Privacy and Utility in Gov 2.0

Latanya Sweeney

latanya@fas.harvard.edu

latanyasweeney.org

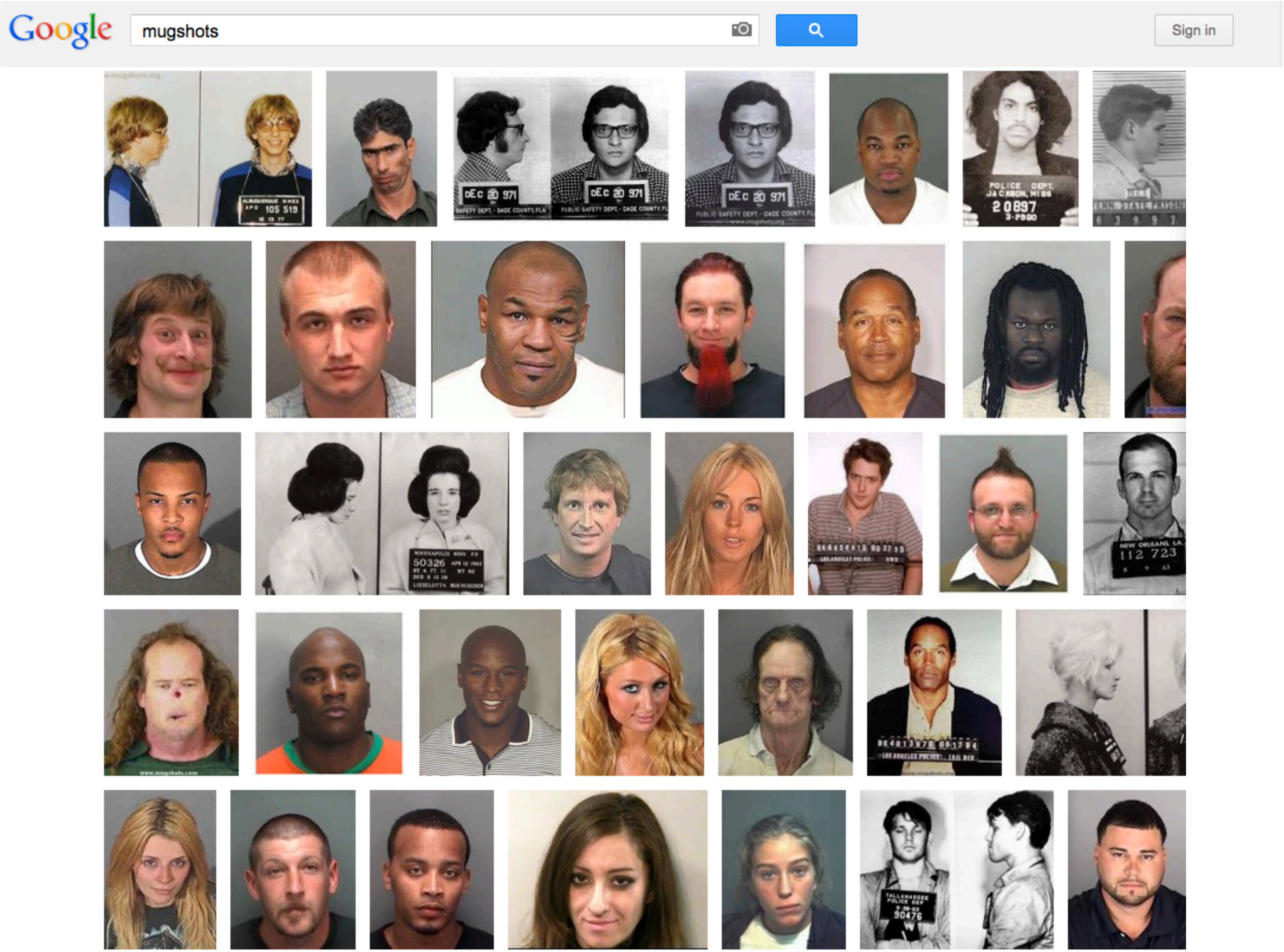
Professor Jim Waldo

Disclaimer

The views and opinions in this presentation represent my own and are not necessarily those of HHS, ONC
or the Obama Administration.
These views are for the benefit
of public discourse and public education,
and are not necessarily an opinion
regarding any position I may take
on related issues decided
by the HIT Policy Committee.

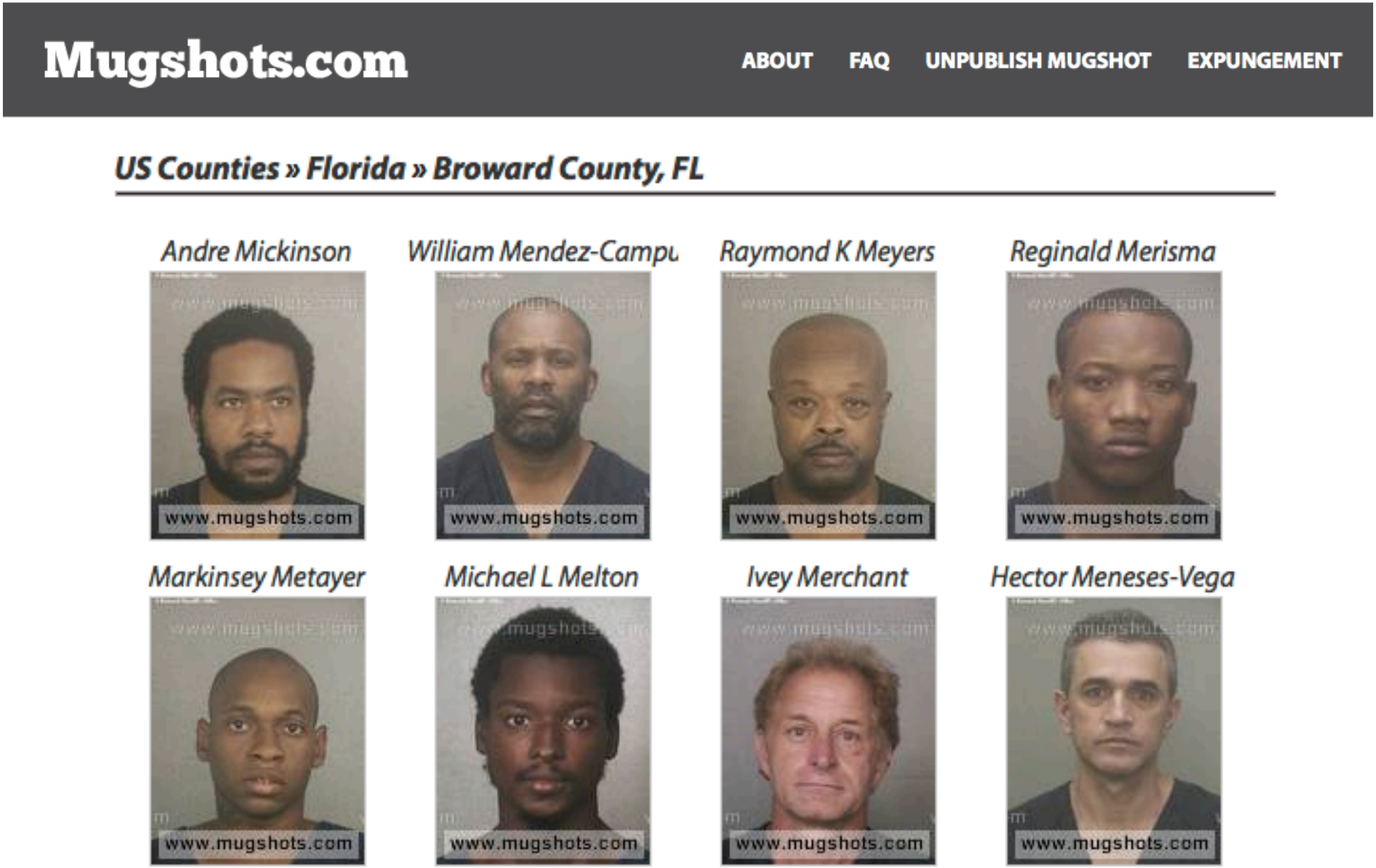
Gov 2.0

Mugshots Online



mugshots

Google
Sign in
ALBUQUERQUE N MEX
105 519
DEC 20 971
PUBLIC SAFETY DEPT. - DADE COUNTY, FLA
POLICE DEPT.
JACKSON, MISS
20897
MINNEAPOLIS MINN PD
50326
NEW ORLEANS, LA.
112 723
LOS ANGELES POLICE
TALLAHASSEE POLICE DEPT
90476



US Counties » Florida » Broward County, FL

Andre Mickinson

William Mendez-Campu

Raymond K Meyers

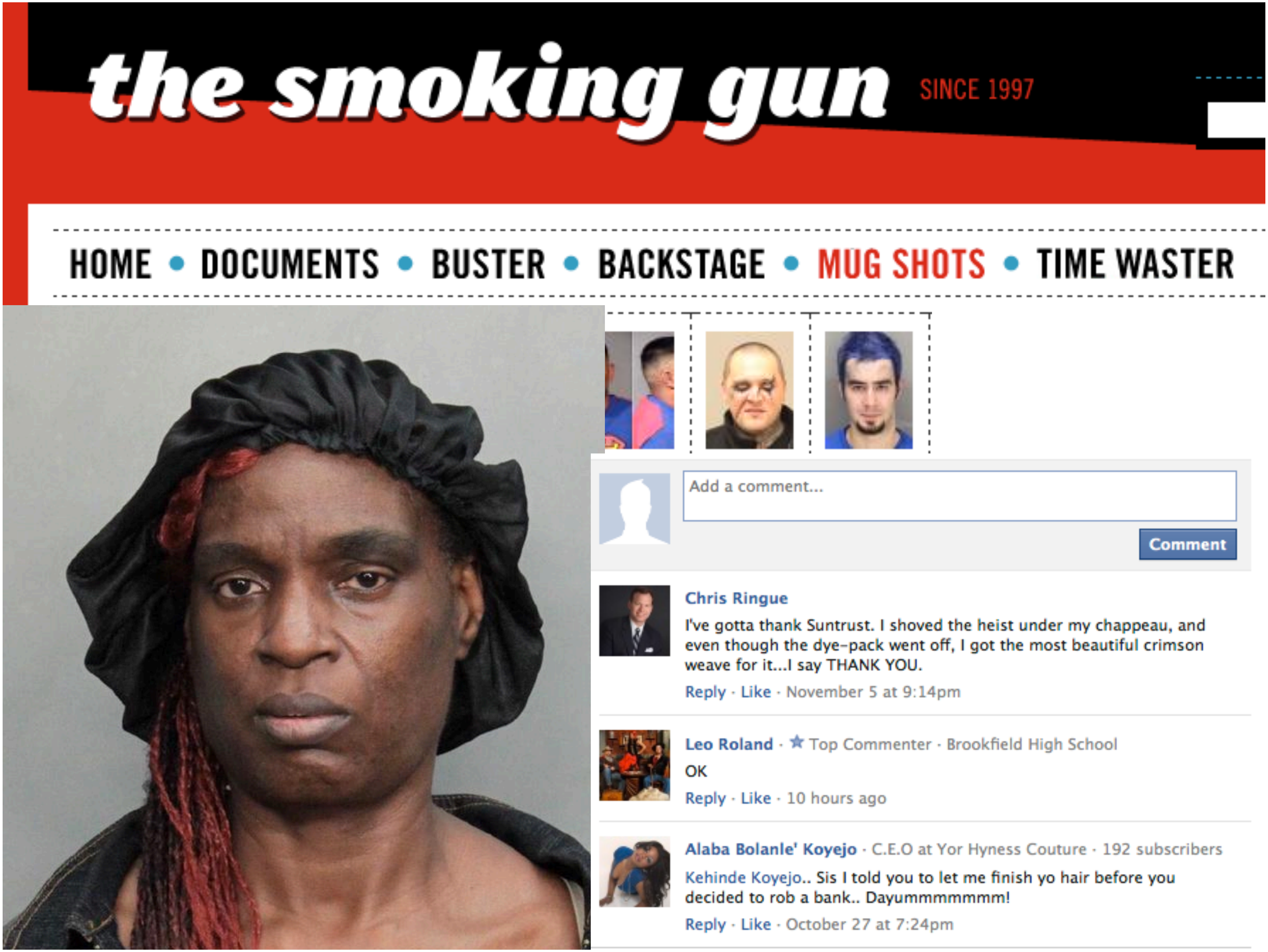
Reginald Merisma

Markinsey Metayer

Michael L Melton

Ivey Merchant

Hector Meneses-Vega


the smoking gun
SINCE 1997
HOME • DOCUMENTS • BUSTER • BACKSTAGE • MUG SHOTS • TIME WASTER
Add a comment...
Comment

Chris Ringue

I've gotta thank Suntrust. I shoved the heist under my chappeau, and even though the dye-pack went off, I got the most beautiful crimson weave for it...I say THANK YOU.

Reply · Like · November 5 at 9:14pm

Leo Roland · ★ Top Commenter · Brookfield High School

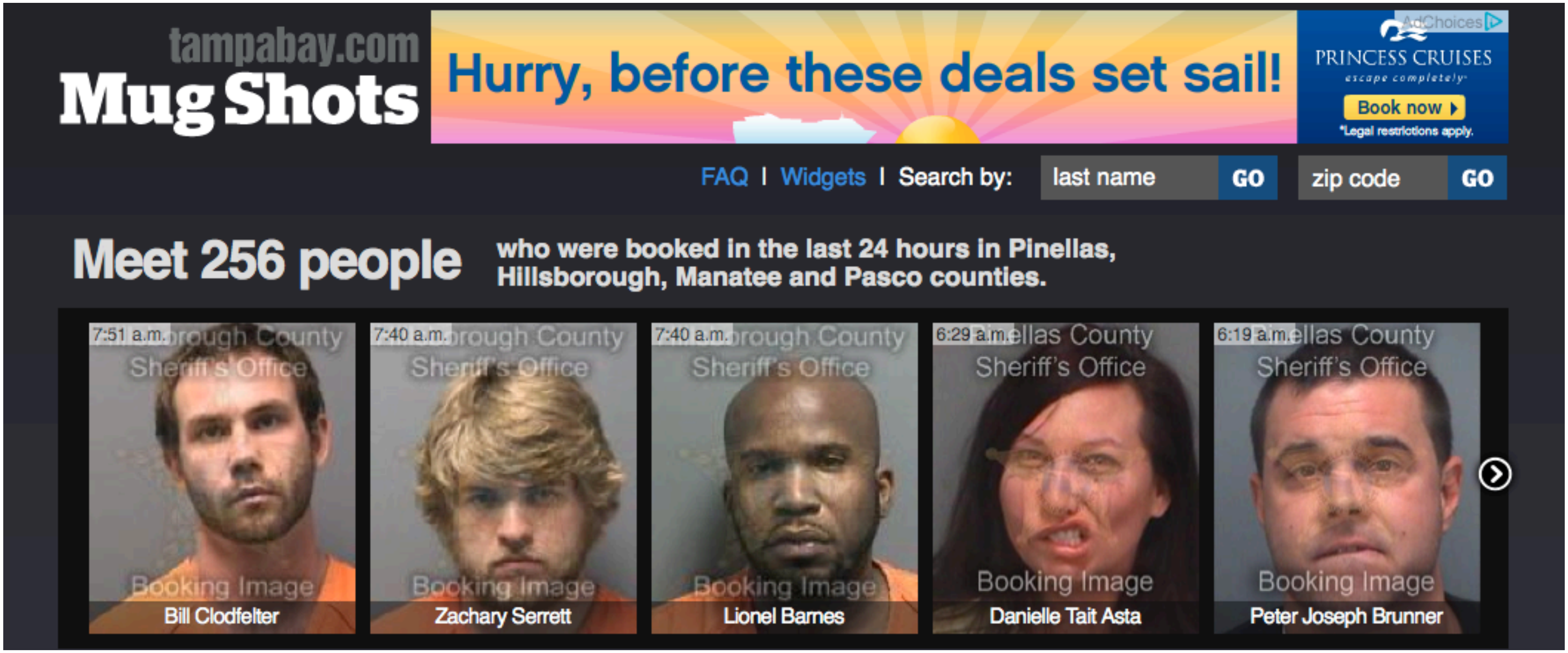
OK

Reply · Like · 10 hours ago

Alaba Bolanle' Koyejo · C.E.O at Yor Hyness Couture · 192 subscribers

Kehinde Koyejo.. Sis I told you to let me finish yo hair before you decided to rob a bank.. Dayummmmmmmm!

Reply · Like · October 27 at 7:24pm



tampabay.com

Mug Shots

FAQ | Widgets | Search by: last name GO zip code GO

Meet 256 people who were booked in the last 24 hours in Pinellas, Hillsborough, Manatee and Pasco counties.

