



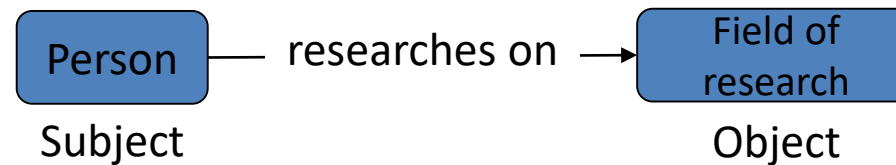
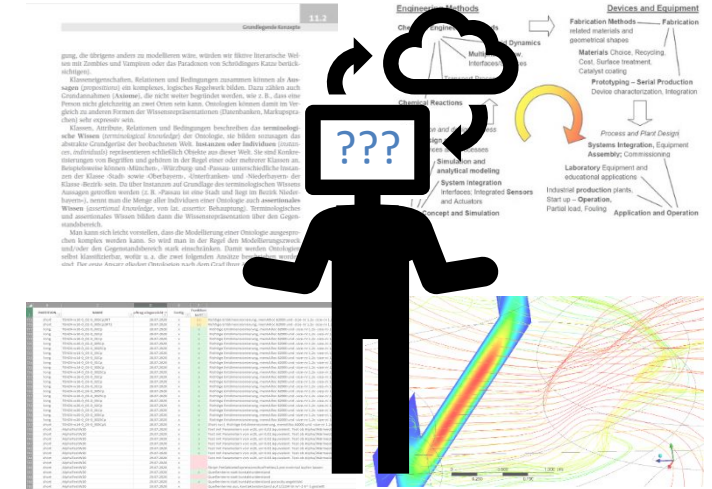
# Introduction to Ontologies – Workflow and Tools in Catalysis

A.S. Behr, N. Kockmann

JCDL - 23.06.2022

# Ontologies – How can we use them?

- Ontologies consist of a network of information with logical relations
- Interconnect (meta) data
- Different data types readable for humans
- Aim: Machine- and human-readable (meta) data
- Information in triplets



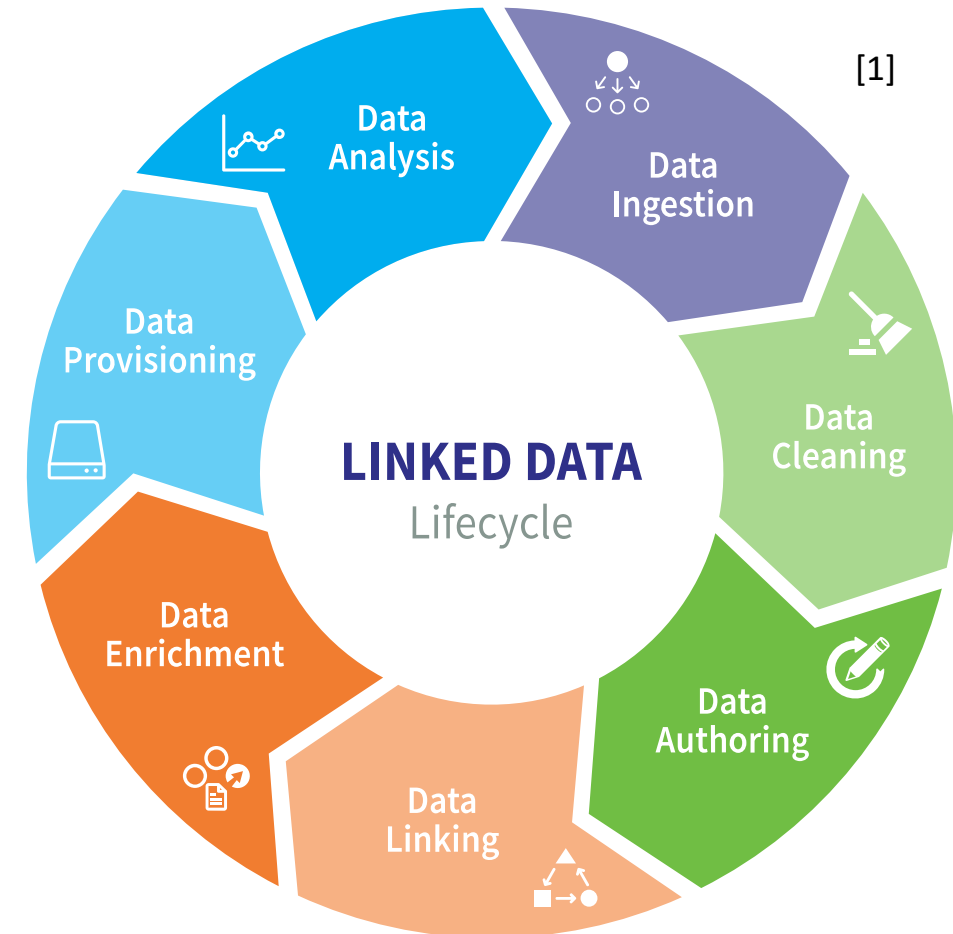
➔ Unified data formats through *ontologies* and standardized *metadata schemes*

