Iceland
Liechtenstein
Norway grants

THESEUS

Connect the Disconnections -
from Disparate Data to Insightful Analysis

Interactive Course Material

University of Malta

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Iceland Liechtenstein Norway grants

THESEUS
Connect the Disconnections -
from Disparate Data to Insightful Analysis

Module 2 – What is Big Data? Where is it used?

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Sudden Surge in Popularity

- Extensively used across security, health care, astronomy, law, advertising, etc.

- Some uses:
  - Searching for terrorists
  - predicting food preferences
  - DNA research
  - Data from the Kepler Space Telescope

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Example

- WhatsApp: 2 billion users exchanging 65 billion messages daily

- Google Maps: 154 million users/month

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Definition of Big Data (NIST)

i. **Volume**, that is, the size of the dataset.

ii. **Variety**, which refers to the structured, unstructured, and semi-structured data that is gathered from multiple sources.

iii. **Velocity**, which refers to the speed at which data is generated, though it can also refer to the speed at which data is electronically processed.

iv. **Variability**, which refers to the changing rates inflow of data, such as the increase in data flow during peak times.

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Causes for the Surge in Use of Big Data

- Exponential increase in computing power (Moore’s Law)
- Very large amounts of devices generating data (In 2020, there were 2.86 billion smartphone users in the world)
- Many places to share data – social media and social networking sites
- Increase in the speed of data transfers – Wi-Fi, 4G, 5G
- Advances in data storage and data mining

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Actual Uses of Big Data

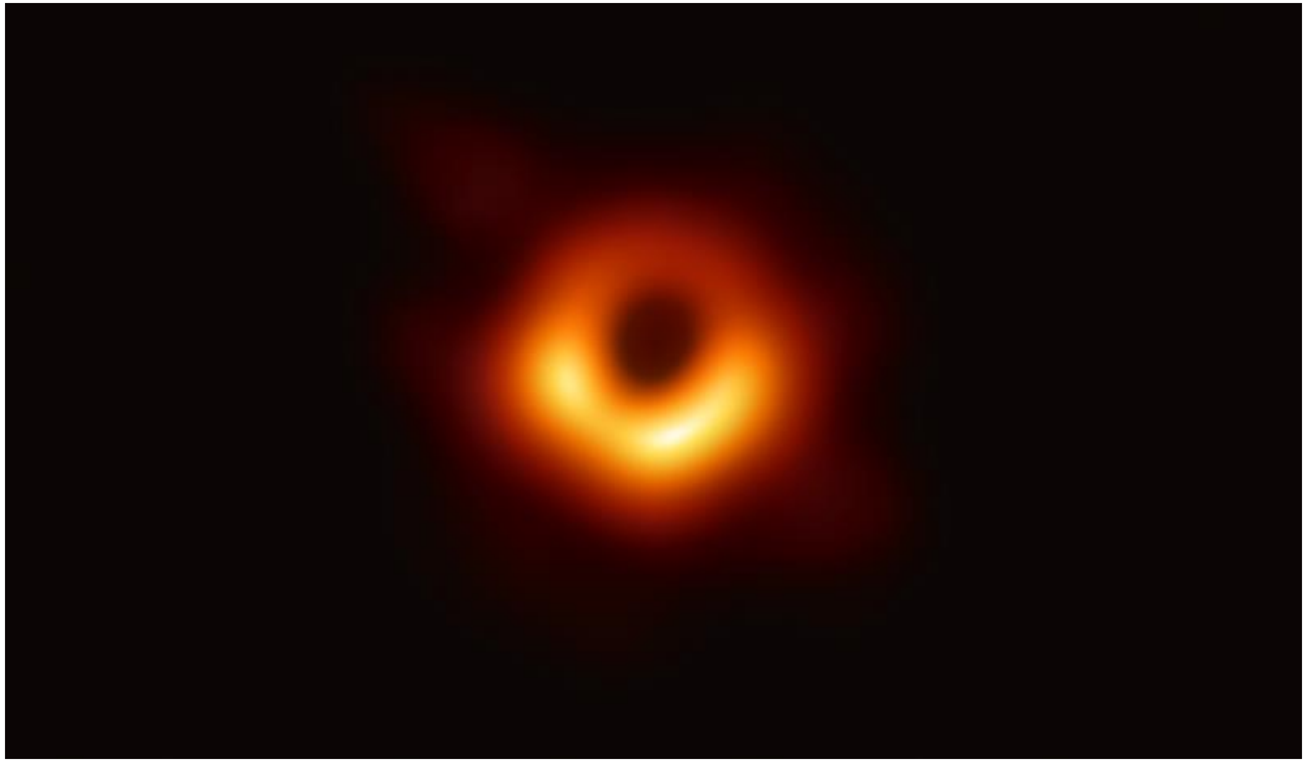
a. Data Mining

b. Predictive Modelling

c. Text Analytics

d. Understanding of human behaviour – correlating attitudes, intentions and actions

e. Science – astronomy, climatology, discovering cures through secondary use of patient data



Left: Image of a black hole based on data that took half a ton of hard drives to store.

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Module 2 – Technical, Legal and Ethical Challenges

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Technical Challenges

- Data storage: even with CCTV cameras, up to 2500 petabytes of data is expected to be generated
- Data transfers: with such large amounts of data, transferring it to different places is a challenge even with Wi-FI/4G/5G
- Speed of computing infrastructure not keeping pace with the growth of data generation
- Databases and Data warehouses are becoming inadequate

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Security Challenges

- Misuse of Data:
  - Human error
  - Whistleblowing
  - Cyberattacks

- Data Processing and Interpretation: How do you identify the signal within all the noise?

