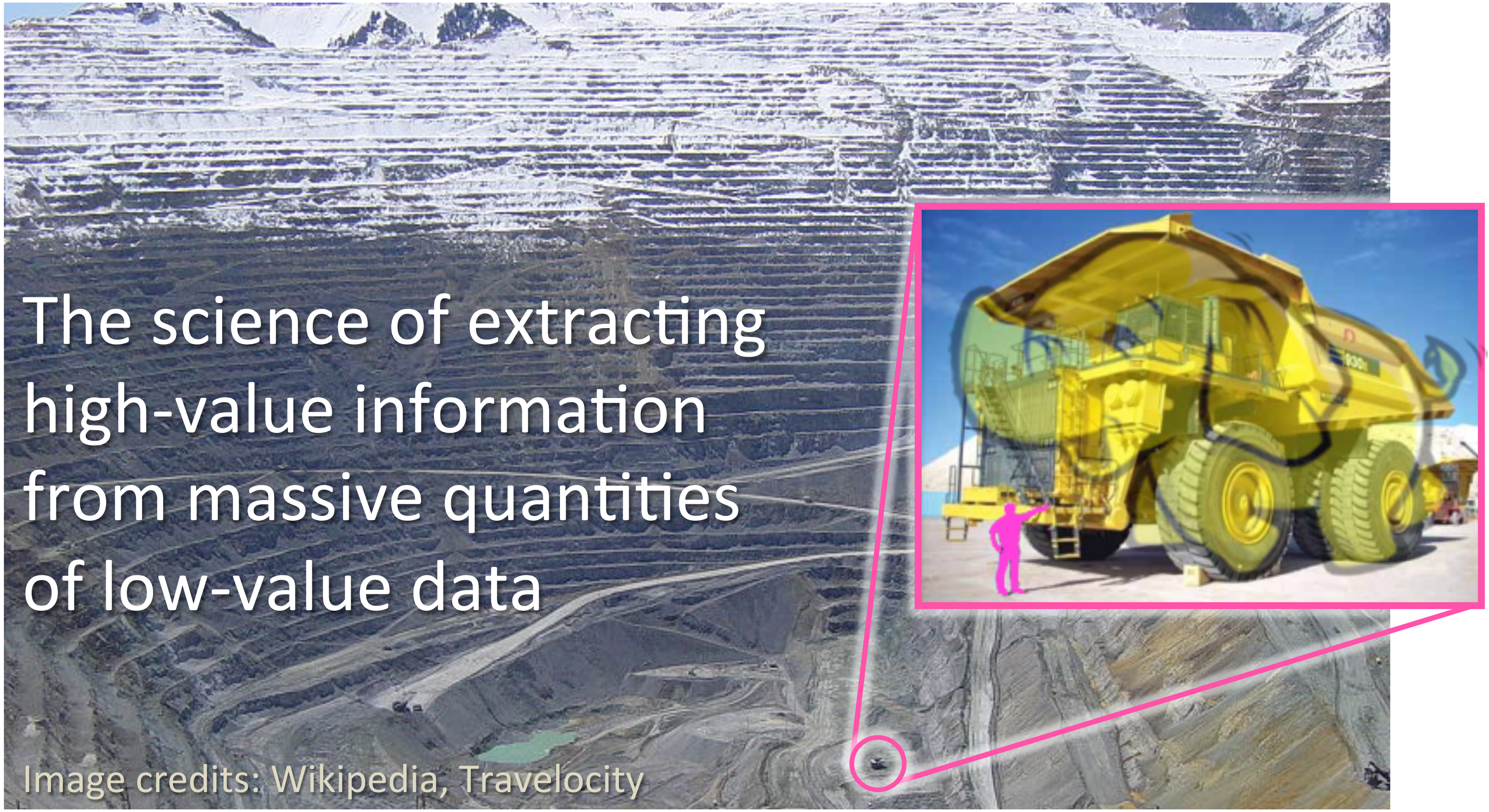
Quis custodiet ipsos Big Data custodes?

Ryan Johnson

HPTS 2013

UNIVERSITY OF
TORONTO

Big data in a nutshell



The science of extracting high-value information from massive quantities of low-value data

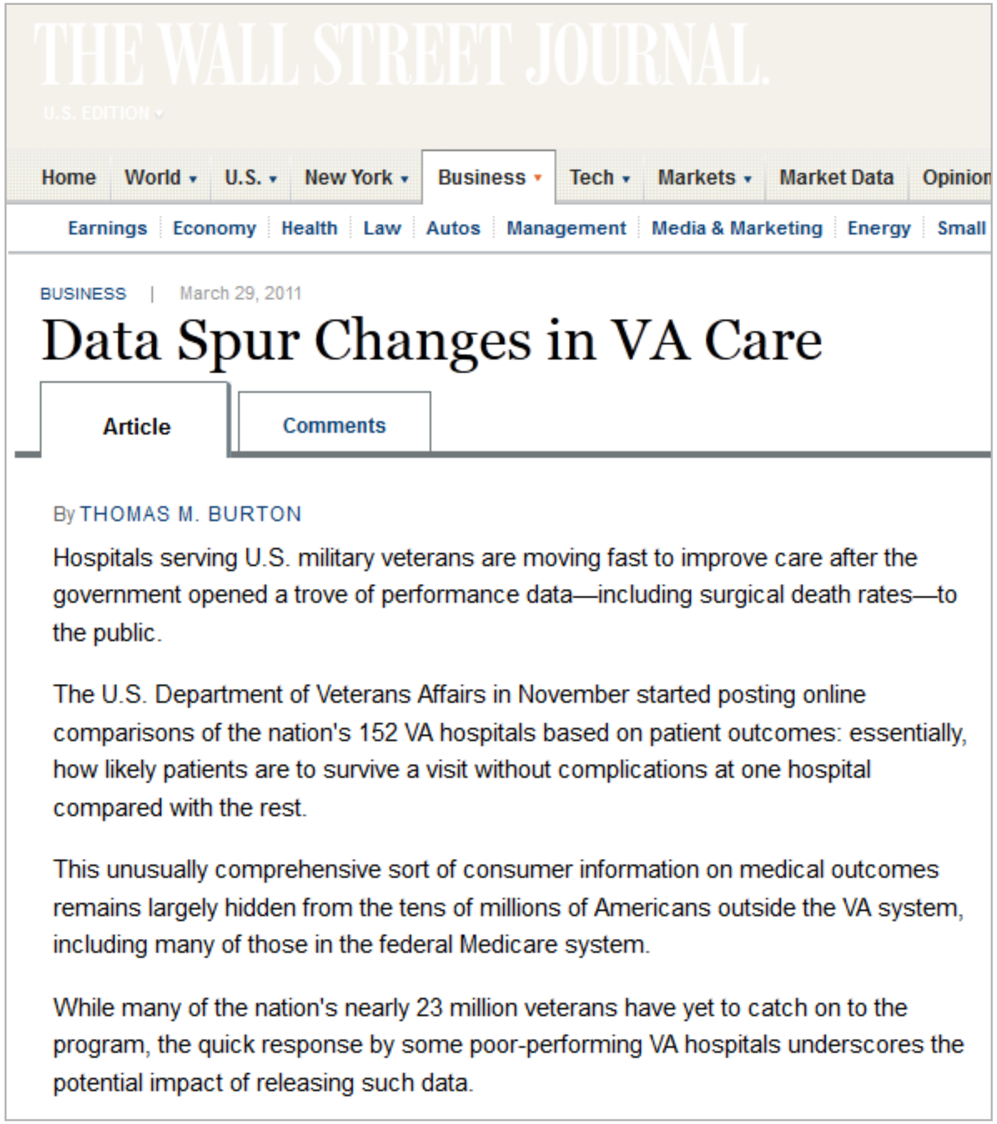
Image credits: Wikipedia, Travelocity

A massive treasure trove awaits those willing to dig deeply and broadly enough

Where is Big Data taking us?

- ~~Opener~~
- The good, the bad, and the shady
- A word of warning
- A few modest proposals

