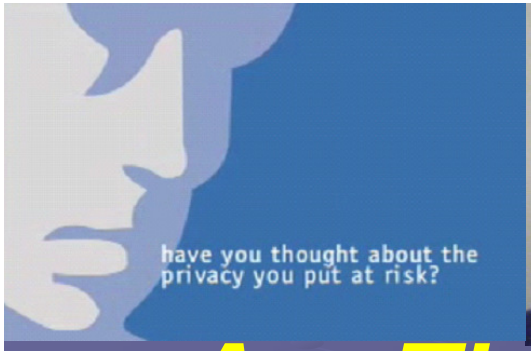
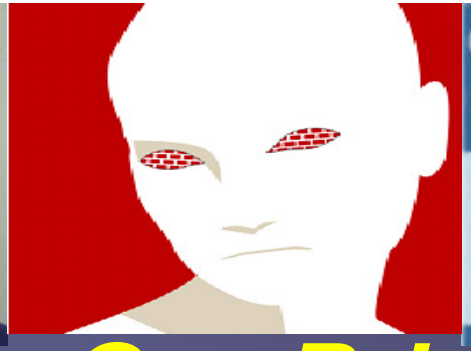
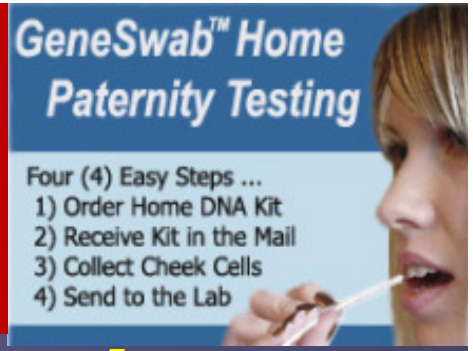


Are They Making Our Privates Public?
Emerging Risks in Governmental Open Data Projects

Dr. Tom Keenan FCIPS, I.S.P., Professor
Faculty of Environmental Design & Dept. of Computer Science
University of Calgary

IFIP Summer School

Trento, Italy, September 8, 2011

UNIVERSITY OF CALGARY

Copyright © 2011
, T.P. Keenan



What is the Open Data Movement?

- Initiative by many governments to make their data openly available
- Driven by:
  - Calls for transparency ("it's the taxpayers money")
  - Sincere desire for innovation ("they'll come up with some great uses for our data"
  - Fear ("Wikileaks is going to get this anyway, it looks better if we release it")
  - Cheapness ("we don't have to pay the developers")
  - Positive sentiments about open source, etc. "data wants to be free"