➔ FAIR data principles: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable

# Ontologies – Why should you use them?

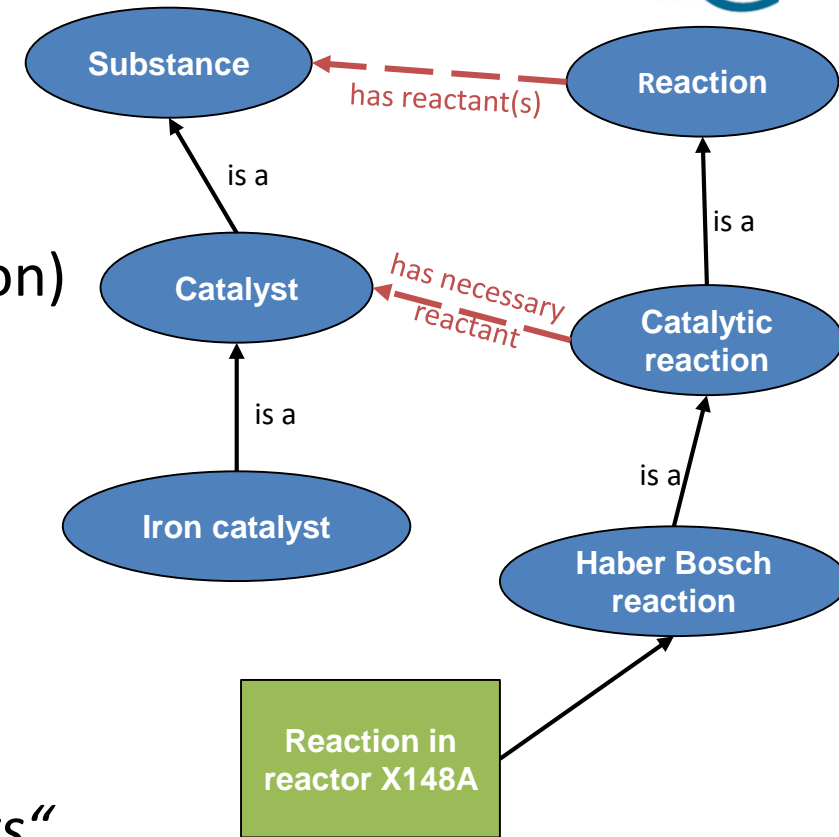
- Describe existing knowledge
- Derive knowledge based on described knowledge
- Annotate data
- Query data
- Normalize data
- Interconnect data

 Gain new insights from your data!

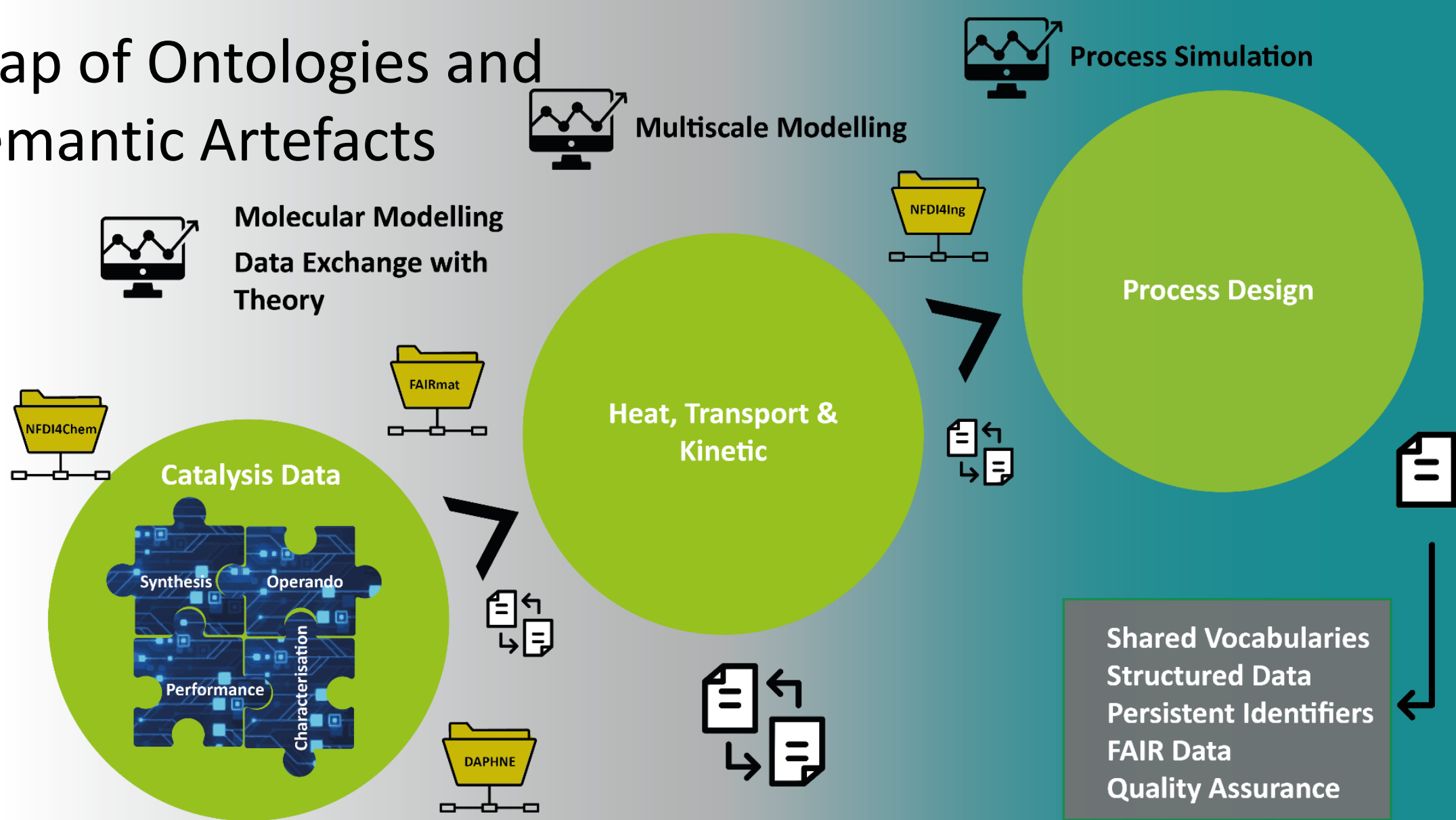


# Ontologies – A simple example

- 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 the 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.“

