



U.S. Department of Justice  
*Office of Justice Programs*

# The Role of Privacy in the Design and Dissemination of National Statistics

Michael Planty  
Bureau of Justice Statistics  
U.S. Department of Justice  
September 29-30, 2016



BUREAU OF JUSTICE STATISTICS

---

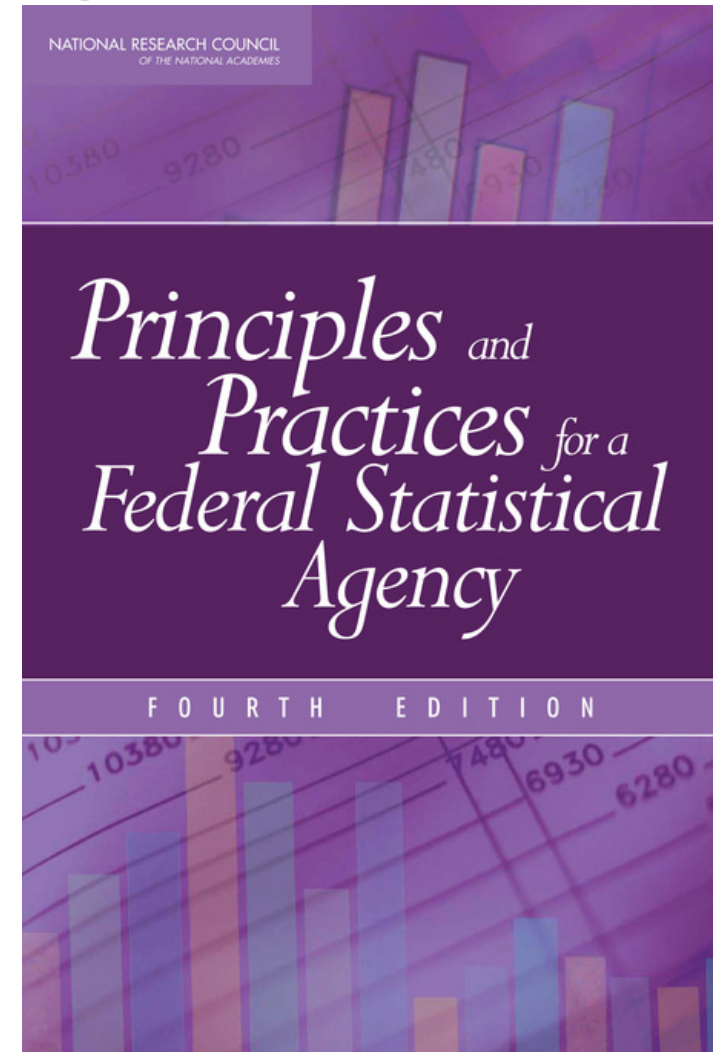
# Overview

## Privacy from a Design Approach

- Principles & Practices
- Framework: Total Survey Error (TSE)
- Statutory requirements
  
- Data Collection Lifecycle: Capturing and Protecting Private Information

# Principles and Practices for a Federal Statistical Agency

- Developing and capturing information by design
- Collection of information for statistical purposes
- Privacy and confidentiality
- Follow Paperwork Reduction Act (PRA) 1980/1995
- OMB Circular A-130



