NFDI4Cat

NFDI for Catalysis-Related Sciences

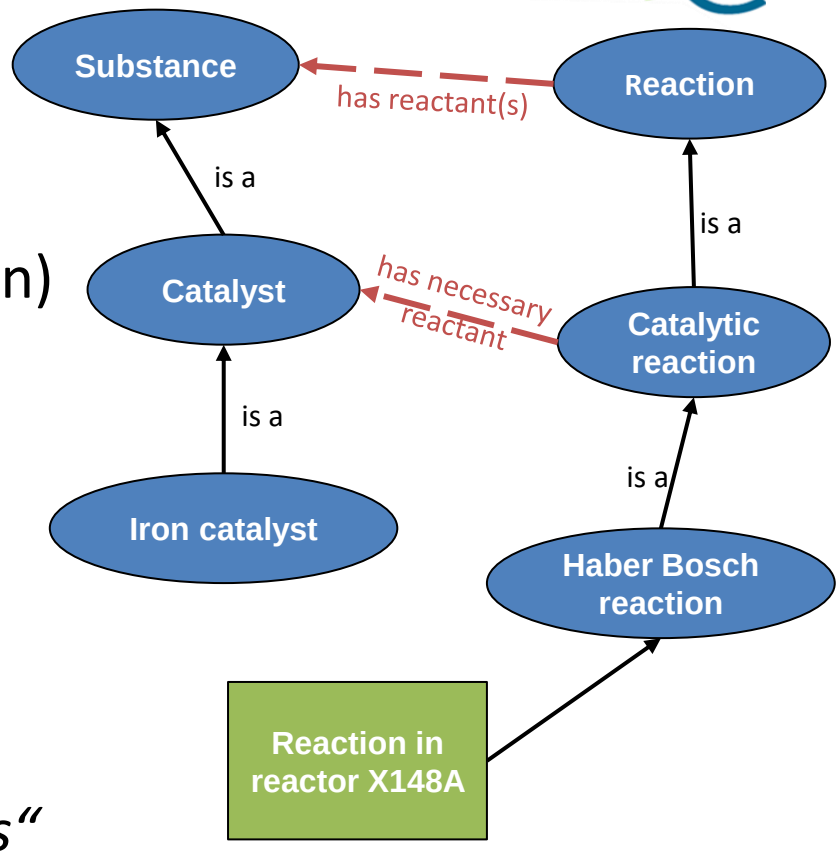
Introduction to Ontologies – Hands on Ontologies

