Data Security and Privacy in Emerging Scenarios

Pierangela Samarati
Dipartimento di Informatica
Università degli Studi di Milano
pierangela.samarati@unimi.it

8th Int'l Conference on Information Systems Security and Privacy (ICISSP 2022)

February 10, 2022

MARSAL

ICT ecosystem

- Advancements in the ICT have changed our society
- Infrastructures and services are more powerful, efficient, and complex

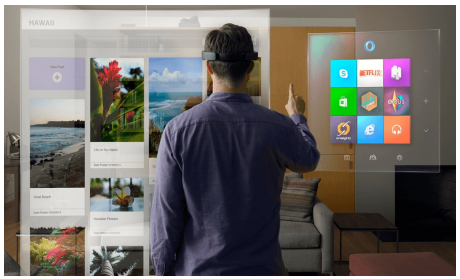
- ICT is the enabling factor for a smart society ...

… Everything is getting smart

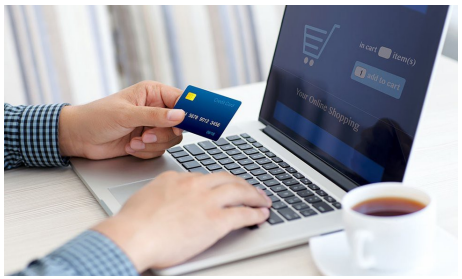
Smart car

Museum and exhibitions

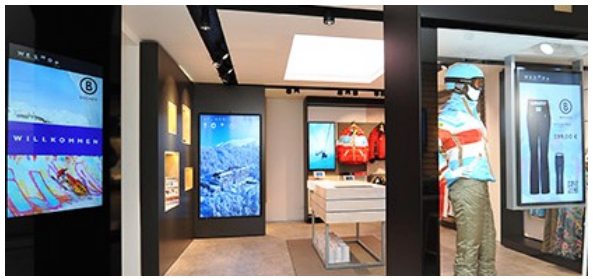
Health Care



Augmented reality



Smart e-commerce



Intelligent shops

Smart entertainment systems

Smart governance

Smart toothbrush

Smart society

Smart society - Advantages

Smart services and security – Advantages

+ Better protection mechanisms
+ Business continuity and disaster recovery
+ Prevention and response

. . . but . . .

Smart services and security – Disadvantages

- More complexity . . .

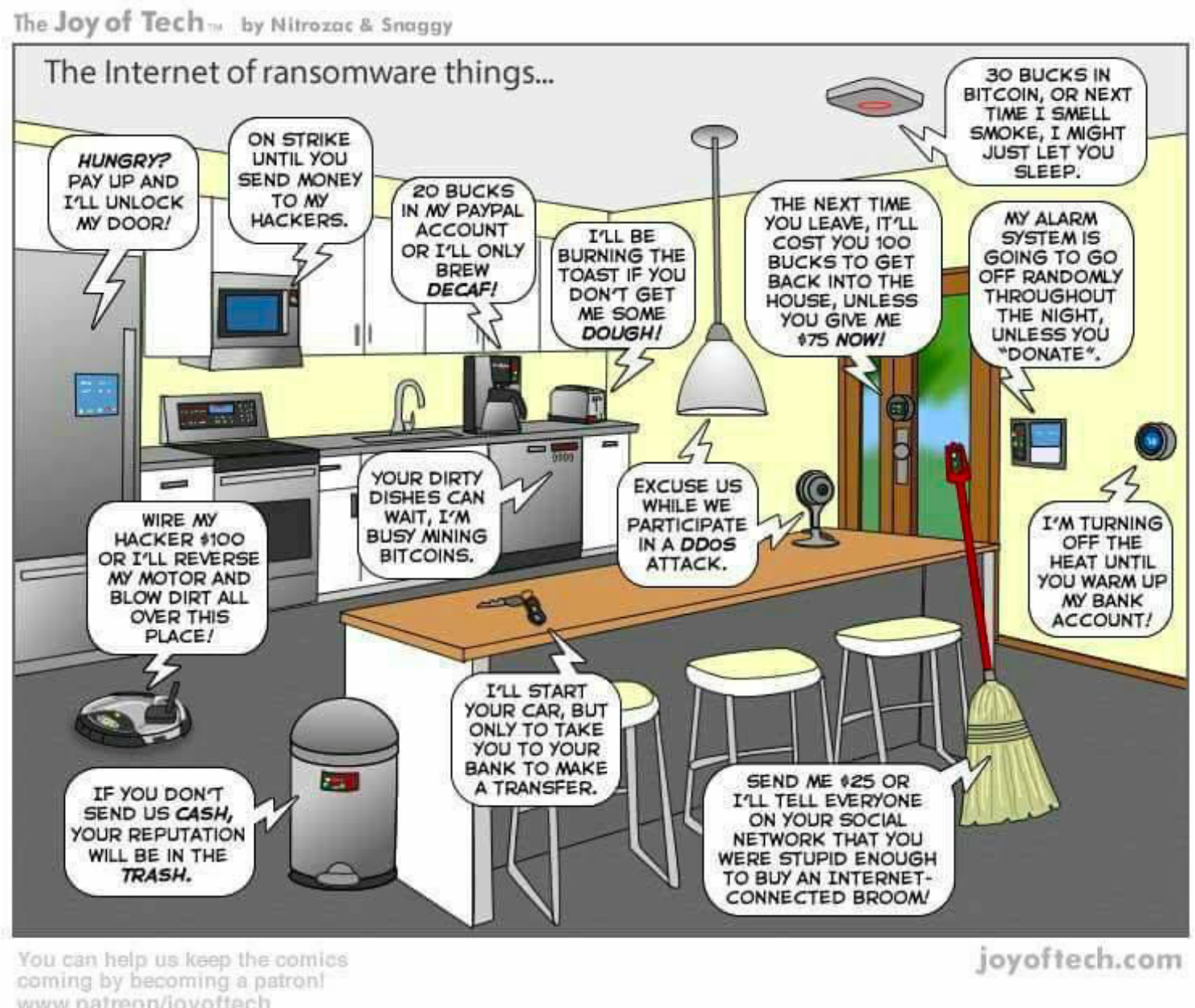
  . . . weakest link becomes a point of attack

