

# Delft Data Science

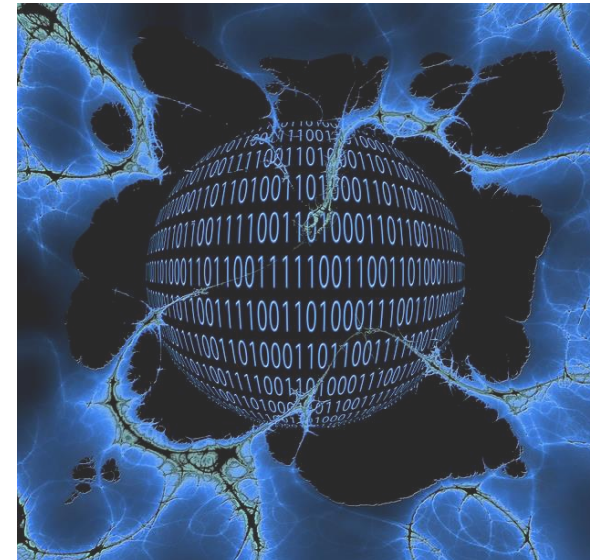
**Seminar  
January 26, 2015**

*Big Data Analytics for  
Cyber Situational Awareness*

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

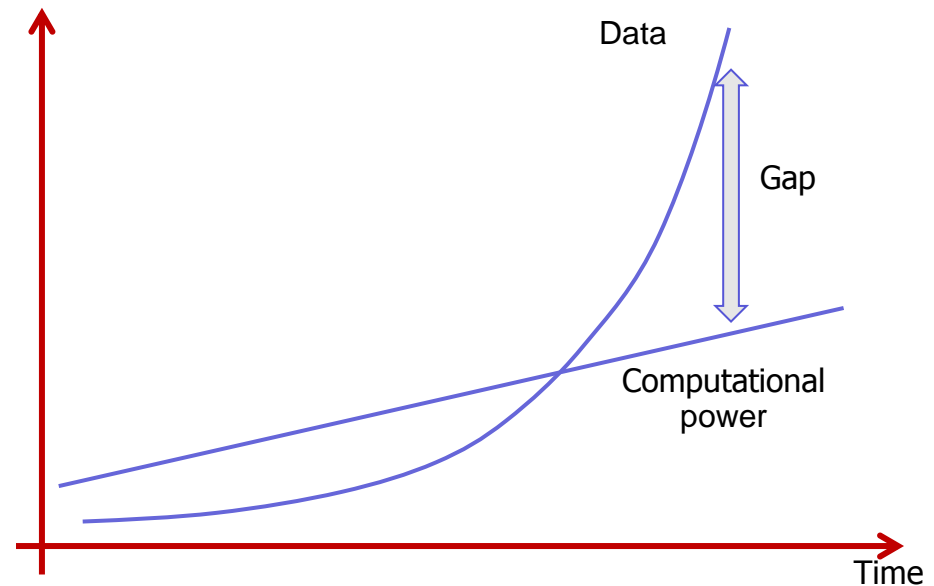
often, (massive amounts of) data from **outside** the system properties that systems are grappling with – “data, too big to handle”



# Big Data can fuel our economy & society

digital (Web) data  
and its descriptions of the world  
bring a new **complexity** –  
for systems to know  
(how to handle) the data

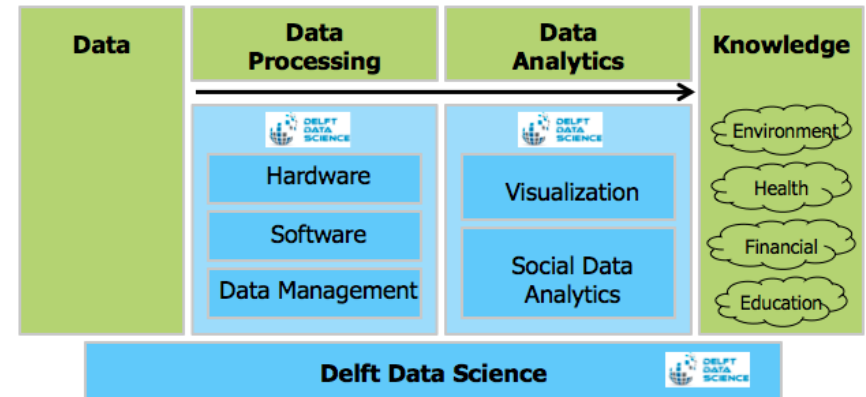
a new scientific discipline for  
scientific **understanding** and  
creating **technology** for  
how to create, process, and  
understand digital data



