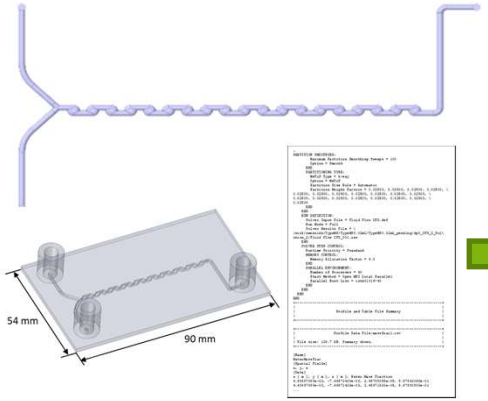
technische universität
dortmund

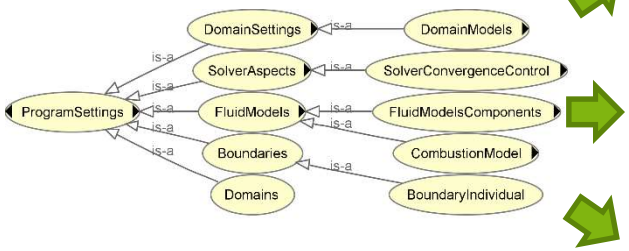
Expanding an Ontology with Semantically Linked CFD-Simulation Data by Segmentation into Reusable Concepts

Hendrik Borgelt, Alexander Behr, Norbert Kockmann
Laboratory of Equipment Design

CFD - Ontology for Catalysis



Ontology

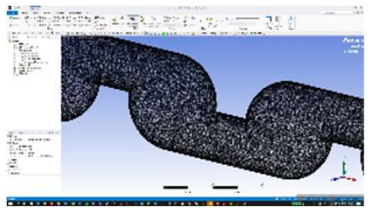


- Aligning data structures (knowledge graphs)
- Expandable (terms/classes, relations/axioms, information/definitions, …)
- Interconnectivity (e.g. SQL database, metadata frameworks, electronic laboratory notebooks,…)

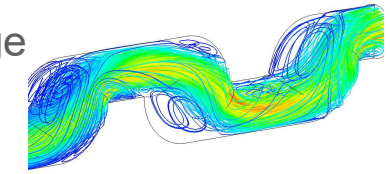
- Storage and Metadata Annotation

```
Microreactor has Y-mixing-element
"Flow Analysis 1": {
  "ANALYSIS TYPE": {...}},
  "DOMAIN": {
    "Default Domain": { ...
      "BOUNDARY": {
        "Default Domain Default": {...
          "BOUNDARY CONDITIONS": {...}}}},
  "DOMAIN MODELS": {...}},
```

- New Simulations / Change Parameters



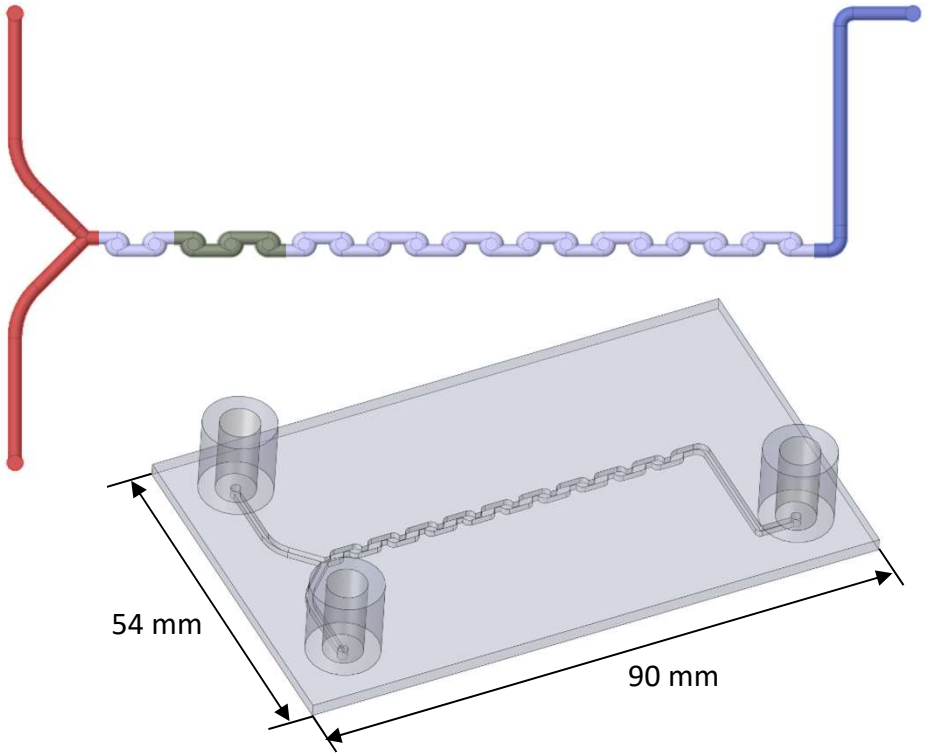
- Infer Knowledge



↑ Eddies & ↓ Reaction → unmixed Pockets

Motivation – Data Origin

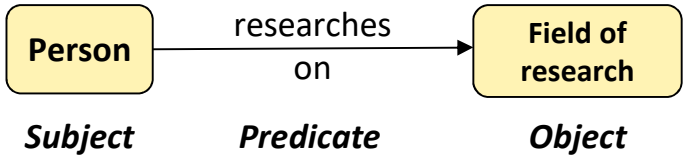
- Data availability
  - Experimental [1]
  - Theoretical [2]
    - Computational fluid dynamics (CFD)
    - Micro reactor of Little Things Factory
- Data from master thesis at TU Dortmund University
  - Subdivision of simulations
  - Variety of settings

