Big Data, sewage monitoring and botnets

BART VAN DER SLOOT
TILBURG INSTITUTE FOR LAW, TECHNOLOGY, AND SOCIETY (TILT)
WWW.BARTVANDERSLOOT.NL

Overview

- ▶ (1) Interactive debate
- ▶ (2) Big Data
- ► (3) Break
- ▶ (4) Interactive debate
- ▶ (5) GDPR
- ▶ (6) Break
- ▶ (7) Interactive debate
- ▶ (8) Sewage monitoring and botnets

Case 1

A mayor of a big city is having an affair outside his marriage. Can a television program secretly film him while he is dating in a bar?



Case 2

Tomtom is collecting real time traffic data without the explicit consent of the users. Can it also sell the data to the police?



Question 1



Would you discuss your income with strangers?

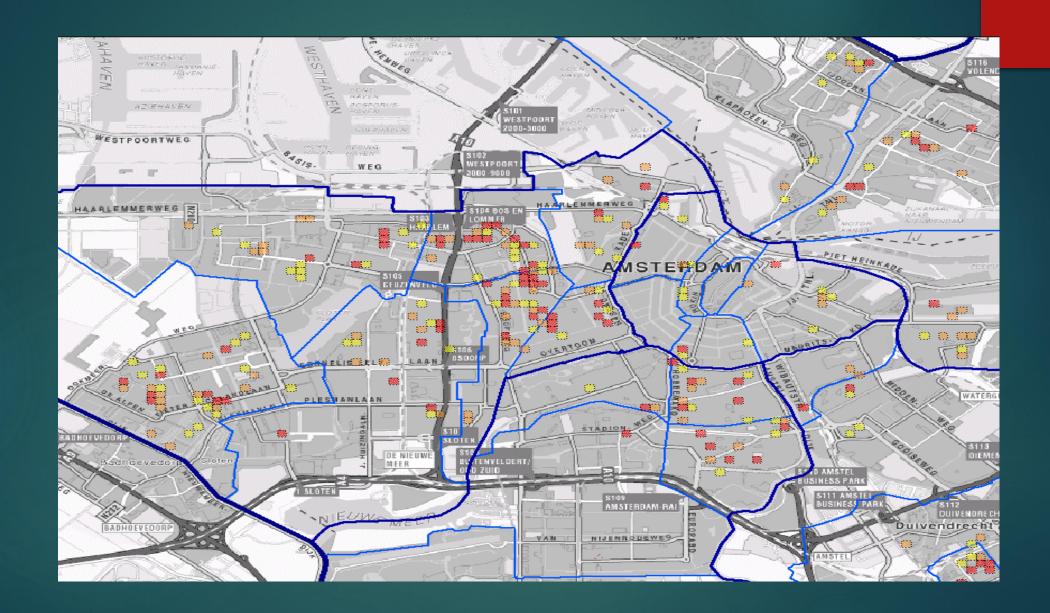
Case 4

The Intelligence agency gathers data about every Dutch citizen. Can it also share this data with American intelligence agencies?



(2) Big Data





DATA LEAKAGE IN ONLINE ADVERTISING

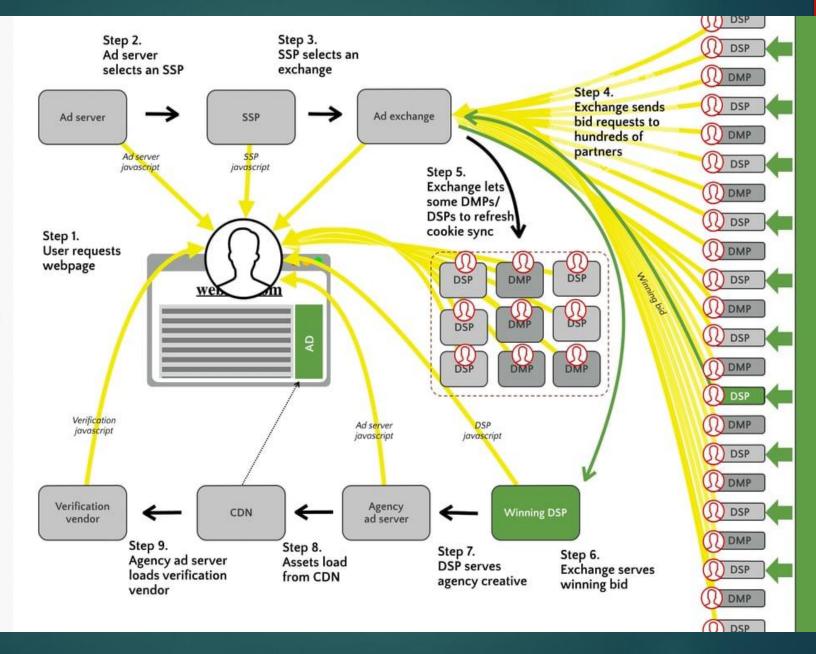
This is the current process of real-time bidding that is used in online behavioural advertising.