# Data Science for advancing technology

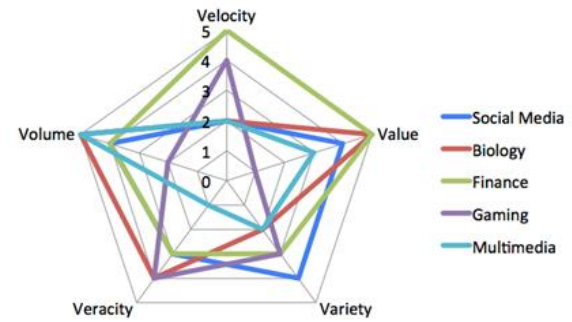


TU Delft coordinating initiative for research, education and training in data science and technology

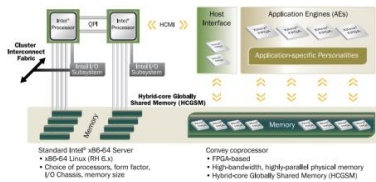


# Delft Data Science – research & education for technology & talent

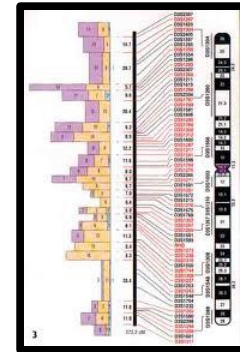
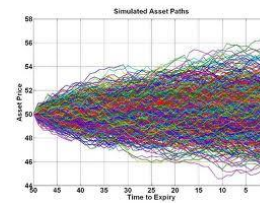
# Hardware for Data Science



**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

## Software for Data Science

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

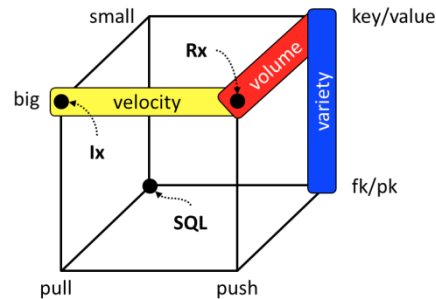
**Solution:** programming languages that abstract from hardware, close to domain experts

**Problem:** data engineers and scientists not trained as software engineers

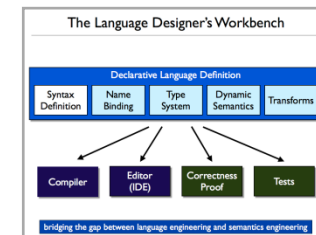
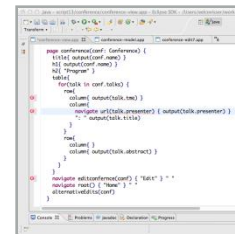
**Cloud Programming:** composing computations using mathematically solid foundations

