

Confidentiality and Data Access: Perspectives on Demographic Data

Pat Doyle

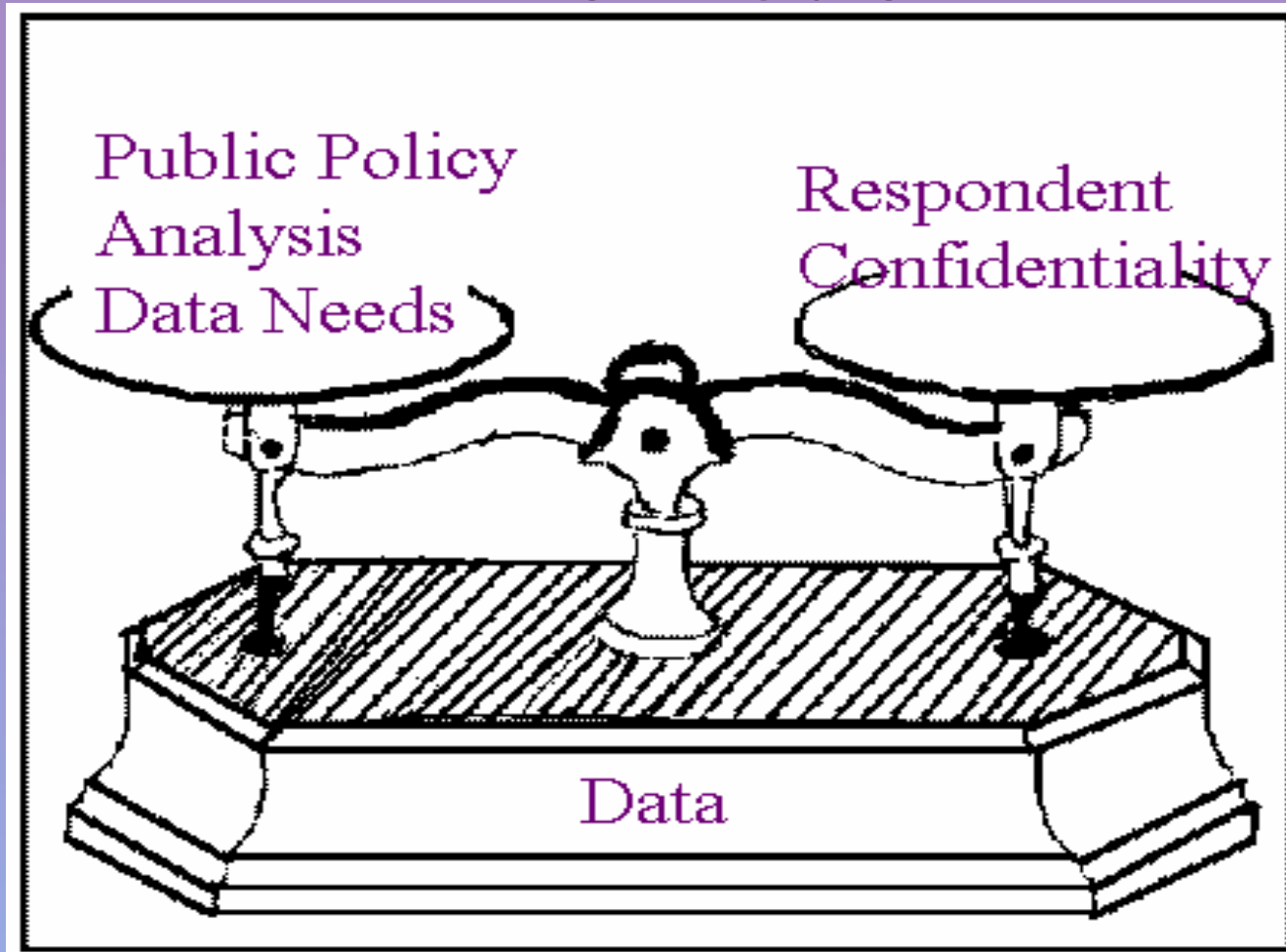
U.S. Census Bureau

Prepared for the IASSIST Annual Conference, University of Connecticut,
Storrs CT, June 2002