  - system hacking
  - improper information leakage
  - data and process tampering

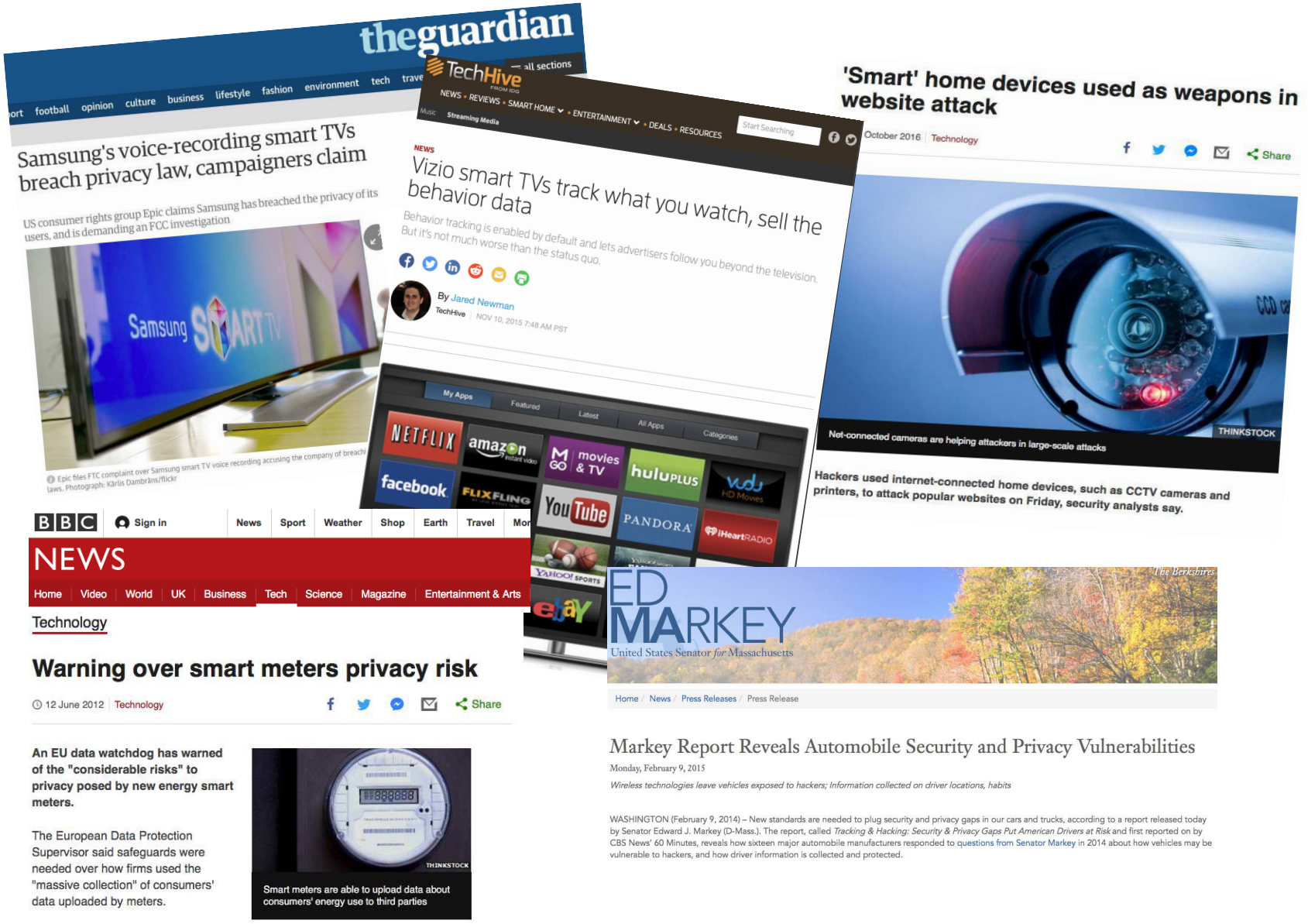
- Explosion of damages and violations

- Loss of control over data and processes

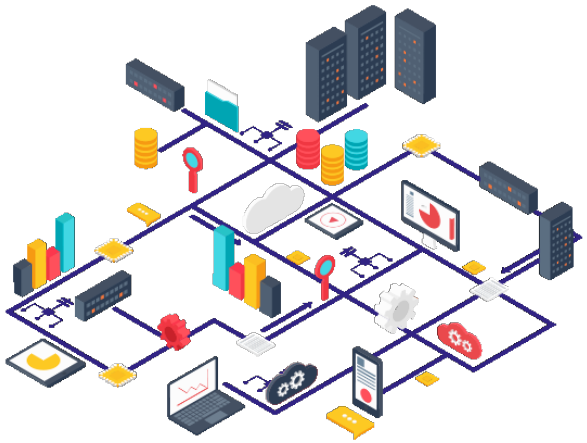
Maybe too smart? – 1



Maybe too smart? – 2



Security … a complex problem



Protection of infrastructure

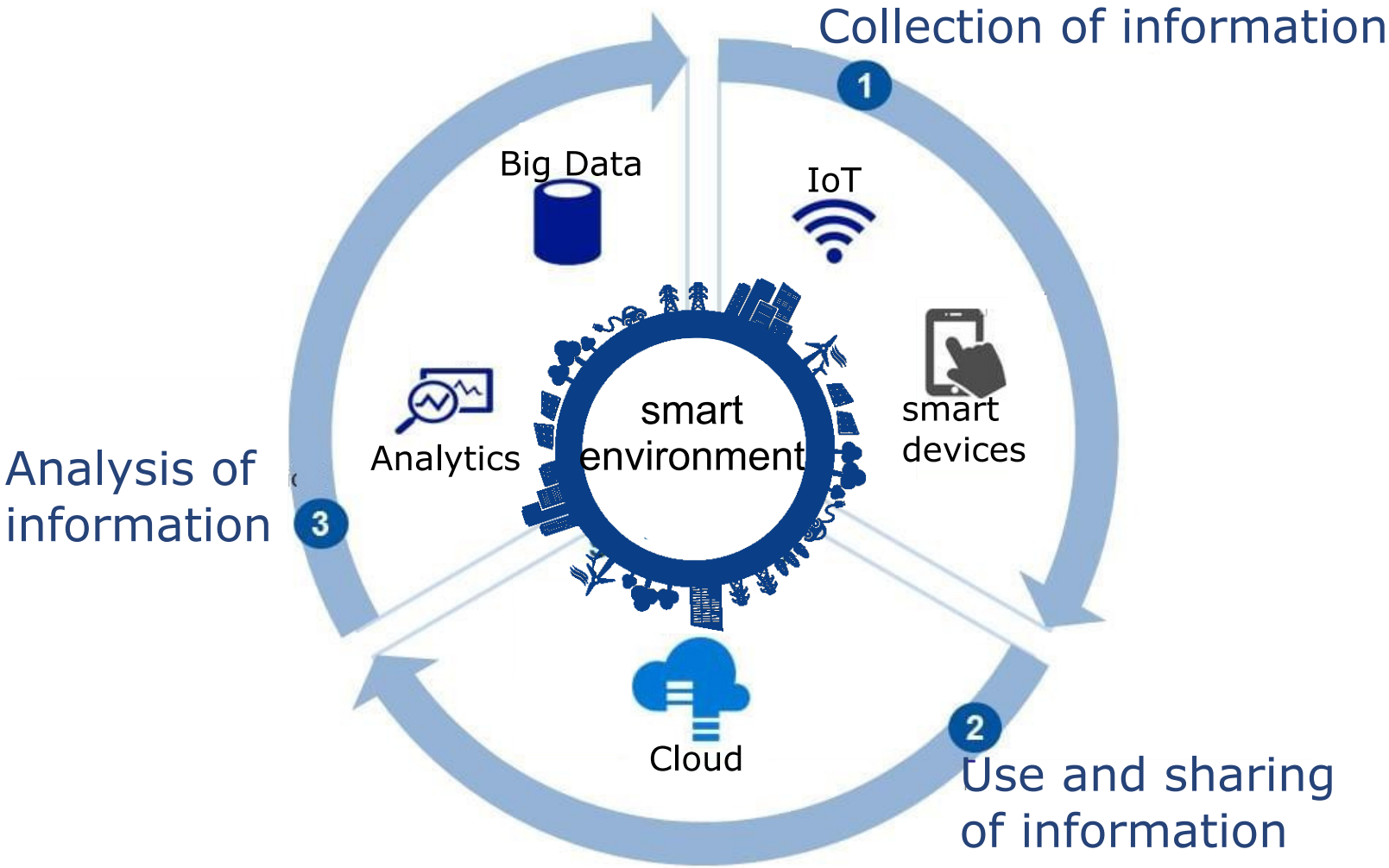
Protection of communication

Protection against malware and attacks

Protection of devices

Protection of data

The role of data in a smart environment

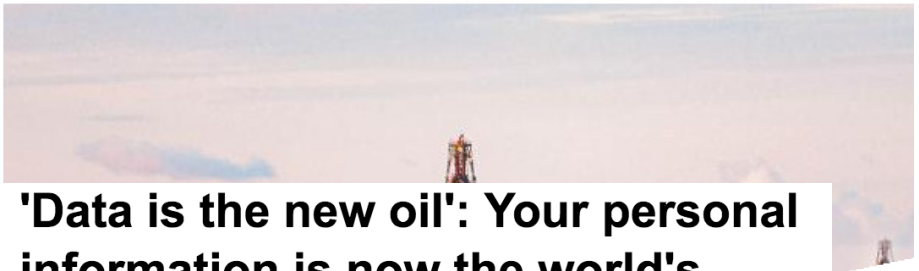


$\Longrightarrow$ better governance and intelligent systems

The most valuable resource - Data

INQUIRER

The new oil: data is the world's most valuable resource



'Data is the new oil': Your personal information is now the world's most valuable commodity

Huge amounts of data are controlled by just 5 global mega-corporations t. bigger than most governments

By Ramona Pringle, CBC News Posted: Aug 25, 2017 5:00 AM ET | Last Updated: Aug 25, 2017 11:28 AM

Fuel of the future

Data is giving rise to a new economy

How is it shaping up?

Why is data protection so important?

BLOG 06 February 2017

digitally needs to be properly protected. From financial
t information for your staff, data usage in the UK is protected by

a legal necessity, but crucial to protecting and maintaining your

Big Data and Analytics Play an Important Role in the Energy Industry

Real-TimeDAILY

AROUND THE NET

Data Is Now The World's Most Valuable Resource

The Economist, Monday, May 8, 2017 6:22 AM

Data is now the world's most valuable resource according to *The Economist,* which reports on antitrust concerns about Alphabet (Google's parent company), Amazon, Apple, Facebook, and Microsoft, all of which have tons of data. The

PARTNER CONTENT ARVIND SINGH

bus

IS BIG DATA THE NEW BLACK GOLD?

Impact on data protection and privacy

Uber reveals 2.7 million UK users of its app were affected by a mass data breach that saw names, emails and phone numbers stolen

- Uber has revealed 2.7m UK users of its app were affected by a 2016 data brea
- The taxi-hailing firm then tried to cover up the breach for more than a year
- It was also found Uber had paid two hackers £75,000 to delete the data

By TH
PUBLI

Computer Scientists Develop a Simple Tool to Tell If Websites Suffered a Data Breach

Published: December 12, 2017.

Uber says data breach compromised 380K users in Singapore

Ride-sharing company reveals 380,000 in Singapore were affected by the massive data breach that compromised 57 million accounts globally, but says no fraud or misuse has been tied to these users.

By Eileen Yu for By The Way | December 16, 2017 -- 10:07 GMT (10:07 GMT) | Topic: Security

The Dutch Data Protection Authority accidentally leaked its employees' data

by MIX — 4 weeks

Approx. 9,000 Penn students affected by security breach that released their private information

SECURITY

By Kelly Hainzerling | 03/12/18 6:50pm

1 DEC 2017 NEWS

Over 100GB of Secret Consumer Credit Data Leaked Online

A collection of 1.4 Billion Plain-Text leaked credentials is available online

December 12, 2017 By Pierluigi Paganini

NEWS

63,500 records breached by misconfigured database

by Jessica Davis | April 12, 2018

MASSIVE

Personal Data of Over 143 Million Americans Stolen from a Credit Reporting Firm

By R
Sr

18 SHARES

SHARE TWEET SUBMIT

MyFitnessPal breach affects millions of Under Armour users

19 DEC 2017 NEWS

Californian Voters Suffer Major Data Breach

My Page Like 2

A 41-gigabyte archive containing **1.4 Billion** credentials in clear text was found in dark web, it had been updated at the end of November

LOCAL

Photo: Jimmy Bernhard, KSDK-TV

Former nursing home employee admits stealing residents' credit card numbers

Shaniece Borney, 29, will be forced to pay the victims back and could face an additional $250,000 fine, 10 years in prison or both.

Mar 01 2018

Equifax discovers another 2.4 million customers hit by data breach

Posted by Dissent at 11:02 am | Business Sector, Hack, U.S.

NEWS

Facebook admits to far higher number of data breaches

Facebook has said personal data on 87 million users was shared with Cambridge Analytica, millions more than it admitted earlier. The social media giant also unveiled new privacy rules, but the whiff of scandal lingers.

Deloitte hit by cyber-attack revealing clients' secret emails

Exclusive: hackers may have accessed usernames, passwords and personal details of top accountancy firm's blue-chip clients

