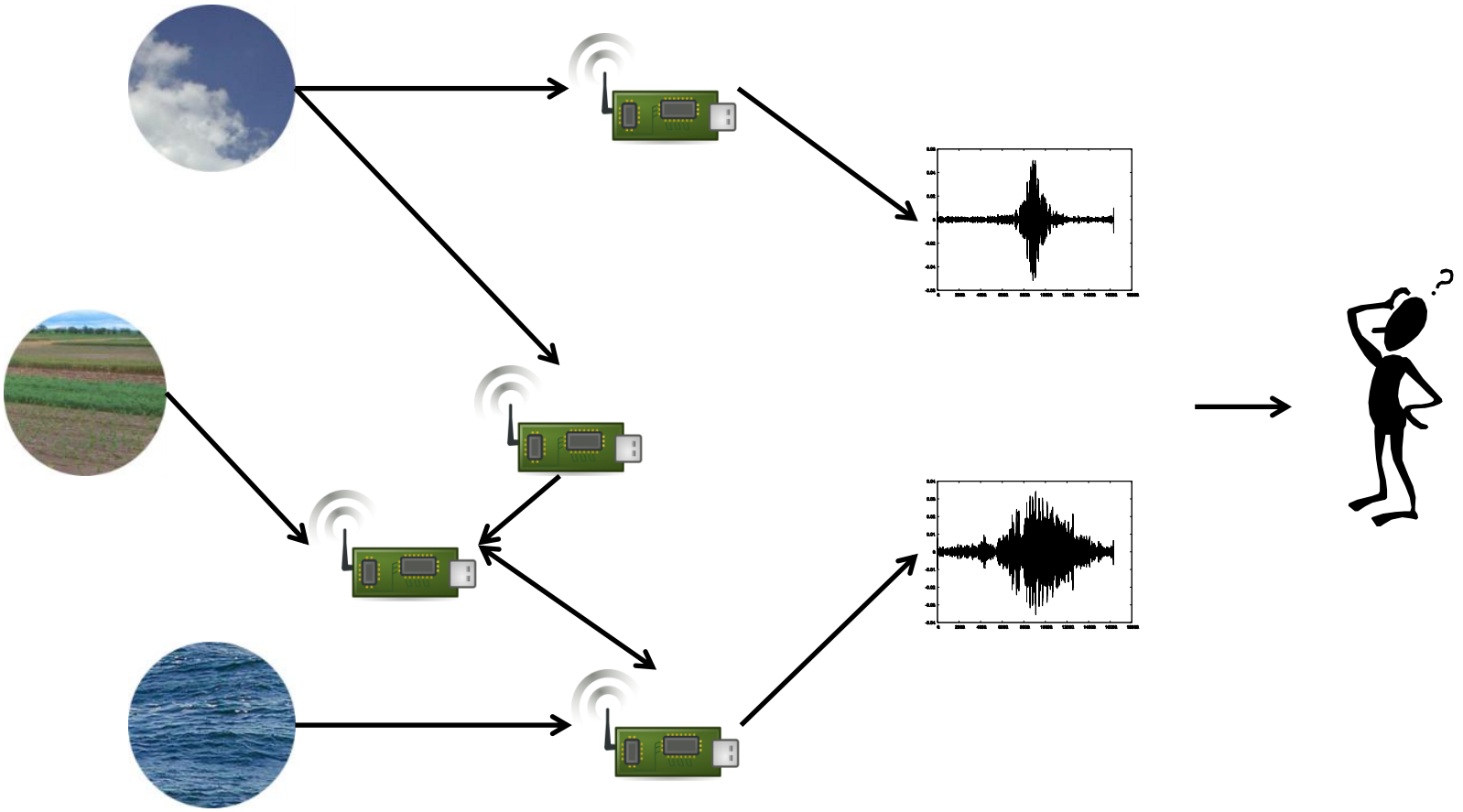
Markus Stocker

Journal Club in Biogeochemistry, Friday 3, 2012

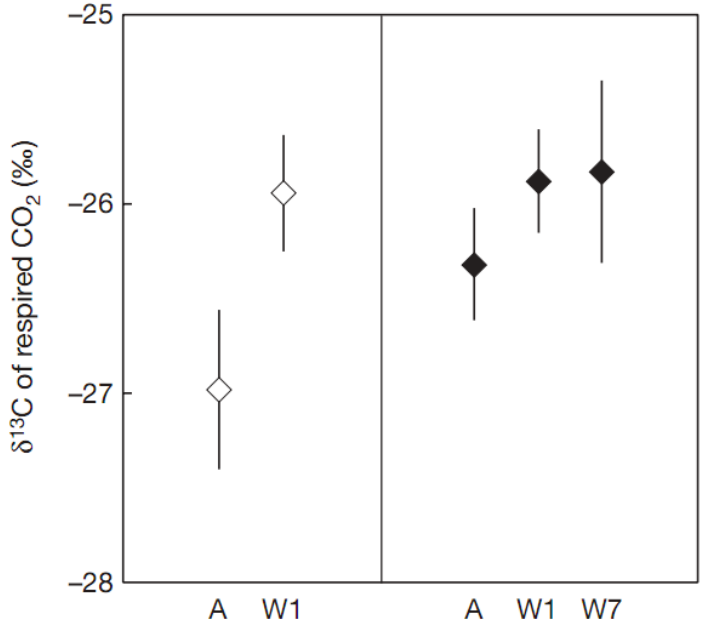
Too much data and not enough knowledge*: What to do with a 100 million data points