Big data in medicine and public sector



BUSINESS | March 29, 2011

Data Spur Changes in VA Care

By THOMAS M. BURTON

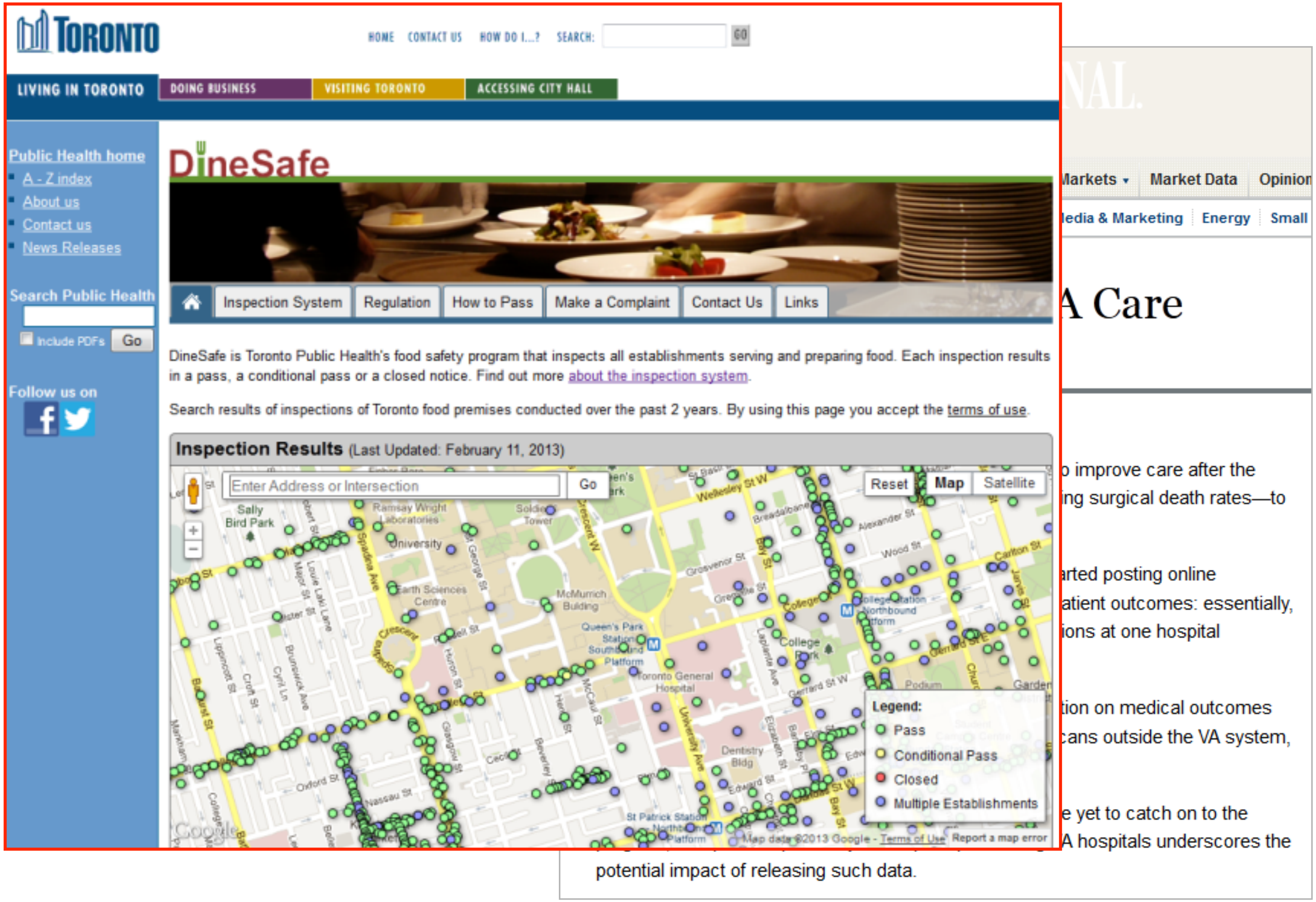
Hospitals serving U.S. military veterans are moving fast to improve care after the government opened a trove of performance data—including surgical death rates—to the public.

The U.S. Department of Veterans Affairs in November started posting online comparisons of the nation's 152 VA hospitals based on patient outcomes: essentially, how likely patients are to survive a visit without complications at one hospital compared with the rest.

This unusually comprehensive sort of consumer information on medical outcomes remains largely hidden from the tens of millions of Americans outside the VA system, including many of those in the federal Medicare system.

While many of the nation's nearly 23 million veterans have yet to catch on to the program, the quick response by some poor-performing VA hospitals underscores the potential impact of releasing such data.

Big data in medicine and public sector



TORONTO

HOME CONTACT US HOW DO I...? SEARCH: GO

LIVING IN TORONTO | DOING BUSINESS | VISITING TORONTO | ACCESSING CITY HALL

Public Health home
- A - Z index
- About us
- Contact us
- News Releases

Search Public Health

☐ Include PDFs Go

Follow us on

DineSafe

Inspection System | Regulation | How to Pass | Make a Complaint | Contact Us | Links

DineSafe is Toronto Public Health's food safety program that inspects all establishments serving and preparing food. Each inspection results in a pass, a conditional pass or a closed notice. Find out more about the inspection system.

Search results of inspections of Toronto food premises conducted over the past 2 years. By using this page you accept the terms of use.

Inspection Results (Last Updated: February 11, 2013)

Enter Address or Intersection Go | Reset | Map | Satellite

Legend:
- ○ Pass
- ○ Conditional Pass
- ○ Closed
- ○ Multiple Establishments

Map data ©2013 Google - Terms of Use Report a map error

NAL.

Markets ▾ | Market Data | Opinion

Media & Marketing | Energy | Small

A Care

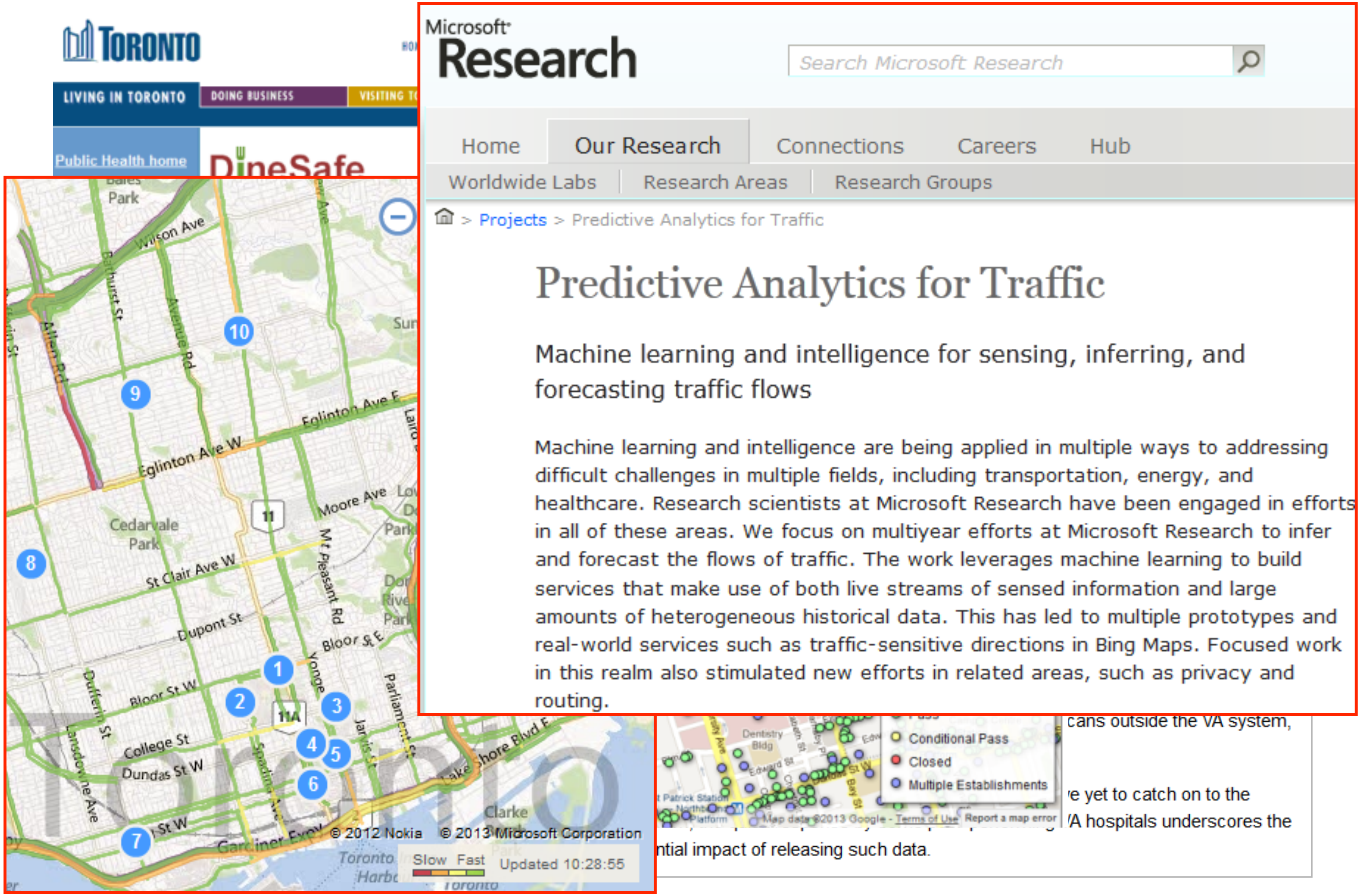
o improve care after the
ing surgical death rates—to

arted posting online
atient outcomes: essentially,
ions at one hospital

tion on medical outcomes
ans outside the VA system,

e yet to catch on to the
A hospitals underscores the
potential impact of releasing such data.

Big data in medicine and public sector



Microsoft Research

Search Microsoft Research

Home | Our Research | Connections | Careers | Hub

Worldwide Labs | Research Areas | Research Groups

> Projects > Predictive Analytics for Traffic

Predictive Analytics for Traffic

Machine learning and intelligence for sensing, inferring, and forecasting traffic flows

Machine learning and intelligence are being applied in multiple ways to addressing difficult challenges in multiple fields, including transportation, energy, and healthcare. Research scientists at Microsoft Research have been engaged in efforts in all of these areas. We focus on multiyear efforts at Microsoft Research to infer and forecast the flows of traffic. The work leverages machine learning to build services that make use of both live streams of sensed information and large amounts of heterogeneous historical data. This has led to multiple prototypes and real-world services such as traffic-sensitive directions in Bing Maps. Focused work in this realm also stimulated new efforts in related areas, such as privacy and routing.