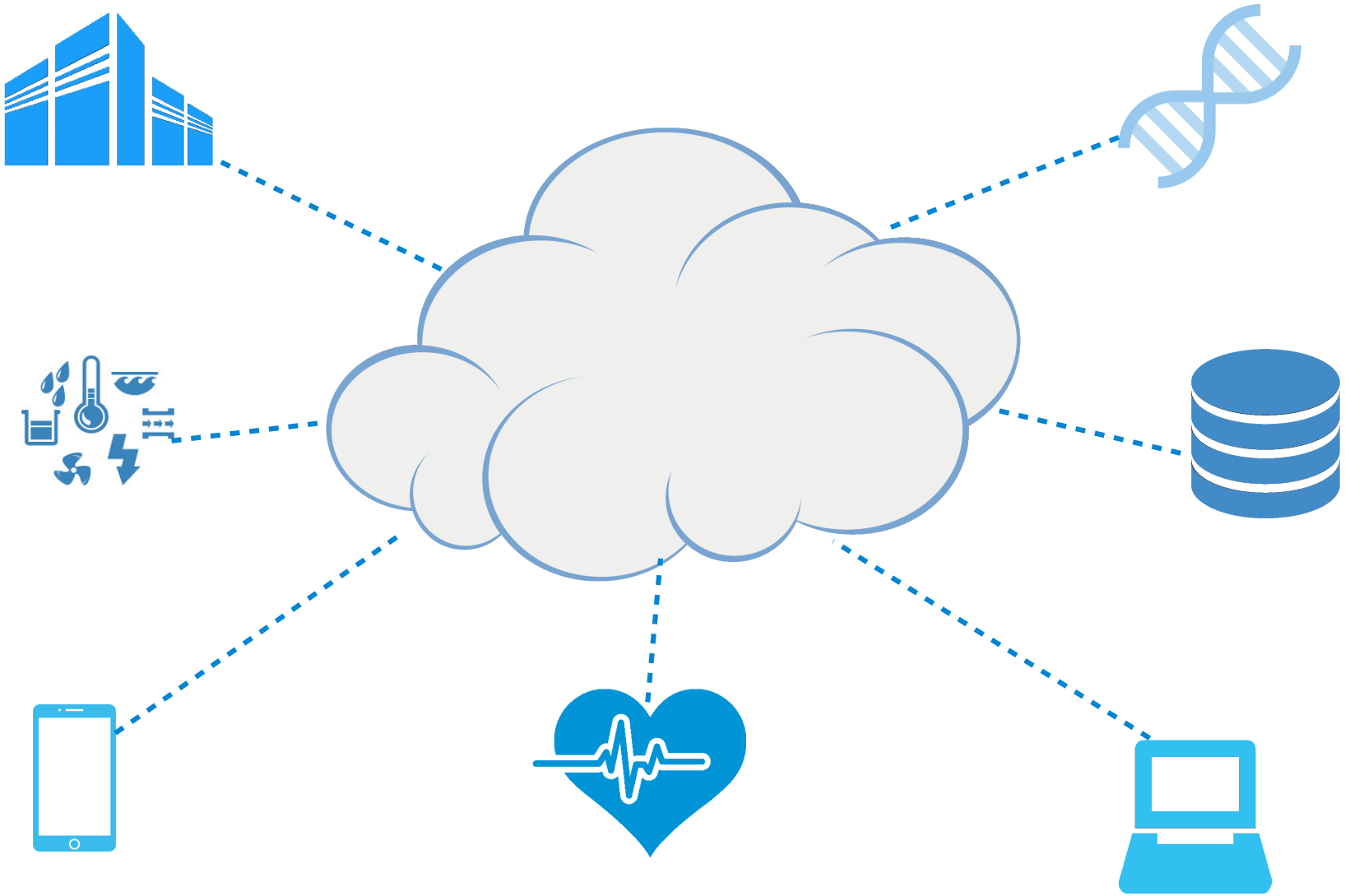
Privacy

Carphone Warehouse Breach: 'Striking' Failures Trigger Fine

Mathew J. Schwartz • January 10, 2018

Mobile phone retailer Carphone Warehouse has been hit with one of the largest fines ever imposed by Britain's data privacy watchdog

Huge amount of data stored at external providers

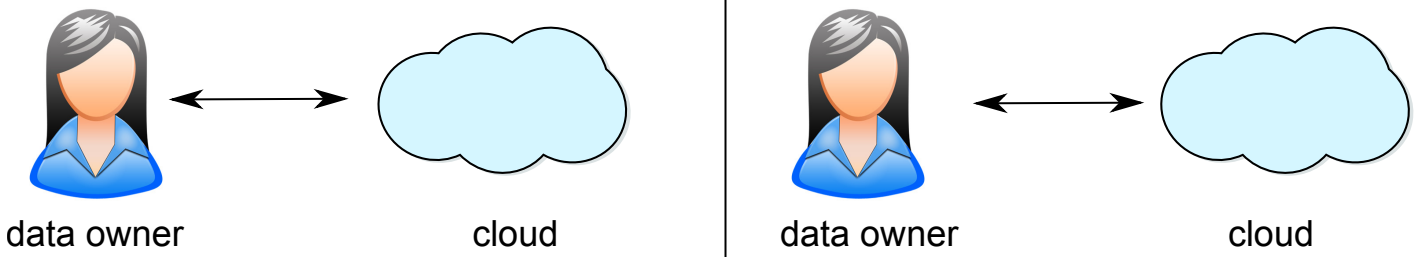


Cloud computing

- The Cloud allows users and organizations to rely on external providers for storing, processing, and accessing their data
  - + high configurability and economy of scale
  - + data and services are always available
  - + scalable infrastructure for applications
- Users lose control over their own data
  - – new security and privacy problems
- Need solutions to protect data and to securely process them in the cloud

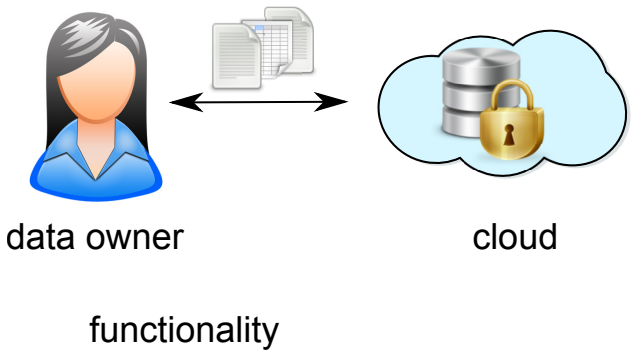
Cloud computing: Today

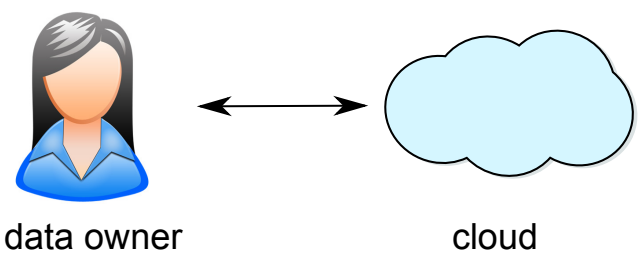
Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



Cloud computing: Today

Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders

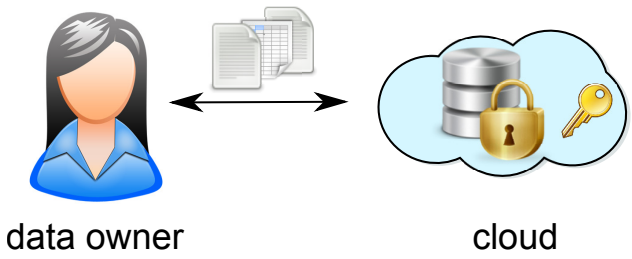




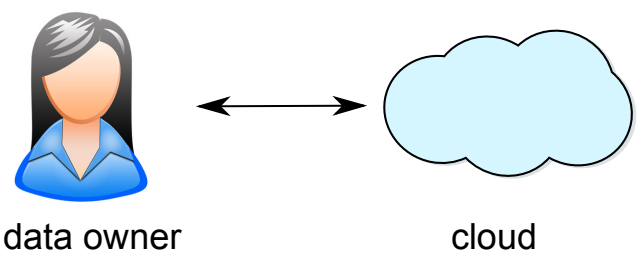
- functionality

Cloud computing: Today

Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



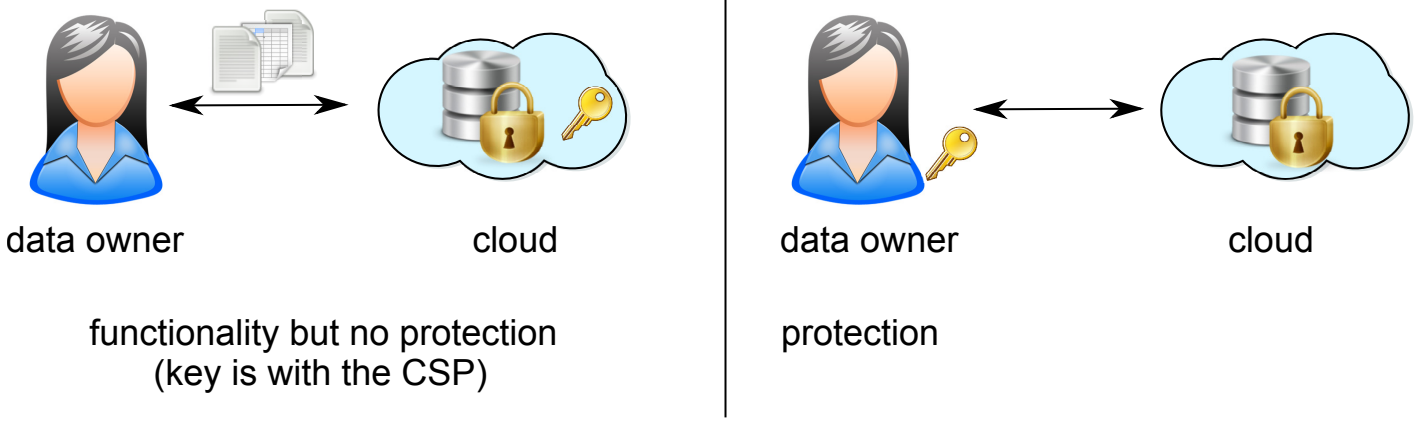
functionality but no protection
(key is with the CSP)



- functionality implies full trust in the CSP that has full access to the data (e.g., Google Cloud Storage, iCloud)

Cloud computing: Today

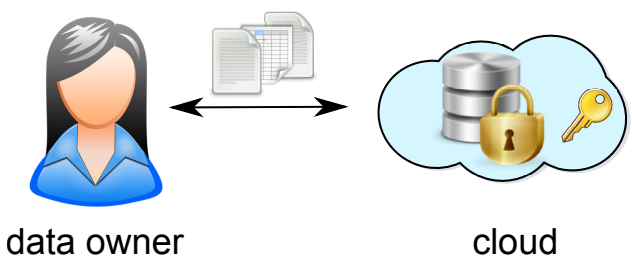
Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



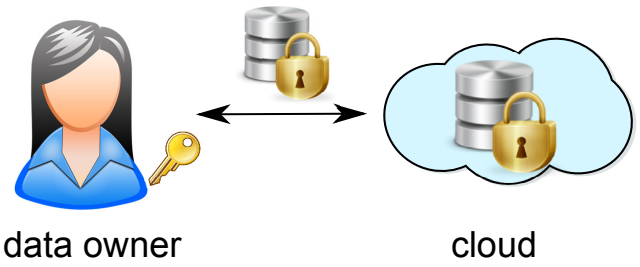
- functionality implies full trust in the CSP that has full access to the data (e.g., Google Cloud Storage, iCloud)
- protection

Cloud computing: Today

Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



functionality but no protection
(key is with the CSP)

