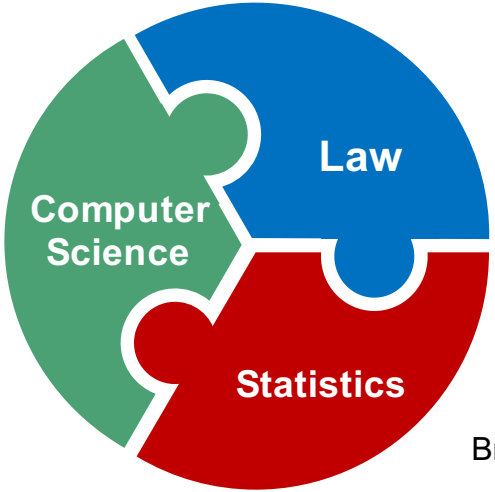
Main Goal: Furthering Census' Use of Formal Privacy Models

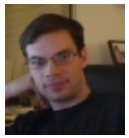
Two important gaps:

- Between formal privacy models and heuristic approaches in current use and existing regulatory and policy frameworks.
- Between theoretical developments and actual use of such analysis and publication techniques.




U. Gasser (Harvard)

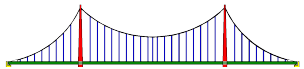

K. Nissim (Georgetown)


D. O'Brien (Harvard)


A. Smith (Boston U)


S. Vadhan (Harvard)


A. Wood (Harvard)


Bridging Privacy Definitions Working Group

How this project was conceived

Privacy Tools work: new collaborations between law and computer science; the *Bridging Privacy Definitions Work Group*.

- *Bridging the Gap between Computer Science and Legal Approaches to Privacy*. A paper bridging differential privacy and the privacy standard in the Family Educational Rights and Privacy Act of 1974 (FERPA).
- *Differential Privacy: A Primer for a Non-technical Audience*. A paper explaining differential privacy without relying on heavy math machinery (yet rigorously).

☹ An unsuccessful proposal: *Legal-Technological Tools for a Robust Regulatory Framework for Privacy.*

From the panel review: "The panelists recommend that the PIs more carefully examine the literature on limitations of differential privacy, e.g. [. . .]

Panelists, *please* see these blogposts:

- Anand Sarwate: **An exercise in careful misreading**, Nov 2014.
- Frank McSherry: **Differential privacy for dummies**, Feb 2016.

How this project was conceived

Privacy Tools work: new collaborations between law and computer science; the *Bridging Privacy Definitions Work Group*.

- *Bridging the Gap between Computer Science and Legal Approaches to Privacy*. A paper bridging differential privacy and the privacy standard in the Family Educational Rights and Privacy Act of 1974 (FERPA).
- *Differential Privacy: A Primer for a Non-technical Audience*. A paper explaining differential privacy without relying on heavy math machinery (yet rigorously).

A meeting in MIT: *2nd MIT-Census workshop*. (Dec 2015, organized by M. Altman, C. Capps, and R. Prevost)

- Presentation with Alex Wood on bridging work.

A BOC solicitation: *Administrative Data and Data Linkages Research Program.*

- Submitted Aug 2016.
- Start date: Jan 2017.

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.
- The Census Bureau operates under a dual statutory mandate to publish data that are suitable for their intended uses and to protect the confidentiality of individual and business respondents.
    - 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 late 90s, computer scientists and legal scholars have noted that traditional statistical disclosure limitation (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 computer science: **Formal privacy models.**
  - In particular, **differential privacy (DP)** [Dwork, McSherry, Nissim, Smith 2006, 2017].

