



Πανεπιστήμιο Αιγαίου

Ειδικά θέματα Δικαίου της Πληροφορίας

BIG DATA - Privacy and Data Protection

Λίλιαν Μήτρου (L.Mitrou@aegean.gr)

Αναπληρώτρια Καθηγήτρια

Τμήμα Μηχανικών Πληροφοριακών & Επικοινωνιακών
Συστημάτων



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





Big Data

- Enormous increase in access to and automated use of information
- Gigantic amounts of digital data controlled by companies, authorities and other large organisations which are subjected to extensive analysis based on the use of algorithms

...with enormous increase

- ✦ Amount of data on the global level is growing by 50 per cent annually
- ✦ 90% of the world's data has been generated within the past two years alone
- ✦ Mostly generated by consumers through interaction with Internet-based services

Uses of Big Data (1)

- ✦ The ability to store and analyse vast quantities of data will prove beneficial to society in many different ways
- ✦ Big Data is, to a certain extent, already used to analyse data in order to identify and predict trends and correlations
- ✦ Big Data may be used, for example, to predict the spread of epidemics, uncover serious side effects of medicines and combat pollution in large cities



Uses of Big Data (2)

- ✦ Big Data analyses do not involve necessarily use of personal data at all, such as analysis of weather data or sensor data from equipment on oil platforms
- ✦ Big Data may also be used in ways that may directly affect individuals
- ✦ It may be used to prepare profiles and predict the behaviour of individuals and groups of individuals by compiling and analysing personal data collected from many different sources

Big Data Collection

- ✦ Collection by mobile phone apps, smart grids, toll tag transponders in vehicles, patient records, location data, social websites, air traffic passenger data, public registers, customer loyalty programs, genome sequencing, sales history -Sensor technology
- ✦ Sometimes collected *automatically and unknowingly* in connection with the use of services (e.g., toll booth transaction data and location data, collection of airport Wi-Fi to track travellers)
- ✦ Personal data derived from various sets of information that appear to be anonymous, *added* from external sources to enrich the (previously collected) data, *shared* with external sources to enrich (personal) data from partner companies



Aggregation-Storage-Correlation

- ✦ Big Data as a tool available to both small and big enterprises, in all sectors of the economy
- ✦ Big Data in combination with Cloud Computing : switch from the traditional storage and processing of data using mainframe computers
- ✦ Merge data from various, different sources to generate profiles and use analysis tools to derive information otherwise not available

Privacy/ Data Protection Concerns/ Implications (1)

- ✦ Big Data involves almost by definition reuse of data
- ✦ Implication with purpose principle :collected data may not be used for purposes *incompatible* with the original purpose
- ✦ Big Data :value in itself based on its potential *future* uses in contrast to the principles of relevance and data minimisation



Privacy/ Data Protection Concerns/ Implications (2)

- ✦ Lack of transparency concerning uses and data controllers and difficulty to exercise access/ deletion rights
- ✦ Compilation of pieces of non sensitive information, which may generate new sensitive information
- ✦ Risk of re-identification of persons through compilation of data from several sources

Data Mining and Profiling

- ✦ Data mining :a process that has as its goal the transformation of raw data into information that can be utilised as strategic intelligence within the context of an organisation's identifiable goals
- ✦ Profiling:The process of 'discovering' correlations between data in databases that can be used to identify and represent a human or nonhuman subject (individual or group) and/or the application of profiles (sets of correlated data) to individuate and represent a subject or to identify a subject as a member of a group or category

Profiling/ a legal definition

- ✦ 'profile' refers to a set of data characterising a category of individuals that is intended to be applied to an individual;
- ✦ e. 'profiling' means an automated processing technique that consists of applying a 'profile' to an individual, particularly in order to take decisions concerning her or him or for analysing or predicting her or his personal preferences, behaviours or attitudes."

The stages of profiling

- Digitised **observations** regarding individuals' behaviour or characteristics are collected and stored on a large scale (data-warehousing). The resulting data may be nominative, coded or anonymous
- Data are analysed and "probed" (data-mining) permitting the determination of **correlations** between different behaviours/characteristics and other behaviours/characteristics
- On the basis of certain observable behavioural variables or characteristics specific to a, generally identified, individual, new past, present or future characteristics or behavioural variables are deduced

Anonymised or identified?

- ✦ The first two stages can be carried out using anonymised or coded data
- ✦ The third stage concerns an individual who is identified or identifiable and is carried in a growing variety of fields and by increasing numbers of actors
- ✦ Important to distinguish profiling techniques from other aids to decision-making

Uses/ Purposes of Profiling

- ✦ Criminalistic and criminological prognosis and techniques, as well as methods of operational case analysis
- ✦ Assessment of future or existing customers' risks (credit scoring) in the bank sector
- ✦ Medical research purposes and detectoon of congenital diseases
- ✦ Profiling techniques to identify taxpayers who are more likely than others to evade tax by fraudulent means
- ✦ In the commercial field, profiling can tailor the price of goods or a service according to a consumer

Implications of Profiling(1)

- ✦ The observation of the behaviour and characteristics of individuals through mining of large quantities
 - ✦ The determination of correlation between characteristics and patterns and the respective classification of data
 - ✦ The ability to predict individual attributes and behavior may infringe fundamental rights and raises major issues
- Lack of transparency in processing and of the data processed –

Implications of Profiling (2)

- ✦ Profiling ends up at treating a person as belonging to a specific category, which in turn indicates what “sort of person” someone is, the category becoming more important than the individual herself
- ✦ De-individualisation
- ✦ De-contextualisation
- ✦ Predictive data mining and profiling (e.g., flagging someone as potential insider threat, offender or inaccurate employee) results to classifications that may have considerable impact

Implications of Profiling (3)

- ✦ Binding application of other peoples' data
- ✦ Inaccuracy: problem of 'false positives' and 'false negatives'- Use of correlation analysis may lead to incorrect results for individuals
- ✦ Possibilities for abuse arise in particular when the profile can be linked to an identified individual : reputational damage, use for fraudulent purposes



Social implications

- ✦ Extensive use of Big Data analytics may increase the imbalance between large corporations and individuals
 - Information asymmetries which may lead to an imbalance in the playing field between government and citizens, and between businesses and consumers
- ✦ Risk of discrimination : Big Data analysis may consolidate existing prejudices and stereotypes and reinforce social exclusion and discrimination
- ✦ Stigmatisation of groups and groups members



Solutions and instruments

- ✚ Control over data and consent as main choice
- ✚ Robust and appropriate anonymisation in order to prevent identification irreversibly
- ✚ Privacy Impact Assessment in order to assess the challenges in terms of protection of privacy as early as possible, and in any case prior to the processing of Big Data
- ✚ Privacy by Design and accountability
- ✚ Enhancement of knowledge and awareness