Legend

Channel of data leakage



Personally identifiable information







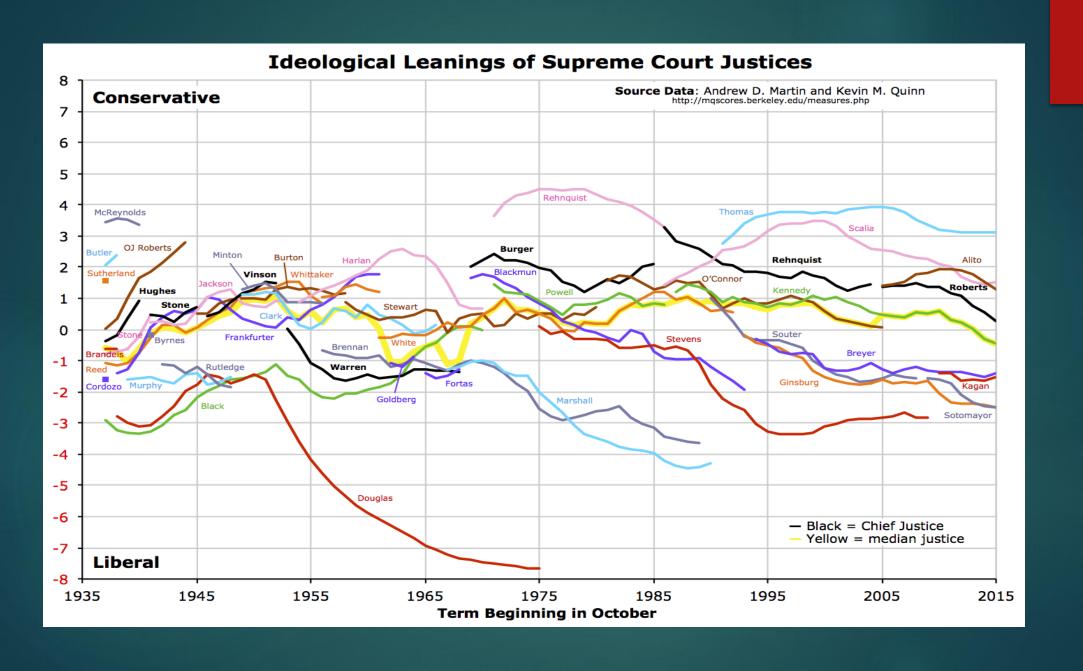


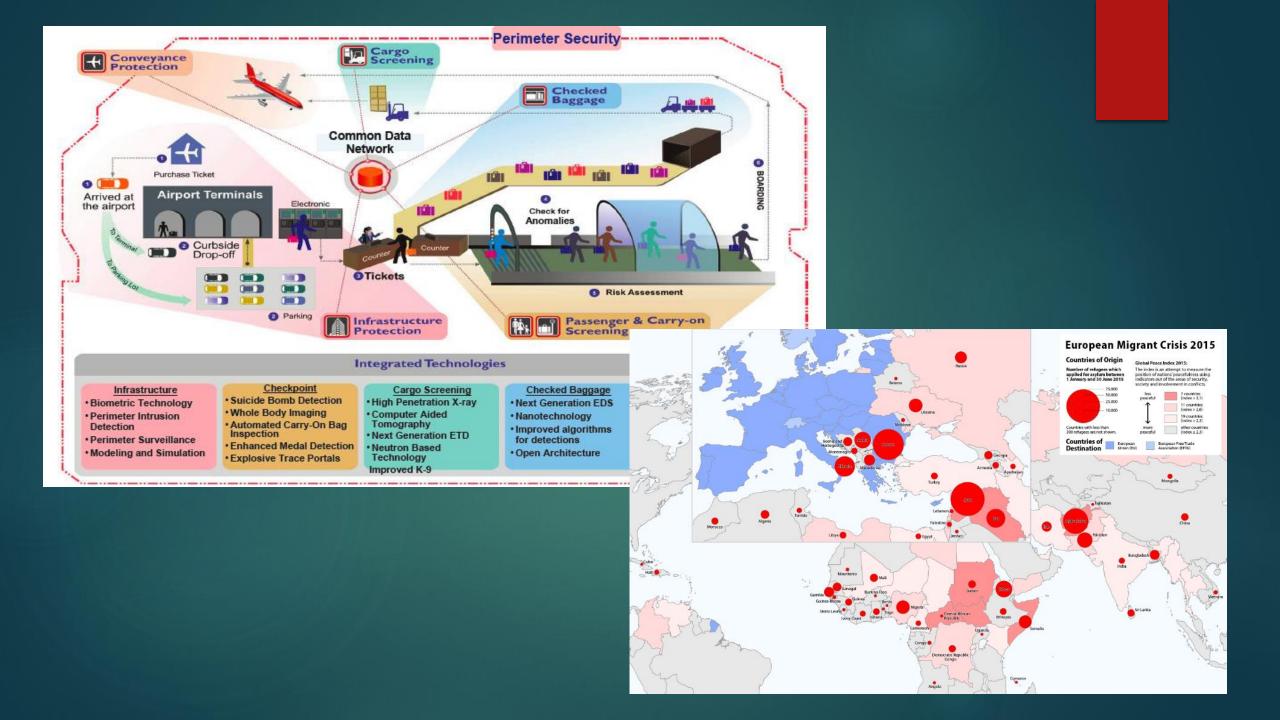


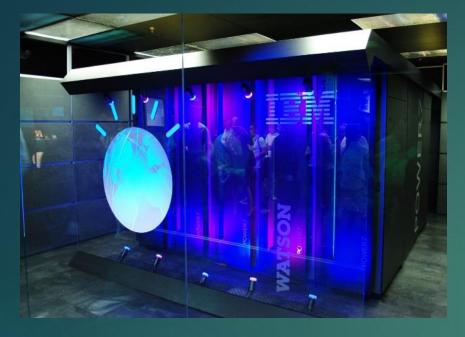




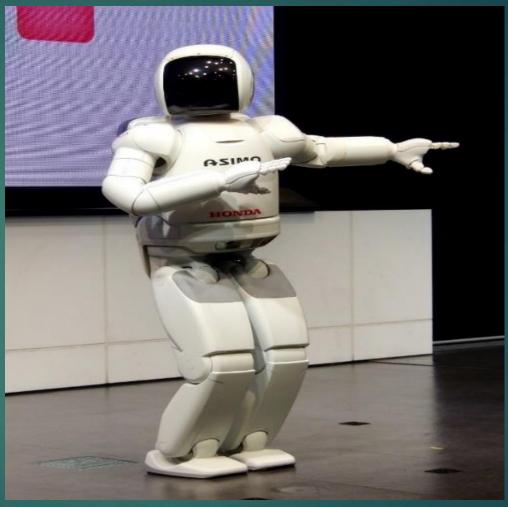






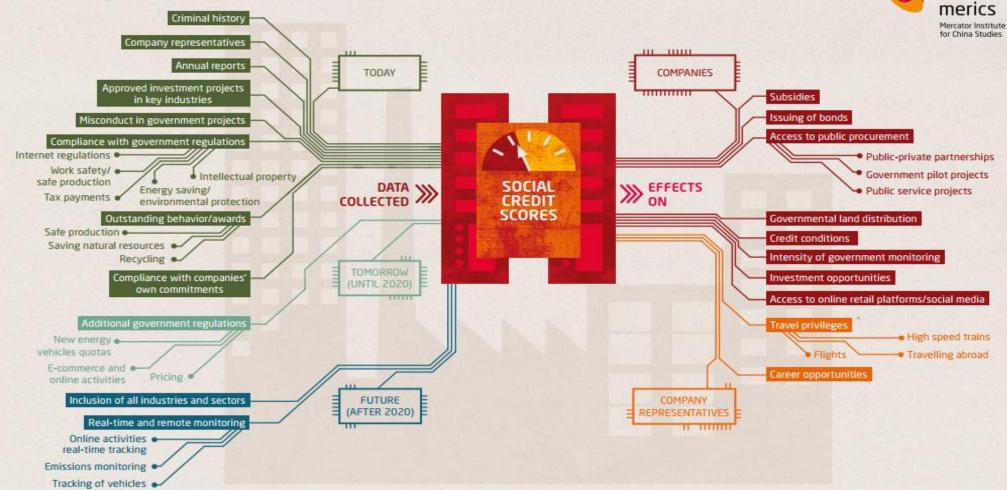






China's tight grip on enterprises

Influencing business decisions via Social Credit Scores*



^{*}Selection of data collected and exemplary effects of Social Credit Scores.

Source: Policy documents and regulations released by the Chinese central government since 2014.

Big Data does not exist

- 'Big Data refers to the exponential growth both in the availability and in the automated use of information: it refers to gigantic digital datasets held by corporations, governments and other large organisations, which are then extensively analysed (hence the name: analytics100) using computer algorithms. Big Data can be used to identify more general trends and correlations but it can also be processed in order to directly affect individuals.'
- Article 29 Working Part
- b 'Big Data is a term which refers to the enormous increase in access to and automated use of information. It refers to the gigantic amounts of digital data controlled by companies, authorities and other large organizations which are subjected to extensive analysis based on the use of algorithms. Big Data may be used to identify general trends and correlations, but it can also be used such that it affects individuals directly.'
- International Working Group on Data Protection in Telecommunication
- "Big Data means large amounts of different types of data produced at high speed from multiple sources, whose handling and analysis require new and more powerful processors and algorithms. Not all of these data are personal, but many players in the digital economy increasingly rely on the large scale collection of and trade in personal information. As well as benefits, these growing markets pose specific risks to individual's rights to privacy and to data protection.'

European Data Protection Supervisor

"There are many definitions of 'Big Data' which may differ depending on whether you are a computer scientist, a financial analyst, or an entrepreneur pitching an idea to a venture capitalist. Most definitions reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data. In other words, 'data is now available faster, has greater coverage and scope, and includes new types of observations and measurements that previously were not available.'

Podesta Report

▶ "Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze."