reactive extensions

interactive extensions

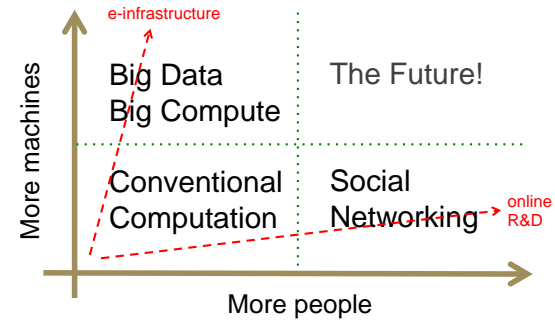


**Domain-Specific Languages:** enabling software engineers to systematically design & apply DSLs



# Enabling programmability of big data analytics

# Data Management for Data Science



## Graph Data Processing: processing graphs at big data and web scale



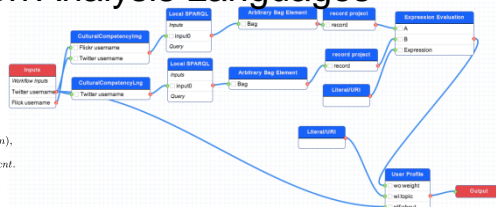
Web-Scale Graph Indexing

### Declarative Graph Analysis Languages

PageRank (iteration  $i + 1$ )

```

int N = 4847571, # # of nodes in LiveJournal data
EDGE (int src: 0..N, int sink);
EDGECOUNT (int src: 0..N, int cnt);
NODES (int n: 0..N);
RANK (int iter: 0..10, (int nodes: 0..N, int rank));
RANK(i + 1, n, SSUM(r)) := NODES(n), r = 0.15/N;
                        := RANK(i, p, r1), EDGE(p, n),
                        EDGECOUNT(p, cnt),
                        cnt > 0, r = 0.85 * r1 / cnt.
    
```

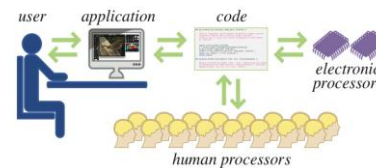


Optimizing Big Data Processing Workflows

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



Crowdsourcing and Human  
Computation



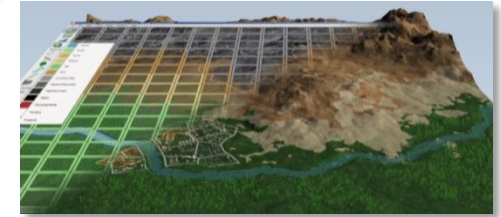
Crowd Capacity

Data Generation & Data Curation

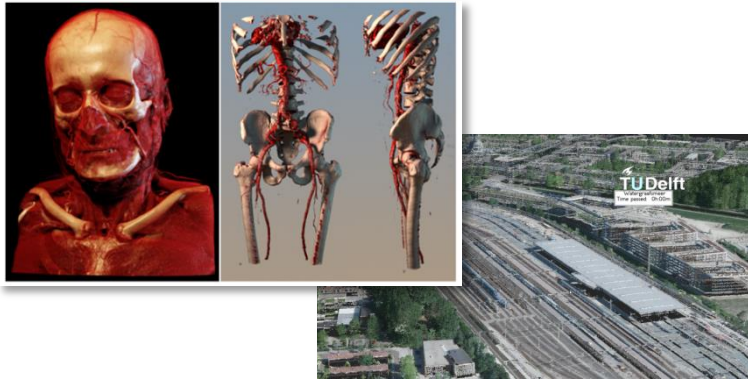
# Enabling big data management at scale and with human interpretation



