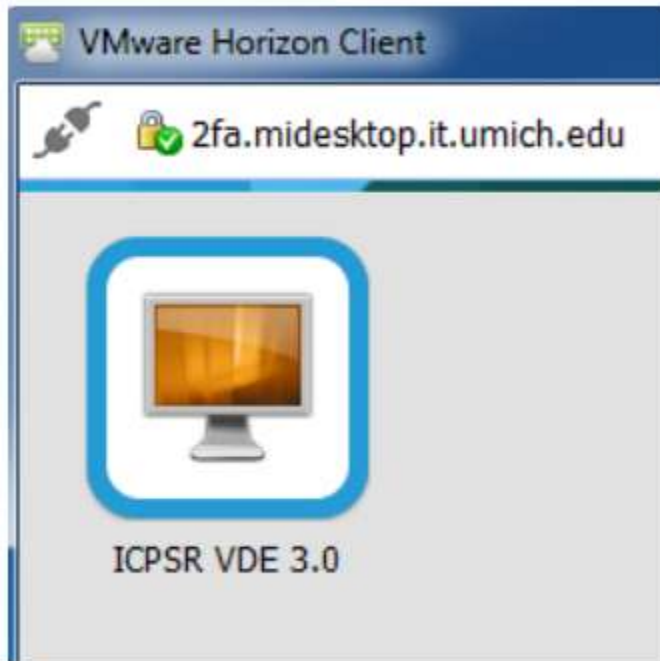


# VDE Options and Software

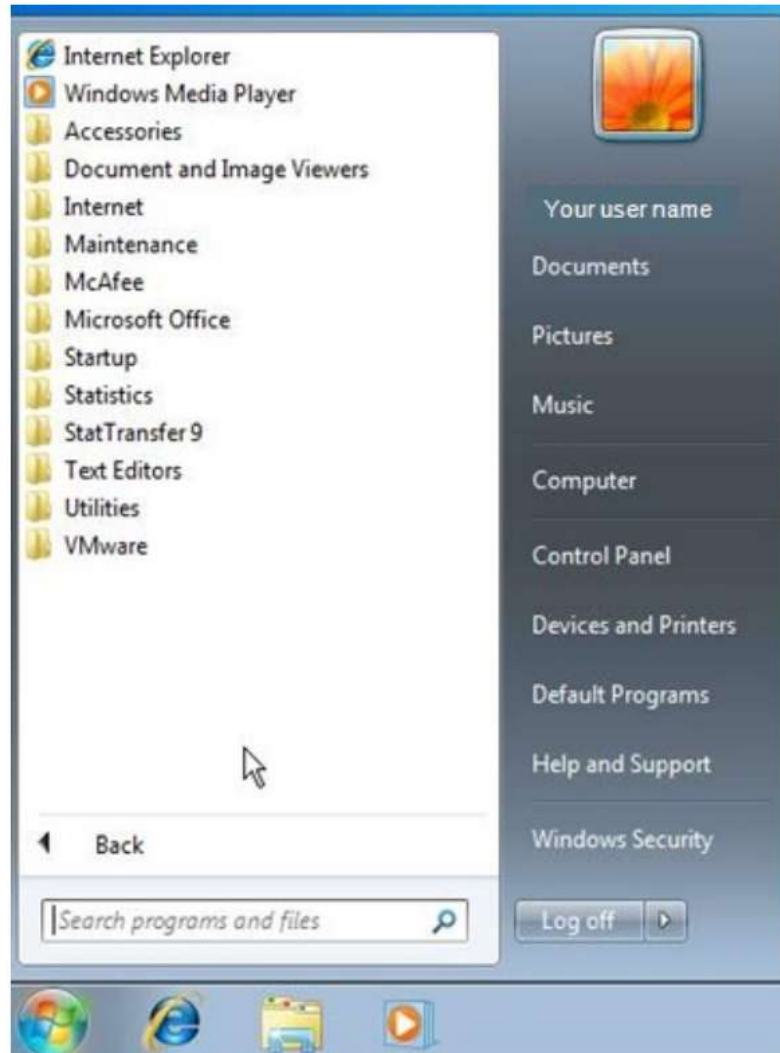
- ICPSR offers options such as
  - Vetting of research results generated in VDE
    - Third-party disclosure review with release of cleared output outside VDE firewall
  - File sharing among project team members
- Software tools and support available for use within VDE include
  - Geospatial analysis tools
  - Statistical analysis software
  - Various documentation programs