7:51 a.m. Bill Clodfelter

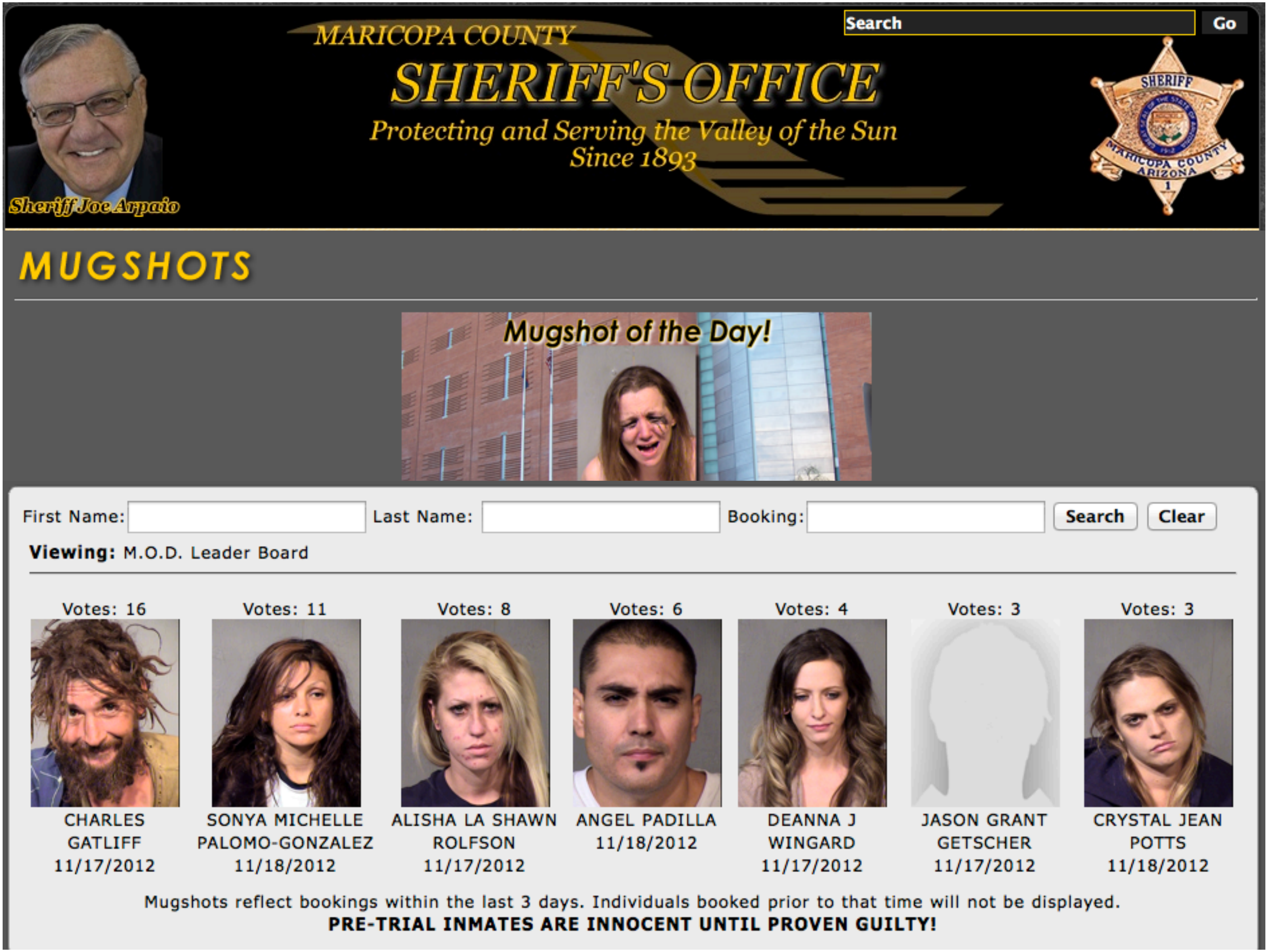
7:40 a.m. Zachary Serrett

7:40 a.m. Lionel Barnes

6:29 a.m. Danielle Tait Asta

6:19 a.m. Peter Joseph Brunner

Hillsborough County Sheriff's Office
Booking Image
Pinellas County Sheriff's Office
Booking Image



MUGSHOTS

First Name: Last Name: Booking: Search Clear

Viewing: M.O.D. Leader Board

Votes: 16	Votes: 11	Votes: 8	Votes: 6	Votes: 4	Votes: 3	Votes: 3
CHARLES GATLIFF	SONYA MICHELLE PALOMO-GONZALEZ	ALISHA LA SHAWN ROLFSON	ANGEL PADILLA	DEANNA J WINGARD	JASON GRANT GETSCHER	CRYSTAL JEAN POTTS
11/17/2012	11/18/2012	11/17/2012	11/18/2012	11/17/2012	11/17/2012	11/18/2012

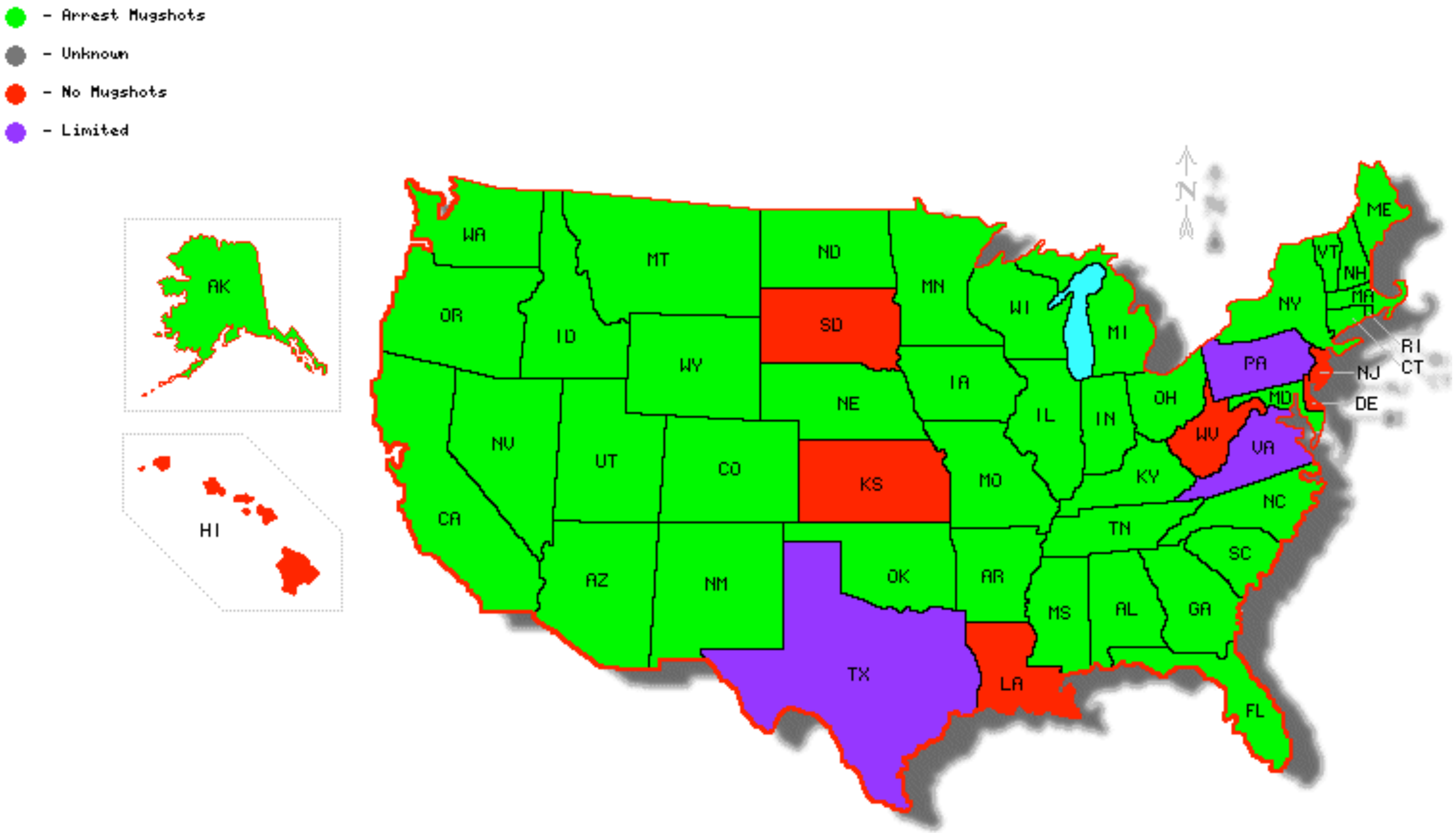
Mugshots reflect bookings within the last 3 days. Individuals booked prior to that time will not be displayed.

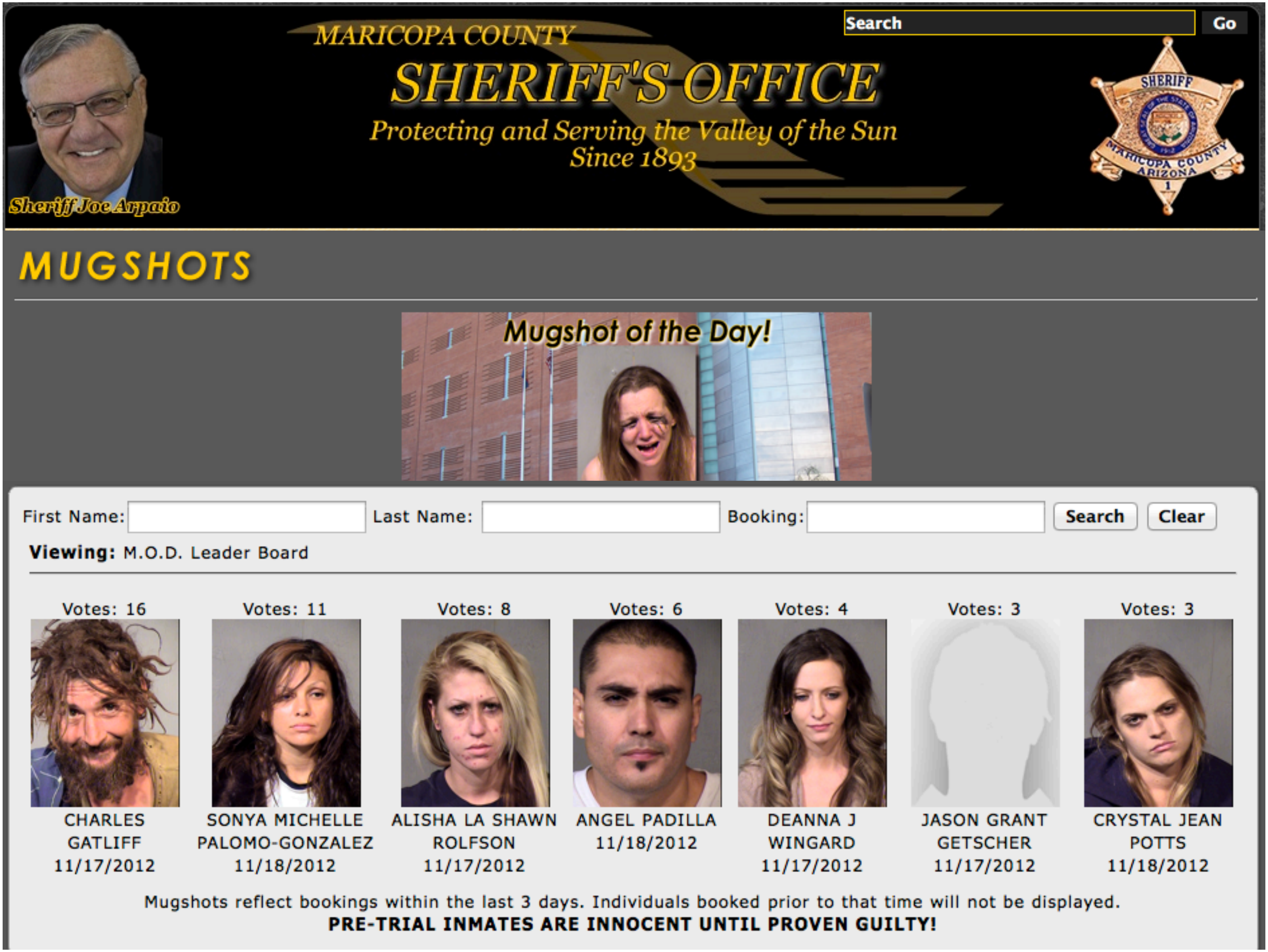
PRE-TRIAL INMATES ARE INNOCENT UNTIL PROVEN GUILTY!

Public Records

Legislative Intent

Conflicting Laws


🟢 - Arrest Mugshots
● - Unknown
🔴 - No Mugshots
🟣 - Limited
N
AK
HI
WA
MT
ND
MN
OR
ID
WY
SD
WI
MI
NY
ME
VT
NH
MA
RI
CT
NJ
DE
MD
PA
IA
NE
IL
IN
OH
NV
UT
CO
KS
MO
KY
WV
VA
CA
AZ
NM
OK
AR
TN
NC
SC
MS
AL
GA
TX
LA
FL



MARICOPA COUNTY
SHERIFF'S OFFICE
Protecting and Serving the Valley of the Sun
Since 1893

MUGSHOTS

Viewing: M.O.D. Leader Board

Votes	Name	Date
Votes: 16	CHARLES GATLIFF	11/17/2012
Votes: 11	SONYA MICHELLE PALOMO-GONZALEZ	11/18/2012
Votes: 8	ALISHA LA SHAWN ROLFSON	11/17/2012
Votes: 6	ANGEL PADILLA	11/18/2012
Votes: 4	DEANNA J WINGARD	11/17/2012
Votes: 3	JASON GRANT GETSCHER	11/17/2012
Votes: 3	CRYSTAL JEAN POTTS	11/18/2012

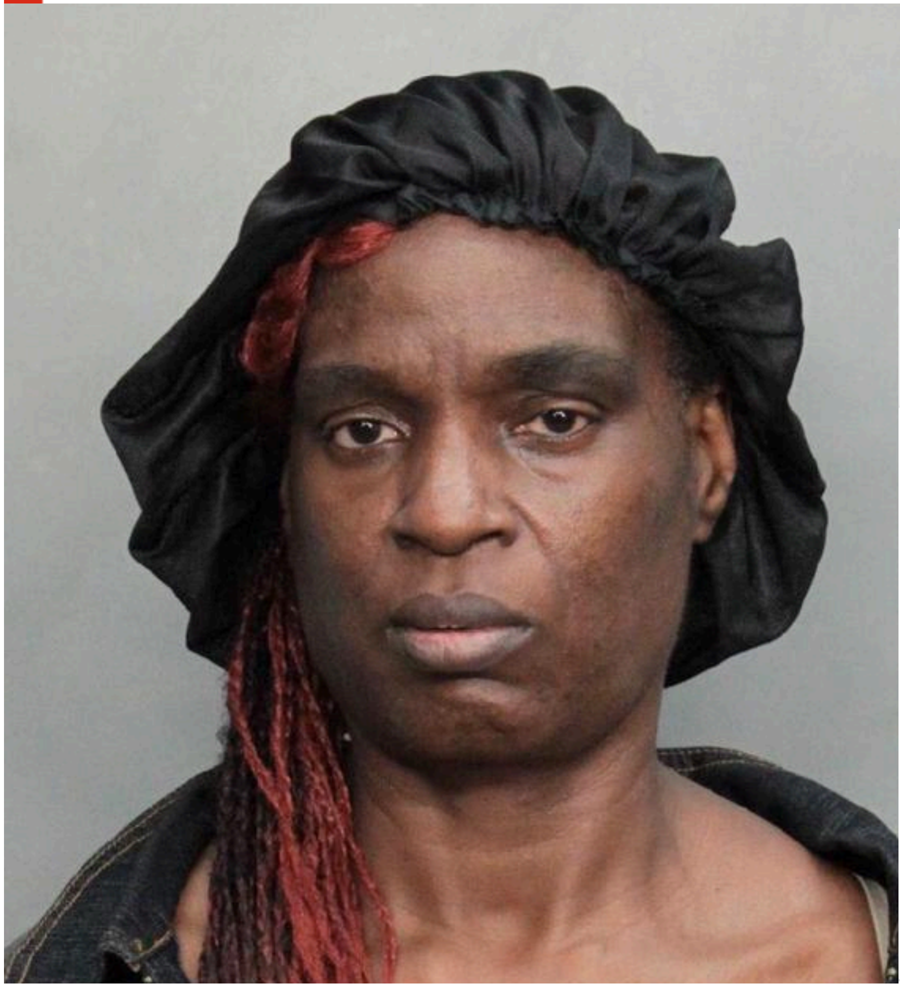
Mugshots reflect bookings within the last 3 days. Individuals booked prior to that time will not be displayed.

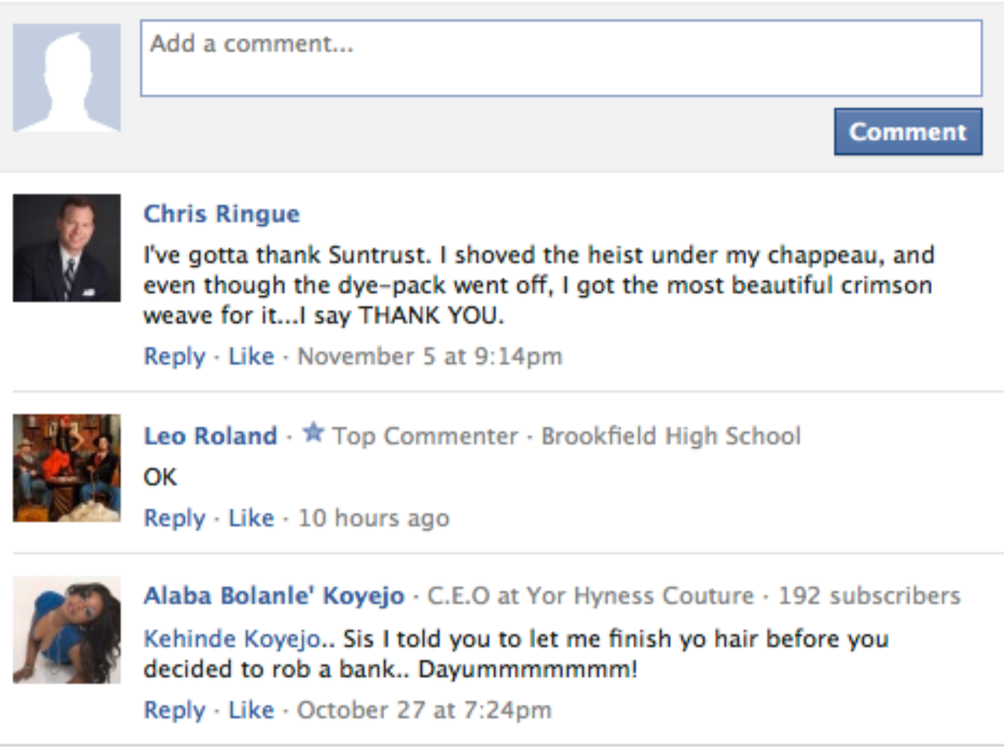
PRE-TRIAL INMATES ARE INNOCENT UNTIL PROVEN GUILTY!

the smoking gun

SINCE 1997

HOME • DOCUMENTS • BUSTER • BACKSTAGE • MUG SHOTS • TIME WASTER





Chris Ringue

I've gotta thank Suntrust. I shoved the heist under my chappeau, and even though the dye-pack went off, I got the most beautiful crimson weave for it...I say THANK YOU.

