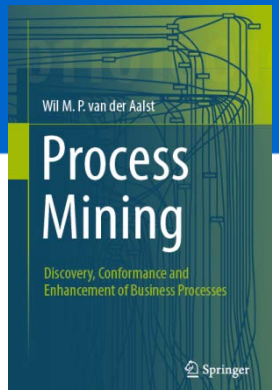
Process Mining: Data Science in Action

Data Science and Big Data

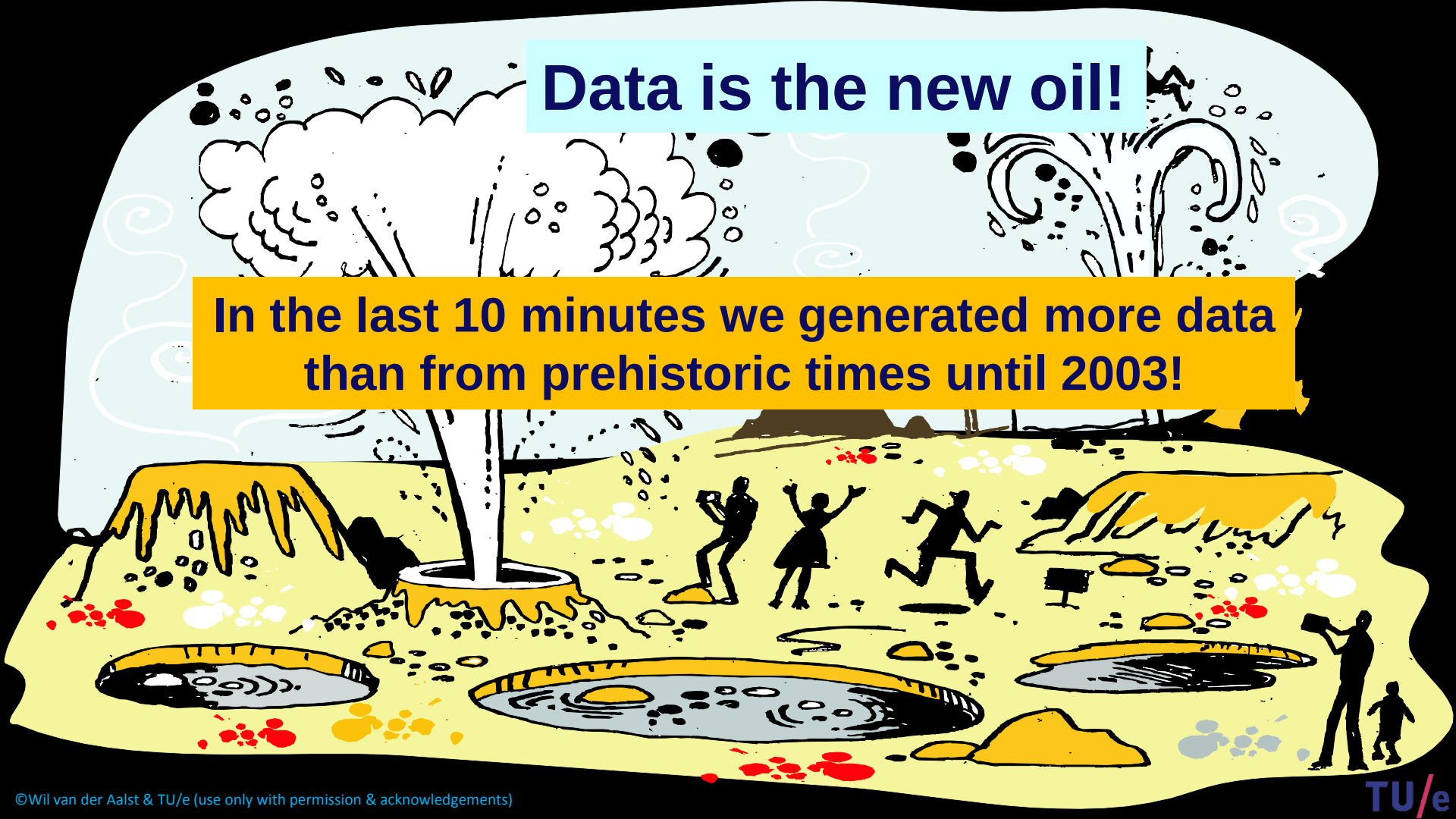


prof.dr.ir. Wil van der Aalst
www.processmining.org

TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

Data is the new oil!



TU/e

We are all generating event data!

taking the train

getting a speeding ticket

sending an e-mail

refueling your car

buying a coffee

making an appointment

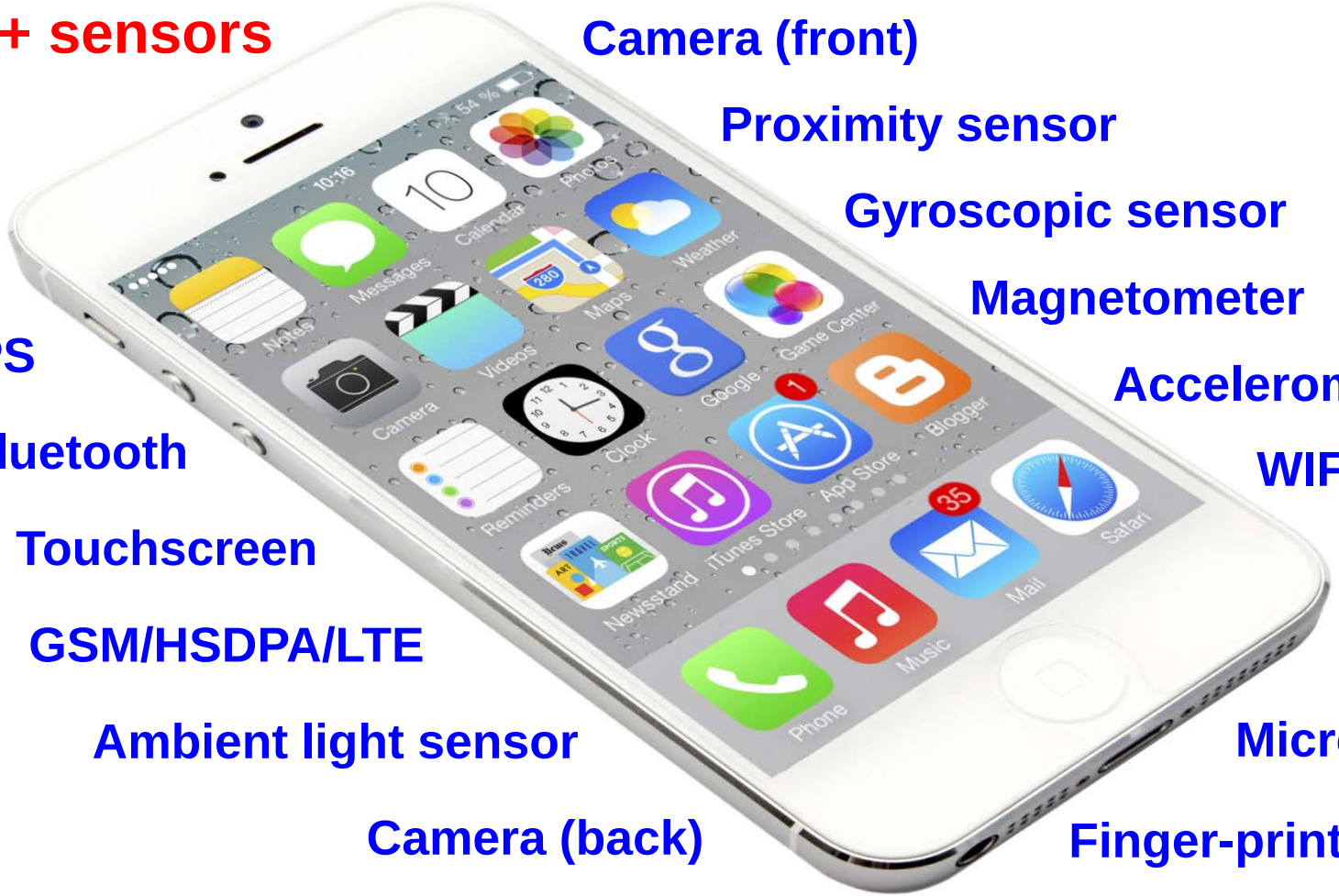
making a phone call

watching this lecture

adjusting the temperature

TU/e

14+ sensors



Camera (front)

