Navigating Data Licenses and Building a Research Plan to Access RDC Data

Andrew S. Hanks, PhD
The Ohio State University

What are barriers for researchers?

- ***Where*** to find the appropriate data
- ***Which*** data are linkable
- ***Accessing*** the data
- ***How*** to link the data

On the Agenda…

Objective: Provide information to reduce access barriers

Primary access barrier – data licenses

My foray into the world of licenses

Steps for obtaining a license

Navigating the “ultimate” license -- FSRDC

Primary Access Barrier – Data License

- Many researchers use publicly available data (see ERS Data Table, ICPSR, IPUMS)
- Excellent restricted access data available
  - Proprietary – owned by organization or company
  - Restricted versions of public access data (FoodAPS, CES, CPS, etc…)
  - Administrative data – usually government agency data

Restricted (non-public) data require a license…

Before Entering the World of Data Licenses

- First ask
  - “Do I need restricted access data for my work?”
- “No” -- Use the public data
- “Yes” – begin the process

It all began when…

Market outcomes for interdisciplinary researchers → Data?

A3. What was the primary field of study for your research doctoral degree?
Do not use acronyms or abbreviations.

Field Name

A4. Was your dissertation research (or performance, project report, or music or literary composition) interdisciplinary?

☐ Yes
☐ No → GO TO A6

A5. *(If Yes to interdisciplinary research)* Please list the fields of study for your dissertation research.

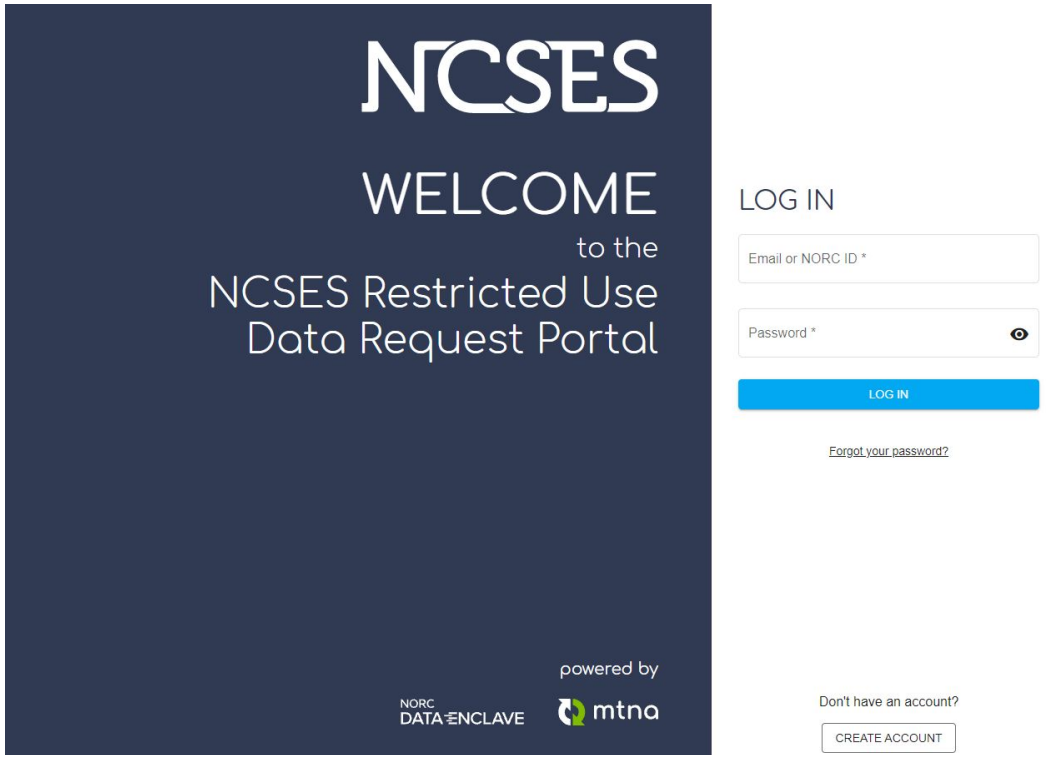
Primary Field: Field Name
Field 2: Field Name
Field 3: Field Name

B15. What will be your basic annual salary for this principal job?
If you are not salaried, please estimate your earned income. Do not include bonuses or additional compensation for summertime teaching or research.
Please enter a whole number without any commas, decimals or special characters.

