

THE REB Wizard Tool

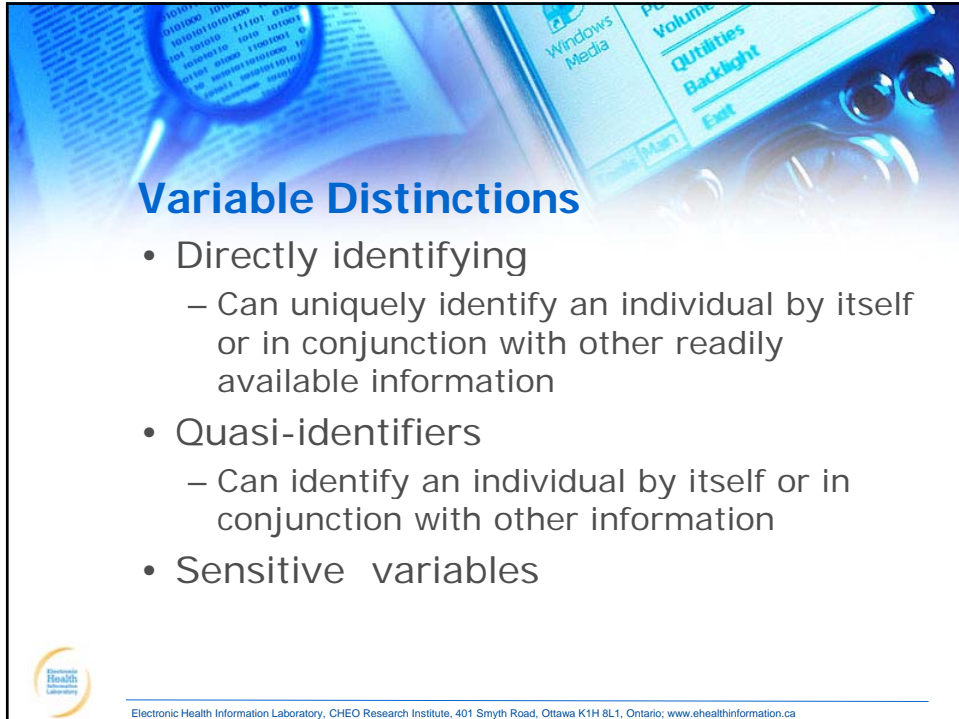
Khaled El Emam, CHEO RI & uOttawa



Context


- There are two general scenarios for de-identification:
 - Before data is collected a decision needs to be made about whether the collected data is de-identified
 - Data is available and it will be used or disclosed and needs to be de-identified beforehand
- Our focus today is on the first one