Proximity sensor

Gyroscopic sensor

Magnetometer

Accelerometer

WIFI

Microphone

Finger-print scanner

Camera (back)

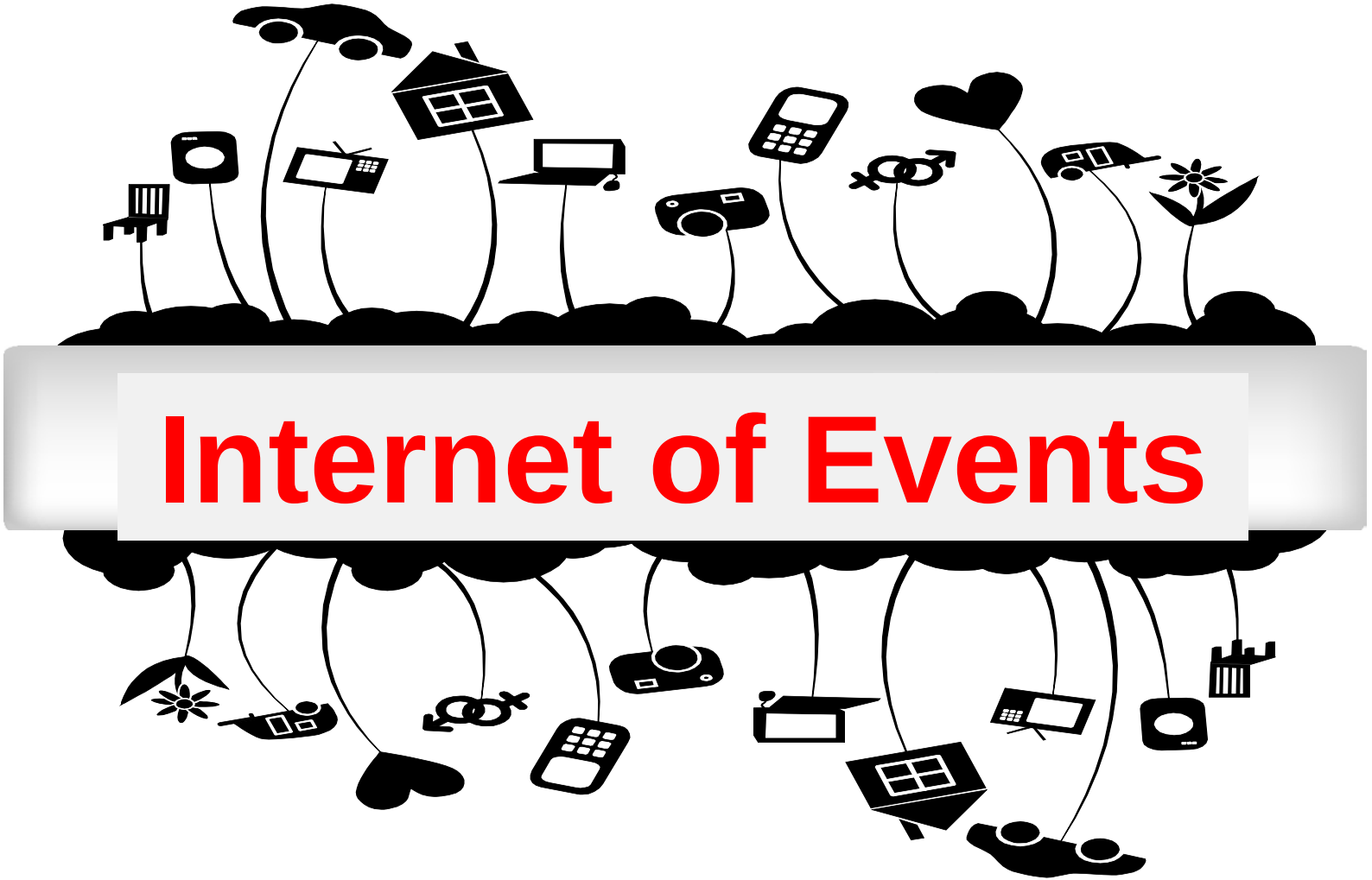
Ambient light sensor

