Introduction to the MnRDC

Gina Erickson
rdcadmin@umn.edu
U.S. Census Bureau
Minnesota Census Research Data Center Administrator
www.mnrdc.umn.edu

January 24, 2011

What are RDCs?

• Collaborative effort between Census and research institutions
  – Secure facilities to access non-publicly available data
  – 9 (+2) locations, 2 more planned, plus HQ

• Dissemination of secure data
  – Agreements with other agencies
    • NCHS
    • AHRQ
What is the MnRDC?

- 27 Willey Hall
  - 8 work stations, 1 conference room, RDCA office

- Computer security
  - Thin-client desktop machines
  - Secure router connects to server in Bowie, MD
  - Closed network (no internet, no physical access)

- Physical security
  - Badged entry, video monitoring, motion and glass-break alarms

Data Availability

- The MnRDC provides access to non-public versions of
  - decennial census and demographic surveys
  - economic censuses and surveys
  - health data
    - National Center for Health Statistics (NCHS)
    - Agency for Healthcare Research and Quality (AHRQ)
Data Availability

• Economic
  – Economic censuses
    • Finance-Insurance-Real Estate, Retail trade, Wholesale trade, Transportation-Communications-Utilities, Services, Mining, Manufactures, Construction industries
  – LBD, ILBD, Business Register
    • Universe of all U.S. business establishments with paid employees from 1976 to present (LBD) and without paid employees (ILBD)
    • Relational database of organizational units of 160,000, multi-establishment businesses (1.8 million establishments), 5 million single establishments, 21 million non-employer businesses (BR)

Data Availability

• Economic
  – LEHD
    • Matched employer-employee plus administrative data
  – MEPS-IC
    • Employer-sponsored health insurance data
  – Commodity Flow Survey and LFTTD
    • Domestic and foreign trade data
  – Numerous surveys...
    • SBO Survey of Business Owners
    • PACE Pollution Abatement Costs and Expenditures
    • R&D Survey Research, Development, Innovation
Data Availability

• Demographic
  – Decennial and ACS
    • Short and long form
  – AHS
    • Housing units and occupants (HUD data)
  – March CPS
    • Social and economic characteristics
  – National Longitudinal Surveys
  – SIPP
    • Government programs and participation

Data Availability

• Health
  – MEPC-HC
  – NHIS
  – NHANES
  – NSFG
Data Potentials

• New data
  – NCVS?
  – Recovered Data
    • CPS monthly supplements from 1962 forward
    • ASM as early as 1954
  – Older CPS supplements

• Data linkages
  – Internal and outside public-use/proprietary files
  – Across datasets
  – Administrative records (SSA)

Text Box 4-1.
RECOVERED DATA: HIGHLIGHTS
The economic and demographic data recovered from the Unisys will be valuable additions to the data already available at the Center for Economic Studies (CES) and the Research Data Centers (RDCs). Most files will require additional work before they can be used for research purposes, and some may require approval by sponsoring agencies. Examples of data recovered from the Unisys include:
• Earlier years of series already available at CES
  - Census of Mining, Retail, Wholesale, and Services
  - Annual Survey of Manufactures (ASM)
  - Survey of Industrial Research and Development (SIRD)
  - Survey of Minority-Owned Business Enterprises
  - Commodity Transport Survey (now called the Commodity Flow Survey)
  - Decennial Census data
  - Puerto Rico sample and complete count files
  - U.S. Possessions sample and complete count files
  - Selected Current Population Survey (CPS) March Supplements
• Series not currently available at CES
  - Agriculture Surveys and Censuses
  - Annual Survey of Oil and Gas
  - Housing Fuel Survey
  - Water Use Survey
  - Survey of Construction
  - Income Surveys Development Program
  - CPS Supplements for months other than March
• New variables for series already available at CES
  - Census of Manufactures (CM) special inquiries data
  - ASM Central Administrative Office data
  - ASM and CM data flags
• Historical analysis files (see Text Box 4-2)
  - Industrial Time Series data
  - Linked CM/SIRD data created for Zvi Griliches
Census vs. Health Data

• NCHS, AHRQ, Census each have own RDCs at HQ locations in DC
  – NCHS and AHRQ use existing network of Census RDCs
  – Census RDCs serve as hosts for other agency data

• Each agency has own proposal review process

RDC Process

• Ensure your project requires RDC data and read through the complete proposal guidelines.
• Discuss your project with the RDC Administrator and enter a brief preliminary proposal on the CES site.
• Work with the RDC Administrator in developing your final proposal.
• When invited by the RDC Administrator, submit your final proposal into the CES system.
  – The final proposal consists of four separate documents:
    • curriculum vitae of all investigators
    • abstract of the proposal
    • project description (full proposal)
    • statement of benefits to the Census Bureau (Predominant Purpose Statement)
• Once approved, you will need to obtain Special Sworn Status (SSS). This includes passing a background check and making a signed, sworn statement about preserving the confidentiality of the data.
**RDC Process**