Annual Salary/Earned Income:

Application Process

- Complete application for restricted use data
  - Access website (not available in 2012)
  - Project proposal – data, variables, RQs, analytical plan
  - Sign NDA and other forms (need notarized form) – includes disclosure review
  - Need university rep signature



Institution:··[Insert·Name]¶
Principal·Researcher:··[Insert·Last·Name·of·Principal·Researcher]¶

Attachment·#1:·Data·Requirements¶

Date:···[please·use·MM/DD/YYYY·format]¶

The·following·files·are·available·for·licensing·from·the·National·Center·for·Science·and·Engineering·Statistics·(NCSES).·Please·select·the·restricted·use·data·files·that·you·would·like·to·license:¶

Doctorate·Records·Files¶

Each·year,·the·data·from·the·Survey·of·Earned·Doctorates·(SED)·are·compiled·into·the·Doctorate·Records·File·(DRF).·The·DRF·is·a·historical·record·of·doctorate-degree·graduates.·Begun·in·1920,·the·DRF·contains·annual·information·used·to·track·the·number·of·graduates·in·various·files·and·other·information·related·to·the·U.S.·doctorate·earning·population.¶

☐···DRF·(through·2020)¶

Early·Career·Doctorate·Survey·¶

The·Early·Career·Doctorates·Survey·(ECDS)·gathers·in-depth·information·about·individuals·who·earned·their·first·doctoral·degree·(PhD,·MD,·or·equivalent)·in·the·past·10·years·and·work·at·academic·institutions·and·federally·funded·research·and·development·centers.·Unique·in·scope,·the·ECDS·includes·professional·and·research·doctorate·holders·from·all·fields·trained·in·the·United·States·and·abroad.·Sponsored·by·NCSES·and·by·the·National·Institutes·of·Health,·the·ECDS·provides·new·data·on·the·demographics,·work·experiences,·and·career·paths·of·individuals·in·the·first·years·following·the·completion·of·their·doctoral·studies.·Data·for·the·ECDS·survey·is·available·for·the·2013·pilot·study·and·the·2017·full·survey.¶

☐···ECDS·¶

National Science Foundation National Center for Science and Engineering Statistics

LICENSE FOR RESTRICTED-USE DATA

Survey files requested:

WHEREAS, the **National Center for Science and Engineering Statistics** (NCSES) of the National Science Foundation (NSF) has collected individually identifiable information in the above named survey(s), the confidentiality of which is protected by the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), 44 U.S.C. 3501 et. sec.; section 14(i) of the National Science Foundation Act of 1950, as amended, 42 U.S.C. 1873(i); and/or the Privacy Act of 1974 as amended, 5 U.S.C. 552a; and

WHEREAS, NCSES wishes to make the data available for use by qualified and capable researchers collaborating with NCSES and/or other divisions of the NSF exclusively for statistical purposes so long as the data are appropriately used and protected in accordance with the terms and conditions stated in this License and all prevailing laws and regulations;

appendix_e.pdf - Adobe Acrobat Pro DC (32-bit)
File Edit View E-Sign Window Help
Home Tools NSF NCSES AFFIDA... appendix_c.pdf appendix_e.pdf
2 / 2 125%

NSF/NCSES AFFIDAVIT OF NONDISCLOSURE

I, _____ , do solemnly swear (or affirm) that I have read and understand the content of the NSF/NCSES LICENSE FOR RESTRICTED-USE DATA are protected under the Confidential Information Protection and Statistical Efficiency Act of 2002, 44 U.S.C. 3501 et. sec.; section 14(i) of the National Science Foundation Act of 1950, as amended, 42 U.S.C. 1873(i); and/or the Privacy Act of 1974 as amended, 5 U.S.C. 552a and when given access to the restricted use data (hereinafter referred to as the subject file) from the Survey of Doctorate Recipients, the Survey of Earned Doctorates, the National Survey of Recent College Graduates and/or the SESTAT Integrated Database:

- I will not use the data in the subject file for any purpose other than statistical reporting and analysis. Information from the file will be released only in statistical summaries that do not disclose information about any individual.