UNIVERSITY OF CALGARY



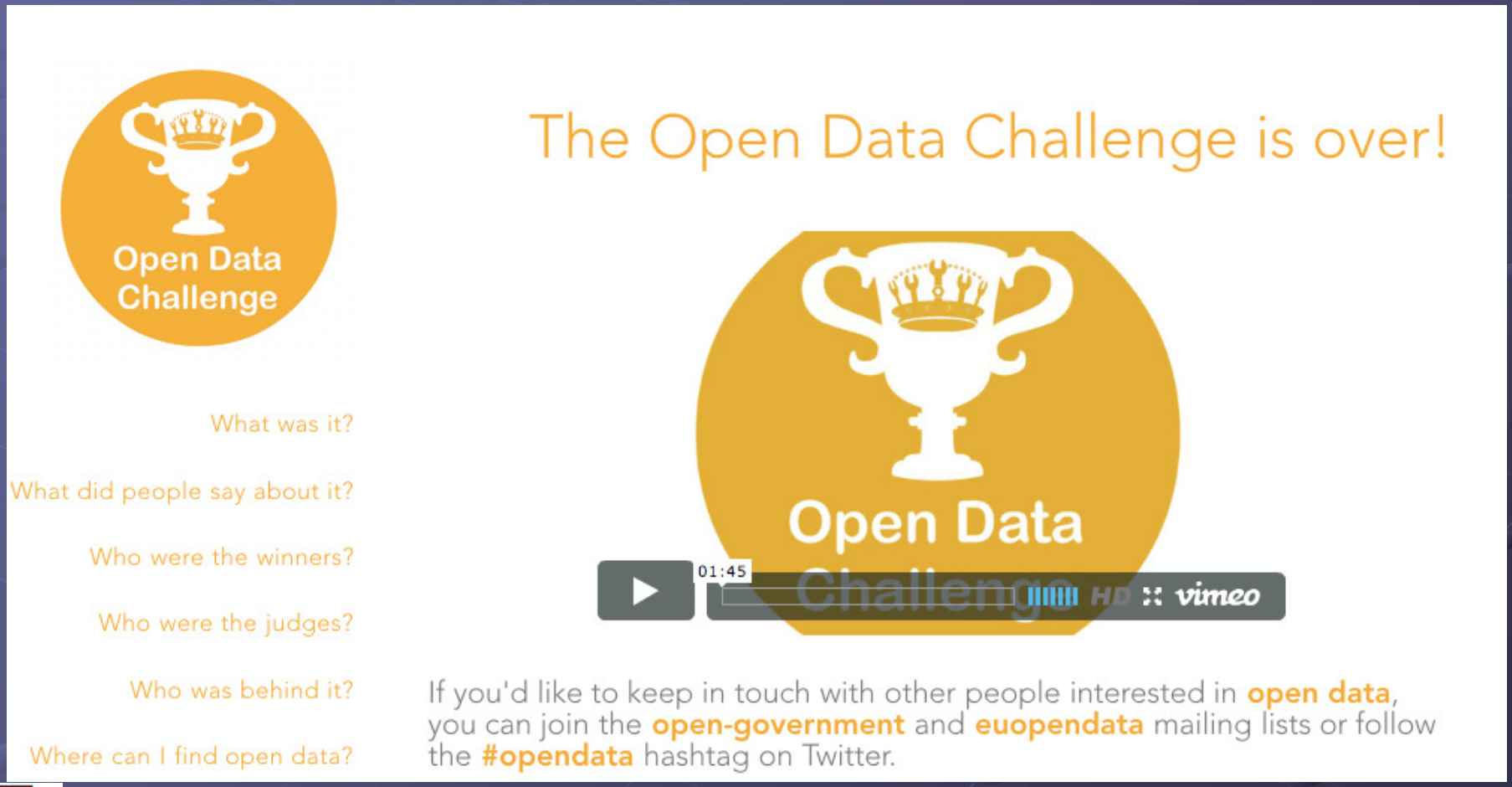
Some notable examples

- NYC Datamine
- Philadelphia OpenDataPhilly
- Open Data Framework (several Canadian cities)
- Open Data Challenge (EU)

UNIVERSITY OF CALGARY



When you offer a 20,000 € prize, things happen!



434 entries from 24 member states



So…Who Won?

PublicData.eu beta — Europe's Public Data

Groups Apps Ideas About

Login Register

Find datasets

ZNasichDani / From Our Taxes

View

Link

http://znasichdani.sk/l?l=en

Developer

Fair-play Alliance and VORSR.sk

License

Closed

Tags

power networks transparency managers politicians

ZNasichDani.sk uncovers who are influential persons (owners, managers, statutories) standing behind companies successful in securing contracts with the state, thus helping uncover conflicts of interest; it also shows how large are the sums of contracts obtained in this way.





www.fair-play.sk

UNIVERSITY OF CALGARY



Who Wouldn't Agree with Such a Wonderful Application?

Open Data Challenge Winner Ordered To Remove Certain Data

Source: spectator.sme.sk

A court has ordered Slovak NGO Fair-Play Alliance to take down data from their award winning open data application.

Bratislava, 28 June 2011

(by Ton Zijlstra)

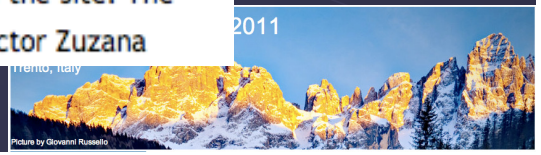
The Slovak NGO Fair-Play Alliance, a transparency watch dog organisation, won the Open Data Challenge with their application Znasichdani.sk during the Digital Agenda Assembly in Brussels on June 16th.

A Slovak judge has now ordered the NGO to take down specific data from their winning application.

The application combines two public registers. The overview of all work contracted by the state, and the company register are combined to show what the total volume of work is that various people, through the businesses they represent or own, do for public bodies.

