NORDTEK 9.6.2022

DATA IS HUMAN

Situating people within data-driven systems

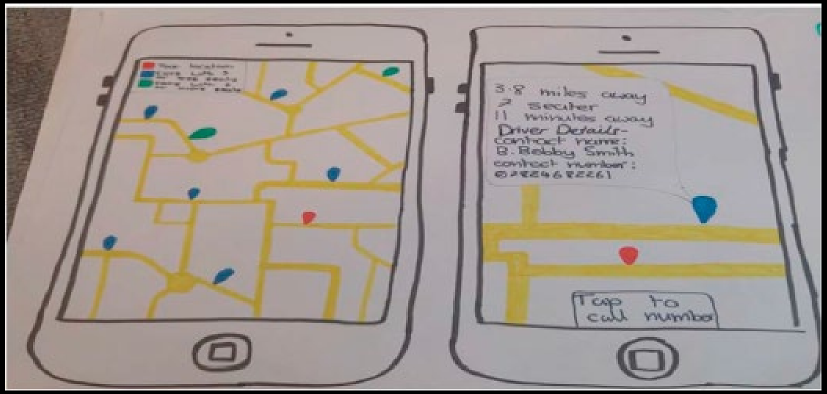
Annika Wolff

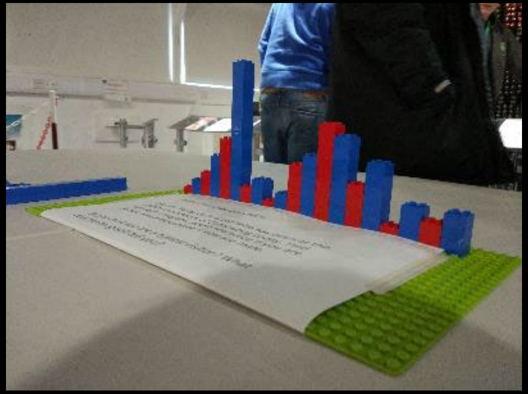
LUT University

OVERVIEW

OVERVIEW OF TALK

- In this talk I would like to demonstrate some of many ways in which data is essentially human. I will:
  - Challenge notions of data objectivity
  - Explore ideas of data equitability and justice – who should benefit from the data? Who should design with it?
  - Propose approaches for widening participation in data-driven systems





Context

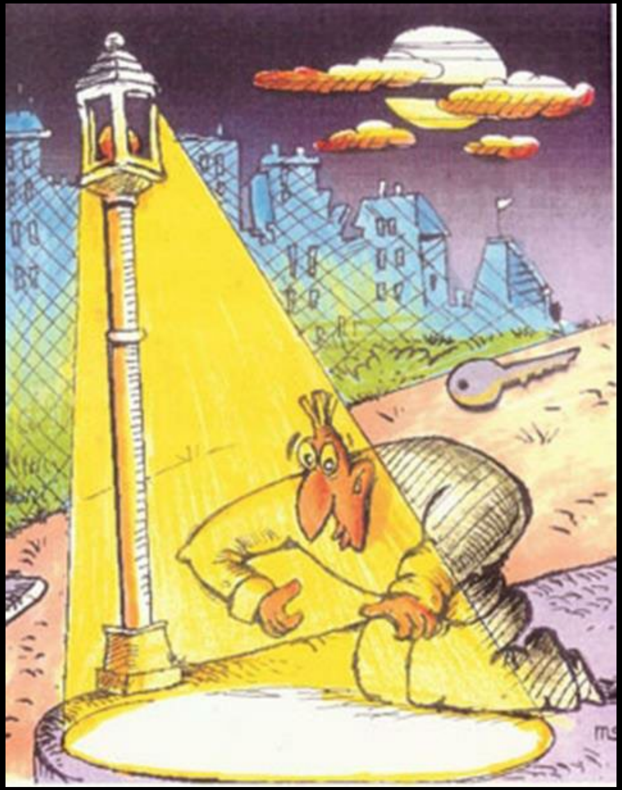
WHAT IS A DATA-DRIVEN SYSTEM?