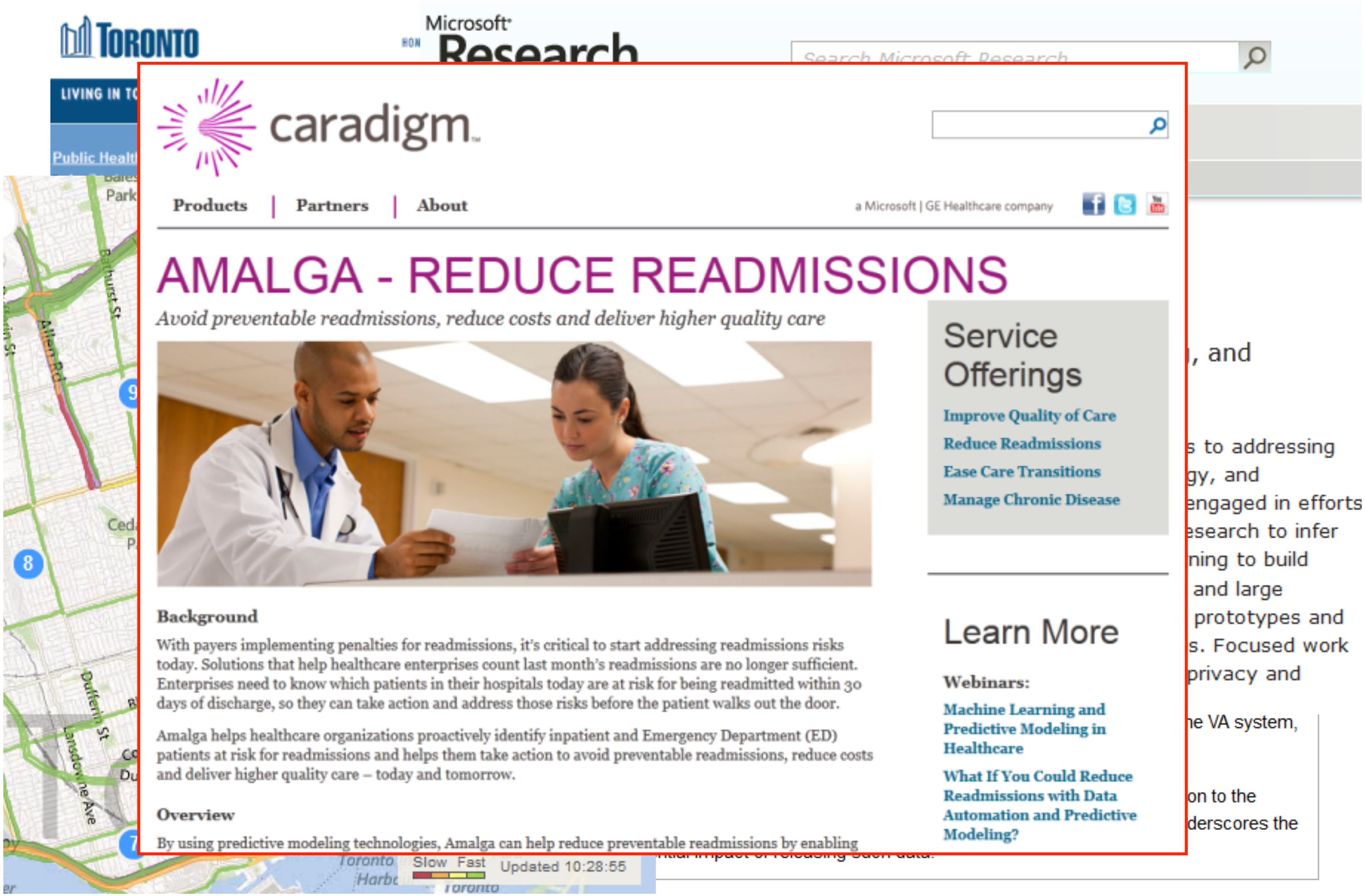
cans outside the VA system,

ve yet to catch on to the

VA hospitals underscores the

ntial impact of releasing such data.

Big data in medicine and public sector



caradigm™

Products | Partners | About

a Microsoft | GE Healthcare company

AMALGA - REDUCE READMISSIONS

Avoid preventable readmissions, reduce costs and deliver higher quality care

Background

With payers implementing penalties for readmissions, it's critical to start addressing readmissions risks today. Solutions that help healthcare enterprises count last month's readmissions are no longer sufficient. Enterprises need to know which patients in their hospitals today are at risk for being readmitted within 30 days of discharge, so they can take action and address those risks before the patient walks out the door.

Amalga helps healthcare organizations proactively identify inpatient and Emergency Department (ED) patients at risk for readmissions and helps them take action to avoid preventable readmissions, reduce costs and deliver higher quality care – today and tomorrow.

Overview

By using predictive modeling technologies, Amalga can help reduce preventable readmissions by enabling

Service Offerings

Improve Quality of Care

Reduce Readmissions

Ease Care Transitions

Manage Chronic Disease

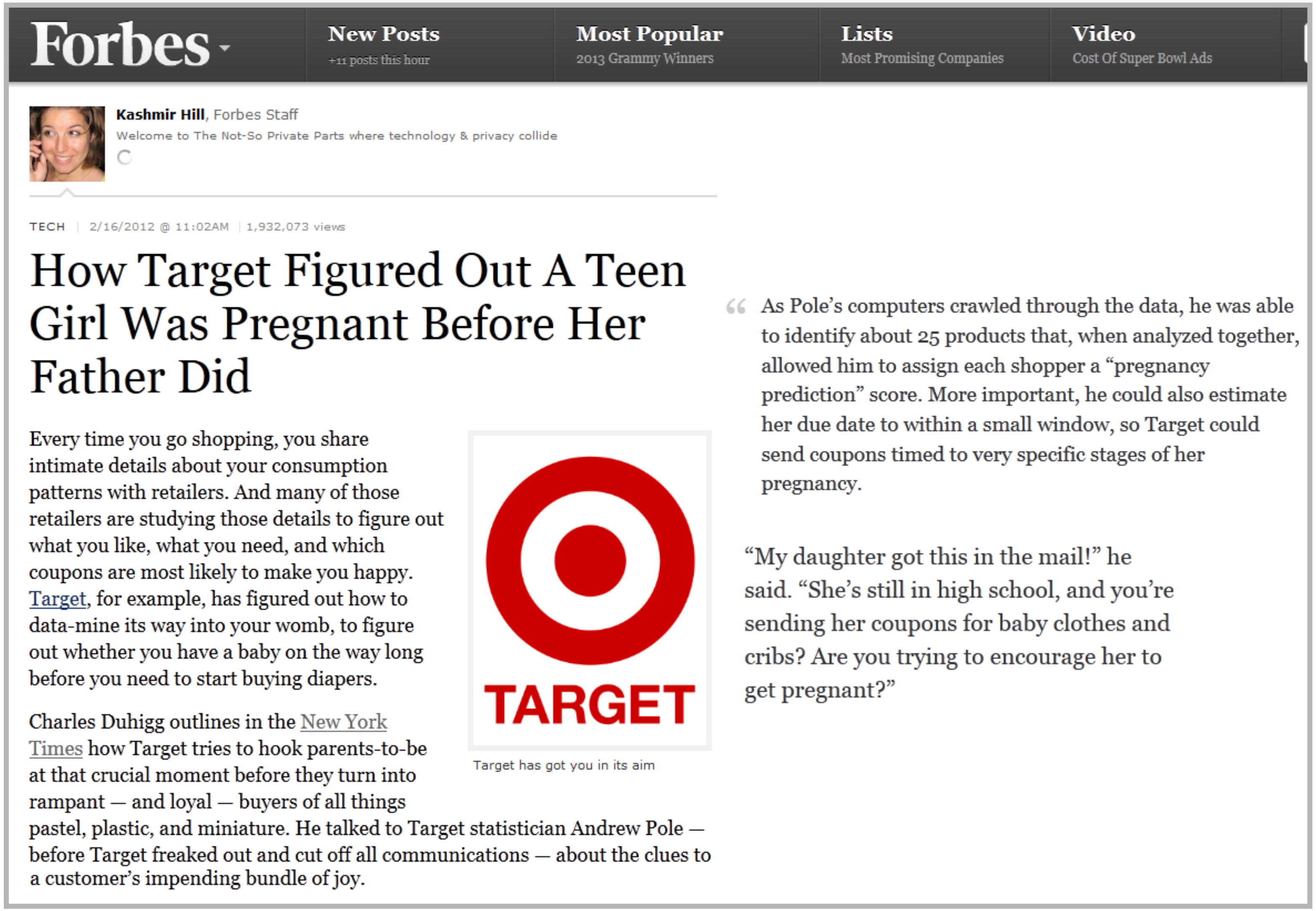
Learn More

Webinars:

Machine Learning and Predictive Modeling in Healthcare

What If You Could Reduce Readmissions with Data Automation and Predictive Modeling?

