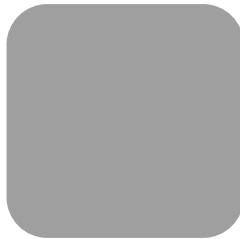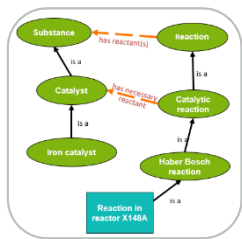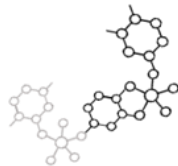# Modelling Knowledge with Ontologies for Catalysis Research

Machine Learning and Modelling Seminar
Charles University, Prague, September 29, 2022
<u>Alexander S. Behr</u>, Norbert Kockmann

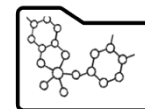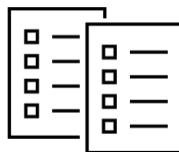Alexander S. Behr | Prague 29.09.2022

Agenda

Motivation → Current State of Ontologies → Terminologies → NLP-Applications → Annotation of Research Data → Conclusion / Outlook

Department of Biochemical and Chemical Engineering

equipment design

What is an ontology?

- An ontology is a
  formal, **explicit specification** of a **shared** **conceptualization**.*

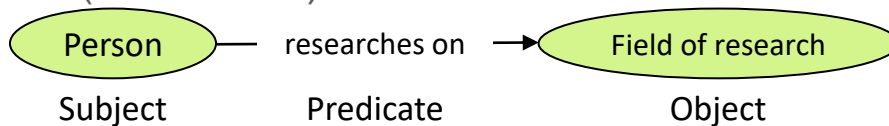| Machine readable	Concepts, properties, relations, functions, constraints, axioms are explicitly defined.	A consensus rather than an individual view.	An abstract model of some phenomenon in the world that we want to represent.

Guarino, Nicola, Daniel Oberle and Steffen Staab. "What Is an Ontology?" Handbook on Ontologies (2009).

Ontologies – A Simple Example

- Ontologies consist of
  - **Classes** (to express concepts)
  - **Relations** between classes
  - **Individuals** representing real existing elements
  - **General Rules**, like
    "All catalytic reactions need one or more catalyst"
- Information is stored as triplets
  in OWL-file (RDF+XML)

Person — researches on → Field of research

Subject   Predicate   Object
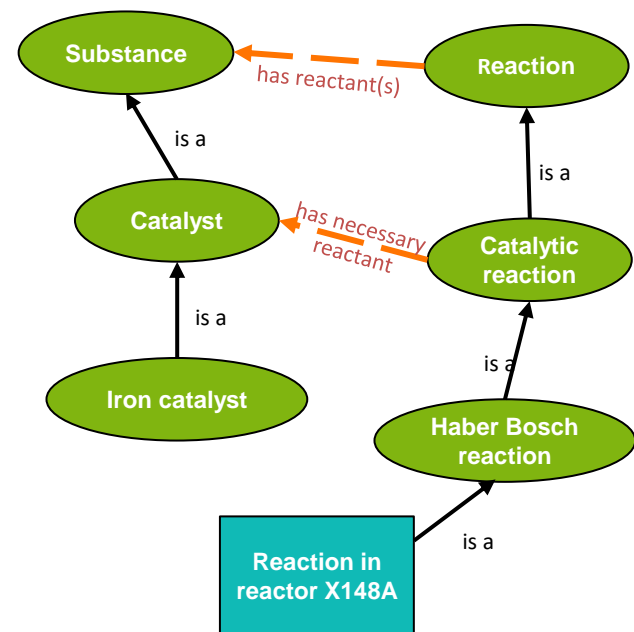
Ontologies – A Simple Example

- Ontologies consist of
  - **Classes** (to express concepts)
  - **Relations** between classes
  - **Individuals** representing real existing elements
  - **General Rules**, like
    "All catalytic reactions need one or more catalyst"
- Information is stored as triplets
  in OWL-file (RDF+XML)

Person	researches on →	Field of research
Subject	Predicate	Object

Substance ←– has reactant(s) –– Reaction

Catalyst ←– has necessary reactant –– Catalytic reaction

Substance – is a – Catalyst

Reaction – is a – Catalytic reaction

Catalyst – is a – Iron catalyst

Catalytic reaction – is a – Haber Bosch reaction

Haber Bosch reaction – is a – Reaction in reactor X148A

Department of Biochemical and Chemical Engineering

equipment design
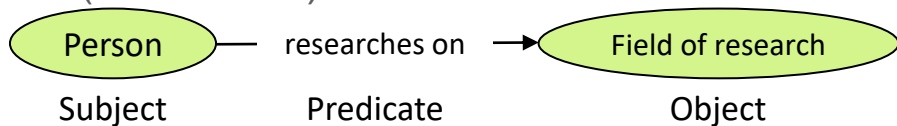
Ontologies – A Simple Example

- Ontologies consist of
  - **Classes** (to express concepts)
  - **Relations** between classes
  - **Individuals** representing real existing elements
  - **General Rules**, like
    "All catalytic reactions need one or more catalyst"
- Information is stored as triplets
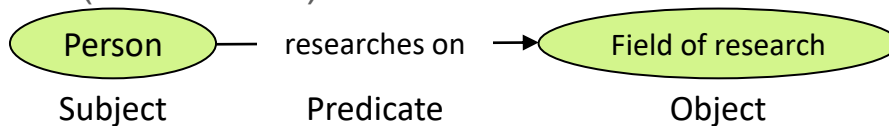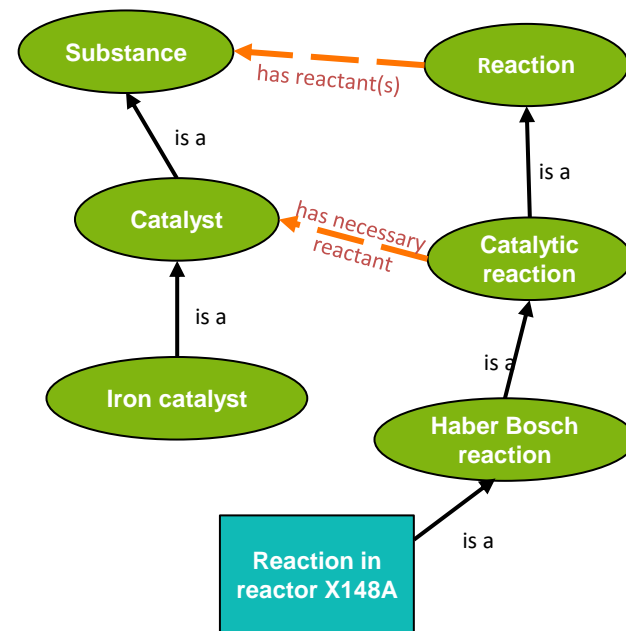 in OWL-file (RDF+XML)

- Reasoning enhances data
  - ➢ Inference can yield: "The reaction in reactor X148A is a Haber-Bosch reaction, which in turn is a catalytic reaction and uses iron catalyst as catalyst."

Department of Biochemical
and Chemical Engineering

equipment
design

Ontologies – How can we use them?

- Interconnect (meta) data
- ➤ Aim: Machine- and human-readable (meta) data

Main advantages:

- ➤ Enhance findability of available data by classification + relations
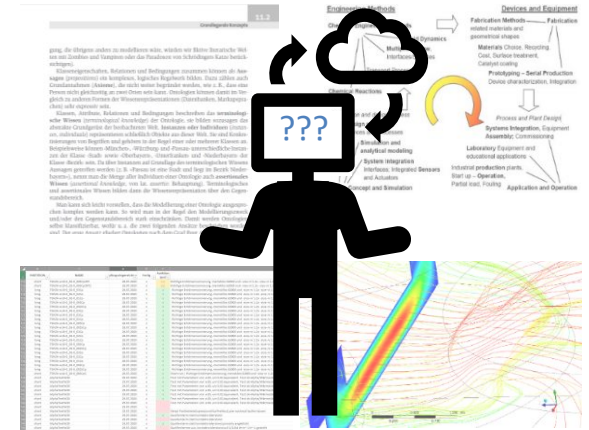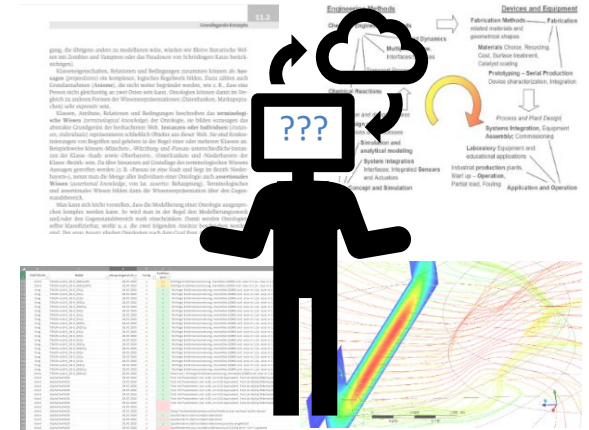- ➤ Provide better interoperability of data

Ontologies – How can we use them?

- Interconnect (meta) data
- ➤ Aim: Machine- and human-readable (meta) data

Main advantages:

- ➤ Enhance findability of available data by classification + relations
- ➤ Provide better interoperability of data

➡ Unified data formats through *ontologies* and standardized *metadata schemes*

Motivation

- Unified data formats through ontologies and standardized metadata schemes

- Work smart, not hard
- Are there reusable ontologies?

Motivation

- Unified data formats through ontologies and standardized metadata schemes

- ➤ Work smart, not hard
- ➤ Are there reusable ontologies?

Situation to be avoided:
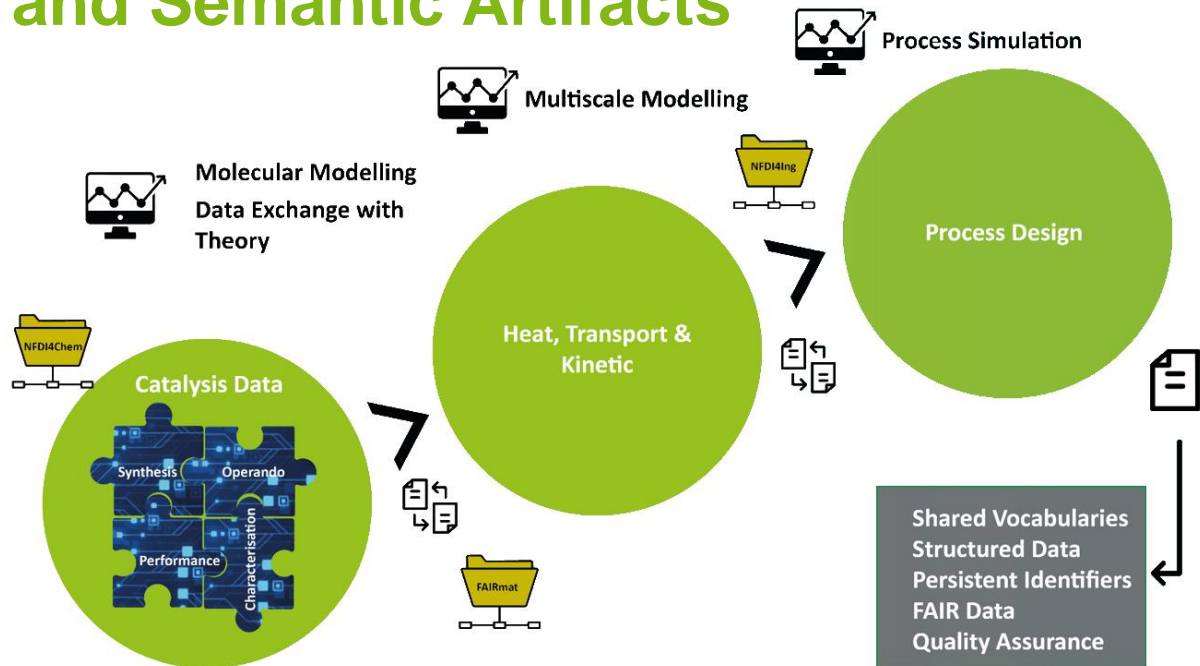
[1]

[1] R. Munroe, xkcd, https://xkcd.com/927

Ontology Collection

One Ontology to Rule Them All?