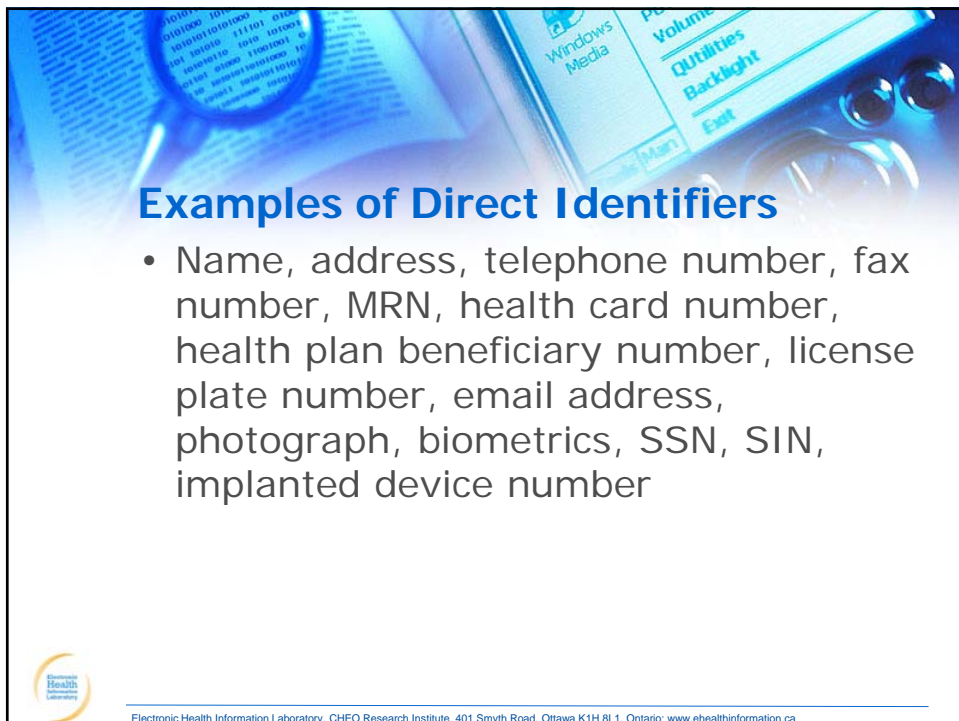




Variable Distinctions


- Directly identifying
 - Can uniquely identify an individual by itself or in conjunction with other readily available information
- Quasi-identifiers
 - Can identify an individual by itself or in conjunction with other information
- Sensitive variables

 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



Examples of Direct Identifiers

- Name, address, telephone number, fax number, MRN, health card number, health plan beneficiary number, license plate number, email address, photograph, biometrics, SSN, SIN, implanted device number

 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



Examples of Quasi-Identifiers

- sex, date of birth or age, geographic locations (such as postal codes, census geography, information about proximity to known or unique landmarks), language spoken at home, ethnic origin, aboriginal identity, total years of schooling, marital status, criminal history, total income, visible minority status, activity difficulties/reductions, profession, event dates (such as admission, discharge, procedure, death, specimen collection, visit/encounter), codes (such as diagnosis codes, procedure codes, and adverse event codes), country of birth, birth weight, and birth plurality



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca

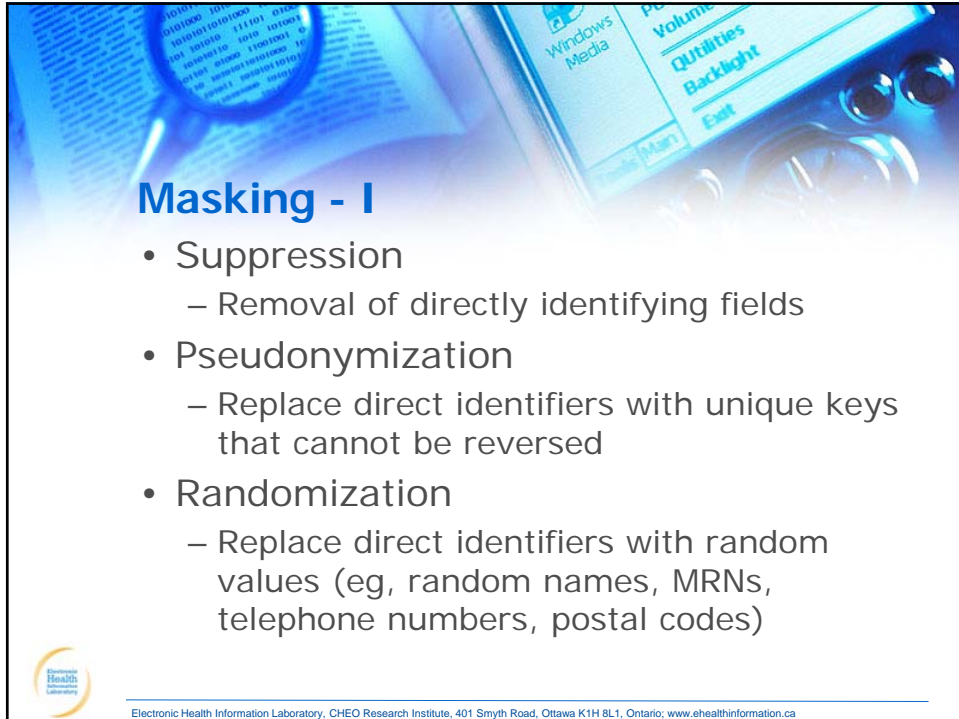


Methods

- Masking
 - Deals with the directly identifying variables
- De-identification
 - Deals with the quasi-identifiers