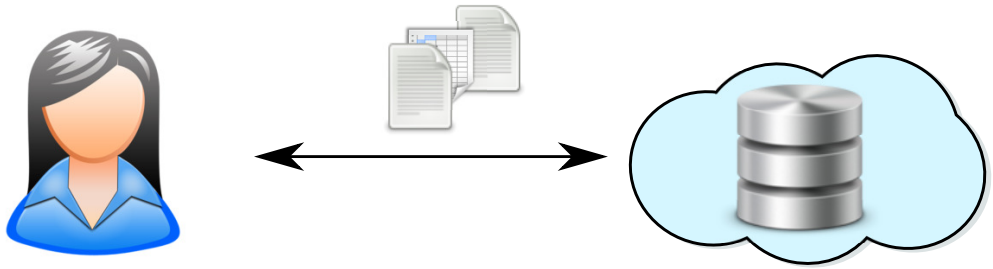


protection but limited functionality
(you cannot access data as you like)

- functionality implies full trust in the CSP that has full access to the data (e.g., Google Cloud Storage, iCloud)
- protection but limited functionality since the CSP cannot access data (e.g., Boxcryptor, SpiderOak)

Cloud computing: New vision

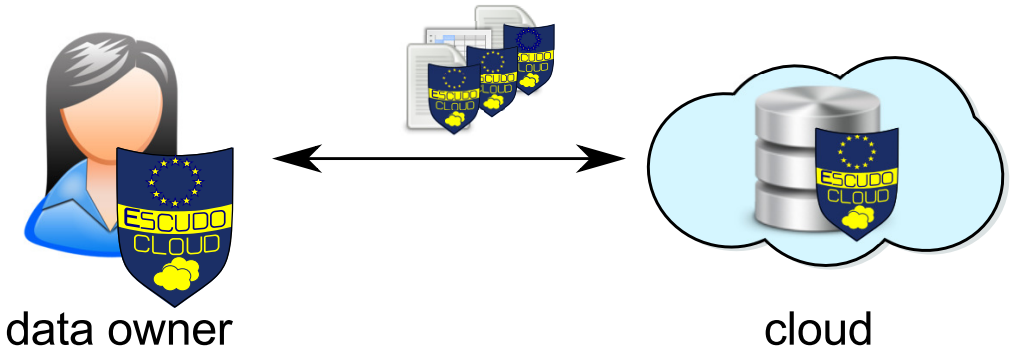
Solutions that provide protection guarantees giving the data owners both: full control over their data and cloud functionality over them



https://mosaicrown.eu https://www.marsalproject.eu

Cloud computing: New vision

Solutions that provide protection guarantees giving the data owners both: full control over their data and cloud functionality over them

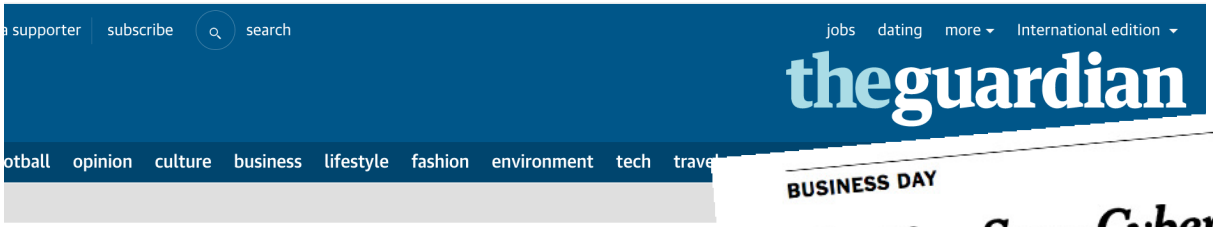


- client-side trust boundary: only the behavior of the client should be considered trusted

  $\Longrightarrow$ techniques and implementations supporting direct processing of encrypted data in the cloud

https://mosaicrown.eu https://www.marsalproject.eu

Data protection – Base level



Yahoo hack: 1bn accounts compromised by biggest data breach in history

The latest incident to emerge - which happened in 2013 - is probably distinct from the breach of 500m user accounts in 2014

Technology

Hackers steal 2.5 million PlayStation and Xbox players' details in major breach

theguardian

Deloitte hit by cyber-attack revealing clients' secret emails

Exclusive: hackers may have accessed usernames, passwo
details of top accountancy firm's blue-chip cli

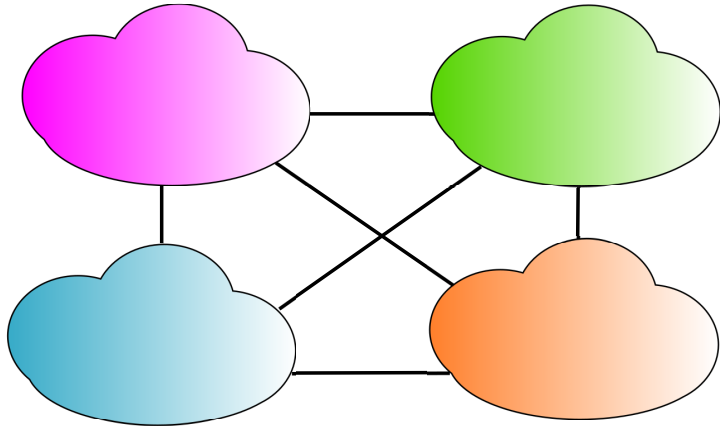
Security

Two million customer records pillaged in IT souk CeX hack attack

Data protection – Base level

Data protection – Regulation

Access and usage control



Selective sharing

General Data Protection Regulation

HIPAA Health Insurance Portability & Accountability Act

Governance and regulation

Data protection – Confidentiality (1)

- Minimize release/exposition
  - correlation among different data sources
  - indirect exposure of sensitive information
  - de-identification $\neq$ anonymization

Data protection – Confidentiality (2)

THREAT LEVEL | privacy

Netflix Spilled Your *Brokeback Mountain* Secret, Lawsuit Claims

BY RYAN SINGEL 12.17.09 4:29 PM

Follow @rsingel

Share 174 Tweet 18 +1 0 Share 5

The Telegraph

Home News World Sport Finance Comment Blogs Culture Travel Life Women

Technology News | Technology Companies | Technology Reviews | Video Games | Technology

HOME » TECHNOLOGY » FACEBOOK

Gay men 'can be identified by their Facebook friends'

Homosexual men can be identified just by looking at their Facebook friends, a
to unpublished research by two students at the Massachusetts Institute of Tec

Print this

Share 576

Facebook 560

Twitter 16

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Aut

News & Comment > News > 2017 > October > Article

NATURE | NEWS

Privacy loophole found in genetic databases

DNA donors' identities can be determined from publicly available records.

Erika Check Hayden

17 January 2013

Rights & Permissions

loophole could allow anyone

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake

SIGN IN TO E-MAIL THIS

PRINT

REPRINTS

THE WAY WAYBACK WATCH TRAILER

Re-identification with any information

- Any information can be used to re-identify anonymous data

  $\Longrightarrow$ ensuring proper privacy protection is a difficult task since the amount and variety of data collected about individuals is increased

- Two examples:
  - AOL
  - Netflix

AOL data release – 1

In 2006, to embrace the vision of an open research community, America OnLine publicly posted queries to AOL's search engine

- 20 million search queries for 658,000 users summarizing 3 months of activity
- obviously identifying information (AOL username, IP address) was removed
- usernames replaced with unique identification numbers

AOL data release – 2

User 4417749:

- numb fingers
- 60 single men
- dog that urinates on everything
- hand tremors
- nicotine effects on the body
- dry mouth
- bipolar
- several people with last name Arnold
- landscapers in Lilburn, Ga
- homes sold in shadow lake subdivision Gwinnett county, Georgia

AOL data release – 2

User 4417749:

- numb fingers
- 60 single men
- dog that urinates on everything
- hand tremors
- nicotine effects on the body
- dry mouth
- bipolar
- several people with last name Arnold
- landscapers in Lilburn, Ga
- homes sold in shadow lake subdivision Gwinnett county, Georgia

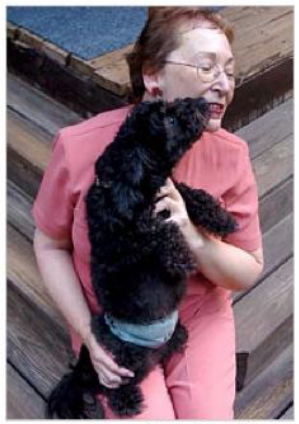
Thelma Arnold, a 62-year-old widow living in Lilburn, Ga

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E-MAIL THIS
PRINT
REPRINTS

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.



Erik S. Lesser for The New York Times

Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.

No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

AOL data release – 3

What about user 17556639?

- how to kill your wife
- how to kill your wife
- wife killer
- how to kill a wife
- poop
- dead people
- pictures of dead people
- killed people
- dead pictures
- dead pictures
- dead pictures
- murder photo
- steak and cheese
- photo of death
- photo of death
- death
- dead people photos
- photo of dead people
- www.murderdpeople.com
- decapatated photos
- decapatated photos
- car crashes3
- car crashes3
- car crash photo

Netflix prize data release – 1

In 2006: “Netflix Prize” of USD 1 million for a movie recommendation algorithm that improved Netflix’s algorithm by 10%

- 100 million records (movie rated, rating, date) for 500,000 users from Oct.’98 to Dec.’05
- only a sample (one tenth) of the database was released
- some ratings were perturbed (but not much, not to alter statistics)
- identifying information (usernames) removed, but a unique user identifier was assigned to preserve rating-to-rating continuity

Netflix prize data release – 2

Netflix Prize dataset + IMDb:

- with 6 movie ratings and dates ($\pm$ 2 weeks), 99% of records uniquely identified
- with 2 movie ratings and dates ($\pm$ 3 days), 68% of records uniquely identified
- 84% of subscribers in the dataset uniquely identified by knowing 6 obscure (outside the top 500) movies

Netflix prize data release – 2

Netflix Prize dataset + IMDb:

- with 6 movie ratings and dates ($\pm$ 2 weeks), 99% of records uniquely identified
- with 2 movie ratings and dates ($\pm$ 3 days), 68% of records uniquely identified
- 84% of subscribers in the dataset uniquely identified by knowing 6 obscure (outside the top 500) movies

THREAT LEVEL | privacy

Netflix Spilled Your *Brokeback Mountain* Secret, Lawsuit Claims

BY RYAN SINGEL 12.17.09 4:29 PM

Follow @rsingel

Share 174 Tweet 18 +1 0 Share 5

An in-the-closet lesbian mother is suing Netflix for privacy invasion, alleging the movie rental company made it possible for her to be outed when it disclosed insufficiently anonymous information about nearly half-a-million customers as part of its $1 million contest to improve its recommendation system.

The suit known as Doe v. Netflix (.pdf) was filed in federal court in California on Thursday, alleging that Netflix violated fair-trade laws and a federal privacy law protecting video rental records, when it launched its popular contest in September 2006.

