Netherlands Second in Data Science?

Time for action!

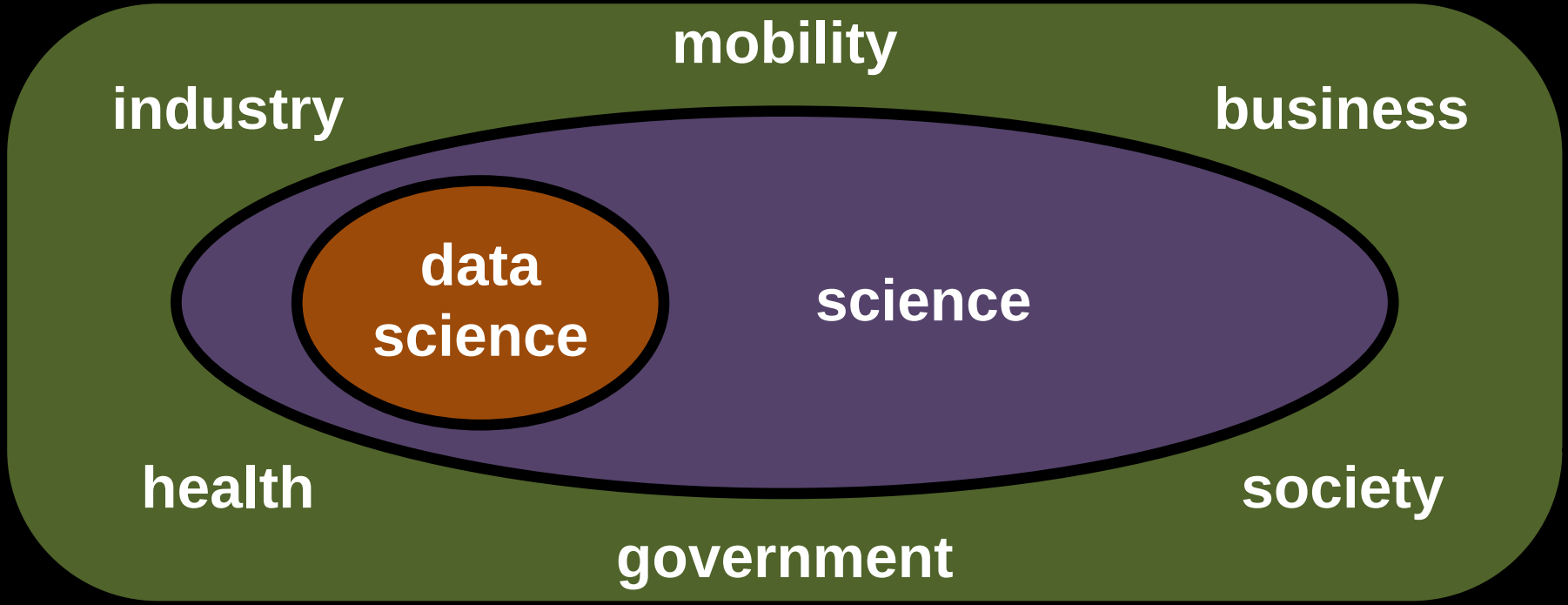
KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

