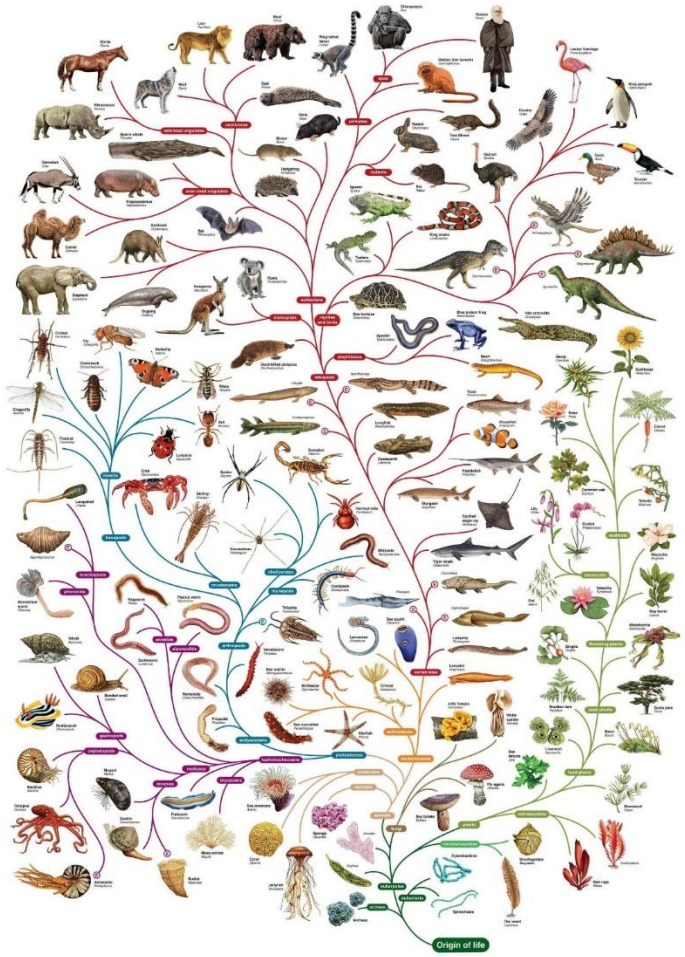
How to agree to disagree

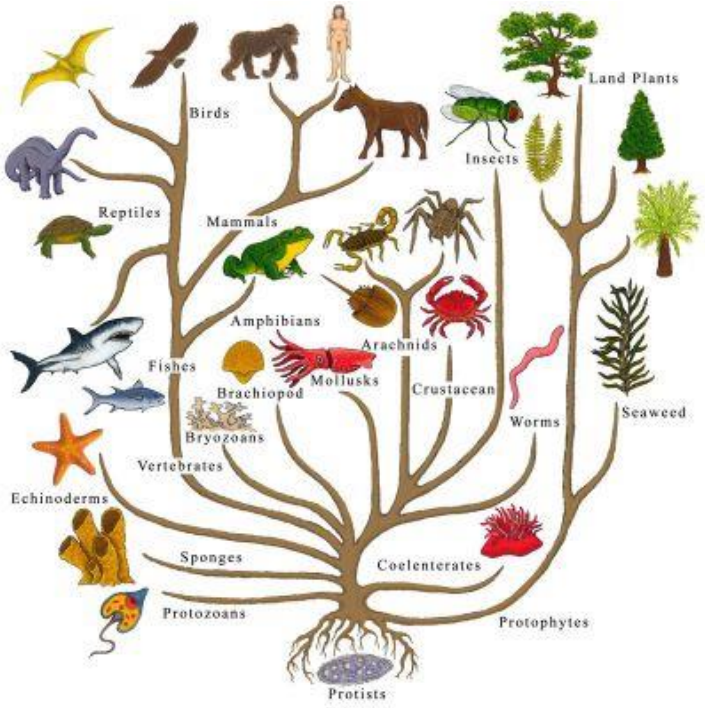
(on data solution automation)

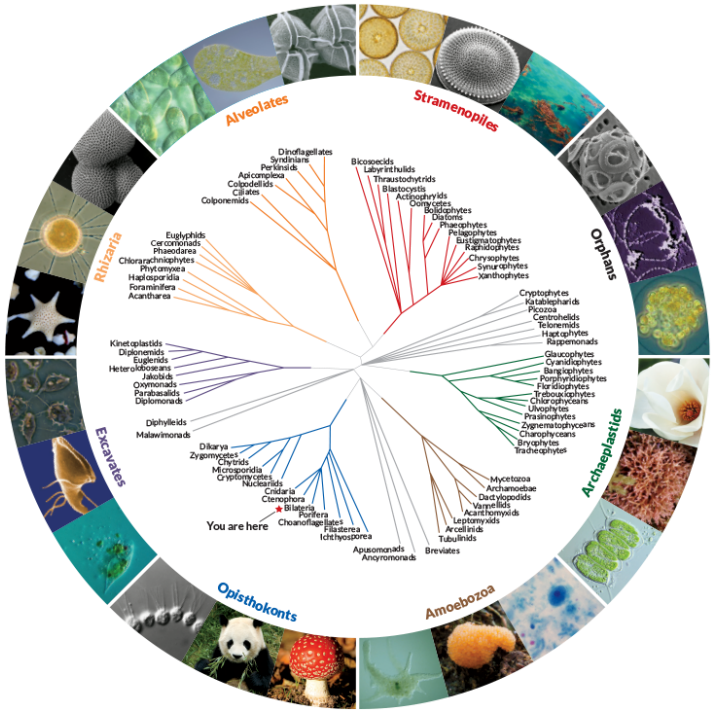
Roelant Vos

Budapest Data Forum - 2022

Reality?


Origin of life


Birds
Reptiles
Mammals
Amphibians
Fishes
Brachiopod
Mollusks
Arachnids
Crustacean
Insects
Land Plants
Worms
Seaweed
Bryozoans
Vertebrates
Echinoderms
Sponges
Protozoans
Coelenterates
Protophytes
Protists


Alveolates
Stramenopiles
Rhizaria
Orphans
Excavates
Archaeplastids
Opisthokonts
Amoebozoa
You are here

‘Inspired laziness’

