VANDERBILT School of Medicine

More than a Matter of Trust: Security & Privacy in Voter Registration Records

Presented at the National Academies Committee on State Voter Registration Databases Workshop II

Bradley Malin, PhD (b.malin@vanderbilt.edu)
Assistant Professor of Biomedical Informatics, School of Medicine
Assistant Research Professor of Computer Science, School of Engineering
Vanderbilt University
November 30, 2007

Questions

- *What principles should guide security decisions? How might these apply to voter registration databases?*
- *What privacy considerations need to be taken into account, especially with the impact of combining and linking data?*
- *What standard, adversarial test could be applied against each state's database? What would you include in such a test?*

Question 1

- *What principles should guide security decisions? How might these apply to voter registration databases?*
- Principles of systems and data protection are dependent on who has access.
- There are multiple environments to consider
  - Private access
    - Collection and use of personal information to maintain accuracy and prevent fraud
  - Public access
    - political and related uses; or non-marketing

Public Information Sharing

- Example: Washington
  - If you are a voter, your name, address, political jurisdiction, gender, date of birth, voting record, date of registration, and registration number are public information under state law.
    (RCW 29A.08.710)
- This is public record by law and does not violate security or privacy, however …

Behind Closed Doors

- More than public information for registration of voters
- NAS Committee: Recommend how to evolve and maintain voter registration databases in a way that enables sharing of information with other states securely and **accurately** in fulfillment of HAVA
- Example: Pennsylvania
  - "Statewide Uniform Registry of Electors", county election officials have direct access to the centralized statewide database
  - The state uses "identifying number, name, and date of birth" for linking to motor vehicle and/or Social Security records
    - Use a hybrid match: the number and first two characters of the last name must match exactly, with discretion left to the county commission to determine if the rest of the record is a match
    - Currently uses the AAMVA (American Association of Motor Vehicles Administrators) criteria to match information with SSN digits: exact match of the SSN-4, first name, last name, month of birth, and year of birth.

Several Principles

- People are inherently curious.
- Account for the hacker, but don't forget about the insider

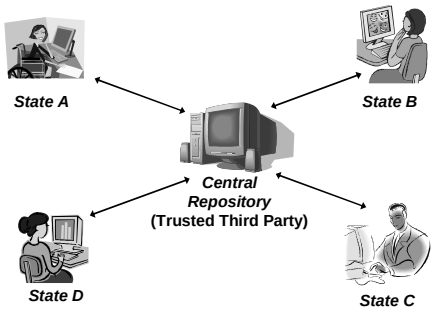
- "Easier" Goals
  - Authenticate the users
  - Access Control
  - Logs and Audits
  - Encrypt data in transmission
  - Encrypt data at rest
  - Try to minimize data on mobile devices
    - Passwords, encrypt, etc.

- "Harder" Challenges
  - Define use cases
  - Specify workflows
  - Define roles
  - Develop Standards
  - Learn or specify baselines of systems use

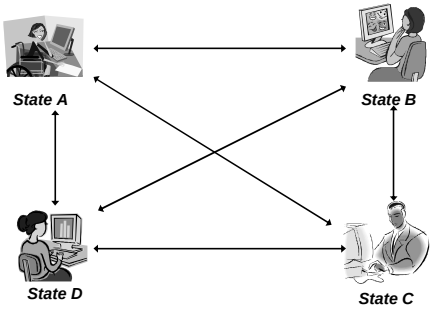
Secure Data Sharing

- NAS Committee: Recommend how to evolve and maintain voter registration databases in a way that enables **sharing** of information with other states **securely** and accurately in fulfillment of the Help America Vote Act of 2002
- Is data sharing centralized or distributed?
- What information is being shared?
  - Is it public information (names, dates of birth, etc.) or private information (Social Security Numbers)?

Private Information Sharing



Private Information Retrieval



What Data is Being Shared?

