Data Security and Privacy in Emerging Scenarios

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

Future Digileaders

Stockholm, Sweden – November 27, 2019

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 ...

Smart home, smart grid, …

NEWS

Home | Video | World | UK | Business | Tech | Science | Magazine | Entertainment & Arts

Building cities of the future now

By Jane Wakefield
Technology reporter

21 February 2013 | Technology

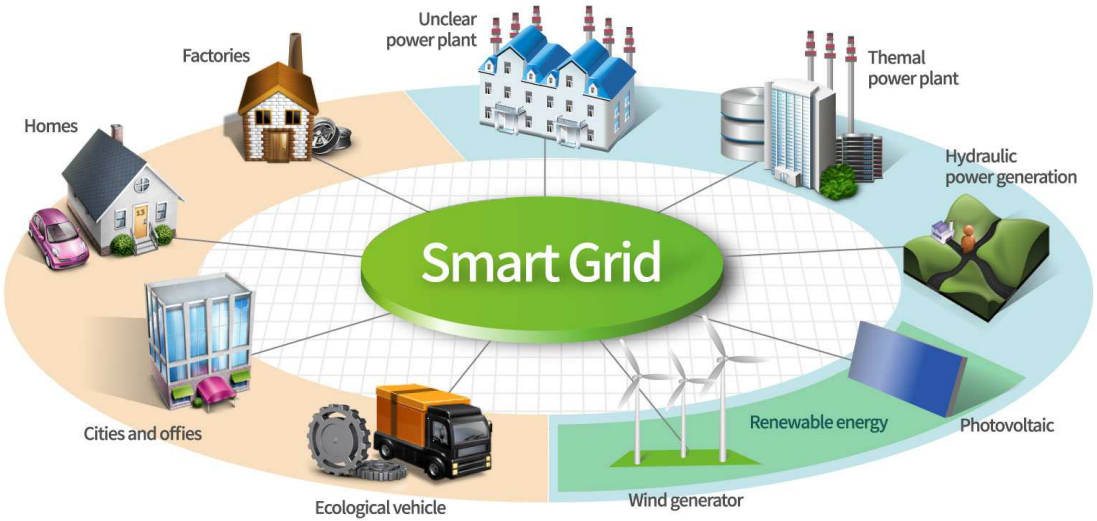
Share

Around the world new cities are being built while those we have lived in for centuries are being upgraded for the future.

It is partly a reaction to over-crowding and pollution and partly because in an ever-connected world it makes increasing sense to hook entire cities up to the network.

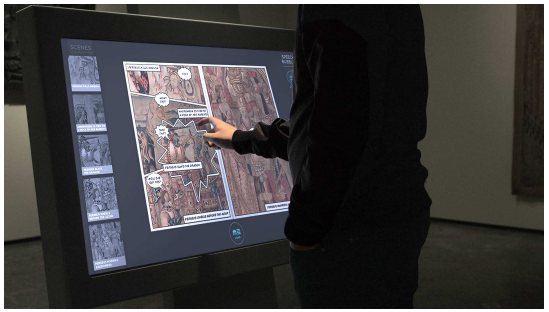
A smarter city may mean one that uses data on traffic to ease congestion or one that aims to join up services to provide better information for citizens. For many it is about

Technology being used in urban communities around the world hints at how we may live in



… Everything is getting smart

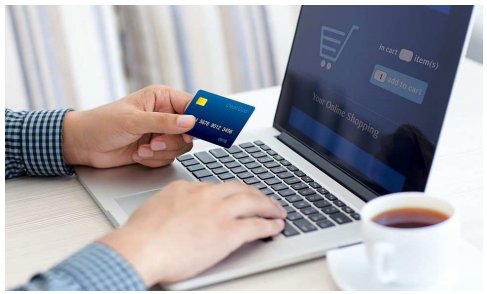
Smart car



Museum and exhibitions

Health Care

Augmented reality



Smart e-commerce

Intelligent shops

Smart entertainment systems

Smart governance

Smart transportation

Smart society

Smart services and security – Advantages

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

. . . but . . .

Smart services and security – Disadvantages

- More complexity . . .

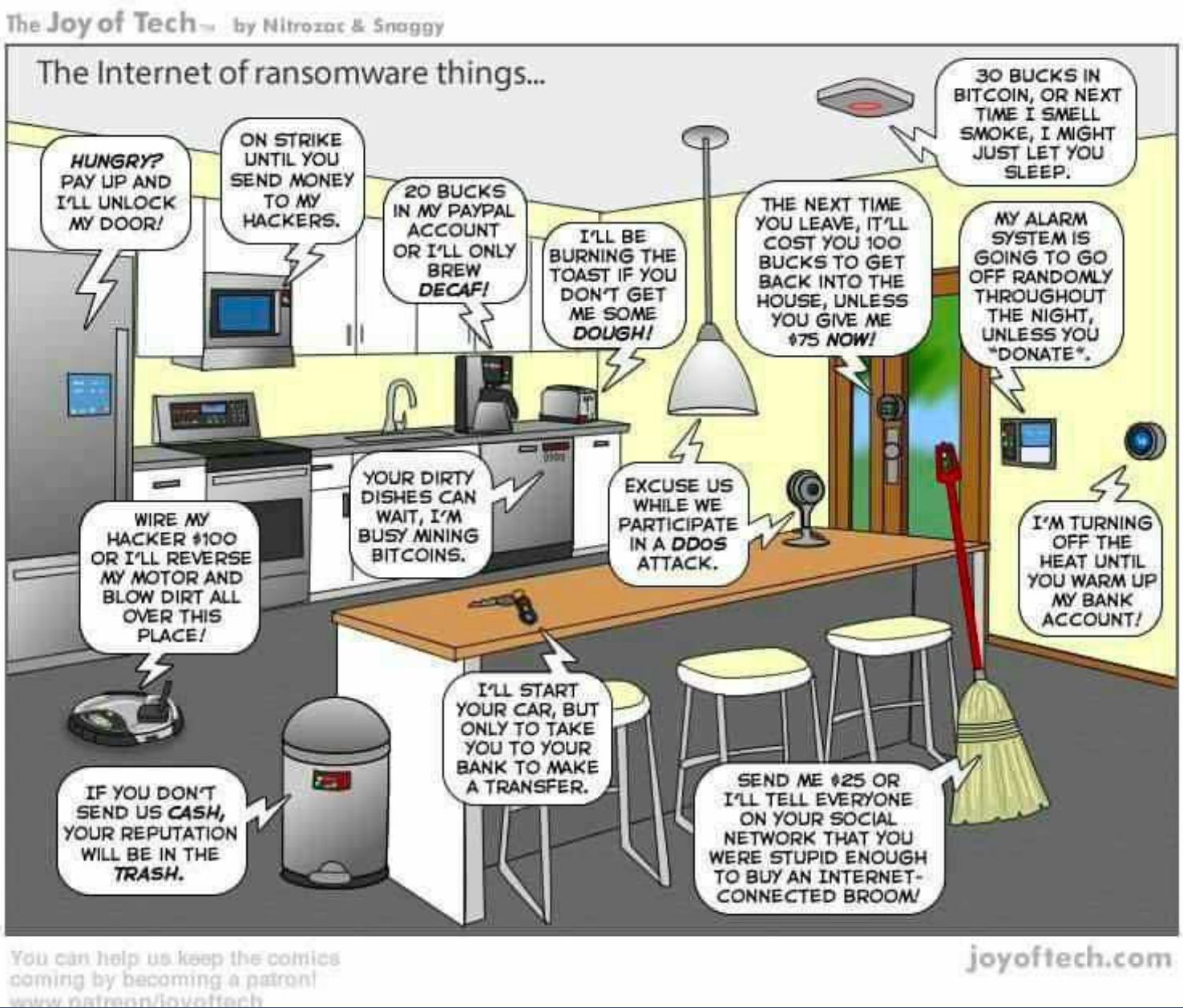
  . . . weakest link becomes a point of attack

