NFDI4Cat

NFDI for Catalysis-Related Sciences

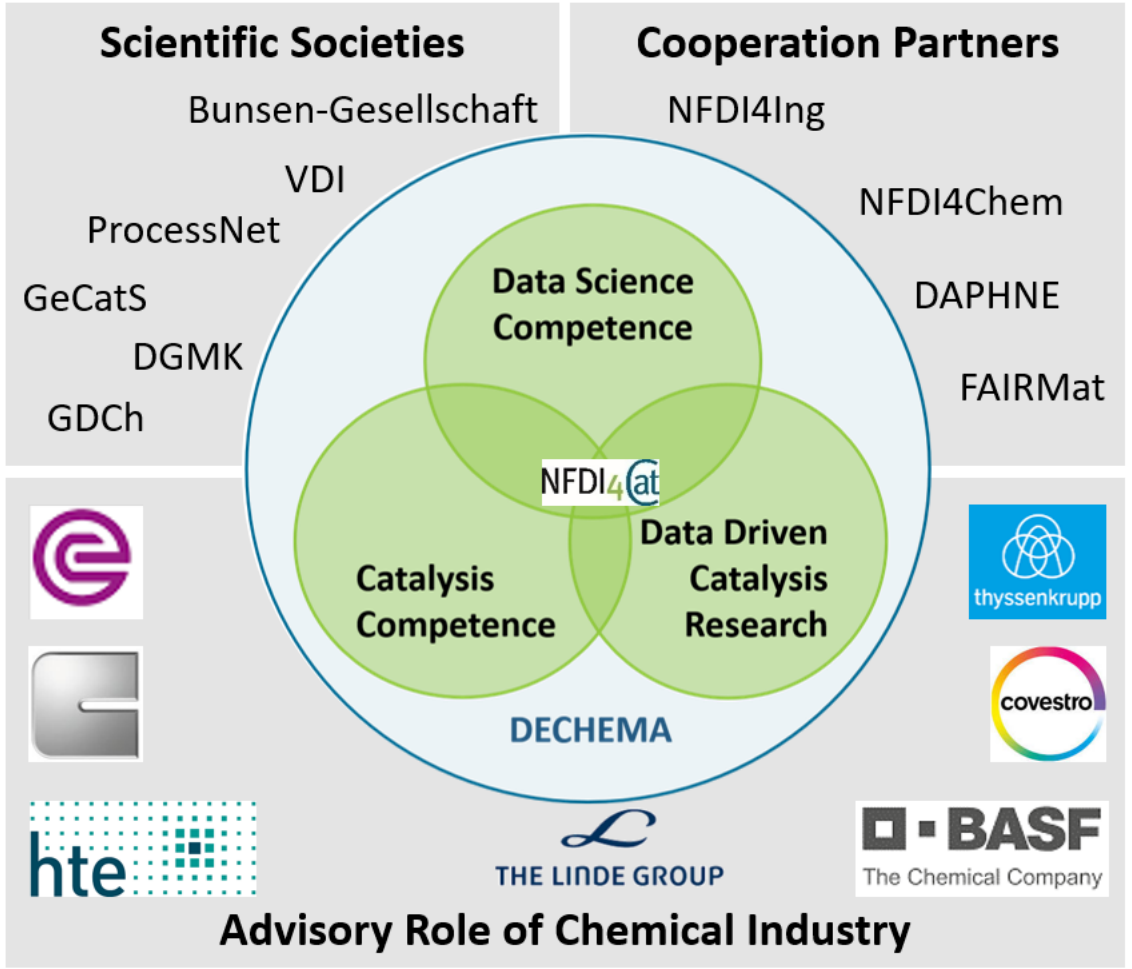
Catalysis Research Data: Structures, Workflows, and Repositories
Results TA1 – TA4

S. Schimmler, N. Kockmann, O. Deutschmann, R. Krähnert, T. Bönisch

Frankfurt, 17.11.2021

DOI: 10.5281/zenodo.5882747

The NFDI4Cat Consortium



DOI: 10.5281/zenodo.5882747

The NFDI4Cat Consortium Governance







Strategic Discussion, Validation, & Evaluation

Steering Board

Budget Planning
Representation
Strategic Decisions
Coordination

M. Beller
M. Resch
A. Förster
O. Deutschmann

Metadata & Ontology Development Board
M. Özaslan

Data Standards & Quality Assurance Board
S. Schimmler

Strategic Decisions, Budget allocation, Integration of new Partners

Project Management Board
Spokesperson & Co-Spokespersons (1 Representative per Organization

Support, Representation & Development of TAs

Project Management

Steering group of the Task Areas (TA)
Heads of TAs & One representative per organization

Advisory Board

Industry representatives

S. Schunk

Definition of Tasks
Project Work
Documentation
Analysis

Project Partners

Scientific institutions and societies that contribute key projects, infrastructure, & scientific results

DOI: 10.5281/zenodo.5882747

Core Development Topics of NFDI4Cat

- **TA1: Ontology Development and Metadata Standards**
- **TA2: Data Standards, Data Collection, Interfaces**
- **TA3: Data Analysis, Quality Management and Re-Use**
- **TA4: Linked Extensible Infrastructure and Access Management**
- TA5: Dissemination and Outreach/ Training
- TA6: Networking with NFDIs, SFBs and International
- TA7: Intellectual Property and Confidentiality, Licences and Reward models
- TA8: Management

Data & Meta Data Standards

Data Science & Information Infrastructure Design

Community & User-related Aspects

DOI: 10.5281/zenodo.5882747

Task Area 4

Fraunhofer FOKUS

TA4: Linked Extensible Infrastructure and Access Management

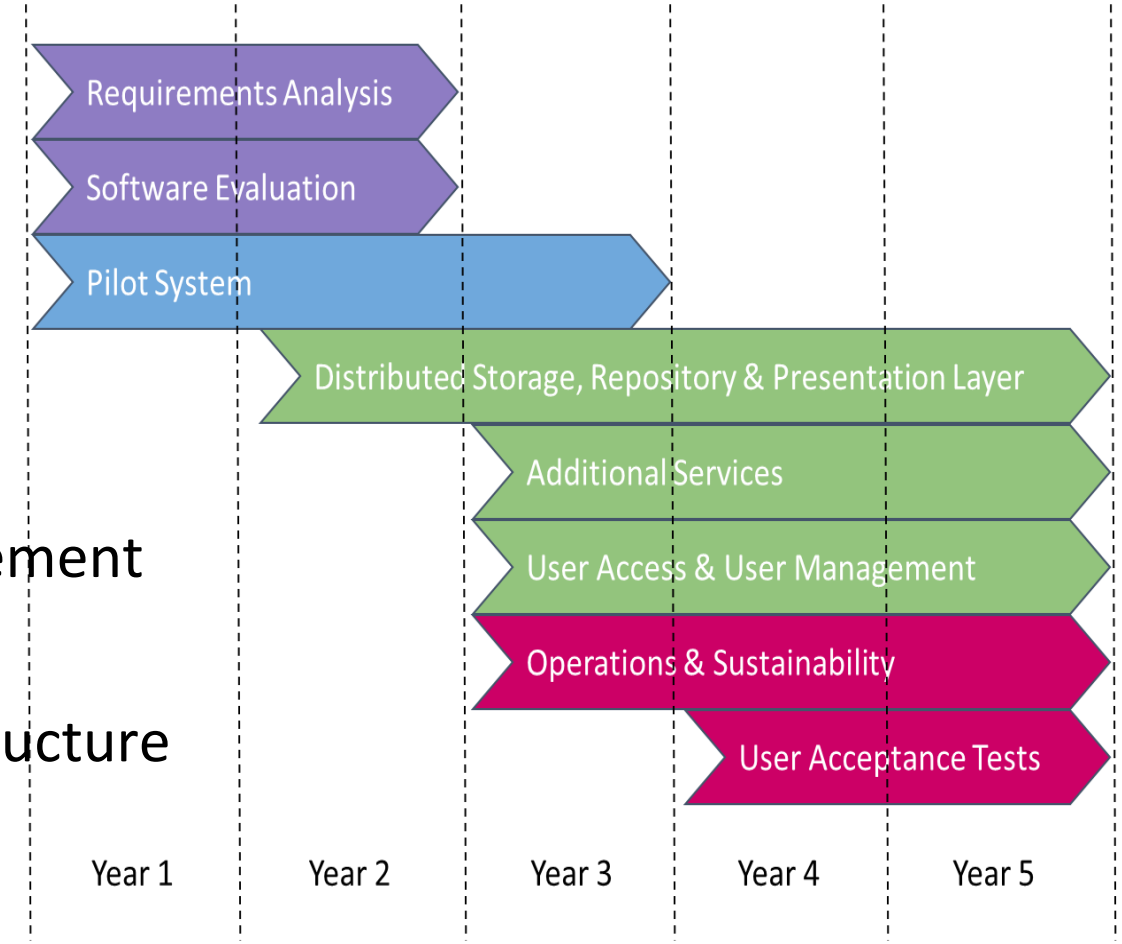
HLRS Höchstleistungsrechenzentrum | Stuttgart

- Requirements analysis
  - Requirements elicitation, based on stakeholder interviews
  - Requirements document
- Next steps
  - Architecture document,
  - Software evaluation, pilot system

DOI: 10.5281/zenodo.5882747

TA4: Agenda and Overview

- Measures
  1. **Initial phase**
     1. **Requirements Analysis**
     2. **Software Evaluation**
     3. **Pilot System**
  2. Development phase
     4. Repository Layer
     5. Presentation Layer
     6. Additional Services
     7. User Access & User Management
  3. Evaluation phase
     8. Operations
     9. Sustainability of the Infrastructure
     10. Specification Review &
     11. User Acceptance Tests

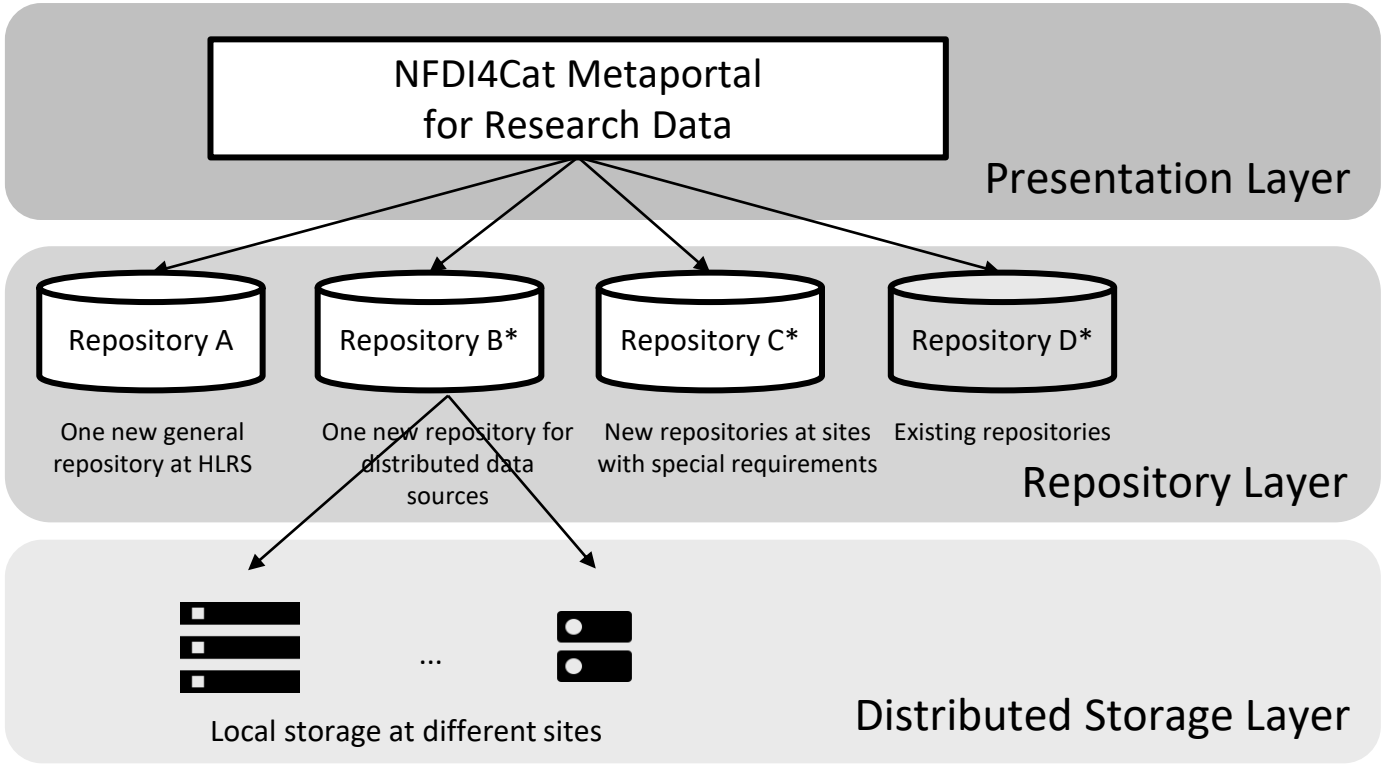


DOI: 10.5281/zenodo.5882747

Requirements Elicitation & Analysis

Proceeding

- Requirements elicitation
  - Interviews with partners
  - Definition of personas, epics and user stories
- Requirements analysis
  - Requirements document
  - Architecture document



Excerpt of the Proposal of NFDI4Cat

DOI: 10.5281/zenodo.5882747

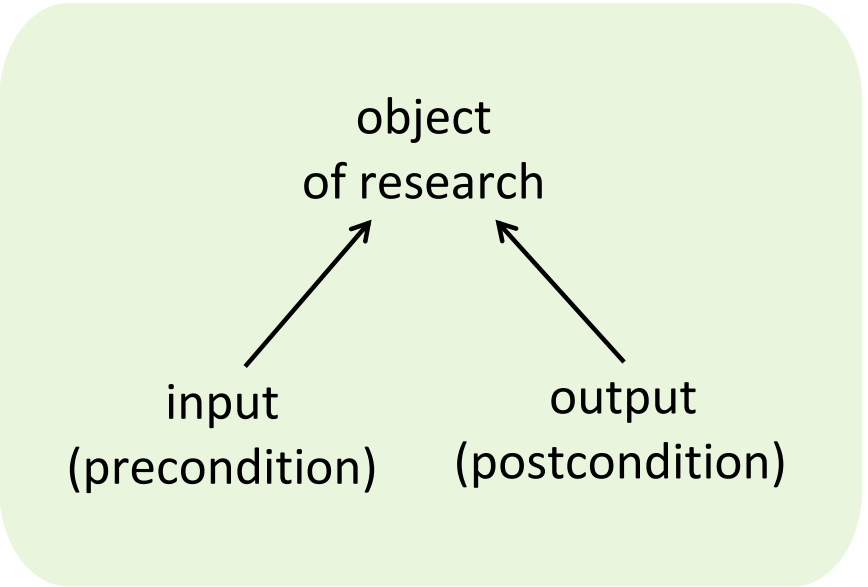
Requirements Elicitation

User Interviews

- Interviews of 30 minutes each were conducted with internal prospective users.

Representative research workflows

- For each research step, we jointly identified:
  - **input**, *i.e.*, all that needs to be present in advance (including equipment);
  - **output**, *i.e.*, all that is generated as an outcome of the research step.



Competency questions

- Useful for deriving requirements and for TA1 ontology development and metadata standards.

