

Carnegie Mellon  
DATA PRIVACY LAB

## Privacy Technology

*Computer Scientists Help Save the World*

*"provable guarantees of privacy protection while allowing information to be widely shared"*



Latanya Sweeney

privacy.cs.cmu.edu

Carnegie Mellon  
DATA PRIVACY LAB

## Abstract for this Panel

1. What new technologies, infrastructures, and interfaces are creating privacy benefits?
2. What future mechanisms will give users better control over who has access to their information and how it can be used?
3. How can privacy protection technologies be made truly usable, cost-effective, and ubiquitous?

privacy.cs.cmu.edu

Carnegie Mellon  
DATA PRIVACY LAB

## Abstract for this Panel

1. What new technologies, infrastructures, and interfaces are creating privacy benefits?
2. What future mechanisms will give users better control over who has access to their information and how it can be used?
3. How can privacy protection technologies be made truly usable, cost-effective, and ubiquitous?

How should a technology developer think about privacy in the technology he builds?

privacy.cs.cmu.edu

Carnegie Mellon  
DATA PRIVACY LAB













## Privacy Technology

1. The Privacy Problem.
2. How computer scientists help solve this problem.
3. Example: video surveillance
4. Example: bio-terrorism surveillance
5. Example: identity theft
6. Example: selective revelation
7. Example: distributed surveillance
8. Example: privacy-preserving surveillance
9. Example: DNA privacy
10. Example: Identity theft protections
11. Example: k-Anonymity
12. Example: Webcam surveillance
13. Example: Text de-identification
14. Example: Policy specification and enforcement
15. Example: Scam Spam

privacy.cs.cmu.edu

Carnegie Mellon  
DATA PRIVACY LAB

## Some People from the Lab

|  |  |
|--|--|
|  Edoardo Airoldi  |  Latanya Sweeney    |
|  Ralph Gross      |  Michael Shamos     |
|  Yiheng Li        |  William Gronim     |
|  Bradley Malin    |  Rolf Holzer        |
|  Brian Carini     |  Kishore Madhava    |
|  Samuel Edho-Eket |  Sherice Livingston |

...and more...

## Definition. Privacy

**Privacy** reflects the ability of a person, organization, government, or entity to control its own *space*, where the concept of space (or "*privacy space*") takes on different contexts

- Physical space, against invasion
- Bodily space, medical consent
- Computer space, spam
- Web browsing space, Internet privacy

Essential for survival; not just for humans but for entities.

## Ubiquitous Technologies

Seek to monitor and/or alter human spaces,

Conflicts emerge because the benefactor or controller of the deployed technology is not necessarily or exclusively the person whose privacy space is being infringed.

*Example:* Data mining; loyalty cards

*Example:* Location tracking (GPS) in car; telematics.

*Example:* Network of webcams

## Society must have Useful Information and Privacy

## The Privacy Problem

(with respect to ubiquitous technologies)

Given a human  $h$  having space  $s$  in which sensitive information about  $h$  is available, and a technology  $T$  that achieves a function  $f$ , how do we construct  $T'$  with function  $f'$  such that  $f'(s) = f(s)$  but  $T'(s)$  does not reveal the sensitive information about  $h$  that  $T(s)$  reveals.

How do we construct a privacy enhancing technology (PET) that performs the same function as the privacy invasive technology (PIT), but does so with provable privacy protection?

## Privacy Technology

1. The Privacy Problem.
2. How computer scientists help solve this problem.
3. Example: video surveillance
4. Example: bio-terrorism surveillance
5. Example: identity theft
6. Example: selective revelation
7. Example: distributed surveillance
8. Example: privacy-preserving surveillance
9. Example: DNA privacy
10. Example: Identity theft protections
11. Example: k-Anonymity
12. Example: Webcam surveillance
13. Example: Text de-identification
14. Example: Policy specification and enforcement
15. Example: Scam Spam

[privacy.cs.cmu.edu](http://privacy.cs.cmu.edu)

## Technically-Empowered Society

L. Sweeney. Information Explosion, Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies, L. Zayatz, P. Doyle, J. Theeuwes and J. Lane (eds), Urban Institute, Washington, DC, 2001.

## Typical Birth Certificate Fields, post 1925

| Field name                                       |
|--|
| Child's first name                               |
| Child's middle name (sometimes or initial)       |
| Child's last name                                |
| Day, month and year of birth                     |
| City and/or County of birth (sometimes hospital) |
| Father's name                                    |
| Mother's name (including maiden name)            |
| Place of birth (address and town/city)           |
| Mother's age and address                         |
| Mother's birthplace (town/city, state, county)   |
| Mother's occupation                              |
| Mother, number of previous children              |
| Father's age and address                         |
| Father's birthplace (town/city, state, county)   |
| Father's occupation                              |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 1-15

| Field# | Size | Field name           |
|--------|------|----------------------|
| 1      | 1    | File Status          |
| 2      | 50   | Baby's First Name    |
| 3      | 50   | Baby's Middle Name   |
| 4      | 50   | Baby's Last Name     |
| 5      | 1    | Baby's Suffix Code   |
| 6      | 3    | Baby's Suffix Text   |
| 7      | 8    | Baby's Date of Birth |
| 8      | 5    | Baby's Time of Birth |
| 9      | 1    | AM/PM Indicator      |
| 10     | 1    | Baby's Sex           |
| 11     | 3    | Blood Type           |
| 12     | 1    | Born Here?           |
| 13     | 40   | Place of Birth       |
| 14     | 1    | Facility Type        |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 16-30

