Less is More :
a Disclosure Control Approach to Data Sharing

Erin Kenneally
CEO, Elchemy
UC San Diego
erin@elchemy.org

Scott Coull
Redjack

What's the Problem?

- Technologists optimize for **EFFICIENCY**
- Lawyers optimize for **CERTAINTY**
- Collective Data Sharing needs to optimize for **TRUST**

Data Sharing Truisms

- Technologists
  - Option A:

  - Option B:

- Lawyers
  - Option A:

  - Option B:

Data (un)Sharing Symptoms

- 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
  - Self-perpetuating scarcity : scalable, transparent, sustainable sharing

Network Data Sharing Causes

- **Difficulty bounding attack risk**
  - New inference attacks being developed
  - Can’t quantify access to 2ndry data sources
  - Privacy applications immature for network data
- **Data Complexity**
  - Huge data volumes and heterogeneity
  - +++ protocols and new ones being developed
  - Many stakeholders to consider
- **Interactions between policy and technology**
  - Different levels of risk tolerance, control IQ & needs
  - Exponential number of unique scenarios to cover

How to Optimize for Trust : Less is More

- **The Data Sharing Weight Loss Plan** (psychic)

  - **Technologists** (less inefficiency)
    - < operational overhead
      (reuse established infrastructure)
    - < legal overhead
    - > ROI : data utility
  - **Lawyers** (less uncertainty)
    - < legal risk (perception & reality)
      - transparency WHO, WHAT, WHY
    - < reputation risk
      - standardized auditing/accountability

Disclosure Control Framework- 3 Components

- **Profiles (Risk & Utility)**

  - associated with a dataset, methodology for considering in concert;
  - DP can create concise, standardized audit trail of the decision-making process underlying the data disclosure

- **Templates (library, reusable)**

  - **Policy**: represent common legal and policy documents with placeholders for scenario-specific answers
  - **Data**: description of released data & method of parsing
  - **Technical**: methods for applying specific technical disclosure controls

- **Environments (library, reusable)**

  - sets of Templates chosen and configured by the DP during the risk assessment and data sharing process
  - includes utility/risk profiles that explicitly state publisher assumptions

Templates e.g.

Document$Template

```
"name": "Privacy Notice",
"description": "A privacy notice and terms of use.",
"categories": ["upstream", "terms", "covenants"],
"text": "We collect [#(COLLECTED_DATA)] kinds of
        information to measure the performance of
        your mobile broadband service.
        [COLLECTED_DATA]
        This data is protected using
        [PROTECTIONS]. You can find more detail in
        the FCC's technical summary of this program.",
"questions": [
        {"question": "Enumerate data items collected.",
        "answer": "COLLECTED_DATA"},

        { "question": "Enumerate protections for raw
                data after collection.",
        "answer": "PROTECTIONS"}
]
```

Technology$Template

```
"name": "Quantize_Location",
"description": "Aggregate location data into blocks.",
"categories": ["aggregation", "location_data"],
"pointer": "http://example.com/quantize_loc"
"parameters":
{
        "k": "10",
        "granularity": "0.5"
}
```

