A distinction between data & information is pointless

Bas van Gils
April 2023

1. Introduction

Prof. dr. Bas van Gils

M +31 (0)6 48432088

E bas.vangils@strategy-alliance.com

IN https://www.linkedin.com/in/basvg/

"in an increasingly digital world, you have to put the people first"

Professional roles

Strategy Alliance : managing partner

DAMA Netherlands: board member

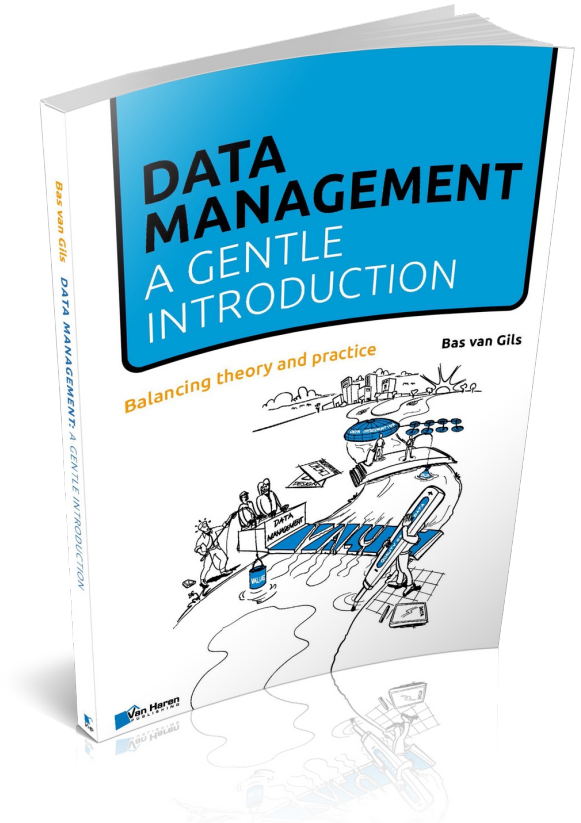
Antwerp Management School: trainer/researcher

Expertise

Business Transformation

Enterprise Architecture

Enterprise Data Management

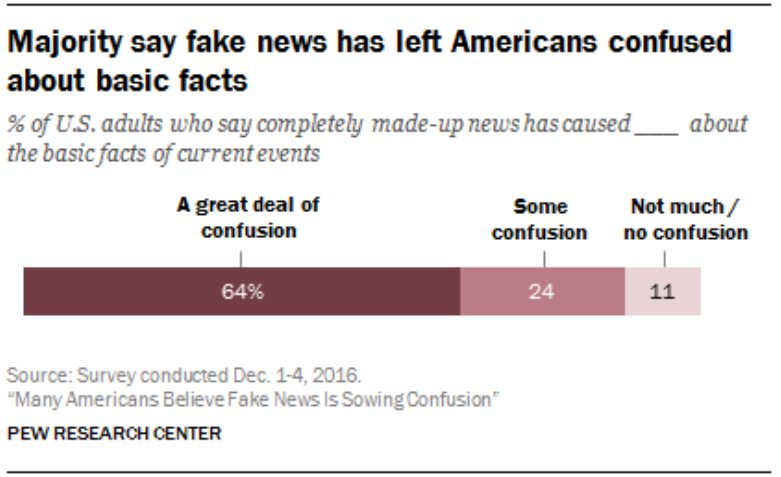


2. Data & information: pointless?

Data

Data is everywhere. The volume grows exponentially. Data-driven is "the" buzzword in business. Unfortunately, data use is not always as succesful as it may seem.

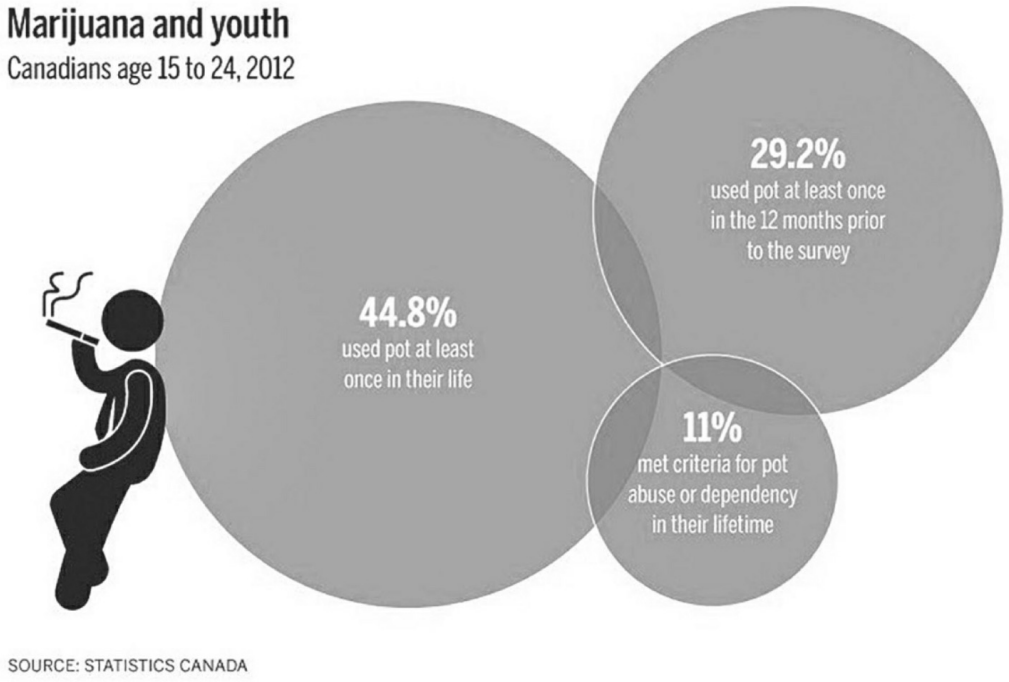
https://museumplantinmoretus.be/en



https://www.pewresearch.org/journalism/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/

