Big Data Science

Prof.dr.ir. Geert-Jan Houben

TU Delft

Web Information Systems

Delft Data Science

KIVI chair Big Data Science

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

big data: it's there, it's important

it is interesting to study it,
to **understand** it,
and to know how to **engineer** it

scientifically, we are driven by many **questions** and unprecedented challenges

Questions driving Big Data Science

TU Delft

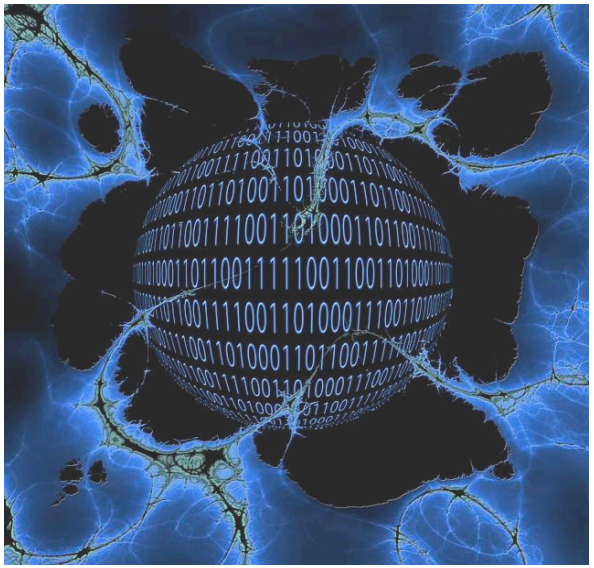
KIVI Engineering Society

DELFT DATA SCIENCE

100 billions in
economic and societal **value**

millions of new jobs and
millions of new talent to educate
in **technology** to get knowledge
and value out of big data

often, (massive amounts of)
data from **outside** the system
with properties that systems are
grappling with –
“data, too big to handle”



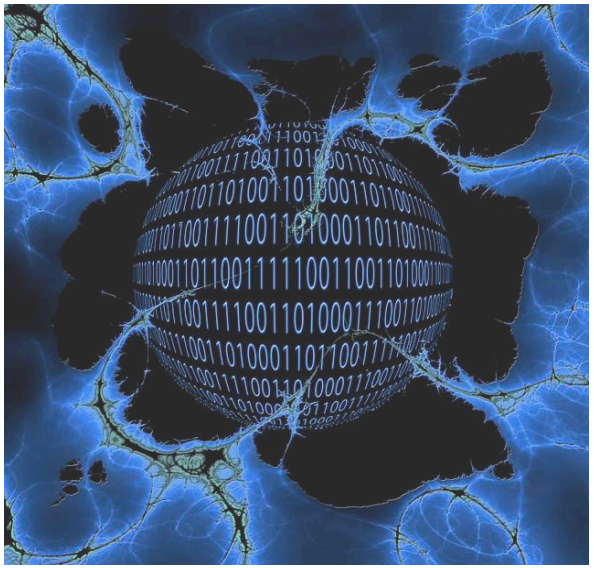
Big Data can fuel our economy & society

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

- decentralized & sustainable energy systems
- smart mobility
- personalized health
- smart industry
- digital society
- smart enterprises
- integral water management
- secure society
- intelligent living environments



Smart sectors rely on Big Data

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

typically information technology, **computing** science,
and a natural focus on software

