Information accountability as the foundation of 21st century privacy protection

Hal Abelson
CSAIL Decentralized Information Group
Massachusetts Institute of Technology

16 October 2013

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 personal

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 personal

Seductive myths about privacy

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

- Reality: Conflating 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

- Reality: Conflating security and privacy is a favorite myth of the computer security industry and of IT organizations everywhere.

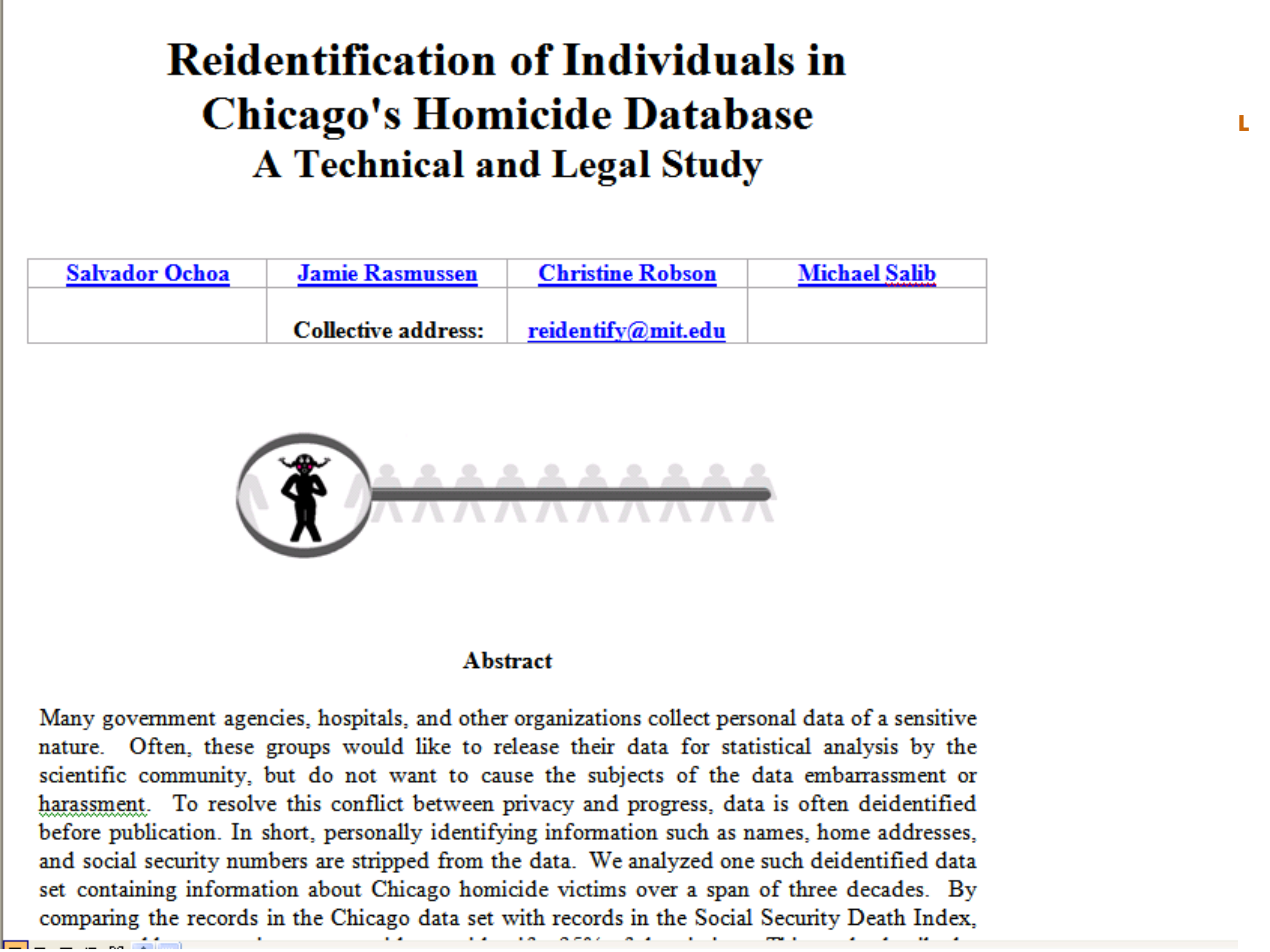
- Misuse by people who have been granted *authorized* access

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

Salvador Ochoa	Jamie Rasmussen	Christine Robson	Michael Salib
	Collective address:	reidentify@mit.edu	

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,

NEWS

Published on Friday, January 21, 2005

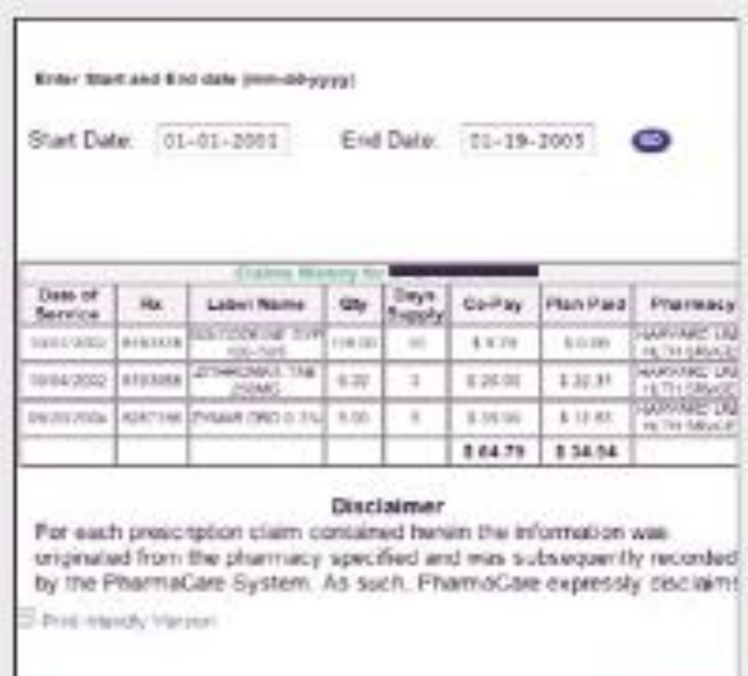
Drug Records, Confidential Data Vulnerable

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.



Enter Start and End date (mm-dd-yyyy)

Start Date: 01-01-2001 End Date: 01-19-2005

Claims History for [redacted]

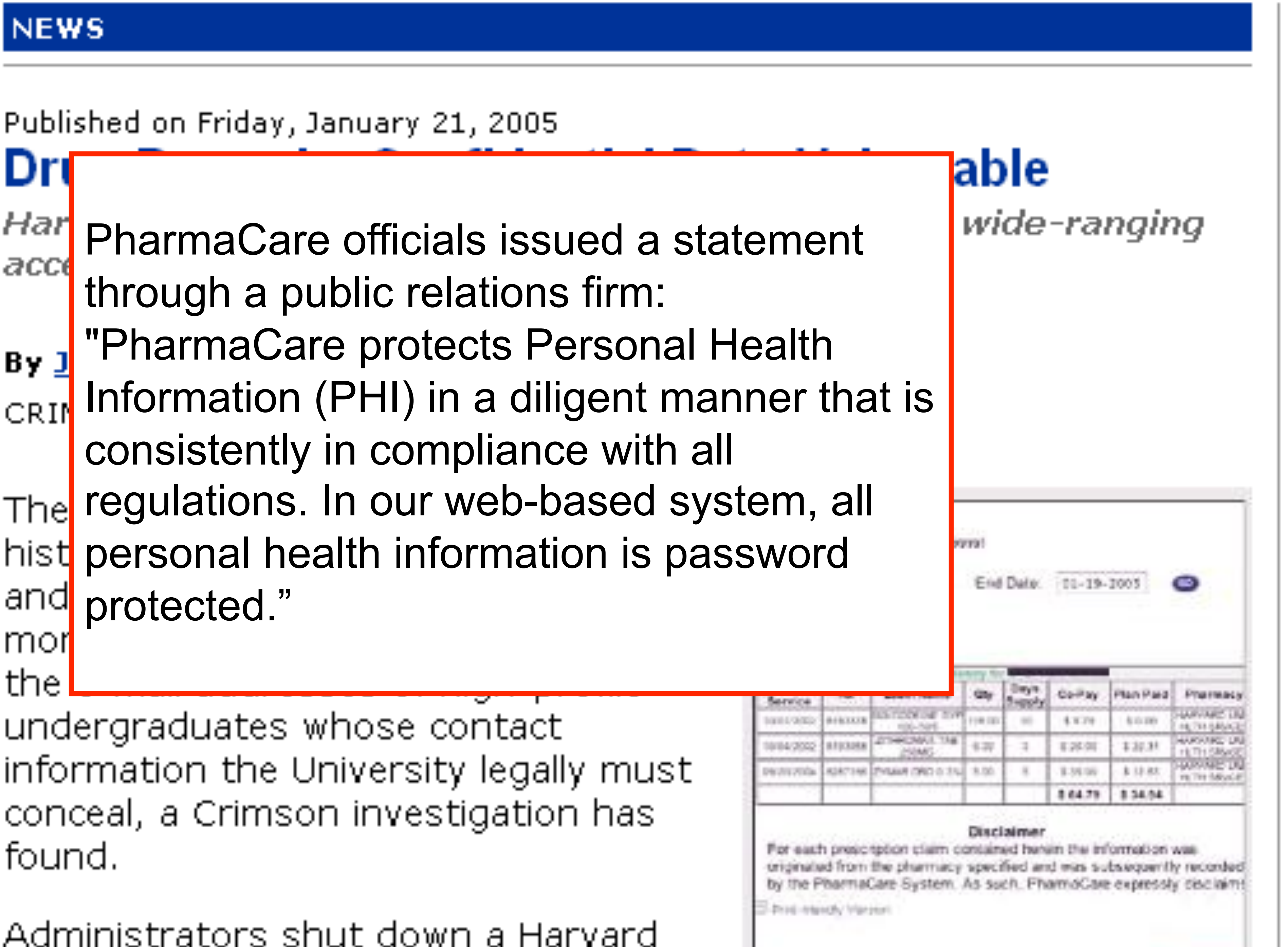
Date of Service	Rx	Label Name	Qty	Days Supply	Co-Pay	Plan Paid	Pharmacy
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$ 9.79	$ 0.00	HARVARD UNI (617) [illegible]
[illegible]	[illegible]	[illegible] TAB 250MG	[illegible]	3	$ 20.00	$ 22.31	HARVARD UNI (617) [illegible]
[illegible]	[illegible]	[illegible]	5.00	5	$ 35.00	[illegible]	HARVARD UNI (617) [illegible]
					$ 64.79	$ 34.54	

Disclaimer

For each prescription claim contained herein the information was originated from the pharmacy specified and was subsequently recorded by the PharmaCare System. As such, PharmaCare expressly disclaims

Print-friendly Version

Administrators shut down a Harvard



NEWS

Published on Friday, January 21, 2005

Dr able

Har wide-ranging acc

By J

CRI

The hist and mor the undergraduates whose contact information the University legally must conceal, a Crimson investigation has found.