GSM/HSDPA/LTE

Touchscreen

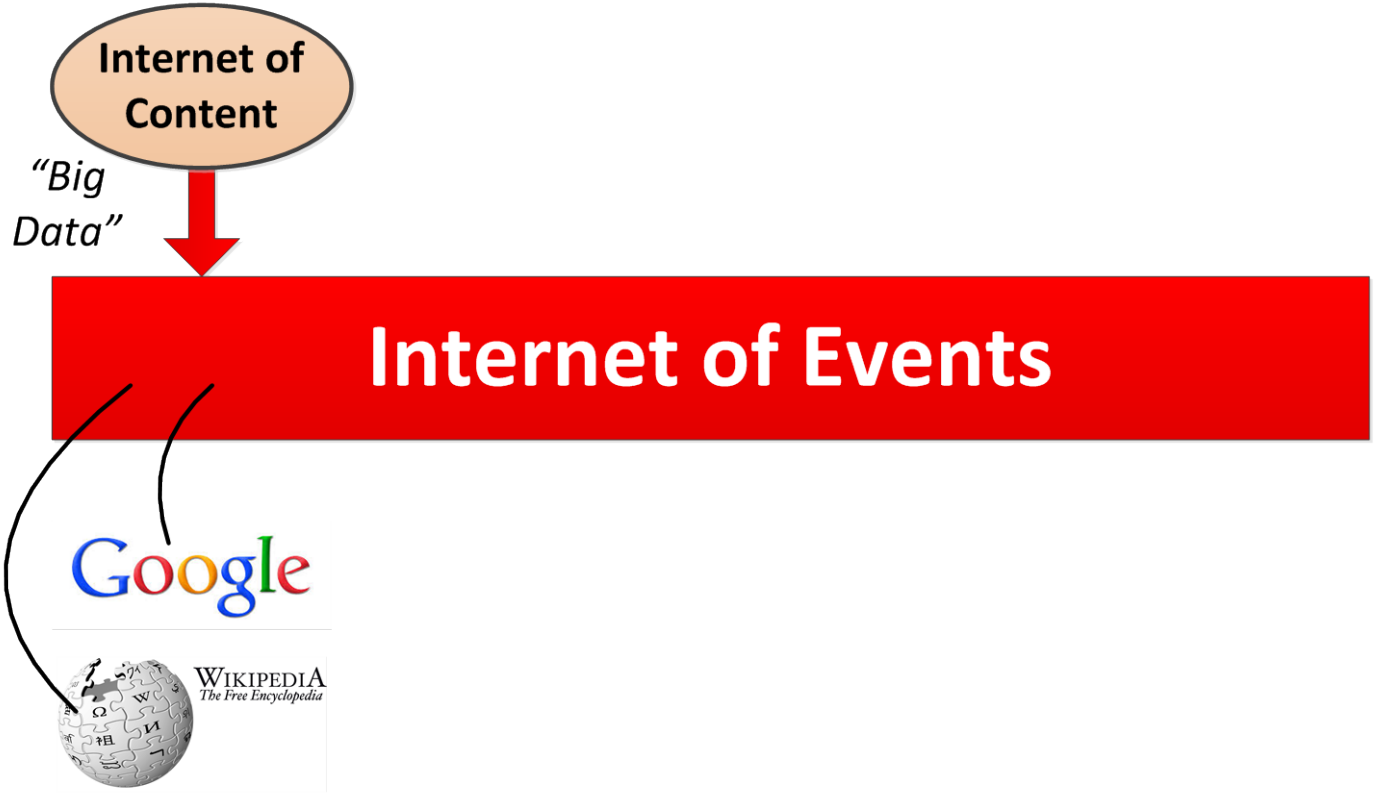
Bluetooth

GPS


Internet of Events

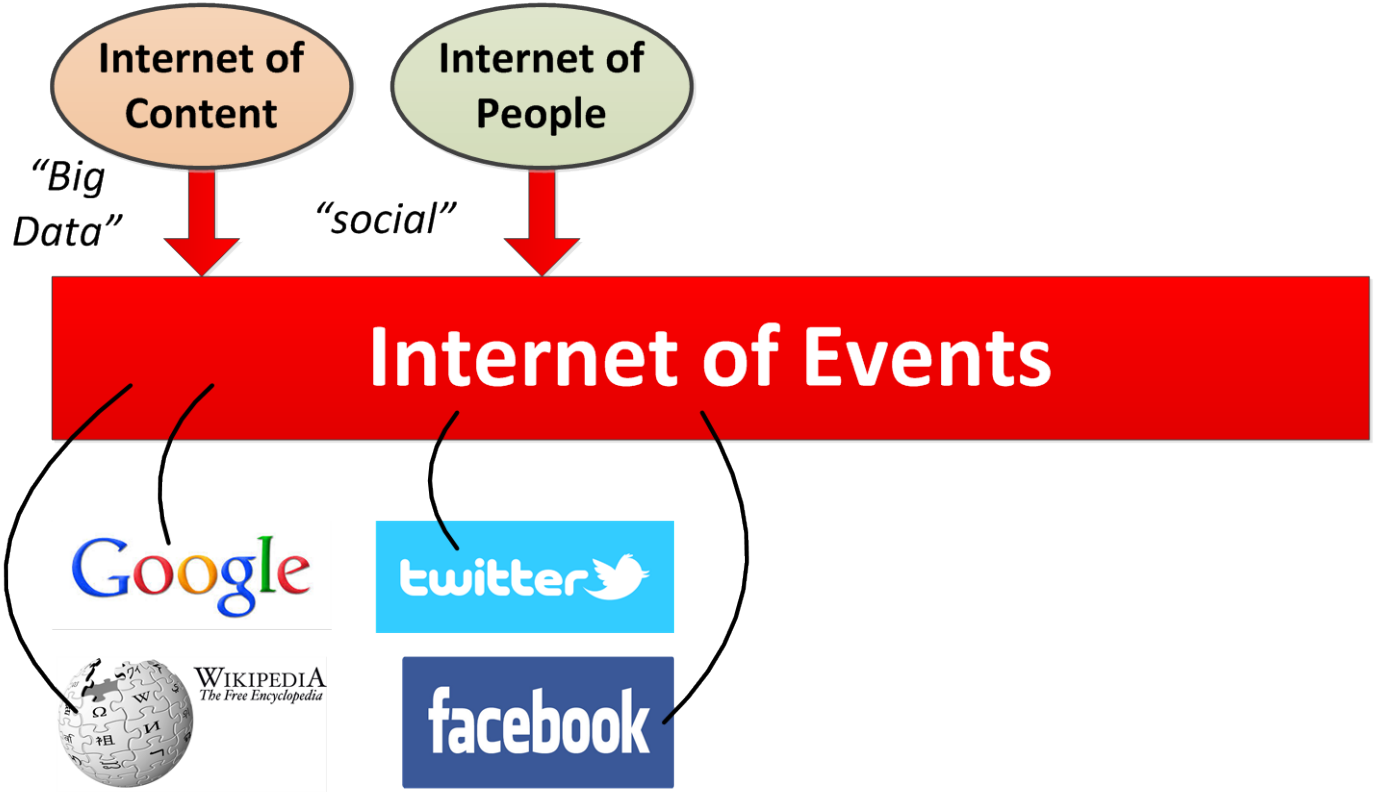
Internet of Events: 4 sources of event data