the complexity is thought to be in **efficiency**

a **prescriptive** design approach: closed, fixed, centralized

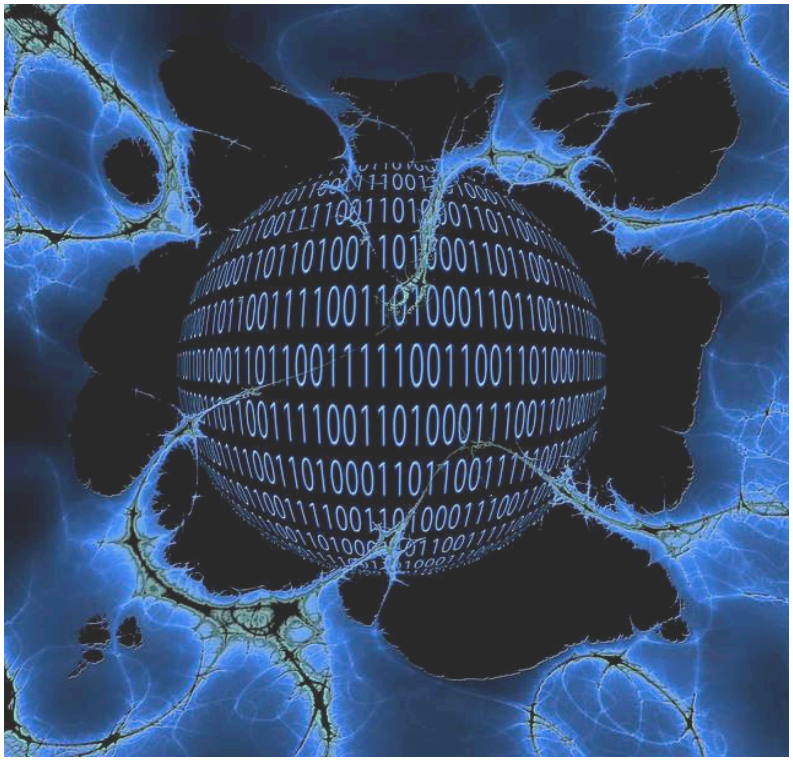
data representing the ‘world’ is **made** to fit the software

Traditional role of data

the Web brought linking **data** & connecting it to **people** for adaptation: utility

a **descriptive** approach: open, dynamic, decentralized

an **unprecedented** source of data about the ‘world’ (that people are part of) - “big data, too big to handle”