Reply · Like · November 5 at 9:14pm

Leo Roland · ★ Top Commenter · Brookfield High School

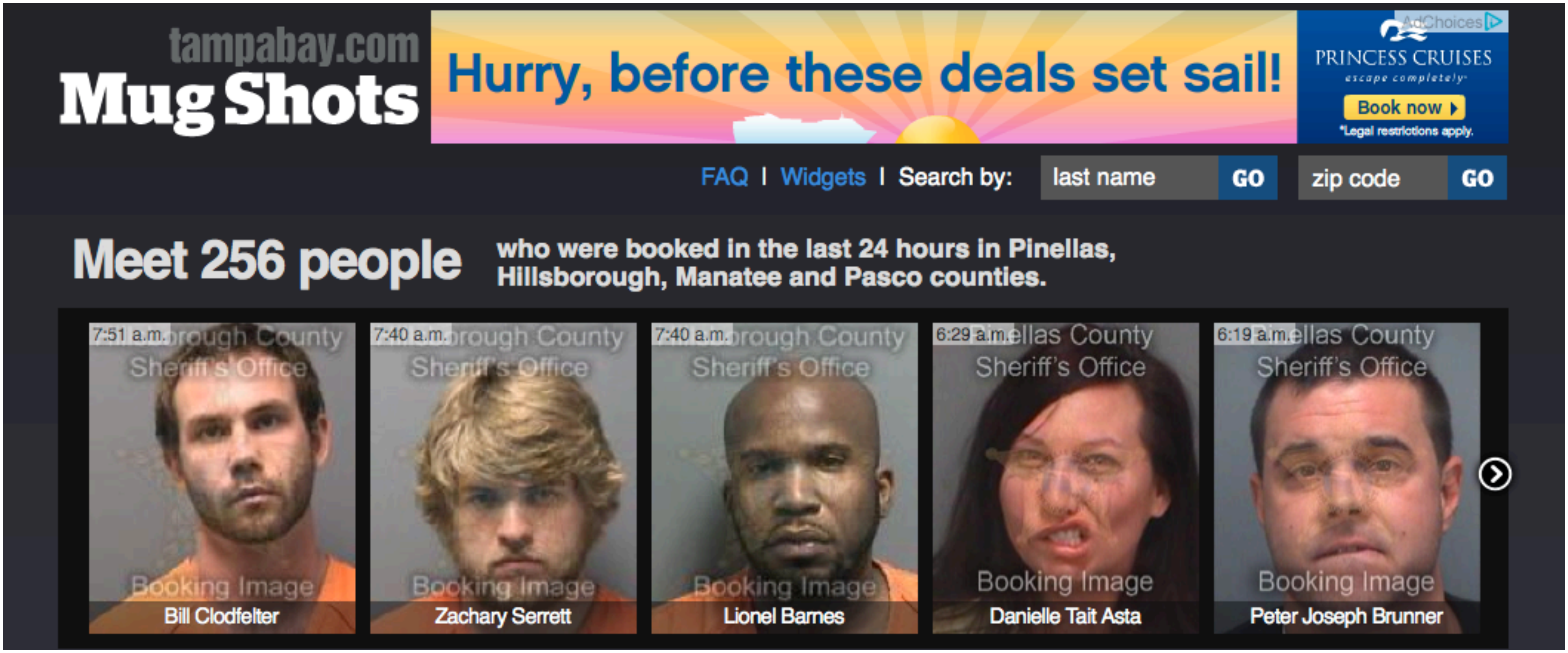
OK

Reply · Like · 10 hours ago

Alaba Bolanle' Koyejo · C.E.O at Yor Hyness Couture · 192 subscribers

Kehinde Koyejo.. Sis I told you to let me finish yo hair before you decided to rob a bank.. Dayummmmmmmm!

Reply · Like · October 27 at 7:24pm



tampabay.com

Mug Shots

Meet 256 people who were booked in the last 24 hours in Pinellas, Hillsborough, Manatee and Pasco counties.

7:51 a.m. Bill Clodfelter

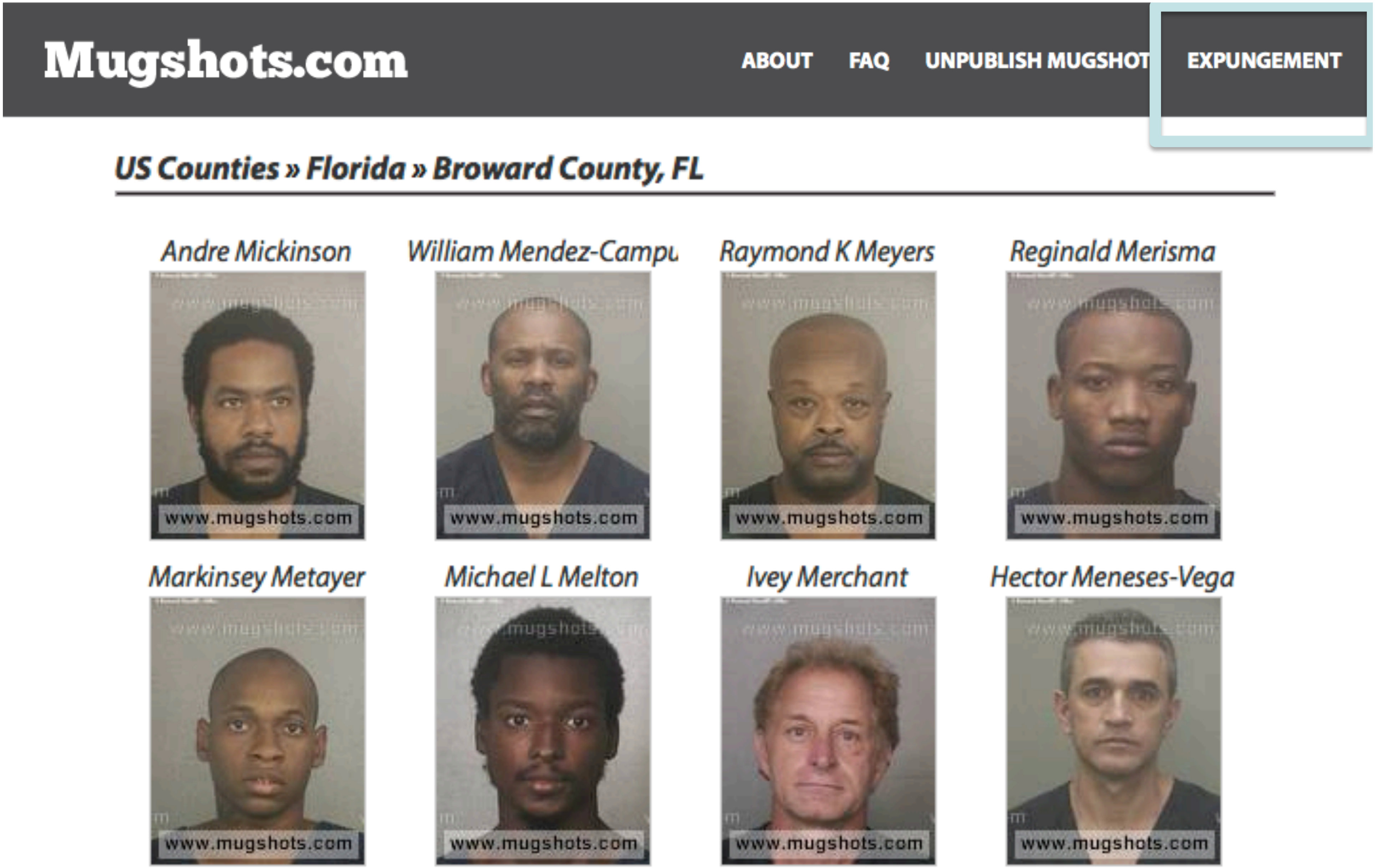
7:40 a.m. Zachary Serrett

7:40 a.m. Lionel Barnes

6:29 a.m. Danielle Tait Asta

6:19 a.m. Peter Joseph Brunner

Hillsborough County Sheriff's Office
Booking Image
Pinellas County Sheriff's Office
Booking Image



US Counties » Florida » Broward County, FL

Andre Mickinson

William Mendez-Campu

Raymond K Meyers

Reginald Merisma

Markinsey Metayer

Michael L Melton

Ivey Merchant

Hector Meneses-Vega

Not Free Speech

Extortion

Technology as Policy Maker

Law

Market forces

Social norms

Technology Design Code for architecture

Lawrence Lessig. *Codev2*. http://codev2.cc/. Earlier version: *Code and Other Laws of Cyberspace*

Law

Technology Design

Market forces

Technology Design

Social norms

Technology Design

Social Security Numbers

123-45-6789

What does it tell you?

123-45-6789

State of Issuance

001-003 New Hampshire
004-007 Maine
008-009 Vermont
010-034 Massachusetts
035-039 Rhode Island
040-049 Connecticut
050-134 New York
135-158 New Jersey
159-211 Pennsylvania
212-220 Maryland
221-222 Delaware
223-231 Virginia

232-236 West Virginia
232 North Carolina
237-246
681-690
247-251 South Carolina
654-658
252-260 Georgia
667-675
261-267 Florida
589-595
766-772
268-302 Ohio

Source: Social Security Administration, http://www.ssa.gov/foia/stateweb.html

123-45-6789

State of Issuance

303-317	Indiana
318-361	Illinois
362-386	Michigan
387-399	Wisconsin
400-407	Kentucky
408-415	Tennessee
756-763*	
416-424	Alabama
425-428	Mississippi
587-588	
752-755*	
429-432	Arkansas
433-439	Louisiana
659-665	
440-448	Oklahoma
449-467	Texas
627-645	
468-477	Minnesota
478-485	Iowa
486-500	Missouri
501-502	North Dakota
503-504	South Dakota
505-508	Nebraska
509-515	Kansas

123-45-6789

State of Issuance

516-517 Montana
518-519 Idaho
520 Wyoming
521-524 Colorado
650-653
525,585 New Mexico
648-649
526-527 Arizona
600-601
764-765
528-529 Utah
646-647

530 Nevada
680
531-539 Washington
540-544 Oregon
545-573 California
602-626
574 Alaska
575-576 Hawaii
750-751*
577-579 District of Columbia
580 Virgin Islands

123-45-6789

State of Issuance

580-584 Puerto Rico
596-599
586 Guam
586 American Samoa
586 Philippine Islands
676-679
691-699*
700-728 Railroad Board**

* Some states may share the same area by transfer or split.

** Railroad employees, discontinued July 1, 1963.

000 will NEVER start a valid SSN.

123-45-6789
Order of Issuance

“Group numbers” assigned in this order:

ODD - 01, 03, 05, 07, 09
EVEN - 10 to 98

After 98:

EVEN - 02, 04, 06, 08
ODD - 11 to 99

123-45-6789

Order of Issuance

The Social Security Administration (SSA) publishes the highest group number assigned per area.

001	98	**002**	98	**003**	96	**004**
007	02	**008**	86	**009**	86	**010**
013	86	**014**	86	**015**	86	**016**
019	86	**020**	86	**021**	86	**022**
025	86	**026**	86	**027**	86	**028**
031	84	**032**	84	**033**	84	**034**
037	68	**038**	68	**039**	68	**040**

Source: http://www.ssa.gov/foia/highgroup.htm (Sample 9/2/2003)

123-45-6789

Order of Issuance

The Social Security Administration (SSA) publishes the highest group number assigned per area.

001	98	**002**	98	**003**	96	**004**
007	02	**008**	86	**009**	80	**010**
013	86					
019	86					
025	86					
031	84					
037	68					

For area 003 (the first 3 digits of an SSN), the highest number used in the 4th and 5th digits is 96.

Source: http://www.ssa.gov/foia/highgroup.htm (Sample 9/2/2003)

123-**45**-6789
Order of Issuance

ODD: 01, 03, 05, 07, 09 then EVEN: 10 to 98

After 98:

EVEN - 02, 04, 06, 08 then ODD - 11 to 99

001	98	**002**	98	**003**	96	**004**
007	02	**008**	86	**009**	86	**010**
013	86	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
019	86	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
025	86	**026**	86	**027**	86	**028**
031	84	**032**	84	**033**	84	**034**
037	68	**038**	68	**039**	68	**040**

003-09-1234 valid SSN?

003-02-1234 valid SSN?

Source: http://www.ssa.gov/foia/highgroup.htm (Sample 9/2/2003)

123-45-6789

Sequential

123-45-6789

What does it tell you?

SSNwatch

On-line SSN validation system. Given the first 3 or 5 digits of an SSN, returns the state in which the SSN was issued along with an estimated age range of the person.

Sample uses:

Job Applications
Apartment Rentals
Insurance Claims
Student Applications

http://privacy.cs.cmu.edu/dataprivacy/projects/ssnwatch/index.html

SSNwatch 078-05-

Geography	**New York**
Date of issuance	Issued before 1993
Year of Birth (5-digit prefix)	64% born 1889 to 1910 **98% born 1879 to 1921**

If the person presenting the SSN is about age 20, then it is extremely unlikely that the provided SSN was issued to that person.

SSNwatch 078-05-

Geography	New York
Date of issuance	Issued before 1993
Year of Birth (5 -digit prefix)	64% born 1889 to 1910 **98% born 1879 to 1921**

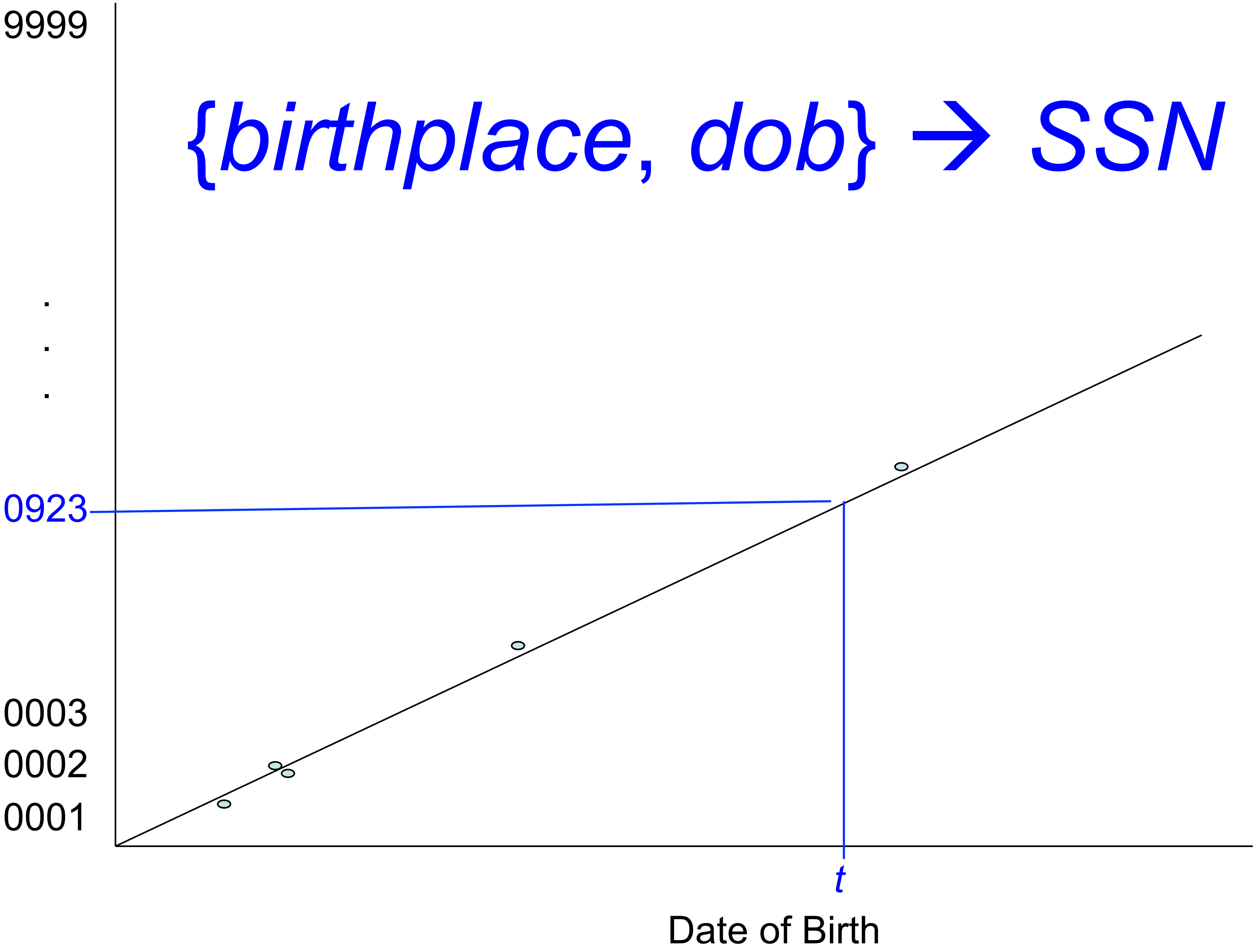
If the person presenting the SSN fails to list or acknowledge New York as a prior residence, then it is extremely unlikely that the provided SSN was issued to that person.

Predicting 6 to 9 digits of a young persons SSN.

Acquisti, A. and Gross, R. Predicting Social Security numbers from public data. Proceedings of the National Academy of Science. Jan 2009. http://www.pnas.org/content/106/27/10975.abstract

078-05-

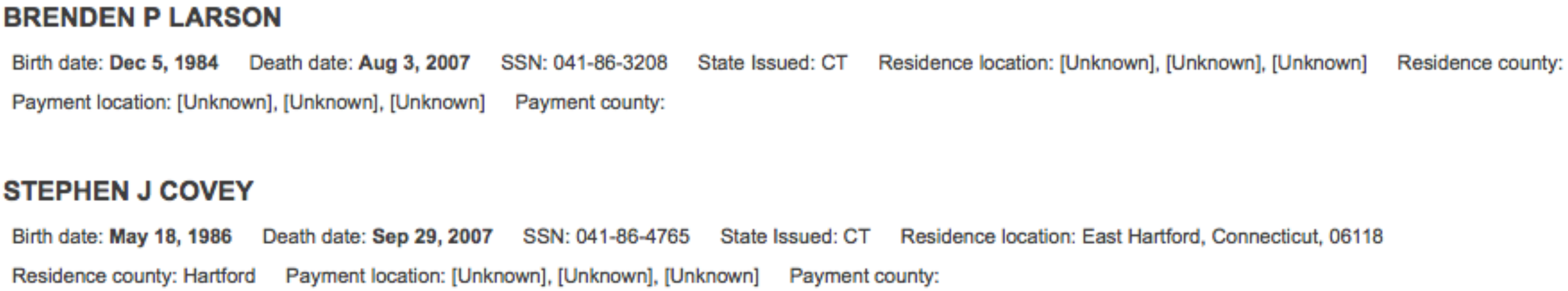
$\{\mathit{birthplace}, \mathit{dob}\} \rightarrow \mathit{SSN}$



When issued?