Alexander S. Behr, Norbert Kockmann

23.06.2022

Ontologies – A simple example

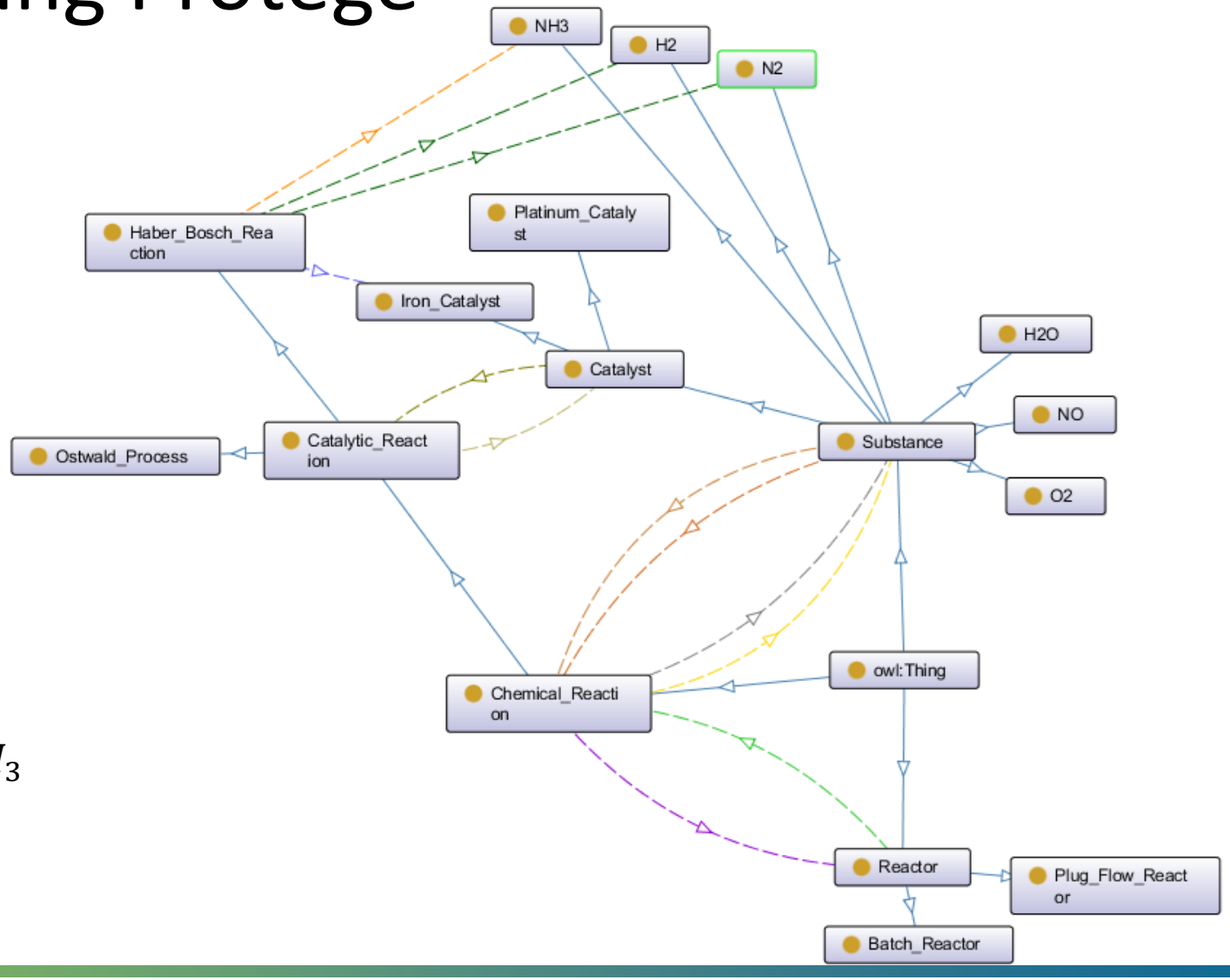
- Ontologies consist of
  - **Classes** (hierarchically structured by „is a“ relation)
  - **Relations** between classes
  - **Individuals** representing real existing elements
  - Rules which are always true, like „Catalytic reactions need one or more catalyst“
- Reasoning enhances data
  - *„The reaction 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.“*



Life Example on Protégé

Minimal Example using Protégé

- Add class hierarchy
- Add properties
- Specify domains and ranges of properties
- Add individuals
- Pose queries

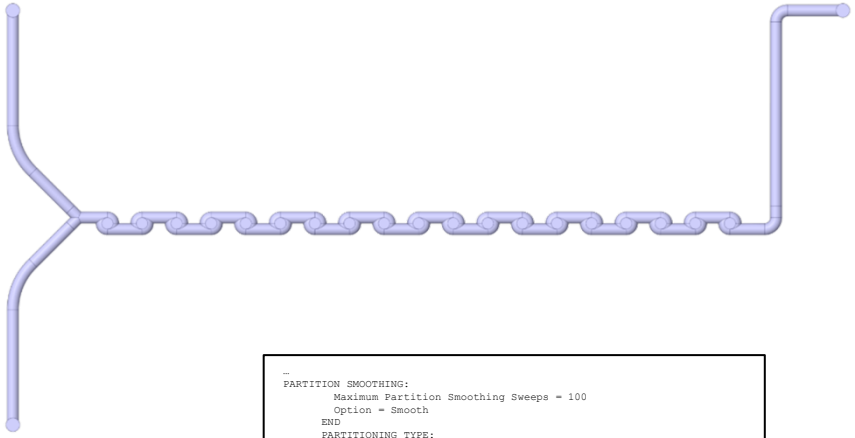


$$3H_2 + N_2 \xrightarrow{Fe-Catalyst} 2NH_3$$

Use-Case on simulation-data

Use Case – CFD Simulations

- Flow through micro-reactor simulated
- Huge amount of data
  - Domain experts (simulators) focus on optimizing model
  - No focus on proper data handling
  - No metadata collection
  - ➢ UnFAIR data handling
  - ➢ (Meta)data contained to each simulation
- ➢ **Can we use ontologies to make this data FAIR again?**