Wil van der Aalst

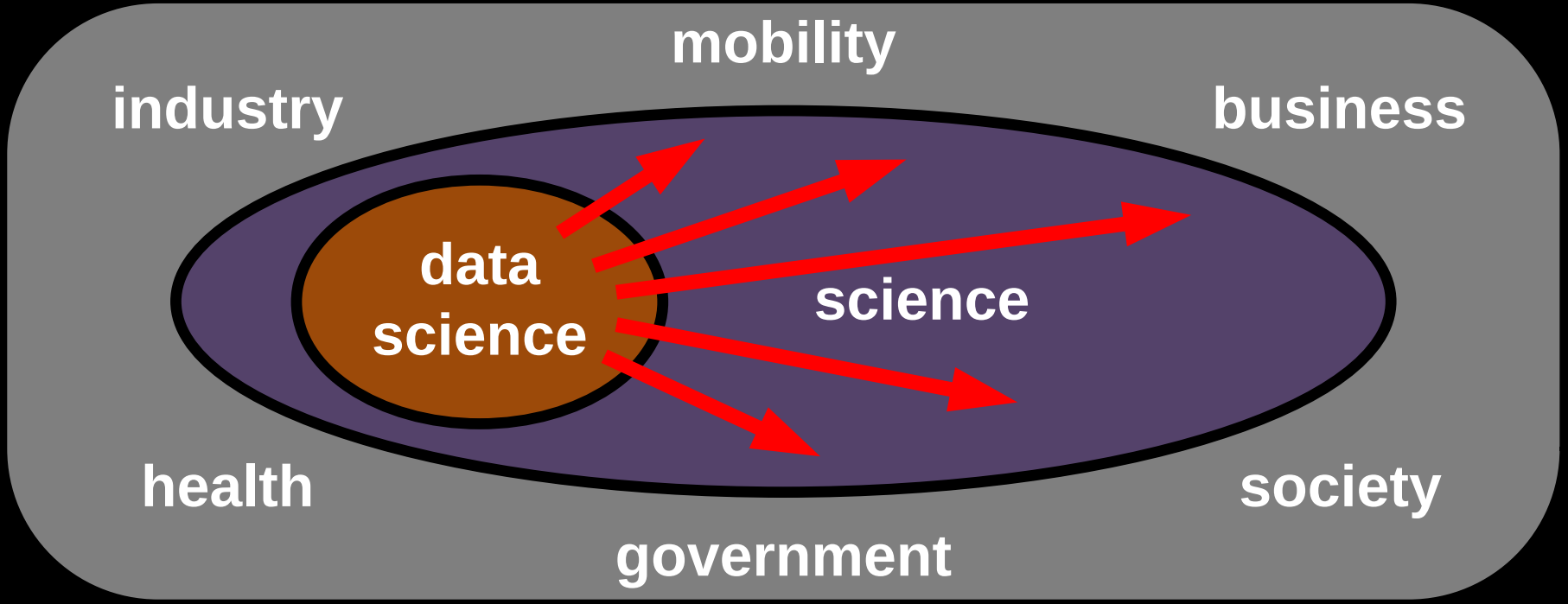
TU/e
DSC/e

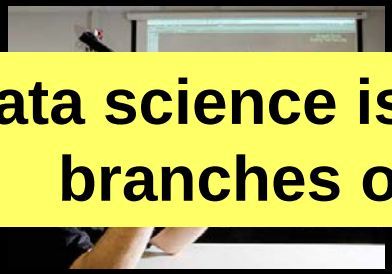
Based on KNAW-NTW lecture 10-4-2017

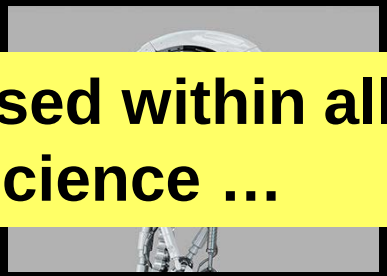

Driver: More data, improved scalability, and smart analytics
cloud
share
BIG DATA
Info
Fast
online
Photo
Music
Data science and Big data are changing everything!