According to a news article in the Slovak Spectator the statutory representative, Jarmila Povazanova, of the Strabag construction company demanded in court that the total value of all public contracts of companies Povazanova represents be removed from the site. The Bratislava II district court ordered the NGO to remove that information. Director Zuzana

UNIVERSITY OF CALGARY



A little technical aside

- This thing runs off multiple databases maintained by the government
- The Fair-Play alliance has read only access
- How are they supposed to remove entries selectively?
- Rule: When it comes to technology, judges are 20 years behind…

UNIVERSITY OF CALGARY



What We Learned

- Public data can be contentious, which is probably a good sign that it needs to be out there!
- The laws ***and policies*** are lagging behind in terms of technology

UNIVERSITY OF CALGARY



Fat Fingers at the Data Office

- "Fat Fingers" –refers to exploiting typing errors (gmial instead of gmail). Gmial.com is a Phishing site
- Artem Dinaberg's papers at BlackHat & Defcon this year

UNIVERSITY OF CALGARY



Bitsquatting Explained

- The binary for the string “microsoft” is:
- 01101101 01101001 01100011 01110010 01101111 01110011 01101111 01100110 01110100
- and the binary encoding for “mic2osoft” is (with the single changed bit highlighted):
- 01101101 01101001 01100011 0**0**110010 01101111 01110011 01101111 01100110 01110100



NYC Datamine

- On October 6, 2009 NYC released, with great fanfare, 103 municipal data sets
- One listed 1100 women's organizations in the city
- Release contained their private email address, secret question like "favorite pet" and answer like "fluffy"