DOI: 10.5281/zenodo.5882747

User Interviews (I)

user interview 1:
photocatalytic
water splitting

user interview 2:
synchrotron QEXAFS
experiments

user interview 3:
CO_2 methanation

user interview 4:
biocatalysis
& multiphase flow

user interview 5:
CO_2 electrolysis

user interview 6:
metal-organic
frameworks

user interview 7:
nanoparticle
& SILP catalysts

user interview 8:
selective NOx
reduction

user interview 9:
photocatalytic
CO_2 reduction

user interview 10:
Fischer-Tropsch
process

user interview 11:
Coupled in-situ
and offline analytics

Representative research workflows

DOI: 10.5281/zenodo.5882747

User Interviews (II)

user interview 21: modelling, simulation, optimization of fixed bed reactors

user interview 20: homogeneous catalysis

user interview 12: syngas-to-ethanol catalyst characterization

user interview 19: synthesis and characterization of high temperature cells

user interview 13: syngas-to-ethanol catalyst performance

user interview 18: syngas-to-ethanol data science

user interview 14: RDM for enzyme activity data

user interview 15: on-line GC /oligomerization reaction

user interview 16: Water-gas shift reaction

user interview 17: FURTHRmind

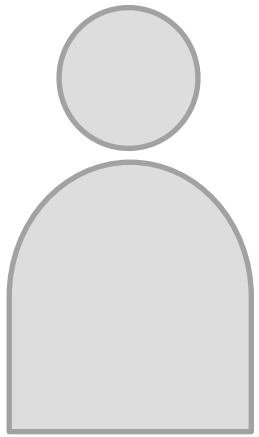
Representative research workflows

DOI: 10.5281/zenodo.5882747

Requirements Analysis – Personas

Personas

- To represent the user groups of the platform
- Characterized by name, age, profession, their overall aim and frequency in using the platform, and their proficiency in using computers and software

Role: Scientific Data Officer	
	**Name:** Julia Huber
	Age: 32
	– scientific data officer – proficient in using software – no intent on scientifically using the platform – uses the platform for managing confidentiality restrictions and is involved in the clearance process of research data trough the platform

Scientist
Local Administrator
Developer
Scientific Data Officer
Externals

DOI: 10.5281/zenodo.5882747

Requirements Analysis – Epics and User Stories

User stories

- A user story is a concept from software development that describes a "functionality that will be valuable" for the user in an actionable way.

Epics

- Some user stories describe a very complex functionality which can be divided into multiple smaller user stories and are called epics.

*As a [**description of user**], I want [**functionality**] so that [**benefit**].*

DOI: 10.5281/zenodo.5882747

Requirements Analysis – Epics Map

Meta Portal	Repository	Storage Harvester	Repository Harvester	Data Security	Metadata	Non-functional Requirements	Components
GUI	FAIR Data Preservation	Data Ingest	Agreement on APIs	Permission Managment Tool	Standardized Annotation	Teaching Materials	Epics
Exploration Tools	Multipurpose Storage and Processing	Unified API for Data Accessing	Harvester API	Technical Security Measures	Standardized Formats and (Meta)Data Schemas	Crediting Researchers	
Analysis Tools	Versioning	Interoperation with various Software Tools		Separation of confidential Data	In-depth Documentation	Performance	
Community Tools	GUI and API			Legal Constraints	Support complex Asserions and Queries	GUI Usability	
Interactive Dashboard	Central Storage			Cool-off- and Data-ownership-models	Providing additional Information	Seamless Integration into Workflow	
Quality Assessement Tool	Data Export				Quality Assurance Tools		
Reward System	Data Publishing	Analysis and Visualization Tools			Userfriendly way of providing metadata		
	Data Managment	Data Exploration Tools					
	API for Management of preexisting Data Servers	Working Spaces					

DOI: 10.5281/zenodo.5882747

Requirements Document

Following an agile approach

- Iterative extension of requirements document by incorporating user feedback
- Total of ~ 230 epics and user stories

Interoperability and architecture requirements analysis and metadata standardization for a research data infrastructure in catalysis

NFDI4Cat: Local and overarching data infrastructures

Current publications

20.01.2022 DOI: 10.5281/zenodo.5882747

Requirements Analysis – TA1 - TA4

*As a **scientist**,*
*I want **that metadata standards support an in-depth documentation of data sets***
*so that **other researchers can work with my data**.*

TA1: Ontology Development and Metadata Standards

TA2: Data Standards, Data Collection, Interfaces

*As a **developer**,*
*I want **an interface that provides access to the available data sets***
*so that **I can harvest the content**.*

*As a **scientific data officer**,*
*I want **algorithms for the automatic evaluation of the quality of data sets***
*so that **I can facilitate data quality assurance**.*

TA3: Data Analysis, Quality Management and Re-Use

TA4: Linked Extensible Infrastructure and Access Management

*As a **scientist**,*
*I want **a meta data portal that provides access to different repositories***
*so that **I can easily search for specific data sets**.*

DOI: 10.5281/zenodo.5882747

Task Area 1

TA1: Ontology Development and Metadata Standards

- overview about existing ontologies
- workflow to match existing ontologies

next steps

- workflow to extend ontologies from existing metadata standards
- gathering shared vocabularies

DOI: 10.5281/zenodo.5882747

Task Area 1: Work Plan / Partners

Develop/extend ontologies for (M1-6)

Metadata standards for (M8-12)

- catalyst synthesis data
- catalyst performance data
- reactor engineering
- catalysis-specific characterisation data
- operando data
- process engineering

Development of basic pilot (M1)

Development of extended pilot (M2)

for consolidated metadata standards

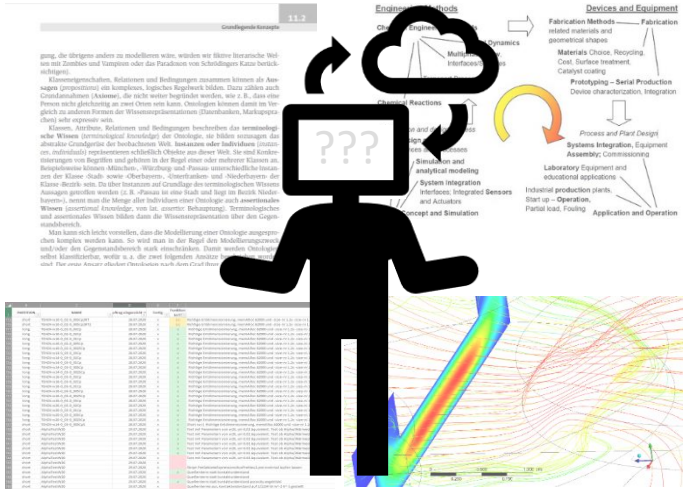
Partners (Year 1)

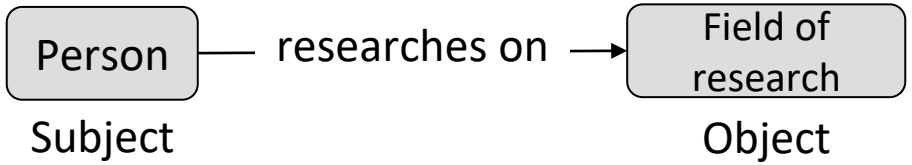
- HLRS
- KIT
- LIKAT
- MPI-CEC
- MPI-DCTS
- FAU
- RWTH
- TUBS
- TUDO
- TUM
- UHGW

DOI: 10.5281/zenodo.5882747

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

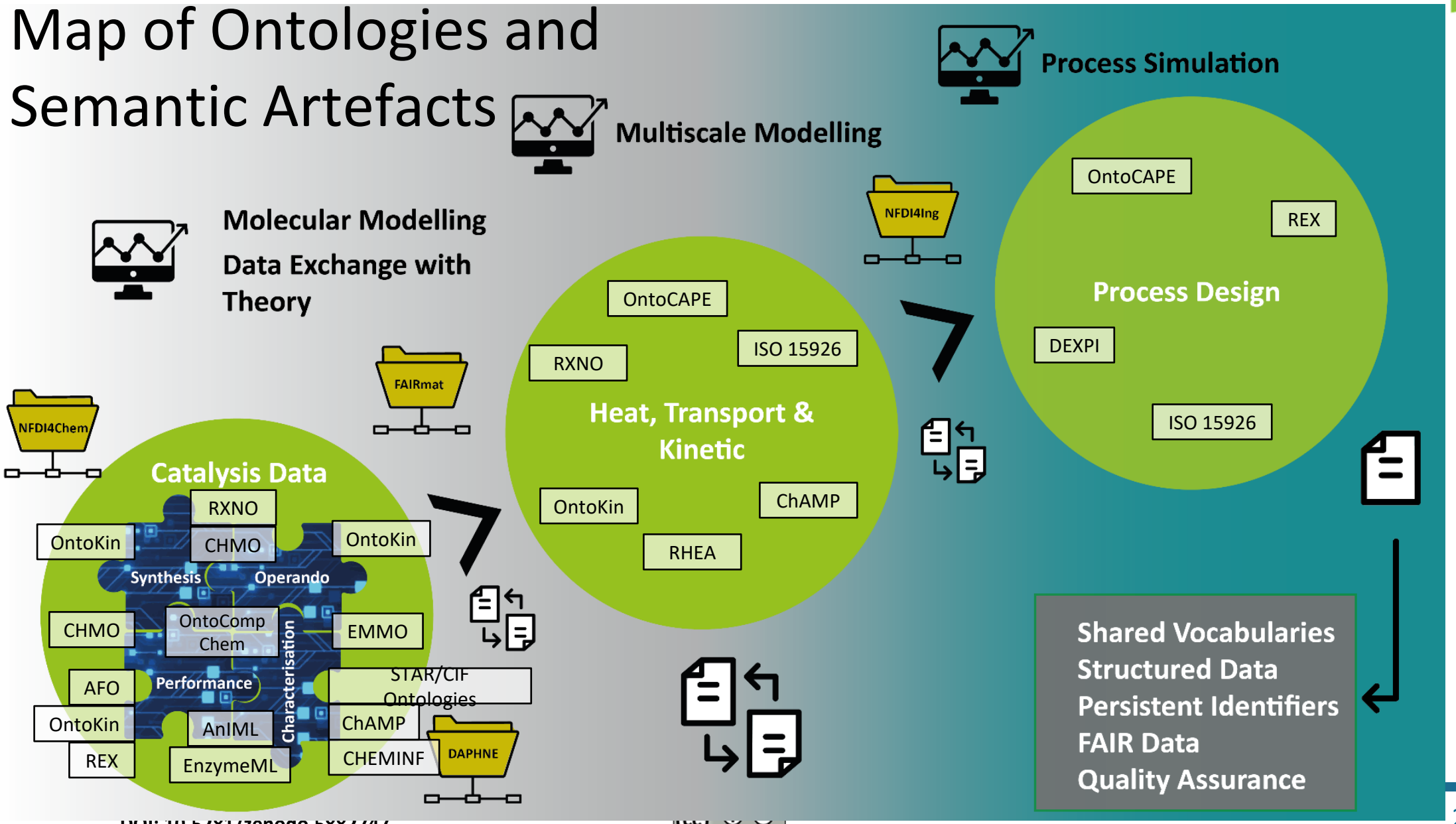




Unified data formats through *ontologies* and standardized *metadata schemes*

DOI: 10.5281/zenodo.5882747

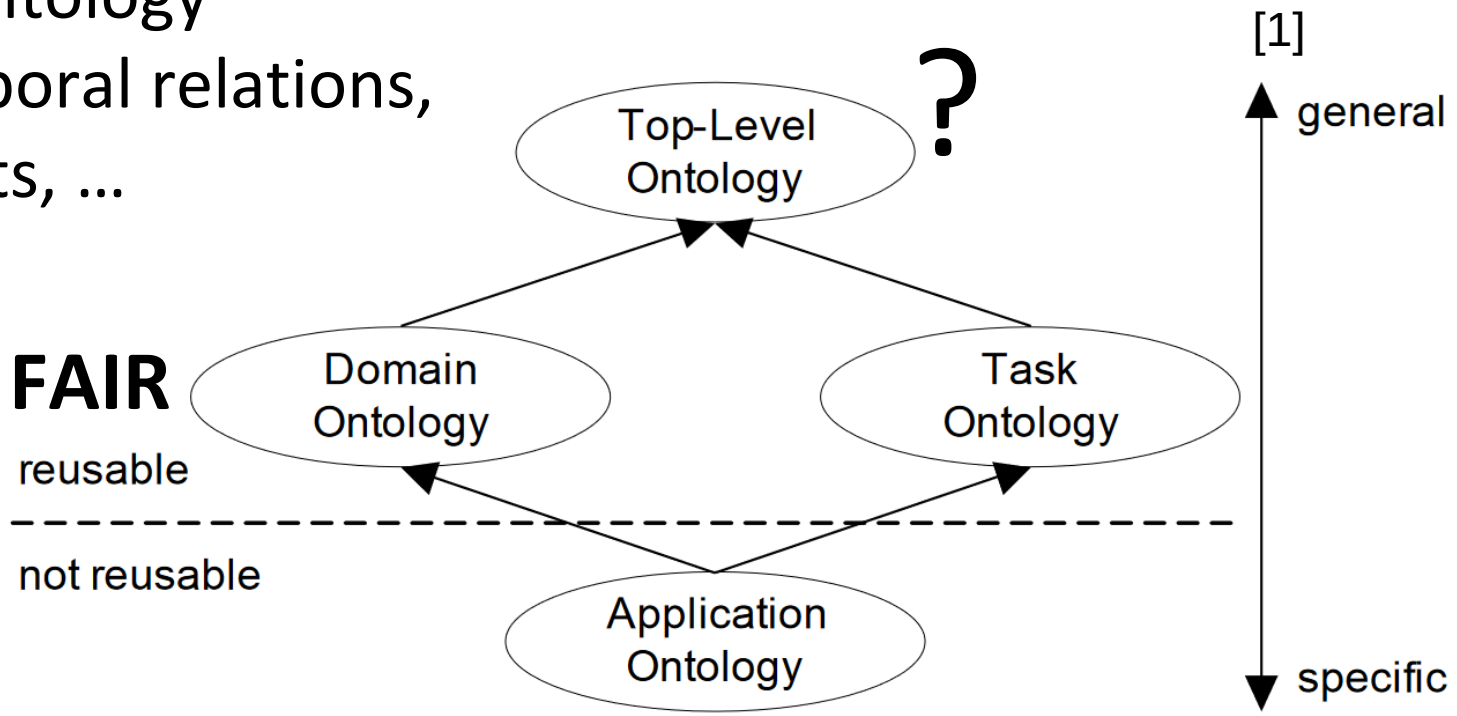
Map of Ontologies and Semantic Artefacts



DOI: 10.5281/zenodo.5882747

Quest of top-level ontology

- Domain and task unspecific ontology, i.e. does not contain any chemical, physical, etc. specific terms
- Most abstract layer of an ontology
- Define e.g. spatial and temporal relations, physical and abstract objects, …