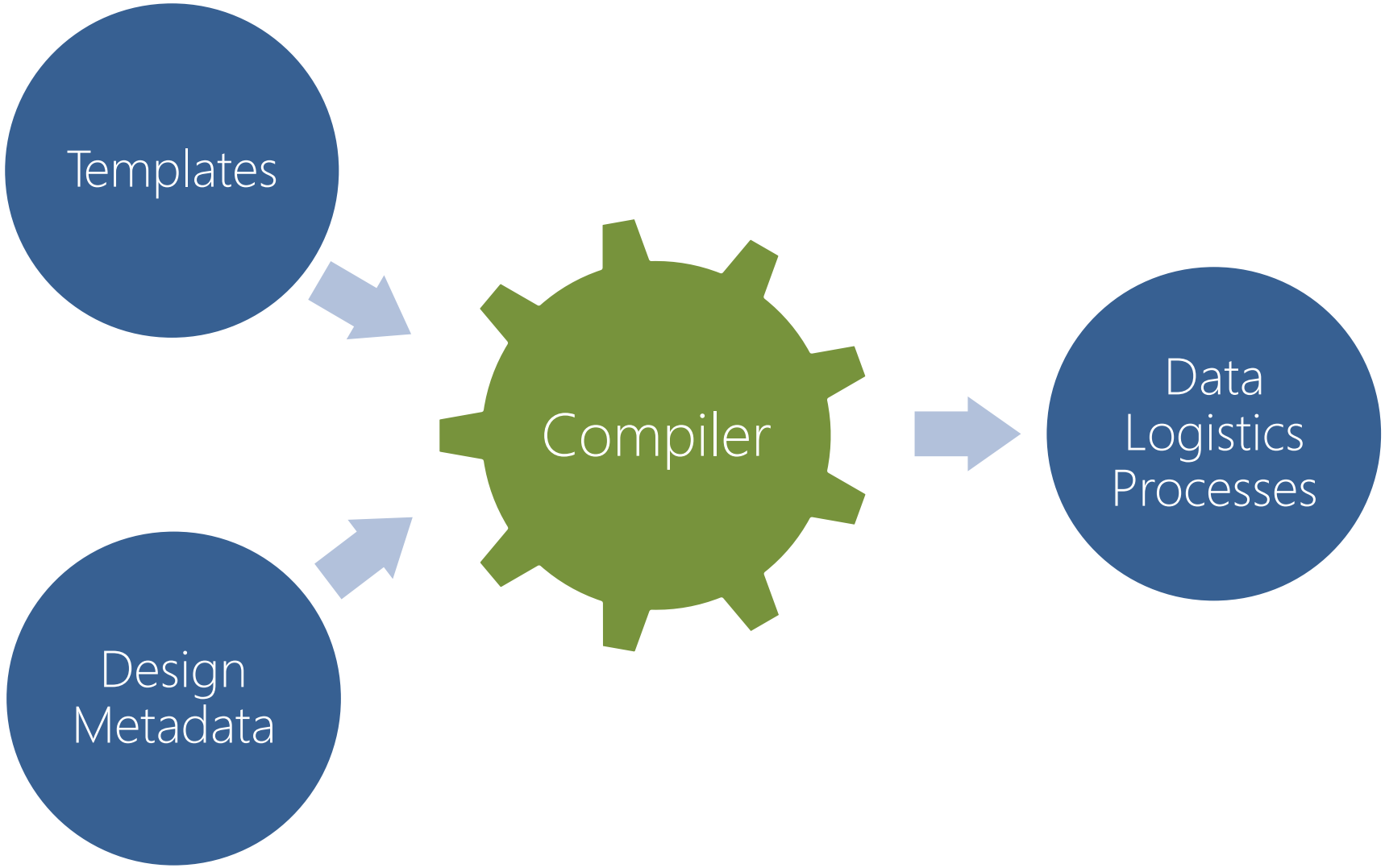
- Only lazy programmers want to write tools that *might replace them*
- Only lazy programmers will *avoid writing monotonous, repetitive code*
- The tools and processes inspired by laziness *speed up production*

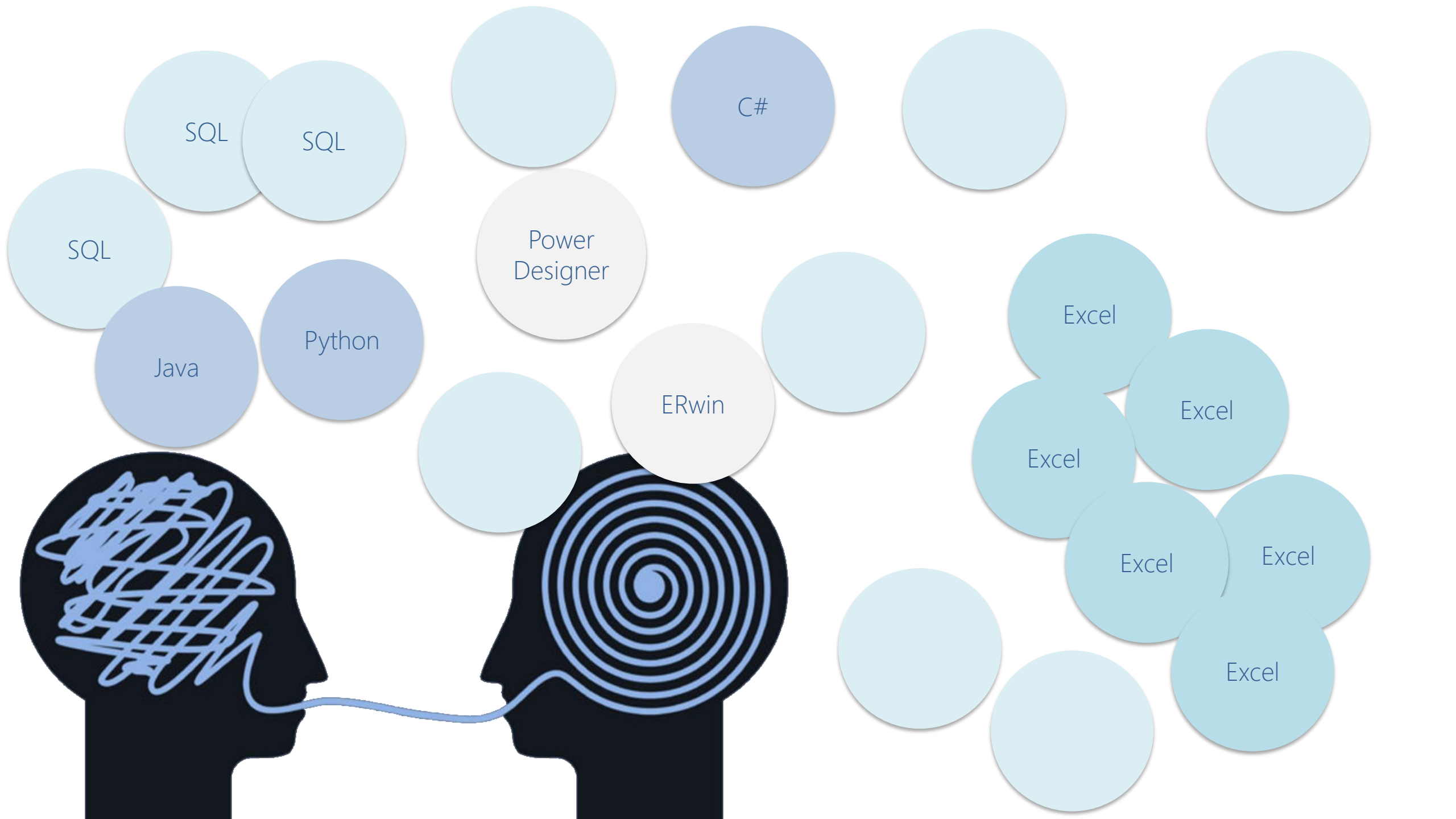
- However, for a lazy programmer to be a good programmer, he or she must be *incredibly ‘unlazy’* when it comes to learning how to stay lazy