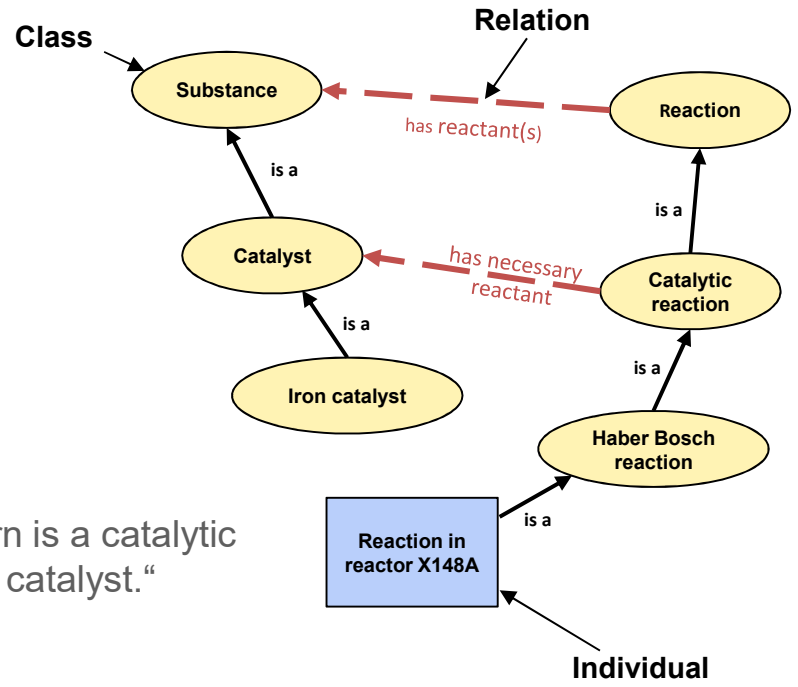


[1] Frede, Timothy Aljoscha; Burke, Inga; Kockmann, Norbert (2021)
[2] Frede, Timothy Aljoscha; Dietz, Marlene; Kockmann, Norbert (2021)

Ontologies – A simple example

- Information is stored in triplets



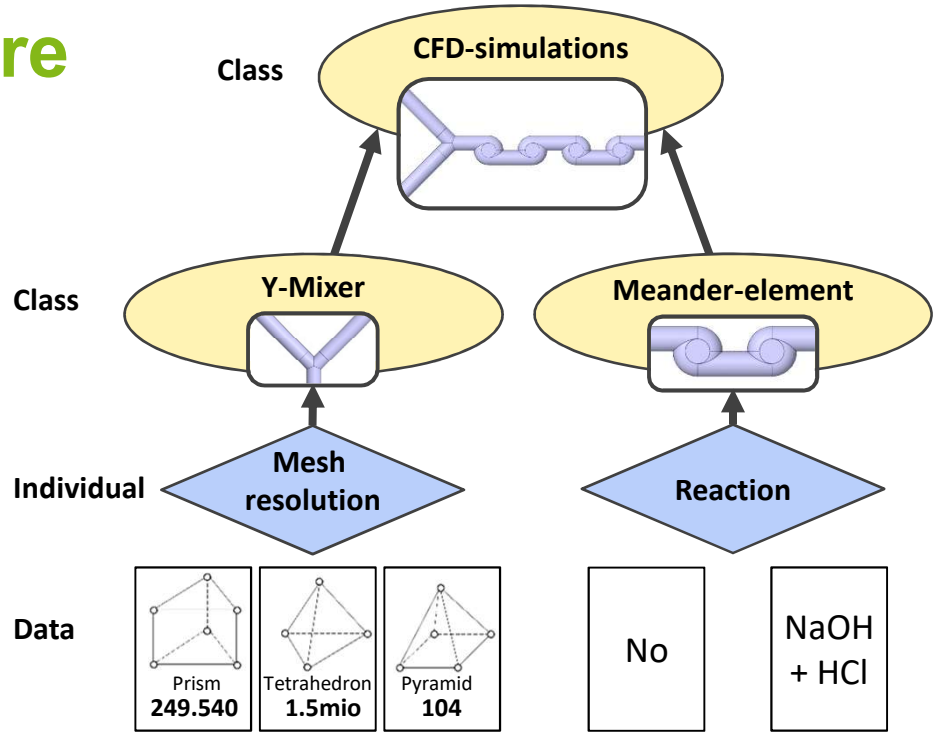


- Reasoning enhances data
  - „The reaction in reactor X148A uses iron catalysts“
  - Inference yields: „The reaction in reactor X148A is a Haber-Bosch reaction which in turn is a catalytic reaction and uses iron catalyst as catalyst.“