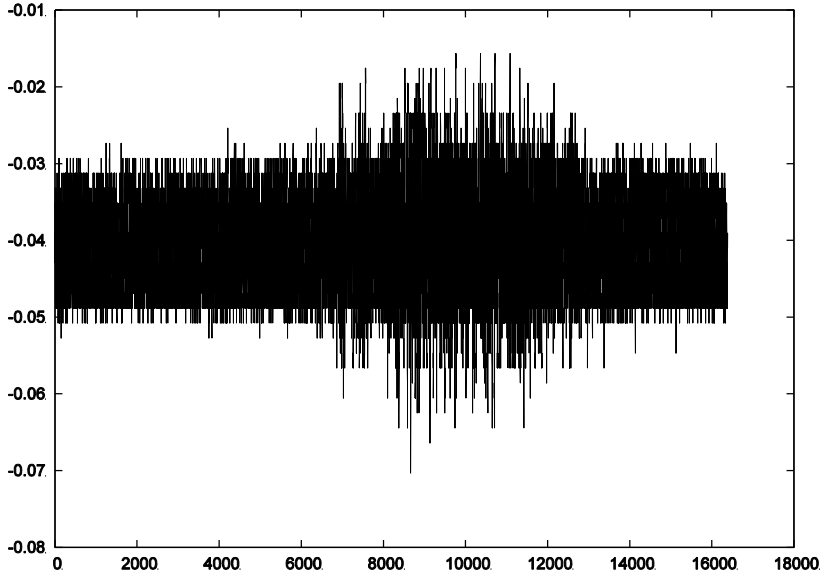
UNIVERSITY OF
EASTERN FINLAND

Domain



Domain (cont'd)



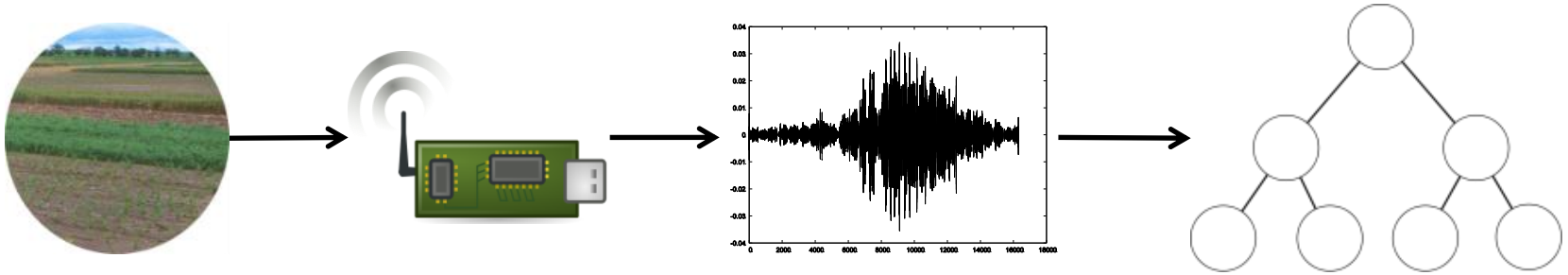


Two values in one year
Three values in seven years [1]

