# Formal Privacy Models and Title 13

A cooperative agreement between the Census Bureau, Georgetown University, Harvard University, and Pennsylvania State University


PI: Kobbi Nissim (Georgetown University)

Co-PIs: Urs Gasser (Harvard) , Adam Smith (Pennsylvania State University), Salil Vadhan (Harvard)
Researchers: David O'Brien (Harvard), Alexandra Wood (Harvard)


NCRN meeting, Census, Apr-24-17

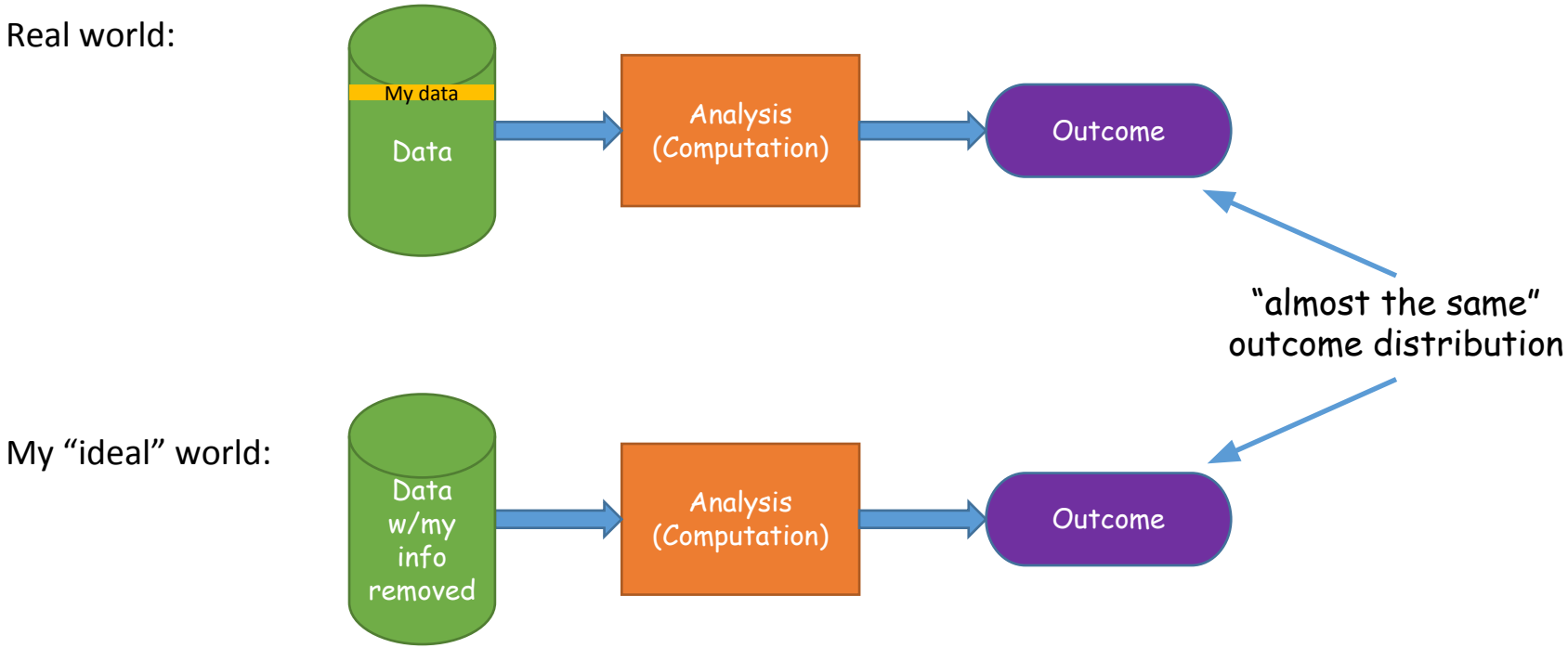
# Background: Census Bureau

- Collects and analyzes data about the nation. Its publications serve as the basis for research and decision making by policymakers, researchers, and businesses.
- Much of these data pertain to individuals, households, and establishments.
- Census has obligation to employ SDL practices that strike a balance between socially beneficial uses of data and protection of individual and establishment privacy interests.
    - Title 13 of the U.S. Code.
    - Other laws: Privacy Act of 1974, Confidential Information Protection and Statistical Efficiency Act, E-Government Act of 2002.
    - These laws protect personal information in government records; allow the release of non-identifiable information to support scientific research and public policy decisions.

# Background: The Evolving Data Privacy Landscape

- Since the 1990s, computer scientists and legal scholars have noted that traditional SDL approaches **often fail to address privacy risks in data sharing**.
    - Traditional SDL techniques often rely on concepts that reflect an information regime very different from the environment today.
    - Traditional approaches are often heuristic and address only a limited scope of potential privacy threats.
    - It is difficult to understand the combined effect of their use on the privacy of surveyed individuals and establishments.
    - There is a patchwork of existing legal, ethical, and procedural requirements for data releases in different contexts, organizations, and jurisdictions.
- A new approach, rooted in theoretical CS: **Formal privacy models.**
    - In particular, **differential privacy** [Dwork, McSherry, Nissim, Smith].

