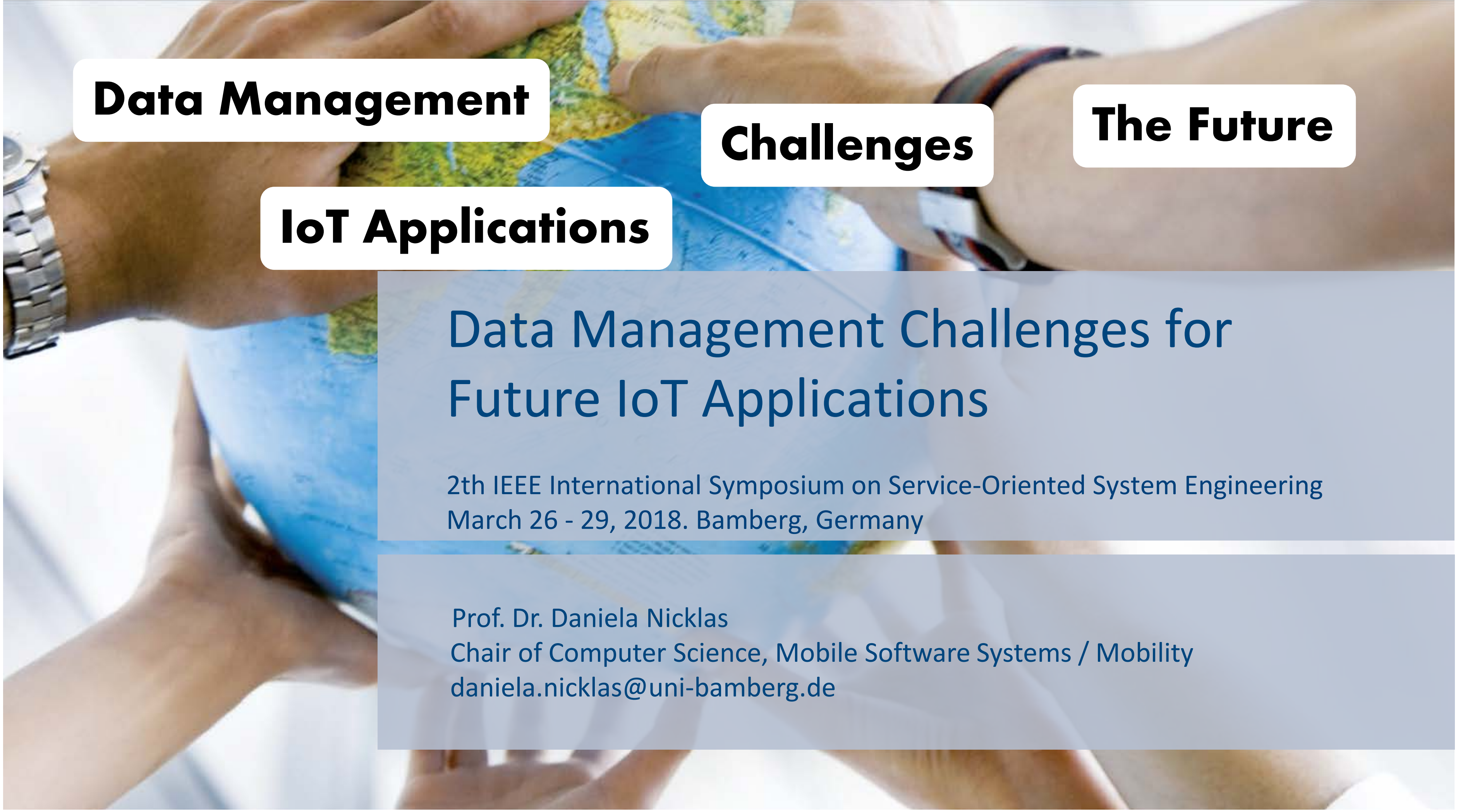
Universität Bamberg



Data Management

Challenges

The Future

IoT Applications

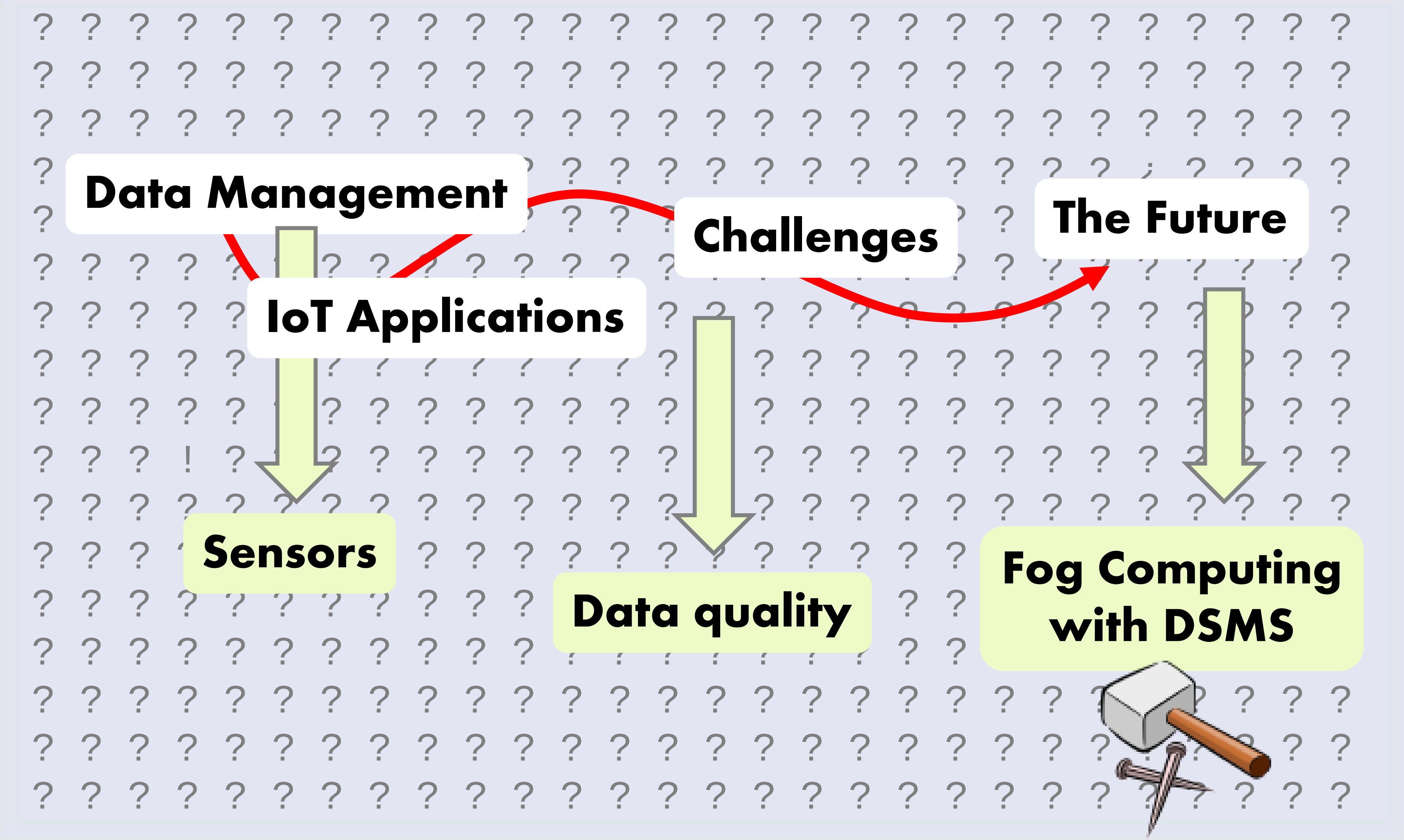
Data Management Challenges for Future IoT Applications

2th IEEE International Symposium on Service-Oriented System Engineering
March 26 - 29, 2018. Bamberg, Germany

Prof. Dr. Daniela Nicklas
Chair of Computer Science, Mobile Software Systems / Mobility
daniela.nicklas@uni-bamberg.de

Overview

Universität Bamberg



Data Management

IoT Applications

Challenges

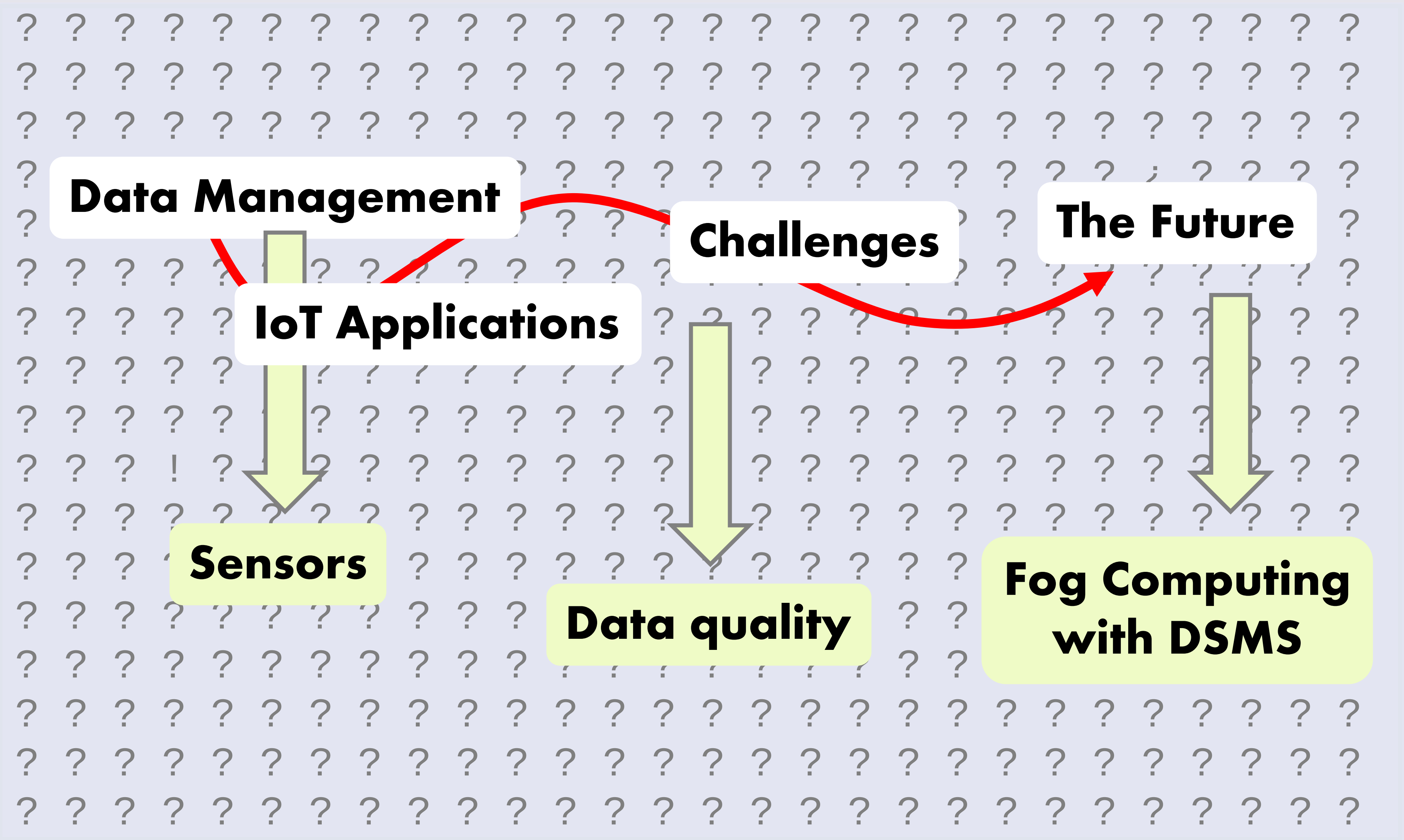
The Future

Sensors

Data quality

Fog Computing with DSMS

Overview



Data Management

Challenges

The Future

IoT Applications

Sensors

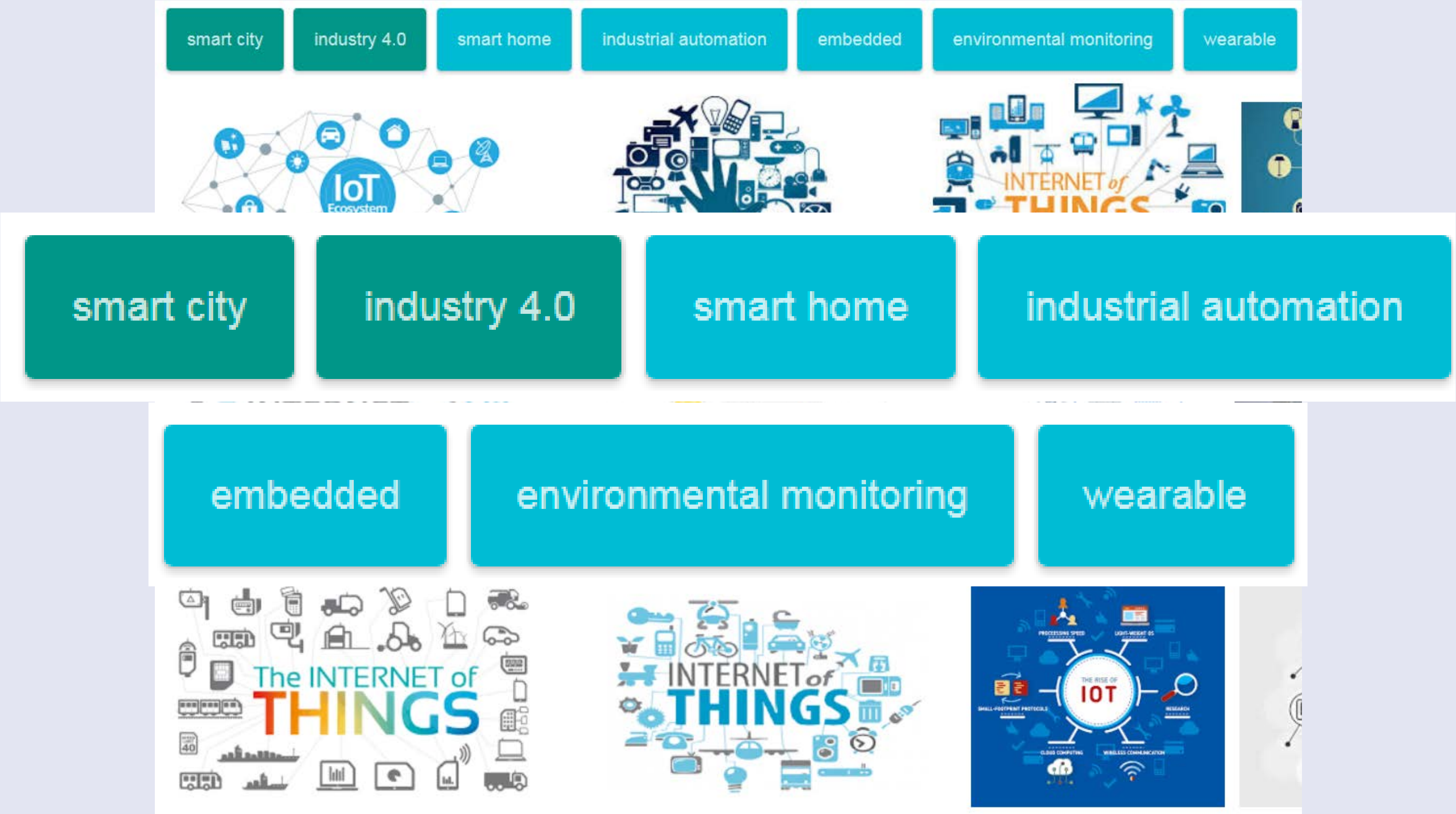
Data quality

Fog Computing with DSMS

OK Google – What is IOT?