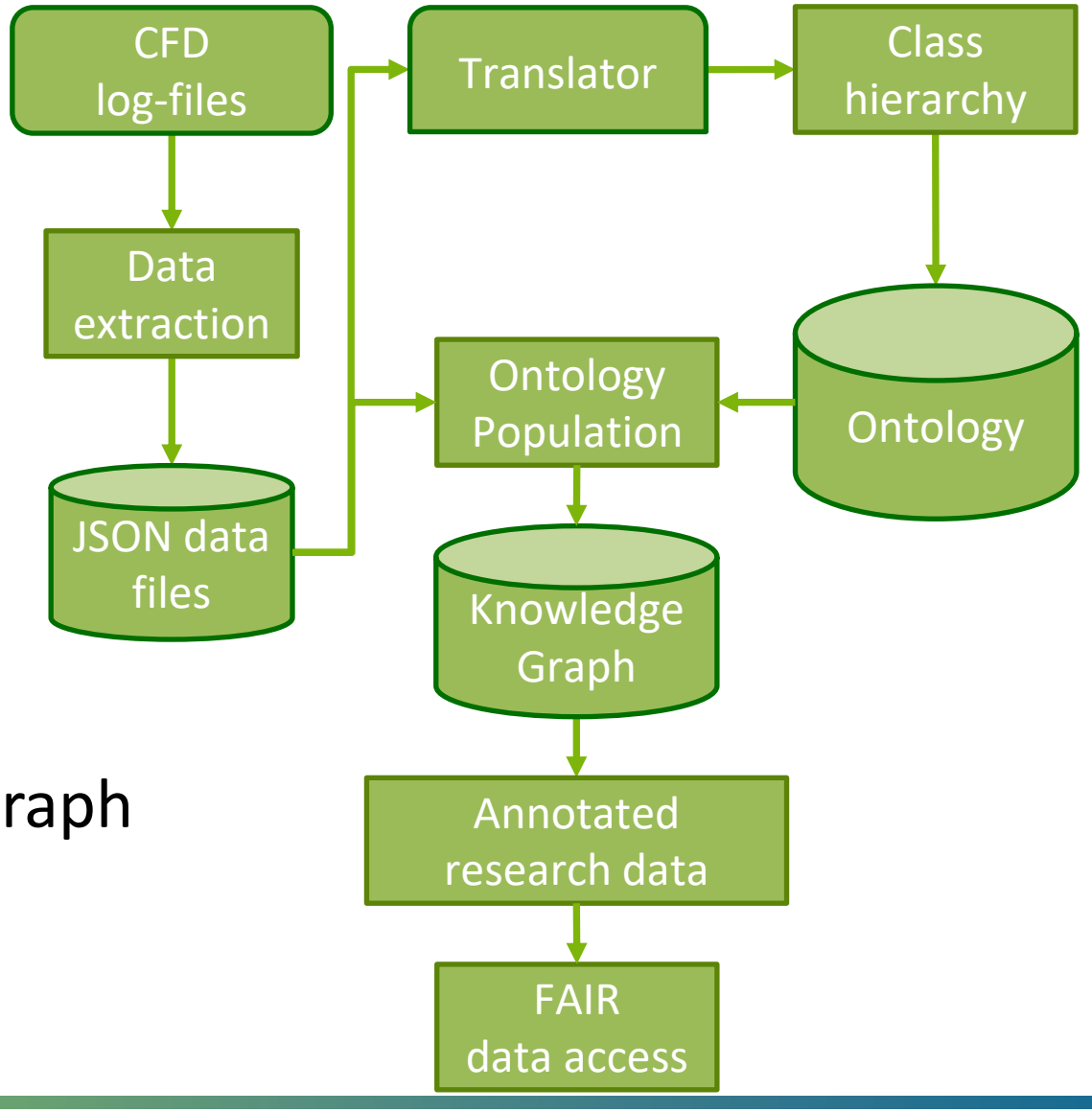
```
…
PARTITION SMOOTHING:
        Maximum Partition Smoothing Sweeps = 100
        Option = Smooth
      END
      PARTITIONING TYPE:
        MeTiS Type = k-way
        Option = MeTiS
        Partition Size Rule = Automatic
        Partition Weight Factors = 0.02500, 0.02500, 0.02500, 0.02500, \
0.02500, 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, 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.                                |
+--------------------------------------------------------------------+
```

