Imperial College
London

COMPUTATIONAL
PRIVACY
GROUP

Privacy Engineering

Week 0 - Intro to Privacy

Dr. Yves-Alexandre de Montjoye

Lecturer in Dept. of Computing jointly with the Data Science Institute

- PhD from the MIT Media Lab
- MScs in Applied Mathematics from Université catholique de Louvain, Ecole Centrale Paris, and Katholieke Universiteit Leuven
- BSc in Engineering from Université catholique de Louvain

Head the Computational Privacy Group (http**s**://cpg.doc.ic.ac.uk/)

1. Design re-identification and profiling algorithms and analysis (“red team”)
2. Develop safe, often interactive, mechanisms for the privacy-conscientious use of large-scale behavioral datasets (“blue team”)
3. Define formal metrics of fairness in AI

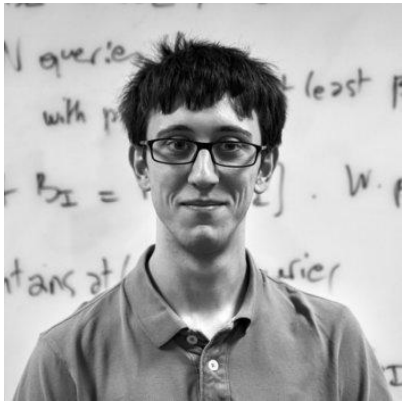
GTAs

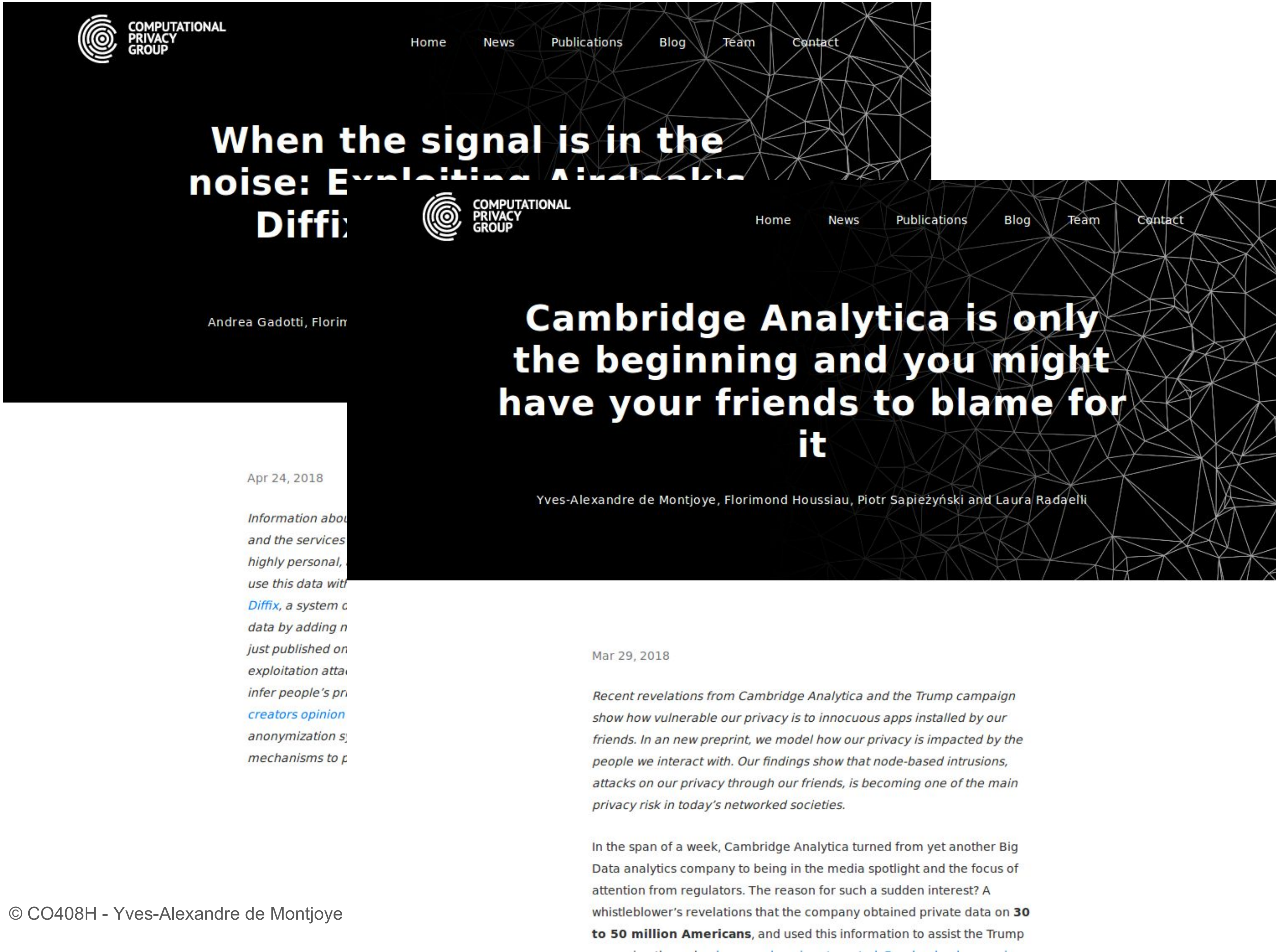
Andrea Gadotti

Andrea received his MSc in mathematical logic from the University of Turin. His research interests include differential privacy, determining vulnerabilities in data-release systems, and designing privacy-preserving mechanisms. E-mail: *gadotti@imperial.ac.uk*

Florimond Houssiau

Florimond received his MSc in applied mathematics from UCLouvain. His research interests include obfuscation and privacy risks in web search data.
E-Mail: *florimond@imperial.ac.uk*





COMPUTATIONAL PRIVACY GROUP

Home News Publications Blog Team Contact

When the signal is in the noise: Exploiting Aircloak's Diffi

Andrea Gadotti, Florim

Apr 24, 2018

Information abou
and the services
highly personal,
use this data wit
Diffix, a system o
data by adding n
just published on
exploitation atta
infer people's pr
creators opinion
anonymization sy
mechanisms to p

COMPUTATIONAL PRIVACY GROUP

Home News Publications Blog Team Contact

Cambridge Analytica is only the beginning and you might have your friends to blame for it

Yves-Alexandre de Montjoye, Florimond Houssiau, Piotr Sapieżyński and Laura Radaelli

Mar 29, 2018

Recent revelations from Cambridge Analytica and the Trump campaign show how vulnerable our privacy is to innocuous apps installed by our friends. In an new preprint, we model how our privacy is impacted by the people we interact with. Our findings show that node-based intrusions, attacks on our privacy through our friends, is becoming one of the main privacy risk in today's networked societies.