Administrators shut down a Harvard

End Date: 01-19-2005

History for [redacted]

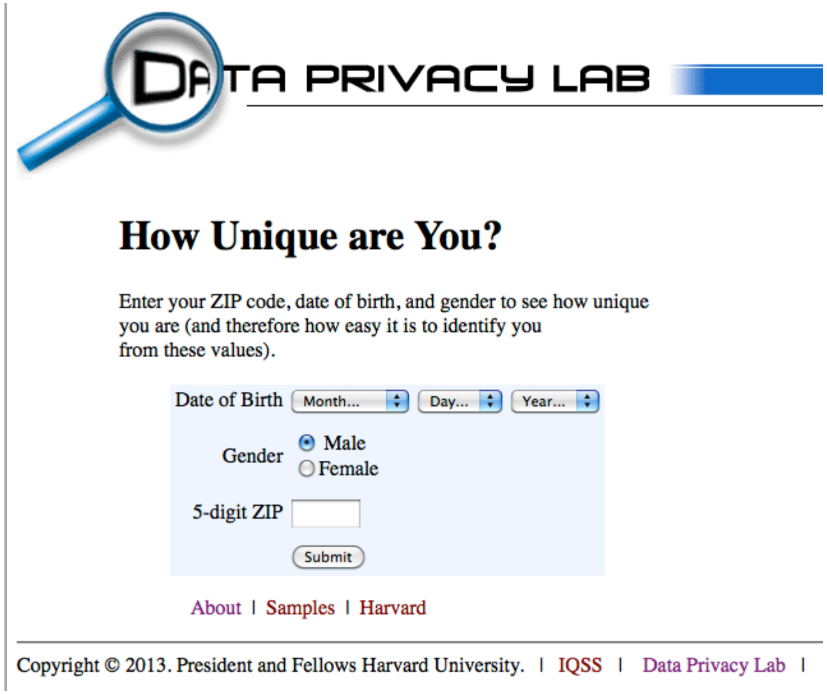
Service			Qty	Days Supply	Co-Pay	Plan Paid	Pharmacy
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$ 9.79	$ 0.00	HARVARD UNI[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$ 20.00	$ 22.31	HARVARD UNI[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	$ 35.00	[illegible]	HARVARD UNI[illegible]
					$ 64.79	$ 34.54	

Disclaimer

For each prescription claim contained herein the information was originated from the pharmacy specified and was subsequently recorded by the PharmaCare System. As such, PharmaCare expressly disclaims

Print-friendly Version

PharmaCare officials issued a statement through a public relations firm: "PharmaCare protects Personal Health Information (PHI) in a diligent manner that is consistently in compliance with all regulations. In our web-based system, all personal health information is password protected."



How Unique are You?

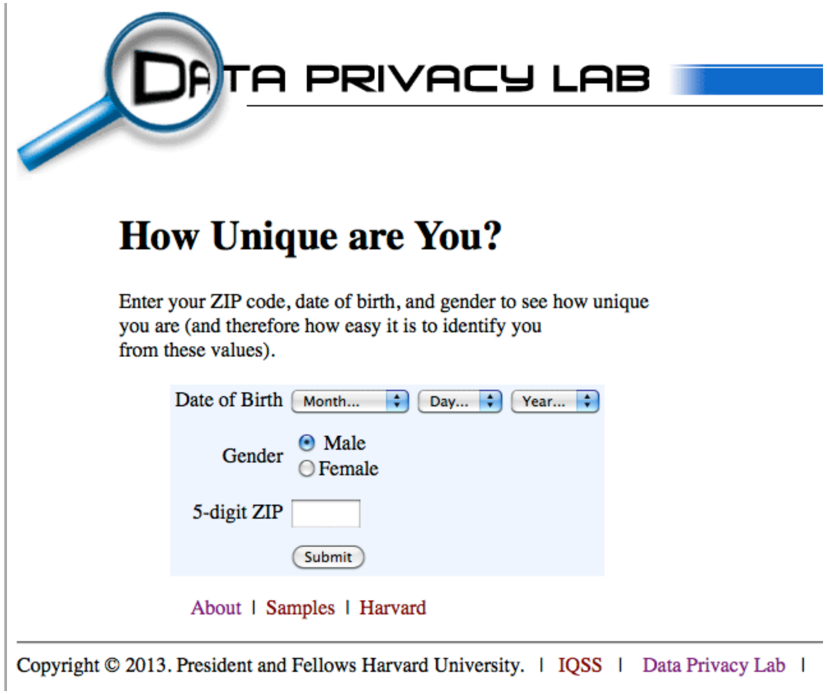
Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

Date of Birth Month... Day... Year...

Gender ● Male ○ Female

5-digit ZIP

About | Samples | Harvard



DATA PRIVACY LAB

How Unique are You?

Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

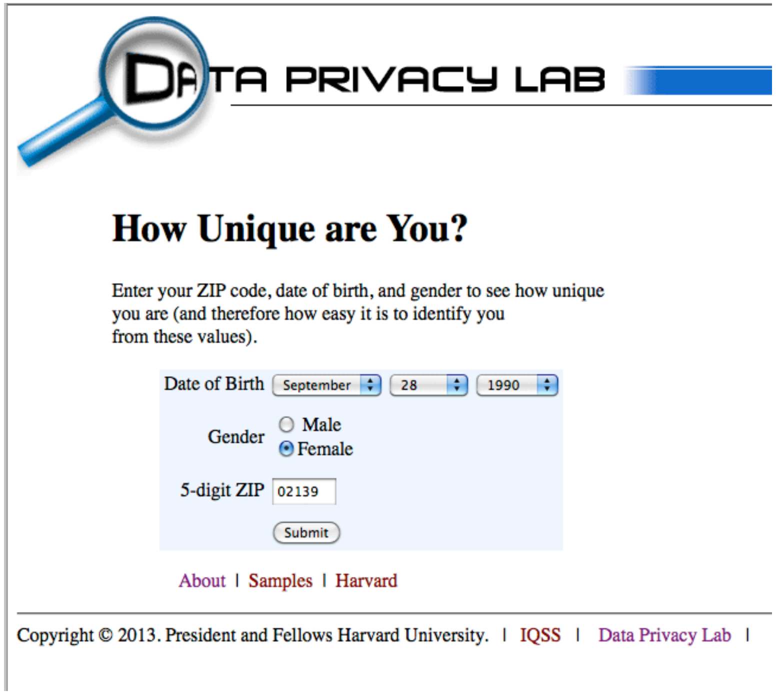
Date of Birth: Month... | Day... | Year...

Gender: ● Male ○ Female

5-digit ZIP:

About | Samples | Harvard

Copyright © 2013. President and Fellows Harvard University. | IQSS | Data Privacy Lab |



DATA PRIVACY LAB

How Unique are You?

Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

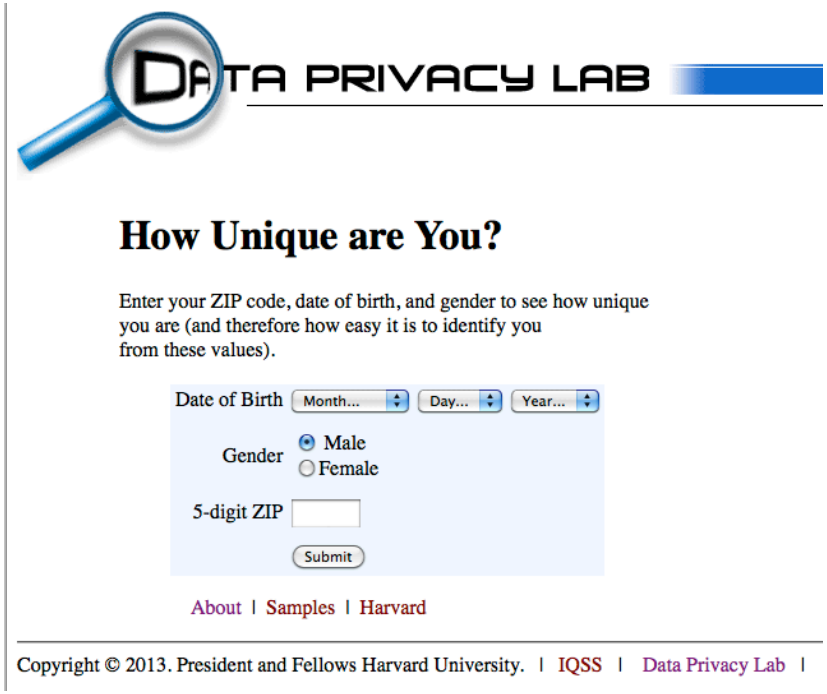
Date of Birth: September | 28 | 1990

Gender: ○ Male ● Female

5-digit ZIP: 02139

About | Samples | Harvard

Copyright © 2013. President and Fellows Harvard University. | IQSS | Data Privacy Lab |



DATA PRIVACY LAB

How Unique are You?

Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

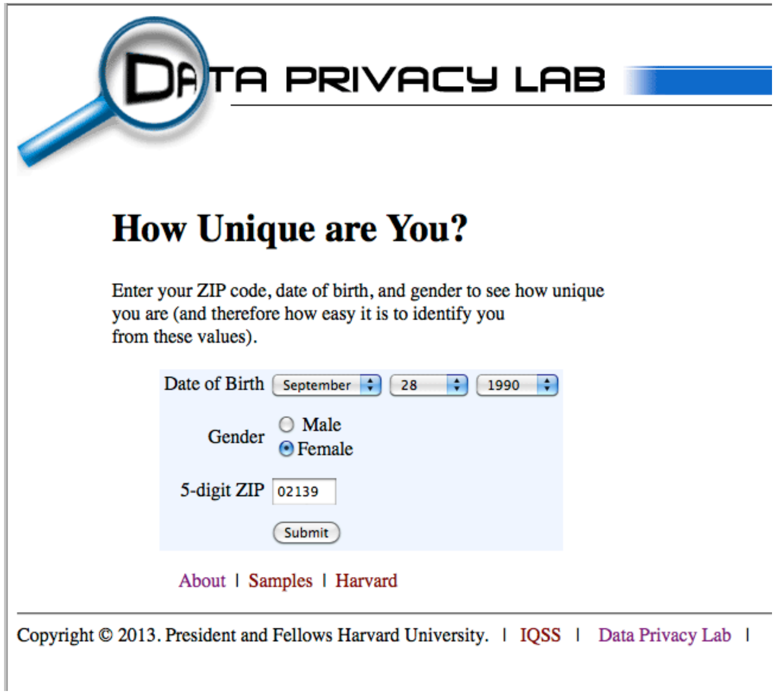
Date of Birth: Month... Day... Year...