Top-level ontologies allow for reuse, interoparability, matching

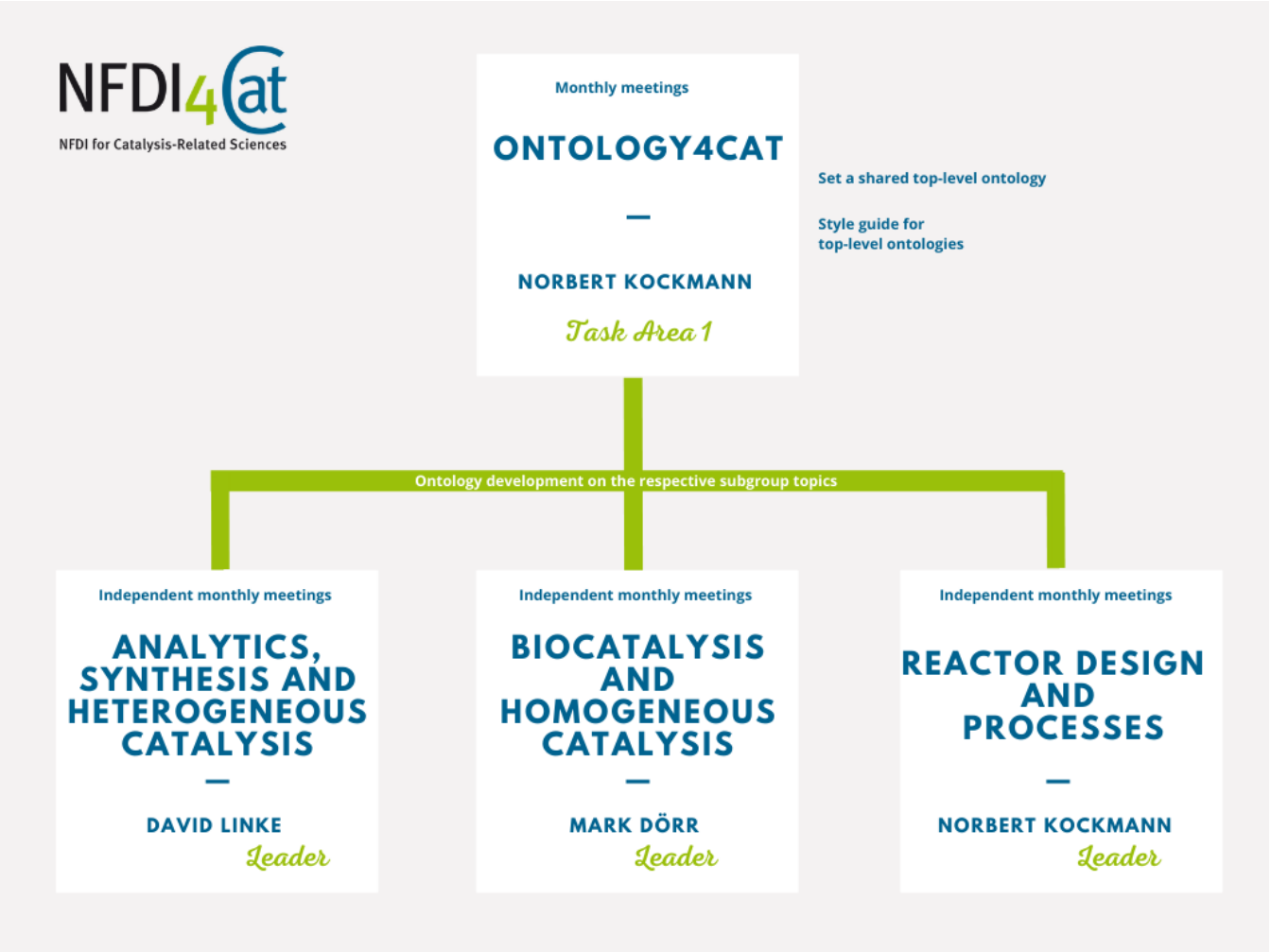
DOI: 10.5281/zenodo.5882747

[1] Guarino: Formal Ontology and Information Systems (1998)

Subgroups for Ontology Development

- Ontology development and refinement of core ontologies in three subgroup topics

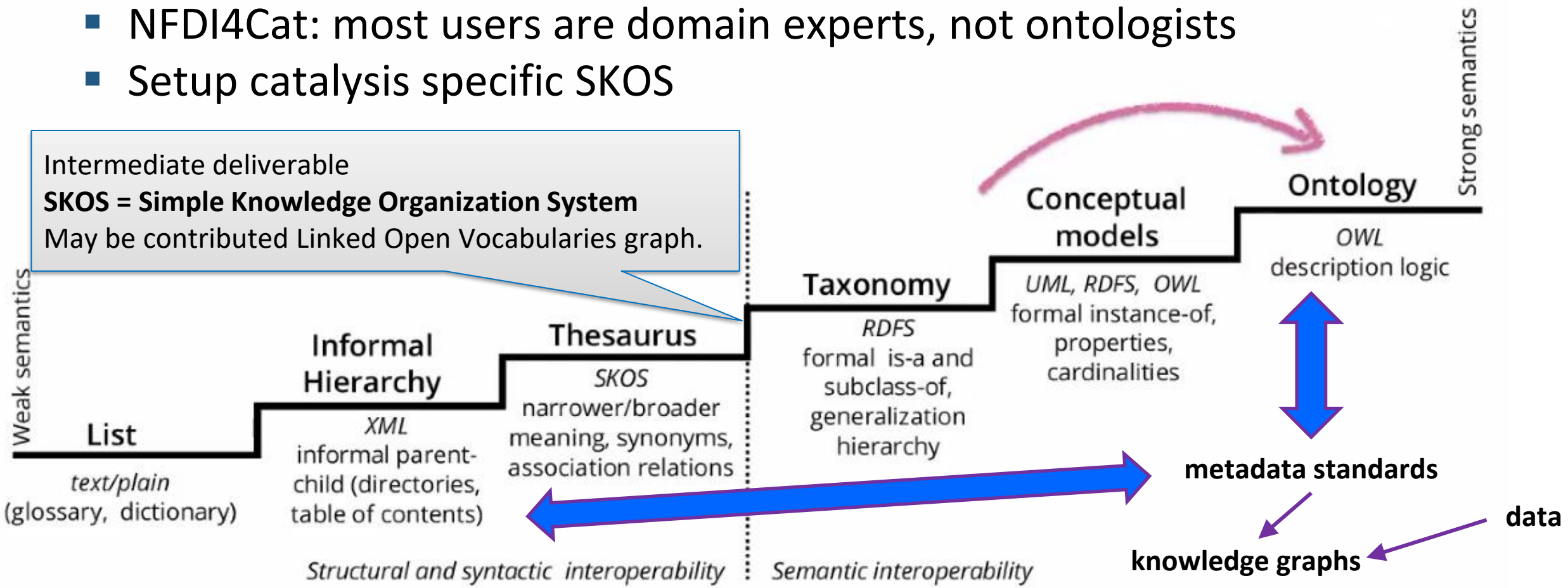
⇨ Determination of core domain and top-level ontologies important



DOI: 10.5281/zenodo.5882747

The way to ontologies

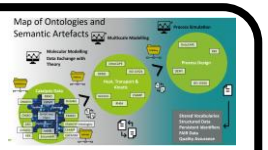
- Many steps with domain and IT knowledge
- NFDI4Cat: most users are domain experts, not ontologists
- Setup catalysis specific SKOS



DOI: 10.5281/zenodo.5882747

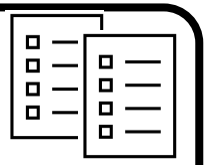
Current workflow of NFDI4Cat – TA1

Existing Ontologies



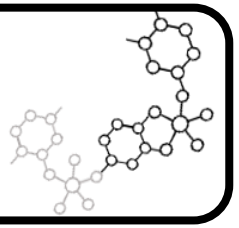
- ✓ Existing ontologies gathered
- ✓ Clustered by topics of catalysis research

Shared vocabularies



- Template for domain experts (similar to VocExcel)
- Workflow for automating SKOS generation

Extended Ontologies

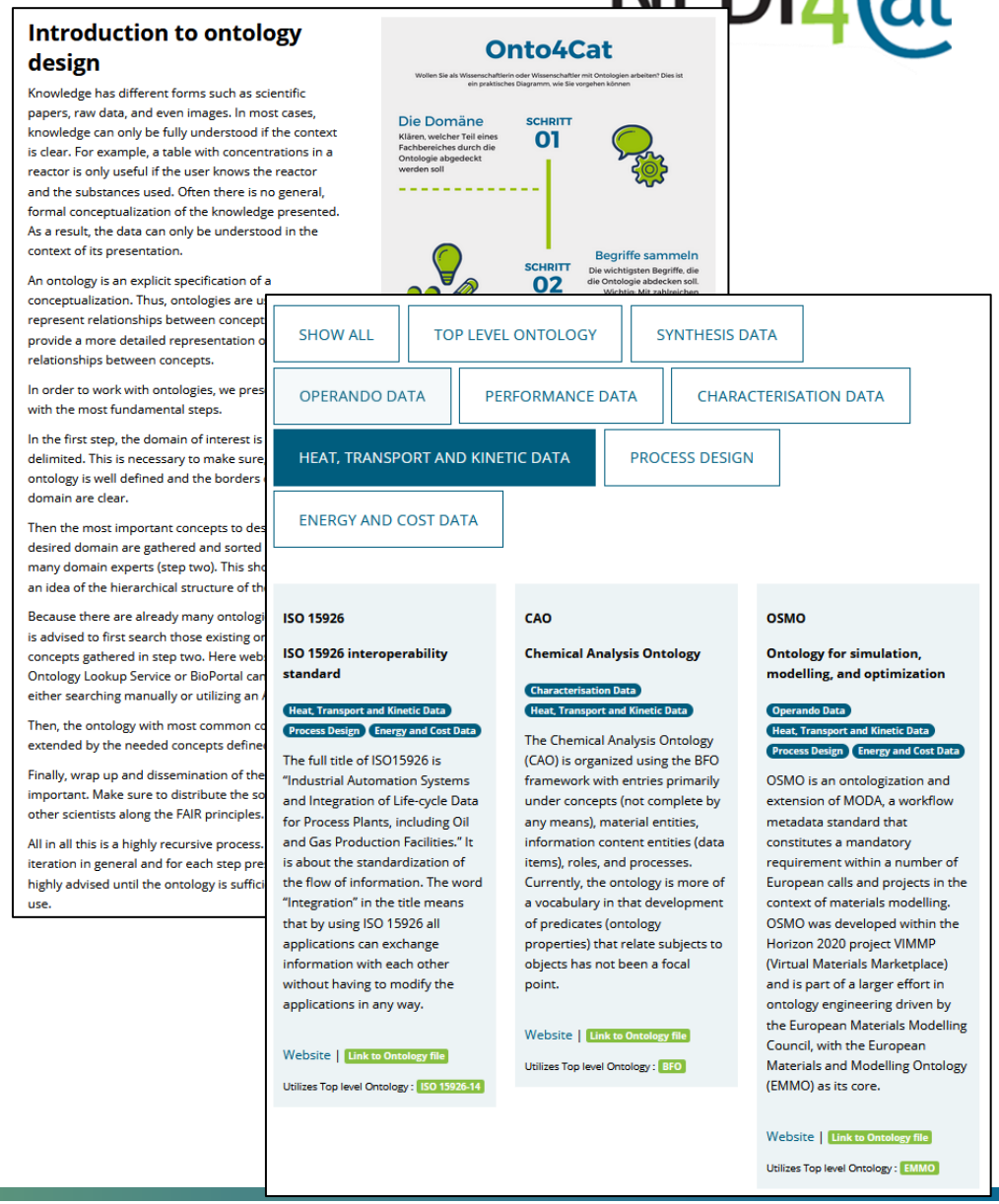


- ➢ Ontologies for catalysis research
- ➢ Extended by concepts of community

DOI: 10.5281/zenodo.5882747

Ontology collection

- Now as website containing short introduction to ontologies!
- Link to ontology documentation and files
- Sort Ontologies by tags relevant to digital value chain in catalysis

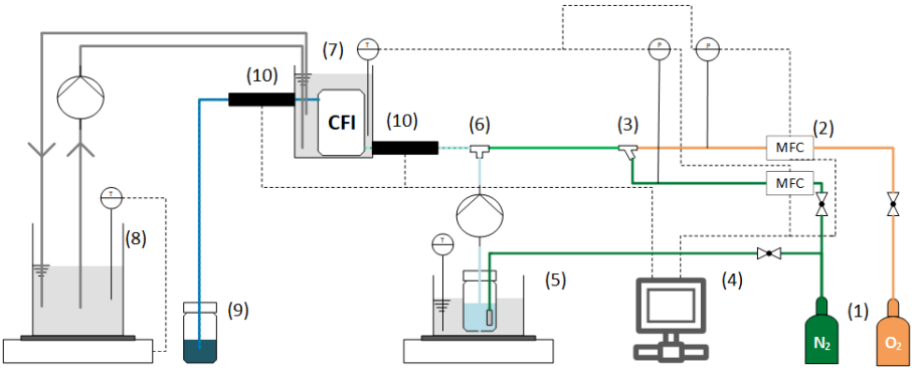


https://nfdi4cat.org/services/ontology-collection

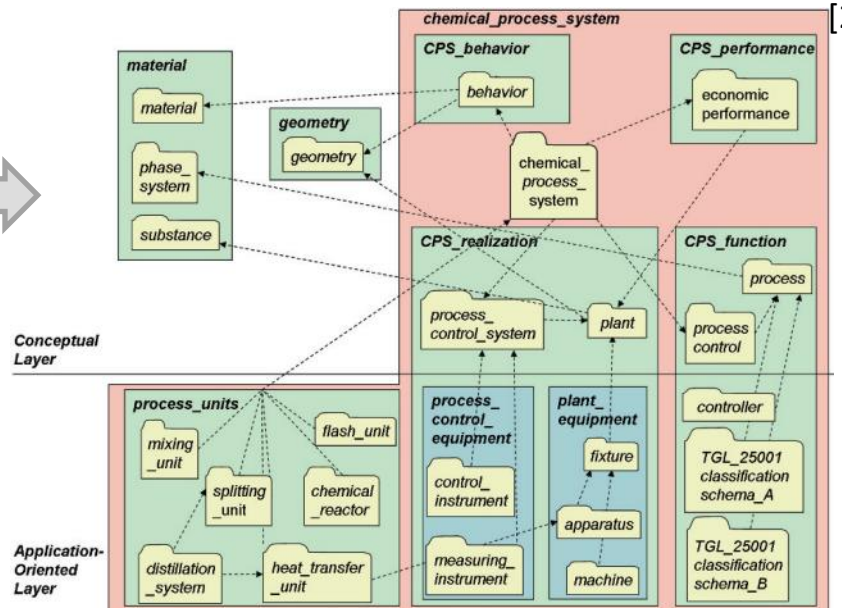
First Example of an Ontology Extension

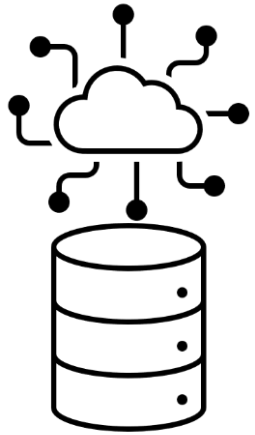
- Laboratory trials investigating Biocatalysis
- Extension of existing Ontologies by needed concepts
- Knowledge graph with raw experimental data in Database

[1]



[2]