# Differential Privacy [Dwork, McSherry, N, Smith 06]



# Background: Formal Privacy Models

- Allow for a *rigorous* and *quantifiable* analysis of privacy and risk
- Properties:
  - Post processing, composition, group privacy
  - Meaningful privacy in presence of auxiliary information and regardless of attacks
  - Utility theoretic interpretation of privacy risk
  - Tool for statistical validity
- Current status of differential privacy:
  - Rich theoretical foundations: algorithms, lower-bounds, relationships with statistics, game theory, machine learning, ..., policy and law
  - First real-world implementation: Census' OnTheMap, Google's RAPPOR, Apple's iOS 10, Harvard's Privacy Tools, ...
  - Research needed to examine how it matches current regulations and policies


* http://privacytools.seas.harvard.edu

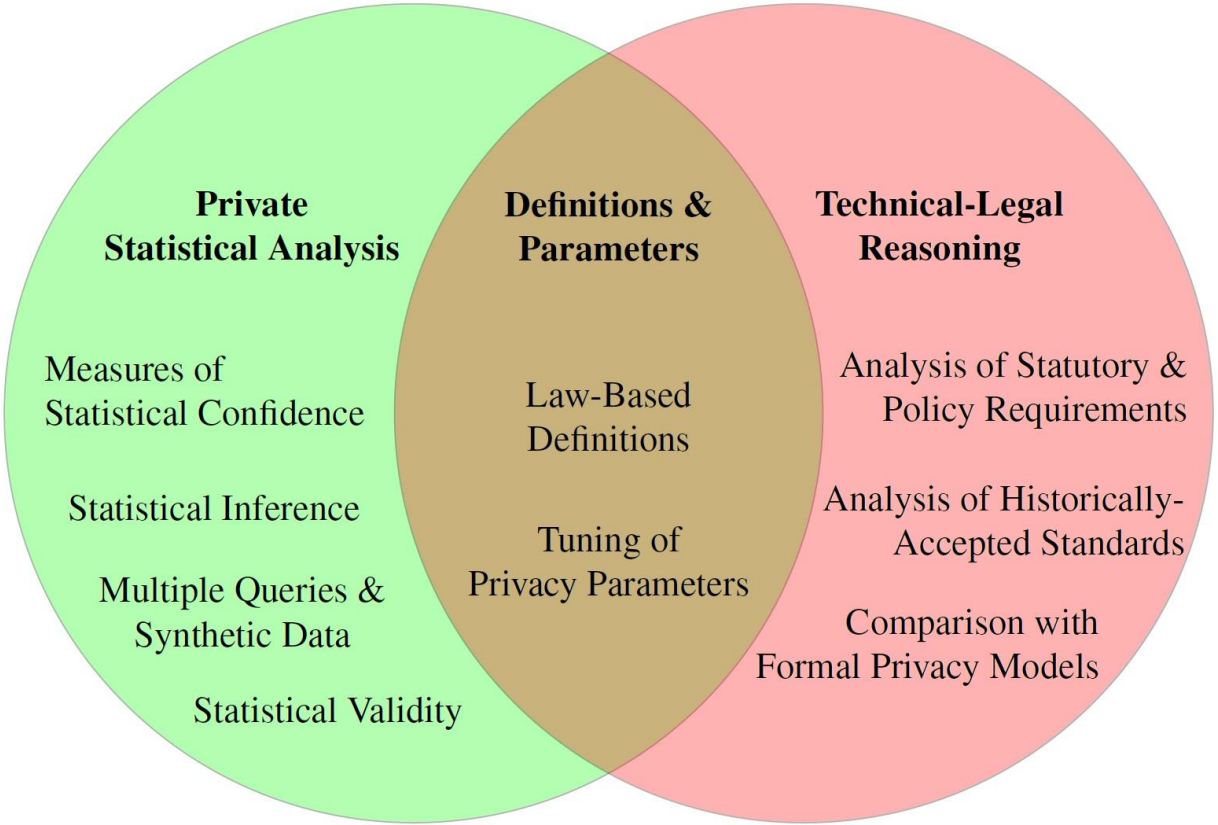
# Furthering Census' Use of Formal Privacy Models

- Two important gaps posing implementation challenges:
  - Wide conceptual and practical gaps between the approaches found in formal privacy models and the heuristic approaches in current use and contemplated by existing regulatory and policy frameworks.
  - Gap between theoretical developments showing that formal privacy models like differential privacy permit, in principle, a wide collection of statistical and machine learning analyses and the actual use of such analysis and publication techniques by the Census Bureau.