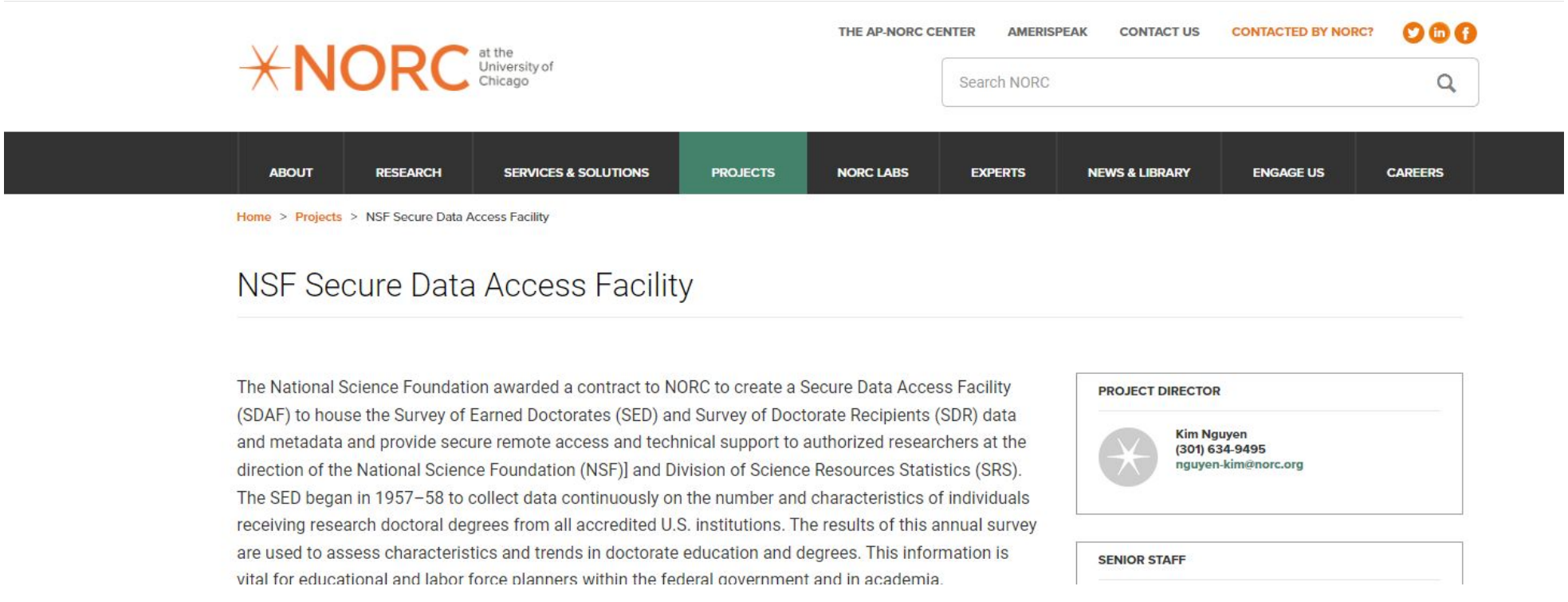
- I will not release the subject file or any part of it to any other person or organization.

- I will not use the subject file to attempt to learn the identity of, or to gain information concerning, any person included in the data set; unless such identification is directly related to survey administrative tasks or quality control, arising from a comparison with NIH administrative data.

If the identity of any individual in the file should be discovered inadvertently, then a) I will make no use of this knowledge, (b) I will advise the Director, NCSES, of the incident, (c) the information about the

