# A Modern Approach to Privacy Law and Policy

Alexandra Wood
Berkman Klein Center for Internet & Society at Harvard University

*Privacy Tools for Data Sharing: Lessons Learned and Directions Forward* Workshop

December 11, 2017

# Motivation: Data Sharing Challenges

- **Variations in Legal Requirements**
  Approaches may require time-consuming, case-by-case review and negotiation processes and result in markedly different treatment.

- **Inadequacy of Current Practice**
  Approaches in common use (e.g., de-identification, notice and consent) are increasingly being shown to fail to provide reasonable protection.

- **Reduced Utility**
  Data are heavily redacted or withheld altogether due to privacy concerns, often overly restricting access and limiting future analyses on the data.

# An Interdisciplinary Collaboration

## Computer Science

Kobbi Nissim, Aaron Bembenek, Mark Bun, Marco Gaboardi, Thomas Steinke, Salil Vadhan

CRCS Center for Research on Computation and Society
at Harvard John A. Paulson School of Engineering and Applied Sciences

## Law & Policy

Urs Gasser, David O'Brien, Alexandra Wood

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

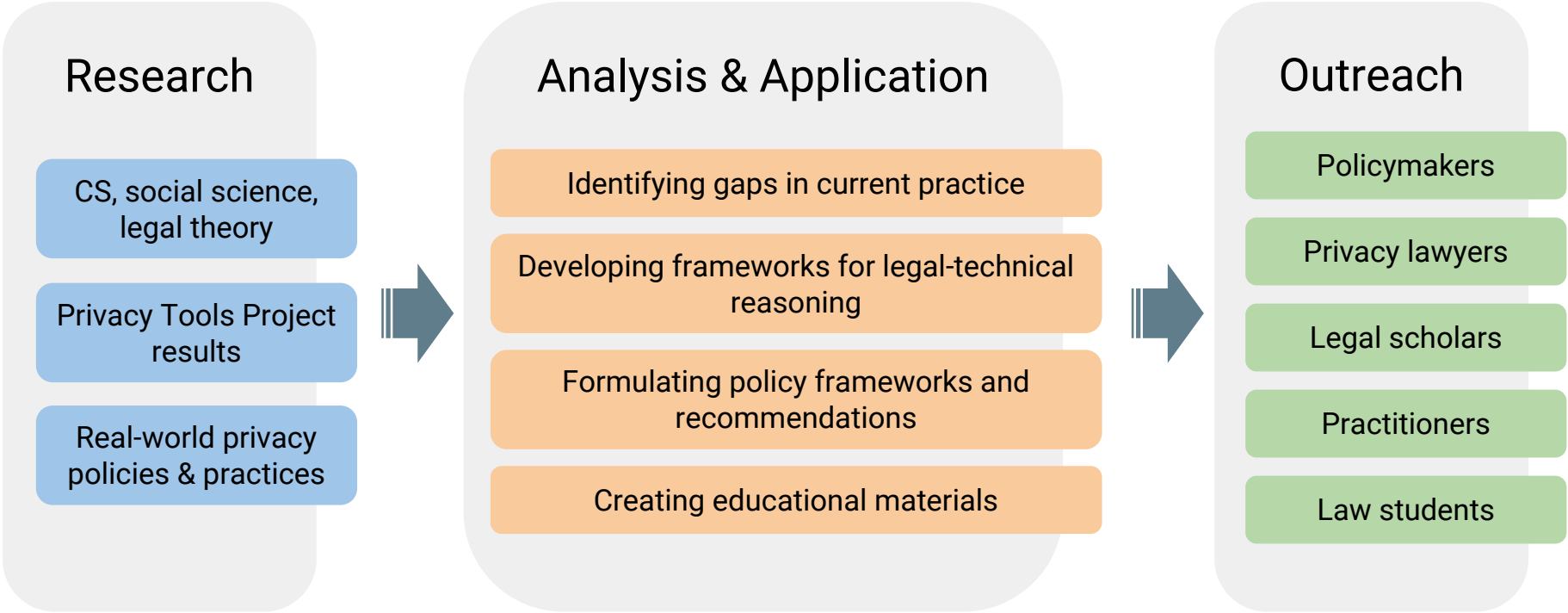
## Social Science & Information Science

Micah Altman

MIT Libraries

# Approach

Transferring project findings and insights from the technical literature to lawyers, policymakers, and practitioners



# Framework for a Modern Privacy Analysis

# Gaps in US Regulatory Framework

- Regulations are sector- and context-specific and highly variable.
- Guidance for implementing privacy safeguards is limited.
- Standards and guidance encourage use of a narrow subset of available privacy safeguards.
- Different actors treat similar privacy risks vastly differently.