www.roelantvos.com/blog

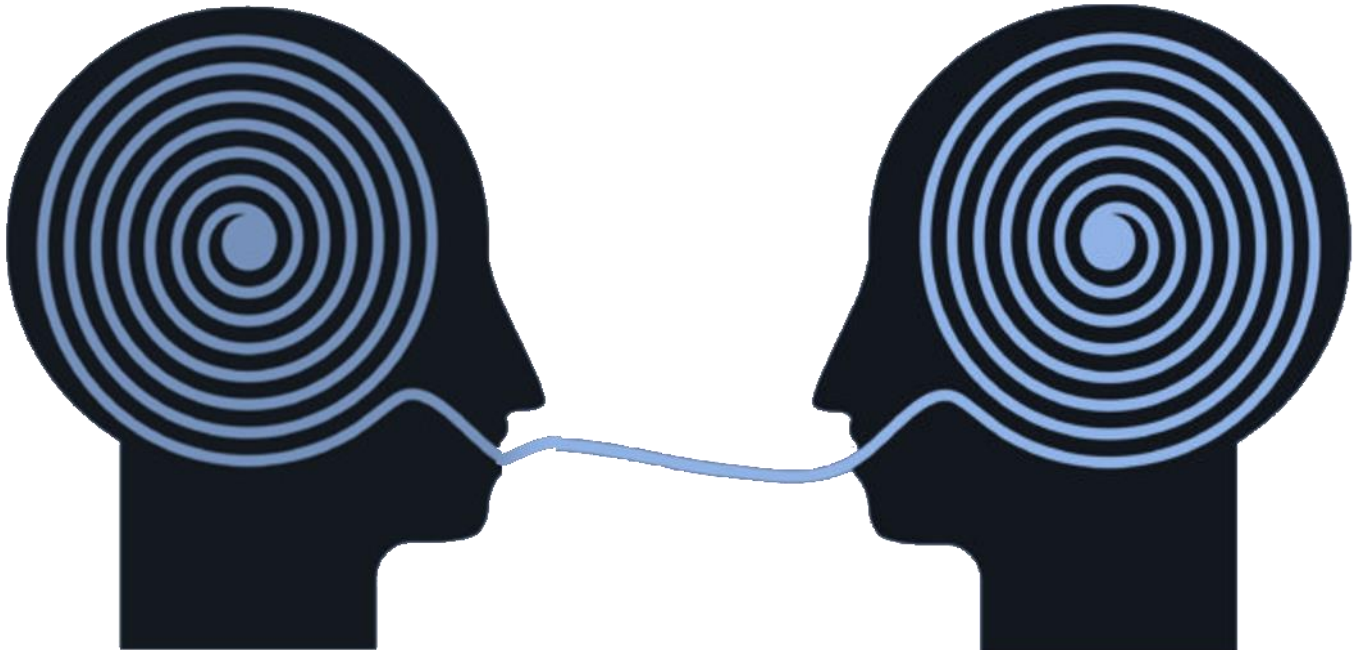
https://github.com/RoelantVos

www.varigence.com/blog


Templates
Design Metadata
Compiler
Data Logistics Processes


SQL
SQL
C#
SQL
Power Designer
Python
Java
ERwin
Excel
Excel
Excel
Excel
Excel
Excel

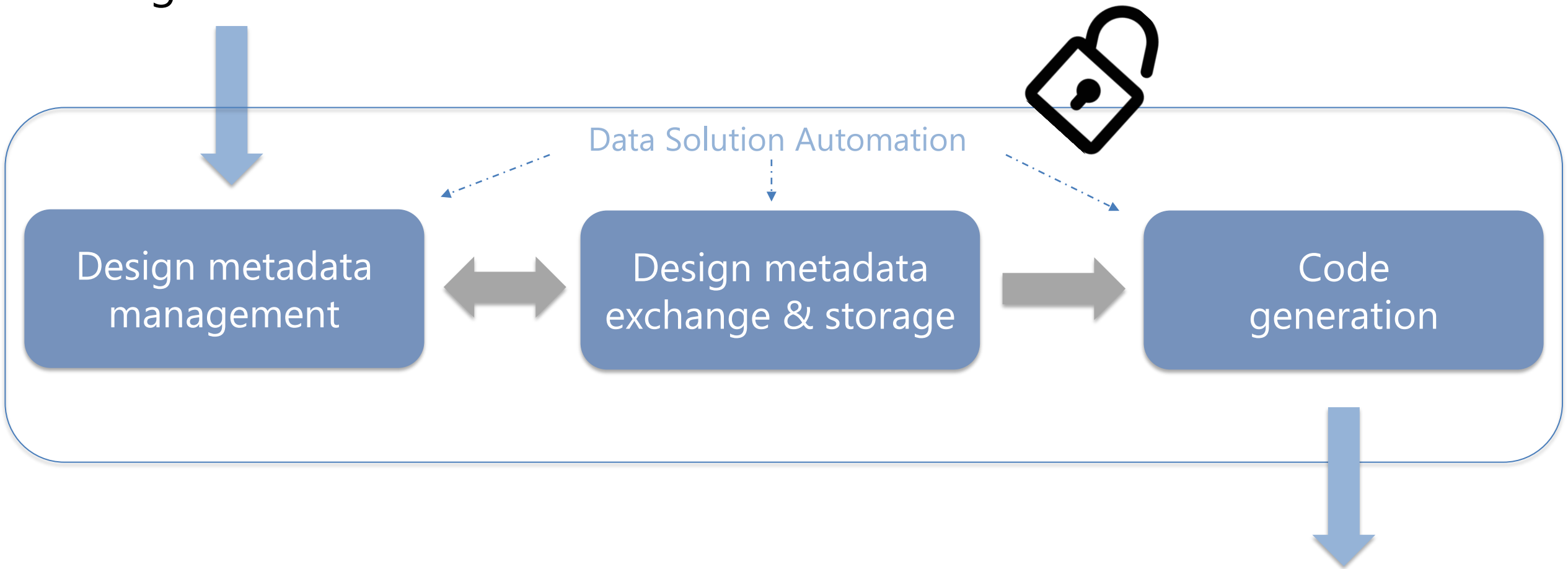
Imagine how much can we accomplish when we can work together?