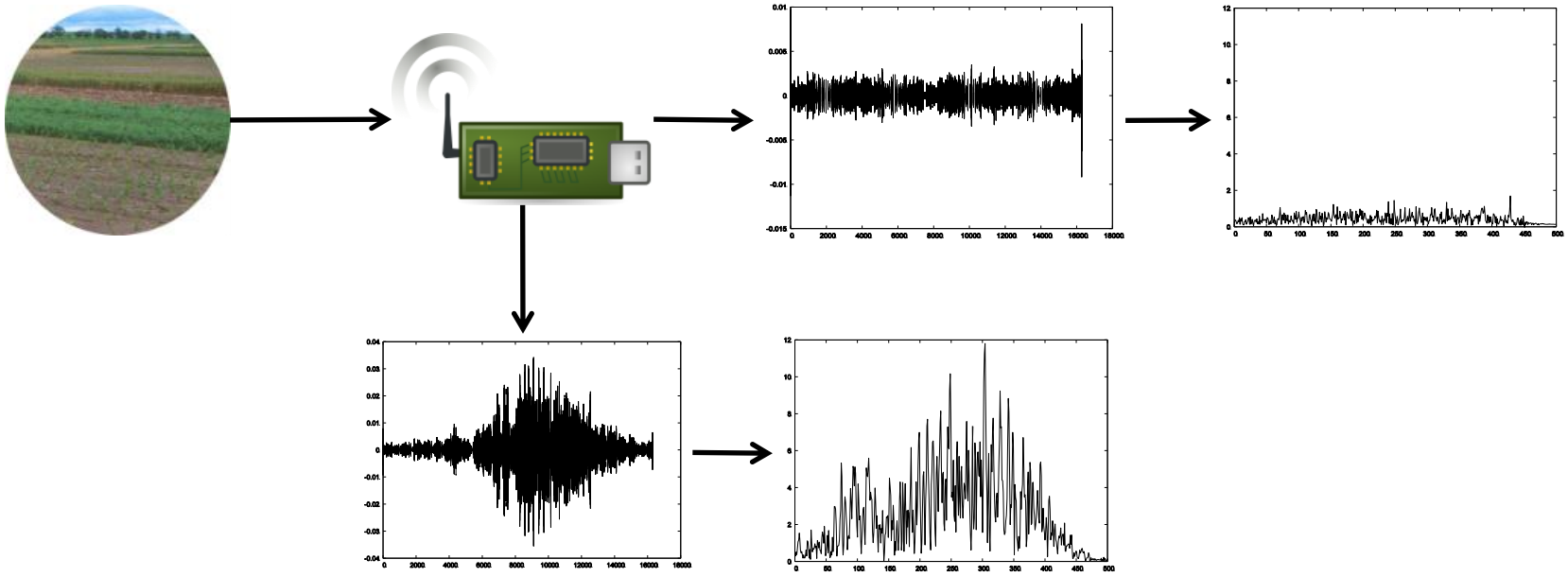
16'384 values in 8.192 seconds

Challenges

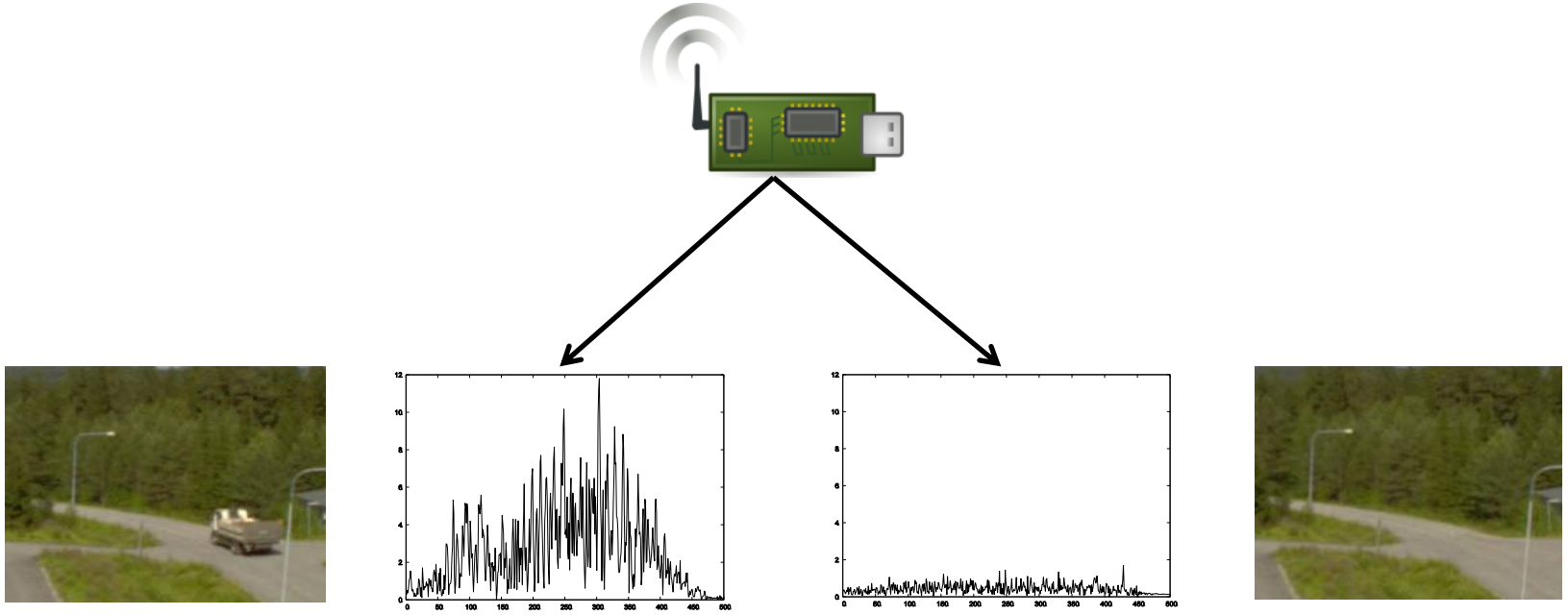
- Data acquisition, processing, and storage
- Making sense of sensor data [2]
  - How can knowledge be acquired
  - How can such knowledge be represented
- Due to the scale this should be automated