Motivation – Data Structure

Not FAIR-data but FAIR-storage

- **F**indable:
  Queryable / SPARQL
- **A**ccessible:
  Classification via metadata / OWL & SHACL
- **I**nteroperable:
  Non-proprietary file formats / RDF
- **R**euseable
  Connection to database and maintaining of metadata / SPARQL & RDF

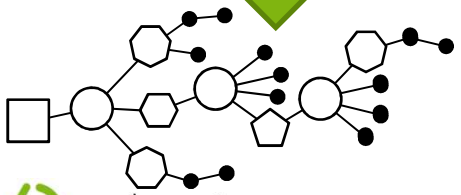


Process Steps

- Summary file as data source
  - Unknown structure
  - Complex syntax
- JSON-archive
  - Human-readable
  - Machine-searchable

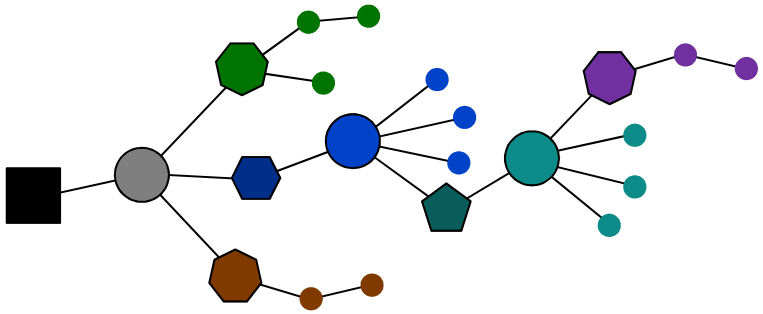
```
…
PARTITION SMOOTHING:
        Maximum Partition Smoothing Sweeps = 100
        Option = Smooth
      END
      PARTITIONING TYPE:
        MeTiS Type = k-way
        Option = MeTiS
        Partition Size Rule = Automatic
        Partition Weight Factors = 0.02500, 0.02500, 0.02500, 0.02500, \
0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, \
0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, \
0.02500
      END
    END
    RUN DEFINITION:
      Solver Input File = Fluid Flow CFX.def
      Run Mode = Full
      Solver Results File = \
/work/smnknikb/TypeMS/TypeMS0.02mL/TypeMS0.02mL_pending/dp0_CFX_2_Sol\
ution_2/Fluid Flow CFX_001.res
    END
    SOLVER STEP CONTROL:
      Runtime Priority = Standard
      MEMORY CONTROL:
        Memory Allocation Factor = 4.0
      END
      PARALLEL ENVIRONMENT:
        Number of Processes = 40
        Start Method = Open MPI Local Parallel
        Parallel Host List = cstd01016*40
      END
    END
  END
END
+---------------------------------------------------------------------+
|                                                                     |
|                  Profile and Table File Summary                     |
|                                                                     |
+---------------------------------------------------------------------+

+---------------------------------------------------------------------+
|                Profile Data File:massfrac1.csv                      |
|                                                                     |
| File size: 129.7 kB. Summary shown.                                 |
+---------------------------------------------------------------------+

[Name]
WaterMassFrac
[Spatial Fields]
x, y, z
[Data]
x [ m ], y [ m ], z [ m ], Water.Mass Fraction
4.48997989e-02, -7.46673439e-03, 2.45798055e-05, 5.67842066e-01
4.48997989e-02, -7.46671436e-03, 2.45971933e-05, 5.67880809e-01
...
```

```
…
            "PARTITION SMOOTHING": {
                "Maximum Partition Smoothing Sweeps": "100",
                "Option": "Smooth"},
            "PARTITIONING TYPE": {
                "MeTiS Type": "k-way",
                "Option": "MeTiS",
                "Partition Size Rule": "Automatic",
                "Partition Weight Factors": "0.02500, 0.02500,
0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500,
0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500,
0.02500, 0.02500, 0.02500, 0.02500, 0.02500, 0.02500"}},
        "RUN DEFINITION": {
            "SOLVER STEP CONTROL": {
                "Runtime Priority": "Standard",
                "MEMORY CONTROL": {
                    "Memory Allocation Factor": "4.0"},
                "PARALLEL ENVIRONMENT": {
                    "Number of Processes": "40",
                    "Start Method": "Open MPI Local Parallel",
                    "Parallel Host List": "cstd01016*40"}}}}},
"Profile and Table File Summary": {
    "Initialization": {
        "0": {
            "box": "Profile Data File:massfrac1.csv \n \nFile size:
129.7 kB. Summary shown. \n"},
        "1": {
            "text": {
                "": [],
                "[Name]": [],
                "WaterMassFrac": [],
                "[Spatial Fields]": [],
                "x, y, z": [],
                "[Data]": [],
                "x [ m ], y [ m ], z [ m ], Water.Mass Fraction",
                "4.48997989e-02, -7.46673439e-03, 2.45798055e-05,
                  5.67842066e-01",
                "4.48997989e-02, -7.40168104e-03, 3.26326626e-05,
                  6.86770558e-01",
                "..."}},
…
```