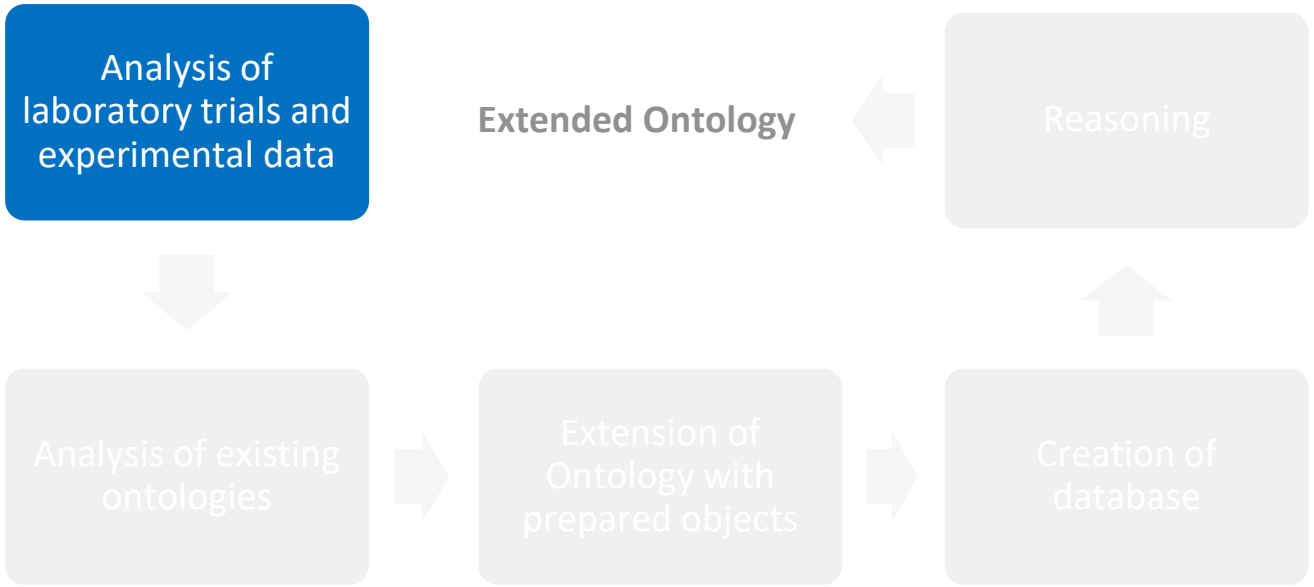
[1] Eroglu, Masterthesis, AG Apparatedesign, Technische Universität Dortmund, 2021

[2] Marquardt W, Morbach J, Wiesner A, Yang A. OntoCAPE. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010

DOI: 10.5281/zenodo.5882747

Workflow – Analysis of data

- Collecting of objects and concepts to be modelled
- Analysis of lab trials
  - Physical components of experimental setup
  - Functional aspects of experiment
  - Classes, Relations, Individuals



Experiments to be Modelled

- 2 different CFI geometries
- 2 different enzyme concentrations
- 3 different oxygen concentration

Parameter	Unit	Value 1	Value 2	Value 3
n_{turns}/n_{bends}	-	3/13	-	10/3
$C_{laccase}$	g/l	0,2	-	0,8
C_{O2}	%	3	7	10

[1]

12 different experiments

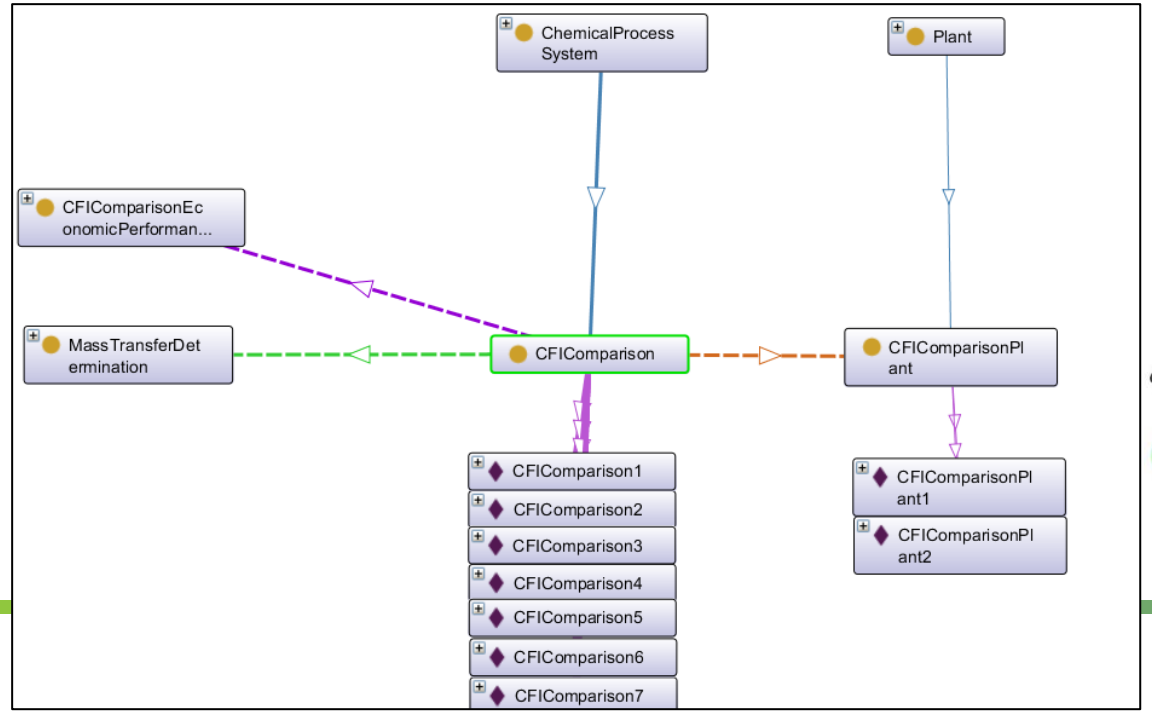
Trial	n_{turns} [-]	$C_{laccase}$ [g/l]	C_{O2} [%]
1	3	0,2	3
2	10	0,2	3
3	3	0,8	3
4	10	0,8	3
5	3	0,2	7
6	3	0,2	10
7	10	0,2	7
8	10	0,2	10
9	3	0,8	7
10	3	0,8	10
11	10	0,8	7
12	10	0,8	10

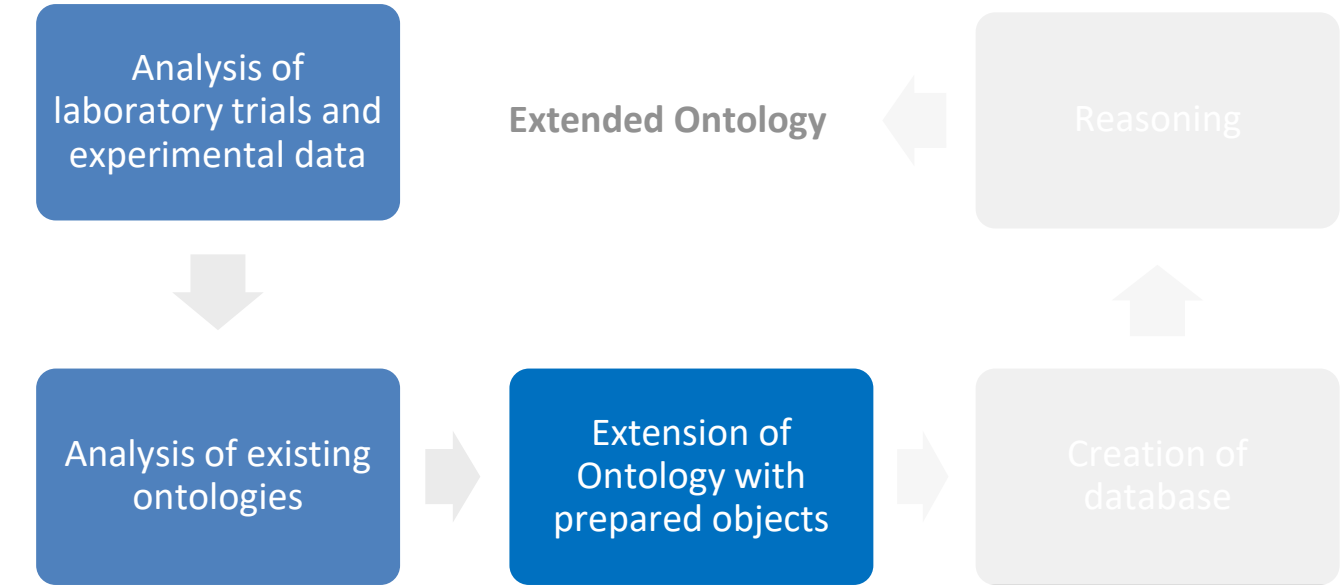
DOI: 10.5281/zenodo.5882747

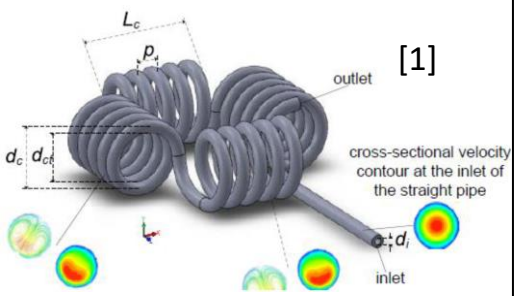
Workflow – Extension of ontologies

- Extension with classes, individuals, relations and attributes modelling the experiments

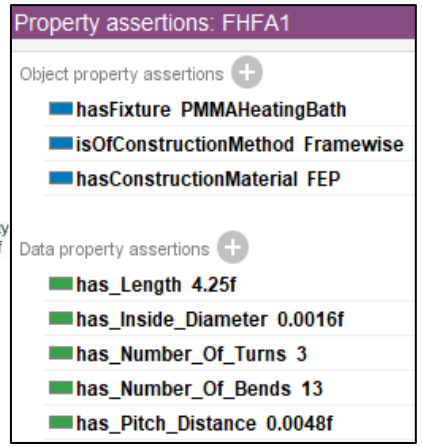
New classes and individuals added to ontology

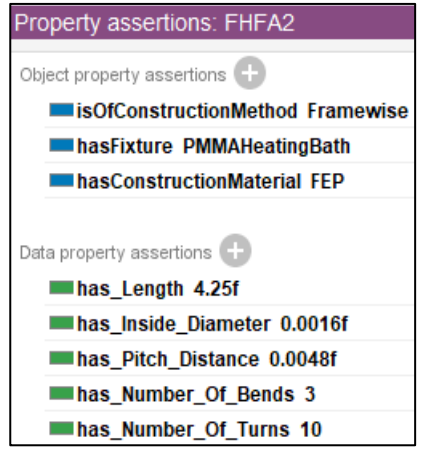






[1]



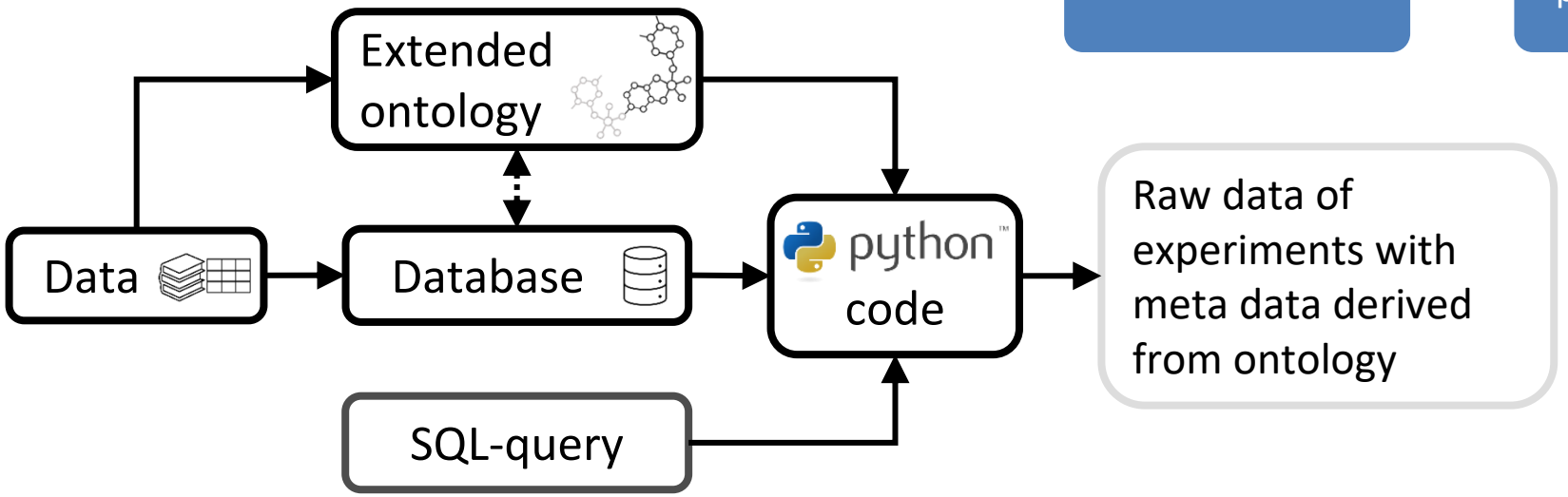


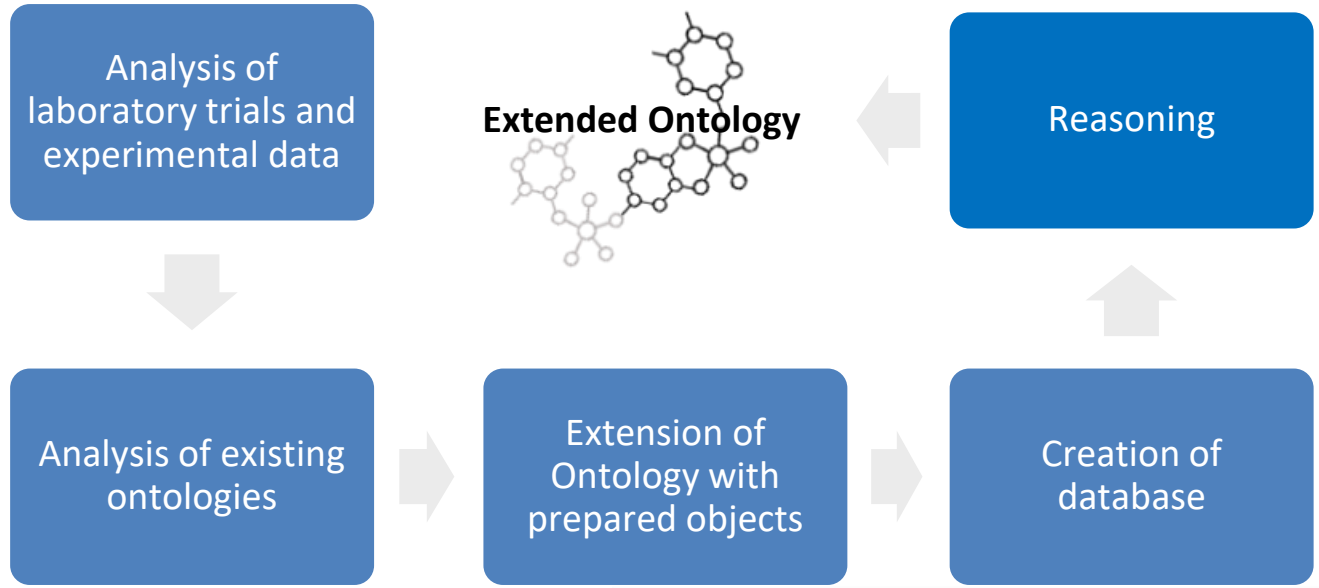
[1] Kurt et al., IEC&R 2017

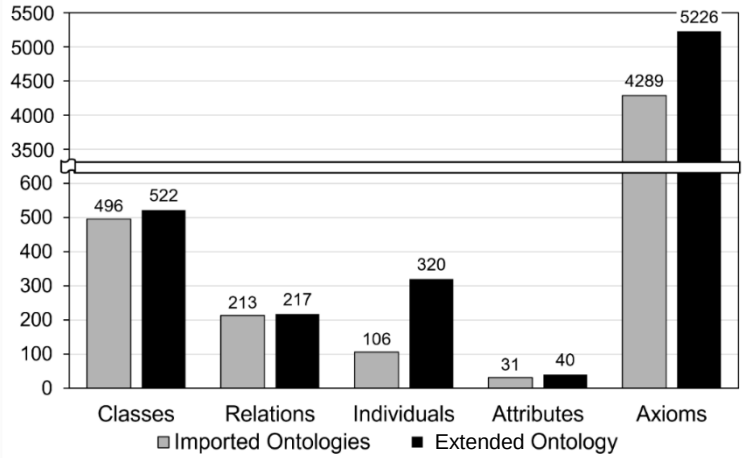
Two individuals with data properties relevant to the experiments