Web & Data

TU Delft

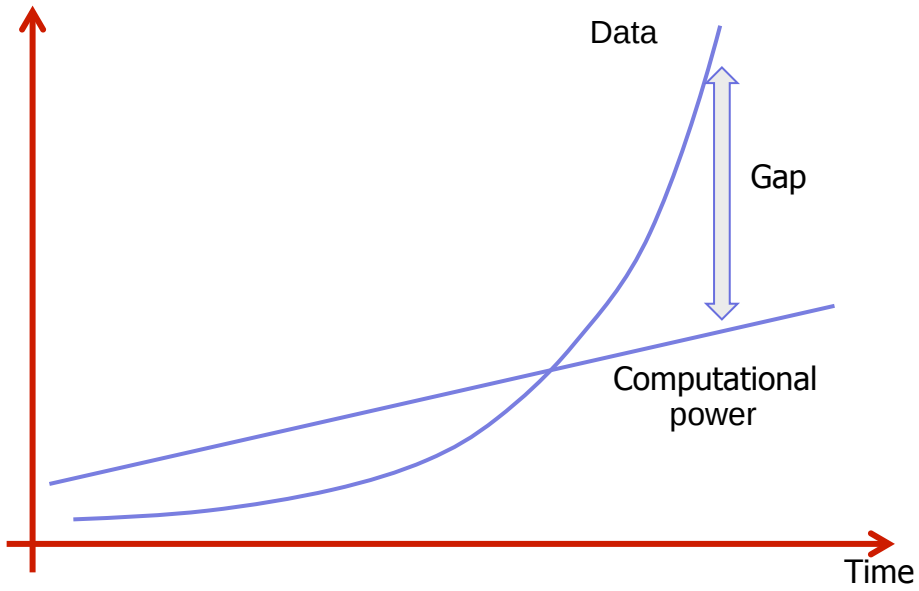
KIVI Engineering Society

DELFT DATA SCIENCE

technology to handle big data asks for a **fundamentally** new computing science

digital (Web) data and its descriptions of the world bring a new **complexity**

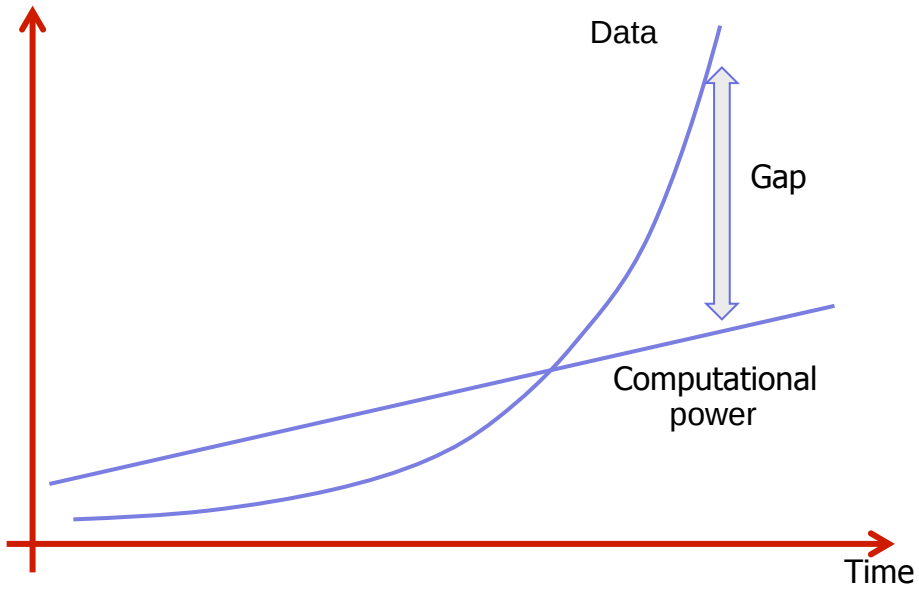
to make sense of the data, data science is all about **S**emantics, w. **S**cale, **S**peed & **S**ustainability



Complexity in data

data science is
a new scientific discipline for
scientific **understanding** and
creating **technology** for
how to create, process, and
understand digital data

data science is the foundational
discipline for **engineering**
data-driven systems



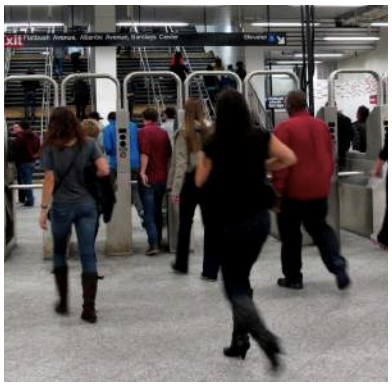
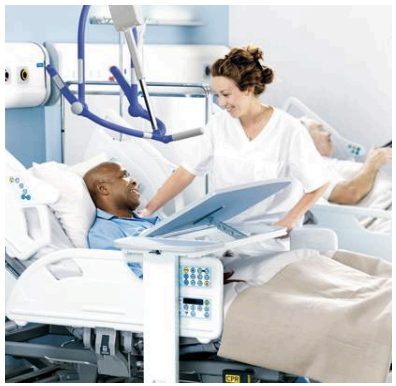
Data Science for advancing technology

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE



big data sets can be interpreted with **new scientific methods**

the world can be observed in **more detail**, leading to new knowledge and insight, and potentially better products, service or decisions

more detailed reflections of the world come with **new** questions

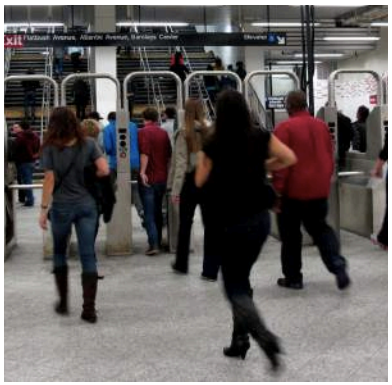
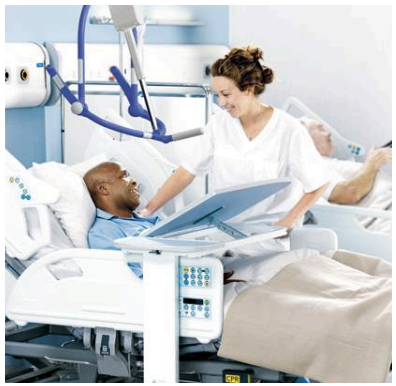
Data reflects the world

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE



how to **conclude** from heterogeneous, incomplete, unverifiable, sensitive, distributed data with **unknown** errors?

how to relate big data to **small** data?

how to **check** and **repeat** analyses?