# VMWare + Two-factor Authentication



# VDE Desktop





# Applying for Access

- Users must apply for access
  - Describe proposed research and why restricted data is required
- All users must sign a Restricted Data Use Agreement
  - Agree to data security requirements and procedures
  - Principal Investigator generally must hold terminal degree
    - Students can be on agreements
    - Some projects provide access to researchers, policymakers, journalists without a terminal degree
  - Be affiliated with legitimate research institution
  - Receive IRB approval or exemption before gaining access to most data



# VDE Pricing Models

- Per User
  - Access to VDE to analyze data for one year
  - Current annual access fee: \$350 per user
    - Supports virtual machine technology and associated software licensing costs
- Per Machine
  - Sponsor pays for certain number of virtual machines that are reserved for particular data sets or research community
  - Each virtual machine can support one simultaneous user



# Researcher Credentialing for Restricted Data Access

- Project Goals:
  - Develop best practices for transferable researcher credential for accessing restricted data
  - Develop and test user interface for repository websites to accept, authorize trusted digital researcher credentials
- Funded by Alfred P. Sloan Foundation
- ICPSR Project Team
  - Principal Investigator: Maggie Levenstein
  - Staff: Johanna Davidson Bleckman
  - Graduate Student Research Assistant: Allison Tyler



# Credentialing Project Research Methodology

- Environmental scan and organizational study of 23 restricted data repositories
  - 16 United States
  - 7 International (Germany, Australia, United Kingdom, the Netherlands)
- Document analysis
  - Policy documents, legal statutes, repository documentation
- Interviews
  - Semi-structured interviews with repository staff



# Convening of Stakeholders

- July 24, 2017: ICPSR and Georgetown convened over two dozen representatives from government agencies, data repositories, and academic institutions
- Researcher Attributes/Research Identity: Professional and institutional affiliations, training certifications, grants received, and publication history
- Feedback from meeting informs White Paper to be published this winter
- Developing and testing software and web interface of credentialing system





# Copyright

## Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

# Fair Use

## NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.




# openICPSR Repositories

- Branded, self-deposit system to demonstrate research impact of a an organization or group
- Demonstrate research transparency by making data available for replication and providing live links to publications
- Provide authors with DOIs and data citations upon deposit
- Restricted-use data shared securely through VDE
- Other Features:
  - Deposit Notifications
  - Project Sharing
  - Delayed Dissemination

# Journal Repository Workflow

- Conditional acceptance of paper
- Authors publish replication data
- Journal notified of data release
- Data citation/DOI added to paper
- Article published
- Article citation/DOI added to related publications section of data



Add/Edit Citation

IMPORT VIA DOI | MANUALLY ENTER CITATION

Note that you can also upload a bibliography file containing multiple citations, rather than entering them individually here.





## Depositing Data with openICPSR to fulfill the Journal of Economic History Archiving and Replication Requirements

The *Journal of Economic History* requires authors to deposit data and materials with the *openICPSR* repository that allow other researchers to replicate analyses in a published article without any additional information from the author. **Authors are also required to include a citation pointing to the data in the reference section of the final version of the article sent to the JEH editors.** The *openICPSR* repository automatically generates a citation when the data are "published."

Deposits should include all data, annotated program code, command files, and documentation that is needed to replicate the findings from a *JEH* article. Each variable in the data collection should have a set of exhaustive, mutually-exclusive codes. Variable labels and value labels should clearly describe the information or question recorded in that variable. Missing data codes should be defined. Identifying information should be removed from the data to protect [confidentiality](#). Program code and command files should be annotated to facilitate replication and ensure clear correspondence between code and figures, tables, and analyses in the *JEH* article. For a discussion of best practice in preparing data for sharing, please refer to ICPSR's [Guide to Social Science Data Preparation and Archiving, 5th Edition](#).