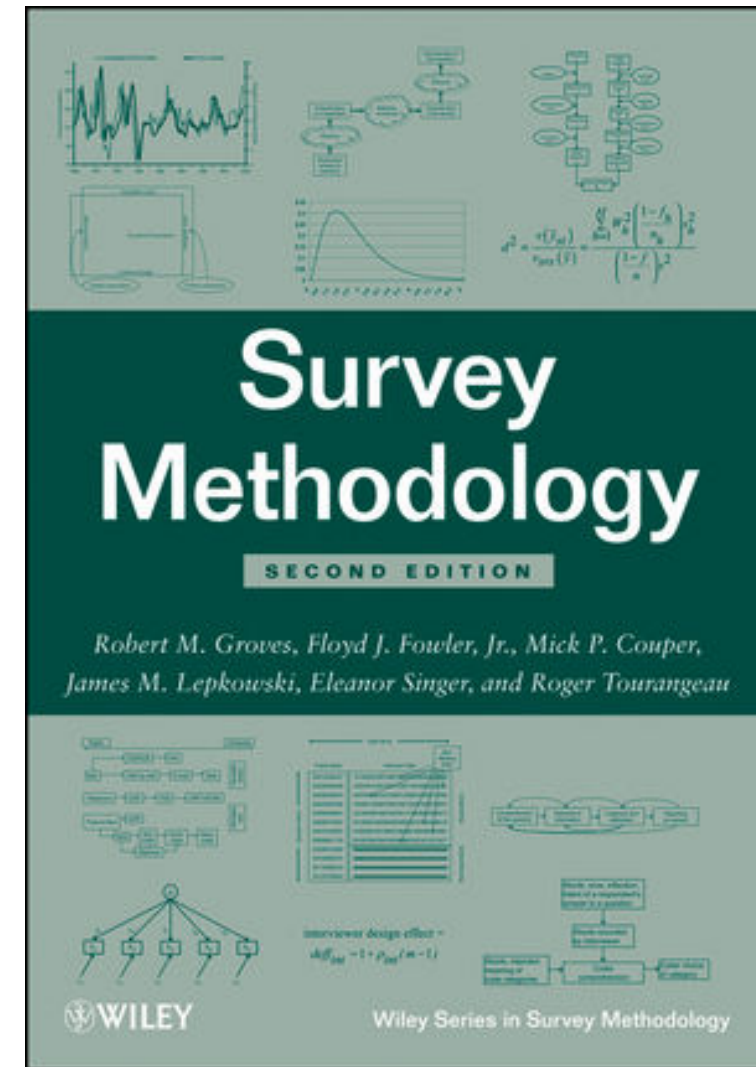
# Total Survey Error Framework

**Total Survey Error:** allocate resources to minimize error for estimates (Platek and Sarndal 2001; Biemer, 2010)

- Designing data collection schemes that balance costs (money, burden) with..
  - data quality (reliability and validity),
  - completeness,
  - credibility
  - timeliness, and
  - relevance.

**Fitness for Use:** The extent to which the data/information serves the purposes of the user.

- Groves et al. *Survey Methodology*
- Dillman et al. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*
- Biemer et al. *Measurement Errors in Surveys*
- Biemer and Lyberg *Introduction to Survey Quality*