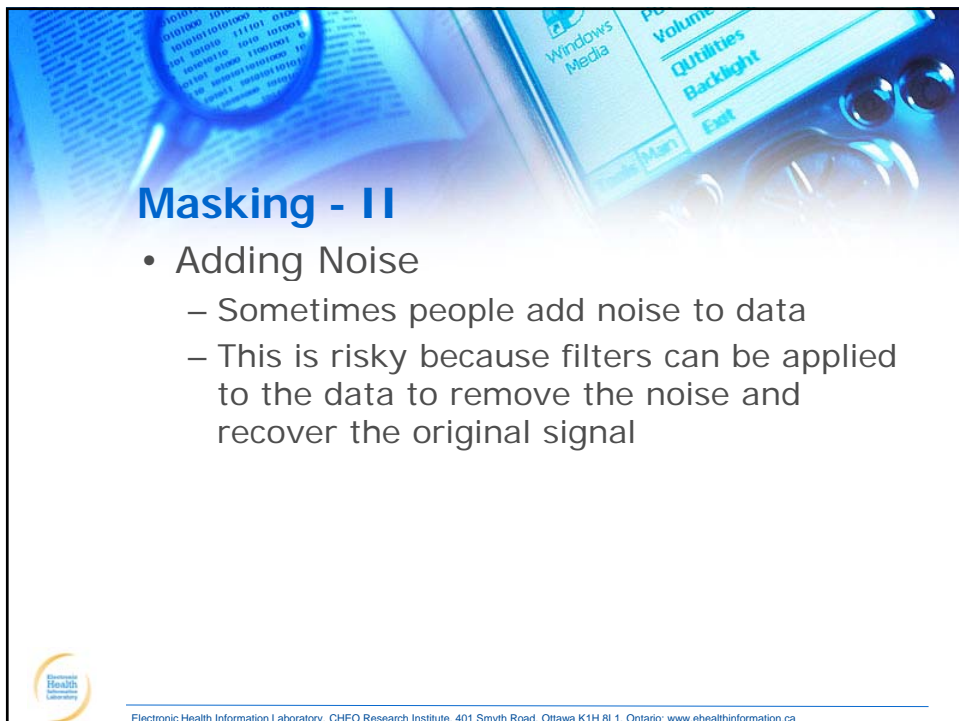
Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca



Masking - I


- Suppression
 - Removal of directly identifying fields
- Pseudonymization
 - Replace direct identifiers with unique keys that cannot be reversed
- Randomization
 - Replace direct identifiers with random values (eg, random names, MRNs, telephone numbers, postal codes)

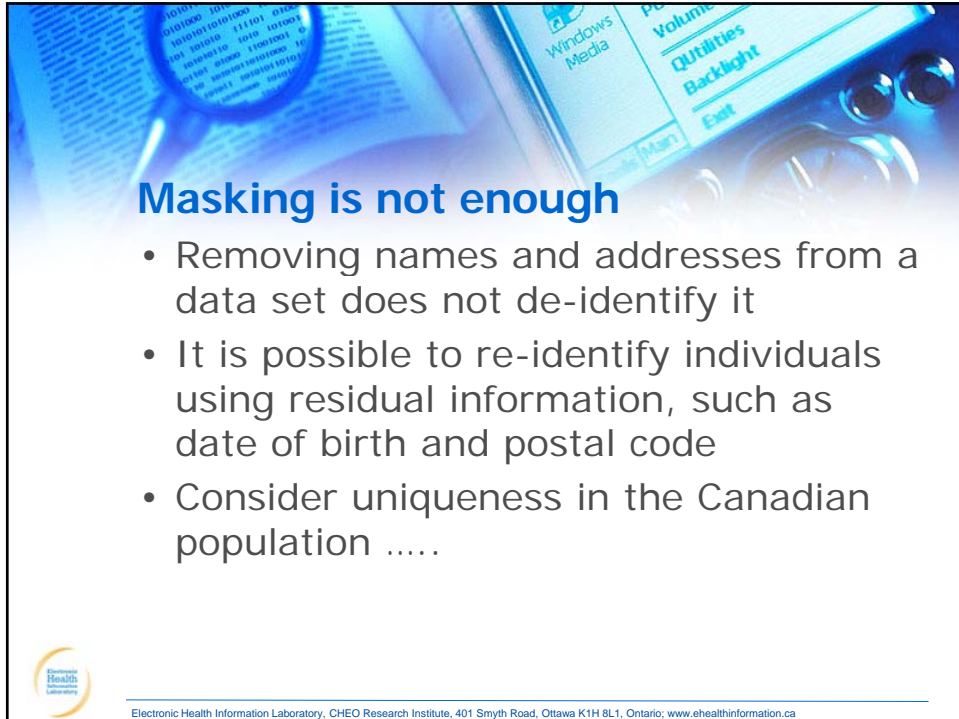
 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



