

Center for Economic Studies Research Data Centers

Arnold P. Reznek

Research Data Center Administrator

Center for Economic Studies

U.S. Census Bureau

Room 207 WP2

Washington, D.C. 20233-6300

301-763-1856

Email: arnold.phillip.reznek@census.gov

Presented July 23, 2003 as part of the Bureau
of Transportation Statistics Confidentiality
Seminar Series

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Acknowledgments and Disclaimer

- **This presentation was developed using material from many sources, and using parts of several presentations by CES staff members. I thank them all.**
- **Senior CES and Census Bureau management have not reviewed this presentation. Opinions expressed are mine and do not necessarily represent official Census Bureau positions. I am responsible for any errors.**

Center for Economic Studies/ Research Data Centers:

- **Background -- history, context, objectives**
- **Growth and change**
- **Current operating model and logistics**
- **Data**
- **Challenges and conclusions**

History:

- **Benefits of Microdata**
 - Research Use
 - Benefits to Census Bureau programs
- **Constraints on Microdata**
 - Title 13 U.S.C. protects confidentiality

Title 13 U.S.C., Sec. 9:

(a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison may...

Title 13 U.S.C., Sec. 9:

- (1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or
- (2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or

Title 13 U.S.C., Sec. 9:

- (3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports. No department, bureau, agency, officer, or employee of the Government, except the Secretary in carrying out the purposes of this title, shall require, for any reason, copies of census reports which have been retained by any such establishment or individual. Copies of census reports which have been so retained shall be immune from legal process, and shall not, without the consent of the individual or establishment concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.

History: (continued)

- **PUMS in household data (demographic)**
 - Restricted data
- **Skewed distributions on business side (economic)**
 - Restricted Access

Title 13 U.S.C., Sec. 23: Special Sworn Status (SSS)

(c) The Secretary may utilize temporary staff, including employees of Federal, State, or local agencies or instrumentalities, and employees of private organizations *to assist the Bureau in performing the work authorized by this title, but only if such temporary staff is sworn to observe the limitations imposed by section 9 of this title.*”

Objectives of RDC Program:

- **Provide access to microdata while maintaining confidentiality**
- **Produce research to improve Census Bureau programs (required by law)**
- **Conduct research in public interest**
- **Improve data and create new data products**
- **Expand Census Bureau's research community**

Historical Vignette: LRD

(How it should work -- Partnership)

- **Longitudinal Research Database (LRD) – Census of Manufactures and Annual Survey of Manufactures**
 - Close cooperation of CES staff and small group external researchers
 - Developed new data product
 - Rich research potential
 - Census Bureau captured and continues to capture knowledge

Impetus for Growth:

- **Program of external access viewed as a success by both the Census Bureau and the academic community.**
- **Interest in establishing remote sites for external access.**

Legal foundation for growth:

- **Title 15 U.S.C., Sec. 1525:**
 - The Secretary of Commerce is authorized, upon the request of any person, firm, organization, or others, public or private, to make special studies on matters within the authority of the Department of Commerce; to prepare from its records special compilations, lists, bulletins, or reports; to perform the functions authorized by section 1152 of this title; and to furnish transcripts or copies of its studies, compilations, and other records; upon the payment of the actual or estimated cost of such special work.

Legal foundation for growth (cont'd)

- **Title 15 U.S.C., Sec. 1525 (cont'd):**
 - In the case of nonprofit organizations, research organizations, or public organizations or agencies, the Secretary may engage in joint projects, or perform services, on matters of mutual interest, the cost of which shall be apportioned equitably, as determined by the Secretary, who may, however, waive payment of any portion of such costs by others, when authorized to do so under regulations approved by the Office of Management and Budget.

