Data Stewards and Data Management Principles  
(LTAT.02.014)

Priit Adler  
Research Fellow, University of Tartu  
Elixir EE, BIIT

Lecture 13
Recap

https://zenodo.org/record/4623713#YlebC30zaL4
Data Literacy

the ability to read, understand, create, and communicate data as information.
Data Practices Courseware

1.5 Analyze and Report - https://datapractices.org/courseware/1_5.html
Big data

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time.
Transactions + Interactions + Observations = BIG DATA

At the same time ...

https://www.slideshare.net/welkaim/introduction-to-big-data-65870623
The Vs of Big Data

https://www.slideshare.net/welkaim/introduction-to-big-data-65870623
The Vs of Big Data

The New York Stock Exchange captures 1 TB of trade information during each trading session.

Modern cars have close to 100 sensors that monitor items such as fuel level and tire pressure.

By 2016, it is projected there will be 18.9 billion network connections – almost 2.5 connections per person on earth.

https://www.slideshare.net/welkaim/introduction-to-big-data-65870623
The Vs of Big Data

As of 2011, the global size of data in healthcare was estimated to be 150 EXABYTES (161 BILLION GIGABYTES).

By 2014, it’s anticipated there will be 420 MILLION WEARABLE, WIRELESS HEALTH MONITORS.

4 BILLION+ HOURS OF VIDEO are watched on YouTube each month.

30 BILLION PIECES OF CONTENT are shared on Facebook every month.

400 MILLION TWEETS are sent per day by about 200 million monthly active users.

https://www.slideshare.net/welkaim/introduction-to-big-data-65870623
The Vs of Big Data

https://www.slideshare.net/welkaim/introduction-to-big-data-65870623
The Vs of Big Data

The ‘Datafication’ of our World:
- Activities
- Conversations
- Words
- Voice
- Social Media
- Browser logs
- Photos
- Videos
- Sensors
- Etc.

Volume
Velocity
Variety
Veracity

Analysis

Analysing Big Data:
- Text analytics
- Sentiment analysis
- Face recognition
- Voice analytics
- Movement analytics
- Etc.

Visualization

Value

https://www.slideshare.net/welkaim/introduction-to-big-data-65870623
Big Data Management (6 ECTS) LTAT.02.003

The objective of this course is to introduce students to the principles and methods of advanced data management and processing. The course will cover the techniques of storing and processing different types of data (structured, semi-structured and unstructured). It will cover the state-of-the-art in different types of big data processing systems (e.g., stream processing, graph data processing, scalable machine and deep learning systems).
Distributed Systems
(6 ECTS) LTAT.06.007

The student will get the main idea of construction and basic principles of work of distributed systems. After passing the course the student is able to create the distributed systems, analyse and improve the existing systems.
Homework 9

Compile detailed DMP

• OP1: Continue with the project from HW7 / HW8
  Improve the plan to reduce bad practices. Write the plan, that would have saved time and effort for the researchers.

• OP2: Write one for one of your own projects
  Ideally for running project and something you can use from now on

• Deadline 31st of May