Gender: ● Male ○ Female

5-digit ZIP:

About | Samples | Harvard

Copyright © 2013. President and Fellows Harvard University. | IQSS | Data Privacy Lab |



DATA PRIVACY LAB

How Unique are You?

Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

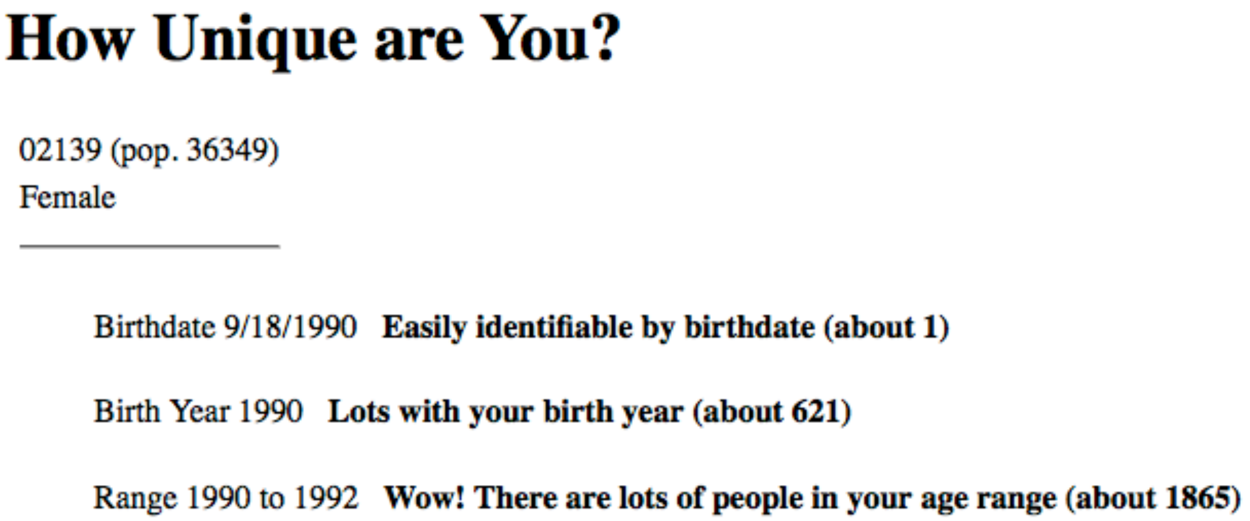
Date of Birth: September 28 1990

Gender: ○ Male ● Female

5-digit ZIP: 02139

About | Samples | Harvard

Copyright © 2013. President and Fellows Harvard University. | IQSS | Data Privacy Lab |



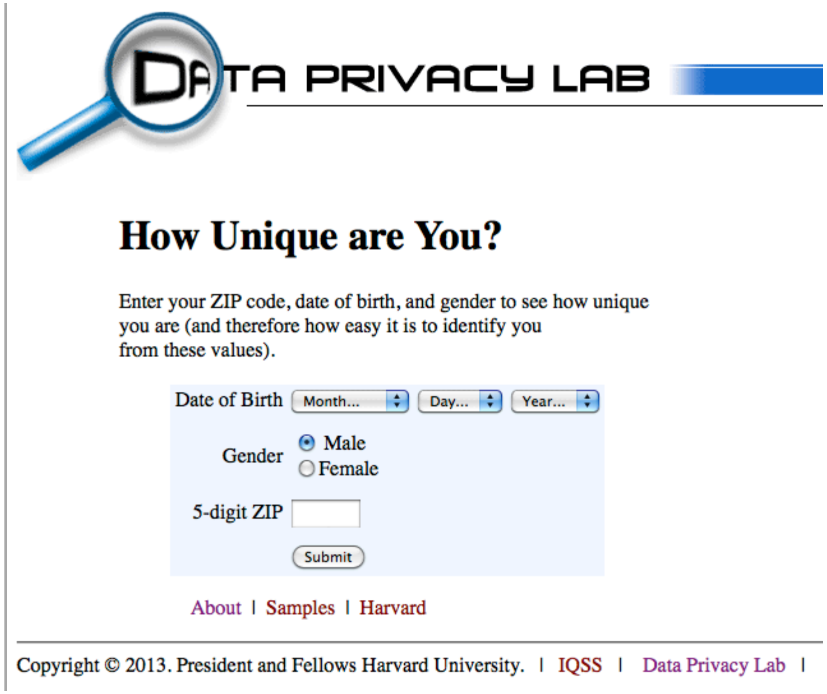
How Unique are You?

02139 (pop. 36349)
Female

Birthdate 9/18/1990 **Easily identifiable by birthdate (about 1)**

Birth Year 1990 **Lots with your birth year (about 621)**

Range 1990 to 1992 **Wow! There are lots of people in your age range (about 1865)**

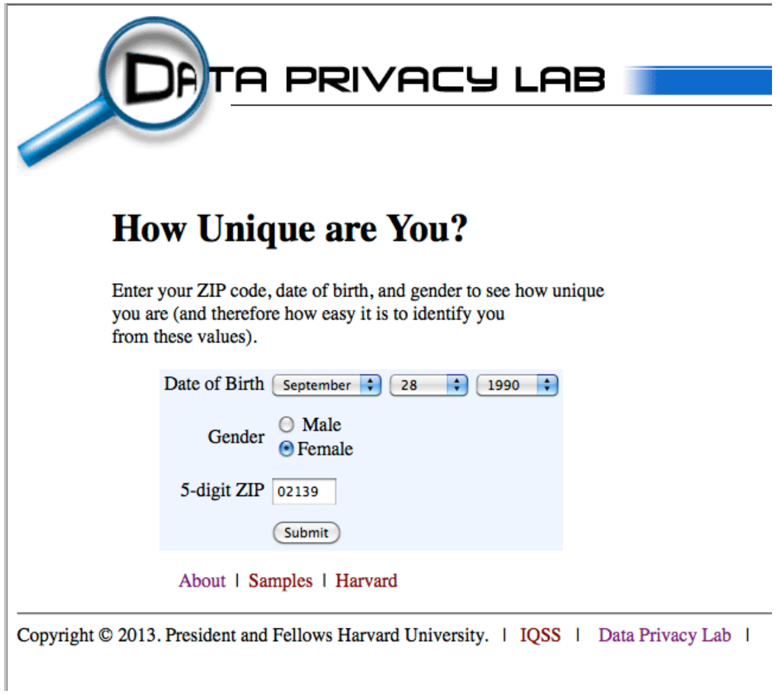


How Unique are You?

Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

Date of Birth	Month... Day... Year...
Gender	● Male ○ Female
5-digit ZIP	
	Submit

About | Samples | Harvard

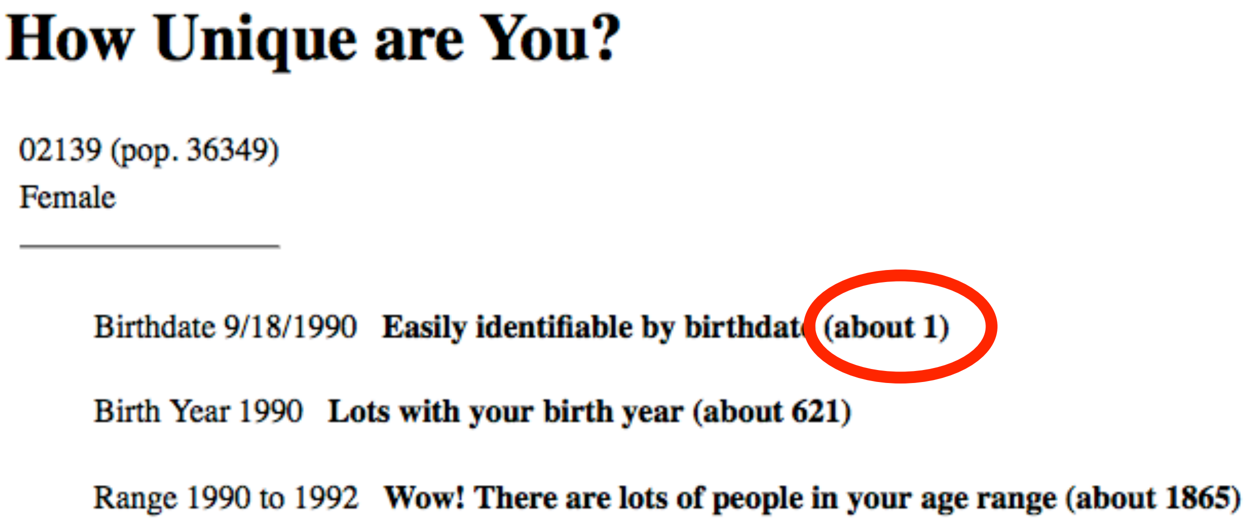


How Unique are You?

Enter your ZIP code, date of birth, and gender to see how unique you are (and therefore how easy it is to identify you from these values).

Date of Birth	September 28 1990
Gender	○ Male ● Female
5-digit ZIP	02139
	Submit

About | Samples | Harvard



How Unique are You?