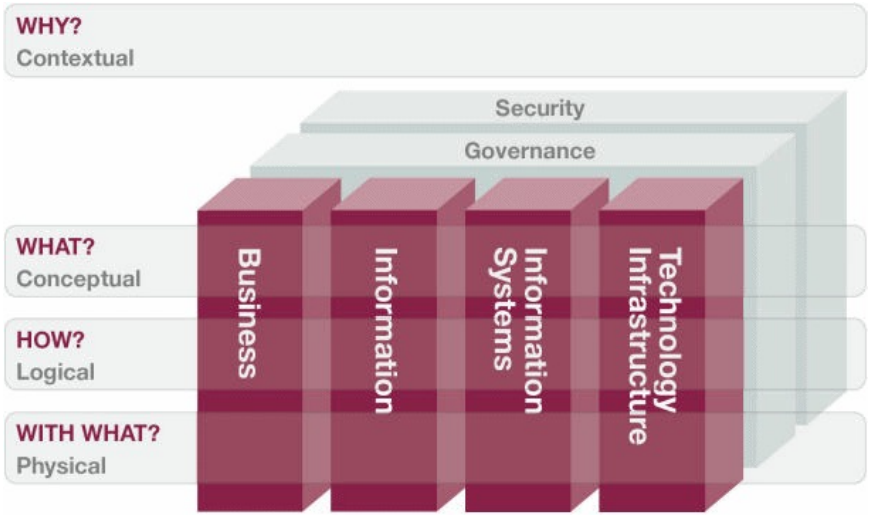
Statistics?

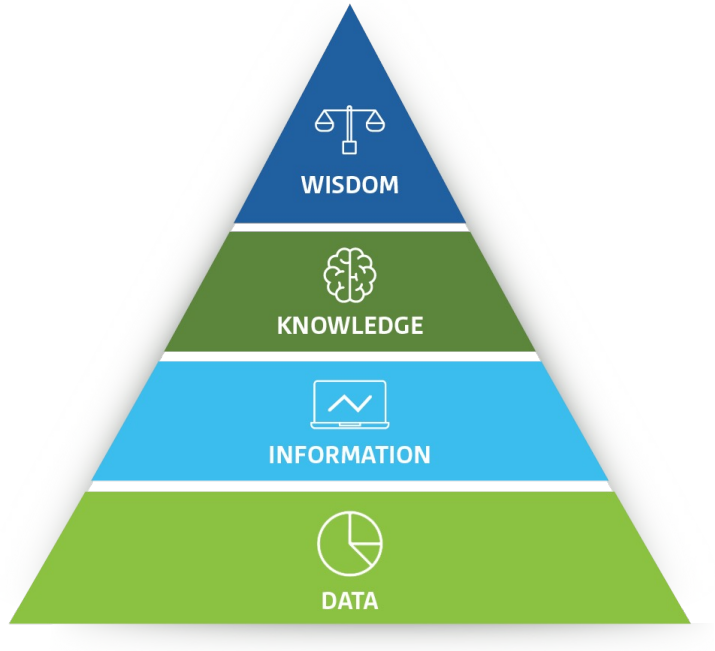
Sometimes we really like to lie with data. Graphs look impressive but are they really true?



Modelling

We typically distinguish between "data modelling" and "information modelling".