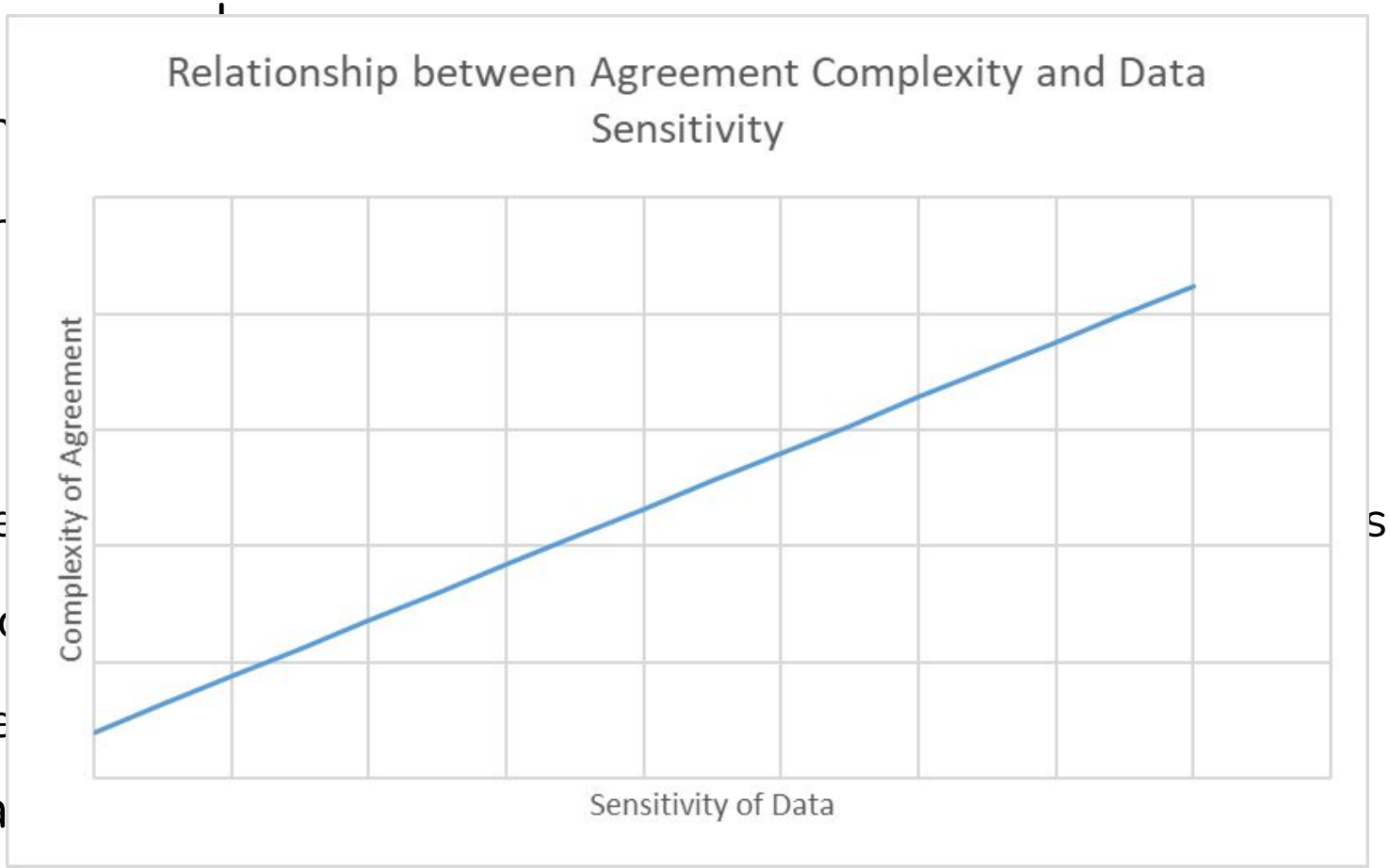
Accessing the Data

- Several encryption layers to access NORC server
- Once in the sever, access data, variety of statistical packages, and other software
- Disclosure process to get results



Unpacking the Process

- Developing
  - Researc
  - Identifyi
  - IRB
- Signing the
  - Data Use
  - May nee
  - Agreeme
- University a
- Receive data or get access to remote server (NORC, RDC)
  - For the RDC, you need to travel to the actual site