- Increasingly instrumenting our environment to generate and collect data, e.g. via IoT, sensor networks, clickstreams and so forth
- Commonly quantify such systems with the term ’smart’
  - Smart home/building
  - Smart city/village
  - Smart mobility
  - Smart grid
  - Smart agriculture
  - Smart health
  - Smart campus/learning environment

Context

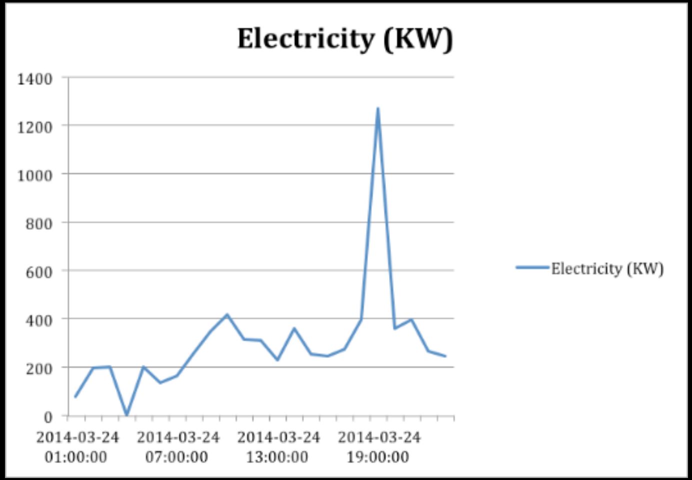
COMMON PROBLEMS OF DATA-DRIVEN SYSTEMS?

- Data quality, subjectivity and (inherent) bias
  - Common example of bias in AI, e.g. recruitment algorithm biased against women and other minorities
- Streetlight effect – we collect the data that is easiest to collect and then we look for solutions in easy places.
- Data first approaches risk missing **real needs of end-users**:
  - What **questions** does the data answer?
  - What **problems** can it solve?
  - Who is the **beneficiary**? (data equitability)



CHALLENGING THE IDEA OF DATA OBJECTIVITY

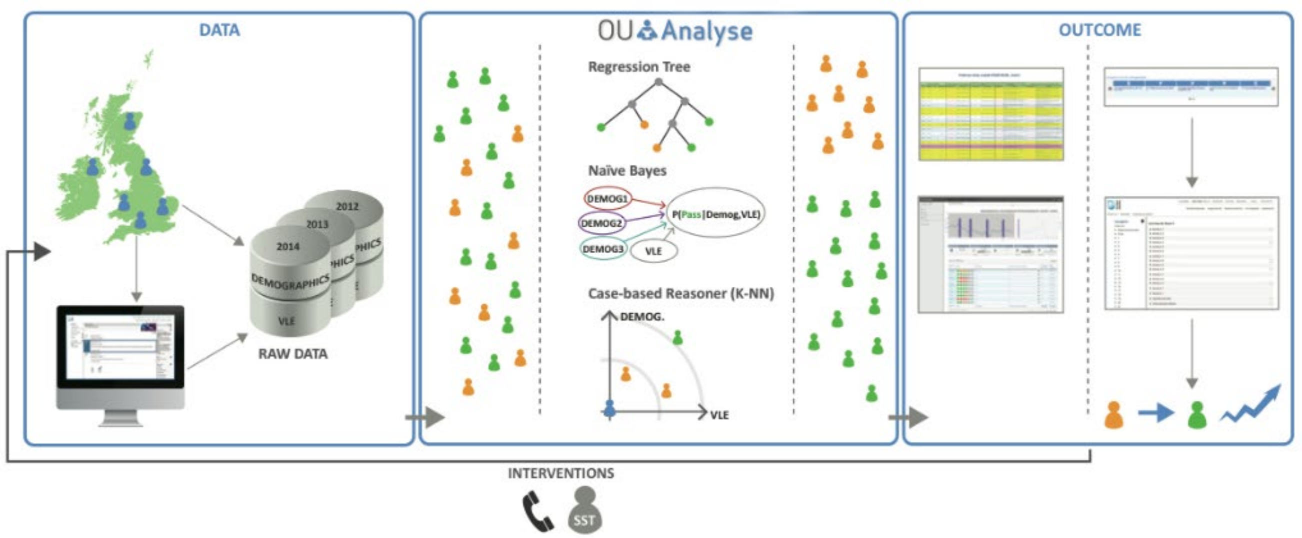
- Many people believe that data stands alone – it is pure and objective. But is this true?
- Example: smart meter data can help us all to reflect on our own energy use. In one initiative we took smart data into primary and secondary schools in the UK as part of an 'urban data school' initiative



- Students notice and speculate about the **0** reading:
  - Power cut – do nearby houses show similar pattern?
  - Power cut caused by key meter running out – how quickly do people normally resolve this?
  - Leaving the house for a while – it is difficult to turn off power to entire house. Would anyone do it for this long?
  - Fault in meter – do meters generally show intermittent faults like this?

