



Introduction

5ARB0: DATA ACQUISITION & ANALYSIS (2022 – 2023)

Uzay Kaymak, Jheronimus Academy of Data Science, u.kaymak@tue.nl

Mastertrack: Artificial Intelligence & Engineering Systems

Outline

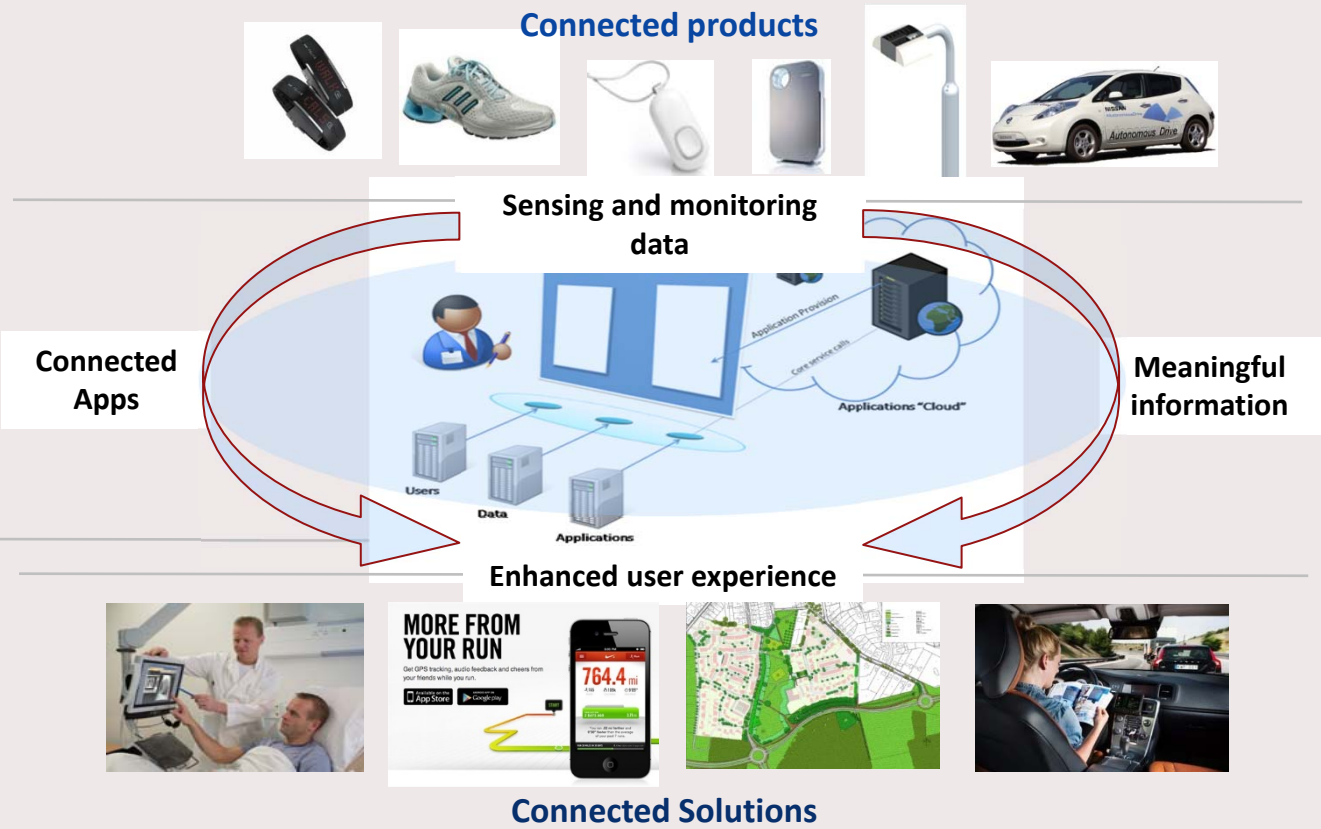
Motivation

Course organization

Artificial intelligence & engineering systems

Data collection preliminaries

The Always-On Society



Everywhere Analytics

From Deloitte



AI systems: self-driving cars



Maker: metamorworks | Bronvermelding: Getty Images/iStockphoto



<https://getfello.com/wp-content/uploads/2017/03/Google-self-driving-car-prototype-front-three-quarters1.jpg>