  - 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

theguardian

all sections

ort football opinion culture business lifestyle fashion environment tech travel

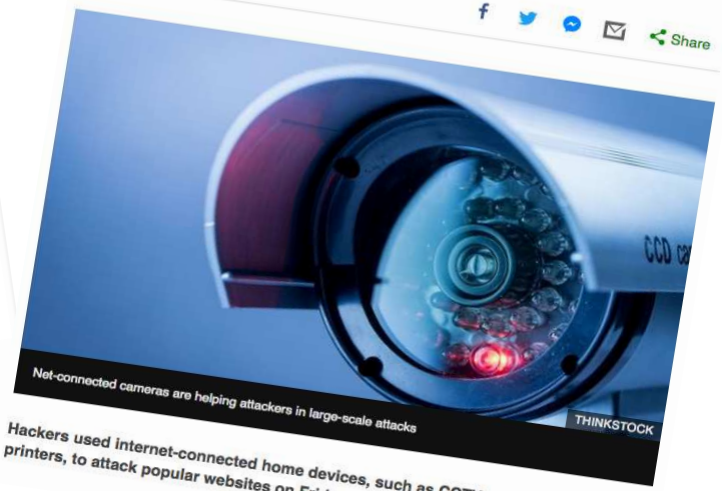
Samsung's voice-recording smart TVs breach privacy law, campaigners claim

US consumer rights group Epic claims Samsung has breached the privacy of its users, and is demanding an FCC investigation

'Smart' home devices used as weapons in website attack

22 October 2016 | Technology

Share



Hackers used internet-connected home devices, such as CCTV cameras and printers, to attack popular websites on Friday, security analysts say.

BBC | Sign in | News | Sport | Weather | Shop | Earth | Travel | Mor

NEWS

Home | Video | World | UK | Business | Tech | Science | Magazine | Entertainment & Arts

Technology

Warning over smart meters privacy risk

12 June 2012 | Technology

Share

An EU data watchdog has warned of the "considerable risks" to privacy posed by new energy smart meters.

The European Data Protection Supervisor said safeguards were needed over how firms used the "massive collection" of consumers' data uploaded by meters.

Smart meters are able to upload data about consumers' energy use to third parties

Home / News / Press Releases / Press Release

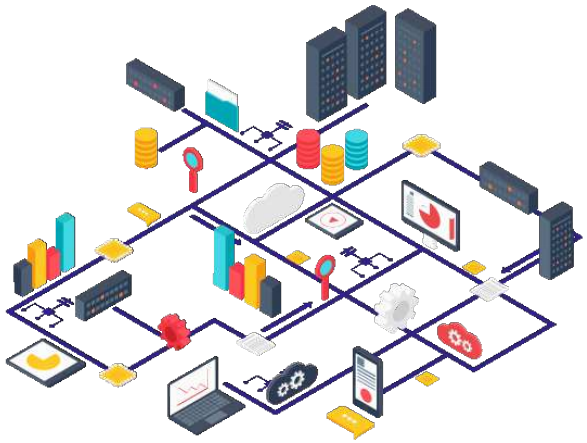
Markey Report Reveals Automobile Security and Privacy Vulnerabilities

Monday, February 9, 2015

Wireless technologies leave vehicles exposed to hackers; Information collected on driver locations, habits

WASHINGTON (February 9, 2014) – New standards are needed to plug security and privacy gaps in our cars and trucks, according to a report released today by Senator Edward J. Markey (D-Mass.). The report, called *Tracking & Hacking: Security & Privacy Gaps Put American Drivers at Risk* and first reported on by CBS News' 60 Minutes, reveals how sixteen major automobile manufacturers responded to questions from Senator Markey in 2014 about how vehicles may be vulnerable to hackers, and how driver information is collected and protected.