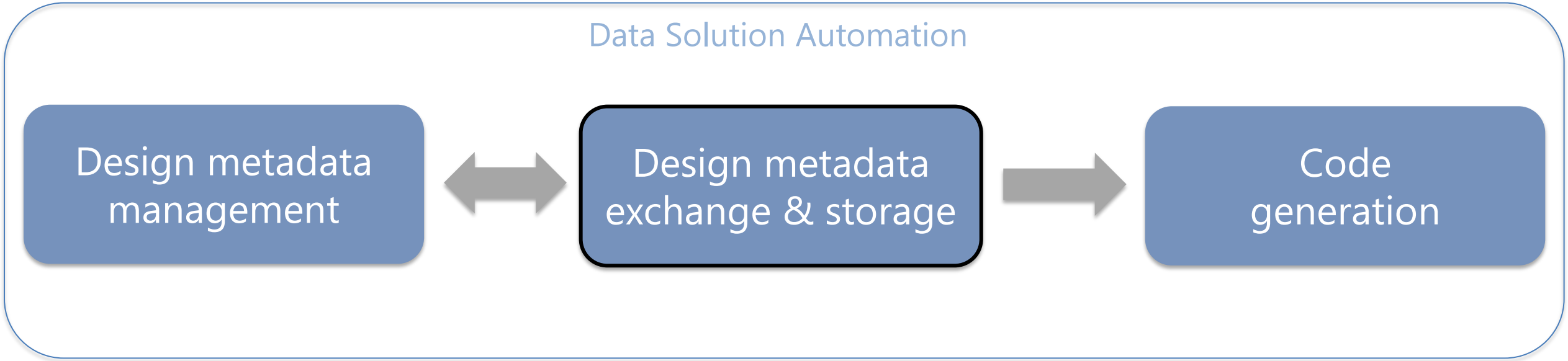
Design
metadata

Artifacts

Design metadata



Artifacts



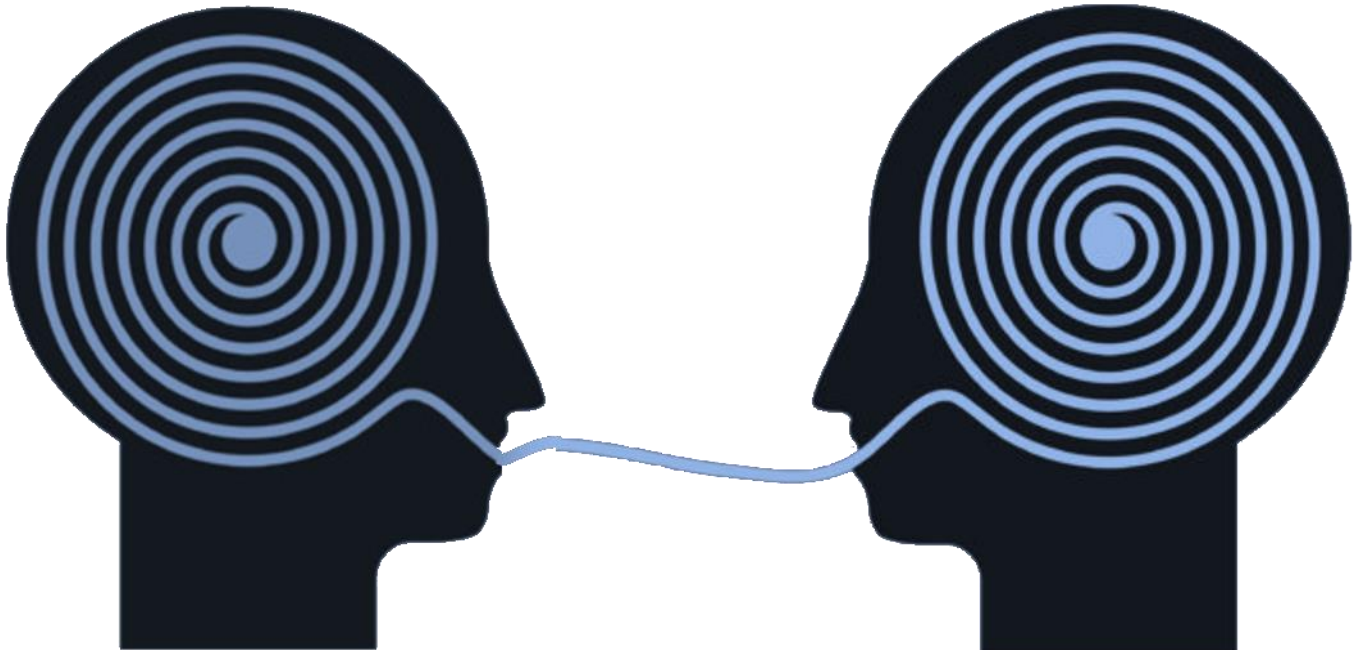
- 3rd party automation software e.g. BimlFlex
- Homebrew e.g. TEAM, Excel
- Database
- Modelling tool with macros e.g. Powerdesigner

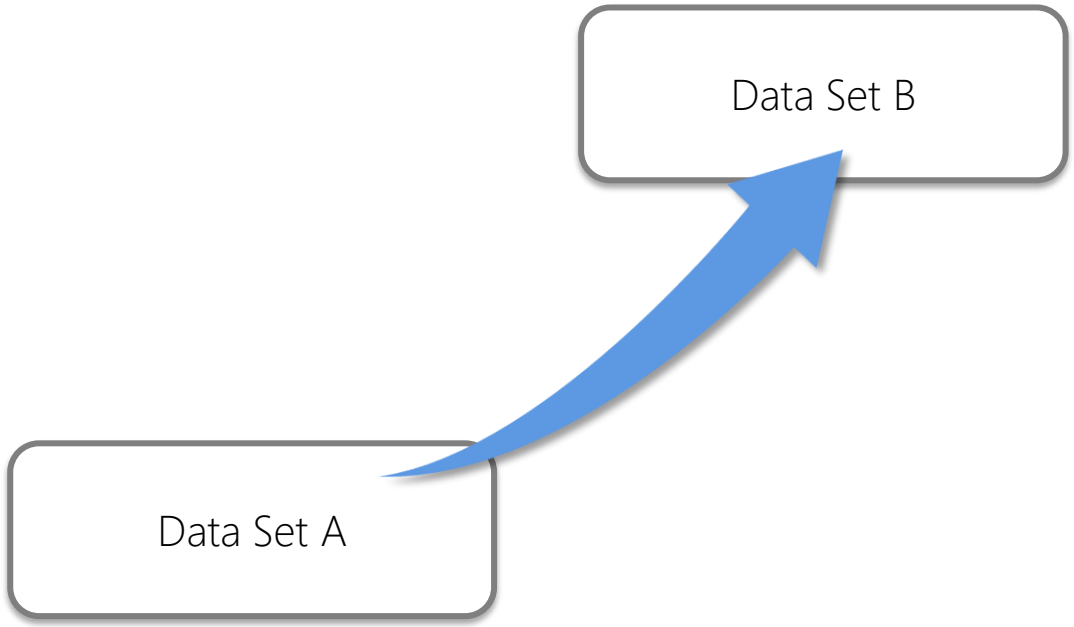
Schema for Data Solution Automation (Json / XML / Yaml / etc.)

As long as it's file based to go into version control.