# Privacy and Survey Design

BJS authorizing statute 42 U.S.C. 3732(b) and 3789g

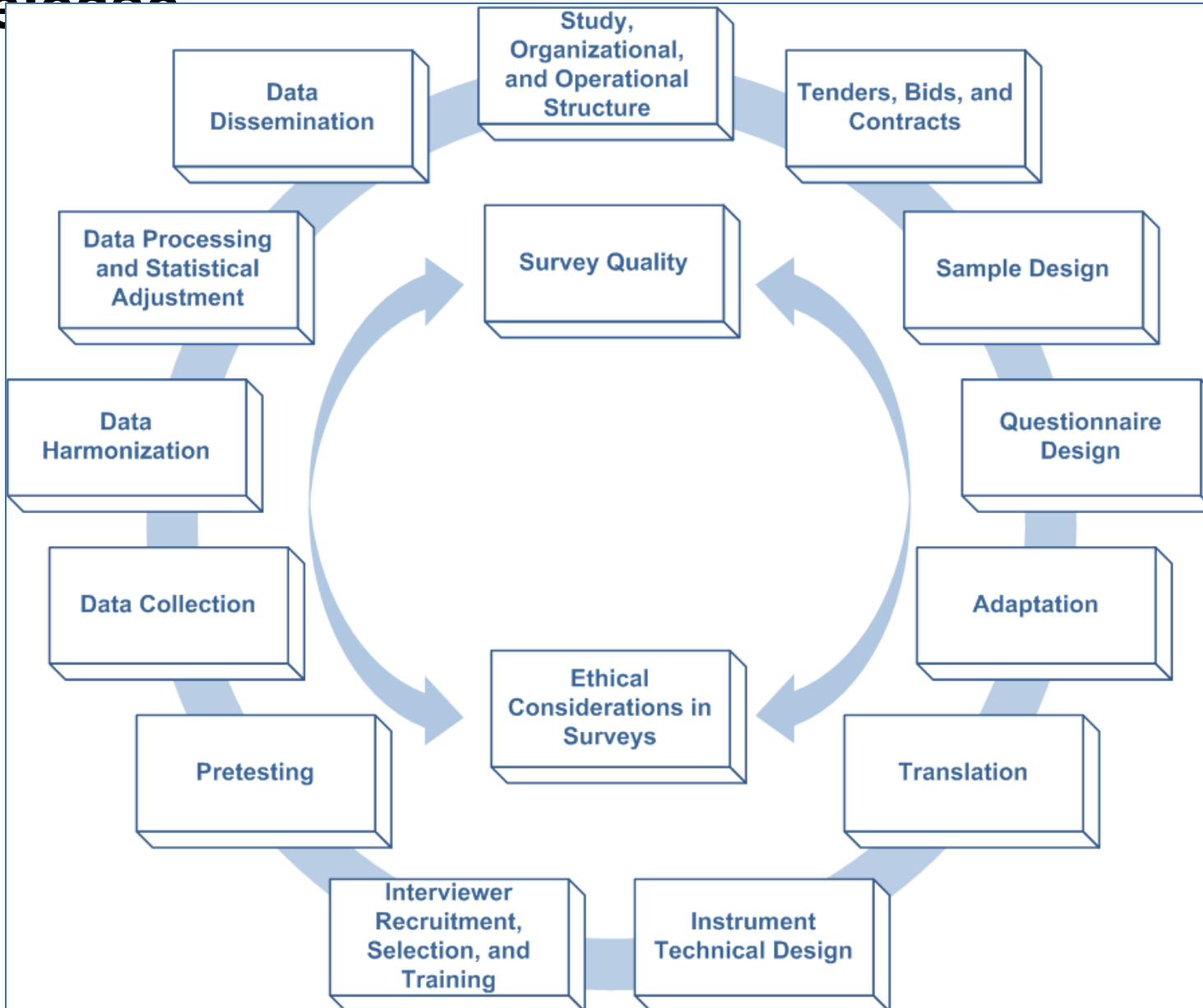
- BJS must maintain the confidentiality of all data it collects
- Director of BJS shall protect against the improper or illegal use or disclosure
- No officer or employee of the federal Government, including BJS employees and data collection agents/contractors may use or reveal any research or statistical information, including data identifiable to any specific private person, for any purpose other than the purpose for which it was obtained.
- Only use the data it collects for statistical or research purposes... precludes their use for law enforcement including administrative, regulatory, or other purpose that affects the respondents, rights, privileges, or benefits.

# Privacy and Survey Design

## Privacy and Confidentiality:

1. Soliciting and capturing “private” information
2. Once captured, ensuring private information is kept “confidential”
3. Physical security issues (CIPSEA/Circular A-130)
4. Information used only to prepare statistical summaries

# Survey Lifecycle and Privacy concerns: Data capture, handle, release



- Corporate capabilities: IRB, data security, interviewer training, laws/regulations
- Instrument design- items asked, information collected
- Data collection/interview process
- Data processing: de-identification, training/access, aggregation
- Dissemination: report and archiving
- Data linking/secondary uses

# Privacy and Survey Design

## PRA/OMB Information Collection Review (ICR) Process

- Justify information collected based on “needs and uses.”
  - Testing
  - Public comment: Federal Register Notice 60/30
- Collect only information that will be used.
- Link to other available sources to reduce burden, improve quality/value.
- Legal and ethical considerations throughout the design.
- Lower risk of disclosure risk by reducing information captured.



# **Example: National Crime Victimization Survey (NCVS)**

## Brief Overview

- Collect information on experiences with criminal victimization
- Household survey
- Interview all persons 12 or older
- Repeated interview every 6 months over 3 years (7 interviews)
- In-person or phone interview
- Violent and property crime core collection
- Special supplements on school crime, identity theft, stalking, fraud
- U.S. Census Bureau data collection agent: 1,200 interviewers

# NCVS Interview Process and Confidentiality

## Initial contact

### U.S. Census Bureau Staff

- Trained field staff on disclosure and protection and take oath
- Title 13 protections: \$250,000 fine and/or up to 5 years imprisonment

### Pledge of confidentiality to Respondent

#### “A Word About Confidentiality

We use the information you provide for statistical purposes only. Anyone who might see your answers has to take an oath and is subject to a fine and/or imprisonment for improperly disclosing any information you provide (U.S. Code, Title 13, Section 9 and 214).

It is illegal for us to provide anyone, including other government agencies, with information about you as an individual.”