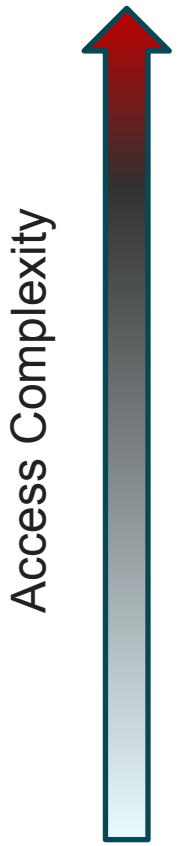


s Agreement

Unpacking the Process Continued

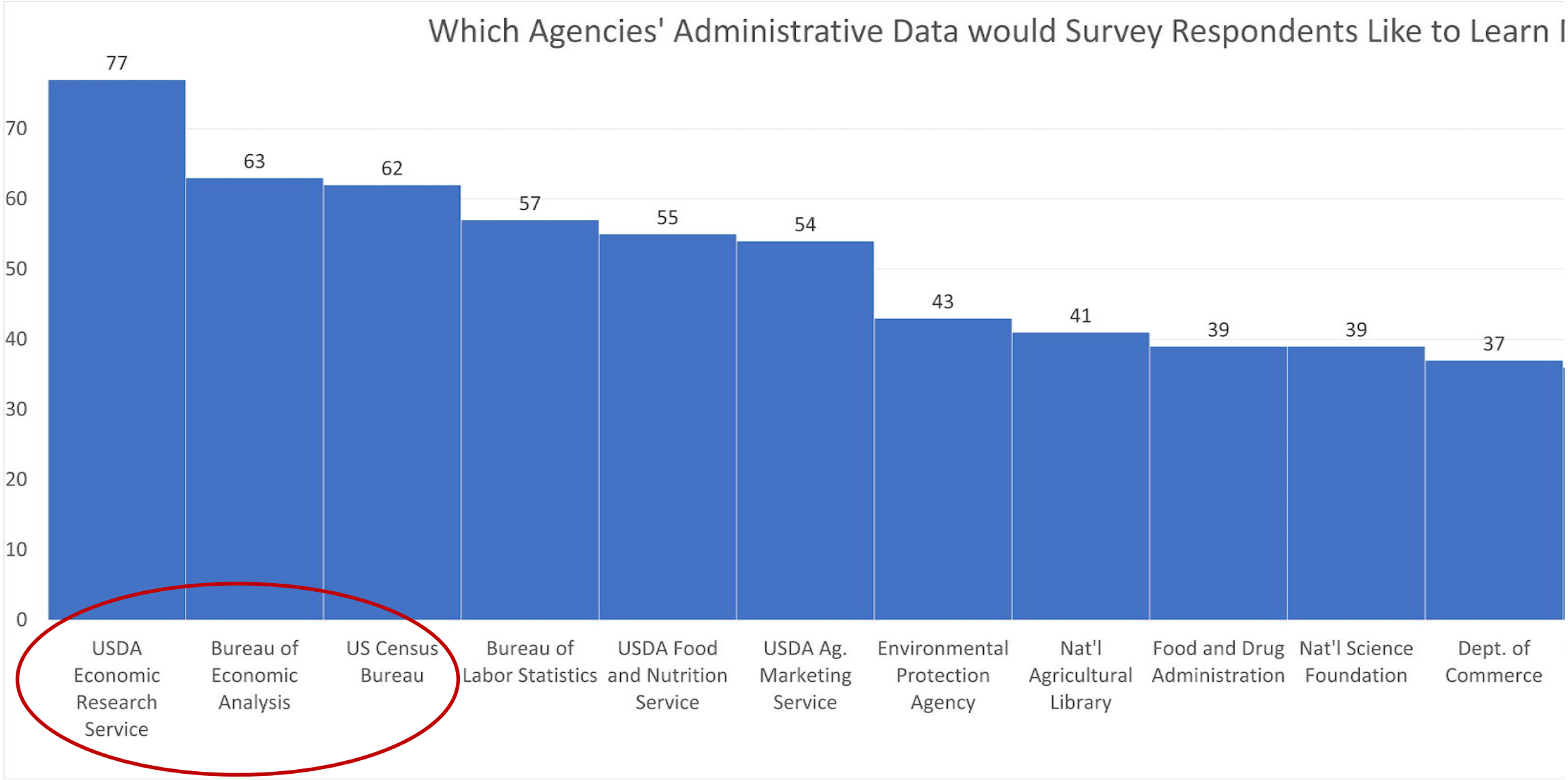
- Disclosure process – getting results
  - Write draft first OR disclose results first
  - Remember, NDAs protect privacy
- Maintaining the license
- CITE DATA, provide disclaimer, report final outputs