041-86-8892

Social Security Death Index



BRENDEN P LARSON

Birth date: **Dec 5, 1984** Death date: **Aug 3, 2007** SSN: 041-86-3208 State Issued: CT Residence location: [Unknown], [Unknown], [Unknown] Residence county: Payment location: [Unknown], [Unknown], [Unknown] Payment county:

STEPHEN J COVEY

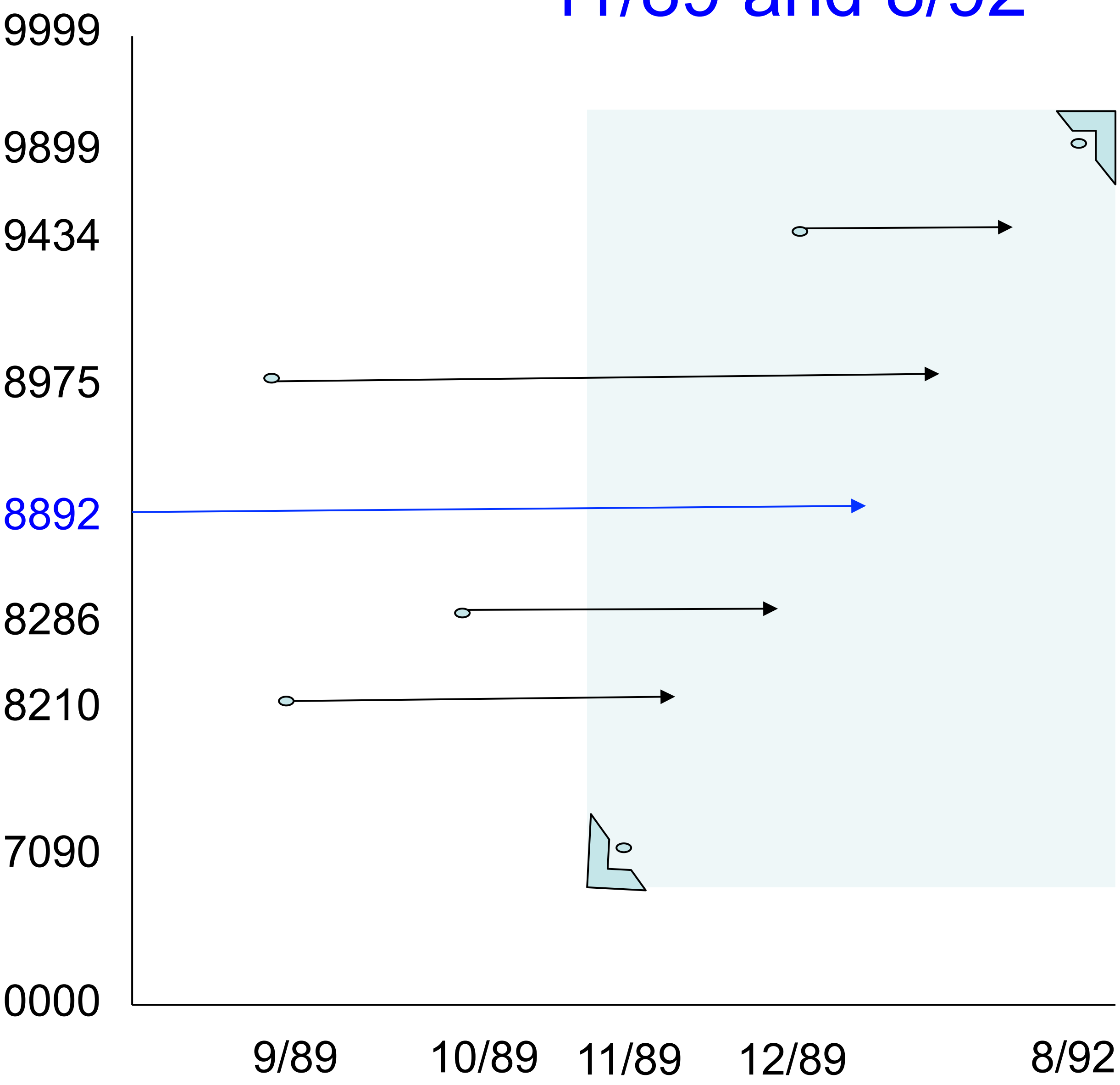
Birth date: **May 18, 1986** Death date: **Sep 29, 2007** SSN: 041-86-4765 State Issued: CT Residence location: East Hartford, Connecticut, 06118 Residence county: Hartford Payment location: [Unknown], [Unknown], [Unknown] Payment county:

http://www.americanancestors.org/search.aspx?Ca=0344&Da=269&Co=191&Re=04186*&Run=1

Area	Group	**Serial**	DOB	DOD
41	86	9936	9-Dec-79	19-May-06
41	86	9899	27-Jun-28	15-Aug-92
41	86	9453	1-Oct-73	21-Feb-93
41	86	9434	3-Dec-89	12-Jan-07
41	86	9378	27-Jul-36	6-Feb-09
41	86	9368	4-Dec-51	18-Feb-10
41	86	9349	16-Dec-13	14-Oct-00
41	86	9342	18-Aug-87	4-Jun-06
41	86	9204	12-Sep-86	19-Aug-02
41	86	9128	27-Nov-47	8-Oct-08
41	86	9000	12-Dec-20	21-Oct-96
41	86	8975	18-Sep-89	2-Apr-01
41	86	8907	23-Jul-32	9-Sep-10
41	86	8899	1-May-07	11-May-98
41	86	8892	???	???
41	86	8770	6-Aug-84	18-May-07
41	86	8289	27-Mar-87	20-Feb-09
41	86	8286	10-Oct-89	4-Mar-07
41	86	8274	9-Jan-78	6-Feb-00
41	86	8210	28-Sep-89	21-Aug-09
41	86	7939	26-Aug-63	23-Aug-06
41	86	7913	15-Jun-84	2-Dec-99
41	86	7725	11-Oct-70	21-Apr-05
41	86	7715	17-Feb-33	9-Jun-07
41	86	7701	31-Jul-81	21-Jan-10
41	86	7543	9-Jan-85	21-Oct-94
41	86	7407	23-Nov-87	11-Jul-07
41	86	7090	30-Nov-89	13-Jun-09
41	86	7015	23-Aug-23	4-Sep-99
41	86	6822	19-Jan-38	24-Apr-08

041-86-

Issued between 11/89 and 8/92

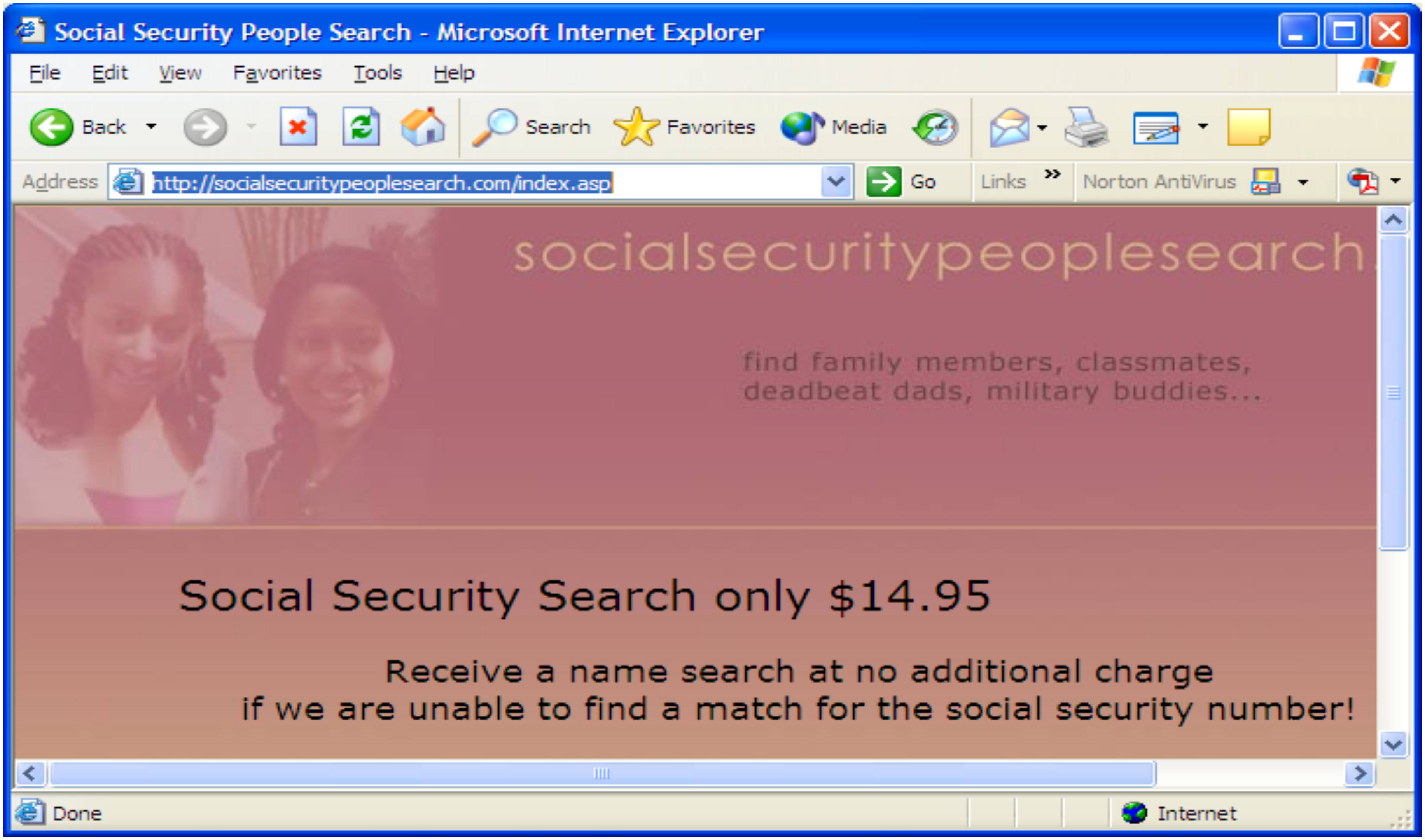


123-45-6789

What is a better way to assign SSNs?

123-45-6789

How can you get someone’s?


Social Security People Search - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address
http://socialsecuritypeoplesearch.com/index.asp
Go
Links
Norton AntiVirus
Done
Internet

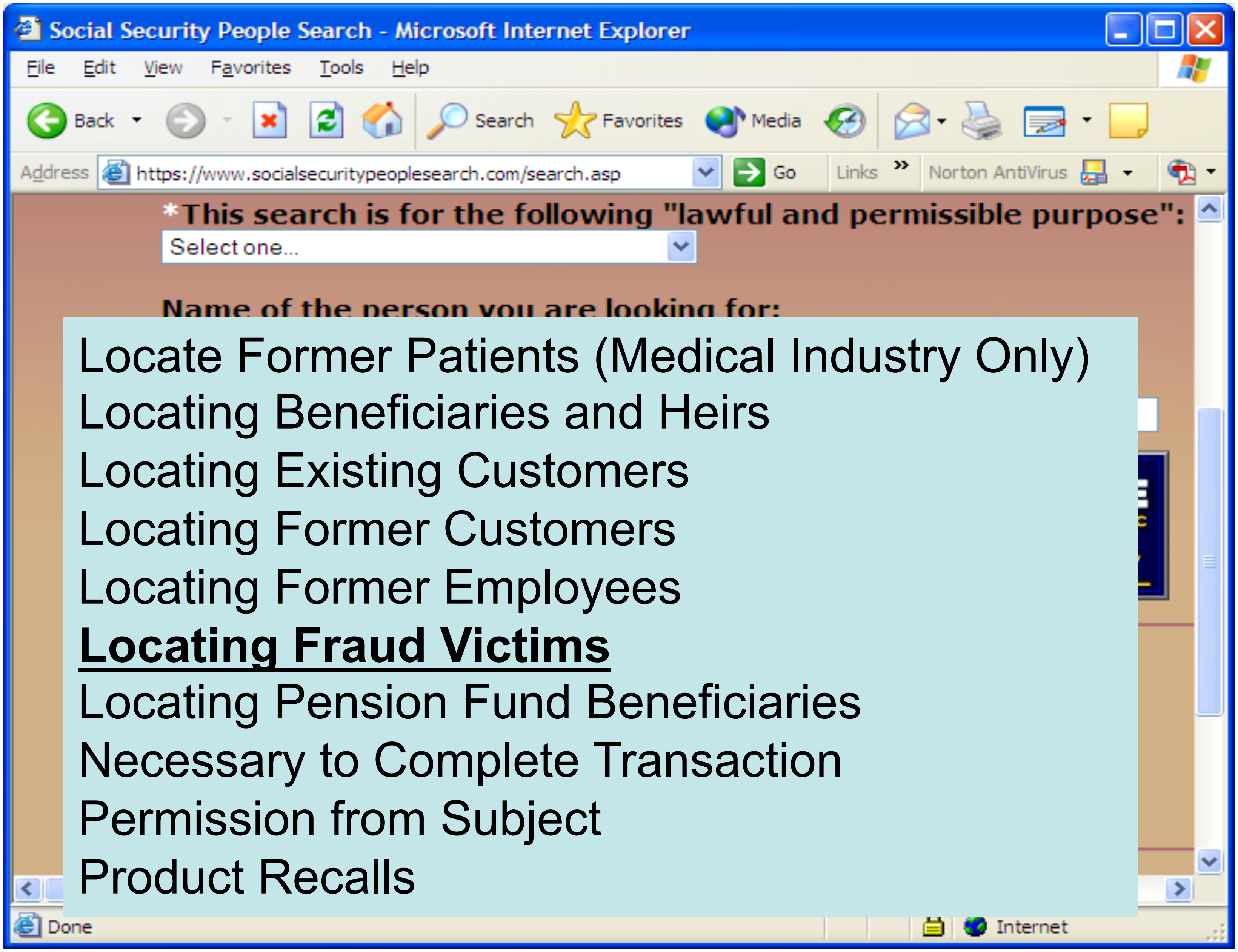
socialsecuritypeoplesearch.

find family members, classmates, deadbeat dads, military buddies...

Social Security Search only $14.95

Receive a name search at no additional charge
if we are unable to find a match for the social security number!

http://socialsecuritypeoplesearch.com/index.asp



https://www.socialsecuritypeoplesearch.com/search.asp

*This search is for the following "lawful and permissible purpose":

Select one...

Name of the person you are looking for:

Locate Former Patients (Medical Industry Only)
Locating Beneficiaries and Heirs
Locating Existing Customers
Locating Former Customers
Locating Former Employees
Locating Fraud Victims
Locating Pension Fund Beneficiaries
Necessary to Complete Transaction
Permission from Subject
Product Recalls

123-45-6789

What evil can others do?

Identity Theft

Federal Trade Commission Report: Victim Complaint Data

IDENTITY THEFT Data Clearinghouse

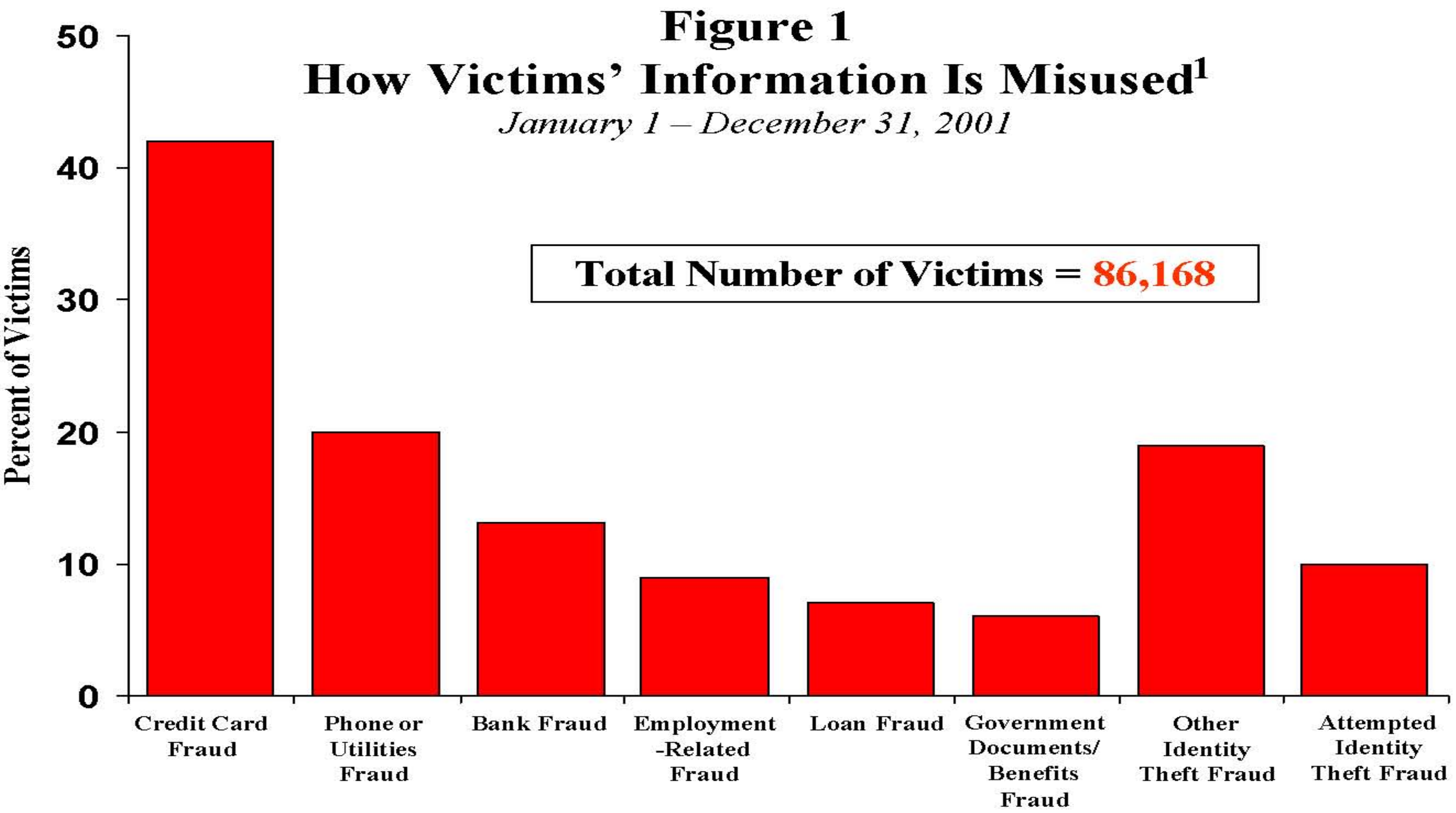


Figure 1

How Victims' Information Is Misused[1]

January 1 – December 31, 2001

Total Number of Victims = 86,168

[1] Approximately 20% of the 86,168 victim complaints received from all sources (FTC Hotline and Internet complaint form, SSA -OIG Fraud Hotline referrals, and a small number from other law enforcement organizations) reported experiencing more than one type of identity theft.