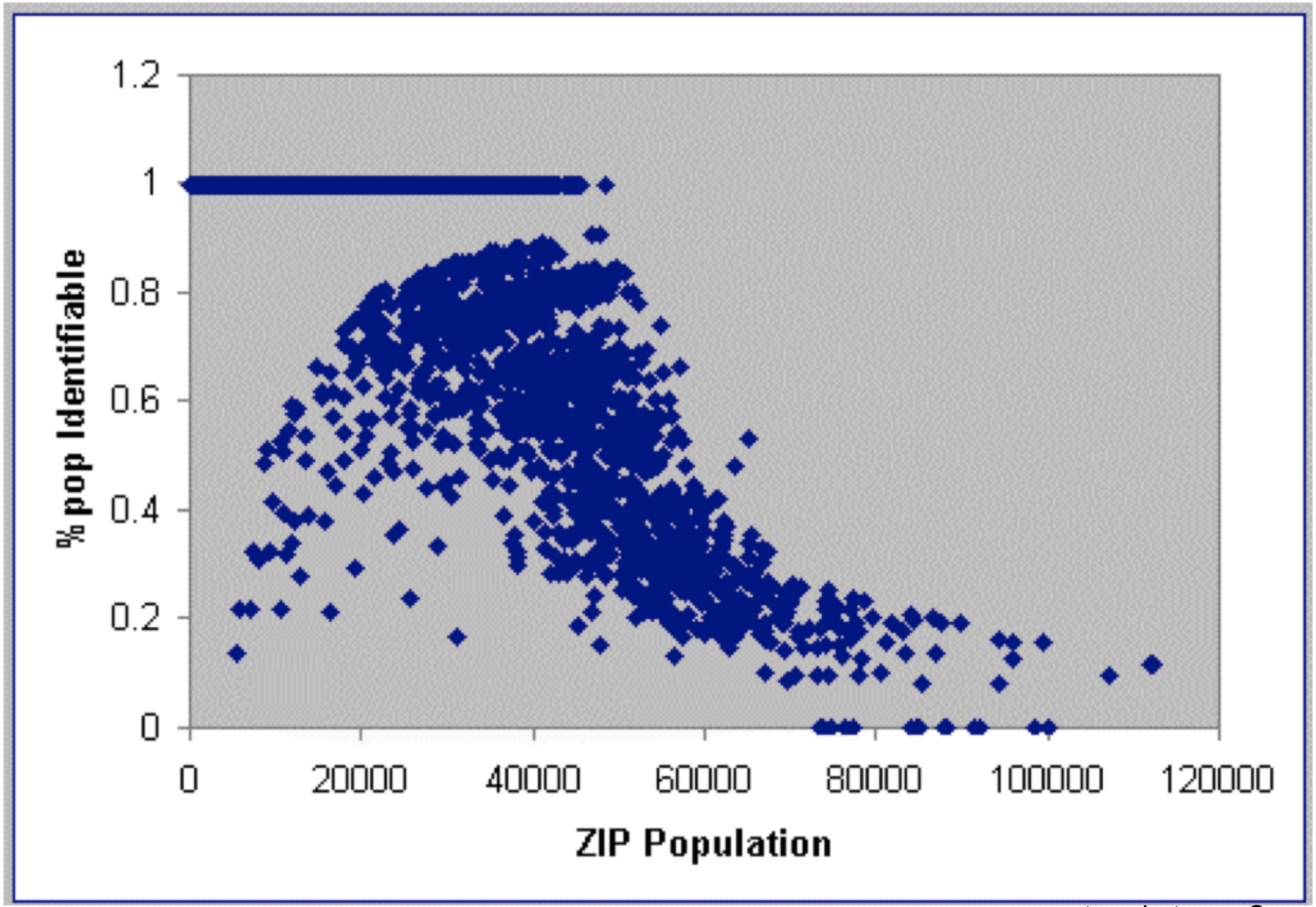
02139 (pop. 36349)
Female

Birthdate 9/18/1990 **Easily identifiable by birthdate (about 1)**

Birth Year 1990 **Lots with your birth year (about 621)**

Range 1990 to 1992 **Wow! There are lots of people in your age range (about 1865)**

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



courtesy Latanya Sweeney, CMU

Seductive myths about privacy

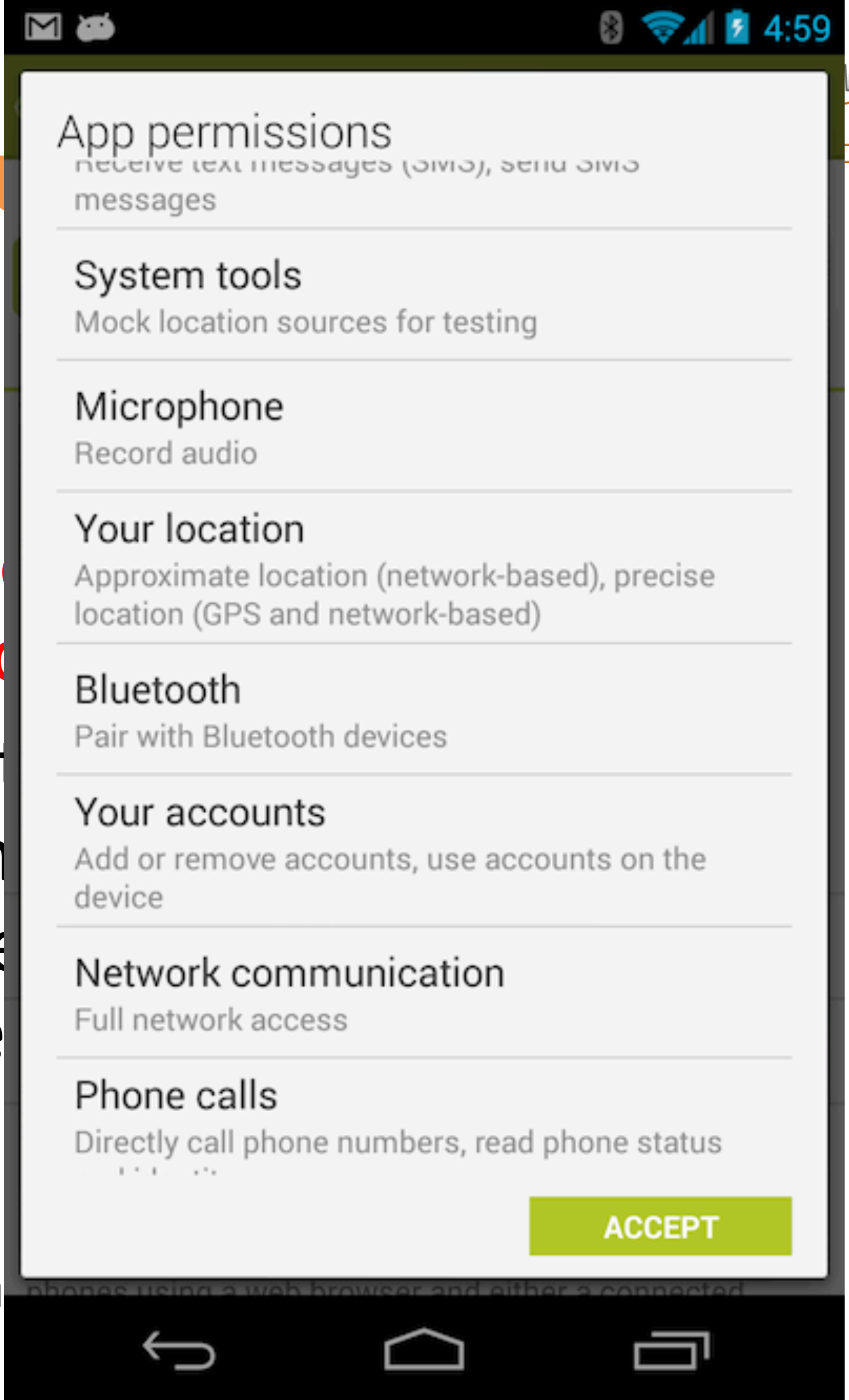
- Myth: Notice and choice is an adequate framework for privacy protection

Seductive myths about privacy

- Myth: Notice and choice is an adequate framework for privacy protection
- Reality: Choice, whether opt-in our opt-out are meaningless if the choice is not informed. "User choice" has become a way for industry to shift blame to users.

Seductive myt

- Myth: Notice and ch framework for priva
- Reality: Choice, whe are meaningless if th "User choice" has be to shift blame to use



4:59

App permissions

Receive text messages (SMS), send SMS messages

System tools
Mock location sources for testing

Microphone
Record audio

Your location
Approximate location (network-based), precise location (GPS and network-based)

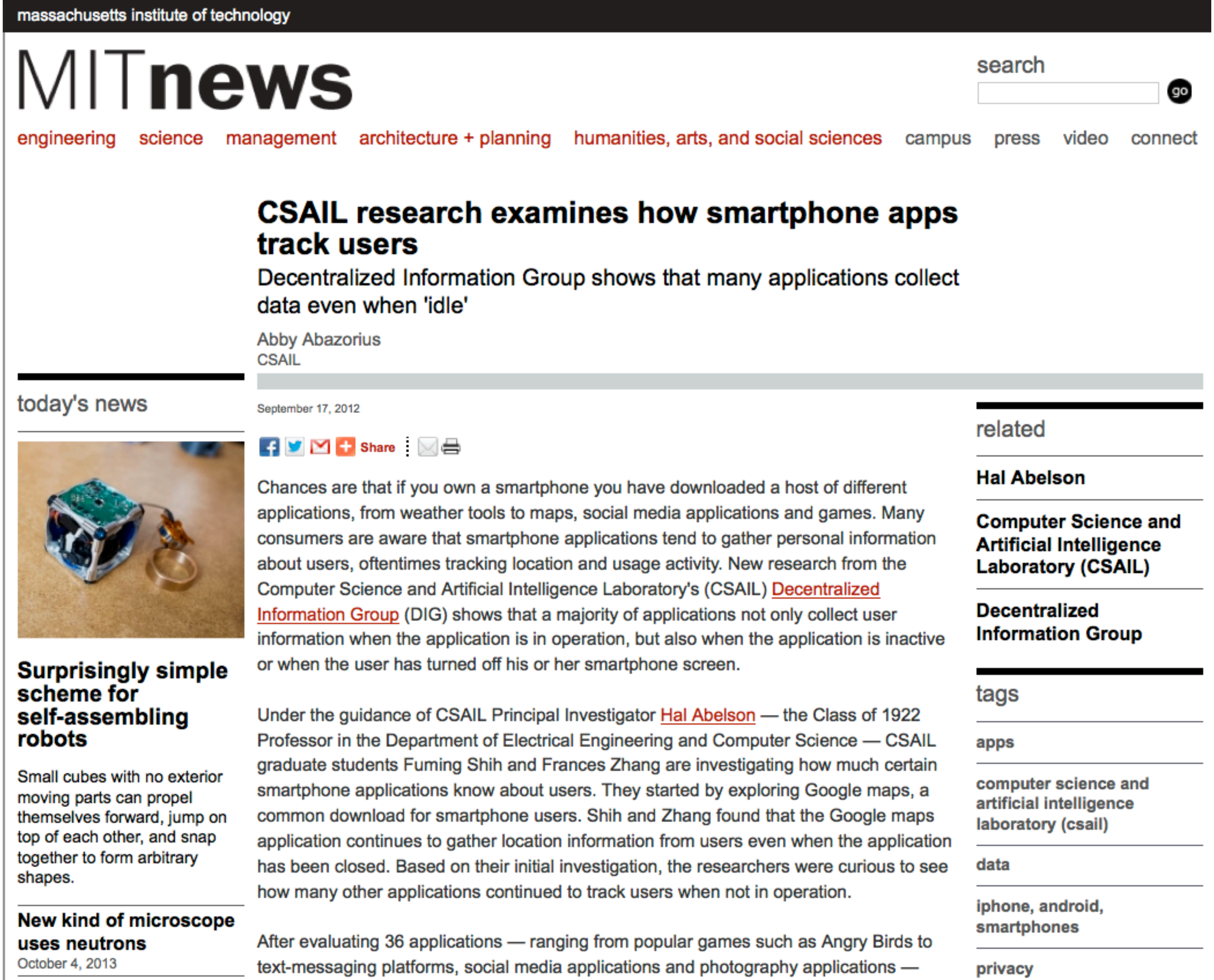
Bluetooth
Pair with Bluetooth devices

Your accounts
Add or remove accounts, use accounts on the device

Network communication
Full network access

Phone calls
Directly call phone numbers, read phone status

phones using a web browser and either a connected



CSAIL research examines how smartphone apps track users

Decentralized Information Group shows that many applications collect data even when 'idle'

Abby Abazorius
CSAIL

September 17, 2012

Chances are that if you own a smartphone you have downloaded a host of different applications, from weather tools to maps, social media applications and games. Many consumers are aware that smartphone applications tend to gather personal information about users, oftentimes tracking location and usage activity. New research from the Computer Science and Artificial Intelligence Laboratory's (CSAIL) Decentralized Information Group (DIG) shows that a majority of applications not only collect user information when the application is in operation, but also when the application is inactive or when the user has turned off his or her smartphone screen.