- Voluntary, respondent can stop answering at any time

# Interview Privacy

## Private Interview Setting

- Keeps respondents' answers to the survey questions confidential from other family members/persons.
- Sensitive questions can lead to social desirability bias or under-reporting.
- In private settings, respondents are more likely to answer honestly about sensitive behaviors and improve the accuracy of reporting on such topics (Turner et al., 1998; Turner et al., 2002).

# Interview Privacy

## Presence of others

- Suppression effects- Persons interviewed with other family present, lower levels of intimate partner violence and rape captured

## Mode effects and Personalization of contact

- Data collection modes such as self-administered, web-based and computer assisted are related to higher reports of violence
- Campus survey personalize contact letter: increase response rates but reduce number of crimes reported

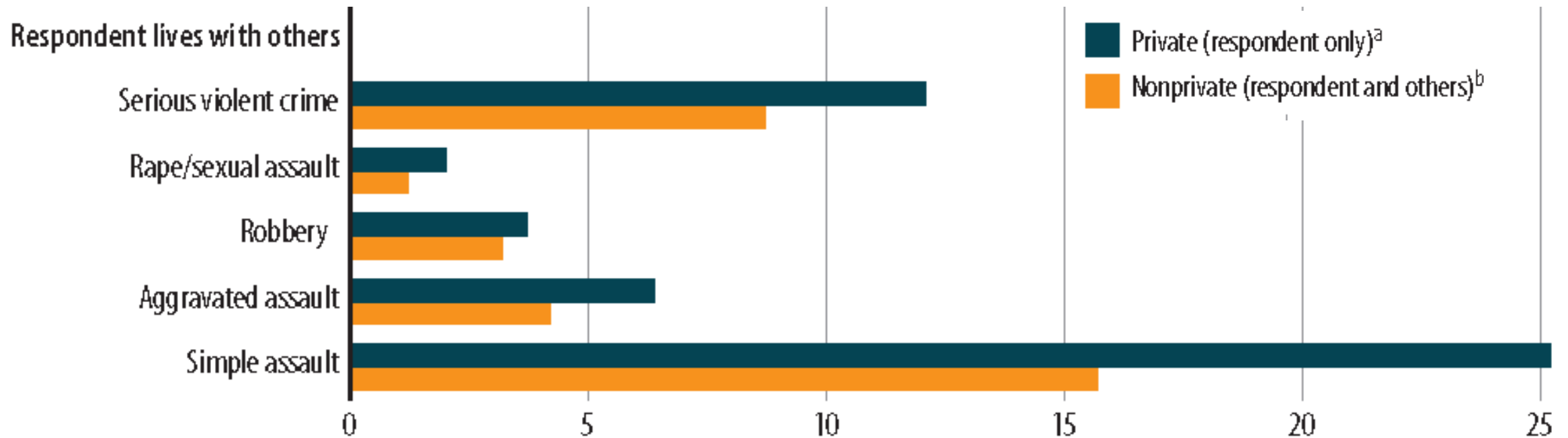
## Interview length: Victims take longer than non-victims

- Prison Rape Eliminate Act (PREA)- Interview inmates, long interview times would potentially reveal victims: solution, build in additional questions for non-victims

# Interview Privacy

## Presence of others

- Suppression effects- Persons interviewed with others present, lower levels of violence captured



# Interview Privacy

Personalization of Contact: Campus Climate Survey

Randomized experiment:

- Personalized greeting: “Dear Sarah”
- Generic Greeting: “Dear University of XXX student”

Personalize contact letter: increased response rates but reduced number of crimes reported

**Table 41. Comparison of participation and sexual assault rates, by greeting assignment and sex, 2014–2015 academic year**

	Generic <sup>a</sup>			Personalized		
	Number	Percent	SE	Number	Percent	SE
<b>Participation</b>						
Males	1,819	43.3 %	0.3 %	2,009	46.9 %*	0.3 %
Females	3,382	57.2	0.1	3,589	60.7 *	0.1
<b>Victimization</b>						
Males	65	3.6	0.1	68	3.4	0.1
Females	443	13.1	0.1	441	12.3 *	0.1

Source: Campus Climate Survey Validation Study (CCSVS), 2015

# Interview Privacy

Personalization of Contact: Campus Climate Survey

Randomized experiment:

- Personalized greeting: “Dear Sarah”
- Generic Greeting: “Dear University of XXX student”

Personalize contact letter: increased response rates but reduced number of crimes reported