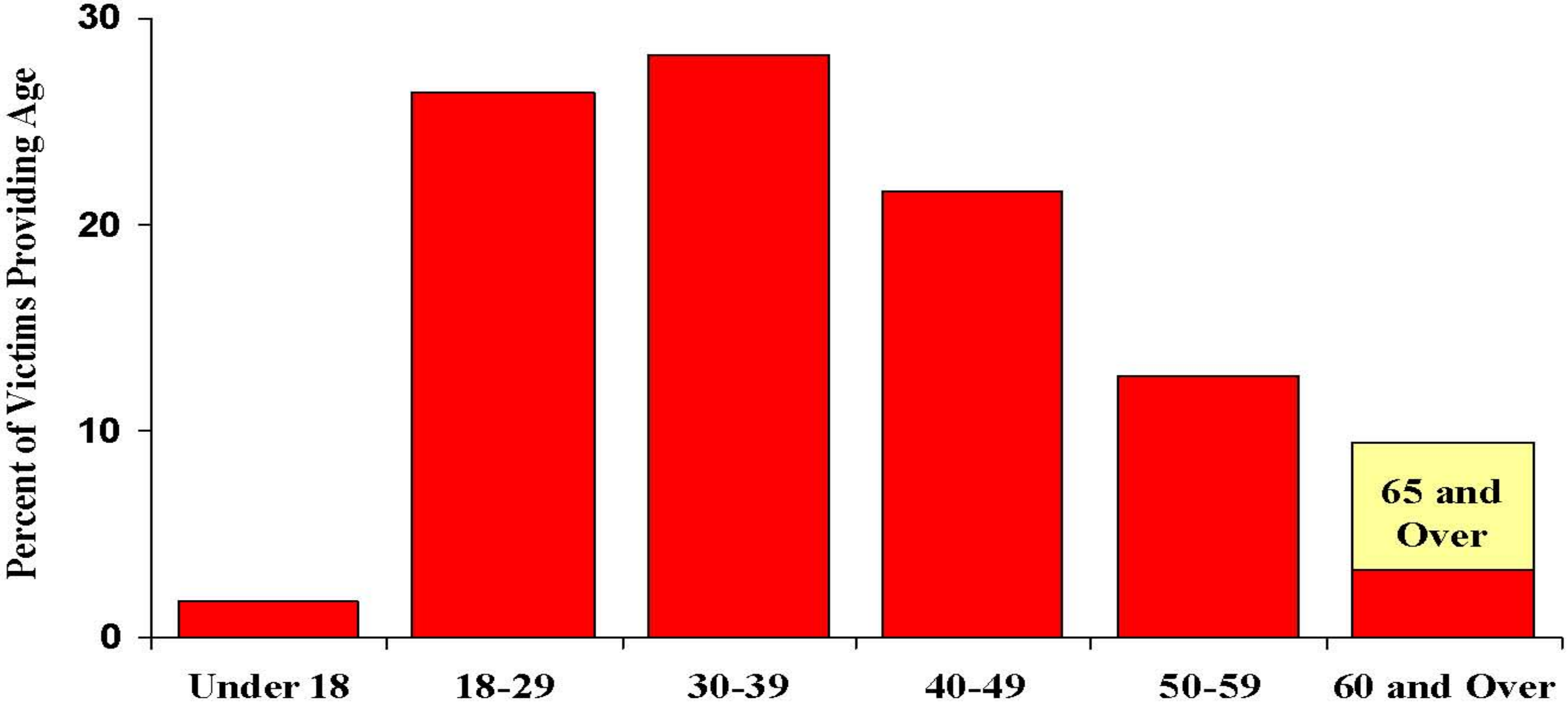
Federal Trade Commission
Created May 15, 2002

Federal Trade Commission Report: Victim Complaint Data

IDENTITY THEFT
Data Clearinghouse

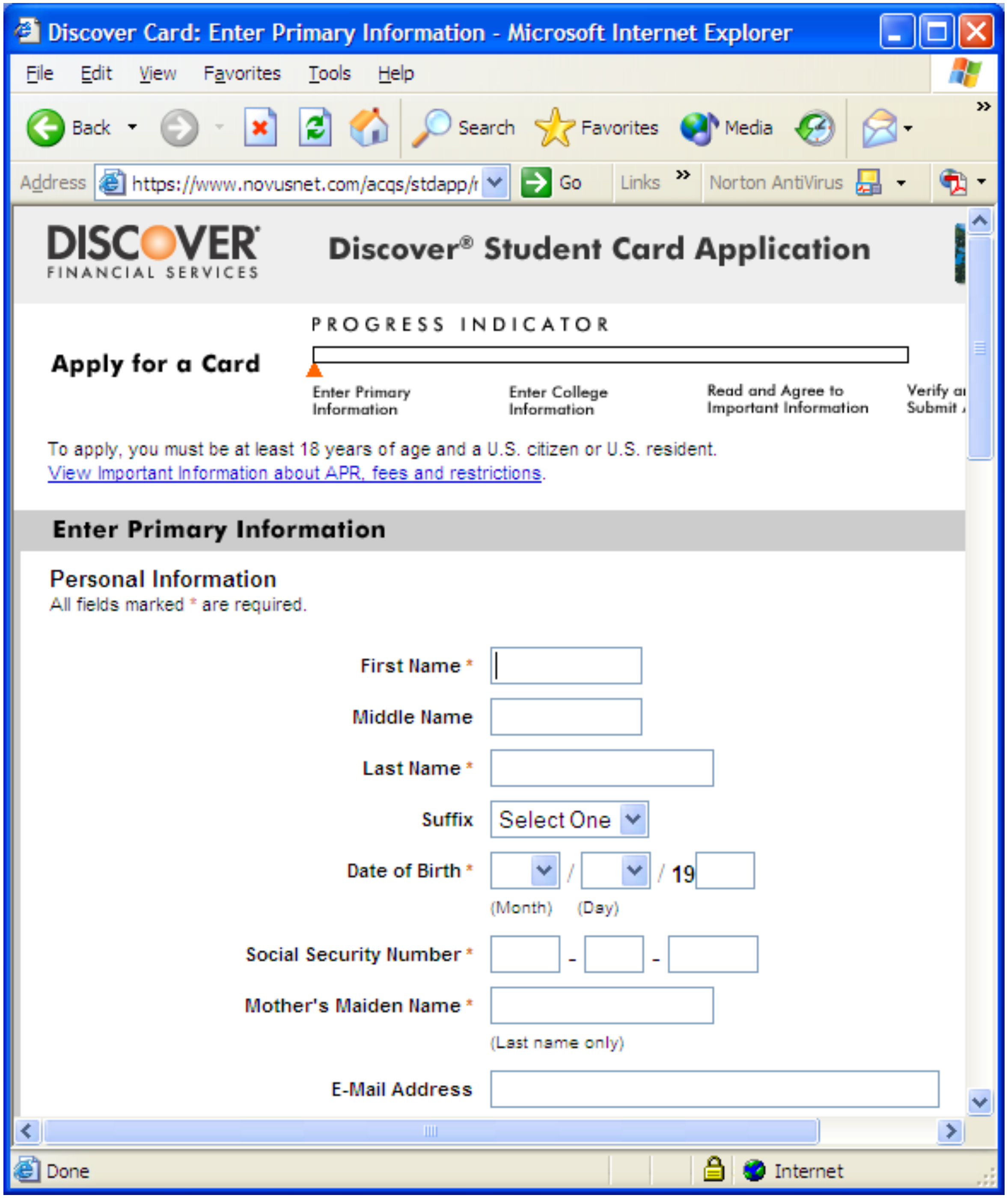
Figure 3
Victim Age Distribution[1]

January 1 – December 31, 2001



[1] This chart is based on reports from victims who contacted the FTC directly (70,540 victims) because external data contributors generally do not provide this information. 88% (61,956) of all victims reporting directly to the FTC provided their age.

Federal Trade Commission
Created May 15, 2002


Discover Card: Enter Primary Information - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Media
Address
https://www.novusnet.com/acqs/stdapp/r
Go
Links
Norton AntiVirus
Done
Internet

DISCOVER
FINANCIAL SERVICES

Discover® Student Card Application

PROGRESS INDICATOR

Apply for a Card

Enter Primary Information — Enter College Information — Read and Agree to Important Information — Verify a Submit

To apply, you must be at least 18 years of age and a U.S. citizen or U.S. resident.
View Important Information about APR, fees and restrictions.

Enter Primary Information

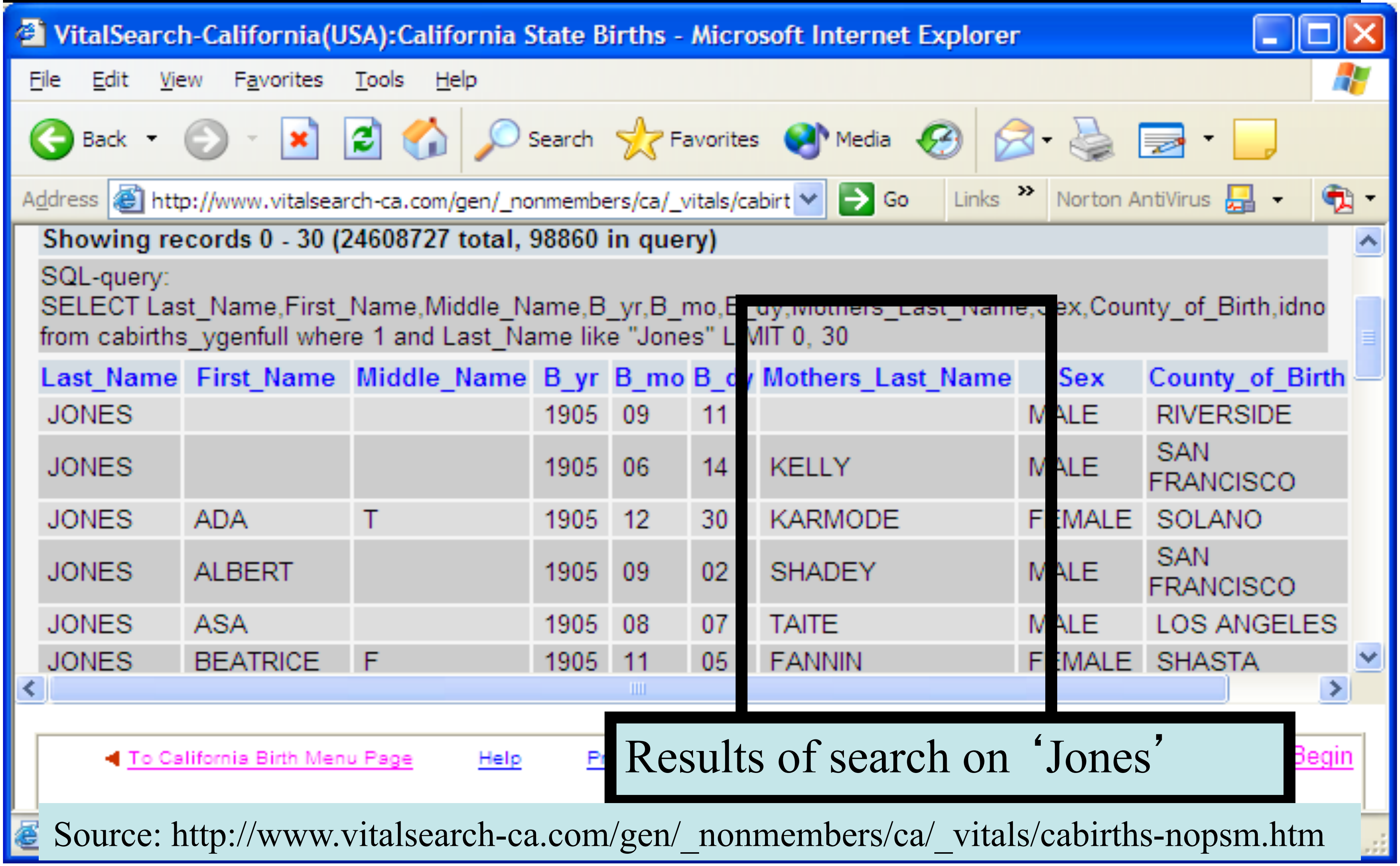
Personal Information
All fields marked * are required.

First Name *	
Middle Name	
Last Name *	
Suffix	Select One
Date of Birth *	/ / 19 (Month) (Day)
Social Security Number *	- -
Mother's Maiden Name *	(Last name only)
E-Mail Address	

Credit Card Application

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

California on-line Birth Records


VitalSearch-California(USA):California State Births - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Search
Favorites
Media
Address
http://www.vitalsearch-ca.com/gen/_nonmembers/ca/_vitals/cabirt
Go
Links
Norton AntiVirus

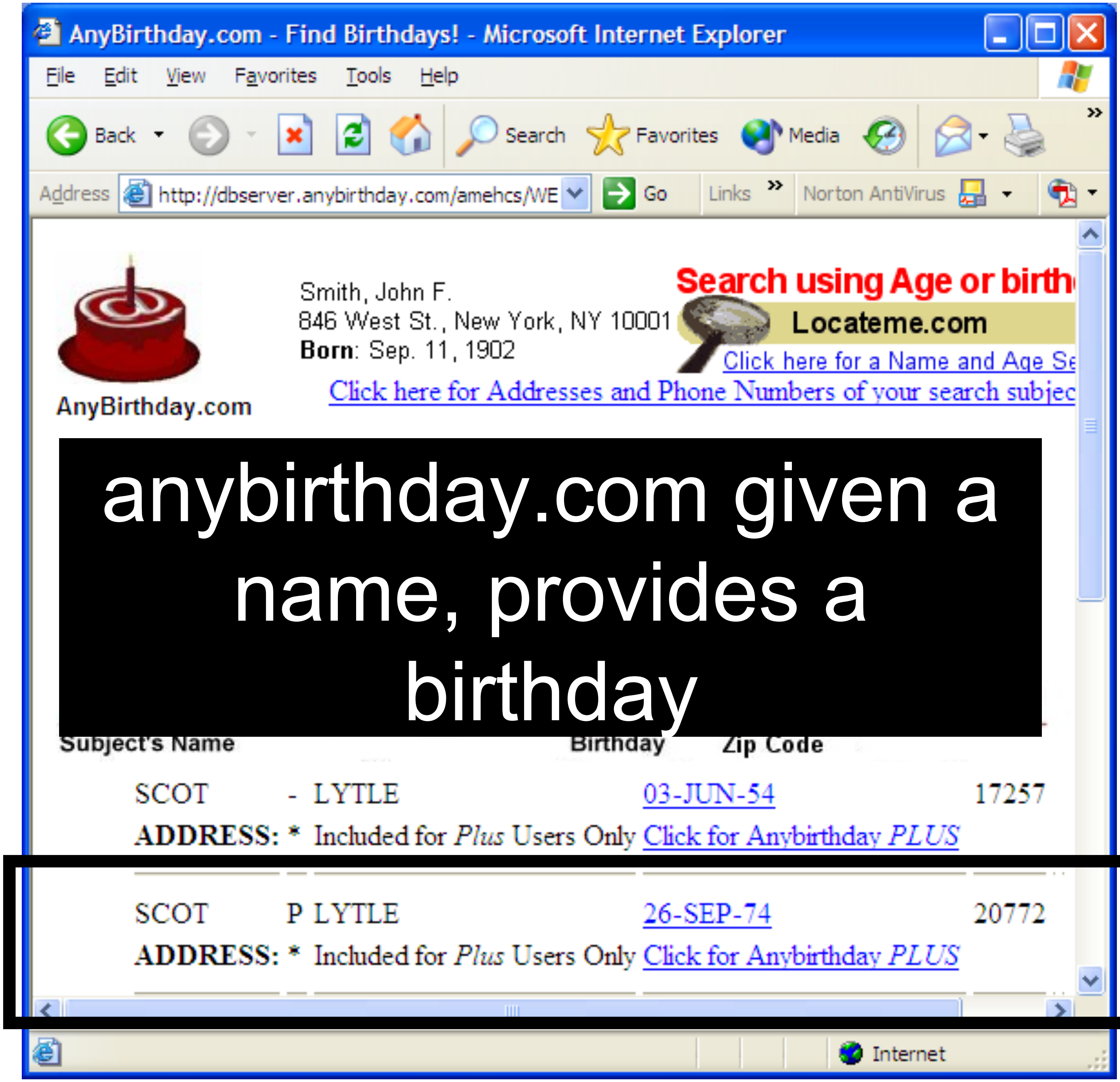
Showing records 0 - 30 (24608727 total, 98860 in query)

SQL-query:
SELECT Last_Name,First_Name,Middle_Name,B_yr,B_mo,B_dy,Mothers_Last_Name,Sex,County_of_Birth,idno from cabirths_ygenfull where 1 and Last_Name like "Jones" LIMIT 0, 30

Last_Name	First_Name	Middle_Name	B_yr	B_mo	B_dy	Mothers_Last_Name	Sex	County_of_Birth
JONES			1905	09	11		MALE	RIVERSIDE
JONES			1905	06	14	KELLY	MALE	SAN FRANCISCO
JONES	ADA	T	1905	12	30	KARMODE	FEMALE	SOLANO
JONES	ALBERT		1905	09	02	SHADEY	MALE	SAN FRANCISCO
JONES	ASA		1905	08	07	TAITE	MALE	LOS ANGELES
JONES	BEATRICE	F	1905	11	05	FANNIN	FEMALE	SHASTA

Results of search on 'Jones'

Source: http://www.vitalsearch-ca.com/gen/_nonmembers/ca/_vitals/cabirths-nopsm.htm anybirthday.com given a name, provides a birthday


AnyBirthday.com - Find Birthdays! - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Search
Favorites
Media
Address http://dbserver.anybirthday.com/amehcs/WE
Go
Links
Norton AntiVirus
Internet

AnyBirthday.com

Smith, John F.
846 West St., New York, NY 10001
Born: Sep. 11, 1902

Search using Age or birth
Locateme.com
Click here for a Name and Age Se
Click here for Addresses and Phone Numbers of your search subjec

Subject's Name			Birthday	Zip Code
SCOT	-	LYTLE	03-JUN-54	17257
ADDRESS:	*	Included for *Plus* Users Only	Click for Anybirthday *PLUS*	
SCOT	P	LYTLE	26-SEP-74	20772
ADDRESS:	*	Included for *Plus* Users Only	Click for Anybirthday *PLUS*	

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

10:17 50°

123-45-6789

Can we solve the problem by just having everyone publish their SSN on the Web?

Surveillance of the Homeless

> “…perform an unduplicated accounting of homeless persons sufficient to provide annual reports … documenting the demographics and utilization patterns of homeless persons.”

Conference Report (H.R. Report 107-272) for the Fiscal Year 2002 HUD Appropriations Act (Public Law 107-73).

Congress:

track homeless → save money.

Conference Report (H.R. Report 106-988) for the Fiscal Year 2001 HUD Appropriations Act (Public Law 106-377).