Internet of Events: 4 sources of event data



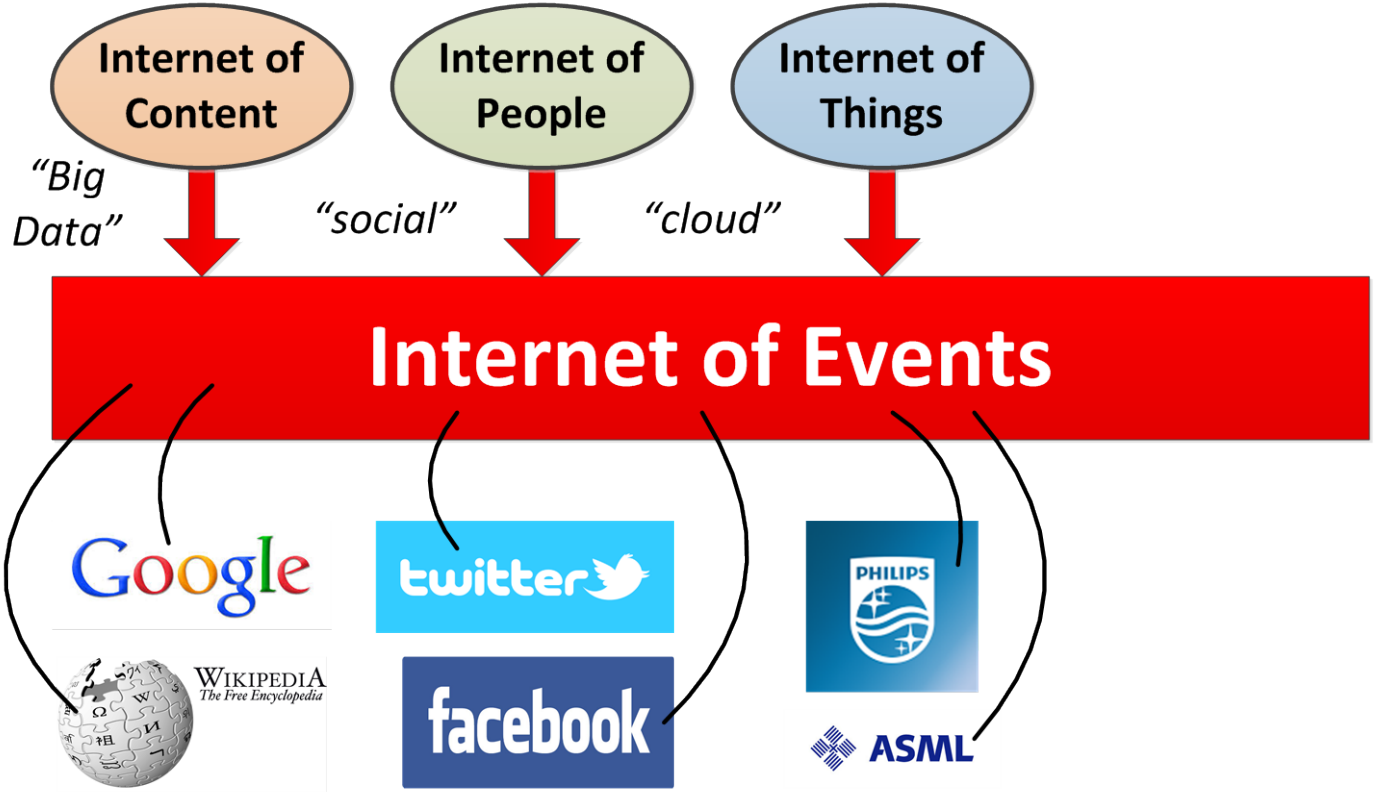
TU/e

Internet of Events: 4 sources of event data



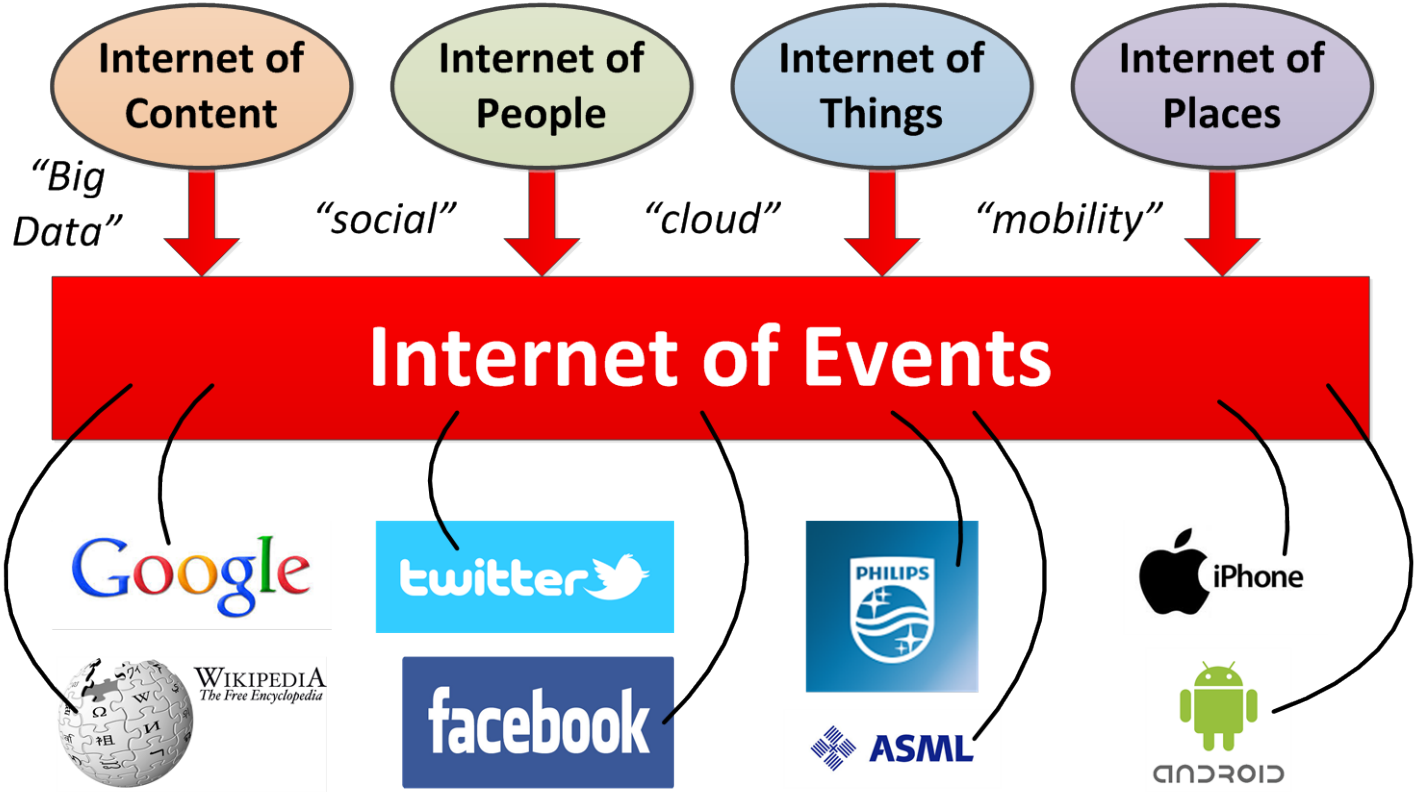
TU/e

Internet of Events: 4 sources of event data

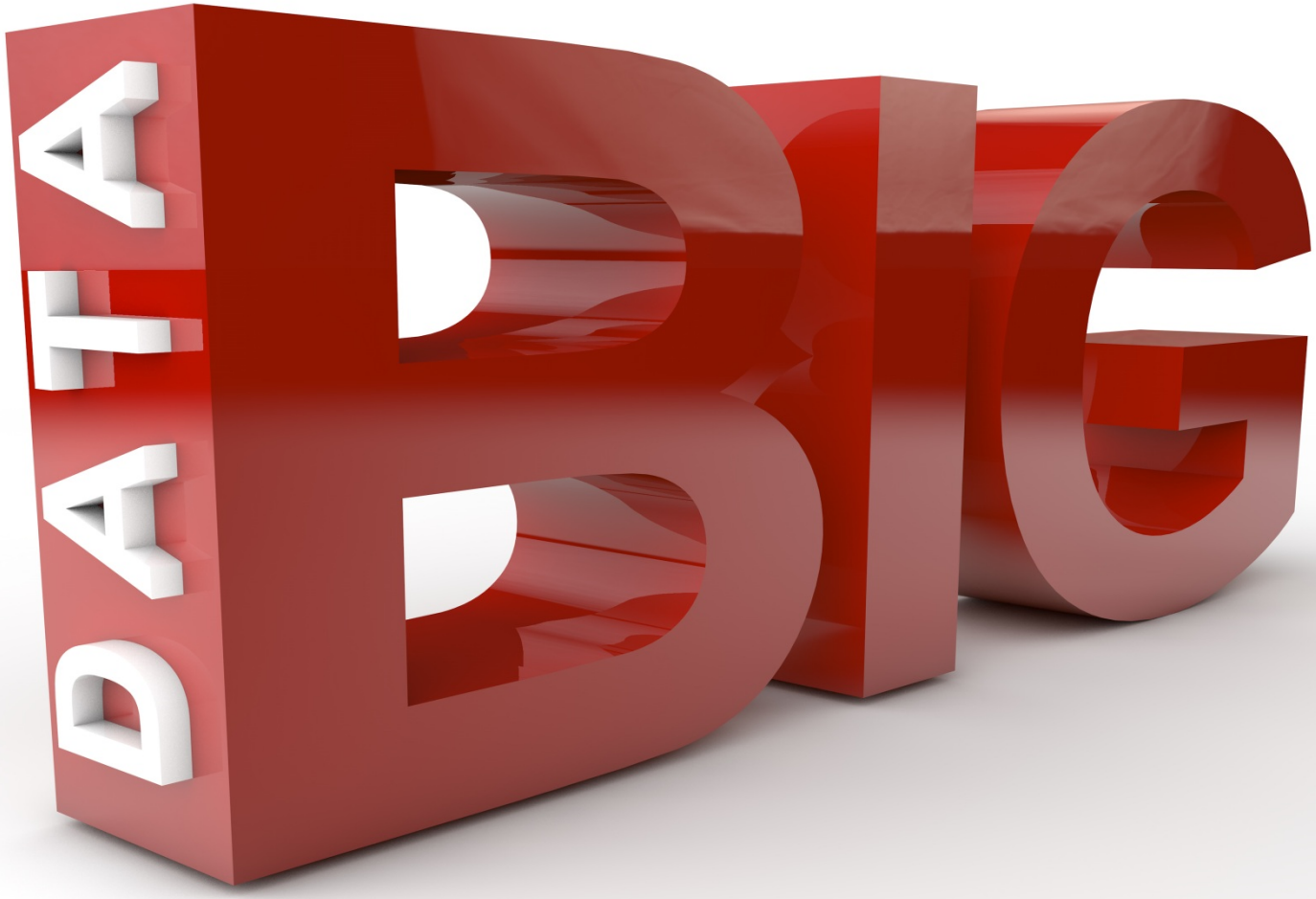


TU/e

Internet of Events: 4 sources of event data

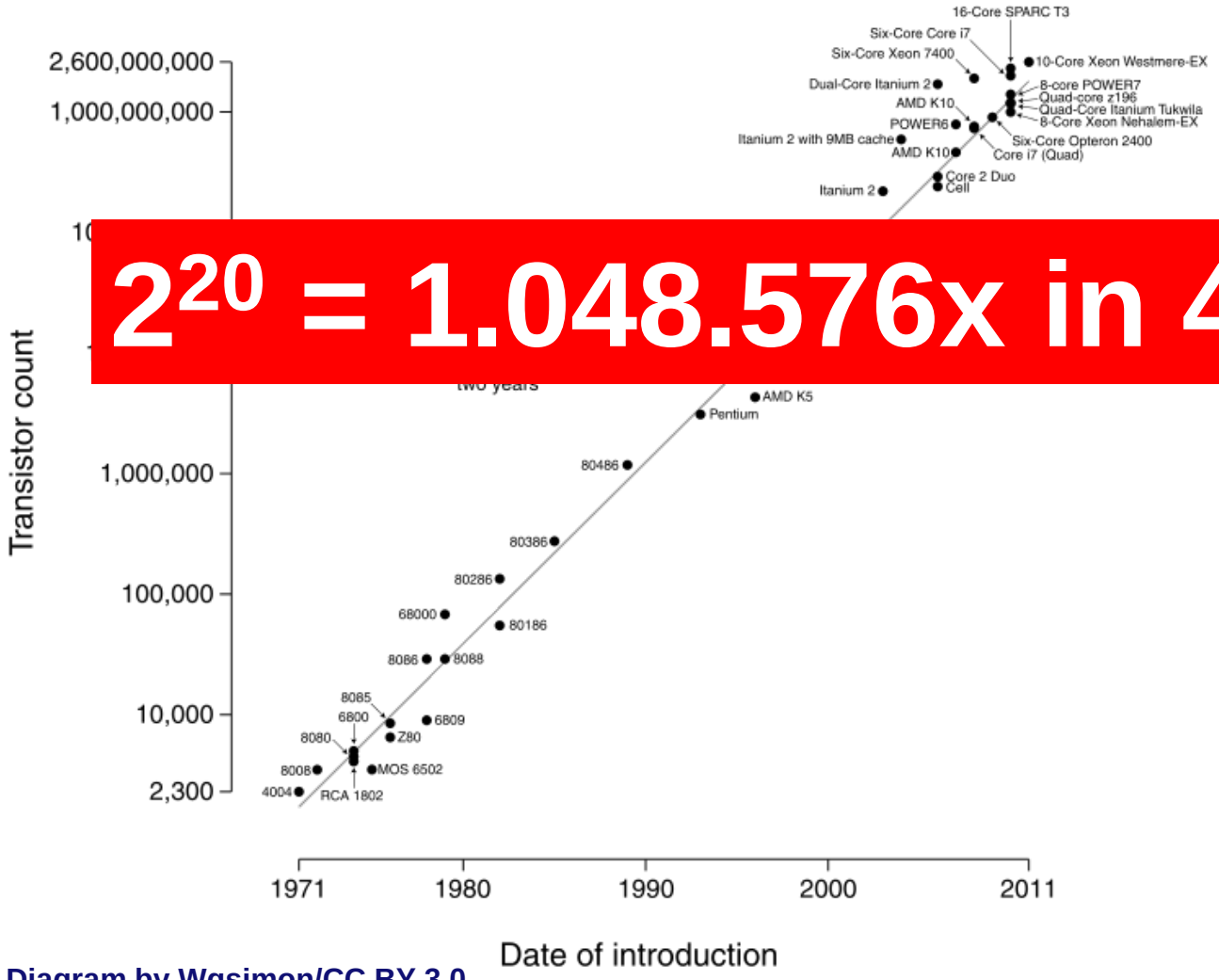


TU/e



TU/e

Moore's law



2^{20} = 1.048.576x in 40 years

Other examples:

- **Computing power**
- **Capacity of disks**
- **Bytes per dollar**
- **…**

Gordon E. Moore, Cramming More Components onto Integrated Circuits, Electronics, pp. 114–117, 1965.

Question