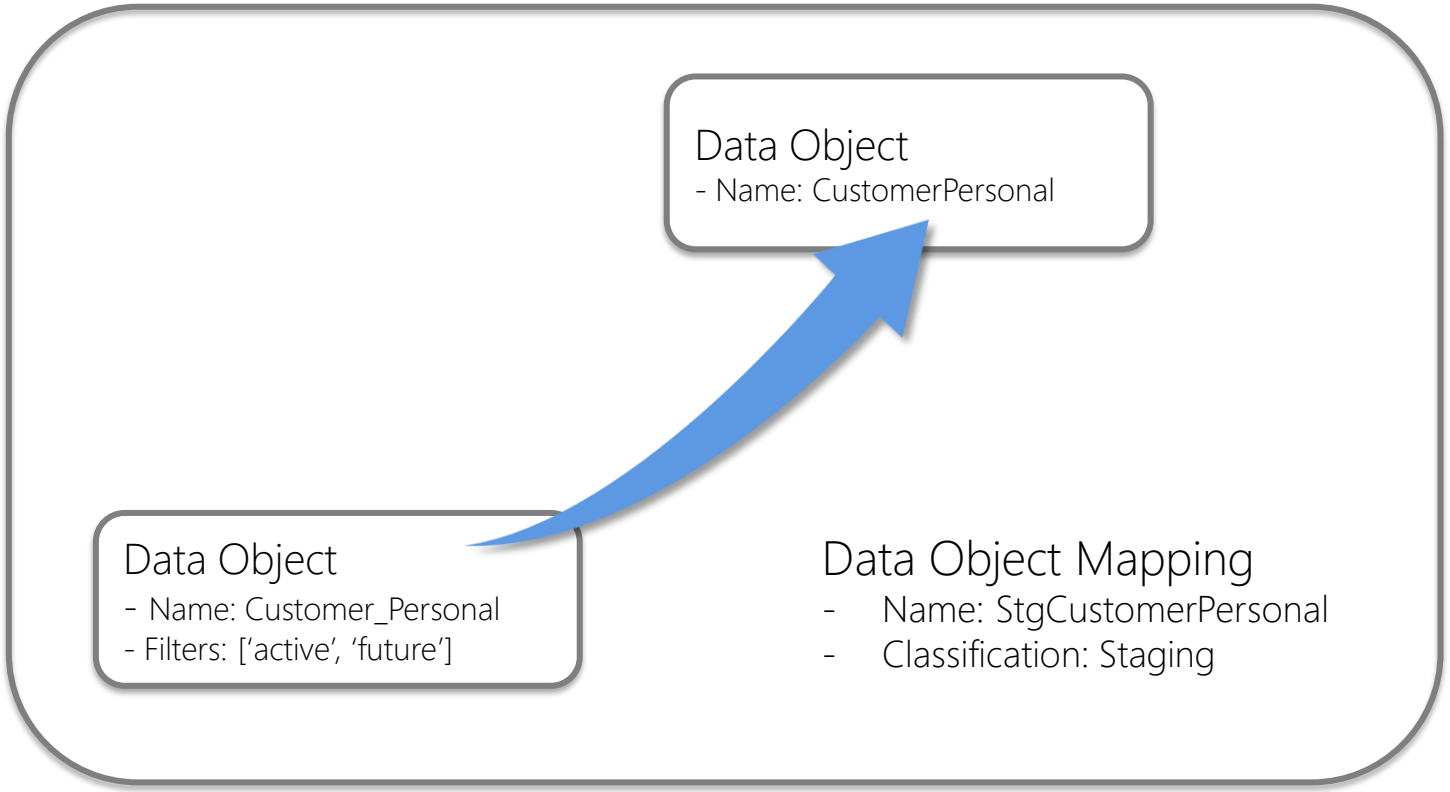
- Templating engines
  - Handlebars
  - Pebble
  - Scriban
  - Jinja
- Compilers
  - Biml
- SDKs
- Proprietary
  - SQL-based
  - C#, Python, Java

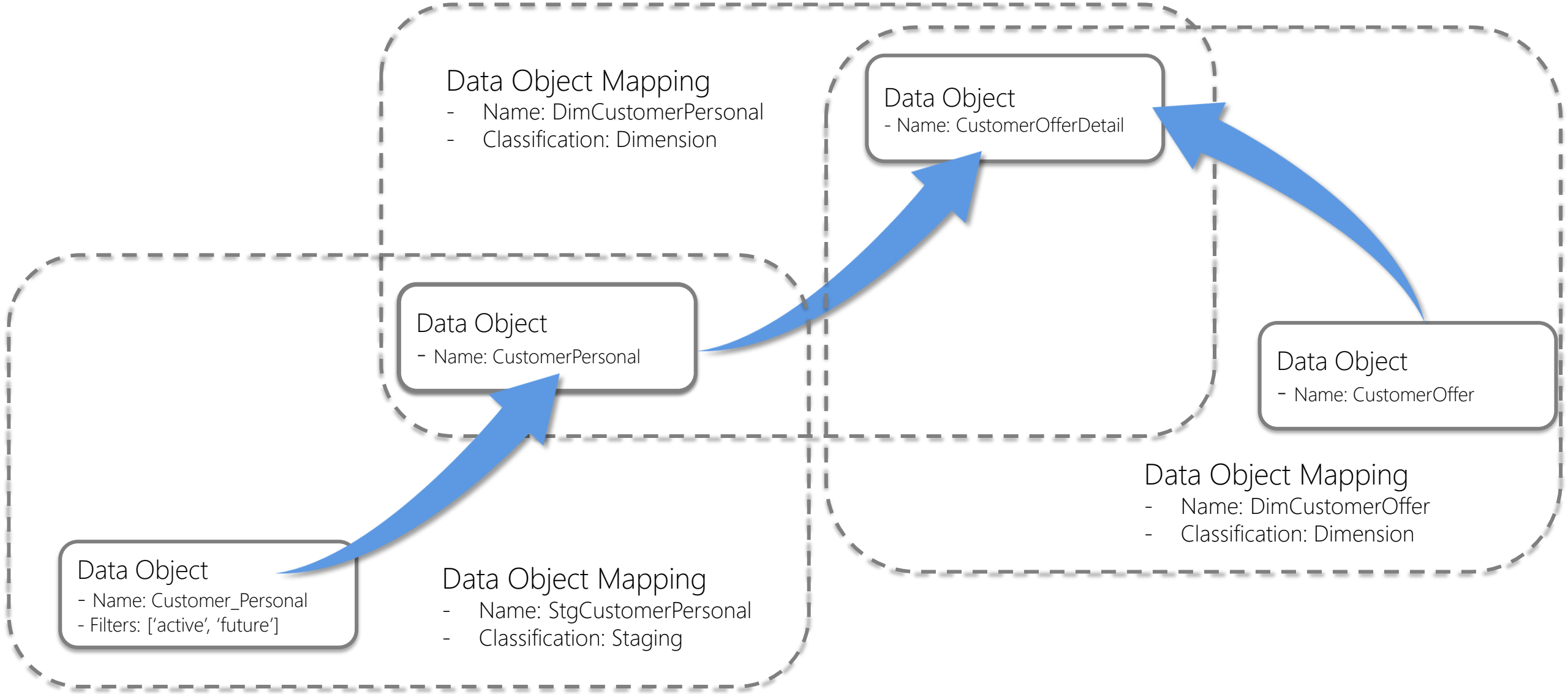
So, what do we need to record?