Masking - II


- Adding Noise
 - Sometimes people add noise to data
 - This is risky because filters can be applied to the data to remove the noise and recover the original signal

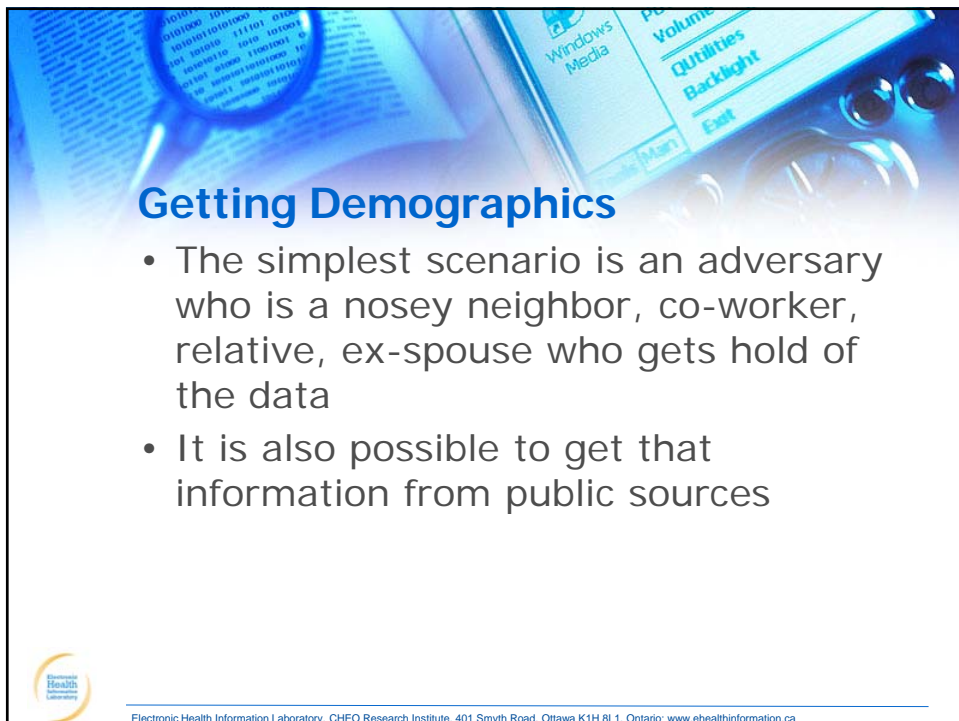
 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



Masking is not enough


- Removing names and addresses from a data set does not de-identify it
- It is possible to re-identify individuals using residual information, such as date of birth and postal code
- Consider uniqueness in the Canadian population

 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca



Getting Demographics

- The simplest scenario is an adversary who is a nosy neighbor, co-worker, relative, ex-spouse who gets hold of the data
- It is also possible to get that information from public sources

 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca



Examples of Public Sources - I

- Canadian public sources of demographics:
 - Obituaries: available from newspapers and funeral homes; there are obituary aggregator sites that make this simple
 - PPSR: Private Property Security Registration; contains information on loans secured by property (e.g., cars)
 - Land Registry: information on house ownership