Map of Ontologies and Semantic Artifacts

- Interdisciplinary field of catalysis research
- Strong connection to other research domains

Map of Ontologies and Semantic Artifacts

- Interdisciplinary field of catalysis research

- Strong connection to other research domains

🌐 nfdi4cat.org/ontology-collection

[1]

[1] Horsch, M. et al. (2022). Interoperability and Architecture Requirements Analysis and Metadata Standardization for a Research Data Infrastructure in Catalysis. Communications in Computer and Information Science (CCIS), vol 1620. Springer, Cham. https://doi.org/10.1007/978-3-031-12285-9_10

Map of Ontologies and Semantic Artifacts

- Interdisciplinary field of catalysis research
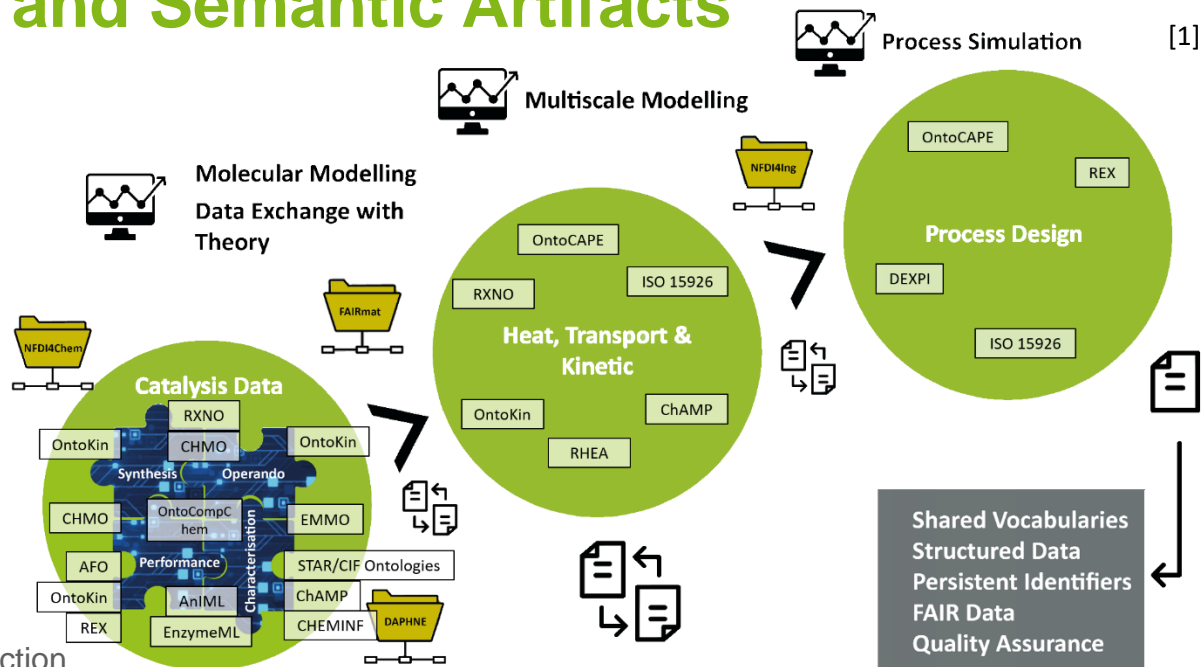
➡️

- Strong connection to other research domains

Some "ontologies" existing, but semantics not sufficient to describe catalysis research

🌐 nfdi4cat.org/ontology-collection

Department of Biochemical and Chemical Engineering

equipment design

Ontology Extension

Domain Description by Concept Collection

Ontologies, Where to Start?

Intermediate deliverable in data format
SKOS = Simple Knowledge Organization System
Part of linked Open Vocabularies graph.

strong semantics

Ontology — description logic
OWL

Conceptual Models — formal instance-of, properties, cardinalities
UML, RDFS, OWL

semantic interoperability

Taxonomy — formal is-a, subclass-of generalization hierarchy
RDFS

- -

structural and syntactic interoperability

Thesaurus — 'has narrower/broader meaning than', synonyms, association relations
SKOS

Informal Hierarchy (weak Taxonomy) — informal parent-child (directories, table of contents, XML ...)
XML

List — glossary, dictionary controlled vocabulary, set of terms, catalog ID

weak semantics

[1] Goldbeck, Gerhard & Simperler, Alexandra (2018). https://doi.org/10.5281/zenodo.1240229 (CC-By 4.0)

Ontology Extension by Thesauri (SKOS)

Existing Ontologies
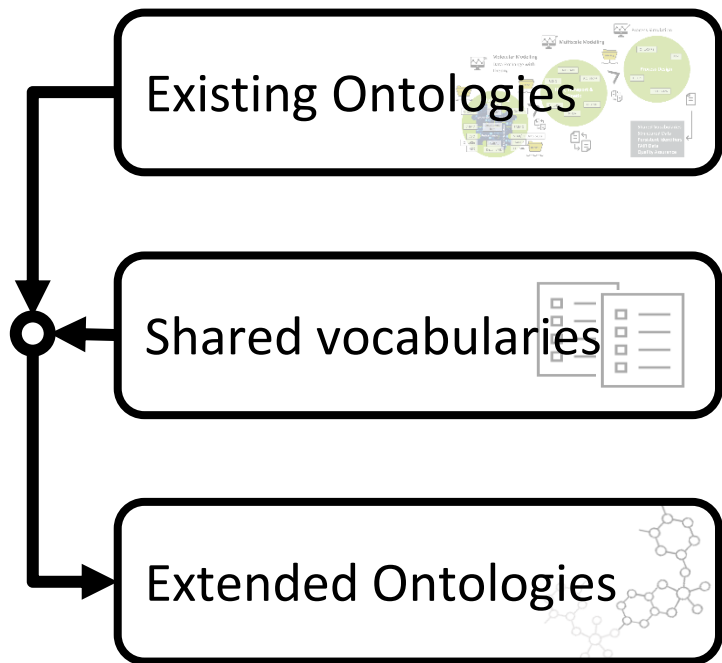
- ✓ Existing ontologies gathered
- ✓ Clustered by topics of catalysis research

Shared vocabularies

- • Concept collection using Excel templates
- • Workflow for automating SKOS file generation

Extended Ontologies

- ➢ Ontologies for catalysis research
- ➢ Extended by concepts of community

Concepts / Taxonomy Collection

- Gather concepts important to different domains in catalysis
- Ease contribution by domain experts by using spreadsheets

Column name	Description
Preferred Label	Best name for a concept
Alternate Labels	Other names for a concept
Definition	The definition of the concept
Children	Subclasses of this concept
Related	Other related concepts, that are not subclasses

Shared Vocabulary: Concept Tables

- Convert vocabulary to machine-readable SKOS / TTL-files

➢ Description of processes and data with shared vocabularies

Preferred Label	Alternate Labels	Definition	Children
Biologic Catalysis			bio catalyst preparation, protein purification, bio catalyst application
Bio Catalyst Preparation		The Catalyst Preparation is a process that is intended to synthesize a catalyst from several components.	cultivation, induction, cell harvesting, cell disruption, gel filtration, protein concentration, transformation
Expression Strain preparation	Strain preparation		
Expression host	expression host, organism, microorganism, bacteria, yeast		
...	...	...	...

Concept-table → SKOS-converter.py → Documentation (html), Taxonomy as SKOS/turtle-file, Dendrogram

[1] M. Menke, A. S. Behr, K. Rosenthal, D. Linke, N. Kockmann, U. Bornscheuer, M. Dörr. Development of an ontology for biocatalysis, Chemie Ingenieur Technik, [In Production], doi: 10.1002/cite.202200066, (2022)

How to Address "Shared" Aspect / Collaboration?

- Needs
  - Enable contributions from the community
  - Liberal license (ideally CC0, no copyright to ensure usage!)
  - Track discussions and resolutions about the conceptualization & terminologies
  - Enable versioning and provenance tracking
  - Keep effort for contributors and maintainers low
  - Publish updated vocabulary quickly after accepted changes

- Open-source development has already solved this challenge!

- Adapt modern open-source contribution workflow to NFDI4Cat

Vocabulary Contribution Workflow
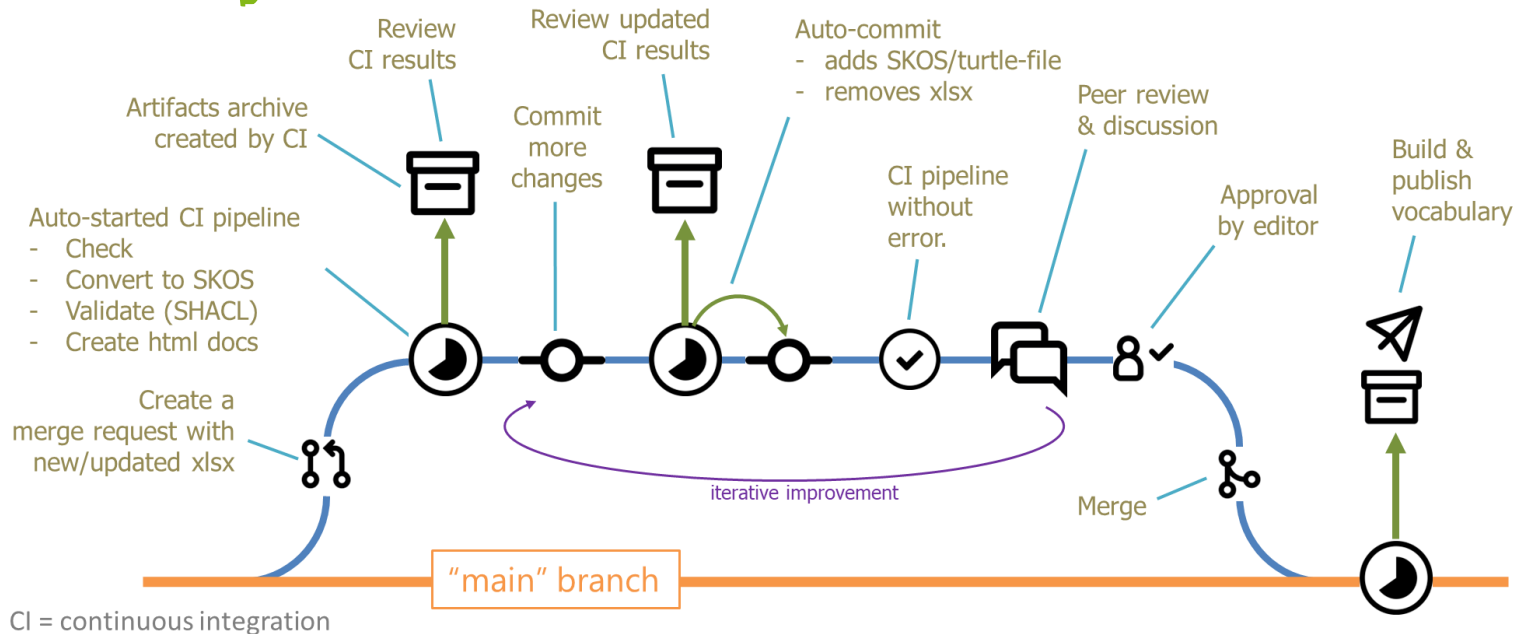
Review CI results

Review updated CI results

Auto-commit
- adds SKOS/turtle-file
- removes xlsx

Peer review & discussion

Build & publish vocabulary

Artifacts archive created by CI

Commit more changes

CI pipeline without error.

Approval by editor

Auto-started CI pipeline
- Check
- Convert to SKOS
- Validate (SHACL)
- Create html docs

Create a merge request with new/updated xlsx

iterative improvement

Merge

"main" branch

CI = continuous integration

Slide adapted from presentation of D. Linke: „Ontology development in Catalysis: Status of NFDI4Cat-TA1", A.S.Behr, D.Linke, M.Dörr, N.Kockmann, held on 55. Jahrestreffen Deutscher Katalytiker, 27.06.2022

Create a Merge Request

- Submit change requests as Excel (or SKOS/turtle if an expert)
- Vocabulary processing pipeline will be started upon submission

Find out more.

