

# Accessing Restricted Public Data

- Why use restricted-access data?
- Reasons for restricting access
- Means for accessing restricted data
- Facilitate access AND maintain safeguards
  - ICPSR Virtual Data Enclave
  - NORC Thin-Client Laptops
  - Cornell Restricted Access Data Center

# Why Use Restricted-access Data?

- Public-use statistics (tabulations, micro-data)
  - Not detailed enough
    - Aggregated when micro-data required
    - Categories coarsened
  - Leave out critical elements
    - Detailed geography
    - Family structure
    - Longitudinal structure

# Reasons for Restricting Access

- Confidentiality protection
  - Because of laws (Title 13, Title 26, HIPAA, FERPA, FISMA, etc.)
  - Because of contracts
  - Because of IRB
- Measures undertaken to protect confidentiality
  - Coarsening/collapsing of categories
  - Aggregation to higher levels
  - Swapping
  - Removal of longitudinal linkages

# Examples

- Coarsening/collapsing of categories
  - Geography on NLSY or HRS
- Aggregation to higher levels
  - County Business Patterns, QCEW, QWI, Decennial Census
- Swapping
  - Decennial Census, ACS
- Removal of longitudinal linkages
  - Canadian Survey of Labor and Income Dynamics (SLID)
  - Decennial Censuses, Current Population Survey
- Top-coding
  - Current Population Survey

# Solutions to Restricted Access

- Contracts
  - Regarding use of the data
  - Regarding local storage of the data
- Restricted access environments
  - Researcher has physical access
  - Researcher has remote access
  - Researcher has remote access by proxy