The suit seeks more than $2,500 in damages for each of more than 2 million Netflix customers.

Target data mining

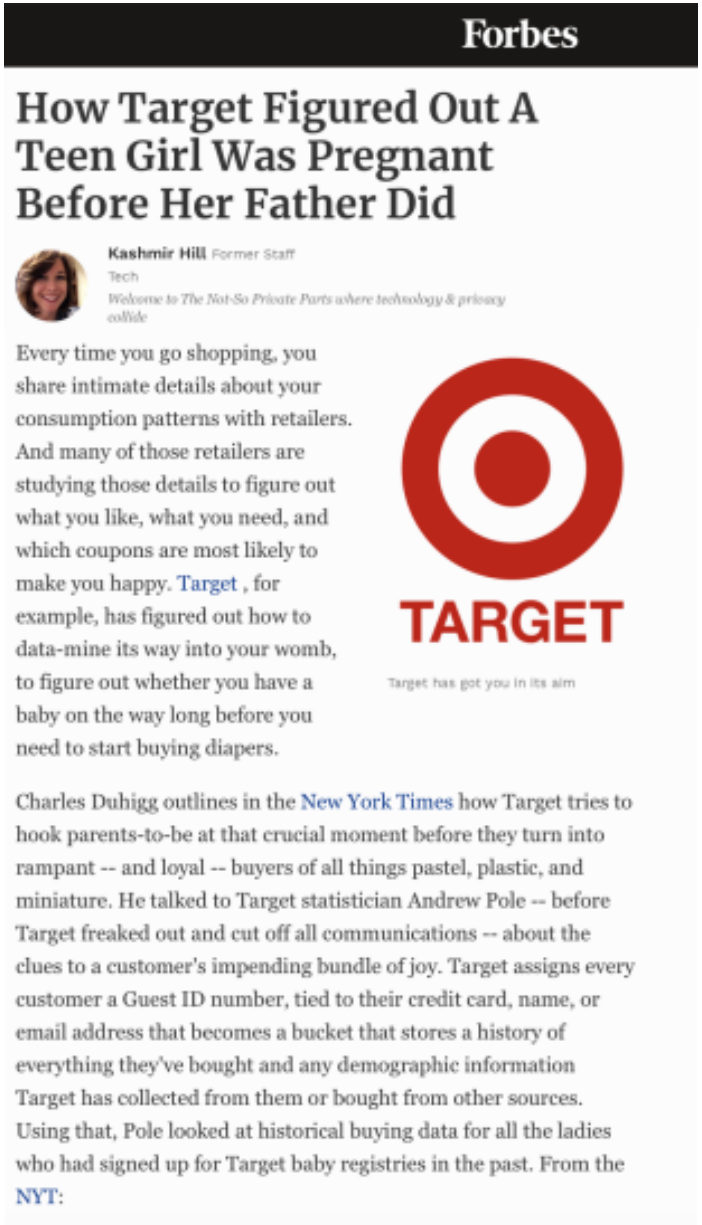
In 2012, Target found to mine customers' data for targeted advertising

- Every customer assigned a Guest ID number:
  - tied to credit card, name, email address, . . .
  - stores history of bought goods and other (bought) information
- Purchase history enables mining to
  - infer major life events
  - predict shopping habits
  - target on expected interest

Target data mining

In 2012, Target found to mine customers' data for targeted advertising

- Every customer assigned a Guest ID number:
  - tied to credit card, name, email address, . . .
  - stores history of bought goods and other (bought) information
- Purchase history enables mining to
  - infer major life events
  - predict shopping habits
  - target on expected interest



Forbes

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

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

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

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. Target assigns every customer a Guest ID number, tied to their credit card, name, or email address that becomes a bucket that stores a history of everything they've bought and any demographic information Target has collected from them or bought from other sources. Using that, Pole looked at historical buying data for all the ladies who had signed up for Target baby registries in the past. From the NYT:

Cambridge Analytica scandal – 1

www.observer.co.uk | Sunday 18 March 2018 | £3.00

The Observer

Revealed: 50m Facebook files taken in record data breach

Exclusive

- Whistleblower tells of bid to influence votes
- Tech giant suspends controversial data firm

Carole Cadwalladr & Emma Graham-Harrison

The data analytics firm that worked with Donald Trump's election team and the winning Brexit campaign harvested millions of Facebook profiles of US voters, in one of the tech giant's biggest ever data breaches, and used them to build a powerful software program to predict and influence choices at the ballot box.

A whistleblower has revealed to the *Observer* how Cambridge Analytica – owned by the hedge fund billionaire Robert Mercer, and headed at the time by Trump's key adviser Steve Bannon – used personal information taken without authorisation in early 2014 to build a system that could profile individual US voters, in order to target them with personalised political advertisements.

Christopher Wylie, who worked with a Cambridge University academic to obtain the data, told the *Observer*: "We exploited Facebook to harvest millions of people's profiles. And built models to exploit what we knew about them and target their inner demons. That was the basis the entire company was built on."

Documents seen by the *Observer*, and confirmed by a Facebook statement, show that by late 2015 the company had found out that information had been harvested on an unprecedented scale. However, at the time it failed to alert users and took only limited steps to recover and secure the private information of more than 50 million individuals.

The *New York Times* is reporting that copies of the data harvested for Cambridge Analytica could still be found online; its reporting team had viewed some of the raw data.

The data was collected through an app called thisisyourdigitallife, built by academic Aleksandr Kogan, separately from his work at Cambridge University. Through his company Global Science Research (GSR), in collaboration with Cambridge Analytica, hundreds of thousands of

Continued on page 7

The Cambr Analyti Files

Full int Whistle Christo Wylie li the lid

Cover sto New Revi

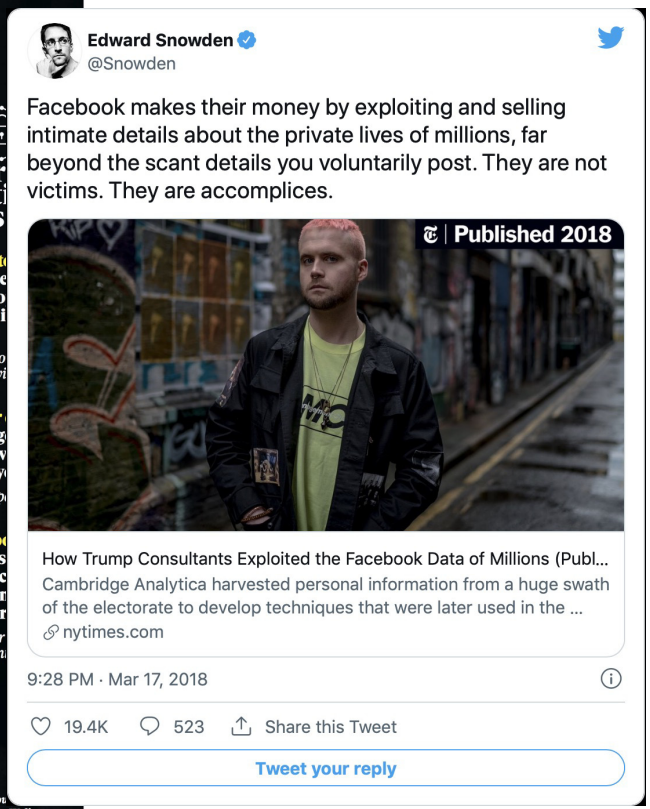
Like or The alg that rev about y

Report, p

Faceboo How its destruc ethos in democr

Observer Commen

Whistleblow Christopher Wylie. Photograph by Antonio Olmos for the Observer



Edward Snowden
@Snowden

Facebook makes their money by exploiting and selling intimate details about the private lives of millions, far beyond the scant details you voluntarily post. They are not victims. They are accomplices.

How Trump Consultants Exploited the Facebook Data of Millions (Publ...
Cambridge Analytica harvested personal information from a huge swath of the electorate to develop techniques that were later used in the ...
nytimes.com

9:28 PM · Mar 17, 2018

19.4K 523 Share this Tweet

Tweet your reply

Cambridge Analytica scandal – 2

- Personality quiz app
  - installed by 330,000 Facebook users who gave permission for accessing their data …
  - … but the app was also collecting data of those users' friends
- Data from 87 million Facebook users retrieved by the app
  - data shared with Cambridge Analytica
  - users profiled through their data

User profiling - Facebook/Cambridge Analytica

OCEAN model

- **O**penness
- **C**onscientiousness
- **E**xtraversion
- **A**greeableness
- **N**euroticism

User profiling - Facebook/Cambridge Analytica

OCEAN model

- **O**penness
  do you enjoy new experiences?
- **C**onscientiousness
  do you prefer plans and order?
- **E**xtraversion
  how social are you?
- **A**greeableness
  do you value others' needs
  and society?
- **N**euroticism
  how much do you tend to worry?

User profiling - Facebook/Cambridge Analytica

OCEAN model

- **O**penness
  do you enjoy new experiences?
- **C**onscientiousness
  do you prefer plans and order?
- **E**xtraversion
  how social are you?
- **A**greeableness
  do you value others' needs and society?
- **N**euroticism
  how much do you tend to worry?

Message to push support for Second Amendment of US Constitution

---

Conscientious individual with high neuroticism:

"The second amendment isn't just a right. It's an insurance policy. Defend the righ to bear arms!"

User profiling - Facebook/Cambridge Analytica

OCEAN model

- **O**penness
  do you enjoy new experiences?
- **C**onscientiousness
  do you prefer plans and order?
- **E**xtraversion
  how social are you?
- **A**greeableness
  do you value others' needs and society?
- **N**euroticism
  how much do you tend to worry?