Data$Template

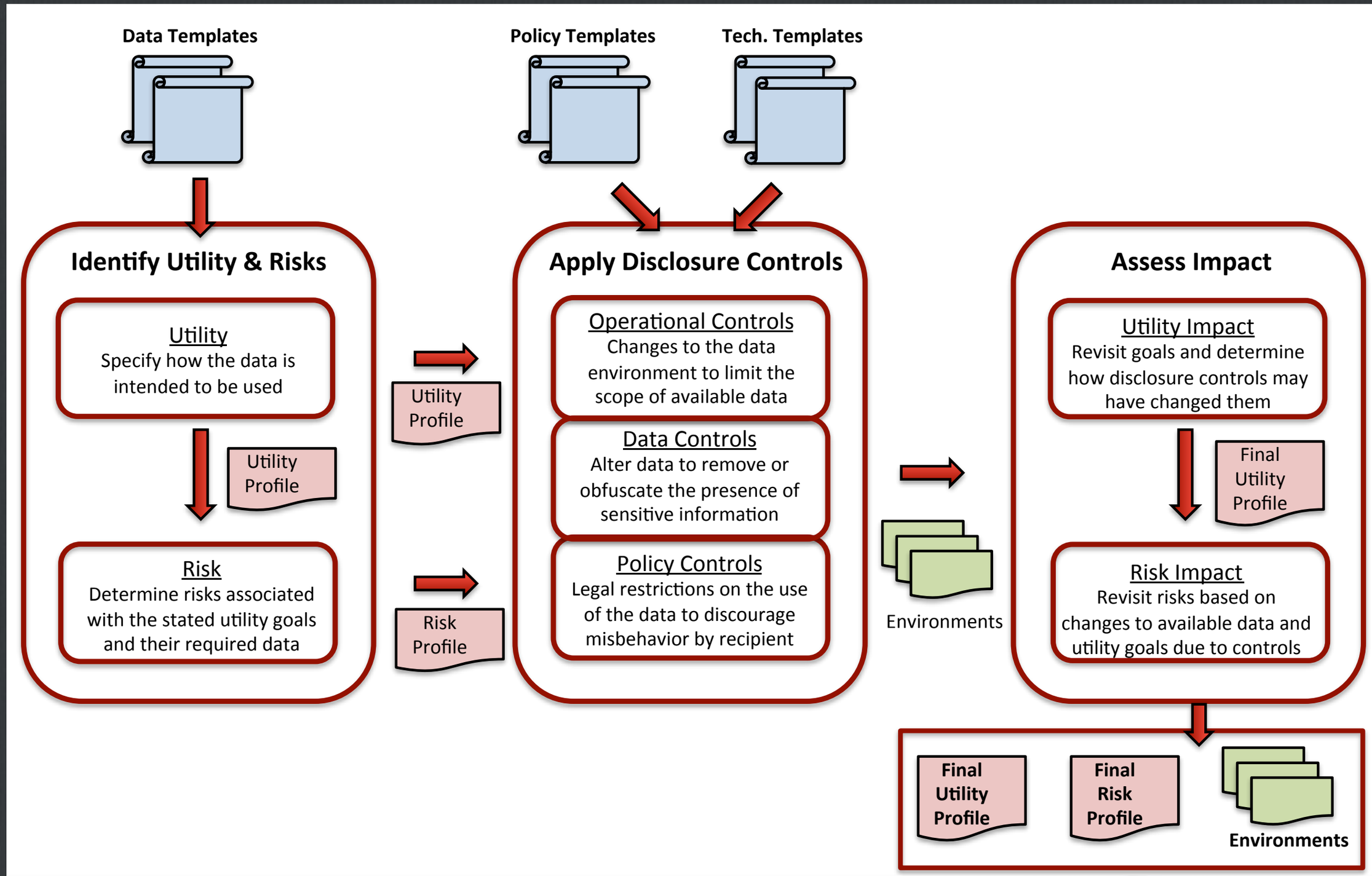
```
"name": "GPS Data",
"description": "Lat. and long. data",
"categories": ["location_data"],
"parser": "http://example.com/gps"
"schema": "
{
        "accuracy": "float",
        "latitude": "float",
        "longitude": "float",
        "timestamp": "datetime"
}"
```

e.g., Profiles

Use Case	Description	Audience	Duration	Timeliness	Detail	Functionality	Output
Signal Strength in Geographic Block	Coverage map of average signal strength by carrier and bearer channel during off- and on-peak timeframes.	Public	Indefinite	Released upon completion (~1 year).	Statistical time-series information for carriers and channel.	Indefinite	Broad geographic maps for awareness and education.

Use Case	Type of Data	Participants		Risk Factors
		Publisher	Recipient	
Signal Strength in Geographic Block	Non-identifying signal strength, cell phone carrier	FCC and contractor(s): Federal agency with the following considerations • Legal/regulatory • Contractual • Ethical	General Public: • Variety of entities. • On average, low knowledge, skills and abilities. • Low motivation and intent for harm or abuse.	-Data: Indirect identifiability viz. quasi-identifiers from carrier and bearer channel -Source: Contractual (Privacy Notice)

DCF- Phases & Components



Not a Magic Pill (dern it!)