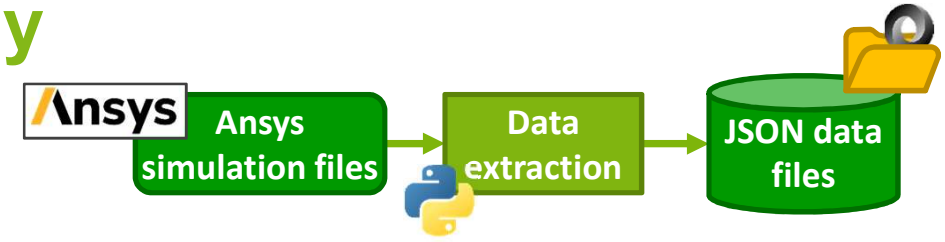
Process Steps

- Summary file as data source
  - Unknown structure
  - Complex syntax
- JSON-archive
  - Human-readable
  - Machine-searchable

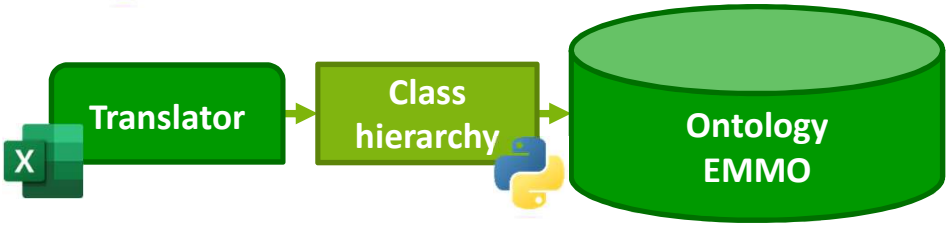



```
"FLOW": {
  "Flow Analysis 1": {
    "ANALYSIS TYPE": {
      "Option": "Steady State",
      "EXTERNAL SOLVER COUPLING": {
        "Option": "None"}},
    "DOMAIN": {
      "Default Domain": {
        "Coord Frame": "Coord 0",
        "Domain Type": "Fluid",
        "Location": "B3105",
        "BOUNDARY": {
          "Default Domain Default": {
            "Boundary Type": "WALL",
            "Location": "F2988.3105",
            "Use Profile Data": "False",
            "BOUNDARY CONDITIONS": {
              "MASS AND MOMENTUM": {
                "Option": "No Slip Wall"}}}},
    "DOMAIN MODELS": {
      "BUOYANCY MODEL": {
        "Option": "Non Buoyant"}},
    ...
```

Ontology



- VIMMP [1]
  - Domain / task
  - x Inconsistent



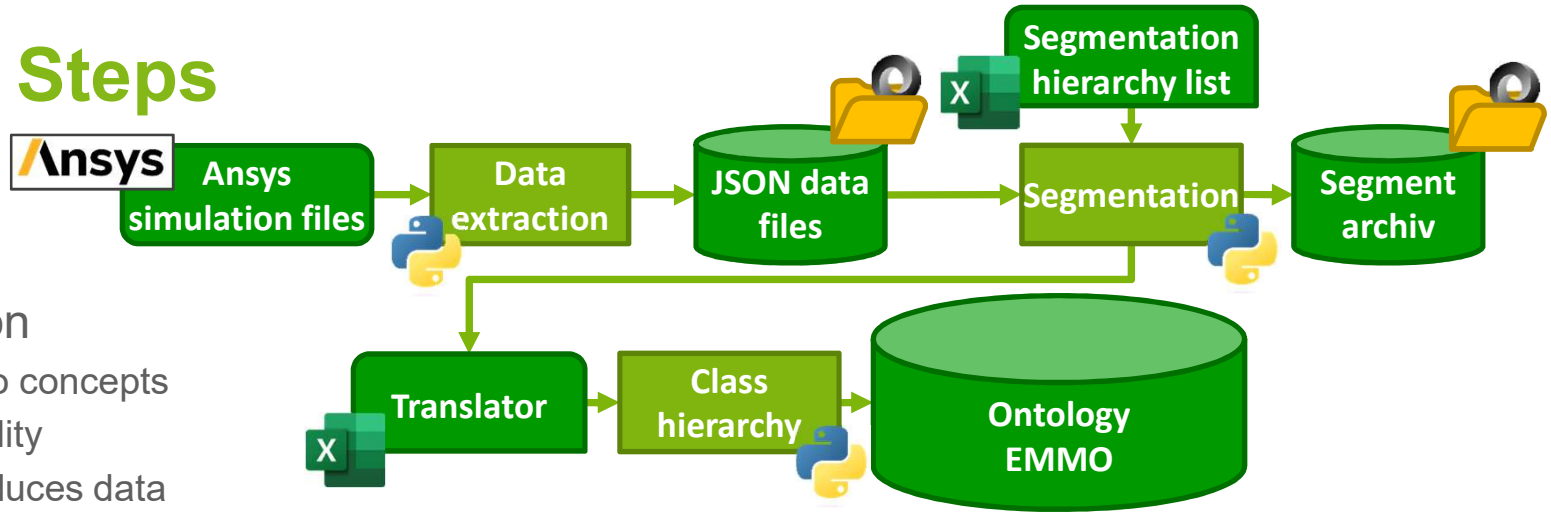
- EMMO [2]
  - Material modeling
  - Conceptualized

- Translator
  - Nested-to-nested-structure
  - Class hierarchy
  - Individual assignment