Under the guidance of CSAIL Principal Investigator Hal Abelson — the Class of 1922 Professor in the Department of Electrical Engineering and Computer Science — CSAIL graduate students Fuming Shih and Frances Zhang are investigating how much certain smartphone applications know about users. They started by exploring Google maps, a common download for smartphone users. Shih and Zhang found that the Google maps application continues to gather location information from users even when the application has been closed. Based on their initial investigation, the researchers were curious to see how many other applications continued to track users when not in operation.

After evaluating 36 applications — ranging from popular games such as Angry Birds to text-messaging platforms, social media applications and photography applications —

today's news

Surprisingly simple scheme for self-assembling robots

Small cubes with no exterior moving parts can propel themselves forward, jump on top of each other, and snap together to form arbitrary shapes.

New kind of microscope uses neutrons
October 4, 2013

related

Hal Abelson

Computer Science and Artificial Intelligence Laboratory (CSAIL)

Decentralized Information Group

tags

apps

computer science and artificial intelligence laboratory (csail)

data

iphone, android, smartphones

privacy

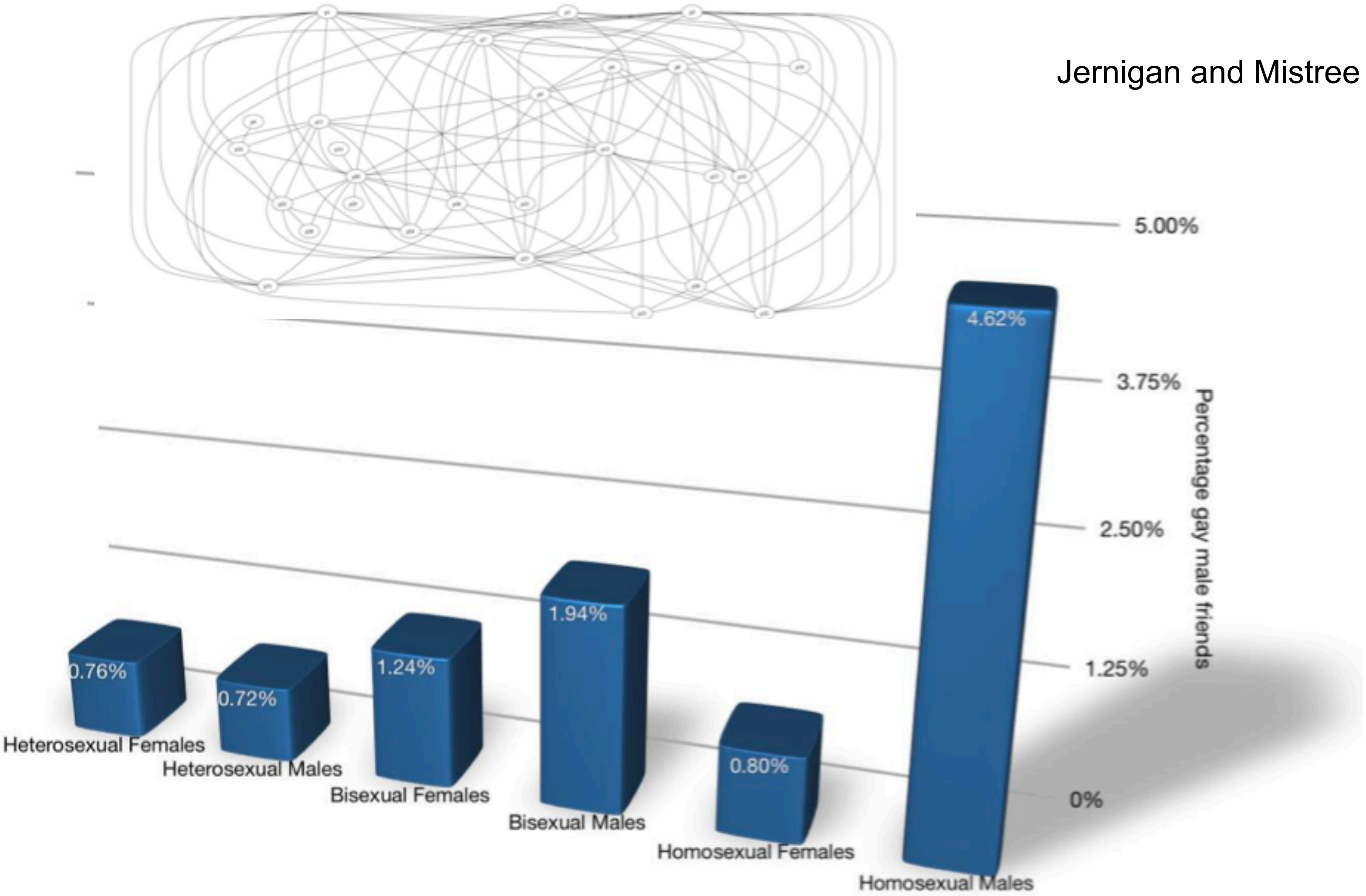
Seductive myths about privacy

- Myth: Personal privacy is personal

Seductive myths about privacy

- Myth: Personal privacy is personal
- A “personal choice” to reveal information about yourself also reveals information about your associates.

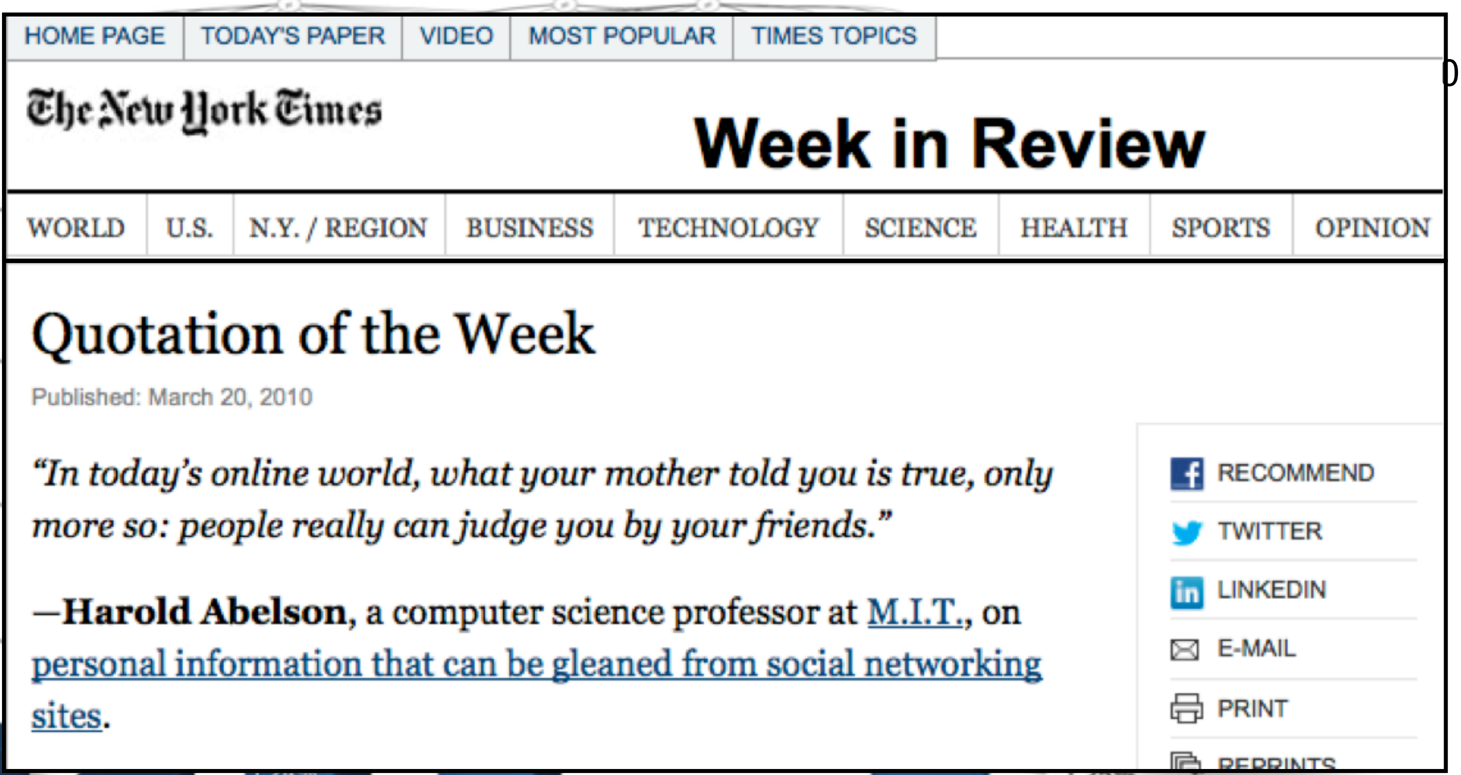
Information Leakage from Social Networks



Jernigan and Mistree (2007)

Information Leakage from Social Networks

007)



The New York Times

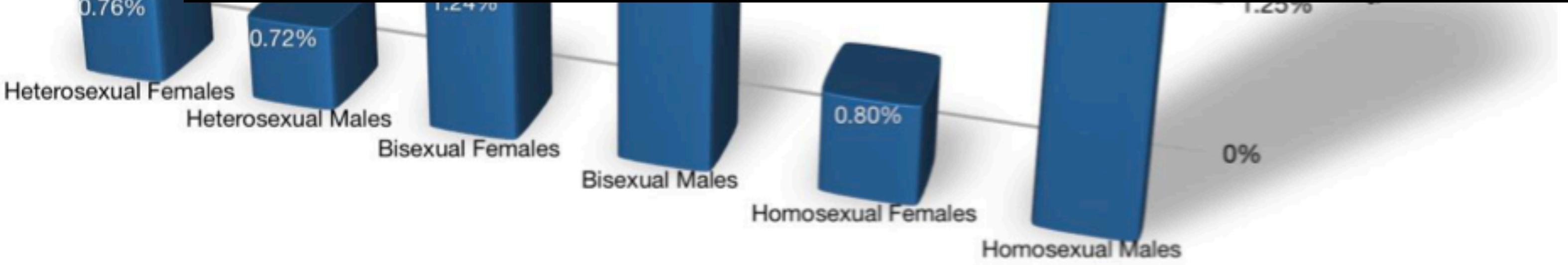
Week in Review

Quotation of the Week

Published: March 20, 2010

"In today's online world, what your mother told you is true, only more so: people really can judge you by your friends."