| Field# | Size | Field name                      |
|--------|------|---------------------------------|
| 16     | 20   | County of Birth                 |
| 17     | 6    | Certifier's Code                |
| 18     | 30   | Certifier's Name                |
| 19     | 1    | Certifier's Title               |
| 20     | 30   | Attendant's Name                |
| 21     | 1    | Attendant's Title               |
| 22     | 23   | Attendant's Address             |
| 23     | 19   | Attendant's City                |
| 24     | 2    | Attendant's State               |
| 25     | 10   | Attendant's Zip Code            |
| 26     | 50   | Mother's First Name             |
| 27     | 50   | Mother's Middle Name            |
| 28     | 50   | Mother's Last Name              |
| 29     | 9    | Mother's Social Security Number |
| 30     | 8    | Mother's Date of Birth          |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 31-45

| field# | Size | Field name                        |
|--------|------|-----------------------------------|
| 31     | 3    | Mother's State of Birth           |
| 32     | 7    | Mother's Residence Address        |
| 33     | 2    | Mother's Residence Direction      |
| 34     | 20   | Residence Street Address          |
| 35     | 10   | Residence Type                    |
| 36     | 2    | Residence Extension               |
| 37     | 10   | Residence Apartment #             |
| 38     | 20   | Mother's Town of Residence        |
| 39     | 1    | Mother's Residence in City Limits |
| 40     | 14   | Mother's County of Residence      |
| 41     | 3    | Mother's State of Residence       |
| 42     | 10   | Mother's Residence Zip Code       |
| 43     | 38   | Mother's Mailing Address          |
| 44     | 19   | Mother's Mailing City             |
| 45     | 2    | Mother's Mailing State            |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 46-60

| Field# | Size | Field name                      |
|--------|------|---------------------------------|
| 46     | 10   | Mother's Mailing Zip Code       |
| 47     | 1    | Mother Married?                 |
| 48     | 50   | Father's First Name             |
| 49     | 50   | Father's Middle Name            |
| 50     | 50   | Father's Last Name              |
| 51     | 1    | Father's Suffix Code            |
| 52     | 9    | Father's Suffix Text            |
| 53     | 9    | Father's Social Security Number |
| 54     | 8    | Father's Date of Birth          |
| 55     | 3    | Father's State of Birth         |
| 56     | 14   | Mother's Origin                 |
| 57     | 14   | Mother's Race                   |
| 58     | 2    | Mother's Elementary Education   |
| 59     | 2    | Mother's College Education      |
| 60     | 11   | Mother's Occupation             |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 61-75

| Field# | Size | Field name                    |
|--------|------|-------------------------------|
| 61     | 11   | Mother's Industry             |
| 62     | 14   | Father's Origin               |
| 63     | 14   | Father's Race                 |
| 64     | 2    | Father's Elementary Education |
| 65     | 2    | Father's College Education    |
| 66     | 11   | Father's Occupation           |
| 67     | 11   | Father's Industry             |
| 68     | 1    | Plurality                     |
| 69     | 1    | Birth Order                   |
| 70     | 2    | Live Births Still Living      |
| 71     | 2    | Live Births Now Dead          |
| 72     | 4    | Month/Year Last Live Birth    |
| 73     | 2    | Number of Terminations        |
| 74     | 4    | Month/Year Last Termination   |
| 75     | 1    | Baby's Weight Unit            |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 76-90

| Field# | Size | Field name                       |
|--------|------|----------------------------------|
| 76     | 5    | Baby's Weight                    |
| 77     | 6    | Date of Last Normal Menses       |
| 78     | 1    | Month Prenatal Care Began        |
| 79     | 2    | Total Number of Visits           |
| 80     | 2    | Apgar Score - 1 Minute           |
| 81     | 2    | Apgar Score - 5 Minute           |
| 82     | 2    | Estimate of Gestation            |
| 83     | 6    | Date of Blood Test               |
| 84     | 22   | Laboratory                       |
| 85     | 1    | Mother Transferred In            |
| 86     | 30   | Facility Mother Transferred From |
| 87     | 1    | Baby Transferred Out             |
| 88     | 30   | Facility Baby Transferred To     |
| 89     | 1    | Tobacco Use During Pregnancy     |
| 90     | 3    | Number of Cigarettes/Day         |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 91-105

| Field# | Size | Field name                   |
|--------|------|------------------------------|
| 91     | 1    | Alcohol Use During Pregnancy |
| 92     | 3    | Number of Drinks/Week        |
| 93     | 3    | Mother's Weight Gain         |
| 94     | 1    | Release Info For SSN         |
| 95     | 6    | Operator Code                |
| 96     | 12   | Hospital ID                  |
| 97     | 1    | Sent to Romans               |
| 98     | 1    | Sent to APORS                |
| 99     | 16   | Other Certifier Specify      |
| 100    | 12   | Temporary Audit Number       |
| 101    | 16   | Other Facility Specify       |
| 102    | 16   | Other Attendant Specify      |
| 103    | 1    | Mother's Race                |
| 104    | 1    | Father's Race                |
| 105    | 2    | Mother's Origin              |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 106-120

| Field# | Size | Field name                   |
|--------|------|------------------------------|
| 106    | 2    | Father's Origin              |
| 107    | 1    | Attendant Same YN            |
| 108    | 1    | Mailing Address Same YN      |
| 109    | 1    | Capture Father's Info YN     |
| 110    | 2    | Mother's Age                 |
| 111    | 2    | Father's Age                 |
| 112    | 12   | Baby's Hospital Med. Rec.    |
| 113    | 1    | High Risk Pregnancy YN       |
| 114    | 1    | Care Giver (For Chicago)     |
| 115    | 1    | Record Selected For Download |
| 116    | 1    | Downloaded                   |
| 117    | 1    | Printed                      |
| 118    | 12   | Form Number                  |
|        |      | <b>MEDICAL RISK FACTORS</b>  |
| 119    | 1    | Anemia                       |
| 120    | 1    | Cardiac Disease              |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 121-135