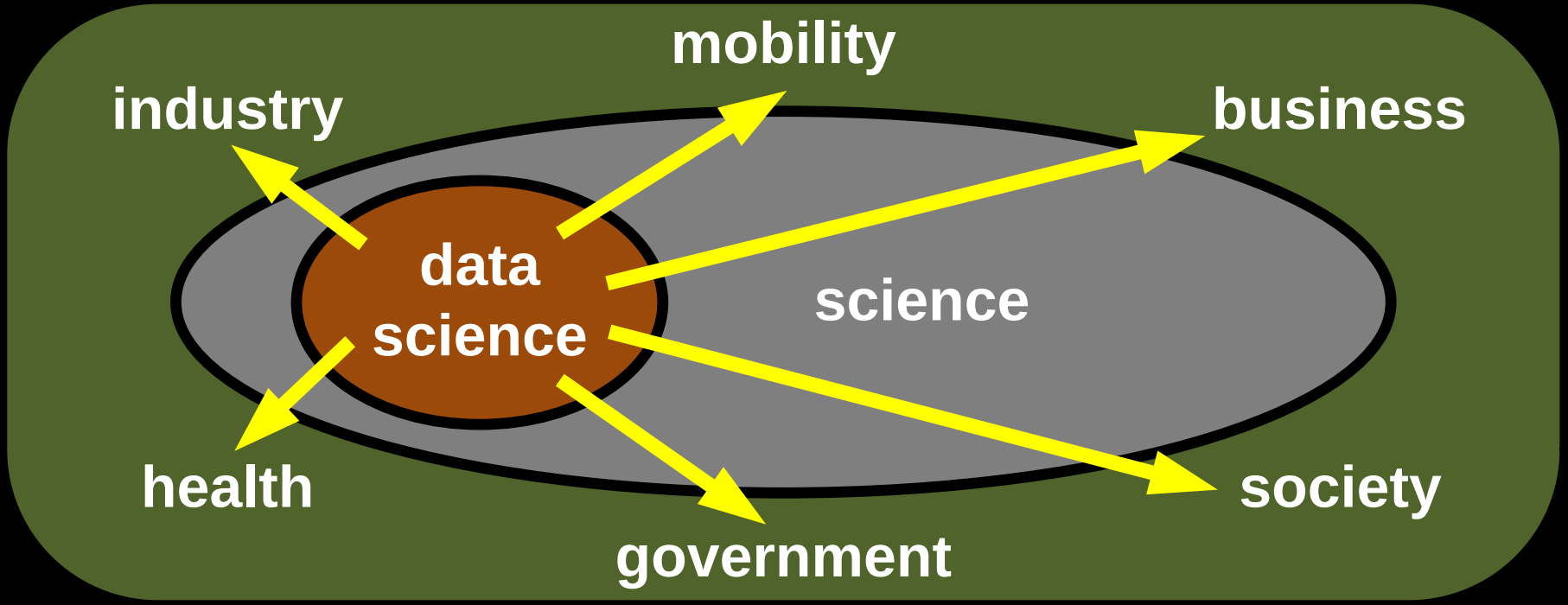
An emerging scientific discipline

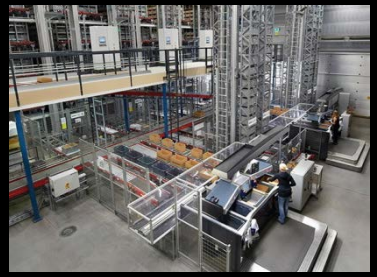






Data science is used within all branches of science …



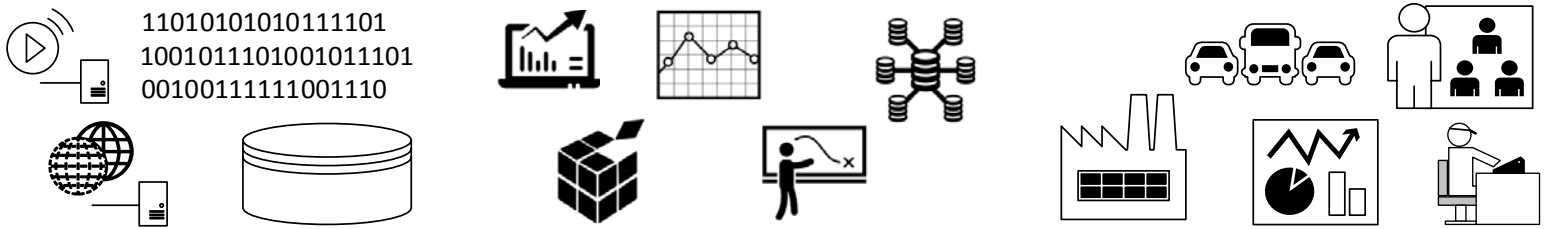


and outside science.