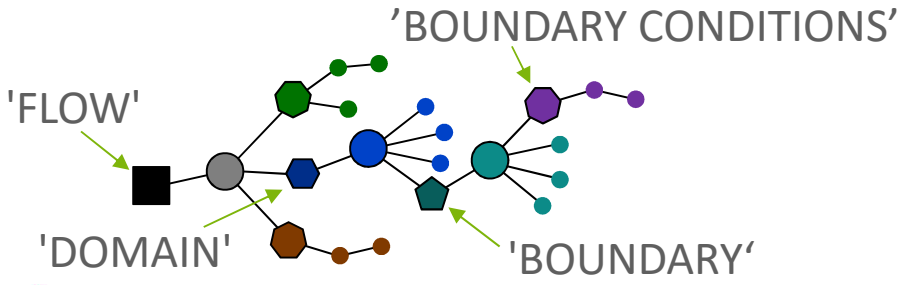
Dictionary key	MeanderElement4	Ansys
Super-class prefLabel in ontology	CfdSimulations	W3C OWL
Class prefLabel in ontology	MeanderElement	W3C OWL

[1] https://www.vimmp.eu/
[2] https://github.com/emmo-repo/EMMO

Process Steps



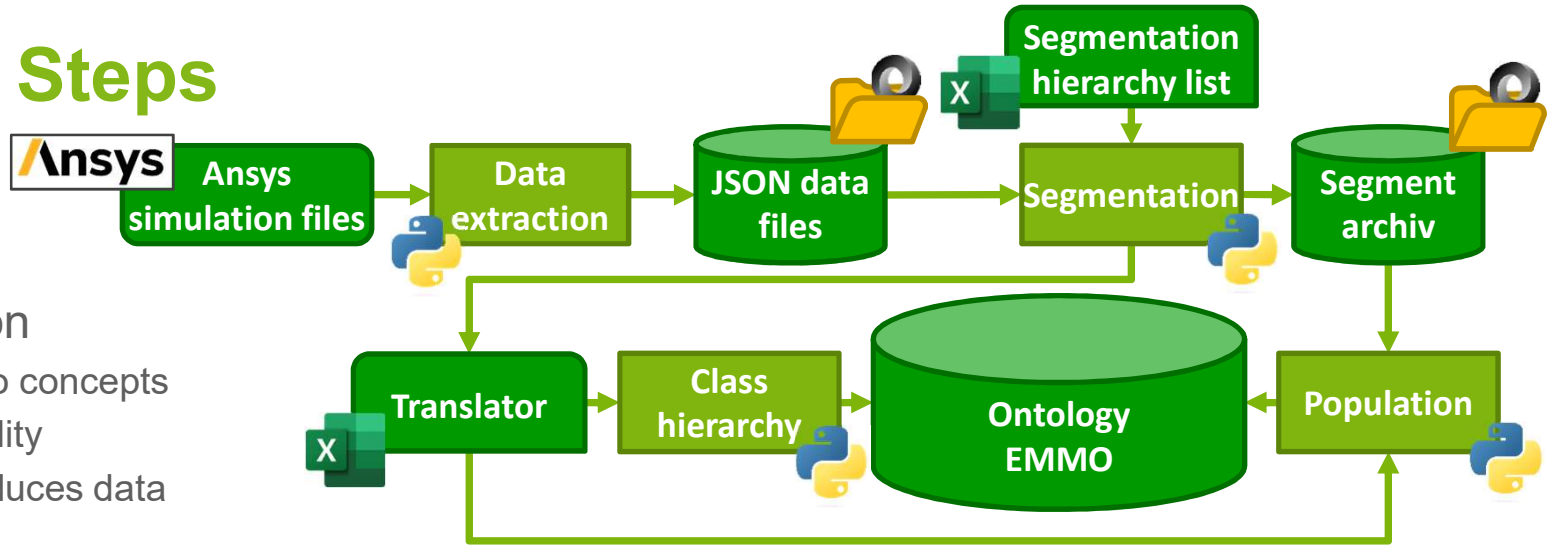
- Segmentation
  - Divided into concepts
  - Comparability
  - Reuses/reduces data



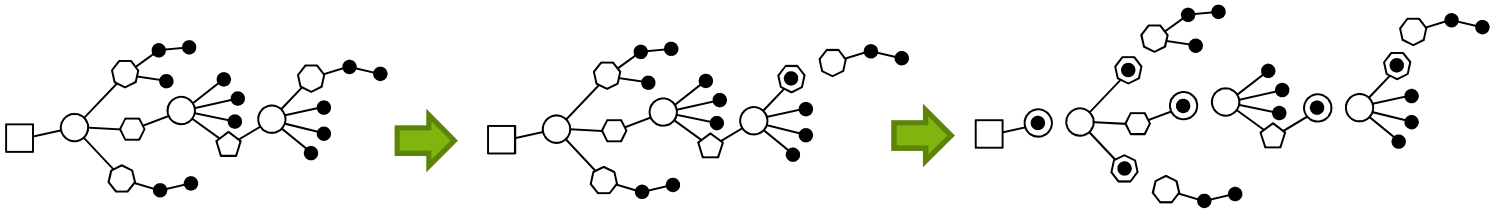
➢Segmentation hierarchy list example

Named:	'FLOW'
Named:	'DOMAIN'
Named:	'BOUNDARY'
Non-named:	'BOUNDARY CONDITIONS'