Security … a complex problem



Protection of infrastructure

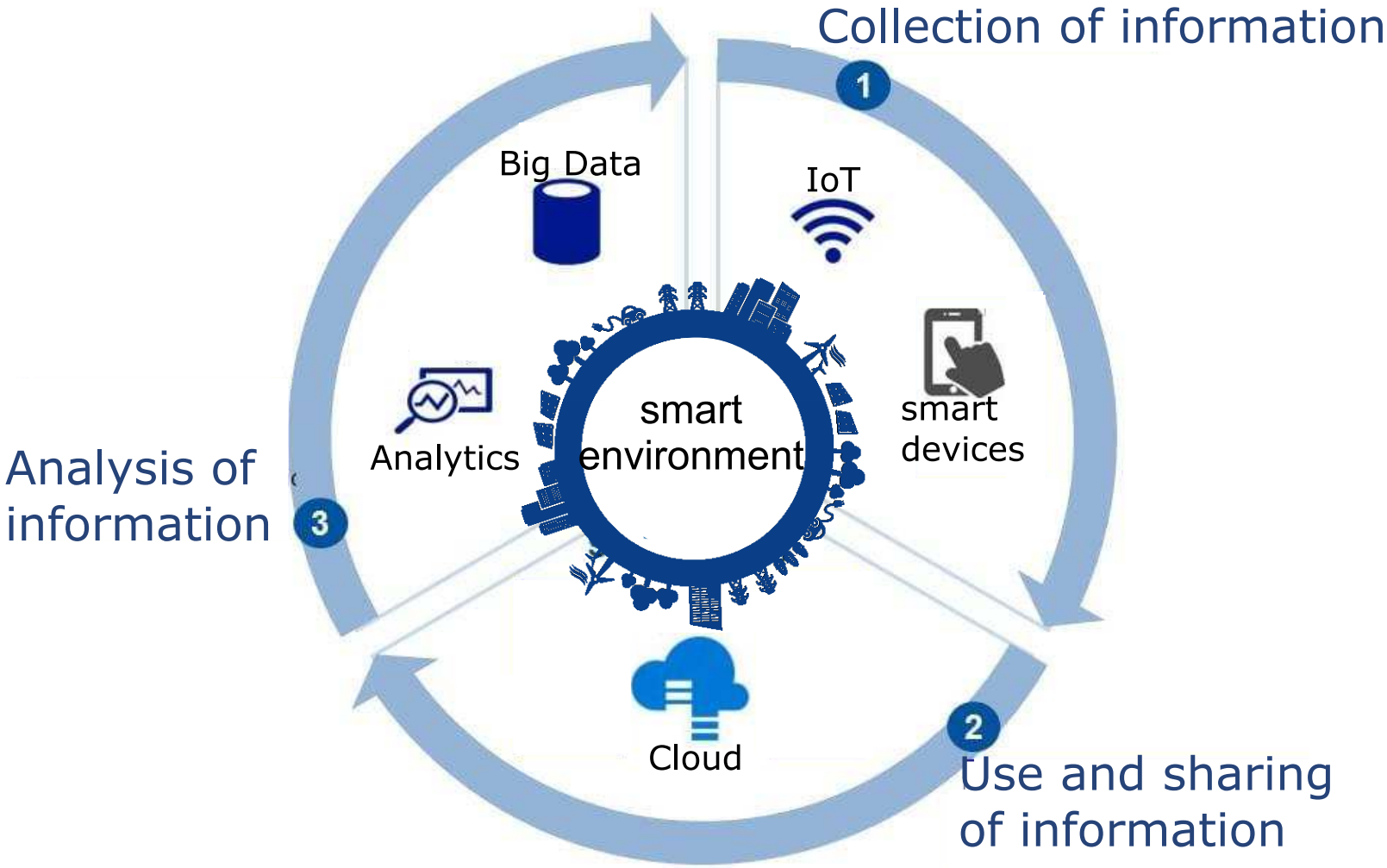
Protection of communication

Protection against malware and attacks

Protection of devices

Protection of data

The role of data in a smart environment

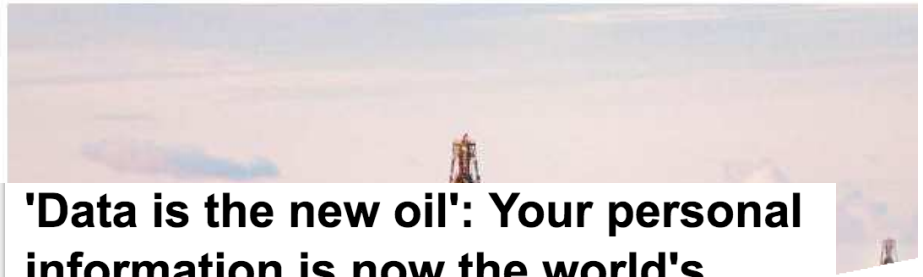


$\Longrightarrow$ better governance and intelligent systems

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
ct 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

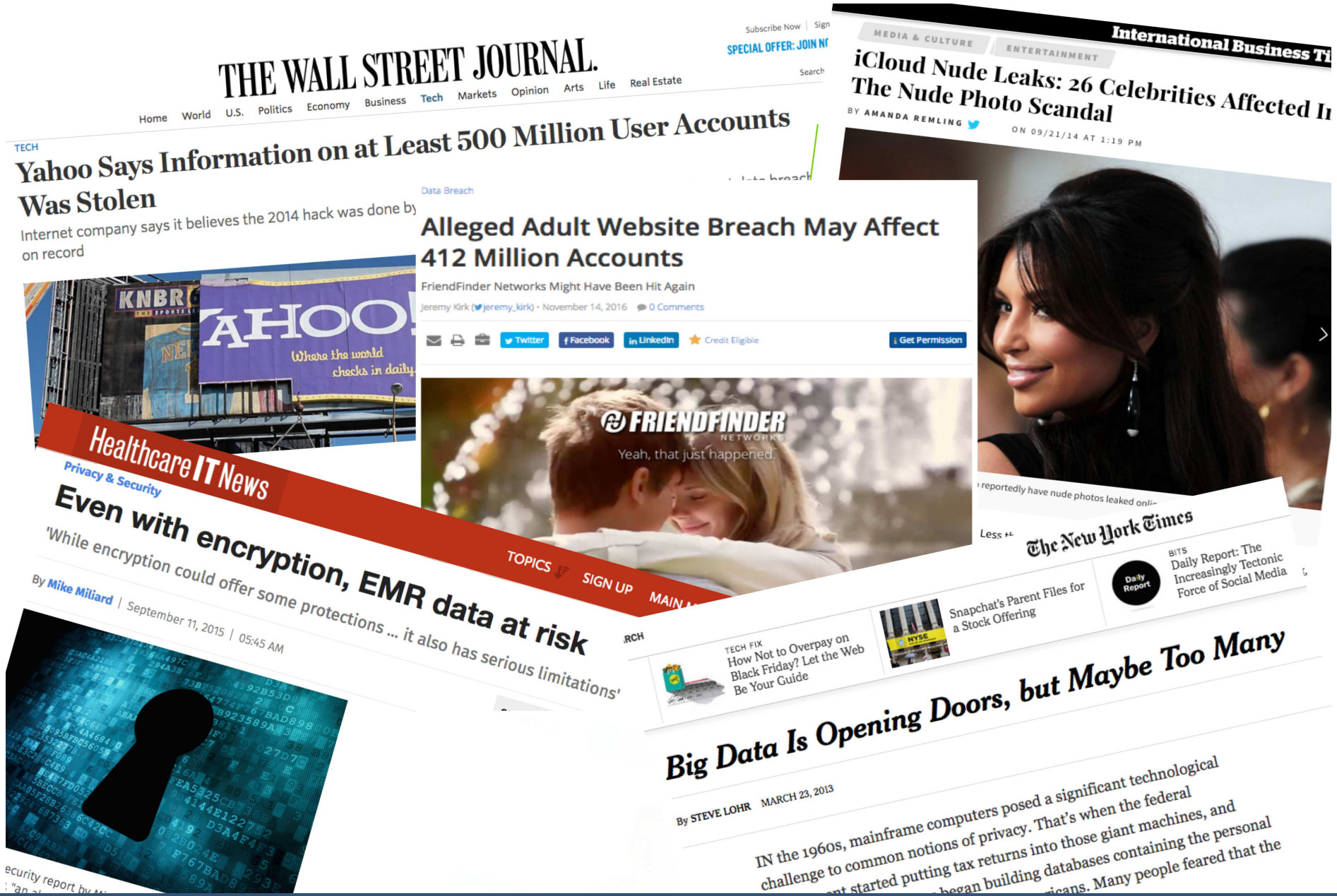
Data is now the world's most valuable resource according to *The Economist,* which reports on antitrust concerns about Alphabet (Google's parent company), Amazon, Apple, Facebook, and Microsoft, all of which have tons of data. The

bus

PARTNER CONTENT ARVIND SINGH

IS BIG DATA THE NEW BLACK GOLD?