Mechanism for growth

- **National Science Foundation as partner in choosing new sites**
- **Federal Register announcement**
- **Potential RDC partners (sites) submit application to NSF**
- **Review of application at NSF and Census Bureau**

Research Data Centers:

- **Boston RDC (NBER) - 1994**
- **Carnegie Mellon RDC - 1997**
- **UC Berkeley RDC- 1999**
- **UCLA RDC - 1999**
- **Triangle RDC (Duke) - 2000**
- **Michigan RDC (University of Michigan) - 2002**
- **Chicago RDC (Northwestern University, University of Chicago, UIC, Argonne Lab, Federal Reserve Bank of Chicago)- 2002**

1999 Internal Revenue Service (IRS) Safeguard Review

- **Issues:**
 - Authorized Use of Federal Tax Information (FTI)
 - Use of Special Sworn Status at CES

Title 26 U.S.C., Section 6103(j)(1)(A)

- **Upon request in writing by the Secretary of Commerce, the Secretary shall furnish**
- **(A) such returns, or return information reflected thereon, to officers and employees of the Bureau of the Census**
- **and as the Secretary may prescribe by regulation for the purpose of, but only to the extent necessary in, the structuring of censuses and national economic accounts and conducting related statistical activities authorized by law.**

Code of Federal Regulations Sec. 301.6103(j)(1)-

(2) Officers or employees of the Service will disclose to officers and employees of the Bureau of the Census for purposes of, but only to the extent necessary in, conducting, as authorized by chapter 5 of title 13, United States Code, demographic, economic, and agricultural statistics programs and censuses and related program evaluation--

Safeguard Review Responses:

- **September 15, 2000 “Criteria Document” (benefits to T13 Ch5 Census Bureau programs)**
- **IRS Party to Proposal Review Process**
- **“Predominant Purpose Statement” (PPS)**
- **“Thin Client” Computing Infrastructure**
- **Post Project Certification**

Benefits to Bureau:

- **Criteria specified in document:**
 - Understanding and/or improving the quality of data produced through a Title 13, Chapter 5 survey, census or estimate
 - Leading to new or improved methodology to collect, measure, or tabulate a Title 13, Chapter 5 survey, census or estimate
 - Enhancing the data collected in a Title 13, Chapter 5 survey or census. For example:
 - Improving imputations for non-response
 - Developing links across time or entities for data gathered in censuses and surveys authorized by Title 13, Chapter 5.

Benefits: (continued)

- **Criteria Document (cont'd)**
 - 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
 - Identifying shortcomings of current data collection programs and/or documenting new data collection needs
 - Constructing, verifying, or improving the sampling frame for a census or survey authorized under Title 13, Chapter 5

Benefits: (continued)

- **Criteria document (cont'd)**
 - Preparing estimates of population and characteristics of population as authorized under Title 13, Chapter 5
 - Developing a methodology for estimating non-response to a census or survey authorized under Title 13, Chapter 5
 - Developing statistical weights for a survey authorized under Title 13, Chapter 5.

Current RDC Operating Model

- **Census Bureau and RDC partners must**
 - Establish physically secure offices and secure computer systems
 - Choose projects that use the data appropriately, benefit Census Bureau programs (as required by law), and present low disclosure risks;
 - Impart to researchers at the RDC the Census Bureau “culture of confidentiality;”
 - Establish policies and procedures that protect confidentiality in the RDC office;
 - Release only research output that is within the scope of approved projects and that does not reveal confidential information.

Operating model (cont'd)

- **Each RDC has a Census Bureau-approved security plan.**
 - Locked office with badges, key cards, keypads, etc.
 - Access limited to researchers with Special Sworn Status (SSS) carrying out active, approved projects at the RDC:
 - Sign written active project agreements
 - Obtain security clearance
 - Sign Census Bureau's standard sworn agreement to preserve the confidentiality of the data.
 - Receive awareness training – T13, T26 (if appropriate), IT security

Operating model (cont'd)

CES employee (the RDC administrator) stationed at each RDC.

- Instills the Census Bureau's "culture of confidentiality" into the researchers
- trains the researchers regarding the security and confidentiality restrictions.
- Carries out disclosure analysis on any research output a researcher wishes to remove from the secure facilities

Operating model (cont'd)

- **Thin client computing environment**

- Completing conversion from isolated local networks of Unix servers with PC (Windows NT) workstations.
- Data stored on secure unix servers at Census Bureau headquarters (Bowie MD). *No confidential data stored at the RDCs.*
- RDCs connected to servers via dedicated T-1 lines.
- Researchers use X-terminals (“thin clients”- no local data storage) to access the data authorized for their projects.
- Access controlled via access control lists
- Researchers are accountable for their computer use, through the use of passwords and system logs.

Operating model (cont'd)

- **Thin client (cont'd). Researchers...**
 - May not upload or download anything to thin client servers (no physical way to do it)
 - Have no access to any non-Census Bureau network (including the Internet) from within the RDC facility.
 - May not bring laptop computers or other portable mass storage devices into the RDC facility.