Universität Bamberg

https://www.google.com/search?q=iot (images)



smart city | industry 4.0 | smart home | industrial automation

embedded | environmental monitoring | wearable

IoT Applications

for X in …

- **Meter**
- **Grid**
- **Factory**
- **Home**
- **City**
- **Phone**
- **Transportation**

X

+ sensors

+ magic

= SmartX

Realized by: IoT

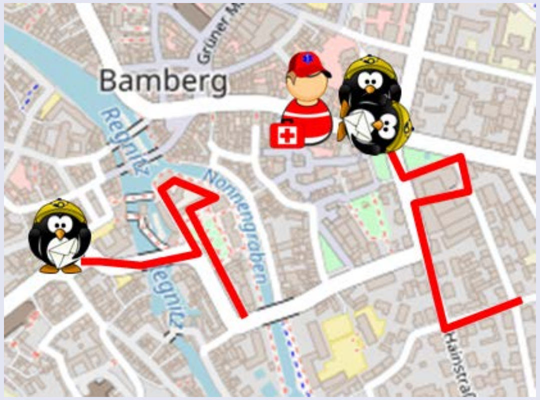
Similar challenges in pervasive computing, ambient intelligence, physical computing, …

Sample FutureIOT applications

FutureIOT

- Pan-Bavarian research projects, 02/2018-01/2021
- 10 research partners, >20 industry partners
- Smart city and smart agriculture applications, e.g.:

Long-term evaluation and new services with inductive parking sensors



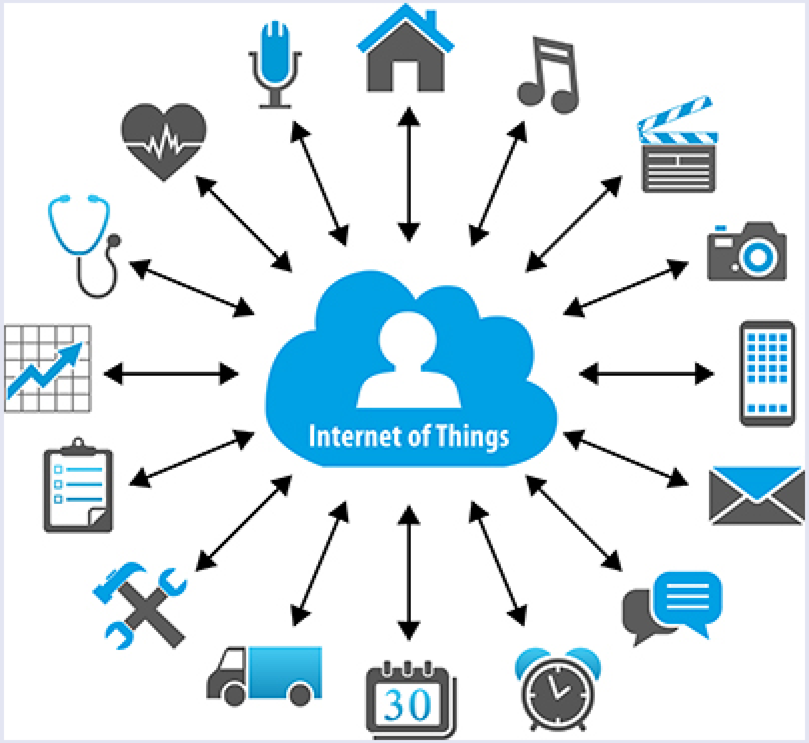
Privacy-aware delivery tracking (with safety and security features)

Environmental sensing – stationary and mobile

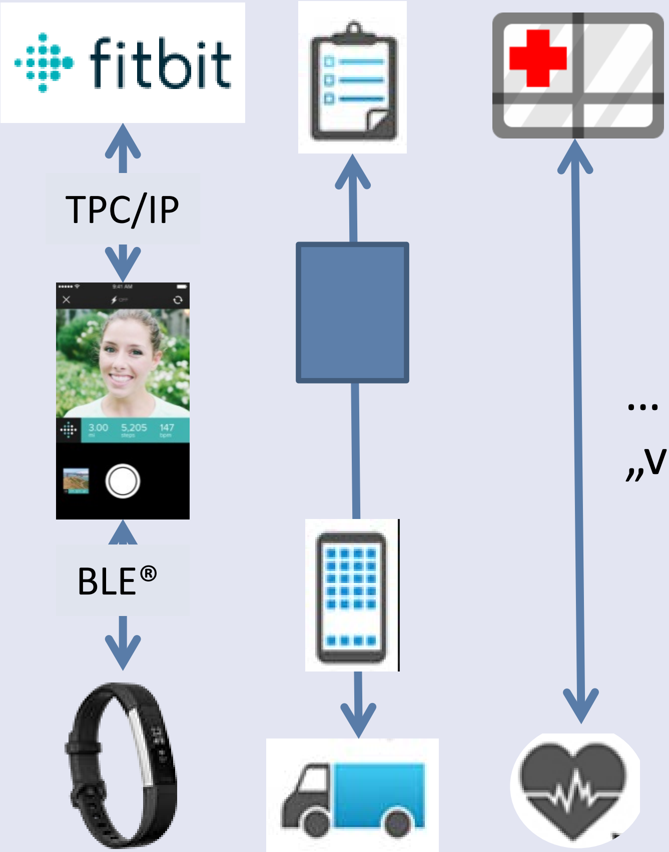
Activity recognition for cows (to detect anormal behaviour)

IoT Architecture

Vision



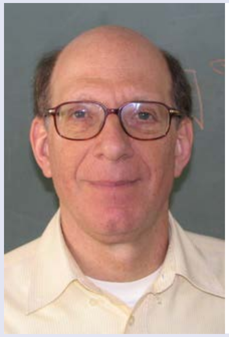
Reality



… many „verticals“

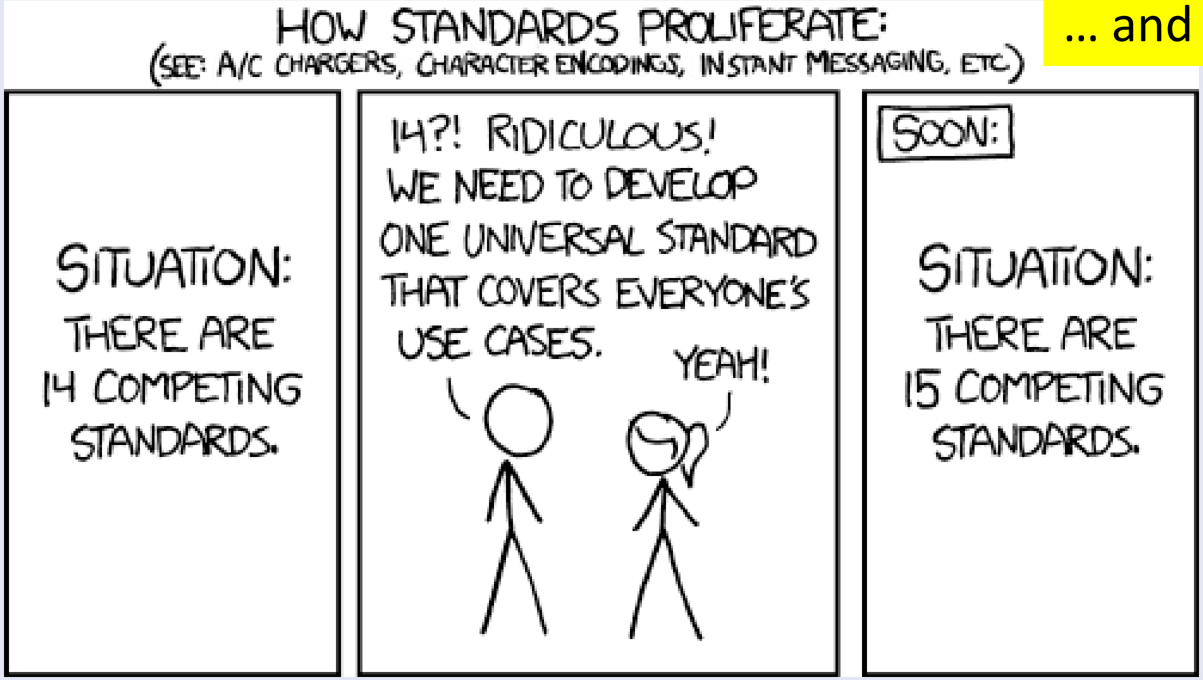
Better architectures?

- Shouldn‘t we help with a standard?



„The nice thing about standards is that you have so many to choose from.“

A. Tanenbaum, *Computer Networks*, 2nd ed., p. 254



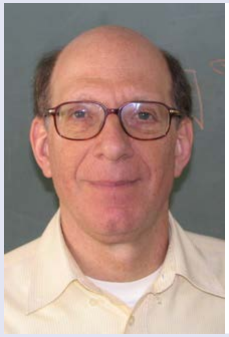
HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.)
SITUATION: THERE ARE 14 COMPETING STANDARDS.
14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.
YEAH!
SOON:
SITUATION: THERE ARE 15 COMPETING STANDARDS.

… and IoT!

https://xkcd.com/927/

Better architectures? *AAS?

- Shouldn‘t we help with a standard?
- From a systematic mapping study [1] on 35 studies on IoT and Cloud:
  - 15 provide Software as a Service (SaaS)
  - 13 provide Platform as a Service (PaaS)
  - 10 provide Infrastructure as a Service (IaaS)
  - 1 provides Network as a Service (NaaS)
  - 2 provide Sensing as a Service (SaaS)
  - 2 provide Sensing and Actuation as a Service (SAaaS)
  - 1 provides Smart Object as a Service (SOaaS)



„The nice thing about standards is that you have so many to choose from.“

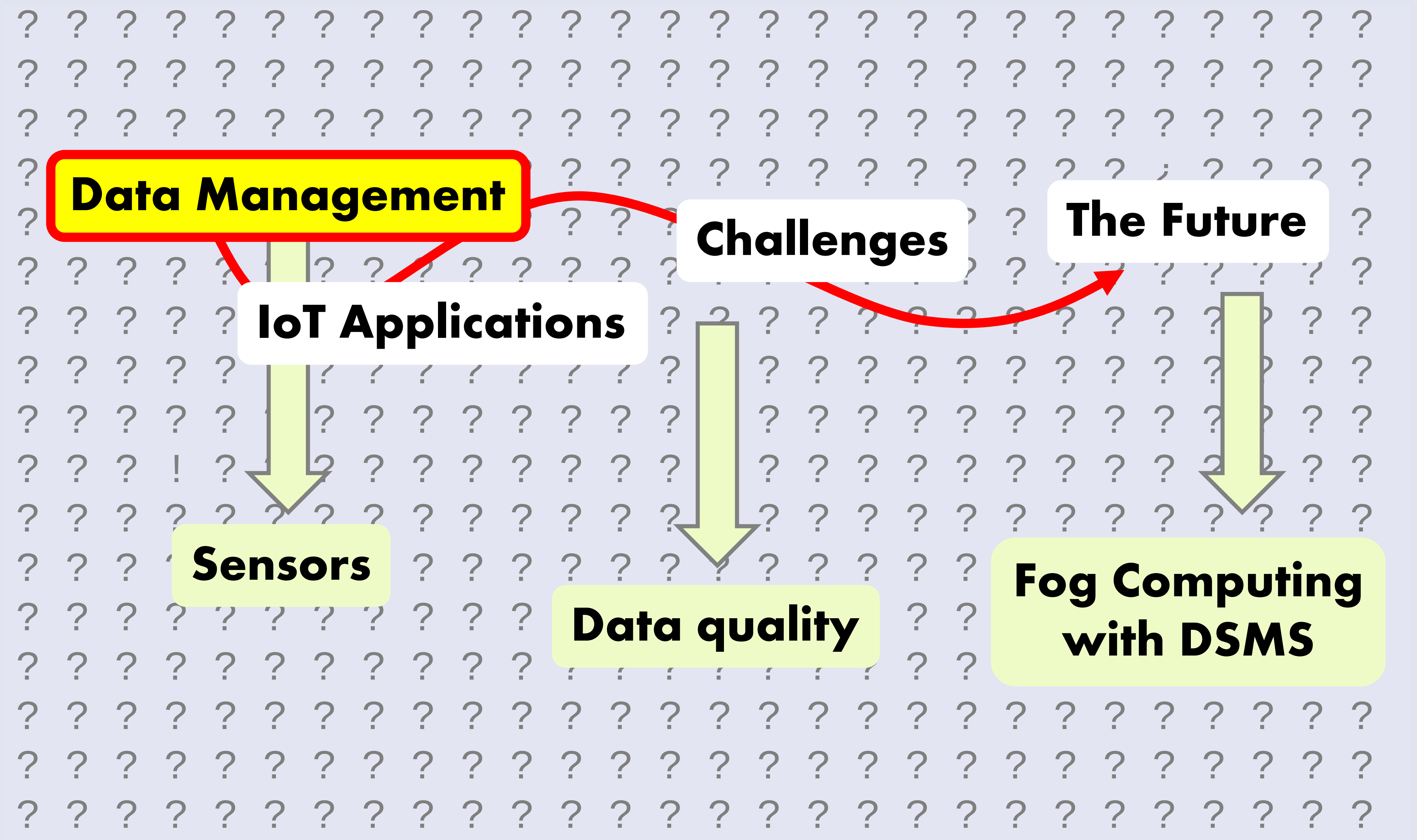
A. Tanenbaum, *Computer Networks*, 2nd ed., p. 254

On 28.3.18, 123 IoT Platforms are listed on Postscapes.

https://www.postscapes.com/internet-of-things-platforms/

[1] E. Cavalcante *et al.*, “On the interplay of Internet of Things and Cloud Computing: A systematic mapping study,” *Computer Communications*, vol. 89–90, pp. 17–33, Sep. 2016.

Overview



Data Management

IoT Applications

Challenges

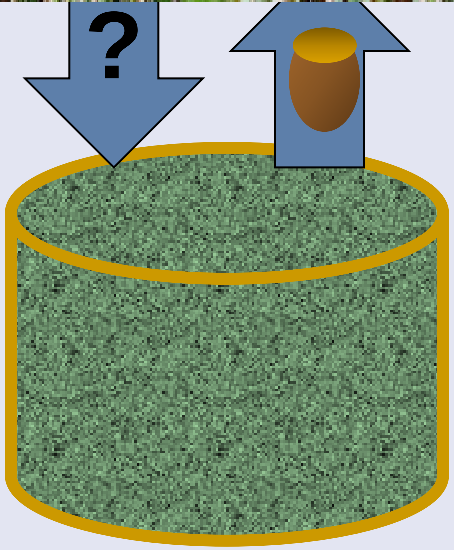
The Future

Sensors

Data quality

Fog Computing with DSMS

Bild: elli60 / pixelio.de



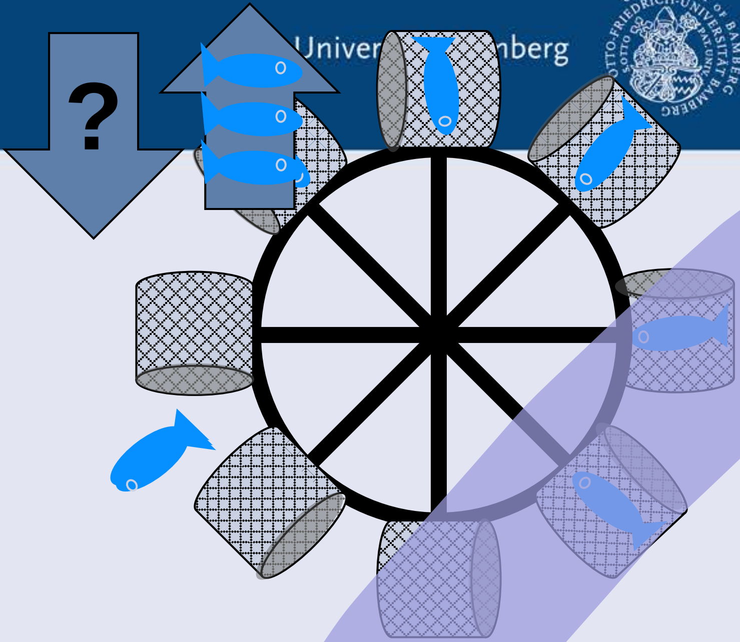


Bild: Ronny Senst / pixelio.de

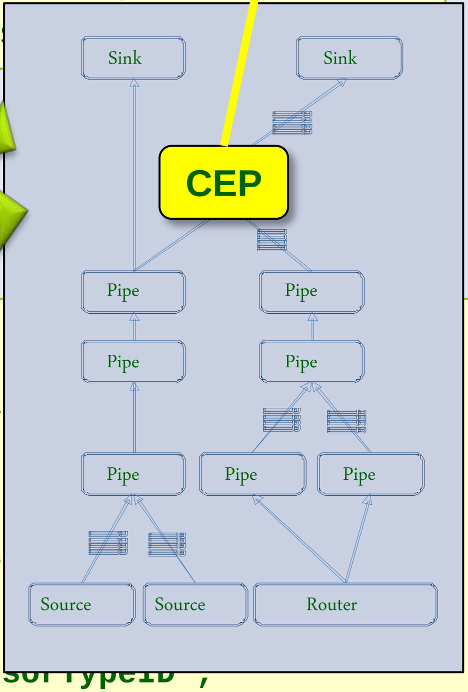
Features of Data Stream Management

- Programming abstraction
  - Declarative: Query
  - Functional: Data flow graph
    - Enables query optimizations
    - Better maintanance of systems