Impact on data protection and privacy – 1



Impact on data protection and privacy – 2

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

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

My Page Like 2

G+

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

NEWS

63,500 records breached by misconfigured database

by Jessica Davis | April 12, 2018

19 DEC 2017 NEWS
Californian Voters Suffer Major Data Breach

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 6:30pm

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

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

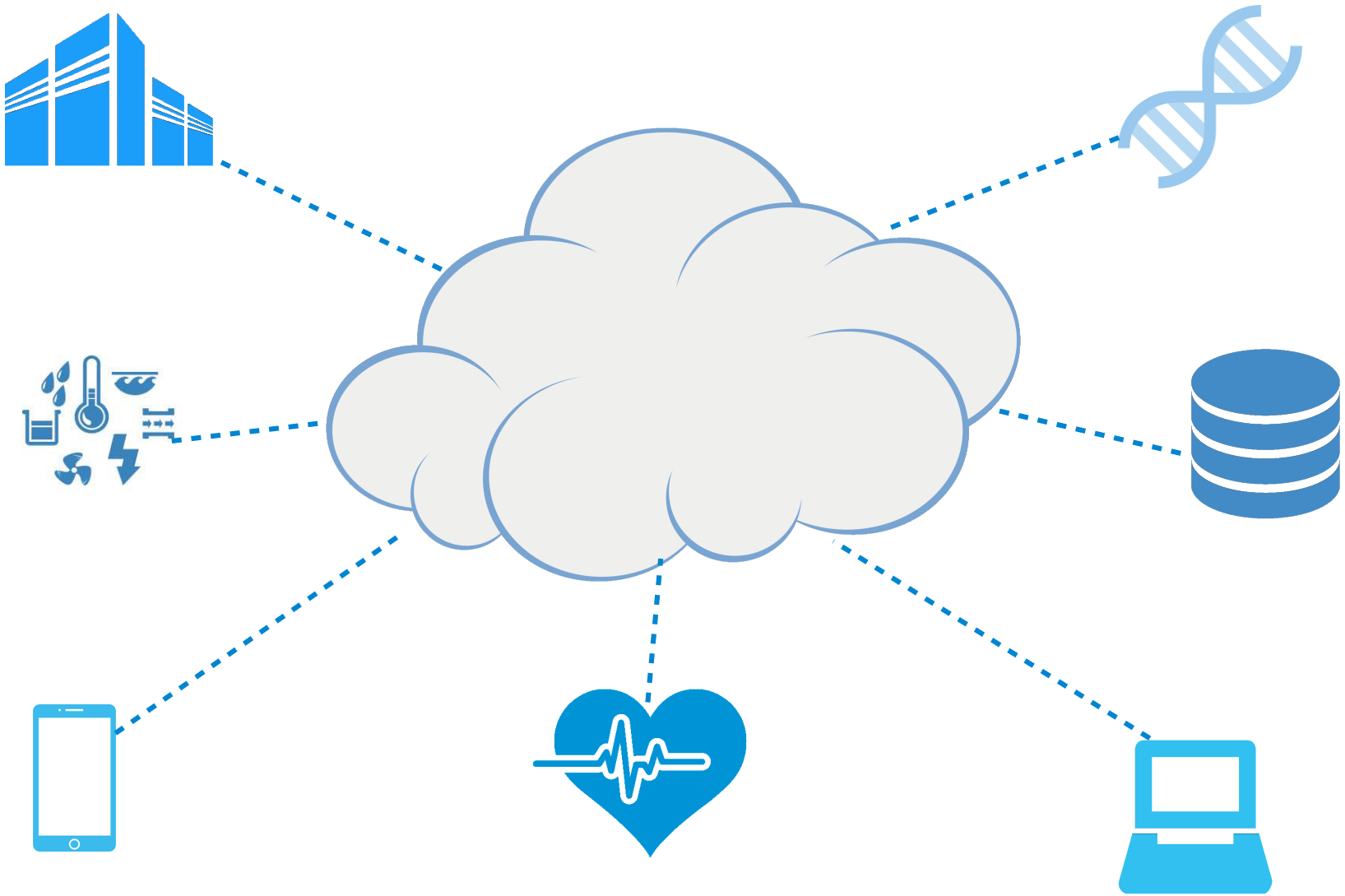
Privacy

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

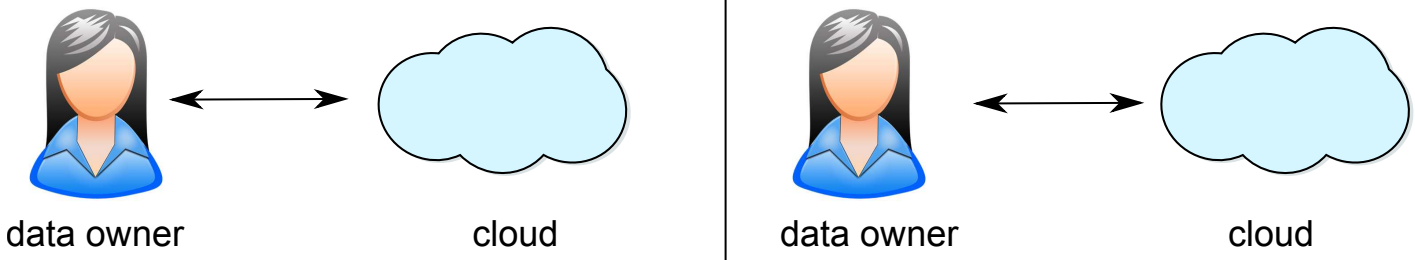


Cloud computing

- 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

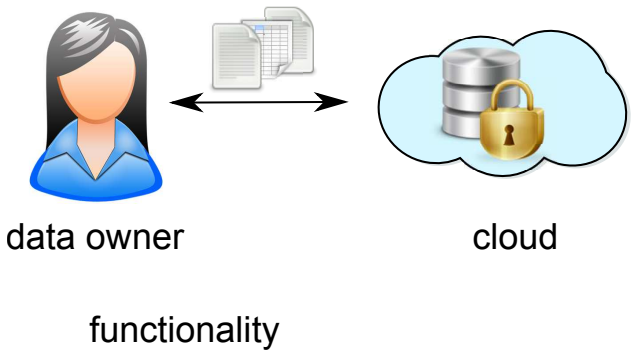
Cloud computing: Today

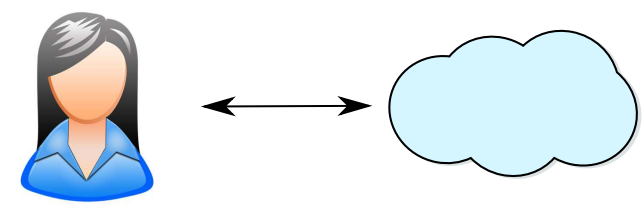
Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