NWA Agenda

- **Questions: About 2500 of the 11,700 questions (more than 20%) were related to data science and Big data.**
- **Many of these questions expressed concerns with respect to irresponsible uses of data: 62 questions related to balancing privacy and information sharing, 77 to digital security, 68 to data and government, 180 to the effects of digitization on society, etc.**
- **Jury acknowledged this in 2015 by creating the Big data route.**
- **Responsible data science has been a common theme in NWA Big Data Route meetings (2-3-2016, 5-4-2016, 13-5-2016, 11-4-2017).**
- **Supported IPN, DSPN, etc.**
- **Main theme in the VSNU Digital Society initiative.**

The main challenges



Data Science as "the enabler"

infrastructure

"volume and velocity"

- instrumentation
- big data infrastructures and distributed systems
- databases and data management
- programming
- security
- ...

analysis

"extracting knowledge"

- statistics
- data/process mining
- machine learning/artificial intelligence
- operations research
- algorithms
- visualization
- ...

effect

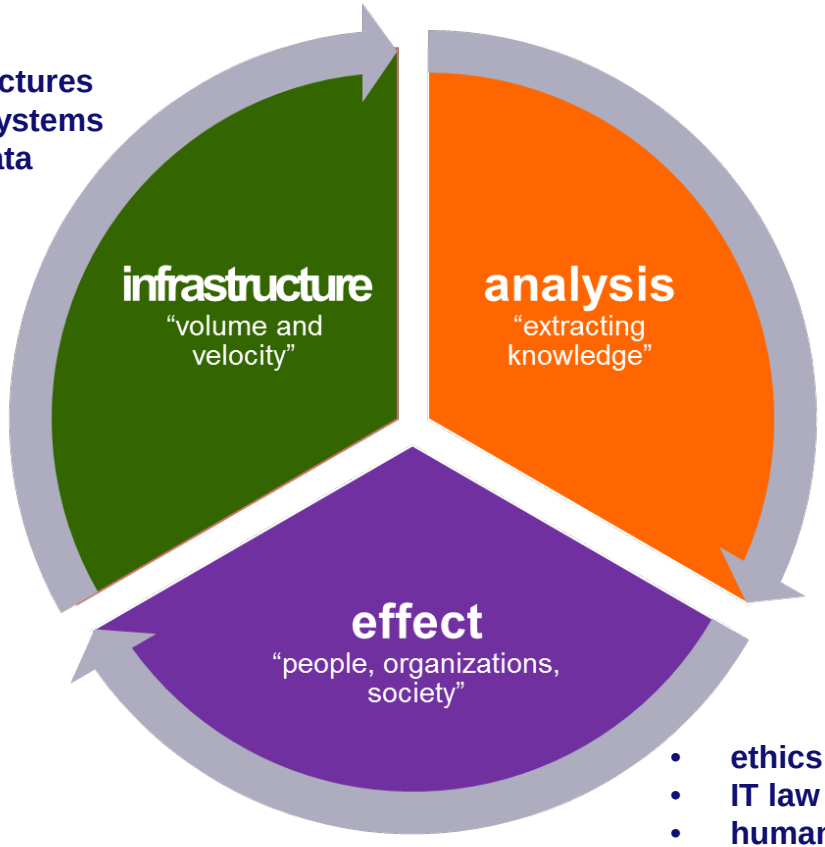
"people, organizations, society"

- ethics & privacy
- IT law
- human technology interaction
- operations management
- business models
- entrepreneurship
- ...

the data science pipeline

TU/e

- instrumentation
- big data infrastructures and distributed systems
- databases and data management
- programming
- security