  → Using a DSMS on data streams is like using a DBMS instead of files
- Easy to combine with complex event processing (CEP)
- Parallel execution of operators in graph → no shared memory
- Data streams can be unbounded: Issues with sorting, joins, aggregation
  - Approximate answers
  - Window semantics

```
SELECT ego.pos
RANGE 10 secon
radar RANGE 15
WHERE ego.speed > 30 AND
radar.speed > 30
AND s2.pos
```

Some DSMS provide CEP operators

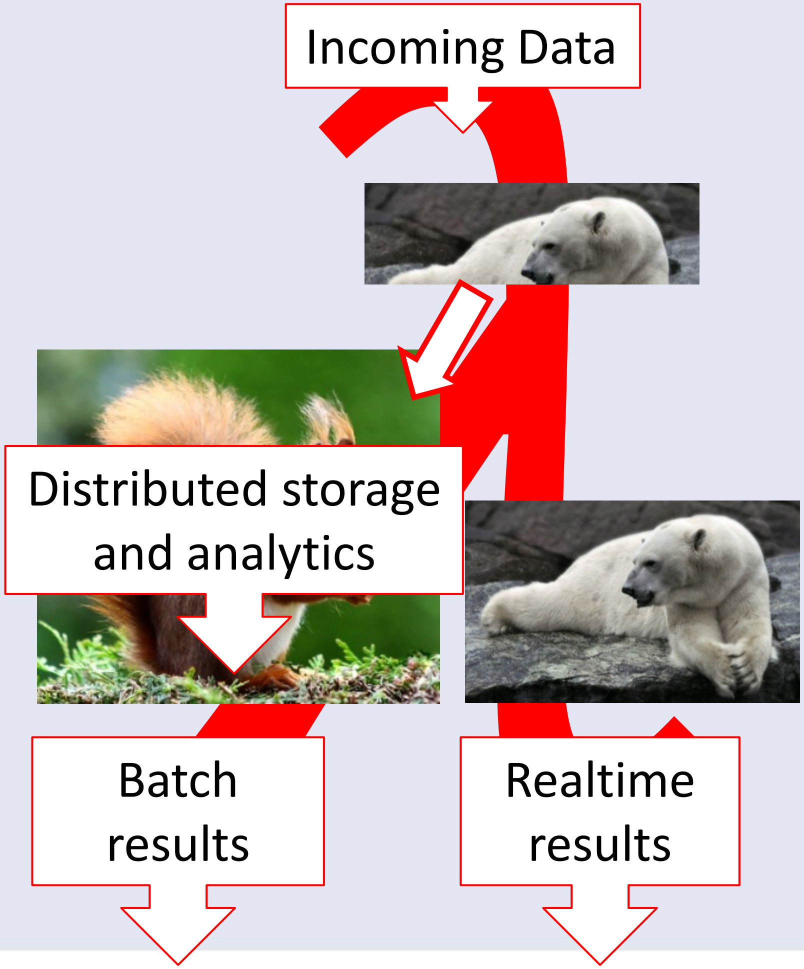


```
stream<uint64 curre
sensorId, uint8 sen
measureValue1, floa
float64 distanceV,
uint8 speed, uint8
direction, uint64 t
Aggregate(HigherGPS

{window HigherGPS :
param groupBy : sen                ,
output MapGPS : avgSpeed = Average(speed)
;}
```

Lambda and Kappa

- Suggested architectures for big data stream processing

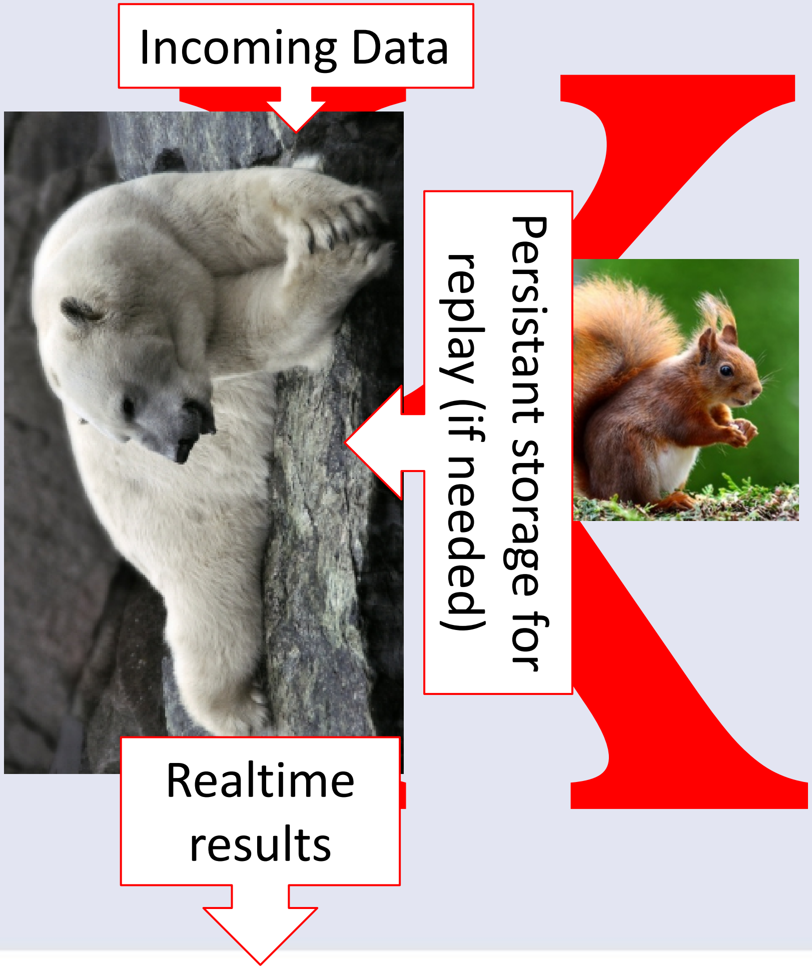


Incoming Data

Distributed storage and analytics

Batch results

Realtime results

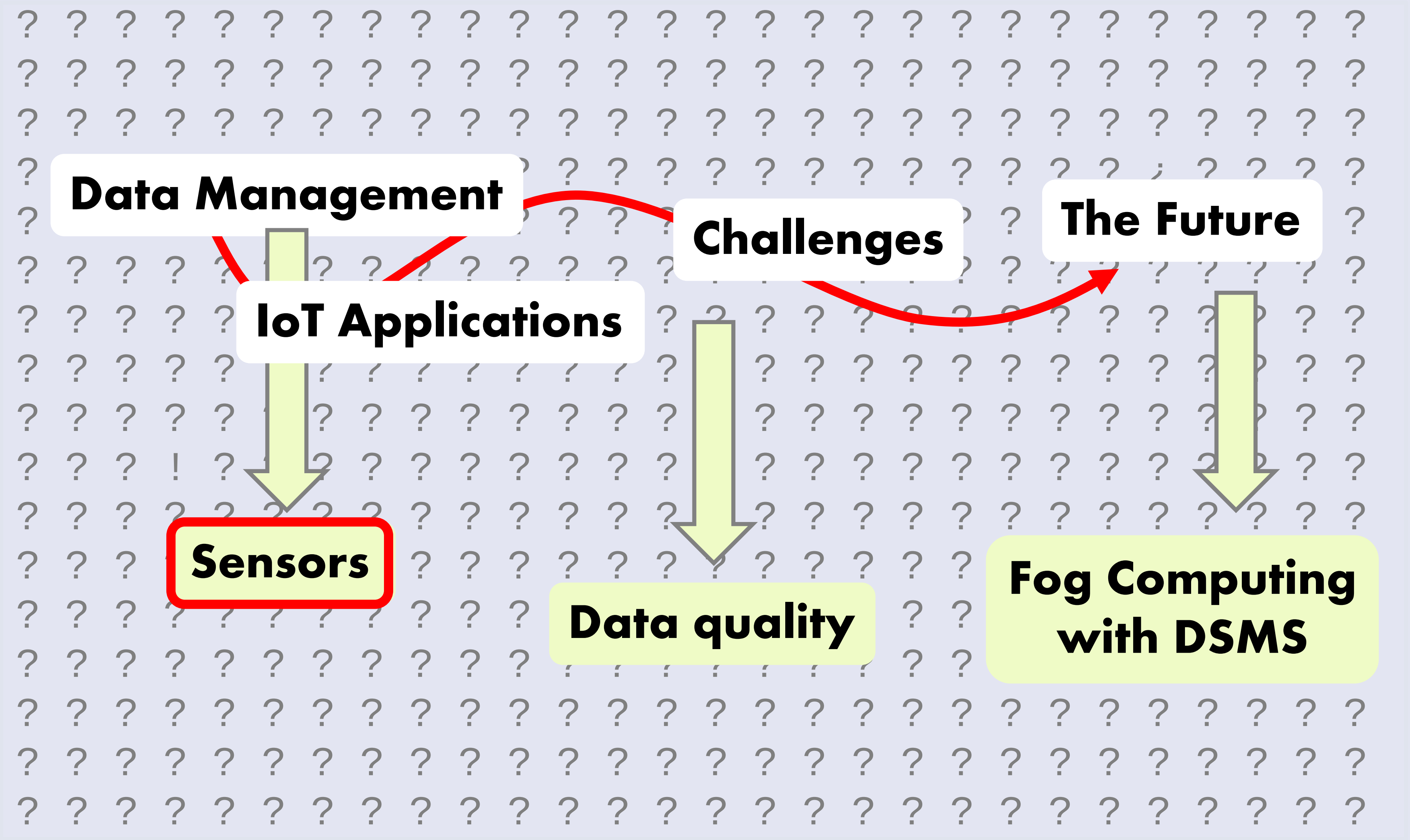


Incoming Data

Persistant storage for replay (if needed)

Realtime results

Overview



Data Management

IoT Applications

Challenges

The Future

Sensors

Data quality

Fog Computing with DSMS

Sensors

A **sensor** is an electronic component, module, or subsystem whose purpose is to detect events or changes in its environment and send the information to other electronics, frequently a computer processor.

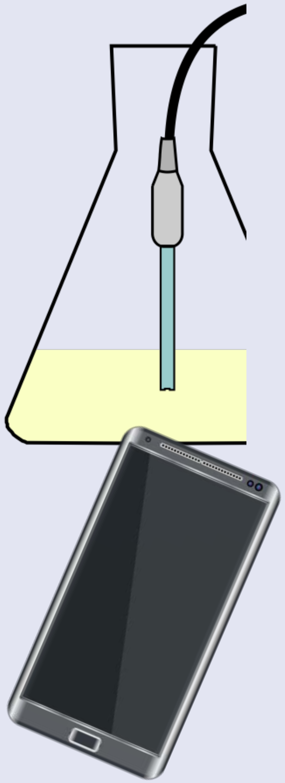
https://en.wikipedia.org/wiki/Sensor

- Technical systems can achieve situational awareness by using data from sensors
- However, sensor data is often …
  - incomplete (not everything can be sensed)
  - late (results do not not arrive in time)
  - inaccurate (values are not exact)
  - mobile (sensed by moving systems)
- To make things worse, sensor data needs to be interpreted

  … and interpretations can cause further errors

Sensor Data

- Sensors implement transfer function:
  - Input: A state of the observed phenomenon
  - Output: A signal (analog or digital) → sensor data
  - Most sensors have a linear transfer function
- Sensitivity of a sensor:
  - How much does the sensor output change when the input changes?
    → Slope of the transfer function
- A sensor system contains:
  - One or many sensors
  - A processing unit (fixed or configurable) to derive data from the sensor signal
  - A communication unit (wired or wireless) to transfer the data to an other system



How to choose a sensor (system)?

Universität Bamberg

- Phenomenon:
  - Physical? (Light, noise, acceleration, radio signals, ..)
  - Chemical? (Substances in gas or fluids)
  - Social? (Behaviour, communication, …)
  - Technical? (Proper operation, …)
- Measurement:
  - Direct or derived?
  - Latency?
- Redundancy:
  - One sensor or many?
  - Same sensors or different?
- Installation:
  - Static or mobile?
  - Wired or wireless?
  - One-hop or multi-hop?
  - Calibration?
- Aging:
  - Battery?
  - Saturation?
  - Re-calibration?
- Cost <-> Quality – Tradeoff!

Types of sensor (system) data

- Format:
  - Structured (e.g. (Timestamp, Value), or (Value, Value, Value))
  - Unstructured (e.g. image stream (video) or audio stream)
  - Semi-structured (e.g. photo + DXF meta data (timestamp, location, resolution, …))
- Semantic levels:
  - Raw: just the signal
  - Feature: a typed attribute of anentity, e.g. the location
  - Object: multiple attributes grouped together for an object
  - Situation: a complex sitation was detected

  → Higher levels are often results of sensor data fusion
- Validity: How long is the sensor value valid?
  1. Only at timestamp
     - if sensor sends with fixed frequency
  2. Fixed until next data comes in
     - if sensor sends when value deviates from last value by threshold
  3. Changing according to model
     - if sensor sends when value deviates from a function of time
     - „dead reckoning“ → often used for moving objects (but can be applied to other phenomena)

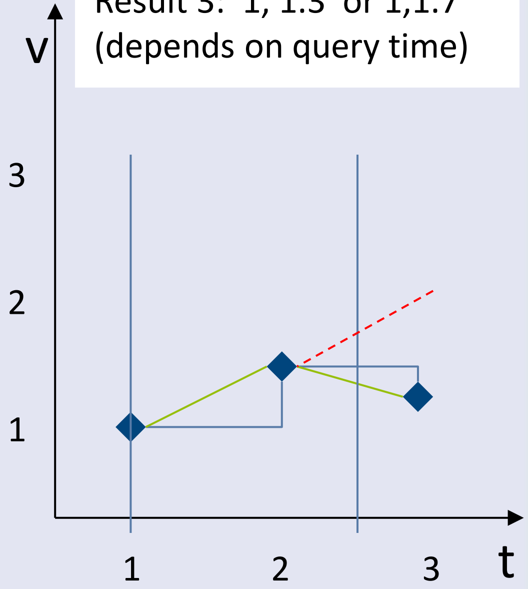
select t, v from sensordata
where t = 1 or t = 2.5