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca

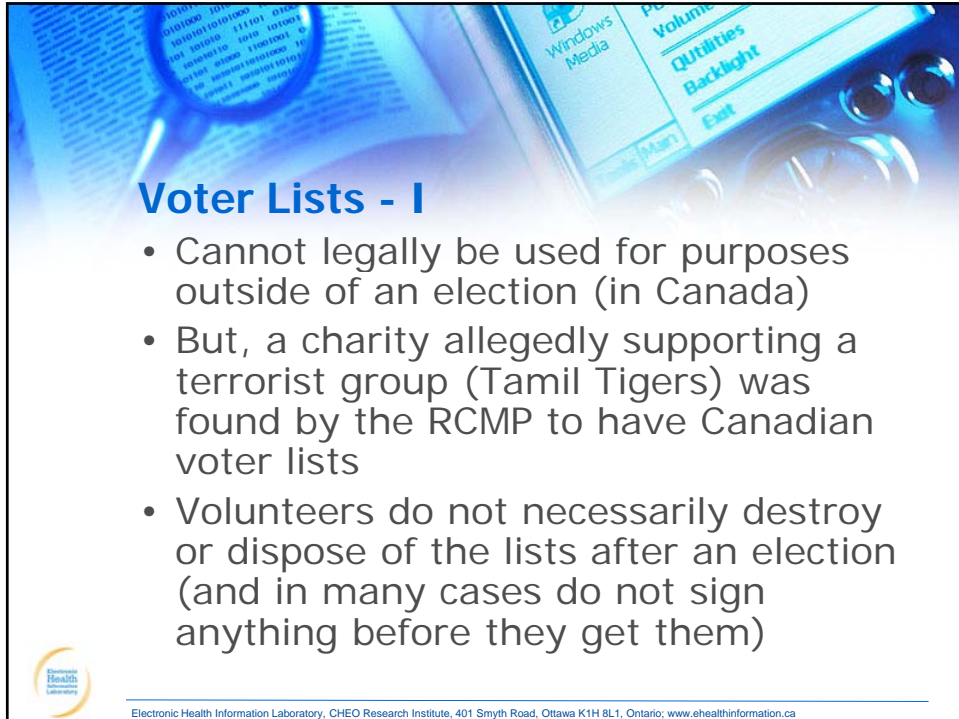


Examples of Public Sources - II

- Membership Lists: provide comprehensive listings of professionals (e.g., doctors, lawyers, civil servants)
- Salary Disclosure Reports: provided by governments for those earning higher than a certain threshold
- White Pages: public telephone directory
- Job Sites: CVs posted in public and closed job web sites
- Donations: Disclosures of donations to political parties (include address)
- Sports Rosters: Include detailed information about team members
- Facebook: Individuals, especially teenagers, post a considerable amount of information on-line




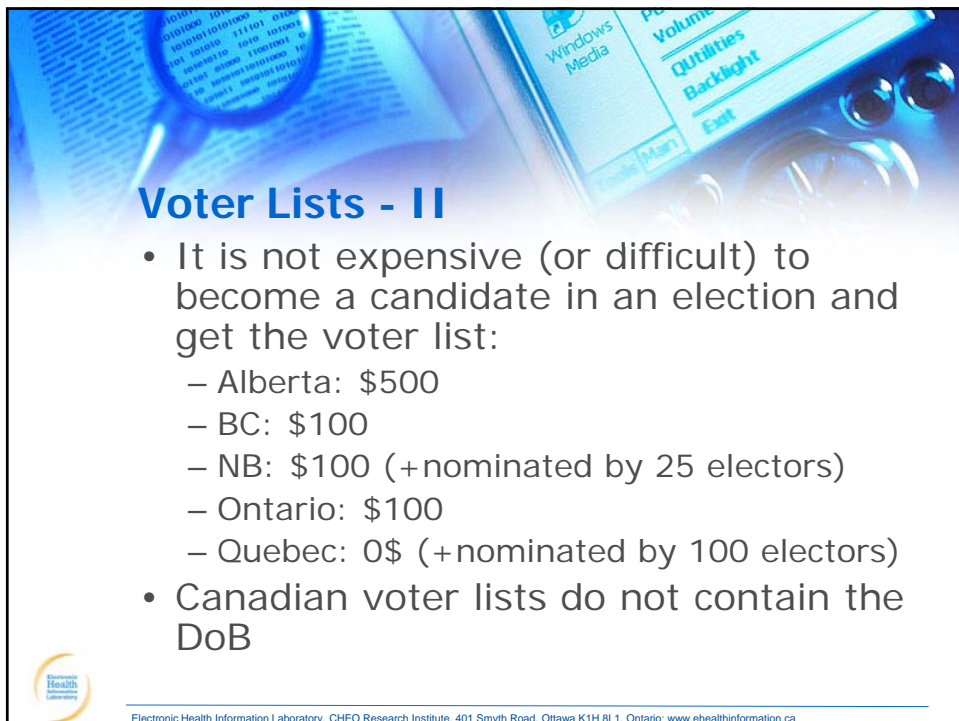
Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



