

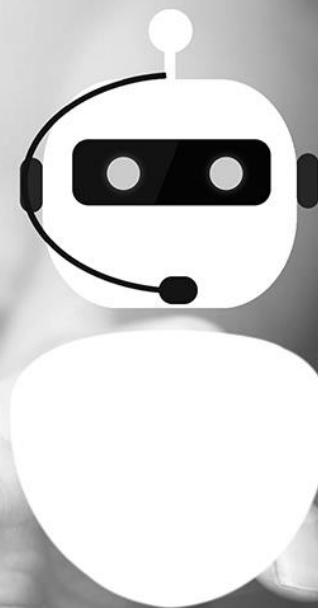


AI4VALUE

Use case-based algorithms

The creation of a data model

- The traditional way of doing things
- The AI way of doing things
- Practical approach to doing/checking a datamodel with the aid of AI
- Summary



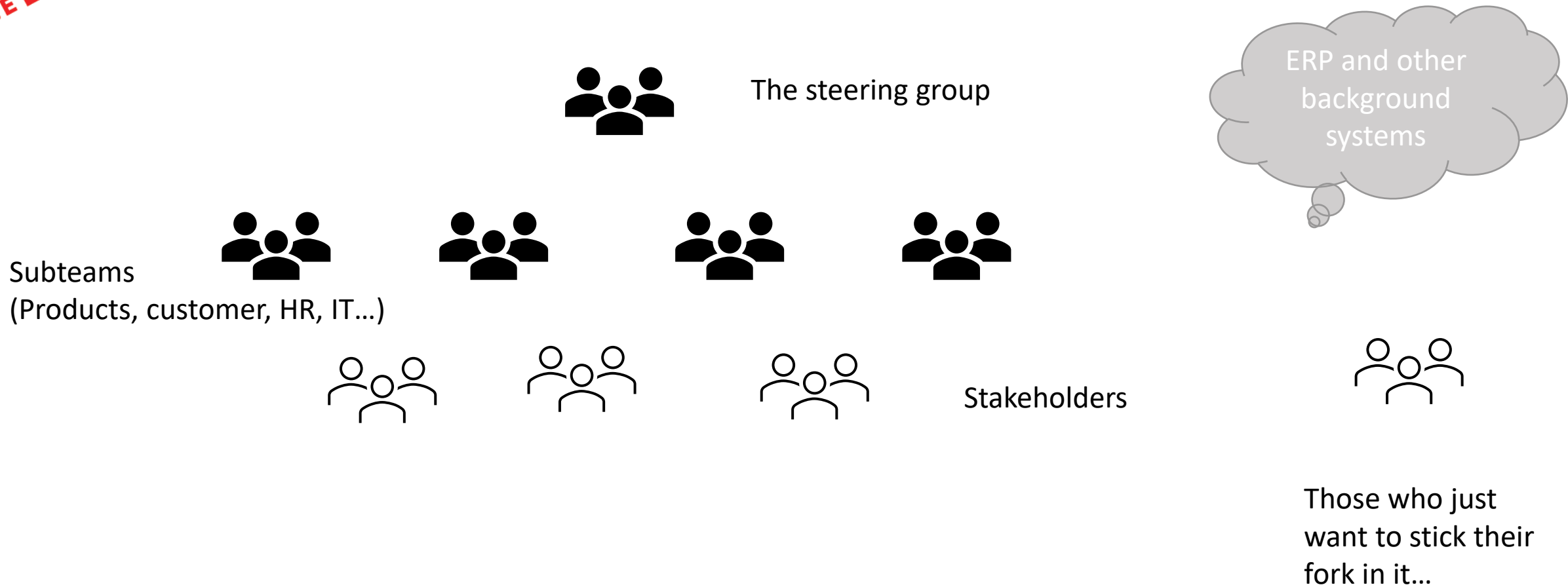
What can i help you with?