—**Harold Abelson**, a computer science professor at M.I.T., on personal information that can be gleaned from social networking sites.



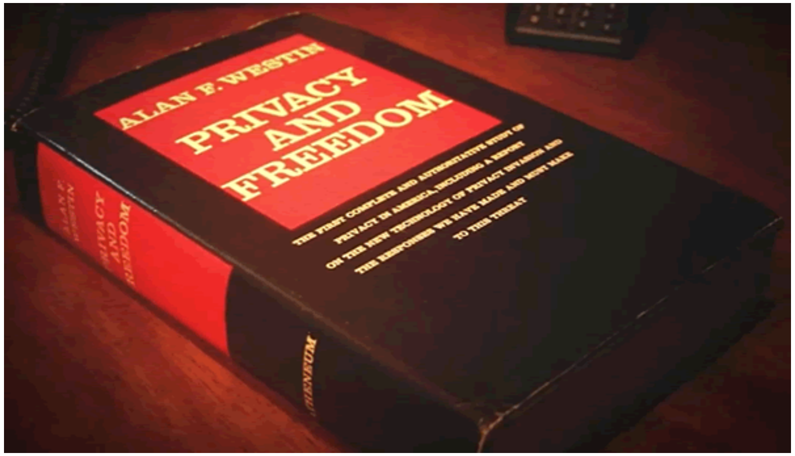
10/

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 personal

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*.**


ALAN F. WESTIN
PRIVACY AND FREEDOM
THE FIRST COMPLETE AND AUTHORITATIVE STUDY OF PRIVACY IN AMERICA, INCLUDING A REPORT ON THE NEW TECHNOLOGY OF PRIVACY INVASION AND THE RESPONSES WE HAVE MADE AND MUST MAKE TO THIS THREAT
ALAN F. WESTIN
PRIVACY AND FREEDOM
ATHENEUM

Alan Westin, *Privacy and Freedom* (1967)

J. Saltzer and M. Schroeder "The Protection of Information in Computer Systems"(CACM 1974)

To a privacy framework for the 21st century

- 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.

Concern with inappropriate disclosure

Concern with inappropriate use

Information accountability

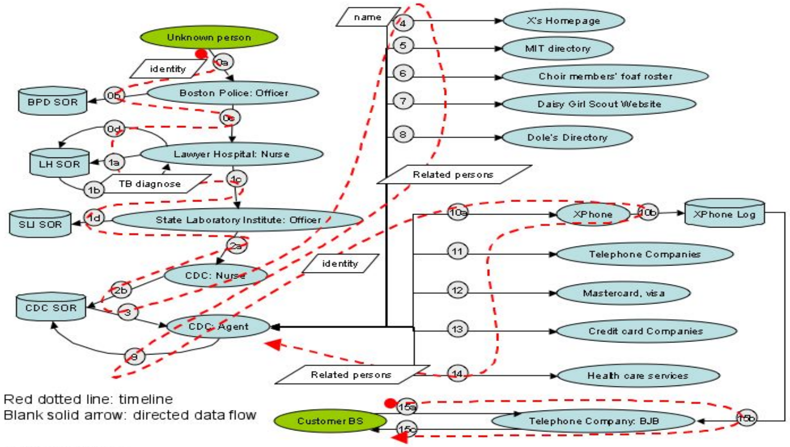
When information has been used, it should to possible to determine what happened, and to pinpoint use that is inappropriate

Technology to support information accountability

- Databases and data providers supply machine-readable policies that govern permissible uses of the data.
- Data transfers and uses are logged so that chains of transfers have audit trails
- Information is annotated with provenance that identifies its source.
- 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.

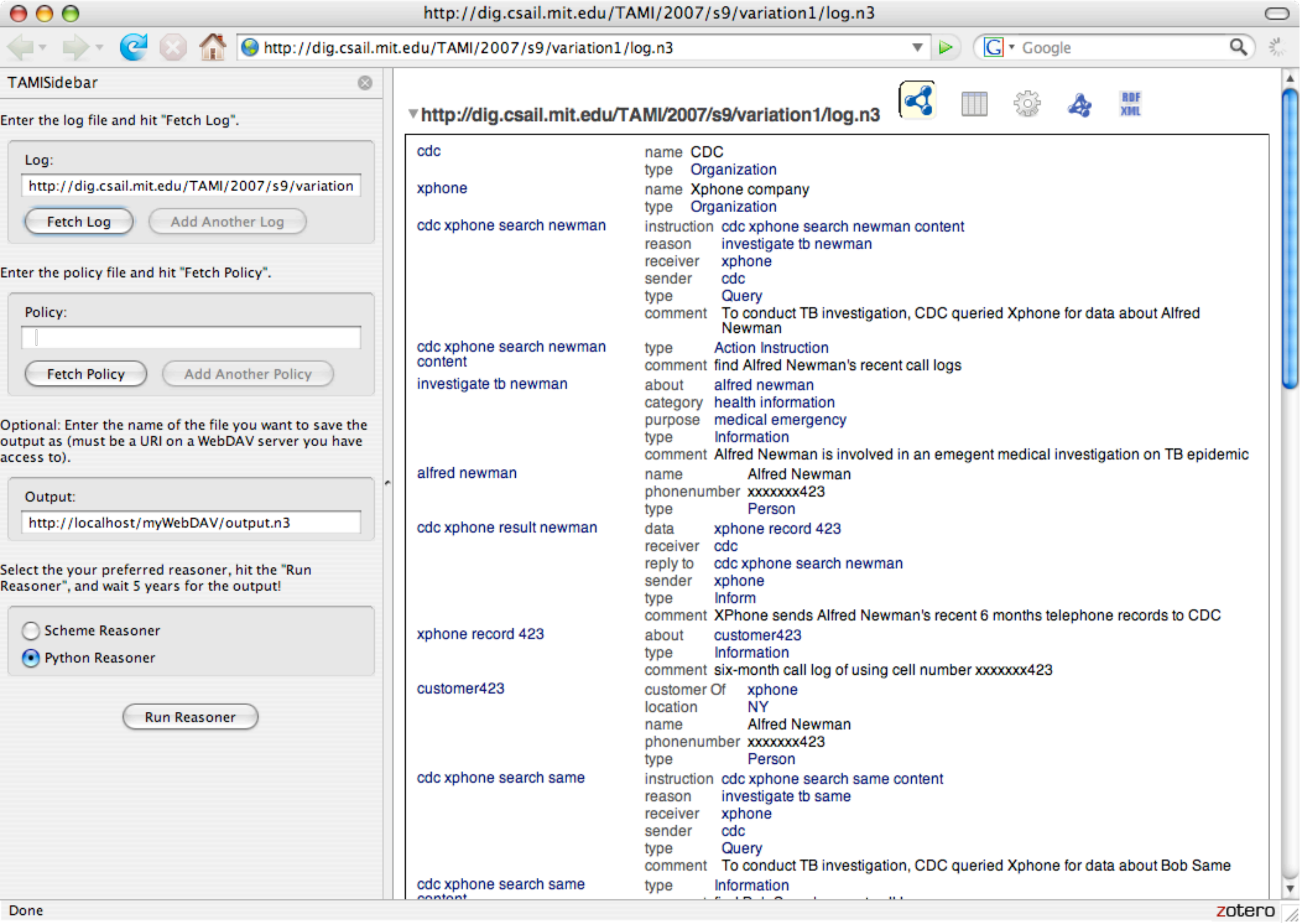
Scenario

- **In order to prevent an epidemic, CDC contacts everyone whom an unconscious tuberculosis patient could have been in contact with**
  - –people he works with, his choir, the members of his scout troop, people he has called, who have called him
- **CDC gets his phone records from**
- **Sometime later Bob Same has phone troubles and calls XPhone to schedule an appt**
- **The customer service operator sees that CDC had obtained his records and infers that he must have some contagious disease**
- **So she refuses to schedule a repairman**



Prepared by Li Ding

Event Log



http://dig.csail.mit.edu/TAMI/2007/s9/variation1/log.n3

TAMISidebar

Enter the log file and hit "Fetch Log".

Log: http://dig.csail.mit.edu/TAMI/2007/s9/variation

Enter the policy file and hit "Fetch Policy".

Policy:

Optional: Enter the name of the file you want to save the output as (must be a URI on a WebDAV server you have access to).

Output: http://localhost/myWebDAV/output.n3

Select the your preferred reasoner, hit the "Run Reasoner", and wait 5 years for the output!

○ Scheme Reasoner
● Python Reasoner

▾http://dig.csail.mit.edu/TAMI/2007/s9/variation1/log.n3

cdc	name	CDC
	type	Organization
xphone	name	Xphone company
	type	Organization
cdc xphone search newman	instruction	cdc xphone search newman content
	reason	investigate tb newman
	receiver	xphone
	sender	cdc
	type	Query
	comment	To conduct TB investigation, CDC queried Xphone for data about Alfred Newman
cdc xphone search newman content	type	Action Instruction
	comment	find Alfred Newman's recent call logs
investigate tb newman	about	alfred newman
	category	health information
	purpose	medical emergency
	type	Information
	comment	Alfred Newman is involved in an emegent medical investigation on TB epidemic
alfred newman	name	Alfred Newman
	phonenumber	xxxxxxx423
	type	Person
cdc xphone result newman	data	xphone record 423
	receiver	cdc
	reply to	cdc xphone search newman
	sender	xphone
	type	Inform
	comment	XPhone sends Alfred Newman's recent 6 months telephone records to CDC
xphone record 423	about	customer423
	type	Information
	comment	six-month call log of using cell number xxxxxxx423
customer423	customer Of	xphone
	location	NY
	name	Alfred Newman
	phonenumber	xxxxxxx423
	type	Person
cdc xphone search same	instruction	cdc xphone search same content
	reason	investigate tb same
	receiver	xphone
	sender	cdc
	type	Query
	comment	To conduct TB investigation, CDC queried Xphone for data about Bob Same
cdc xphone search same content	type	Information