Process Steps

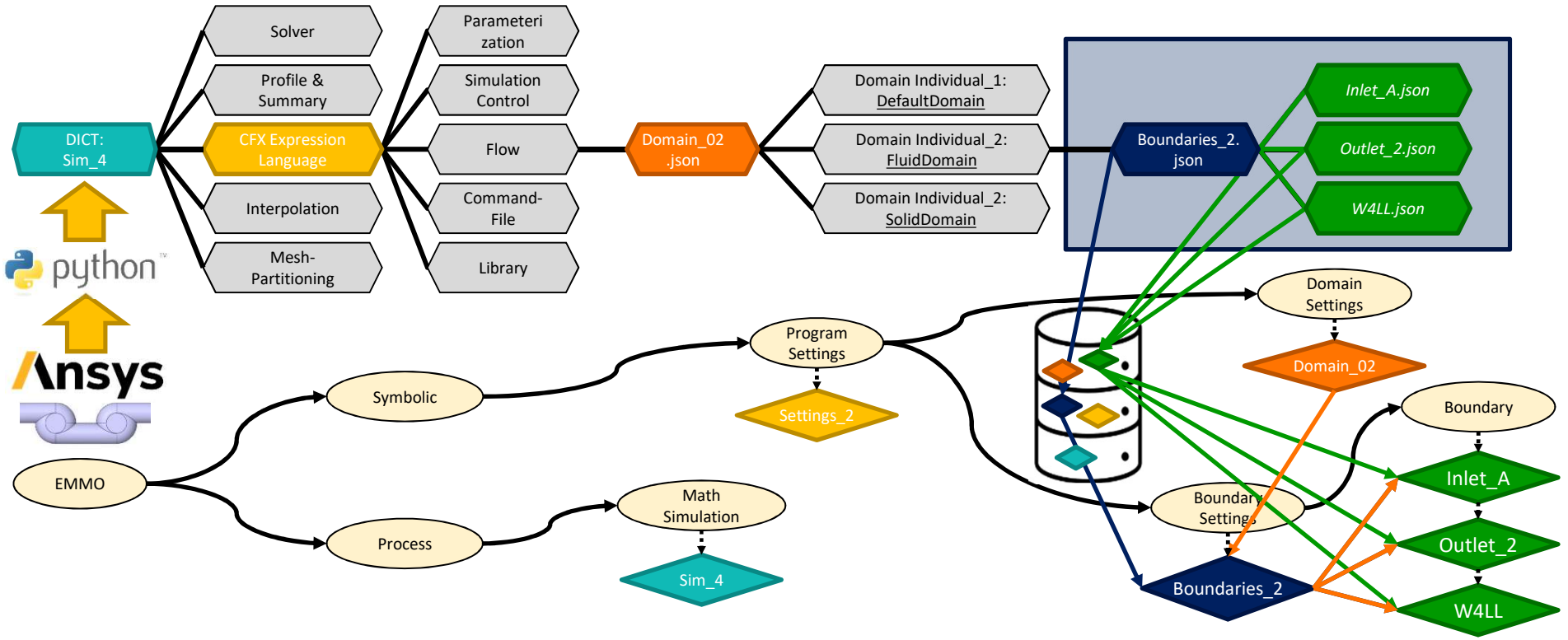


- Segmentation
  - Divided into concepts
  - Comparability
  - Reuses/reduces data
- Population [1,2]
  - Automation
  - Adaptable



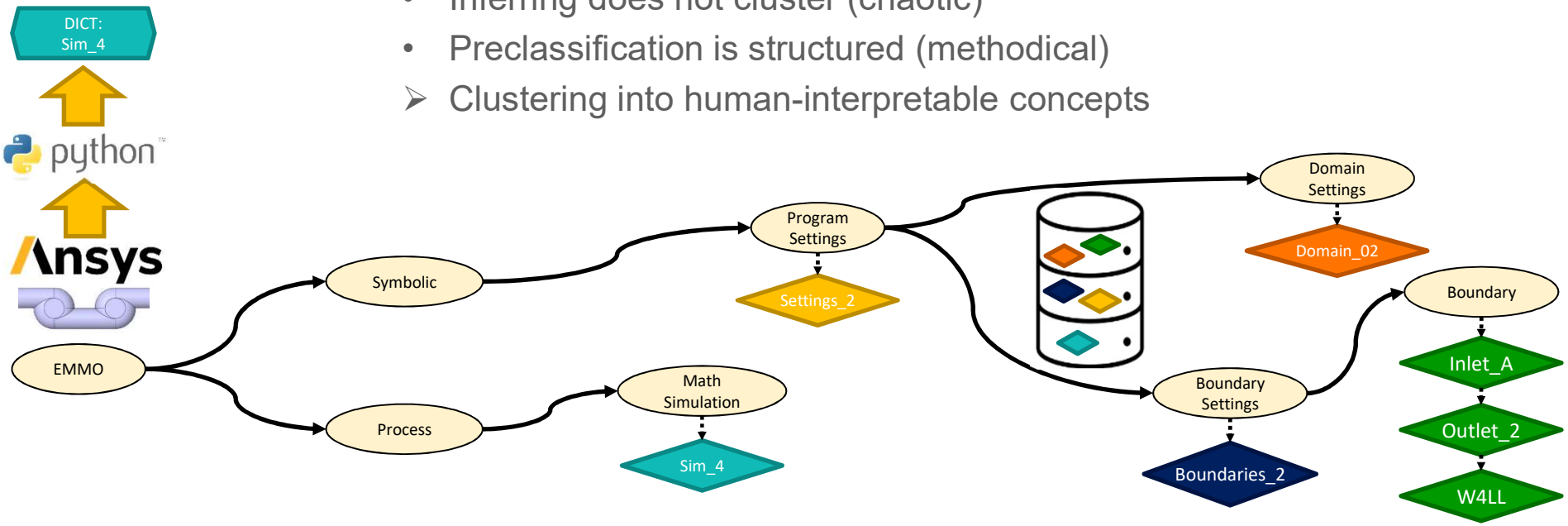
[1] Lamy, Jean-Baptiste (2017), owlready2-python-package
[2] EMMOntopy-python package, https://emmo-repo.github.io/EMMO-python/1.0.1