**Table 41. Comparison of participation and sexual assault rates, by greeting assignment and sex, 2014–2015 academic year**

	Generic <sup>a</sup>			Personalized		
	Number	Percent	SE	Number	Percent	SE
<b>Participation</b>						
Males	1,819	43.3 %	0.3 %	2,009	46.9 %*	0.3 %
Females	3,382	57.2	0.1	3,589	60.7 *	0.1
<b>Victimization</b>						
Males	65	3.6	0.1	68	3.4	0.1
Females	443	13.1	0.1	441	12.3 *	0.1

Source: Campus Climate Survey Validation Study (CCSVS), 2015

# Private Information and Disclosure

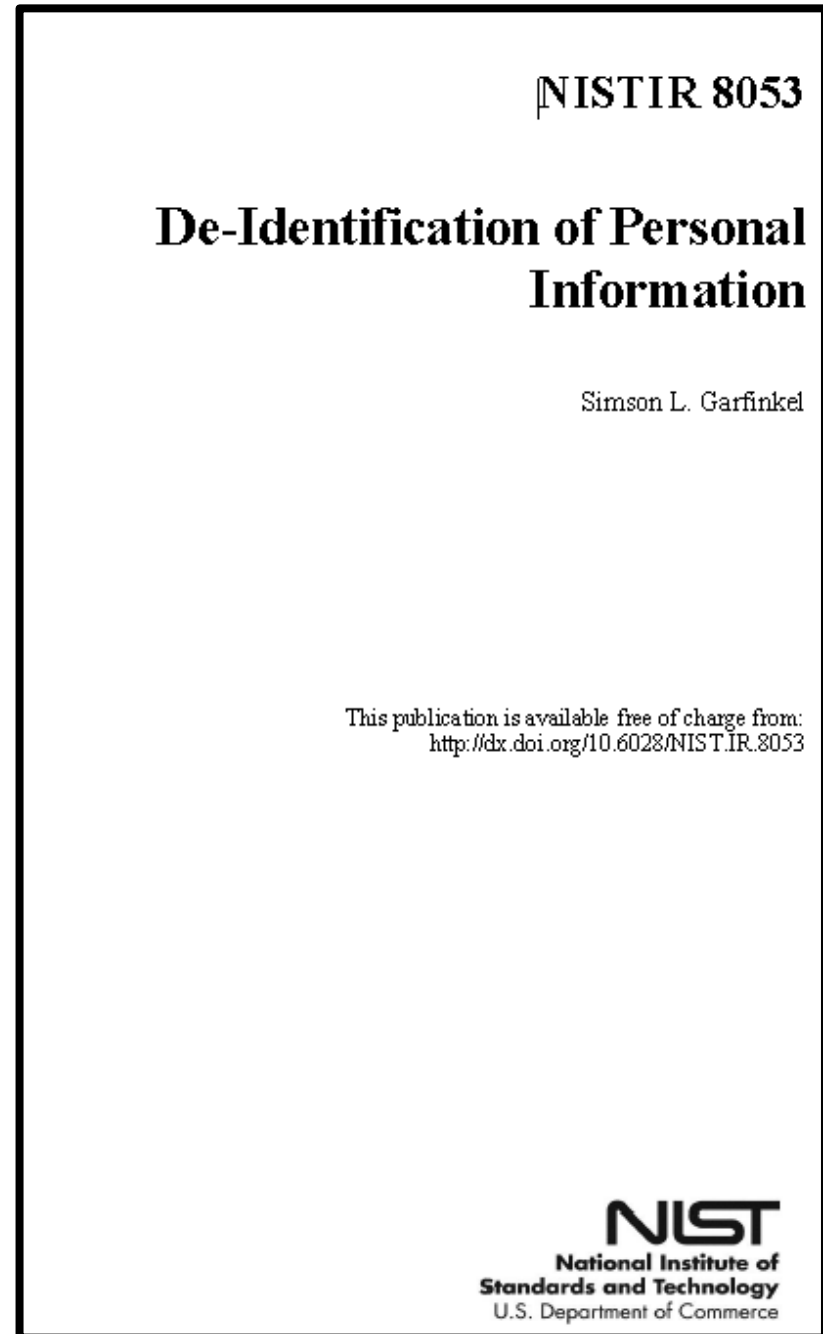
- Disclosure risk: ensure that the data maintain the greatest potential usefulness while putting in place the strongest possible protection to the confidentiality of the individual respondents.
- Disclosure types (Hundepool et al., 2012):
  - 1) identity disclosure
  - 2) attribute disclosure