# Framework for Privacy-Aware Data Releases

*Modeled on information security and lifecycle frameworks:*

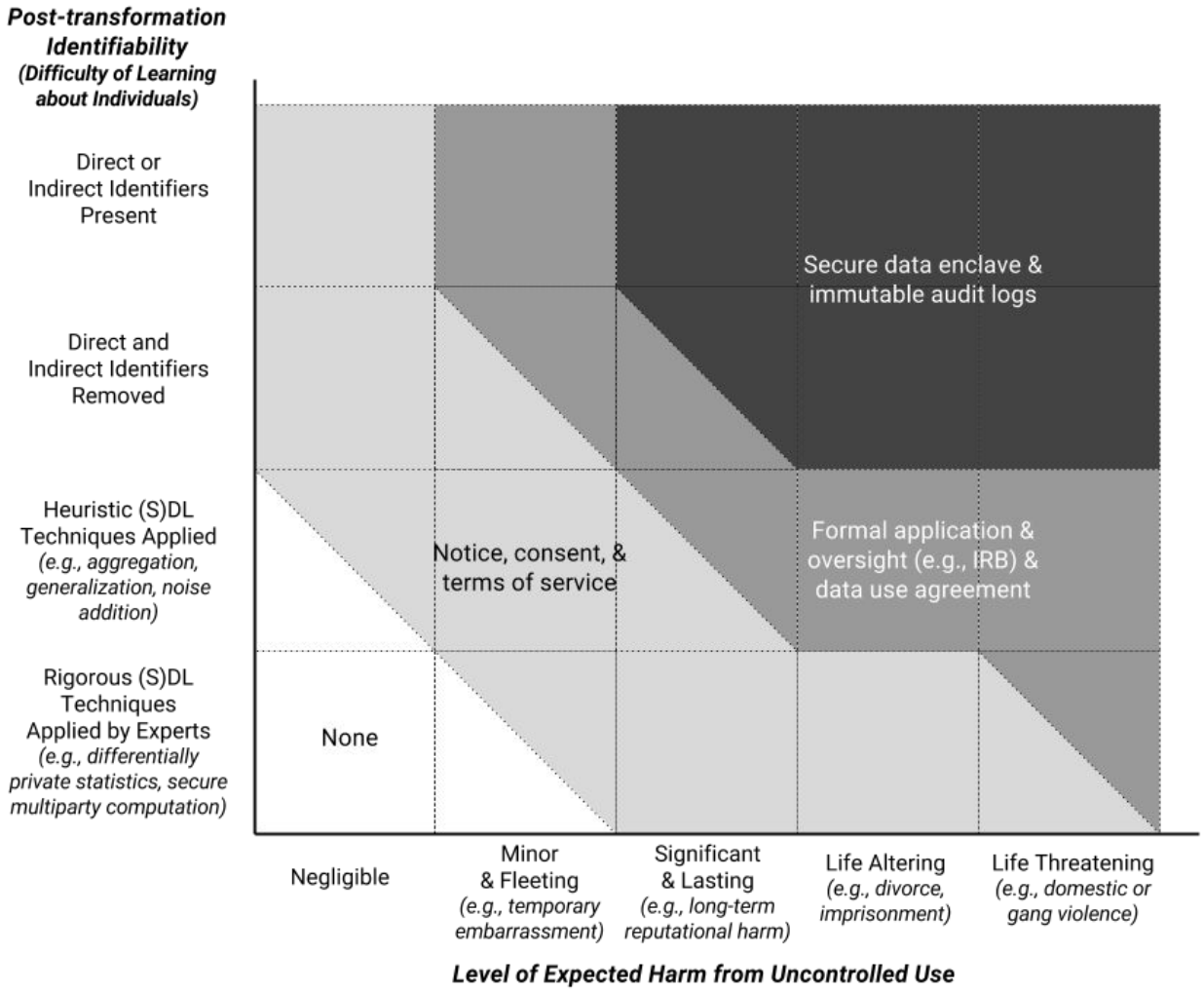
1. Developing a catalog of privacy controls
2. Identifying information uses, threats, and vulnerabilities
3. Designing data releases by aligning uses and risks with controls—at each stage of the information lifecycle

# Catalog of Privacy Controls

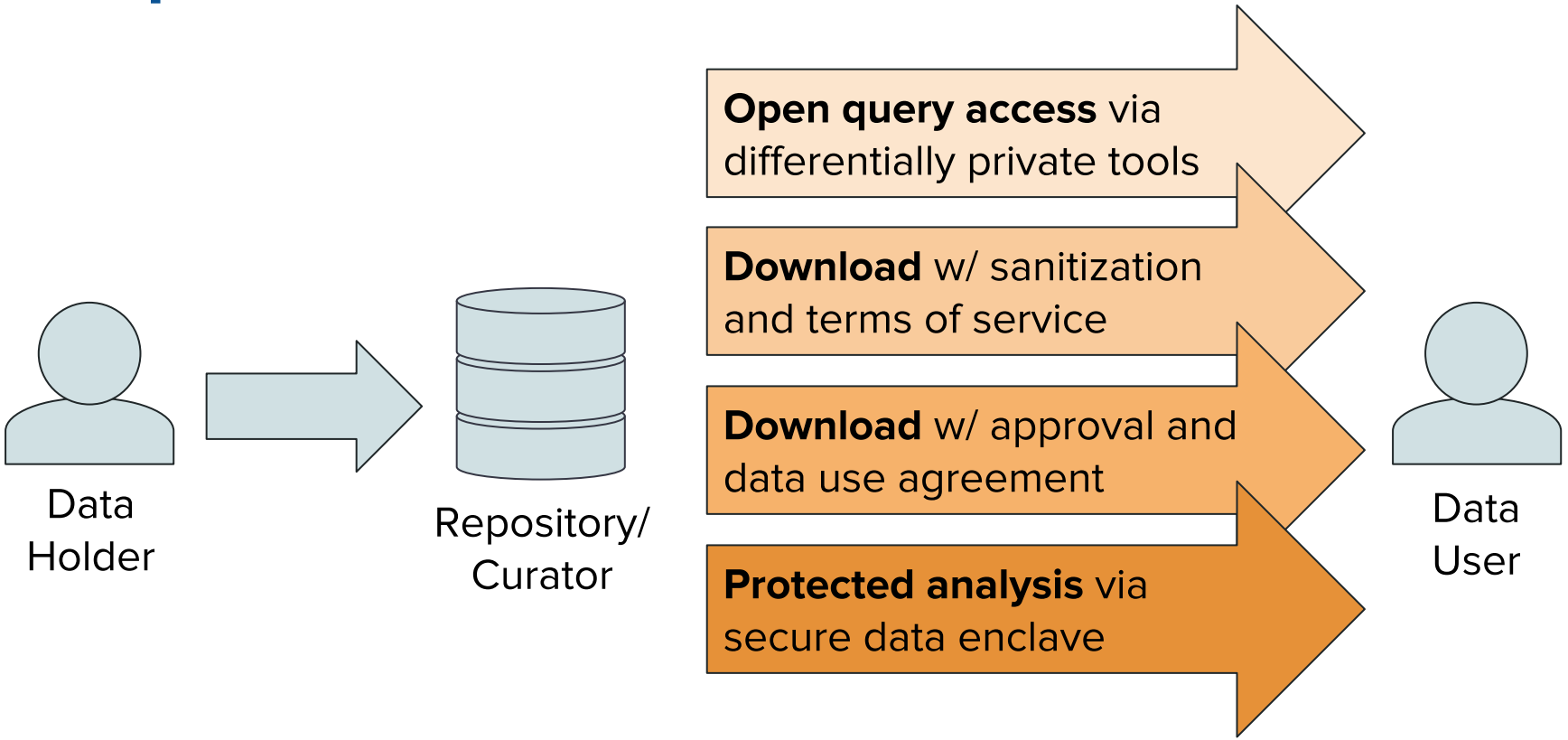
*Procedural, technical, educational, economic, and legal means for enhancing privacy—at each stage of the information lifecycle*