Operating model (cont'd)

- **Thin client (cont'd)**

- Unix windowing environment - CDD
- Software available (unix versions)
 - SAS
 - Stata
 - Other stat packages (not SPSS)
 - Office suite – OpenOffice (“clone” of MS office)
 - File translation software
- Help available
 - SAS on line documentation
 - Software guides, unix tutorials & help docs
 - “Trouble” reporting

Operating Model (cont'd)

- **Access to RDC lab**
 - 24/7
 - Printing
 - Available only when RDC administrator present and logged into local print server
 - User print queues always put headings on every page saying output is confidential
 - RDC admin can print without these headings

Project Logistics

- **Proposal review at CES – 3 times a year**
 - Work with RDC administrator to develop proposals
 - Submit via CES website
 - Components
 - Abstract
 - Proposal
 - Benefits Statement
 - Review Criteria
 - Benefit to Bureau
 - Scientific Merit
 - Disclosure Risk
 - Resources/feasibility
 - Selection meeting about 2 months after proposal deadline

Project logistics (cont'd)

- **Possible other proposal review**
 - Projects using FTI – after CES approval
 - Develop “Predominant Purpose Statement” & submit to IRS
 - PPS discusses benefits in terms of benefits criteria
 - IRS must approve
 - Survey sponsors – case by case, usually concurrent with CES review

Project logistics (cont'd)

- **After project approval**

- Obtain SSS – user id
- Develop project agreement – start & end dates
- Generate Computer account and data request
- CES establishes computer accounts & grants access to approved data
- Swearing in, badge, etc – Census Security
- RDC admin administers T13, T26, IT security awareness training
- CES turns accounts on, gives researcher codes to lab door

Project logistics (cont'd)

- **During project**

- Researchers schedule visits
- Submit clearance requests to RDC admin
- Draft of benefits statements/memo required before releasing results
- Final benefits memo required before releasing final research output
- Only RDC admin can release results – usually via email

Project logistics (cont'd)

- **After project ends**
 - Researchers submit working papers and published papers
 - Project reactivation process – to address referee comments
 - Future proposals accepted only if requirements from past or current project have been met – particularly benefits memos & papers

Data: (“Economic”) – Establishment & Business

- **Economic Censuses**
 - Every 5 years
 - Manufacturing, Retail, Wholesale, Services, FIRE, Mining, Construction
 - Variation in detail collected
- **Annual Surveys**
 - Variation in detail
 - Variation in sampling units
 - Variation in coverage

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Economic Data: (continued)

- **Ancillary Surveys**

- Survey of Manufacturing Technology
- Research and Development
- Pollution Abatement Cost Expenditure
- Manufacturing Energy Consumption Survey

Economic Data: (continued)

- **Longitudinal Business Database**
 - Business Register
 - 1975-1998
 - Coverage of All Sectors
 - Industry, Geography, Employment, Payroll

Data: (“Demographic”) – Households and Individuals

- **Historical focus on “economic” data**
- **Requests for “demographic” data**
 - Higher geographical resolution
 - Thicker samples
- **Obtained permission to provide access to demographic data in RDCs in 1997**

Demographic Data in Use:

- **1990 Decennial Long Form**
- **March CPS Earnings Supplement**
- **SIPP**
- **National Crime Victimization Survey**
- **American Housing Survey**

Demographic Data:

- **2000 Decennial Census**
 - Not in place yet
 - Expected late 2003-2004

Linked Data:

- *Severe restrictions and administrative hurdles*
 - Requires sign-off of multiple agencies
 - Time consuming process
 - Might require additional resources

Current challenges

- **Number of RDCs - From 2 to 8 since 1999**
- **Data**
 - More economic data sets
 - Demographic & Decennial
 - Administrative & linked
 - Documenting standardizing & updating all this
- **Increased oversight and resulting administrative requirements**
- **Infrastructure to manage all this**
 - Staff – no increase in CES headquarters staff
 - Computing
 - Management systems

Conclusion:

- **RDCs present enormous opportunity**
- **Not for everyone**
 - Significant constraints on the program
 - Legal, regulatory, policy, resource, etc.
- **More information:**
- **CES web site: www.ces.census.gov**