Message to push support for Second Amendment of US Constitution

---

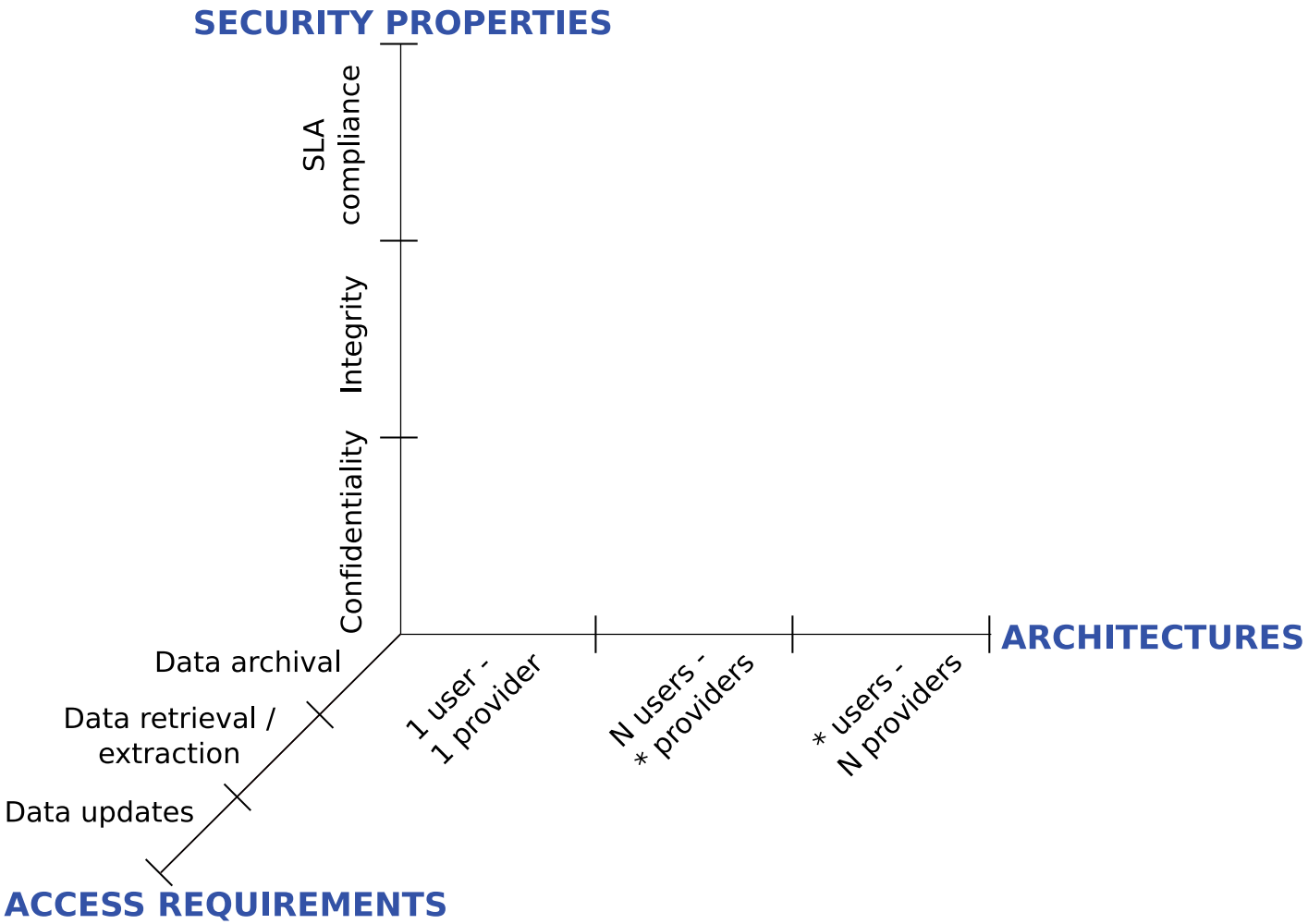
Close and agreeable individual:

"From father to son,
since the birth of our Nation.
Defend the second amendment."

Characterization of Data Protection Challenges in Cloud Scenarios

Scientific and technical challenges

Three dimensions characterize the problems and challenges



Security properties

Confidentiality

- data externally stored
- users identities
- actions that users perform on the data

Integrity

- data externally stored
- computation and query results

SLA compliance

- assurance and certification

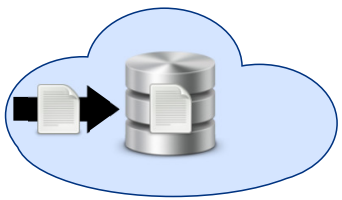
Access requirements

Data archival

- upload/download
- protection of data in storage

Data retrieval/extraction

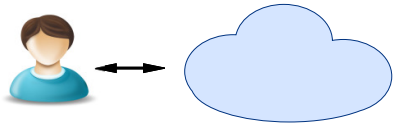
- support for fine-grained data retrieval and queries
- protection of computations and query results



Data update

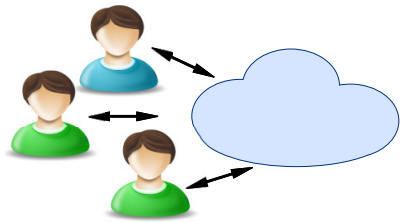
- support for access retrieval and enforcement of updates
- protection of the actions and of their effects on the data

Architectures



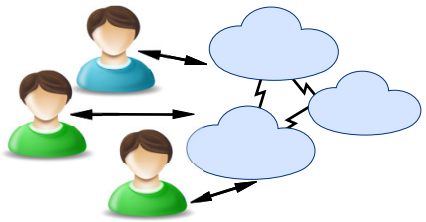
1 user - 1 provider

- protection of data at rest
- fine-grained retrieval
- query privacy/integrity



n users - * providers

- authorizations and access control
- multiple writers



*** users - n providers**

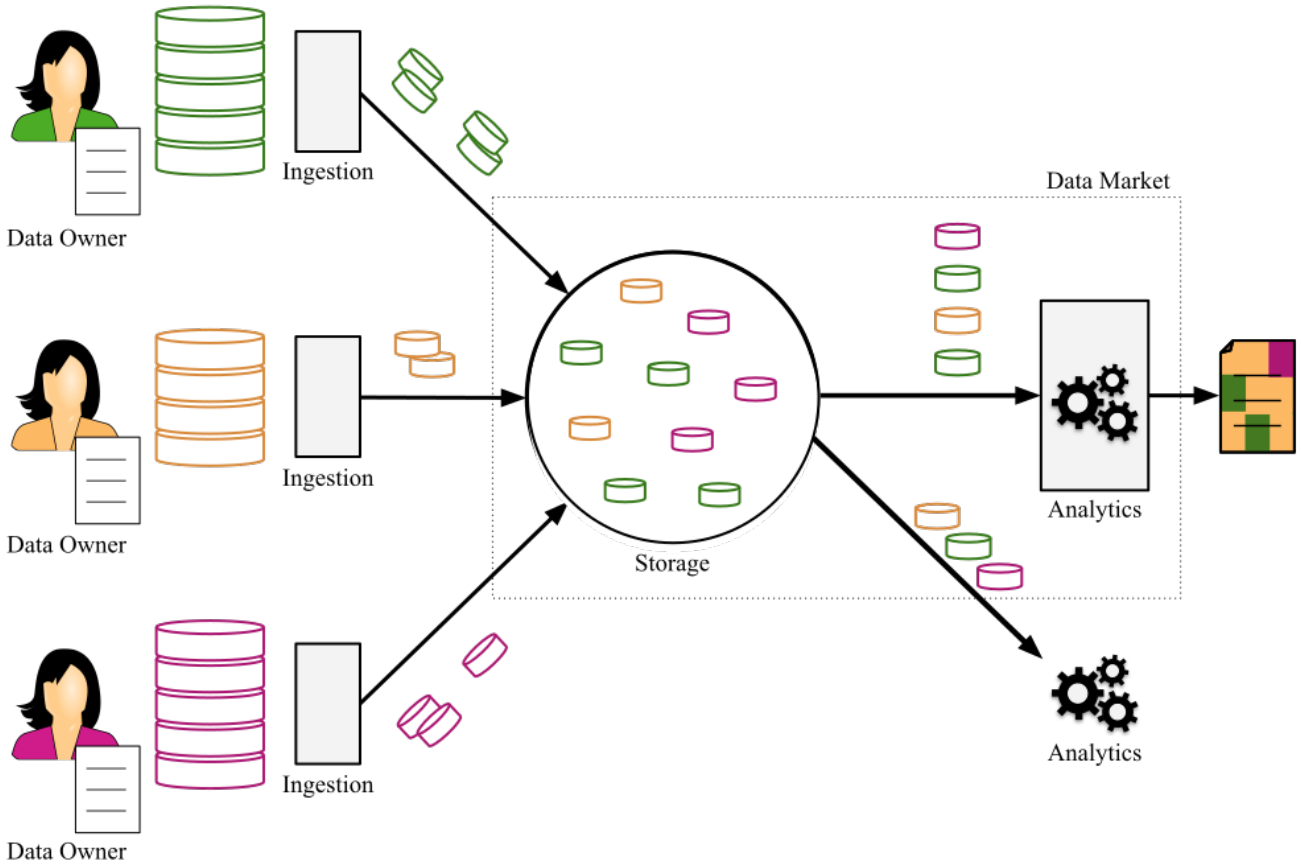
- controlled data sharing and computation

Combinations of the dimensions

- Every combination of the different instances of the dimensions identifies new problems and challenges
- The security properties to be guaranteed can depend on the access requirements and on the trust assumption on the providers involved in storage and/or processing of data
- Providers can be:
  - curious
  - lazy
  - malicious

Digital Data Market

Digital Data Market



Dimensions of the problem and challenges

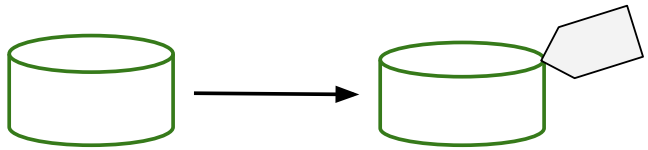
- Requirements capturing and representation

  policies regulating access, sharing, usage and processing

Dimensions of the problem and challenges

- Requirements capturing and representation

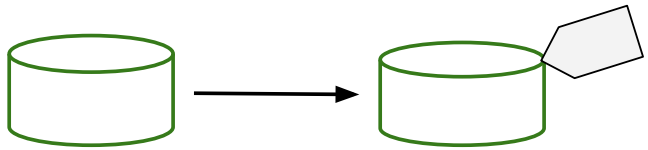
  policies regulating access, sharing, usage and processing



Dimensions of the problem and challenges

- Requirements capturing and representation

  policies regulating access, sharing, usage and processing



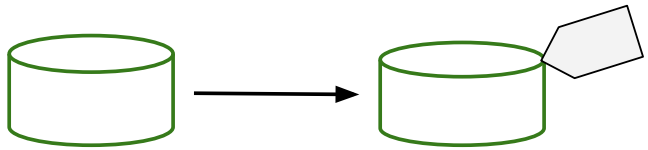
- Enforcing technologies

  data wrapping / sanitization

Dimensions of the problem and challenges

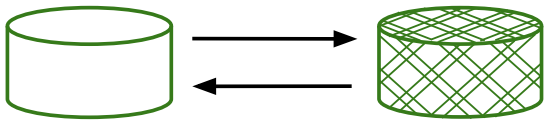
- Requirements capturing and representation