Het McKinsey Global Institute (2011: 1) http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/

Big Data is defined "collected and processed open datasets, which are defined by quantity, plurality of data formats and data origination and processing speed."

De autoriteit gegevensbescherming van Estland

"In August 2014, a definition of the term 'Big Data' was adopted by the French General Commission on terminology and neology (Commission générale de terminologie et de néologie). The official translation of this term in French is 'mégadonnées' and the definition is "data, structured or otherwise, whose very large volume require appropriate analytical tools."

De autoriteit gegevensbescherming van Frankrijk

"Big Data stems from the collection of large structured or unstructured datasets, the possible merger of such datasets as well as the analysis of these data through computer algorithms. It usually refers to datasets which cannot be stored, managed and analysed with average technical means due to their size. Personal data can also be a part of Big Data but Big Data usually extends beyond that, containing aggregated and anonymous data."

De autoriteit gegevensbescherming van Luxemburg

"Big Data is all about collecting as much information as possible; storing it in ever larger databases; combining data that is collected for different purposes; and applying algorithms to find correlations and unexpected new information."

De autoriteit gegevensbescherming van Nederland

"Big Data is a broad term for processing of large amounts of different types of data, including personal data, acquired from multiple sources in various formats. Big Data revolves around predictive analytics – acquiring new knowledge from large data sets which requires new and more powerful processing applications."

De autoriteit gegevensbescherming van Slovenië

Big Data is "repurposing data; using algorithms to find correlations in datasets rather than constructing traditional queries; and bringing together data from a variety of sources, including structured and unstructured data."

De autoriteit gegevensbescherming van het Verenigd Koninkrijk

"the concept is used for situations where large amounts of data are gathered in order to be made available for different purposes, not always precisely determined in advance."

De autoriteit gegevensbescherming van Zweden

Volume

- Tiered storage/hub and spoke
- Selective data retention
- Statistical sampling
- Redundancy elimination
- Offload "cold" data
- Outsourcing

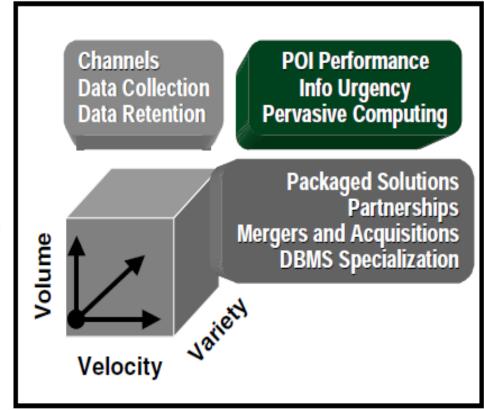
Velocity

- Operational data stores
- Data caches
- Point-to-point data routing
- Balance data latency with decision cycles

Variety

- Inconsistency resolution
- XML-based "universal" translation
- Application-aware EAI adapters
- Data access middleware and ETLM
- Distributed query management
- Metadata management

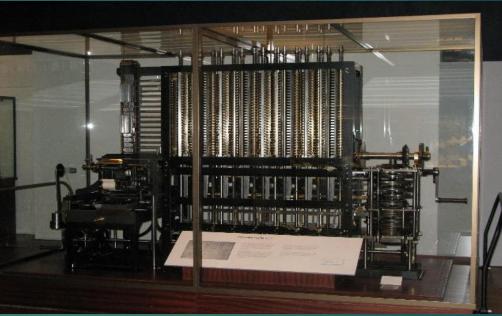
E-Business-Driven Information Explosion Factors

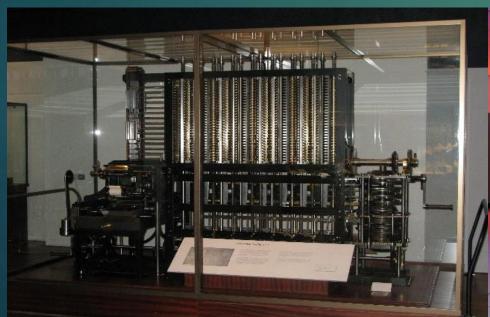


Umbrella term

- ▶ 'Open Data',
- 'Hergebruik',
- 'Internet der Dingen',
- 'Slimme Toepassingen',
- 'Profileren',
- ► 'Algoritmes'
- 'Cloud Computing'
- Machine learning
- Datafication
- Commodification
- Securitisation









Gradual factors

- ▶ Few data ------ Many data
- ▶ Few datasources ------ Many datasources
- ▶ Low speed ----- High speed