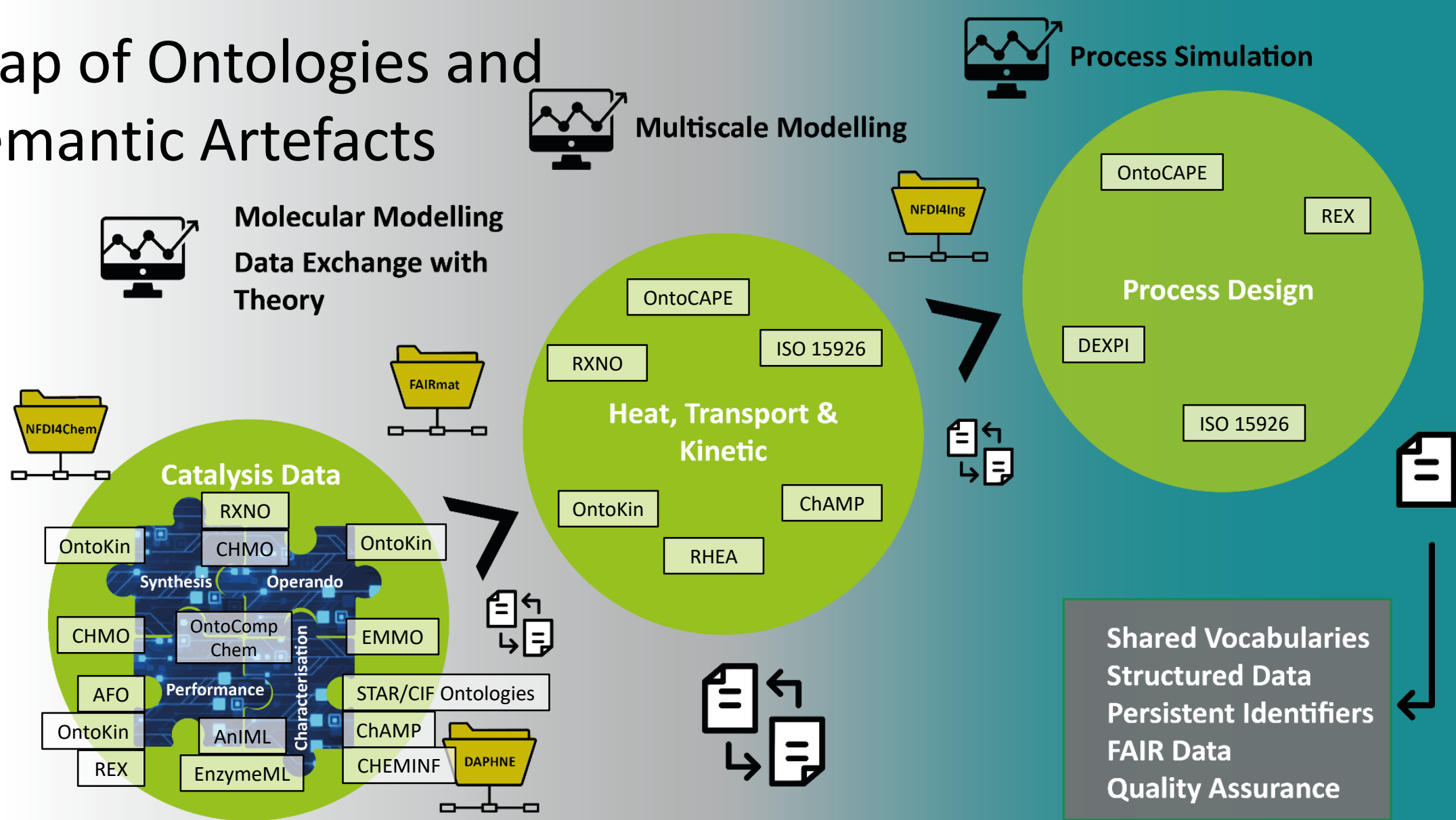


# Map of Ontologies and Semantic Artefacts



See also: [nfdi4cat.org/ontology-collection/](https://nfdi4cat.org/ontology-collection/)