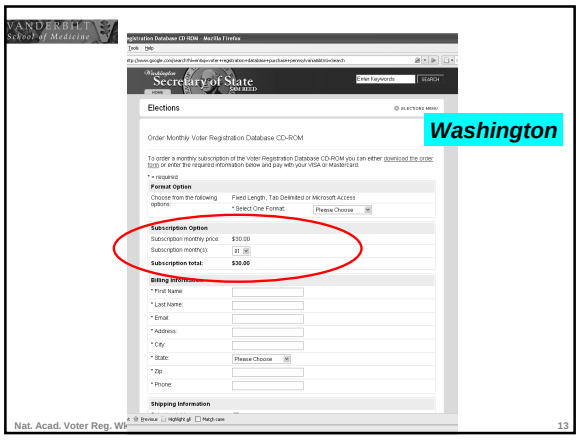
- Does voter registration, and other personal information, need to be compared in the clear?
- Computer science research has shown that comparison of encrypted records is possible.
  - Secure multiparty computation (e.g. Clifton; Wright; Nissim)
  - Need for standardized tools

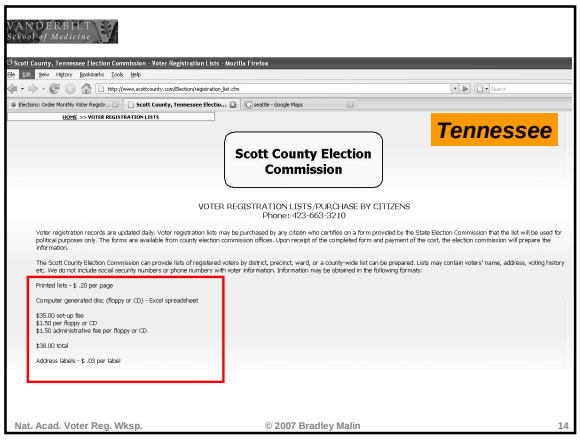
Questions

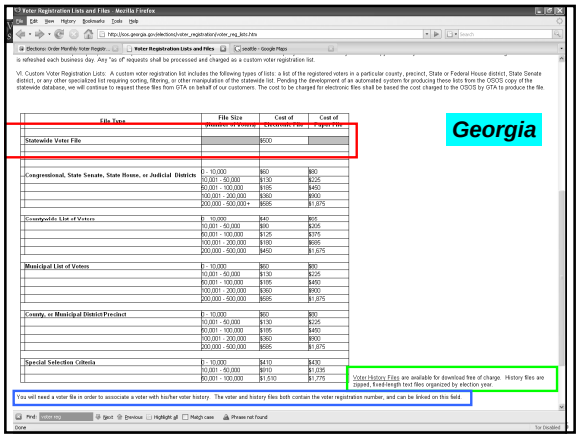
- *What privacy considerations need to be taken into account, especially with the impact of combining and linking data?*
- Concerns are dependent on the realm
  - Information disclosed to the public is different than information used by the private data holders
- Threats exist primarily because we do not consider the availability of other databases

Privacy Concerns

- Privacy violations via inference: *voter registration history*
- Privacy violations via linkage: *voter registration lists*
- Untargeted Identification (UI): Identify any individual
  - Mass UI: Identify as many individuals as possible
- Targeted Identification (TI): Identify a specific individual
  - Mass TI: Identify as many pre-specified individuals as possible
- Ability to achieve attacks is dependent, in part, on the economics of data access







Privacy Violations By Inference

Name	Date	Number of Voters
Catoosa	9/18/2007	1
Cobb	9/18/2007	1
Clayton	9/18/2007	1
Lee	11/06/2007	1
Gwinnett	9/18/2007	1
Dekalb	9/18/2007	3
Chattaho	11/06/2007	3
Sumter	11/06/2007	3
Seminole	11/06/2007	4
Charlton	11/06/2007	5
Dodge	9/18/2007	6
Mitchell	3/20/2007	7
Dawson	9/18/2007	7

Registration number ***76686***

When it is revealed how Catoosa county voted in this election (aggregate results), then we uniquely link this voter to their vote.