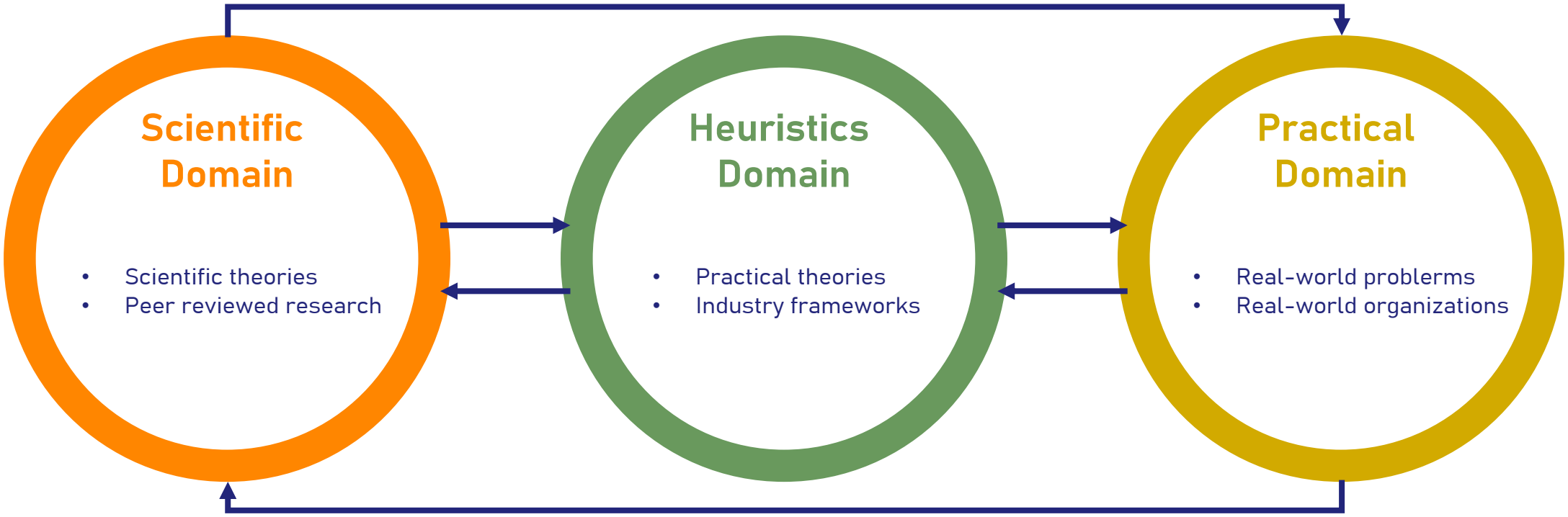




3. Data & information: a scientific perspective

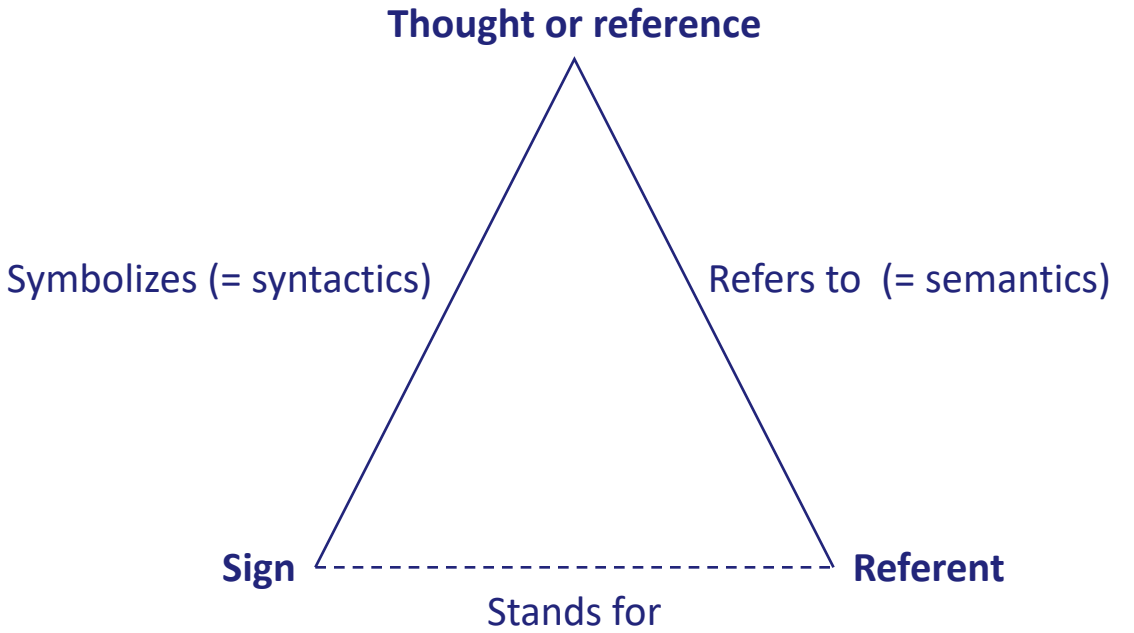
Warning

We are switching from the "heuristic domain" to the "scientific domain"

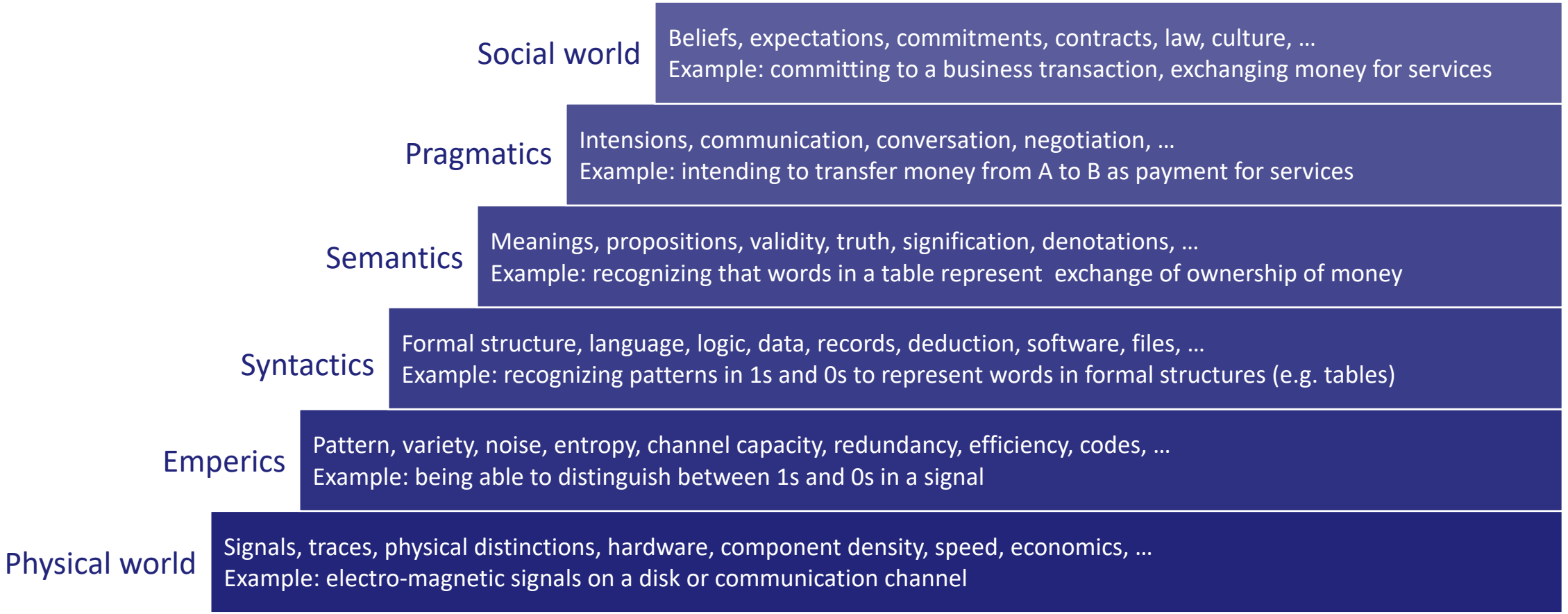


Semiotic triangle

Data symbolizes our understandings referring to / about referents, and therefore stand for that referent.