UNIVERSITY OF CALGARY

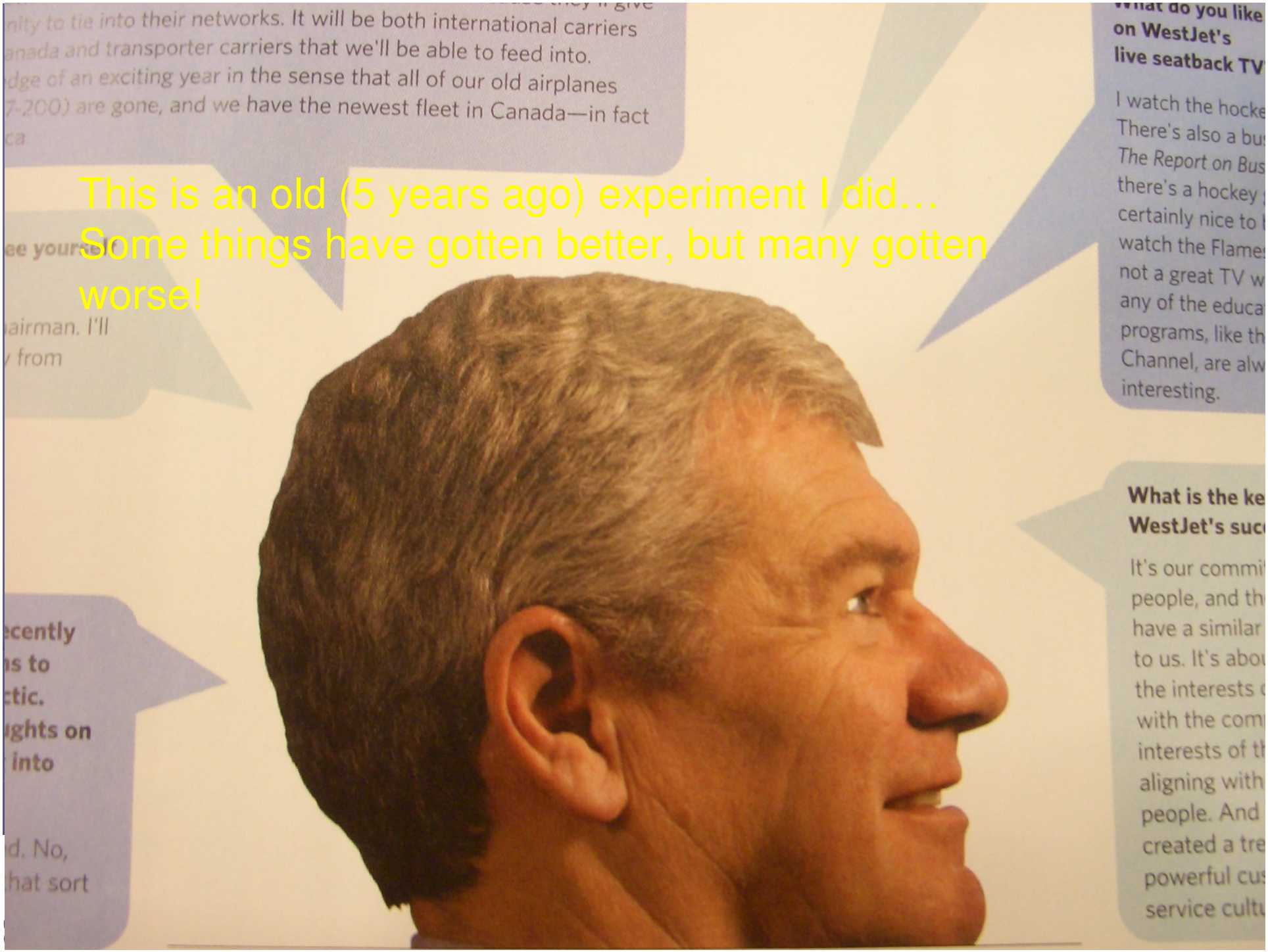


Lessons from that

- Those in charge of releasing public data need to have very skinny (and careful fingers)
- Think carefully how it could be combined with other data to compromise privacy
- Result of an experiment I did in Calgary, Canada five years ago…

UNIVERSITY OF CALGARY




nity to tie into their networks. It will be both international carriers
anada and transporter carriers that we'll be able to feed into.
dge of an exciting year in the sense that all of our old airplanes
7-200) are gone, and we have the newest fleet in Canada—in fact
ca
This is an old (5 years ago) experiment I did…
Some things have gotten better, but many gotten worse!
ee yours
airman. I'll
from
ecently
ns to
ctic.
ghts on
into
d. No,
hat sort
What do you like
on WestJet's
live seatback TV
I watch the hocke
There's also a bu
The Report on Bus
there's a hockey
certainly nice to
watch the Flame
not a great TV w
any of the educa
programs, like th
Channel, are alw
interesting.
What is the ke
WestJet's suc
It's our commi
people, and th
have a similar
to us. It's abou
the interests
with the com
interests of t
aligning with
people. And
created a tre
powerful cu
service cultu

Search: "beddoe, AB" 6 Total Listings

New search | Modify search

BEDDOE, C 1 of 6
Calgary, AB T1X 1E1
(403) 255-2686
More information

Check out C Beddoe with Web Detective!
Search School Yearbooks in Calgary AB
Find Romance in Calgary at Date.ca

BEDDOE, CLIVE 2 of 6
1315 70 Ave SW
Calgary, AB T2V 0R2
(403) 255-8396
More information

Check out Clive Beddoe with Web Detective!
Search School Yearbooks in Calgary AB
Find Romance in Calgary at Date.ca