| Field# | Size | Field name                     |
|--------|------|--------------------------------|
| 121    | 1    | Acute/Chronic Lung Disease     |
| 122    | 1    | Diabetes                       |
| 123    | 1    | Genital Herpes                 |
| 124    | 1    | Hydramnios/Oligohydramnios     |
| 125    | 1    | Hemoglobinopathy               |
| 126    | 1    | Hypertension, Chronic          |
| 127    | 1    | Hypertension, Preg. Assoc.     |
| 128    | 1    | Eclampsia                      |
| 129    | 1    | Incompetent Cervix             |
| 130    | 1    | Previous Infant 4000+ Grams    |
| 131    | 1    | Previous Preterm or SGA Infant |
| 132    | 1    | Renal Disease                  |
| 133    | 1    | Rh Sensitization               |
| 134    | 1    | Uterine Bleeding               |
| 135    | 1    | No Medical Risk Factors        |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 136-150

| Field# | Size | Field name                            |
|--------|------|---------------------------------------|
| 136    | 40   | Other Medical Risk Factors            |
|        |      | <b>OBSTETRIC PROCEDURES</b>           |
| 137    | 1    | Amniocentesis                         |
| 138    | 1    | Electronic Fetal Monitoring           |
| 139    | 1    | Induction of Labor                    |
| 140    | 1    | Stimulation of Labor                  |
| 141    | 1    | Tocolysis                             |
| 142    | 1    | Ultrasound                            |
| 143    | 1    | No Obstetric Procedures               |
| 144    | 40   | Other Obstetric Procedures            |
|        |      | <b>COMPLICATIONS OF LABOR &amp; I</b> |
| 145    | 1    | Febrile (>100 or 38C)                 |
| 146    | 1    | Meconium Moderate, Heavy              |
| 147    | 1    | Premature Rupture (>12 Hrs)           |
| 148    | 1    | Abruptio Placenta                     |
| 149    | 1    | Placenta Previa                       |
| 150    | 1    | Other Excessive Bleeding              |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 151-165

| Field# | Size | Field name                       |
|--------|------|----------------------------------|
| 151    | 1    | Seizures During Labor            |
| 152    | 1    | Precipitous Labor (<3 Hrs)       |
| 153    | 1    | Prolonged Labor (>20 Hrs)        |
| 154    | 1    | Dysfunctional Labor              |
| 155    | 1    | Breech/Malpresentation           |
| 156    | 1    | Cephalopelvic Disproportion      |
| 157    | 1    | Cord Prolapse                    |
| 158    | 1    | Anesthetic Complications         |
| 159    | 1    | Fetal Distress                   |
| 160    | 1    | No Complications of L&D          |
| 161    | 40   | Other Complications of L&D       |
|        |      | <b>METHOD OF DELIVERY</b>        |
| 162    | 1    | Vaginal                          |
| 163    | 1    | Vaginal After Previous C-Section |
| 164    | 1    | Primary C-Section                |
| 165    | 1    | Repeat C-Section                 |

Typical Electronic Birth Certificate Fields  
in 1999 -starting fields 166-180

| Field# | Size | Field name                           |
|--------|------|--------------------------------------|
| 166    | 1    | Forceps                              |
| 167    | 1    | Vacuum                               |
|        |      | <b>ABNORMAL CONDITIONS OF NEWBO</b>  |
| 168    | 1    | Anemia                               |
| 169    | 1    | Birth Injury                         |
| 170    | 1    | Fetal Alcohol Syndrome               |
| 171    | 1    | Hyaline Membrane Disease/RDS         |
| 172    | 1    | Meconium Aspiration Syndrome         |
| 173    | 1    | Assisted Ventilation <30             |
| 174    | 1    | Assisted Ventilation >30             |
| 175    | 1    | Seizures                             |
| 176    | 1    | No Abnormal Conditions of Newborn    |
| 177    | 40   | Other Abnormal Condition of Newborn  |
|        |      | <b>CONGENITAL ANOMALIES OF CHILD</b> |
| 178    | 1    | Anencephalus                         |
| 179    | 1    | Spina Bifida/Meningocele             |
| 180    | 1    | Hydrocephalus                        |

Typical Electronic Birth Certificate Fields in 1999 -starting fields 181-195

| Field# | Size | Field name                         |
|--------|------|------------------------------------|
| 181    | 1    | Microcephalus                      |
| 182    | 40   | Other CNS Anomalies                |
| 183    | 1    | Heart Malformations                |
| 184    | 40   | Other Circ./Resp. Anomalies        |
| 185    | 1    | Rectal Atresia/Stenosis            |
| 186    | 1    | Tracheo-Esophageal Fistula/Esophag |
| 187    | 1    | Omphalocele/Gastroschisis          |
| 188    | 40   | Other Gastrointestinal Ano.        |
| 189    | 1    | Malformed Genitalia                |
| 190    | 1    | Renal Agenesis                     |
| 191    | 40   | Other Urogenital Anomalies         |
| 192    | 1    | Cleft Lip/Palate                   |
| 193    | 1    | Polydactyly/Syndactyly/Adactyly    |
| 194    | 1    | Club Foot                          |
| 195    | 1    | Diaphragmatic Hernia               |