Main Goal: Furthering Census' Use of Formal Privacy Models

What makes the project interesting:

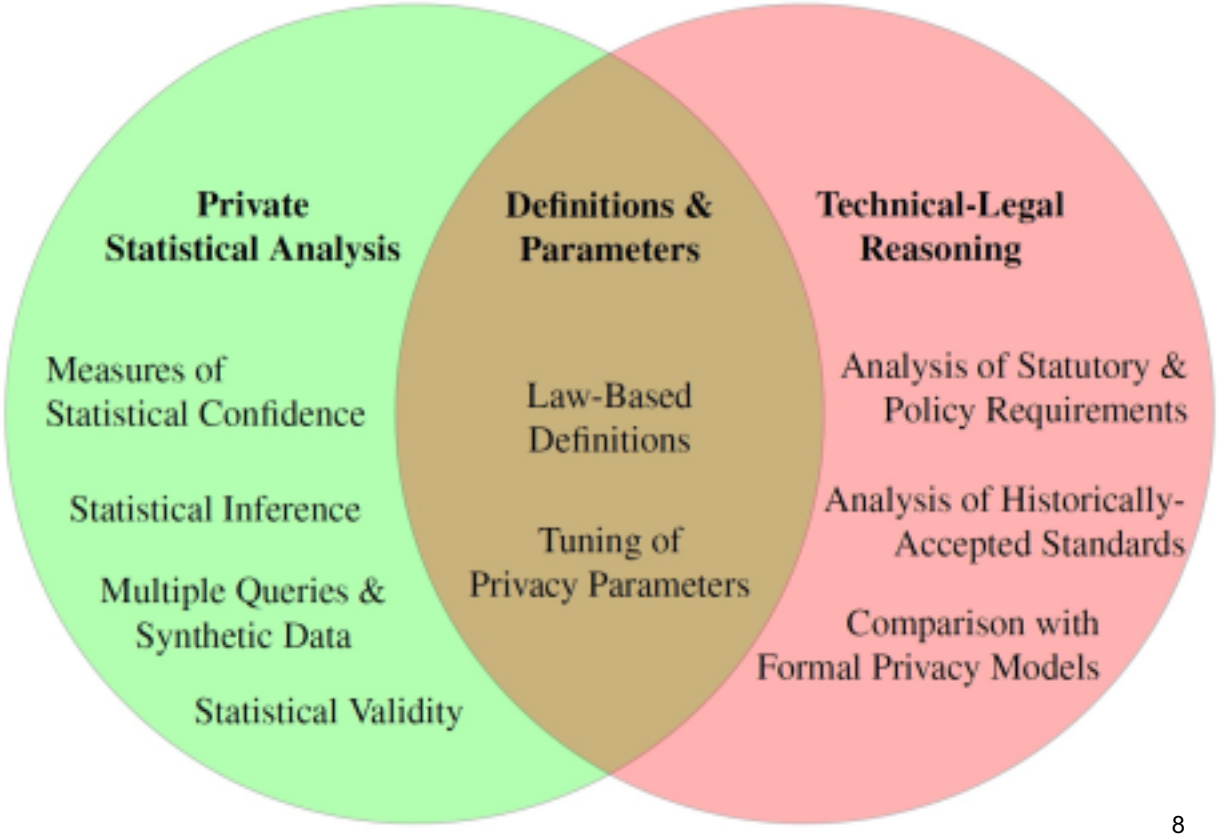
- Technically:
  - Census collects massive data, merges with data from other government entities, releases much data.
  - A real concern: how successful are their current disclosure limitation techniques, and are they sufficient for the future.
- Legally:
  - Census is bound by one of the strictest privacy law: Title 13 protects a broad range of information about respondents, probably broader than any other privacy standard appearing in the US law.
  - Section 9 of Title 13 could be interpreted as not to allow any risk of identification.
  - Title 13 a criminal statute. Other laws, e.g., HIPAA, state a fine, not jail.

Main Goal: Furthering Census' Use of Formal Privacy Models

What makes the project interesting:

- Collaboration with the US Census:
  - Census has extensive technical and legal expertise.
  - Will deploy differential privacy in 2020 census, and our work supports this initiative.
  - Making sure that use of DP satisfies legal requirements and explaining DP to decision makers and public is key.