Three phases

- Gathering
- Analysing
- Using

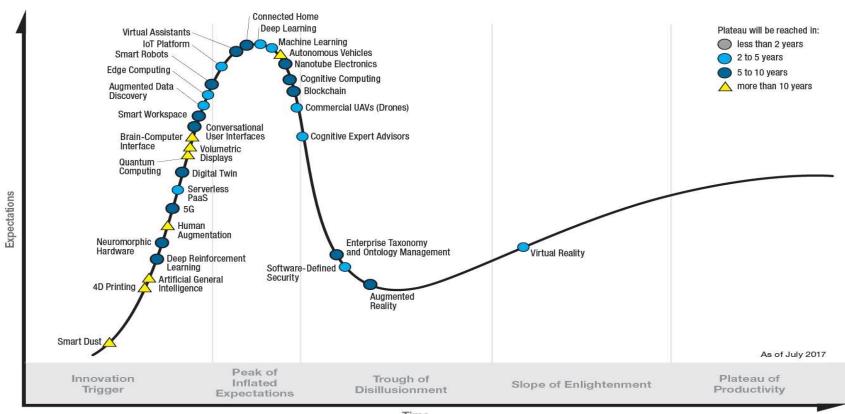
Ten catch phrases

- ▶ The more, the merrier:
- Quantity over quality:
- Dirty data is hot:
- Old data is new data:
- Gathering data is cheap, storing is cheaper, analysing is cheapest:
- Correlation over causation:
- Anything can say something about everything and everything can say something about anything:
- ► The future is now:
- No harm in trying:
- No mountain too high:

Risoco's

- Effectiviteit
 - ▶ Mass Surveillance
 - ▶ Predictive Policing: 'Uit analyse van de 42 projecten in de Investeringsagenda-projectportfolio blijkt dat 36% van de projecten de status 'vertraagd' of 'geannuleerd' heeft. Als we meer in detail kijken naar wijzigingen van planningsdata (ook van inmiddels afgeronde projecten), dan blijkt dat bijna 60% van de projecten vertraging heeft opgelopen of is geannuleerd.'
 - Reclames
 - ▶ Google Flu

Gartner Hype Cycle for Emerging Technologies, 2017



Time

gartner.com/SmarterWithGartner

Source: Gartner (July 2017) © 2017 Gartner, Inc. and/or its affiliates. All rights reserved.