# Map of Ontologies and Semantic Artefacts



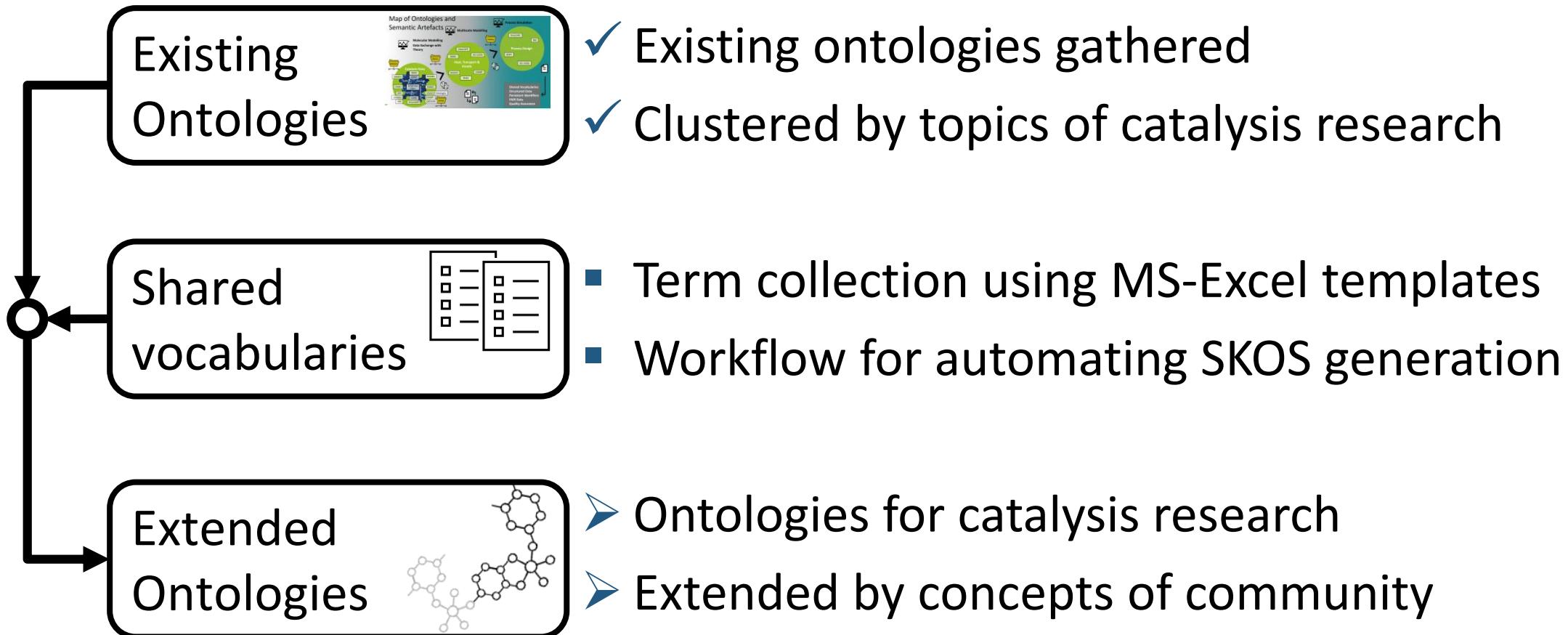
See also: [nfdi4cat.org/ontology-collection/](https://nfdi4cat.org/ontology-collection/)

# Map of Ontologies and Semantic Artefacts



See also: [nfdi4cat.org/ontology-collection/](https://nfdi4cat.org/ontology-collection/)

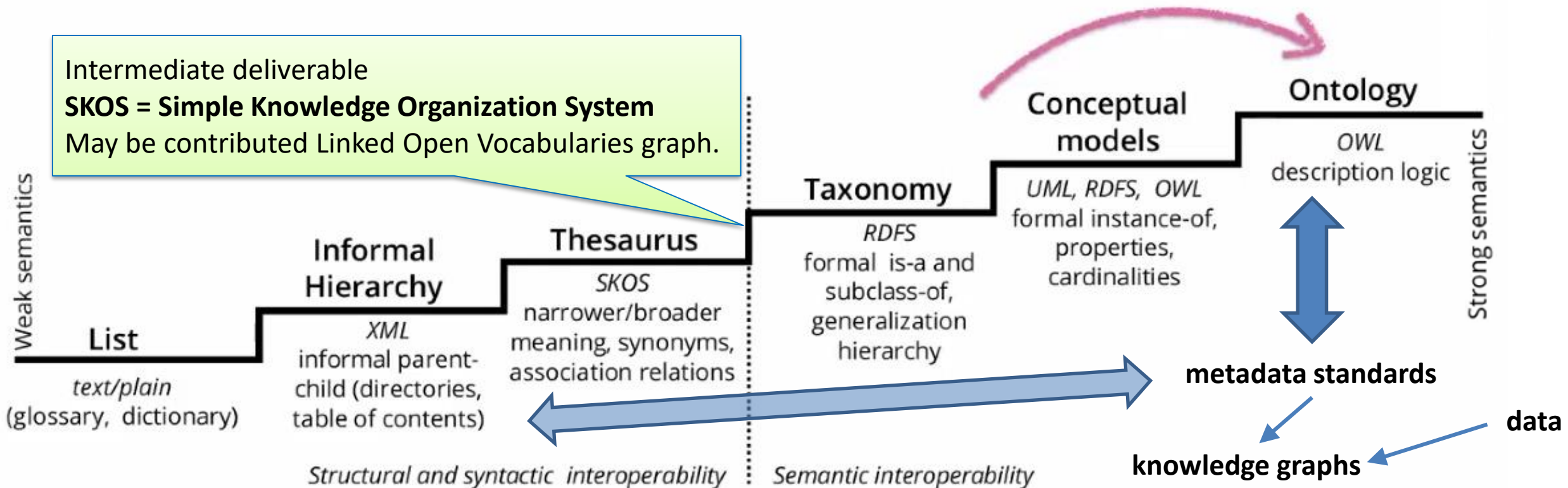
# Overall workflow of NFDI4Cat – TA1





# The way to ontologies

- Many steps with domain and IT knowledge needed
- Intermediate goal: Setup catalysis research specific SKOS <sup>[1], [2]</sup>