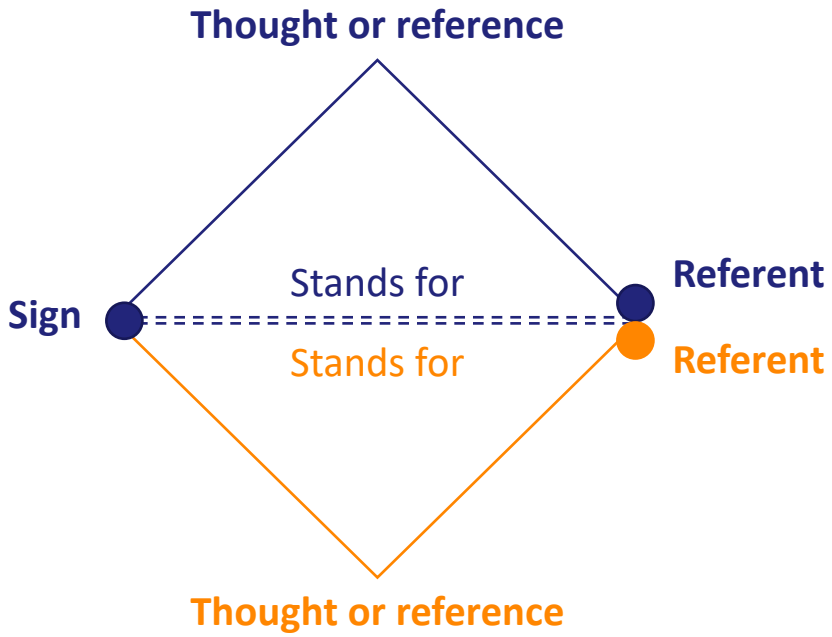


Semiotic ladder



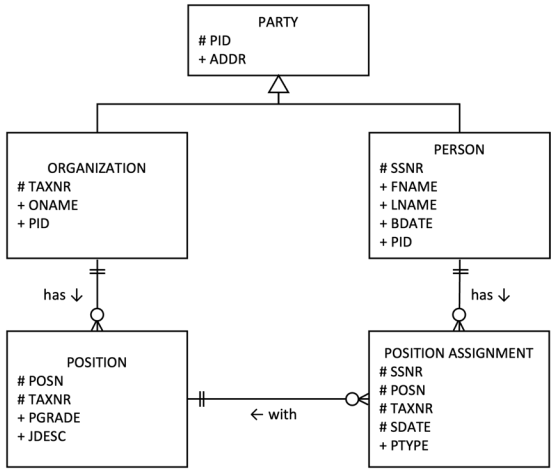
A chalenge that remains

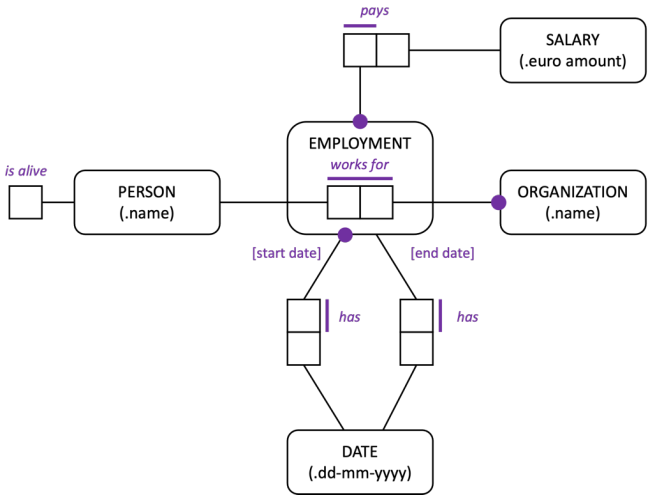
My (intended) semantics versus your (observed) semantics

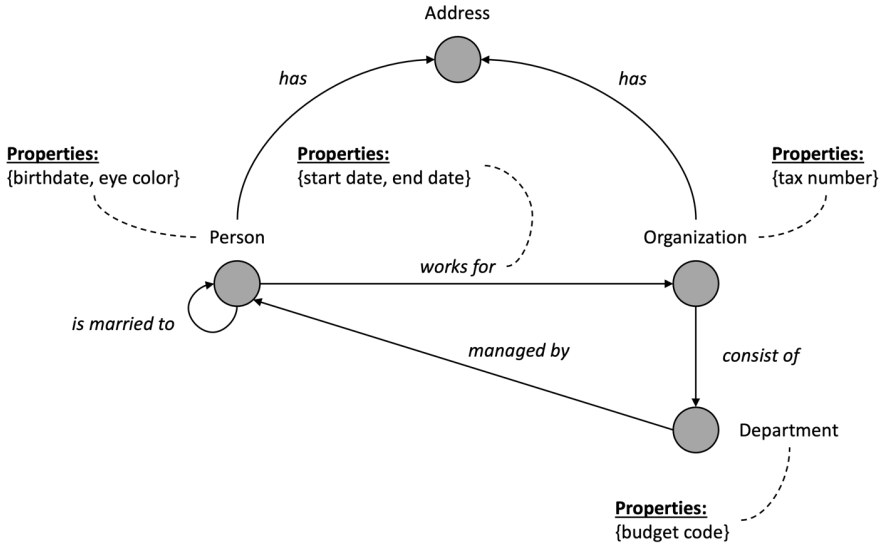


Why bother with data modeling?

Modeling data helps to (a) understand the structure and meaning of data, and (b) to design effective data structures. This can be done in different paradigms.