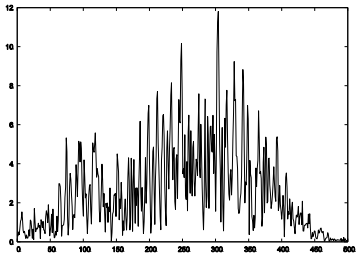
Data acquisition and processing

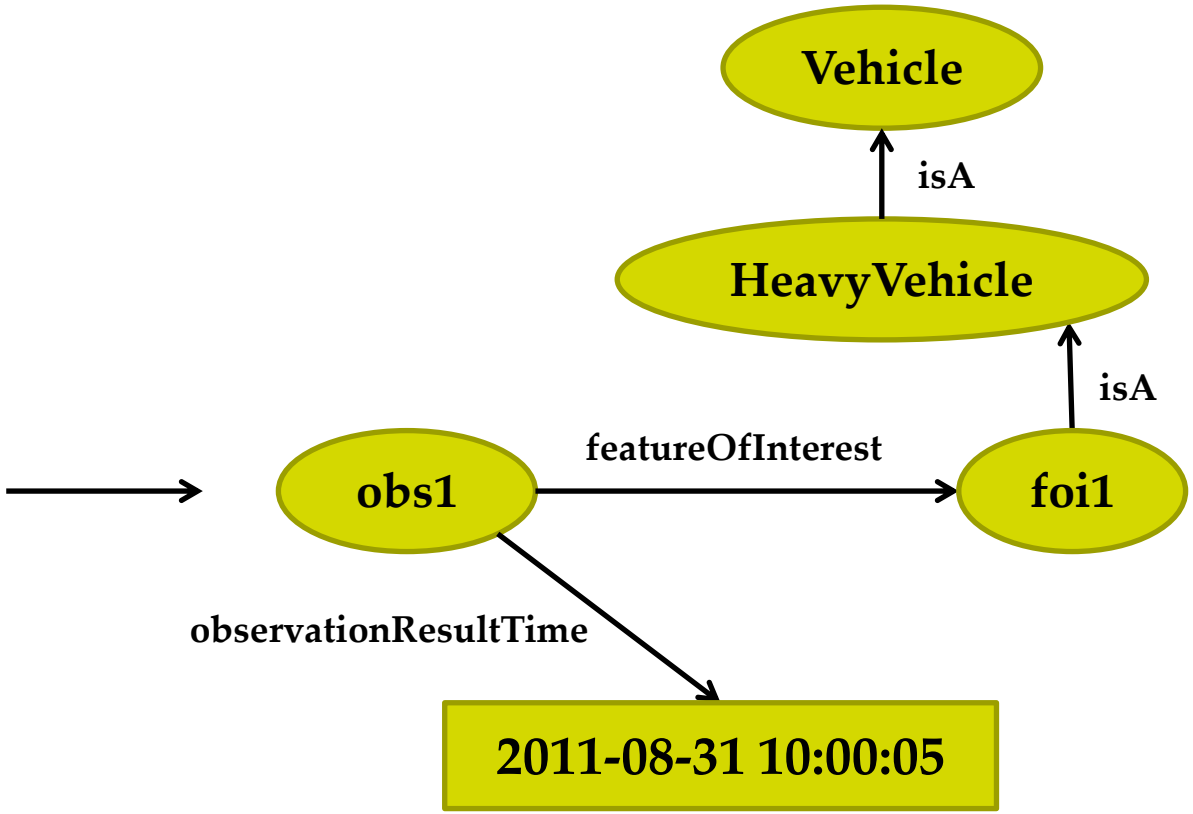


Patterns of interest in sensor data

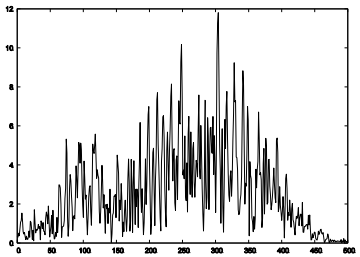


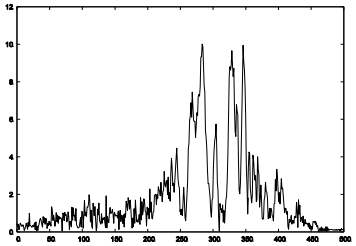
Knowledge acquisition and representation