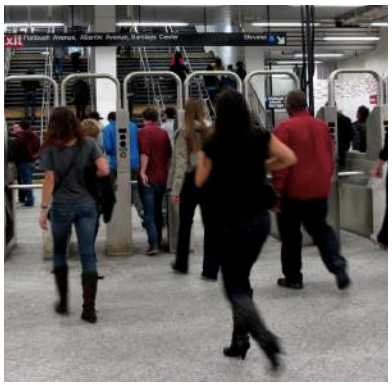
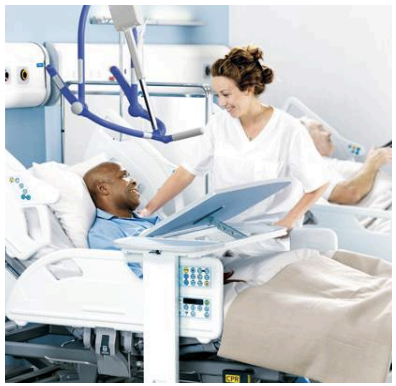
Unprecedented reflections

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE



how to prevent **false** conclusions?

how to answer questions without revealing **secrets**?

how to answer questions without stopping **time**?

how to handle **ownership** of data and satisfy personal and legal **contexts**?

Unprecedented reflections

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

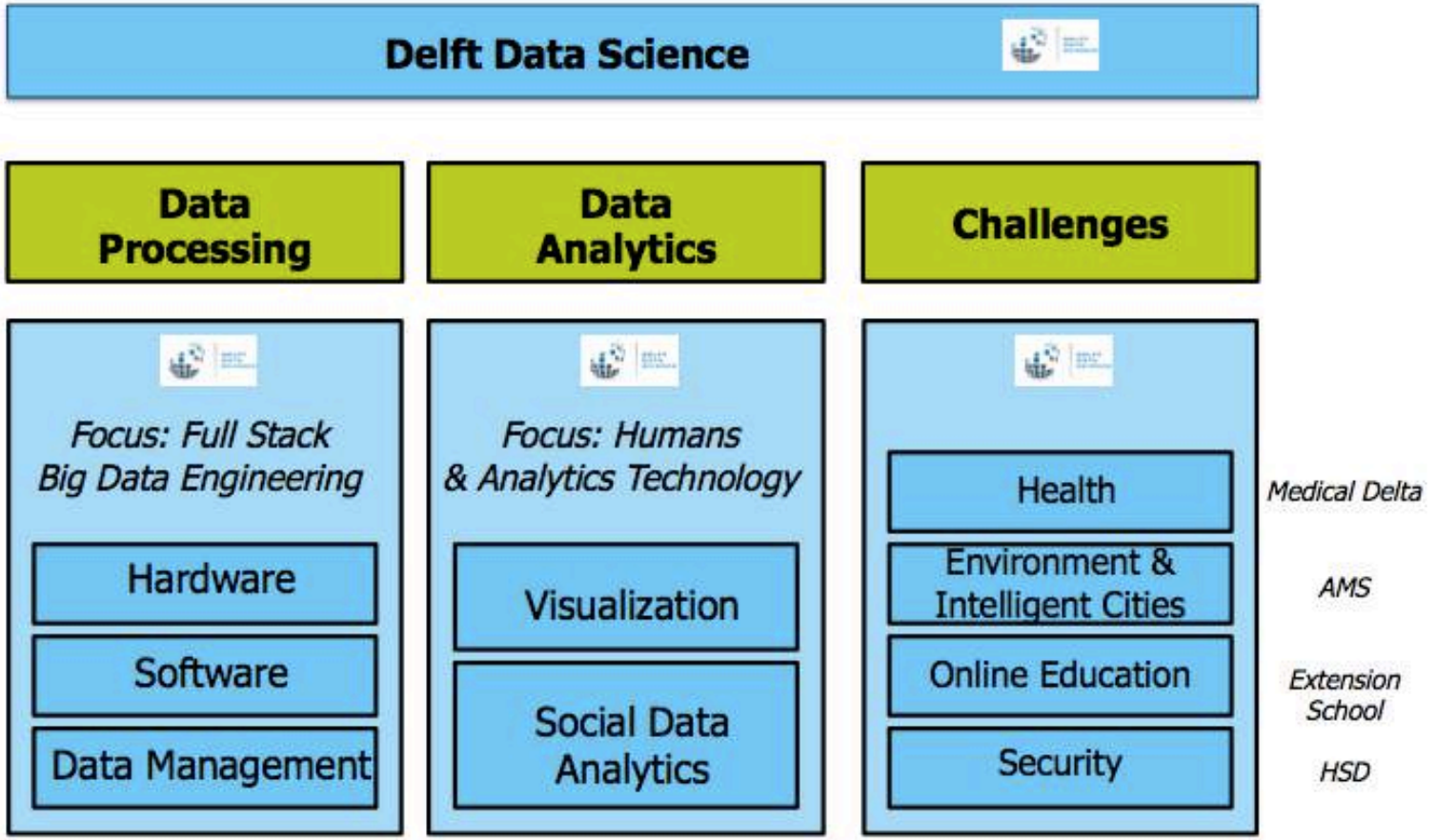
TU Delft coordinating initiative for research, education and training in data science and technology