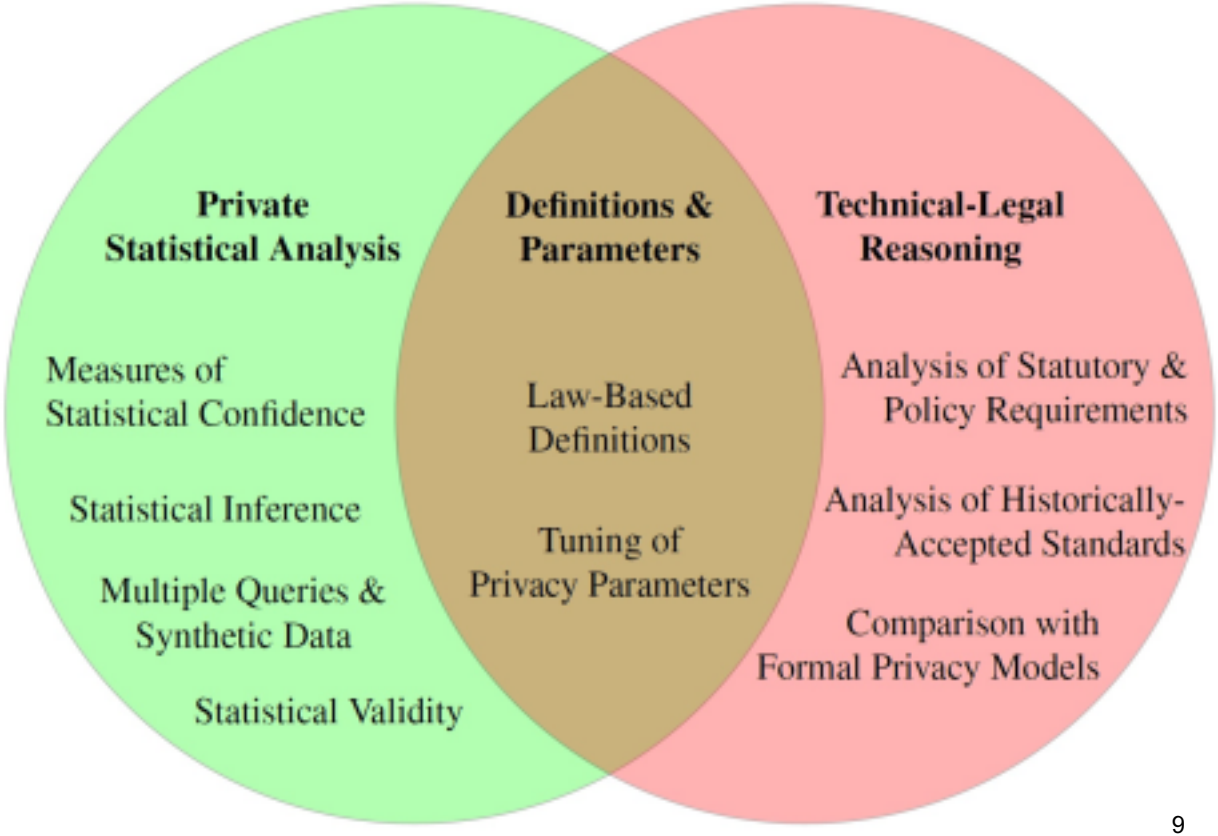
Project components



Project components

Private Statistical Analysis.

