

Laboratory for International Data Privacy at CMU

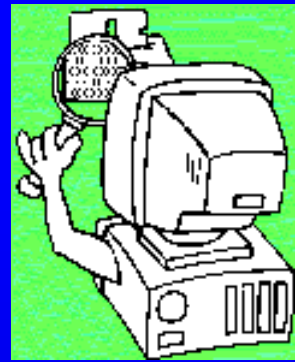
Work with real-world stakeholders:

- public health agencies
- government agencies
- private corporations

Kinds of projects currently underway:

- health data
- web data
- video surveillance data
- genetic data
- census surveys
- crime data
- grocery data, and so on...





Laboratory for International Data Privacy at CMU

Real-world anonymity problems:

- FOIA requests
- producing public-use files
- bio-terrorism surveillance
- distributed data sharing

Technical Assessments:

- re-identification experiments
- measures of privacy risks
- measuring identifiability of data
- measurements of data quality



Disease profiles

1. Anthrax
2. Brucellosis
3. Plague
4. Q Fever
5. Tularemia
6. Smallpox
7. Viral encephalitides
8. Viral hemorrhagic fevers
9. Botulinum toxins
10. Staphylococcal enterotoxin B
11. Salmonella

Short-term (1-2 days)
behavioral patterns of disease
syndromes resulting from
exposure to related agents.



Analyses based on data sources



Data sources

- hospital inpatient claims
- emergency room visits
- physician office visits
- lab reports
- prescriptions
- assisted living reports
- school absentism
- business absentism
- animal reports
- grocery purchases
- web clickstream data
- death data
- videos of public spaces
- GPS auto data

“traditional” for
bioterrorism
surveillance

extended list of
sources

For each separately:
confounders
population basis
presence & strength of signal

Scenario: Inhalation anthrax released at a sporting event

Stage I: *lasts several days or as little as 24-48 hours*

Mild fever, malaise, fatigue, myalgia, headache, congestion, dry cough, occasional sensation in chest of oppression or pressure → flu-like

Stage II: *lasts 24-48 hours and results in death*

Sever shortness of breath, cyanosis, hypotension and shock, abnormal temperature (high or low), profuse perspiration, ...

Causes of death:

Non-treatable pneumonia, progressive respiratory distress, sometimes myocardial infarction, sometimes hemorrhagic meningitis, mediastinitis, ...



Sources of person-specific data concerning scenario

1. Population registers such as birth of child

Birth of child: hospital records^L, birth certificate^I, health care cost data^{D,N}, newspaper birth notice^N, national health care cost database^N, insurance records^{I,L,D,N}, etc. Other events: local census data and city registers^{D,N}, taxes^I, religious membership^N, warranty cards, school attendance^{I,L,D,N}, etc.

2. Sporting event attendance

credit card purchases^N, parking tickets^D, video surveillance cameras^I, etc.

3. Stage I. Flu symptoms treatment

credit card purchases^N, grocery purchases^{D,N}, pharmacy purchases^{D,N}, etc.

4. Stage II. Hospital care

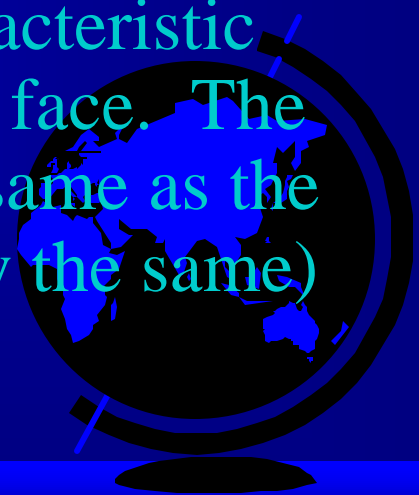
hospital records^L, physician records^L, health care cost data^{D,N}, insurance records^{I,L,D,N}, cancer registry^L, pharmacy purchases^{D,N}, medical research databases^{L,D}, etc.

5. Death records

death data bases^{I,L}, autopsy databases^{L,D,N}, newspaper death notices^N, etc.

Basic Approach in Eigenfaces

1. Use a training set to identify a set of characteristic faces.
2. Given a gallery of known faces, compare each known face to the characteristic faces to get a distance measure for each.
3. Compare an unknown face to the characteristic faces to get a distance measure for the face. The unknown face is considered to be the same as the known face having the same (or nearly the same) distance.



Questions Answered in This Session

1. Can very non-traditional data be useful to bio-terrorism surveillance?
 - web data (definitely)
 - grocery data (yes)
 - video surveillance data (in progress)

2. What about privacy, public health authority and HIPAA?
 - working with anonymous data can be done
 - public health authority is not guaranteed
 - HIPAA plays a role