Done

Policy and Policy Language

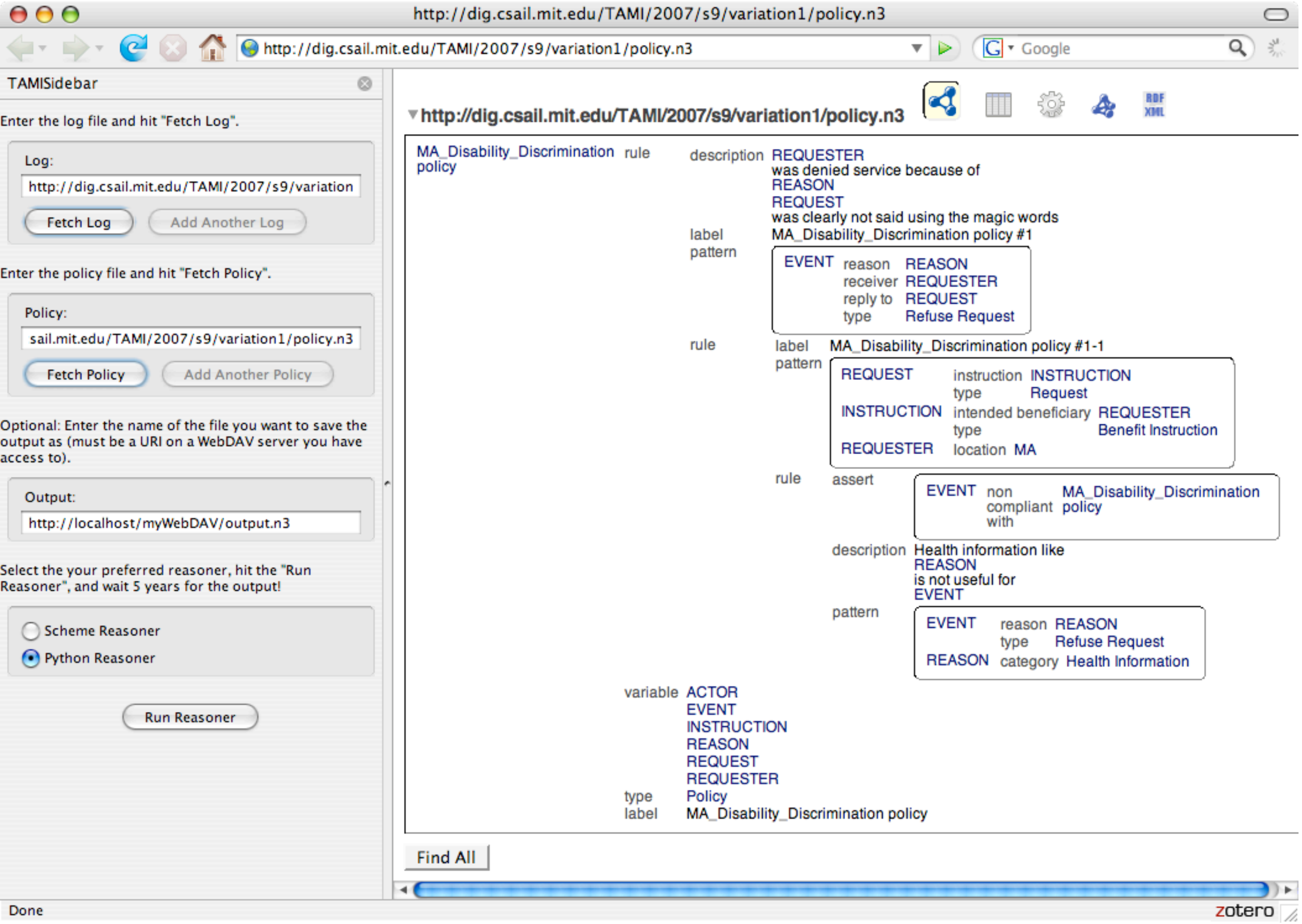
MA Disability Discrimination Policy

No otherwise qualified handicapped individual shall, solely by reason of his handicap, be excluded from participation in, be denied the benefits of, or be subject to discrimination under any program or activity within the Commonwealth

More info: http://www.mass.gov/legis/const.htm#cart114.htm

```
MA_Disability_Discrimination_Policy a air:Policy;
  air:variable :EVENT, :REQUESTER, :ACTOR, :REASON,
  :REQUEST, :INSTRUCTION;
  air:rule [
     air:pattern {
       :EVENT a tami:RefuseRequest;
          tami:reply-to :REQUEST;
          tami:receiver :REQUESTER;
          tami:reason :REASON.
     };
     air:rule [
        air:pattern {
            :REQUEST tami:instruction :INSTRUCTION;
                           a tami:Request.
             :INSTRUCTION tami:intended_beneficiary :REQUESTER;
                           a tami:BenefitInstruction.
             :REQUESTER tami:location tami:MA.
        };
        air:rule [
           air:pattern { :EVENT a tami:RefuseRequest;
                       tami:reason :REASON.
                     :REASON tami:category tami:HealthInformation };
           air:assert { :EVENT air:non-compliant-with :MA_Disability_Discrimination_Policy }
        ]
     ];
  ].
```

Accountability Reasoning



http://dig.csail.mit.edu/TAMI/2007/s9/variation1/policy.n3

TAMISidebar

Enter the log file and hit "Fetch Log".

Log: http://dig.csail.mit.edu/TAMI/2007/s9/variation

Enter the policy file and hit "Fetch Policy".

Policy: sail.mit.edu/TAMI/2007/s9/variation1/policy.n3

Optional: Enter the name of the file you want to save the output as (must be a URI on a WebDAV server you have access to).

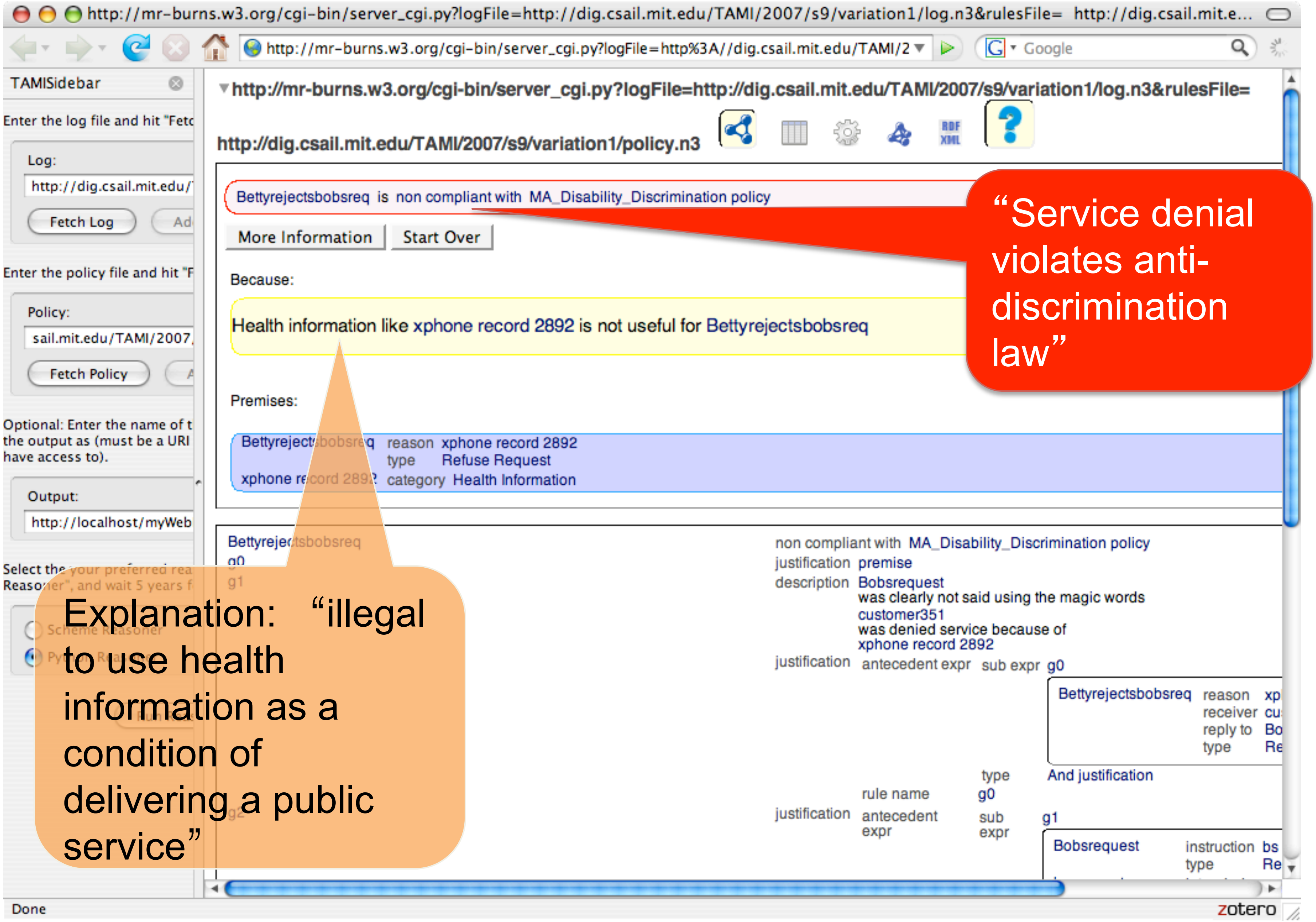
Output: http://localhost/myWebDAV/output.n3

Select the your preferred reasoner, hit the "Run Reasoner", and wait 5 years for the output!

○ Scheme Reasoner
● Python Reasoner

▾http://dig.csail.mit.edu/TAMI/2007/s9/variation1/policy.n3

```
MA_Disability_Discrimination  rule      description  REQUESTER
policy                                               was denied service because of
                                                     REASON
                                                     REQUEST
                                                     was clearly not said using the magic words
                                        label        MA_Disability_Discrimination policy #1
                                        pattern      EVENT  reason    REASON
                                                            receiver  REQUESTER
                                                            reply to  REQUEST
                                                            type      Refuse Request
                                        rule         label    MA_Disability_Discrimination policy #1-1
                                                     pattern  REQUEST      instruction  INSTRUCTION
                                                                           type         Request
                                                              INSTRUCTION  intended beneficiary  REQUESTER
                                                                           type                  Benefit Instruction
                                                              REQUESTER    location  MA
                                                     rule     assert       EVENT  non          MA_Disability_Discrimination
                                                                                  compliant    policy
                                                                                  with
                                                              description  Health information like
                                                                           REASON
                                                                           is not useful for
                                                                           EVENT
                                                              pattern      EVENT   reason    REASON
                                                                                   type      Refuse Request
                                                                           REASON  category  Health Information
                              variable  ACTOR
                                        EVENT
                                        INSTRUCTION
                                        REASON
                                        REQUEST
                                        REQUESTER
                              type      Policy
                              label     MA_Disability_Discrimination policy
```



http://mr-burns.w3.org/cgi-bin/server_cgi.py?logFile=http://dig.csail.mit.edu/TAMI/2007/s9/variation1/log.n3&rulesFile= http://dig.csail.mit.e...

http://mr-burns.w3.org/cgi-bin/server_cgi.py?logFile=http%3A//dig.csail.mit.edu/TAMI/2

TAMISidebar

Enter the log file and hit "Fetc

Log: http://dig.csail.mit.edu/

Enter the policy file and hit "F

Policy: sail.mit.edu/TAMI/2007,

Optional: Enter the name of t the output as (must be a URI have access to).