A CASE FROM LEARNING ANALYTICS

- Scenario: developing a large scale learning analytics programme across Open University Moodle and other data sources to improve student retention.
- Premise: data predicts students at risk of dropping out or failing an exam, based on models built from historical data



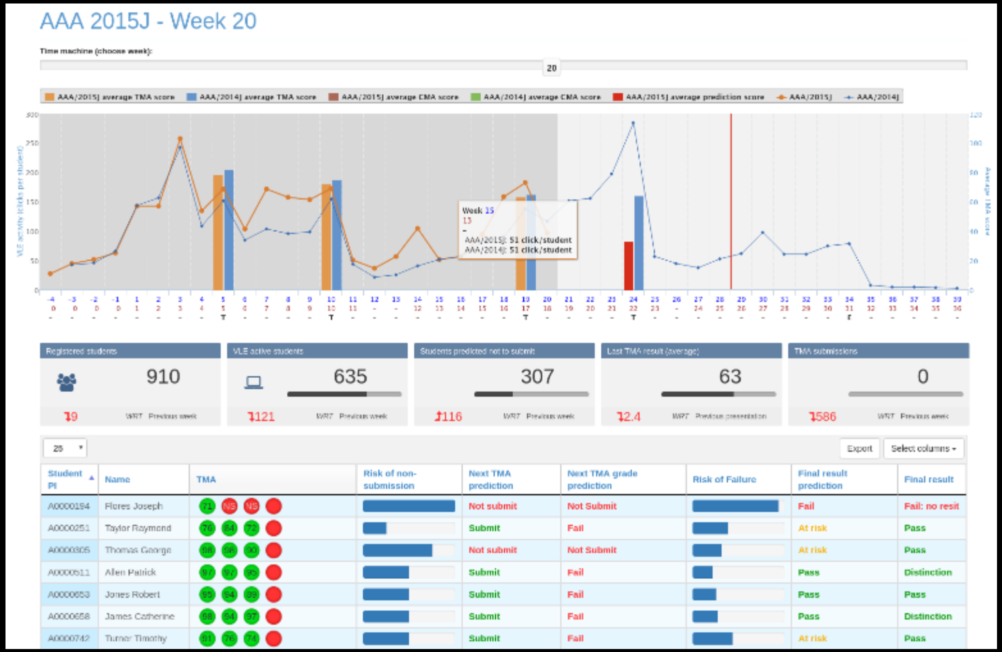
DATA QUALITY, SUBJECTIVITY & BIAS

A CASE FROM LEARNING ANALYTICS

DATA TYPES

- **Static data**
  - **Demographics:** if you rely on demographics alone, you find that people with certain characteristics are inclined to fail. BUT further exploration shows this can be overcome –> Dangerous to suggest that demographics predict learning outcome
  - **Student status:** relies on manual input and categorization of data. Lots of human error found in processes – data may be biased or inaccurate from the moment of its collection -> cannot assume good data practice is followed, must understand who is involved in data collection and their motivations
- **Dynamic data**
  - **Moodle** – based on clicks into the VLE. Less accurate than imaged – did one student really learn 24 hours a day?

EXAMPLE DASHBOARD – *FOR TUTORS*



DATA JUSTICE

WHO BENEFITS? WHO IS IMPACTED?

- Despite some student-centred learning analytics approaches still, in many LA approaches *tutors are given insight from student's own data - that **may not directly benefit** the student, or at least where they do not decide the benefit.* Is this fair?
- Should the people who are closest to the data have some priority, in both its interpretation and use?

DATA JUSTICE

FURTHER EXAMPLE: SMART BUILDING/WINDOWS

- Automated windows on a previous campus– but no one understood on what basis they worked.
- Sometimes the environment was not comfortable
- No way to manually control windows
- Asking building services were just told it ‘worked correctly’ according to how the algorithm was defined
- Yarosh & Zave /2017) showed that similarly people did not understand their smart home security

Yarosh, S. and Zave, P., 2017, May. Locked or not? mental models of iot feature interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2993-2997).