Result 1: 1, NULL

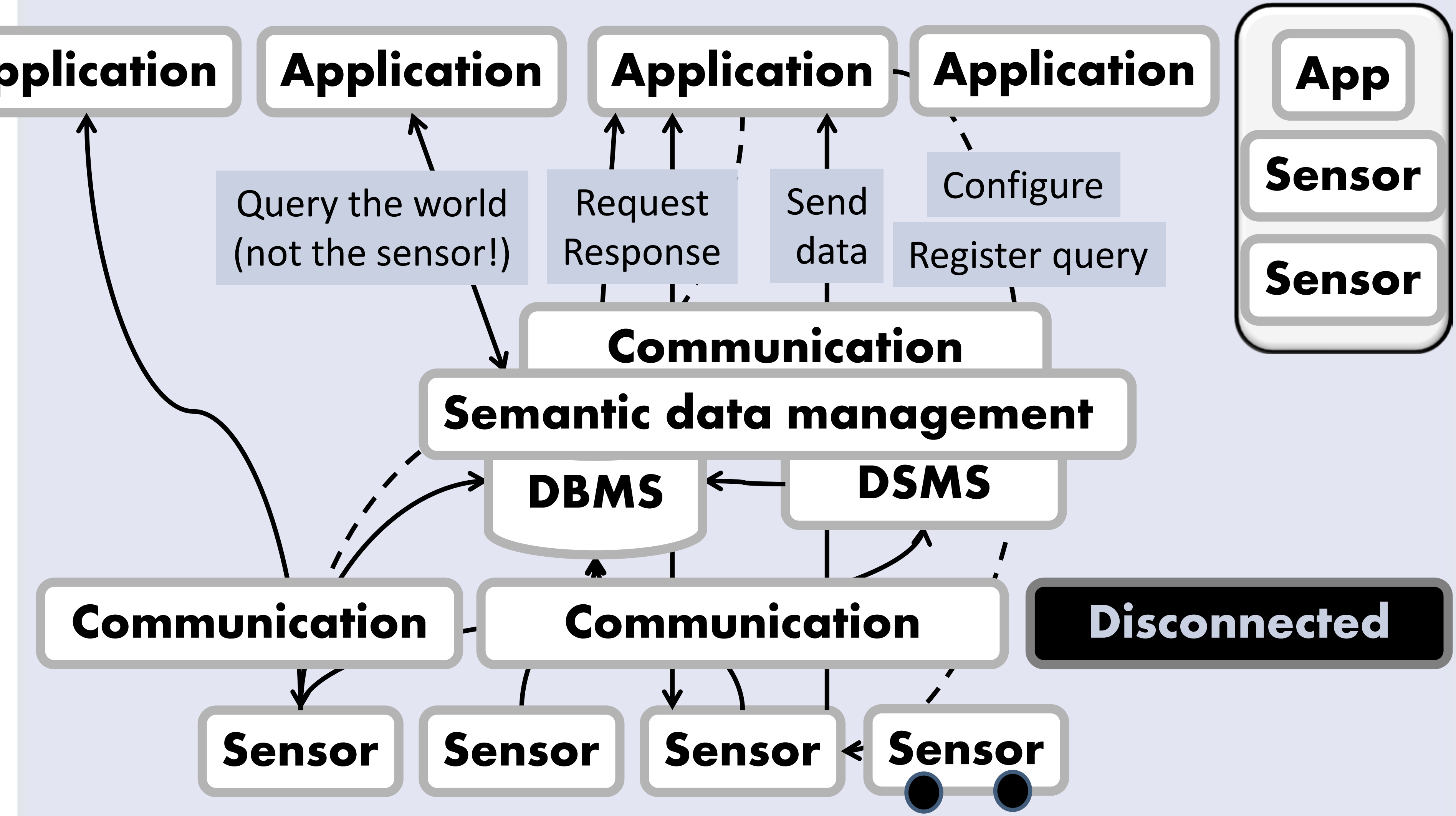
Result 2: 1, 1.5

Result 3: 1, 1.3 or 1,1.7

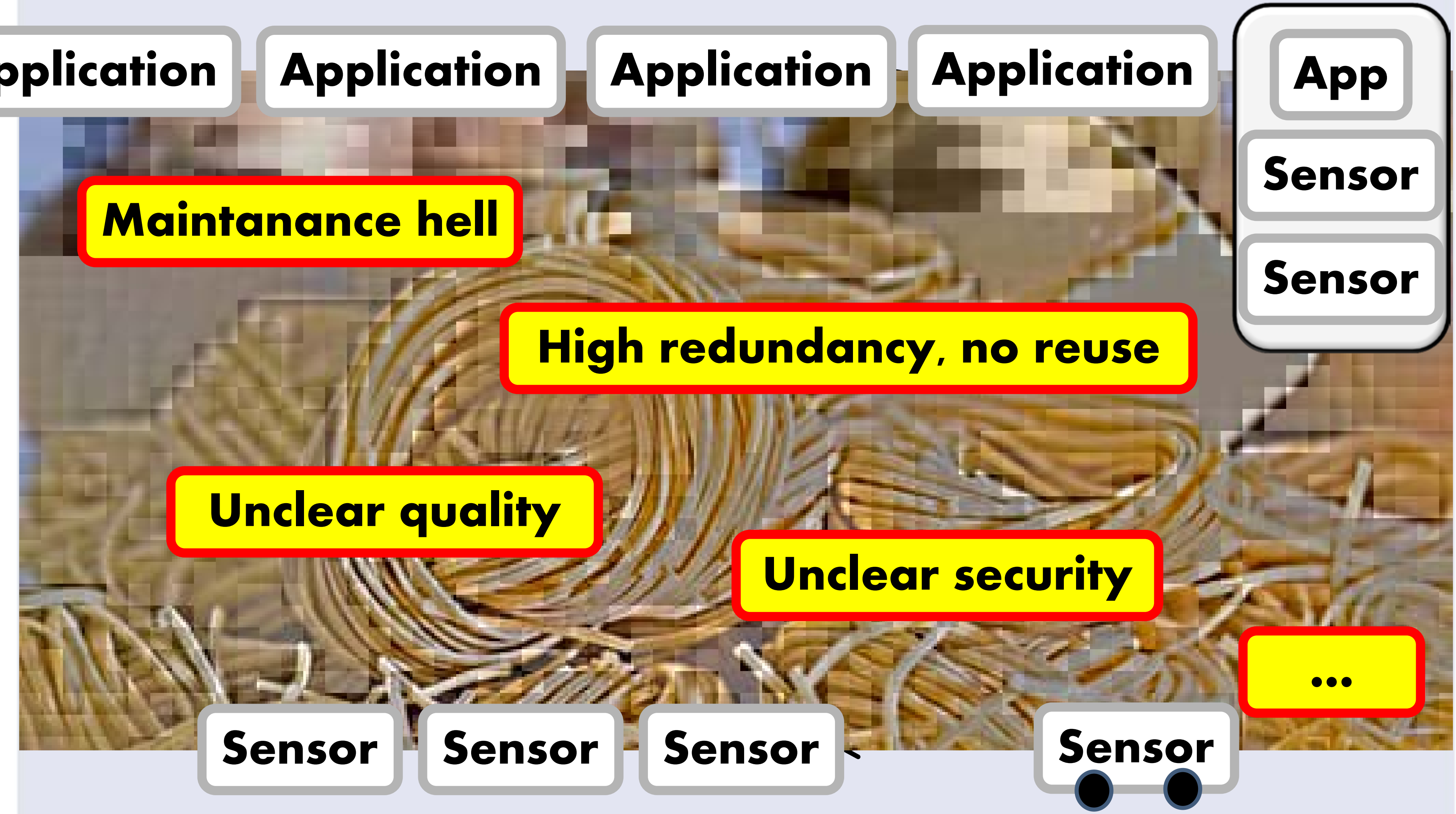
(depends on query time)



Evolution of a sensor-based system



Evolution of a sensor-based system



pplication

Application

Application

Application

App

Sensor

Sensor

Maintanance hell

High redundancy, no reuse

Unclear quality

Unclear security

...

Sensor

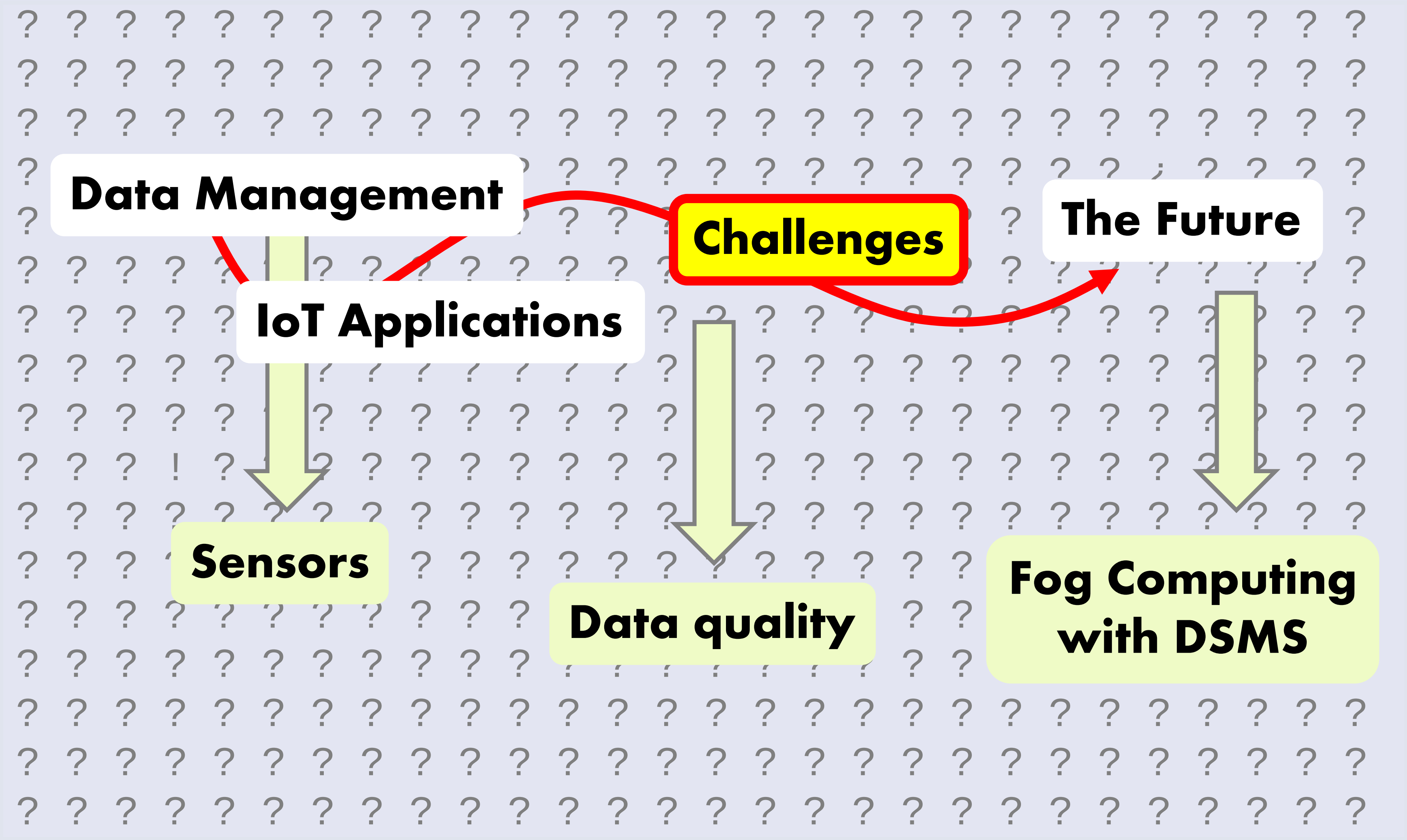
Sensor

Sensor

Sensor

Overview

Universität Bamberg



Data Management

IoT Applications

Challenges

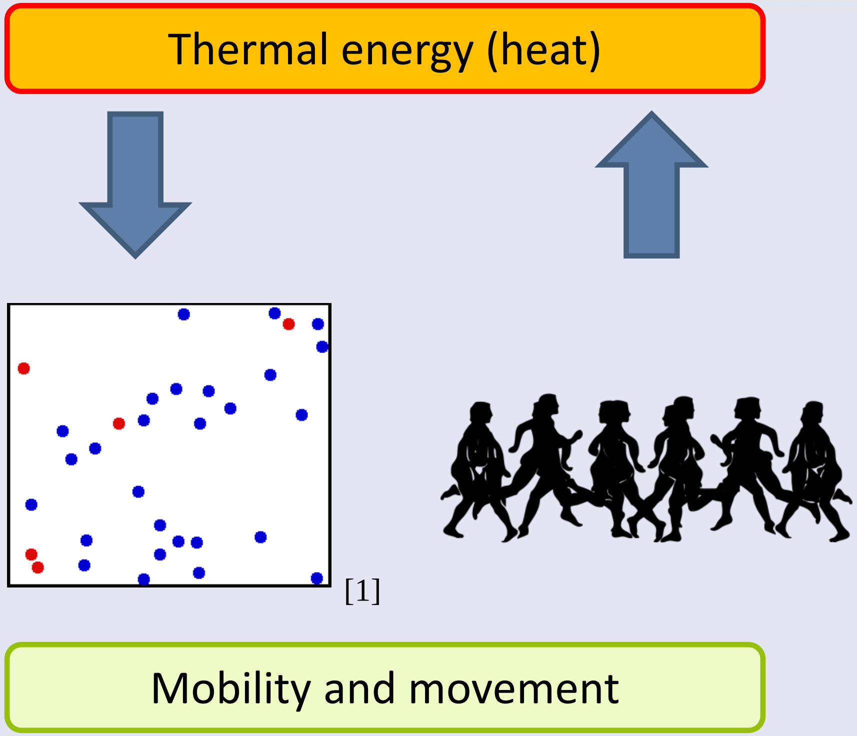
The Future

Sensors

Data quality

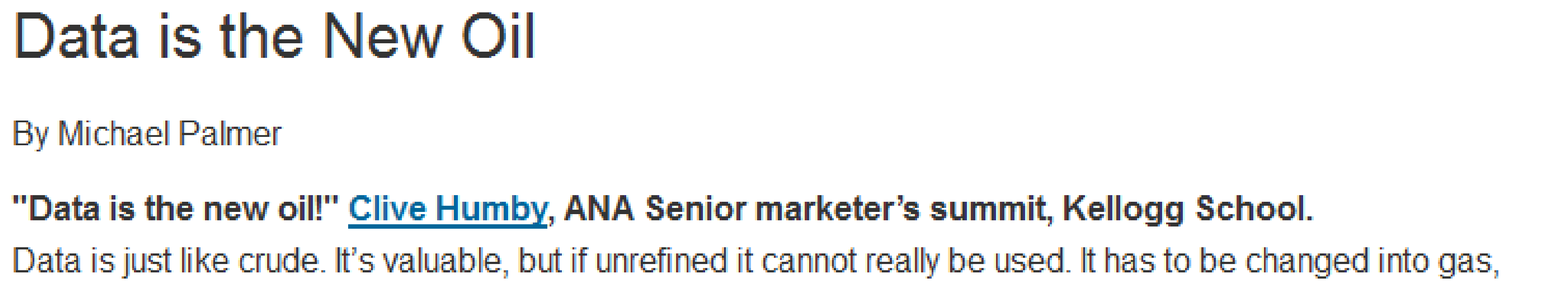
Fog Computing with DSMS

Data and mobility



Katelyn Murray @K9_Murray · 18. März

Officially on **vacation**! **St. Louis**, here I come!



2006, http://ana.blogs.com/maestros/2006/11/data_is_the_new.html, retrieved 21.3.2015

Data is the New Oil

By Michael Palmer

"Data is the new oil!" Clive Humby, ANA Senior marketer's summit, Kellogg School.

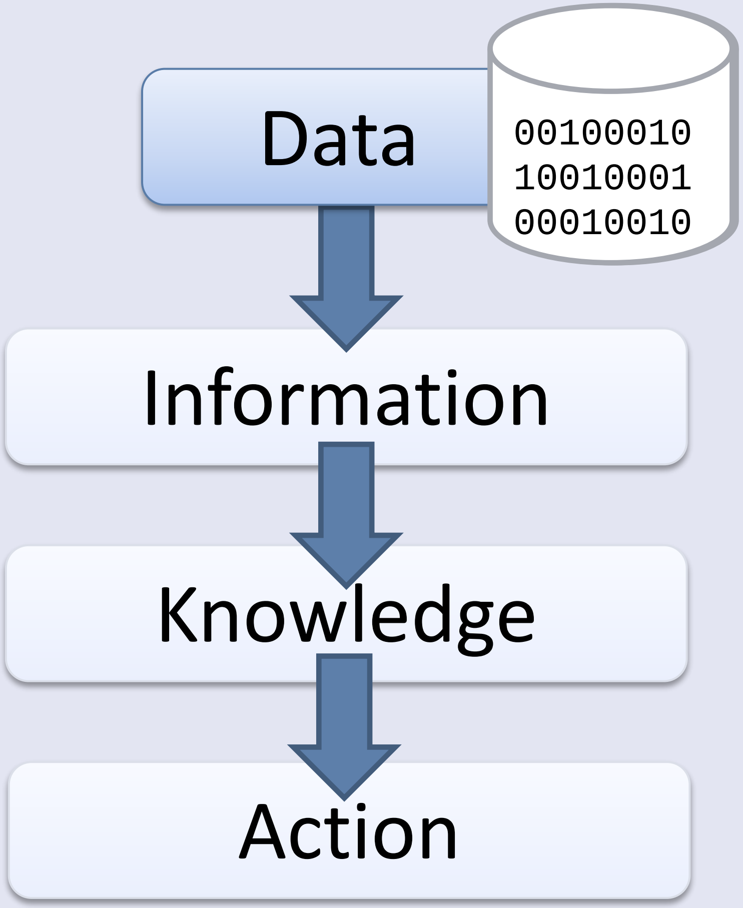
Data is just like crude. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas,

Data is the new crude oil ...