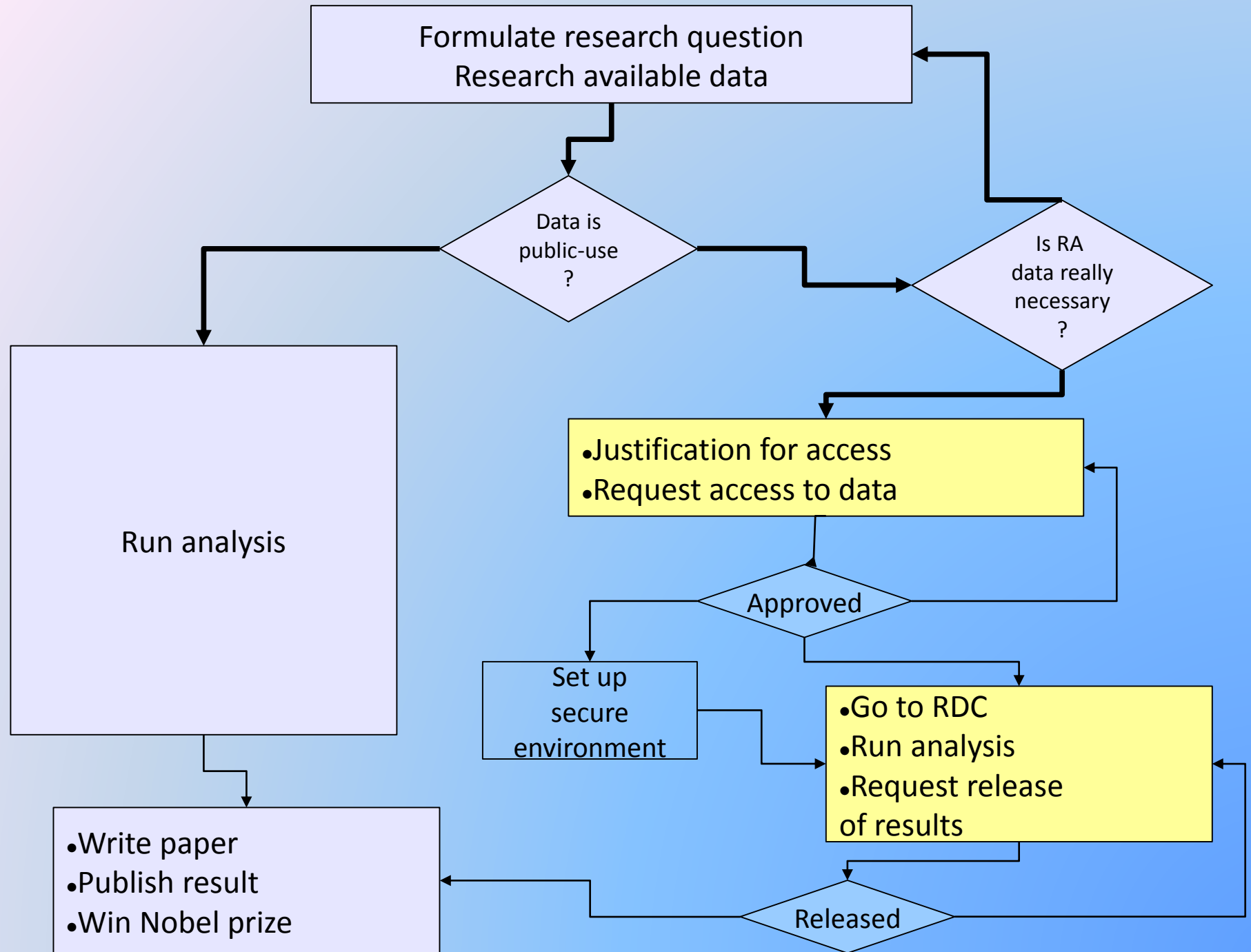
# Solutions to Restricted Access

## Contracts

- Regarding use of the data
  - *NLSY releases more detailed geodata with contractual engagement*
- Regarding local storage of the data
  - *HRS releases more detailed geodata, or linked data, with requirement to set up secure local environment*

# Solutions to Restricted Access

- Restricted-access environments
  - Researcher has physical access
    - RDC at Census, BLS, NCHS
  - Researcher has remote access by proxy
    - Remote processing facility at Statistics Canada, NCHS, IAB, LIS
    - Staff assisted (NCHS)
  - Researcher has flexible “direct” remote access
    - Cornell restricted access data center (one of the first)
    - NORC data enclave
    - ICPSR virtual data enclave





# Obtaining access to restricted data

- Identify and learn about the data
  - Often public-use documentation
  - Not always easy
- Familiarize yourself with the data custodian's requirements
  - Physical access rules
  - Legitimate use rules
- Contact the data custodian

# Cornell Restricted Access Data Center

## CISER Secure Data Services

---

\*\*\*\*\*  
You are attempting to access a computer system with restricted-access data operated by the Cornell Institute for Social and Economic Research (CISER) as a part of the Cornell Restricted Access Data Center (CRADC). If you do not have appropriate permissions you should not proceed. Unauthorized use of these data is subject to penalties imposed by Cornell University and by the providers of the data.

---

OK

# Secure Data Services



- **Assistance preparing data security plan**
- **Secure computing environment**
- **Access to sophisticated statistical tools**

# Restricted Versions of Popular Public Use Files

- Bureau of Labor Statistics NLSY series
- National Longitudinal Study of Adolescent Health
- New Immigrant Survey
- AHRQ's HCUP NIS and SID
- CMS Medicare
- State Vital Statistics Records
- Health and Retirement Study
- General Social Survey
- Fragile Families and Child Wellbeing Study
- Panel Survey of Income Dynamics

<http://ciser.cornell.edu/NYCRDC/>

 <b>Baruch</b> COLLEGE <small>The City University of New York</small>	<b>NYCRDC</b>	 Cornell University
<i>New York Census Research Data Center</i>		
<b>Welcome to the New York Census Research Data Center</b>		
<p>The New York Census Research Data Center (NYCRDC) provides academic researchers a unique opportunity to access selected Census confidential microdata in physically secure facilities at <a href="#">Cornell University</a> and at <a href="#">Baruch College</a>. This access follows approval of a suitable research project that benefits both the scientific community and the US Census Bureau.</p> <p>The NYCRDC is a consortium of 12 universities and research institutions in New York, New Jersey and Connecticut and it operates two locations, one at Baruch College and one at Cornell University as part of <a href="#">CISER</a> (Cornell Institute for Social and Economic Research). The NYCRDC at Baruch and at Cornell are 2 of 15 Census Research Data Centers in the US.</p> <p>The network of Census RDCs is managed by the <a href="#">Census Bureau's Center for Economic Studies (CES)</a>, which directs all proposal reviews. The NYCRDC is partially supported by the National Science Foundation (grant SEC-0322902 and ITR-0427889) and by research contracts from the US Census Bureau. The NYCRDC is committed to facilitate research that provides tangible benefits to the Bureau.</p>		
<hr/> <b>Announcements:</b>		
<p style="text-align: center;"><a href="#">RDC Administrator's contact hours</a></p>		
<small>Copyright © 2005-2013 New York Census Research Data Center  <a href="#">sitemap</a></small>		

<a href="#">Home</a>
<a href="#">News</a>
<a href="#">NYCRDC Consortium Members</a>
<a href="#">RDC Proposal Guidelines</a>
<a href="#">Research Policies and Procedures</a>
<a href="#">RDC Data</a>
<a href="#">Contacts</a>
<a href="#">Helpful Links</a>

# Cornell Course INFO 7470

**Link:**

[www.vrdc.cornell.edu/info7470/](http://www.vrdc.cornell.edu/info7470/)

**Audience:**

Doctoral level researchers interested in accessing confidential data of the US Census Bureau, NCHS and AHRQ

**Instructors:**

John M. Abowd and Lars Vilhuber

**Materials:**

Video recordings of lectures; slides; lab exercises