Workflow – Reasoning

- Checking for inconsistencies using reasoner HermiT
- Revision of ontology
- Obtain extended ontology

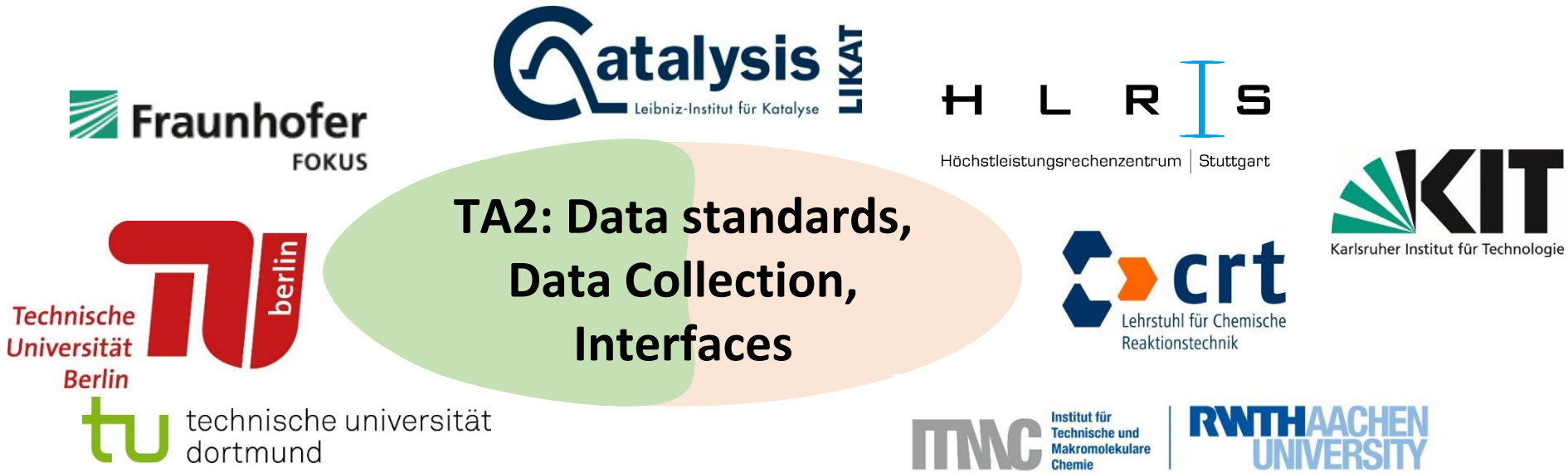






DOI: 10.5281/zenodo.5882747

Task Area 2



- definition of data standards for catalyst performance data
- development of data connections tools
- evaluation of different ELNs

- next steps:
  - coupling ELNs and data connections tools

DOI: 10.5281/zenodo.5882747

Task Area 2: Work Plan / Partners

Data standards (M1-6)

Data collection tools (M8-12)

catalyst synthesis data
catalyst performance data
reactor engineering
catalysis-specific characterisation data
operando data
process engineering

Development of local pilots (M13-17)

specific pilots, **cross-cutting pilots**

Interface specification for local repositories (M7)

Partners (Year 1)

- FOKUS
- HLRS
- KIT
- LIKAT
- FAU
- RWTH
- TUB
- TUDo

M1-6 Data standards: ELNs Task Force Overview

- Purpose:
  - To understand and document the requirements, user concerns and organizational concerns in implementing an ELN for catalysis research
- Activities:
  - Monthly meetings
  - Distribute and analyze user surveys
  - Organize ELN testing
  - Discuss with organizations who already use ELNs

DOI: 10.5281/zenodo.5882747

M1-6 Data standards: ELNs from the User's Perspective

User Stories

- As a researcher, I want:
  - easy to use tool
  - link experiments with results
  - find prior research
  - access from anywhere
  - work collaboratively
  - find chemicals quickly

User Concerns

- *'Will I need to change the way I work?'*
- *'Will this ELN be supported long term?'*
- *'Can I access the notes if I move to a new organization?'*

M2 Data standards for catalyst performance data: Archiving of Catalyst Data including Metadata

Experimentally measured catalyst performance data including metadata

- Fixed bed flow reactor
- Tubular flow reactor
- Monolithic flow reactor

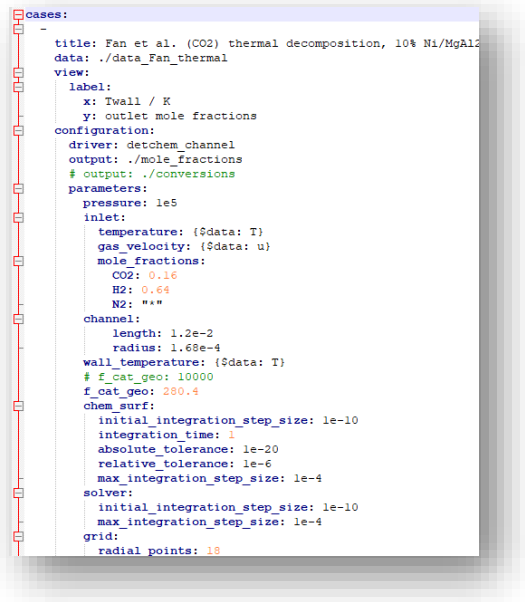
- Data archive
- Extraction of metadata for setting up input files for models and simulation

Reactor data

	Parameter		Value	Unit
EXPERIMENT	type of experiment	*	end-of-pipe	
FIXED BED	length	*	1.07E-01	m
	diameter	*	8.00E-03	m
	total mass	*	0.5	g
	bed porosity	*	38.1	%
	particle shape	*	sphere	
	particle diameter	*	6.30E-04	m
	particle thermal conductivity	1*		W/m K
	radial thermal conductivity			W/m K
	surface area to volume ratio			1/m
	material density			kg/m³
WALL	wall temperature			K
	wall thickness			m
	wall thermal conductivity	1*		W/m K
	outside heat transfer coeff			W/m² K
	overall heat transfer coeff	2*		W/m² K
	heat flux	3*		W/m²
CATALYST	active catalyst/metal	*	Ni	
	metal loading	*	20	%
	particle porosity			%

Performance

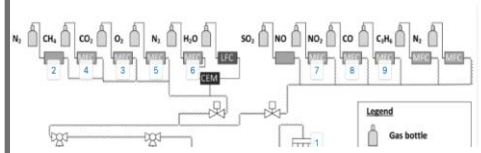
T	H2	CO2	CO	CH4	H2O
405.098	0.051253	0.041878	0.00E+00	0.00E+00	0.00E+00
481.252	0.050993	0.041886	0	0	0
545.448	0.047249	0.040464	0	0.001178	0.002785
632.301	0.029937	0.033865	0.001901	0.005651	0.013151
688.944	0.023335	0.031192	0.003335	0.007264	0.018335
746.846	0.025841	0.029769	0.006376	0.005662	0.017805
806.007	0.032097	0.027811	0.011739	0.002632	0.016918
863.909	0.035317	0.024959	0.016209	0	0.017281
929.364	0.034966	0.022644	0.019073	0.000323	0.019073
986.007	0.0339	0.021043	0.020686	0	0.020686
1048.944	0.032299	0.019442	0.022299	0	0.022299
1114.399	0.031234	0.017663	0.024627	0	0.023734



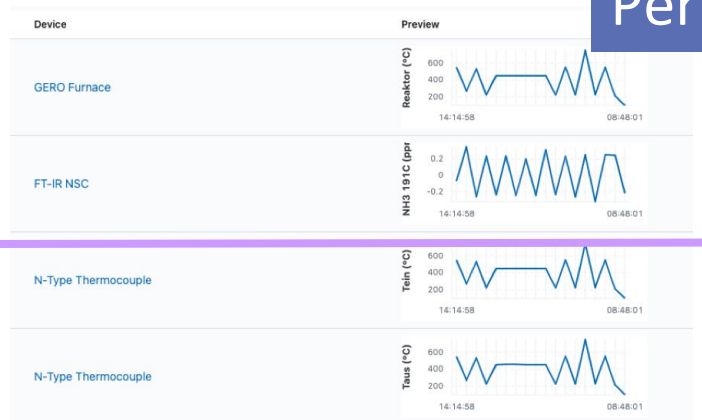
R. Chacku, S. Angeli, D. Schmider, H. Gossler, O. Deutschmann

M9 Data collection tool for catalyst perormance data

Monitor and document processing of original data including analytics and catalyst information



Reactor data



Analytics data

Performance

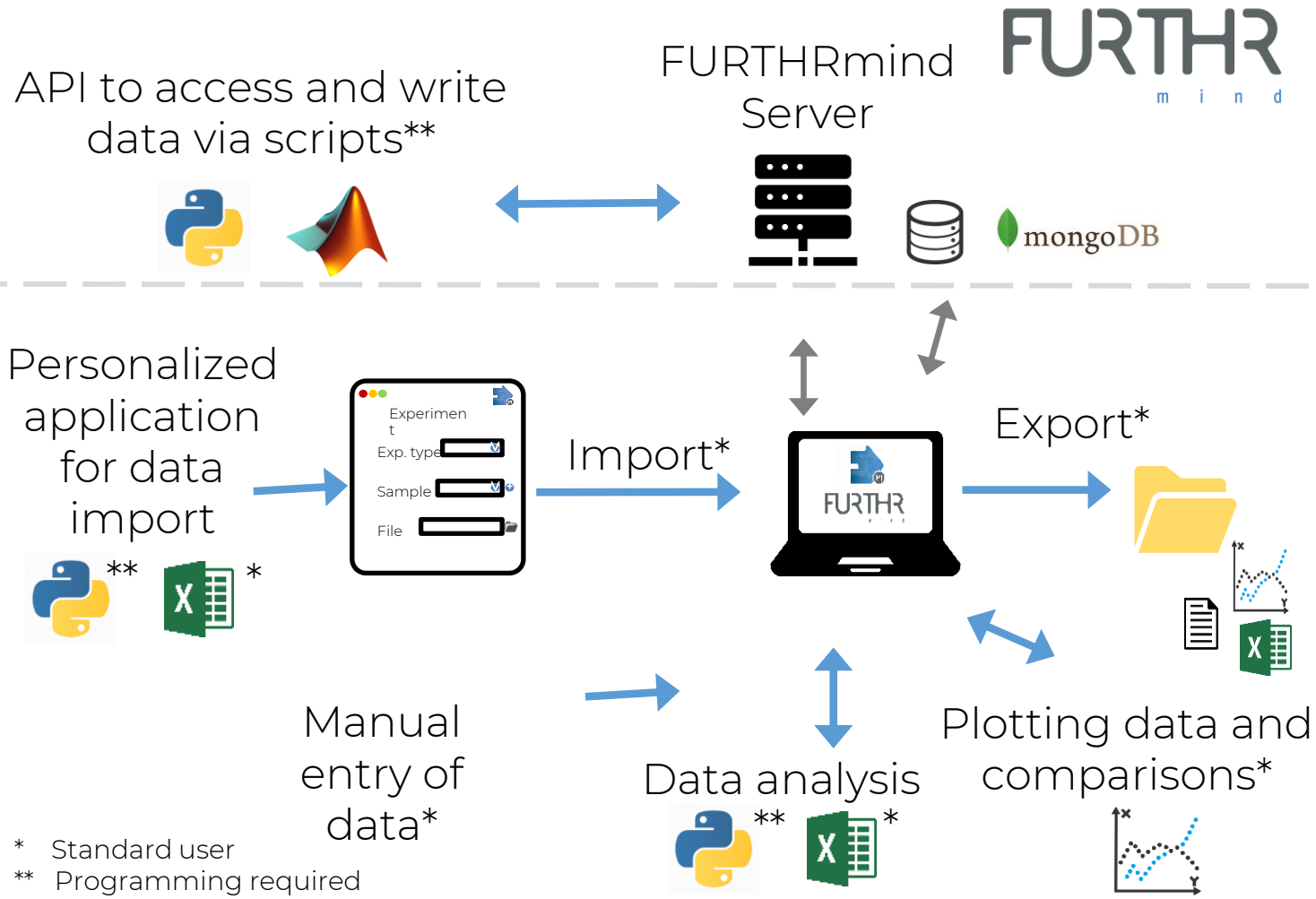
Y1(X0)	Y2(X0)	Y3(X0)	Y4(X0)	Y5(X0)	Y6(X0)	Y7(X0)
Nitrogen diatomi	Carbon monoxid	Methane (CH4)	Air	Nitrogen diatomi	Sulfur dioxide (S	Water (
ml/min	ml/min	ml/min	l/min	ml/min	ml/min	ml/min
N2 Traeger	CO 500ml/min	CH4	5l/min Air	N2	SO2 50 mL	H2O 75
0	0	3,889379	0,450598	0,479986	0	0
0	0	3,889308	0,449088	0,479947	0	0
0	0	3,888306	0,451563	0,479987	0	0

Data archive and extraction of metadata for setting up input files for models and simulation

H. Gossler et al., PhysChemChemPhys 20 (2018) 10857; Catalysts 9 (2019) 227.

DOI: 10.5281/zenodo.5882747

M8-M12 Data Workflow



* Standard user
** Programming required

Targets:

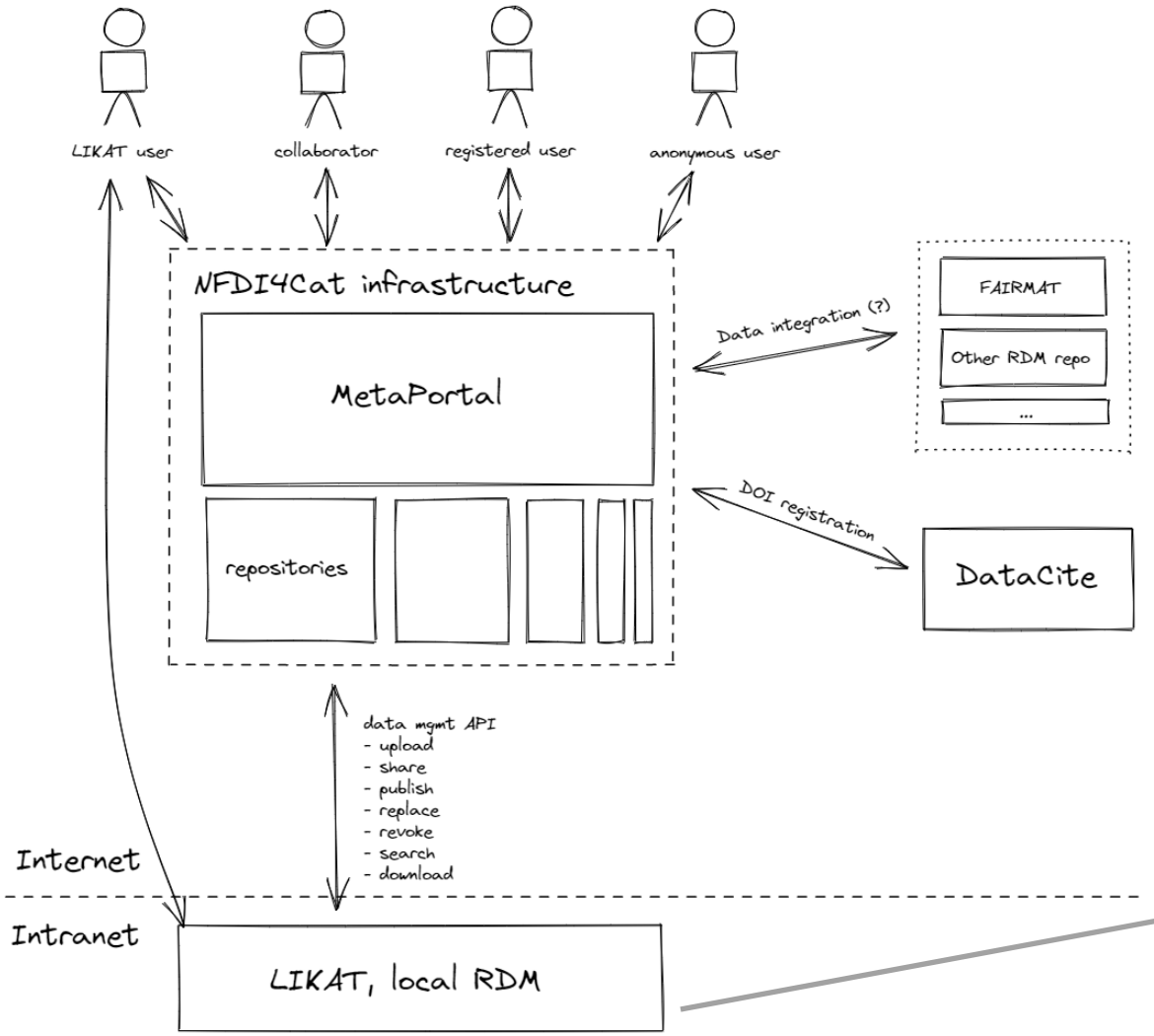
- Find a suitable ELN
- Testing in the working group
- Accessibility
- Automatisation

Sub-targets:

- Training of students
- Get a „feeling“ for data
- Make it attractive

S. Palkovits

M14 Develop basic local cross-cutting pilots

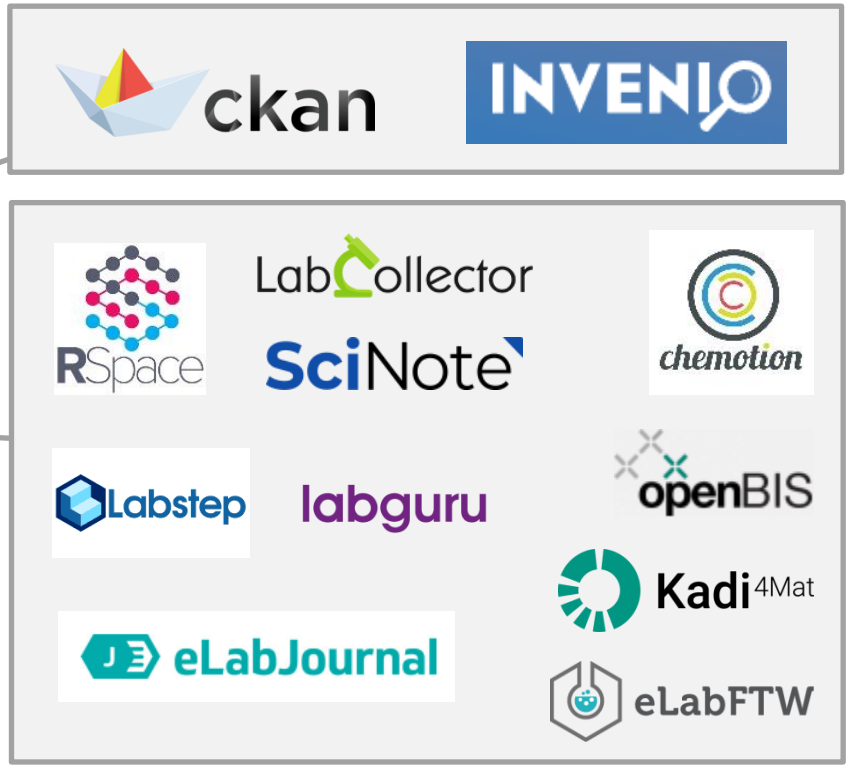


- Scope:
  - A system to support researchers of different catalysis sub-disciplines to locally handle their data
- Goals:
  - Provide benefits to users in daily research routine
  - Linked-data compatible FAIR data storage
  - Enable sharing & publishing via NFDI4Cat meta portal
  - Support complex access control schemes

DOI: 10.5281/zenodo.5882747

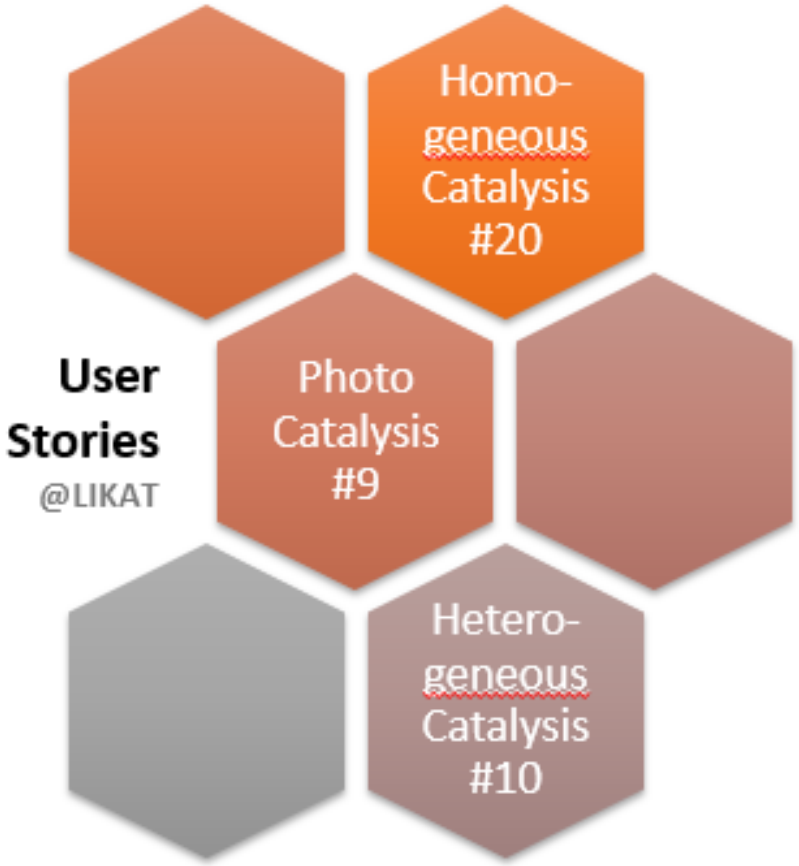
M14 Local cross-cutting pilots: LIKAT´s approach

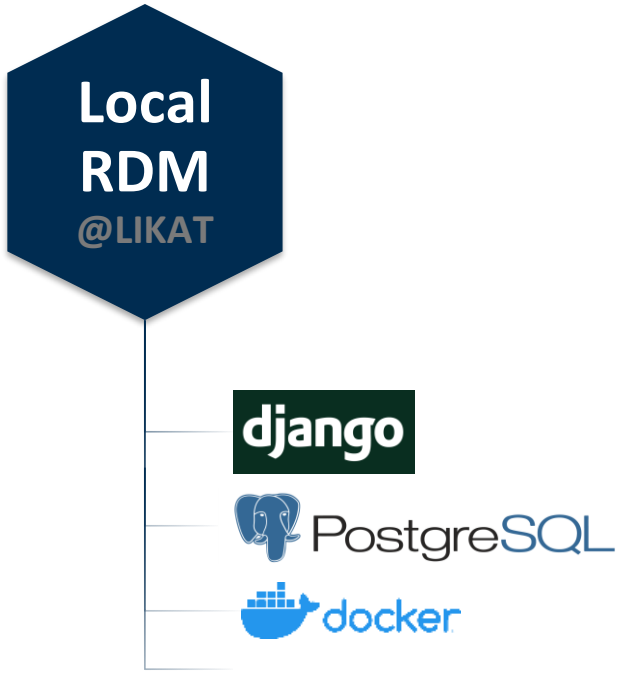
- ✓ LIKAT internal cross-cutting user survey
- ✓ Collection of user stories
- ✓ Review & evaluation of tools landscape
  - Data-repository software
  - ELN and LIMS systems
  - Data formats
  - Web development technology stacks
- ✓ Decision about technical basis
- ○ Implementation ongoing



DOI: 10.5281/zenodo.5882747

M13-17: Development of local pilots: Intranet web application for local RDM





DOI: 10.5281/zenodo.5882747

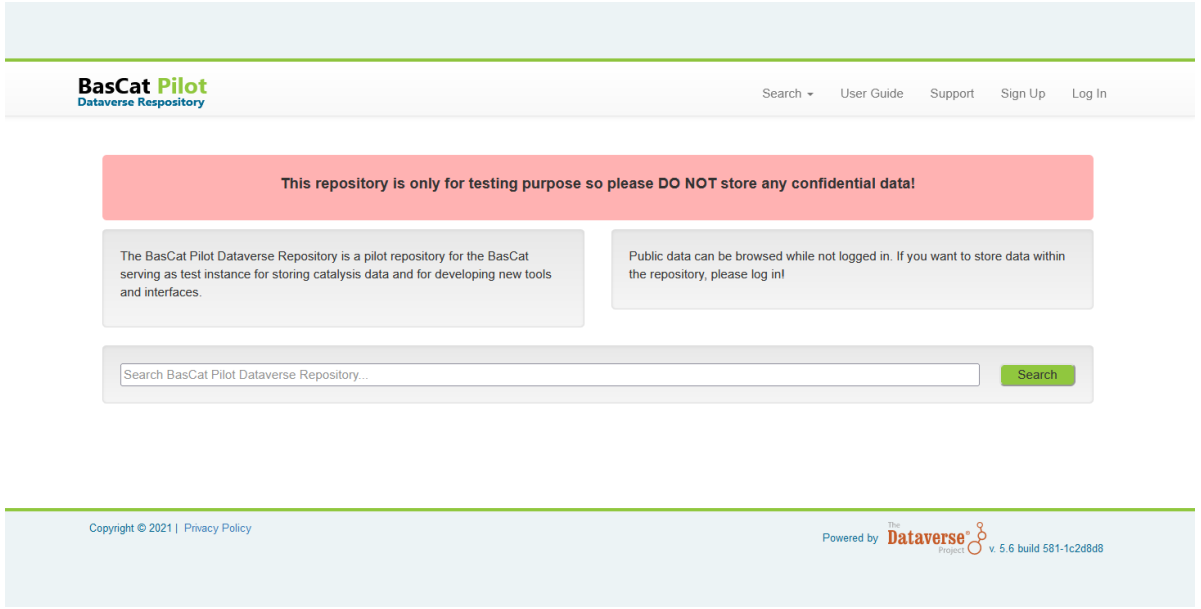
M13-17: Development of local pilots: BasCat Pilot

General Goals

- Pilot will address the conversion of synthesis gas (CO + hydrogen) into larger hydro carbons and oxygenates.
- Development of methods and software tools for heterogeneous catalysis that support the whole data work flow from defined experimental data, data import, data storage as well as further data processing in terms of curation and visualization.

- Instance of repository for BasCat at FOKUS

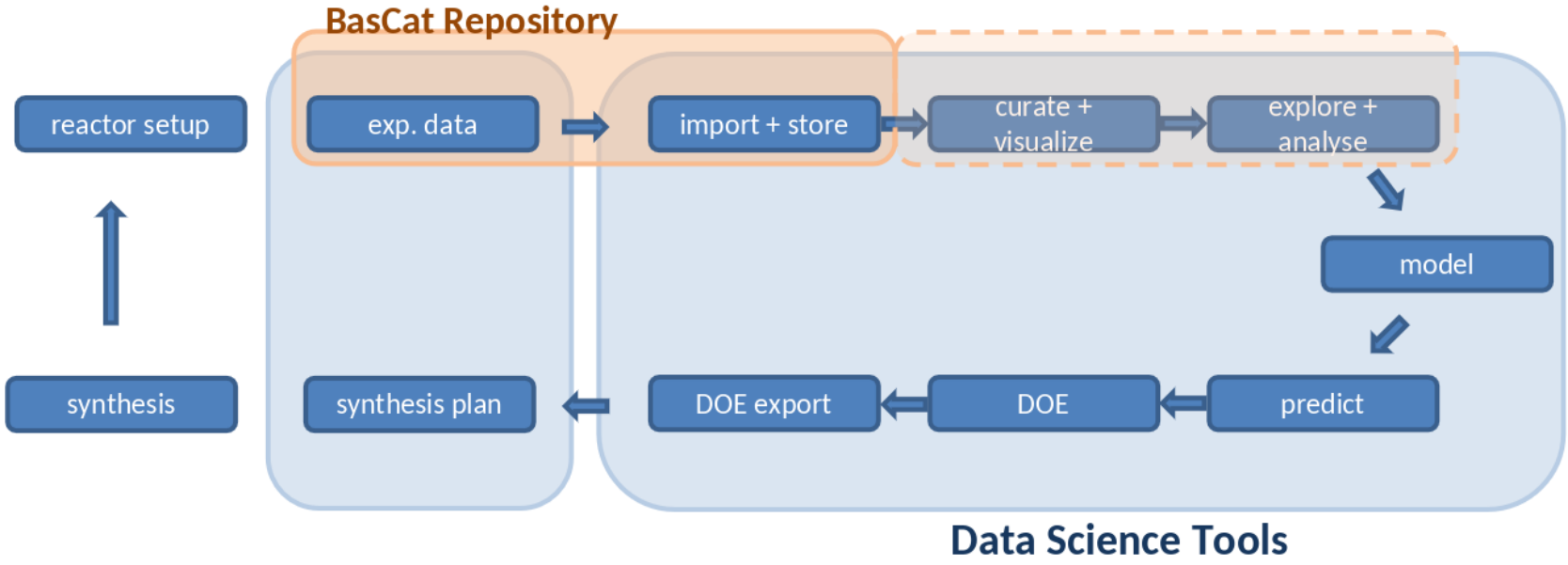
- **First step:**
- Exploration of in-built functionalities:
  - Data import (GUI, API)
  - Data storage
  - Data publishing and permission management
  - Data exploration