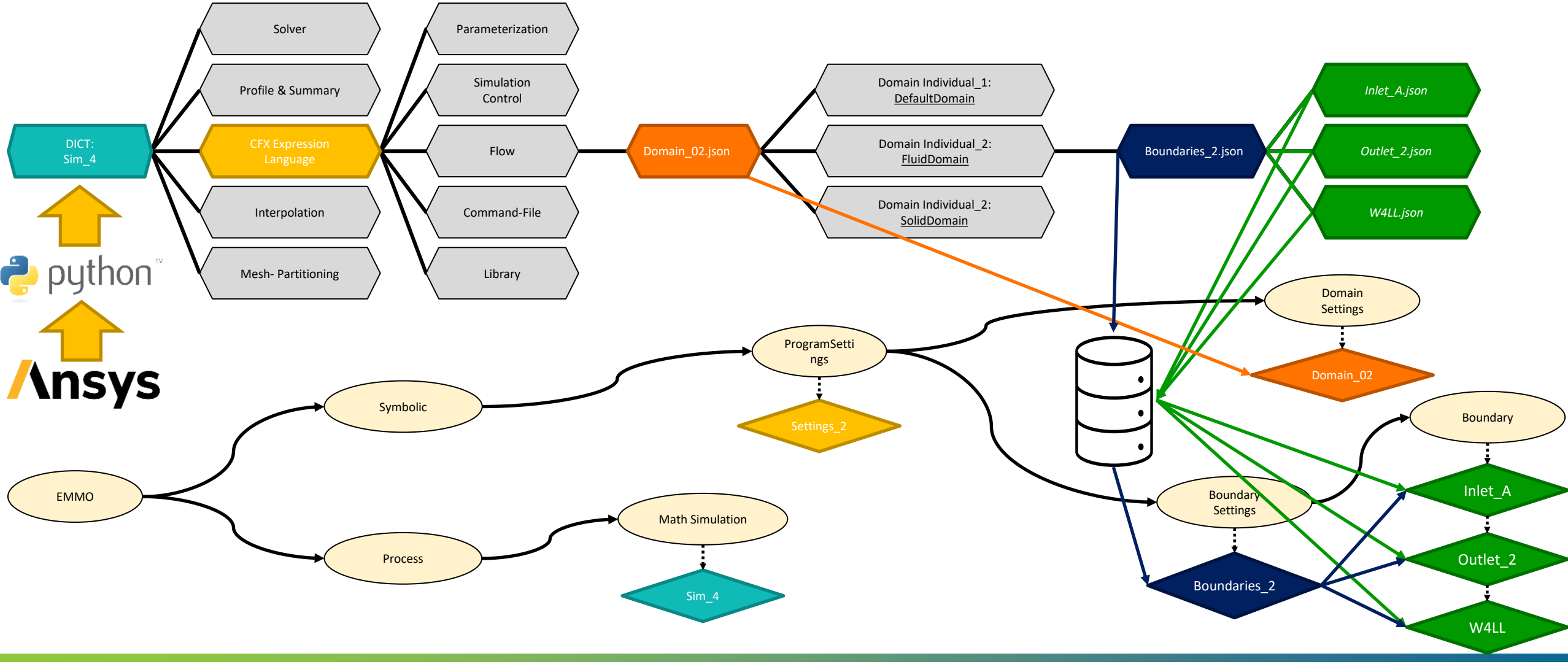
Workflow

- Structuring (meta)data files as JSON-data
- Translation of the data needed to include entries as classes into ontology
- Existing ontology reused and extended by classes
- Ontologies + data = Knowledge Graph