40 years ago it took approximately 1.5 hours to go from Eindhoven to Amsterdam by train.

How long would it take today if transportation technology would have followed Moore's law?

TU/e

Answer

$1.5 \times 60 \times 60 / 2^{20} =$

0.00515 seconds

Question

40 years ago it took approximately 7 hours to go from Amsterdam to New York by airplane.

How long would it take today if transportation technology would have followed Moore's law?

TU/e

$7 \times 60 \times 60 / 2^{20} = 0.0240$ seconds

Answer

Question

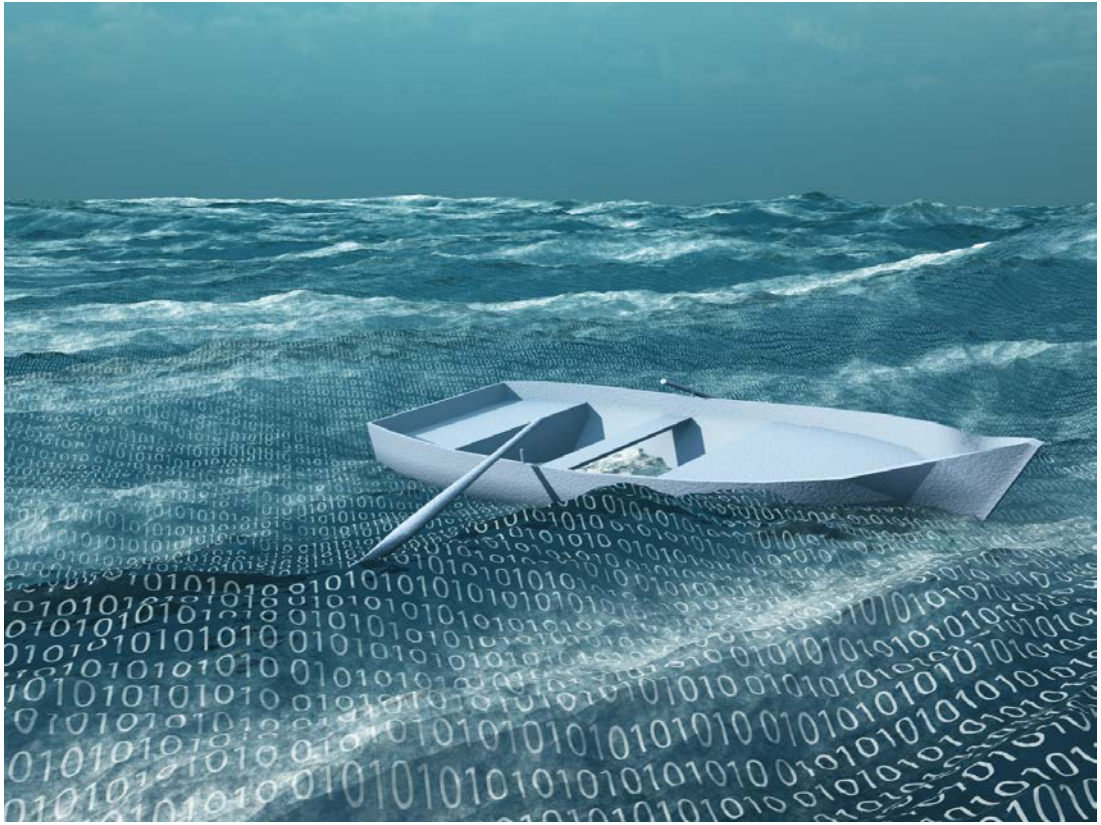
40 years ago it took approximately 4000 liters of petrol to drive around the world.