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Legal Challenges

- Big Data challenges the basic legal principles of information privacy laws
  - Collection Limitation principle
  - Knowledge and Consent of the Individual
  - Purpose Limitation
  - Use Limitation
  - Data Minimization

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Ethical Challenges

- Brings images of Orwell’s 1984 to mind
  - Constant surveillance (CCTV, social media, etc.)
- Individuals leave large digital footprints that are combined to reveal important information about them: a digital DNA
- Snowden revelations: collection of data on millions of people by governments
- Anonymization is not perfect: data can be re-identified

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Final Thoughts

- Correlation Does Not Imply Causation
- Ask the Right Questions
- Garbage In – Garbage Out
- Analysis generated from big data is not necessarily complete, accurate or true

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Module 14 – Legal and Regulatory Implications of Big Data

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Big Data a Unique Phenomenon

- Requires ALL of the below that were not present before:
  - Combination of large sets of data
  - Sets of instructions for analysing data (algorithms)
  - Computing power

- Each existed before, but never together:
  - Datasets: for example, administrative datasets in ancient China
  - Algorithms for parsing data have also existed for decades if not centuries
  - Computing power has only been available for the last few decades, and not at the scale available now

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Privacy and Data Protection

- Big data undermines principles of privacy and data protection laws:
  - Lawfulness, fairness and transparency
  - Purpose limitation
  - Data minimization
  - Accuracy
  - Storage Limitation
  - Integrity and Confidentiality
- Creates challenges beyond the scope of current privacy legislation
  - Just because personal data is not involved does not mean challenges to privacy do not exist

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Transparency and Accountability

- Article 8 of the EU Charter:

> "Everyone has the right of access to data which has been collected concerning him or her."

- Big Data poses unique challenges to transparency:
    - Algorithmic decision making is too opaque/complex to understand
    - Shrouded in physical, technical and legal secrecy
    - Dynamic process driven by massive volumes of data
- Without transparency, identifying harms is difficult, so correcting mistakes and holding authorities accountable also becomes difficult

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Discrimination and Bias

- Most pressing adverse consequences of big data analytics and algorithmic decision-making
- Example: United States' Correctional Offender Management Profiling for Alternative Sanctions (COMPAS):
  - Used in courts to assess likelihood of recidivism
  - Found to have racial bias and have low accuracy

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Chilling Effect on Freedom of Expression and Association

- If people are uncertain about potential future uses of information shared online, they tend to change their behaviour
- Studies following Snowden revelations show that justification of surveillance practices create a chilling effect on democratic discourse, stifles minority political views
- Reduces autonomy and freedom of association, as well as freedom of movement
- People may alter online activity to protect their privacy

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Module 14 – Larger Issues in Complying with Big Data Legal Frameworks

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Contextuality of Legal Frameworks

- No specific laws and regulations in place for big data
- Depends on laws for:
  - Fundamental rights and freedoms
  - Privacy and Data Protection Laws
  - National Laws for intelligence and law enforcement agencies dealing with national security, criminal prosecution, public interest
  - International conventions such as Europe’s Convention 108+, European Charter of Fundamental Rights

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Types of Data

- **Personal Data:** Subject to a variety of regulations, starting from the manner in which it is collected, how it is stored, the legal bases that can be used for processing such data, the amount of time it can be retained, transparency and accountability for those responsible for storing and processing it, and others.
- **De-identified Data:** Just because data appears pseudonymous or anonymous at face value does not mean that one is free to process or analyse it however one wishes. Using analytical techniques to re-identify specific individuals from large datasets may result in the data once again being subject to different regulatory frameworks.
- **Sensitive Data:** Associated with the most stringent forms of regulation. Even with explicit consent or for substantial public interest, safeguards protecting subject’s rights, freedoms and legitimate interests must be in place.
- **Non-Personal Data:** Least controversial from a legal point of view and can, save for certain intellectual property rights, be generally deemed completely free from any legal implications.

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Types of Databases

- **Open-Source Databases:** “public material that is not explicitly published but is still publicly or commercially available, such as commercial satellite imagery”; “material that can be lawfully obtained through request, purchase, or observation by a member of the public”. Subject to the law based on type of data.
- **Open Data and Government:** data that can be freely used, reused and redistributed by anyone – subject only, at most, to the requirements to attribute and share alike. Subject to the law based on type of data.
- **Data Collected for Commercial Use:** private and commercial entities must comply with every data protection principle. If made publicly available for purchase and is then acquired by intelligence or law enforcement agencies, specific national laws may be in place.
- **Data Collected by Intelligence and Law Enforcement Agencies:** If created through deception, covert or misleading tactics, or ‘backdoor’ access, they are typically subject to national laws and internal oversight mechanisms such as independent ex-ante authorisation/approval of data collection, the presence of a separate intermediary body responsible for first filter/selection process, central management of data access, automated controls, etc.

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).

Intelligence Function and Legal Frameworks

- Intelligence Function is separate from Intelligence Agencies
- Could be carried out by law enforcement agencies, companies, journalists, etc.
- Legal framework is highly contextual:
  - intelligence and law enforcement agencies, as competent authorities, are not regulated by the GDPR
  - Law enforcement is regulated by the Police Directive
  - Same for Convention 108+, which allows national legal frameworks
  - Such national laws cannot be general and indiscriminate in this regard and must be in keeping with the European Charter of Fundamental Rights (respect for private and family life, right to protection of personal data and freedom of expression and information)

This publication was realised with the EEA Financial Mechanism 2014-2021 financial support. Its content (text, photos, videos) does not reflect the official opinion of the Programme Operator, the National Contact Point and the Financial Mechanism Office. Responsibility for the information and views expressed therein lies entirely with the author(s).