Balancing Public Good and Personal Privacy

Emilie Corcoran  
B. Eng, Electrical

Eugene Kang  
H. BSc.; M. Information
Roadmap

Open Data is Complex
Take Responsibility
Overview of Framework
Big Data

Internet Connections

30.7B

30.7 billion devices will be connected to the internet in 2020

Existing Data

2.7Z

There are 2.7 zetabytes of data in the digital universe today

Data Creation

90%

90% of the world’s data has been created in the last two years
Vehicle Data
Vehicle Data: Hazardous Driving
Vehicle Data: Idling
Vehicle Data: Weather
Sports Data: Tennis

WOMEN'S SINGLES - ROD LAVER ARENA - QF
A. Ivanovic vs. E. Bouchard

SET 1 2 3 4 5

GAME SUMMARY: RECEIVING POINTS WON

(resume live feed)

GAME 6: A. Ivanovic is serving.

Ivanovic

Bouchard

24% 27%
Sports Data: Football
Major organizations experience an average of 20 data loss incidents per day.
Traditional Frameworks

- Rigid
- Narrowly Defined
- Outdated
- Lack of analysis
Taking Responsibility

Proactive Privacy
Responsibility for Product
Planning Ahead
Responsible Release Framework

1. Baseline
2. Analyze
3. Evaluate (Pass/Fail)
   - Fail: discard
   - Pass: Release
4. Release
5. Monitor
Baseline

1. Baseline
2. Analyze
3. Evaluate
   - Pass
   - Fail
   - discard
4. Release
5. Monitor

RSAC Conference 2018
Privacy

Less Identifying Data

Raw Data

Utility
Analyze
Analysis Steps

1. Get to Know Your Data
2. Determine Threats & Identify Vulnerabilities
3. Set Desired Level of De-Identification
4. Select & Implement De-Id Techniques
5. Re-Assess Impact on Data
De-Identification Categories

- Raw Data
  - No processing done
- Masked Data
  - Direct identifiers processed
- Obscured Data
  - Direct and indirect identifiers processed
- Aggregate Data
  - Summarized data points
De-Identification Techniques

Suppression
- Field Removal
- Record Removal
- Pseudonymization
- Truncation
- Masking

Generalization
- Aggregation
- Grouping
- Weighting

Randomization
- Noise
- Random Perturbation
- Data Swapping
- Substitution

Statistical Privacy
- k-Anonymity
- Differential Privacy Methods

De-Identification Families
Evaluate
Release

Baseline → Analyze → Evaluate → Release → Monitor

- Evaluate: Pass → Release
- Evaluate: Fail → discard

RSAC 2018
Monitor

Baseline → Analyze → Evaluate

Evaluate:
- Pass → Release
- Fail → discard

Release → Monitor
Summary

Open Data is Complex

Take Responsibility

Overview of Framework
Check Out Our Paper!

www.geotab.com/blog/big-data-privacy
Thank You