In the span of a week, Cambridge Analytica turned from yet another Big Data analytics company to being in the media spotlight and the focus of attention from regulators. The reason for such a sudden interest? A whistleblower's revelations that the company obtained private data on **30 to 50 million Americans**, and used this information to assist the Trump

Why Privacy?

Please raise your hand

Now, lower it if you would not want this data about you to be publicly available:

1. The title of the last 5 films you have watched
2. All journeys you've made using your Oyster card
3. All your credit card purchases
4. The URL of every website you’ve visited
5. The list of all the drugs and procedures you received at a hospital

All of these happened...

1. The title of the last 5 films you have watched **(Narayanan, 2009)**
   - https://arxiv.org/pdf/cs/0610105.pdf
2. All journeys you've made using your Oyster card **(Lavrenovs, 2016)**
   - http://ieeexplore.ieee.org/document/7821808/
3. All your credit card purchases **(de Montjoye, 2015)**
   - https://bits.blogs.nytimes.com/2015/01/29/with-a-few-bits-of-data-researchers-identify-anonymous-people/
4. The URL of every website you’ve visited **(Eckert, 2016)**
   - http://www.bbc.com/news/technology-40770393
5. The list of all the drugs and procedures you received at a hospital **(La Libre, 2016)**
   - http://www.lalibre.be/actu/belgique/les-donnees-de-patients-vendues-par-des-hopitaux-elles-doivent-etre-protegees-reagit-de-block-59d71006cd70be70bcd3ef3b

In this course, you'll learn how to perform attacks on privacy (datasets, protocolos, etc) and defend against them

Week 0: What privacy is and why (we think) it matters in today's world

Week 1: Data pseudonymization and anonymization

Secret formulas, hash functions and lookup tables, k-anonymity, etc

Week 2: Anonymization revisited, privacy of big data

Unicity, data generalization, matching and profiling attacks, privacy of unstructured data, etc

Week 3: Query-based systems

Privacy of query-based systems (e.g. SQL), averaging and intersection attacks, etc

Week 4: Formal guarantee for privacy

Differential privacy (2017 Gödel Prize), Laplace mechanism, group privacy, etc

Lectures on Thu (2-4pm in Huxley 139) and exercise sessions the next Tue (11am-2pm in Huxley 202/206)