AI Systems: IBM Watson



[IBM Watson Health focused on value-based care, physician burnout, personalized medicine | Healthcare IT News](#)



[IBM Watson moet helpen bij uitlezen röntgenfoto's - ICT&health \(icthealth.nl\)](#)

AI Systems: robot football



8

TU/e

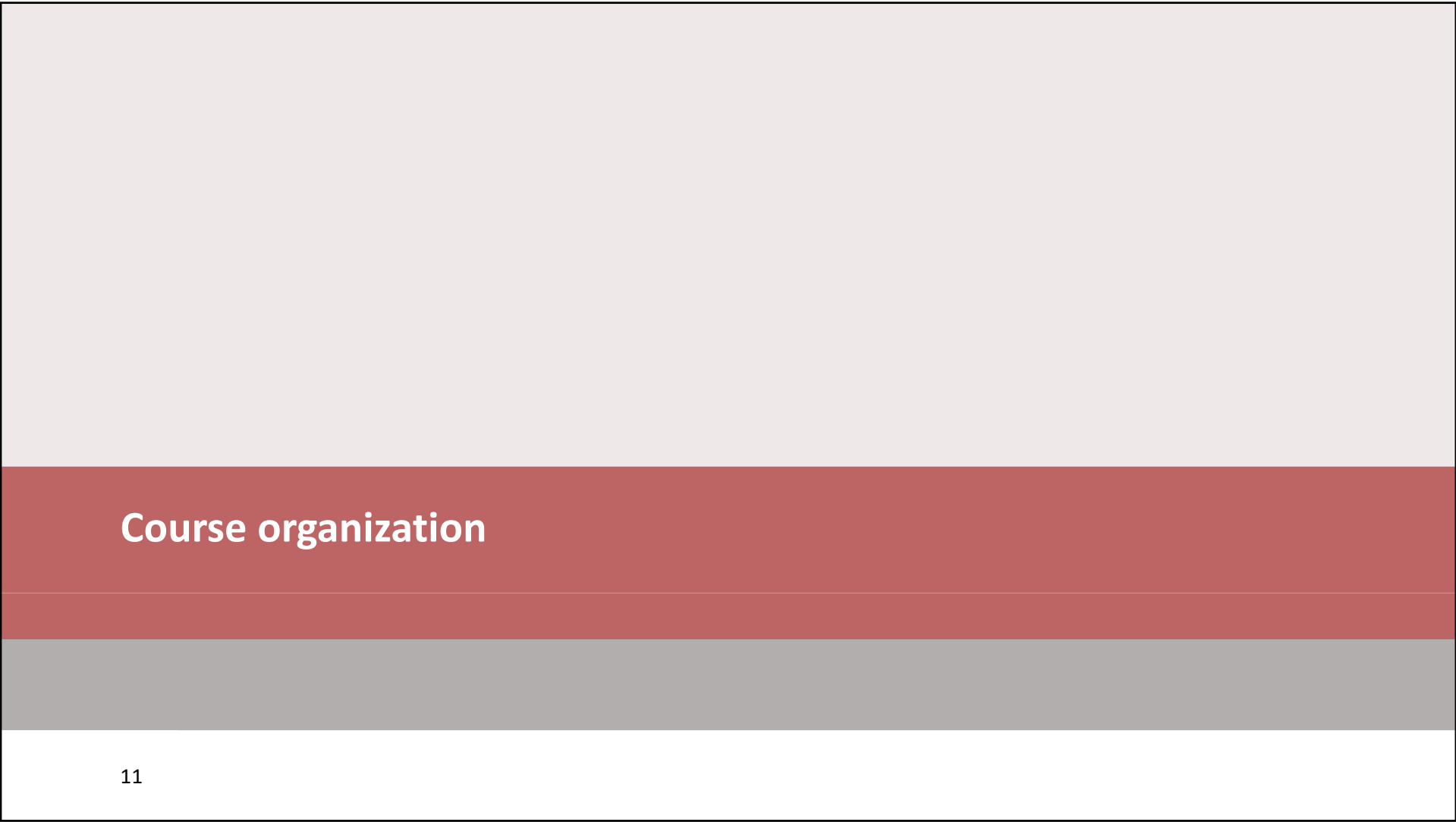
Overall message

Although there are many and varied uses of data in AI applications,

data needs effort to

- **collect,**
- **process,**
- **store, and**
- **analyze.**

Data does not come cheap!




