Seductive myths about privacy

• Myth: The major privacy risk is from unauthorized access to information
• Myth: Privacy can be adequately protected by removing personally identifying information (PII) from records to be released.
• Myth: Notice and choice is an adequate framework for privacy protection
• Myth: Personal privacy is about individuals
Seductive myths about privacy

• Myth: The major privacy risk is from unauthorized access to information
  • Myth: Privacy can be adequately protected by removing personally identifying information (PII) from records to be released.
• Myth: Notice and choice is an adequate framework for privacy protection
• Myth: Personal privacy is about individuals
Seductive myths about privacy

• Myth: The major privacy risk is from unauthorized access to information

• Reality: Confounding security and privacy is a favorite myth of the computer security industry and of IT organizations everywhere.
Seductive myths about privacy

• Myth: The major privacy risk is from unauthorized access to information

• Myth: Privacy can be adequately protected by removing personally identifying information (PII) from records to be released.
Seductive myths about privacy

• Myth: Privacy can be adequately protected by removing personally identifying information (PII) from records to be released.

• Reality: The belief that information can be de-identified is the basis for much current privacy regulation. But information can be readily re-identified.
Reidentification of Individuals in Chicago's Homicide Database
A Technical and Legal Study

<table>
<thead>
<tr>
<th>Salvador Ochoa</th>
<th>Jamie Rasmussen</th>
<th>Christine Robson</th>
<th>Michael Salib</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collective address:</td>
<td><a href="mailto:reidentify@mit.edu">reidentify@mit.edu</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abstract

Many government agencies, hospitals, and other organizations collect personal data of a sensitive nature. Often, these groups would like to release their data for statistical analysis by the scientific community, but do not want to cause the subjects of the data embarrassment or harassment. To resolve this conflict between privacy and progress, data is often deidentified before publication. In short, personally identifying information such as names, home addresses, and social security numbers are stripped from the data. We analyzed one such deidentified data set containing information about Chicago homicide victims over a span of three decades. By comparing the records in the Chicago data set with records in the Social Security Death Index, we were able to identify 85% of the victims. This approach is...
Drug Records, Confidential Data Vulnerable

Harvard ID numbers, PharmaCare loophole provide wide-ranging access to private data

By J. HALE RUSSELL and ELISABETH S. THEODORE
CRIMSON STAFF WRITERS

The confidential drug purchase histories of many Harvard students and employees have been available for months to any internet user, as have the e-mail addresses of high-profile undergraduates whose contact information the University legally must conceal, a Crimson investigation has found.
\{date of birth, gender, 5-digit ZIP\} uniquely identifies 87.1\% of USA pop.

courtesy Latanya Sweeney, CMU
Seductive myths about privacy

• Myth: Notice and choice is an adequate framework for privacy protection
Seductive myths about privacy

• Myth: Notice and choice is an adequate framework for privacy protection

• Reality: Choice, whether opt-in or opt-out are meaningless if the choice is not informed. “User choice” has become a way for industry to shift blame to users.
Seductive myths about privacy

- Myth: Personal privacy is about individuals
Seductive myths about privacy

• Myth: Personal privacy is about individuals
• Reality: On the internet, people really can judge you by your friends (your mother was right).
• A “personal choice” to reveal information about yourself also reveals information about your associates.
Information Leakage from Social Networks

Jernigan and Mistree (2007)
Seductive myths about privacy

- Myth: The major privacy risk is from unauthorized access to information
- Myth: Privacy can be adequately protected by removing personally identifying information (PII) from records to be released.
- Myth: Notice and choice is an adequate framework for privacy protection
- Myth: Personal privacy is about individuals
Moving from an old privacy framework …

• Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.
To a privacy framework for the information age

• Privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others.

• Privacy is the claim of individuals, groups, or institutions to determine when, how, and to what extent information about them is used by others in ways that affect them.
The RMP restrictions

• We currently offer five RMP restrictions:
  ○ no-commercial
  ○ no-depiction
  ○ no-employment
  ○ no-financial
  ○ no-medical

• A user is able to choose any combination of these restrictions to apply on their personal information.

• The user is then given an icon, similar to the Creative Commons icon, that can be publicly posted on their profile pages.
RMP on Facebook/OpenSocial

- RMP applications for Facebook and OpenSocial.
- The applications allow users to create and display restrictions on their private information.
- An icon is created from their choices that is displayed on a user's profile page and links to a page containing more information.
Information Accountability: When information has been used, it should be possible to determine what happened, and to pinpoint use that is inappropriate.
Technology to support information accountability

• Information is annotated with provenance that identifies its source.
• Data transfers and uses are logged so that chains of transfers have audit trails.
• Databases and data providers supply machine-readable policies that govern permissible uses of the data.
• Automated reasoning engines use policies to determine whether data use is appropriate.
• Users manipulate information via policy-aware interfaces that can enforce policies and/or signal non-compliant uses.
Use Case: Data sharing in Fusion Centers

• Current CSAIL research for DHS

• Example
  – Sender: Mia Analysa of Massachusetts Commonwealth Fusion Center
  – Data: Request for Information regarding Robert Guy
  – Receiver: Fedd Agenti of DHS
  – Is this allowed under policies of the involved parties?
Automated policy reasoning

so long as Mia Analyza's Document is necessary for actual performance of Criminal Justice Duties, and Feddy Agenti is actually performing Criminal Justice Duties, as required by MGL 6-172, Para. 2.

Premises:

Feddy Agenti

certified by: The Board
actually performing: Criminal Justice Duties
whose principle function is: prosecution of criminal offenders
type: Male
label: Feddy Agenti
office: g3
lat: 40.7532
long: -74.0007
authorized purpose: law enforcement
family name: Agenti
given name: Fred
Myth: The major privacy risk is from unauthorized