BEDDOE, J 3 of 6
Calgary, AB T1X 1E1
(403) 521-2260
More information

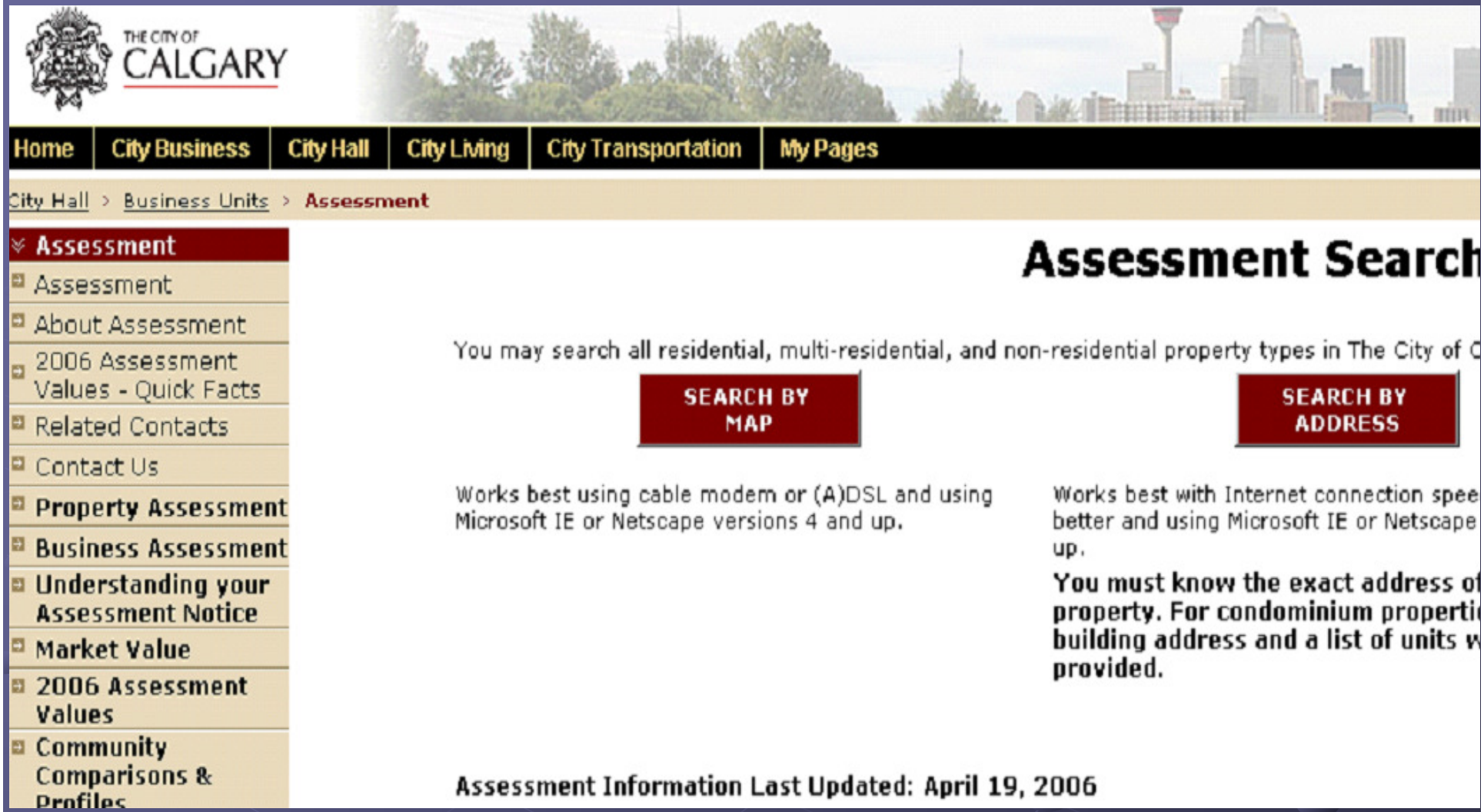
Check out J Beddoe with Web Detective!
Search School Yearbooks in Calgary AB
Find Romance in Calgary at Date.ca

BEDDOE, K & S 4 of 6
Calgary, AB T1X 1E1
(403) 212-2352
More information

Check out K & S Beddoe with Web Detective!
Search School Yearbooks in Calgary AB
Find Romance in Calgary at Date.ca

UNIVERSITY OF CALGARY





UNIVERSITY OF CALGARY



Print Neighbourhood Map

Your property assessment is based on The City of Calgary's estimate of the market value December 31, 2005.

Roll number:	111156600
Property Location:	1315 70 AV SW
Property Assessment:	710,500
Assessment Class:	Residential
Property Type:	Land and Improvement
Property Use:	Single Residential
Taxation Status:	Taxable
Community Name:	Kelvin Grove
Land Use Designation:	RESIDENTIAL SINGLE-DETACHED DISTRICT
Parcel Area:	8,095 sq. ft. / 752 sq. m.
Building Description:	House/2 Storey Garage
Approx. Living Area above grade:	3,260 sq. ft. / 303 sq. m.
Finished Area below grade:	Yes
Garage:	Yes
Fireplace(s):	Yes
Walkout Basement:	No