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

The Problem

2

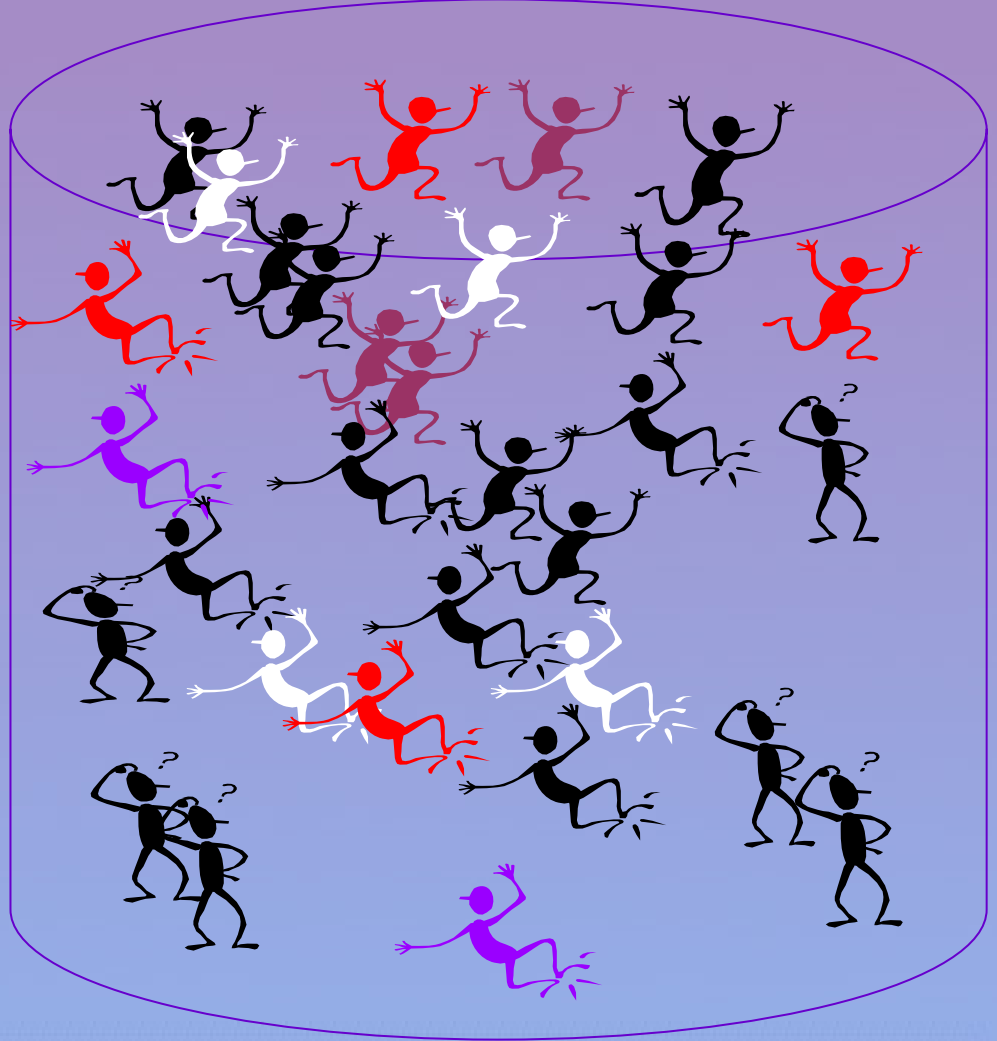


Seeking Balance

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

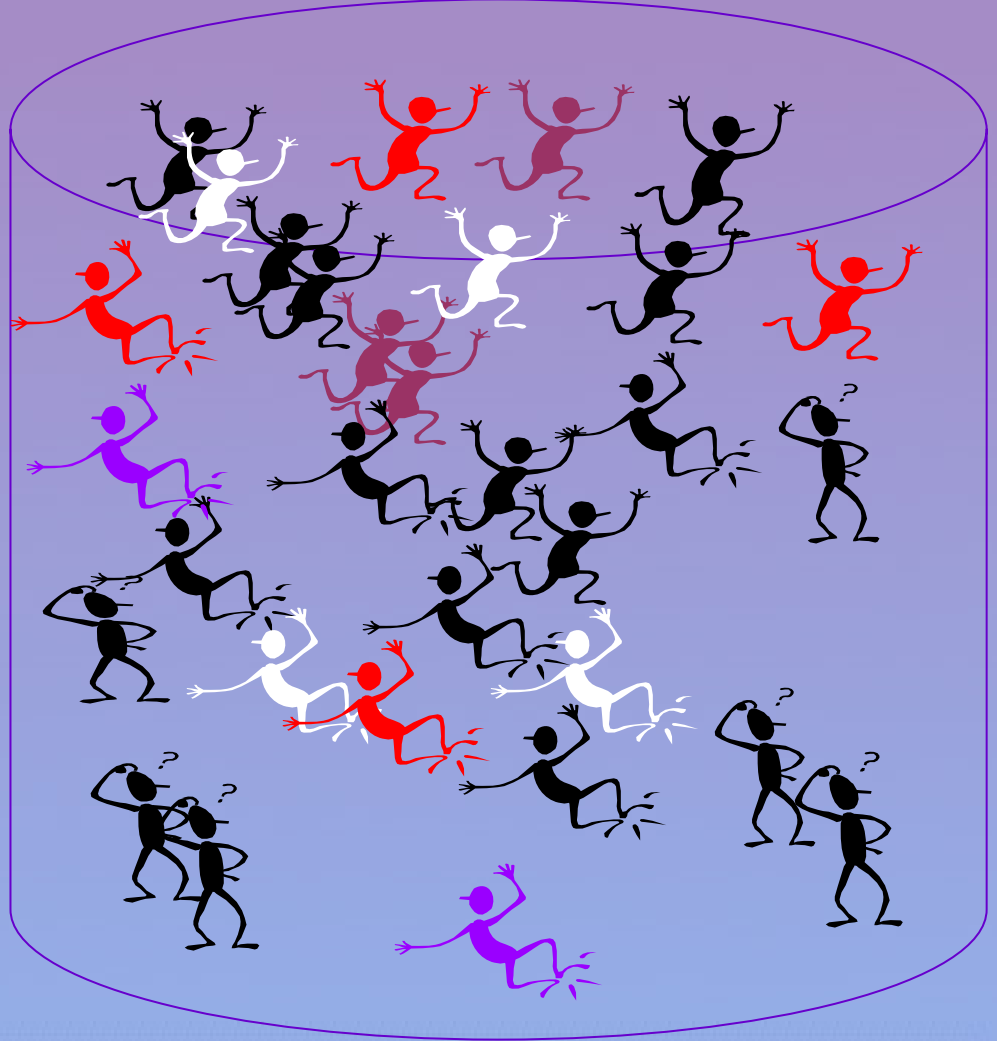