Active Agreements I Hold/Am On



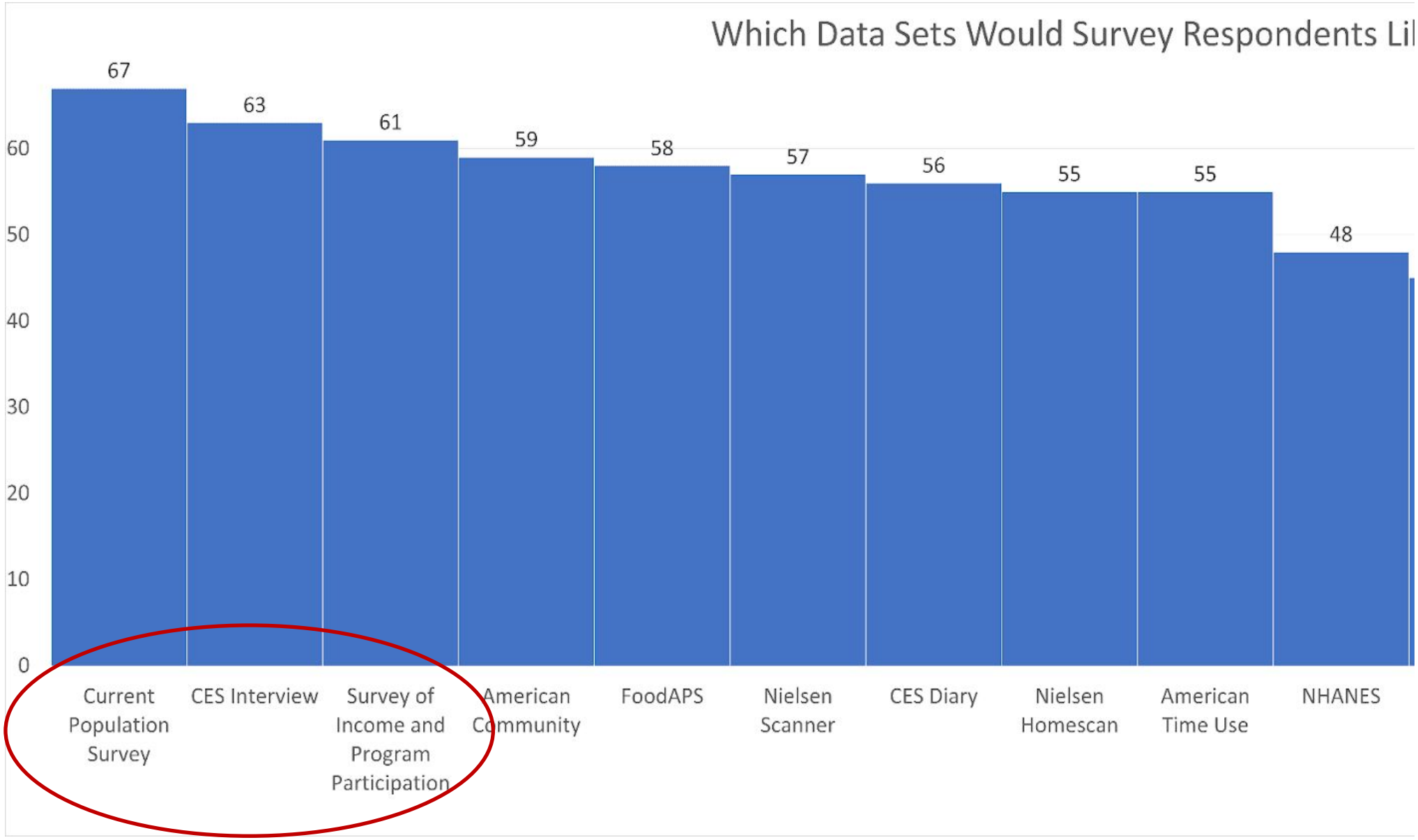
Agency	Data	Remote Access	Local Access (hard drive)	$$$
RDC	SED/LEHD/UMETRICS	x		Maybe
NSF	SED, SDR, ISDR	x		
IRIS	UMETRICS	x		Maybe
Kilts Center for Marketing	Nielsen Consumer Panel		x	x
OH WIC Agency	OH WIC Admin Data		x	
OSU Center for Human Resource Research	Ohio Food Bank Data		x	
Rural Action	Chester Hill Produce Auction Data		x	

Interest in Agencies…



USDA Data

Interest in Data



Federal Statistical Research Data Centers (FSRDC)

- Why access RDCs?
  - Lots of donuts
  - Link admin data, such as employment records (LEHD), to survey data
  - Large samples
  - Detailed individual-level information
- Considerations
  - Time – must go to RDC
  - Locations

https://www.census.gov/about/adrm/fsrdc.html

Which Data Are Available Through an RDC?

- https://www.census.gov/about/adrm/fsrdc/about/available_data.html

Considerations to Gain Access to RDC

https://www.census.gov/about/adrm/fsrdc/apply-for-access.html

Project must:

- Provide benefits to Census Bureau programs – give examples
- Demonstrate scientific merit
- Require non–public data
- Be feasible given the data
- Pose no risk of disclosure

Process to Gain Access to RDC Data (the Census approach)

Research Proposal

- Contact RDC administrator to discuss project
- Proposal
  - Project description
- Predominant Purpose Statement
- If interested in Census data, Benefit to Census statement
- How long does it take?
  - From beginning to end 6-8 months (timing is variable)

What I Learned

Research Proposal

- Need enough detail in your methods, description, and analytical plan
  - Feasibility
  - Need non-public data
  - Confidentiality protection
  - Can add collaborators later

Benefit to Census

- My project – PSEO
- Health and food
  - Census snapshots
  - Utility of Census data for analyzing food assistance programs – develop new tool, new chart, etc.