"artifacts":
- Files created by the pipeline
- log

Slide adapted from presentation of D. Linke: „Ontology development in Catalysis: Status of NFDI4Cat-TA1", A.S.Behr, D.Linke, M.Dörr, N.Kockmann, held on 55. Jahrestreffen Deutscher Katalytiker, 27.06.2022

Understanding and Fixing a Pipeline Failure

Excel file from artifacts zip-archive

```
246  Saved updated file as outbox/Photocatalsis_LIKAT_template043.xlsx
247  Running check of Concepts sheet for file outbox/Photocatalsis_LIKAT_template043.xlsx
248  ERROR: Same Conce
249  ERROR: Same Conce
250  ERROR: Same Conce
251  ERROR: Same Conce
252  ERROR: Same Conce
253  Saved file with h
254  Calling VocExcel
255  Processing file outbox/Photocatalsis_LIKAT_template043.xlsx
```

Concept IRI	Preferred Label	Definition
ex:photo/gas-chromatography	Gas chromatography	A separation technique in which the mobile phase is a gas. Gas chromatography is always carried out in a column.
ex:photo/gas-chromatography	Gas chromatography	An analytic technique in which the mobile phase is a gas. Gas chromatography is always carried out in a column.

Alexander S. Behr | Prague 29.09.2022

Understanding and Fixing a Pipeline Failure

Excel file from artifacts zip-archive

```
246  Saved updated file as outbox/Photocatalsis_LIKAT_template043.xlsx
247  Running check of Concepts sheet for file outbox/Photocatalsis_LIKAT_template043.xlsx
248  ERROR: Same Conce
249  ERROR: Same Conce
250  ERROR: Same Conce
251  ERROR: Same Conce
252  ERROR: Same Conce
253  Saved file with h
254  Calling VocExcel
255  Processing file outbox/Photocatalsis_LIKAT_template043.xlsx
```

Concept IRI	Preferred Label	Definition
ex:photo/gas-chromatography	Gas chromatography	A separation technique in which the mobile phase is a gas. Gas chromatography is always carried out in a column.
ex:photo/gas-chromatography	Gas chromatography	An analytic technique in which the mobile phase is a gas. Gas chromatography is always carried out in a column.

- Here, validation failed because the same concept has been defined more than once

➢ Remove the respective rows and push the updated Excel file

Slide adapted from presentation of D. Linke: „Ontology development in Catalysis: Status of NFDI4Cat-TA1", A.S.Behr, D.Linke, M.Dörr, N.Kockmann, held on 55. Jahrestreffen Deutscher Katalytiker, 27.06.2022

Merge Request View for the Changes Made

- Changes are very easy to review for maintainers…
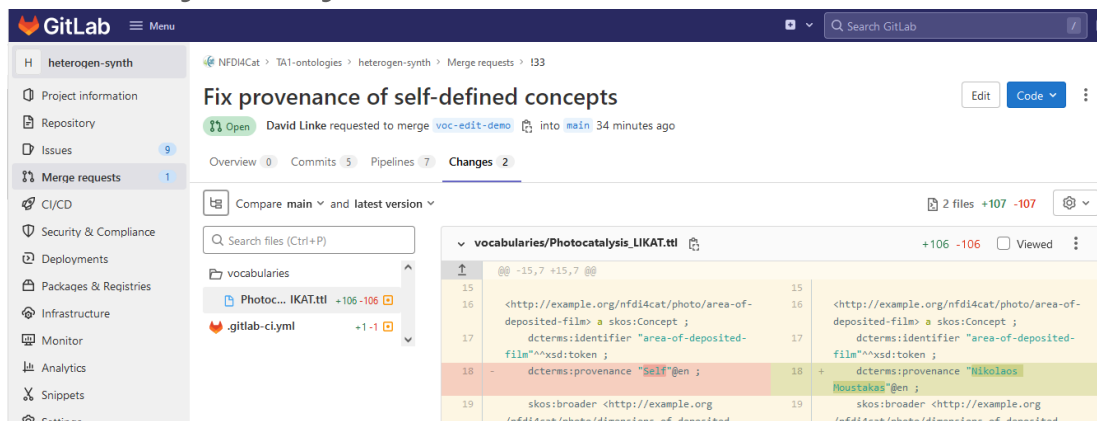
…and can be discussed directly at the merge request using all nice features that gitLab offers

Slide adapted from presentation of D. Linke: „Ontology development in Catalysis: Status of NFDI4Cat-TA1", A.S.Behr, D.Linke, M.Dörr, N.Kockmann, held on 55. Jahrestreffen Deutscher Katalytiker, 27.06.2022

Manual Terminology Creation?

There Must be a Better Way!

From Ontologies to Knowledge Graphs

- First steps taken in ontology analysis and extension

- Tedious work, a lot of manual input needed

➢ Need definition of each concept in order to implement correct semantics

➢ Can we automate tedious tasks?
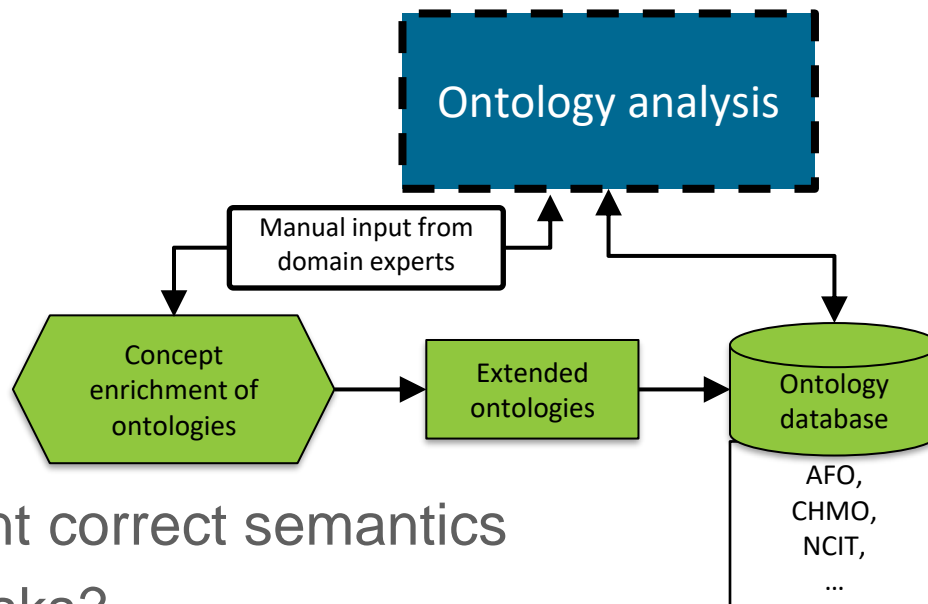
Concept Extraction – Should Be Simple?

- Adding terms manually to concept-tables is a lot of work!
- Semi-automated workflow using natural-language processing (NLP) tools
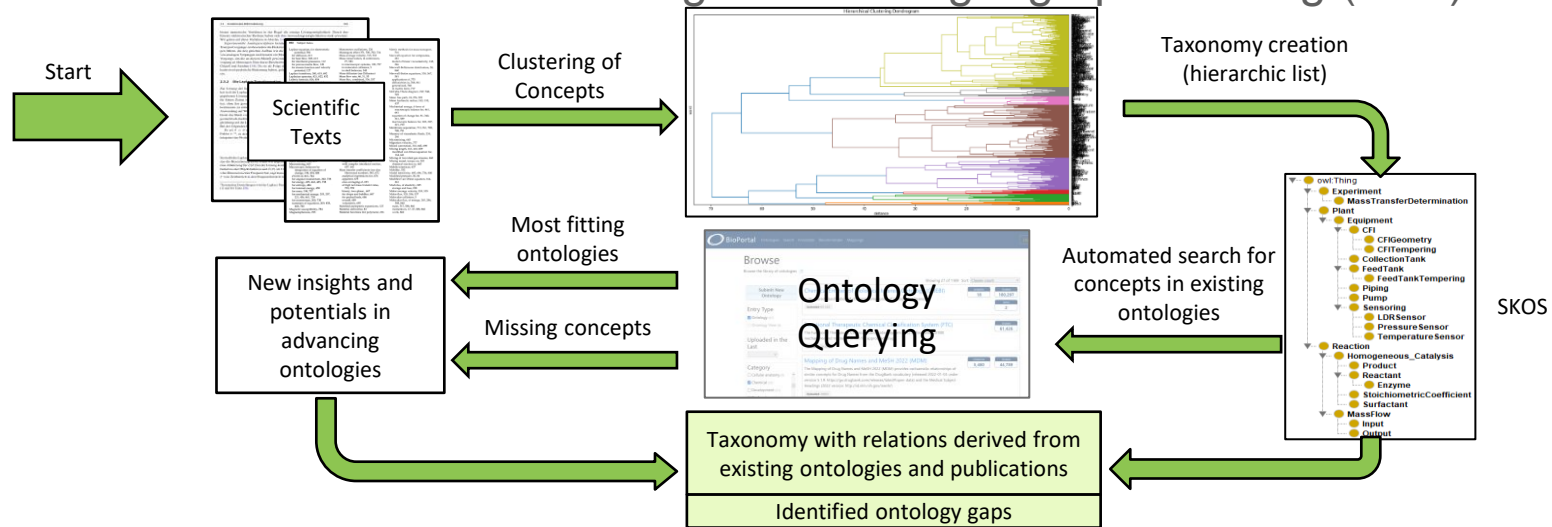
Clustering with Word2Vec

- Preprocessing of texts
  - Data cleaning
  - Pythons SpaCy for Tokenizing, Lemmatizing, POS-Tagging
- Vectorization of words with Word2Vec

```
Text Extraction → Preprocessing → Word2Vec Training → Cluster Analysis & Data Storing
```
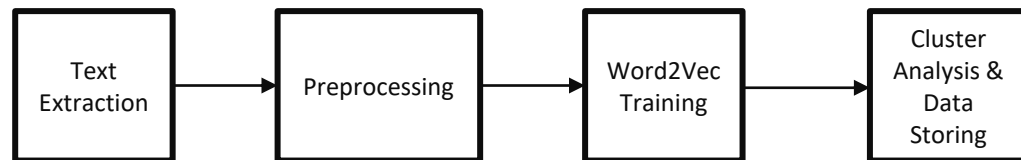
Clustering with Word2Vec

- Preprocessing of texts
  - Data cleaning
  - Pythons SpaCy for Tokenizing, Lemmatizing, POS-Tagging
- Vectorization of words with Word2Vec
- Hierarchial clustering
  - Iterative, agglomeration clustering
- ➢ Concepts relevant domain of knowledge

[1] 10.12.2021 https://towardsdatascience.com/hierarchical-clustering-explained-e58d2f936323
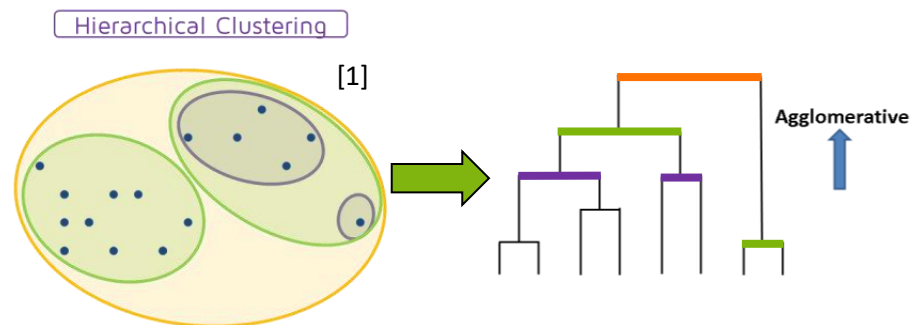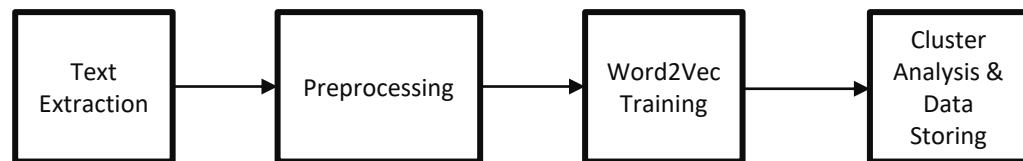
Clustering with Word2Vec

- Storage of hierarchies in JSON-files database

- Names for unknown nodes via hypernym search with WordNet

➢ Search for concepts in existing ontologies
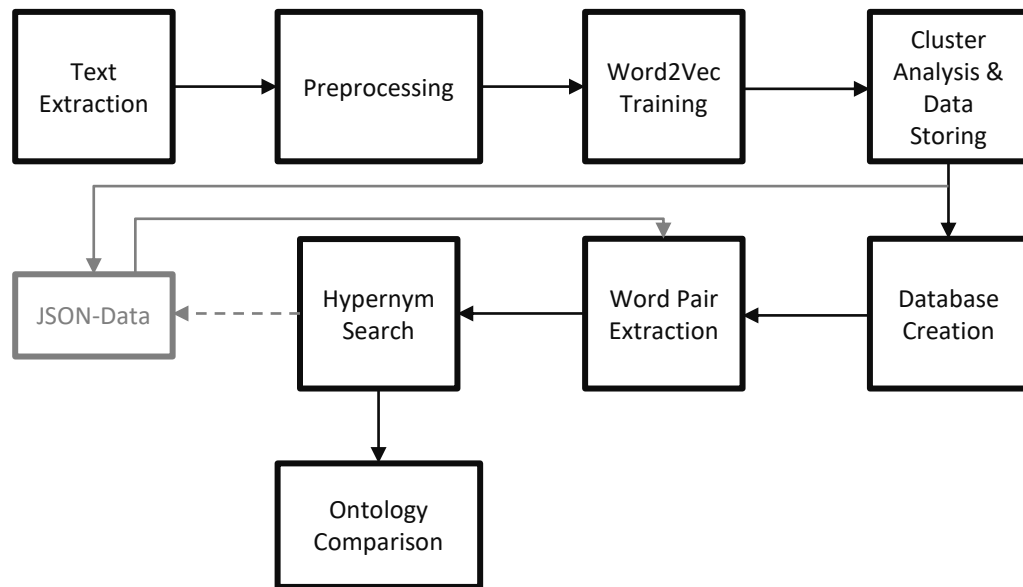
How Did It Perform?