But it's not all good news…



Kashmir Hill, Forbes Staff
Welcome to The Not-So Private Parts where technology & privacy collide

TECH | 2/16/2012 @ 11:02AM | 1,932,073 views

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

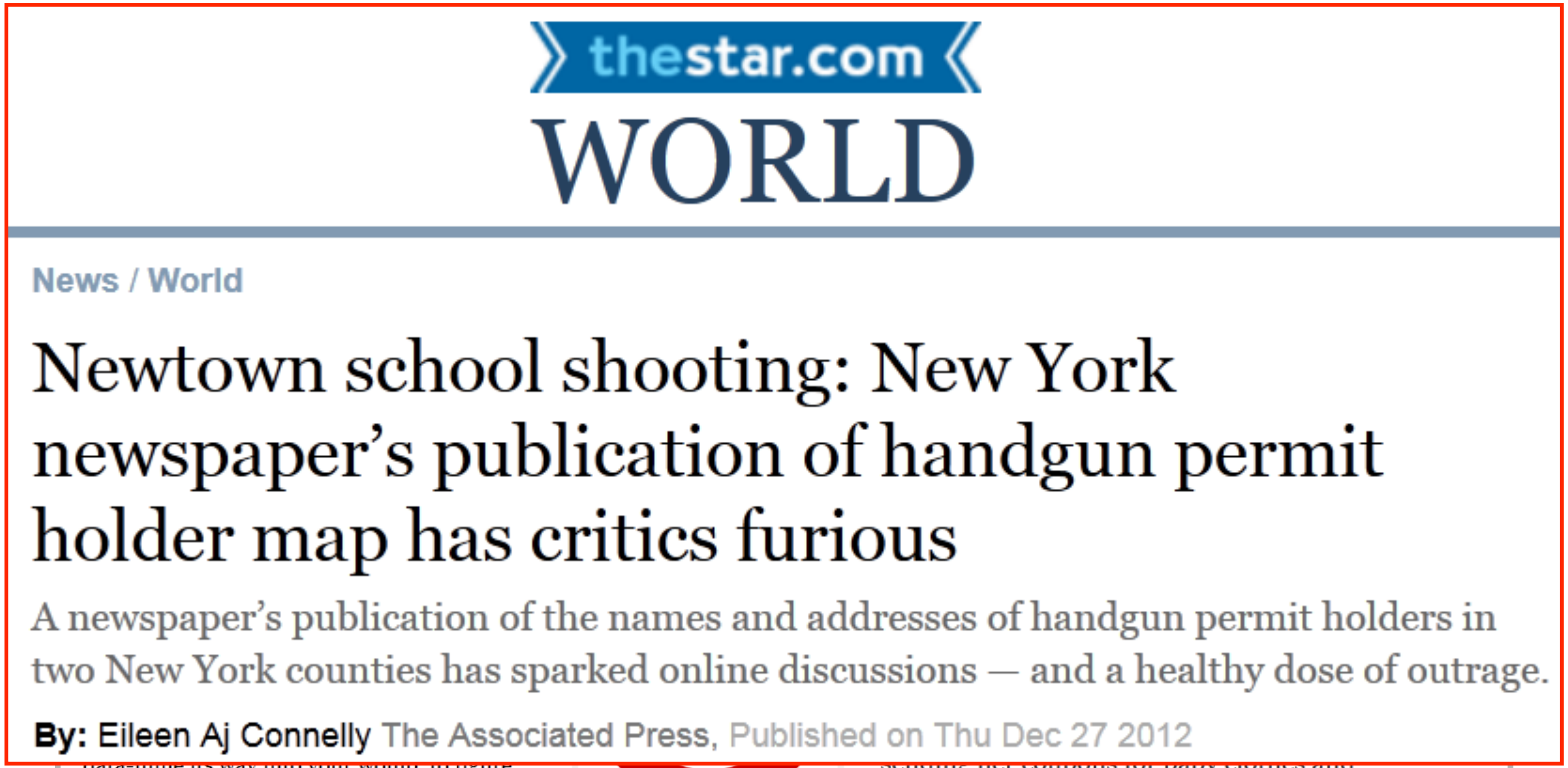
Target has got you in its aim

Charles Duhigg outlines in the New York Times how Target tries to hook parents-to-be at that crucial moment before they turn into rampant — and loyal — buyers of all things pastel, plastic, and miniature. He talked to Target statistician Andrew Pole — before Target freaked out and cut off all communications — about the clues to a customer's impending bundle of joy.

> "As Pole's computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a "pregnancy prediction" score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

"My daughter got this in the mail!" he said. "She's still in high school, and you're sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?"

But it’s not all good news…



thestar.com

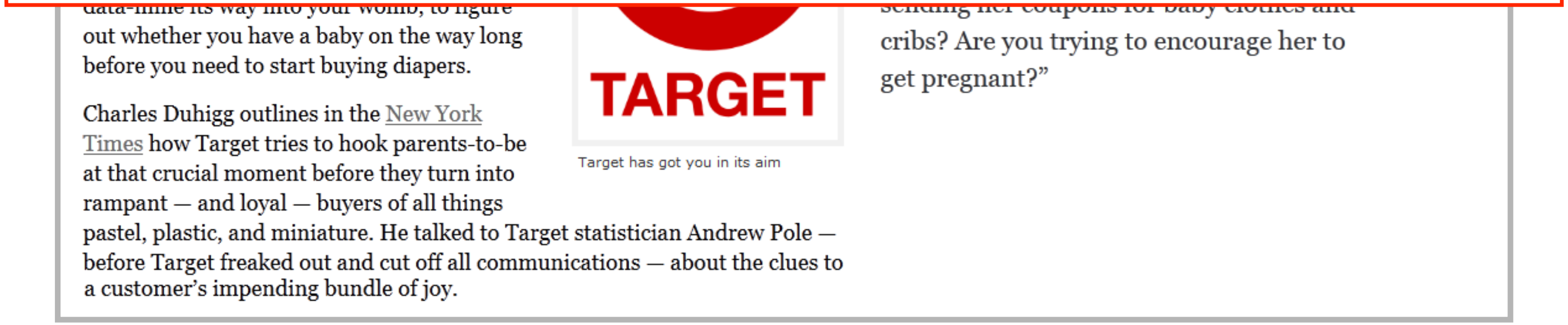
WORLD

News / World

Newtown school shooting: New York newspaper’s publication of handgun permit holder map has critics furious

A newspaper’s publication of the names and addresses of handgun permit holders in two New York counties has sparked online discussions — and a healthy dose of outrage.

By: Eileen Aj Connelly The Associated Press, Published on Thu Dec 27 2012



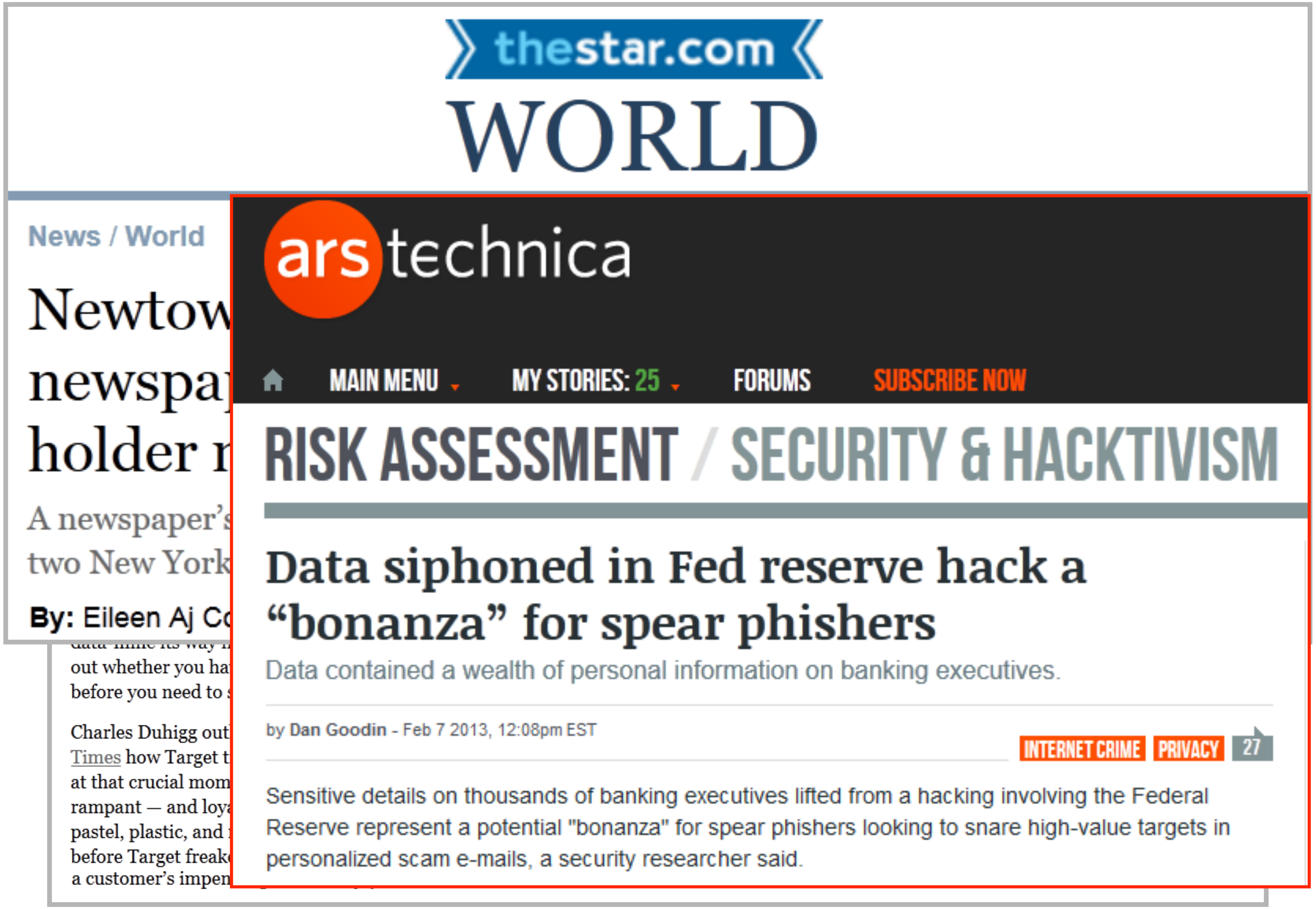
data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

Charles Duhigg outlines in the New York Times how Target tries to hook parents-to-be at that crucial moment before they turn into rampant — and loyal — buyers of all things pastel, plastic, and miniature. He talked to Target statistician Andrew Pole — before Target freaked out and cut off all communications — about the clues to a customer’s impending bundle of joy.

Target has got you in its aim

sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?”