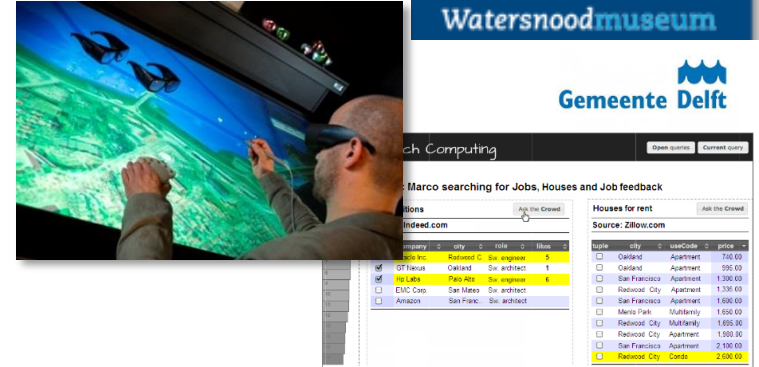
# Visualisation for Data Science



**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



## Science of Social Data

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

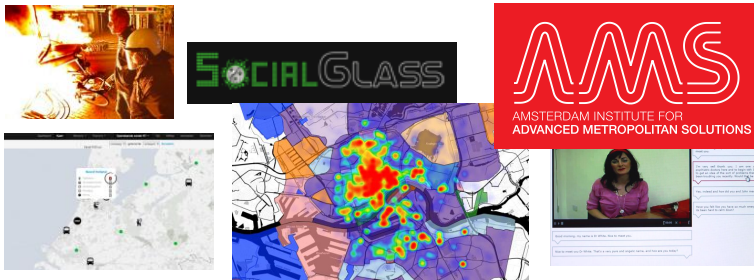


**Challenge:** largest source  
ever made, with yet-to-  
discover semantics



**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

# Example: Intelligent Cities



## **Urban analytics**

*with Amsterdam Institute for  
Advanced Metropolitan Solutions*

*seeing the city through  
**open** social urban data  
for modeling city flows*

*data can tell different stories from  
different perspectives  
and different **semantics***

***engaging** people in collecting data  
about their living environment, and  
**complementing** physical sensing  
and 'official' data*

# 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***

**Playback Selection**

- Good morning, my name is Dr White. Nice to meet you.
- I'm very well thank you, I am one of the psychiatric doctors here and to begin with I'd like to get an idea of the sort of problems that have been troubling you recently. Would that be ok?
- Yes, indeed and how did you and John meet?
- Have you felt like you have so much energy that its been hard to calm down?

Good morning, my name is Dr White. Nice to meet you.

Nice to meet you Dr White. That's a very pure and angelic name, and how are you today?

**Handshake**  
From Wikipedia, the free encyclopedia

A **handshake** is a short ritual in which two people grasp one of each other's hands, in most cases accompanied by a brief and down movement of the grasped hands.

**Summary links**

- 1 History
- 2 Related subjects
- 3 History
- 4 See also
- 5 References
- 6 External links

**History** [edit]

Revision	Summary
1	Handshake
2	Handshake
3	Handshake
4	Handshake
5	Handshake
6	Handshake
7	Handshake
8	Handshake
9	Handshake
10	Handshake
11	Handshake
12	Handshake
13	Handshake
14	Handshake
15	Handshake
16	Handshake
17	Handshake
18	Handshake
19	Handshake
20	Handshake
21	Handshake
22	Handshake
23	Handshake
24	Handshake
25	Handshake
26	Handshake
27	Handshake
28	Handshake
29	Handshake
30	Handshake
31	Handshake
32	Handshake
33	Handshake
34	Handshake
35	Handshake
36	Handshake
37	Handshake
38	Handshake
39	Handshake
40	Handshake
41	Handshake
42	Handshake
43	Handshake
44	Handshake
45	Handshake
46	Handshake
47	Handshake
48	Handshake
49	Handshake
50	Handshake
51	Handshake
52	Handshake
53	Handshake
54	Handshake
55	Handshake
56	Handshake
57	Handshake
58	Handshake
59	Handshake
60	Handshake
61	Handshake
62	Handshake
63	Handshake
64	Handshake
65	Handshake
66	Handshake
67	Handshake
68	Handshake
69	Handshake
70	Handshake
71	Handshake
72	Handshake
73	Handshake
74	Handshake
75	Handshake
76	Handshake
77	Handshake
78	Handshake
79	Handshake
80	Handshake
81	Handshake
82	Handshake
83	Handshake
84	Handshake
85	Handshake
86	Handshake
87	Handshake
88	Handshake
89	Handshake
90	Handshake
91	Handshake
92	Handshake
93	Handshake
94	Handshake
95	Handshake
96	Handshake
97	Handshake
98	Handshake
99	Handshake
100	Handshake