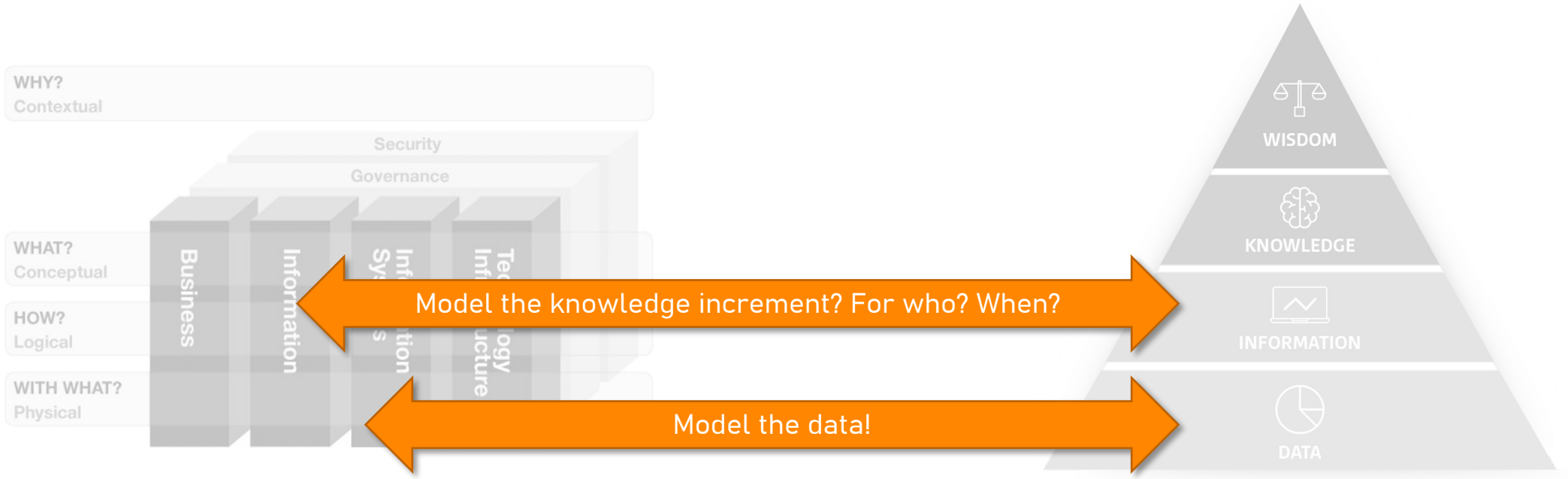
My position

- *Data is a representation of what we believe to be true in a domain, such that it can stand for that domain.*
- *Data can and is useful to be modelled.*

- *Information is the "knowledge increment" that occurs in the mind of an observer. It is highly subjective and situational.*
- *Information can be modelled, yet it is unclear how (hint: needs a representation)*

Modeling revisisted

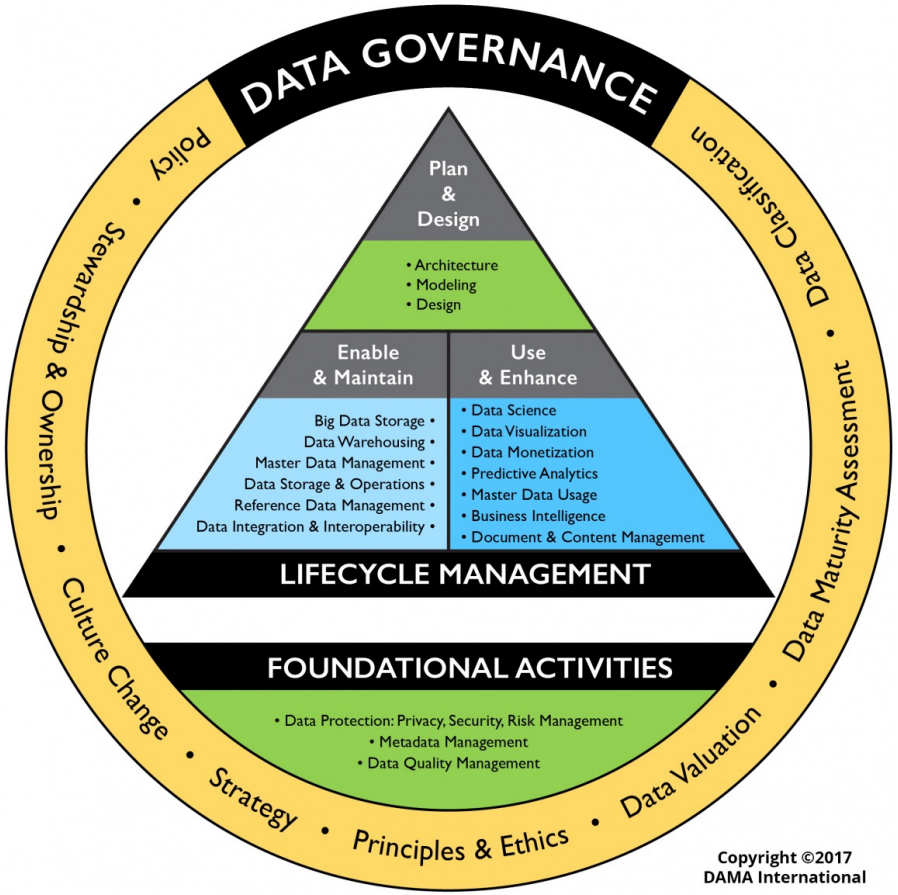
In the context of modeling/ understanding data/ doing cool things with data: avoid the term "information". It is meaningless and pointless (in this context)



