

INFO 7470/ILRLE 7400

Access to restricted data

John M. Abowd and Lars Vilhuber
February 8, 2011

Outline

- Why use restricted-access data?
- Two examples of access
- Elements of the RDC proposal process
- Elements of proposal process for other RDC-accessible data
- What to expect once approved
- What is expected of you

Credits

- Rosemary Hyson
- Lynn Riggs

Why use restricted-access data

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

Reasons for restricting access

- Confidentiality protection
 - Because of laws
 - Because of contracts
 - Because of IRB
- Measures undertaken to protect confidentiality
 - Coarsening/collapsing of categories
 - Aggregation to higher levels
 - Swapping
 - Removal of longitudinal linkages
 - More sophisticated mechanisms later in this class

Examples

- Coarsening/collapsing of categories
 - Geography on NLSY or HRS
- Aggregation to higher levels
 - County business patterns
- Swapping
 - Decennial Census, ACS
- Removal of longitudinal linkages
 - Canadian Survey of Labor and Income Dynamics (SLID)
 - Decennial Censuses
- Topcoding
 - 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

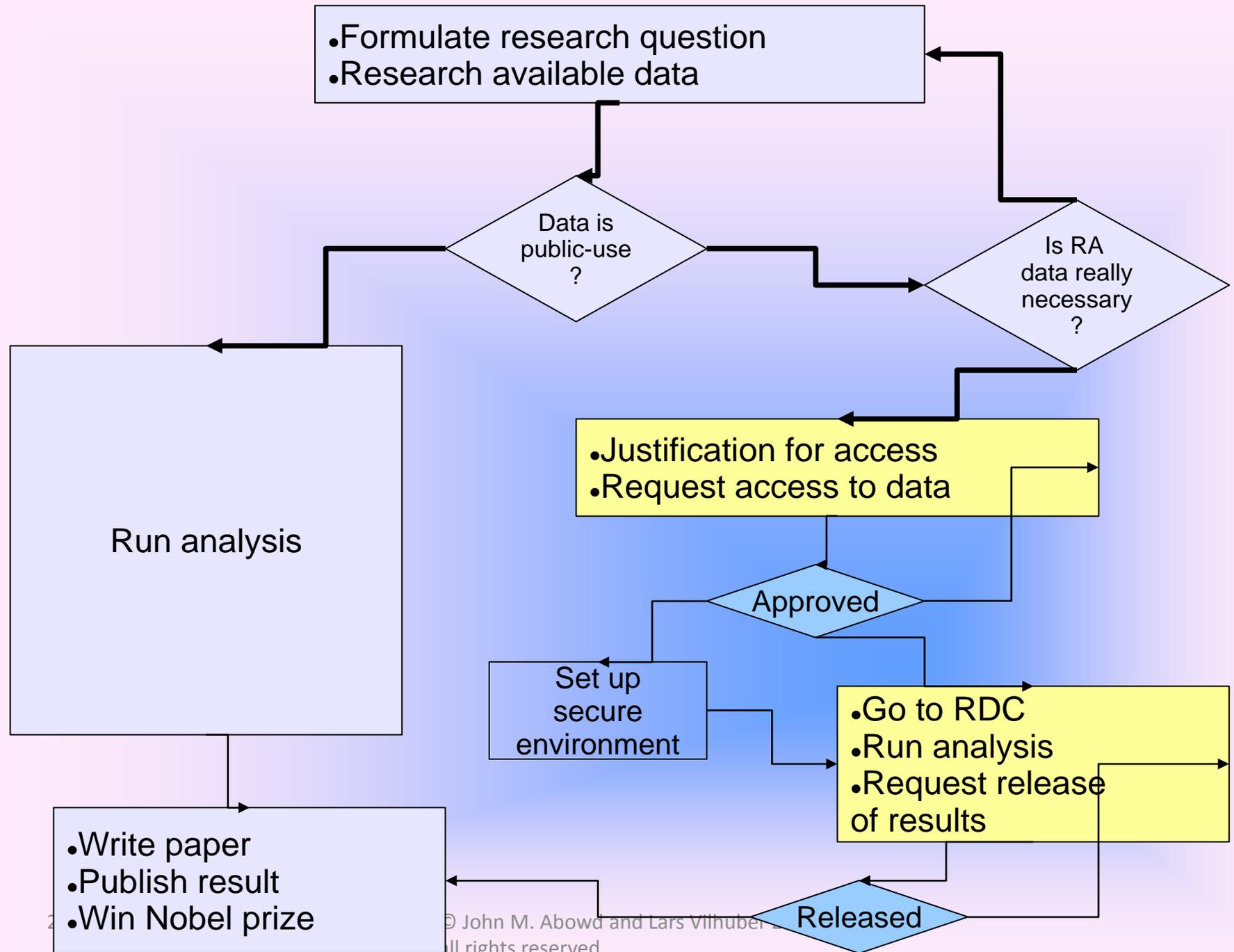
Solutions to restricted access

- 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
 - CRADC (one of the first)
 - NORC data enclave
 - German IAB
 - Researcher has remote access by proxy
 - Remote processing facility at Statistics Canada, at NCHS
 - Staff assisted (NCHS)

Workflow



Obtaining access to restricted data

- Identify and learn about the data
 - Often public-use documentation
- Familiarize yourself with the data custodian's requirements
 - Physical access rules
 - Legitimate use rules
- Contact the data custodian

Obtaining access to RDC data

- Identify and learn about the *RDC* data
 - Often public-use documentation (but not for Census)
- Familiarize yourself with the RDC program requirements
 - What does use of the RDC imply
 - How to justify access to the data
- Contact the RDC administrator
- Create an account on the CES web site

Steps for RDC proposal

- Initial proposal development
- Submission of formal RDC proposal
- Proposal review
- Approval (or not)
- Post approval process
- Post project activities

Initial proposal development

1. Identify and learn about RDC data

- CES Website
 - Data descriptions
- General Census website
 - Survey or program pages
- General research
 - Papers and articles that use RDC data (see CES Website for discussion papers, annual report)
- Personal contacts
 - RDC, CES, Census program staff
 - RDC and other researchers
- VirtualRDC

Things to keep in mind

- External data *can* be used with RDC data
 - (you may still need approval from other data custodians, though)
 - Timing, unit, link records need to be compatible
- Data not currently listed on CES website may also be available (-> personal contacts)
 - Space may be a restriction

Initial proposal development

2. Familiarize yourself with RDC requirements