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

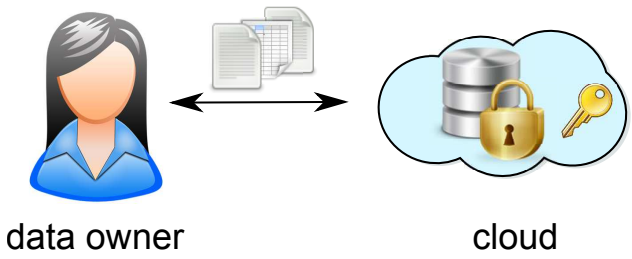




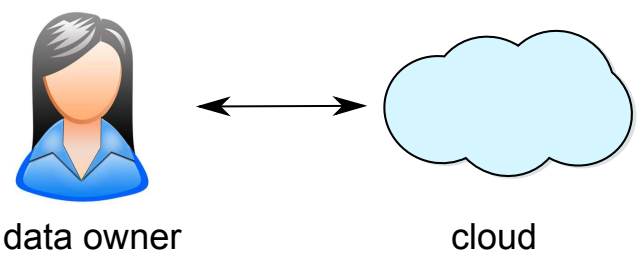
- functionality

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



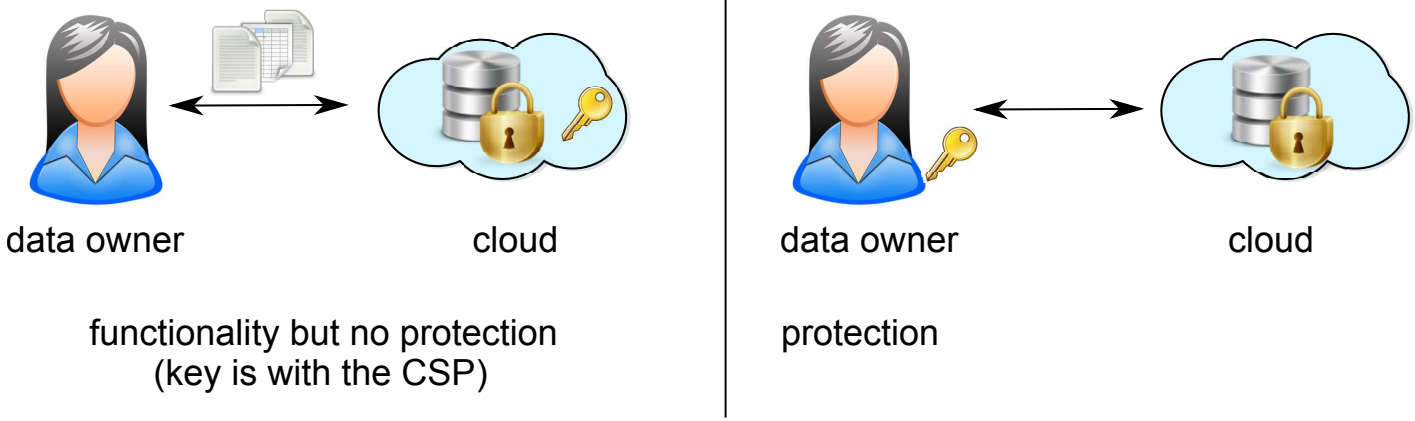
functionality but no protection
(key is with the CSP)



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

Cloud computing: Today

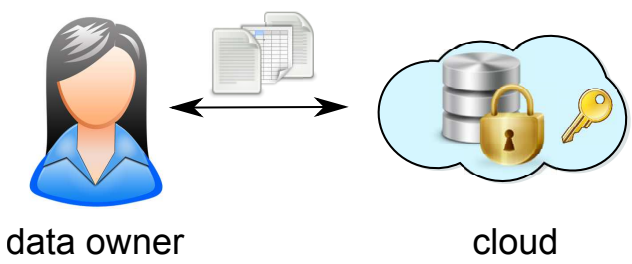
Cloud Service Providers (CSPs) apply security measures in the services they offer but these measures protect only the perimeter and storage against outsiders



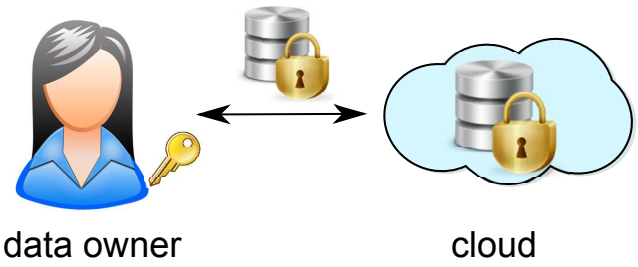
- functionality implies full trust in the CSP that has full access to the data (e.g., Google Cloud Storage, iCloud)
- protection

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)

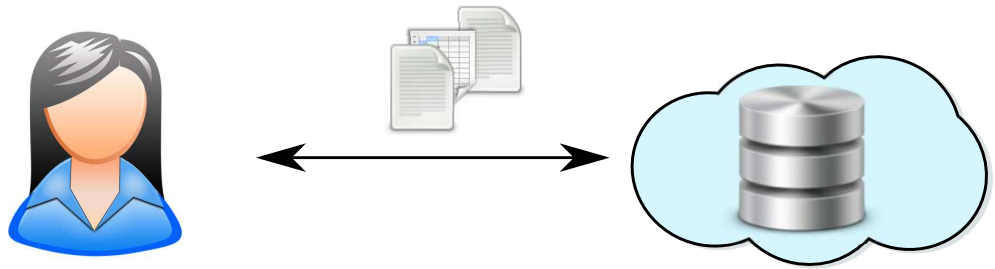


protection but limited functionality
(you cannot access data as you like)

- 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

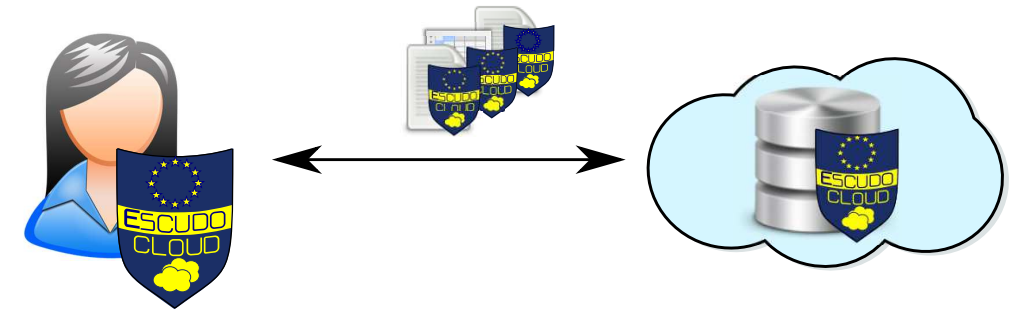


Enforceable Security in the Cloud to Uphold Data Ownership - http://www.escudocloud.eu/

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

---

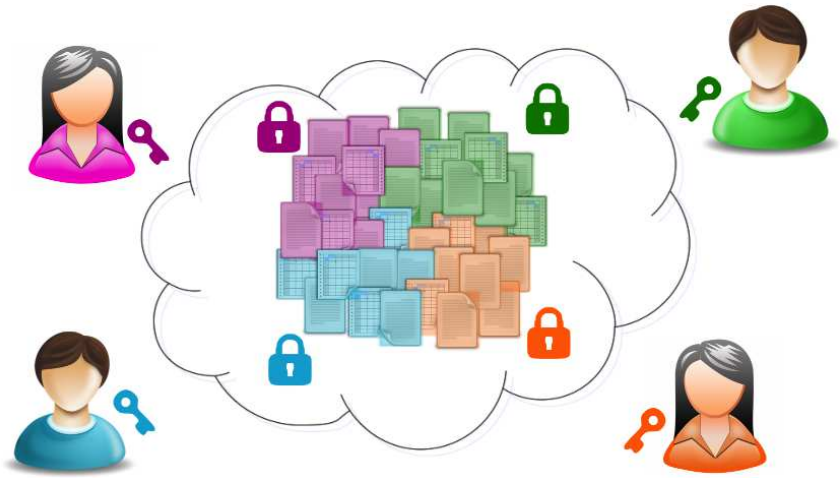
Enforceable Security in the Cloud to Uphold Data Ownership - http://www.escudocloud.eu/

