Towards the Science of Privacy in support of the Art of Privacy

NSA Civil Liberties & Privacy Office

May 2015

Agenda

- Civil Liberties & Privacy at NSA: Vision & Mission
- Civil Liberties & Privacy Assessments, Big Data, & Privacy Risk
- Back to Basics: What *is* Privacy?
- Towards the Science of Privacy
  - Challenge Questions
  - Proposed Framework

Office of Civil Liberties & Privacy: Vision & Mission

Trusted by the public to uphold civil liberties and privacy values as NSA protects America and its allies.

1. Primary Civil Liberties & Privacy advisor to DIRNSA
2. Build meaningful Civil Liberties & Privacy processes
3. Improve protections through Civil Liberties & Privacy research
4. Enhancing meaningful transparency

Building a Systemic & Holistic Approach to Civil Liberties & Privacy

- Developing Civil Liberties & Privacy Assessments
- Focusing on *data* and *use*
- Conducting necessary analysis of activities

Creating a Common Lexicon

DATA

- **Type**
  - Personal Information
    - Biographic
    - Biometric
    - Contextual
- **Bulk or Targeted**
  - *Targeted*: Known, identified target
  - *Bulk*: Known targets, unknown targets, and non-targets intermixed
  - *Gradation as well!*

USE

- **Purpose**
  - Counter-Terrorism
  - Counter-Proliferation
  - Counter-Intelligence & Intents of Foreign Governments
  - Cyber-security
  - Transnational Criminal Threats
  - Threats to Military & Allies
- **Analytical Activities**
  - Discovery
  - Targeted Collection
- **Technological Function**
  - Correlate, filter, format, etc.
  - *Continuing to study what constitutes a set of technological functions…*

- Benefits of both the *Art* and *Science* of Privacy and Civil Liberties analysis

- Big Data and Big Data Analytics challenge existing methodologies to evaluate privacy risk and protections.
  - Every newly introduced data set can upend prior assumptions of privacy risk.
  - Every new analytic or combination of analytics in a workflow can upend prior assumptions of privacy risk

How can one build a scalable and manageable CLP assessment process in the Era of Big Data?

- While generally understood, "Privacy" is a contested concept
  - The Right to be Forgotten? The Right to Hide? The Right to Conceal? No Right at All?

How can personal information be effectively identified and protected?

Developing a *Practicable* and *Scalable* approach to privacy protections

- **Establish a *common lexicon* for *data* and *use*.**
- ***Assumption:* Privacy is a *Data-Driven* and *Use-Driven calculation*.**
- ***Assumption:* Built upon existing compliance and security framework**
- ***Assumption:* Privacy is the means by which one protects Civil Liberty (i.e. Individual Liberty/Free Will/Self-Determination are foundational principles of the U.S. Constitution and Declaration of Independence)**

Data + Use ➔ Identify and Quantify Privacy Risk

Back to Basics

- What are the key attributes of a person that need to be protected?
- What is ***use*** and what does it mean to *use* personal information?
- What are the ***risks to individual's privacy?***
- How to handle ***context***?

Towards the Science of Privacy

- Identify and develop rigorous, repeatable, and scalable ***methodologies*** to empirically evaluate the risks and protections of Big Data and Big Data analytics on individual privacy.

- Create a ***framework*** to underpin a ***Privacy Decision Support Process***.

- Identify & Understand Risks & Protections with respect to individual privacy in my digital daily life.

Challenge Questions

1. What are the actual individual privacy risks that need to be considered?

2. Can methods be developed to evaluate privacy risk based on the type of personal information present and the type of use(s) of that personal information?

3. How can an *Accountable Privacy Framework*† be created for Big Data, building upon an existing compliance and security framework, that evaluates privacy risk based on *the type of personal information* and *type of use(s) applied*?

4. How can we apply current advances in privacy engineering? (e.g., Differential Privacy, Homomorphic Encryption, Secure Multi-Party Computation)

† An *Accountable Privacy Framework* is a capability for big data processing systems (e.g., Cloud) that evaluates privacy risk based on data (i.e., personal information) and use.

Initial Thoughts: Towards the Science of Privacy

- Have begun initial investigations into potential ways to quantify individual privacy risk in big data.
- Following is a *proposed* methodology.
- This initial methodology is a *work in progress.*
- Intended to provoke discussion on appropriate methodologies and concepts for further research.
- Is *not* intended to proscribe alternate approaches.

Initial Thoughts: Attributes of a Person

- Focusing on a what are the attributes of a person that may impact their privacy:

Any tangible information that can be used to identify an aspect of a person. (To include specific facts such as a name or address as well as patterns of behavior.)

- Attempting to apply consistent taxonomy for personal information:
  - **Biometric**: Measurable, physical characteristics of an individual. (e.g., fingerprint, blood type, gait, gender).
  - **Biographic**: Attestable facts about an individual's life. (e.g., name, address, religion).
  - **Contextual**: Identity data from individual's transactions. (e.g., financial, commercial transactions, personal patterns).
- Investigating methods to evaluate/assign relative privacy risk for each category.

Initial Thoughts: Use Taxonomy

- Have identified a handful of hierarchies of "use".
  - *Purpose & Analytical Activities* – more *subjective*, per business needs.
  - *Technological Functions* – more *objective*, per analytical processes.
- Focusing initially on *Technological Functions*:
  - Analytics decompose into atomic technological functions (e.g., filter, correlate, etc.).
  - Composite analytic workflows can be constructed from individual analytics, each consisting of atomic technological functions.
- Need to identify the set of technological functions and accompanying semantic definitions.
- Investigating how to assign/evaluate privacy risk to types of uses of technological functions.

Initial Thoughts: Conceptual Diagram of Data & Analytic Use