... it needs to be refined to be valuable.

Crude oil

- Fuel oil → mobility
- Chemical products:
  pharmaceuticals → health
  fertilizers → increase growth
  pesticides → kill insects
- ...



Data

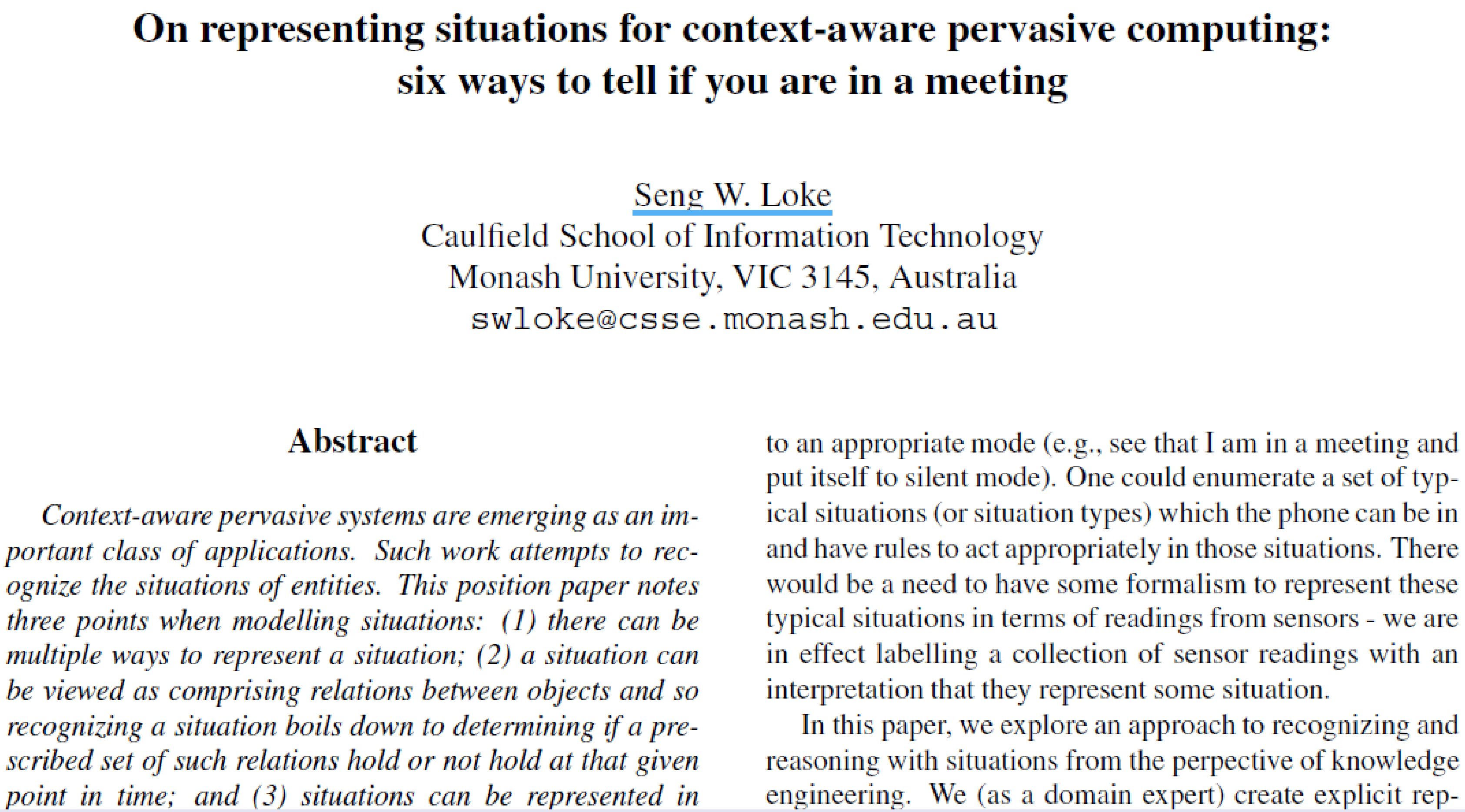
00100010
10010001
00010010

Information

Knowledge

Action

For a given use case, many solutions



On representing situations for context-aware pervasive computing: six ways to tell if you are in a meeting

Seng W. Loke
Caulfield School of Information Technology
Monash University, VIC 3145, Australia
swloke@csse.monash.edu.au

Abstract

Context-aware pervasive systems are emerging as an important class of applications. Such work attempts to recognize the situations of entities. This position paper notes three points when modelling situations: (1) there can be multiple ways to represent a situation; (2) a situation can be viewed as comprising relations between objects and so recognizing a situation boils down to determining if a prescribed set of such relations hold or not hold at that given point in time; and (3) situations can be represented in

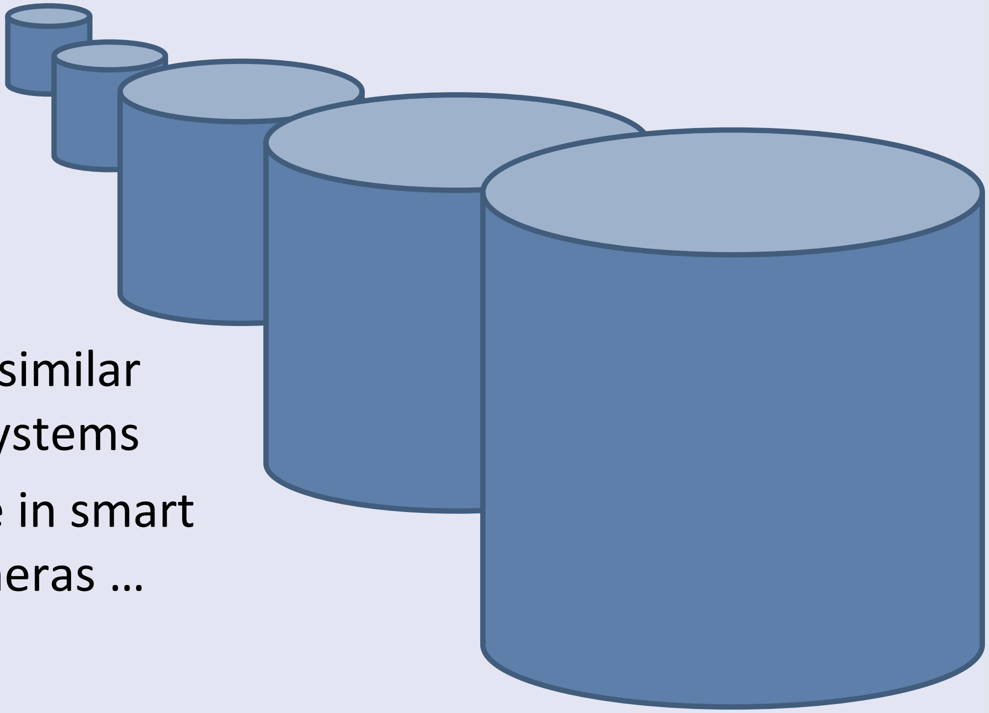
to an appropriate mode (e.g., see that I am in a meeting and put itself to silent mode). One could enumerate a set of typical situations (or situation types) which the phone can be in and have rules to act appropriately in those situations. There would be a need to have some formalism to represent these typical situations in terms of readings from sensors - we are in effect labelling a collection of sensor readings with an interpretation that they represent some situation.

In this paper, we explore an approach to recognizing and reasoning with situations from the perpective of knowledge engineering. We (as a domain expert) create explicit rep-

S. W. Loke, “On representing situations for context-aware pervasive computing: six ways to tell if you are in a meeting,” 2006, pp. 35–39.

DATA and mobility

- Googles self-driving car: nearly 1MB data per second[1]
  - Per day: 85 GB
  - Per year: ~30 TB
  - If 10% of the cars would be like this, or 50%, or …
    (> 1 Billion cars on the world)
  - … not only by self-driving cars, similar for advanced driver assistant systems
  - … plus data from infrastructure in smart cities, like induction loops, cameras …



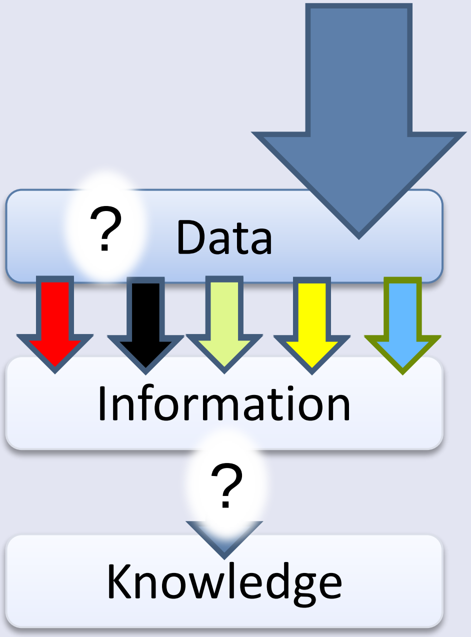
→Big Data!

[1] Bill Gross, Founder and CEO of Idealab
https://www.linkedin.com/today/post/article/20130502024505-9947747-google-s-self-driving-car-gathers-nearly-1-gb-per-second

Big Data Challenges

- Many definitions, often by a numer of 3-5 "V" challenges:

Volume	A lot of data (amount varies)
Variety	Data differs in structure
Variability	Structure changes
Velocity	Many updates
Veracity	Unclear source or quality



Not in list of challenges: Pricacy

(analysis of sensible data, how to adhere to legal / societal constraints?)

Data stream management and Big Data

- More "velocity", less "volume"
- Direct processing
    - Online, (hard/soft) real time, "right time"
- More information, less data
    - Enrichment of data streams
        - E.g., product information for an RFID tag
    - Interpretation and reasoning
        - E.g., classification ("this is a car")
    - Data cleansing
        - Remove redundancy, anomaly detection
- Online quality assessment
- Enables built-in privacy methods
    - Online pseudonmization and anonymization
    - Data economy
    - Certify and/or publish your query plans

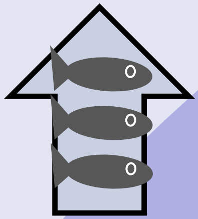
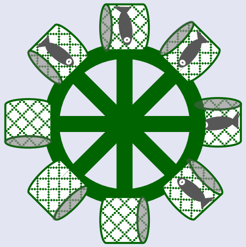
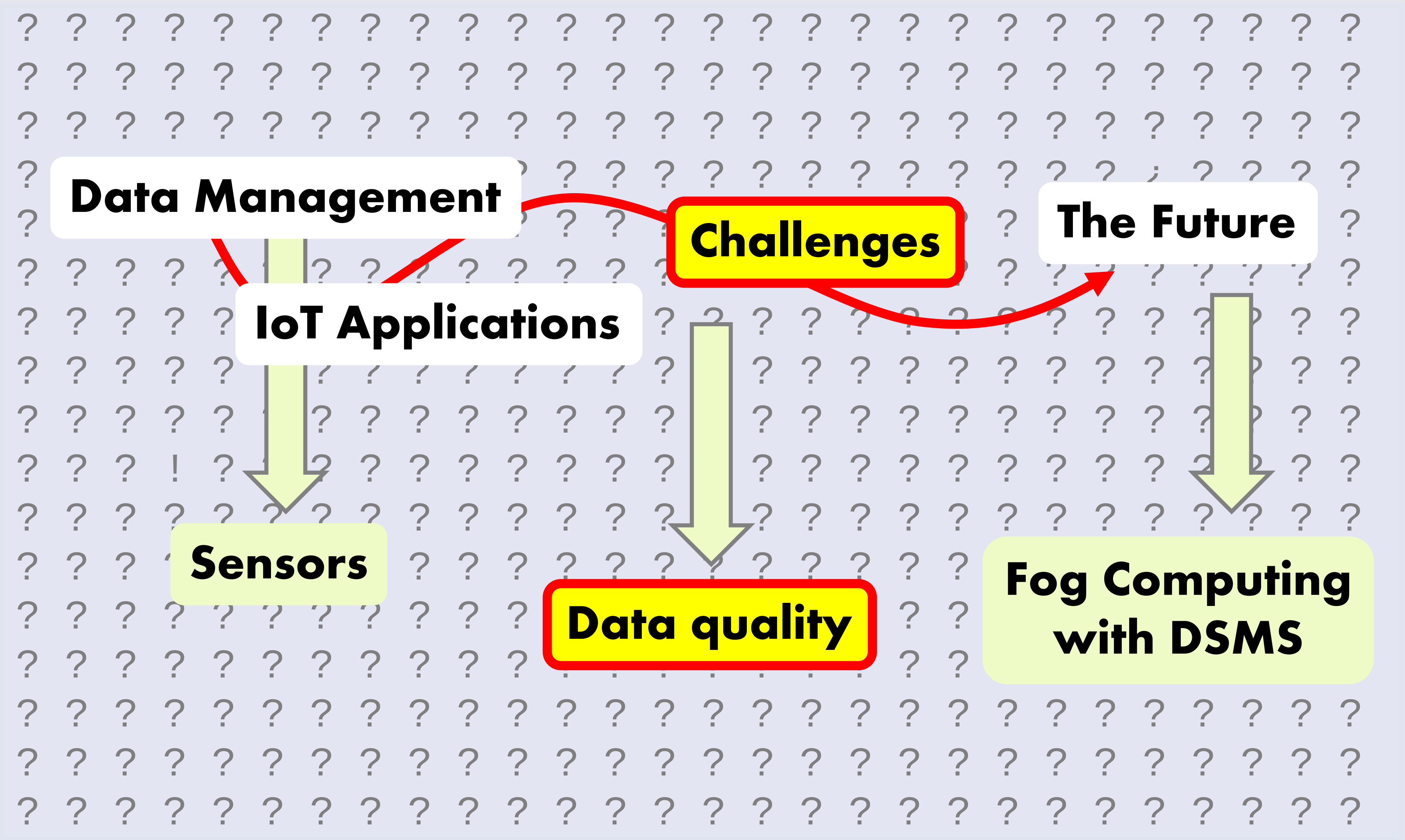


Bild: Ronny Senst / pixelio.de



Overview



Data Management

IoT Applications

Challenges

The Future

Sensors

Data quality

Fog Computing with DSMS

Some common quality issues

- Data source
  - Measurement method, e.g. low frequency of sensor for fast moving objects
  - Environment, e.g., temperature too high for good measurements
  - …
- Data processing
  - Wrong training data for classifier
  - Over-simplified models or missing concepts
  - Not enough input data for algorithm
  - Stale models (due to concept drift)
  - …
- Some can be detected after installation of system, some occur later

→ Decisions based on inpresise data

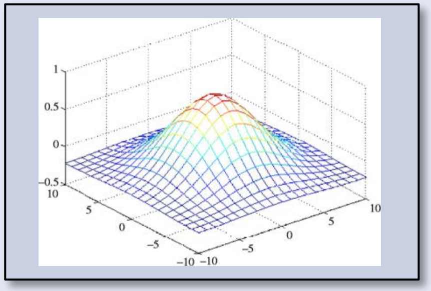
Features for a quality-aware DSMS

Goal: programming abstractions for dealing with non-perfect data

Approach:

1. develop unified data model to represent data quality
2. consider data quality in operators

→ data management can attach combined quality metadata to result



- How to determine data quality and correlations?
  - given by data source / sensor (e.g., accuracy)
  - given by algorithm (e.g., confidence)
  - learned by observation (requires redundancy)