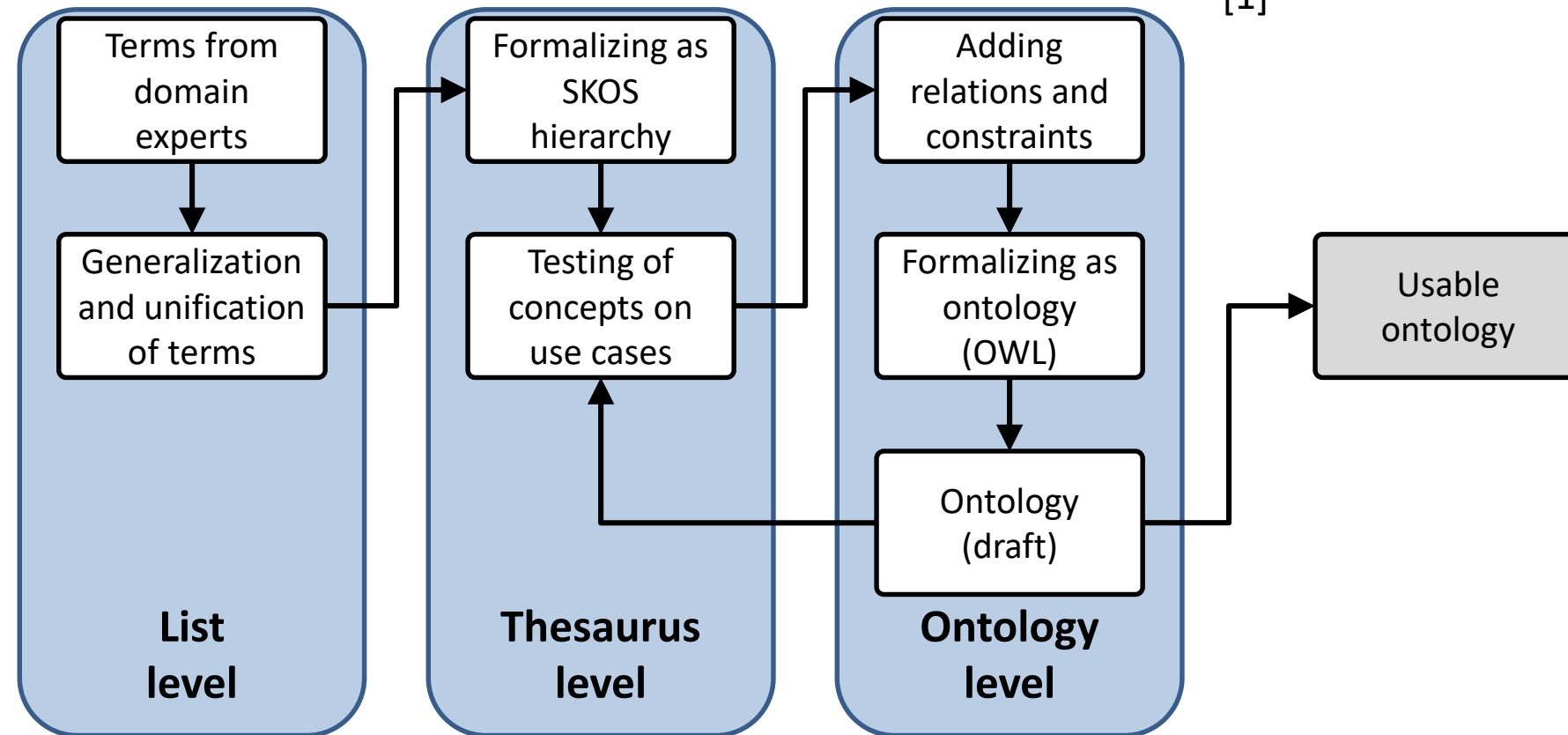


[1] adapted from [https://th.fhi-berlin.mpg.de/meetings/fairdi2020/index.php?n=Meeting.PosterDetails&poster\\_id=18](https://th.fhi-berlin.mpg.de/meetings/fairdi2020/index.php?n=Meeting.PosterDetails&poster_id=18)

[2] SKOS-Reference, W3C, <https://www.w3.org/TR/2009/REC-skos-reference-20090818/>, accessed 20.06.2022