| | Procedural | Economic | Educational | Legal | Technical |
|---|---|---|---|---|---|
| **Access/Release** | Access controls; Consent; Expert panels; Individual privacy settings; Presumption of openness vs. privacy; Purpose specification; Registration; Restrictions on use by data controller; Risk assessments | Access/Use fees (for data controller or subjects); Property rights assignment | Data asset registers; Notice; Transparency | Integrity and accuracy requirements; Data use agreements (contract with data recipient)/ Terms of service | Authentication; Computable policy; Differential privacy; Encryption (incl. Functional; Homomorphic); Interactive query systems; Secure multiparty computation |

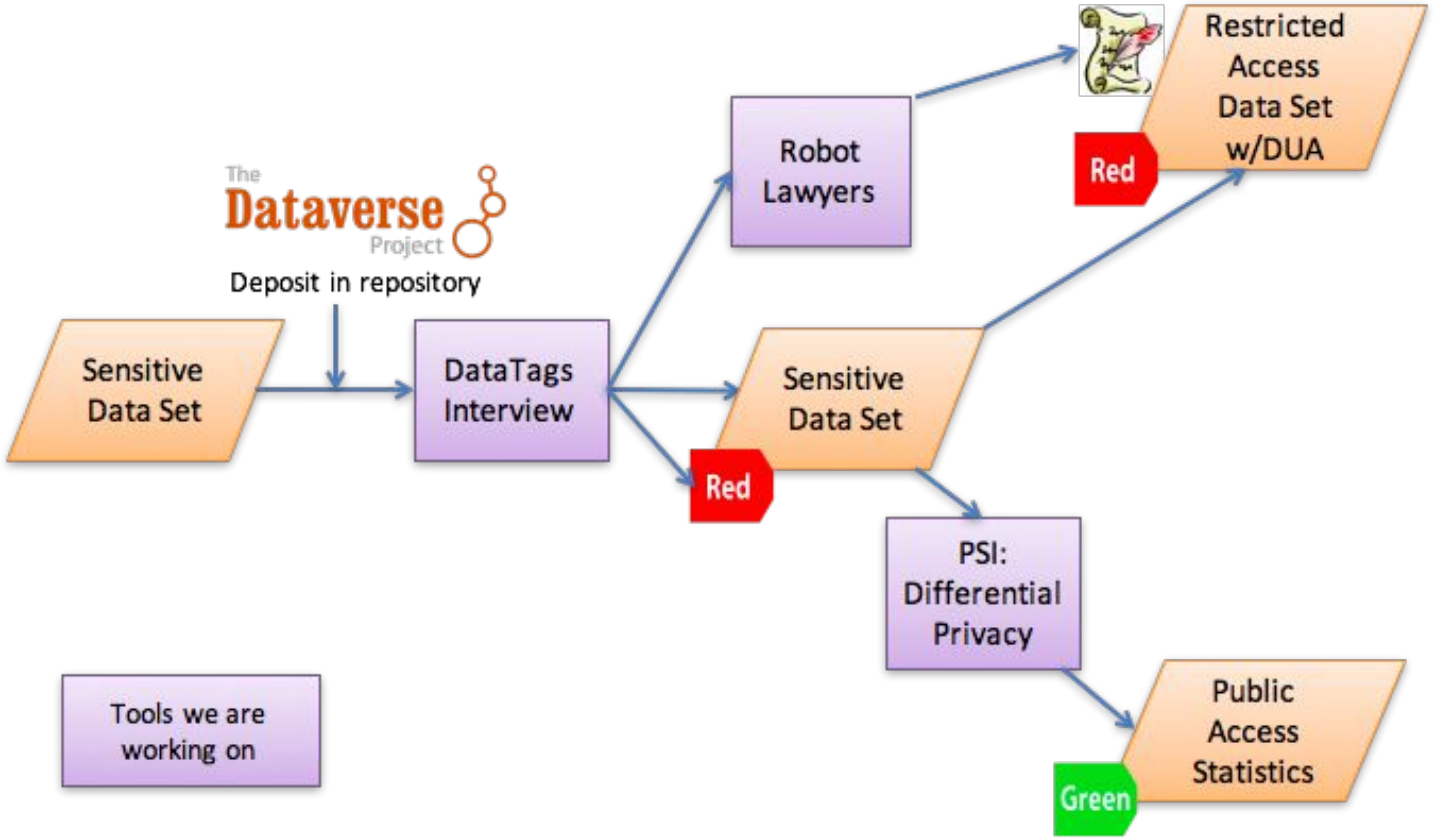
# Guide to Selecting Appropriate Privacy Controls