- *Predominant* purpose must be to increase the utility of Title 13, Chapter 5 programs
 - ... always!
- Scientific merit
 - Funding award from a competitive peer review process is sufficient
- Clear need for non-public data
- Feasibility
 - Given methodology, requested data, and environment
- No risk of disclosure (ex ante)
 - Still ex-post evaluation in release process

Things to keep in mind

- RDC environment is different
 - Linux OS (see SSG with KDE environment to get a feel)
 - Restricted set of software
 - No free upload of programs or software to the system
 - Software very difficult
 - Programs through web-based staging process with review
 - All output, including screen output
 - Can only be viewed within the RDC
 - Can only be viewed by authorized persons (this includes your thesis advisor!) - in a narrow sense (listed and authorized on your project, and only that project)

Initial proposal development

3. Contact your RDC administrator

- The RDC administrator will accompany you the rest of the way
- The RDC administrator will also be responsible for
 - Reviewing output (round-robin assignment)
 - Reviewing input (program uploads, etc.)
- Bring them cookies...

Initial proposal development

4. Account in the CES project system

- Now called “*RDCProject*”
- At a minimum, upload CV
- Handles
 - Proposal review
 - Disclosure avoidance
 - Project follow-up

Search

whole site ▾

[GO](#)

- Quick Links**
- [Main CES Website](#)
 - [Employment Opportunities](#)
 - [Research Guidelines](#)
 - [Research Data](#)
 - [Research Locations](#)
 - [Discussion Papers](#)
 - [Forms & Documents](#)
 - [Site Map](#)

- Tasks**
- [Account Maintenance](#)
 - [Start a New Proposal](#)
 - [Change Password](#)

- Site Map**
- [Main Menu](#)
 - [Contact Us](#)
 - [Sign Out](#)
 - [My Account](#)
 - [News](#)
 - [Home](#)
 - [Research Program](#)
 - [Proposal Guidelines](#)
 - [Research Locations](#)
 - [Research Data](#)
 - [Publications](#)
 - [Discussion Papers](#)
 - [Census Home](#)
 - [Forms & Documents](#)
 - [Seminar Schedule](#)
 - [Site Map](#)
 - [Mission Statement](#)
 - [Policy Statements](#)

- [Research Program](#)
- [Publications](#)
- [Discussion Papers](#)
- [Clearance Request](#)
- [My Proposals](#)
- [Affiliated Proposals](#)
- [My Reviews](#)

Account Profile For: **Lars Vilhuber** Last Modified: 2011-02-08 01:43

[Change Password](#)

Choose Password

Confirm Password

Send Email Confirmation

[Change Password](#)

[Account Information](#)

Name	Lars Vilhuber
Email Address	lars.vilhuber@cornell.edu
Organization	Cornell University
Address	345 East Ives Hall
City	Ithaca
State	New York
Postal Code	14850
Country	United States

Change Account: Success

[Edit...](#)

[Upload Curriculum Vitae / Resume](#)

If you will be participating in a CES research project you must upload your **Curriculum Vitae / Resume** prior to submitting the research proposal. The absence of this document from any member of the research team will delay processing of the research proposal.