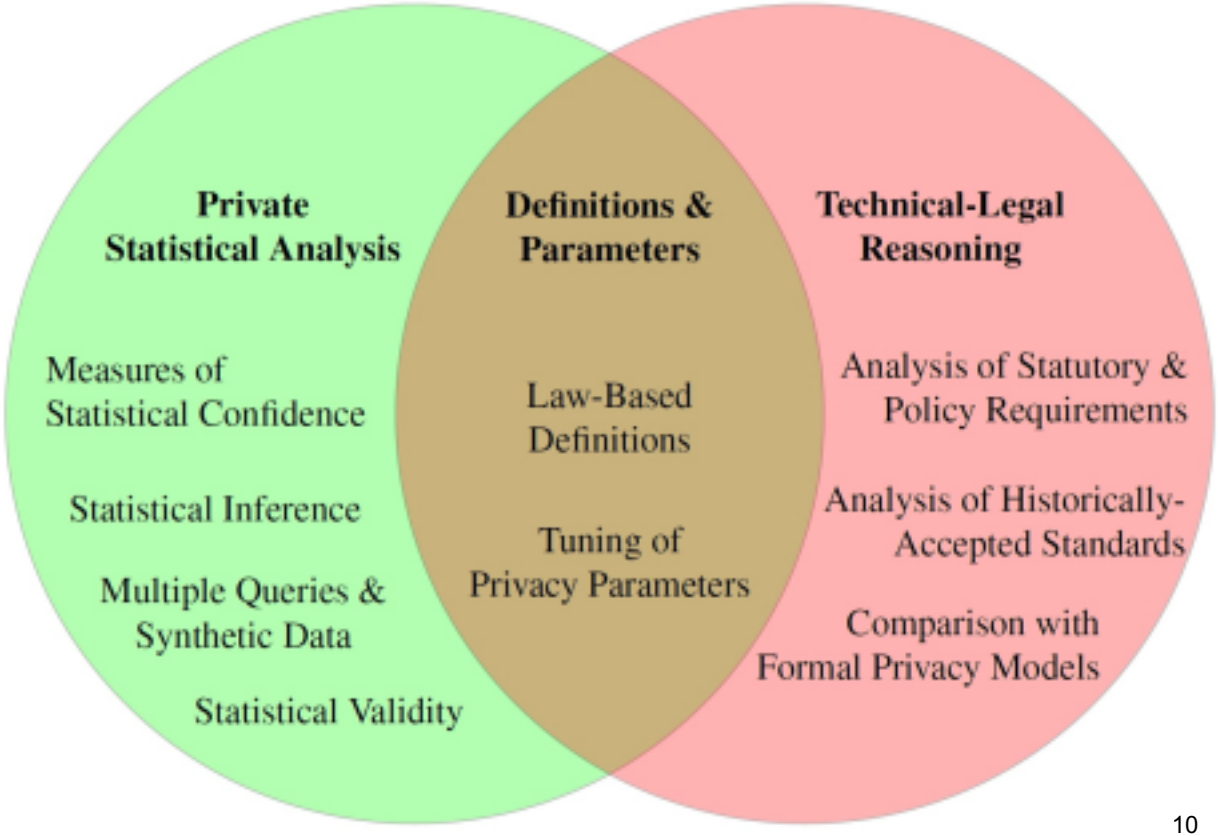
- Research is needed to map the limits of what can be achieved in terms of privacy - utility - efficiency tradeoffs for statistical inference tasks using differential privacy.



Project components

Legal analysis.