- Dataset: 28 papers on methanation of CO_2
- Found 4170 different words in dataset which occurred more than 10 x (nouns only)
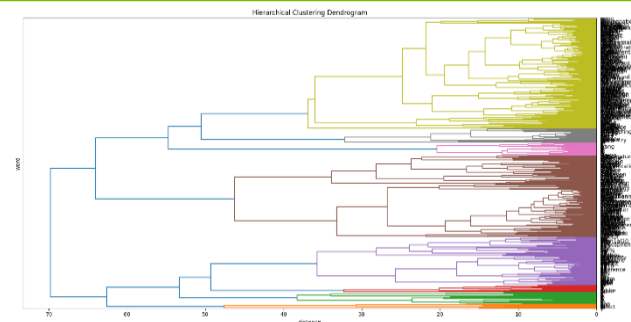
Dendrogram for a min_count of 10 (Word has >10 repetitions in dataset)

How Did It Perform?

- Dataset: 28 papers on methanation of CO_2
- Found 4170 different words in dataset which occurred more than 10 x (nouns only)

- Clustering is not that helpful due to approach
  - Only two concepts at a time combined as siblings
  - Semantic similarity detected by Word2Vec useful only to extend
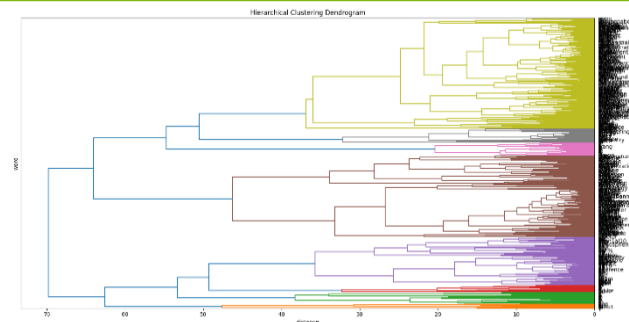
Dendrogram for a min_count of 10 (Word has >10 repetitions in dataset)
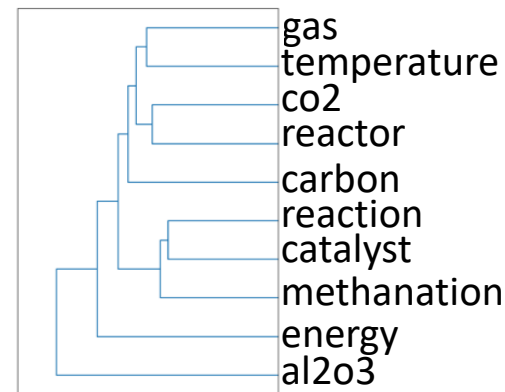
Dendrogram for a min_count of 500 (Word has >500 repetitions in dataset)

How Did It Perform?

- Dataset: 28 papers on methanation of CO_2
- Found 4170 different words in dataset (nouns only)
- Listed words, which exist more than 10 times in dataset
  - ➢ Found 535 domain specific words
- Searched for definition of those words in four ontologies / thesauri (exact matches of class label and word within paper)
  - ➢ Overall, 63.6 % matches of definitions (340 definitions found)
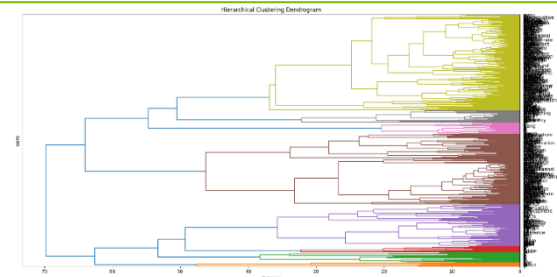
Ontology	CHMO	AFO	CHEBI	NCIT	Overall
# of definitions found	21	98	27	311	340
% of total words	3.9	18.3	5.0	58.1	63.6

Preliminary Summary

✓ SpaCy and Word2Vec are helpful tools for automating word extraction from text data base

✓ Searching for words in ontologies gives (rough) idea of appropiate ontology, but no semantics!

✗ Clustering of word pairs does not help to compile hierarchical lists

Ontology	CHMO	AFO	CHEBI	NCIT	Overall
# of definitions found	21	98	27	311	340
% of total words	3.9	18.3	5.0	58.1	63.6

Hierarchical Clustering Dendrogram
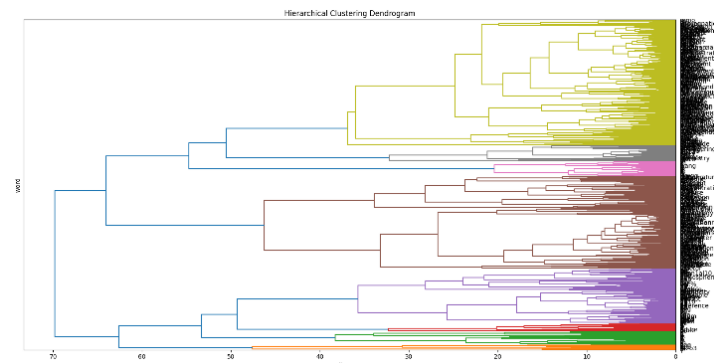
If NLP Could Already Help Us

Can It Help Even More?

From Ontologies to Knowledge Graphs

- First steps taken in ontology analysis and NLP-supported term extraction

➢ Need extended ontologies for catalysis research!

➢ Can we use even more NLP?

PDFs

Scientific theses, papers, books, …

NLP-based term extraction

Ontology analysis

Manual input from domain experts

Concept enrichment of ontologies

Extended ontologies

Ontology database(s)

AFO, CHMO, NCIT, …

Text to Graph with ChEBI[1]

„The gas-liquid ratio affects the behavior of the packed column, removing H_2S from biogas at the active surface of the packing material"

```
┌──────────────┐         ┌──────────────────────┐
│  Input:      │ ──────▶ │   NLP with SpaCy     │
│  sentence    │         └──────────────────────┘
└──────────────┘                    │
                                    ▼
                         ┌──────────────────────┐
                         │    Preprocessing     │
                         └──────────────────────┘
                                    │
                                    ▼
                         ┌──────────────────────┐
                         │ Filtering of irrelevant and │
                         │ unprocessable sentences │
                         └──────────────────────┘
```

[1] ChEBI – Chemical Entities of Biological Interest ; dictionary of molecular entities focused on 'small' chemical compounds, https://www.ebi.ac.uk/chebi/

Department of Biochemical and Chemical Engineering

equipment design

Text to Graph with ChEBI[1]

„The gas-liquid ratio affects the behavior of the packed column, removing H_2S from biogas at the active surface of the packing material"

[1] ChEBI – Chemical Entities of Biological Interest ; dictionary of molecular entities focused on 'small' chemical compounds, https://www.ebi.ac.uk/chebi/

Text to Graph with ChEBI[1]

„The gas-liquid ratio affects the behavior of the packed column, removing H_2S from biogas at the active surface of the packing material"

[1] ChEBI – Chemical Entities of Biological Interest ; dictionary of molecular entities focused on 'small' chemical compounds, https://www.ebi.ac.uk/chebi/

Graph Representation of Sentences

„The gas-liquid ratio affects the behavior of the packed column,
removing H$_2$S from biogas at the active surface of the packing material"

Graph Representation of Sentences

„The gas-liquid ratio affects the behavior of the packed column,
removing H$_2$S from biogas at the active surface of the packing material"

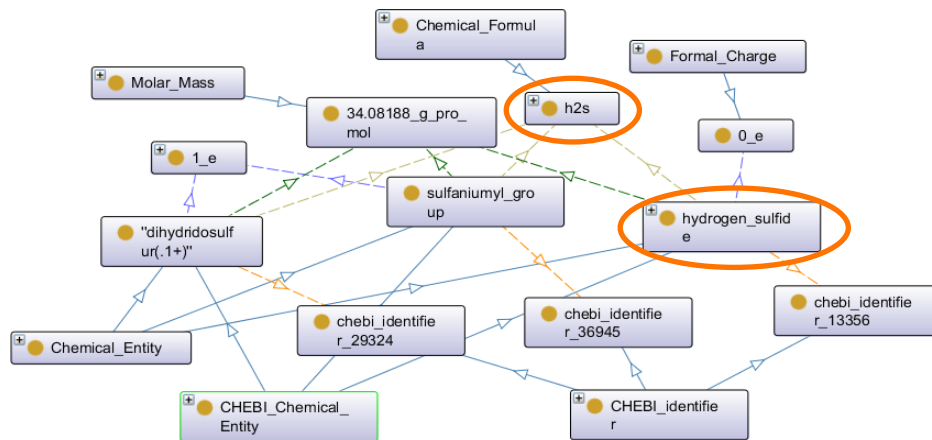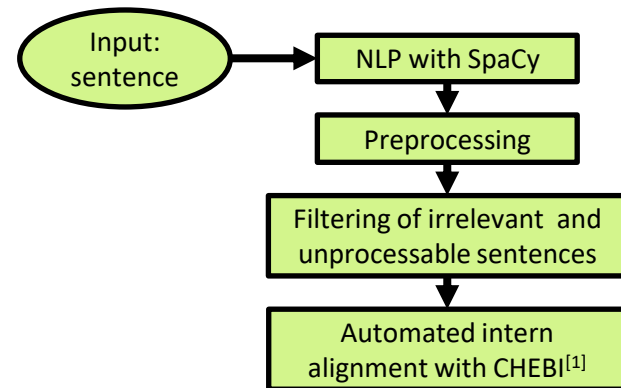

Independent Clause

Graph Representation of Sentences

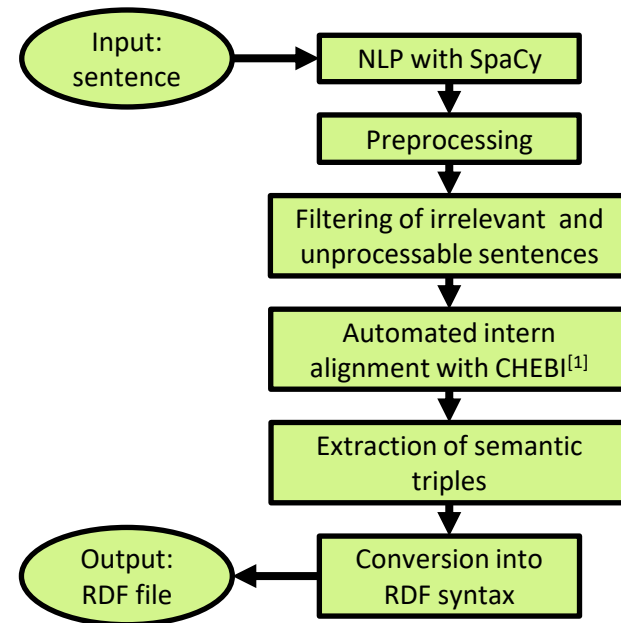*„The gas-liquid ratio affects the behavior of the packed column, removing H_2S from biogas at the active surface of the packing material"*