What I Learned

Once the proposal is accepted

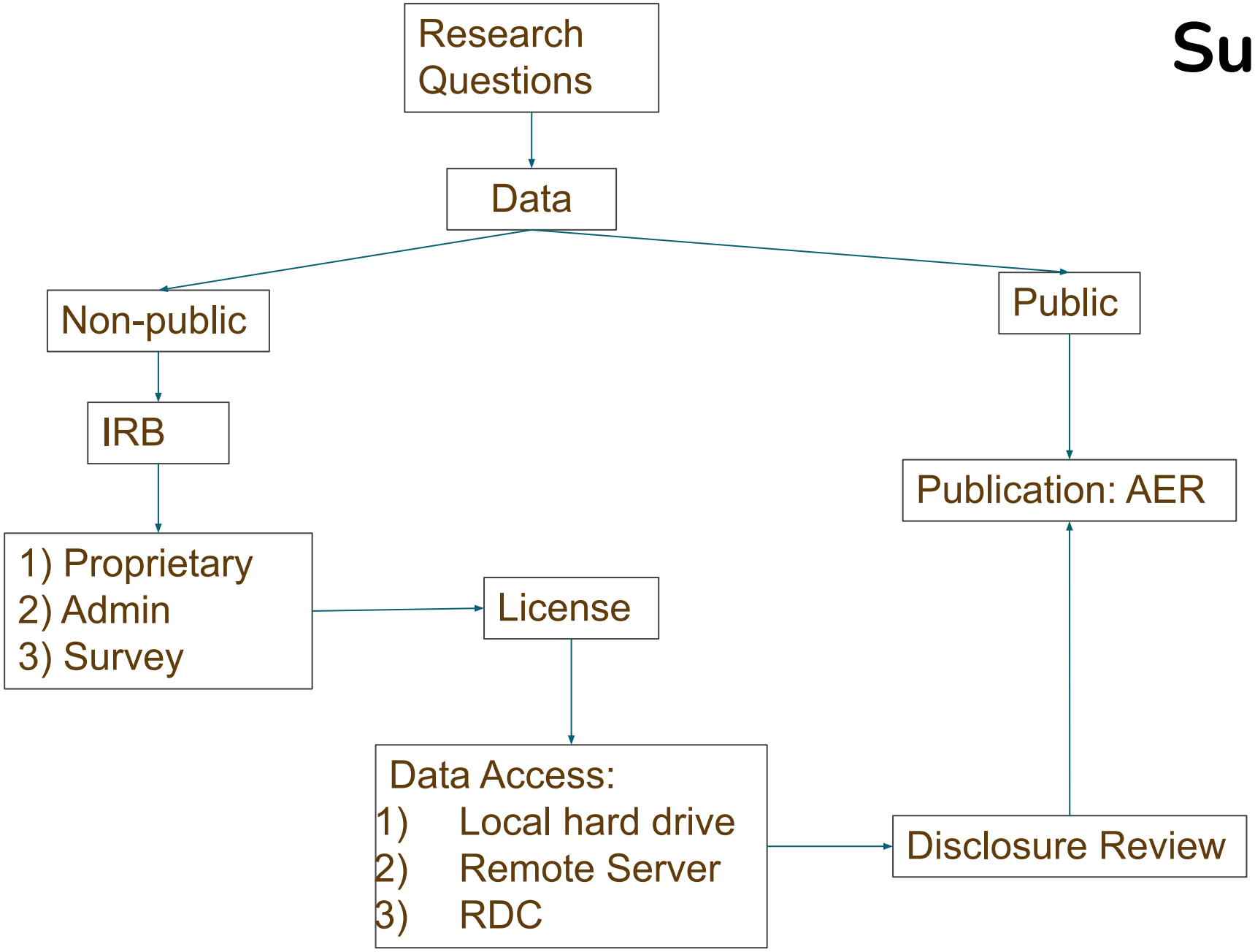
- More documents
- Training
- Complete as quickly as possible

Once you get SSS

- Access site (may need to pay access fee)
- Learn “ropes” of accessing remote server from the RDC
- Disclosure process – pretty big upfront time costs
    - Two layers – RDC administrator, Non-Disclosure officer

- Plan ahead – time is proportionate to complexity
- Don’t recreate the wheel
- Set reminders for license renewals
- Learn process at your university/institution

Summary


Research Questions
Data
Non-public
Public
IRB
Publication: AER
1) Proprietary
2) Admin
3) Survey
License
Data Access:
1) Local hard drive
2) Remote Server
3) RDC
Disclosure Review

Appendices

Acronyms

- DUA – data user agreement
- NDA – non-disclosure agreement
- TPAA – third party access agreement
- NORC – National Opinion Research Center
- UMETRICS – Universities: Measuring the Impacts of Research on Innovation, Competitiveness and Science
- ADRF – Administrative Data Research Facility
- FSRDC – Federal Statistical Research Data Facility

How to Access UMETRICS, Administered by IRIS

Overview of Application Process

- A research group leader or Principal Investigator should apply for data access using the Data Access Application Form found on the IRIS website (https://iris.isr.umich.edu/research-data/access/)
- Review of an application is typically completed within two weeks once all materials are received. If approved, all research group members will be asked to:

1. Sign the IRIS Data Use Agreement
2. Pay the administrative seat fee (applicable only for researchers not affiliated with an IRIS member university)
3. Complete brief online data security training
4. As part of the training, sign the IRIS Virtual Data Enclave Acceptable Use Policy
    - https://iris.isr.umich.edu/research-data/access/

Is access to the IRIS UMETRICS research data release available for non-research purposes such as program evaluation?