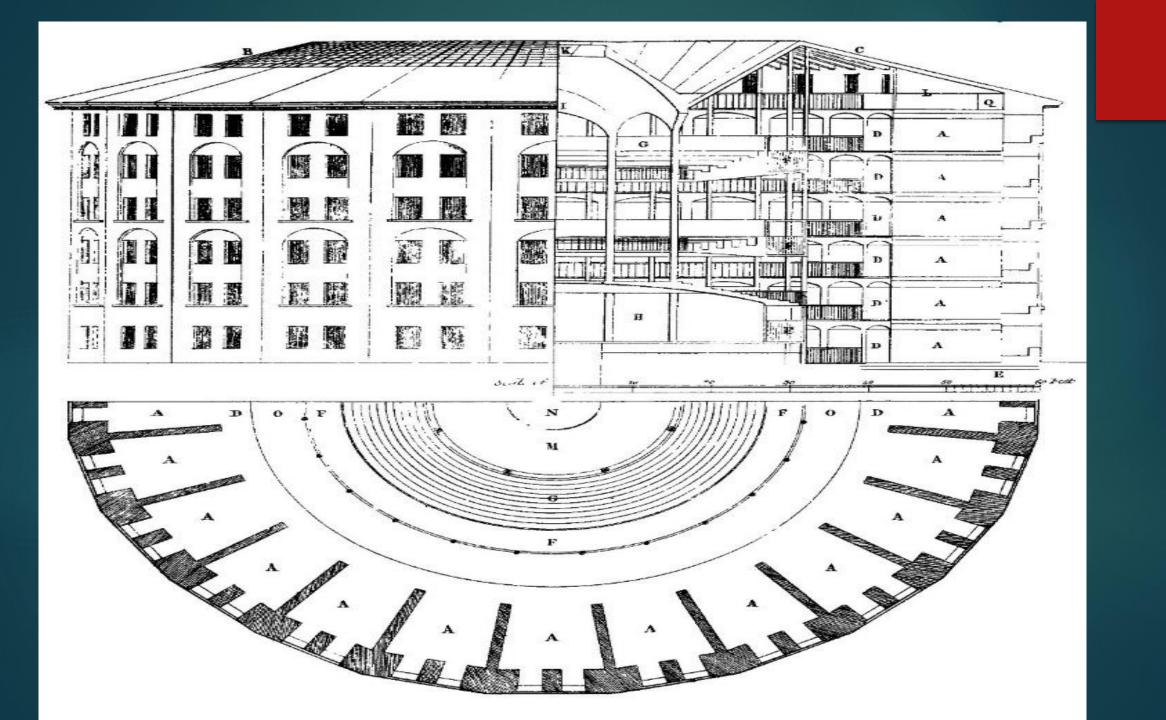


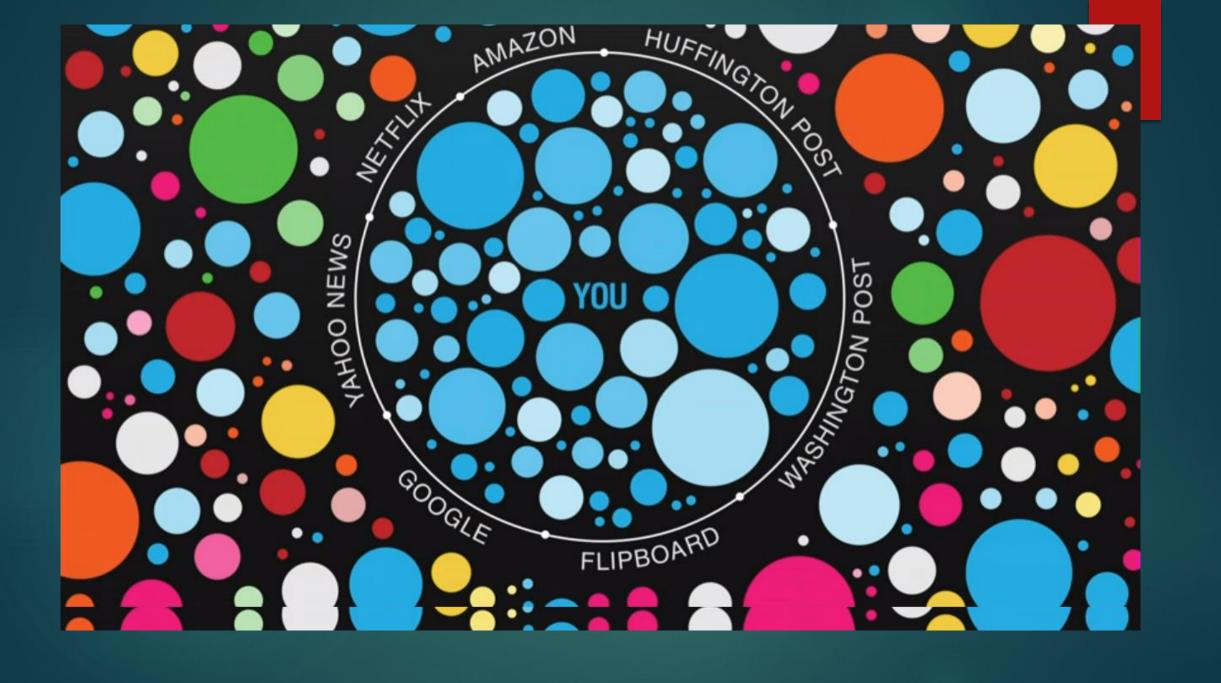
Gebrekkige analyse

- de representativiteit van de data
- Onderzoeksmethodologie
- Wijze van vragen stellen > open sources
- Algoritmes zelf gebiased
- updaten van datasets
- ▶ Bij veel databases ontbreken de zogenoemde meta-data
- Categoriseren van data is niet neutraal
- Dirty Data werkt niet
- Statische correlaties zijn geen causale verbanden
- Falsificatie ontbreekt

Voorspellingen

- ▶ Het ideaal van Borges
- Predictive
 - ▶ Valse positieven
 - ▶ Valse negatieve
- Verabsolutering van voorspellingen







FRANK PASQUALE

THE BLACK BOX Society

The Secret Algorithms That Control Money and Information

Social Dimensions of Privacy

Interdisciplinary Perspectives



Edited by

Beate Roessler and Dorota Mokrosinska

CAMBRIDGE





DATA LEAKAGE







Break

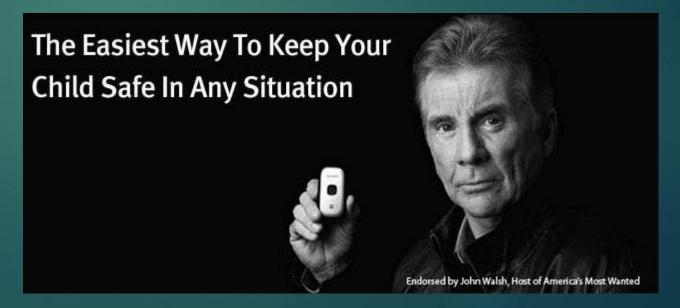


Privacy is dead, get over it

A mother suspects that her 15 year old son is smoking pot.

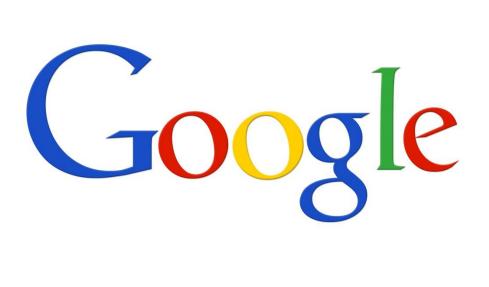
Should she search his room, read his mails and place a tracking device on his telephone so as to keep track of his

movements?



I have nothing to hide

Google is scanning the e-mails sent and received via Gmail. Can it use those data for personalised advertisements?