# Furthering Census' Use of Formal Privacy Models

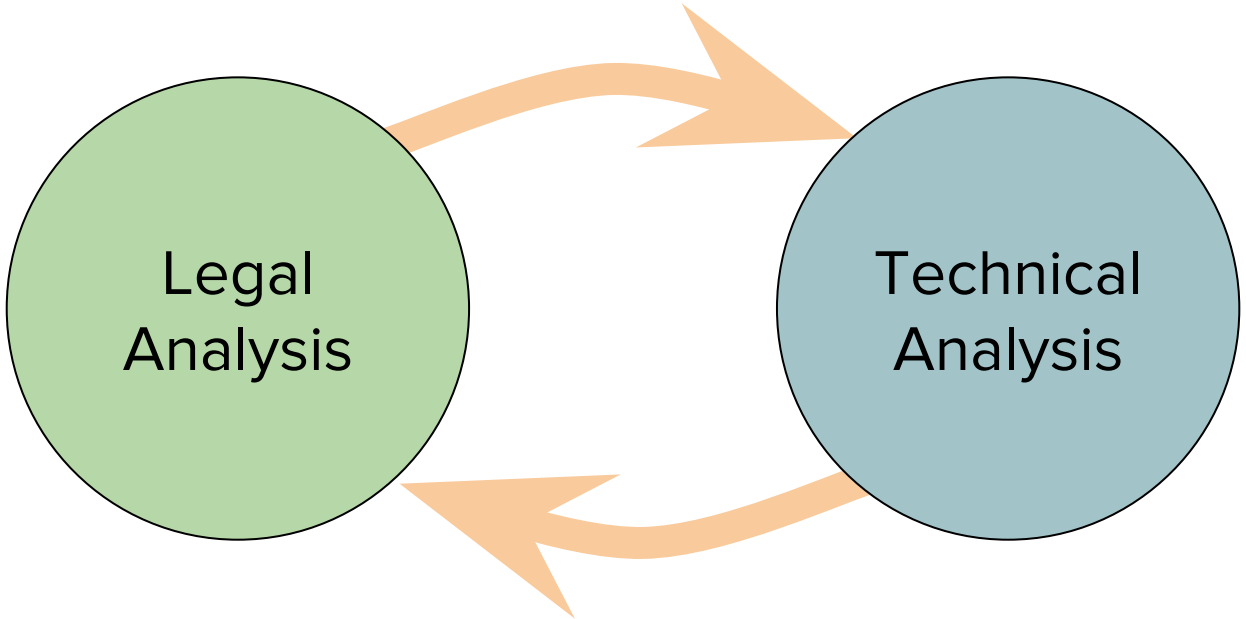
- Efficient statistics. Research is needed to map the limits of what can be achieved in terms of privacy - utility - efficiency tradeoffs for statistical inference tasks using differential privacy.
- Tuning privacy parameters. A barrier to using formal privacy models is the need to set parameters, like differential privacy's privacy loss parameter ($\epsilon$), based on normative judgments about privacy.
- Policy context. Bringing formal privacy models to practice requires tailoring them to the regulations and policies that apply in a given real-world setting.