Dictionary to Ontology



Dictionary to Ontology

- Inferred classification vs preclassification
  - Inferring does not cluster (chaotic)
  - Preclassification is structured (methodical)
  - ➢ Clustering into human-interpretable concepts



Metrics / Program Performance

- Code stable
  - Test set: 128 simulations
  - Validation set: 783 simulations
- Population code
  - Optimizable
    - Optimization by inventory
    - Benchmarking and cross verification
  - Data condensation
    - Reducing pseudo duplicates
    - Extending instead of reconstruction

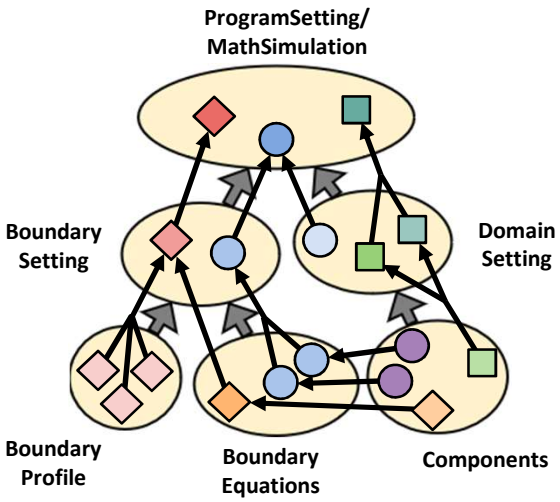
Program	Performance for 911 Simulations	
Data Extraction	69s	8.8M Lines
Segmentation	7min 03s	615 Segments of 36 Concepts
Population	14min 44s	

Performed with 8GB RAM and Intel i5-7200U CPU

Ontology	EMMO *	EMMO * + 911 Simulations	
Classes	470	846	Δ 376
Object Prop.	47	464	Δ 417
Data Prop.	3	258	Δ 255
Individuals	1	4,894	Δ 4893
Axioms	3,363	528,931	Δ 525,568

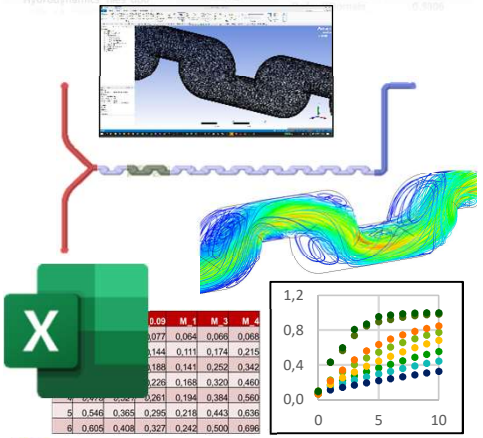
* Currently only EMMO_1.0.0_beta_0 inferred in 6min 52s