- **Ensure your project requires RDC data and read through the complete proposal guidelines.**
- Discuss your project with the RDC Administrator and enter a brief preliminary proposal on the CES site.
- Work with the RDC Administrator in developing your final proposal.
- When invited by the RDC Administrator, submit your final proposal into the CES system.
  - The final proposal consists of four separate documents:
    - curriculum vitae of all investigators
    - abstract of the proposal
    - project description (full proposal)
    - statement of benefits to the Census Bureau (Predominant Purpose Statement)
- Once approved, you will need to obtain Special Sworn Status (SSS). This includes passing a background check and making a signed, sworn statement about preserving the confidentiality of the data.

**Need for RDC data**

- **Demographic and health data**
  - Lower levels of geography than public data
  - Less/no top-coding or recodes
  - Larger samples
  - More detailed disease codes
- **Many economic datasets not publically available**
Linking Datasets

- **Types of links/merges**
  - Unique individual identifiers (SSN, name, and/or address)
  - Probabilistic match at the individual level
  - Linking with other agency data
    - Ex: SSA data
  - Using other variables
    - Geography

- **Considerations:**
  - Lengthy process
    - Ex: Linking 2 mandatory surveys (ACS and Decennial)
  - Small samples
    - Ex: Survey to survey
  - Expensive
    - Ex: Merging with SSNs - external data must be PIK'd to match at individual level

Levels of Sensitivity

- **Low Sensitivity**
  - standard for most economic RDC projects

- **Moderate Sensitivity**
  - linkage between a mandatory demographic and a voluntary project

- **High Sensitivity**
  - linkage between 2 mandatory demographic projects

- **Sensitive Populations**
  - Children, elderly, limited English proficiency, non-citizens, group quarters, small minority groups

- **Sensitive Topics**
  - Abortion, illegal immigration, same sex partners, sexual behavior, sexual orientation
Feasibility: GIS limitations

- GIS software currently available to RDC users
  - Only LINUX based
    - SAS: GIS Mapping/Geocoding (9.2)/Spatial Regression (limited to point pattern analysis-Kriging)
    - GRASS: GIS Mapping/Spatial Modeling
    - R-spdep: New. Comprehensive Spatial Regression tools
    - STATA: Mapping/Spatial Regression (Pisati ado)
    - GeoDa?: Spatial Regression (basic)

RDC Process

- Ensure your project requires RDC data and read through the complete proposal guidelines.
- Discuss your project with the RDC Administrator and enter a brief preliminary proposal on the CES site.
- Work with RDC administrator to develop final proposal.
- When invited by the RDC Administrator, submit your final proposal into the CES system.
  - The final proposal consists of four separate documents:
    - curriculum vitae of all investigators
    - abstract of the proposal
    - project description (full proposal)
    - statement of benefits to the Census Bureau (Predominant Purpose Statement)
- Once approved, you will need to obtain Special Sworn Status (SSS). This includes passing a background check and making a signed, sworn statement about preserving the confidentiality of the data.
Discuss Project with RDC Administrator

- Feasibility
- Data availability
- Project timeline
- Costs

Costs

- Substantial costs associated with running an RDC
- Currently: fees covered by NSF grant
- Data-related fees
  - NCHS charges set-up fee
  - Merging data fees
    - National Death Index
    - PIKing outside data
RDC Process

- Ensure your project requires RDC data and read through the complete proposal guidelines.
- Discuss your project with the RDC Administrator and enter a brief preliminary proposal on the CES site.
- Work with the RDC Administrator in developing your final proposal.
- When invited by the RDC Administrator, submit your final proposal into the CES system.
  - The final proposal consists of four separate documents:
    - Curriculum vitae of all investigators
    - Abstract of the proposal
    - Project description (full proposal)
    - Statement of benefits to the Census Bureau (Predominant Purpose Statement)
- Once approved, you will need to obtain Special Sworn Status (SSS). This includes passing a background check and making a signed, sworn statement about preserving the confidentiality of the data.