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

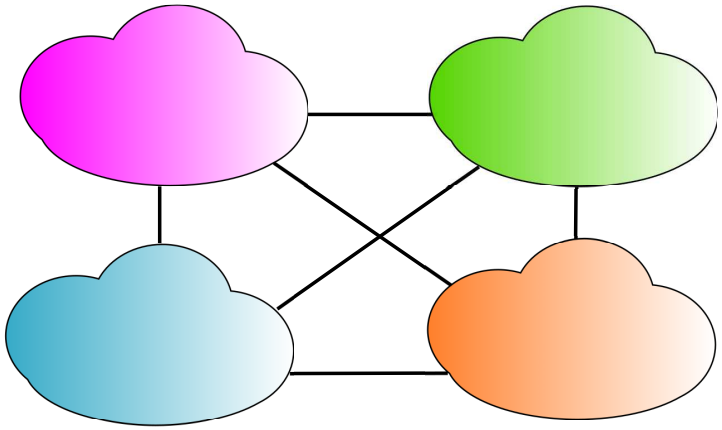
Data protection – Base level

Data protection – Base level

Data protection – Regulation



Access and usage control



Selective sharing

General Data Protection Regulation

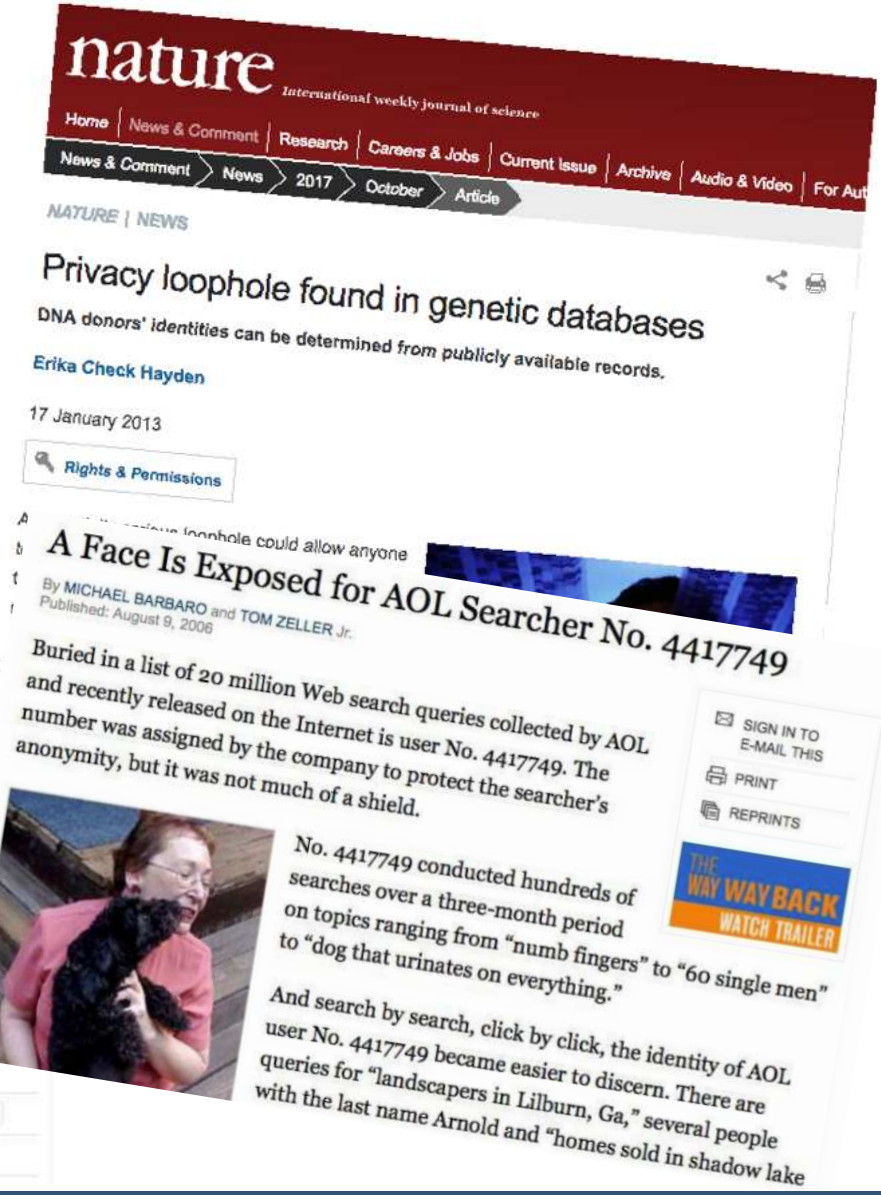
HIPAA Health Insurance Portability & Accountability Act

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)



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, AOL (America OnLine) publicly posted to a website 20 million search queries for 650,000 users of AOL's search engine summarizing three months of activity
- AOL suppressed any obviously identifying information such as AOL username and IP address
- AOL replaced these identifiers with unique identification numbers (this made searches by the same user linkable)

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”, and “bipolar”
  - “Arnold” (several people with this last name)
  - “landscapers in Lilburn, Ga”, “homes sold in shadow lake subdivision Gwinnett county, Georgia”

  $\Longrightarrow$ Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga

- She was re-identified by two New York Times reporters

- She explained in an interview that she has three dogs and that she searched for medical conditions of some friends

AOL data release – 3

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.

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 – 4

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 study – 1

- In 2006, Netlix (the world largest online movie rental service), launched the "Netflix Prize" (a challenge that lasted almost three years)
  - Prize of USD 1 million to be awarded to those who could provide a movie recommendation algorithm that improved Netflix's algorithm by 10%
- Netflix provided 100 million records revealing how nearly 500,000 of its users had rated movies from Oct.'98 to Dec.'05
- In each record Netflix disclosed the movie rated, the rating assigned (1 to 5), and the date of the rating

Netflix prize data study – 2

- Only a sample (one tenth) of the database was released
- Some ratings were perturbed (but not much to not alter statistics)
- Identifying information (e.g., usernames was removed), but a unique user identifier was assigned to preserve rating-to-rating continuity

Netflix prize data study – 3

- De-identified Netflix data can be re-identified by linking with external sources (e.g., user ratings from IMDb users)
  - Knowing the precise ratings a person has assigned to six obscure (outside the top 500) movies, an adversary is able to uniquely identify that person 84% of the time
  - Knowing approximately when ($\pm$ 2 weeks) a person has rated six movies (whether or not obscure), an adversary is able to reidentify that person in 99% of the cases
  - Knowing two movies a user has rated, with precise ratings and rating dates ($\pm$ 3 days), an adversary is able to reidentify 68% of the users

Another example of privacy issue

Movies may reveal your political orientation, religious views, or sexual orientations (Netflix was sued by a lesbian for breaching her privacy)

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 Target case – 1

- Target is the second-largest discount retailer in the U.S.
- Target assigns every customer a Guest ID number:
  - tied to credit card, name, email address, . . .
  - stores history of bought goods and other (bought) information
  - mining on these data for targeted advertising

The Target case – 2

Forbes · Real Time · Most Popular (Most Reputable Companies) · Lists (The World's Billionaires) · Vid (Most)