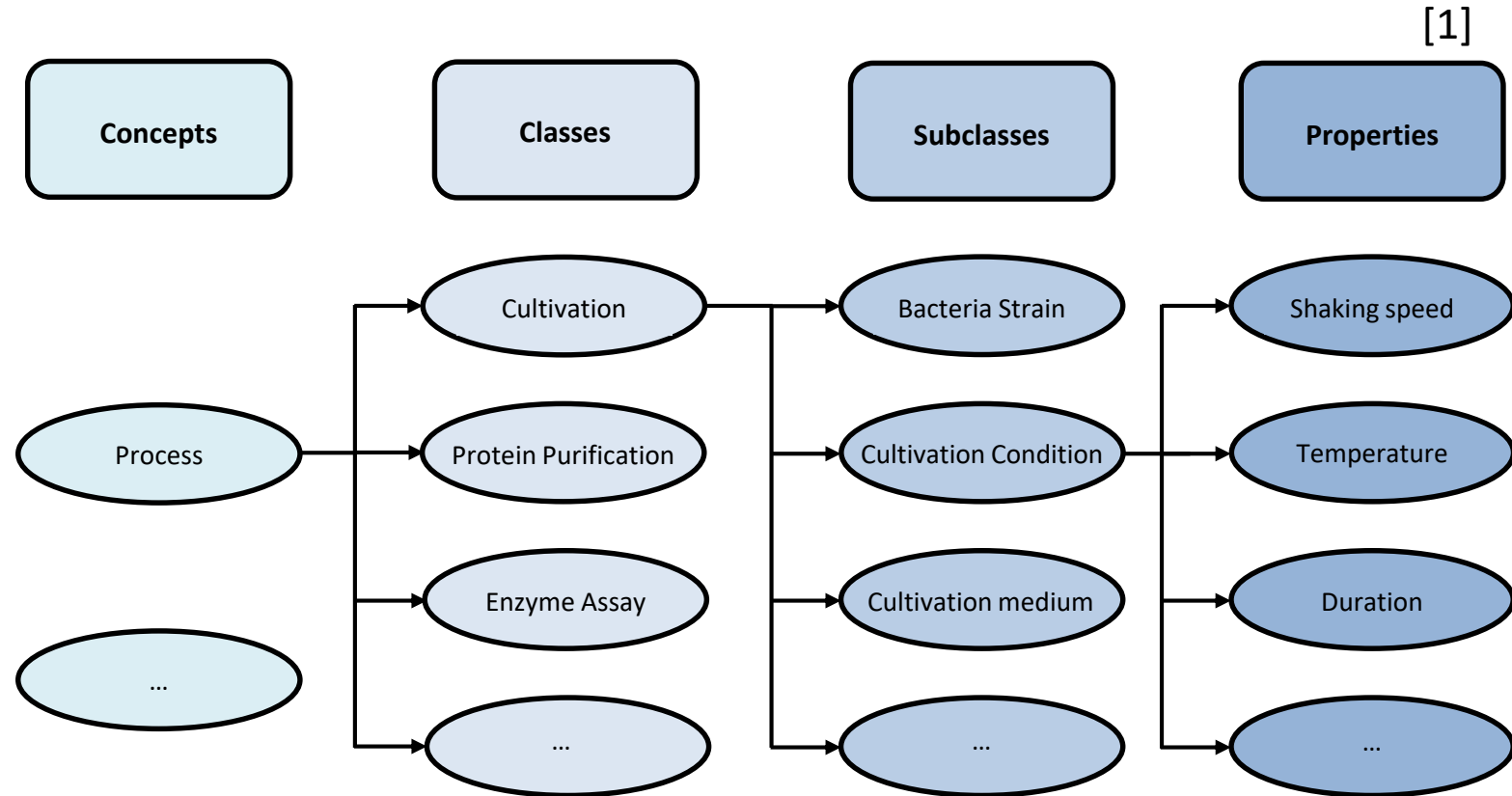
# From List to Ontology Level

- General workflow
- Recursive process
- Formalizing using SKOS-hierarchies of classes as intermediate goal



# List Development – Example on Biocatalysis

- Starting with collection of different processes as concepts
- Different process steps as classes
- Concepts related to classes gathered as subclasses
- Each subclass contains set of properties



# Current workflow – Taxonomy generation

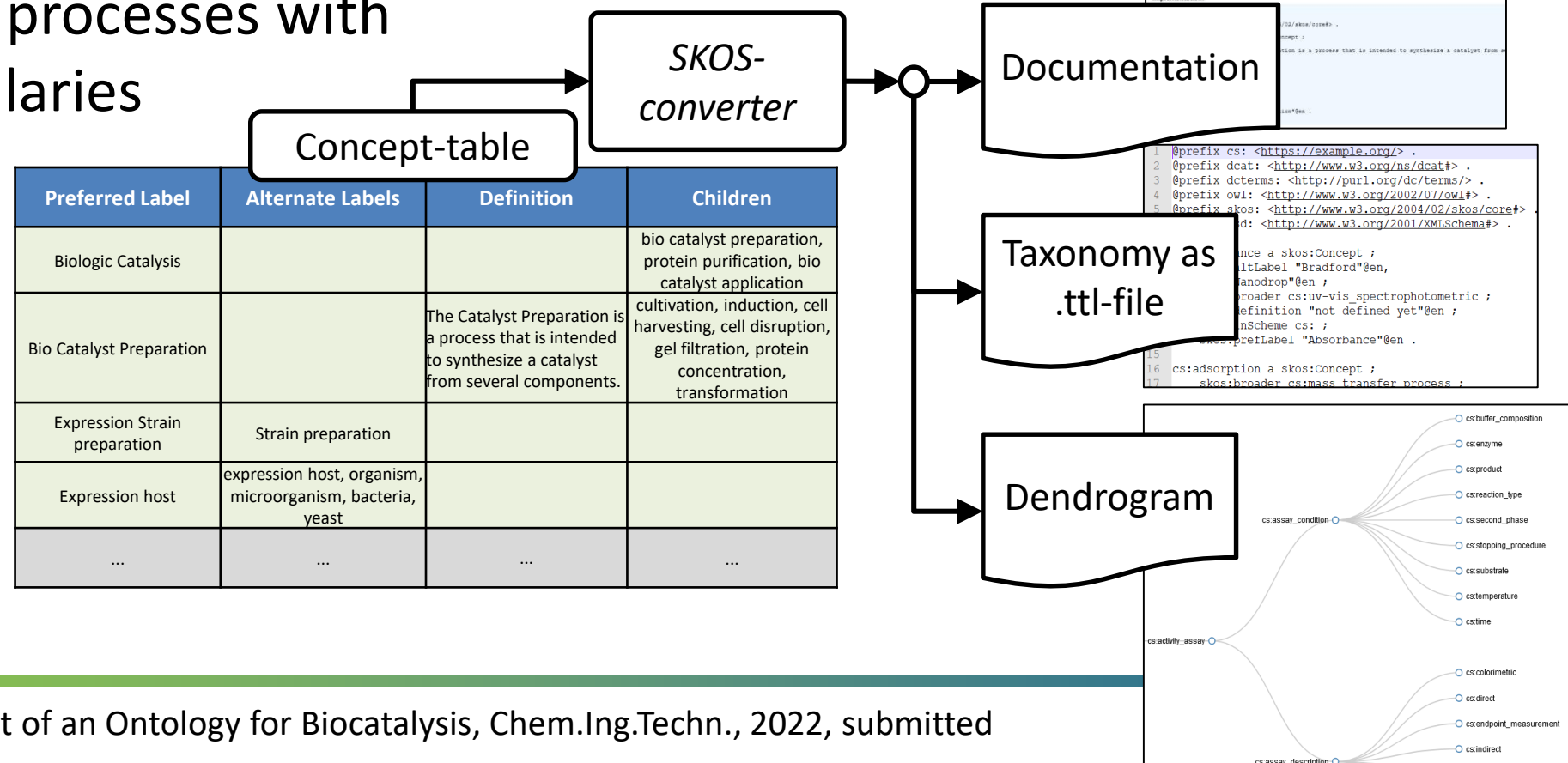
- Gathering concepts important to knowledge domains and processes in concept tables (easy access for domain experts)
  - SKOS-files (Taxonomies) on different domains as 1<sup>st</sup> milestone

