

PRIVACY HORIZONS: TERRA INCOGNITA

29th International Conference of
Data Protection and Privacy Commissioners

September 25 to 28, 2007
Montreal, Canada



LES HORIZONS DE LA PROTECTION DE LA VIE PRIVÉE : TERRA INCOGNITA

29^e Conférence internationale des commissaires
à la protection des données et de la vie privée

du 25 au 28 septembre 2007
Montréal, Canada

Measuring and Managing Re-identification Risk

by

Khaled El Emam

University of Ottawa

Managing Re-id Risk- I

- Before data is collected:
 - Scenarios
 - When preparing a protocol
 - For review by ethics boards
 - When formulating new policies and procedures
 - When writing data sharing agreements
 - Tools
 - Heuristics
 - Simulations

Managing Re-id Risk - II

- After data is collected:
 - Scenarios
 - Providing data to administrators, researchers or government departments
 - Responding to an access to information request
 - Tools
 - Masking
 - Risk-based anonymization

Heuristics, Masking, Anon

- The 20k rule, 70k rule, 100k rule
- Decision tools from matching experiments
- Around 18 tools for masking on the market
- Deciding on a risk threshold for anonymization

Acceptable Re-id Risk

- What databases does an attacker have access to for record linkage ?
- What does an attacker know beforehand ?
- What is the verification cost ?
- How do we account for privacy tradeoffs by the public ?
- What is the impact of consent model ?

Databases

- Public information and registries
- Commercial but generally available databases
- Confidential and proprietary databases

Verification Cost

- At some point the verification cost becomes too high compared to the benefit for the attacker
- The proportion of data that is population unique is important
- The extent of overall matching success is also important
- You can control both through anonymization

Tradeoffs

- The public is willing to trade their privacy for personal benefits/gains
- What they tell us is not necessarily how they will behave
- To what extent is the public willing to trade their privacy for societal gain ?

Consent Models

- Is the impact on recruitment rates and bias a function of the consent model or how it is implemented ?
- There are many factors that influence consent – were all of these controlled for when comparing consent models ?