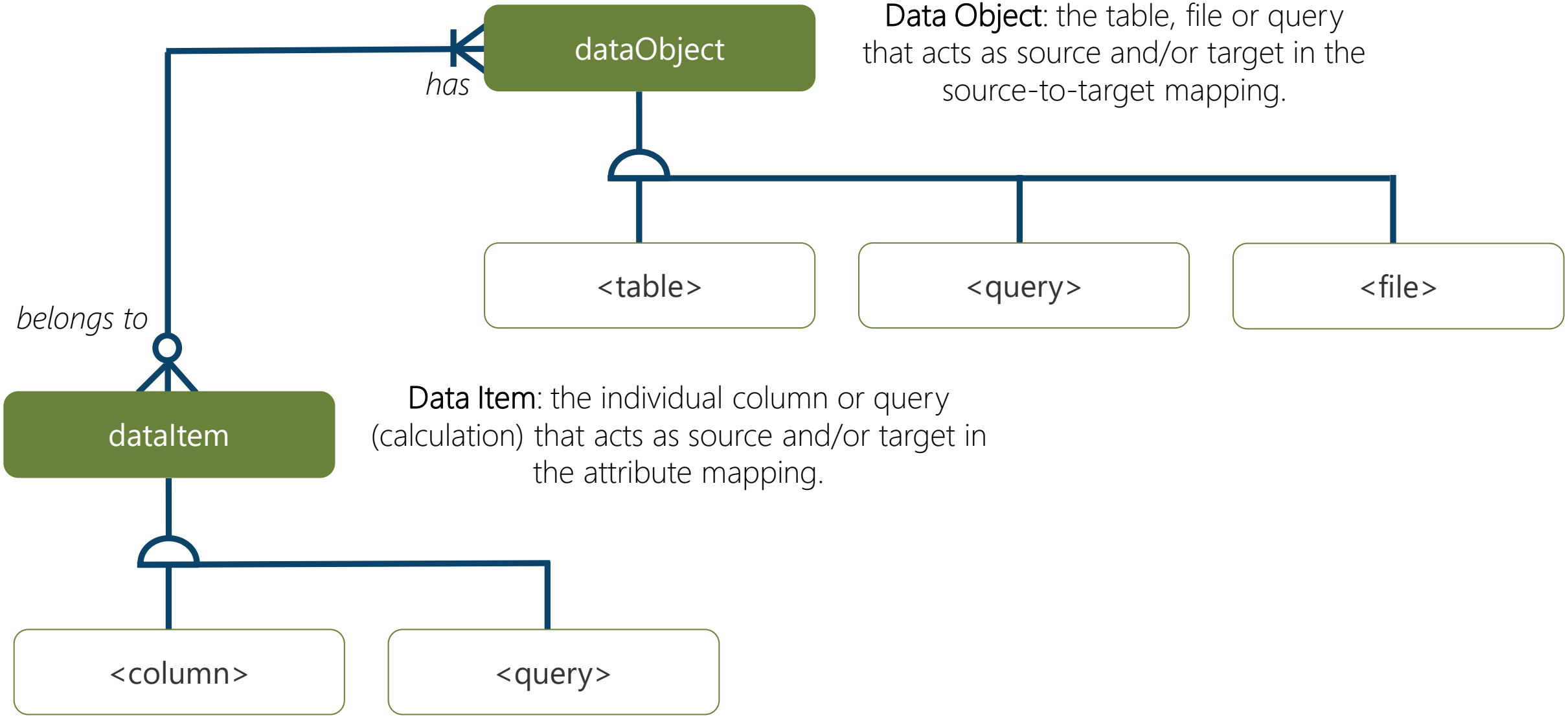


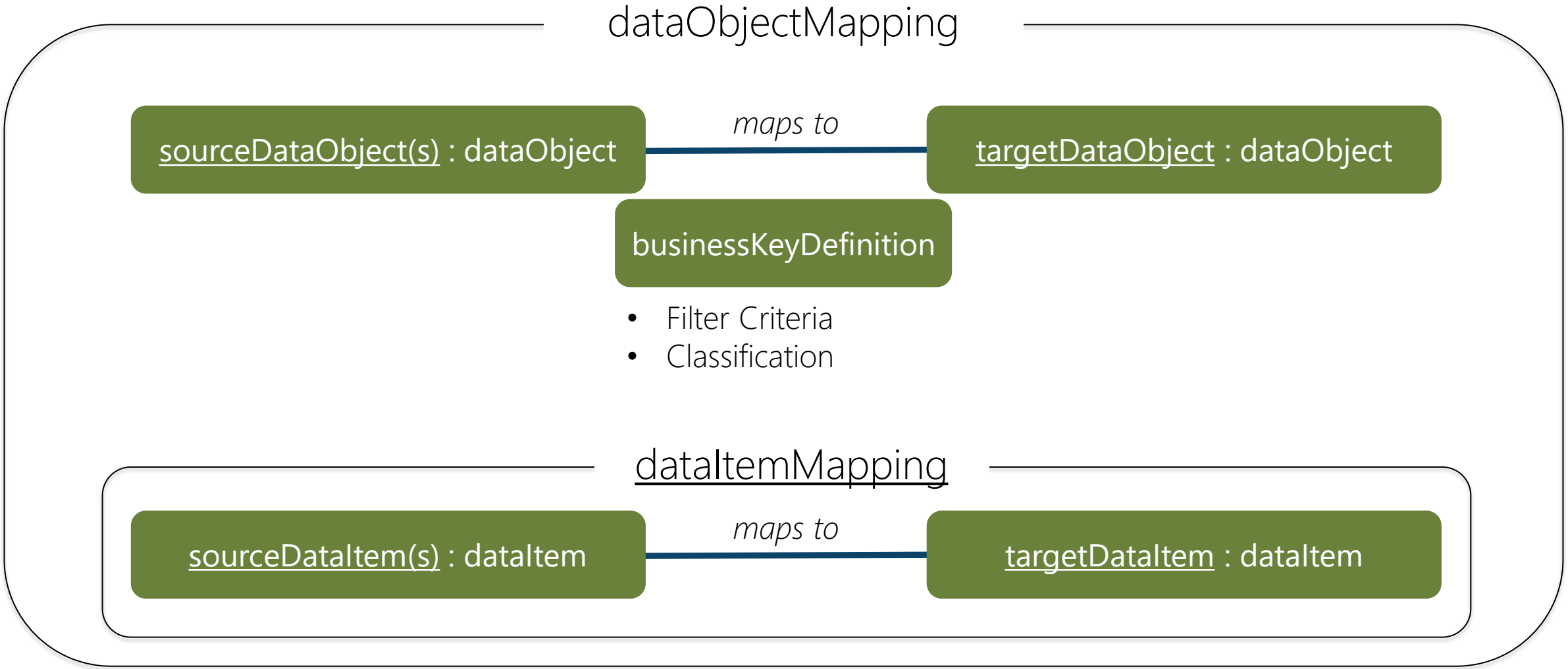


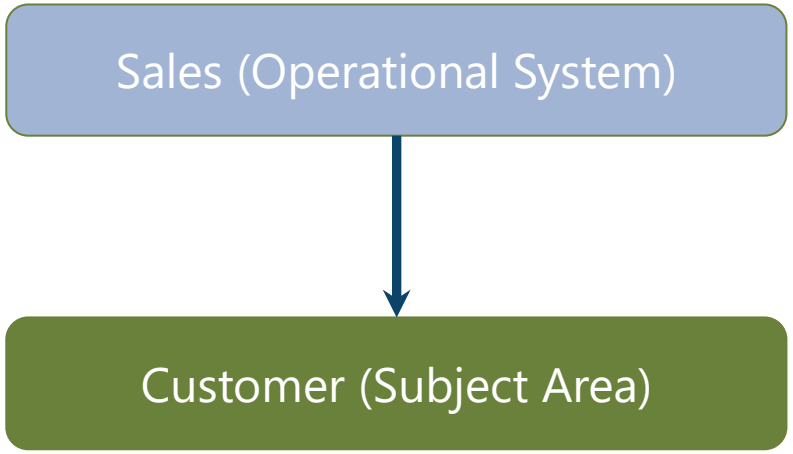
But, what is A and what is B?

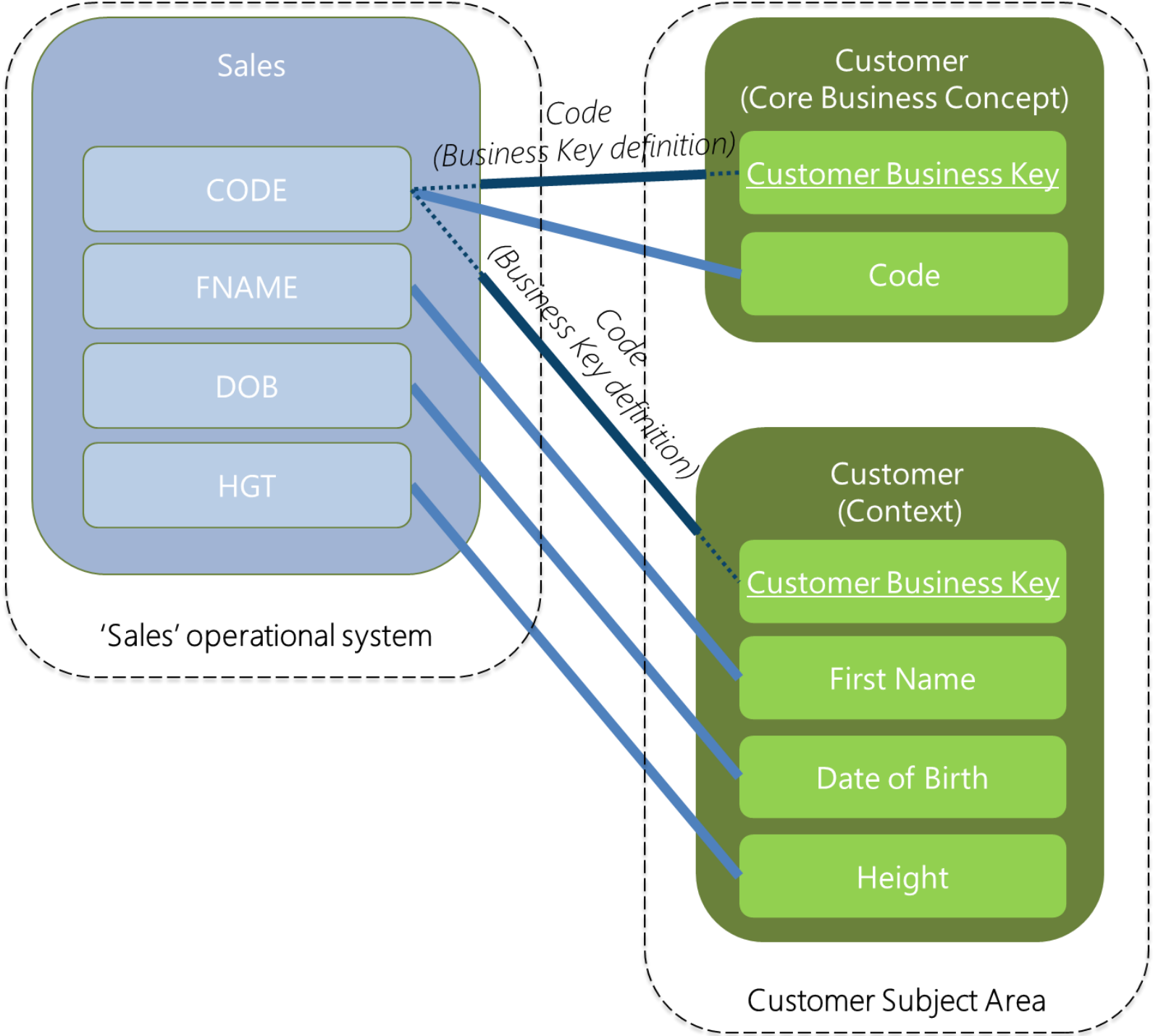

Data Object
- Name: CustomerPersonal
Data Object
- Name: Customer_Personal
- Filters: ['active', 'future']
Data Object Mapping
- Name: StgCustomerPersonal
- Classification: Staging


Data Object Mapping
- Name: DimCustomerPersonal
- Classification: Dimension
Data Object
- Name: CustomerOfferDetail
Data Object
- Name: CustomerPersonal
Data Object
- Name: CustomerOffer
Data Object Mapping
- Name: DimCustomerOffer
- Classification: Dimension
Data Object
- Name: Customer_Personal
- Filters: ['active', 'future']
Data Object Mapping
- Name: StgCustomerPersonal
- Classification: Staging