1	Vocabulary URI	<a href="https://example.org/">https://example.org/</a>	The URI for the vocabulary	required: a URI	
2	Title	vocabulary for processes	The title of the vocabulary as a whole	required: text	
3	Description	A 'process' is defined as a temporal part of a 'physical' that is categorized in a primitive process subclass according to what type of process we want to represent. A 'process' is always a 'physical', since a 'void' does not have elements that evolves in time." [EMMO]	A general description for the vocabulary as a whole	required: text. Can have paragraph breaks	
4	Created		When was this vocabulary first created?	required: date (yyyy-mm-dd)	
5	Modified	2021-11-19	When was this vocabulary last modified?	optional: date (yyyy-mm-dd)	
6	Creator	NFDI4Cat	Defaults to today.		
7	Publisher	NFDI4Cat	A organisation	required	
8	Version		A organisation	required	
9	Provenance	NFDI4Cat - TA1	1. A version number for this vocabulary, e.g. 1.1	optional: text	
10	Custodian	TA1 subgroup homogeneous catalysis and biocatalysis	A note on what the source of this vocabulary is	optional: text	
11	Catalogue PID	TA1 subgroup homogeneous catalysis and biocatalysis	The person managing this vocabulary's content	optional	
12			A catalogue PID or DOI, e.g. eCat ID., if the vocab has one	optional	
13					
14			See notes sheet for an explanation of the following headings		
15	Concept URI	Preferred Label	Alternate Labels	Definition	Children
20		Plasmid			antibiotic resistance, vector, gene
21		Antibiotic Resistance			
22		Vector	Backbone		
23		Transformation	biotransformation		expression host, plasmid, chemical transformation,
24		Chemical transformation	would change transformations into 'low level'-categories.		
25		Electroporation			
26		Cultivation	Incubation		bacteria strain, cultivation condition, cultivation medium
27		Bacteria strain			
28		Cultivation condition			shaking speed, temperature, shaking speed, temperature, length of time, optical density
29		Induction condition	should we add the cultivation volume?		
30		Shaking speed			
31		Temperature			
32		Length of time	Duration, cultivation time		
33		Optical density	OD600, biomass concentration, cell concentration, cell dry weight		
34		Volume			
35		Cultivation medium			autoinduction, non-autoinduction