From $10 to $115mil in 10 years

U.S. Department of Housing and Urban Development. Emergency Shelter Grants Allocation History. www .hud.gov/utilities/intercept.cfm?/offices/cpd/homeless/ budget/esghistory.pdf as of September 2005

New York City spent $1bil

New York City Independent Budget Office. Give ‘Em Shelter: Various City Agencies Spend Over $900 Million on Homeless Services. Fiscal Brief, March 2002

County Funding Soon Exhausted

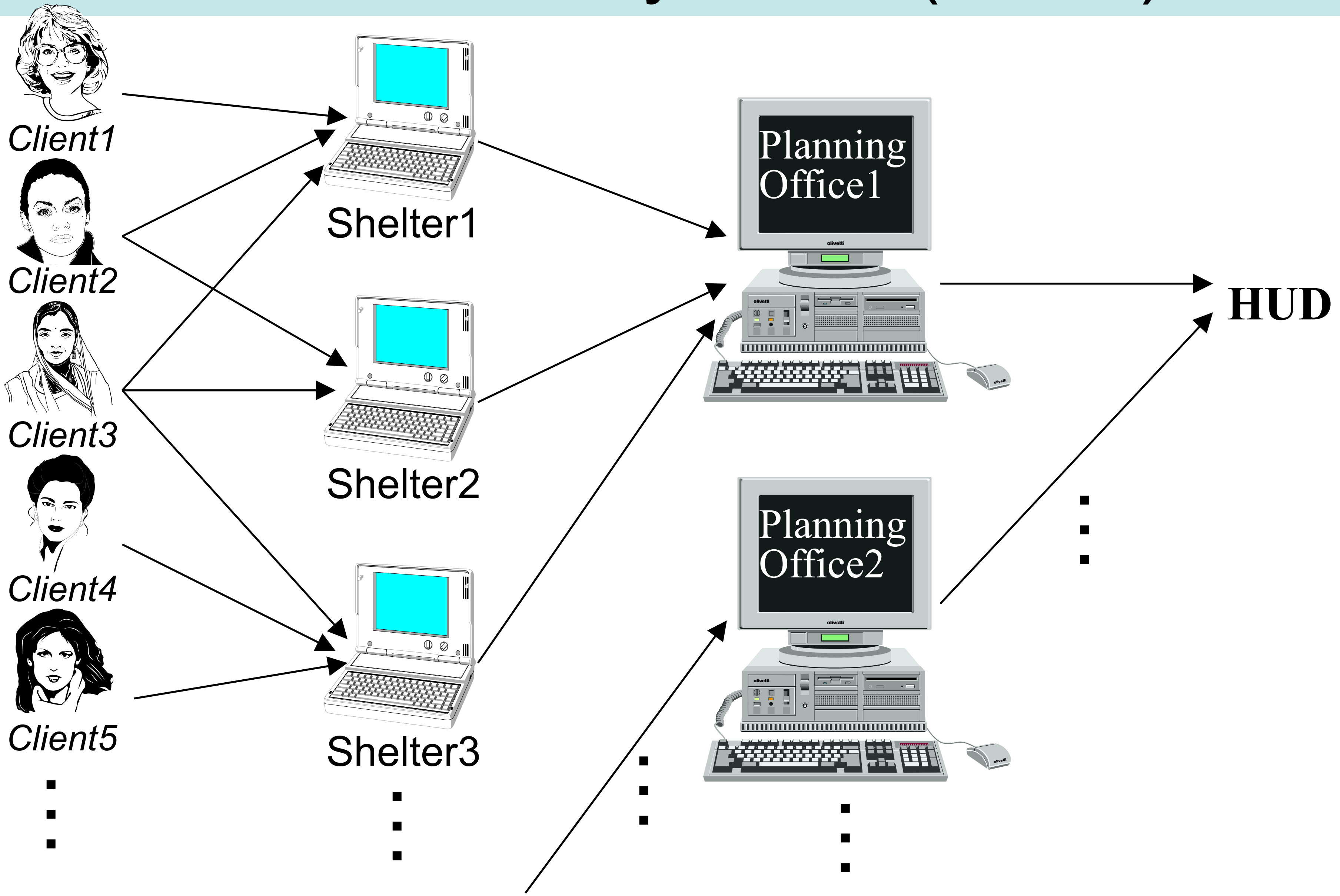
Northeast Ohio Coalition for the Homeless. Overflowing Shelters: a history and recommended solutions. April 9, 2005. www.neoch.org/what_to_do_overflowing.htm

Unmet shelter requests: 37% overall, 52% family

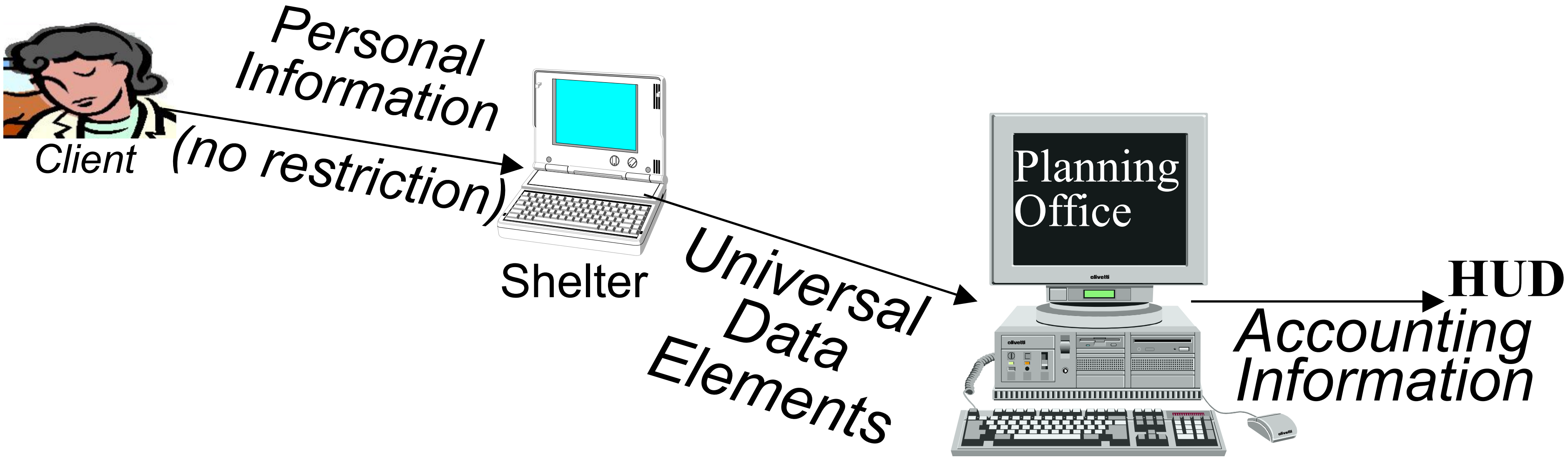
U.S. Conference of Mayors, A Status Report on Hunger and Homelessness in America's Cities 2001. www. usmayors.org/uscm/hungersurvey/2001/ hungersurvey2001.pdf as of Sept 2005.

Why? Dunno.

HUD: Homeless Management Information Systems (HMIS)



HMIS Data Flow



Client

Personal Information (no restriction)

Shelter

Universal Data Elements

Planning Office

HUD

Accounting Information

Universal Data Elements

Unique Identifier (“UID”)

~~Name~~

~~Social Security Number~~

Date of Birth

Ethnicity and Race

Gender

Veteran Status

Disabling Condition

Residence Prior to Program Entry

Code of Last Permanent Address

Program Entry Date

Program Exit Date

Unique Person Identification Number

Program Identification Number

Household Identification Number

Program-Specific Data Elements

Income Sources

Non-cash Benefits

Physical Disability

Developmental Disability

HIV/AIDS

Mental Health

Substance Abuse

Domestic Violence

Services Received

Destination Type

Reasons for Leaving

Employment

Education

General Health Status

Pregnancy Status

Veterans Information

Children’ s Education

HUD Reporting (Sample)

Question #	AHAR Questions: Emergency Shelter -Individuals
1	How many people used emergency shelters at __ time?
2	What is the distribution of family sizes using emergency shelters?
3	What are the demographics of individuals using emergency shelters?
3	distribution by gender?
3	distribution by race and ethnicity?
3	distribution by age group?
3	distribution by household size?
3	distribution by veteran status? By disabling condition?
4	What was the living arrangement the night before entering the emergency shelter?
4	within/outside geographical jurisdiction?
5	What is distribution of the number of nights in an emergency shelter?
5	distribution by gender?
5	distribution by age group?

Common Techniques for Constructing Made-up Identifiers

Encoding

$$\text{“}\underbrace{\text{09121960}}_{\text{Date of birth}}\underbrace{\text{F}}_{\text{Sex}}\underbrace{\text{37213}}_{\text{ZIP}}\text{”}$$

Hashing

“8126r1329ws”

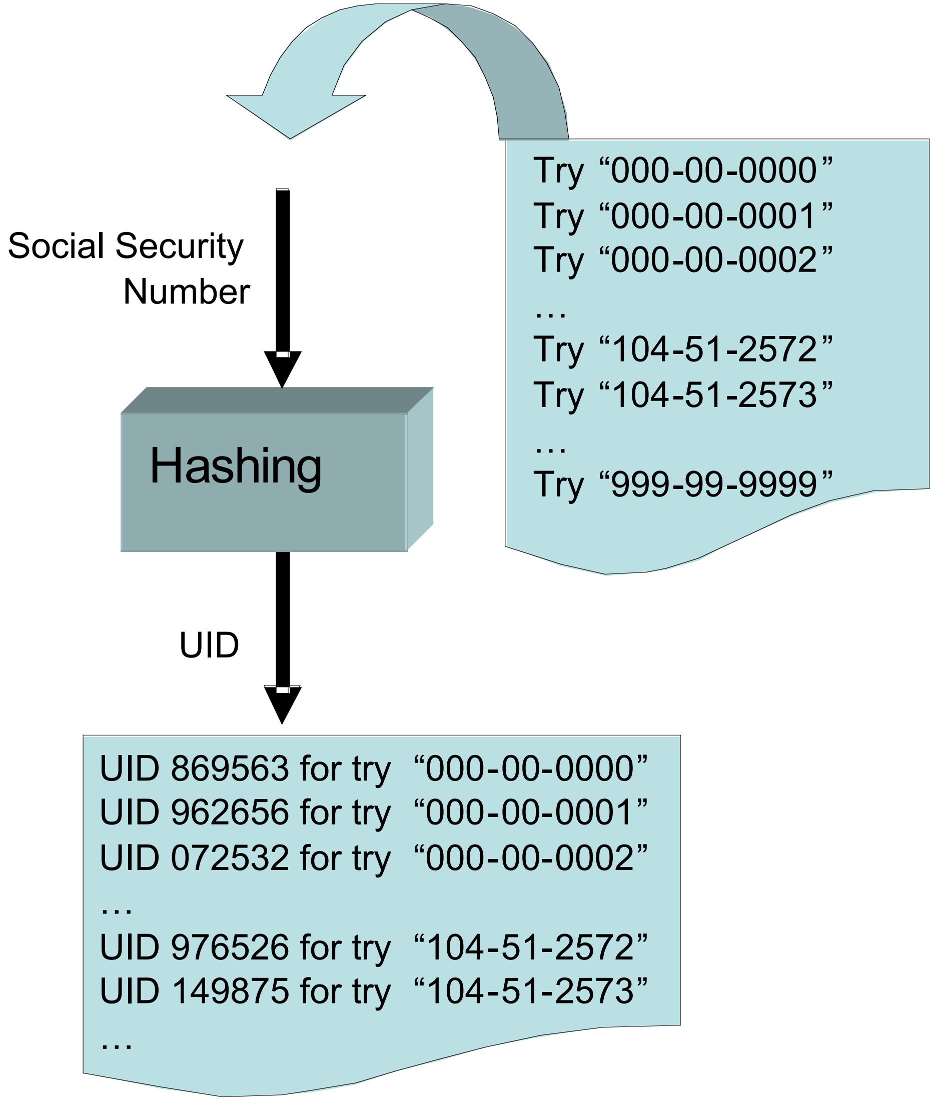
“986s594652”

Date of birth   Sex   ZIP

Problems with Consistent Hashing

Dataset

UID
149875
072532
976526

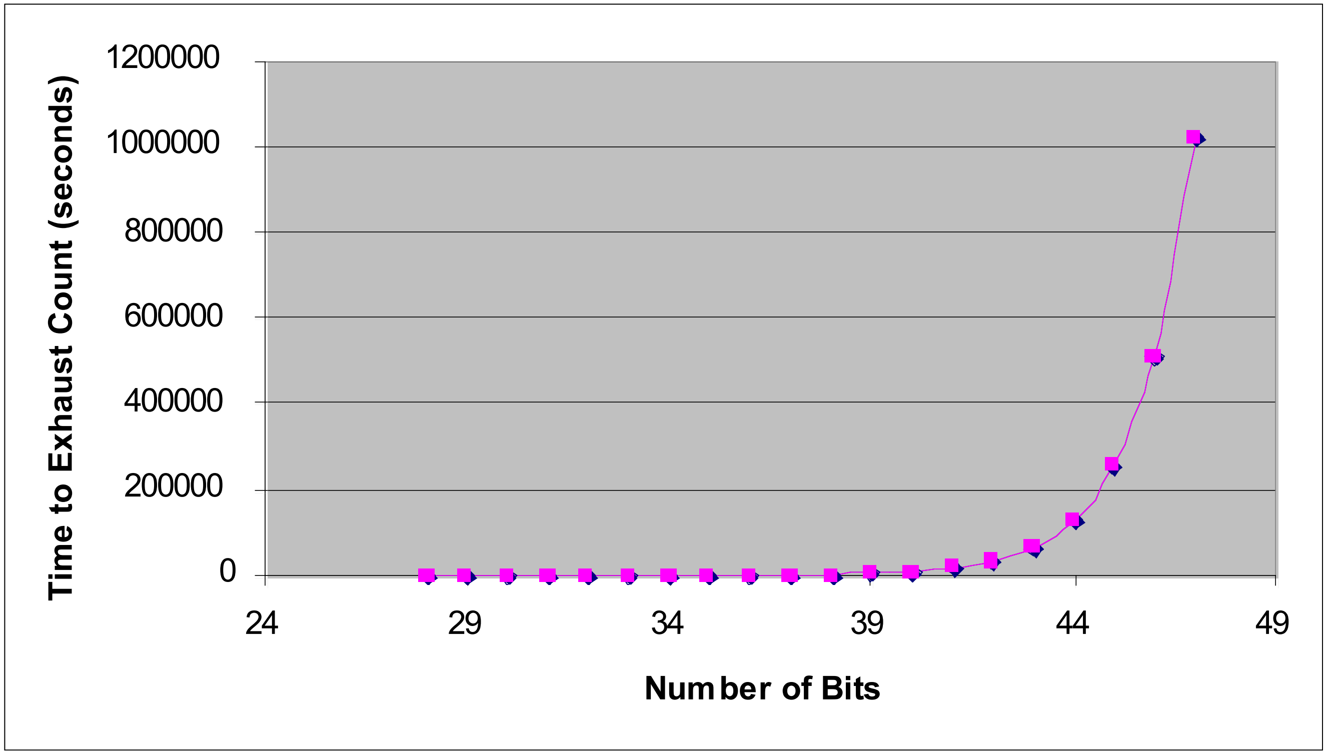


Try “000-00-0000”
Try “000-00-0001”
Try “000-00-0002”
…
Try “104-51-2572”
Try “104-51-2573”
…
Try “999-99-9999”

UID 869563 for try “000-00-0000”
UID 962656 for try “000-00-0001”
UID 072532 for try “000-00-0002”
…
UID 976526 for try “104-51-2572”
UID 149875 for try “104-51-2573”
…

Problems with Consistent Hashing

bits	seconds
28	1
29	3
30	7
31	15
32	31
33	62
34	124
35	249
36	499
37	998
38	1996
39	3993
40	7986
41	15963
42	31926
43	63888
44	127725
45	255463
46	510774
47	1021463



Encryption

“8126r1329ws”

“8126r1329ws” + key

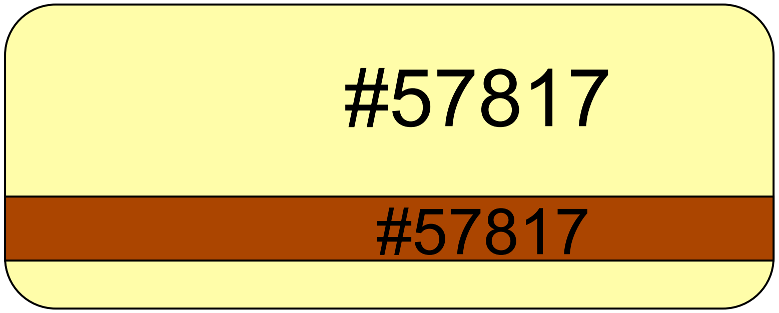
= “9/12/1960, F, 37213”