But it’s not all good news…



thestar.com

WORLD

News / World

Newtow

newspa

holder r

A newspaper’s

two New York

By: Eileen Aj Co

out whether you ha

before you need to s

Charles Duhigg out

Times how Target t

at that crucial mom

rampant — and loya

pastel, plastic, and

before Target freak

a customer’s impen

ars technica

RISK ASSESSMENT / SECURITY & HACKTIVISM

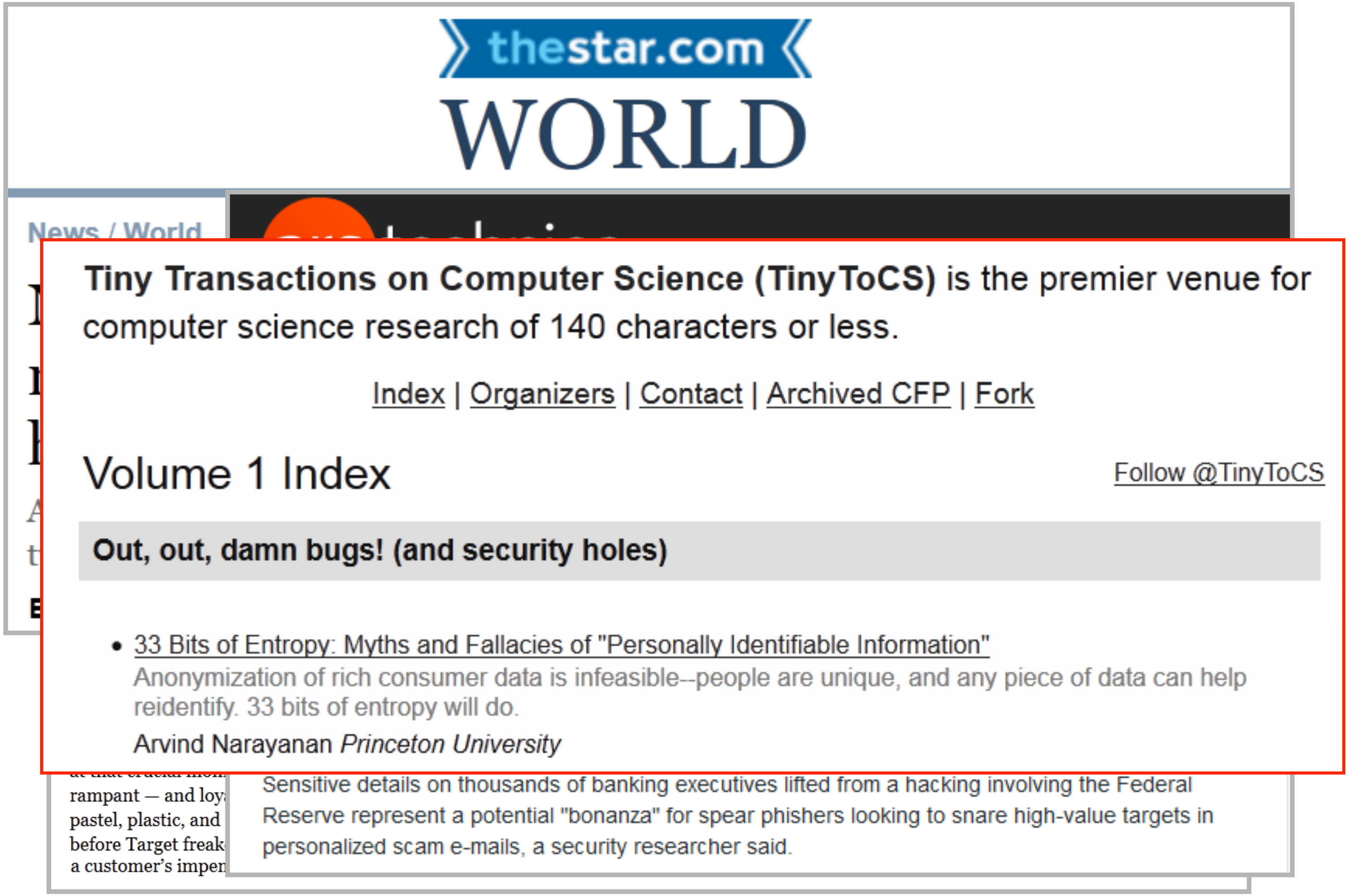
Data siphoned in Fed reserve hack a “bonanza” for spear phishers

Data contained a wealth of personal information on banking executives.

by Dan Goodin - Feb 7 2013, 12:08pm EST

Sensitive details on thousands of banking executives lifted from a hacking involving the Federal Reserve represent a potential "bonanza" for spear phishers looking to snare high-value targets in personalized scam e-mails, a security researcher said.

But it's not all good news…



Tiny Transactions on Computer Science (TinyToCS) is the premier venue for computer science research of 140 characters or less.

Index | Organizers | Contact | Archived CFP | Fork

Volume 1 Index

Follow @TinyToCS

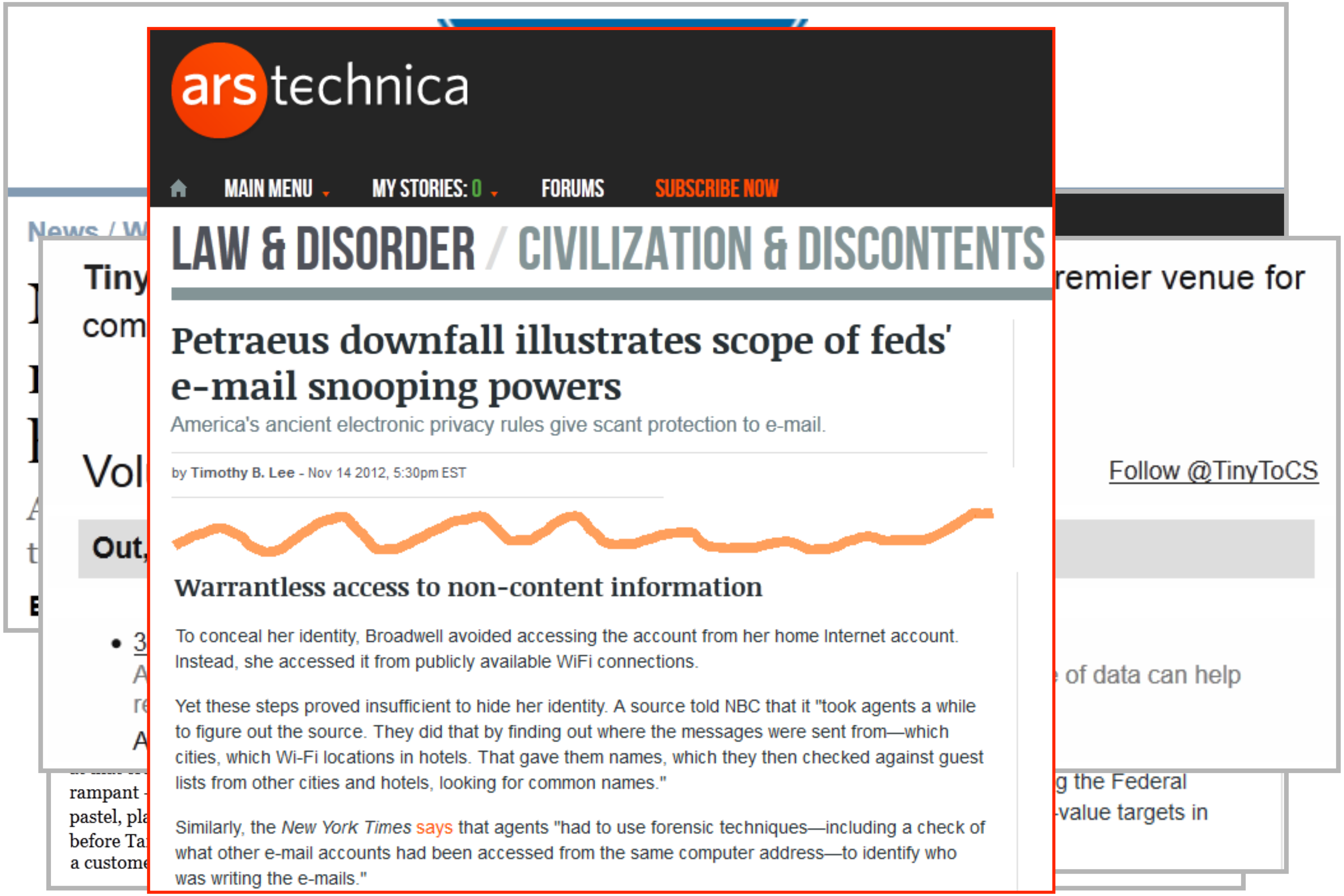
Out, out, damn bugs! (and security holes)

- 33 Bits of Entropy: Myths and Fallacies of "Personally Identifiable Information"
  Anonymization of rich consumer data is infeasible--people are unique, and any piece of data can help reidentify. 33 bits of entropy will do.
  Arvind Narayanan *Princeton University*

rampant — and loy
pastel, plastic, and
before Target freak
a customer's imper

Sensitive details on thousands of banking executives lifted from a hacking involving the Federal Reserve represent a potential "bonanza" for spear phishers looking to snare high-value targets in personalized scam e-mails, a security researcher said.