Census: Proposal Writing

- Proposals generally different in form than traditional (NSF, etc.) proposals
  - Lit review a small part
    - No strong need to prove “it hasn’t been done before”
  - Reviewers trying to anticipate any disclosure issues in methodology section
    - Lots of dummy variables problematic
    - Small sample sizes can be an issue
  - Work through project details
    - Duration and intensity of project
    - Data needed, etc.
- Goal: show need for data!
Uniqueness of Census Data

- No IRB
- No informed consent
- Mandatory participation

- Laws govern use of non-publicly available data (title 13 and title 26)
  - Require “predominant purpose” be to benefit Census Bureau programs

Census: The PPS

- 13 possible criteria:
  1. Evaluating concepts and practices underlying Census Bureau statistical data collection and dissemination practices, including consideration of continued relevance and appropriateness of past Census Bureau procedures to changing economic and social circumstances;
  2. 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;
  3. Developing means of increasing the utility of Census Bureau data for analyzing public programs, public policy, and/or demographic, economic, or social conditions;
  4. Conducting or facilitating 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;
  5. Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census, or estimate;
  6. Leading to new or improved methodology to collect, measure, or tabulate a Title 13, Chapter 5 survey, census, or estimate;
  7. Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example: improving imputations for non-responsive, developing links across time or entities for data gathered in censuses and surveys authorized by Title 13, Chapter 5;
  8. Identifying the limitations of, or improving, the underlying Business Register, Master Address File, and industrial and geographical classification schemes used to collect the data;
  9. Identifying shortcomings of current data, collection programs and/or documenting new data collection needs;
  10. Constructing, verifying, or improving the sampling frame for a census or survey authorized under Title 13, Chapter 5;
  11. Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5;
  12. Developing a methodology for estimating non-response to a census or survey authorized under Title 13, Chapter 5;
  13. Developing statistical weights for a survey authorized under Title 13, Chapter 5.
Census: Proposal Review

- Benefit to Census Bureau programs under Title 13
- Scientific Merit in that the research will contribute to existing knowledge
- Clear Need for Non-Public Data
- Feasibility of success
- Acceptance of all confidentiality protection and disclosure avoidance review requirements

NCHS and AHRQ Proposal Review

- Work with agency Data Center coordinators/administrators
  - Ray Kuntz at AHRQ
  - Peter Meyer at NCHS
- Process generally more iterative than Census
RDC Process

- Ensure your project requires RDC data and read through the complete proposal guidelines.
- Discuss your project with the RDC Administrator and enter a brief preliminary proposal on the CES site.
- Work with the RDC Administrator in developing your final proposal.
- When invited by the RDC Administrator, submit your final proposal into the CES system.
  - The final proposal consists of four separate documents:
    - Curriculum vitae of all investigators
    - Abstract of the proposal
    - Project description (full proposal)
    - Statement of benefits to the Census Bureau (Predominant Purpose Statement)

- Once approved, you will need to obtain Special Sworn Status (SSS). This includes passing a background check and making a signed, sworn statement about preserving the confidentiality of the data.

Special Sworn Status

- Online trainings and certifications
- Off-line paperwork and documentation
- Background check
  - Submitted online and followed with interview
    - Residential history
    - Foreign travel
    - Education and employment history
    - References
- Fingerprints and passport photos (for badges)
  - Requires trip (or 2) downtown to USDA
Disclosure Avoidance Review

• Performed by RDCA and agency disclosure officer
  – No output can leave RDC without review
  – Review process worked out in proposal stage
  – Researchers should plan for time
    • Turn around generally 1 week
  – Intermediary data generally not released
  – Descriptive data can be problematic
  – Sensitive topics and populations will likely require approval from Disclosure Review Board

RDC Process

• Timeline?
  – Proposal writing
    • Multiple drafts between RDCA and researchers
    • PPS
  – Plan at least 6 months from submission
    • 3-4 months for proposal review
    • 2 months for SSS processing
  – Project duration
    • 36-60 months
AHRQ and NCHS Process

- Ensure your project requires restricted NCHS data or AHRQ data and read through the complete proposal guidelines at NCHS and AHRQ.
- Contact the MnRDC Administrator to ensure your project is feasible for completion in a Census RDC.
- Work with the NCHS RDC Director or AHRQ Data Center Coordinator to prepare your proposal.
- Notify the MnRDC Administrator once your project has been approved to begin the process of data access in the MnRDC (obtaining Special Sworn Status)
- Disclosure done by servicing agency

More Information

- Course on RDC data
  – Cornell University course on social and economic data
  – Distance learning classroom at U of M
  – Tuesdays from 3:00 to 5:30 pm. – January 18 through May 3.
  – Contact Gina (rdcadmin@umn.edu) if interested
RDC Funding Opportunity

- **MnRDC small grants program**
  - Proposals must be submitted to the relevant federal agency by March 31, 2011

Questions?

- [www.mnrdc.umn.edu](http://www.mnrdc.umn.edu)
- [www.ces.census.gov](http://www.ces.census.gov)