The traditional way





The traditional way – a committee and subteams



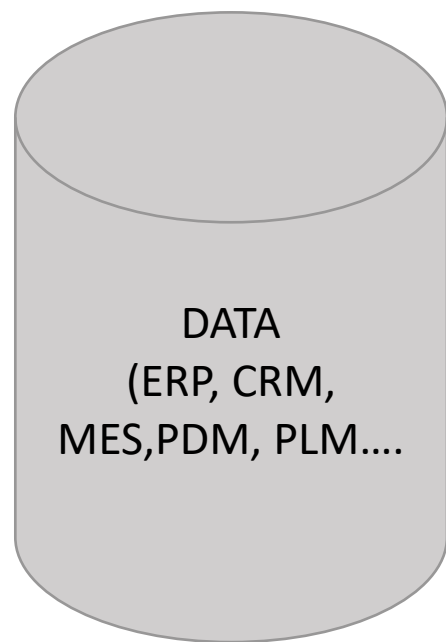
What happens?

- It takes YEARS
- Subteams fight among each other, "our model is the best model", "Our business line has very specific needs", "This change would ruin our reporting"...
- Steering team tries to compromise, has to trust what the different subteams and stakeholders tell them and realize just how difficult it is to consolidate.
- At the same time the ERP or some other technology roll-out is in full force and desperately need the new datamodel.
- 18 months and the data is again non-standard





The AI way



A sample of upper-level classification

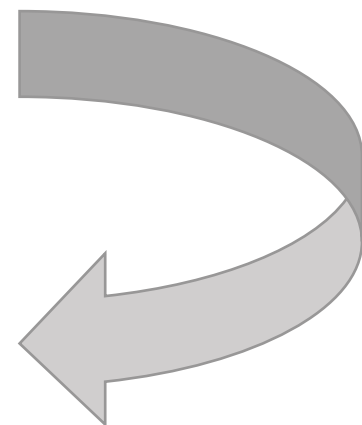


An ontology



A datamodel

Continuous cleansing





1) Data dump

- This, surprisingly, is often the most time-consuming part of the process
 - Data is scattered in various back-end systems, according to various system-required or legacy data models
 - There may be data warehouses – but even they are usually divided by function, not an all-encompassing datalake.
 - Getting data out of some legacy systems can be tricky
 - Output formats vary
 - IT personnel are too busy to look into ways to getting the export
 - Secure transfer to a service provider – a lot of companies want to do this via SharePoint – we prefer Secure FTP.



2) Upper-level classification – a sample

- There should be a rough idea of “what good looks like”
- This is usually not a problem, typically it is something like:
 - “product item”, “manufacturer”, “measurements”, “standards”
- The best results are achieved when the algorithm is “fed” a sample which on a high-level looks pretty good. The algorithm may disagree with this in the end, but it is still a good starting point.



3) An ontology.

- Be honest – how many of you have a defined ontology for your data?
- The definition of an ontology: “a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.”
- This is something we ALWAYS do in each and every NLP project. Automatically.
- It gives a completely neutral view of what the **data** tells us, not just what people assume.
- Example: [Ai4Value Ontology \(coronaontology.com\)](http://coronaontology.com)



4) Training the algorithm

- Based on the ontology, the algorithm is trained by giving the algorithm feedback.
- In data model cases this is typically done so, that the algorithm creates a suggestion – “does this look right to you”.
- In practise this is a subset of the data, an Excel sheet in its’ most basic form, that a human reviews and provides feedback.
- The feedback is fed to the algorithm and it starts to learn – based on both the ontology and the human feedback.
- It becomes more clever by each iteration and in the end it can classify the data based on the content itself.



5) The datamodel

- As the end result you have a data-based datamodel. That describes what you actually have, not what you assume the data should be like.
- There are usually always surprises, these wow-moments when people look at the model and say “ohhh, right, of course” ...



6) The maintenance

- Unlike with human resources, the algorithm can also keep your data clean and according to the data model.
- People will continue to make mistakes and ignore the data model, but the algorithm can keep this in check.
- The algorithm can be, for instance, made to run through all new data on a weekly/monthly basis and either point out to data owners that it has found an anomaly – or eventually, once trusted enough it can make the correction by itself.



Summary

- AI is a great tool for supporting your data model projects.
- The throughput time of such development is circa 2 – 4 months.
- The effort it requires from users is typically 2 hrs kick-off, 30-min scrum meetings each week – the algorithms do the rest.
- The AI approach can also take care of the maintenance phase and make sure that the data model keeps intact.
- That which used to be a multi-year challenge can now be solved quickly and painlessly – and mathematically, without emotion.

The end!

- Questions & comments?





Be in touch!
katriina@ai4value.com
+358505015403