11

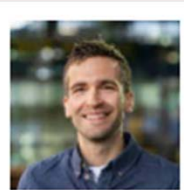
Learning objectives

- reflect on non-technical aspects of data acquisition, including privacy, safety, ethical and multi-disciplinary aspects;
- design a basic data acquisition system based on FAIR principles;
- work with basic requirements of data curation;
- apply experiment design and system excitation principles for goal-targeted data collection;
- assess quality of data and handle missing data;
- analyze a data set by implementing and using basic analysis techniques, e.g. principal components analysis, clustering;
- present the results of data analysis verbally and in writing.


Lecturers



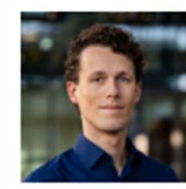
Prof.dr.ir. U. Kaymak
Course Coordinator; Lecturer
u.kaymak@tue.nl




Dr. W. Kouw
Lecturer
w.m.kouw@tue.nl



Dr. M. van Gilst
Lecturer
m.m.v.gilst@tue.nl



Ir. B. van Erp
Lecturer
b.v.erp@tue.nl



MSc. S. Hommerson
Lecturer
s.m.hommerson@tue.nl

& a number of guest lecturers from
research institutes and the industry

Teaching assistants

- D. Ferreira de Carvalho, MSc.
- G. M. Grońska, MSc.
- P.-E. Simon, MSc.

Meetings

16 sessions

(1 x 2 hrs./week lecture, 1 x 2 hrs./week instruction, 8 weeks long)

Tuesday 1+2, [IPO 0.98](#)

Friday 5+6, [Lune 1.050](#)

Lectures: introduce and explain main concepts

Instructions for practice exercises and working on the assignment

Q&A session at the end of the quartile

Further questions can be asked through Canvas (preferred method) or during instructions

Planning – 1 (tentative)

Date	Time	Type	Room	Content
Tue Sep 6	08:45 – 10:45	Lecture	IPO 0.98	Introduction Data Collection I
Fri Sep 9	13:30 – 15:30	Instruction	Luna 1.050	Working in teams (p.m.) Software installation Intro assignment 1 (p.m.)
Tue Sep 13	08:45 – 10:45	Lecture	IPO 0.98	Data Collection II
Fri Sep 16	13:30 – 15:30	Instruction	Luna 1.050	Working in teams (p.m.) Intro assignment 1 (p.m.) Work on Assignment 1
Tue Sep 20	08:45 – 10:45	Lecture	IPO 0.98	Data Analysis I
Fri Sep 23	13:30 – 15:30	Instruction	Luna 1.050	Work on Assignment 1
Tue Sep 27	08:45 – 10:45	Lecture	IPO 0.98	Data Analysis II
Thu Sep 29	20:00			Deadline Assignment 1

Planning – 2 (tentative)

Tue Oct 4	08:45 – 10:45	Instruction	IPO 0.98	Intro Assignment 2
Fri Oct 7	13:30 – 15:30	Lecture	Luna 1.050	Protocols & Privacy
Tue Oct 11	08:45 – 10:45	Lecture	IPO 0.98	Data Analysis III
Fri Oct 14	13:30 – 15:30	Instruction	Luna 1.050	Work on Assignment 2
Tue Oct 18	08:45 – 10:45	Lecture	IPO 0.98	Data Quality
Fri Oct 21	13:30 – 15:30	Instruction	Luna 1.050	Work on Assignment 2
Tue Oct 25	08:45 – 10:45	Lecture	IPO 0.98	Data and model curation Overarching case
Fri Oct 28	13:30 – 15:30	Instruction	Luna 1.050	Review & Q&A
Sun Oct 30	20:00			Deadline Assignment 2
Wed Nov 9	13:30 – 16:30	Exam	TBA	
Wed Feb 1	18:00 – 21:00	Resit	TBA	