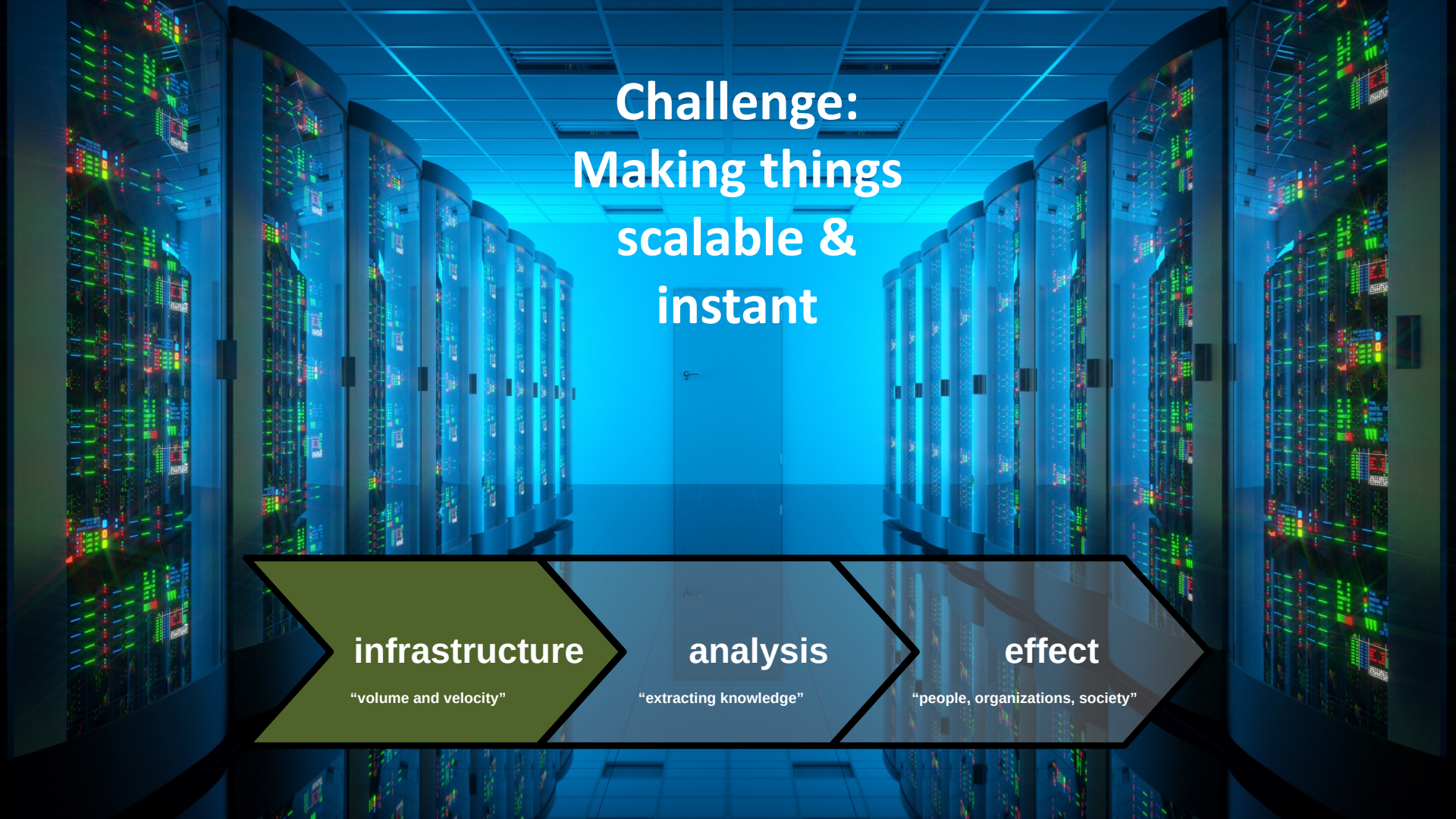
- statistics
- data/process mining
- machine learning/artificial intelligence
- operations research
- algorithms
- visualization

Closed loop

- ethics & privacy
- IT law
- human technology interaction
- operations management
- business models
- entrepreneurship

TU/e

The data science landscape


Challenge:
Making things
scalable &
instant
infrastructure
“volume and velocity”
analysis
“extracting knowledge”
effect
“people, organizations, society”

Challenge: Providing answers to known and unknown unknowns

United States Secretary of Defense
Donald Rumsfeld (12-2-2002)

Challenge: Doing all of this in a responsible manner!

www.responsibledatascience.org

Importance is clear, but we have a major problem!