Delft Data Science – research & education for technology & talent

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

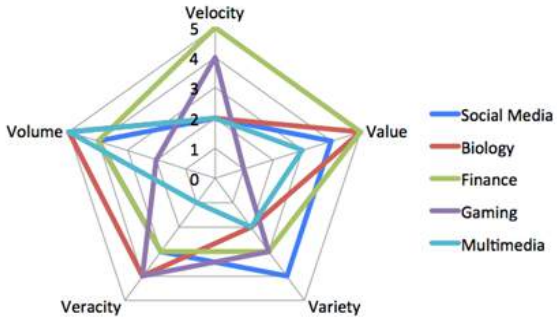


TU Delft

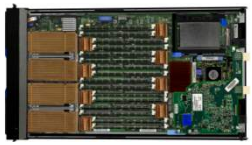
KIVI Engineering Society

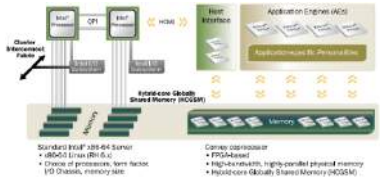
DELFT DATA SCIENCE

Hardware for Data Science

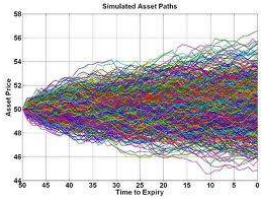


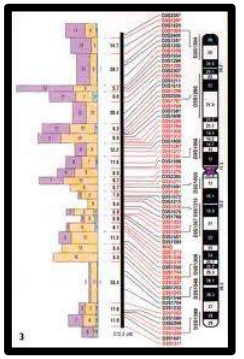
Big Data Computing Systems:
application specific
computing systems and hardware





New Algorithms & Architectures:
application and domain specific
e.g. finance/bio-informatics/seismic





Enabling big data computing systems to adapt to the challenges

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Software for Data Science

Problem: programming multi-core distributed cloud machines with Von Neumann programming languages

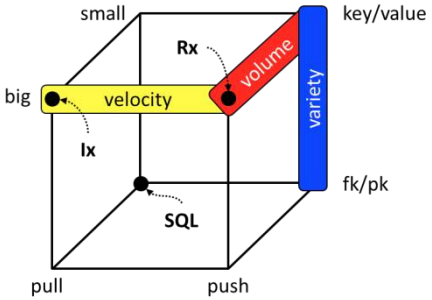
Problem: data engineers and scientists not trained as software engineers

Solution: programming languages that abstract from hardware, close to domain experts

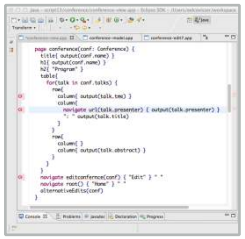
Cloud Programming:
composing computations using mathematically solid foundations

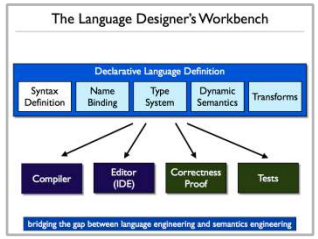
reactive extensions

interactive extensions



Domain-Specific Languages:
enabling software engineers to systematically design & apply DSLs





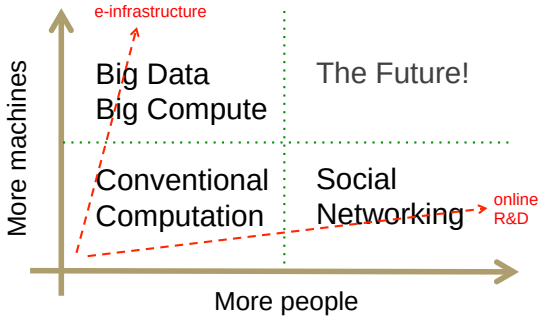
Enabling programmability of big data analytics