Querying



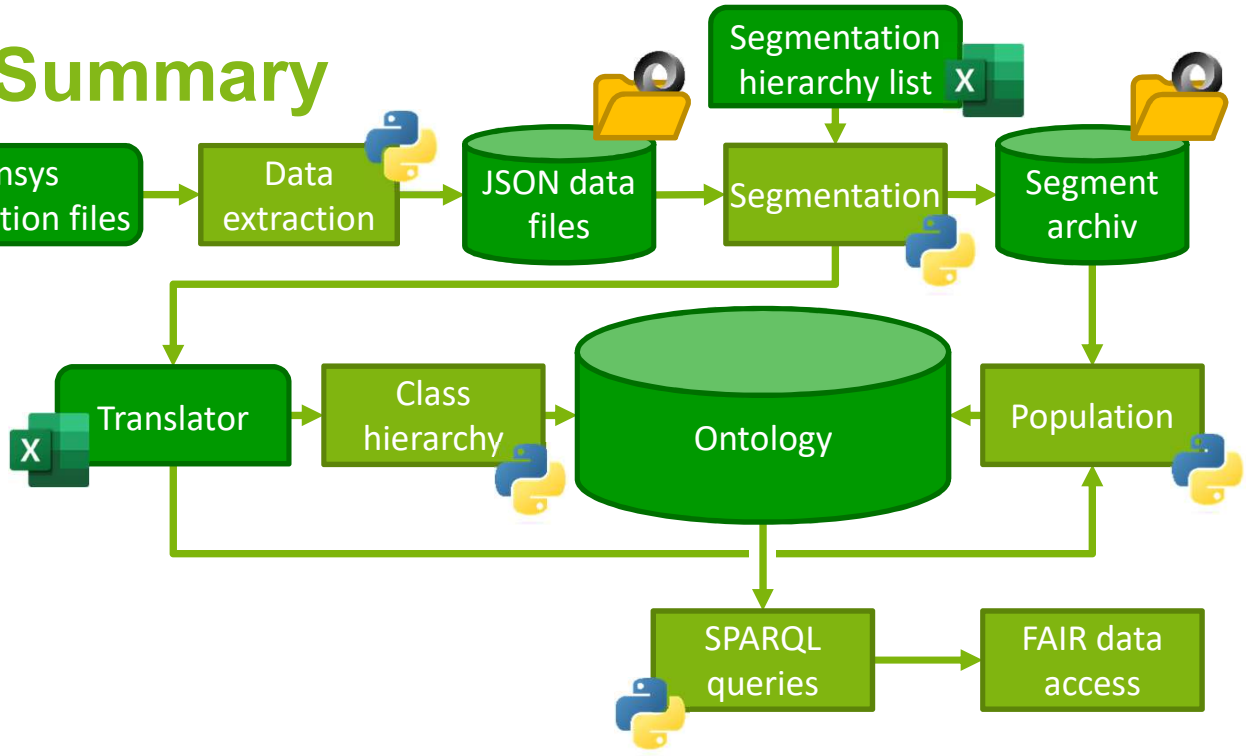

```
1  PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
2  PREFIX owl:  <http://www.w3.org/2002/07/owl#>
3  PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
4  PREFIX xsd:  <http://www.w3.org/2001/XMLSchema#>
5  PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
6  PREFIX emmo: <http://emmo.info/emmo#>
7
8  SELECT  ?SimPrefL  ?Boundary  ?indivPrefL  ?Value
9
10 WHERE  { ?indivCls    skos:prefLabel         ?IndivClsPrefL.
11          ?individual  a                      ?indivCls.
12          ?individual  skos:prefLabel         ?indivPrefL.
13          ?individual  emmo:hasSimulation     ?Sim.
14          ?Sim         skos:prefLabel         ?SimPrefL.
15          ?Sim         emmo:hasUmomKpiMinConv ?Value.
16          ?Boundary    emmo:hasDefIndividual  ?individual.
17          ?BoundaryTyp emmo:hasIndividual     ?Boundary.
18          ?BoundaryTyp emmo:hasBoundaryType   ?Typ.
19
20 FILTER (STR(?IndivClsPrefL) = "BoundaryComponentIndividual")
21 FILTER (?Value <0.95)
22 FILTER (STR(?Typ)="INLET")
23 }
```

Simulation Name	Boundary Name	Component Name	Convergence KPI
Mesh_independence_study_Mesh1_Mesh1_files_dp0_CFX_1_CFXFluid_Flow_CFX_006	FlowAnalysis1_DefaultDomain_Inlet_0	FlowAnalysis1_DefaultDomain_Inlet_Water_0	0.9489
	FlowAnalysis1_DefaultDomain_Inlet_1		



Process Steps - Summary

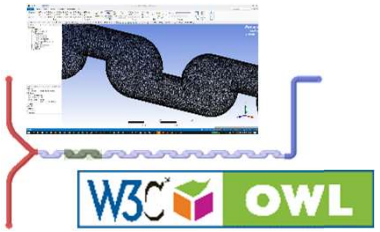
- FAIR data access by FAIR data storage
- Dictionary-to-ontology
  - Nested-to-nested
  - Unknown/varying structure
- Ontology
  - Knowledge graph
  - Classifiers due to segments
- Validated



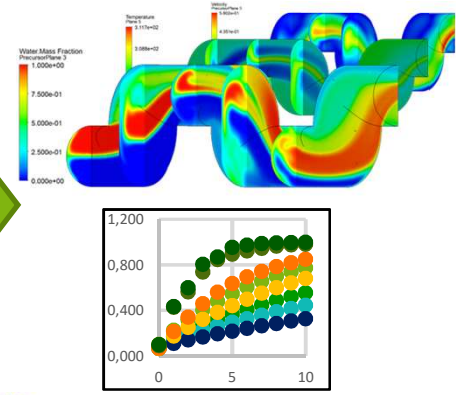
Outlook

- Connection to database
  - Ontology extension instead of recreation
  - Connect Knowledge Graph to database
- From Excel- to RDF-translator
  - Creating complex relations swiftly
  - Decoupling ontology creating from data population

- Usecases outside of database
  - Reperforming simulation with minimal dataset (RDF, mesh, additional inputs)



	M_0.03	M_0.06	M_0.09	M_1	M_3	M_4
0	0,102	0,083	0,077	0,064	0,066	0,068
1	0,227	0,166	0,144	0,111	0,174	0,215
2	0,320	0,223	0,188	0,141	0,252	0,342
3	0,403	0,274	0,226	0,168	0,320	0,460
4	0,478	0,321	0,261	0,194	0,384	0,560
5	0,546	0,365	0,295	0,218	0,443	0,636
6	0,605	0,408	0,327	0,242	0,500	0,696



Special thanks to:
Colleagues at Laboratory of Equipment Design,
Members of NFDI4Cat - TA1,
Deutsche Forschungsgemeinschaft (DFG)