But it's not all good news…



LAW & DISORDER / CIVILIZATION & DISCONTENTS

Petraeus downfall illustrates scope of feds' e-mail snooping powers

America's ancient electronic privacy rules give scant protection to e-mail.

by Timothy B. Lee - Nov 14 2012, 5:30pm EST

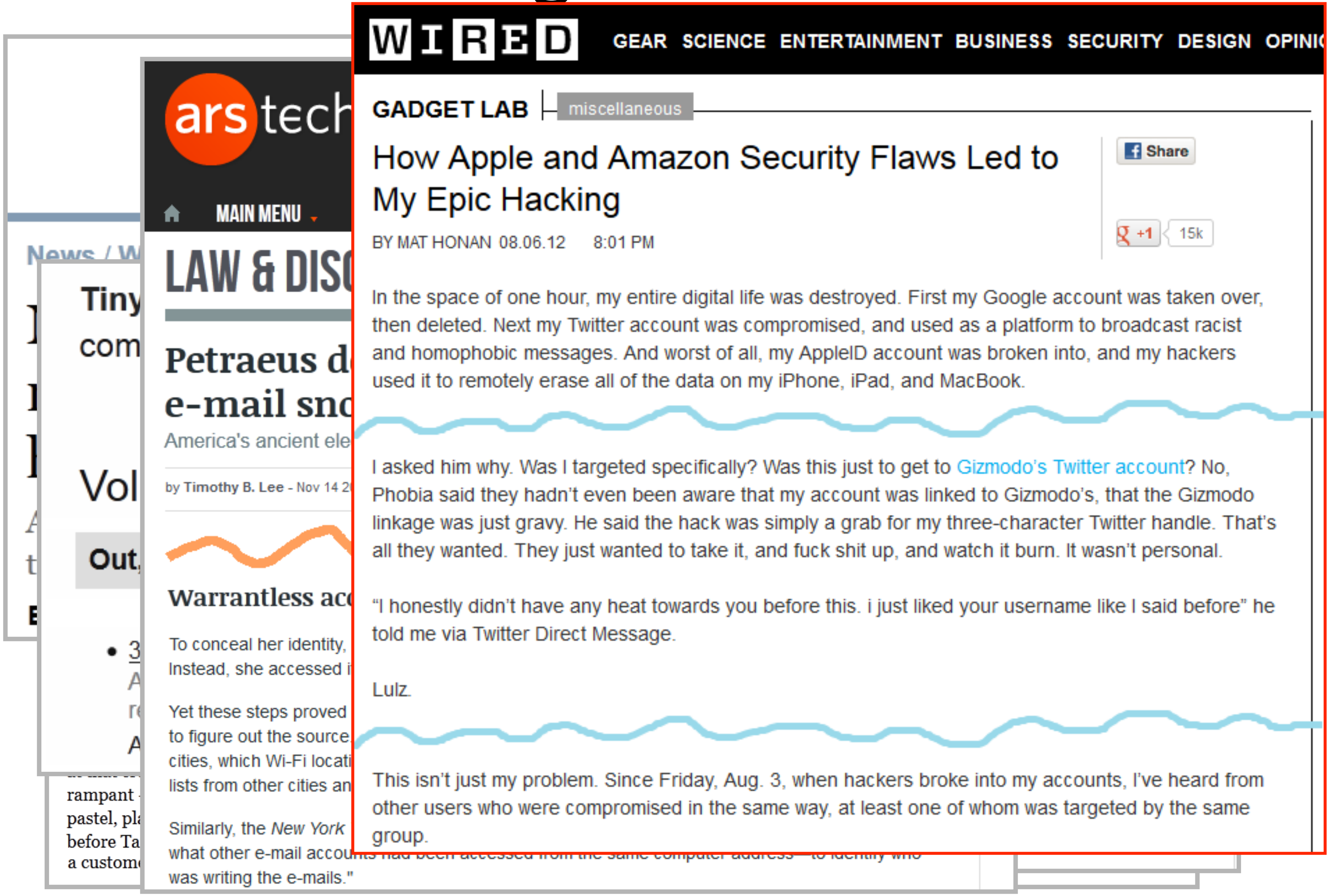
Warrantless access to non-content information

To conceal her identity, Broadwell avoided accessing the account from her home Internet account. Instead, she accessed it from publicly available WiFi connections.

Yet these steps proved insufficient to hide her identity. A source told NBC that it "took agents a while to figure out the source. They did that by finding out where the messages were sent from—which cities, which Wi-Fi locations in hotels. That gave them names, which they then checked against guest lists from other cities and hotels, looking for common names."

Similarly, the *New York Times* says that agents "had to use forensic techniques—including a check of what other e-mail accounts had been accessed from the same computer address—to identify who was writing the e-mails."

But it's not all good news…



News / W

Tiny

com

Vol

Out,

rampant

pastel, pl

before Ta

a custom

ars tech

MAIN MENU

LAW & DISC

Petraeus d e-mail snc

America's ancient ele

by Timothy B. Lee - Nov 14 2

Warrantless ac

To conceal her identity, Instead, she accessed i

Yet these steps proved to figure out the source cities, which Wi-Fi locati lists from other cities an

Similarly, the *New York* what other e-mail accounts had been accessed from the same computer address—to identify who was writing the e-mails."

WIRED

How Apple and Amazon Security Flaws Led to My Epic Hacking

BY MAT HONAN 08.06.12 8:01 PM

In the space of one hour, my entire digital life was destroyed. First my Google account was taken over, then deleted. Next my Twitter account was compromised, and used as a platform to broadcast racist and homophobic messages. And worst of all, my AppleID account was broken into, and my hackers used it to remotely erase all of the data on my iPhone, iPad, and MacBook.

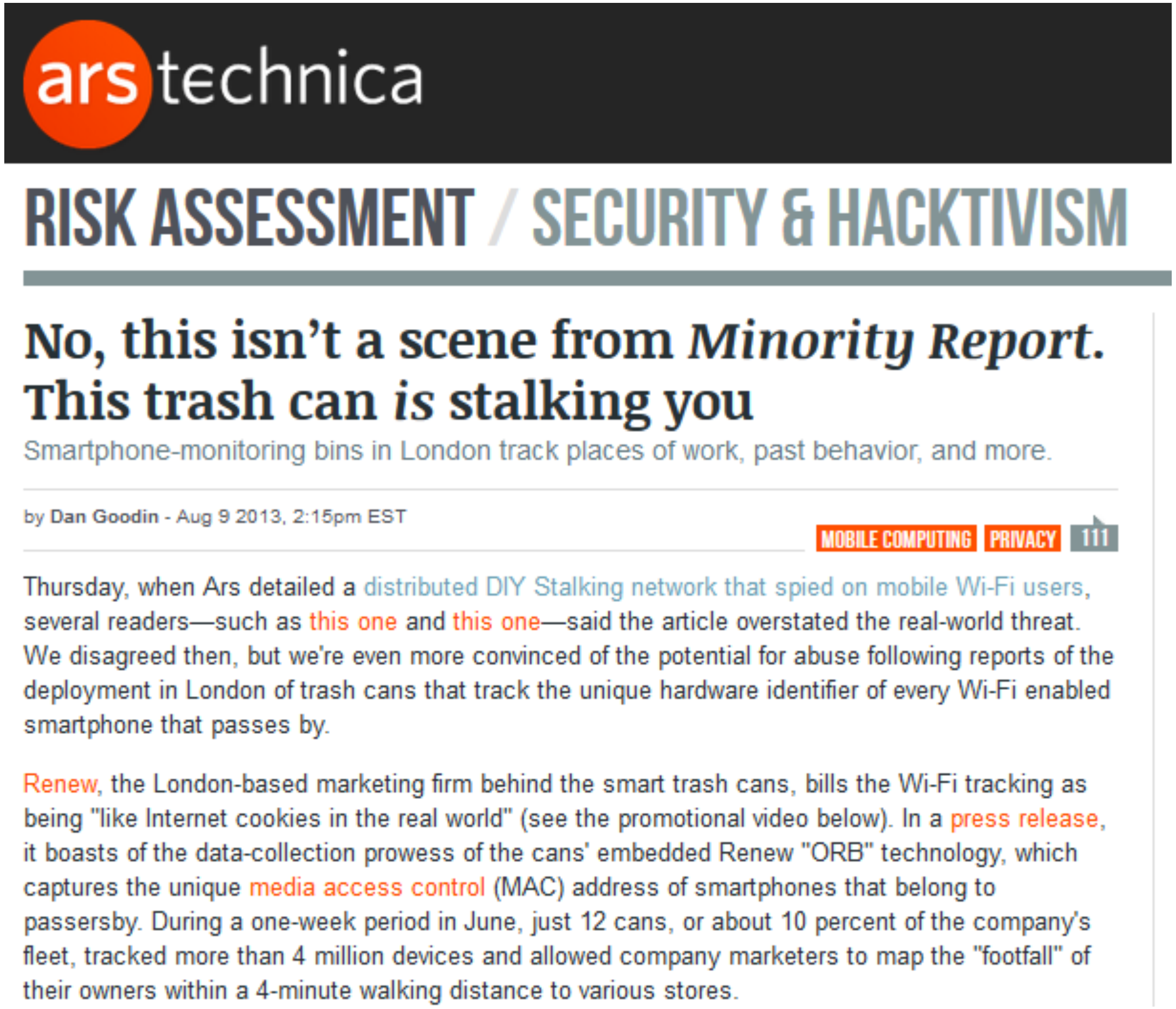
I asked him why. Was I targeted specifically? Was this just to get to Gizmodo's Twitter account? No, Phobia said they hadn't even been aware that my account was linked to Gizmodo's, that the Gizmodo linkage was just gravy. He said the hack was simply a grab for my three-character Twitter handle. That's all they wanted. They just wanted to take it, and fuck shit up, and watch it burn. It wasn't personal.

"I honestly didn't have any heat towards you before this. i just liked your username like I said before" he told me via Twitter Direct Message.

Lulz.

This isn't just my problem. Since Friday, Aug. 3, when hackers broke into my accounts, I've heard from other users who were compromised in the same way, at least one of whom was targeted by the same group.