How much petrol would it take today if transportation technology would have followed Moore's law?

$4000 / 2^{20} = 0.0038$ liters

Answer

Drowning in data



How to extract real value from event data?

TU/e

4 V's of Big Data

VOLUME

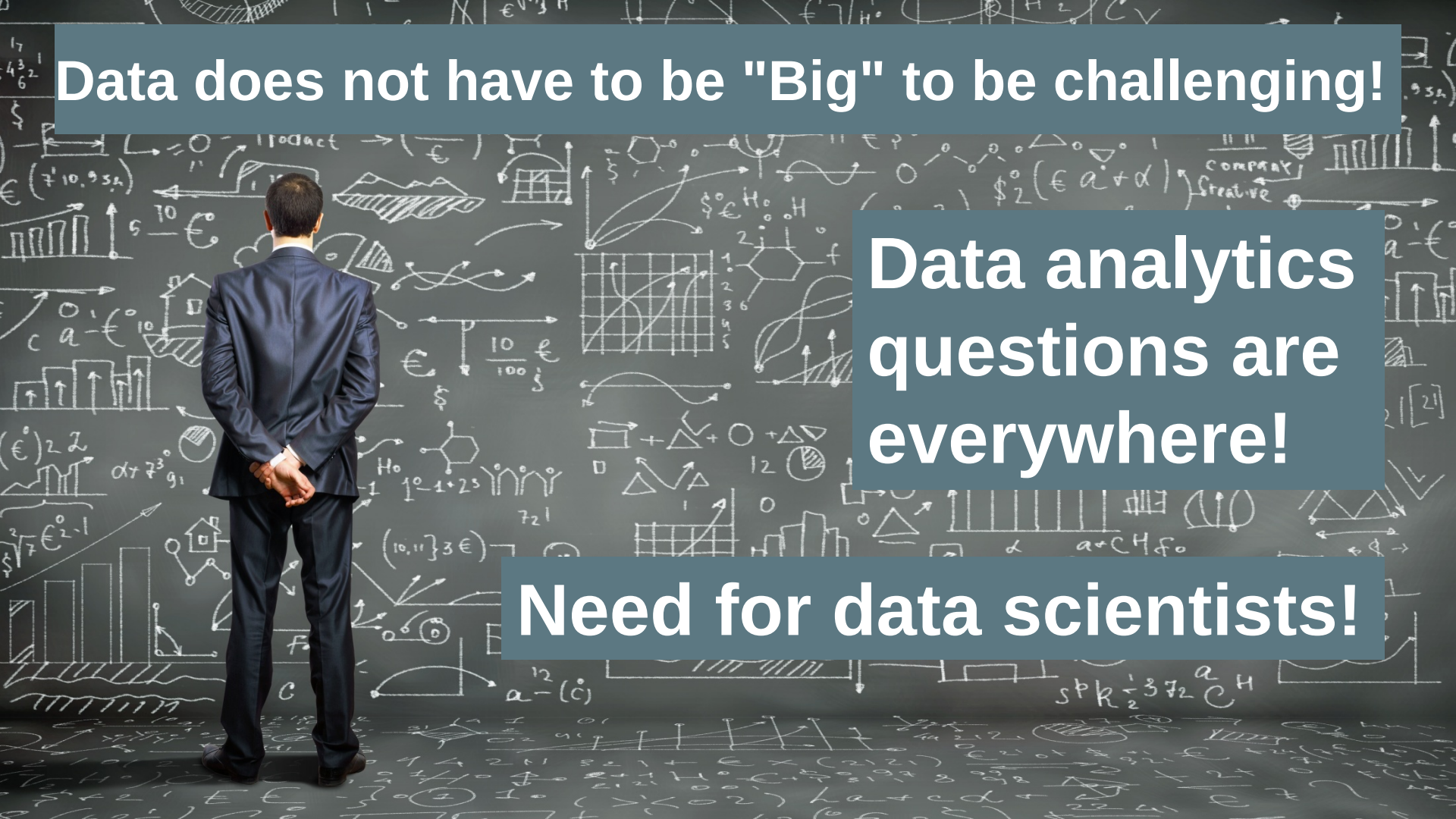
VOLUME
DATA SIZE

VELOCITY
SPEED OF CHANGE

VARIETY
DIFFERENT FORMS OF DATA SOURCES

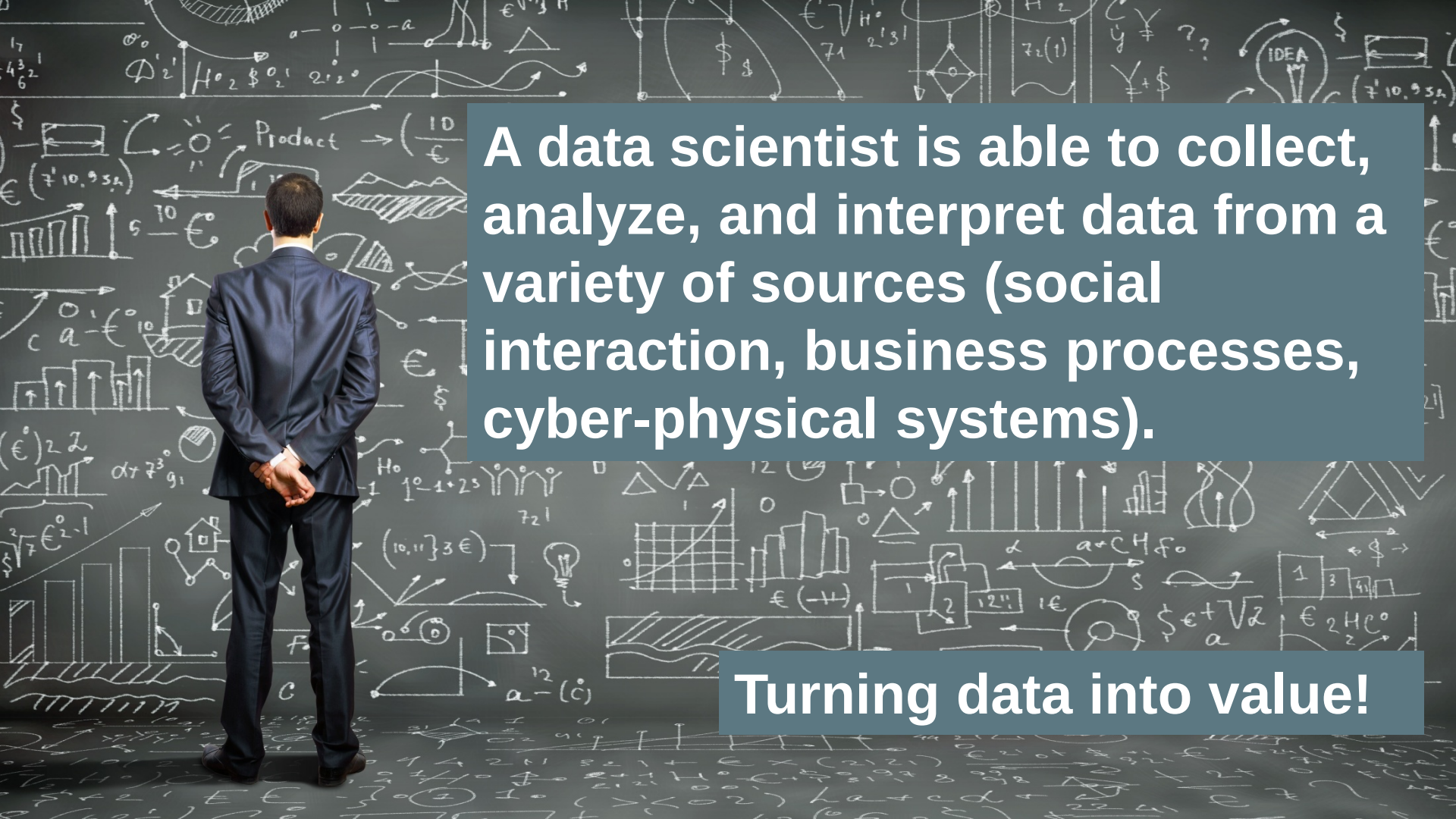
VERACITY
UNCERTAINTY OF DATA

Data does not have to be "Big" to be challenging!