# Example Tiered Access Model



# Inspired by Privacy Tools Project Model



# Applications of the Framework

## Policy Commentaries

- Occupational Safety & Health Administration
- White House Office of Science & Technology Policy
- Federal Trade Commission
- Department of Health & Human Services
- National Committee on Vital and Health Statistics
- National Institute of Standards & Technology
- Commission on Evidence-based Policymaking
- Future of Privacy Forum

## Scholarly Publications

- Open government data
- Long-term commercial and government big data activities
- Research ethics in commercial settings

# Applications of the Framework

## Policy Commentaries

- Occupational Safety & Health Administration
- White House Office of Science & Technology Policy
- Federal Trade Commission
- Department of Health & Human Services
- National Committee on Vital and Health Statistics
- National Institute of Standards & Technology
- Commission on Evidence-based Policymaking
- Future of Privacy Forum

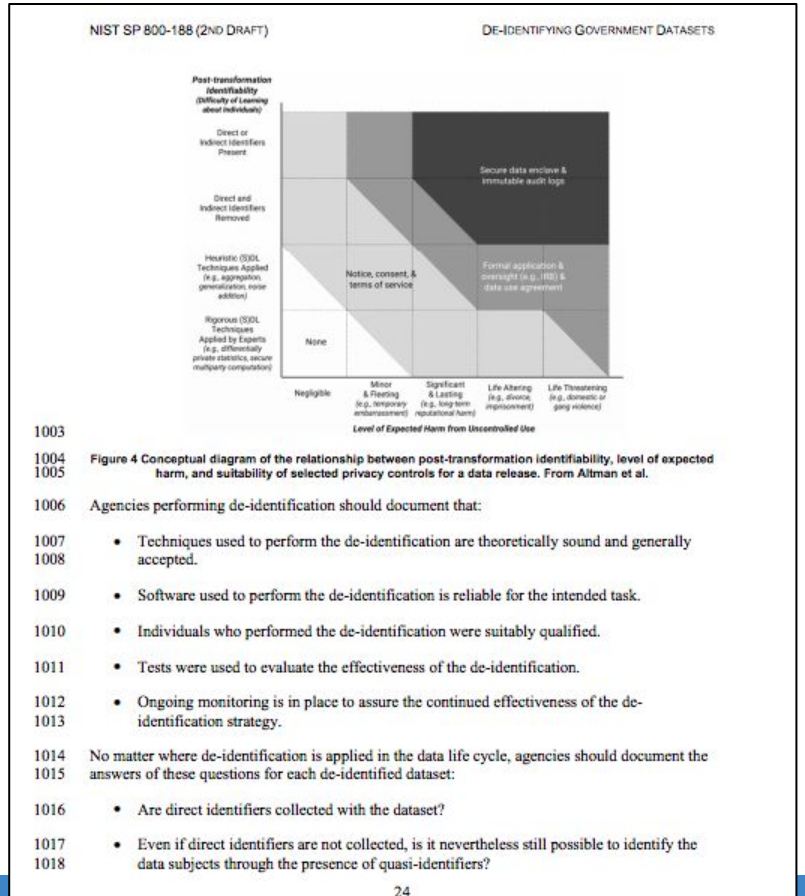






**Figure 4 Conceptual diagram of the relationship between post-transformation identifiability, level of expected harm, and suitability of selected privacy controls for a data release. From Altman et al.**