Plenty of gray in the middle, too



RISK ASSESSMENT / SECURITY & HACKTIVISM

No, this isn't a scene from *Minority Report*. This trash can *is* stalking you

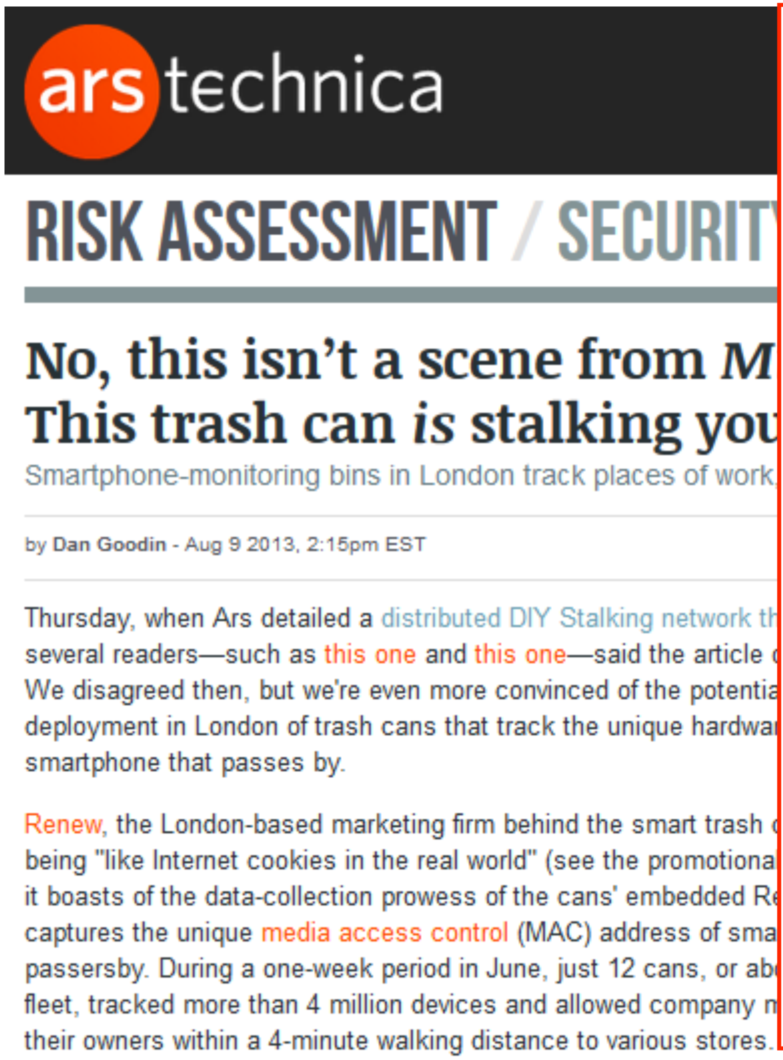
Smartphone-monitoring bins in London track places of work, past behavior, and more.

by Dan Goodin - Aug 9 2013, 2:15pm EST

Thursday, when Ars detailed a distributed DIY Stalking network that spied on mobile Wi-Fi users, several readers—such as this one and this one—said the article overstated the real-world threat. We disagreed then, but we're even more convinced of the potential for abuse following reports of the deployment in London of trash cans that track the unique hardware identifier of every Wi-Fi enabled smartphone that passes by.

Renew, the London-based marketing firm behind the smart trash cans, bills the Wi-Fi tracking as being "like Internet cookies in the real world" (see the promotional video below). In a press release, it boasts of the data-collection prowess of the cans' embedded Renew "ORB" technology, which captures the unique media access control (MAC) address of smartphones that belong to passersby. During a one-week period in June, just 12 cans, or about 10 percent of the company's fleet, tracked more than 4 million devices and allowed company marketers to map the "footfall" of their owners within a 4-minute walking distance to various stores.

Plenty of gray in the middle, too



ars technica

RISK ASSESSMENT / SECURIT

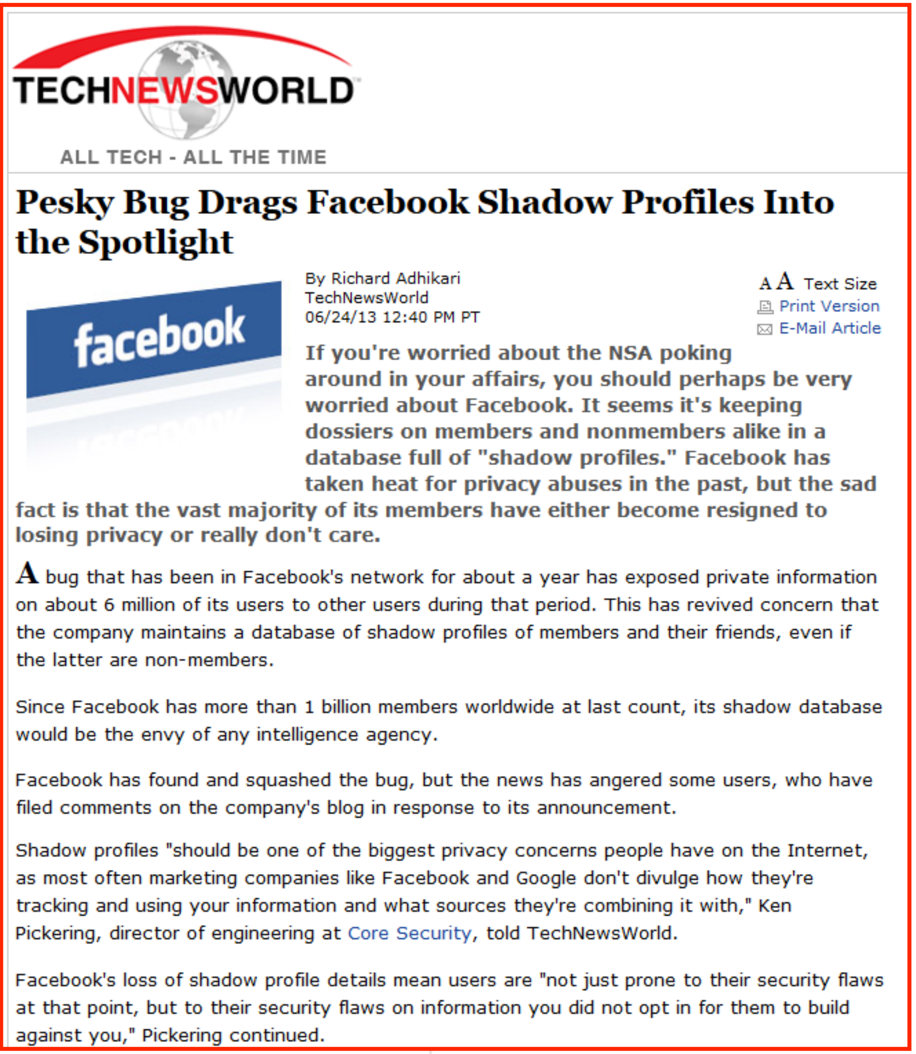
No, this isn't a scene from M This trash can *is* stalking you

Smartphone-monitoring bins in London track places of work

by Dan Goodin - Aug 9 2013, 2:15pm EST

Thursday, when Ars detailed a distributed DIY Stalking network th several readers—such as this one and this one—said the article c We disagreed then, but we're even more convinced of the potentia deployment in London of trash cans that track the unique hardwa smartphone that passes by.

Renew, the London-based marketing firm behind the smart trash c being "like Internet cookies in the real world" (see the promotiona it boasts of the data-collection prowess of the cans' embedded Re captures the unique media access control (MAC) address of sma passersby. During a one-week period in June, just 12 cans, or ab fleet, tracked more than 4 million devices and allowed company m their owners within a 4-minute walking distance to various stores.



TECHNEWSWORLD

ALL TECH - ALL THE TIME

Pesky Bug Drags Facebook Shadow Profiles Into the Spotlight

By Richard Adhikari
TechNewsWorld
06/24/13 12:40 PM PT

If you're worried about the NSA poking around in your affairs, you should perhaps be very worried about Facebook. It seems it's keeping dossiers on members and nonmembers alike in a database full of "shadow profiles." Facebook has taken heat for privacy abuses in the past, but the sad fact is that the vast majority of its members have either become resigned to losing privacy or really don't care.

A bug that has been in Facebook's network for about a year has exposed private information on about 6 million of its users to other users during that period. This has revived concern that the company maintains a database of shadow profiles of members and their friends, even if the latter are non-members.