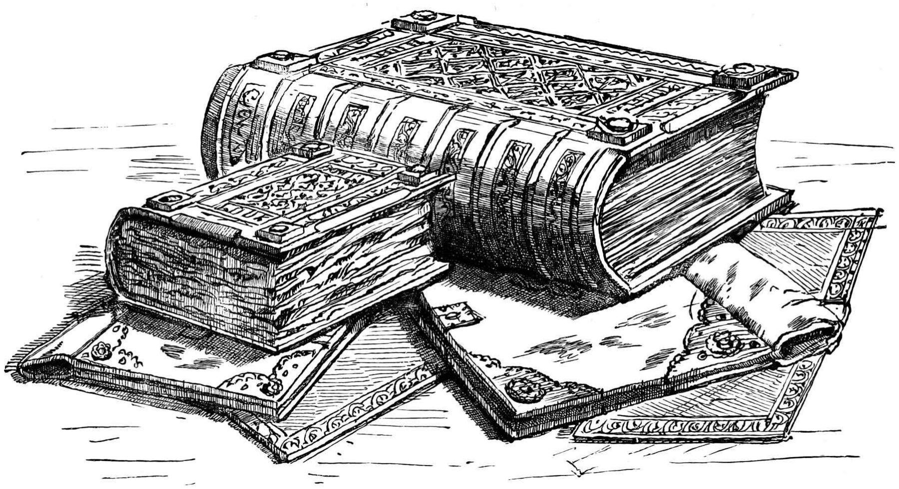
A step back

Privacy, a new notion?

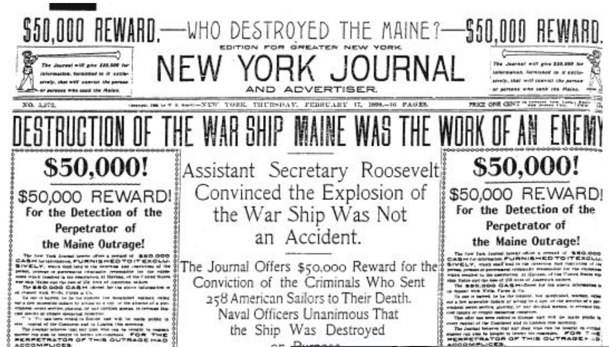
The Doomesday Book is a manuscript record of the "Great Survey" of much of England and parts of Wales completed in 1086 by order of King William the Conqueror.

The survey's main purpose was to determine what taxes were owed to the king.

The book is metaphorically called, Domesday, i.e., the Day of Judgement. For as the sentence of that strict and terrible last account that cannot be evaded



Privacy and technology



$50,000 REWARD.—WHO DESTROYED THE MAINE?—$50,000 REWARD.

EDITION FOR GREATER NEW YORK.

NEW YORK JOURNAL

AND ADVERTISER.

NEW YORK, THURSDAY, FEBRUARY 17, 1898.—16 PAGES. PRICE ONE CENT

DESTRUCTION OF THE WAR SHIP MAINE WAS THE WORK OF AN ENEMY

$50,000!

$50,000 REWARD! For the Detection of the Perpetrator of the Maine Outrage!

Assistant Secretary Roosevelt Convinced the Explosion of the War Ship Was Not an Accident.

The Journal Offers $50,000 Reward for the Conviction of the Criminals Who Sent 258 American Sailors to Their Death. Naval Officers Unanimous That the Ship Was Destroyed on Purpose

$50,000!

$50,000 REWARD! For the Detection of the Perpetrator of the Maine Outrage!

NAVAL OFFICERS THINK THE MAINE

Hidden Mine or a Sunken Torpedo Believed to Have Been
and Men Tell Thrilling Stories of Being Blown Into
Shells---Survivors Brought to Key West Sc
test Too Much---Our Cabinet Ord
Divers to Havana to Report

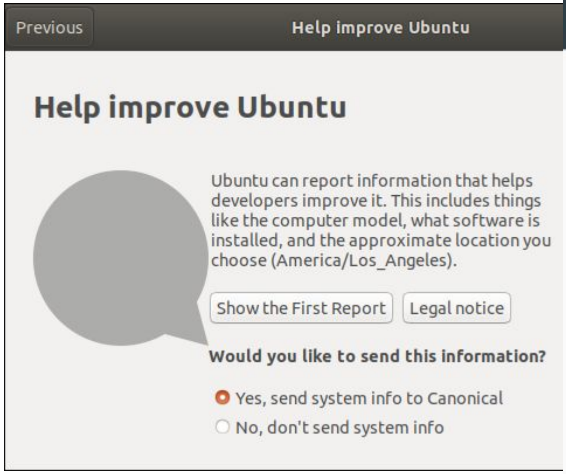
With the law adapting to new technologies

- 1604: Home privacy as early as in the Semayne's Case and later in the Fourth amendment
  - "*tender a regard to the immunity of a man's house that it stiles it his castle, and will never suffer it to be violated with impunity.*"
- 1782: Privacy of (snail) mail with the US Congress passing the a law to protect the secret of correspondences
- 1880: The Telegraph
  - After the Civil War, Congress began to seek access to telegraph messages maintained by Western Union for various investigations. This raised a considerable outcry among some members of Congress with a New York Times editorial calling the practice as "an outrage upon the liberties of the citizen". -- Solove
- 1890: The right to be let alone, first theory of privacy by Samuel Warren and Louis Brandeis in Harvard Law Review
- Modern privacy: the right to selective disclosure

Modern privacy: Right to selective disclosure

"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." - A. Westin and L. Blom-Cooper, *Privacy and freedom*, 1970.