TU Delft

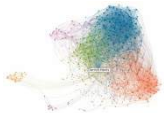
KIVI Engineering Society

DELFT DATA SCIENCE

Data Management for Data Science

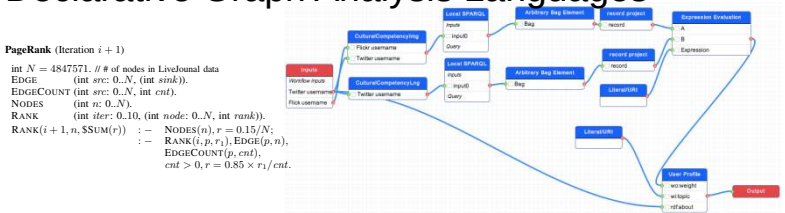


Graph & Network Data Processing:
processing graphs and networks
at big data and web scale



Web-Scale Graph Indexing

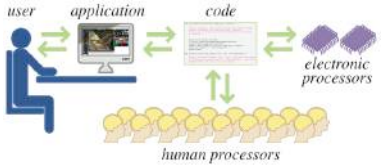
Declarative Graph Analysis Languages



Optimizing Big Data Processing Workflows

Humans Interacting in the Process:
enabling systems to include
human computation & interpretation

Crowdsourcing and Human Computation



Crowd Capacity

Data Generation & Data Curation

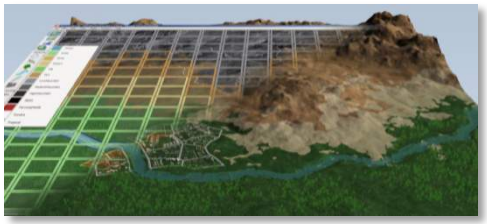
Enabling big data management at scale and with human interpretation

TU Delft

KIVI Engineering Society

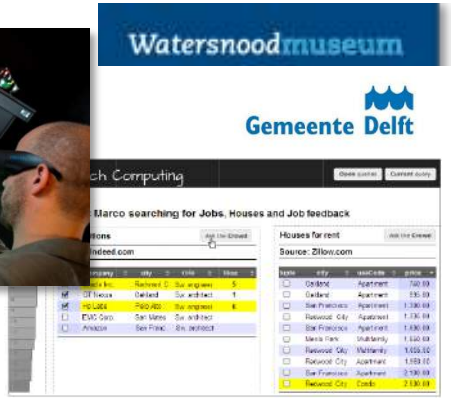
DELFT DATA SCIENCE

Visualisation for Data Science



Big Data Visualisation:
for real-time visual analytics
e.g. medicine, environment

Big Data Interaction:
for intuitive big data
exploration and manipulation



Enabling big data visual analytics

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

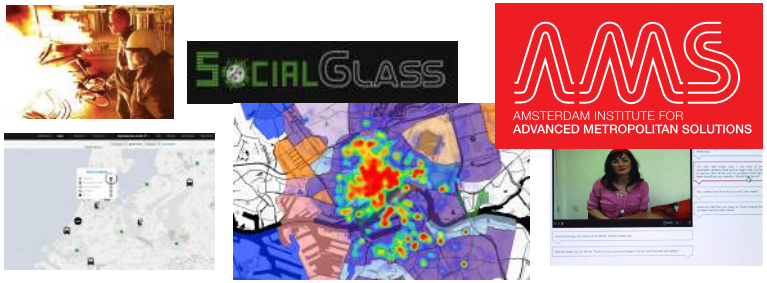
Science of Social Data

Opportunity:
data generated by humans, (re)presenting their take on the world

Challenge: largest source ever made, with yet-to-discover semantics

Social Data Analytics Machines:
repurposing social data that is out there, in controlled & well-understood manner

- Emergencies & incidents
- Intelligent cities
- Massive online education



Data Creation & Interpretation Machines:
including humans & human computing helping software in pro-actively creating & interpreting social data