Survey

Administrative Records



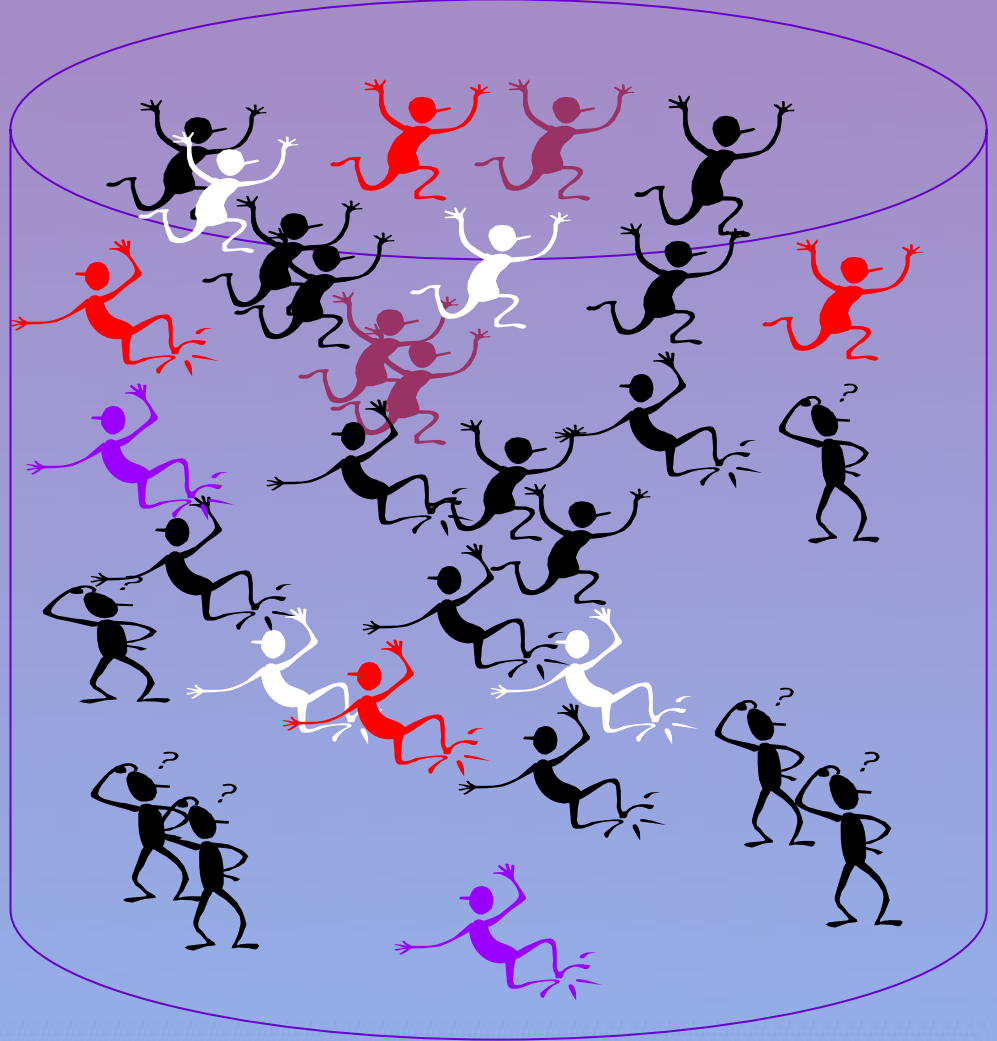
Survey

Administrative Records



Survey

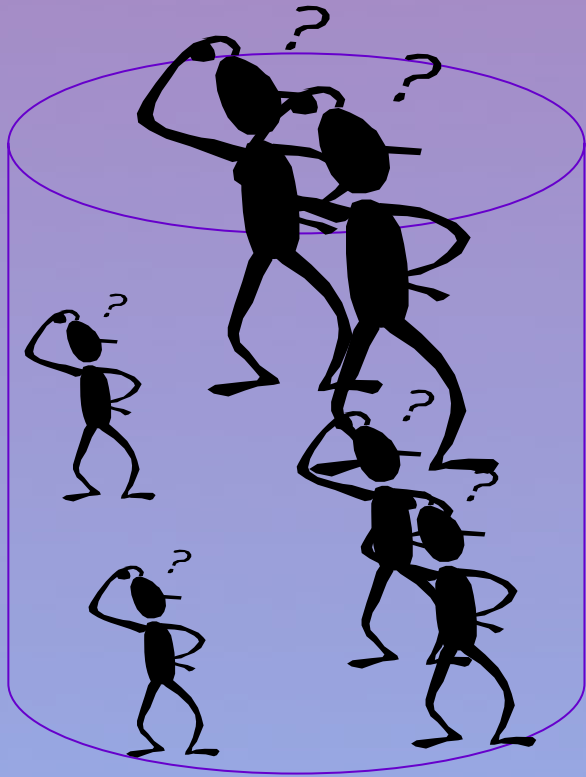
Administrative Records