28639 Share

26688 Tweet

2.4k +1

4662 Share

9313 reddit

329 Submit

Kashmir Hill, Forbes Staff

Welcome to The Not-So Private Parts where technology & privacy collide

+ Follow (631) Subscribe 112k

TECH | 2/16/2012 @ 11:02AM | 1,544,998 views

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

259 comments, 150 called-out + Comment now

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

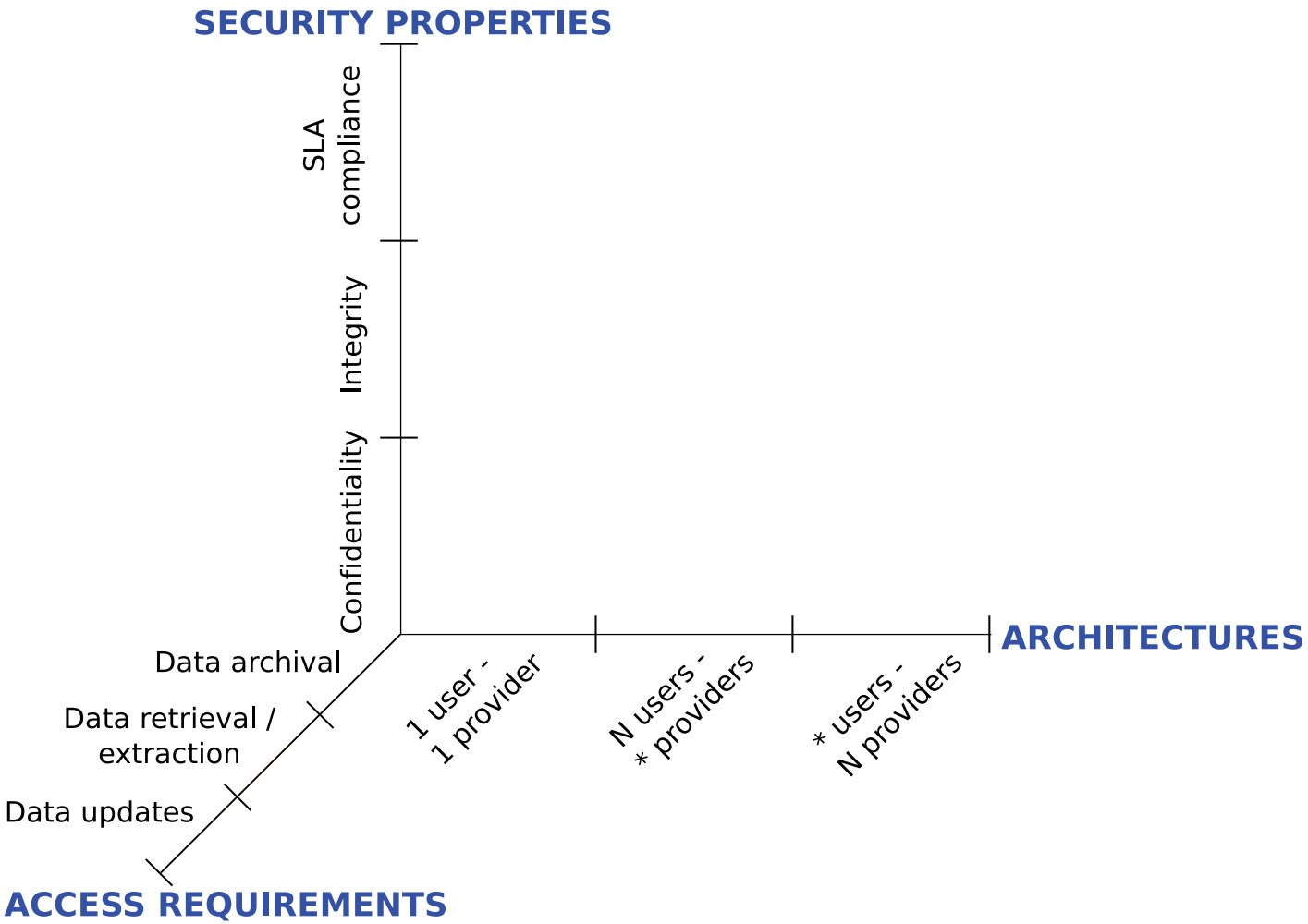
The Target case – 3

- Analysts at Target identified $\sim$ 25 products that assign each shopper a pregnancy prediction score
  - e.g., woman, 23 y.o., buying in March cocoa-butter lotion, a purse large enough to double as a diaper bag, zinc and magnesium supplements and a bright blue rug $\Longrightarrow$ 87% due late August
  - due time in a small window to send coupons timed to very specific stages of a pregnancy
- Mining data reveals customers' major life events (e.g., graduating from college or getting a new job or moving to a new town)
  - shopping habits became flexible, predictable, and potential gold mines for retailers
  - between 2002 (starting of similar campaigns) and 2010 Target's revenues grew from USD 44B to USD 67B

Characterization of Data Protection Challenges in Cloud Scenarios

Scientific and technical challenges

Three dimensions characterize the problems and challenges



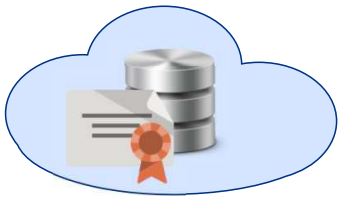
Security properties

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

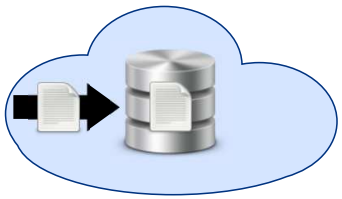
Access requirements

Data archival

- upload/download
- protection of data in storage

Data retrieval/extraction

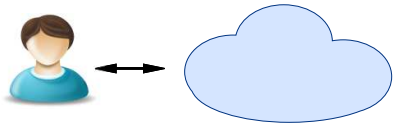
- support for fine-grained data retrieval and queries
- protection of computations and query results



Data update

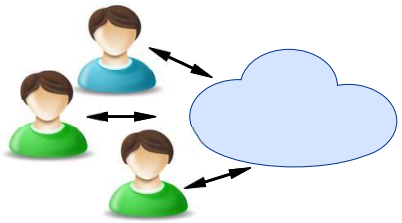
- support for access retrieval and enforcement of updates
- protection of the actions and of their effects on the data