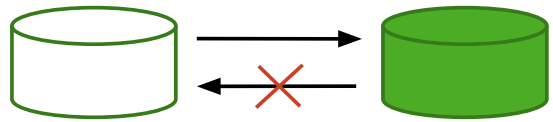
  policies regulating access, sharing, usage and processing



- Enforcing technologies

  data wrapping / sanitization

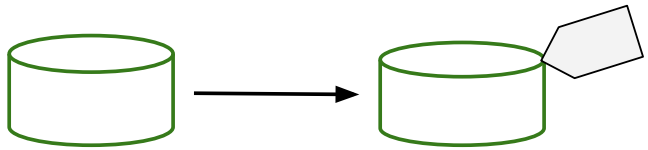




Dimensions of the problem and challenges

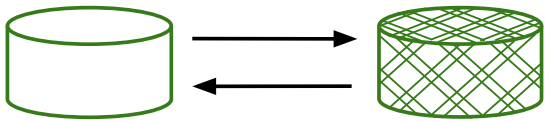
- Requirements capturing and representation

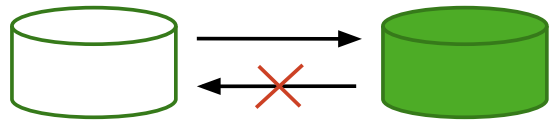
  policies regulating access, sharing, usage and processing