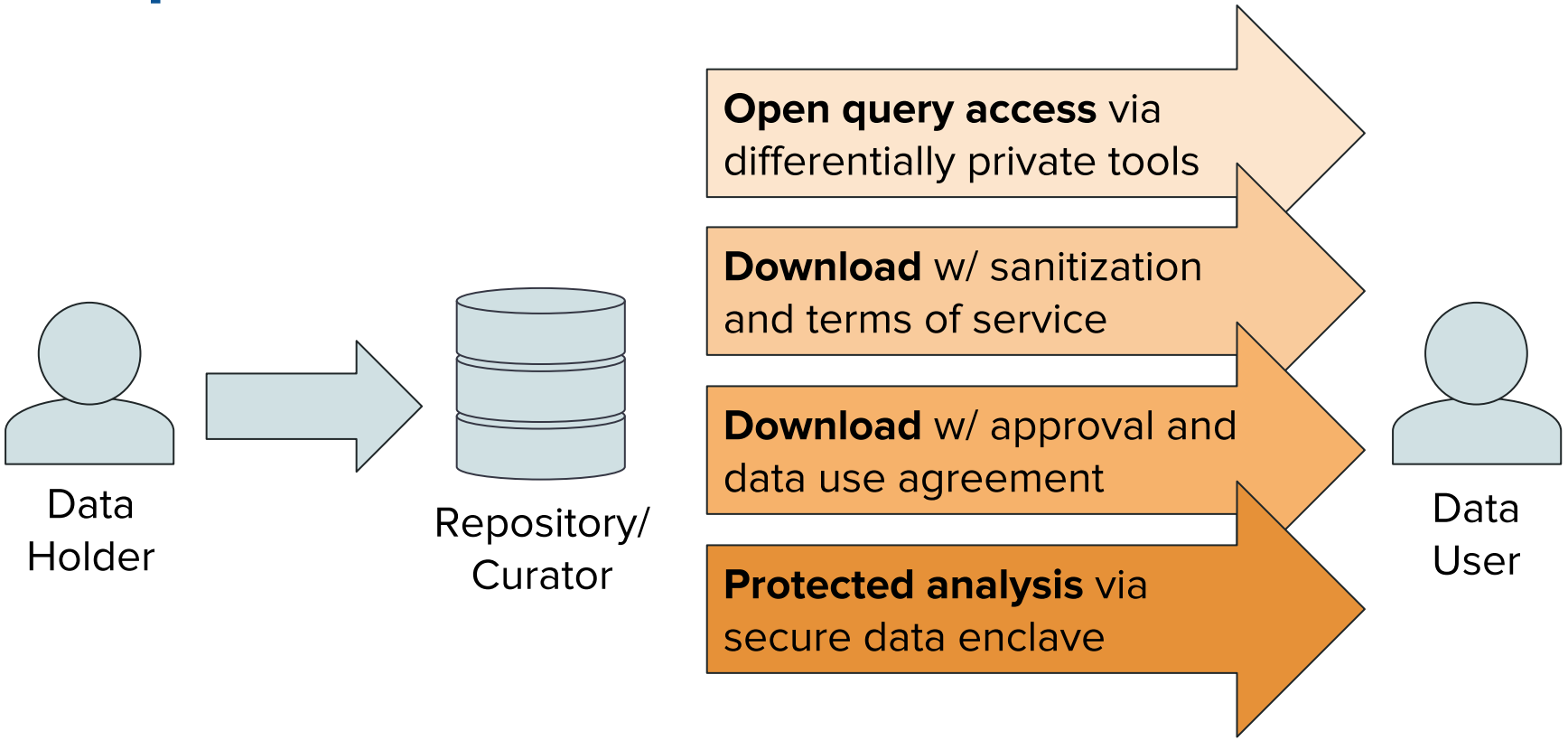
Agencies performing de-identification should document that:

- Techniques used to perform the de-identification are theoretically sound and generally accepted.
- Software used to perform the de-identification is reliable for the intended task.
- Individuals who performed the de-identification were suitably qualified.
- Tests were used to evaluate the effectiveness of the de-identification.
- Ongoing monitoring is in place to assure the continued effectiveness of the de-identification strategy.

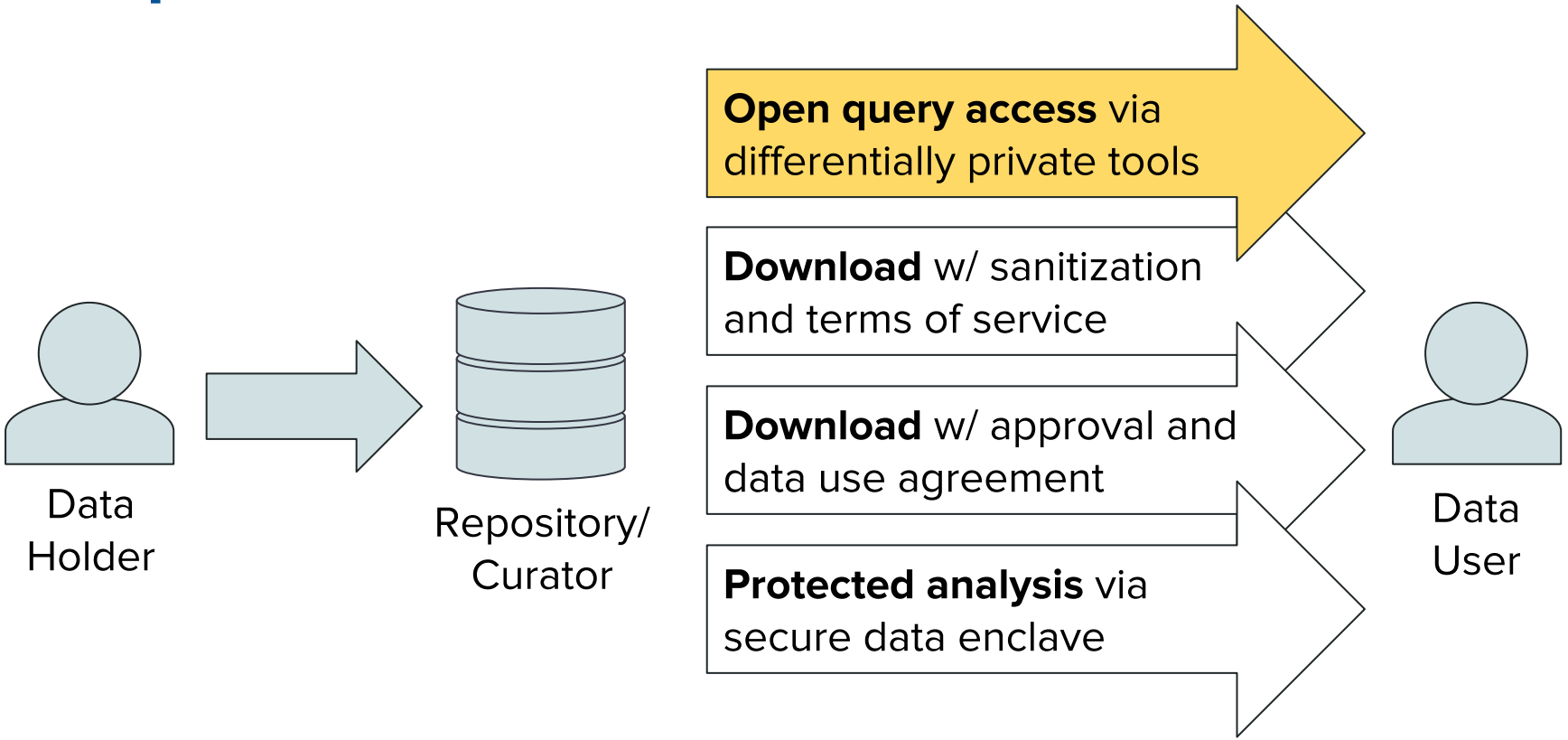
No matter where de-identification is applied in the data life cycle, agencies should document the answers of these questions for each de-identified dataset:

- Are direct identifiers collected with the dataset?
- Even if direct identifiers are not collected, is it nevertheless still possible to identify the data subjects through the presence of quasi-identifiers?

# Example Tiered Access Model



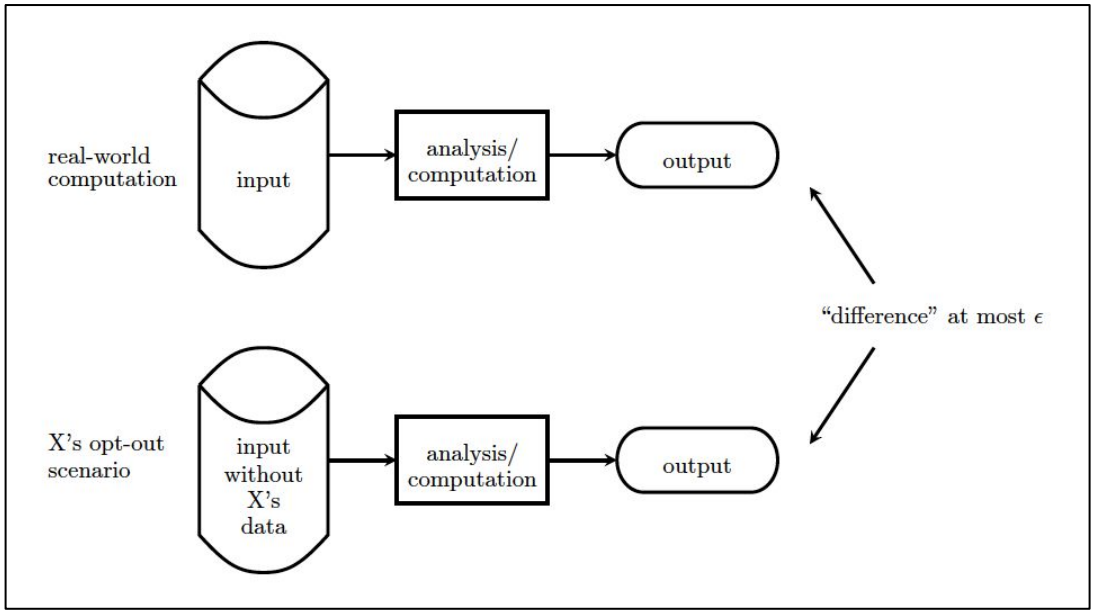
# Example Tiered Access Model



# Explaining Differential Privacy to Non-technical Audiences

# Primer for a Non-technical Audience

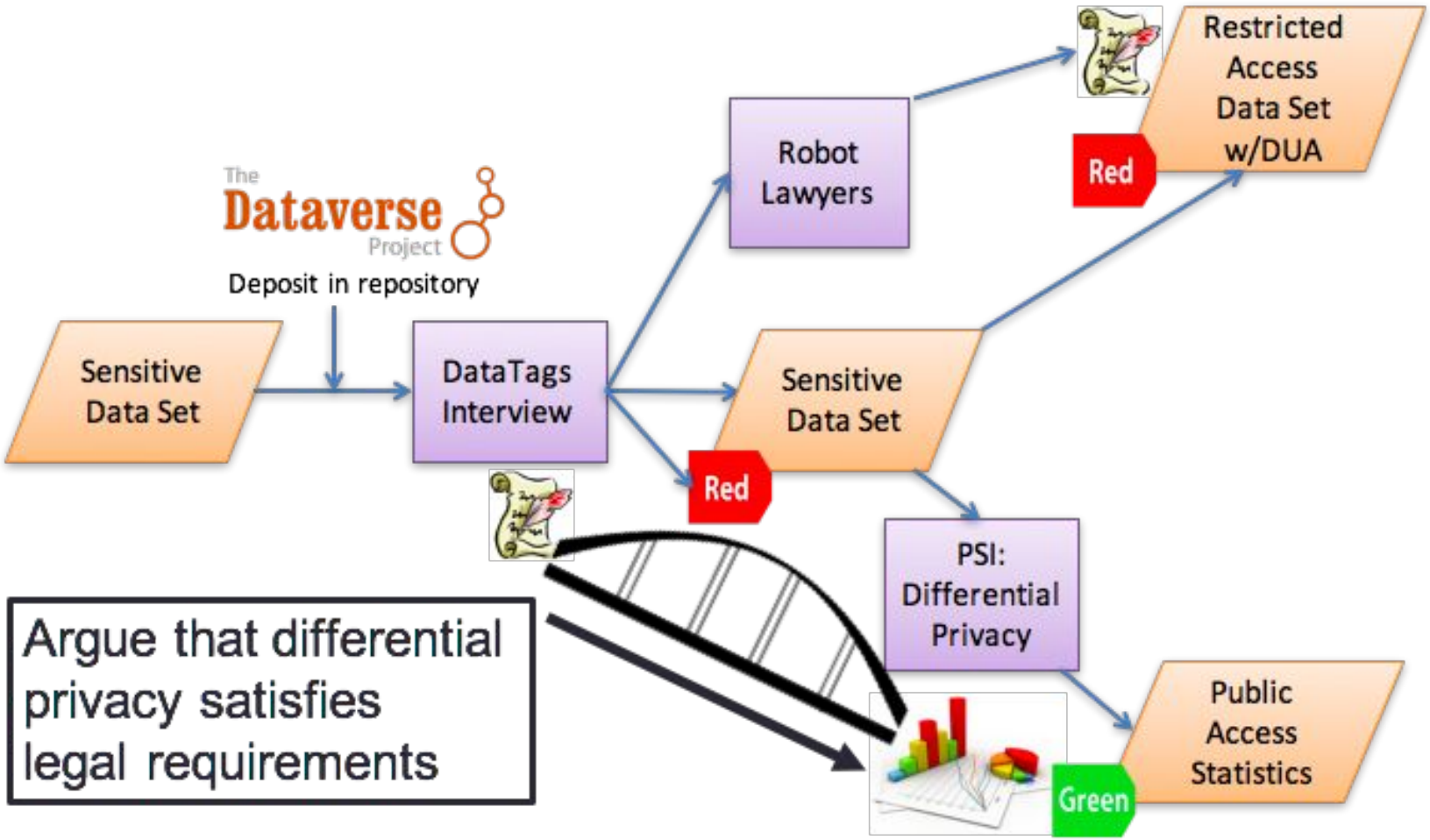
*Introduces differential privacy using simplified and intuitive, but mathematically accurate, illustrations.*