CHEBI Alignment
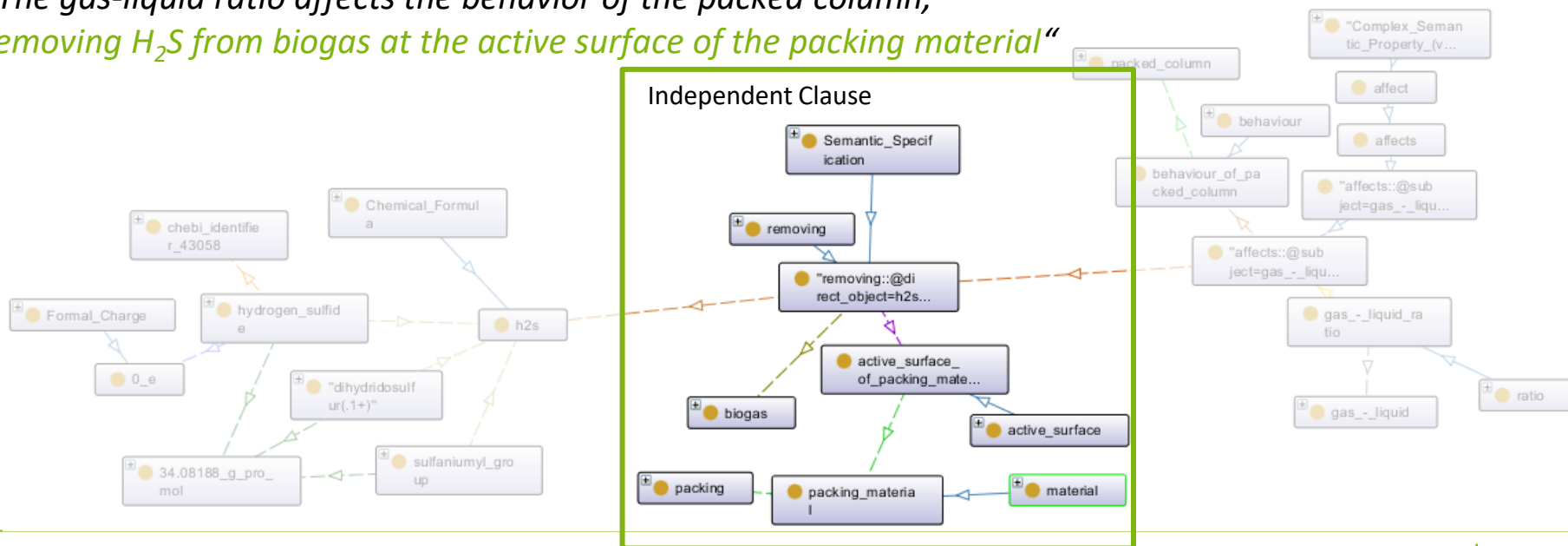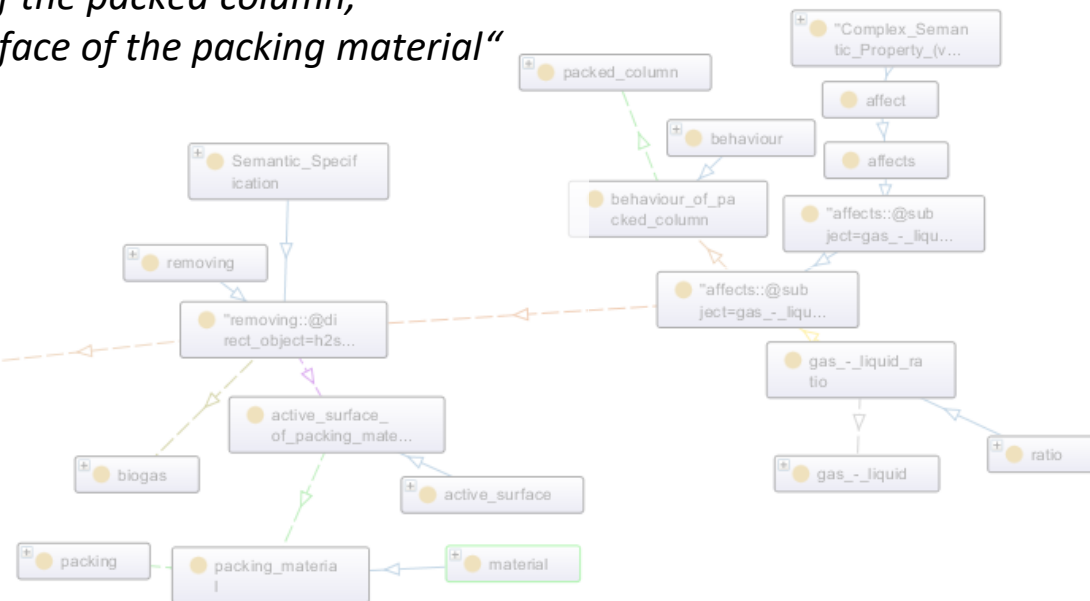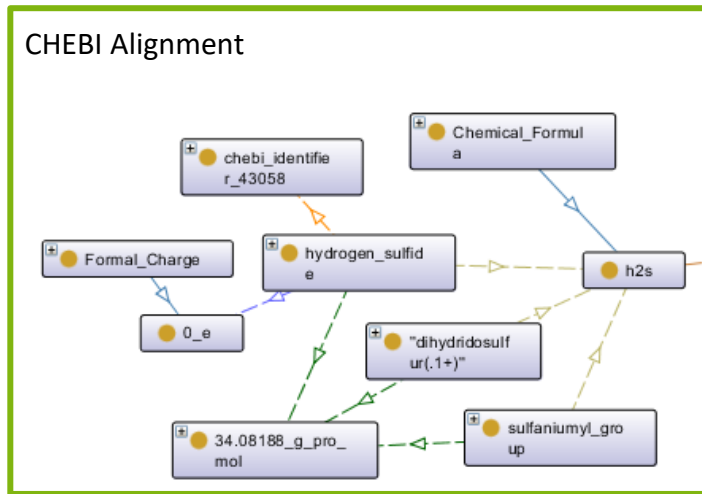
Intermediate Summary

✓ Pythons NLTK, RDFlib, SpaCy and Word2Vec are helpful tools for automating word extraction from text data base

✓ Automatically linking concepts to ontology classes

✓ Relations from text can be converted to RDF graphs
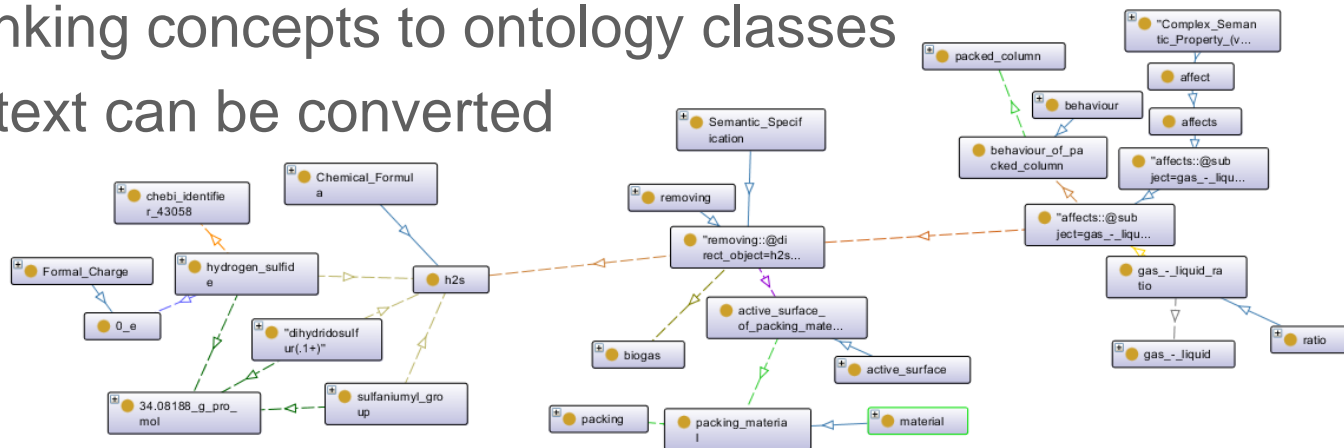
Okay, Now We Have Some Semantics.

But How Can Ontologies
Help On Research Data?

From Ontologies to Knowledge Graphs
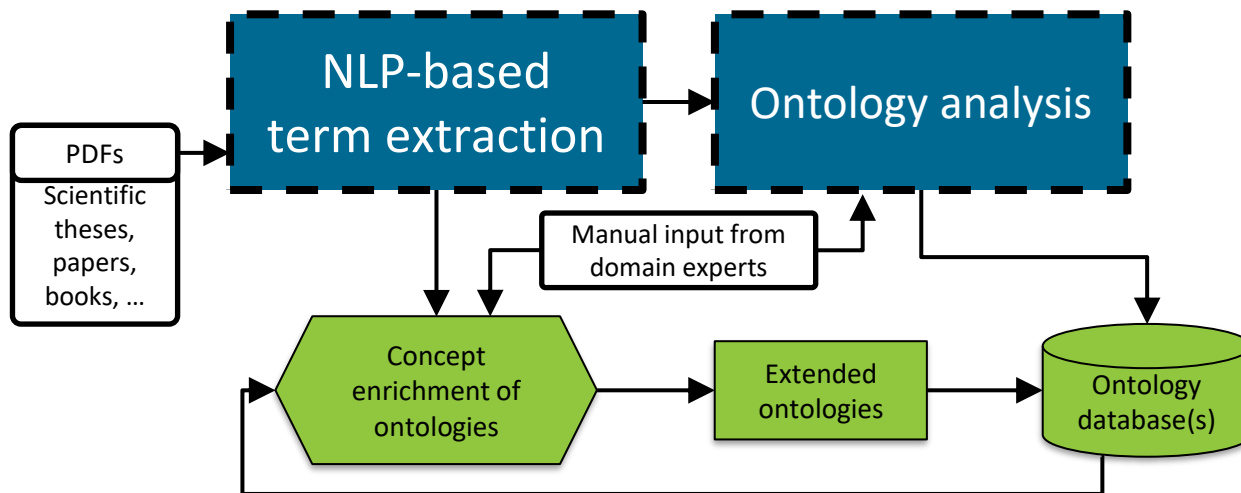
From Ontologies to Knowledge Graphs
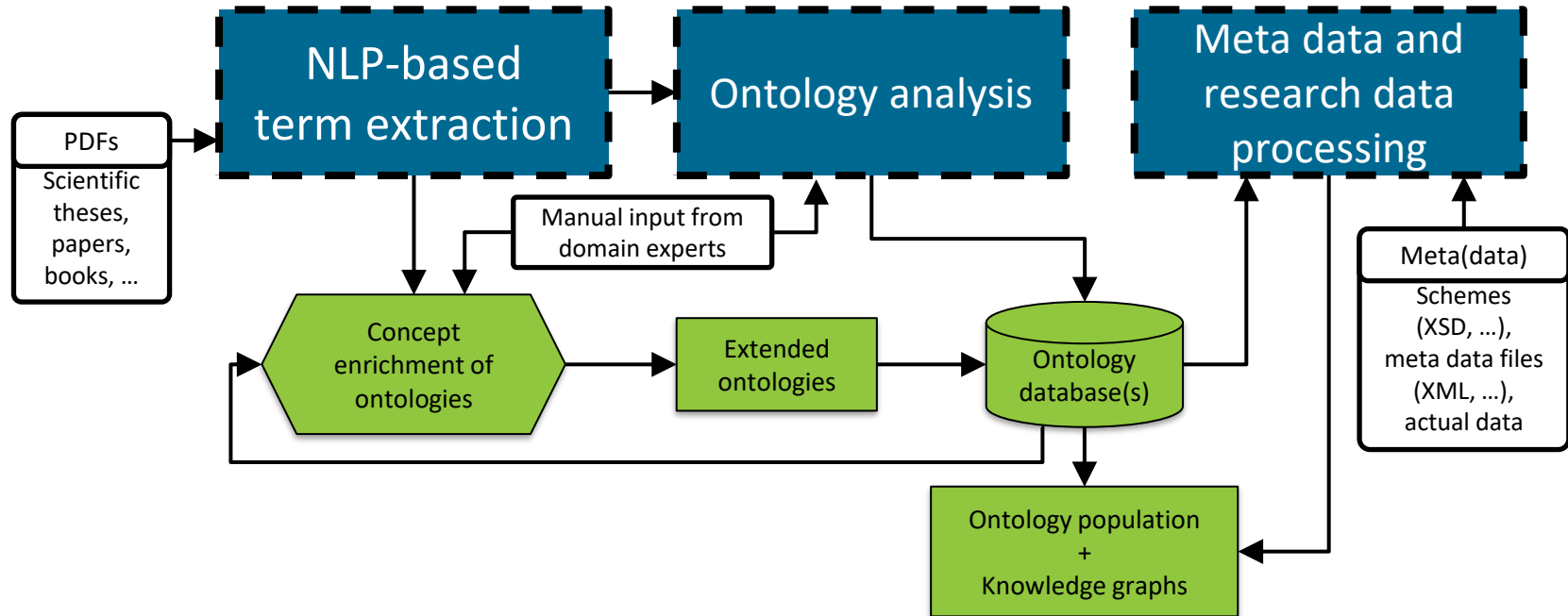
Motivation - Data Origin