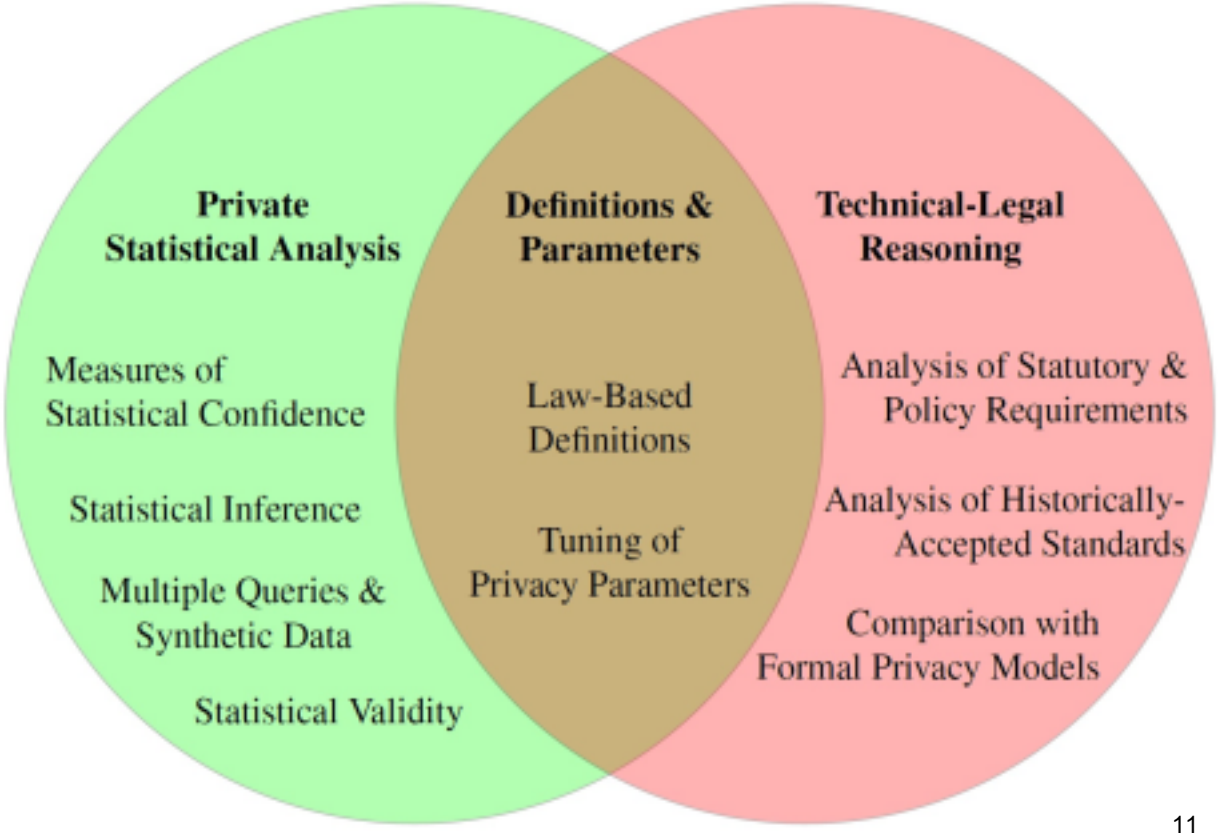
- Bringing formal privacy models to practice requires tailoring them to the regulations and policies that apply in a given real-world setting.



Project components

Law-Based Definitions.

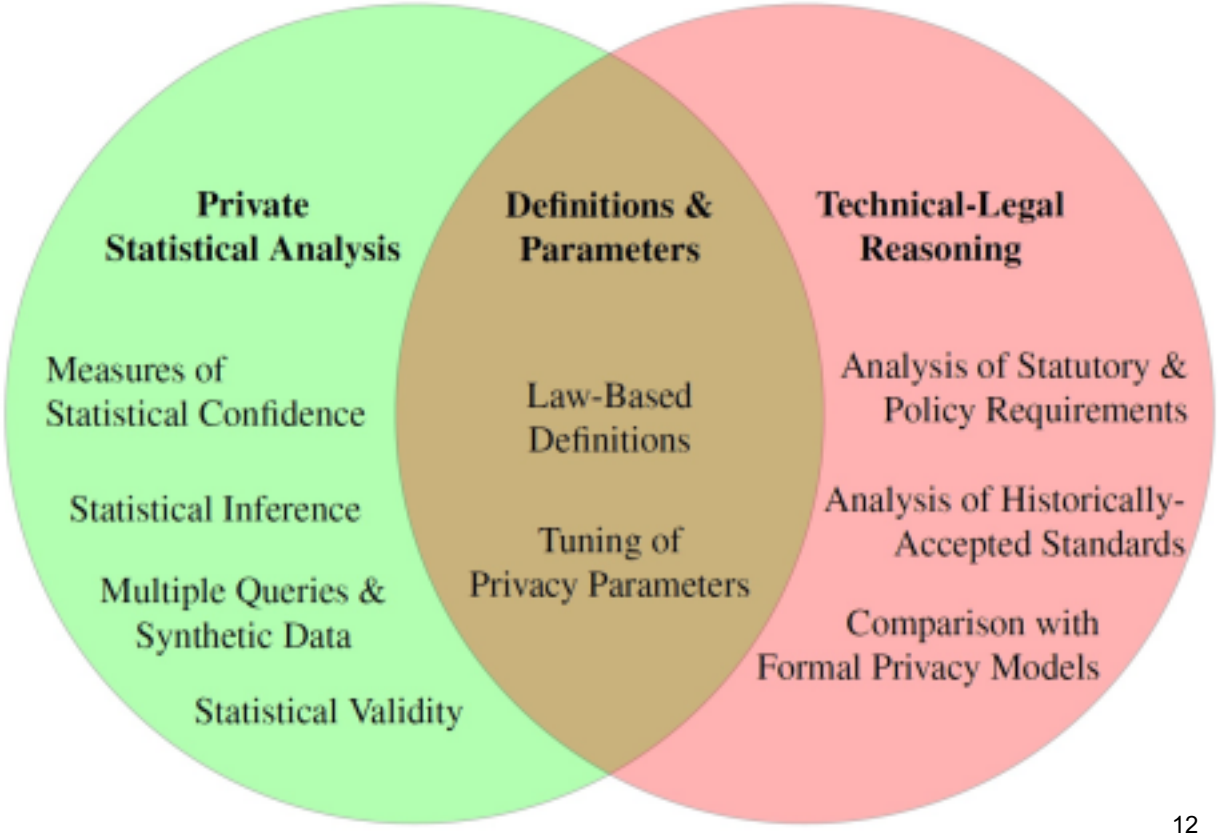
- Where differential privacy is insufficient for satisfying the requirements of the law, the legal analysis can direct search for variants of differential privacy or even other concepts.



