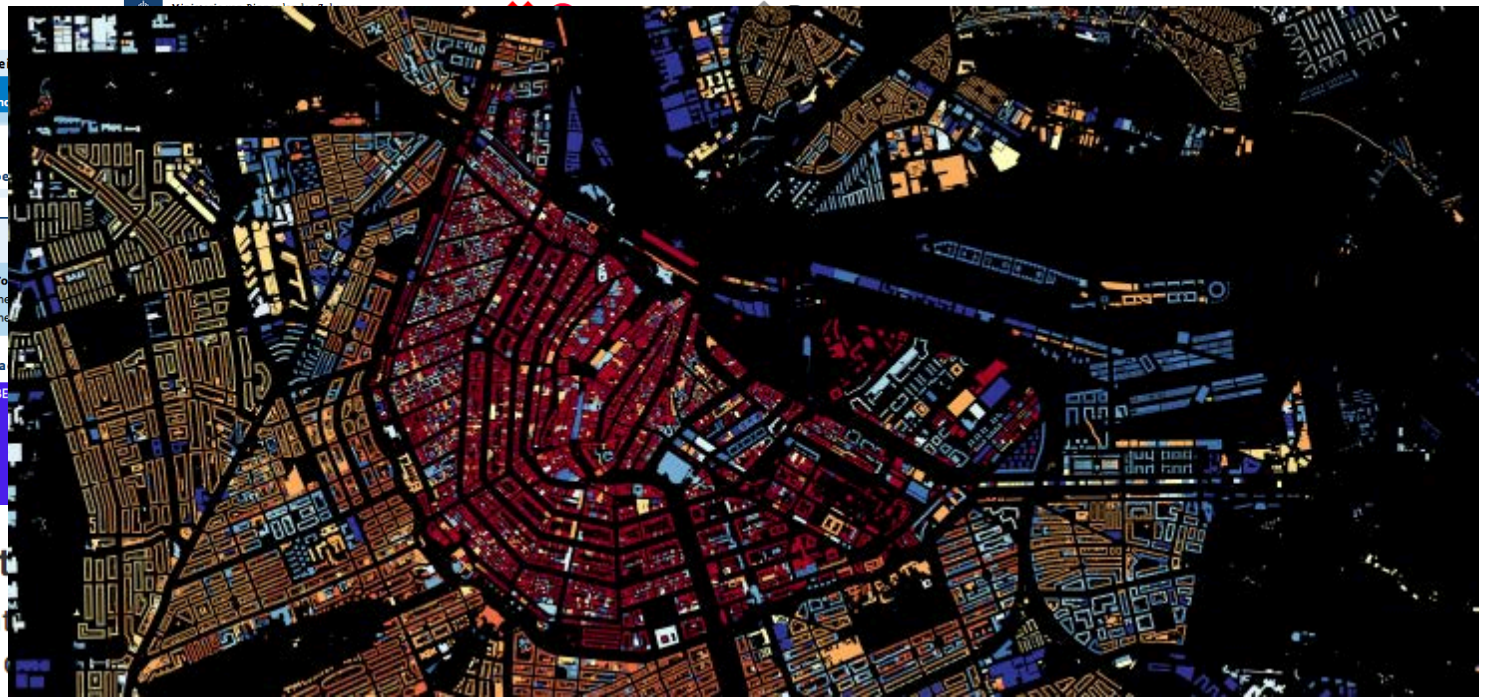
Privacy
Who cares and Who should care?

Big Data and Privacy
Seaside Match Making Event, October 15, 2016

Assist. Prof. dr. Zeki Erkin

Cyber Security Group
Delft University of Technology

Open Data



Source: http://dev.citysdk.waag.org/buildings/

What can go wrong?

Sito Veracruz, Lilly Lam

Makkie klauwe

I like it | Share | Like 1 | Tweet

Makkie klauwe confronts citizens by putting their properties in danger. The app combines public data so that thieves see the location where they are best able to steal specific property. Citizens are thus awakened and forced to think about the role of open data in the city.

For more information see this article

Tag people Enlarge

makkieklauwe

Privacy

- Case 1: Service provider is not trustworthy
  - process data for other purposes
  - sell data to third parties

- Case 2: Service provider is trustworthy
  - corporation take-over/bankruptcy
  - law and regulations prohibit storing sensitive data (medical data)
  - physical security/forgetful employees
  - competing service providers

Questions…

- From whom should we protect our data?
- How can we protect data?
- What are the best practices?
- How can we keep the balance between privacy and utility?