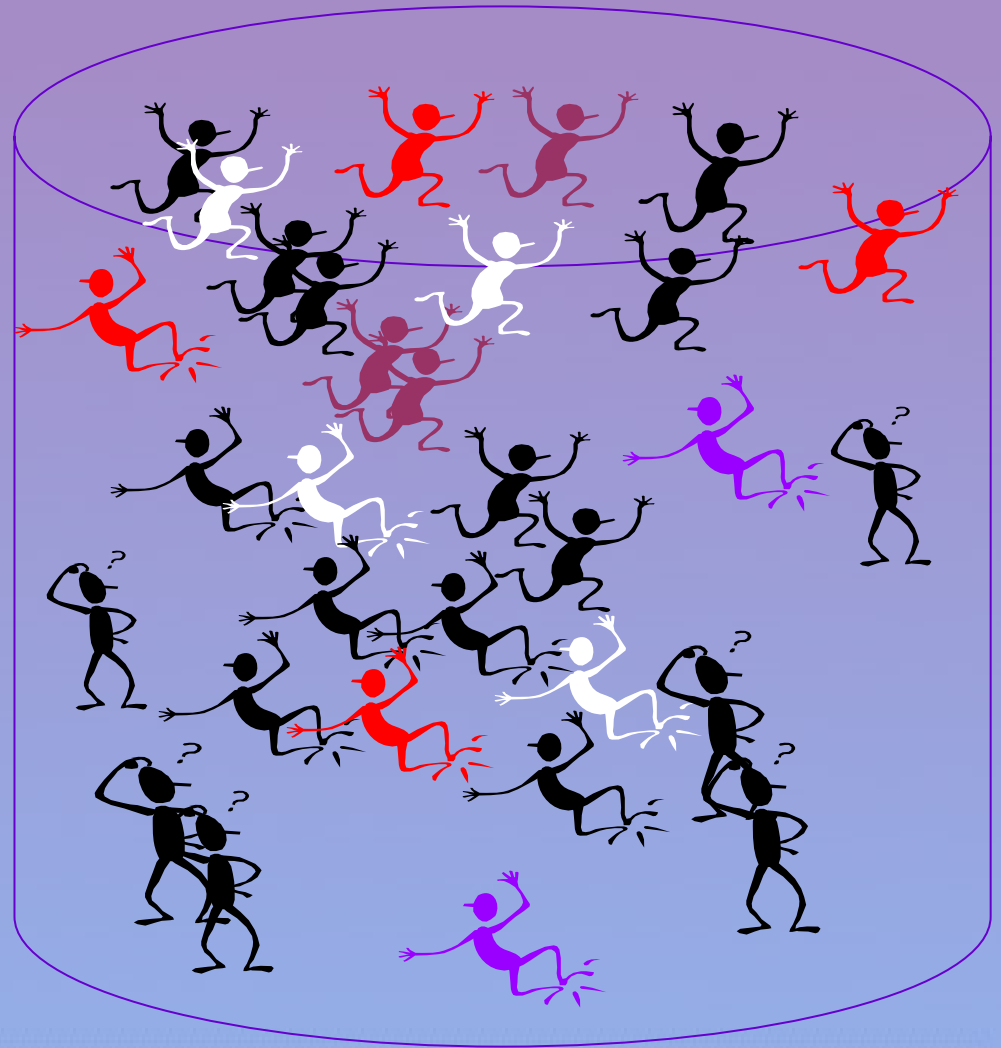
Administrative Data is Necessary but Not Sufficient

- Full re-identification of sample cases not verifiable because
 - Noise/error in administrative data
 - Noise/error in survey data
 - Insufficient number of matching keys
- Risk of re-identification slightly higher now than in the past due to reduced cost of accessing and using administrative lists
- But what about selected re-identification?