Don't fall in the common fallacy that privacy is about data not being collected or shared. Modern (informational) privacy is about the individual having (meaningful) control over information about him or her (incl. "baseline")



Previous | Help improve Ubuntu

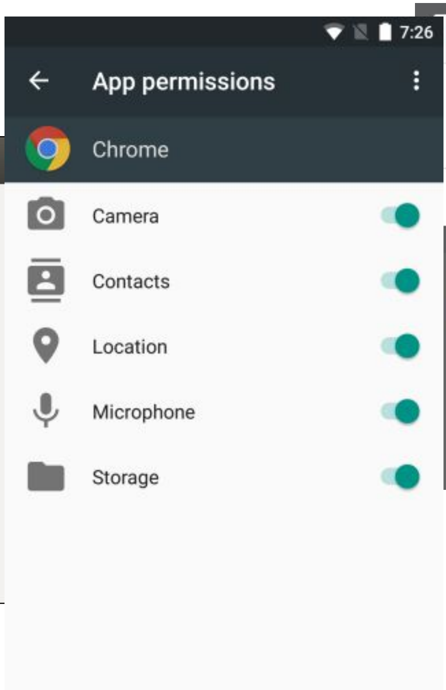
Help improve Ubuntu

Ubuntu can report information that helps developers improve it. This includes things like the computer model, what software is installed, and the approximate location you choose (America/Los_Angeles).

Show the First Report | Legal notice

Would you like to send this information?

● Yes, send system info to Canonical
○ No, don't send system info

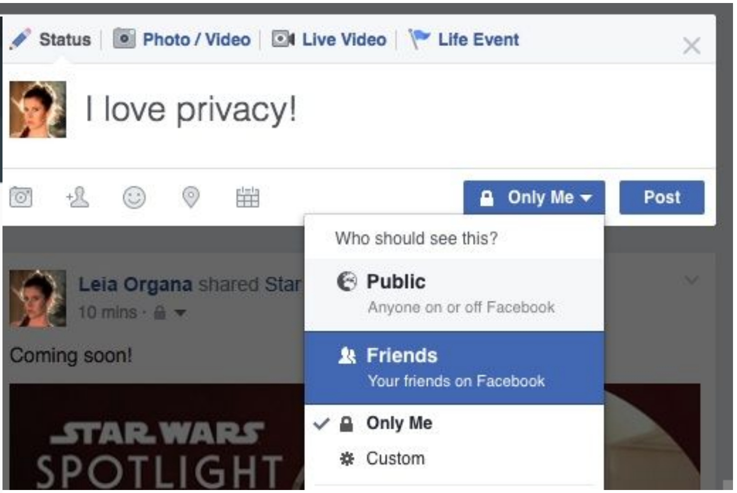


7:26

App permissions

Chrome

- Camera ●
- Contacts ●
- Location ●
- Microphone ●
- Storage ●



Status | Photo / Video | Live Video | Life Event

I love privacy!

Only Me | Post

Who should see this?

- Public — Anyone on or off Facebook
- Friends — Your friends on Facebook
- ✓ Only Me
- Custom

Leia Organa shared Star

10 mins ·

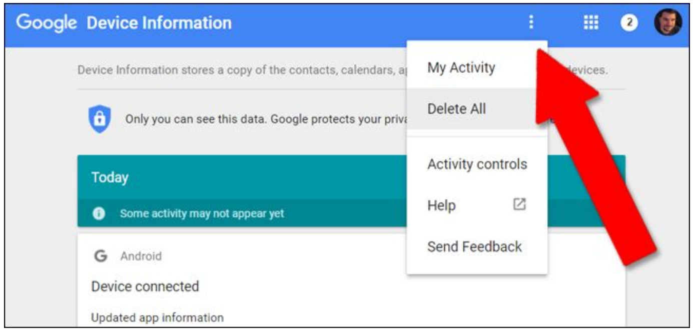
Coming soon!

Modern privacy: Right to selective disclosure

Incl. when used against someone

Privacy is the right of people to "conceal information about themselves that others might use to their disadvantage" - R. Posner, *The economics of justice*, 1983

Incl. after the information has been disclosed, the right to erasure or right to be forgotten



Device Information stores a copy of the contacts, calendars, a evices.