[Start Your Deposit](#)

### ICPSR's Role

ICPSR does not approve or alter datasets deposited in openICPSR in any way. Data are distributed in the same condition and format submitted by the depositor. Data approved for public use are distributed at no cost to users. ICPSR may charge a fee for the costs of managing data that have been restricted to protect confidential information about research subjects.

### Confidentiality and Copyright

Investigators submitting data to ICPSR are entirely responsible for ensuring the confidentiality of research subjects. The presence of identifiable or sensitive information must be indicated during the submission process. ICPSR does not review submissions for disclosure risk. Investigators are also required to affirm that they have the right to publish this material. ICPSR requires a license for distribution of data, but copyright remains with the author.

### About JEH

The Journal of Economic History is devoted to the multidisciplinary study of history and economics, and is of interest not only to economic historians but to social and demographic historians, as well as economists in general. The journal has broad coverage, in terms of both methodology and geographic scope. Topics covered include money and banking, trade, manufacturing, technology, transportation, industrial organisation, labour, agriculture, servitude, demography, education, economic growth, and the role of government and regulation.

Filters

Subject Terms ▸

- [Census of India](#) 1
- [Economic History](#) 1
- [Financial history](#) 1
- [econometrics](#) 1
- [education](#) 1
- [history](#) 1

[view all](#)

Geography ▸

Access Level ▸

Search Results

8 results.

Per page:  Sort by:

	Project Title/Investigator	Released/Updated
1.	<a href="#">The (South) American Dream: Mobility and Economic Outcomes of First- and Second-Generation Immigrants in 19th-Century Argentina</a> Perez, Santiago	2017-07-31
2.	<a href="#">Common tongue: The impact of language on educational outcomes</a> Jain, Tarun	2017-06-22
3.	<a href="#">Marketing and Pricing Risk in Marine Insurance in Sixteenth-Century Antwerp data</a> Puttevils, Jeroen.; Deloof, Marc	2017-06-22
4.	<a href="#">The Effects of World War II Military Service: Evidence from Australia</a> Siminski, Peter.; Ville, Simon.; Cousley, Alex	2017-06-19
5.	<a href="#">Political institutions, economic liberty and the great divergence</a> Cox, Gary W	2017-05-16
6.	<a href="#">Human Capital and Knowledge-Intensive Industries Location: Evidence from Soviet Legacy in Russia</a> Ivanov, Denis	2017-05-01
7.	<a href="#">Lifespans of the European Elite, 800-1800</a> Cummins, Neil	2017-05-01
8.	<a href="#">Local Economic Impacts of Coal Mining in the United States 1870 to 1970</a> Matheis, Mike	2016-12-02



# Project Sharing

- Collaborate on data deposit prior to publication of data
- Use-cases:
  - Co-authors
  - Graduate students and PIs
  - Authors and journal editors



+ CREATE NEW PROJECT

Search within workspace, enter 3 or more characters

Hide inactive

Workspace

- OPENICPSR-100501


Shared with me

+ OPENICPSR-100647

## Test APSA Data Deposit [Edit Project Header](#)

Create Folder Upload Files Import From Zip more

DELETE SELECTION

Name	Type	Size	Last Modified	Actions
 APSA_Codebook.pdf	application/pdf	3 MB	3/10/2017 8:17:26 AM	
 APSA_code.sps	text/x-spss-syntax	3 KB	8/18/2016 5:10:44 AM	
 APSA_data.sav	application/x-spss-s...	433 KB	3/9/2017 9:49:35 AM	

### Project

+ Expand All

+ Project Description

+ Scope of Project

+ Methodology

+ Related Publications

Project Communication Log

+ add entry watch

OPENICPSR-100501

PUBLISH PROJECT

Share Project

Change Owner

View Log

Report a Problem

#### Storage

Space Used: 3 MB

Sub Folders: 0

Files: 3

# Test APSA Data Deposit [Edit Project Header](#)

Create Folder

Upload Files

Import From Zip

more ▾



DELETE SELECTION

<input type="checkbox"/>	Name	Type	Size	Last Modified	Actions
<input type="checkbox"/>	APSA_Codebook.pdf	application/pdf	3 MB	3/10/2017 8:17:26 AM	
<input type="checkbox"/>	APSA_code.sps	text/x-spss-syntax	3 KB	8/18/2016 5:10:44 AM	
<input type="checkbox"/>	APSA_data.sav	application/x-spss-s...	433 KB	3/9/2017 9:49:35 AM	