# NCVS Private Information and Disclosure

Concerns with de-identification, re-identification, and data sharing models

<http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf>



# NCVS Private Information and Disclosure

## 2.3 De-Identification Data Flow Model

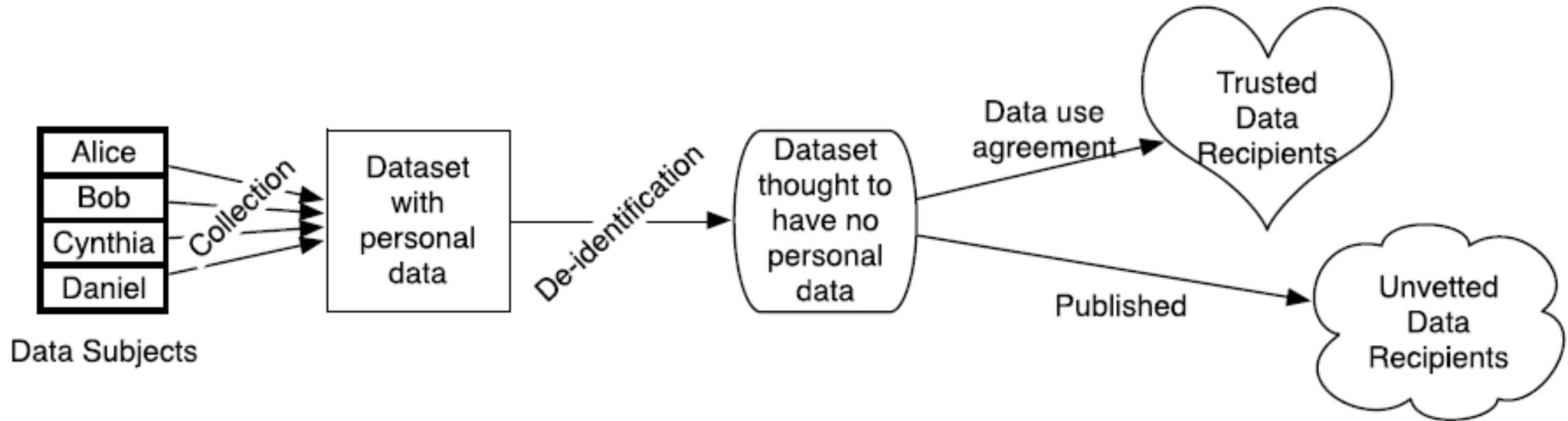


Figure 2: Data Collection, De-Identification and Use

<http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf>

# NCVS Private Information and Disclosure

NCVS concerns:

- Sensitive information, revealing one household member's information can reveal the others, implications for other BJS/ Census collections if respondent revealed.

NCVS solution:

- No direct PII on data file (names, address removed)
- Limits level of geography in public-use: national, region
- Lower levels of geography down to census tract available on restricted-use files
- Data linkage with other files for statistical purposes

# NCVS Dissemination and Disclosure

## Statistical reports/products

- Primarily tabular presentations
- Suppress estimates based on small sample sizes
- Use multiple years to produce estimates for rare events/small subgroups
- Limit level of geography (State/City estimates)

# Research/Data Linking

## Linking design data to other sources, 2 examples

### Google Street View- capture systematic observations for measuring neighborhood/block/household characteristics

- Potential to reduce burden on respondent and interviewer, improve measurement; save money
- Potential to reveal addresses that are in sample (searches are conducted in public space)
- Limited to physical disorder, may still need to ask respondent about other types

### Data linkage and Mosaic Effect

- Center for Administrative Records Research and Applications (CARRA)- opportunities to link data
- Potential to reduce burden/costs, improve measurement, improve relevance but increase risk of disclosure
- How do you integrate the errors of multiple data sets to assess fitness for use in the context of disclosure risk?

# Summary

- Privacy concerns are present and must be addressed in all parts of survey design.
- Like the Total Survey Error framework, there's a complicated balancing act: privacy, data quality and fitness for use.
- A breakdown in one aspect has potential impacts on other areas of a survey collection.
- Growth of data linkage opportunities across data collections multiplies concerns about data quality, fitness for use, and privacy.