When uploading documents to our system, please make sure the documents are formatted as either **Adobe PDF (Portable Document Format)** or **RTF (Rich Text Format)**. If the system detects an invalid file format, the file will be ignored.

[Open Your Current Curriculum Vitae / Resume](#)

Curriculum Vitae / Resume

[Choose File](#) No file chosen

[Start Upload](#)

- [Census Home](#)
- [Forms & Documents](#)
- [Seminar Schedule](#)
- [Site Map](#)
- [Mission Statement](#)
- [Policy Statements](#)

Census Bureau Links: [Home](#) • [Search](#) • [Subjects A-Z](#) • [FAQs](#) • [Data Tools](#) • [Catalog](#) • [Census 2010](#) • [Quality](#) • [Privacy Policy](#) • [Contact Us](#)

Formal proposal submission

1. Elements of a formal proposal

- Preliminary proposal information in CES proposal management system
- Project description
- Predominant purpose statement
- Abstract

Formal proposal submission

2. Preliminary proposal information

- you must have:
 - Created a CES web site account
 - Read the RDC Proposal Guidelines
- Sign in to the CES web site and select “Start a new proposal”
- Read and agree to the prerequisite checklist

Search

whole site ▾

GO

Quick Links

- [Main CES Website](#)
- [Employment Opportunities](#)
- [Research Guidelines](#)
- [Research Data](#)
- [Research Locations](#)
- [Discussion Papers](#)
- [Forms & Documents](#)
- [Site Map](#)

Tasks

- [Account Maintenance](#)
- [Start a New Proposal](#)
- [Change Password](#)

Site Map

- [Main Menu](#)
- [Contact Us](#)
- [Sign Out](#)
- [My Account](#)
- [News](#)
- [Home](#)
- [Research Program](#)
 - [Proposal Guidelines](#)
 - [Research Locations](#)
 - [Research Data](#)
- [Publications](#)
- [Discussion Papers](#)
- [Census Home](#)
- [Forms & Documents](#)
- [Seminar Schedule](#)
- [Site Map](#)
- [Mission Statement](#)
- [Policy Statements](#)

[Research Program](#)

[Publications](#)

[Discussion Papers](#)

[Clearance Request](#)

[My Proposals](#)

[Affiliated Proposals](#)

[My Reviews](#)

Create a Proposal

To **begin** a new proposal select **Start Proposal**. If you are unaware of the research conducted by CES, or are not sure of the mechanics of the research proposal process, please review the [Research Guidelines](#)

[Start Proposal](#) →

Continue a Proposal

To **continue** working on a proposal, choose it from the list below. When uploading documents to our system, please make sure the documents are formatted as either **Adobe PDF (Portable Document Format)** or **RTF (Rich Text Format)**. If the system detects an invalid file format, the file will be ignored.

Action	PID	Title	Status	Date Created
			ACTIVE PROJECT	
Track	658	Older Workers' Displacement and Mobility	Abstract	11067
			Project Description	106635
			Predominant Purpose Statement (PPS)	38850
			ACTIVE PROJECT	
Track	538	LEHD Data Infrastructure Project (CES-445)	Abstract	48657
			Project Description	104365
			Predominant Purpose Statement (PPS)	33983

[Census Home](#)

[Forms & Documents](#)

[Seminar Schedule](#)

[Site Map](#)

[Mission Statement](#)

[Policy Statements](#)

Census Bureau Links: [Home](#) · [Search](#) · [Subjects A-Z](#) · [FAQs](#) · [Data Tools](#) · [Catalog](#) · [Census 2010](#) · [Quality](#) · [Privacy Policy](#) · [Contact Us](#)

USCENSUSBUREAU

Helping You Make Informed Decisions

Formal proposal submission

2. Preliminary proposal information

- Summarizes proposal information:
 - Researchers involved and affiliations
 - Data sets requested—RDC and external
 - Criteria met by proposed benefits
- Short proposal abstract
- Duration and funding for project
 - Note: extensions rarely granted, except for journal revisions
- RDC location(s) for project

Formal proposal submission

2. Preliminary proposal information

- Criteria met by proposed benefits
 - To the Census Bureau, benefits might be
 - New estimates derived from existing data
 - Expanded measurement capabilities
 - Improved documentation
 - Consultation with leading researchers
- (from Riggs presentation, 2009, in Atlanta)
- What is the catch?