Voter Lists - I


- Cannot legally be used for purposes outside of an election (in Canada)
- But, a charity allegedly supporting a terrorist group (Tamil Tigers) was found by the RCMP to have Canadian voter lists
- Volunteers do not necessarily destroy or dispose of the lists after an election (and in many cases do not sign anything before they get them)

 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca



Voter Lists - II

- It is not expensive (or difficult) to become a candidate in an election and get the voter list:
 - Alberta: \$500
 - BC: \$100
 - NB: \$100 (+nominated by 25 electors)
 - Ontario: \$100
 - Quebec: 0\$ (+nominated by 100 electors)
- Canadian voter lists do not contain the DoB

 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca

Public Registries

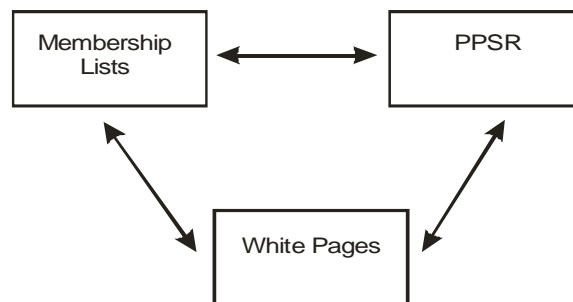
- In the following slides I will explain how to use public sources to create demographic profiles of individuals



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca

Professional Groups - I

We can construct identification databases for specific professional groups



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca



Professional Groups - II

- College of Physicians and Surgeons of Ontario
- Law Society of Upper Canada
- Professional Engineers Ontario
- College of Occupational Therapists
- College of Physical Therapists
- Public servants (eg, GEDS)
-



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



What is the success rate ?

	CPSO	LSUC
• Ability to get home postal codes (source: PPSR and telephone directory)	60%	45%
• Ability to get practice/firm postal codes (source: CPSO/LSUC)	100%	100%
• Ability to get date of birth (source: PPSR)	40%	45%
• Ability to get gender (source: CPSO/ genderizing LSUC)	100%	100%
• Ability to get initials (source: CPSO/LSUC)	100%	100%



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca

What is the success rate by gender?

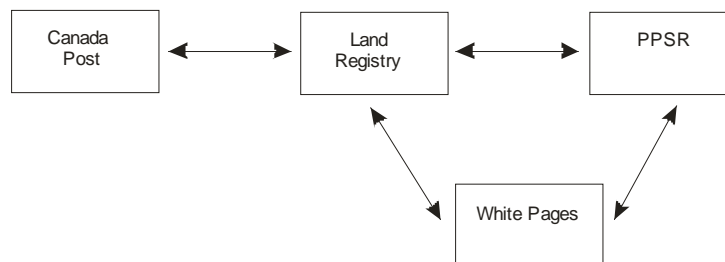
	CPSO	LSUC
MALE		
• Ability to get home postal codes (source: PPSR and telephone directory)	63%	48%
• Ability to get date of birth (source: PPSR)	45%	48%
FEMALE		
• Ability to get home postal codes (source: PPSR and telephone directory)	49%	40%
• Ability to get date of birth (source: PPSR)	29%	40%



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca

Homeowners

- We can construct identification databases for specific postal codes



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



What is the success rate ?