- Data availability
  - Experimental [1,2]
  - Theoretical [3,4]
    - Computational fluid dynamics (CFD)
    - Hydrodynamic & reaction in LTF Type-S microreactor (Little Things Factory, DE)
- CFD-Data
  - Segmentation of domain
  - Variety of settings
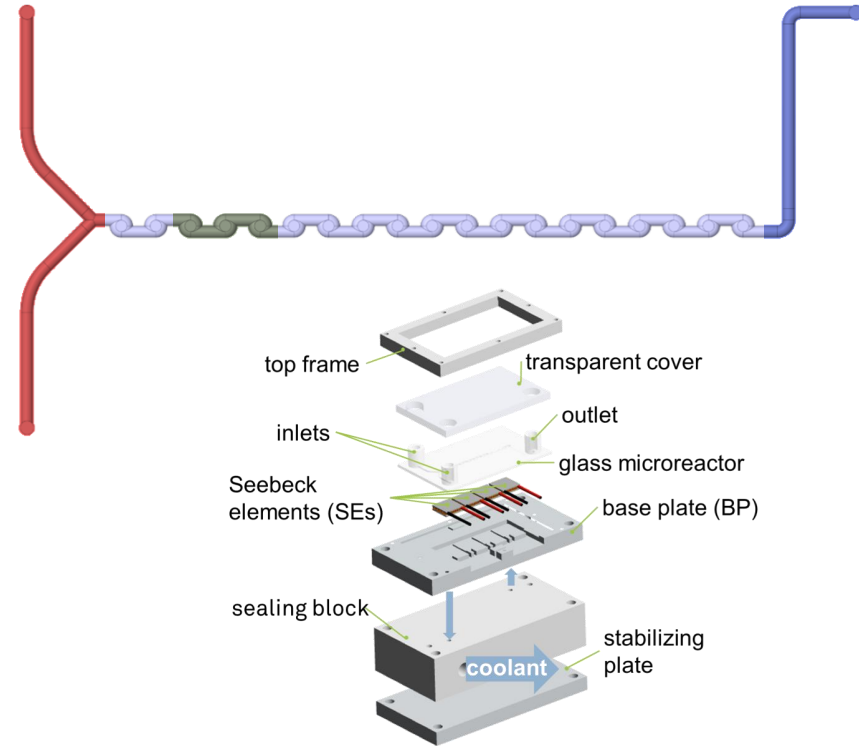
All work supervised by T. A. Frede, PhD-Student at Laboratory of Equipment Design:
[1] H. Köster, Master Thesis (2021), TU Dortmund
[2] N. Link, Master Thesis (2021), TU Dortmund
[3] M. Dietz, Master Thesis (2021), TU Dortmund
[4] N. Nickbin, Master Thesis (2022), TU Dortmund

top frame
transparent cover
inlets
outlet
glass microreactor
Seebeck elements (SEs)
base plate (BP)
sealing block
coolant
stabilizing plate

Motivation - Data Structure

- FAIR-Data storage
  - Findable
  - Accessible
  - Interoperable
  - Reuseable

- Ontologies
  - Knowledge Graph
  - Classifier
  - Expandable
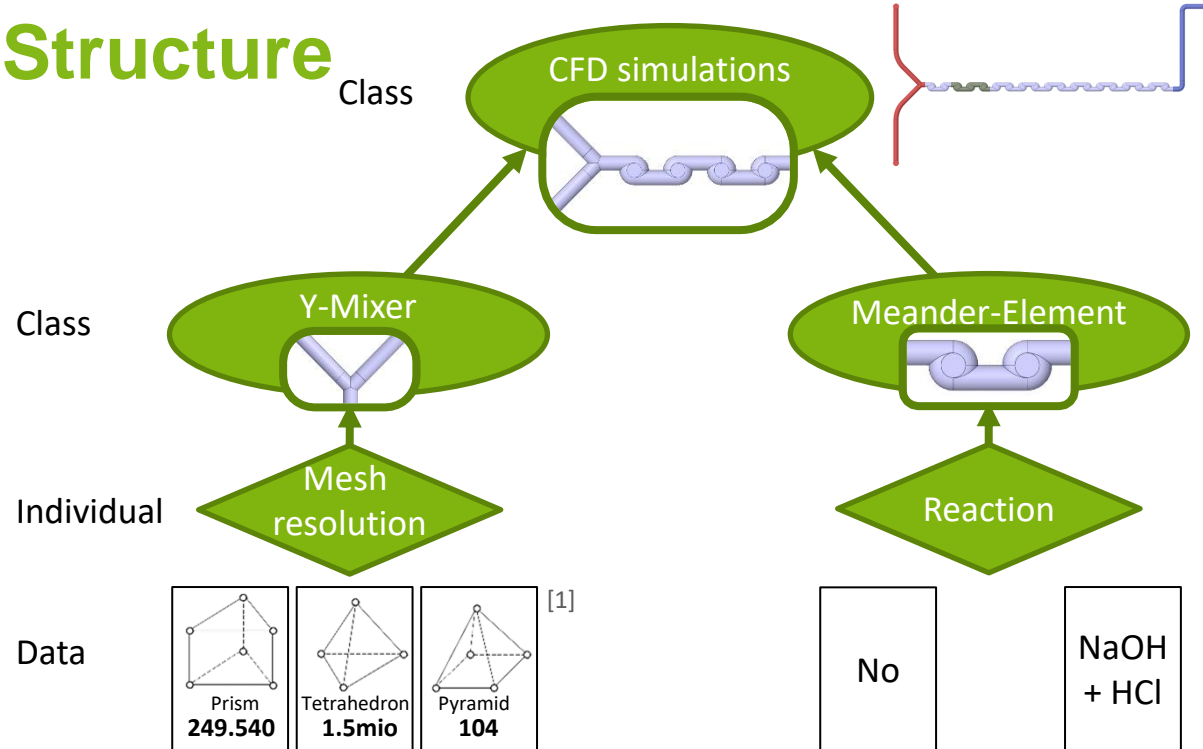
[1] Khaydarov, Valentin (2018) Dissertation. TU Dresden.
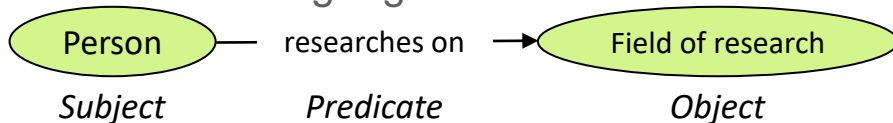
Class — CFD simulations

Class — Y-Mixer — Meander-Element

Individual — Mesh resolution — Reaction

Data —
Prism **249.540** | Tetrahedron **1.5mio** | Pyramid **104** [1]
No | NaOH + HCl

Ontology

- RDF/XML-Language

Subject	Predicate	Object
Person	researches on	Field of research

Meander-Element	isPartOf	CFD-Simulation
MeshResolutionIndividual1	hasPrisms	104

- ➢ Classes, Individuals, Object-, Data-, Annotation- properties
- Hierarchical

Class

CFD-simulations

isSuperclass isSubclass

isPartOf

Class

Meander-Element

isMeshResolution Individualof

Individual

MeshResolution Individual1

Data

hasPrisms 104

Ontology

- RDF/XML-Language

Person	researches on →	Field of research
Subject	*Predicate*	*Object*
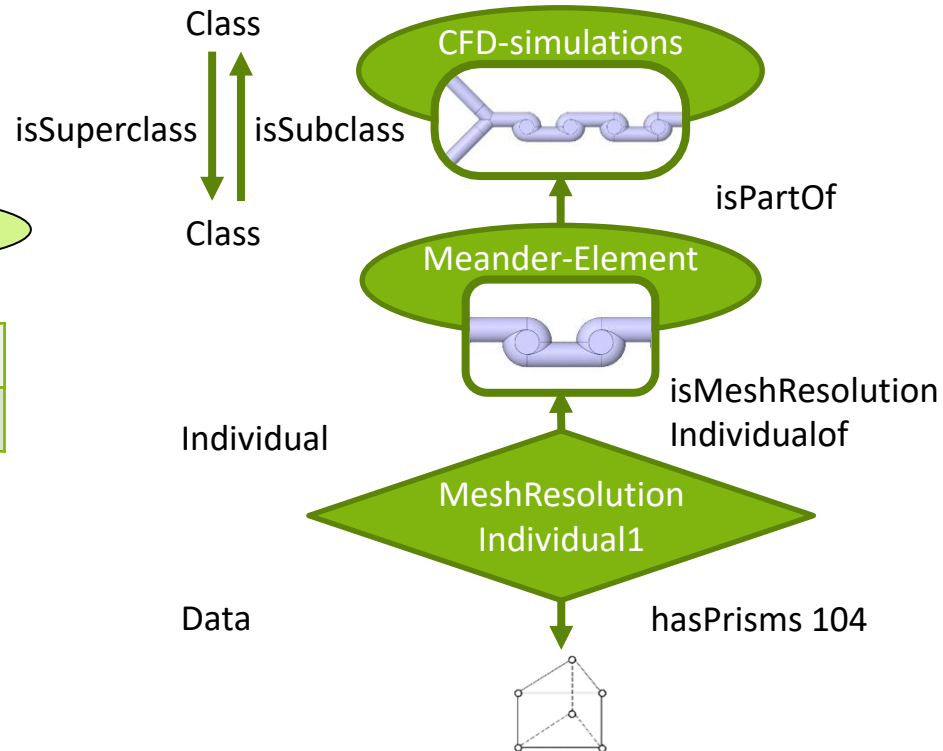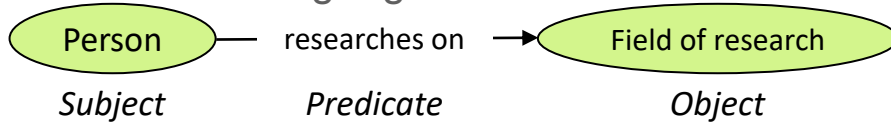
Meander-Element	isPartOf	CFD-Simulation
MeshResolutionIndividual1	hasPrisms	104

- ➤ Classes, Individuals, Object-, Data-, Annotation- properties
- Hierarchical
- Reasoning / inferring its structure