Only you can see this data. Google protects your priv

Some activity may not appear yet

Device connected

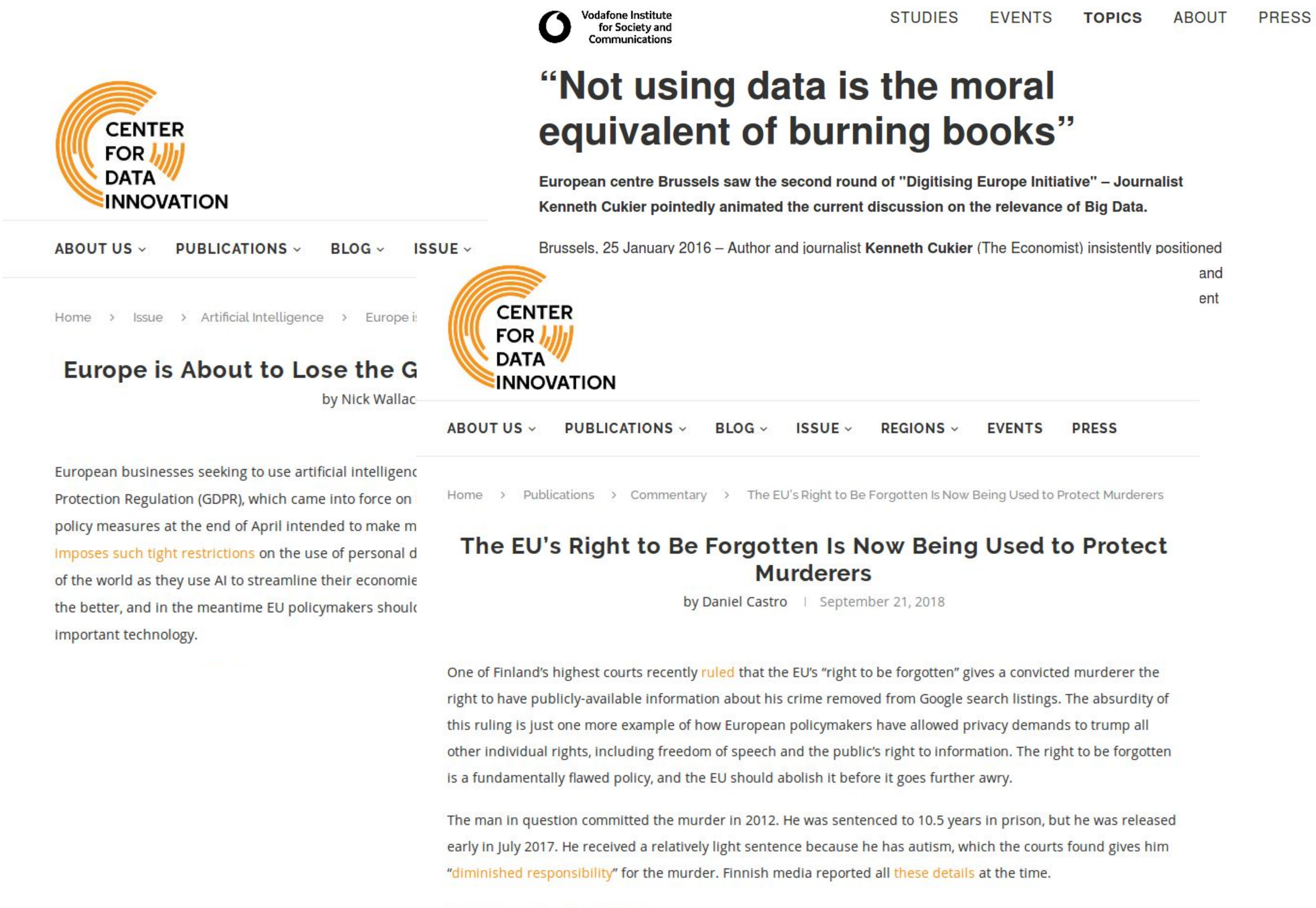
Updated app information

Lots of laws worldwide are protecting privacy, e.g.:

- **USA** - HIPAA (Health data), FERPA (education data)
- **UK** - Data Protection Act (1998). Updated to DPA 2018
- **EU** - General Data Protection Regulation [GDPR] (2016)

Privacy vs
Innovation, Security, etc

What do you think?



CENTER FOR DATA INNOVATION

ABOUT US ⌄ PUBLICATIONS ⌄ BLOG ⌄ ISSUE ⌄

Home › Issue › Artificial Intelligence › Europe i

Europe is About to Lose the G

by Nick Wallac

European businesses seeking to use artificial intelligenc
Protection Regulation (GDPR), which came into force on
policy measures at the end of April intended to make m
imposes such tight restrictions on the use of personal d
of the world as they use AI to streamline their economie
the better, and in the meantime EU policymakers shoul
important technology.