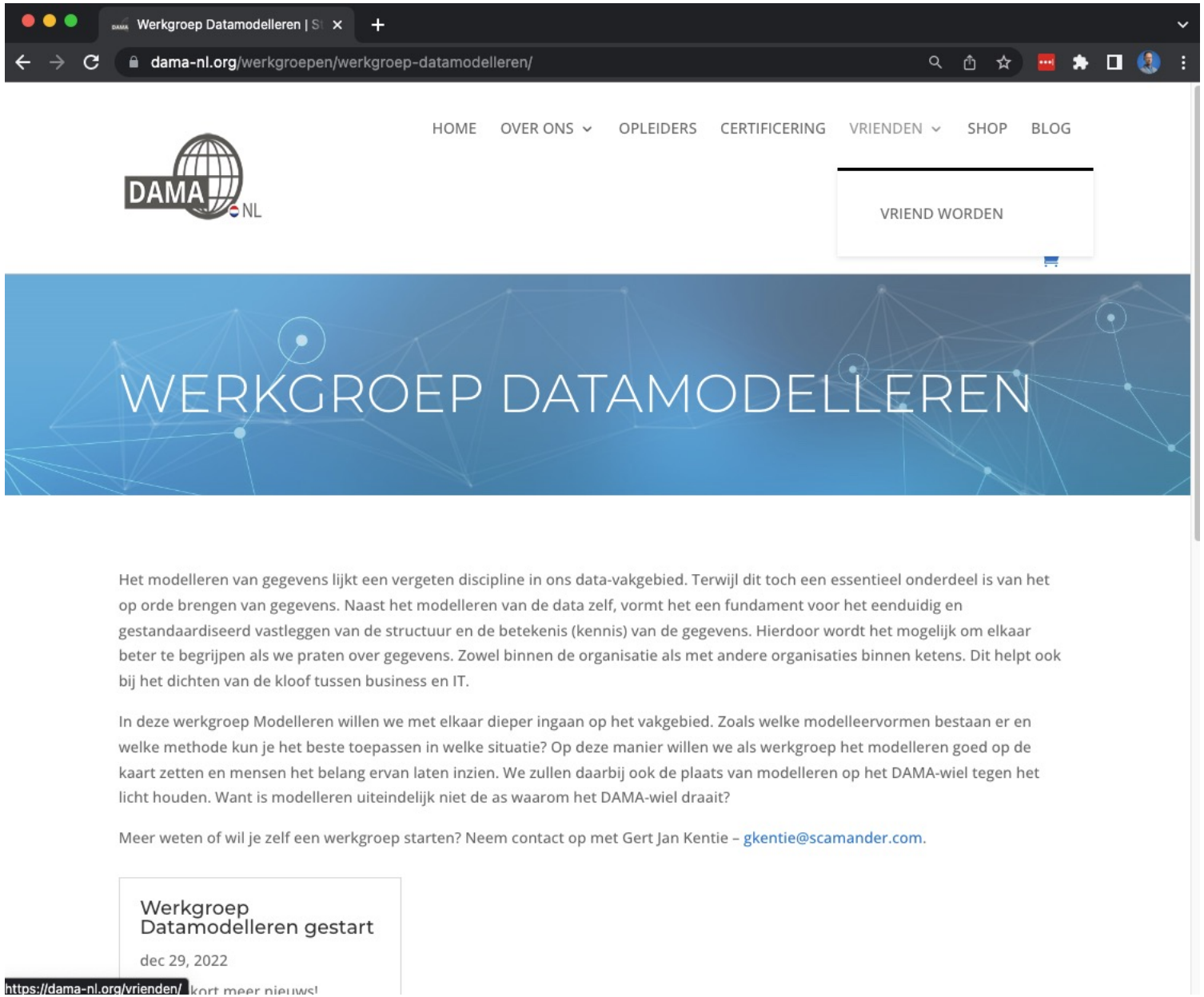
4. Data management

Data modeling as enabler

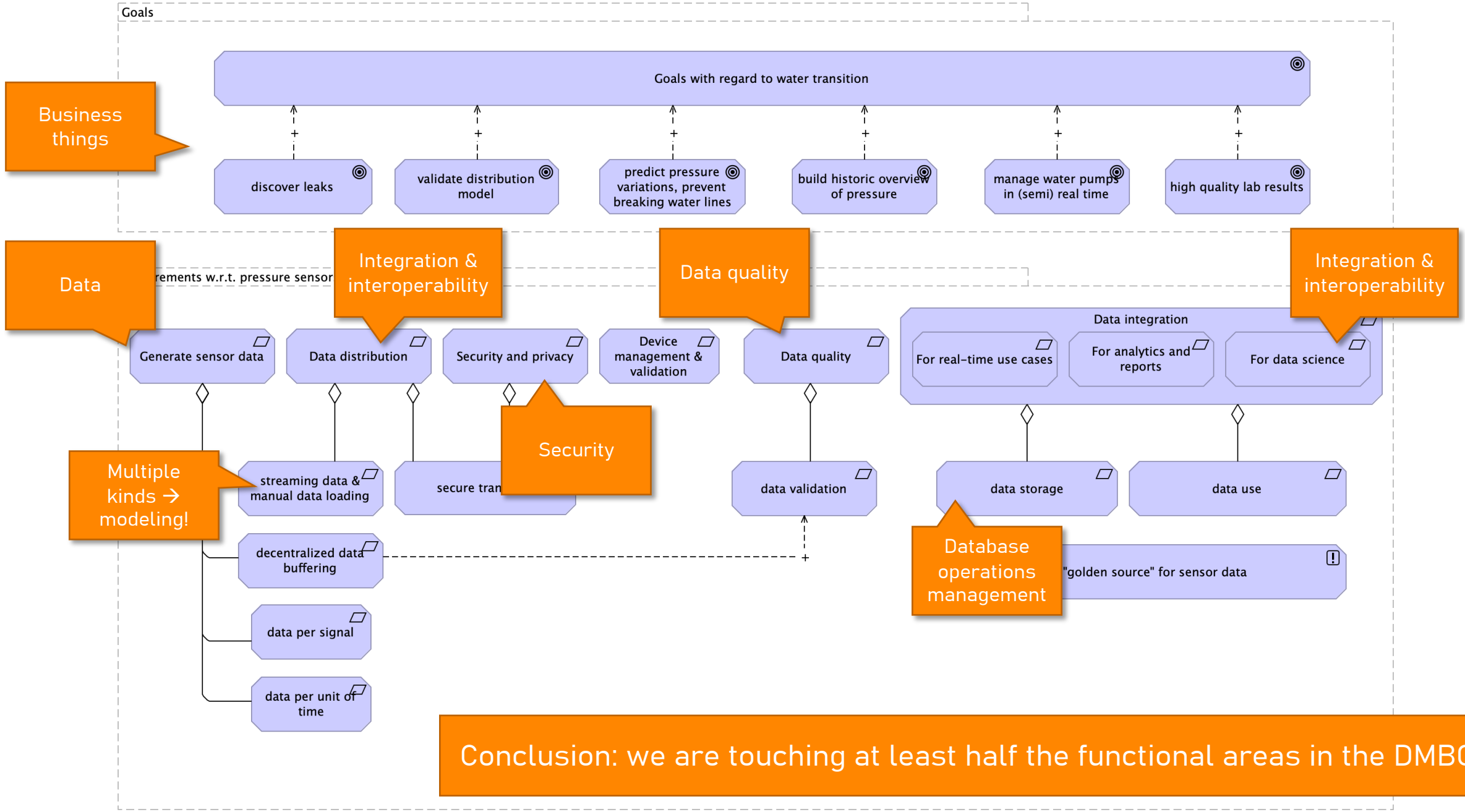
If you want to use data then you have to take good care of it: data management. Data modeling is crucial for understanding data. Without this understanding you can neither manage nor use it.