	Ott	To
• Ability to get initials	93%	100%
• Ability to get DoB	33%	40%
• Ability to get telephone number	80%	50%
• Ability to get gender	87%	95%



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca

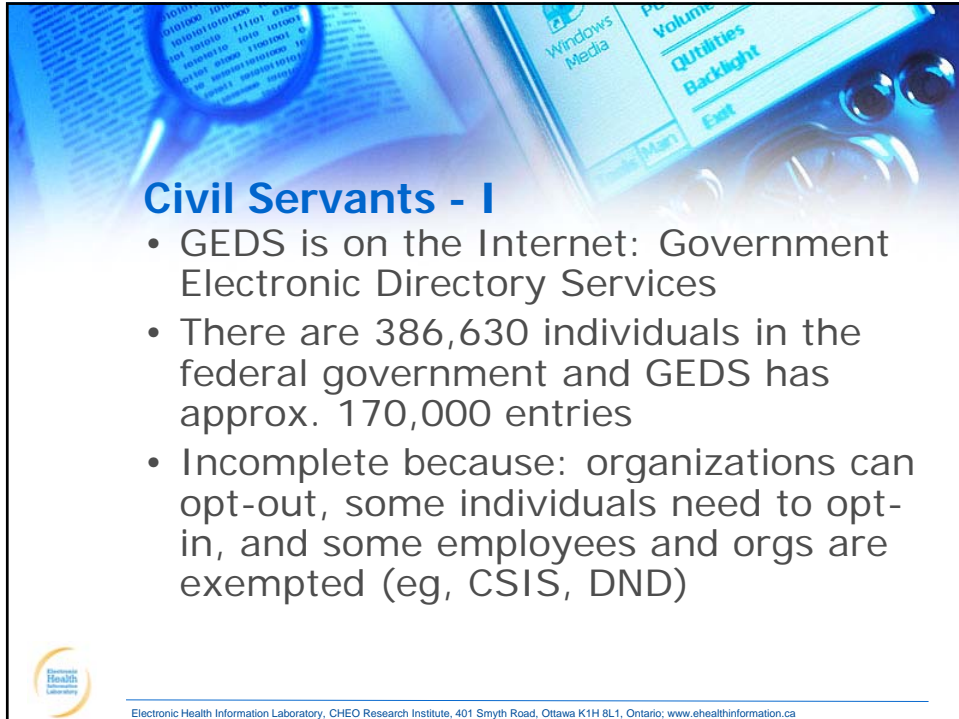


Re-id Risk for Homeowners

- The number of households per postal code is quite small (Ott: 15; To: 20)
- The individuals (homeowners) were unique on common combinations of quasi-identifiers (eg, gender and DoB)
- For these individuals re-identification risk is very high




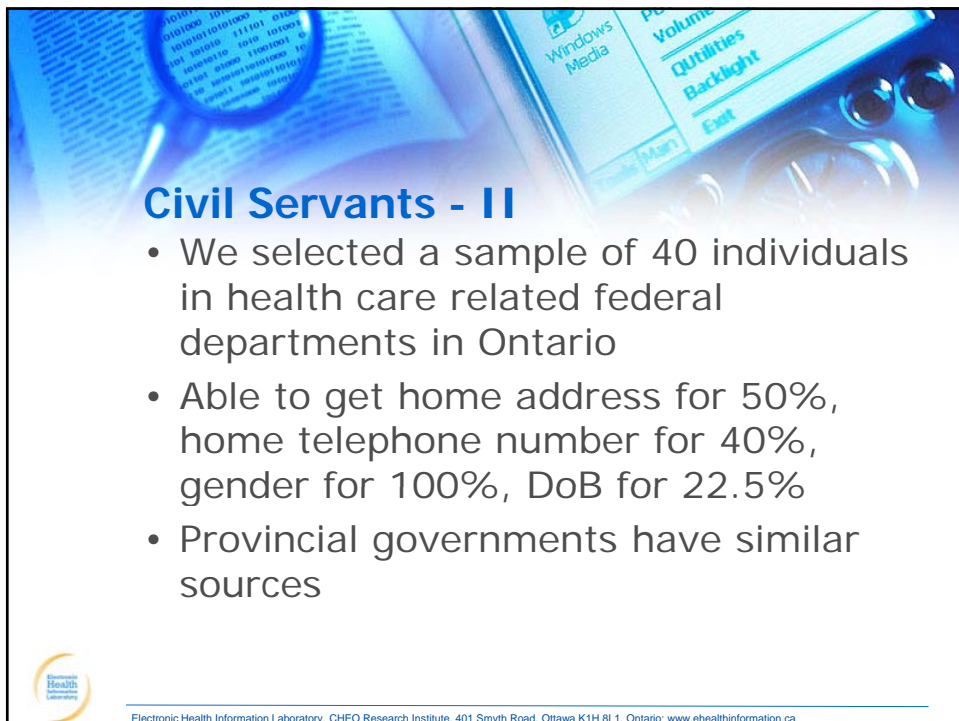
Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