Course material (literature and tools)

- Material provided by lecturers (reader "Data Acquisition and Analysis"), e.g. slides, handouts, etc.
- Scientific papers
- Industry white papers
- Self-study (video) tutorials (recommended)
- Python
- Momotor

Assessment

Components:

Assignment 1 – 30%

- Deadline: 29/09/2022

It is not possible to re-sit assignments

Assignment 2 – 40%

Assignments are valid only in the current academic year

- Deadline: 30/10/2022

Written exam – 20%

Peer review – 10%

Assignments will be made in groups of 3.

Register through Canvas as soon as possible

Exam

Type: written, closed book

Date: 9 November 2022, 13.30 – 16.30

Re-sit: 1 February 2023, 18.00 – 21.00

Position in curriculum

Topic: data science

Course name: Data acquisition and analysis

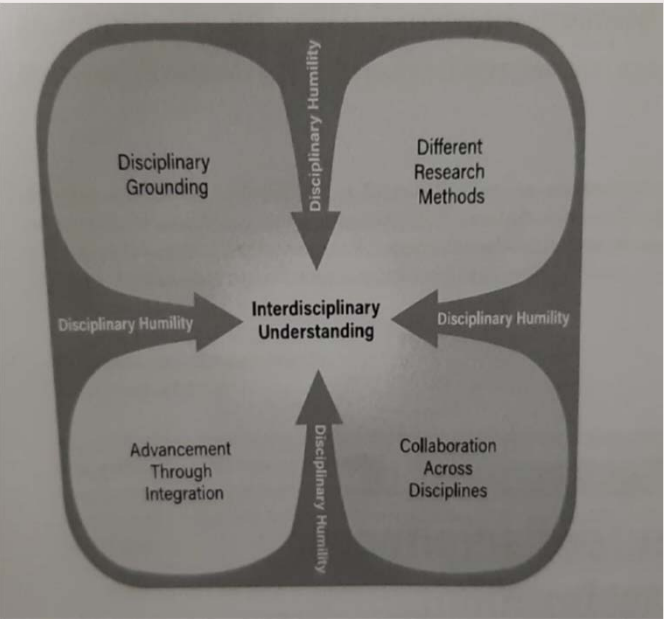
Credits: 5 ECTS

Core course in Master program Artificial Intelligence and Engineering Systems

- **Year 1, Q1**

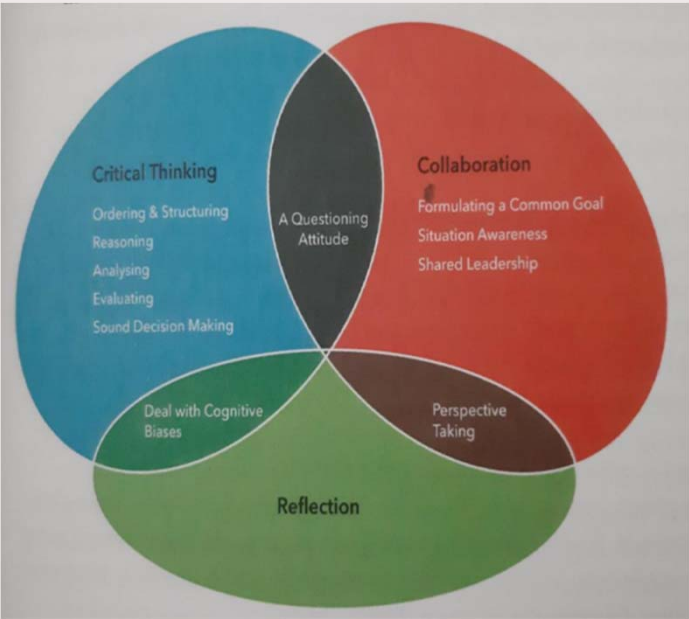
Input towards multiple courses (core and elective)