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

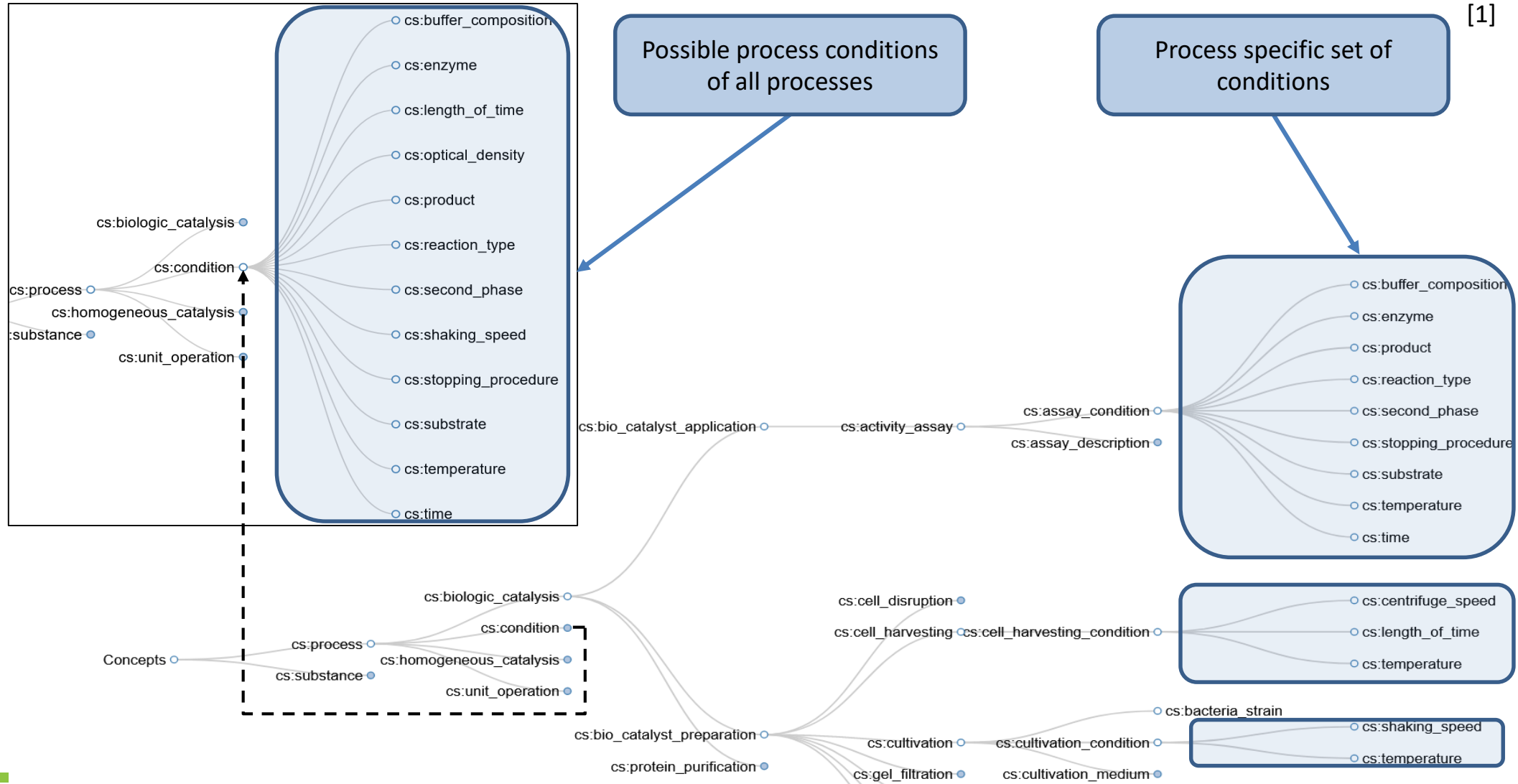
# Shared vocabulary: Concept tables

- Gather important concepts in tables
- Convert vocabulary to machine-readable taxonomy-files
- Description of processes with shared vocabularies
- Description of data with shared vocabularies

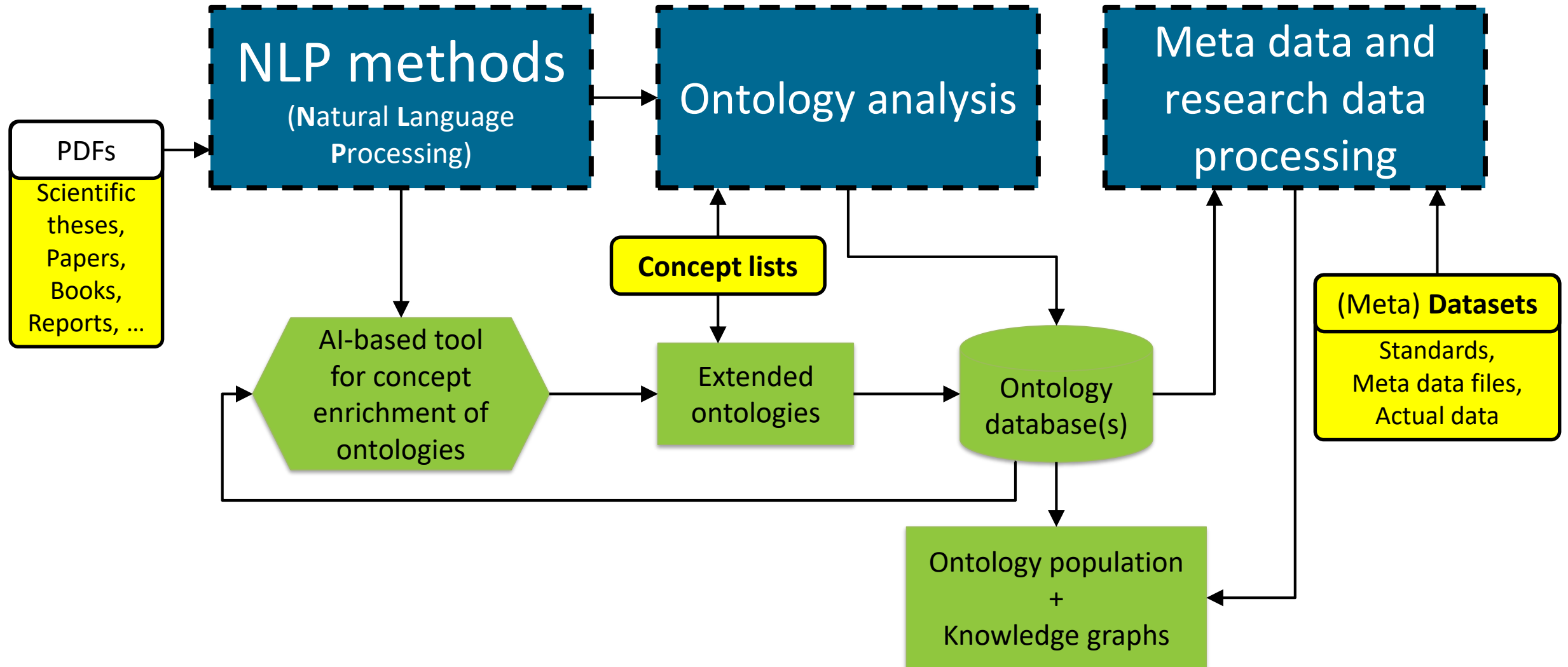


# Class Hierarchy – Example on Biocatalysis

[1]



# Proposed Workflow on Ontologies4Cat



# Wrap up

- Full potential of data only possible with linked data
- Ontologies needed for catalysis domain
- Shared vocabularies (concept lists) are first step for ontology development
- Further steps are necessary on
  - Concepts relevant for catalysis research + applications
  - Data and metadata for annotation + UseCases
  - Already applied data standards of research





**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](mailto:sara.espinoza@dechema.de)



<http://www.nfdi4cat.org/>



[info@nfdi4cat.org](mailto:info@nfdi4cat.org)



[/linkedin.com/company/nfdi4cat](https://www.linkedin.com/company/nfdi4cat)



[@NFDI4Cat](https://twitter.com/NFDI4Cat)

**Alexander Behr, Norbert Kockmann**

**JCDL – 23.06.2022**