Date of birth

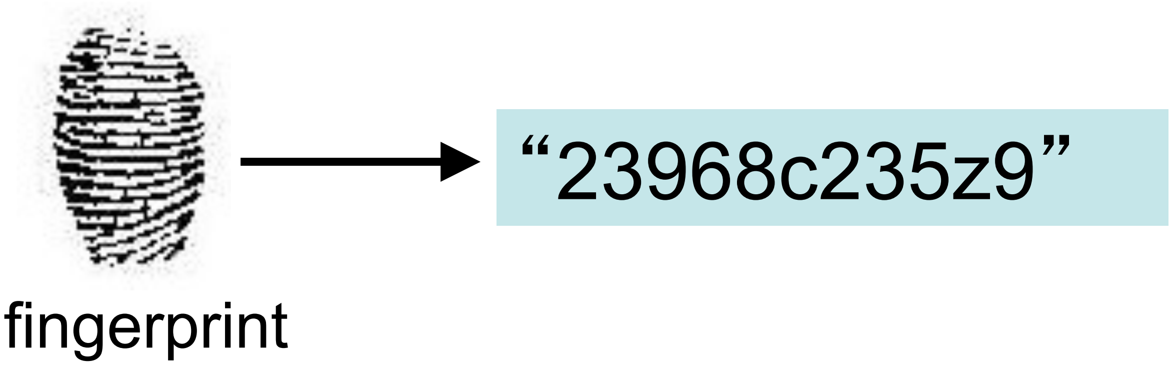
Sex

ZIP

Scan Cards / RFID Tags



Biometrics

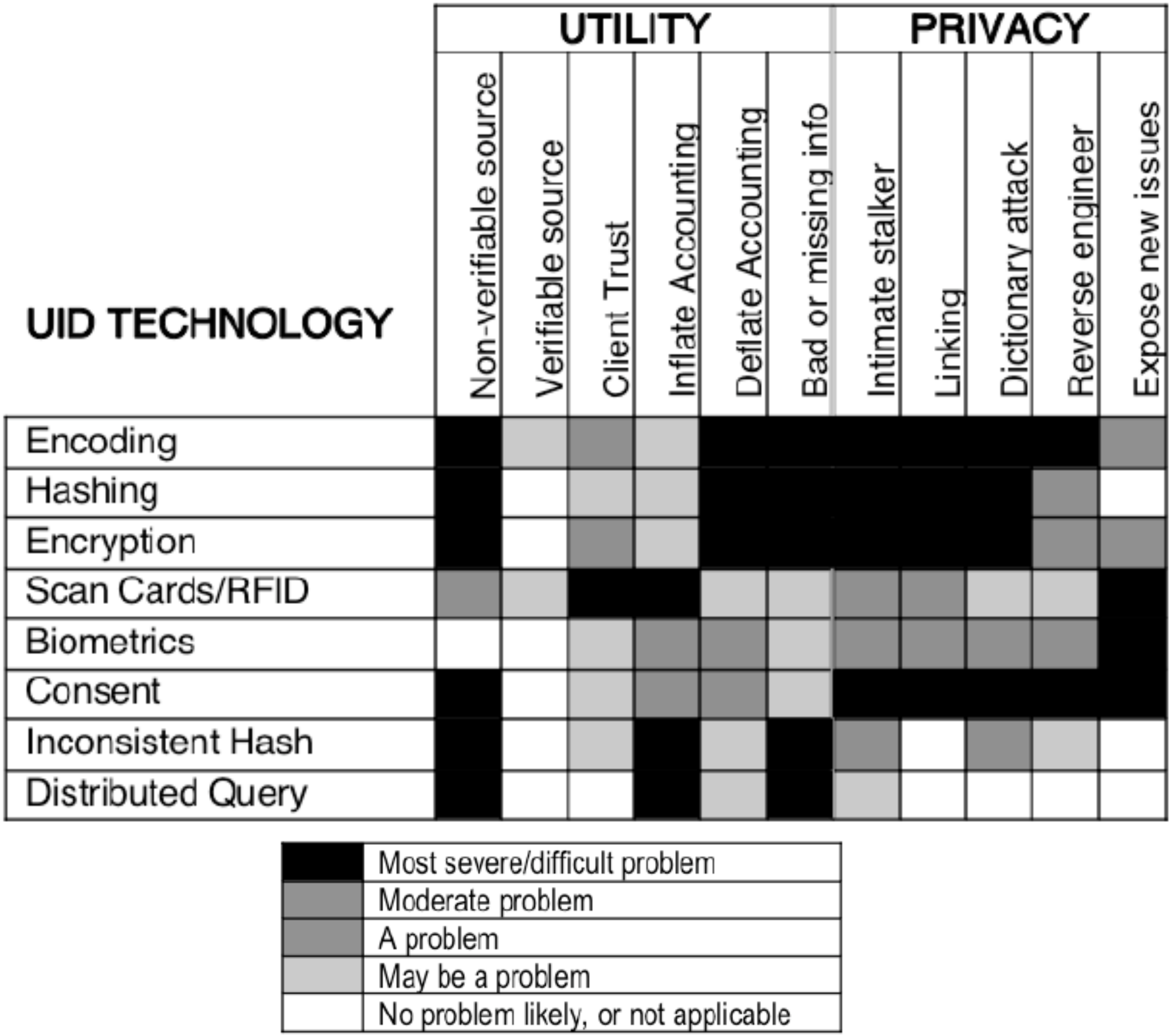


fingerprint

“23968c235z9”

Consent

“Jane Smith, 9/12/1960, F, 37213”



UID TECHNOLOGY	UTILITY						PRIVACY				
	Non-verifiable source	Verifiable source	Client Trust	Inflate Accounting	Deflate Accounting	Bad or missing info	Intimate stalker	Linking	Dictionary attack	Reverse engineer	Expose new issues
Encoding	■	░	▓	░	■	■	■	■	■	■	▓
Hashing	■	□	░	░	■	■	■	■	■	▓	□
Encryption	■	□	▓	░	■	■	■	■	■	▓	▓
Scan Cards/RFID	▓	░	■	■	░	░	▓	▓	░	░	■
Biometrics	□	□	░	▓	▓	░	▓	▓	▓	▓	■
Consent	■	□	░	▓	▓	░	■	■	■	■	■
Inconsistent Hash	■	□	░	■	░	■	▓	□	▓	░	□
Distributed Query	■	□	□	■	░	■	░	□	□	□	□

■	Most severe/difficult problem
▓	Moderate problem
▓	A problem
░	May be a problem
□	No problem likely, or not applicable

L. Sweeney. *Risk Assessments of Personal Identification Technologies for Domestic Violence Homeless Shelters*. Carnegie Mellon Tech Report CMU-ISRI-05-133 Pittsburgh: November 2005.

Threats?

Intimate Stalker “Threat”

- Knows detailed information about a targeted client
- Is highly motivated to locate a targeted client
- Can compromise a shelter or Planning Office

Planning Office “Threat”

- Has lots of other information that may contain the client.
- Motivated to learn information about clients by combining available client-specific data
- Can combine data to construct client profiles

Privacy Standard?

De-Identification Under HIPAA Scientific Standard

Based on generally accepted nal, statistical and scientific pr d methods, a person certifies that

“the **risk is ve** t the information could be used, alone ination with other reasonably available info on, by an anticipated recipient to identify an ind vidual who is a subject of the information.”

Minimal risk.

U.S. Health and Human Services; Standards for Privacy of Individually Identifiable Health Information; Final Rule, 45 CFR Parts 160 and 164. *Federal Register*, vol 67, no 157, August 14, 2002.

VAWA Privacy Standard

[not share] identifying information … about an individual ...likely to disclose the locati including —

(I) first and last name;
(II) home or physical add
(III) contact informati or IP address, or tele
(IV) a social); and
(V) any other uding date of birth, racial or ethnic id, … that, in combination with any other n ersonally identifying information would serve to identify any individual.

Provable anonymity.

Violence Against Women and Department of Justice Reauthorization Act of 2005, H.R. 3402. ("VAWA")

Solution

Secure Multi-Party Computation

1. Diffie and Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644-654, 1976.

2. Adi Shamir. How to share a secret. *Communications of the ACM*, 22(11):612-613, 1979

3. D. Chaum. Blind signatures for untraceable payments. In *Proceedings of Advances in Cryptology*, pages 199-203. Plenum Press, 1982

4. A. Yao. Protocols for secure computations. In *Proceedings, IEEE Symposium on Foundations of Computer Science*, pages 160-164. IEEE Computer Society, 1982.

For survey list, see privacy.cs.cmu.edu/dataprivacy/papers/multipartycomputation/index.html

Secure Multi-Party Computation

DNA databases

→ Malin's PhD Thesis '06
(Malin and Sweeney '01, '05)

General

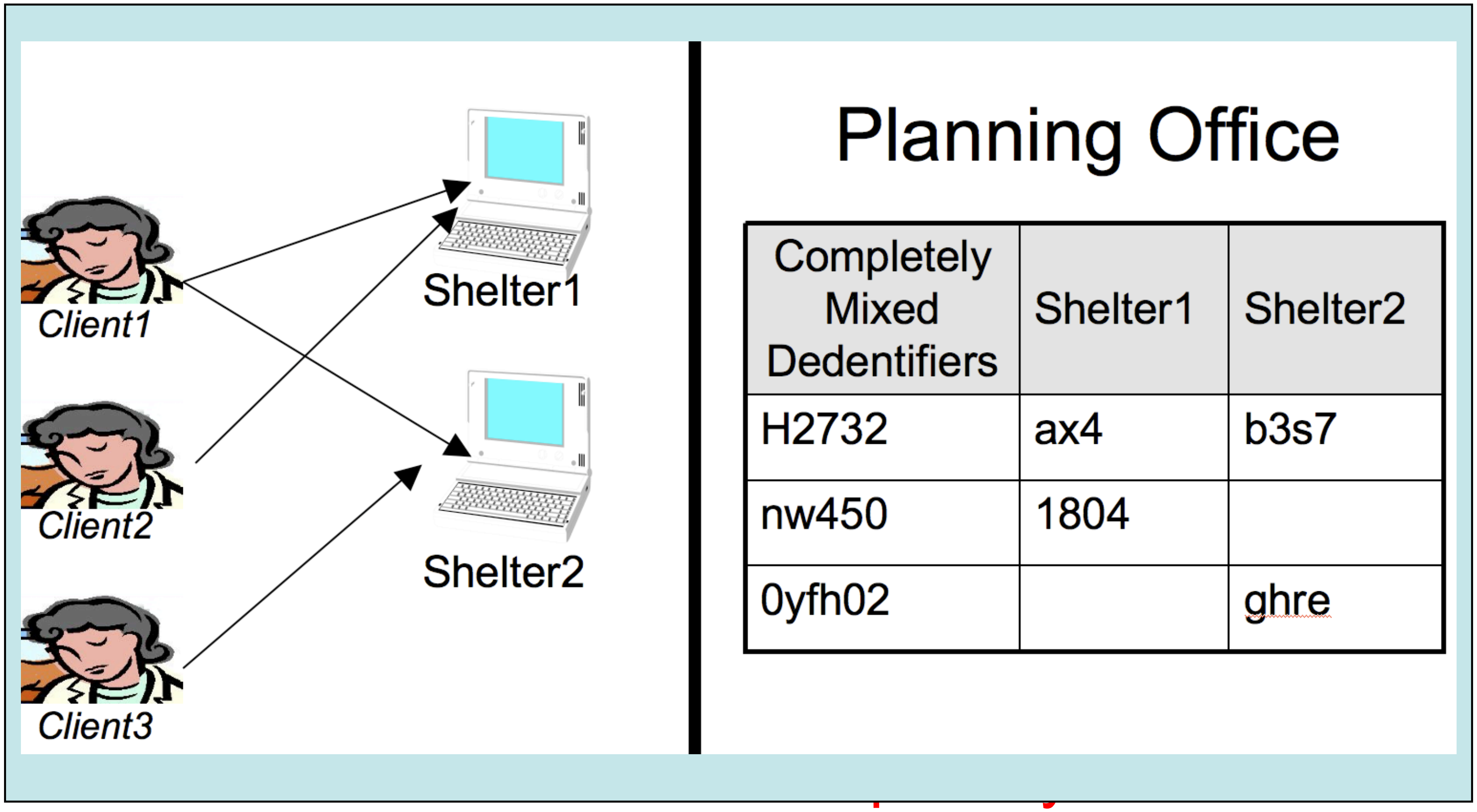
→ RandomOrder Protocol
(Sweeney and Shamos '04)

Bioterrorism Surveillance

→ PrivaSum Protocol
(Edo-Eket and Sweeney '04)

PrivaMix (This Work)

multiple sources produce an anonymous linked dataset



Planning Office

Completely Mixed Dedentifiers	Shelter1	Shelter2
H2732	ax4	b3s7
nw450	1804	
0yfh02		ghre

The Big Idea in 3 Steps

1. Compute “dedentifiers.”

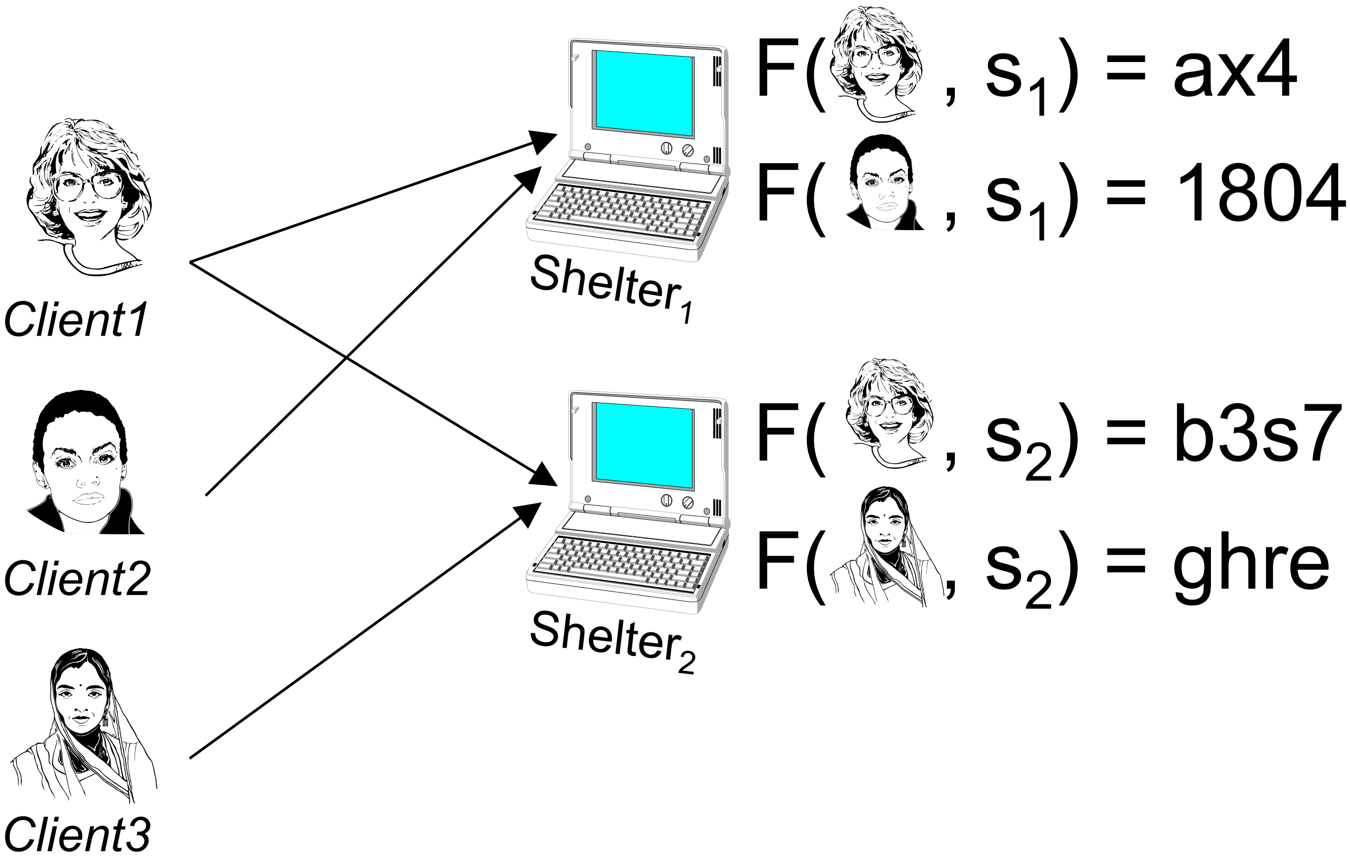
2. Send data and dedentifiers to Planning Office.

3. Use network to “mix” dedentifiers.

Step 1. Compute Dedentifiers

For each C_i visiting Shelter S_j,

S_j computes $\mathbf{F}(SSN_i, s_j) = D_{ij}$



$F(\quad , s_1) = ax4$

$F(\quad , s_1) = 1804$

$F(\quad , s_2) = b3s7$

$F(\quad , s_2) = ghre$

Step 2. Data to Planning Office

P creates a table $\{S_j, D_{ij}, UDE_{ij}\}$

for Client C_i at Shelter S_j.

Planning Office

Shelter	Dedentifier	UDE
S_1	ax4	…
S_1	1804	…
S_2	b3s7	…
S_2	ghre	…

Step 3. Mixing Dedentifiers

3.1 Randomly select Shelter order & send D’s to S_1.

For k = 2 to n do:

3.2. P computes:

$$\mathbf{Z_1} = \left\{ x \mid x = F^{a}\left(F\left(SSN_i, s_j\right), \ldots\right), x \in \mathbf{Z}, a \geq 0, j \neq k \right\}$$

$$\mathbf{Z}_2 = \left\{ x \mid x = F^{a}\left(F\left(SSN_i, s_j\right), \ldots\right), x \in \mathbf{Z}, a \geq 0, j = k \right\}$$

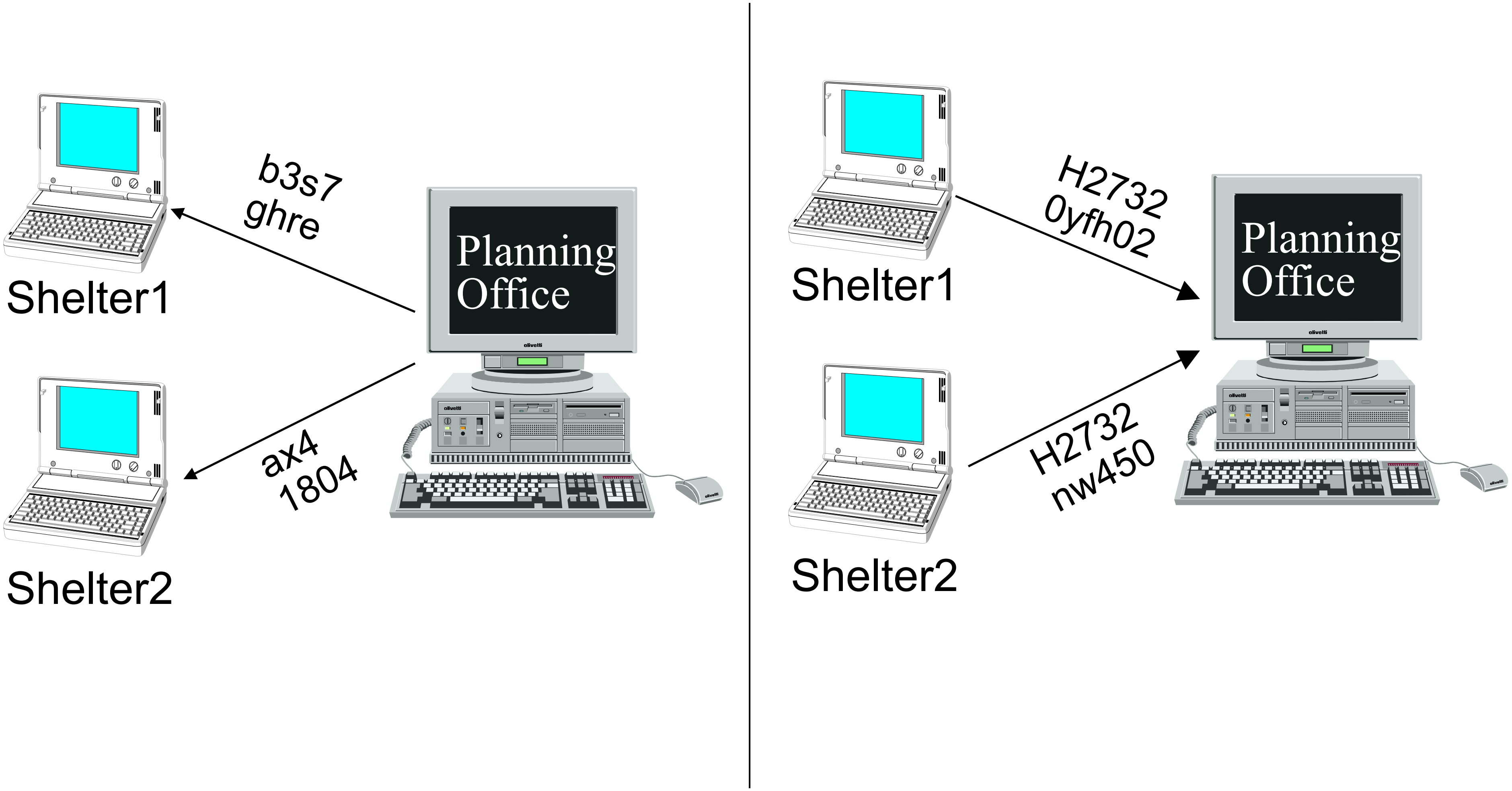
3.3. P sends $\mathbf{Z}_1$ to S_k

3.4. S_k sends P:

$$\mathbf{Z}_{Sk} = \left\{ F\left(x, s_k\right) \mid x = F^{a}\left(F\left(SSN_i, s_j\right), \ldots\right), \forall x \in \mathbf{Z_1} \right\}$$