Interdisciplinary integration

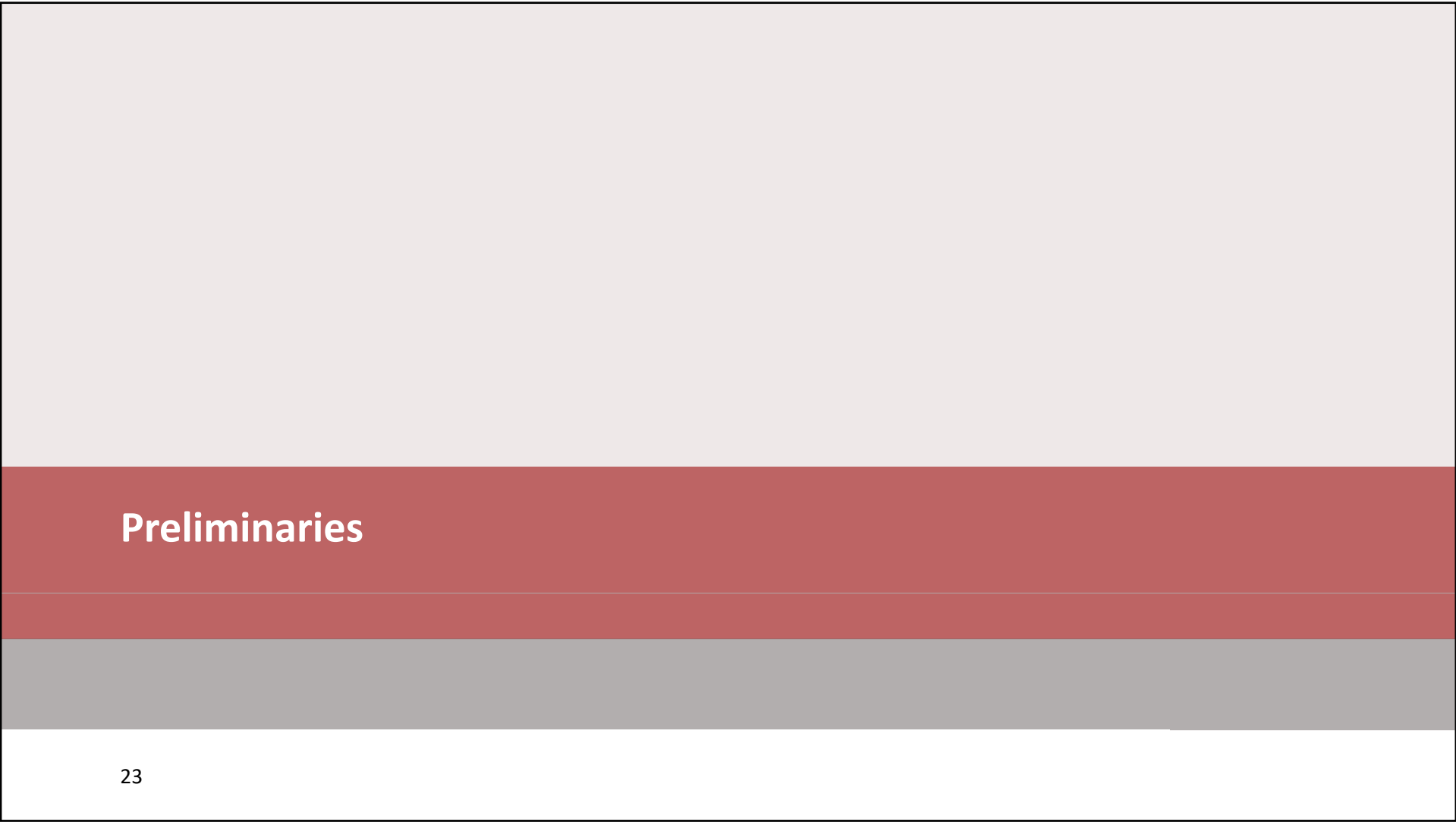


Tipp & Shortlidge (2019) Interdisciplinary Science Framework (IDSF)

Competencies



(Kirschner & Van Merriënboer (2013)