*Previous semester:*

*DDS seminar on Data Science for Environmental Monitoring*

*DDS seminar on Data Science for Finance*

*DDS seminar on Data Science for Health*

*Next quarter:*

*DDS seminar on Data Science for Online Education*

*DDS seminar on Data Science for Open (Government) Data*

*Today:*

*DDS seminar on Data Science for Cybersecurity*

# Delft Data Science Seminars

Cybersec@TUDelft

Data Science & Cybersecurity





# Cybersec@TUDelft mission



“Being an *open* **collaboration platform** for scientists, BSC-, MSc- and PhD-students interested in (*research & education* in) the domain of **Cyber Security**”

# Sections connected

[EWI/CYS](#): section cybersecurity (Van den Berg/Lagendijk)

[TBM/POLG](#): policy, organisation, law, and gaming (Van Eeten)

[EWI/NAS](#): network architectures and services (Van Mieghem/Kooij)

[TBM/ICT](#): cybersecurity within ICT (Marijn/Tan/Vanden Berg)

[EWI/SERG](#): software eng. research group (Van Deursen, Visser)

[TBM/SystemEng](#): systems engineering (Brazier, Verbraeck)

[EWI/PDS](#): parallel & distributed systems (Sips, Epema)

[TBM/SSS](#): safety & security science (Van Gelder, Reniers)

[TBM/Philosophy](#): ethics/philosophy of technology (van den Hoven)

Support by Central Board (~ safety & security science)



# Cybersec@TUDelft cooperation (national)

- Cooperation with other universities
  - LDE Center Safety & Security
  - 3TU partners
  - The Hague University of Applied Sciences
- Partnering within “golden triangle”  
(like partners of The Hague Security Delta = HSD)

# New education initiatives, ex.'s

- 3TU regular MSc program CyS  
(starting Sept 2015)
- Professional MSc program CyS  
(at Cyber Security Academy The Hague, started January 2015)
- Other arrangements
  - MOOC on Cyber Security Economics
  - Applied crypto course (continuation of TopTech course)
  - Other dedicated courses
  - ...

So ?!

Interested in Cyber@TUDelft ? →



**Please contact us during the breaks!**

# Agenda

Cybersec@TUDelft

Data Science & Cybersecurity



# Data Science & Cybersecurity challenges

## Two big research directions:

### 1. **Cyber Security** through (big) data:

Monitoring & Analytics for securing **cyber space**, e.g.

of data flows, data processing in devices, ...

in social networks, dark web, ...

of financial e-transactions

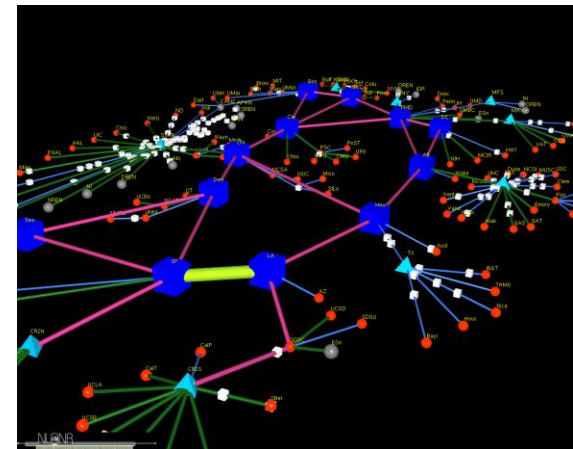
### 2. **Securing Big Data**:

Big Data for improved **decision making**, e.g.

marketing, healthcare, finance, ...

environment, sustainable future, ...

individuals, companies, politics,



# Cyberspace defined

A **global, man-made 'nervous system'** with

high speed *connectivity*

huge distributed *data processing* capabilities

huge *data storage* capabilities

millions of '*intelligent systems*' autonomously taking decisions

+/- 3 billion interconnected *human actors*

**enabling "cyberacting"**

# Cyber activities of all kind...



## *Basic cyber activities*

*Communication, information retrieval, watching, listening, ...*

## *More advanced cyber activities*

*Searching, (automatic) transacting, social gathering, gambling, educating, monitoring & surveillance, supporting business processes of all kind, controlling critical infrastructures, ...*