Civil Servants - I


- GEDS is on the Internet: Government Electronic Directory Services
- There are 386,630 individuals in the federal government and GEDS has approx. 170,000 entries
- Incomplete because: organizations can opt-out, some individuals need to opt-in, and some employees and orgs are exempted (eg, CSIS, DND)

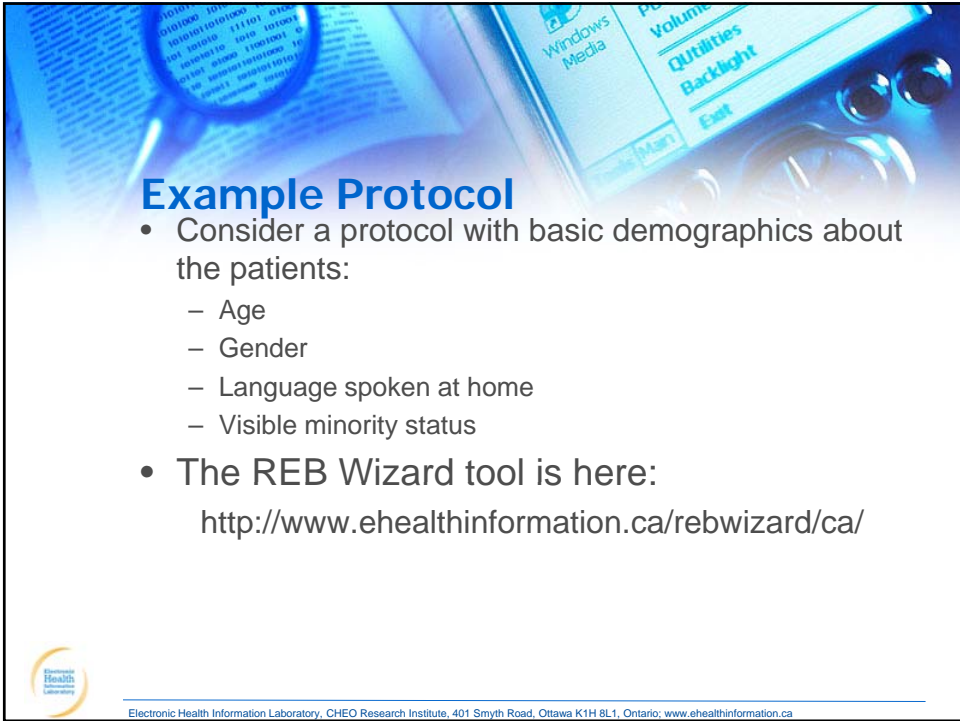
 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca



Civil Servants - II


- We selected a sample of 40 individuals in health care related federal departments in Ontario
- Able to get home address for 50%, home telephone number for 40%, gender for 100%, DoB for 22.5%
- Provincial governments have similar sources

 Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario, www.ehealthinformation.ca



Example Protocol

- Consider a protocol with basic demographics about the patients:
 - Age
 - Gender
 - Language spoken at home
 - Visible minority status
- The REB Wizard tool is here:
<http://www.ehealthinformation.ca/rebwizard/ca/>



Electronic Health Information Laboratory, CHEO Research Institute, 401 Smyth Road, Ottawa K1H 8L1, Ontario: www.ehealthinformation.ca



kelemam@uottawa.ca

www.ehealthinformation.ca

www.ehealthinformation.ca/knowledgebase