Formal proposal submission

2. Preliminary proposal information

- Criteria met by proposed benefits
 - Under Title 13, Chapter 5, benefits **are**
 - New estimates derived from existing data - **yes**
 - Expanded measurement capabilities - **yes**
 - Improved documentation - **yes**
 - Consultation with leading researchers - **no**

(from Riggs presentation, 2009, in Atlanta)

Research Benefits (choose one or more)

In this step, you will be asked to indicate how your research project best benefits the Census Bureau. Proposals must demonstrate that the research is likely to provide one or more Title 13 benefits to the Census Bureau. A research project must demonstrate that its predominant purpose is to benefit Census Bureau programs. If a project has as its predominant purpose one or any combination of the following, it will be considered to have as its predominant purpose increasing the utility of Title 13, Chapter 5 data.

- Evaluate Concepts/Practices Re Statistical Data Collection/Dissemination**
Evaluating concepts and practices underlying Census Bureau statistical data collection and dissemination practices, including consideration of continual relevance and appropriateness of past Census Bureau procedures to changing economic and social circumstances. (This is not to be used when you have FTI data sets)

- Analyze Demo/Social/Econ Processes Affecting Programs and/or Improve Product Quality**
Analyzing demographic and social or economic processes that affect Census Bureau programs, especially those that evaluate or hold promise of improving the quality of products issued by the Census Bureau. (This is not to be used when you have FTI data sets)

- Increase Utility of Data for Analysis**
Developing means of increasing the utility of Census Bureau data for analyzing public programs, public policy, and/or demographic, economic, or social conditions. (This is not to be used when you have FTI data sets)

- Conduct or Facilitate Data Collection/Processing/Dissemination**
Conducting or facilitating Census Bureau census and survey data collection, processing or dissemination, including through activities such as administrative support, information technology support, program oversight, or auditing under appropriate legal authority. (This is not to be used when you have FTI data sets)

- Understanding or Improving Data Quality**
Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census, or estimate.

- Improved Methodology**
Leading to new or improved methodology to collect, measure, or tabulate a Title 13, Chapter 5 survey, census, or estimate.

- Enhancing Data**
Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example, improving imputations for non-response, or developing links across time or entities for data gathered in censuses and surveys authorized by Title 13, Chapter 5.

- Business Register**
Identifying the limitations of, or improving, the underlying Business Register, Household Master Address File, and industrial and geographical classification schemes used to collect the data.

- Data Collection**
Identifying shortcomings of current data collection programs and/or documenting new data collection needs.

- Sample Frame Maintenance**
Constructing, verifying, or improving the sampling frame for a census or survey authorized under Title 13, Chapter 5.

- New Estimates**
Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5.

- Estimating Non-Response**
Developing a methodology for estimating non-response to a census or survey authorized under Title 13, Chapter 5.

- Statistical Weights**
Developing statistical weights for a survey authorized under Title 13, Chapter 5.

Comments

Formal proposal submission

3. Project description

Typical components of any research proposal

- Background, discussion of relevant literature
- Proposed methodology
- Data sets

Components specific to the RDC program

- Summary of proposed benefits to Census Bureau programs (again)
- Description of the research outputs you plan to request be disclosed (used to assess ex-ante disclosure risk)
- Description of how you plan to construct the analysis data set if combining multiple data sets, including external ones (feasibility)
- Expected project duration (and cost)

Formal proposal submission

4. Predominant Purpose Statement

PPS and benefits to Census Bureau Title 13, Chapter 5 programs

- PPS describes and documents proposed benefits
- Benefits must be the predominant purpose of project
- PPS is template for certifying benefits at project end (Post Project Certification-PPC)

Formal proposal submission

Benefit criteria

- Evaluating concepts and practices
- Analyzing demographic/social/economic processes affecting Census

Bureau programs and/or Improve Product Quality

- Increasing the utility of Census Bureau data
- Conduct/facilitate data collection/processing/ dissemination
- Understanding and / or improving quality of data
- New or improved methodology to collect, measure, tabulate

Formal proposal submission

Benefit criteria (cont.)

