Why Are We Talking About Data Sharing?

Is There a Problem? Do we care?

Erin Kenneally, CAIDA, Elchemy

Why Are We Talking About Data Sharing?

- Lack of scale re: resources and exchange solutions
- Lack of shared skills and knowledge/ understanding between stakeholders
- Easy to stonewall, risk wins (easy for risk manager to say "no")

Is There a Problem?

- Data complexity
  - Massively heterogeneous data
  - Huge volume of data
  - Many different types of actors to consider
- Difficulty bounding attack risk
  - Cannot quantify access to secondary data sources
  - Privacy definitions are immature for network data
- Interactions between policy and technology
  - Not just PII → intellectual property, network security, etc.
  - Lack of legal precedent or guidance for network data

Reality of Data Sharing

- Uncertainty of legal risk
- Understated value of potential benefits
- One-size-fits-all approach to disclosure controls
- Implicit assumption that **any** sharing increases risk

- Results in:
  - Data rich vs. data poor
  - Sharing through ad-hoc, interpersonal relationships
  - Scarcity of scalable, transparent, sustainable sharing

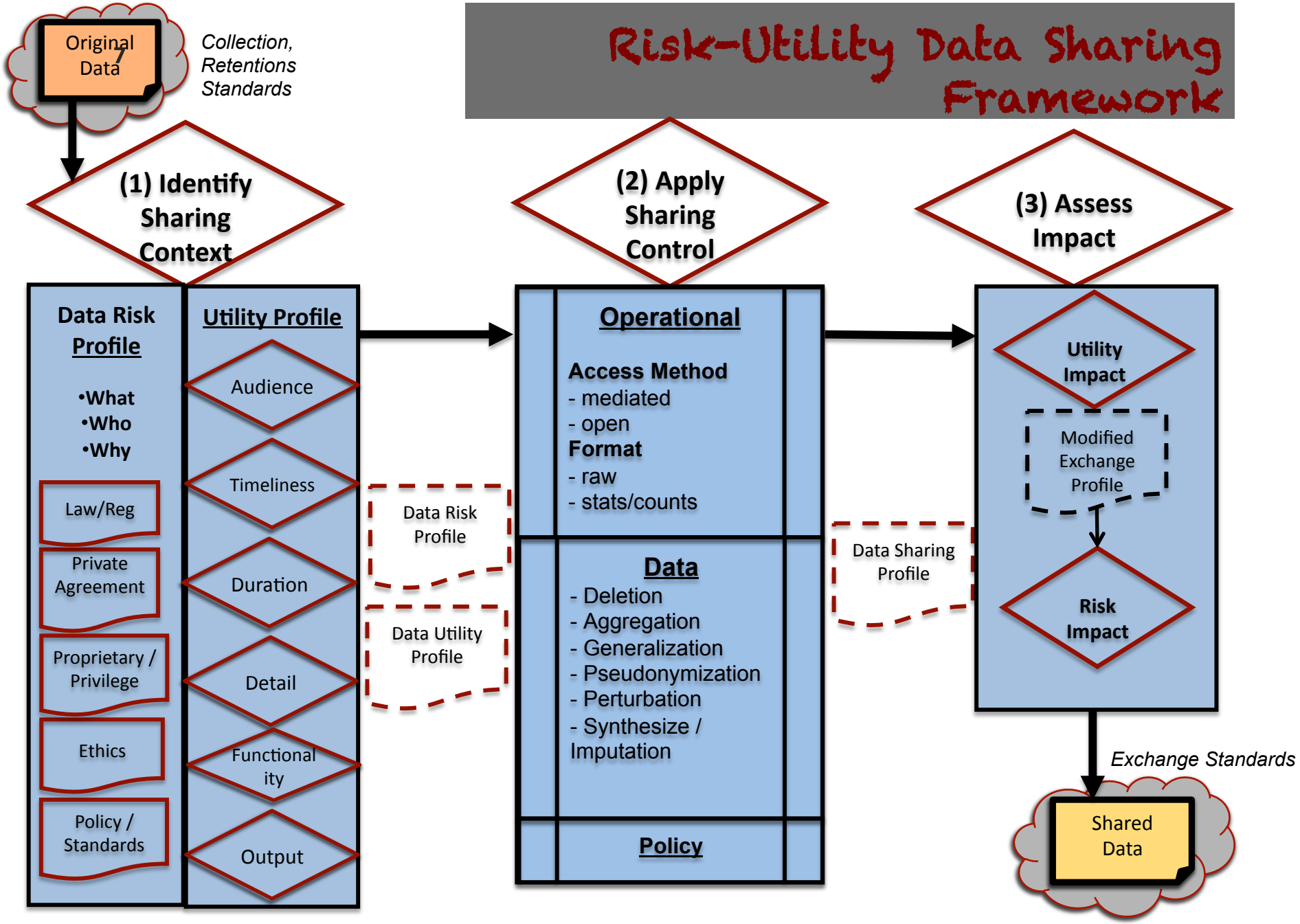
What to do?

- Qualitative framework for:
  1. Identifying specific utility goals and related risks
  2. Choosing disclosure controls to address risks
  3. Assessing effects of those controls
- Generalizable across all network data & scenarios
- Enable data providers to:
  - Better understand sources of risk
  - Tailor controls to intended utility
  - Justify choices and explicitly state assumptions
- Promote the social value of shared data & process

This is NOT Death or Taxes

- Does NOT provide yes/no answers
  - Data sharing is a risk management process
  - Appetite for risk varies significantly

- Attacks may exist or information may be leaked
  - Understand what risks exist
  - Justify disclosure control choices



<http://ssrn.com/abstract=2032315>

© 2012 Elchemy | Redjack

Your Turn: Data Sharing

- **Drivers:**
  (1) What are the major factors in your decision to collect and share network measurement data?

- Risk IQ:
  (2) Do you feel like you have a strong understanding of the risks (legal, contractual, etc.) of sharing network data?

- (3) Do you feel like you have a strong understanding of the available controls for mitigating those risks (both technical and policy)?

Your Turn: Data Sharing

- Incentives:
  (4) What (if anything) would motivate you to collect and share more network data with the research and operational community?
  - (a) Better understanding of best practices (both technical and policy)?
  - (b) A community-driven best practices document? Format if not doc?
  - (c) How detailed should the guideline be? General categories? specific implementation details?
  - (d) Would you expect the guidelines to provide a quantifiable risk score, or is general discussion of the concepts sufficient?

Your Turn: Data Sharing

- Components:

(5) What should a Best Practices Guide include to improve data sharing?

- (a) Description of policy risks (e.g. laws, contracts, and ethical guidelines)?
- (b) Description of intended utility objectives (e.g. publicly available research, private operational release)?
- (c) Description of available disclosure controls, their benefits, and potential pitfalls? Technical? Policy?
- (d) Description of threat considerations and how they impact how well disclosure controls will work?
- (e) What are we missing?

Resources

- A Qualitative Risk Assessment Framework for Sharing Computer Network Data, Coull, Kenneally (2012)

  https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2032315

- Cyber-security Research Ethics Dialogue & Strategy

  IEEE Security & Privacy Symposium, San Francisco, May 23, 2013
  http://www.caida.org/workshops/creds