# Primer for a Non-technical Audience

- Provides a foundation to guide future decisions when:
  - Analyzing and sharing statistical data about individuals,
  - Informing individuals about the privacy protection they will be afforded, and
  - Designing policies and regulations for robust privacy protection.

# Bridging Gaps between Differential Privacy and Legal Standards for Privacy


The Dataverse Project
Deposit in repository
Sensitive Data Set
DataTags Interview
Robot Lawyers
Red
Restricted Access Data Set w/DUA
Sensitive Data Set
Red
PSI: Differential Privacy
Public Access Statistics
Green
Argue that differential privacy satisfies legal requirements

# Challenges for Formal Privacy Models

*With the emergence of new technologies based on formal privacy models, can we claim they satisfy existing regulatory requirements?*

Information privacy laws create uncertainty because they are generally:

- **context-specific**,
- **subject to interpretation**,
- allow for some **degree of flexibility**, and
- rely on **traditional, often heuristic, conceptions of privacy**.

# *Is it possible to bridge these very different languages?*


$M: X^n \rightarrow T$ satisfies $\epsilon$-differential privacy if

$\forall x, x' \in X^n$ s.t. $dist_H(x, x') = 1$ $\forall S \subseteq T$,

$$\Pr_{M}[M(x) \in S] \leq e^{\epsilon} \Pr_{M}[M(x') \in S].$$

# Approach #1: Formal Modeling

**Goal**: Rigorously arguing that a technological privacy solution satisfies the requirements of a particular law.

**Proposed approach**:

1. Extracting a formal mathematical requirement from the law.
2. Proving mathematically that a technological privacy solution satisfies the requirement derived from the law.

# Illustration: Formally Modeling FERPA

We extracted a formal model of the privacy desiderata for the Family Educational Rights & Privacy Act (FERPA).

We used a **game-based privacy definition**:

- This provides a concise and fairly intuitive abstraction of FERPA's requirements.
- If a formal model, such as differential privacy, satisfies the definition, then we have a strong argument that it satisfies the requirements of FERPA.

# Modeling FERPA: The Adversary

**Personally identifiable information**: “information that, alone or in combination, is linked or linkable to a specific student that would allow a reasonable person in the school community, who does not have personal knowledge of the relevant circumstances, to identify the student with reasonable certainty.”

This is FERPA’s **implicit adversary**.