Vodafone Institute for Society and Communications

STUDIES EVENTS **TOPICS** ABOUT PRESS

“Not using data is the moral equivalent of burning books”

European centre Brussels saw the second round of "Digitising Europe Initiative" – Journalist Kenneth Cukier pointedly animated the current discussion on the relevance of Big Data.

Brussels, 25 January 2016 – Author and journalist **Kenneth Cukier** (The Economist) insistently positioned
and
ent

CENTER FOR DATA INNOVATION

ABOUT US ⌄ PUBLICATIONS ⌄ BLOG ⌄ ISSUE ⌄ REGIONS ⌄ EVENTS PRESS

Home › Publications › Commentary › The EU's Right to Be Forgotten Is Now Being Used to Protect Murderers

The EU's Right to Be Forgotten Is Now Being Used to Protect Murderers

by Daniel Castro | September 21, 2018

One of Finland's highest courts recently ruled that the EU's “right to be forgotten” gives a convicted murderer the right to have publicly-available information about his crime removed from Google search listings. The absurdity of this ruling is just one more example of how European policymakers have allowed privacy demands to trump all other individual rights, including freedom of speech and the public's right to information. The right to be forgotten is a fundamentally flawed policy, and the EU should abolish it before it goes further awry.

The man in question committed the murder in 2012. He was sentenced to 10.5 years in prison, but he was released early in July 2017. He received a relatively light sentence because he has autism, which the courts found gives him “diminished responsibility” for the murder. Finnish media reported all these details at the time.

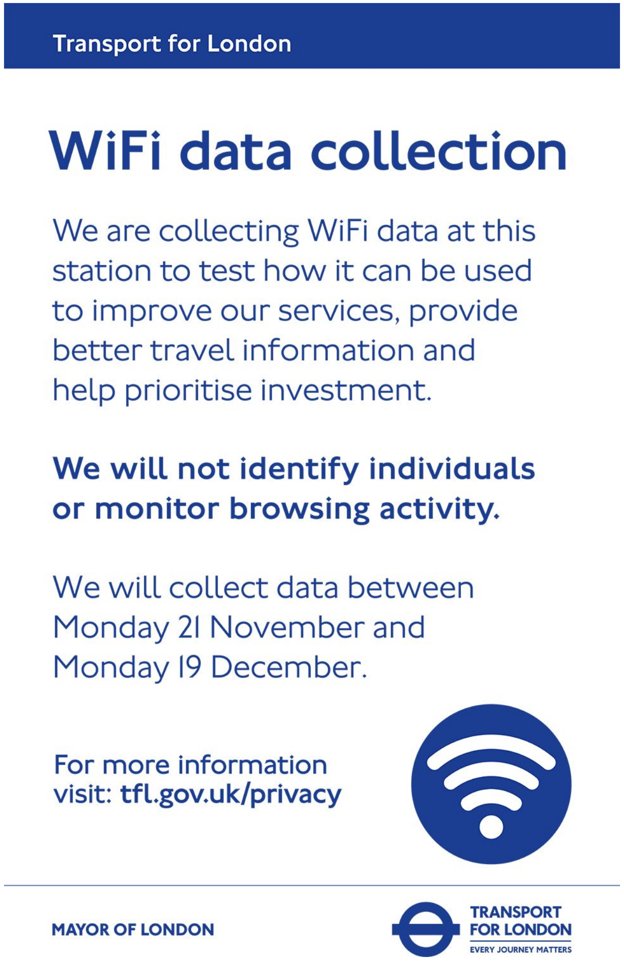
Data collection and opt-out: Transport for London

At the end of 2016, Transport for London started a pilot to collect data about commuters in the London underground using WiFi access points to track commuters’ location.

TfL data was used to better understand crowding and “collective travel patterns” so that “we can improve services and information provision for customers”. But it emerged that they might use it to study in-station footfall and increase marketing revenue.

Commuters had to turn off WiFi to prevent data collection (opt-out)

Locations were collected with MAC address (unique identifier) but pseudonymized (more on this next week)



Transport for London

WiFi data collection

We are collecting WiFi data at this station to test how it can be used to improve our services, provide better travel information and help prioritise investment.

We will not identify individuals or monitor browsing activity.

We will collect data between Monday 21 November and Monday 19 December.