Y DR SW
1408
1404
1320
1316
1312
636000
1308
531500
1304
512500
70 AV SW
642000
1403
462500
1327
369000
1323
435500
1319
414000
1315
710500
1311
634500
1307
371000
1303
592500
1219
450500
7252
653500
7248
400500
7244
462000
7240
476500
7236
472500
7232
392000
7228
463000
7224
503000
7220
483000
KANANASKIS DR SW

This one is still current (verified yesterday)



Hernando County Sheriff's Office

Sheriff Al Nienhuis

- HCSO Home
- Explore HCSO
- Hernando County Jail
- Emergency Management
- Traffic Information
- Alert Bulletins
- Crimestoppers
- Media Releases
- CASE Partnership
- Fraud Alerts
- General Information
- Public Records Search
- Crime Statistics
- Sexual Offender Info
- Job Line
- eLert Notification

About Hernando County Sheriff's Office

Hernando County is located on the West coast of Central Florida. The Sheriff's Office provides law enforcement services to a population of approximately 173,000. This is an increase of nearly 330% since 1980, and Hernando continues to be regarded as one of the fastest growing counties in the United States.

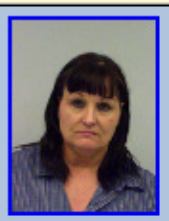
This decade will bring new challenges to our county, but the commitment of the Hernando County Sheriff's Office will not change. We will continue to provide our community with professional, efficient law enforcement through well-trained employees, technology, crime prevention and a proactive approach to public safety.



Highlights

- Click here for Info on how you can purchase t-shirts from the Mecklenburg Memorial Motorcycle Ride
- Run for New Beginnings Youth Shelter 5K Run/Walk
- Click here for more information about the Sheriff's Office proposed BUDGET for 2012
- Public Notice - Edward Byrne Memorial Justice Assistance Grant
- HCSO Citizen Volunteer Corps
- Pictures from 2011 Law Enforcement Special Olympics Torch Run
- Hernando County Community

Records Returned = 30 | Criteria Given : Between 2/6/2009 and 2/6/2009

Report#:	ArrestDate:	Name:	DOB:	Offense:	Arrest Location:	Latest Photo Available:
2009-04928/1	02/06/2009 11:44 PM	SIMMONS, CATHY L	02/19/1961	831.02 - UTTERING FORGED INSTRUMENT	6242 CALIFORNIA ST	
2009-04923/1	02/06/2009 11:05 PM	WILBUR, WILLIAM LEO III	01/26/1941	88.231 - CONTEMPT COURT/CHILD SUPPORT	16160 APPENNINE DR	
2009-04913/1	02/06/2009 09:41 PM	LEDE, JEANETTE MICHELE	10/01/1981	322.34 - DRIVER'S WITH LICENSE REVOKED OR SUSPENDED	16425 SPRING HILL DR	
2009-04912/1	02/06/2009 09:41 PM	LEDE, JEANETTE MICHELE	10/01/1981	948.06 - VIOLATION OF PROBATION	16425 SPRING HILL DR	
2009-04909/1	02/06/2009 08:55 PM	ASKEW, ROBERT JARMAN	10/31/1959	784.03 - DOMESTIC BATTERY	16096 MEDRICK RD	

UNIVERSITY OF CALGARY



What if this was Your Son?

- 01/08/2010 04:00PM
- COU***EY, BOBBY
- DOB 11/10/1997
- 810.02 - BURGLARY DWELLING
- 18900 CORTEZ BV



UNIVERSITY OF CALGARY

Philadelphia's OpenDataPhilly.org

- Unnecessary data provided
- Example: campaign contributions
- "Results are freely viewable and printable but not available for download."

UNIVERSITY OF CALGARY



Search for Smith in 2010

phila.gov | records department

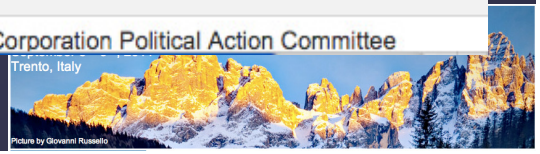
Document Search | Logout public

New Search | Modify Search | Printable V

You searched for: Year = 2010 and approximate search in ContributorID for Smith and =CFR - Schedule I - Part A - Contributions Received From Political Committees ($50.01 $250.00), CFR - Schedule I - Part B - All Other Contributions ($50.01 - $250.00), CFR - Schedule I - Part C - Contributions Received From Political Committees (Over $250.00), Schedule I - Part D - All Other Contributions (Over $250.00), CFR - Schedule I - Part E - Other Receipts, CFR - Schedule II - Part F - In-Kind Contributions Received (Value of $5 $250.00), CFR - Schedule II - Part G - In-Kind Contributions Received (Value Over $250.00)