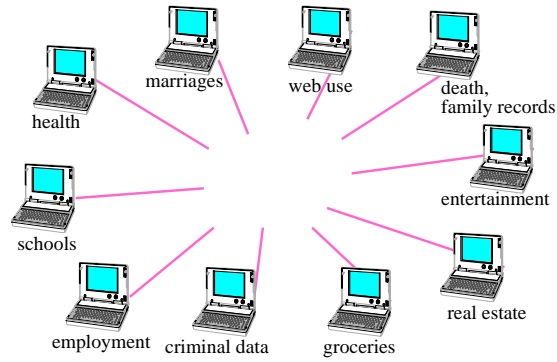
Typical Electronic Birth Certificate Fields in 1999 -starting fields 196-210

| Field#            | Size | Field name                           |
|-------------------|------|--------------------------------------|
| 196               | 40   | Other Musculoskeletal/Integumental A |
| 197               | 1    | Down's Syndrome                      |
| 198               | 40   | Other Chromosomal Anomalies          |
| 199               | 1    | No Congenital Anomalies              |
| 200               | 40   | Other Congenital Anomalies           |
| <b>CODE STRIP</b> |      |                                      |
| 201               | 1    | Record Complete YN                   |
| 202               | 1    | Record Type                          |
| 203               | 4    | Facility ID                          |
| 204               | 4    | City of Birth                        |
| 205               | 3    | County of Birth                      |
| 206               | 2    | Mother's State of Birth              |
| 207               | 2    | Mother's State of Residence          |
| 208               | 4    | Mother's Town of Residence           |
| 209               | 3    | Mother's County of Residence         |
| 210               | 2    | Father's State of Birth              |

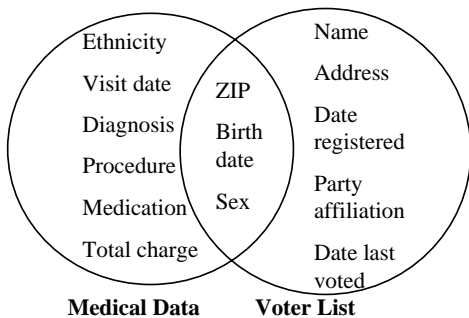
Typical Electronic Birth Certificate Fields in 1999 -starting fields 211-226.

| Field# | Size | Field name                  |
|--------|------|-----------------------------|
| 211    | 14   | Certifier's License Number  |
| 212    | 6    | Laboratory ID Number        |
| 213    | 4    | Mother Xfer Code            |
| 214    | 3    | Mother Xfer County Code     |
| 215    | 4    | Baby Xfer Code              |
| 216    | 3    | Baby Xfer County Code       |
| 217    | 4    | Year of Birth               |
| 218    | 7    | Certificate #               |
| 219    | 1    | Unique Code                 |
| 220    | 8    | File Date                   |
| 221    | 2    | Community Area              |
| 222    | 4    | Census Tract                |
| 223    | 2    | Century of Last Live Birth  |
| 224    | 2    | Century of Last Termination |
| 225    | 2    | Century of Last Menses      |
| 226    | 2    | Century of Blood Test       |

Numerous Efforts Underway to Fuse Available Data Together on Individuals

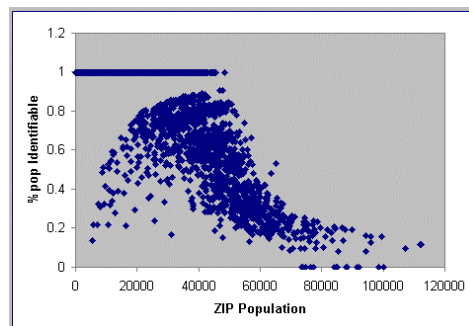


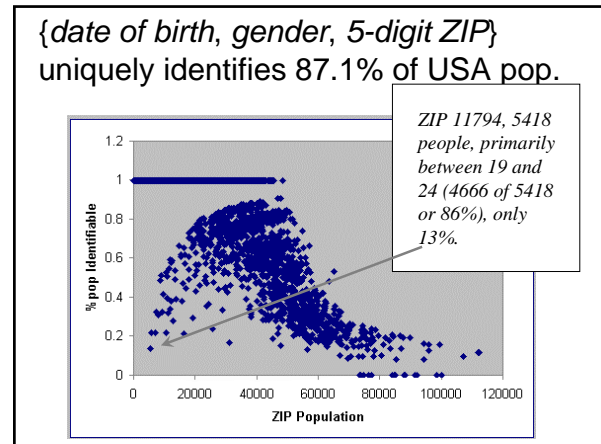
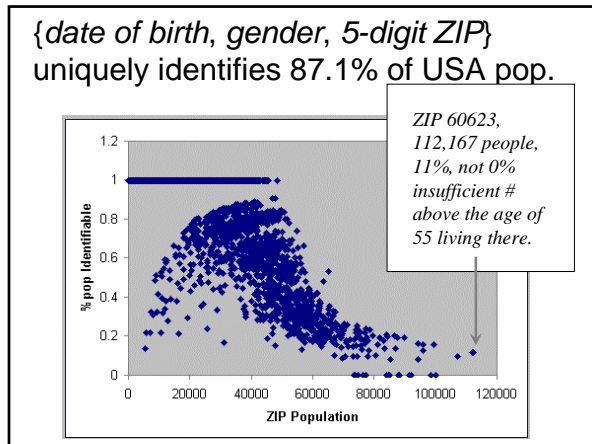
Linking to re-identify data



L. Sweeney. Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine and Ethics*. 1997; 25:98-110.

{date of birth, gender, 5-digit ZIP} uniquely identifies 87.1% of USA pop.





**Computer Scientists Must Help Save the World**

**Policy** is limited by that which can be expressed in words, tends to be crude descriptions in absence of technical refinement.

Traditional research based on **statistical and economic** approaches to policy problems tend to be retrospective and descriptive and assume technology is relatively static.