DOI: 10.5281/zenodo.5882747

M13-17: Development of local pilots: BasCat Pilot

- Import of performance data for storage in BasCat repository (explore import/export functionality)
- Embedding of BasCat tools for processing of performance data (explore data science methods)
- Setting up pilot of NFDI4Cat meta dataportal (piveau)
- Connection of BasCat repository
- Stepwise extension of pilot to cover steps within the work flow (synthesis, characterization, catalytic tests)



DOI: 10.5281/zenodo.5882747

Task Area 3

TA3: Data Analysis, Quality Management and Re-Use

- Overall data workflow concept developed for heterogeneous catalysis research
- Data tools developed as proof-of-concept with local data
- First link established BasCat, CaRMeN, local pilot
- Use case data measured and documented for heterogeneous catalysis

DOI: 10.5281/zenodo.5882747

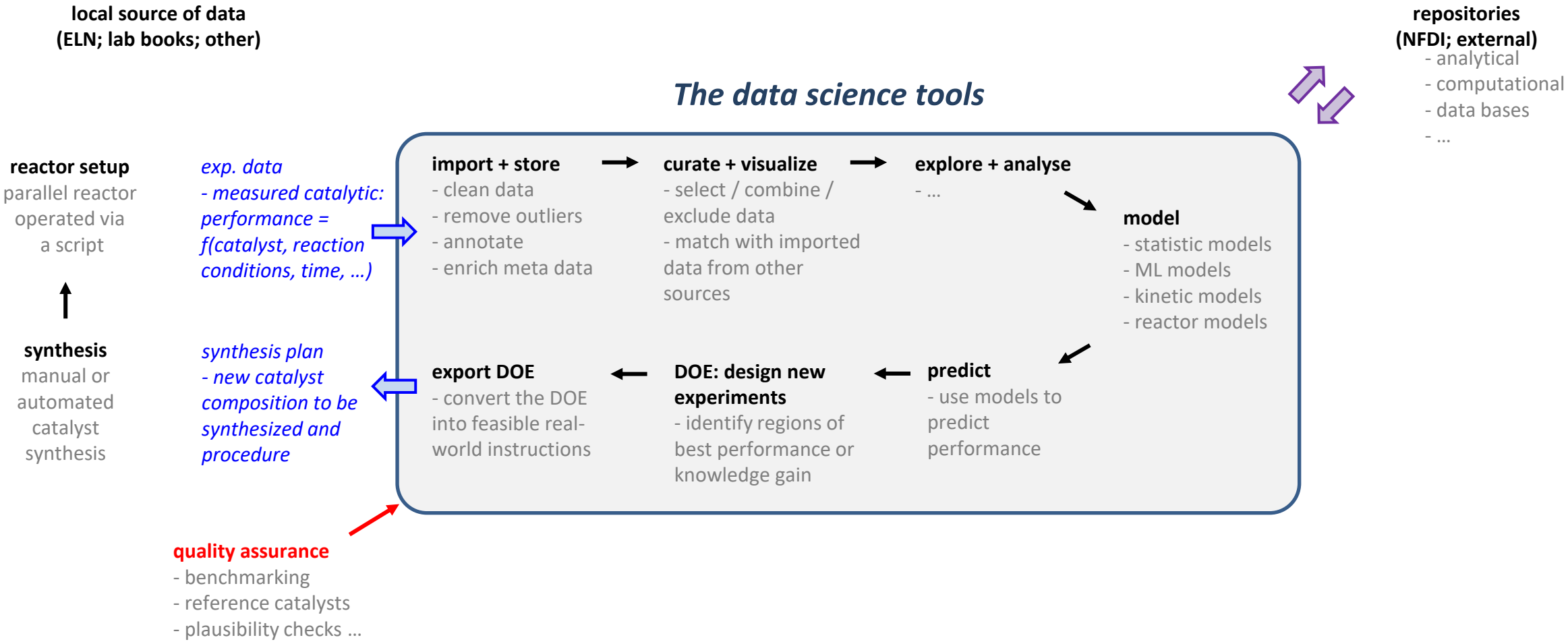
Task Area 3: Work Plan / Partners

Novel algorithms (M1-3)	integrate existing tools into basic pilots algorithms to rate data quality
Tool development (M4-10)	tools for data analysis and re-use tools for quality monitoring
Connecting the tools (M11-14)	interface external repositories
Use cases (M15-18)	use cases in heterogeneous catalysis

Partners (Year 1)

- TU Berlin
- KIT

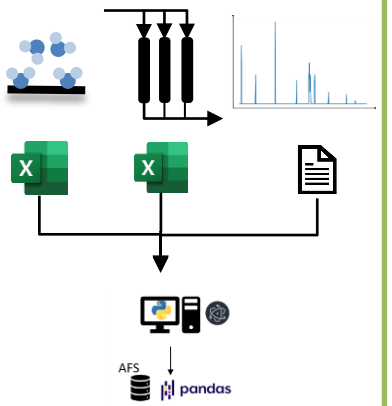
The overall data workflow concept



Tool development as proof-of-concept

import / store

From files to data frames

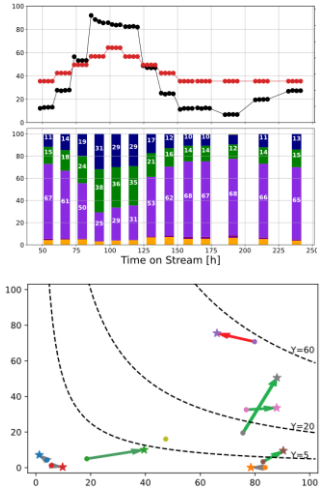


Importing & storing tools

- Data from preparation, testing and analytic is semi-automatically imported and stored in structured data-frame

curate / visualize

Plot multiple experiments

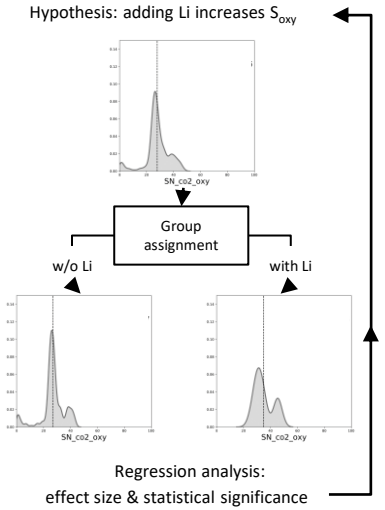


Visualization tools

- Calculation of experimental performance indicators
- Visualization and comparison of catalyst performance

explore / analyze

Meta analysis

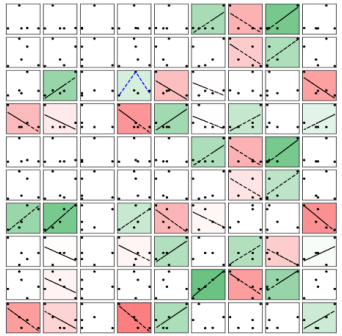


Analysis tools

- Twinning
- Regression
- Meta analysis

model

Searching for correlations



Correlation tools

- Build a set of potential descriptors from experiment and theory
- Identify and visualize possible correlations between different descriptors

quality assessment

Quality analysis

Descriptive statistics

- Min / max / average / frequency distribution

Device limitations

- Detection limits
- Calibration range & error

Replicates for error estimates

- Search for data points that are repetitions and estimate std deviation (GC run; reactor channel; stability along one run; run reproduction; synthesis reproduction; …)
- Suggest experiments to improve data quality
- <> Reference catalysts?

… towards fundamental criteria

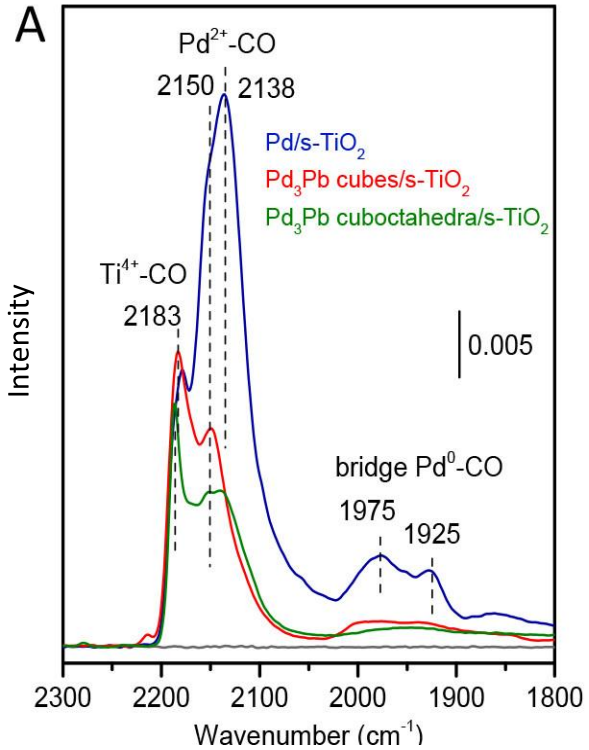
- …

Connecting the tools

Experimentally measured IR Spectrum

- catalyst
- state of catalyst (e.g. prereduced)
- temperature/pressure
- inlet gas composition

- IR Spectrum
- adsorption orientation
- vibrational mode



DFT calculations

- model surface (facet, xyz file)
- state of surface
- temperature/pressure
- inlet gas composition

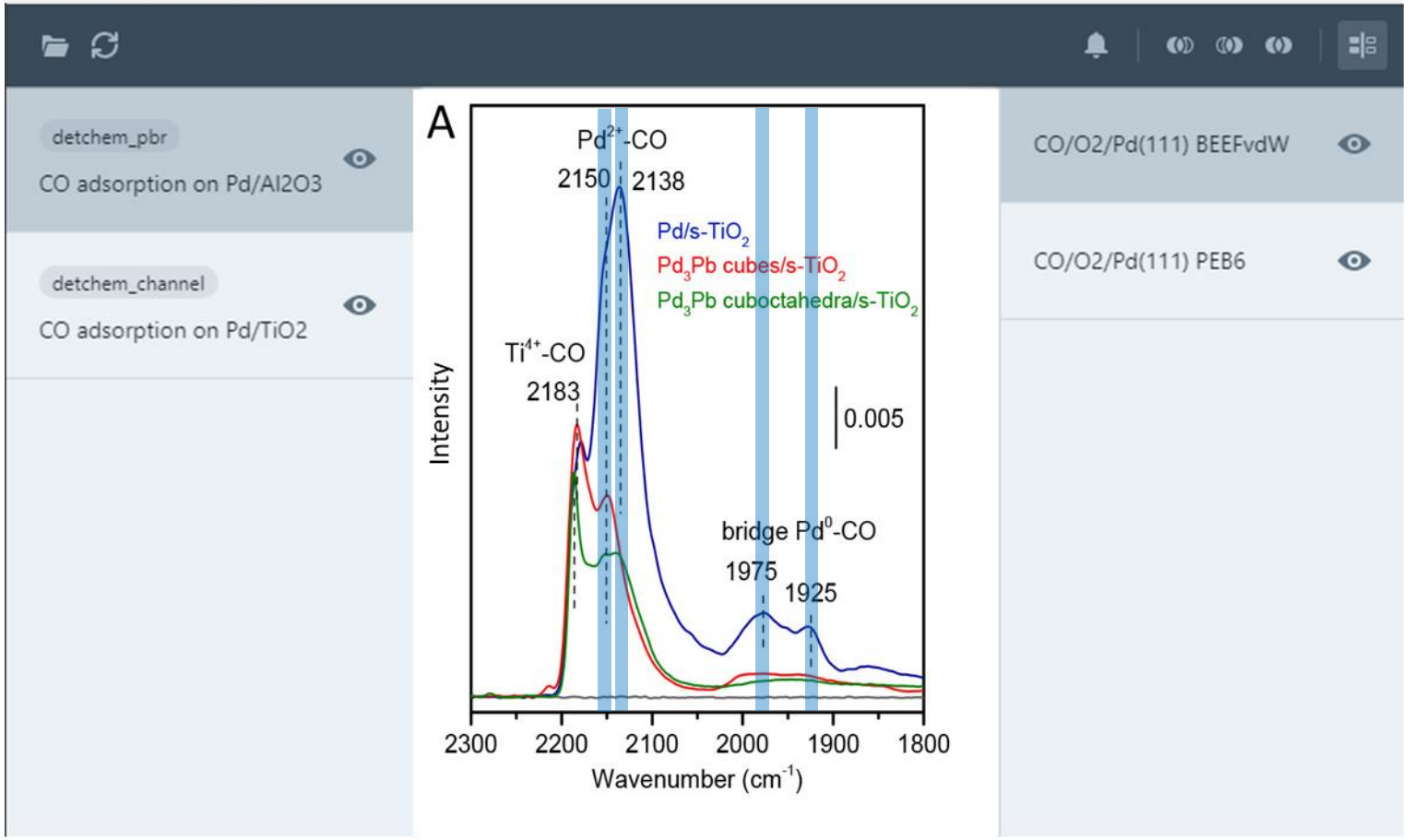
- Vibrational frequencies for CO adsorption
- adsorption orientation
- vibrational mode

V. R. Naina, S. Wang, D. I. Sharapa, M. Zimmermann, M. Hähsler, L. Niebl-Eibenstein, J. Wang, C. Wöll, Y. Wang, S. K. Singh, F. Studt and S. Behrens, *ACS Catal.*, 2021, **11**, 2288–2301.

Connecting the tools

- Experimental data/metadata in .yml and .csv data packages

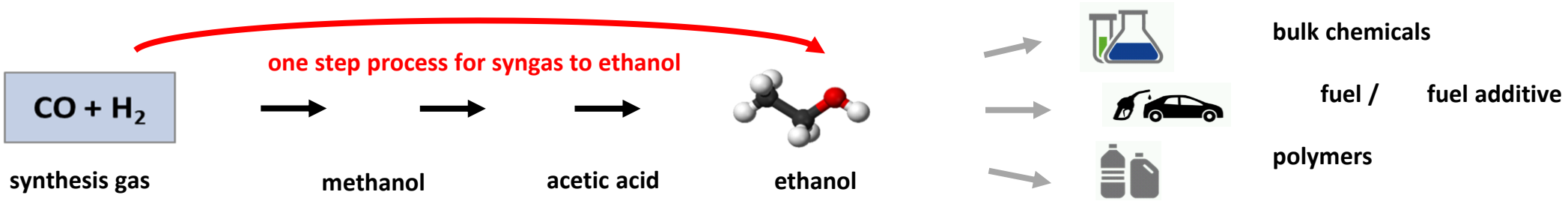
CaRMeN



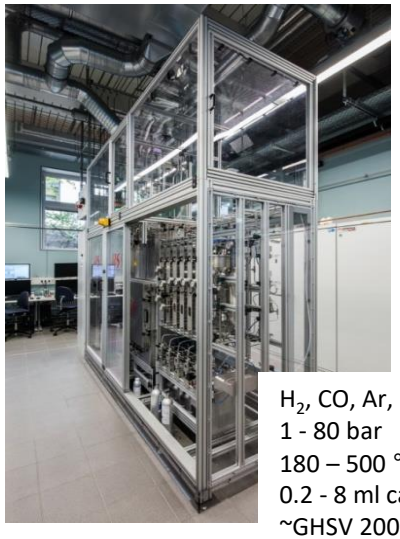
- theoretical data/metadata in the form of lines (if intensity is not known)
- Development of drivers to connect to simulation software (e.g. VASP, TURBOMOLE)

V. R. Naina, S. Wang, D. I. Sharapa, M. Zimmermann, M. Hähsler, L. Niebl-Eibenstein, J. Wang, C. Wöll, Y. Wang, S. K. Singh, F. Studt and S. Behrens, *ACS Catal.*, 2021, **11**, 2288–2301.