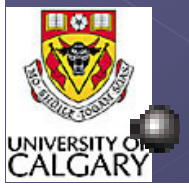
Download Results as CSV

Contribution By	Address	Amount	Date	Recipient
D. Smith	Two Liberty Place 1601 Chestnut St Philadelphia, PA 19192	$6.83	May 13, 2010	CIGNA Corporation Political Action Committee
Joseph Smith	1719 Spring Garden Street Philadelphia, PA 19130	$101.20	Nov 25, 2009	Local Union #98 IBEW COPE
D SMITH	1534 E 80TH ST LOS ANGELES, CA 90001-3404	$40.00	Apr 22, 2010	Communication Workers of America
LETITIA SMITH	16368 KENSINGTON PLACE MORENO VALLEY, CA 92551-0000	$10.00	Apr 26, 2010	Communication Workers of America
Mark Smith	Two Liberty Place 1601 Chestnut St Philadelphia, PA	$5.00	Aug 19,	CIGNA Corporation Political Action Committee



Very good source of home addresses

- The donor files for 2010 were downloaded from OpenDataPhilly as .CSV files, imported into Microsoft Excel and sorted by the contributor name field.
- “Potential Duplicates” were defined as additional names that were shown identically as a name in the database, but with a different address..





Results from most common US names

Name	Number of Entries	"Dups"
Smith	588	8
Johnson	426	3
Williams	400	4

UNIVERSITY OF CALGARY



Are those addresses real home addresses?

- A significant number of contributors (86/588 of Smith, 63/426 of Johnson, 77/400 of Williams) listed the same address, 1719 Spring Garden Street, Philadelphia, PA., 19130.
- Viewing this location on Google Maps Streetview shows that it is the office of the I.B.E.W. Electricians Union.

UNIVERSITY OF CALGARY



What We Learned

- Excellent place to get very good data information on some people
- Possibly compromising their privacy rights
- Not good for its intended purpose which was probably to disambiguate donors

UNIVERSITY OF CALGARY



Provocative Question: Should the Rich Have Less Privacy?

- People who run companies that get contracts, give contributions (especially large ones) have money
- In the US there is a huge sentiment that the EBTA food stamp program is abused
- In theory we could post Mrs. X bought junk food and tried to buy beer..but we don't

UNIVERSITY OF CALGARY



Indirect Risks of Government Disclosure

- You can't get it back. Companies like ChoicePoint Systems (now part of Lexis/Nexis) are grabbing and storing it
- New technologies (think DNA matching)
- "Genetic Geneaology"

UNIVERSITY OF CALGARY



Ancestry.com

- 14 years of operation
- 6,000,000,000 records, mostly from public sources
- 20,000,000 family trees

UNIVERSITY OF CALGARY



Beware of Time Travelling Robots from the Future

PRIVACY INTERNATIONAL

DONATE

ABOUT US | RESEARCH | KEY RESOURCES | BLOG | PRESS AND NEWS | EPHR | PRIVASIA

Home

Search

PI FILES COMPLAINT ABOUT ONLINE DNA GENEALOGICAL TESTING FIRM

OCTOBER 22, 2007 BY *PI*

The leading online genealogical site Ancestry.com has embarked on a new service offering to test the DNA of its customers. This commercial service has been provided to assist genealogical tracing through familial lines.

However there are a number of troubling questions that Privacy International believes must be resolved before people can have confidence in this new service. Until these questions have been satisfactorily answered, we believe the service presents a number of substantial dangers for customers.