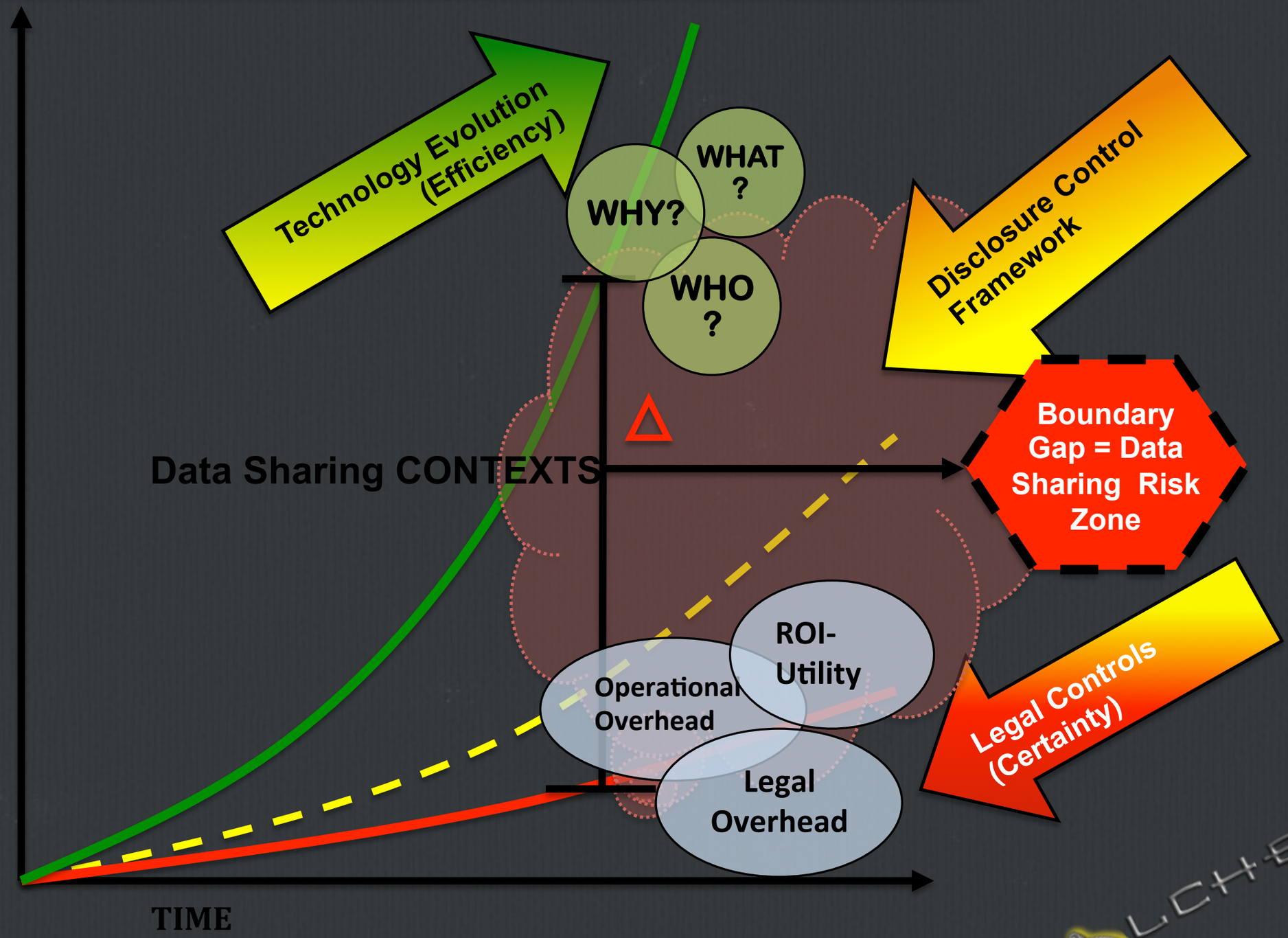
- Does not provide yes/no answers
  - Data sharing is a risk management process
  - Appetite for risk varies significantly

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

Disclosure Control → Optimizing Trust

- Framework can be a change agent for current data sharing
  - (articulate risk & benefit, vanilla response, skewed risk-benefit calculation)
- Enables data providers to trust recipients and be trusted by oversight entities
- Provides a unified methodology to enhance certainty and efficiency
- Tacitly state justifications and assumptions for choices
- Facilitates interaction between technology and policy
- Generalizable across all network data & scenarios
- Re-conceptualizes risk

How Does Disclosure Control Optimize for Trust?



Status

- **Paper Publication:**
  - IEEE Homeland Security Technology Conf (www.ssrn.com/abstract=2032315)
  - Technology Policy Research Conference
- **FCC Case Study:**
  - Help advise Mobile Broadband Measurement project
  - Privacy analysis of collected data and policy controls
- **Standardization**
  - Used utility/risk assessment methodology to choose disclosure control options
  - Developed initial drafts of profile, template, environment structures