Our data-driven research is like a pizza …

What makes a good pizza?

The focus is mostly on the toppings (= applications)

…, but the base actually matters!

TU/e

by @lorilewis

Exception !

Sweden saves us!

Do we accept that most of our pizza bases come from the US?

Are our scientists (i.e. we) doing any better ?

Nope!

TU/e

Automated ranking based on Google Scholar (H-index).

Guide 2 Research

Ranking on 9-4-2017

Top H-Index for Computer Science and Electronics :

We list only scientists having H-Index>=40. If you or other scholars are not listed, we appreciate if you can contact us

Search by name [redacted] View by country [redacted]

World	National Ranking	Scholar	Country	Citations	H-index
1	1	Herbert Simon, Carnegie Mellon University	United States	298,593	167
2	2	Anil K. Jain, Michigan State University	United States	168,787	165
3	3	Jiawei Han, University of Illinois	United States	129,114	146
4	4	Scott Shenker, University of California, Berkeley	United States	113,032	141
5	5	Mark Gerstein, Yale University	United States	92,584	141
6	6	Terrence Sejnowski, Salk Institute	United States	104,561	139
7	7	Takeo Kanade, Carnegie Mellon University	United States	96,858	139
8	8	David Haussler, University of California, Santa Cruz	United States	125,759	136
9	9	robert tibshirani, Stanford University	United States	243,695	135
10	10	Michael I. Jordan, University of California, Berkeley	United States	113,817	134
11	11	Philip S. Yu, University of Illinois at Chicago	United States	81,087	132
12	1	Wil van der Aalst, Eindhoven University of Technology	Netherlands	76,301	131
13	1	Andrew Zisserman, University of Oxford	United Kingdom	106,944	130
14	12	Sebastian Thrun, Stanford University	United States	77,097	130
15	1	Geoffrey Hinton, University of Toronto	Canada	165,322	128

Exception !

Exception !

Exception !

Automated ranking based on Google Scholar (H-index).

Ranking on 9-4-2017

Also the top 10 computer science and electrical engineering researchers are based in US!

Netherlands 2nd ?

Not really: Netherlands 8th

Relatively OK, note Belgium, China, and Scandinavian countries

Dutch researchers

- 1 in top 250 (WvdA 12th)
- 5 in top 500
- 14 in top 1000
- 38 in top 1635 (H-index > 40)

?

United States	879
United Kingdom	105
Germany	90
Canada	85
France	59
Switzerland	51
Italy	44
Netherlands	38
Australia	35
Israel	33
Hong Kong	24
Spain	23
Belgium	21
Japan	18
Singapore	18
Austria	12
China	10
Sweden	10
Greece	10
South Korea	7
Finland	7
India	7
Ireland	7
Poland	5
Taiwan	5
New Zealand	4
Portugal	4
Mexico	3
Denmark	3
Luxembourg	2
Brazil	2
Czech Republic	2
Chile	2
Qatar	1
Saudi Arabia	1
Iceland	1
Estonia	1
South Africa	1
Turkey	1
United Arab Emirates	1
Norway	1
Iran	1
Serbia	1

TU/e

Let's put more resources on the pizza base of our digital society!

… or stop making pizzas

TU/e


Call for Action!
databases/distributed systems
machine learning/data mining
big data algorithms
visualization

Edsger Dijkstra (1930-2002), Turing award winner (1972), inventor of the shortest path algorithm, first implementation of Algol, semaphores, structured programming, etc. Professor at TU/e from 1962 to 1984.

Once upon a time ...


MADE IN
HOLLAND
MADE IN
Let’s make the Netherlands great again!