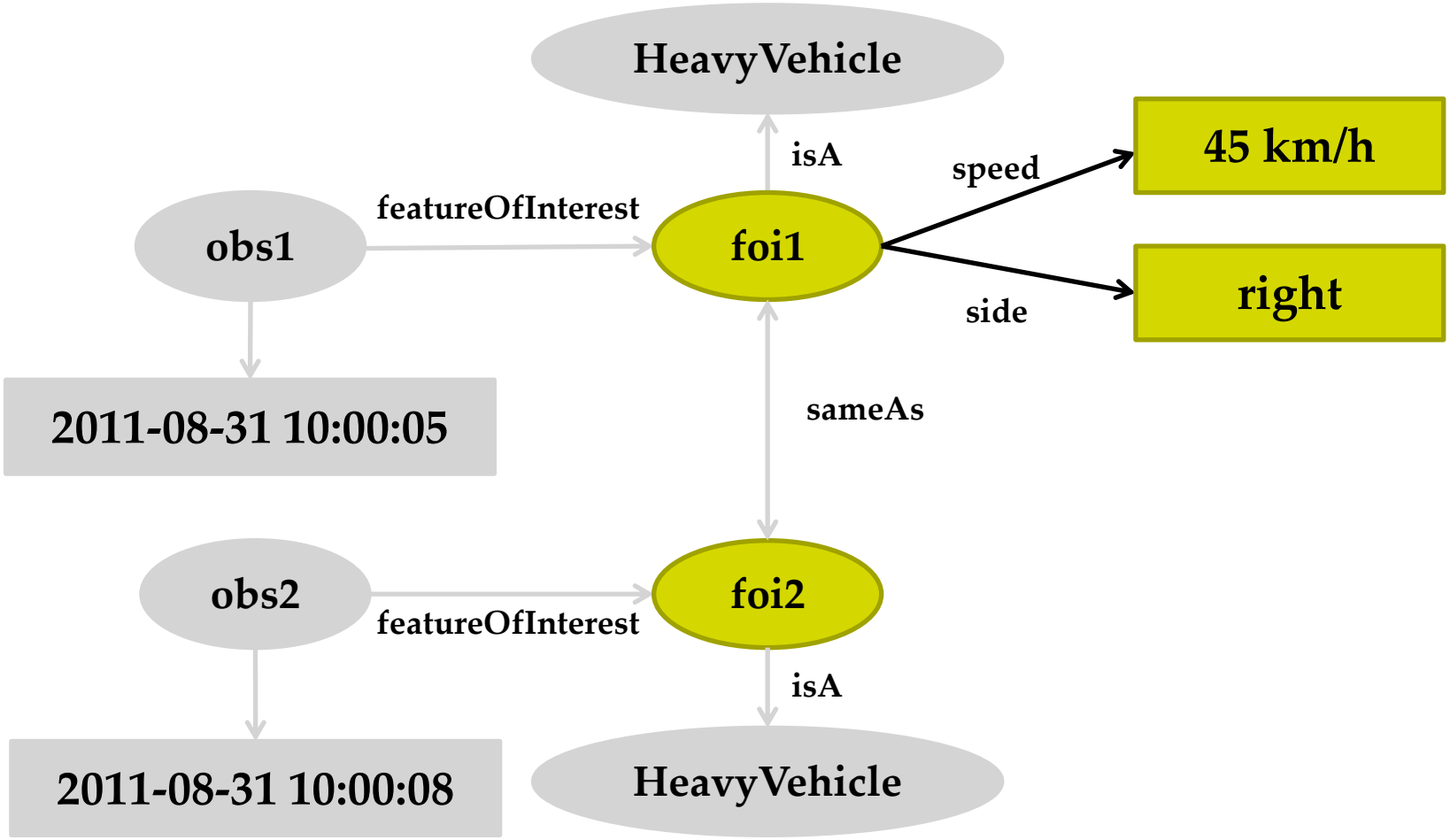


Knowledge acquisition and representation

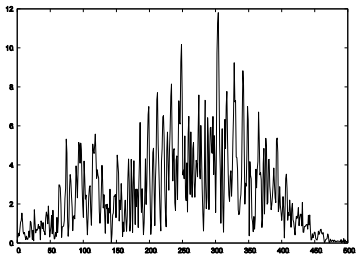


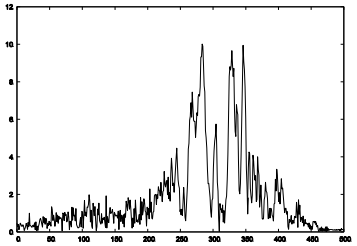


Knowledge acquisition and representation



Knowledge acquisition and representation





Conclusions

- Sensors measure the signal of phenomena in the environment
  - Hence, we operate in a specific domain, an area of interest
- Sensors and sensor networks can produce a lot of data
  - We are typically not interested in such data, but what does tell
  - Due to the scale, manual processing and analysis difficult
  - Hence, automatic methods are needed
- Automated workflow
  - Sensor data acquisition and processing
  - Knowledge acquisition and representation
- Knowledge represented formally may lead to
  - Benefits such as inference, consistency checking, query, visualization

References

(*) Sheth, A., C. Henson, and S. Sahoo (2008). Semantic Sensor Web. Internet Computing, IEEE, 12(4):78-83.

[1] Dorrepaal, E., et al. (2009). Carbon respiration from subsurface peat accelerated by climate warming in the subarctic. Nature 460: 616-619.

[2] http://www.nature.com/news/2011/110809/full/476135a.html

[3] Finkelstein, L. Theory and Philosophy of Measurement. In Sydenham, P. and Thorn, R., Handbook of Measurement Science, Volume 1, Theoretical Fundamentals, pages 1-30. John Wiley & Sons, 1982.