- Enhance data
 - Improving imputations
 - Developing links across time or entities
- Business Register
- Identifying shortcomings of data collection
- Sampling frame maintenance
- New estimates
- Estimating non-response
- Statistical weights

Formal proposal submission

Benefit criteria (cont.)

- Useful links:
 - [“Research Opportunities at the U.S. Census Bureau”](#)
 - [“Writing Benefit statements”](#) (CES)
 - For projects involving Title 26 (tax) data:
[“IRS Criteria Document”](#)
- Other sources
 - RDC Administrator, Census Bureau program contacts
 - CES Annual Report, CES Discussion Papers
 - Summary of Census Bureau Research Problems Identified by Senior Staff

Formal proposal submission

Writing the PPS

- RDC administrator
 - Is alerted when preliminary proposal information is entered
 - Approves preliminary information
- Researcher
 - Clicks on “generate PPS” which
 - Lists benefit criteria from checklist
 - Lists FTI (Title 26 information) in requested data sets
 - Other proposal information
- Remember: cookies...

Formal proposal submission

The IRS...

- For projects involving FTI
 - IRS will review the PPS only – need to make sure it is clearly written
 - Group benefits in a logical way

Formal proposal submission

5. Processing documents

- RDC administrator follows the process
 - Considers proposal complete
- RDC-specific Project review also has to approve for submission
- Remember: cookies...

Formal proposal submission

6. Approval process

Reviews by

- CES and Census bureau subject matter and data experts
- external researchers (scientific merit)
- relevant Census Bureau programs/offices (benefits, feasibility)
- RDC Disclosure Officer and other disclosure experts (disclosure)
- other data custodians/sponsors:
IRS, HUD, BLS (benefits, feasibility)

Formal proposal submission

6. Approval process

- Duration
 - May be longer if involving FTI
 - At least 3-6 months as a guideline, but check with your RDC administrator
- Other issues
 - Special Sworn Status process

Post project approval

2. Badge

- Census Bureau computer ID
- Project account setup
- CES Account and Data Request Form

Post project approval

3. External Data

- external data form
- CD/media with external data
- use permission documentation

4. Researcher RDC orientation

--> Project finally starts!

Post project approval

1. SSS application

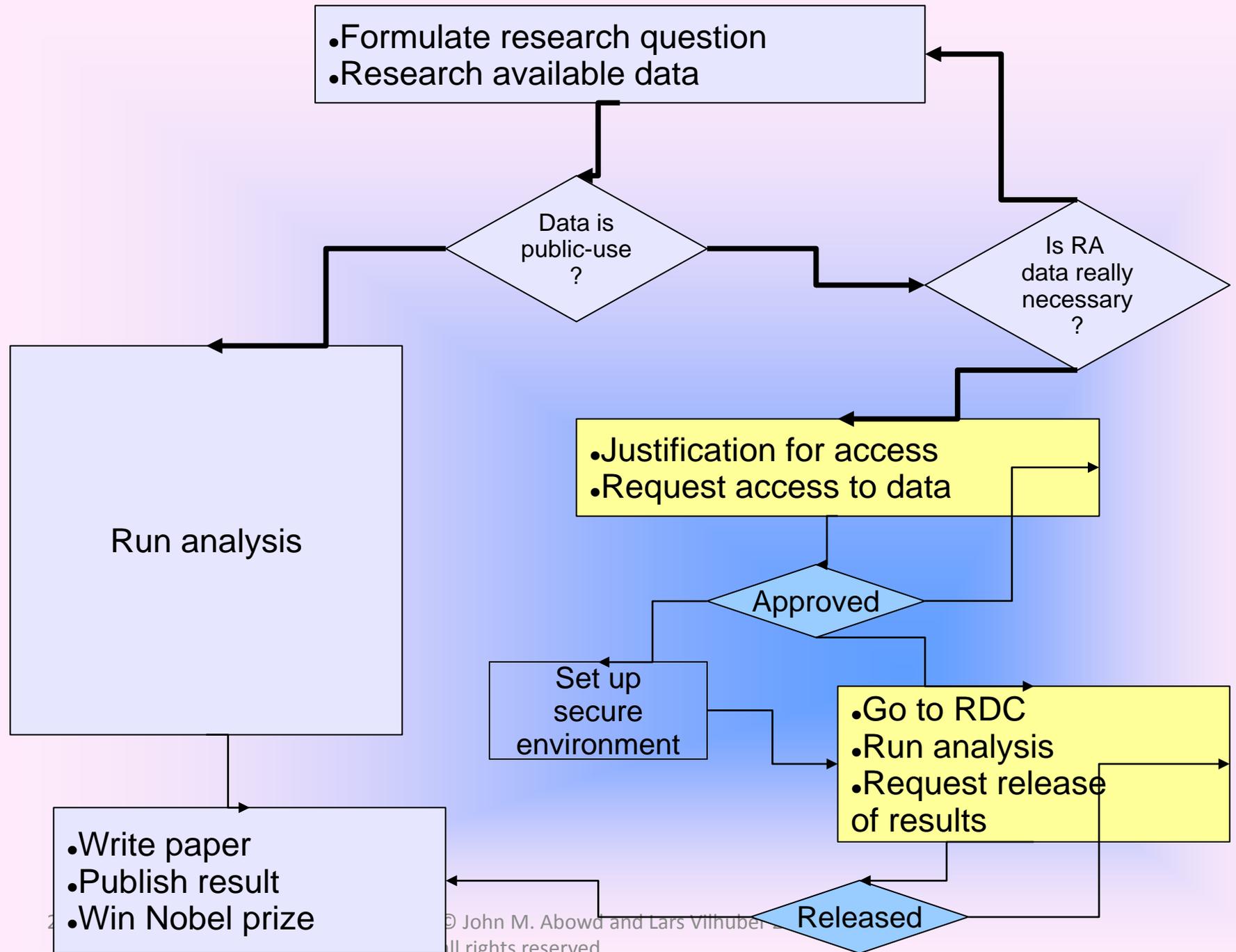
- Forms, fingerprinting and documentation
- Training
- RDC Administrator checks forms
- SSS package sent to Office of Security
- May send back corrections/clarifications