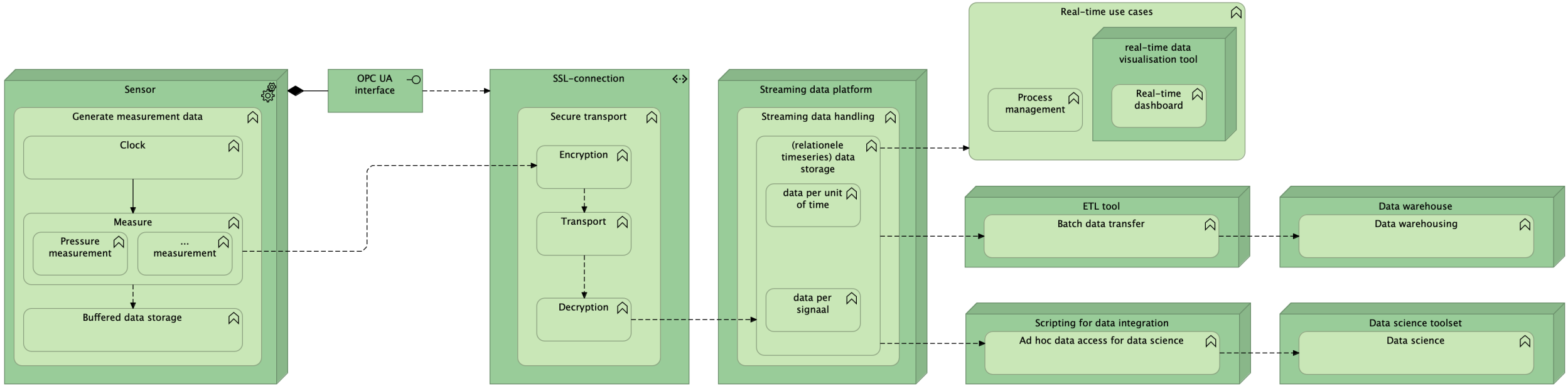


Copyright ©2017
DAMA International

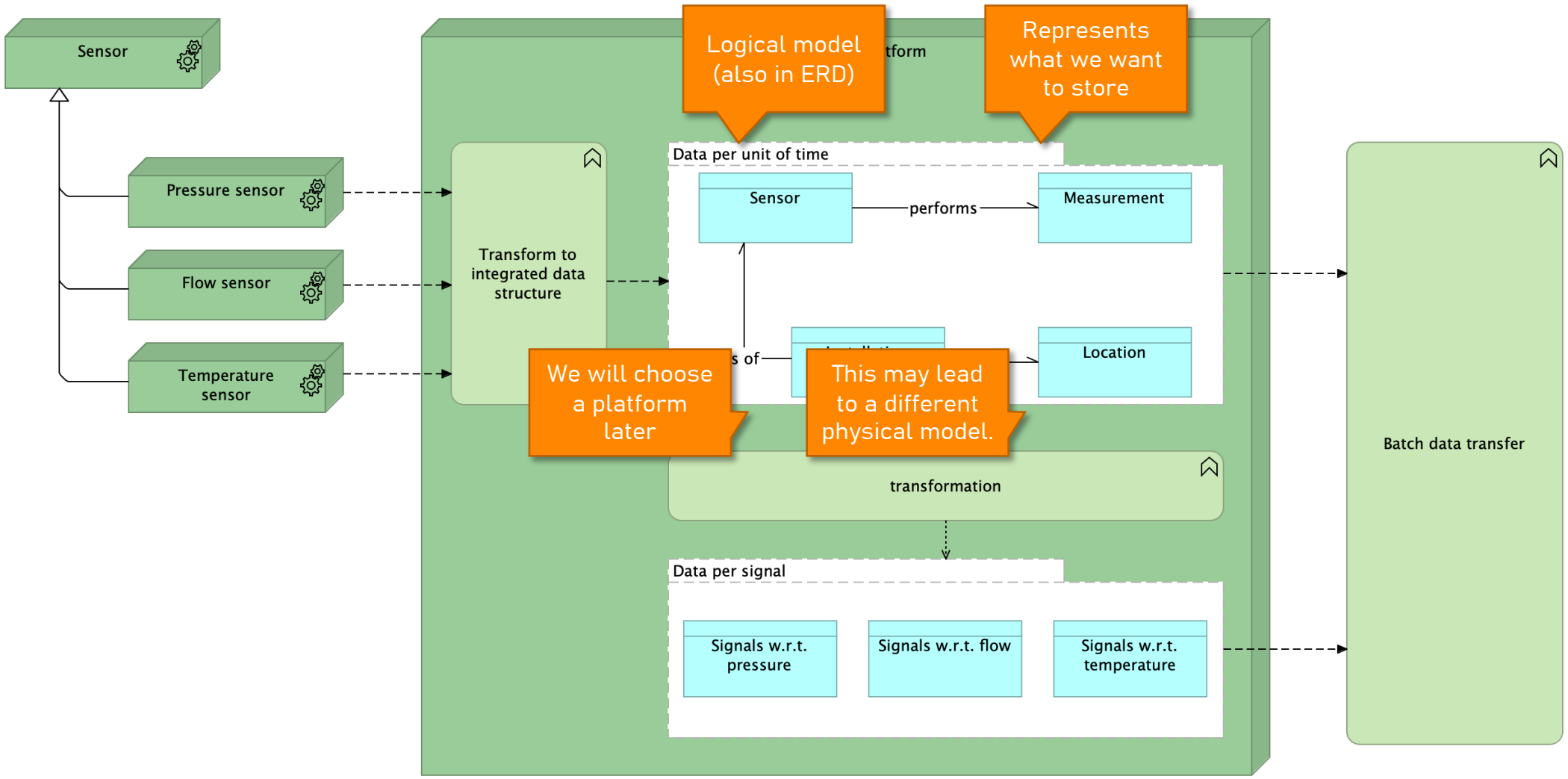


5. Case study

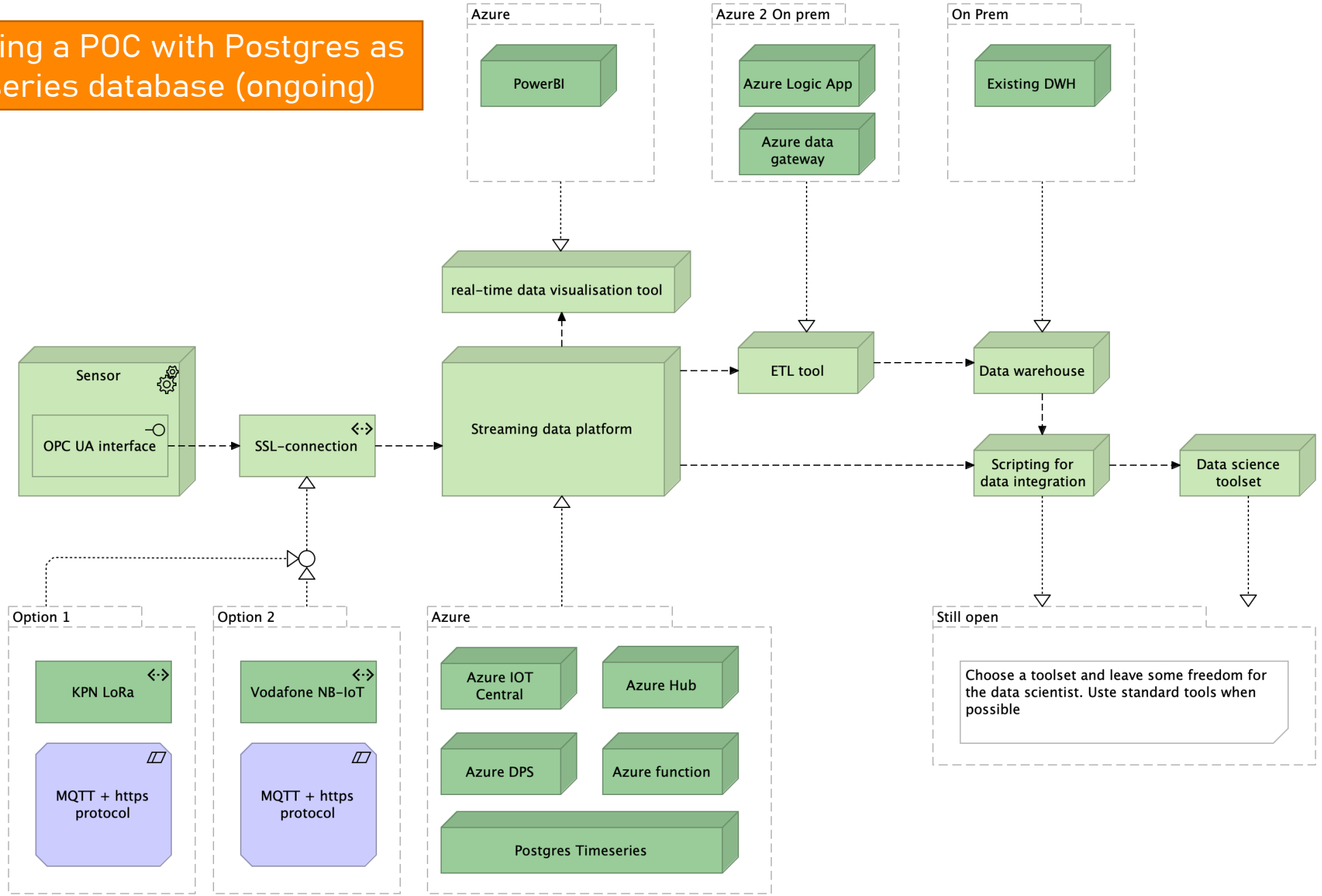

Goals
Goals with regard to water transition
discover leaks
validate distribution model
predict pressure variations, prevent breaking water lines
build historic overview of pressure
manage water pumps in (semi) real time
high quality lab results
Business things
Data
Integration & interoperability
Data quality
Integration & interoperability
rements w.r.t. pressure sensor
Generate sensor data
Data distribution
Security and privacy
Device management & validation
Data quality
Data integration
For real-time use cases
For analytics and reports
For data science
Security
Multiple kinds → modeling!
streaming data & manual data loading
secure tran
data validation
data storage
data use
Database operations management
decentralized data buffering
"golden source" for sensor data
data per signal
data per unit of time
Conclusion: we are touching at least half the functional areas in the DMBOK.



First function, then construction. Translate requirements to a solution


Sensor
Pressure sensor
Flow sensor
Temperature sensor
Logical model (also in ERD)
Represents what we want to store
Transform to integrated data structure
Data per unit of time
Sensor
performs
Measurement
Location
We will choose a platform later
This may lead to a different physical model.
transformation
Data per signal
Signals w.r.t. pressure
Signals w.r.t. flow
Signals w.r.t. temperature
Batch data transfer

We are doing a POC with Postgres as a time-series database (ongoing)



Reflection