- Social sensing
- Workforce engagement
- Crowd annotation & knowledge creation

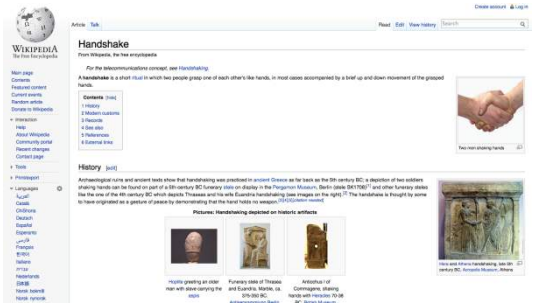
Unlocking human-generated data

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Example: Online Education



Learning analytics

with Delft Extension School for Open & Online Education

*analytics to make online education truly **learner-centric** and to adapt to the students & their backgrounds*

***massive** online education is about massively adapting to the context of use*

*with increasing diversity comes importance of social and cultural features: **inclusion***

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Unlocking Social Data:
Software & Human-enhanced Machines with Well-Understood Properties

TU Delft

KIVI Engineering Society

WIS - Web Information Systems

DELFT DATA SCIENCE

Science of social data

1. Social data gives us one of the **largest reflections** of the world, but/and it is a **man-made** reflection
   'unique opportunity turning into interesting research problem'

2. Sense & value come from big data, but even more so from what (software and human-enhanced) **machines** can make of the data
   *'V = M * D'*

3. The **power** of what machines can do with the data needs to be well-understood and transparent for solid engineering and uptake
   'what machines can do and what they cannot do'

4. Science and technology follow the principles of the **Web**
   'fundamental & experimental'

Data Science for **Intelligent Cities**
Data Science for **Environmental Monitoring**
Data Science for **Finance**
Data Science for **Health**
Data Science for **Online Education**
Data Science for **Cybersecurity**
Data Science in **Open Data**
Data Science for **Workforce Management**
…

DELFT DATA SCIENCE

Delft Data Science & domains

TU Delft

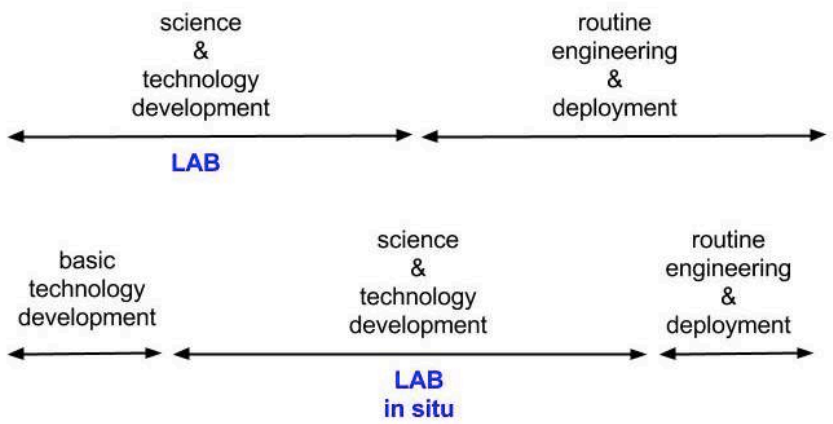
KIVI Engineering Society

DELFT DATA SCIENCE

data science comes with **new methods** for doing science and new **ways** to **collaborate**

data science labs are **in-situ**

data science research shows the importance of **local understanding** of how to apply [seemingly global] technology



New scientific methods & collaborations

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Research Agenda

Over the coming years, researchers engage with engineers for inspiration, experimentation, and valorization.

We invite to collaborate.

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Next?

Today, the first three master classes.

Get introduced to three branches of data science.

Engage with this research after today.

