Breakout group 1: Technical challenges and opportunities for transnational data repositories: How to represent, reason about, and enforce diverse security, privacy, and retention policies?

Co-leads:
Anupam Datta, Carnegie Mellon University
Artur Hecker, Telecom-ParisTech
## Starting questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the purpose of data sharing?</td>
<td>What do we win? What can we do what we cannot do otherwise?</td>
</tr>
<tr>
<td>What are typical applications?</td>
<td>Critical infrastructure protection, enterprise IT security improvement, law enforcement…</td>
</tr>
<tr>
<td>What are examples of scenarios for international data exchange?</td>
<td>Network attack signatures, health records, electrical grid faults, …</td>
</tr>
<tr>
<td>What incentives promote/deter data sharing?</td>
<td>Broadening views, collaboration of defenders vs. difficulties collecting and evaluating data</td>
</tr>
</tbody>
</table>

Not everything that can be counted counts and not everything that counts can be counted.  

* A. Einstein
Starting questions – Policies

What are acceptable policies governing such exchanges?

How do we express and enforce such policies?

Contextual Integrity,…

Preventive policy enforcement, e.g., access control

Privacy through Accountability (Incentives, Audit, …)

Support privacy-preserving transfer of and computation over aggregate information
More Detailed Questions

- What is the purpose and mechanism of the exchange?
  - What useful function can be computed over the released aggregate data? (e.g., technical results in differential privacy)
  - What is an appropriate data exchange architecture?

- What are acceptable policies (information sharing, use, retention) governing such exchanges?
  - How can we systematically determine what policies are socially acceptable? (e.g., studies in contextual integrity)

- How do we express policies?
  - Design of policy languages with relevant concepts (what are the concepts, what is new) (e.g., for access control, privacy, provenance, data deletion)
More Detailed Questions

- How can such policies be enforced?
  - Preventive policy enforcement methods (access control, history-based policies, future obligations, analysis of organizational processes) [logic, programming languages, formal methods]
  - Privacy through accountability and audit (what kind of policies can be enforced in this manner?) [formal methods, game theory]
  - Respecting privacy in release of and computation over aggregate information (what are the inherent limits of privacy vs. utility of data?) [statistics, cryptography, behavioral studies]
- What incentives promote/deter data sharing? [economic, political, legal issues]
Example 1: Privacy-Preserving Healthcare Systems

**Goal:** Respect privacy expectations in the transfer and use of individual and aggregate personal information within and across organizational boundaries
What Can go Wrong?

Insurance company

Network adversary

Billing info

Patient

Public

Health statistics

Irresponsible insider

Patient database

Nurse

Doctor

Admin

Linking adversary
Security and Privacy Mechanisms

- Cryptographic network protocols
- Privacy-preserving data release
- Regulation compliant business processes
- Audit and accountability

- Patient
- Public
- Nurse
- Doctor
- Admin
- Insurance company
- Patient database
- Billing info
Example 2: Network Attack Data

- What is the data used for?
  - Understanding the underground economy
  - Attack attribution

- Why is data not shared? Legal, political, social constraints? Incentives misaligned? What is needed to enable sharing of network attack data?

- What policies should govern such data exchanges?
  - Can sanitized (privacy-preserving) data be used for the same purpose?
  - Prevent further dissemination (share summaries)
  - Cannot just publish data – privacy concerns, inform adversaries
  - Who, when, where to share data with and with what expectations of further flow and use?

- How can we enforce such policies?
Synthesized Notes from Breakout
Synthesized notes

- What are some examples in which international data exchange is relevant?
  - Data that countries are putting in for compliance towards international regulations (emissions etc.)
  - Cyber-crime (attack and defense patterns), global pandemics, financial data sharing, rapid incident-response

- What is different about international data exchange?
  - Cultural norms, laws
  - International technology platforms
  - Incentive-based enforcement
  - Laws are different
  - Authorization and credentials different
  - Distance between administrative domains (co-operation and input across countries in standardization efforts)
  - Standardization of formats (semantic web etc.) + best practices for data sharing
Synthesized notes

- Focus on specific domains and identify general abstractions and technology goals
  - Data, policies, mechanisms, architectures
  - Use and purpose of data
  - What are the abstractions? Perhaps, “meta-formalisms” using which different formalisms used by different countries can be integrated.
  - What specific policies do people care about in the real world?
  - Are these policies in conflict? Can we develop formalisms and tools that help manage inconsistencies?
Some directions

1. Representing and structuring data
   - Standards for structured text

2. Policy representation and understanding
   - Understanding the law, social norms (higher level concepts, meta-language?)
   - Expressive policy languages
   - Policy analysis and conflict detection
   - Formalization and semi-automated enforcement
   - Usability
Some directions

3. Architectures and enforcement

- crypto, private data analysis, watermarking (aggregate information)
- Identifying anomalous behavior
- Audit and accountability
- Managing risks (economic analysis)
- Access control, preventive
- Integrity of data
- Undo decisions, provenance, retraction, update
- Process centric vs. data-centric architectures
- Usability
- Compositional enforcement of policies under constraints (e.g., limited monitoring in some countries, data retention policies, support accountability) What are the design principles? Driven by specific applications.
- Work in legal requirements and social norms into different points of the architecture (e.g., integrity of data could be a legal obligation)
Some directions

4. Specific test-beds

- Cloud platforms in which data move around and associated policies change
- Create a platform:
  Provide an interface for people to contribute data to a data repository that can be used as a testbed (e.g., network attack data, DHS, clinical studies data, vulnerability data)
- Types of data, data ontology
- Cloud computing platform offers generality is international by default; could be specialized to different application domains DRM, health care, … (e.g., Google file system) What should the design of such a system be to support enforcement of policies across boundaries?