The GDPR

Overview

- ▶ (1) Background
- (2) Applicability
- ▶ (3) Backbone
- ▶ (4) Duties
- ► (5) Rights

Background

- Who
- ▶ What
- Where
- ▶ When
- Why

Applicability

- ▶ 1. Personal data
- ▶ 2. Are being processed
- ▶ 3. By a data controller
- ▶ 4. The EU has competence
- ▶ 5. No exception

Personal data

'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

Processing

- 'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;
- This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system

Data Controller

- ▶ (7) 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law;
- ▶ (8) 'processor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

EU competence

- ► Article 3
- ▶ Territorial scope
- ▶ 1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.
- 2. This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:
- (a) the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or
- (b) the monitoring of their behaviour as far as their behaviour takes place within the Union.
- ▶ 3. This Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law.

Exceptions

- 2. This Regulation does not apply to the processing of personal data:
- (a) in the course of an activity which falls outside the scope of Union law;
- ▶ (b) by the Member States when carrying out activities which fall within the scope of Chapter 2 of Title V of the TEU;
- (c) by a natural person in the course of a purely personal or household activity;
- (d) by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.

- ▶ 1. De reikwijdte van de verplichtingen en rechten als bedoeld in de artikelen 12 tot en met 22 en artikel 34, alsmede in artikel 5 kan, voor zover de bepalingen van die artikelen overeenstemmen met de rechten en verplichtingen als bedoeld in de artikelen 12 tot en met 20, worden beperkt door middel van Unierechtelijke of lidstaatrechtelijke bepalingen die op de verwerkingsverantwoordelijke of de verwerker van toepassing zijn, op voorwaarde dat die beperking de wezenlijke inhoud van de grondrechten en fundamentele vrijheden onverlet laat en in een democratische samenleving een noodzakelijke en evenredige maatregel is ter waarborging van:
- a) de nationale veiligheid;
- b) landsverdediging;
- c) de openbare veiligheid;
- d) de voorkoming, het onderzoek, de opsporing en de vervolging van strafbare feiten of de tenuitvoerlegging van straffen, met inbegrip van de bescherming tegen en de voorkoming van gevaren voor de openbare veiligheid;
- e) andere belangrijke doelstellingen van algemeen belang van de Unie of van een lidstaat, met name een belangrijk economisch of financieel belang van de Unie of van een lidstaat, met inbegrip van monetaire, budgettaire en fiscale aangelegenheden, volksgezondheid en sociale zekerheid;
- f) de bescherming van de onafhankelijkheid van de rechter en gerechtelijke procedures;
- g) de voorkoming, het onderzoek, de opsporing en de vervolging van schendingen van de beroepscodes voor gereglementeerde beroepen;
- h) een taak op het gebied van toezicht, inspectie of regelgeving die verband houdt, al is het incidenteel, met de uitoefening van het openbaar gezag in de in de punten a), tot en met e) en punt g) bedoelde gevallen;
- ▶ i) de bescherming van de betrokkene of van de rechten en vrijheden van anderen;
- j) de inning van civielrechtelijke vorderingen.

Speciale regime

- Article 89 Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes
- ▶ 1. Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner.
- 2. Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.
- ▶ 3. Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.
- ▶ 4. Where processing referred to in paragraphs 2 and 3 serves at the same time another purpose, the derogations shall apply only to processing for the purposes referred to in those paragraphs.

Speciale regime

- Artikel 44. Uitzonderingen inzake wetenschappelijk onderzoek en statistiek
- Indien een verwerking wordt verricht door instellingen of diensten voor wetenschappelijk onderzoek of statistiek, en de nodige voorzieningen zijn getroffen om te verzekeren dat de persoonsgegevens uitsluitend voor statistische of wetenschappelijke doeleinden kunnen worden gebruikt, kan de verwerkingsverantwoordelijke de artikelen 15, 16 en 18 van de verordening buiten toepassing laten.

Speciale regime

- VSNU Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek
- versie 0.9 Dit is de consultatieversie van de Gedragscode. Deze versie wordt bij verschillende relevante groepen getoetst.
- ▶ Deze versie zal nog veranderen en is nog niet vastgesteld door het bestuur van de VSNU.
- https://www.vsnu.nl/files/documenten/Domeinen/Governance/Consultatieversie%20-%20VSNU%20Gedragscode%20voor%20gebruik%20van%20persoonsgegevens%20in%20wetenschappelijk%20onderzoek.pdf

The backbone

- ▶ 1. Fundamental right
- ▶ 2. Fair Information Principles
- ➤ 3. Legitimate interest
- ▶ 4. Sensitive data
- ▶ 5. Transfer of data

Article 8 ECHR

- ARTICLE 8 Right to respect for private and family life
- 1. Everyone has the right to respect for his private and family life, his home and his correspondence.
- ▶ 2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Charter of Fundamental Rights of the EU

- Article 7 Respect for private and family life
- Everyone has the right to respect for his or her private and family life, home and communications.
- Article 8 Protection of personal data
- 1. Everyone has the right to the protection of personal data concerning him or her.
- ▶ 2. 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.
- ➤ 3. Compliance with these rules shall be subject to control by an independent authority.