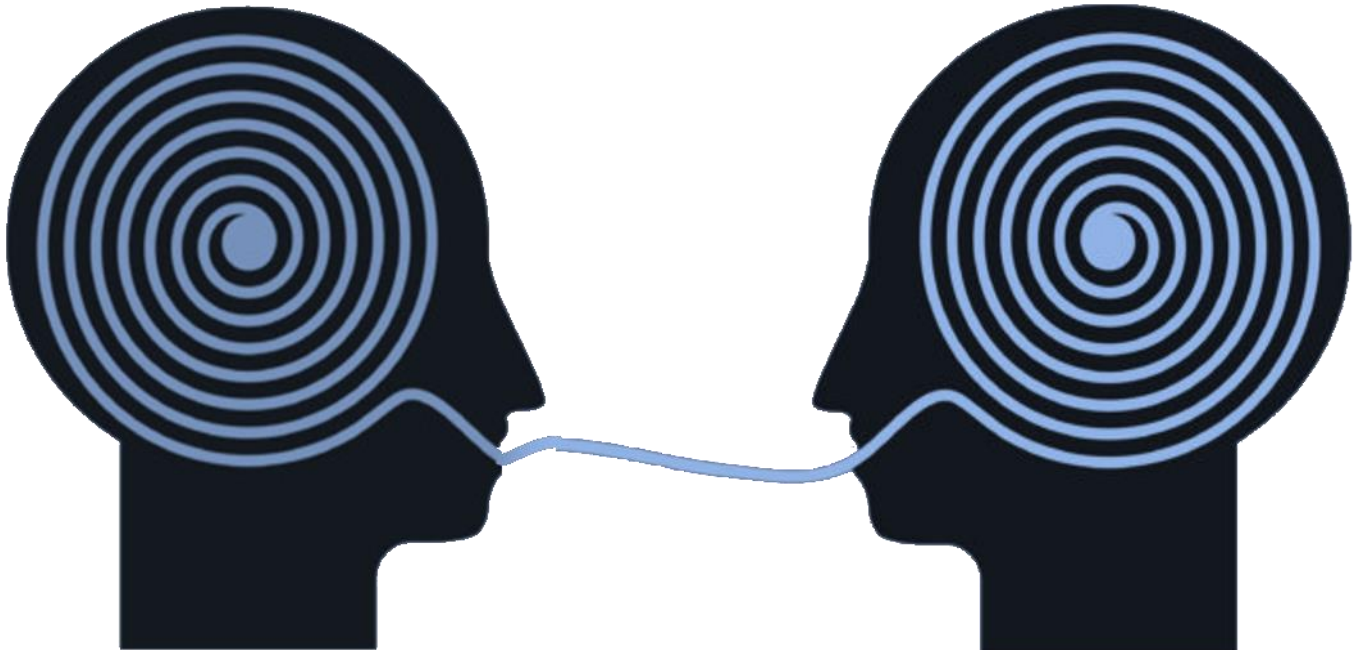

dataObject
has
Data Object: the table, file or query that acts as source and/or target in the source-to-target mapping.
<table>
<query>
<file>
belongs to
dataItem
Data Item: the individual column or query (calculation) that acts as source and/or target in the attribute mapping.
<column>
<query>


dataObjectMapping
sourceDataObject(s) : dataObject
maps to
targetDataObject : dataObject
businessKeyDefinition
• Filter Criteria
• Classification
dataItemMapping
sourceDataItem(s) : dataItem
maps to
targetDataItem : dataItem

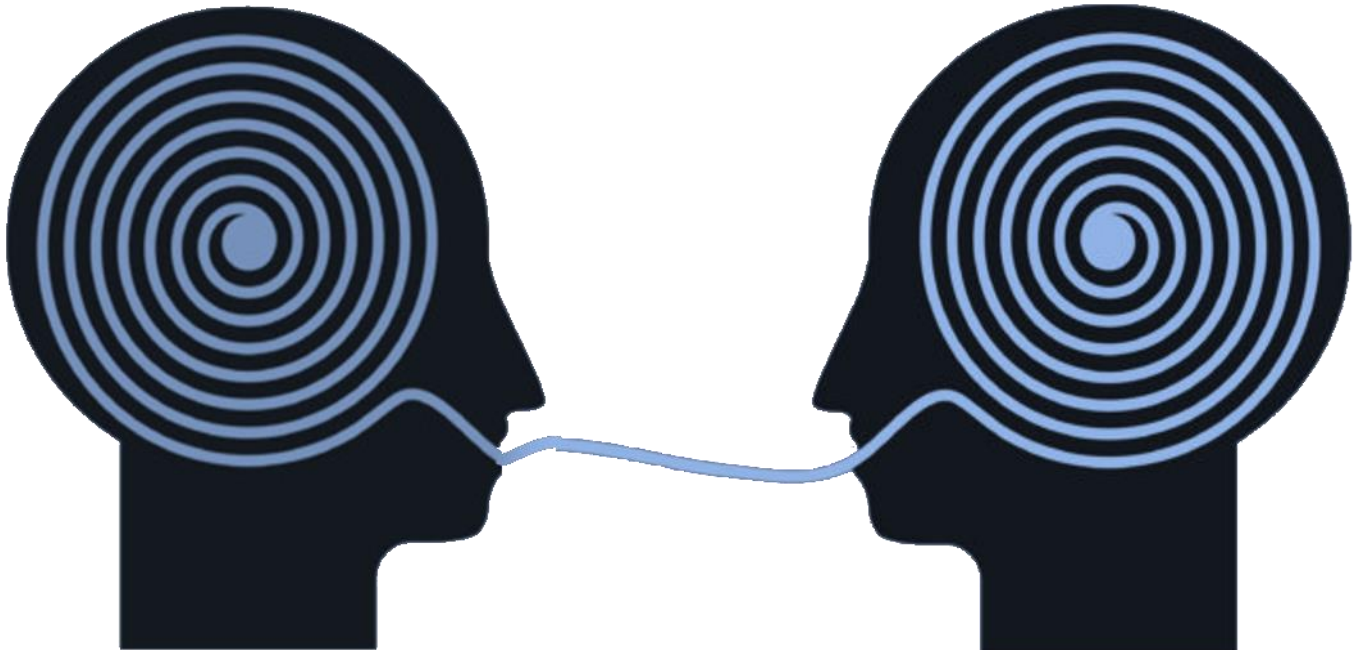

Sales (Operational System)
Customer (Subject Area)


Sales
CODE
FNAME
DOB
HGT
'Sales' operational system
Code
(Business Key definition)
Code
(Business Key definition)
Customer
(Core Business Concept)
Customer Business Key
Code
Customer
(Context)
Customer Business Key
First Name
Date of Birth
Height
Customer Subject Area

Demonstration



Missing something?



```
"extension": {
     "$id": "#/definitions/extension",
     "title": "Key/Value pair extensions",
     "description": "Free-format key/value pair that can be used for additional context.",
     "type": "object",
     "required": [ "key" ],
     "properties": {
       "key": {
         "$id": "#/definitions/extension/properties/key",
         "title": "Extension key",
         "description": "The Key in a Key/Value pair.",
         "type": "string"
       },
       "value": {
         "$id": "#/definitions/extension/properties/value",
         "title": "Extension value",
         "description": "The Value in a Key/Value pair.",
         "type": "string"
       },
       "description": {
         "$id": "#/definitions/extension/properties/description",
         "title": "Extension description",
         "description": "Any additional information to explain the intent of extension key/value pair.",
         "type": "string"
       }
     }
```

Thank you!

Roelant Vos
roelant@ravos.nl
www.roelantvos.com/blog
https://github.com/RoelantVos/Data-Warehouse-Automation-Metadata-Schema