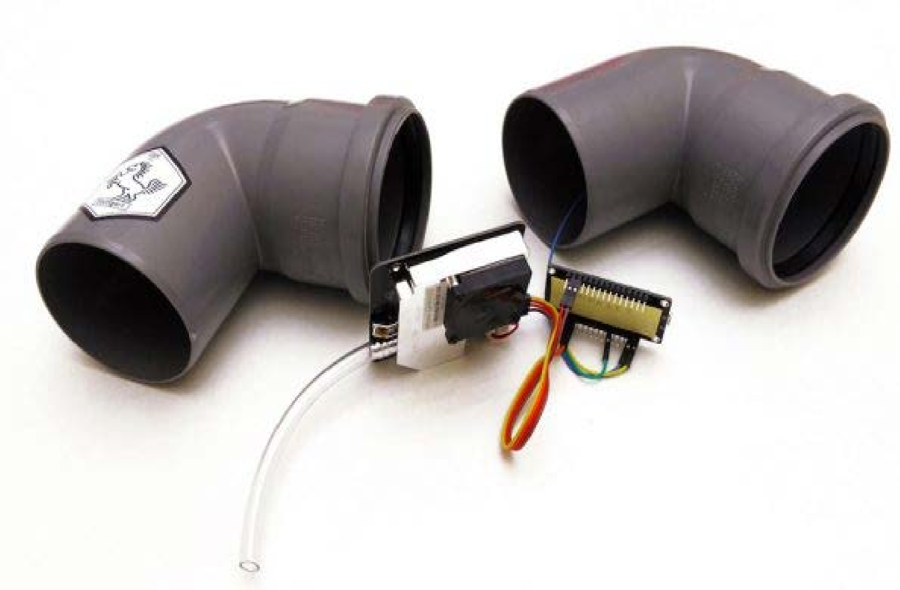
→ store in sensor relationship model

type (probability)
bicycle (0.8)
pedestrian (0.1)
other (0.1)

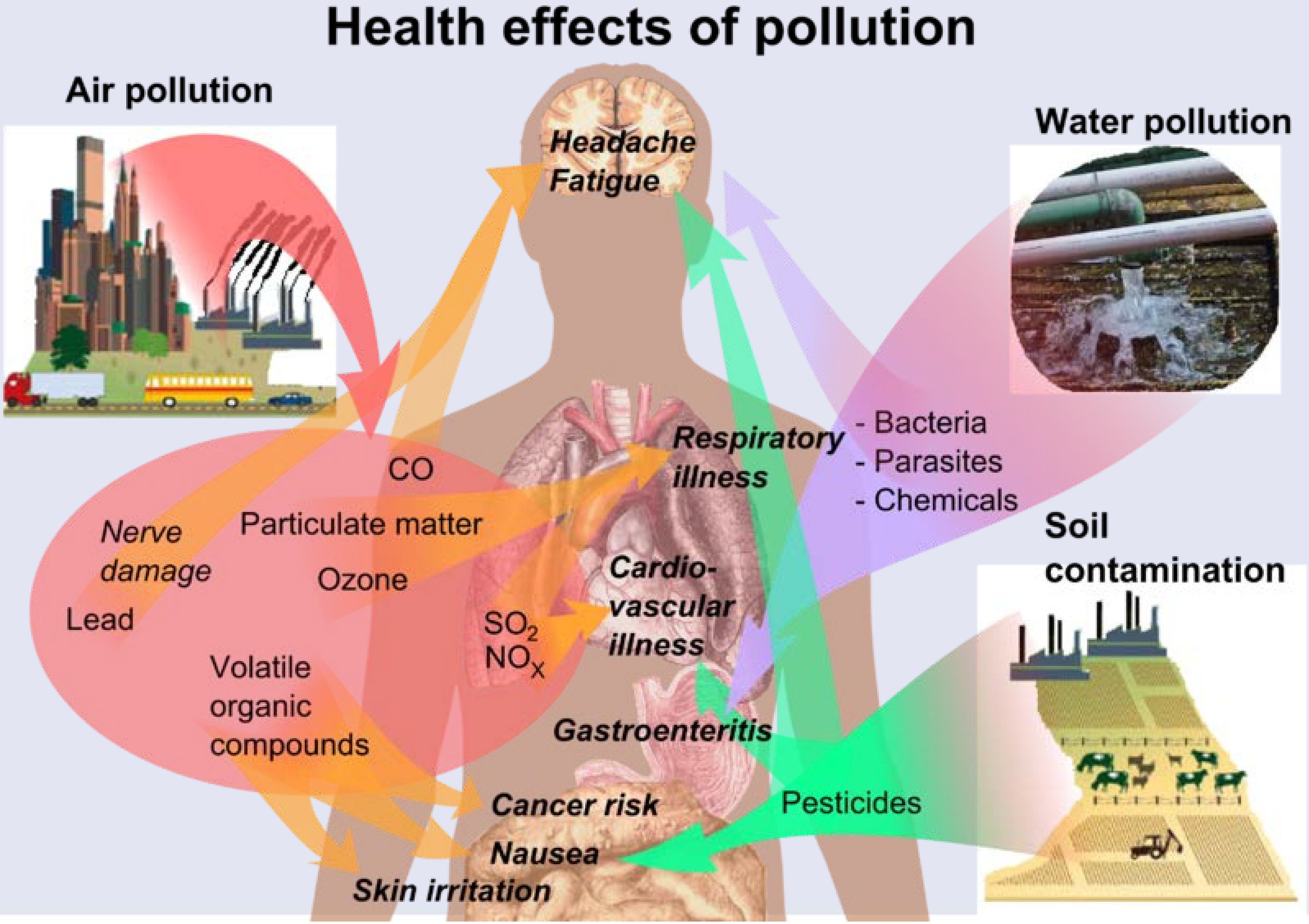
Quality Matters: Supporting Quality-aware Pervasive Applications by Probabilistic Data Stream Management, DEBS2014

Example:
Citizen Science Project luftdaten.info

- Founded by Open Knowledge Lab Stuttgart
  - A group of ten people working mostly on *Citizen Science* projects
- Start: June 2015
- Access for all people:
  - Feasible components
  - Easy assembly
  - Regular workshops and talks throughout Germany
- Goals:
  - Monitor the air quality in Stuttgart
  - Involve citizens in the process
  - Increase the coverage to other areas in Germany and other countries



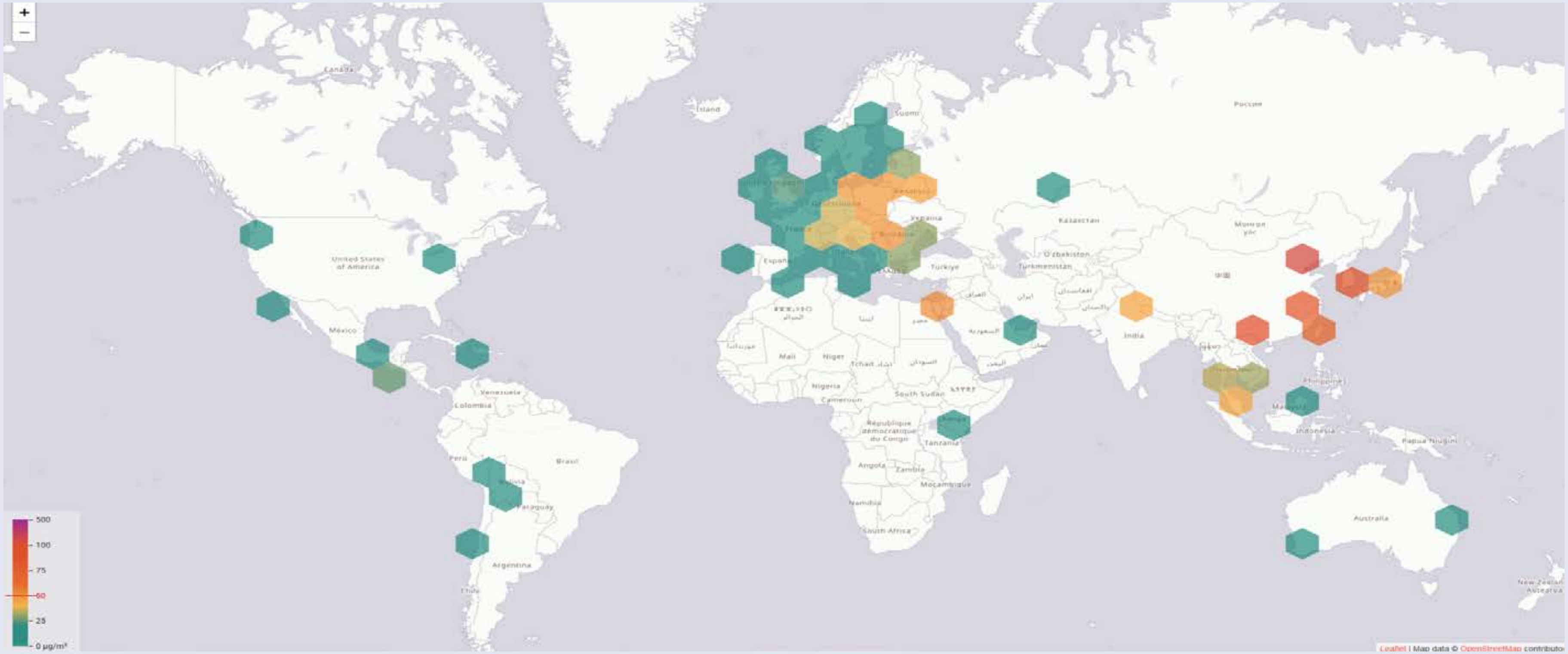
Why should we care?



Current status of sensor installations

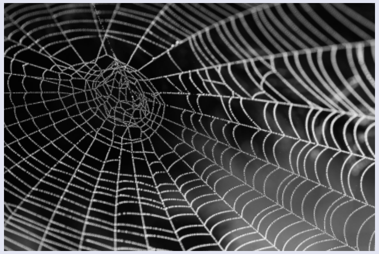
Universität Bamberg

- 4230 registered sensors
- Data from over 15 countries
- Average sensor measurements produced per day: ~80000 readings (as of 02/2018)



Quality issues

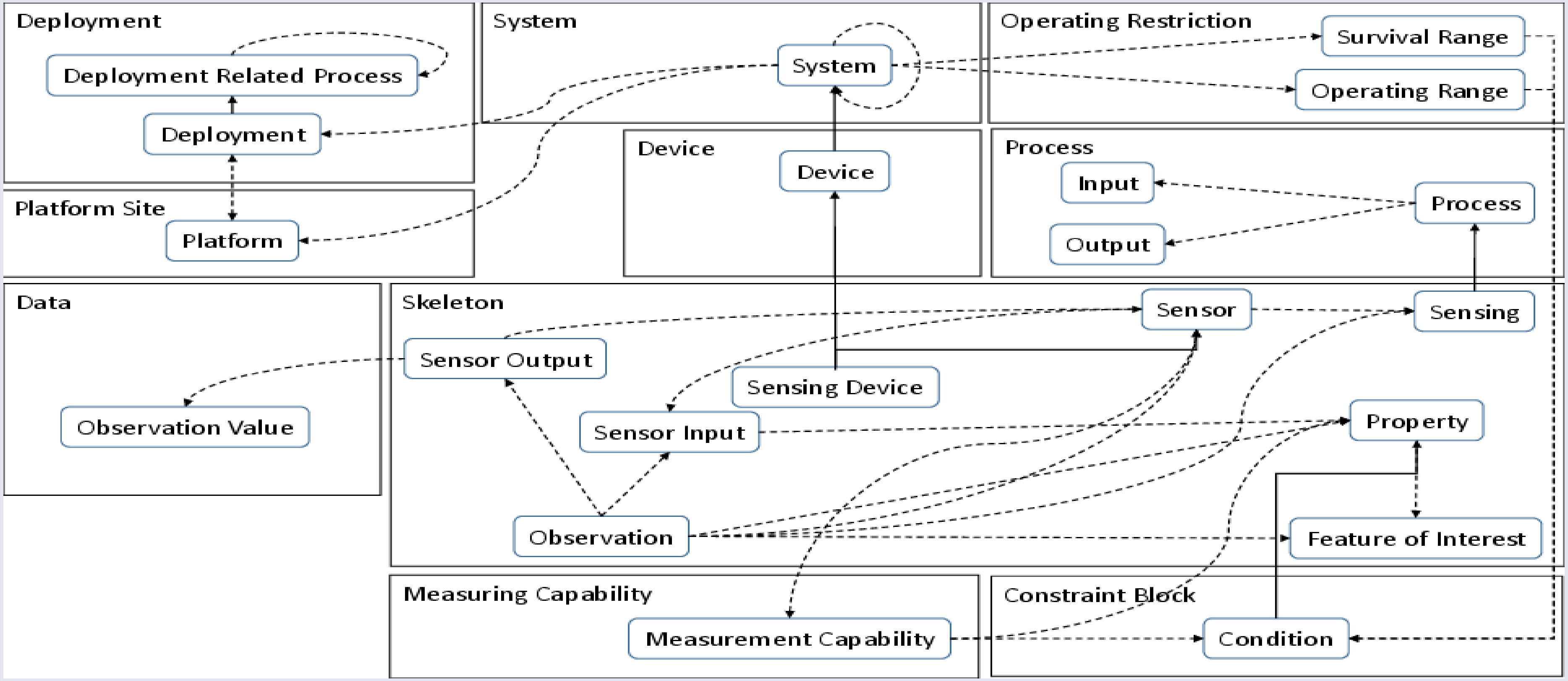
- Sensors can have faulty behaviour and anomalies
  - PM values are no longer reliable when the humidity rises above 70%
  - Some sensors are installed indoors → data leads to wrong conclusions
  - Strange effects might occur later (e.g., spiders in the sensor box)



- Offline detection of the PM quality issues
  - Fusion of the sensor data with data from the neighbouring weather stations
- Online detection of the PM quality issues
  - Use of the on board humidity sensor as a reference
  - Live Stream processing of the sensors with the related weather satations based on the underlying semantic model

SSN ontology: Overview

- Created by W3C Semantic Sensor Network Incubator Group (2011) for
  - Syntactic interoperability
  - Semantic compatibility of sensors and their measurements

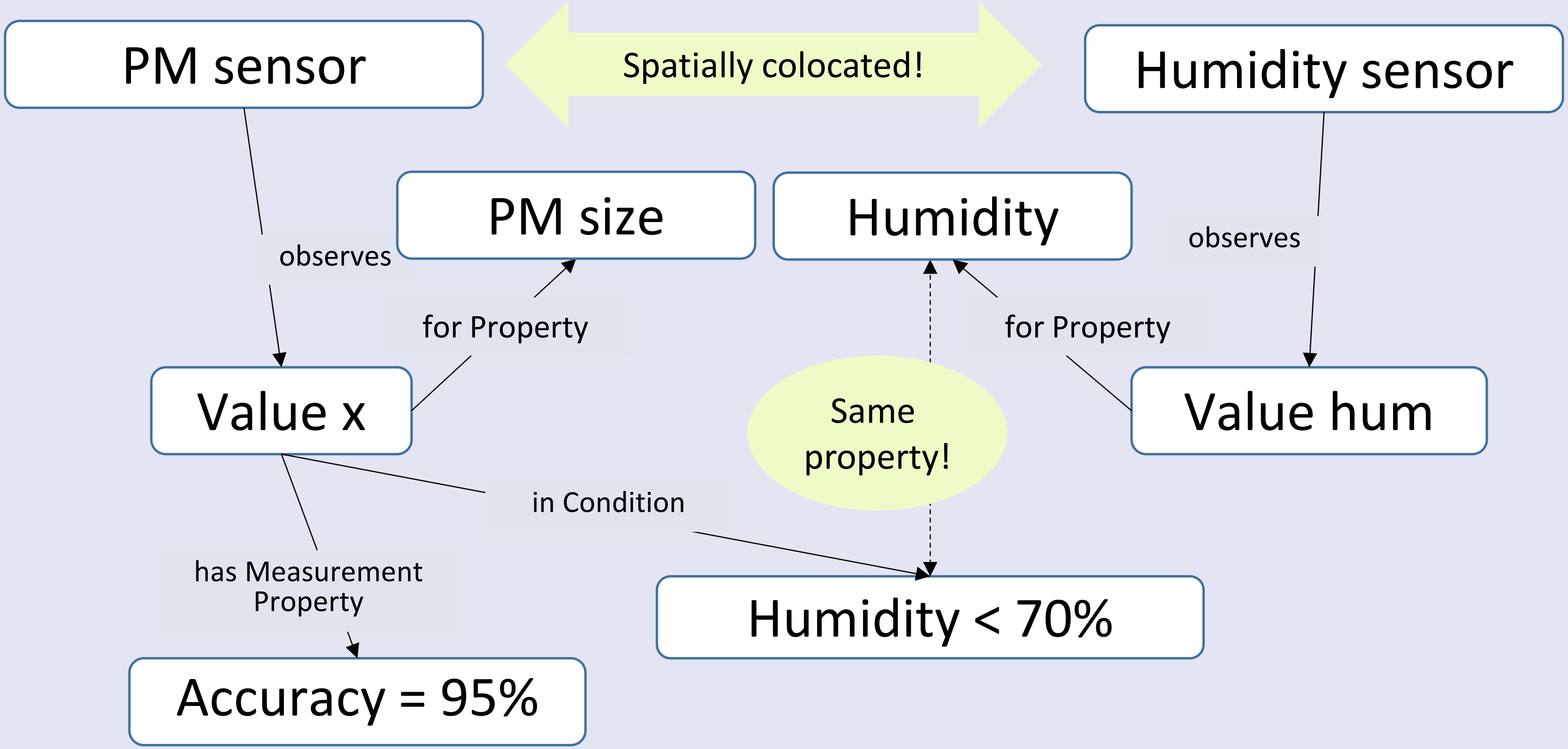


Christian Kuka: Qualitätssensitive Datenstromverarbeitung zur Erstellung von dynamischen Kontextmodellen, PhD, Universität Oldenburg, 2015