For more information visit: **tfl.gov.uk/privacy**

MAYOR OF LONDON

TRANSPORT FOR LONDON
EVERY JOURNEY MATTERS

Data leak: Ashley Madison

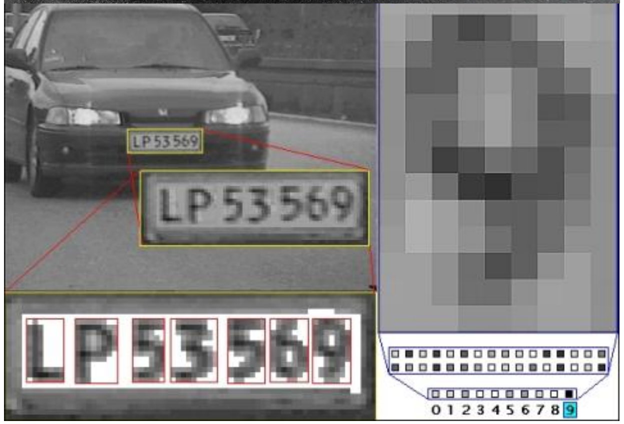
Ashley Madison is a online dating website for people who are married or in relationships. In July 2015, a group of hackers stole and published on Torrent 20+ GB of private data about users of dating website Ashley Madison.

Soon after, one company started offering a "search engine" to find people who had an account on the service. What do you think? Do they deserve it?

- Extortionists began targeting people whose details were included in the leak with several suicides potentially linked to the breach
- 1,200 Saudi Arabian .sa email addresses were in the leaked database. In Saudi Arabia adultery can be punished by death.

Automated license plate readers

Police car now have computer vision systems that automatically capture all license plate numbers that come into view (with location and time)



Is this good or bad?

Information collected can be used by police to, e.g.:

- find out where a plate has been in the past
- determine whether a vehicle was at the scene of a crime
- discover vehicles that may be associated with each other.

But also...

Repossession companies

“A Vigilant spokesperson said Friday that the dataset now includes 4.2 billion sightings, and is growing at a rate of 120 million data points a month.”

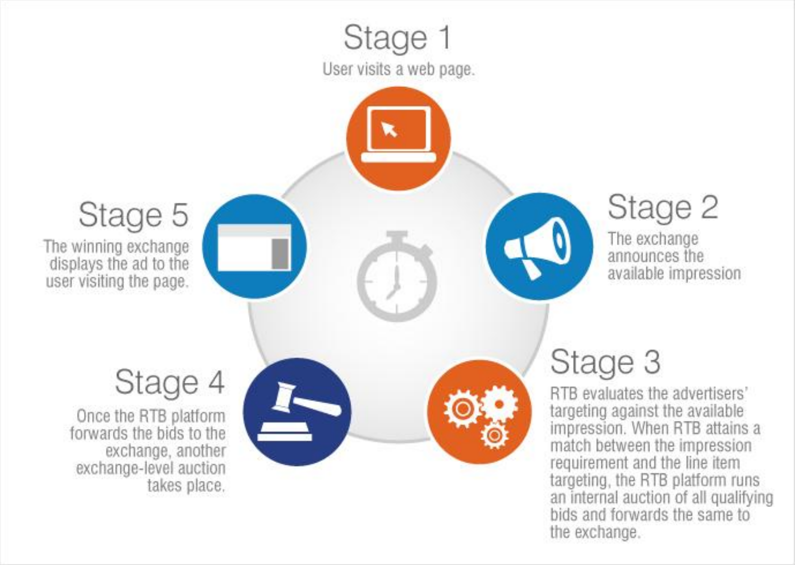
“In a 2014 investigation into automatic plate readers for The Boston Globe, Shawn Musgrave found at least ten repossession companies in Massachusetts that used license-plate readers to do their job.”

“And with 200 to 400-dollar bounties for locating cars that were stolen or are in default, some of those companies focused their search on the most lucrative neighborhoods. Two Massachusetts companies told Musgrave that they expressly targeted low-income housing developments, since it’s likely that a disproportionate number of residents in those areas are behind on auto payments, their cars ripe for repossession.”

---Kaveh Waddell on Apr 22, 2016 for the Atlantic

Real-time bidding (RTB) for online advertising

If you start looking for a new pair of shoes online, you are likely to see a lot of shoe-related ads on every website you visit. Ever wondered why that happens?


Stage 1
User visits a web page.
Stage 2
The exchange announces the available impression
Stage 3
RTB evaluates the advertisers' targeting against the available impression. When RTB attains a match between the impression requirement and the line item targeting, the RTB platform runs an internal auction of all qualifying bids and forwards the same to the exchange.
Stage 4
Once the RTB platform forwards the bids to the exchange, another exchange-level auction takes place.
Stage 5
The winning exchange displays the ad to the user visiting the page.

When you visit a webpage, you are assigned a unique identifier by the ad network that manages that page (we'll call it B). Every time you visit a page managed by B, B records and stores your activity on that page.