Yes. To apply for access to the IRIS UMETRICS data for purposes other than research, please submit the following materials to nbirk@umich.edu:

1. **A statement of the purpose for which access to data is required, including plans to post findings on websites or social media and publication (generally no more than 1 page)**
2. **Signature on the IRIS Restricted Data Use Agreement by user and an institutional official**
3. **An IRB determination letter from the IRIS member institution's Institutional Review Board**
4. **Completion of the IRIS VDE Acceptable Use Policy**
5. **Completion of IRIS VDE training**

How to Obtain a Restricted-Use Data License from NCSES

1. Navigate to the NCSES restricted-use data access online portal: https://nsfportal.norc.org/de-management-portal/account/login.
2. Create an account and submit a project application. The application should include data requirements, a research plan, and a list of the specific restricted data-use variables required. The data requirements and research plan form is available here: Data Requirements and Research Plan (DOCX 37KB).
3. Once the research plan and data requirements from the project application have been approved, applicants can download and complete the full license application from the online portal. For a license application to be considered complete, the researcher must (1) fill out the appropriate application forms, (2) obtain signatures from a senior official at the university who has the authority to enter the institution into a legal agreement with NSF and a system security officer who can attest to the security of the institution's systems, (3) complete the NCSES security training, and (4) execute affidavits of nondisclosure.
4. Once the license application is completed, applicants should submit it via the online portal. After approval, NCSES will provide instructions on accessing the restricted-use data through our Secure Data Access

How to Obtain Access to IRI Data

- Access to proprietary IRI data is limited to researchers collaborating on USDA-sponsored projects. USDA-sponsored projects include USDA grants, USDA cooperative agreements, and/or direct collaboration with USDA researchers on an issue of interest to the Department of Agriculture. For a sponsored research project, the institution affiliated with the research collaborator must enter into a Third Party Agreement (TPA) with the data provider. The language of the TPA is specified by the Vendor and must be signed as is.
- Steps to obtain a TPA:
- 1. The ERS point of contact works with collaborators to complete the USDA TPA Input Sheet. The form is returned to the ERS Data Steward to review and forward the form to IRI.
- 2. IRI will populate the actual TPA based on the information provided in the Input Sheet and send the populated TPA in pdf form to ERS for signatures.
- 3. A senior official needs to sign the TPA for the institution. The confidentiality clause holds the collaborator/institution responsible for the use of the licensed materials by any third party that they share the data with. ERS and collaborators shall maintain a record of each recipient requiring access to the data.
- https://www.ers.usda.gov/topics/food-markets-prices/food-prices-expenditures-costs/using-scanner-data/

How to Obtain Access FSRDC

Overview

- The Census Bureau considers proposals from qualified researchers in social science disciplines consistent with the subject matter of the surveys and censuses collected by the Census Bureau. Proposals may be submitted at any time. All researcher access to restricted–use data occurs at secure Federal Statistical Research Data Centers (FSRDC).
- 1. Proposed projects must provide benefits to Census Bureau programs
- 2. Demonstrate scientific merit
- 3. Require non–public data
- 4. Be feasible given the data
- 5. Pose no risk of disclosure
- Selected restricted-access data from the Agency for Healthcare Research and Quality (AHRQ), Bureau of Economic Analysis (BEA), Bureau of Labor Statistics (BLS), and the National Center for Health Statistics (NCHS) can also currently be accessed in the FSRDCs. Proposals to use those data must meet the requirements of those agencies.

Developing and Submitting a Research Proposal

- Before preparing your research proposal, consult
- 1. RDC administrator at the location where you want to base your project regarding access fees and the content and form of a proposal
- 2. Research Proposal Guidelines
- 3. Proposal Registration Form
- 4. RDC Predominant Purpose Statement Template
- 5. Writing Benefit Statements For Projects Accessing Confidential Data
- 6. The IRS Criteria Document
- https://www.census.gov/programs-surveys/ces/data/restricted-use-data/apply-for-access.html