➢ FAIRer data access



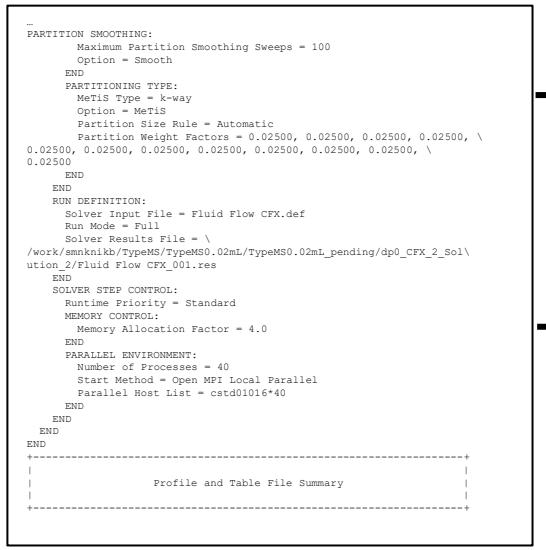
Data to ontology connection



Results

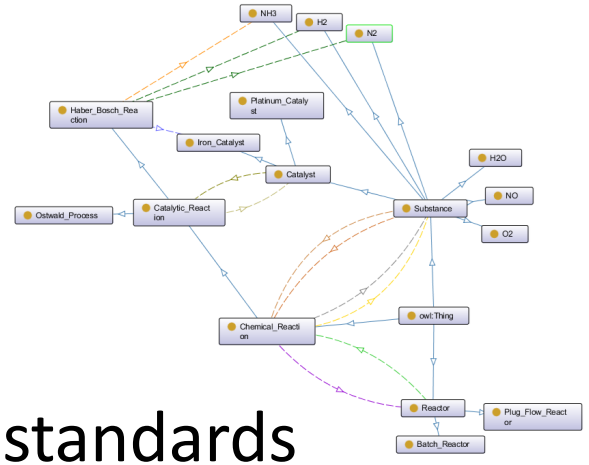
- Raw data annotated using ontologies
- Most promising simulations obtained by ontology-based knowledge graph
- ➢ Automated annotation of simulation results with corresponding metadata independent of simulator

Ontology	EMMO	EMMO + 911 Simulations	
Classes	470	788	Δ 318
Object Prop.	47	190	Δ 143
Data Prop.	3	244	Δ 241
Individuals	1	28,818	Δ 28,817
Axioms	3,363	756,719	Δ 753,356

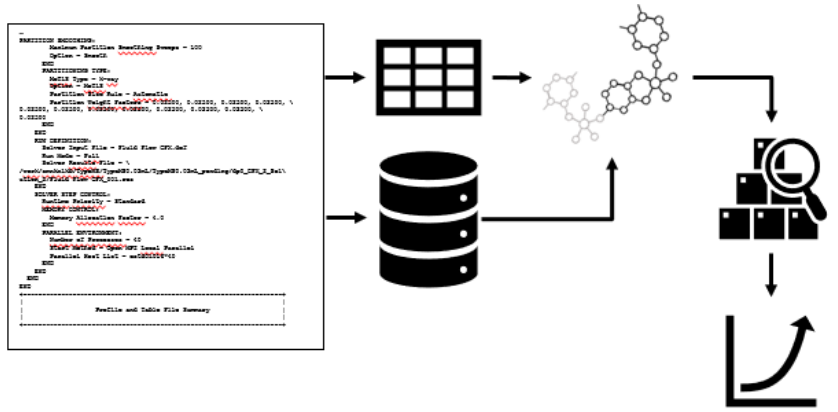


Wrap-Up

- Ontologies
  - can be set up with, e.g. Protégé
  - allow for classification of data
  - can provide for concepts needed for metadata standards



- Data-FAIRness can be enhanced by (among others)
  - annotations based on ontologies
  - automated metadata extraction based on ontologies



Further Starting Points for Starters

- Protege
  - https://protege.stanford.edu/
- Ontology101
  - https://ontology101tutorial.readthedocs.io/en/latest/StartingProtege.html
- Ontology Collection and Introduction to Ontologies
  - https://nfdi4cat.org/ontology-collection/
- W3C documentation on Web Ontology Language – OWL
  - https://www.w3.org/TR/owl-features/

NFDI4Cat

NFDI for Catalysis-Related Sciences

Thank you for your attention!

Would you like to exchange ideas or become a part of the community?

Dr. Sara Espinoza

Phone +49-69-7564-354

Email sara.espinoza@dechema.de

http://www.nfdi4cat.org/ info@nfdi4cat.org /linkedin.com/company/nfdi4cat @NFDI4Cat

Alexander Behr, Norbert Kockmann

JCDL – 23.06.2022