- Enforcing technologies

  data wrapping / sanitization



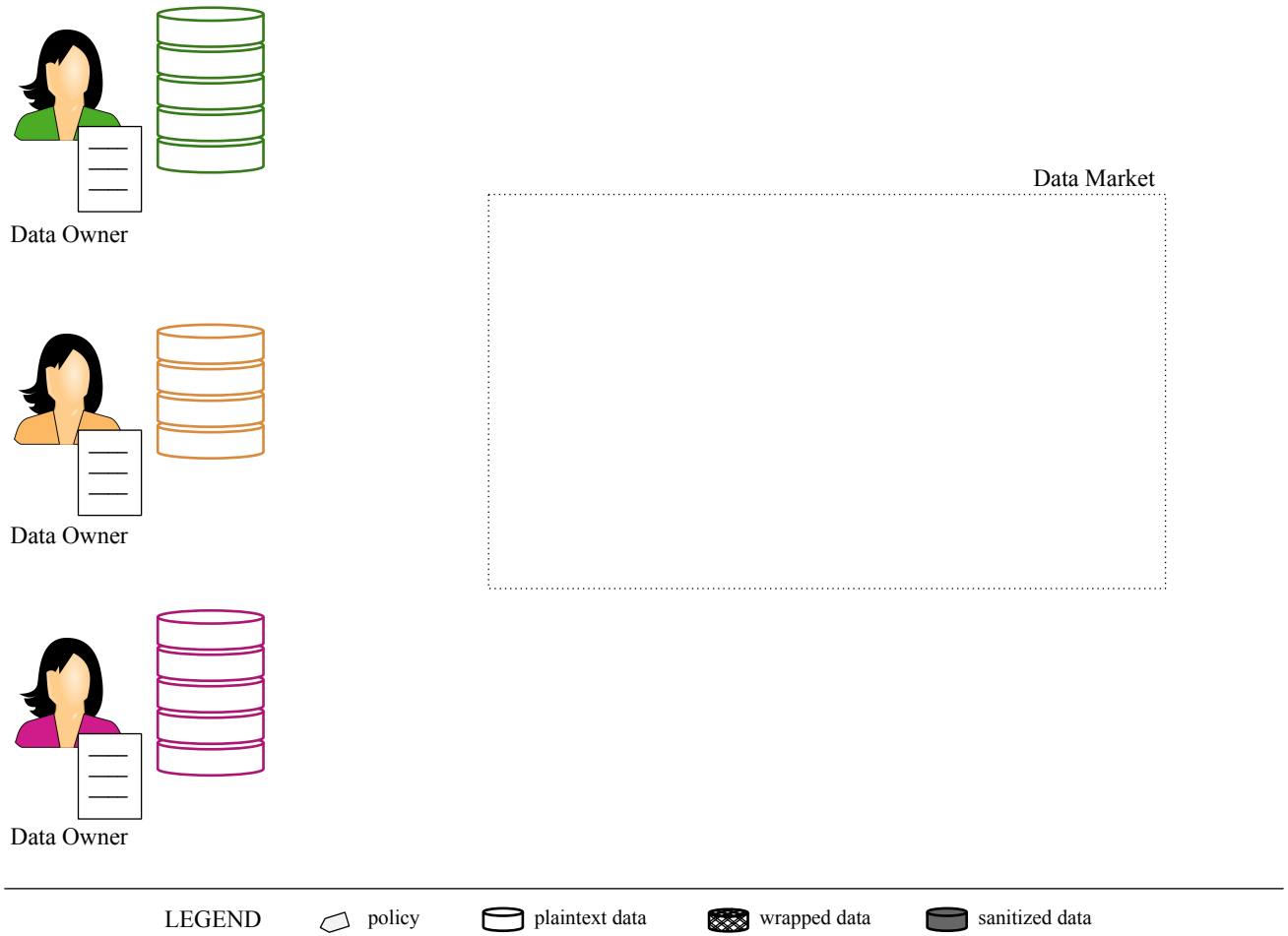


- Enforcement phase

  ingestion / storage / analytics

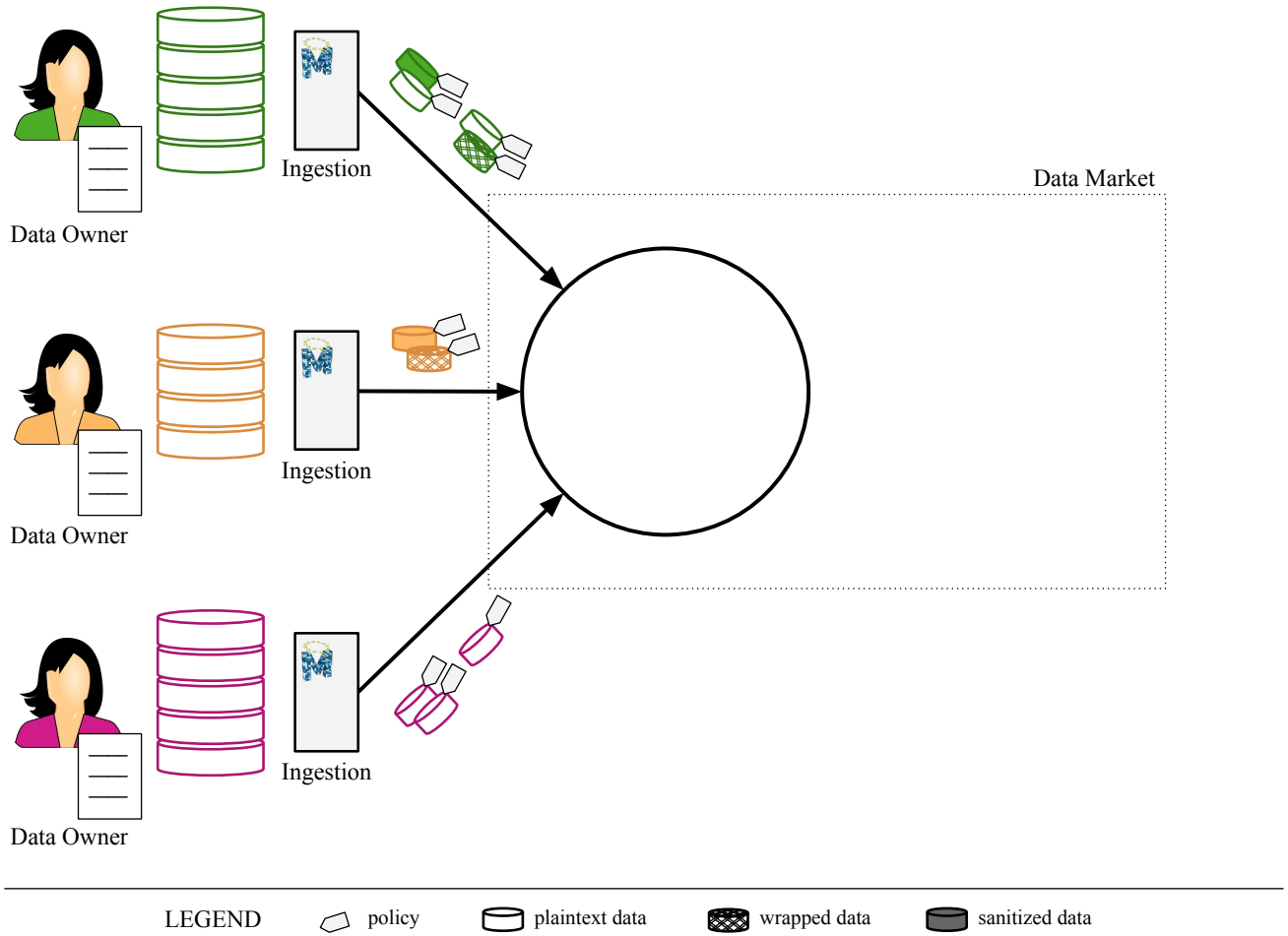
Enforcement phase

- Ingestion / Storage / Analytics



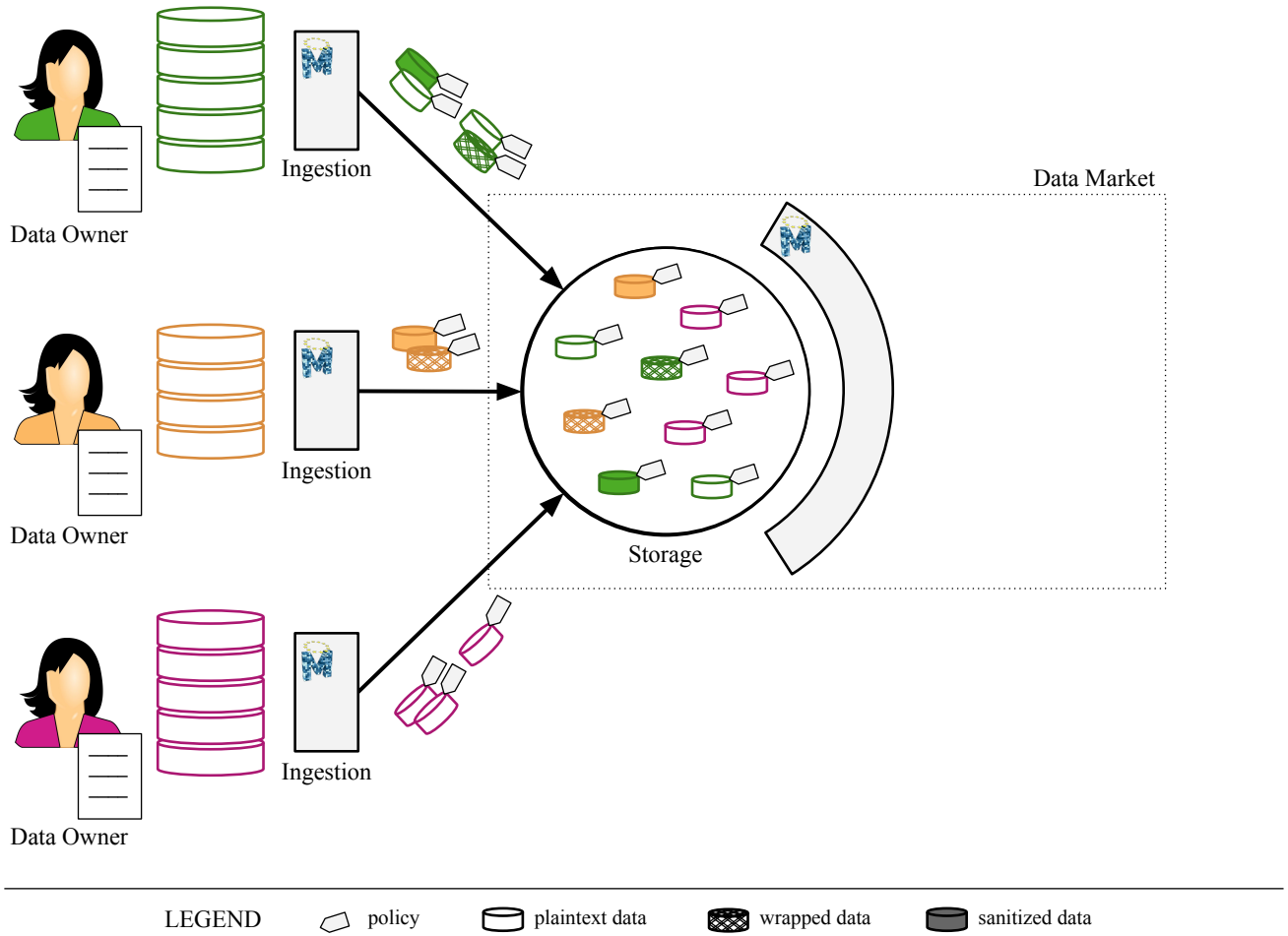
Enforcement phase

- Ingestion / Storage / Analytics



Enforcement phase

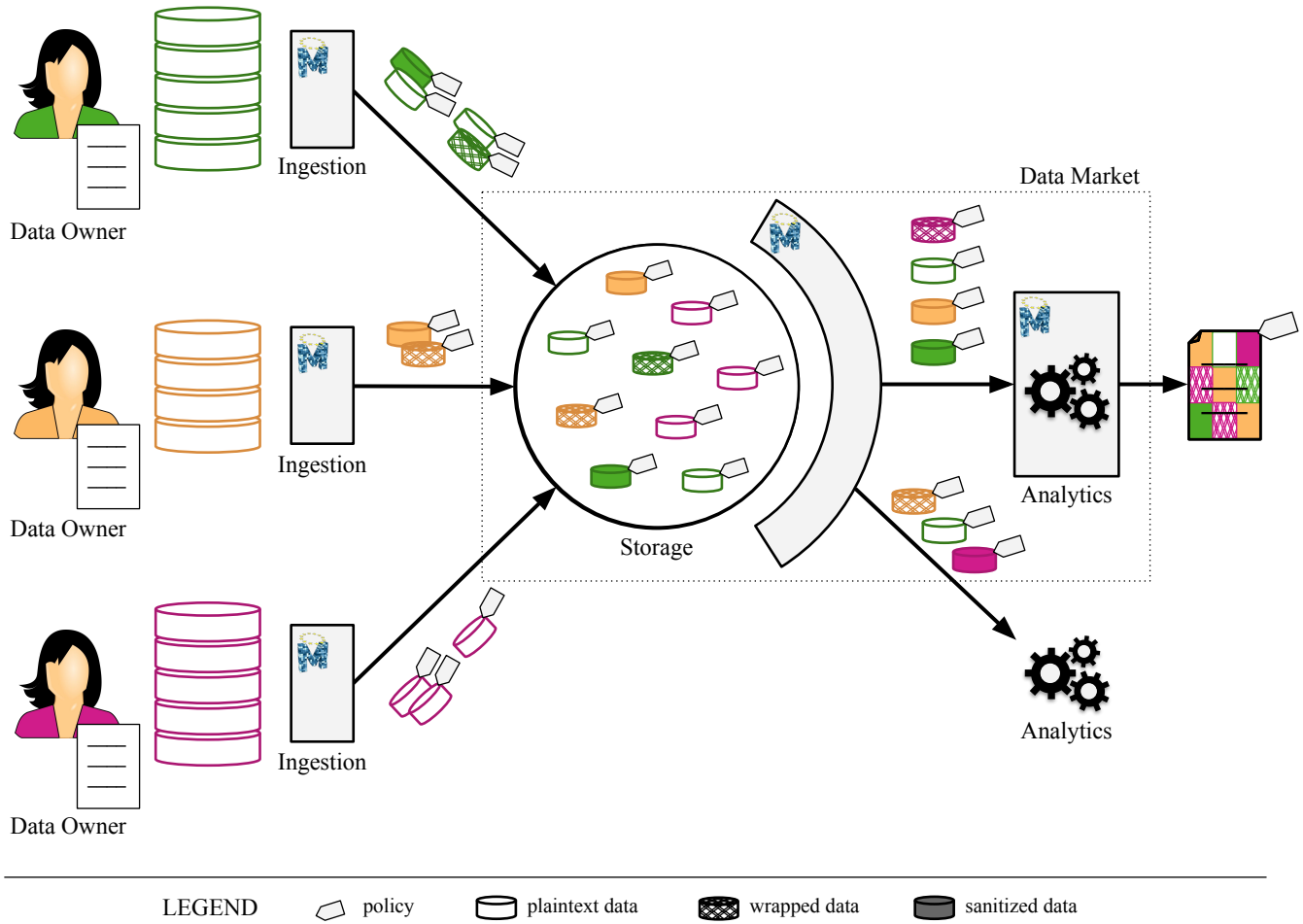
- Ingestion / Storage / Analytics



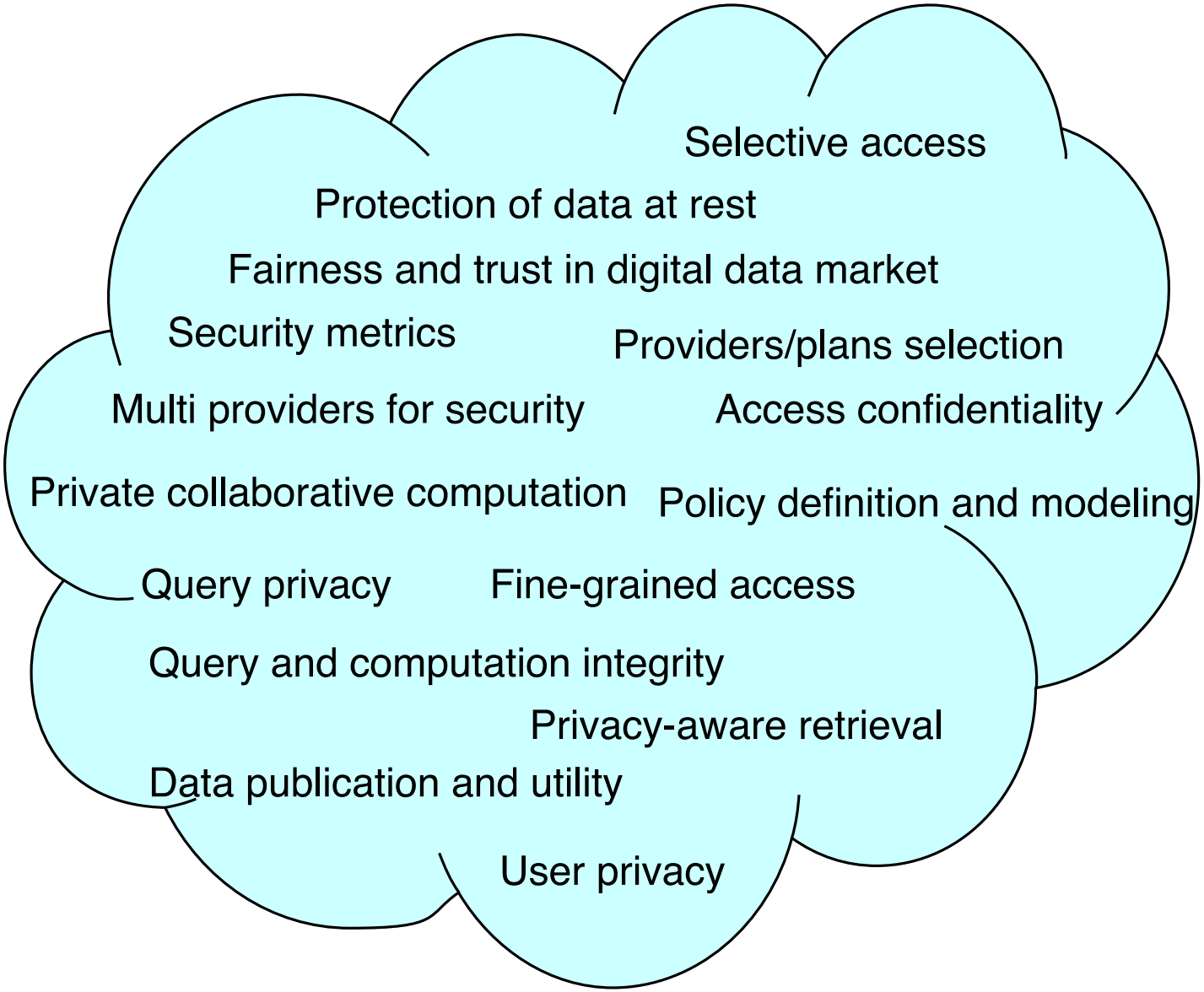
Multi-Owner data Sharing for Analytics and Integration respecting Confidentiality and OWNer control - https://mosaicrown.eu

Enforcement phase

- Ingestion / Storage / Analytics



Other open issues



Conclusions

- ICT advancements introduces:
  - new needs and risks for privacy
  - new opportunities for protecting privacy
- Lots of opportunities for new open issues to be addressed

. . . towards allowing society to fully benefit from information technology while enjoying security and privacy

"Before I write my name on the board, I'll need to know how you're planning to use that data."