23

Two MSc programs

Data Science and Artificial Intelligence

Artificial Intelligence and Engineering Systems

24

TU/e

Types of data collection

Primary data collection – data collected for the first time, for a particular purpose

Secondary data collection – re-use of data that has already been collected for another purpose

Knowledge pyramid

Data

Information

Knowledge

Wisdom

Data collection influenced by:

- task
- perspective
- goal

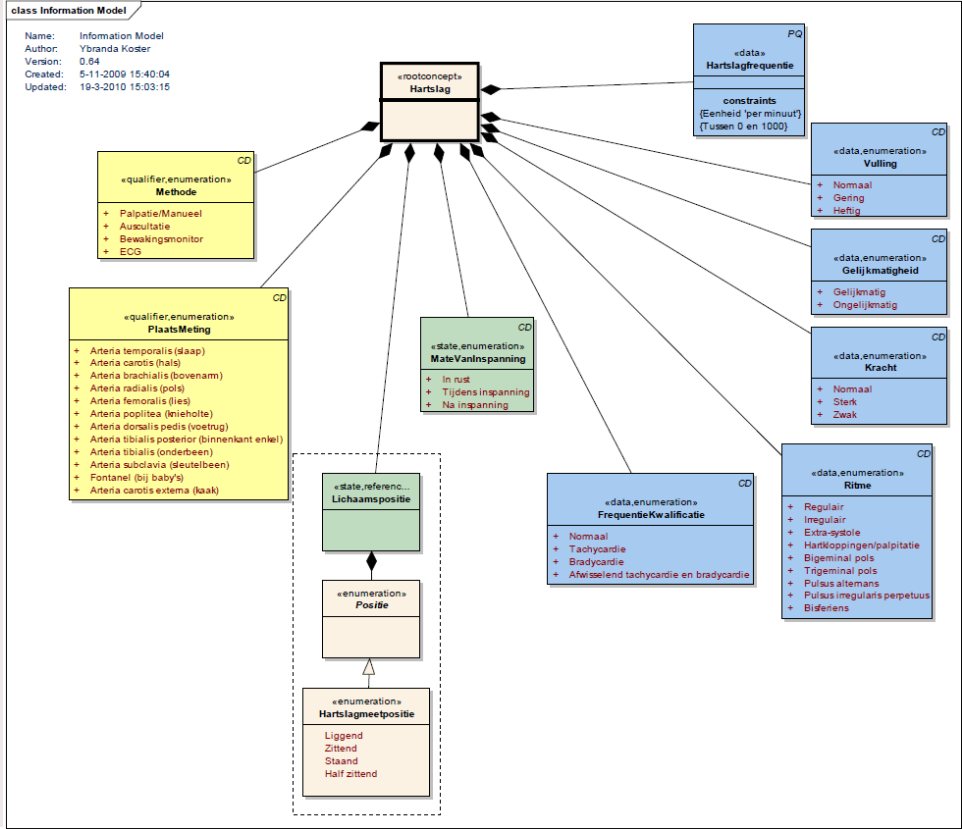
Based on Ackoff, R. L., "From Data to Wisdom", Journal of Applied Systems Analysis, Volume 16, 1989 p 3-9.

Data

- What is a datum?
- When is something data?
- What do you need to make sense from the data?
- Components to record?

→ Data models

Example: data model for measuring heart rate



- Information elements:
- Heart rate frequency
 - Heart condition qualification
 - Heart rhythm
 - Method of measurement
 - Place of measurement
 - Body posture
 - ...

Types of data

Narrative data (text)

Name, address, description symptoms

Numerical data

Weight, temperature

Analog data (recorded signals)

Accelerometer data

Sequence data

Time series

Images

Photos and drawings

Video recordings

Audio recordings

Etc.

Categories of data

- **Attribute – value pairs**
- **Unstructured data**
- **Sequence data**
- **Graph data**

30

TU/e

Aspects of data collection



Illustration: iStock



Royal Geographical Society

Recap

- **Data has multiple facets**
- **Distinction between primary data collection and secondary data collection**
 - In engineering systems, primary data collection has focus
 - Data science deals also with secondary data collection
- **Six aspects to data collection**

What	When
Why	Where
Who	How