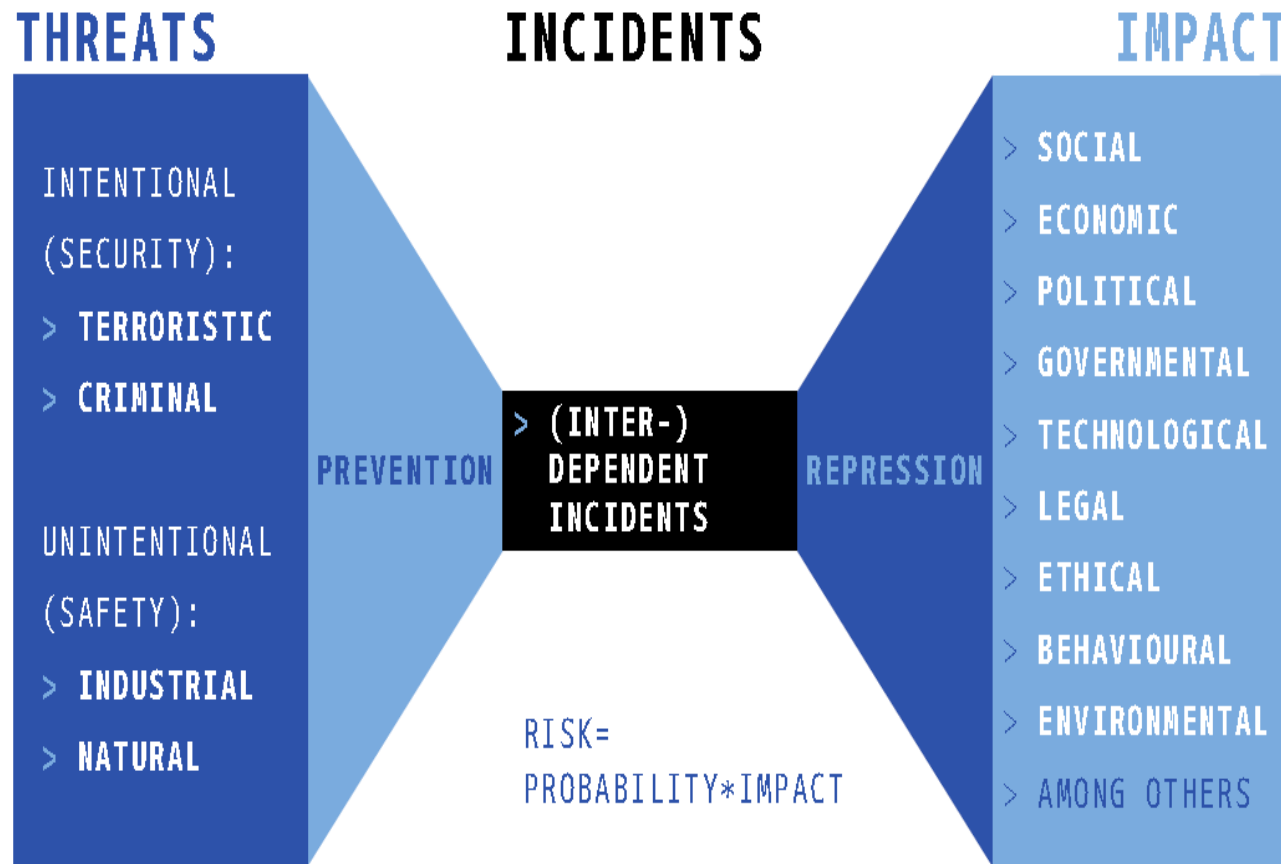
## *Less favourable cyber activities*

*Cyber crime (dark markets), cyber warfare, ...*



# Goal: reduction of cyber risks to 'acceptable levels'

for all stakeholder groups



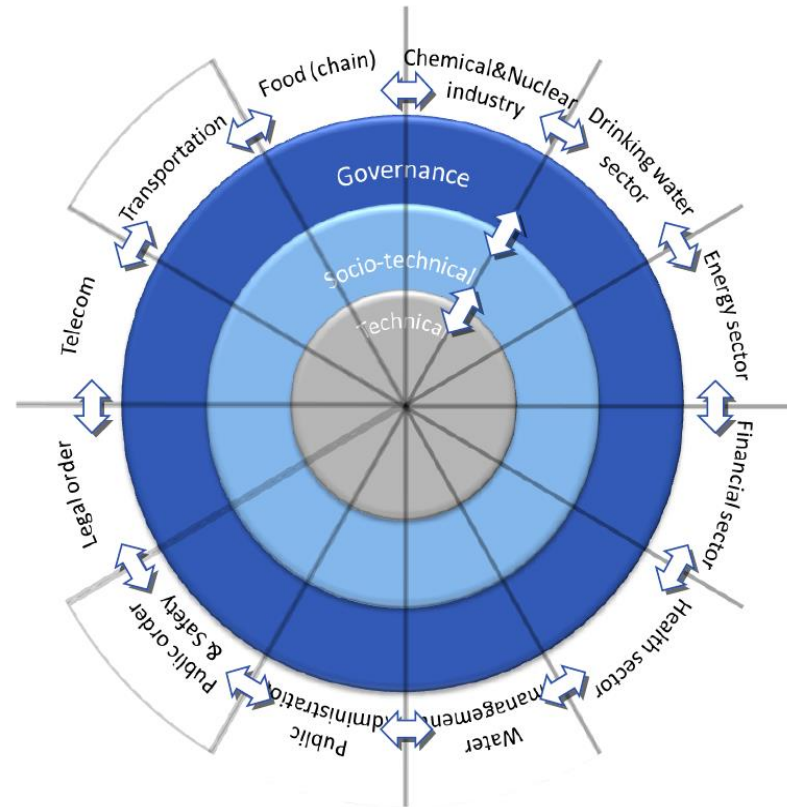
# Decomposing cyberspace: layers

***Technical layer:*** **IT services** (Internet, applications, servers, cloud, IoT, BYOD, ..., many different devices), **World of information security** ~ **CIA** ~ **information security breaches**

***Socio-tech layer:*** **cyber activities** by **3 billion end-users** and (?) **billion autonomous IT systems**, at home/work, in organisations, critical infrastructures, states, ... ~ **'Business/Societal Values'** ~ **cyber activity breaches (!)**

***Governance layer:*** governance and management activities

- **Stakeholder groups:** end-users/organisations/sectors/states, ...



# Consequences

- *Cybersecurity* concerns securing cyberspace where
  - Technical layer *enables* cyber activities
  - Governance layer *steers (& constraints)* cyber activities
  - *Cyber activities* focus on the *complex interaction of (autonomous) IT-systems and their users*
- **Proposition: in the end, any fruitful discussion on cyber security should pay attention to all 3 layers!**

# 2. Securing big data

- Classical *information security* criteria  
CIA(A)
- Also *protecting privacy* in all big data application areas like ...
  - ➔ Privacy enhancing technologies
  - ➔ Homomorphic encryption
  - ➔ Enabling personal control over data
  - ➔ Enabling anonymity (TOR, TRIBLER)
  - ➔ Enforcing the right to become forgotten
  - ➔ ...

# Master classes

- *Ad 2: Securing (big) data:*  
***Security and Privacy: Conflict or Harmony***  
***Zekeriya Erkin***
- *Ad 1: Securing cyberspace through big data:*  
***The Challenge of Bot Detection***  
***Sicco Verwer and Pieter Burghouwt***
- *Ad 1 & 2:*  
***Why have a lock at all? On the fallacy of ICT security***  
***Christian Doerr***

# Thank you

After the masterclasses the plenary program will  
start again right here in Hall Boole at 15.00  
sharp