**IS/IT** can provide glue technology, but heavily relies on existing technology.

Laws can change and **lawyers** often lack understanding of ways technology will continue to change. Rivest: technology changes in months, laws change in years

**Computer scientists** construct tomorrow's machines, and can do so with privacy as part of their problem definition, so that new technology can be deployed and easily adopted.

**The Privacy Problem**  
(with respect to ubiquitous technologies)

Given a human  $h$  having space  $s$  in which **sensitive information** about  $h$  is available, and a technology  $T$  that achieves a function  $f$ , how do we construct  $T'$  with function  $f'$  such that  $f'(s) = f(s)$  but  $T'(s)$  does not reveal the **sensitive information** about  $h$  that  $T(s)$  reveals.

If the sensitive information varies based on the human deciding who gets what or additionally varies from person-to-person  
→ *policy specification and enforcement.*

**The Privacy Problem**  
(with respect to ubiquitous technologies)

Given a human  $h$  having space  $s$  in which sensitive information about  $h$  is available, and a technology  $T$  that achieves a function  $f$ , how do we construct  $T'$  with function  $f'$  such that  $f'(s) = f(s)$  but  $T'(s)$  **does not reveal the sensitive information about  $h$**  that  $T(s)$  reveals.

If there exists a uniform notion of the sensitive information, which is constant from person-to-person → *anonymization technologies.*

**Privacy Enhancing Technology**  
has 2 Basic Areas in 2 Settings

| Data Anonymity<br>( <i>anonymity</i> )                    | Policy Specification & Enforcement<br>( <i>rights mgt</i> )           | Database Security<br>( <i>security</i> )                           | Distributed & Ubiquitous Environments<br>( <i>distributed</i> ) |
|---|---|--|---|
| Methods for detecting and controlling inferences in data. | Methods for language design with automated detection and enforcement. | Methods for controlling access to and protecting database content. | Methods related to having a network of data sources.            |

## 6-Prong Approach to Developing and Assessing a Privacy Solution's Appropriateness for an Emerging Technology

1. Examine potential **benefits** of the technology.
2. Consider **privacy concerns** voiced about the technology.
3. Understand **the setting** (laws, regulations, practices, expectations) in which the technology will reside.
4. Conduct a systematic analysis of some subset of #1, #2 and #3 → **enhanced problem statement**.
5. Synthesize a proposed **solution**.
6. For a proposed solution, Prove **compliance** to #2 and #3 ("protected") and prove **warranty** to #1 ("useful").

Carnegie Mellon

## DATA PRIVACY LAB

# Privacy Technology


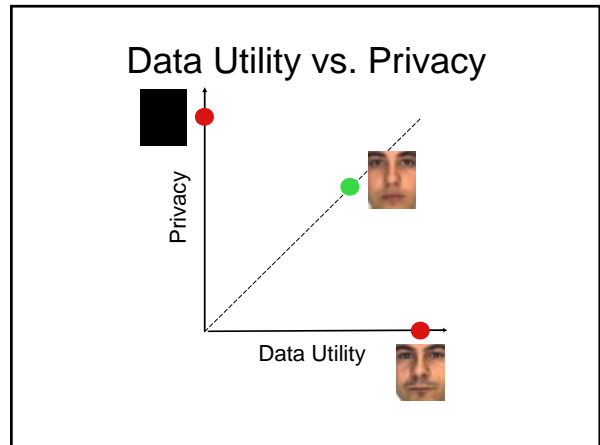
1. The Privacy Problem.
2. How computer scientists help solve this problem.
3. Example: video surveillance
4. Example: bio-terrorism surveillance
5. Example: identity theft
6. Example: selective revelation
7. Example: distributed surveillance
8. Example: privacy-preserving surveillance
9. Example: DNA privacy
10. Example: Identity theft protections
11. Example: k-Anonymity
12. Example: Webcam surveillance
13. Example: Text de-identification
14. Example: Policy specification and enforcement
15. Example: Scam Spam

privacy.cs.cmu.edu


Probable Cause "catch 22:" something useful MAY be on a video recording related to a crime, but without viewing the video cannot get a search warrant to access the video.

Want to retain privacy protections afforded by US Constitution and the need for search warrant yet enable more sharing of video.


Can we share video with law-enforcement such that no matter how good face recognition software might become, people cannot be re-identified without due process?

## Controlling Face Recognition




De-identify faces such that no face recognition software can be successful, even if face recognition software is perfect.




Newton, Sweeney, and Malin. Preserving Privacy by De-identifying Facial Images. *IEEE Transactions on Knowledge and Data Engineering*, Feb 2005.

## Ad Hoc Schemes that Don't Work I




Single Bar Mask




T-Mask

Mouth Only


- ✓ Grayscale
- ✓ Black & White



Ordinal




Threshold



In each of these, face recognition software can be trained on the method and the result yields successful re-identifications!

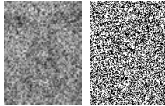
### Ad Hoc Schemes that Don't Work II



Pixelation

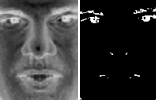
Random

- ✓ Grayscale
- ✓ Black & White




Negative

- ✓ Grayscale
- ✓ Black & White



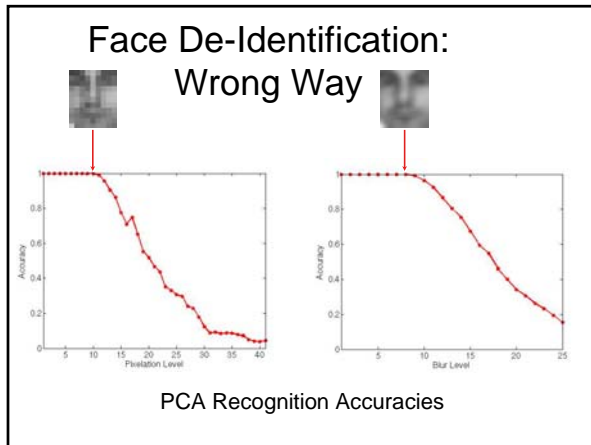
In each of these, face recognition software can be trained on the method and the result yields successful re-identifications!