Class

isSuperclass ⇅ isSubclass

Class

CFD-simulations

isPartOf

Meander-Element

isMeshResolutionIndividualof

Individual

Mesh resolution individual1

Data

hasPrisms 104

Department of Biochemical and Chemical Engineering

equipment design
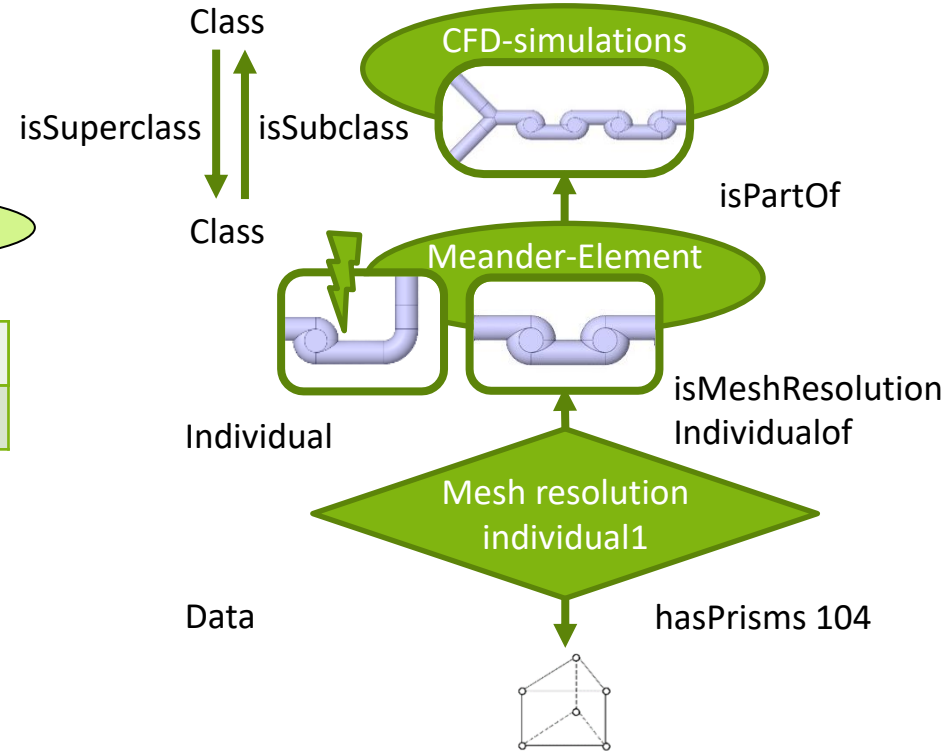
Process Steps

- Summary file as data source
  - Non-machine readable data
  - Complex syntax

- Aim: 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, 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
...
```

Ansys simulation log-file

Data Converter

Headers	→	Subheaders	→	Special Segments	→	Regular Segments

- Transfer into accessible file
- Flexible and automated
- Only scripted where needed

- Implemented & running
  - Programmed for 128 files
  - Verified on 783 files
- FAIR data conversion

```
...
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
    END
    RUN DEFINITION:
        Solver Input File = Fluid Flow CFX.def
        Run Mode = Full
        Solver Results File = \
/work/smknikb/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
...
```

- CFX Command Language for Run