Architectures



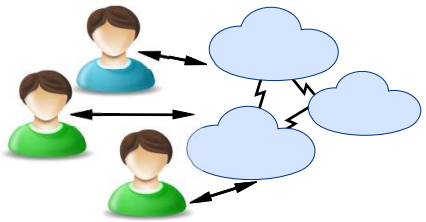
1 user - 1 provider

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



n users - * providers

- authorizations and access control
- multiple writers



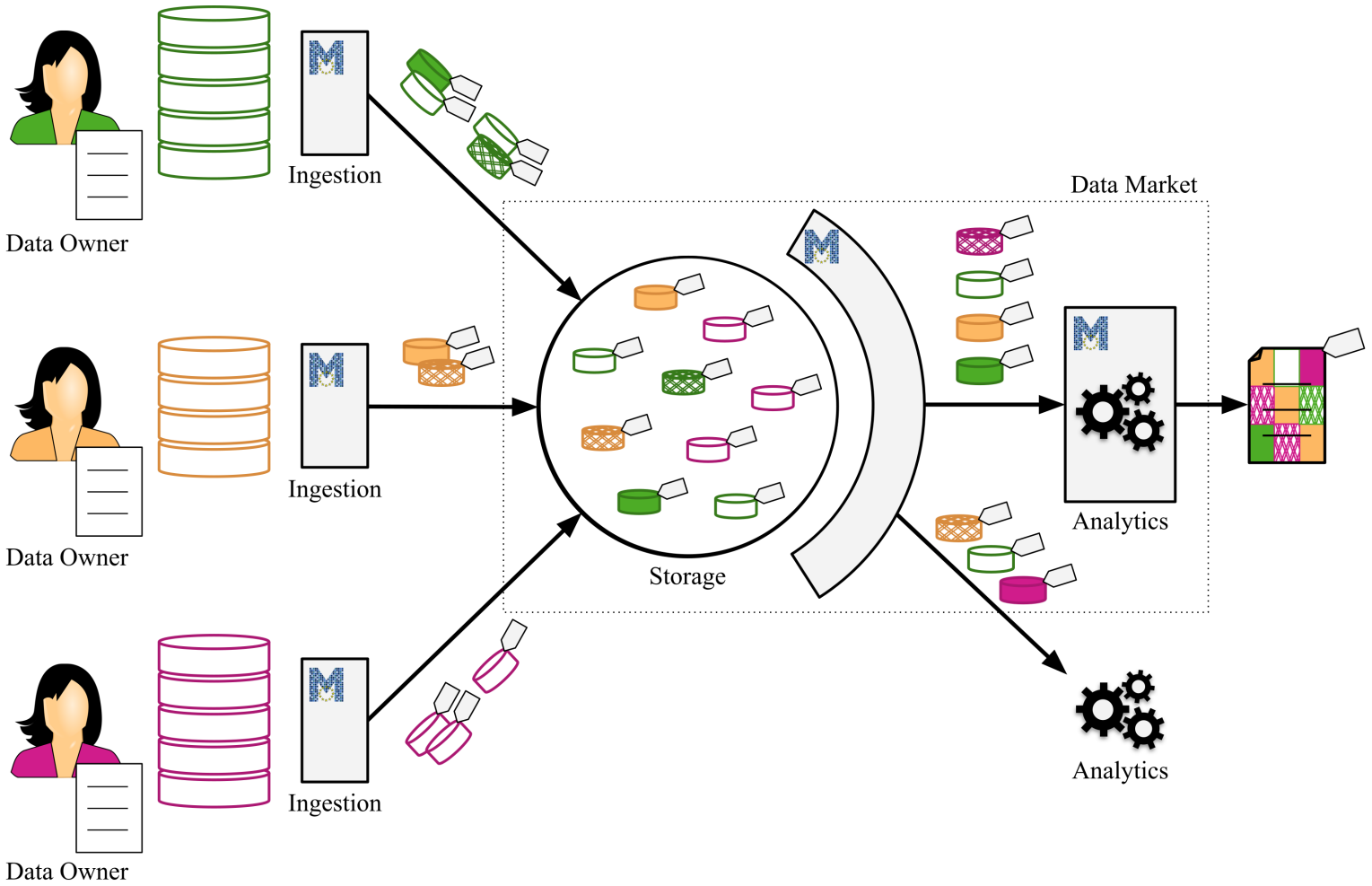
*** users - n providers**

- controlled data sharing and computation

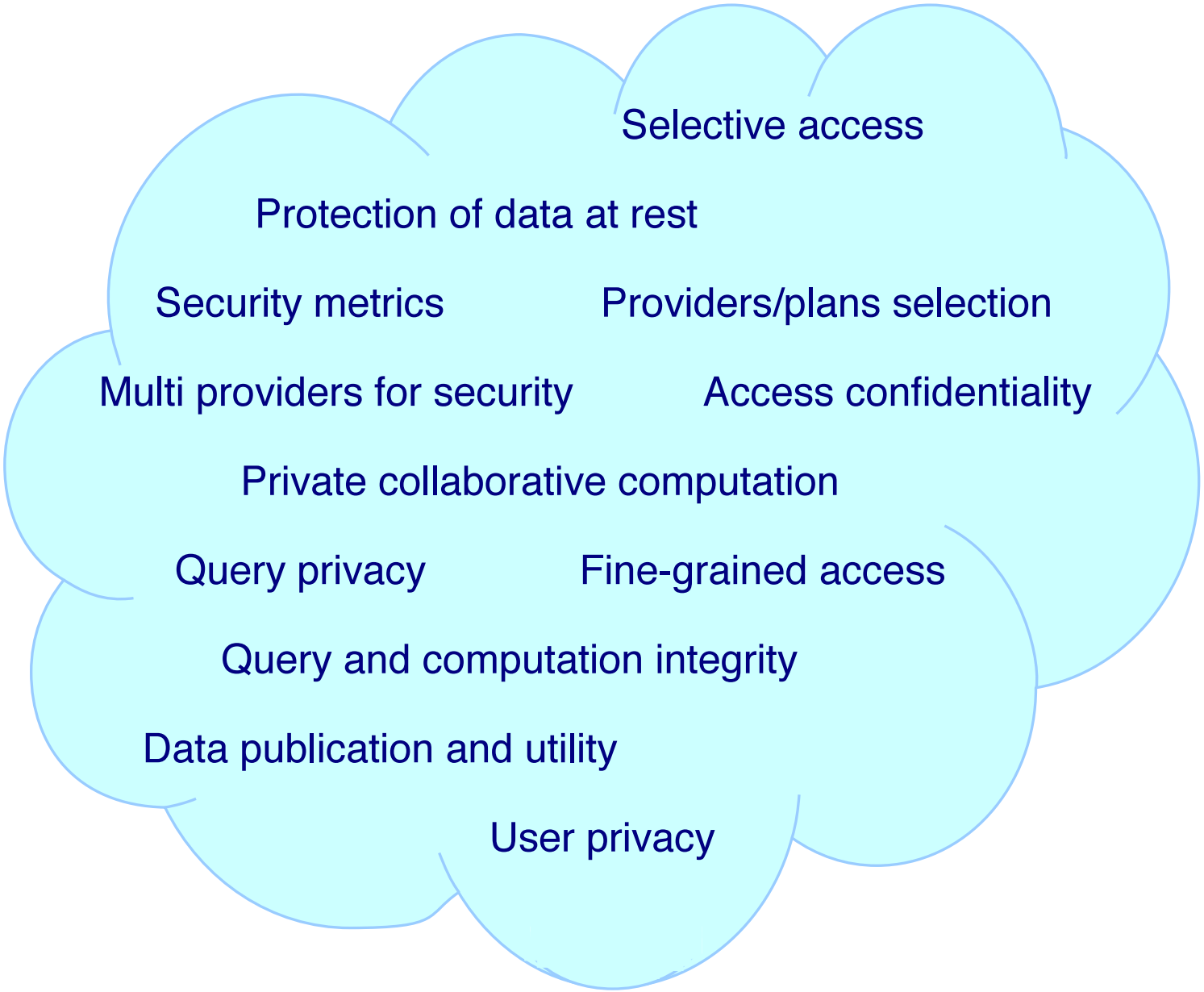
Combinations of the dimensions

- 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

Data protection in the digital data market



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."