Mr. Potato Head


### Pixelation – recognition improved!



Carnegie Mellon  
**DATA PRIVACY LAB**


### k-Same anonymizing faces



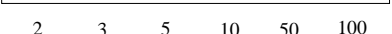
Ralph Gross  
Elaine Newton  
Latanya Sweeney  
Bradley Malin

*Thwarts face recognition while many facial details remain!*

-Pixel

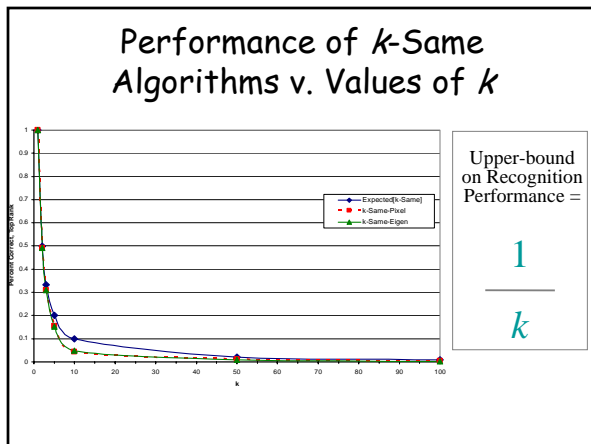


-Eigen



$k =$

2 3 5 10 50 100




### k-Same Works even as Face Recognition Software Improves

Theorem. There cannot exist any face recognition software for which a subject's k-Samed image can be correctly recognized better than  $1/k$  probability.

Note. Theorem above loosely specified.  $H$  is the de-identified face set,  $|H| \geq k$  and  $k > 1$ .




### Face Tracking



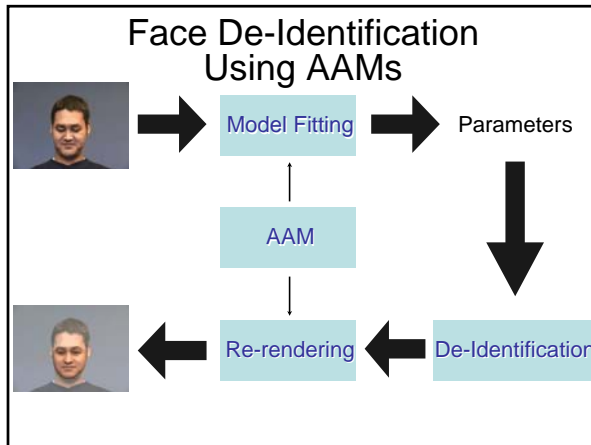
- Model fit at 230 frames per second
- Accurately captures non-rigid facial motions

[I. Matthews and S. Baker, Active Appearance Models Revisited, IJCV, 60(2), 2004]

### Fitting Models With Occlusion



[R. Gross, I. Matthews and S. Baker, Constructing and Fitting Active Appearance Models with Occlusion, IEEE Workshop on Face Processing in Video, 2004]



Carnegie Mellon  
**DATA PRIVACY LAB**

## Lessons Learned

1. Ad hoc techniques don't work.
2. Must provide a "warranty" - proof that the result is still useful.
3. Must provide a "compliance statement" - proof that a privacy constraint is satisfied.

<http://privacy.cs.cmu.edu/dataprivacy/projects/>

Carnegie Mellon  
**DATA PRIVACY LAB**

## Privacy Technology

1. The Privacy Problem.
2. How computer scientists help solve this problem.
3. Example: video surveillance
4. Example: bio-terrorism surveillance
5. Example: identity theft
6. Example: selective revelation
7. Example: distributed surveillance
8. Example: privacy-preserving surveillance
9. Example: DNA privacy
10. Example: Identity theft protections
11. Example: k-Anonymity
12. Example: Webcam surveillance
13. Example: Text de-identification
14. Example: Policy specification and enforcement
15. Example: Scam Spam

[privacy.cs.cmu.edu](http://privacy.cs.cmu.edu)

Carnegie Mellon  
**DATA PRIVACY LAB**

## Identity Angel

Latanya Sweeney

Scans the Web determining whether there is sufficient publicly-available information on an individual to fraudulently represent the person in financial and credentialing transactions.

L. Sweeney. AI Technologies to Defeat Identity Theft Vulnerabilities. AAAI Spring Symposium on AI Technologies for Homeland Security, 2005.  
<http://privacy.cs.cmu.edu/dataprivacy/projects/idangel/index.html>

## What is Identity Theft?

### **Identity theft**

Identity theft occurs when a person uses another person's personally-identifying information such as name, Social Security number, credit card number or other explicitly identifying information, without permission to commit fraud or other crimes.

Source: Federal Trade Commission, <http://www.consumer.gov/idtheft/>

## Problems Posed by Identity Theft

Identity theft is a serious crime.

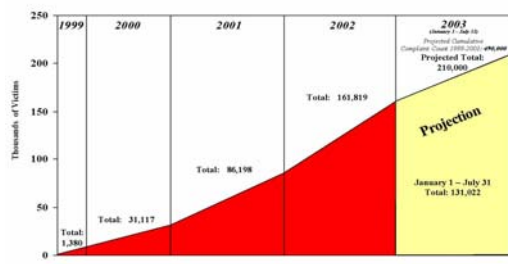
People whose identities have been stolen can spend months or years - and their hard-earned money - cleaning up the mess thieves have made of their good name and credit record.