Modeling online sensor quality in SSN

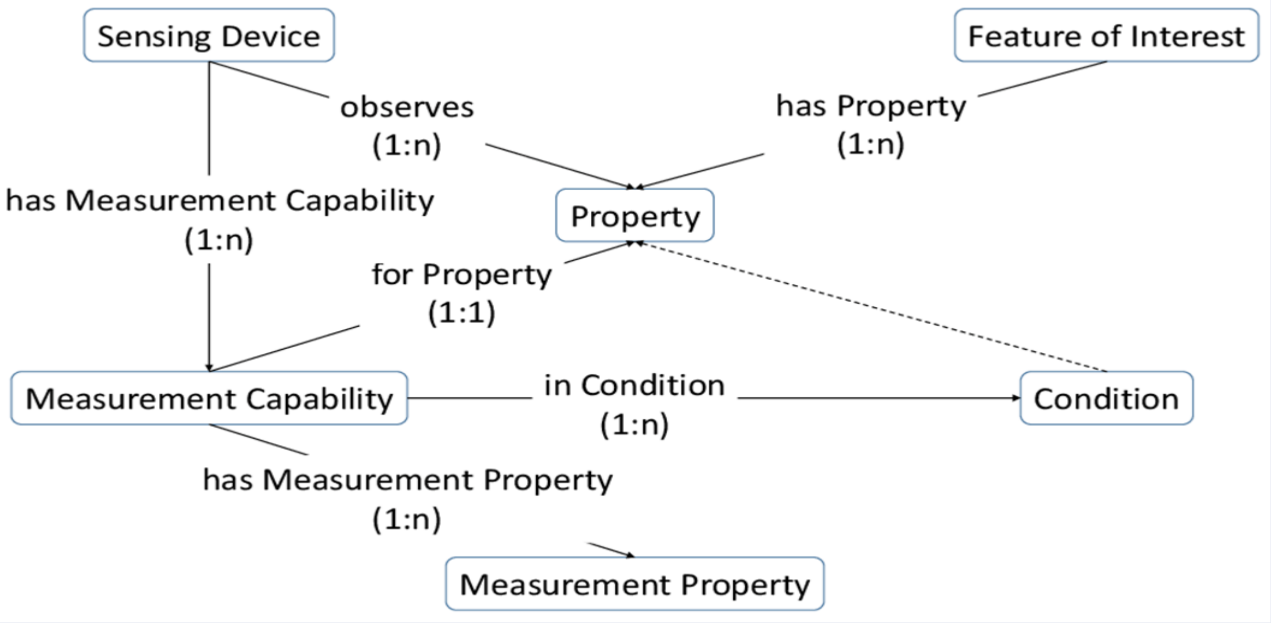
Universität Bamberg

- Measurement capabilities can have conditions
- Conditions can be measured / observed by other sensors!



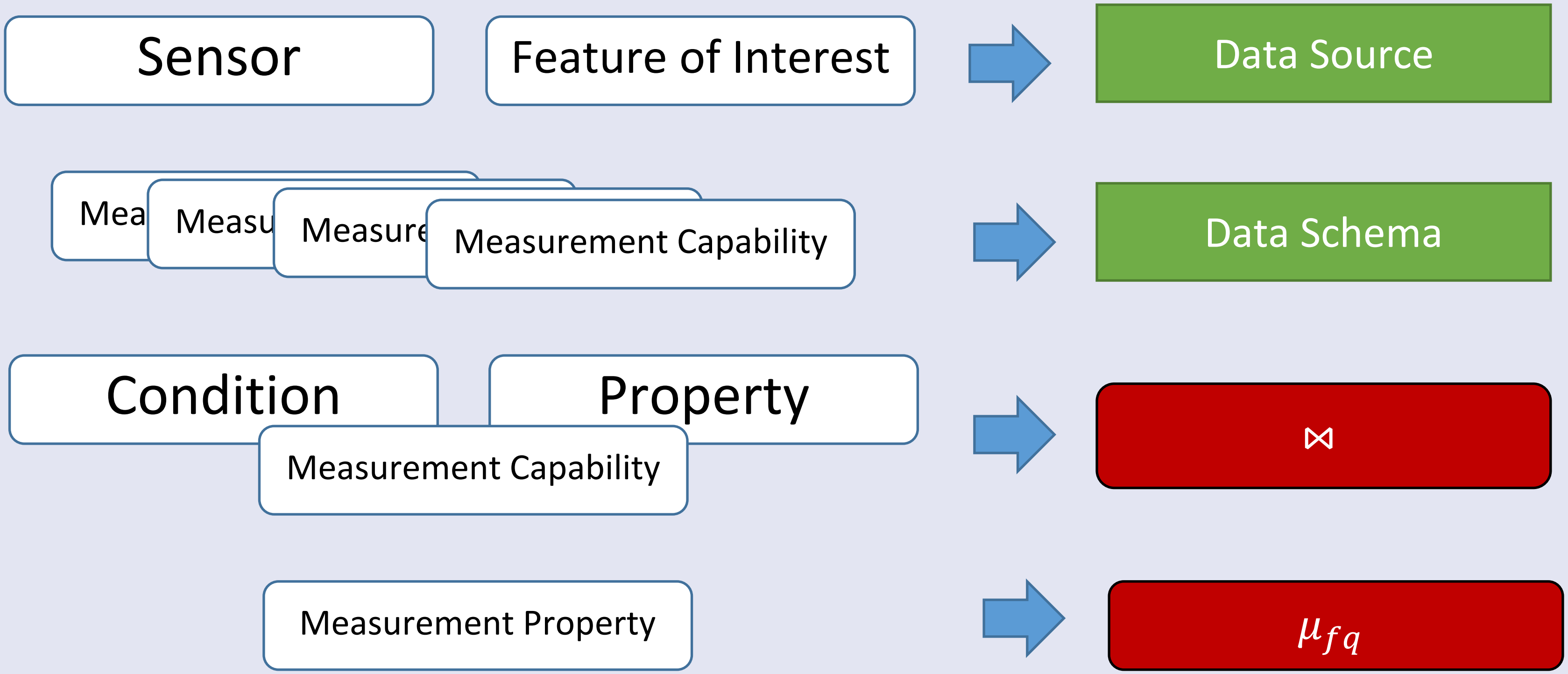
From ontology to stream processing (1)

- Mapping of nodes in the ontology to sources and schemas in a DSMS
  - *Sensors and Feature of Interest to active sources*
  - *Measurement Capability* to data stream schemas
- Transformation of links in the ontology into partial processing queries
  - Conditions are transformed into select predicates and *Measurement* Properties into mapping expressions



From ontology to stream processing (2)

- Mapping linked sensor data to stream processing operators



Sensor

Feature of Interest

Data Source

Mea

Measu

Measure

Measurement Capability

Data Schema

Condition

Property

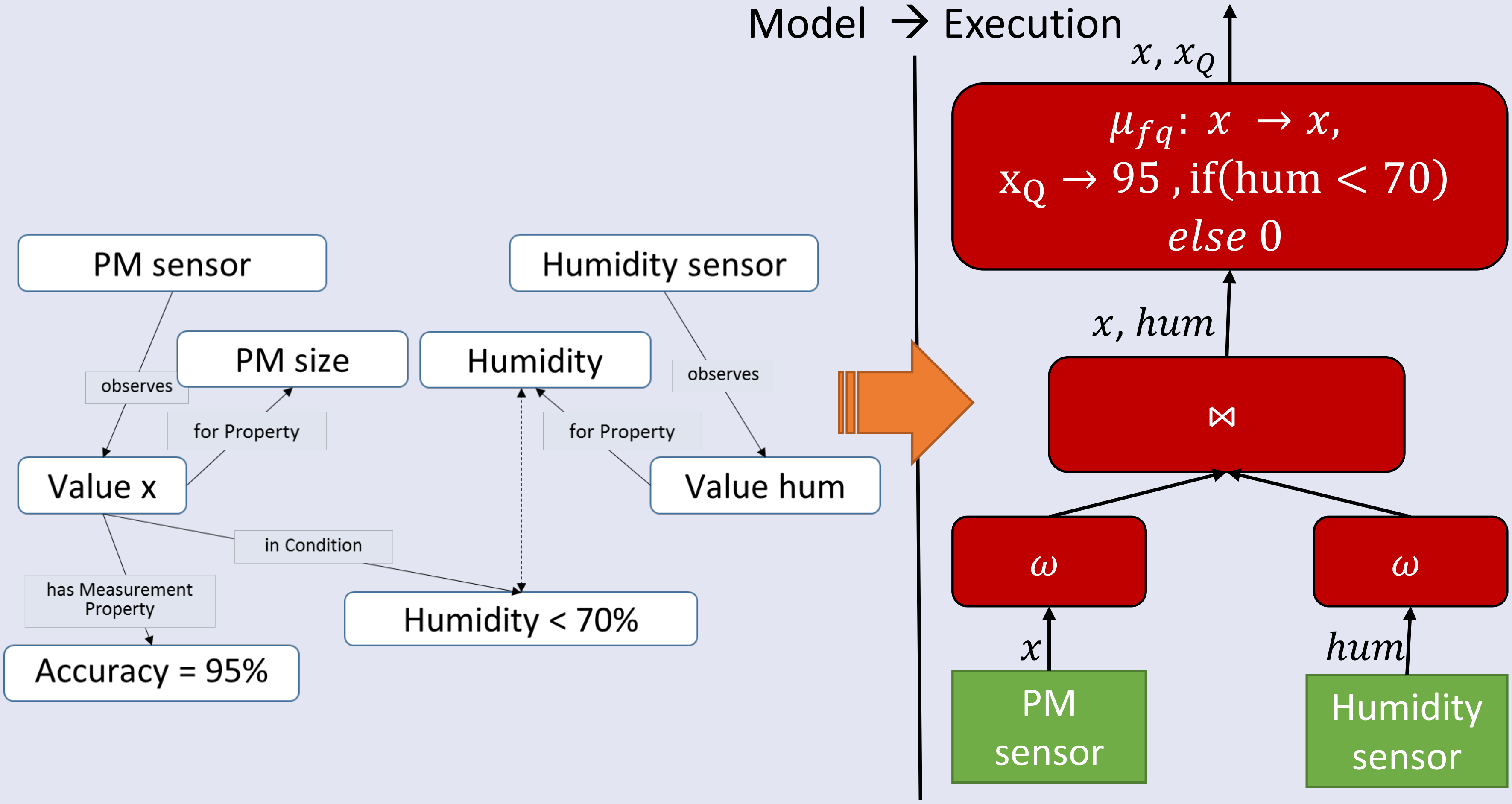
Measurement Capability

$\bowtie$

Measurement Property

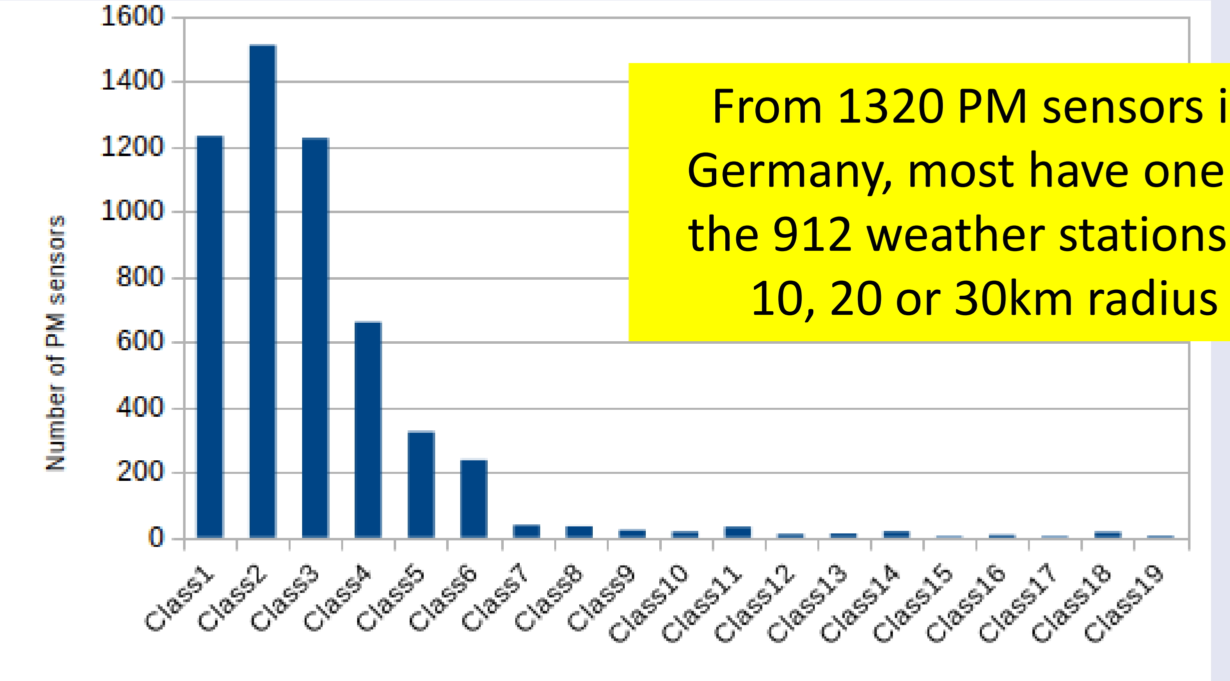
μ_{fq}

From ontology to stream processing (3)



Which humidity sensors can we use?