Principles

- Necessity
- Proportionality
- Subsidiarity
- ▶ Effectiveness

Fair information principles

- Legitimate:
- Responsible:
- Fair:
- Purpose specification:
- Purpose limmitation:
- Data minimalisation:
- Correct:
- ▶ <u>Up-to-date</u>:
- Storage limmigation:
- ► <u>Technological security:</u>
- Organisational security:
- ▶ Transparancy:

Legitimate interest

- Article 6
- Lawfulness of processing
- ▶ 1. Processing shall be lawful only if and to the extent that at least one of the following applies:
- (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
- (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
- ▶ (c) processing is necessary for compliance with a legal obligation to which the controller is subject;
- (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
- (e) processing is necessary for the performance of a task carried out in the public interest or in the
 exercise of official authority vested in the controller;
- (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.
- ▶ Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks.

Consent

- ▶ Free
- ► Specific
- ▶ Informed
- Unambigious
- Proof
- Minors

Sensitive data

- Article 9
- Processing of special categories of personal data
- ▶ 1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

Bijzondere persoonsgegvens

- Vital interests data subject
- Curch, political party, union
- Empoyee's health
- Social security
- Explicit consent
- Explicitly made public
- Statistical or scientific research
- Legal procedure
- ▶ Public health
- Important public interest

Transfer of data

- Article 44
- General principle for transfers
- ▶ Any transfer of personal data which are undergoing processing or are intended for processing after transfer to a third country or to an international organisation shall take place only if, subject to the other provisions of this Regulation, the conditions laid down in this Chapter are complied with by the controller and processor, including for onward transfers of personal data from the third country or an international organisation to another third country or to another international organisation. All provisions in this Chapter shall be applied in order to ensure that the level of protection of natural persons guaranteed by this Regulation is not undermined.
- **4**

Transfer of data

- Article 45
- Transfers on the basis of an adequacy decision
- ▶ 1. A transfer of personal data to a third country or an international organisation may take place where the Commission has decided that the third country, a territory or one or more specified sectors within that third country, or the international organisation in question ensures an adequate level of protection. Such a transfer shall not require any specific authorisation.

Transfer of data

- Article 46
- Transfers subject to appropriate safeguards
- ▶ 1. In the absence of a decision pursuant to Article 45(3), a controller or processor may transfer personal data to a third country or an international organisation only if the controller or processor has provided appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available

- Article 49
- Derogations for specific situations
- ▶ 1. In the absence of an adequacy decision pursuant to Article 45(3), or of appropriate safeguards pursuant to Article 46, including binding corporate rules, a transfer or a set of transfers of personal data to a third country or an international organisation shall take place only on one of the following conditions:
- (a) the data subject has explicitly consented to the proposed transfer, after having been informed of the possible risks of such transfers for the data subject due to the absence of an adequacy decision and appropriate safeguards;
- ▶ (b) the transfer is necessary for the performance of a contract between the data subject and the controller or the implementation of pre-contractual measures taken at the data subject's request;
- (c) the transfer is necessary for the conclusion or performance of a contract concluded in the interest of the data subject between the controller and another natural or legal person;
- (d) the transfer is necessary for important reasons of public interest;
- (e) the transfer is necessary for the establishment, exercise or defence of legal claims;
- (f) the transfer is necessary in order to protect the vital interests of the data subject or of other persons, where the data subject is
 physically or legally incapable of giving consent;
- (g) the transfer is made from a register which according to Union or Member State law is intended to provide information to the public and which is open to consultation either by the public in general or by any person who can demonstrate a legitimate interest, but only to the extent that the conditions laid down by Union or Member State law for consultation are fulfilled in the particular case.
- Where a transfer could not be based on a provision in Article 45 or 46, including the provisions on binding corporate rules, and none of the derogations for a specific situation referred to in the first subparagraph of this paragraph is applicable, a transfer to a third country or an international organisation may take place only if the transfer is not repetitive, concerns only a limited number of data subjects, is necessary for the purposes of compelling legitimate interests pursued by the controller which are not overridden by the interests or rights and freedoms of the data subject, and the controller has assessed all the circumstances surrounding the data transfer and has on the basis of that assessment provided suitable safeguards with regard to the protection of personal data. The controller shall inform the supervisory authority of the transfer. The controller shall, in addition to providing the information referred to in Articles 13 and 14, inform the data subject of the transfer and on the compelling legitimate interests pursued.

Obligations

- ▶ 1. Register
- ▶ 2. Data Protection Officer
- ▶ 3. Data Protection Impact Assessement
- ▶ 4. Security
 - ► Technological security
 - Organisational secuirty
 - ▶ By Design
- ▶ 5. Transparancy
 - General transparancy
 - ▶ Information to subject
 - Report data breach

Rights data subject

- ► Right to information
- ▶ Right to copy
- ▶ Right to rectification
- ▶ Right to complete incomplete data
- ▶ Right to be forgotten
- ▶ Right to data portability
- ▶ Right to resctriction
- ▶ Right to object
- ▶ Prohibition of automatic decision making
- ▶ Right to file a complaint

Break



Tension with Big Data

- Purpose
- Purpose limmitation
- Data minimalisation
- Storage limmigation
- Security and confidentiality
- Data correct and up to date
- Transparancy
- Individual rights
- Personal data

Sewage monitoring



Botnets