Victims may lose job opportunities, be refused loans, education, housing or cars, or even get arrested for crimes they didn't commit.

Source: Federal Trade Commission, <http://www.consumer.gov/idtheft/>

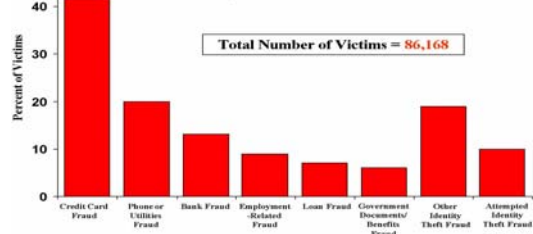
## Federal Trade Commission Report: Overview of the Identity Theft Program, Oct 1998 – Sep 2003

Number of Complaints Entered Into the IDT Data Clearinghouse 1999-2003



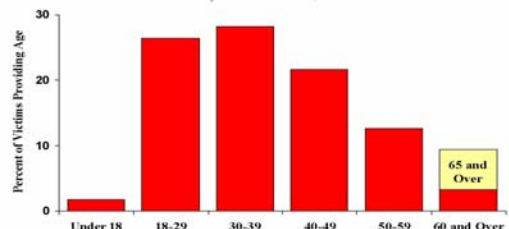
## Federal Trade Commission Report: Victim Complaint Data

Figure 1  
How Victims' Information Is Misused<sup>1</sup>  
January 1 – December 31, 2001



## Federal Trade Commission Report: Victim Complaint Data

Figure 3  
Victim Age Distribution<sup>1</sup>  
January 1 – December 31, 2001



## New Finding: Using Online Resumes for Fraudulent New Credit Cards

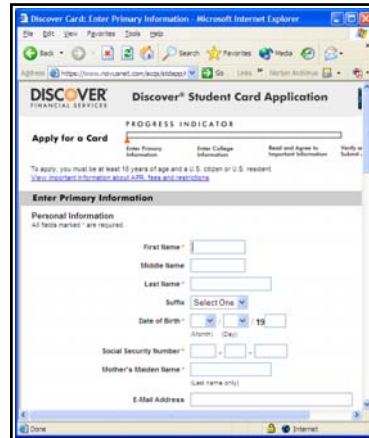
An algorithm that shows how thousands of fraudulent credit cards could be issued to malicious parties using only FREE on-line resumes and other information.

- If so, thousands of Americans are at risk to identity theft immediately!
- What's needed:
  - Credit card application requirements
  - Finding Social Security numbers on-line
  - Finding dates of birth on-line

### Basic Information Necessary For a Credit Card Application

- Name
- Social Security number
- Address
- Date of birth
- Mother's maiden name

The most sensitive combination is name and Social Security number.



Student application Basic information and School Information

### Basic Information Necessary For a Credit Card Application

Do these first.

- Name
- Social Security number
- Address
- Date of birth
- Mother's maiden name

Strategy: if one can identify these fields for a person, they have the basic information needed to acquire a credit card in that person's name. Therefore, we need only demonstrate how this information can be obtained on-line.

### Identity Angel - resumes

1. Locate on-line resumes (using Filtered Searching)
2. Extract sensitive values (using regular expressions)
3. Email subjects about their risks

L. Sweeney. AI Technologies to Defeat Identity Theft Vulnerabilities. AAAI Spring Symposium on AI Technologies for Homeland Security, 2005. <http://privacy.cs.cmu.edu/dataprivacy/projects/idangel/index.html>

### ID Angel, Sample Resume 1

[DOC]RESUME  
File Format: Microsoft Word 2000 - View as HTML  
RESUME. RICHARD ALLEN BROWN. Richard Allen Brown. PO Box 782. Kayenta, AZ 86033.  
Home Telephone-520-697-3513. NAU Telephone-520-523-4099. DOB: 03-10-77. SSN: 527-71 ...  
[dana.ucc.nau.edu/~rab39/RAB%20Resume.doc](http://dana.ucc.nau.edu/~rab39/RAB%20Resume.doc)

100's found. One is shown above. But the actual resumes are amidst lots of non-resume pages!

### ID Angel, Sample Resume 2

resume  
... 2843. DOB: 10-10-48 New Britain, CT 06050-4010. F: (860) 832-3753.  
SSN: 461-84-8245 H: (203) 740-7255 C: (203) 561-8674.  
Education. Ph. ...  
[www.math.ccsu.edu/vaden-goad/resume.htm](http://www.math.ccsu.edu/vaden-goad/resume.htm)

A second example.

### ID Angel, Sample Resume 3

Scot Lytle's Resume  
 Scot Patrick Lytle. Home: (301)-249-5330 2116 Blaz Court  
 School: (410)-455-1662  
 Upper Marlboro, MD 20772 SSN: 578-90-8915  
 OBJECTIVE. ...  
[userpages.umbc.edu/~slytle1/resume.html](http://userpages.umbc.edu/~slytle1/resume.html)

We emailed warnings to these people that this is not a good practice!  
 One claimed to have been the victim of a identity theft recently.

### Basic Information Necessary For a Credit Card Application

|         |   |
|---------|---|
| Done.   | <ul style="list-style-type: none"> <li>• Name</li> <li>• Social Security number</li> <li>• Address</li> </ul> |
| Next... | <ul style="list-style-type: none"> <li>• Date of birth</li> <li>• Mother's maiden name</li> </ul>             |

Strategy: if one can identify these fields for a person, they have the basic information needed to acquire a credit card in that person's name. Therefore, we need only demonstrate how this information can be obtained on-line.