Despite claiming that a) data = IT and b) we don't need it, we definitely wanted to achieve a specific and crucial business goal.
Understanding 1) what data is needed and 2) how it flows was crucial.

Side note: and it didn't take us all that long to get these insights!

6. Conclusion

Take-away

There is a difference between data & information. In the context of managing data as an asset in order to use it for value creation, this distinction is meaningless. Recommendation: try to understand your data as an enabler for good data management.

Don't forget to have some fun ☺

Sources

- Ackoff, R. L. (1989). From data to wisdom. *Journal of applied systems analysis*, 16 (1), 3–9.
- Date, C. J. (2019a). *Database design and relational theory: Normal forms and all that jazz*. Apress.
- Date, C. J. (2019b). *E. F. Codd and relational theory: A detailed review and analysis of Codd's major database writings*. Lulu Publishing Services.
- Gils, B. v. (2020). *Data management, a gentle introduction. Balancing theory and practice*. Van Haren Publishing.
- Gils, B. v. (forthcoming). *Data in context: Using models as an enabler for managing and using data.*
- Henderson, D. (Ed.). (2017). *DAMA DMBOK - data management body of knowledge*. 2nd editon. Technics Publications.
- Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning – a study of the influence of language upon thought and of the science of symbolism*. Magdalene College, University of Cambridge.
- Stamper, R. K. (1991). The semiotic framework for information systems research. *Information systems research: Contemporary approaches and emergent traditions*, 515–528

https://www.stuv.la/wp-content/uploads/2020/04/QA_Handwriting-QdA-rszd-800x423.png