SOLUTIONS – WIDENING PARTICIPATION

EMPOWERING USERS THROUGH CO-DESIGN

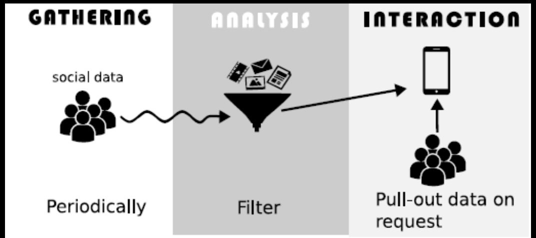
- Common principles of co-design:
  - Involves the empowerment of end-users within a design process
  - Is based on collaborative activities that build mutual understandings and knowledge exchange
    - Technologists move their mindset closer towards end-users
    - End-users learn about and influence the direction of new technologies to closer fit their needs
- Benefits: Solutions better fit the needs of everyone involved. Can discuss and agree trade-offs and get buy-in for use of data

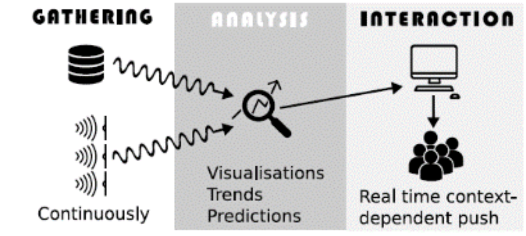
Common barrier: the difficulty to develop a shared understanding amongst people with different knowledge and skills, especially with complex data – how to support data as a design material?

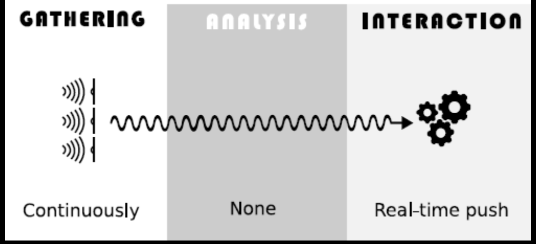
LUT University

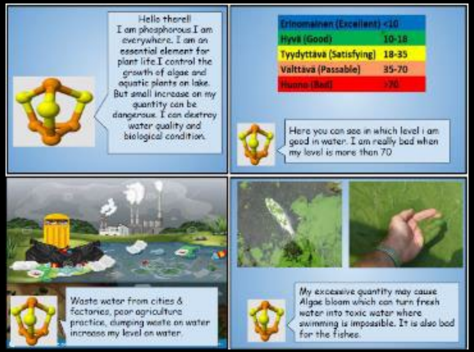
SOLUTIONS – WIDENING PARTICIPATION

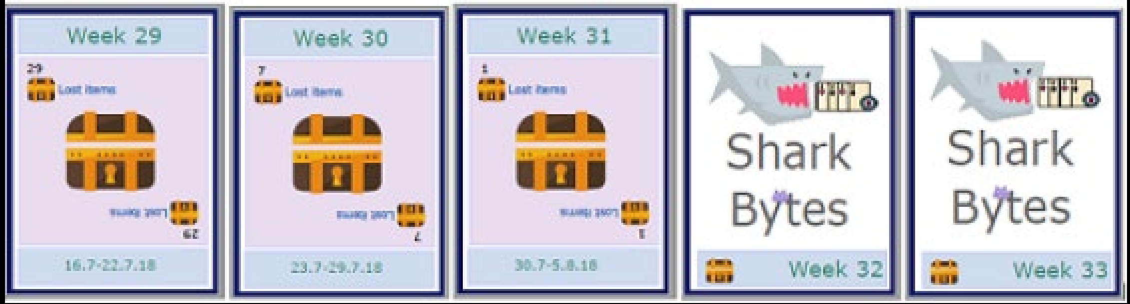
METHODS FOR UNDERSTANDING PROCESSES AND DATA

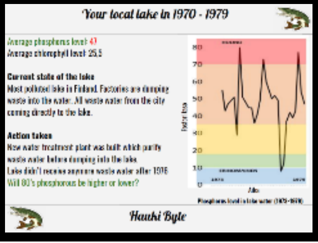












SUMMARY

SO WHERE ARE PEOPLE SITUATED IN DATA-DRIVEN SYSTEMS?

- People define the problems, either on behalf of themselves or on behalf of others (but is everyone always included?)
- People collect and record data about themselves and about others, e.g. into databases or as part of their own scientific endeavours (and sometimes introduce mistakes)
- People choose the data that machines will collect (and sometimes introduce bias)
- People decide how data will be analysed (but the analysis may benefit only a few)
- **When we are engineering complex, data driven and hyperconnected system we should understand that people working together and sharing knowledge can mitigate against all of the above!**


LUT
University