Project components

Tuning of privacy parameters.

- A barrier to using formal privacy models is the need to set parameters, like differential privacy's privacy loss parameter (ϵ), based on normative judgments about privacy.



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. 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.

Some current research focuses

- Differentially private statistics, focus on non-asymptotic inference.
  - Near-optimal confidence intervals for mean of a normal population.
  - Bootstrap mechanism for costless standard errors and confidence intervals.
  - Possibilities and limitations of private hypothesis tests.
- Updating our non-technical primer on differential privacy.
  - Using intuitive illustrations and limited math to help social scientists, statisticians, regulators, and policymakers conceptualize the guarantees provided by DP.
- Legal analysis of Title 13's privacy requirements.

Analysis of Title 13

- Our starting point: Our previous analysis bridging between privacy requirements of the the *Family Educational Rights and Privacy Act* (FERPA) and *differential privacy*.
  - A rigorous argument that use of differential privacy satisfies a large class of potential 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.

Can the same approach work for bridging between Title 13 and Differential privacy?

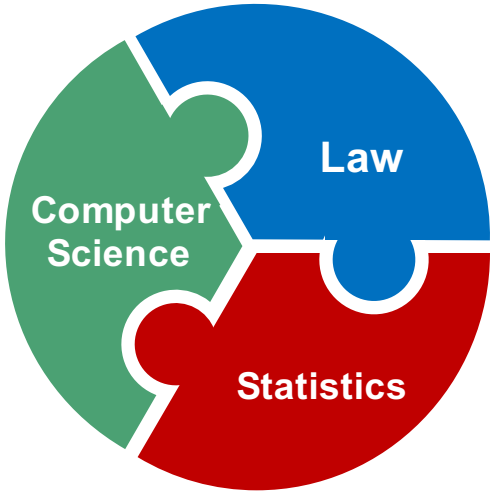
- However: Significant differences between Title 13 and FERPA, analysis needs to be modified/extended significantly:
  - Statute with criminal vs. civil penalties.
  - Lean description of privacy goals found in Sections 8 & 9.
  - Substantial deference to analysis and decisions of the Disclosure Review Board.

Analysis of Title 13

- Research to develop a formal legal-technical argument for Title 13:
  - Review of Title 13, policy documents, past decisions, and other documents provided by Census Policy Coordination Office.
  - Discussion is part of weekly interdisciplinary working group meetings.
- Some interesting questions emerge:
  - Bridging between differential privacy and concepts from Census law and policy, e.g.:
    - Personally identifiable information (PII).
    - De-identification.
    - Linkage.
    - Inference.
    - Identification risk.
  - Notions of 'specificity' in data.
  - Relationships with Nissenbaum's framework of contextual integrity.
  - Differential privacy and equal protection of the law.

Summary

- A collaboration between academia and the US Census Bureau to further the Bureau's use of formal privacy models.



- Main components of this project:
  - 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.