# Modeling FERPA: Directory Information

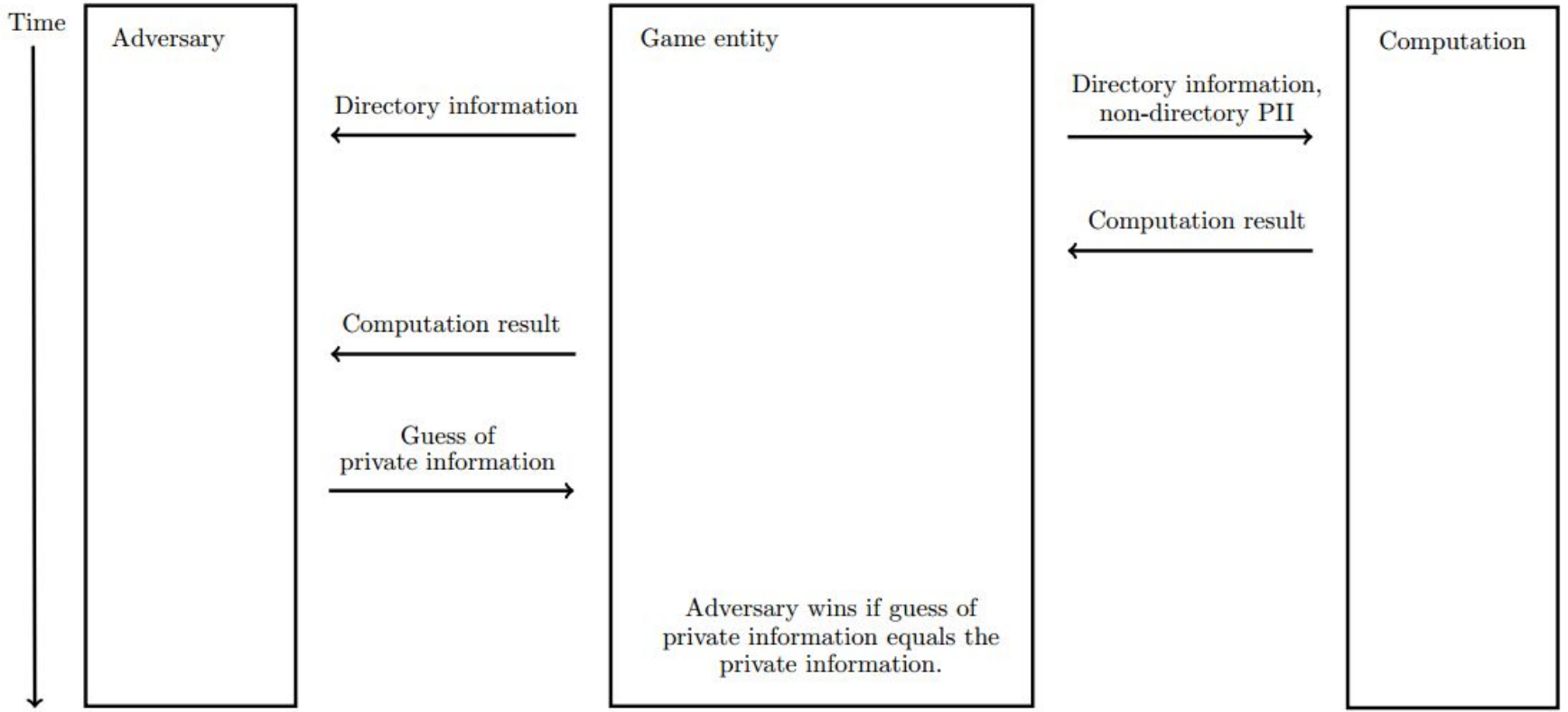
- Regulatory language interpreted conservatively, i.e., erring towards what is most beneficial for the adversary.

- E.g., the definition of **directory information** is ambiguous.

  - If we made assumptions about the definition, new interpretations could call these assumptions into question.

  - Instead, **we let the attacker to choose** what constitutes directory information.

# Components of a FERPA Privacy Game



# Approach #2: Interpreting the Differential Privacy Guarantee

- Legal requirements relevant to issues of privacy in computation rely on an understanding of a range of different privacy concepts.

- None of the privacy concepts that appear in the law refer directly to differential privacy.

  - However, the differential privacy guarantee can be interpreted in reference to these concepts—while accommodating differences in how these concepts are defined across contexts.

# Illustration: Personally Identifiable Information

- Legal protections typically extend only to PII.
  - e.g., FERPA, HIPAA Privacy Rule, Massachusetts data security regulation, OMB memorandum
- Definitions vary significantly, but generally understood to refer to the presence of pieces of information that are linkable to the identity of an individual or to an individual's personal attributes.

# PII: Interpreting Differential Privacy Guarantee

- PII does not have a precise technical meaning.
- It can be difficult to determine whether information is PII.
- Regardless of the definition of PII that is used, differential privacy can be interpreted as (essentially) ensuring that using an individual's data will not reveal any PII that is specific to her.
  - Here, **specific** is used to refer to information that cannot be inferred unless the individual's information is used in the analysis.

# Lessons Learned

- There are **significant, seemingly irreconcilable, gaps** between computer science and legal approaches to privacy.

- **Progress through long-term, deep interdisciplinary collaboration:**
  - Framework for a modern privacy analysis
  - Educational materials explaining recent computer science advances
  - Novel approaches to legal-technical privacy reasoning

- **Rich area full of potential research directions:**
  - New research through cooperative agreement with the US Census Bureau
  - Opportunities to explore differential privacy and contextual integrity, approaches to setting formal privacy parameters

# Selected References

Kobbi Nissim, Aaron Bembenek, Alexandra Wood, Mark Bun, Marco Gaboardi, Urs Gasser, David R. O'Brien, & Salil Vadhan, **Bridging the Gap between Computer Science and Legal Approaches to Privacy**, 31 Harvard Journal of Law & Technology __ (forthcoming 2018).

Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien, & Salil Vadhan, **Differential Privacy: A Primer for a Non-technical Audience** (Preliminary Version) (2017).

Alexandra Wood, Edo Airoldi, Micah Altman, Yves-Alexandre de Montjoye, Urs Gasser, David O'Brien, & Salil Vadhan, **Comments on the Proposed Rules to Revise the Federal Policy for the Protection of Human Subjects** (2016).

Effy Vayena, Urs Gasser, Alexandra Wood, David R. O'Brien, & Micah Altman, **Elements of a New Ethical Framework for Big Data Research**, 72 Washington & Lee Law Review Online 420 (2016).

Micah Altman, Alexandra Wood, David R. O'Brien, Salil Vadhan, & Urs Gasser, **Towards a Modern Approach to Privacy-Aware Government Data Releases**, 30 *Berkeley Technology Law Journal* 1967 (2015).

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