### ID Angel, Sample Resume 1

[DOC]RESUME  
 File Format: Microsoft Word 2000 - View as HTML  
 RESUME. RICHARD ALLEN BROWN. Richard Allen Brown. PO Box 782. Kayenta, AZ 86033.  
 Home Telephone-520-697-3513. NAU Telephone-520-523-4099. DOB: 03-10-77. SSN: 527-71 ...  
[dana.ucs.nau.edu/~rab39/RAB%20Resume.doc](http://dana.ucs.nau.edu/~rab39/RAB%20Resume.doc)

This on-line resume, located earlier, actually listed date of birth too!

### ID Angel, Sample Resume 2

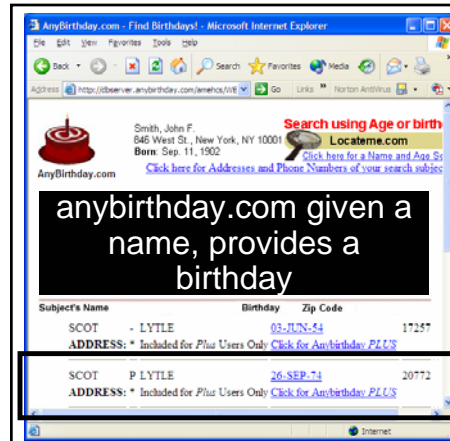
resume  
 ... 2843 (DOB: 10-10-48 New Britain, CT 06050-4010. F: (860) 832-3753  
 SSN: 461-84-8245 H: (203) 740-7255 C: (203) 561-8674.  
 Education. Ph. ...  
[www.math.ccsu.edu/vaden-goad/resume.htm](http://www.math.ccsu.edu/vaden-goad/resume.htm)

This on-line resume, found earlier, also listed date of birth!

### ID Angel, Sample Resume 3

Scot Lytle's Resume  
 Scot Patrick Lytle. Home: (301)-249-5330 2116 Blaz Court  
 School: (410)-455-1662  
 Upper Marlboro, MD 20772 SSN: 578-90-8915  
 OBJECTIVE. ...  
[userpages.umbc.edu/~slytle1/resume.html](http://userpages.umbc.edu/~slytle1/resume.html)

The third resume did not have his DOB listed.



Had several hits matching name, but only one in his ZIP.

### Basic Information Necessary For a Credit Card Application

|         |   |
|---------|---|
| Done.   | <ul style="list-style-type: none"> <li>• Name</li> <li>• Social Security number</li> <li>• Address</li> </ul> |
| Done.   | <ul style="list-style-type: none"> <li>• Date of birth</li> </ul>   |
| Next... | <ul style="list-style-type: none"> <li>• Mother's maiden name</li> </ul>                                      |

Strategy: if one can identify these fields for a person, they have the basic information needed to acquire a credit card in that person's name. Therefore, we need only demonstrate how this information can be obtained on-line.

Carnegie Mellon  
**DATA PRIVACY LAB**

### Identity Angel –resume findings

1. 1000 resume hits on Google using filteredSearch, revealed 150 resumes, of which 140 (or 93%) had complete 9-digit SSNs.

10 resumes had partial, invalid, or some other country's SSN.

L. Sweeney. AI Technologies to Defeat Identity Theft Vulnerabilities. AAAI Spring Symposium on AI Technologies for Homeland Security, 2005. <http://privacy.cs.cmu.edu/dataprivacy/projects/idangel/index.html>

Carnegie Mellon  
**DATA PRIVACY LAB**

### Identity Angel –resume findings

2. All email addresses (113 of 113 or 100%) were found. The '@' and dot (.) notation worked well. All dates of birth (110 of 110 or 100%) were found, but some dates, which were not dates of birth were incorrectly reported as such; this happened in 20 cases (but only 7 where the proper DOB was not also found).

L. Sweeney. AI Technologies to Defeat Identity Theft Vulnerabilities. AAAI Spring Symposium on AI Technologies for Homeland Security, 2005. <http://privacy.cs.cmu.edu/dataprivacy/projects/idangel/index.html>

Carnegie Mellon  
**DATA PRIVACY LAB**

### Identity Angel –resume findings

3. In terms of combinations:
  - 104 (or 69%) resumes had {SSN, DOB};
  - 105 (or 70%) had {SSN, email},
  - 76 (or 51%) had {SSN, DOB, email}.

L. Sweeney. AI Technologies to Defeat Identity Theft Vulnerabilities. AAAI Spring Symposium on AI Technologies for Homeland Security, 2005. <http://privacy.cs.cmu.edu/dataprivacy/projects/idangel/index.html>

Carnegie Mellon  
**DATA PRIVACY LAB**

### Identity Angel –resume findings

4. A single email message was sent to each of the 105 people having {SSN, email} alerting them to the risk. Within a month, 42 (or 55% of all of DBB) no longer had the information publicly available. A year later, 102 (or 68% of all of DBA) no longer had the information available.

L. Sweeney. AI Technologies to Defeat Identity Theft Vulnerabilities. AAAI Spring Symposium on AI Technologies for Homeland Security, 2005. <http://privacy.cs.cmu.edu/dataprivacy/projects/idangel/index.html>

Carnegie Mellon  
**DATA PRIVACY LAB**

### Privacy Technology

1. The Privacy Problem.
2. How computer scientists help solve this problem.
3. Example: video surveillance
4. Example: bio-terrorism surveillance
5. Example: identity theft
6. Example: selective revelation
7. Example: distributed surveillance
8. Example: privacy-preserving surveillance
9. Example: DNA privacy
10. Example: Identity theft protections
11. Example: k-Anonymity
12. Example: Webcam surveillance
13. Example: Text de-identification
14. Example: Policy specification and enforcement
15. Example: Scam Spam

[privacy.cs.cmu.edu](http://privacy.cs.cmu.edu)