3.5. P computes: $\mathbf{Z} = \mathbf{Z}_{Sk} \cup \mathbf{Z_2}$

Mixing Rounds



Shelter1

b3s7
ghre

Planning Office

Shelter2

ax4
1804

Shelter1

H2732
0yfh02

Planning Office

Shelter2

H2732
nw450

Mixing Results

$$\boxed{\text{H2732} = F(F(SSN_1, s_2), s_1)}$$

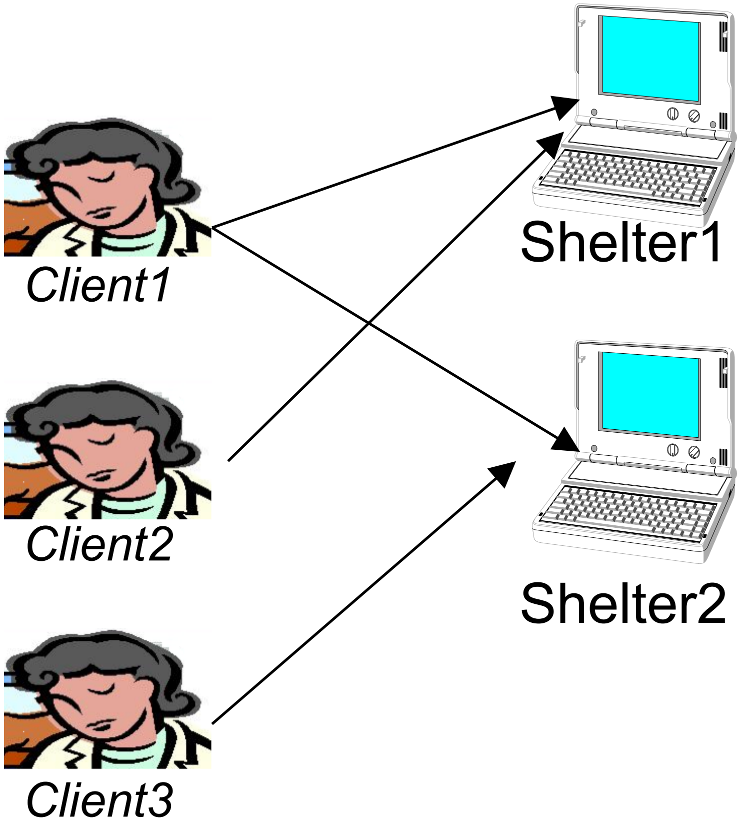
$$\text{0yfh02} = F(F(SSN_3, s_2), s_1)$$

$$\boxed{\text{H2732} = F(F(SSN_1, s_1), s_2)}$$

$$\text{nw450} = F(F(SSN_2, s_2), s_2)$$

The Client whose SSN was the same has the same final mix, notwithstanding the order of mixing.

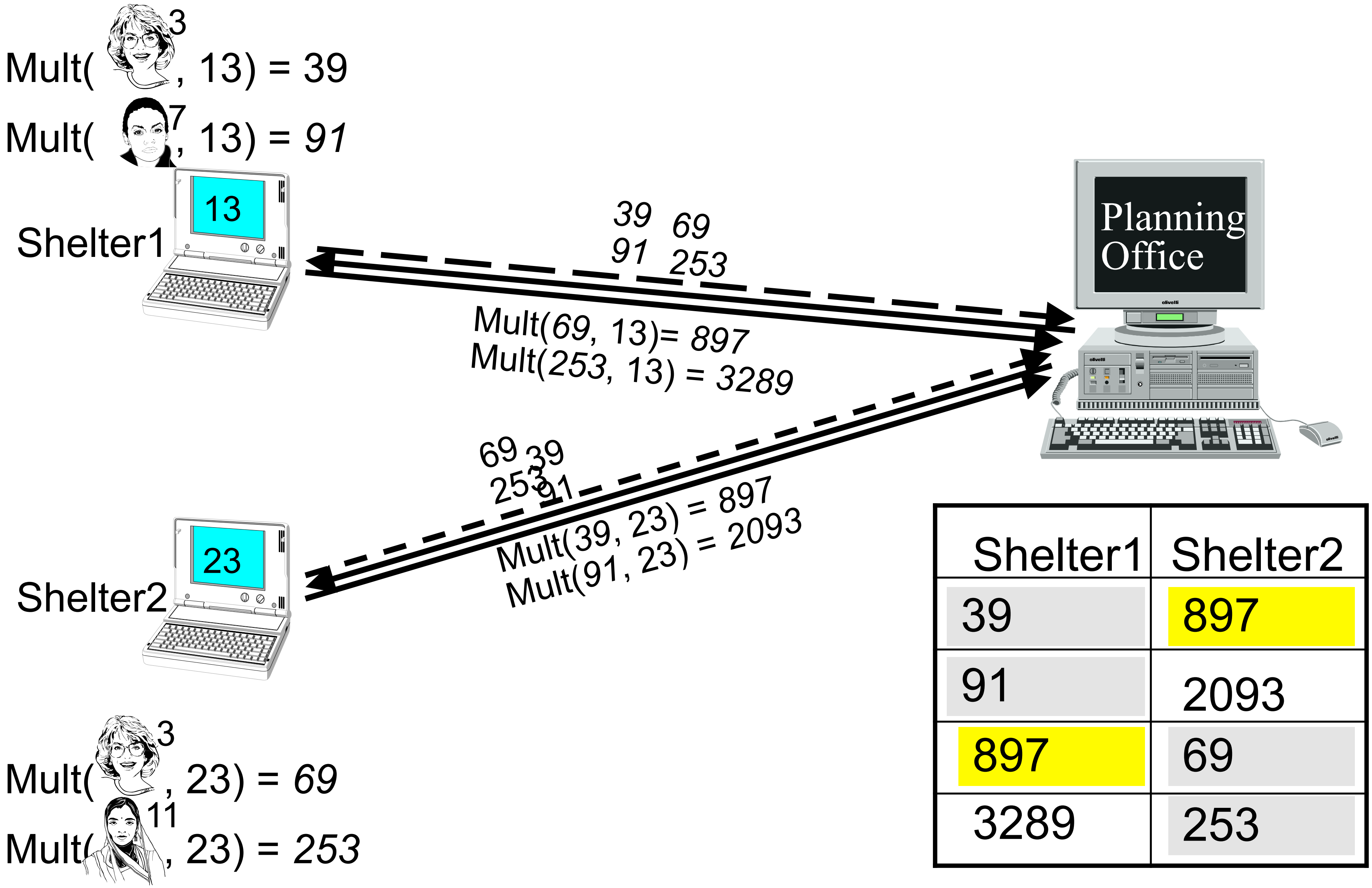
Planning Office Learns



Planning Office

Completely Mixed Dedentifiers	Shelter1	Shelter2
H2732	ax4	b3s7
nw450	1804	
0yfh02		ghre

Simplified Multiplication Example



Mult(3, 13) = 39

Mult(7, 13) = *91*

Shelter1

Mult(3, 23) = *69*

Mult(11, 23) = *253*

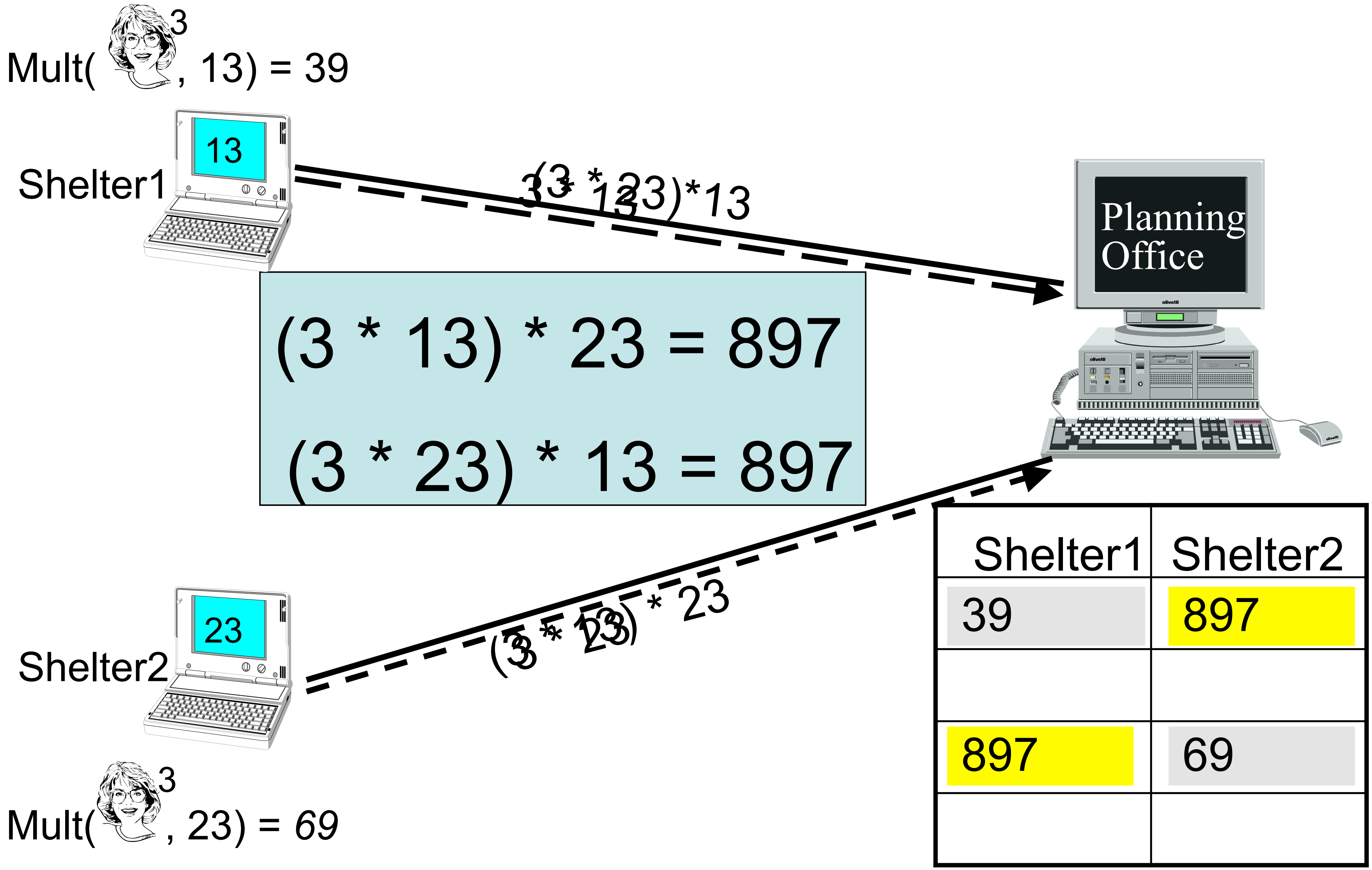
Shelter2

Shelter1	Shelter2
39	897
91	2093
897	69
3289	253

Planning Office Learns

Completely Mixed Dedentifiers	Shelter1	Shelter2
897	39	69
2093	91	
3289		253

Simplified Multiplication Example



Mult(3, 13) = 39

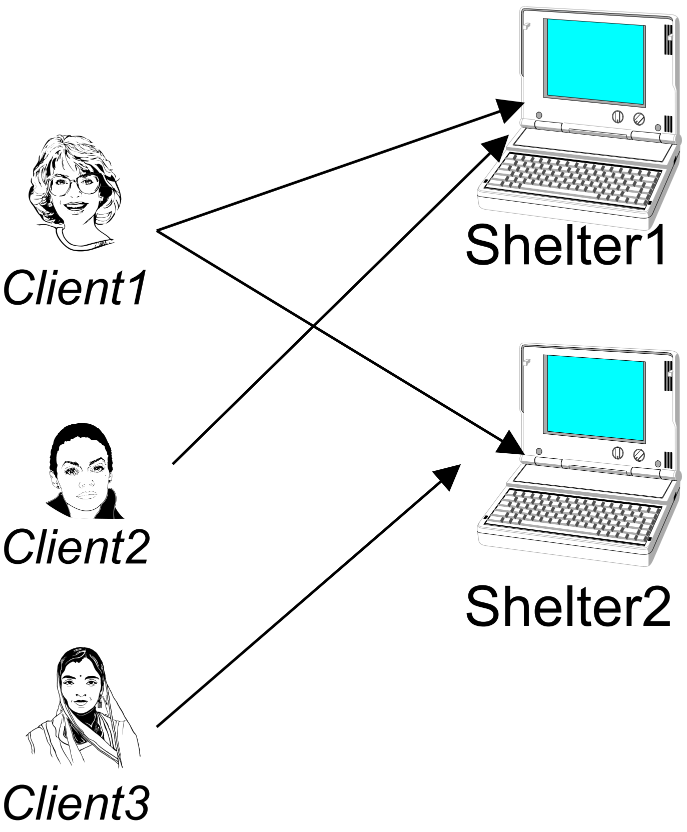
Mult(3, 23) = *69*

(3 * 13) * 23 = 897

(3 * 23) * 13 = 897

Shelter1	Shelter2
39	897
897	69

Planning Office Learns



Planning Office

Completely Mixed Dedentifiers	Shelter1	Shelter2
897	39	69
2093	91	
3289		253

PrivaMix Function Requirements

1. Inconsistent assignment
2. One-way (hard to revers
3. Commutative (enable re
4. Privacy: secret client inf complete or partial infor
5. Collision-free
6. Correct: same final values iff same client

$$z_i^1 = \left(c_i^{s1} \bmod p\right)$$

$$z_i^{12} = \left(c_i^{s1} \bmod p\right)^{s2} \bmod p$$

$$z_i^{21} = \left(c_i^{s2} \bmod p\right)^{s1} \bmod p$$

L. Sweeney. Requirements for and Examples of PrivaMix Functions: a research notebook. April 2006.

How good is it?

Utility

If a client's private value is unique and consistently used, then her utilization pattern will be accurate.*

Privacy

Compromising a shelter will not help the intimate stalker learn where a targeted client has been.*

	UTILITY						PRIVACY				
UID TECHNOLOGY	Non-verifiable source	Verifiable source	Client Trust	Inflate Accounting	Deflate Accounting	Bad or missing info	Intimate stalker	Linking	Dictionary attack	Reverse engineer	Expose new issues
Encoding	■	░	▓	░	■	■	■	■	■	■	▓
Hashing	■		░	░	■	■	■	■	■	▓	
Encryption	■		▓	░	■	■	■	■	■	▓	▓
Scan Cards/RFID	▓	░	■	■	░	░	▓	▓	░	░	■
Biometrics			░	▓	▓	░	▓	▓	▓	▓	■
Consent	■		░	▓	▓	░	■	■	■	■	■
Inconsistent Hash	■		░	■	░	■	▓		▓	░	
Distributed Query	■			■	░	■	░				

PrivaMix (client-level)	■			■	░	■	░	░			
PrivaMix (aggregate)	■			■	░	■					

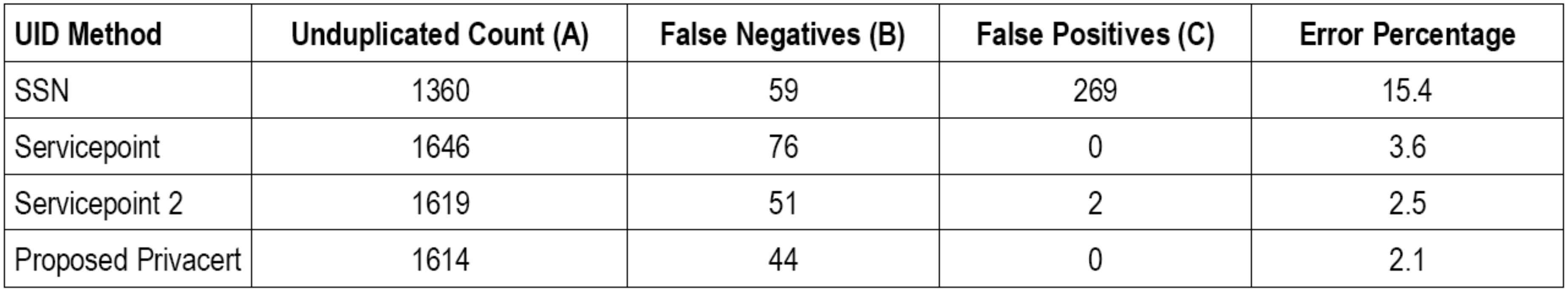
Index

Real-world Iowa Experiment

Conducted by Abt Associates in DesMoines

PrivaMix source: {first name, date of birth}

6 months of data



UID Method	Unduplicated Count (A)	False Negatives (B)	False Positives (C)	Error Percentage
SSN	1360	59	269	15.4
Servicepoint	1646	76	0	3.6
Servicepoint 2	1619	51	2	2.5
Proposed Privacert	1614	44	0	2.1

Index

What did we learn?

- Avoid ad hoc approaches
- Brainstorm about threats
- Understand governing privacy standard
- Solve for privacy “and” utility