Data analytics questions are everywhere!

Need for data scientists!

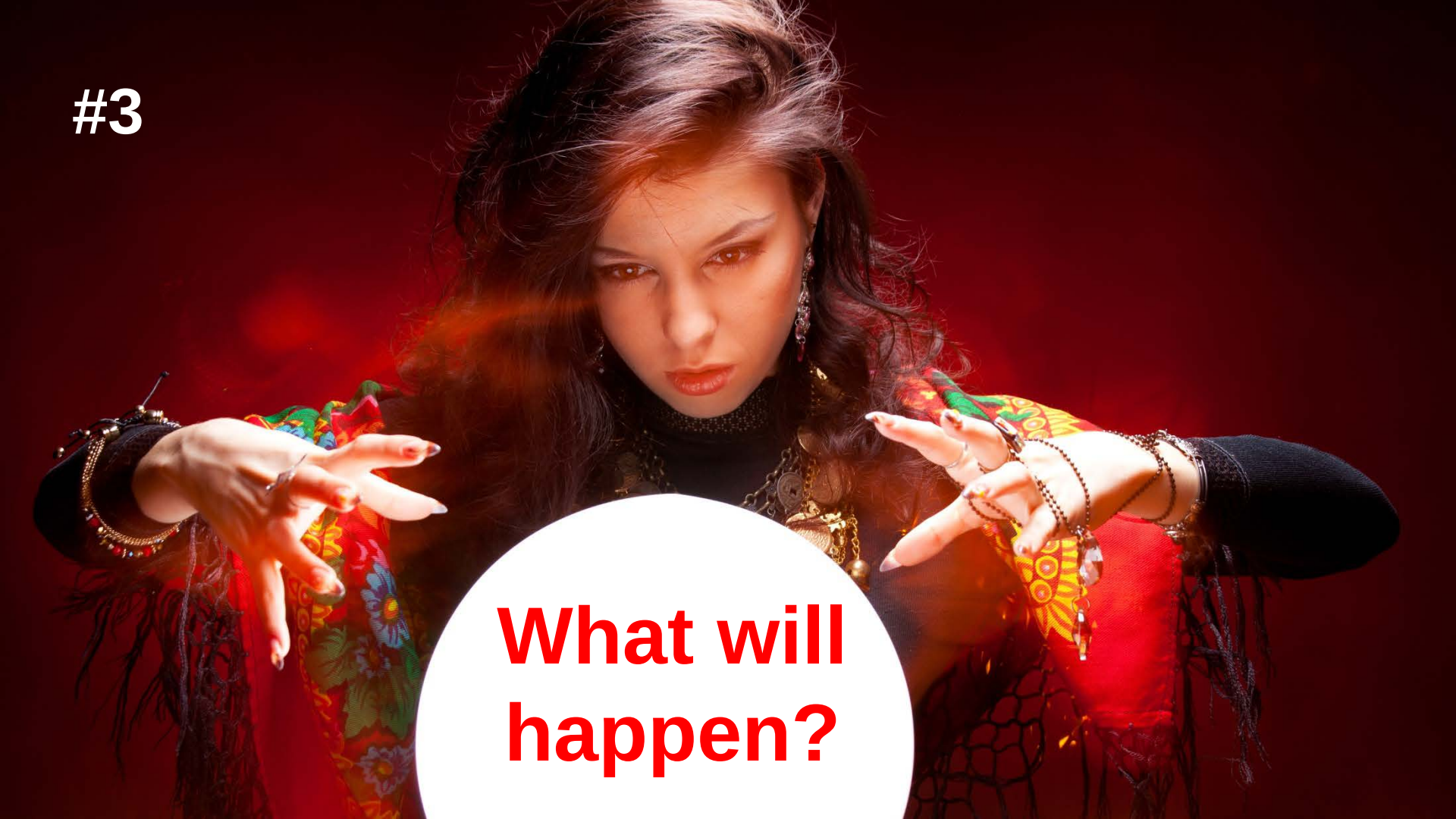

A data scientist is able to collect, analyze, and interpret data from a variety of sources (social interaction, business processes, cyber-physical systems).
Turning data into value!


Four generic
data science
questions

#1

#2

#3



#4

Why do patients have to wait so long?

Do doctors follow the guidelines?

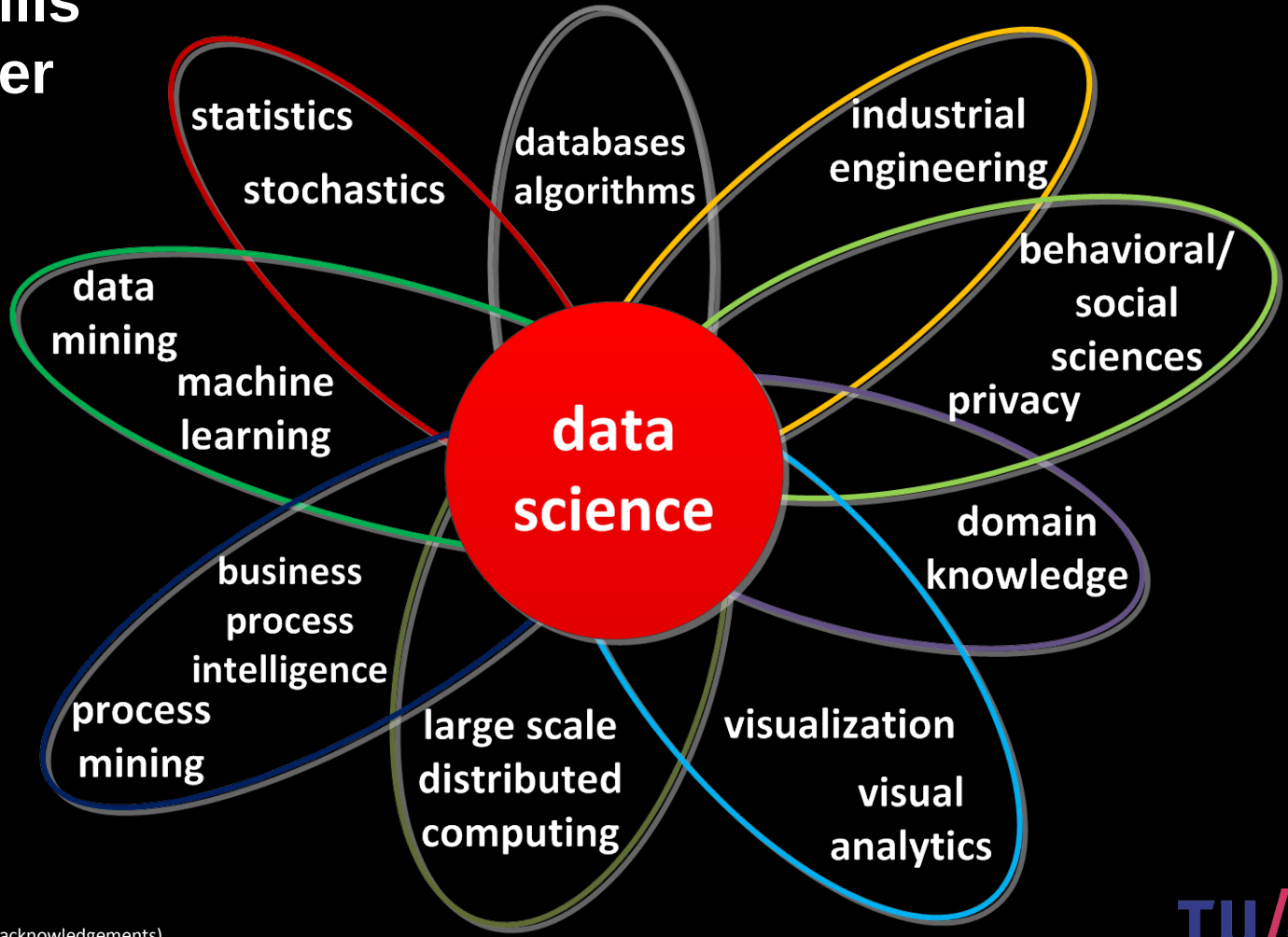
Can we predict waiting times?