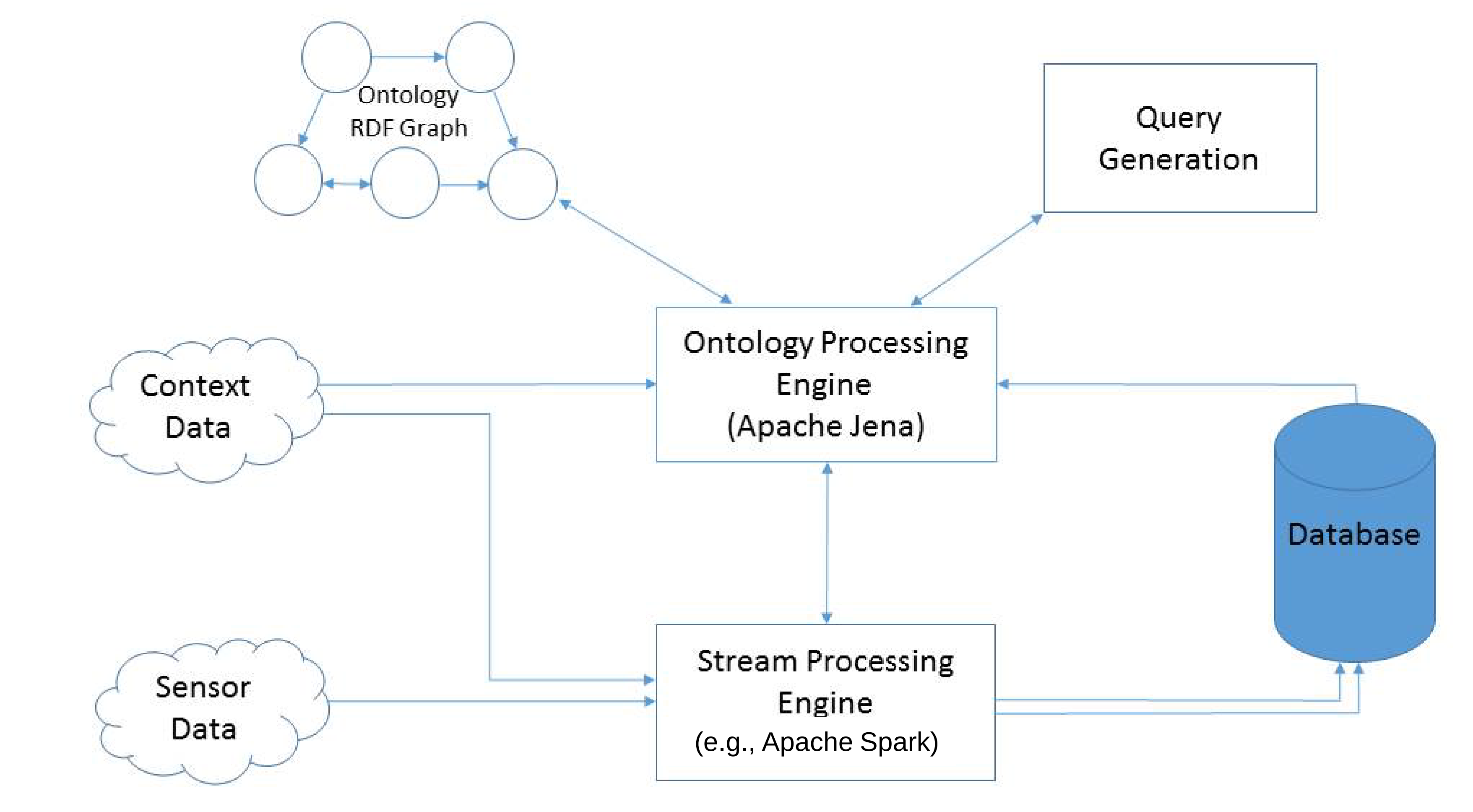
- Can we trust an on-board huminidity sensors?
- Or: Use weather stations as data quality reference sources?
- Data from the weather stations used to assess the accuracy of sensors



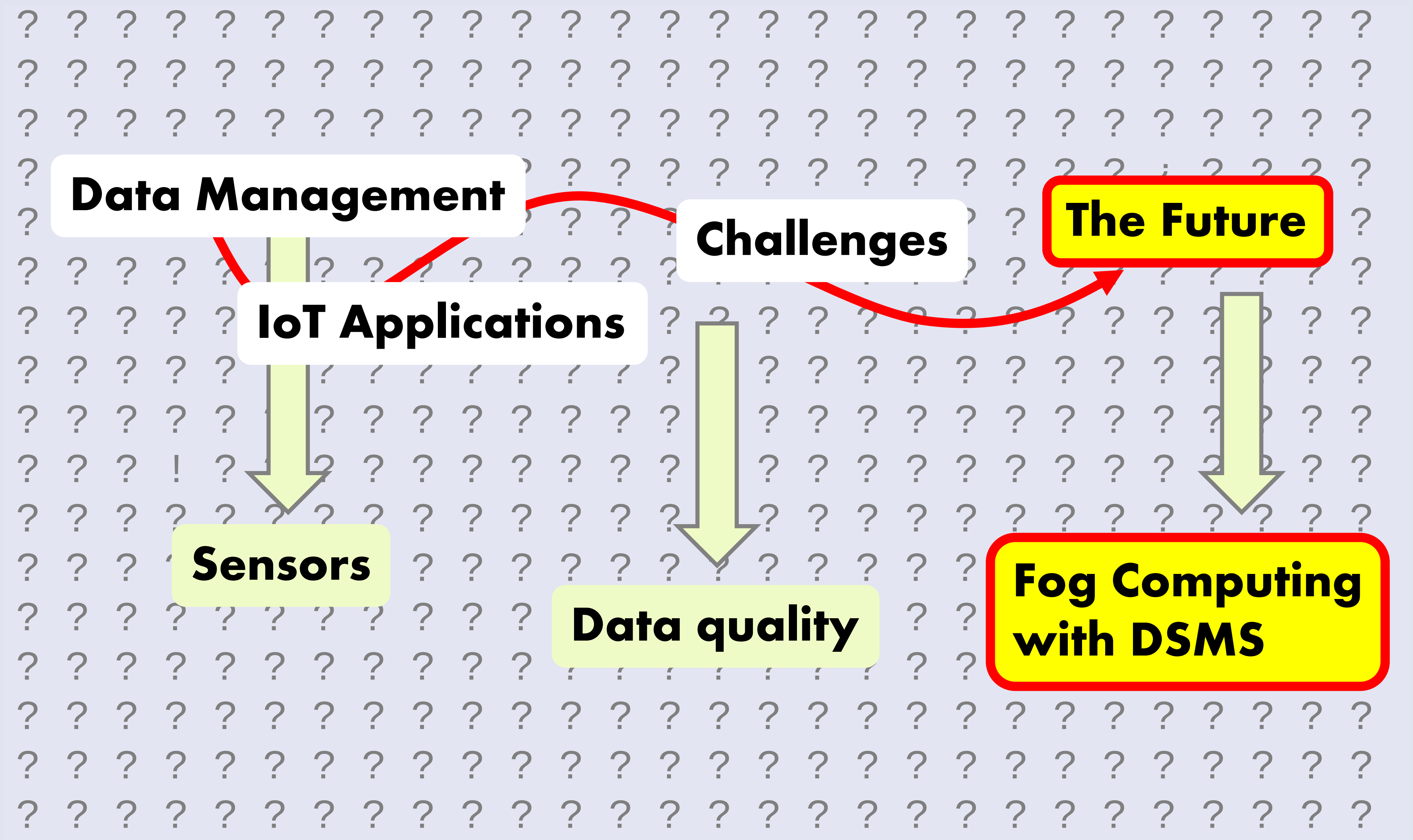
From 1320 PM sensors in Germany, most have one of the 912 weather stations in 10, 20 or 30km radius

Number of sensors covered by at least one weather station in 10km radius classes

Resulting architecture



Overview



Data Management

IoT Applications

Sensors

Challenges

Data quality

The Future

Fog Computing with DSMS

Perfect data quality for future mobility?

Rush Hour by Fernando Livschitz, Black Sheep films
http://vimeo.com/106226560

Next step in architectures: Fog Computing

Universität Bamberg

- Sending all raw sensor data to the cloud cannot be the final solution:
  - Bandwith
  - Energy comsumption
    - (computing needs less than communication)
  - Application needs, e.g., privacy
- Edge computing:
  - Move the processing to the edge of the network
- Fog computing:
  - Utilize further processing nodes on the way

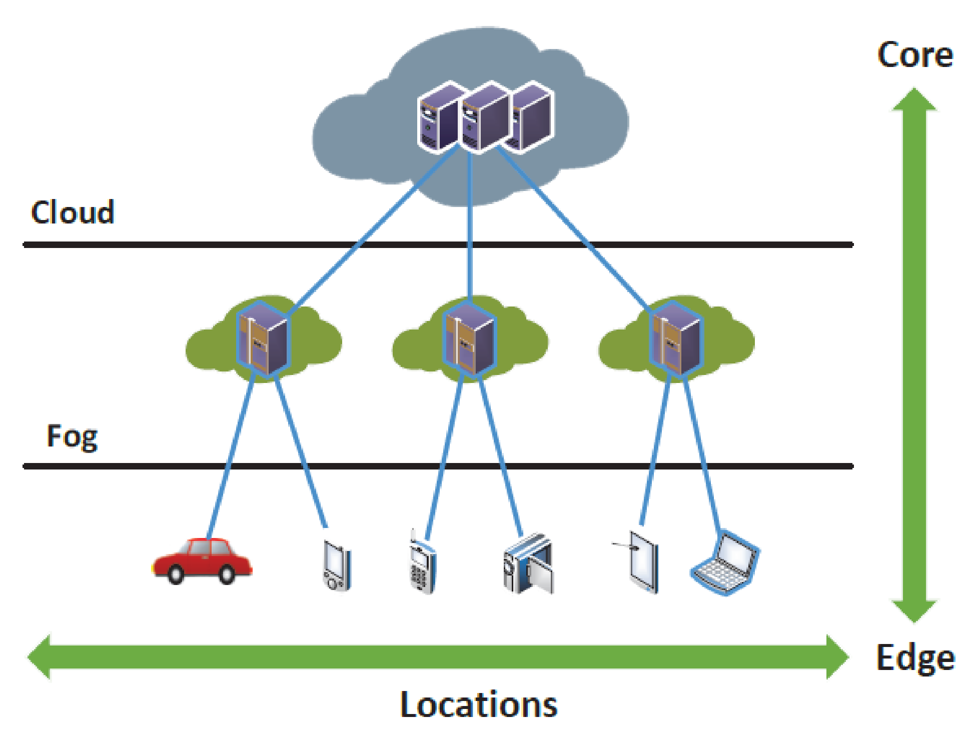


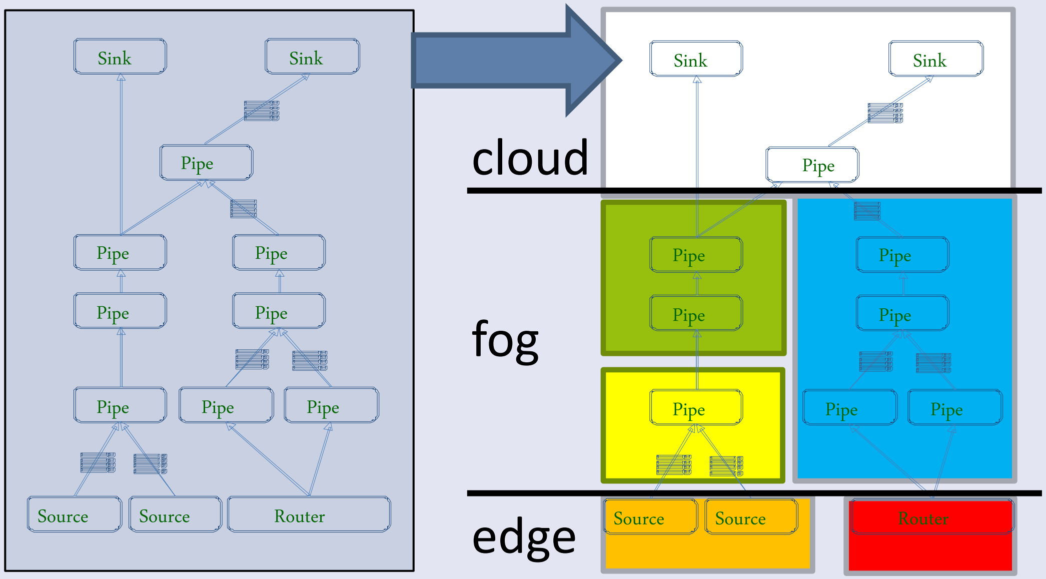
Fig. 1. Fog between edge and cloud.

[1]

[1] I. Stojmenovic and S. Wen, “The Fog Computing Paradigm: Scenarios and Security Issues,” 2014, pp. 1–8.

Fog computing and distributed data stream management

- Data stream management:
  - Provides a higher-level abstraction to stream-based data processing
- Distributed stream management:
  - Distributes the execution of the data stream processing over nodes
  - Finds an optimized query execution plan
  - Can adapt to changing situations and migrate the execution



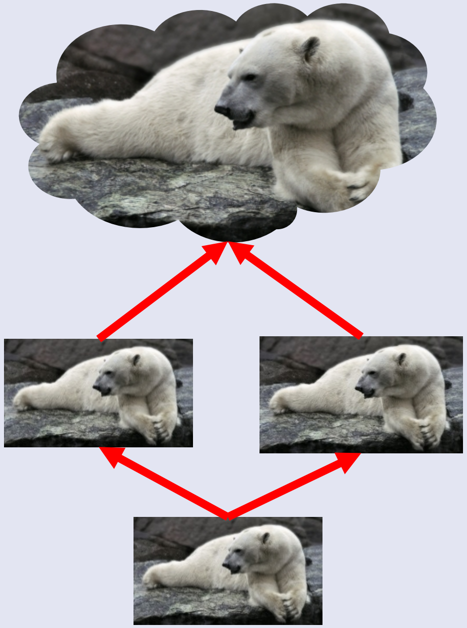
We can use distributed DSMS to implement sensor data management in a fog-computing architecture

Summary and outlook

Universität Bamberg

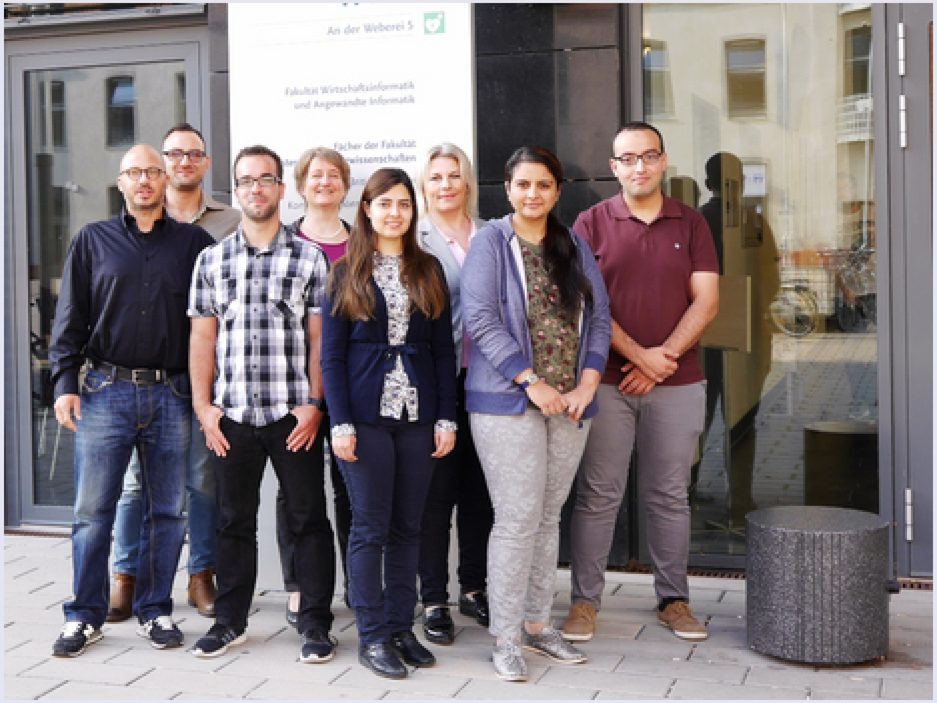
- IoT applications can lead to large-scale sensor data management systems
- Issues to solve:
  - The „V“ challenges → maybe you do not need to store everything in the cloud
  - The „P“ challenge → maybe you can anonymize or aggregate at the edge or in the fog
  - The „Q“ challenge → know thy quality, before and during operation
- IoT platforms can help, but are only slowly moving towards fog architectures („who owns the data?“)

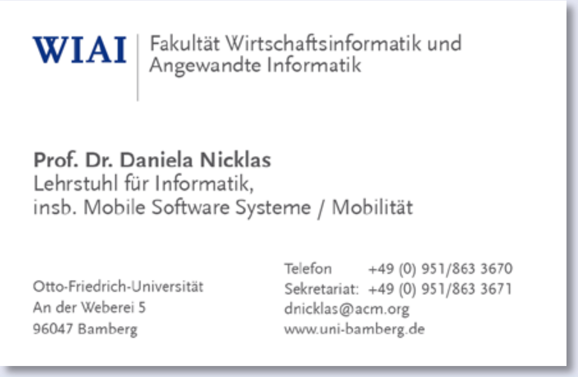
  → Distributed data stream processing revisited?



Ronny Senst / pixelio.de

Thank‘s for all the fish!





WIAI | Fakultät Wirtschaftsinformatik und Angewandte Informatik

Prof. Dr. Daniela Nicklas
Lehrstuhl für Informatik,
insb. Mobile Software Systeme / Mobilität

Otto-Friedrich-Universität
An der Weberei 5
96047 Bamberg

Telefon	+49 (0) 951/863 3670
Sekretariat:	+49 (0) 951/863 3671
dnicklas@acm.org	
www.uni-bamberg.de	

Bild: Ronny Senst / pixelio.de