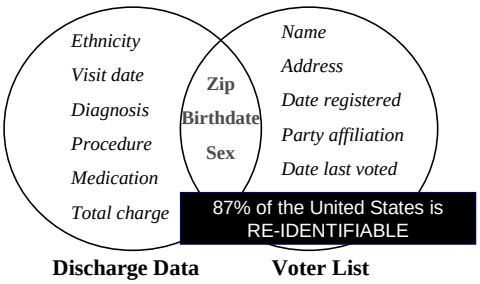
Georgia Voting 2007 History

Principle

- Census and other government agencies have developed tools for analysis and protection of contingency tables for almost half a century
- Disclosure control
  - Statistical: suppress and/or permute voter histories with small cell counts (e.g. Fienberg; Duncan; Domingo-Ferrer)
  - Computational: generalize geocodings until cell counts are above minimal threshold (e.g. Sweeney)

Privacy Violations By Linkage

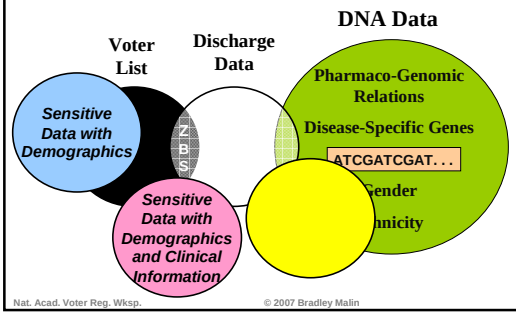
(Sweeney 1997, 1998)



87% of the United States is RE-IDENTIFIABLE

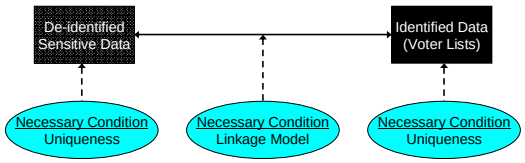
Chain of Links

(Malin & Sweeney, 2000 - 2007)

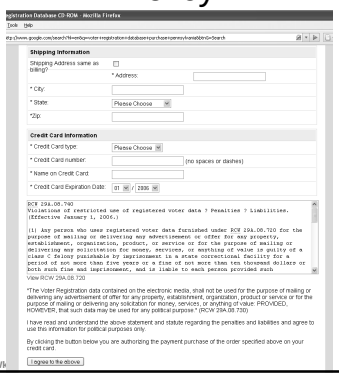


Questions

- *What standard, adversarial test could be applied against each state's database? What would you include in such a test?*
- Central Dogma of Re-identification



Policy



Policy

RCW 29A.08.740

Violations of restricted use of registered voter data ? Penalties ? Liabilities. (Effective January 1, 2006.)

(1) **Any person who uses registered voter data furnished under RCW 29A.08.720 for the purpose of mailing or delivering any advertisement or offer for any property, establishment, organization, product, or service or for the purpose of mailing or delivering any solicitation** for money, services, or anything of value is guilty of a class C felony punishable by imprisonment in a state correctional facility for a period of not more than five years or a fine of not more than ten thousand dollars or both such fine and imprisonment, and is liable to each person provided such advertisement or solicitation, without the person's consent, for the nuisance value of such person having to dispose of it, which value is herein established at five dollars for each item mailed or delivered to the person's residence. …

(2) Each person furnished data under RCW 29A.08.720 shall take reasonable precautions designed to assure that the data is not used for the purpose of mailing or delivering any advertisement or offer for any property, establishment, organization, product, or service or for the purpose of mailing or delivering any solicitation for money, services, or anything of value. However, the data may be used for any political purpose. Where failure to exercise due care in carrying out this responsibility results in the data being used for such purposes, then such person is jointly and severally liable for damages under subsection (1) of this section along with any other person liable under subsection (1) of this section for the misuse of such data.

[2005 c 246 § 19. Prior: 2003 c 111 § 249; 2003 c 53 § 176; 1999 c 298 § 2; 1992 c 7 § 32; 1974 ex.s. c 127 § 3; 1973 1st ex.s. c 111 § 4. Formerly RCW 29.04.120.]

Technology + Policy

- Must combine security and privacy risk evaluations with economic and policy deterrents
- Current regulations are insufficient: nothing to prevent the linkage of this information with other records.
- Legal protection from marketing, but does not prevent an adversary from linking a voter registration record to financial, health, and other types of information

Principles

- *What standard, adversarial test could be applied against each state's database? What would you include in such a test?*
  - Develop a threat model
  - Model what other databases in the state are available?
  - Document what type of information the records contain
  - Be aware of who has access to these records
  - Rank risks by
    - Accessibility – How many people potentially have access to the data?
    - Sensitivity – What do the available records communicate about the individuals?
    - Cost – What are the economic barriers to access and linkage?