Output: http://localhost/myWeb

Select the your preferred rea Reasoner", and wait 5 years f

○ Scheme Reasoner

● Python Reasoner

http://mr-burns.w3.org/cgi-bin/server_cgi.py?logFile=http://dig.csail.mit.edu/TAMI/2007/s9/variation1/log.n3&rulesFile= http://dig.csail.mit.edu/TAMI/2007/s9/variation1/policy.n3

Bettyrejectsbobsreq is non compliant with MA_Disability_Discrimination policy

Because:

Health information like xphone record 2892 is not useful for Bettyrejectsbobsreq

Premises:

Bettyrejectsbobsreq	reason	xphone record 2892
	type	Refuse Request
xphone record 2892	category	Health Information

Bettyrejectsbobsreq non compliant with MA_Disability_Discrimination policy

g0

g1

justification premise

description Bobsrequest was clearly not said using the magic words customer351 was denied service because of xphone record 2892

justification antecedent expr sub expr g0

Bettyrejectsbobsreq reason xp receiver cu reply to Bo type Re

type And justification

rule name g0

justification antecedent expr sub expr g1

Bobsrequest instruction bs type Re

g2

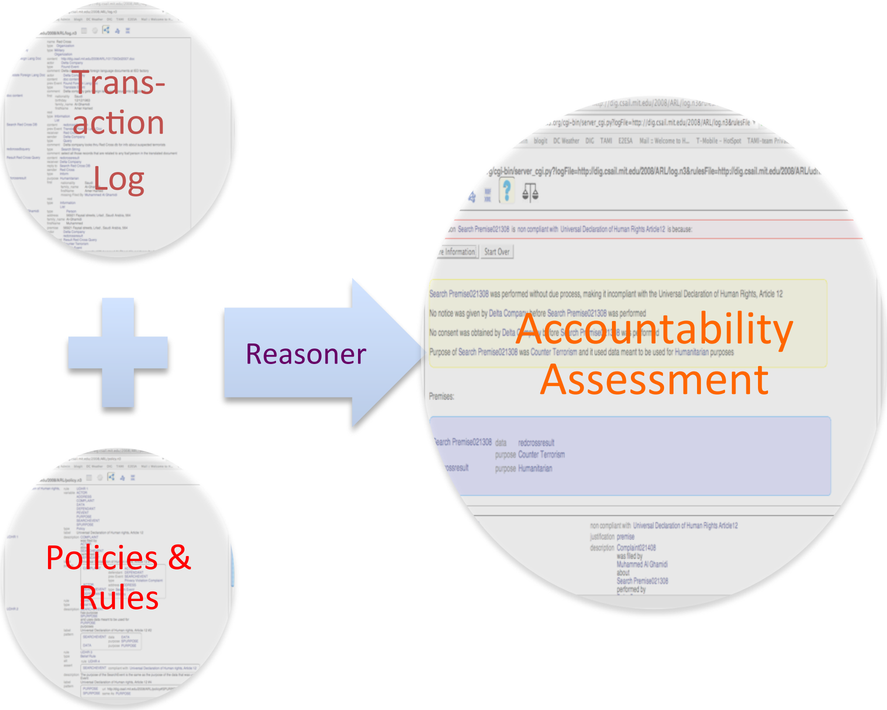
Done

“Service denial violates anti-discrimination law”

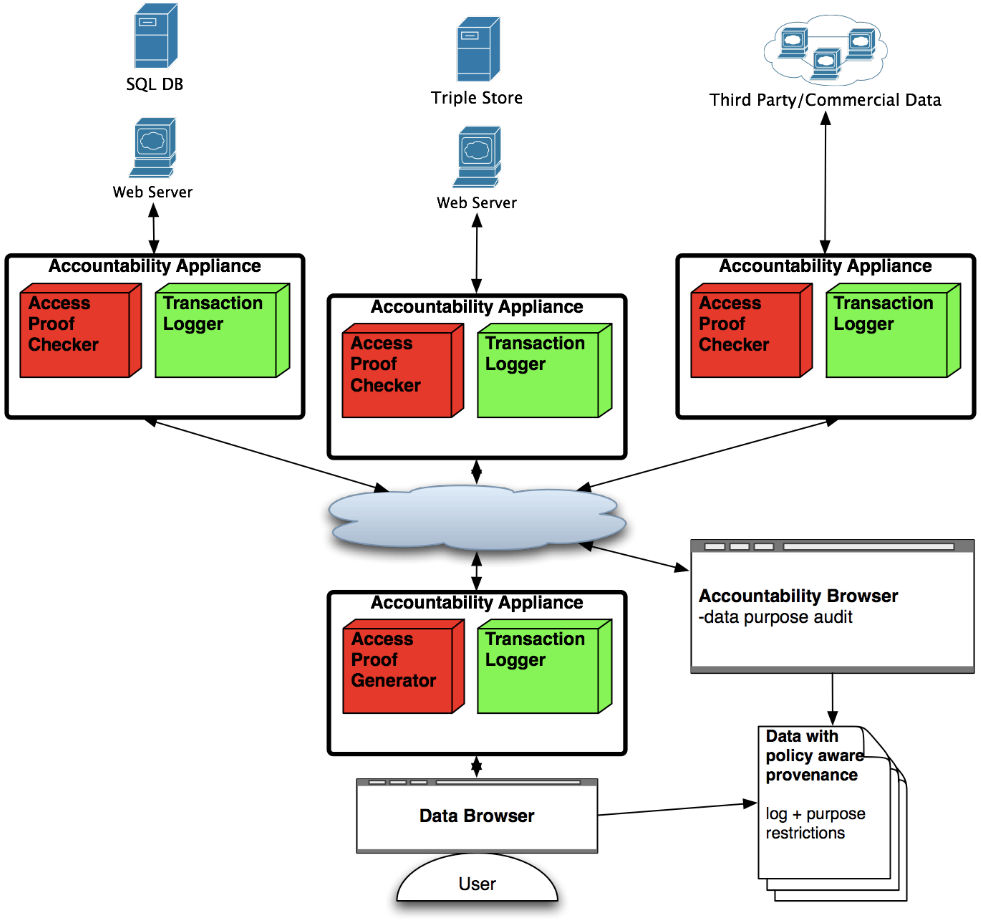
Explanation: “illegal to use health information as a condition of delivering a public service”

Properties of Accountable Systems

- Expressivity
- Evaluation of usage post-collection & analysis
- Explanation
- Support incompleteness and inconsistency



Accountability architecture



- Access control through Decentralized Authentication Proofs based on access rules expressed over data semantics
- *Transparent* data usage logging for real-time compliance hints and *a posteriori accountability*
- Engineered as Web architecture components

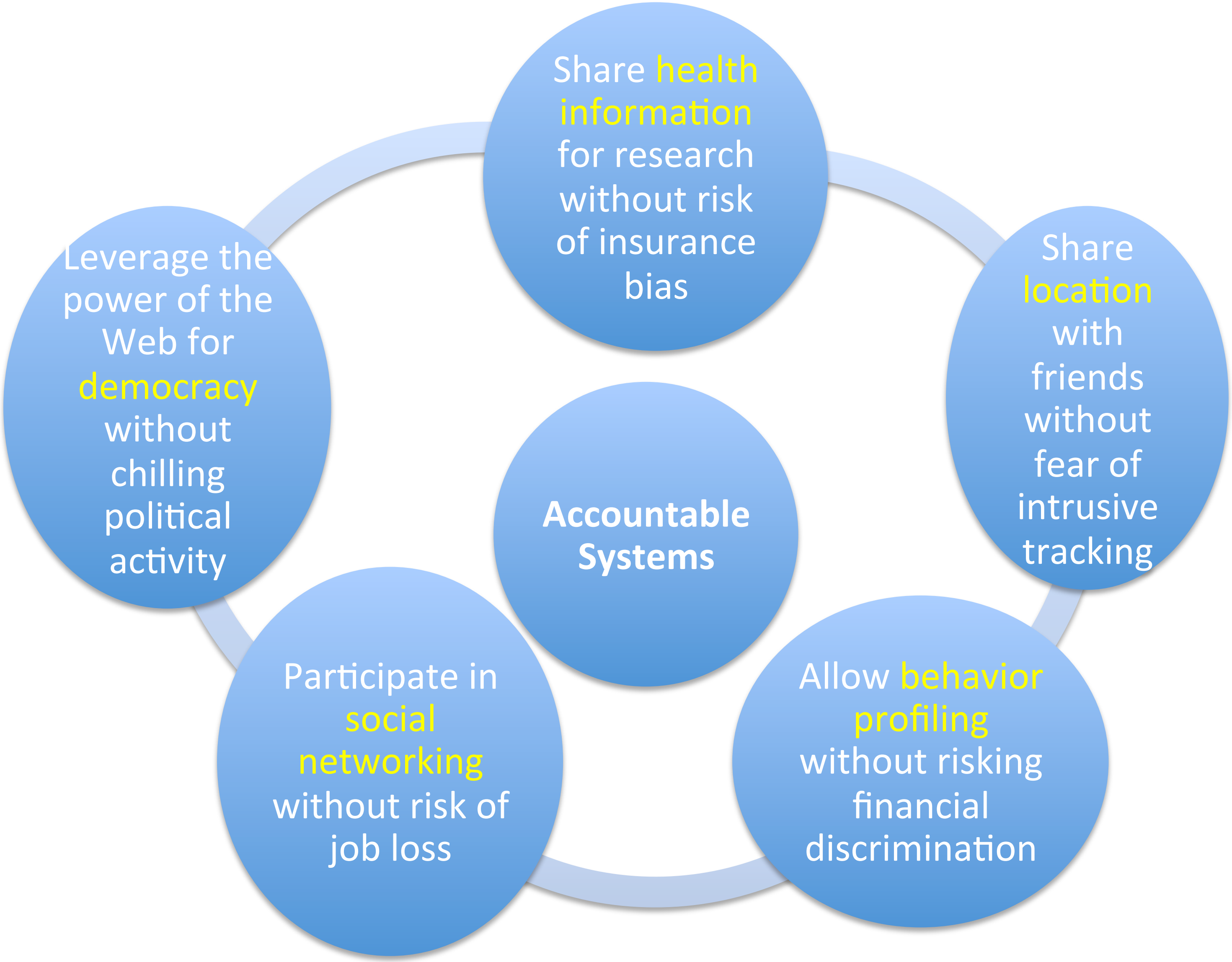
Information Accountability an as alternative to secrecy

- Rules and law should govern how information is used:

  "It is illegal to consider health status of applicant or her family in hiring decisions"

- Interactions with data are logged in order to provide possibility of machine-assisted human-driven accountability

A World of Accountable Systems



Accountable Systems

Share health information for research without risk of insurance bias

Share location with friends without fear of intrusive tracking

Allow behavior profiling without risking financial discrimination

Participate in social networking without risk of job loss

Leverage the power of the Web for democracy without chilling political activity