## Project

+ Expand All

### - Project Description

**Principal Investigator(s) (required)** + add value

Justin Noble, Inter-university Consortium for Political and Social Research edit remove

**Summary (required)**

This is the replication package for...

edit remove



OPENICPSR-100501

[PUBLISH PROJECT](#)

[Share Project](#)

[Change Owner](#)

[View Log](#)

[Report a Problem](#)

**Storage**

---

Space Used: 3 MB

---

Sub Folders: 0

---

Files: 3

Share project ✕

User Email (s), comma separated.

- Allow user to read contents
- Allow user to edit contents
- Allow user to delete contents
- Allow user to download contents
- Allow user to control permissions and publish contents

SHARE

# Terms and Conditions

- Disclosure Risk
- Distribution Method
- Delayed Dissemination
- Deposit Agreement



# Disclosure Risk

Can individuals be identified from information in this Data Collection? If the data were made public, could someone use a combination of variables (e.g. age, sex, race, occupation, geography) to find individuals in a publicly available database?

Yes  No (required)

Does this Data Collection include sensitive information? Would the release of individually identifiable information create a risk of harm (e.g. psychological distress, social embarrassment, financial loss) greater than the risks that people experience in everyday life?

Yes  No (required)

# Distribution Method

Data Collections that are unsuitable for public download due to disclosure risk and data sensitivity issues may be published as restricted data. Users will be required to apply for access to the data.

Public download  Restricted Access

Unsure about whether to publish as restricted data or public download? Read our [FAQ](#) on the pros and cons of restricted data.



# Delayed Dissemination

Do you wish to delay the release of your data? openICPSR allows depositors to delay release of data for up to three years.

Yes  No Release Date:





# Local Economic Impacts of Coal Mining in the United States 1870 to 1970

Principal Investigator(s): Mike Matheis, Saint Anselm College

Version: V1

Name	File Type	Size	Last Modified
 <a href="#">Analysis1870to1976051816runOpenICPSRPost.do</a>	text/x-stata-syntax	44.6 KB	12/02/2016 05:18:AM
 <a href="#">AnalysisStateLevel010416C</a>			
 <a href="#">AnalysisStateLevel051816ru</a>			
 <a href="#">MiningManfUScounty1870to1976</a>			
 <a href="#">WhiteFBornAgeDist010416CoalProdrunNC.dta</a>	application/x-stata	3.5 MB	12/02/2016 05:20:AM

Release of this data has been delayed until 12/01/2018 at the request of the depositor.

Release of this data has been delayed until 12/01/2018 at the request of the depositor.

## Project Citation:

Matheis, M. (2016). Local Economic Impacts of Coal Mining in the United States 1870 to 1970 [Data set]. Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E100364V1>

**Persistent URL:** <http://doi.org/10.3886/E100364V1>

## Report a Problem

Found a serious problem with the data, such as disclosure risk or copyrighted content? Let us know.

## Project Description

**Project Title:** Local Economic Impacts of Coal Mining in the United States 1870 to 1970

**Summary:** This article expands upon the current "resource curse" literature by using newly collected county data, spanning over a century, to capture the short- and long-run effects of coal mining activity. It provides evidence that increased levels of coal production had positive net impacts on county-level population and manufacturing activity over an initial ten-year span, which become negative over the subsequent decades. The results provide evidence that any existence of a "resource curse" on local areas due to coal mining is a long-run phenomenon, and in the short run there are potential net benefits.



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

openICPSR data are distributed exactly as they arrived from the data depositor. ICPSR has not checked or processed this material. Users should consult the investigator(s) if further information is desired.

# ICPSR

SHARING DATA TO ADVANCE SCIENCE



## Thank you!

Justin Noble, Acquisitions Manager ([noblejus@umich.edu](mailto:noblejus@umich.edu))

[www.openicpsr.org](http://www.openicpsr.org)