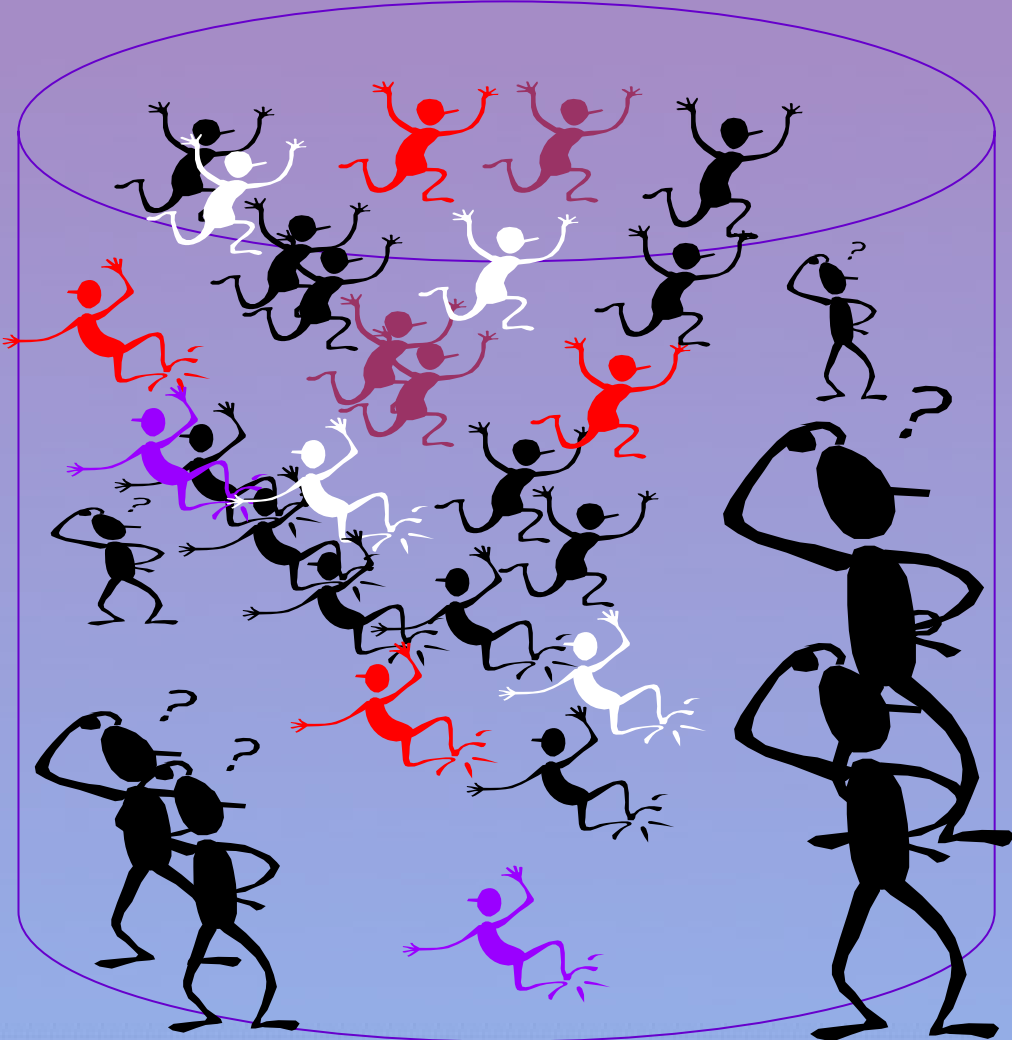
Local Announcements



Administrative Records⁷



Administrative and Local Records



Administrative and Local Records

Survey



Local Information Increases Likelihood of Selected Re-Identification

- Local information increases known attributes of people with rare characteristics
- More verifiable information for re-identification purposes
- In the absence of change in disclosure methods--increased risk of re-identification of selected situations with more and more accessible local information
- Families at higher risk than individuals

Growth in Publicly Available Administrative Data

11

(from Sweeney, 2001)

Data Source	Number of items by Year	
	1983	1996
Birth Certificates	280	1864
Hospital Discharges	0	663
Grocery Purchases	32	1272

Change in Disclosure Methods at the Census Bureau

- Eliminate sub-state geography, or
- Suppress information on rare characteristics or events, or
- Perturb information on rare characteristics or events

Highlights of the Book¹³

**Confidentiality, Disclosure, and
Data Access:**

**Theory and Practical Applications
for Statistical Agencies**

By

Pat Doyle, Julia Lane, Laura Zayatz and Jules
Theeuwes

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

Current Methods and Data

- Felsö et al.—Survey of disclosure methods in use
 - Most statistical agencies do not release demographic microdata to the public
 - Many agencies providing demographic microdata have limited access
- Sweeney—Re-identification methods
 - Points out the accessibility and richness of private data
 - Illustrates the increased risk associated with data integrated across administrative data sources

Risk and Data Quality Assessments for Public Demographic Data

- Elliot—Review of state-of-the-art in quantifying risk of disclosure
 - Absolute risk measure not feasible, relative risk may be
 - Risk is a function of sampling rate, level of detail, number of match keys, compatibility of keys
- Domingo-Ferrer and Torra—Disclosure methods for microdata, information loss and risk
 - Summarize disclosure methods and information loss measures
 - Evaluates disclosure methods based on re-identification risk and information loss

New Disclosure Methods for Demographic Microdata

- Abowd and Woodcock adopt Rubin's multiple imputation strategy for demographic surveys to mask the underlying data

Access to Non-Public Data

- Seastrom—Licensing and its use at NCES
- Dunne—Secure research sites
- Blakemore—Remote access

Perceptions

- Singer—Individuals' perceptions and attitudes toward confidentiality
- Gerber—An ethnographic study of the relationship between privacy attitudes and response