Health Data Available at the RDCs

- NCHS & AHRQ Data Extracts available
 - More relevant for non-DC RDCs
 - Possible to link NCHS and/or AHRQ data to Census data
 - May be more costly and time-intensive
 - May be difficult to justify the benefits
- MEPS-IC is collected by Census for AHRQ and considered Census data (Census application process)
 - Asks employers about provision of health insurance

NCHS version

- Several more [options](#)
 - NCHS RDC (Hyattsville, MD; Atlanta, GA)
 - More software choices
 - Census RDC
 - More locations, higher security approval burden
 - Remote access/submission
 - only SAS/Sudaan
 - Staff assisted
 - \$\$\$



NCHS version

- Similar process
 - Why not public use data
 - Proposal development
 - Benefit statement
 - CVs
 - Proposal submission
 - By email
 - Proposal review
- New choice
 - Access mode



Shell - Konsole

Session Edit View Bookmarks Settings Help

```
bash-3.1$
```

Shell

A terminal window titled 'Shell - Konsole' with a menu bar containing 'Session', 'Edit', 'View', 'Bookmarks', 'Settings', and 'Help'. The main area shows a shell prompt 'bash-3.1\$' with a cursor. The window has standard window controls and a taskbar at the bottom with a 'Shell' icon.

2
3 4

Shell - Konsole
Ganglia Cluster Toolkit: Clu



15:24
Tuesday
2011-02-08

SAS: Explorer

File Edit View Tools Solutions Help

Contents of 'SAS Environment'

- Libraries
- File Shortcuts
- Favorite Folders

SAS: Log-Untitled

File Edit View Tools Solutions Help

```

WARNING: DMS bold font metrics fail to match DMS font.

NOTE: AUTOEXEC processing beginning; file is /home/lv39/autoexec.sas.

NOTE: Libref HERE was successfully assigned as follows:
      Engine:          V9
      Physical Name:   /export/home/admin/extras

NOTE: AUTOEXEC processing completed.
    
```

Stata/IC 11.1

File Edit Data Graphics Statistics User Window Help

Review Command

Results

```

      _____ (R)
     / / / / /
    / / / / /
   / / / / /
  / / / / /
 / / / / /
/ / / / /
Statistics/Data Analysis

11.1 Copyright 2009 StataCorp LP
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.
979-696-4600 stata@stata.com
979-696-4601 (fax)

5-user Stata network perpetual license:
Serial number: 50110530211
Licensed to: Cornell University
Social Science Gateway

Notes:
1. (-m# option or -set memory-) 10.00 MB allocated to data
    
```

Variables

Target: Command window

SAS: ToolBox: Explorer

I