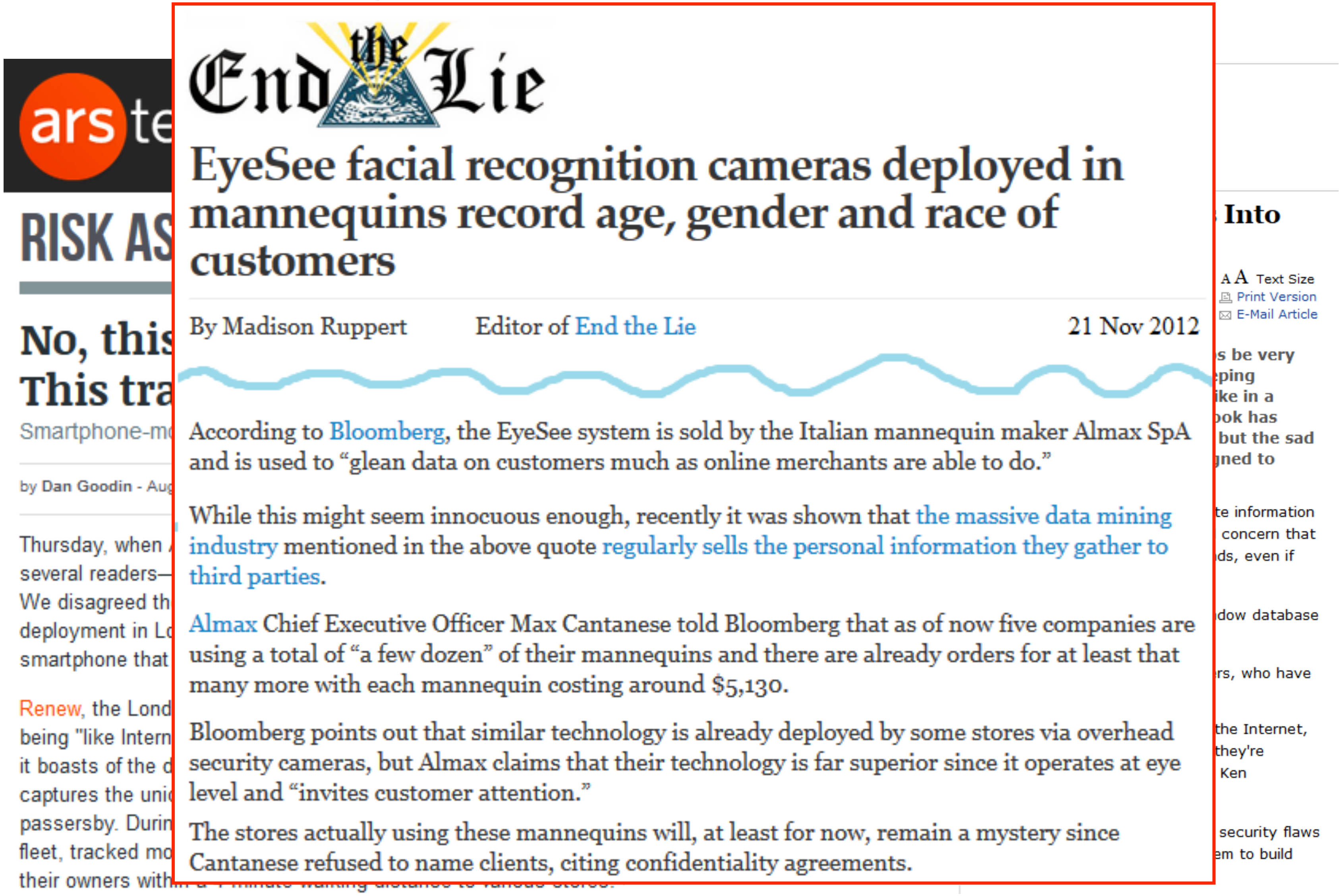
Since Facebook has more than 1 billion members worldwide at last count, its shadow database would be the envy of any intelligence agency.

Facebook has found and squashed the bug, but the news has angered some users, who have filed comments on the company's blog in response to its announcement.

Shadow profiles "should be one of the biggest privacy concerns people have on the Internet, as most often marketing companies like Facebook and Google don't divulge how they're tracking and using your information and what sources they're combining it with," Ken Pickering, director of engineering at Core Security, told TechNewsWorld.

Facebook's loss of shadow profile details mean users are "not just prone to their security flaws at that point, but to their security flaws on information you did not opt in for them to build against you," Pickering continued.

Plenty of gray in the middle, too



End the Lie

EyeSee facial recognition cameras deployed in mannequins record age, gender and race of customers

By Madison Ruppert Editor of End the Lie 21 Nov 2012

According to Bloomberg, the EyeSee system is sold by the Italian mannequin maker Almax SpA and is used to “glean data on customers much as online merchants are able to do.”

While this might seem innocuous enough, recently it was shown that the massive data mining industry mentioned in the above quote regularly sells the personal information they gather to third parties.

Almax Chief Executive Officer Max Cantanese told Bloomberg that as of now five companies are using a total of “a few dozen” of their mannequins and there are already orders for at least that many more with each mannequin costing around $5,130.

Bloomberg points out that similar technology is already deployed by some stores via overhead security cameras, but Almax claims that their technology is far superior since it operates at eye level and “invites customer attention.”

The stores actually using these mannequins will, at least for now, remain a mystery since Cantanese refused to name clients, citing confidentiality agreements.

ars te

RISK AS

No, this
This tra

Smartphone-m

by Dan Goodin - Aug

Thursday, when
several readers—
We disagreed th
deployment in Lo
smartphone that

Renew, the Lond
being "like Intern
it boasts of the d
captures the uni
passersby. Durin
fleet, tracked mo
their owners with a 1-minute walking distance to various stores.

Into

A A Text Size
Print Version
E-Mail Article

s be very
ping
ike in a
ook has
but the sad
ned to

te information
concern that
ds, even if

dow database

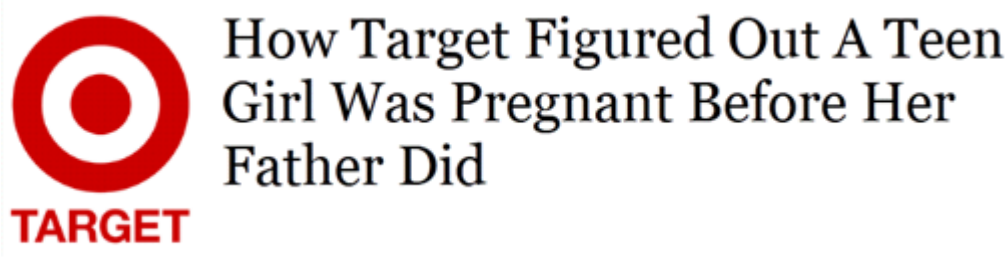
rs, who have

the Internet,
they're
Ken

security flaws
m to build

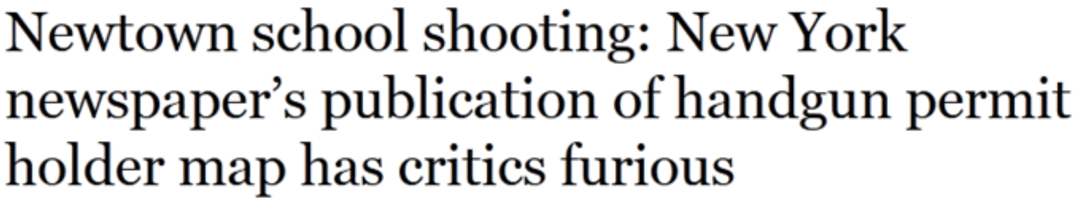
Quis custodiet ipsos big data custodes?

- Privacy



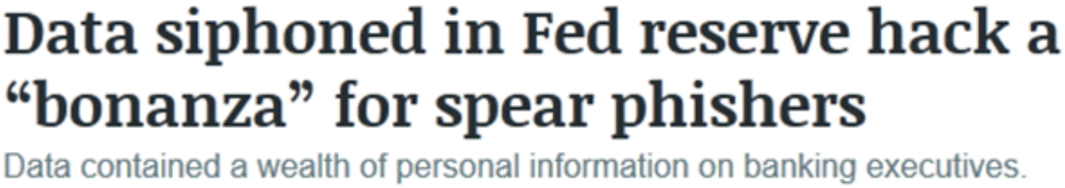
How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

- Transparency



Newtown school shooting: New York newspaper's publication of handgun permit holder map has critics furious

- Security



Data siphoned in Fed reserve hack a "bonanza" for spear phishers

Data contained a wealth of personal information on banking executives.

Victim has little recourse if anything goes wrong

Where is Big Data taking us?

- Opener
- ~~The good, the bad, and the shady~~
- A word of warning
- A few modest proposals

The [data] dwarves dug too greedily and too deep…

© New Line Cinema (The Lord of the Rings)

You know what they awoke … shadow and flame

Facing the Big Data Balrog

- Flame (collateral damage, backlash)
  - Users, corporations, government all victims
  - Burdensome regulations, hostile users
- Shadow (disincentives to transparency)
  - Data sharers vulnerable to real harm
  - Trend is to hide/hoard data

Where is Big Data taking us?

- Opener
- ~~The good, the bad, and the shady~~
- ~~A word of warning~~
- A few modest proposals

Lots of non-tech issues

- Expectations of privacy vs. reality
  - Everyone in celebrity spotlight
  - We all “leak” info
- Jurisdiction
  - Multi-continent crimes
  - Governments wear multiple hats
- Legal/regulatory
  - “Sunshine laws”
  - Inconsistent/conflicting regulations

*How can big data technology **help**?*

Three modest proposals

- Strong privilege separation
- External audit/monitoring
- Statistically plausible ignorance

Protections, recourse for data sharers

1. Privilege separation

- Homomorphic encryption (e.g. CryptDB++)
  - Data curator: log everything, decrypt nothing
  - Analyst: capped query frequency/size
  - Intruder: limited flexibility

- See also
  - Perfect forward secrecy
  - Commutative encryption, secret sharing
  - Trusted HW (e.g. CipherBase)

No one holds "keys to the kingdom" *[SIGMOD'13 keynote]*

2. Monitoring and auditing

- Monitoring
  - Mine those logs!
  - Suspicious queries, data transfers
  - How to train it? False positives/negatives?
- Privacy-preserving audit
  - Auditor: sees queries, data touched, no results
  - Composable encryption?

Point big data at itself!

3. Statistically plausible ignorance

- "Consistency is not accidental"
- Mine for actions driven by forbidden data
  - Targeted ads based on shadow profiles?
  - Discriminatory hiring practices?
  - Biased/selective enforcement of rules?
- Builds on previous techniques

What you don't know can't hurt me

Conclusions

- Big Data: real gold... real risks
- Issues: **P**rivacy, **S**ecurity, **T**ransparency

- Three proposals
  - Encryption for privilege separation (P/S)
  - Monitoring and auditing (S/T)
  - Statistically plausible ignorance (P/T)

- Point big data at itself!