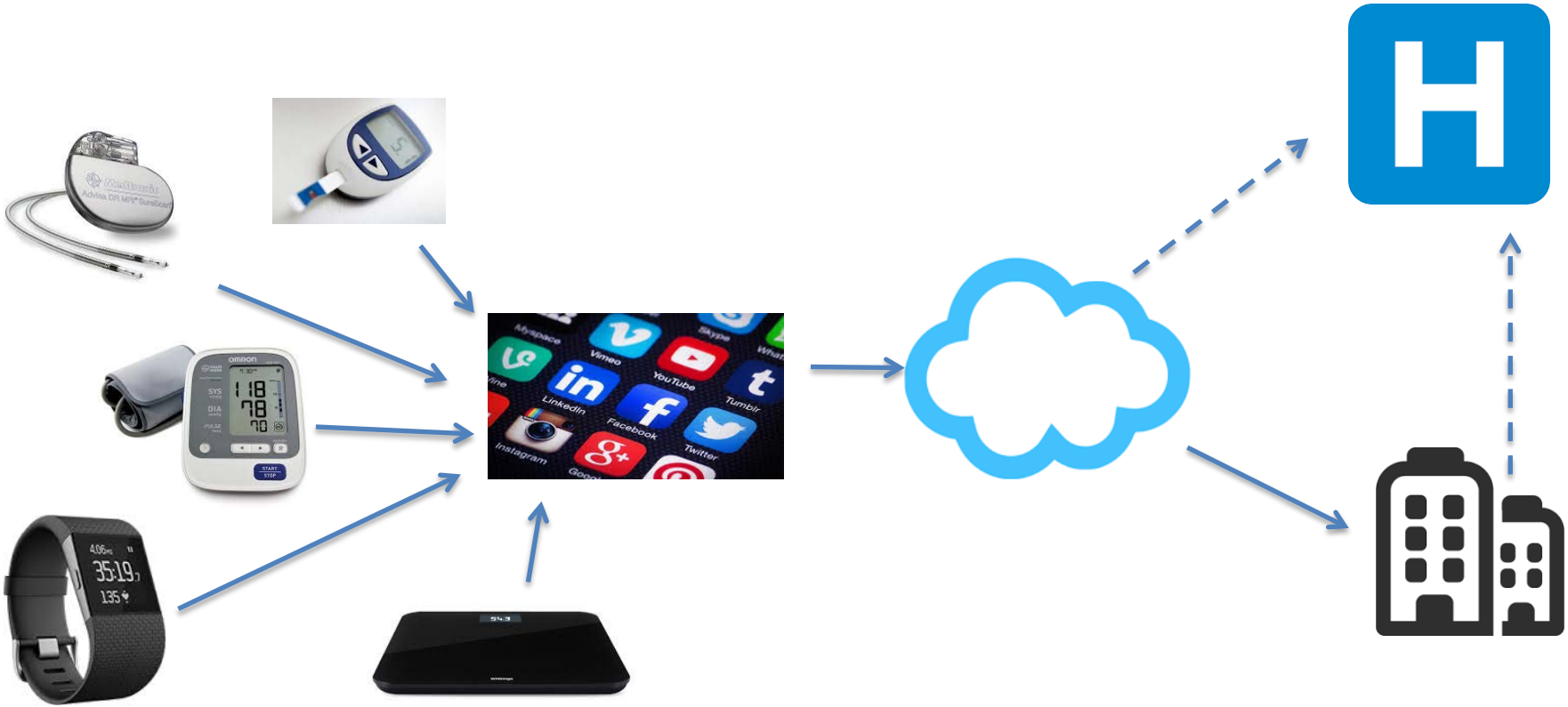
Solution

- Legislations and regulations
- Physical security
- Access control
- But not enough…

Big Data and Finance

- Machine learning for predictive models
  - Explicit and implicit data collected from different resources
  - Application specific parameter and indicators
- The more data, the better results
  - Fraud detection
  - But banks cannot combine datasets!

Big Data and Well-being



Challenges

- Multi-players
- Trust
- Limitations (medicine)
- Geographical distance (logistics)
- Real-time services

Scientific Approach

- Anonymization
  - Generalisation
  - Differential Privacy

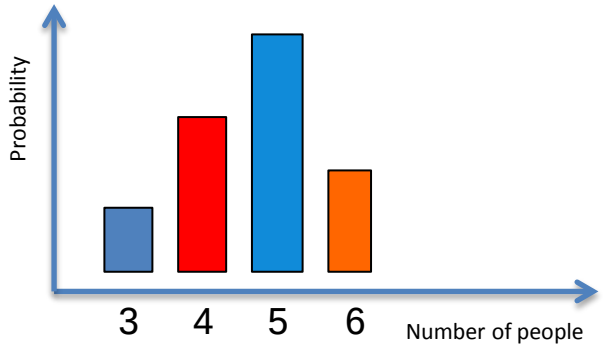
- Provably Secure Systems using Cryptography
  - Security by Design
  - Privacy by Design

Anonymization

- Identifiers: Remove them
- Quasi-Identifiers: Generalize/Supress
  - K-anonymity
  - L-diversity
  - T-closeness
- Easy to de-anonymize [using big data]

Differential Privacy

- Query results is published, e.g. 5 people
- But now we output a number of outcomes with certain probabilities.



- And if you leave the population, the outcome does not change significantly

Cryptographic Approach

- Based on cryptographic tools
- Provably secure
- Custom-design
- Overhead

And…

- Privacy is important
- We all should care
- Privacy should be protected!
  - General Data Protection Regulation
  - Penalty 4% of worldwide turnover