B can send part of this profile to advertisers (e.g. shoe shops) that can bid in real time (with automatic systems) how much they are willing to pay to show you an ad, based on your profile.

Good or Bad?

A few arguments on why privacy (still) matters

1. It is a basic human right

"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation."

United Nations Universal Declaration of Human Rights

“Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.”

Article 8 of the EU Charter of Fundamental Rights

2. There is no freedom of thought without privacy

“The right to privacy is often understood as an essential requirement for the realization of the right to freedom of expression. Undue interference with individuals’ privacy can both directly and indirectly limit the free development and exchange of ideas. […] An infringement upon one right can be both the cause and consequence of an infringement upon the other.

[...]

Communications surveillance should be regarded as a highly intrusive act that potentially interferes with the rights to freedom of expression and privacy and threatens the foundations of a democratic society”

Frank La Rue (2013), United Nations Special Rapporteur on Freedom of Expression and Opinion

3. The “nothing-to-hide” argument

"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."
-- Eric Schmidt

The nothing-to-hide argument is flawed on many levels.

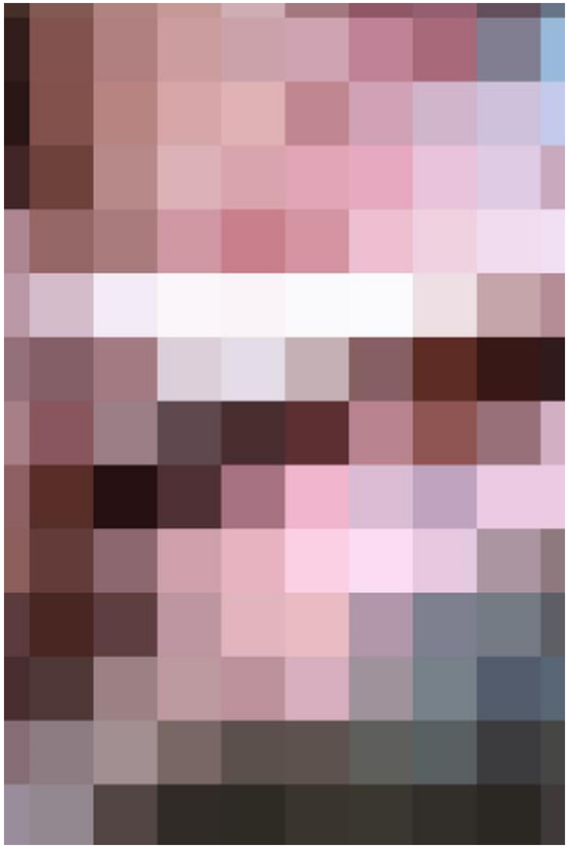
1. Sometimes you want to be let alone even if you’re doing nothing bad. That’s why we have doors.
2. Privacy is essential to protect freedom and dignity against fear of shame.
3. The nothing-to-hide argument implicitly assumes that there are only two kinds of people: good citizens and bad citizens. But are dissidents bad? And whistleblowers? And journalists? **Bad according to who and when?**

See .e.g https://www.schneier.com/essays/archives/2006/05/the_eternal_value_of.html

4. Surveillance as a political weapon

NSA spied on porn habits of Muslim 'radicalizers' in effort to discredit them
By Rich McCormick Nov 27, 2013 - The Verge

The US National Security Agency reportedly spied on the online sexual activity of Muslim "radicalizers" as part of a plot to undermine their reputation and authority. The Huffington Post says that the internet proclivities of six individuals were monitored in order to find "personal vulnerabilities" such as the viewing of pornographic material that could be exploited to "shape the perception of the messenger as well as that of his followers."



© CO408H - Yves-Alexandre de Montjoye

5. Chilling effect and access to information

When people know they are being (or might be) watched, they behave differently. This is this idea behind Jeremy Bentham's panopticon

More recently:

Using lists of possibly troublesome search terms from the Department of Homeland Security's Social Media Monitoring Unit and elsewhere, we identified 282 search terms that were then independently evaluated for whether raters thought they would get you in trouble with the government or with a friend if it became known you had used this search term. [...] We found that U. S.-based search traffic falls by around 5% in the Google Trends index for government-sensitive terms **after the PRISM revelations. This is the first academic empirical evidence of a chilling effect on users' willingness to enter search terms that raters thought would get you into trouble with the U. S. government**.

When we look outside the U. S., at the effects on its top ten trading partners, we find that Google users in **these countries on average searched less not only on government-sensitive search terms such as "anthrax" but also on personally-sensitive terms like "eating disorder."** So we know that there is that much of an effect.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564

Questions?

Now, we’ll be heading to the lab for an exercise session on pandas and matplotlib