How much staff is needed tomorrow?

How can we reduce costs?


How are X-ray machines really used?
Can we predict that the machine will break down next week?
Why and when do X-ray machines malfunction?
Which components should be replaced?
Which parts need to be improved?
from the organizational level
to the hardware/software level

Data science skills needed to answer such questions



TU/e


DSC/e
http://www.tue.nl/dsce/

It is the process stupid!

In the end it is the process that matters (and not the data or the software).

Not just patterns and decisions, but end-to-end processes.

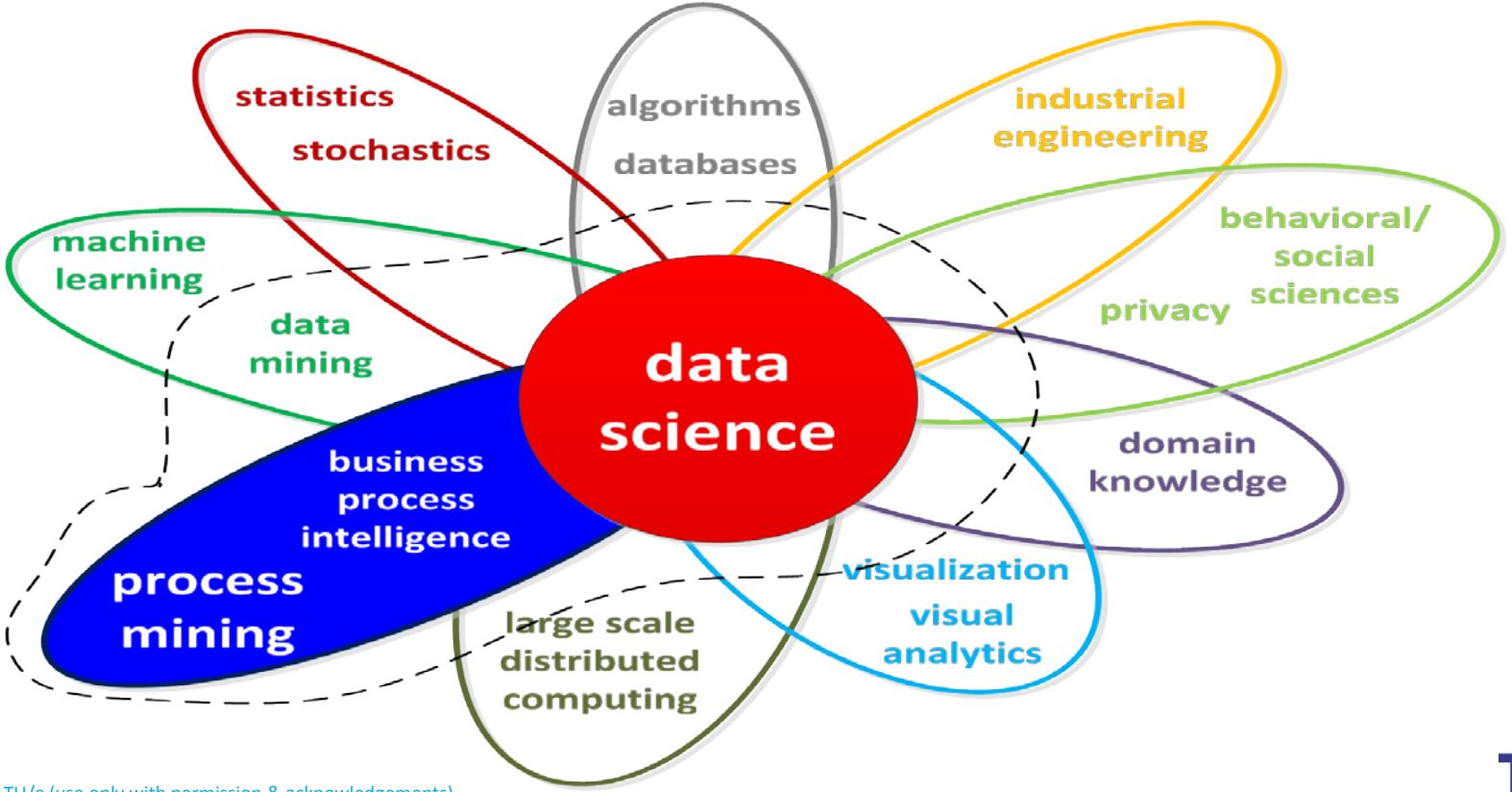
Process-centric view on data science

Understand and improve "care flows" in a hospital by intelligently using event data scattered over hundreds of database tables with patient data.

Understand and improve the behavior and performance of X-ray machines in the field using terabytes of low-level event data collected through a remote services network.

TU/e

Focus of this course



TU/e

Process mining use cases

- **What is the process that people *really* follow?**
- **Where are the bottlenecks in my process?**
- **Where do people (or machines) deviate from the expected or idealized process?**

TU/e

Many more use cases for process mining

- What are the "highways" in my process?
- What factors are influencing a bottleneck?
- Can we predict problems (delay, deviation, risk, etc.) for running cases?
- Can we recommend countermeasures?
- How to redesign the process / organization / machine?
- …

TU/e

Process Mining: Data Science in Action

Part I: Preliminaries

- Chapter 1: Introduction
- Chapter 2: Process Modeling and Analysis
- Chapter 3: Data Mining

Part II: From Event Logs to Process Models

- Chapter 4: Getting the Data
- Chapter 5: Process Discovery: An Introduction
- Chapter 6: Advanced Process Discovery Techniques

Part III: Beyond Process Discovery

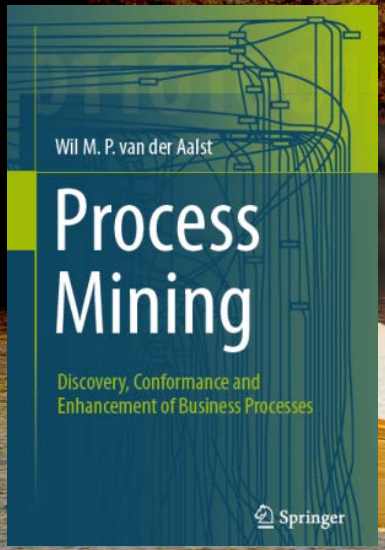
- Chapter 7: Conformance Checking
- Chapter 8: Mining Additional Perspectives
- Chapter 9: Operational Support

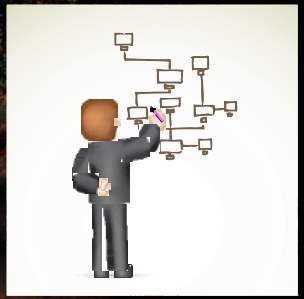
Part IV: Putting Process Mining to Work

- Chapter 10: Tool Support
- Chapter 11: Analyzing "Lasagna Processes"
- Chapter 12: Analyzing "Spaghetti Processes"

Part V: Reflection

- Chapter 13: Cartography and Navigation
- Chapter 14: Epilogue





TU/e