MORE LIKE THIS

- Genewatch campaign to 'Reclaim your DNA' from growing UK database
- PI Files complaint about expansion of CCTV on Toronto transit network
- Police are allowed to Keep DNA of Innocent, say UK Law Lords
- Key report damns UK National DNA Database Policy
- PI files complaint about Google Street View

https://www.privacyinternational.org/

Tools

- Scripting languages
- ScraperWiki, collection of Ruby, Python, PHP scripts and plenty of ideas
- SQL Lite and other databases
- Hackathons

UNIVERSITY OF CALGARY



“Super Public Data”

- Data that used to be public but buried in a dusty courtroom basement
- Now instantly available, searchable, storable
- “if it happened in the past ten years, it might be online. If it happened in the past five years, it’s probably online. And if it happened in the past two years, it’s almost certainly online.” (Fertig and Thompson)

UNIVERSITY OF CALGARY



Even how we walk is being studied



- "Human gaits, for example, can provide a lot of information about people's intentions. At the American Army's Aberdeen Proving Ground in Maryland, a team of gait analysts and psychologists led by Frank Morelli study video, much of it conveniently posted on the internet by insurgents in Afghanistan and Iraq."

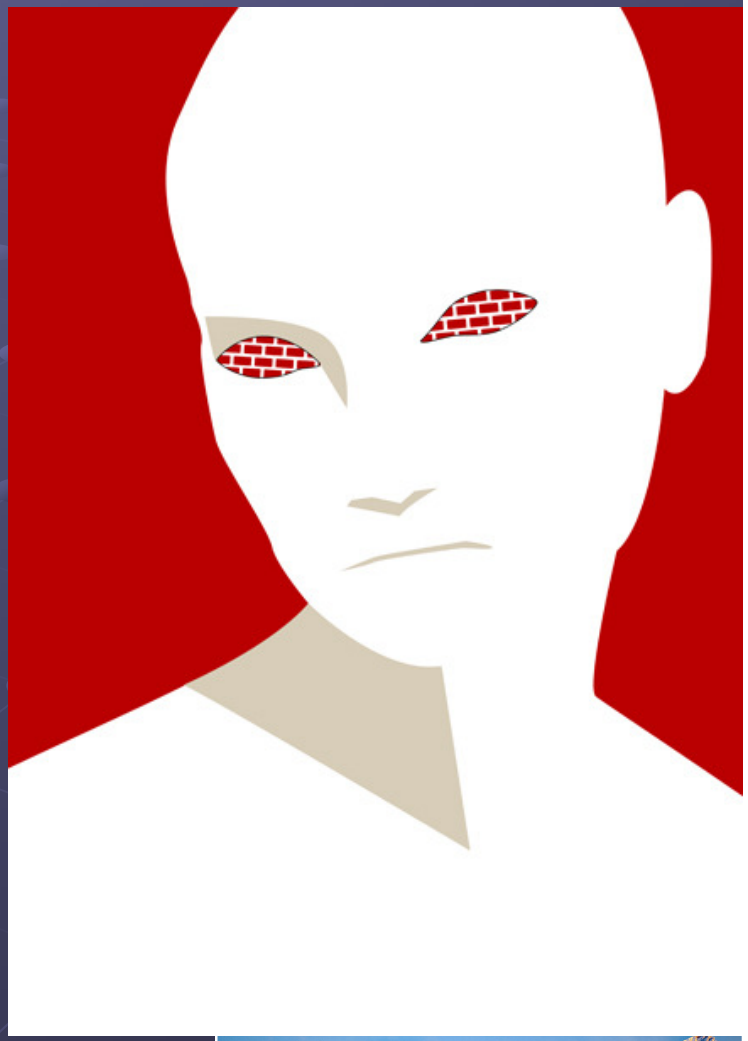
© The Economist, October 23, 2008

Some recommendations

Scan files carefully for direct PII that may be included;
-Consider ways in which PII may be revealed indirectly;
-Act promptly to remove database that are shown to reveal PII, and retain clear legal rights to do so;
-Anticipate the cross-correlation of government data with other databases, public and private;
-Provide a convenient mechanism for users to express privacy concerns and follow up;
-Sponsor hackathons before the data is released to try to foresee unanticipated uses;
-Negotiate strong privacy protection on data provided to the private sector;

Thank you!

- Dr. Tom Keenan, FCIPS
- keenan@ucalgary.ca
- www.ucalgary.ca/evds



This image courtesy
Elise Pelchat
www.conceptsforall.uqam.ca

UNIVERSITY OF CALGARY