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

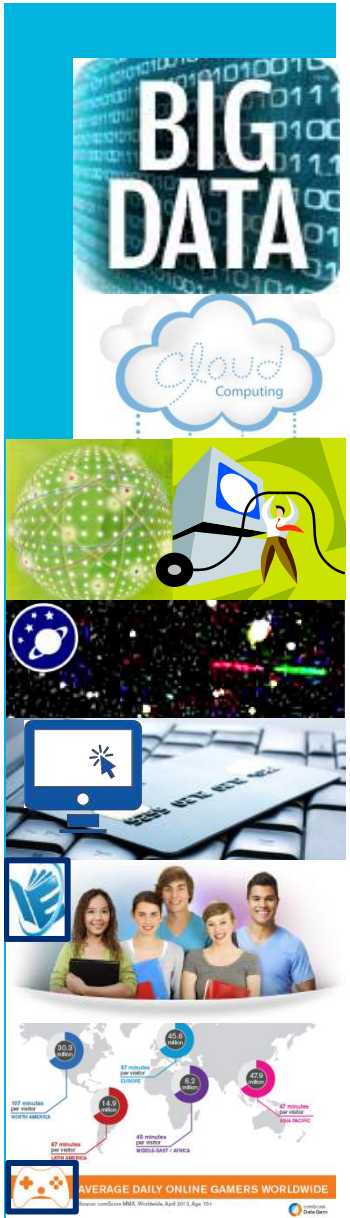
Alexandru Iosup

Scalable High Performance Systems

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE



Scientific and Societal Challenges

How to massivize datacenters?

- Super-scalable, super-flexible, yet efficient
- End-to-end automation
- Dynamic workloads
- Evolving hardware and software
- Strict performance, cost, energy, reliability, and fairness requirements

The quadruple helix: prosperous society & blooming economy & inventive academia & wise governance

- Enable data access & processing as a fundamental right in Europe
- Enable big science and engineering (2020: €100 bn., 1 mil. jobs)
- “To out-compute is to out-compete”, but with energy footprint <5%
- Keep Internet-services affordable yet high quality in Europe
- The Schiphol of computation: Netherlands as a world-wide ICT hub

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Alessandro Bozzon

Crowdsourcing in Enterprise Environments

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

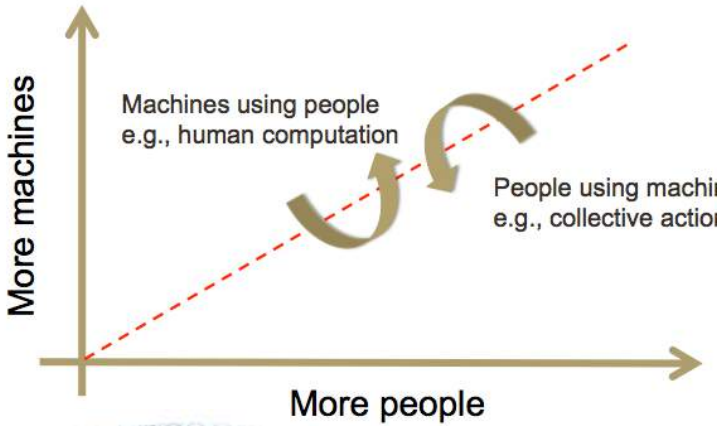
Challenges

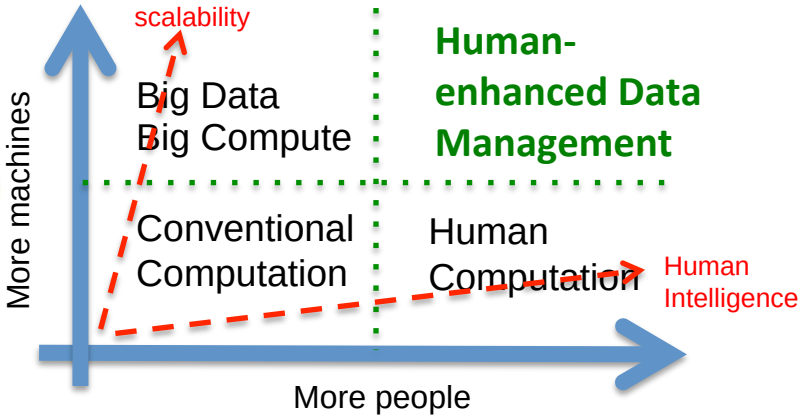
Scientific Challenge:

- How can humans and machines better collaborate in computation problems?

Societal (and business) challenges

- Knowledge Creation
- Inclusion and Well-being
- Employment





TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Zaid Al-Ars

Acceleration of Personalized Medicine Applications

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

Scientific and societal challenges

- Urgent clinical diagnostics, for example
  - Targeted cancer & neo-natal diagnostics
  - ➔ We provide techniques to ***reduce compute time***
- Cost prohibitive for society
  - More patients & diseases to be treated
  - ➔ We provide techniques to ***reduce cost***

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE

gjhouben.nl

wis.ewi.tudelft.nl

delftdatascience.tudelft.nl

TU Delft

KIVI Engineering Society

DELFT DATA SCIENCE