```
"DOMAIN": {
  "Default Domain": {
    "Coord Frame": "Coord 0",
    "Domain Type": "Fluid",
    "Location": "B156, B178, B29, B76",
    "BOUNDARY": {
      "Default Domain Default": {
        "Boundary Type": "WALL",
        "Location": "F100.156,F167.178,…",
        "Use Profile Data": "False",
        "BOUNDARY CONDITIONS": {
          "ADDITIONAL VARIABLE": {
            "Age": {
              "Option": "Zero Flux"}},
          "MASS AND MOMENTUM": {
            "Option": "No Slip Wall"}}},
      "Inlet": …
```

- Partitioning
  - Initial Memory Allocation
    - Text
      ```
      "text": {
        "Run mode": "partitioning run",
        "Host computer:  cstd01-016 (PID:16311)": [],
        "Job started:    Tue Feb  1 09:52:31 2022": []
        }
      ```
    - Box
      ```
      "box": {
        The MeTiS partitioning method allocates additional memory. \n
        Total memory usage will therefore exceed the values shown above.
        }
      ```
    - Table
      ```
      "table": {
        "":        ["Real", "Integer", "Character", "Logical", "Double"],
        "Mwords": ["8.40", "101.20", "22.92", "0.36", "2.70"],
        "Mbytes": ["32.05", "386.04", "21.86", "1.37", "20.61"]
        }
      ```
  - Mesh Statistics
  ...
- Solver
  - Convergence
   -History
      ```
      "6": {
        "OUTER LOOP ITERATION": "6",
        "CPU SECONDS": "3.578E+02",
        "Equation": ["U-Mom", "V-Mom", "W-Mom", "P-Mass", "Age", "Water-Mass Fraction"],
        "Rate": ["0.59", "0.58", "0.57", "0.78", "0.83", "0.99"],
      ```
....

Department of Biochemical and Chemical Engineering

equipment design

Ontology

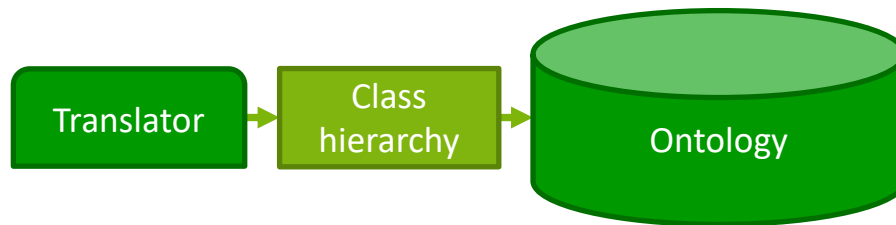Ansys simulation files → Data extraction → JSON data files

- VIMMP
  - Domain of simulation
  - x Reasoner infers as inconsistent

- EMMO
  - Material modelling
  - Conceptualized

Ontology

- VIMMP
  - Domain of simulation
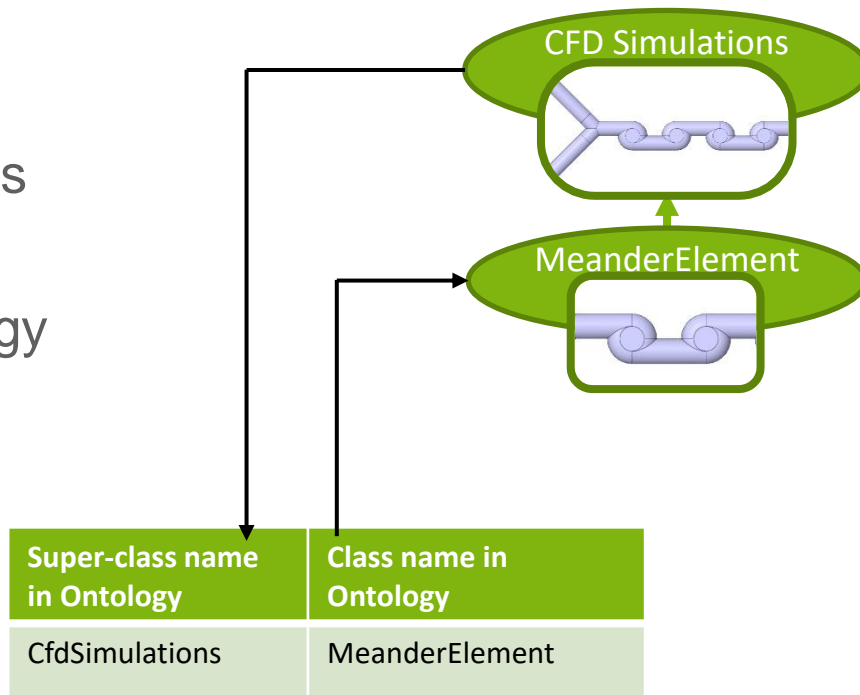 - x Reasoner infers as inconsistent

- EMMO
  - Material modelling
  - Conceptualized

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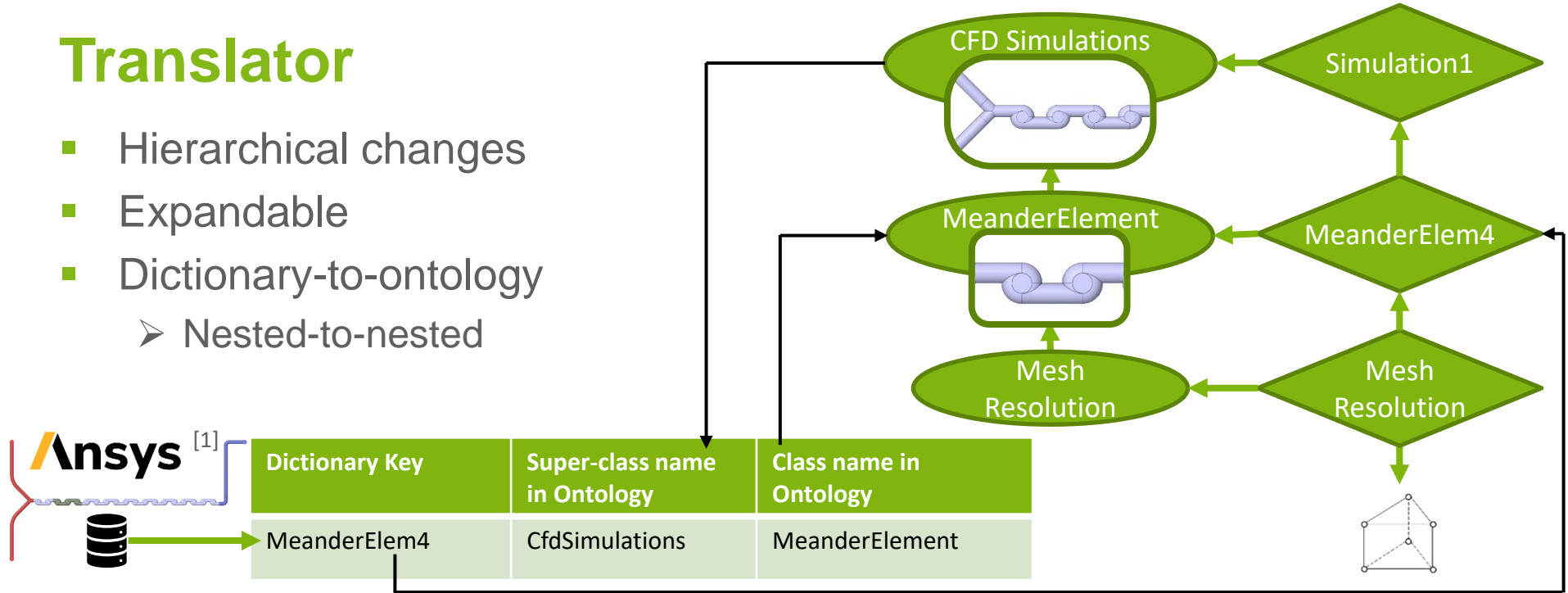 - Individual assignment

Translator

- Hierarchical changes
- Expandable
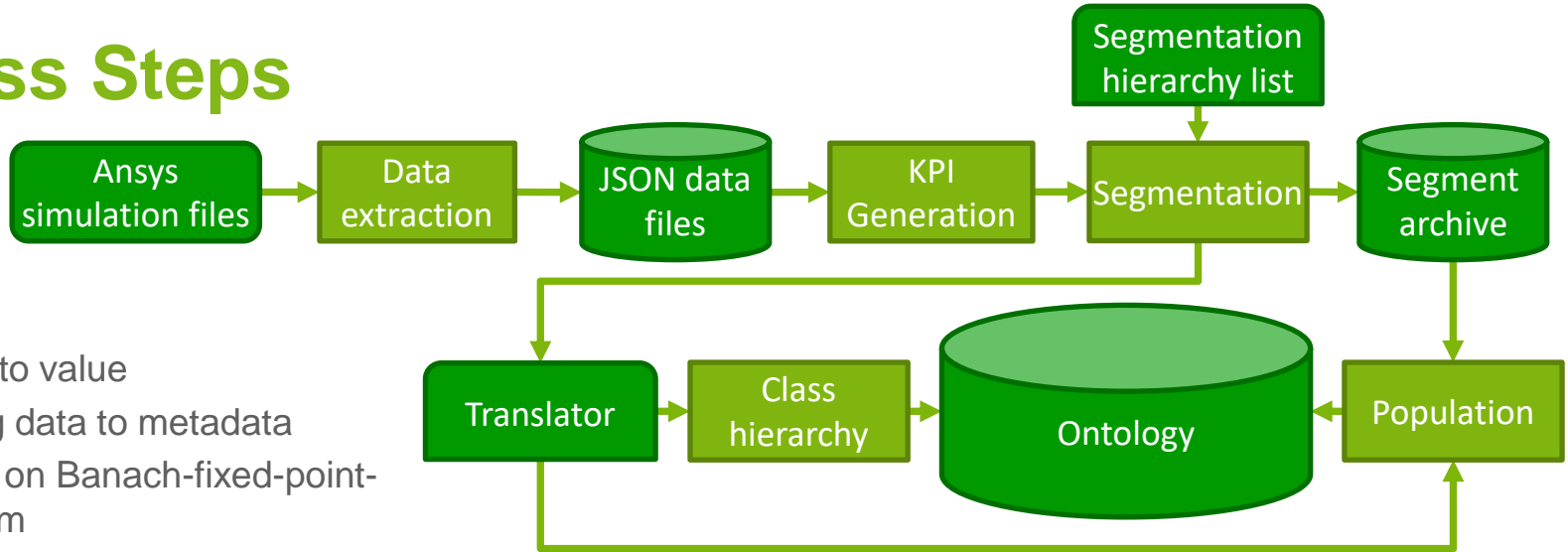- Dictionary-to-ontology
  - Nested-to-nested

Super-class name in Ontology	Class name in Ontology
CfdSimulations	MeanderElement

Department of Biochemical and Chemical Engineering
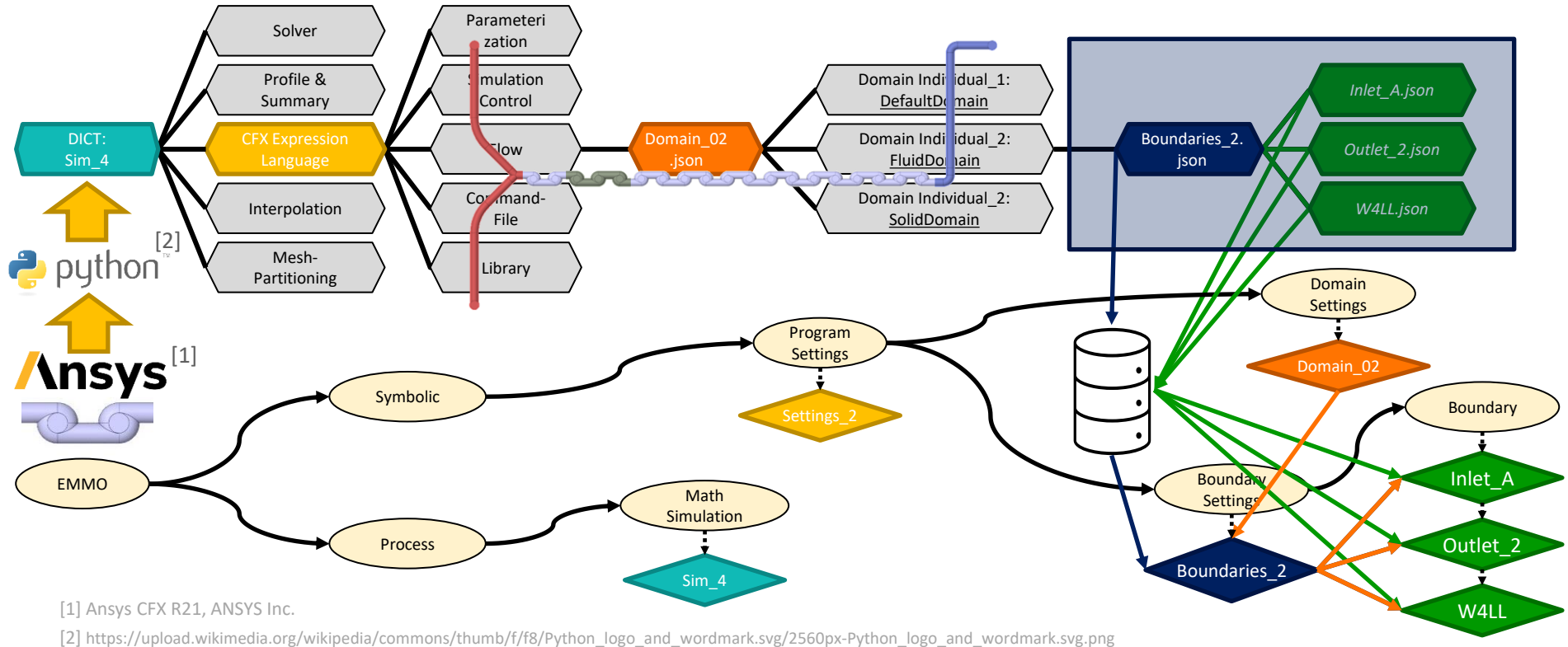
equipment design

Process Steps

- **KPIs**
  - Curve to value
  - Adding data to metadata
  - Based on Banach-fixed-point-theorem

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

- **Population of ontology**
  - Automated
  - Adaptable

[1] Ansys CFX R21, ANSYS Inc.

[2] https://upload.wikimedia.org/wikipedia/commons/thumb/f/f8/Python_logo_and_wordmark.svg/2560px-Python_logo_and_wordmark.svg.png

Metrics/Program Performance

- Code stable
  - Test set: 128 simulations
  - Validation set: 783 simulations

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

Program	Performance for 911 Simulations	
Data Extraction	69.23 s	8.8M lines
Segmentation	3 min 57 s	3.278 segments
Population	15 min 16 s	

Department of Biochemical and Chemical Engineering

equipment design

Metrics/Program Performance

- Code stable
  - Test set: 128 simulations
  - Validation set: 783 simulations

- Population code
  - Least optimized
  - Data condensation

- SPARQL
  - Validation of segmentation
  - Metadata and data

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

Program	Performance for 911 Simulations	
Data Extraction	69.23 s	8.8M lines
Segmentation	3 min 57 s	3.278 segments
Population	15 min 16 s	

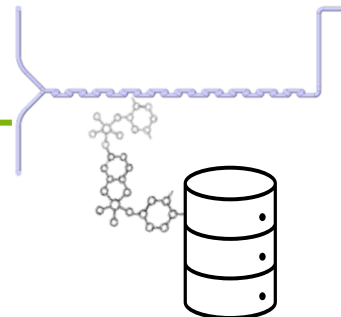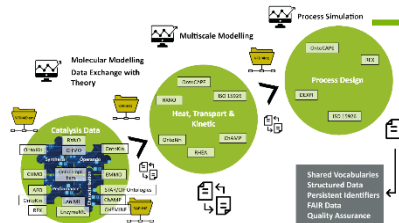| Ontology	EMMO	EMMO + 911 Simulations	
Classes	470	788	Δ 318
Object Prop.	47	190	Δ 143
Data Prop.	3	244	Δ 241
Individuals	1	28,818	Δ 28,817
Axioms	3,363	756,719	Δ 753,356

Process Steps
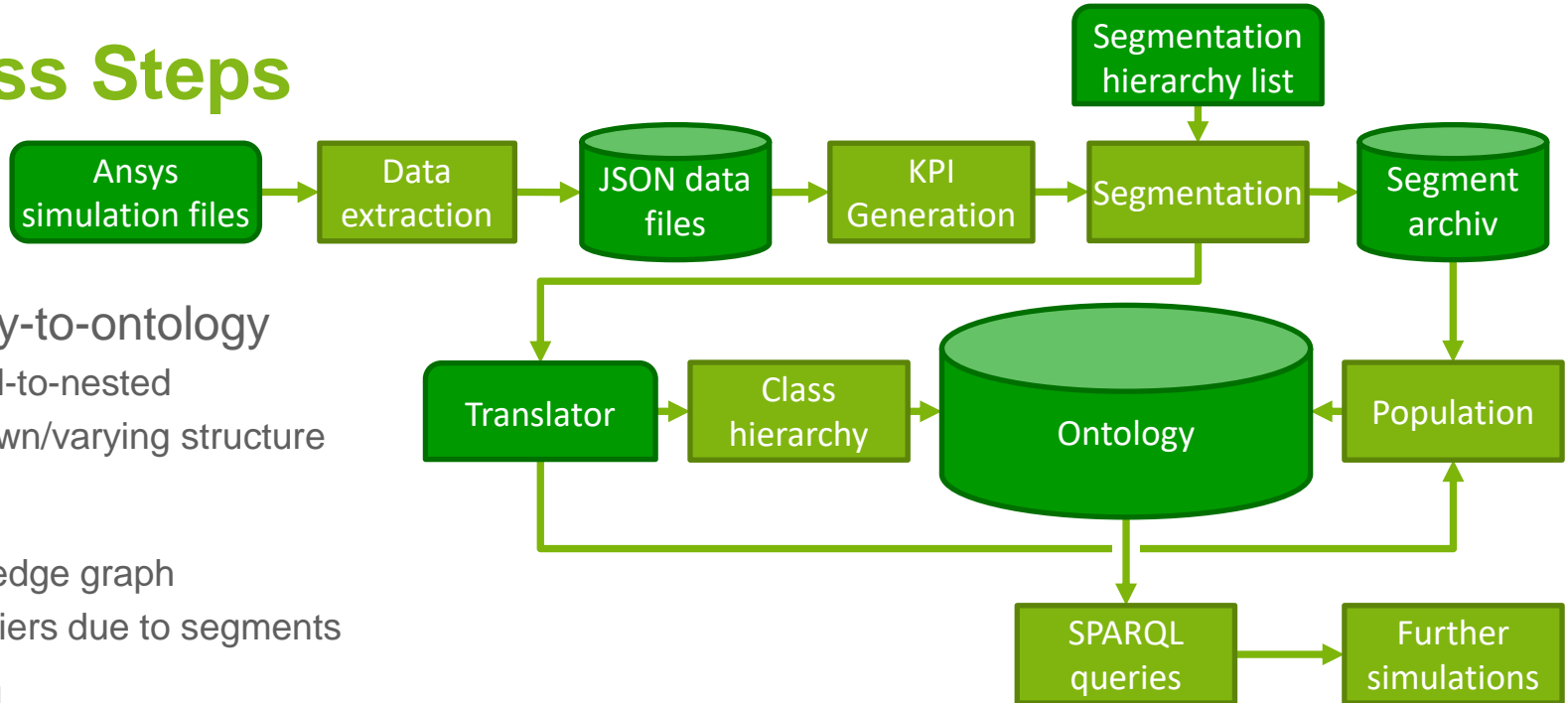
- **Dictionary-to-ontology**
  - Nested-to-nested
  - Unknown/varying structure

- **Ontology**
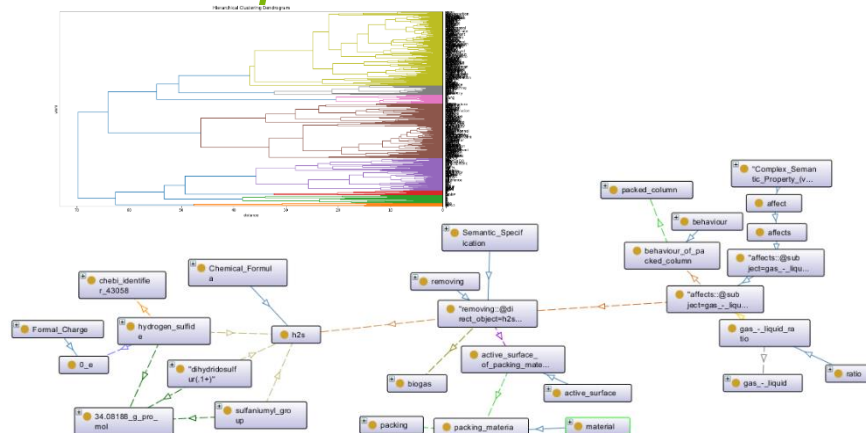 - Knowledge graph
  - Classifiers due to segments

- **Validated**

Process flow: Ansys simulation files → Data extraction → JSON data files → KPI Generation → Segmentation → Segment archiv. Segmentation hierarchy list → Segmentation. JSON data files → Translator → Class hierarchy → Ontology → Population. Segment archiv → Population. Ontology → SPARQL queries → Further simulations → Population.

Conclusion

Outlook

Conclusion

Preferred Label	Alternate Labels	Definition	Children
Biologic Catalysis			bio catalyst preparation, protein purification, bio catalyst application
Bio Catalyst Preparation		The Catalyst Preparation is a process that is intended to synthesize a catalyst from several components.	cultivation, induction, cell harvesting, cell disruption, gel filtration, protein concentration, transformation
Expression Strain preparation	Strain preparation		
Expression host	expression host, organism, microorganism, bacteria, yeast		
...	...	...	...

Outlook

- Intensify text mining
  - Enhance sentence to RDF triplets pipeline
  - One (useful) RDF graph per text corpus
  - Reduce the „fuzzyness" of resulting graphs
  - Integration to existing ontologies
- Mapping of ontologies
  - Graph theory / Graph learning
- Alignment of RDF triplets from text with top-level ontology
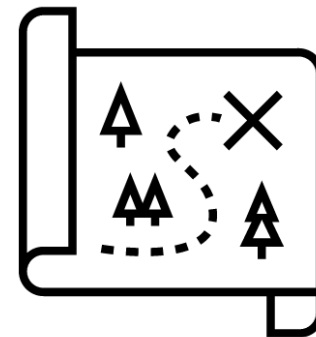
Thank you for your attention!
Questions?

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

www.ad.bci.tu-dortmund.de