# Project Structure

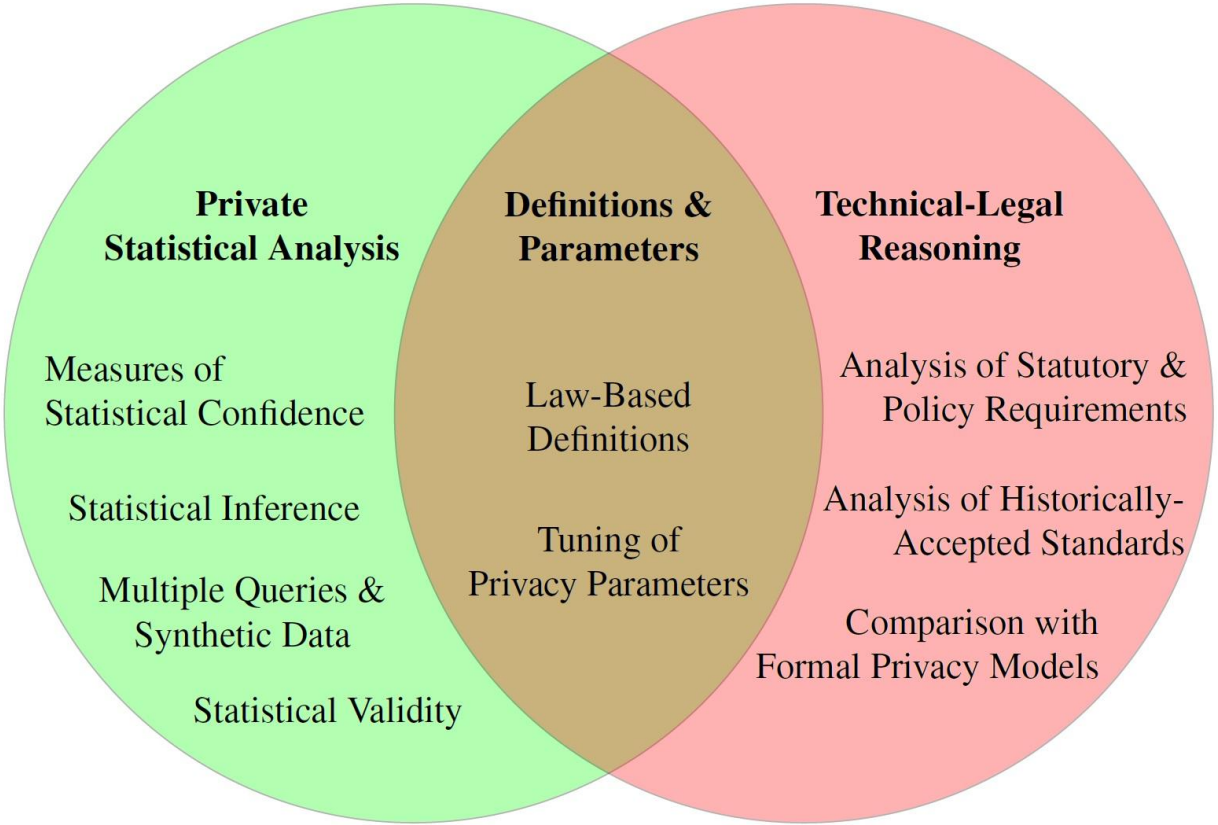


# Project Structure

Essential components of the project lie in the intersections and cross-disciplinary feeds between these research directions.



# Project Structure



# Expected Outcomes

- Bridging definitions. Developing an understanding of the relationship between mathematical and legal notions of privacy in the Census context; proposals to bridge the gap between these notions in ways that support Census activities; a methodology for setting initial privacy parameters in formal privacy models.
- Theoretical results. Reporting on the performance of statistical analyses satisfying formal privacy requirements; new measures of statistical confidence incorporating the noise added for privacy.
- Evaluation. Assessing the practical performance and usability of a variety of algorithms for analyzing and sharing privacy-sensitive data.
- Knowledge transfer and education. Sharing knowledge on current interdisciplinary research in privacy with Census personnel; Publications in academic venues and on the project website; Training of postdocs and graduate students.

# Snapshot of Current Work

- Project start date: January 1, 2017
- Current research focuses:
  - Legal analysis of Title 13's privacy requirements
  - Differentially private statistics, focus on finite sample size
  - Updating our forthcoming non-technical primer on differential privacy
    - The primer uses intuitive illustrations and limited math to help social scientists, statisticians, regulators, policymakers conceptualize the guarantees provided by differential privacy
    - Synergistic with the current project, the paper will include a chapter that discusses the relationship between differential privacy and the privacy standards from several US regulations, including Title 13, as well as concepts from the statistical disclosure limitation literature.

# Analysis of Title 13

- Background: Our previous analysis bridging between FERPA's privacy requirements and differential privacy
  - Goal: A rigorous proof that differential privacy satisfies a large class of interpretations of FERPA's privacy standard
  - Approach: A close analysis of FERPA yielded a formal model of the implicit adversary contemplated by the regulators. Mathematical tools were used to address ambiguous language used in the regulations and implementation guidance.
  - An initial version of this paper is available on the Privacy Tools website: http://privacytools.seas.harvard.edu.

# Analysis of Title 13

- Research to develop a formal legal-technical argument for Title 13:
  - Reviewed Title 13, policy documents, past decisions, and other documents provided by Census Policy Coordination Office
  - Hosted Ashwin Machanavajjhala to learn about the process used by the OnTheMap team to formally model Title 13
  - Discussed Title 13 as part of weekly interdisciplinary working group meetings
- Title 13 is significantly different from FERPA for purposes of our analysis:
  - Statute with criminal vs. civil penalties
  - Lean description of privacy goals found in Sections 8 & 9
  - Substantial deference to analysis and decisions of the Disclosure Review Board
- Our plans:
  - Explore ways to supplement analysis with additional Census documentation (e.g., DRB decisions and underlying rationale)
  - Work on complementary approaches

# Conclusion

- A new collaboration between academia and the Census Bureau to further the Bureau’s use of formal privacy models
- Main components of this research:
  - Theoretical algorithmic developments in private statistical analysis and statistical validity
  - Formal analysis of the privacy protection required by the applicable legal regime
  - Bringing the approach to privacy under the law and current practice in line with a formal, technical understanding of privacy
- Project is currently in its initial stages - stay tuned!