Use case: CO hydrogenation on $Rh/X/SiO_2$

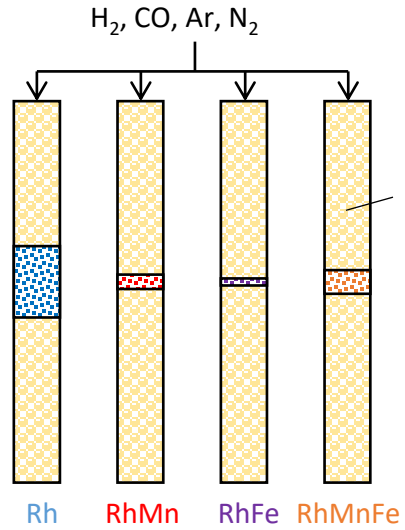


industrially relevant conditions



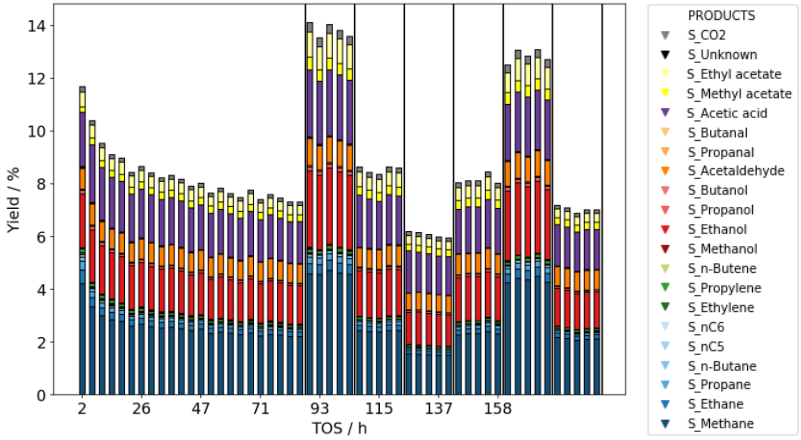
H_2, CO, Ar, N_2, CO_2
1 - 80 bar
180 – 500 °C
0.2 - 8 ml catalyst
~GHSV 200 – 35000 h^{-1}

parallel reactor technology



complex product spectrum

product yields vs. time-one-stream (expl.)



current scope of data

- detailed synthesis protocols
- performance data of catalytic testing
- analytical data: ICP-OES, TPRed, TGA, CHN, FTIR, electron microscopy

P. Preikschas, J. Bauer, K. Knemeyer, R.Naumann d'Alnoncourt, R. Kraehnert, and F. Rosowski, Catal. Sci. Technol. 2021, 11, 5802-5815

Next steps

- **Overall data workflow concept developed applicable to heterogeneous catalysis research**
  - extend the concept (higher complexity, other fields of catalysis)
- **Data tools developed as proof-of-concept with local data**
  - expand the tools (heat and mass transfer, kinetics, design-of-experiments)
  - integrate into local and overarching pilots
- **First link established BasCat, CaRMeN, local pilot**
  - connect with ELNs, other repositories and data sources
- **Use case data measured and documented for heterogeneous catalysis**
  - develop ontologies and meta data standards based on this example
  - transform the data into the new standardized structures and formats
  - use for development and testing of ELN, local pilots, overarching infrastructure

Task Area 4

Fraunhofer FOKUS

TA4: Linked Extensible Infrastructure and Access Management

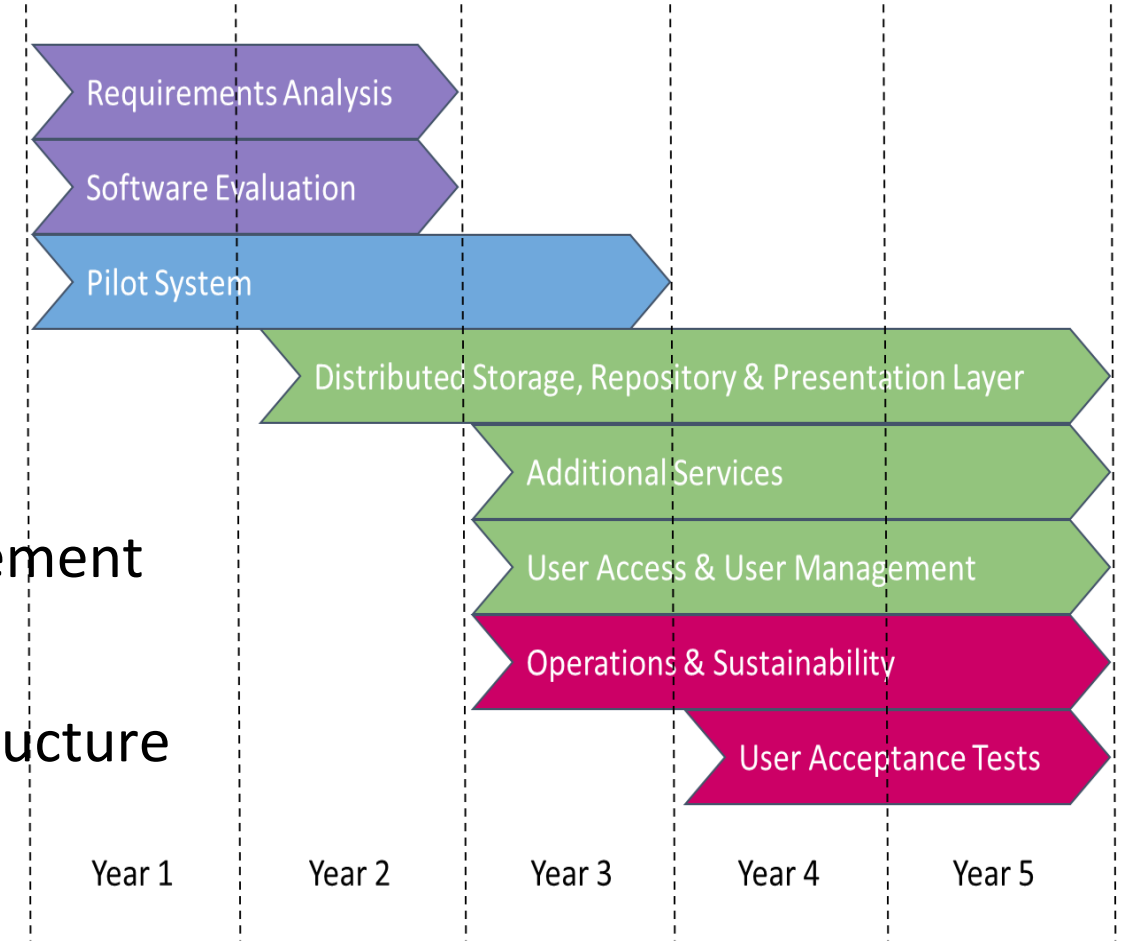
HLRS Höchstleistungsrechenzentrum | Stuttgart

- Requirements analysis
  - Requirements elicitation, based on stakeholder interviews
  - Requirements document
- Next steps
  - Architecture document
  - Software evaluation, pilot system

DOI: 10.5281/zenodo.5882747

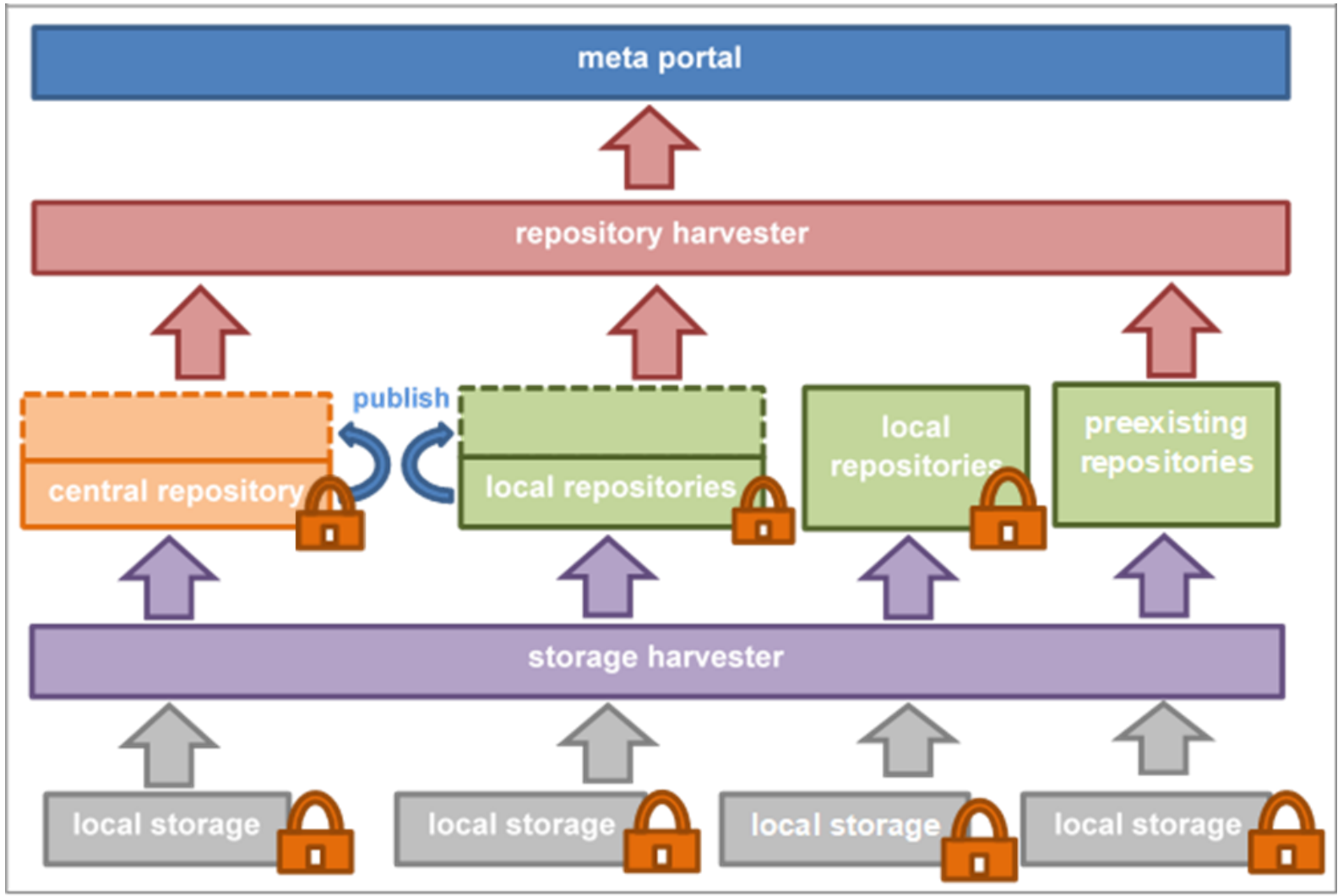
Agenda and Overview

- Measures
  1. **Initial phase**
     1. **Requirements Analysis**
     2. **Software Evaluation**
     3. **Pilot System**
  2. Development phase
     4. Repository Layer
     5. Presentation Layer
     6. Additional Services
     7. User Access & User Management
  3. Evaluation phase
     8. Operations
     9. Sustainability of the Infrastructure
     10. Specification Review &
     11. User Acceptance Tests



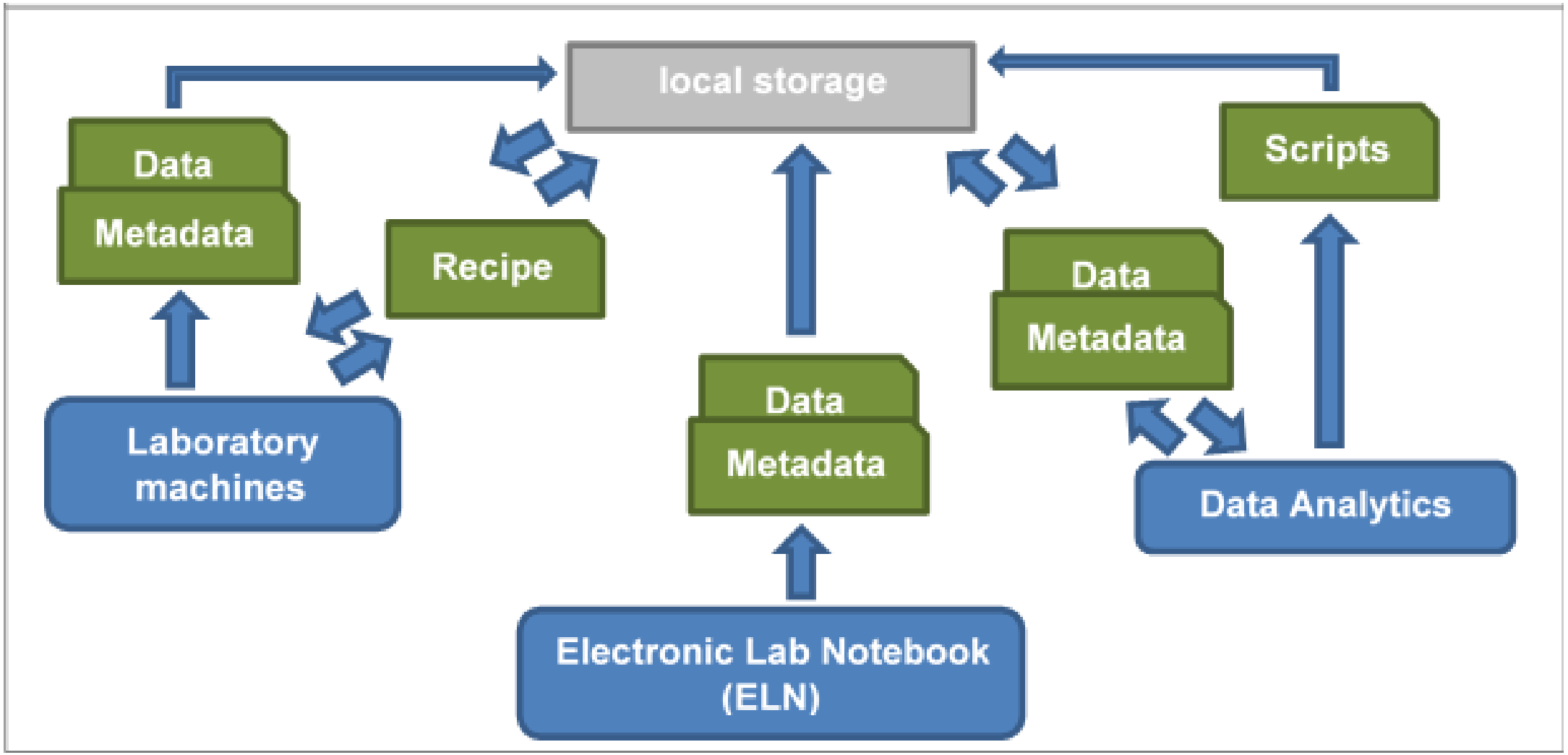
DOI: 10.5281/zenodo.5882747

NFDI4Cat Research Data Infrastructure



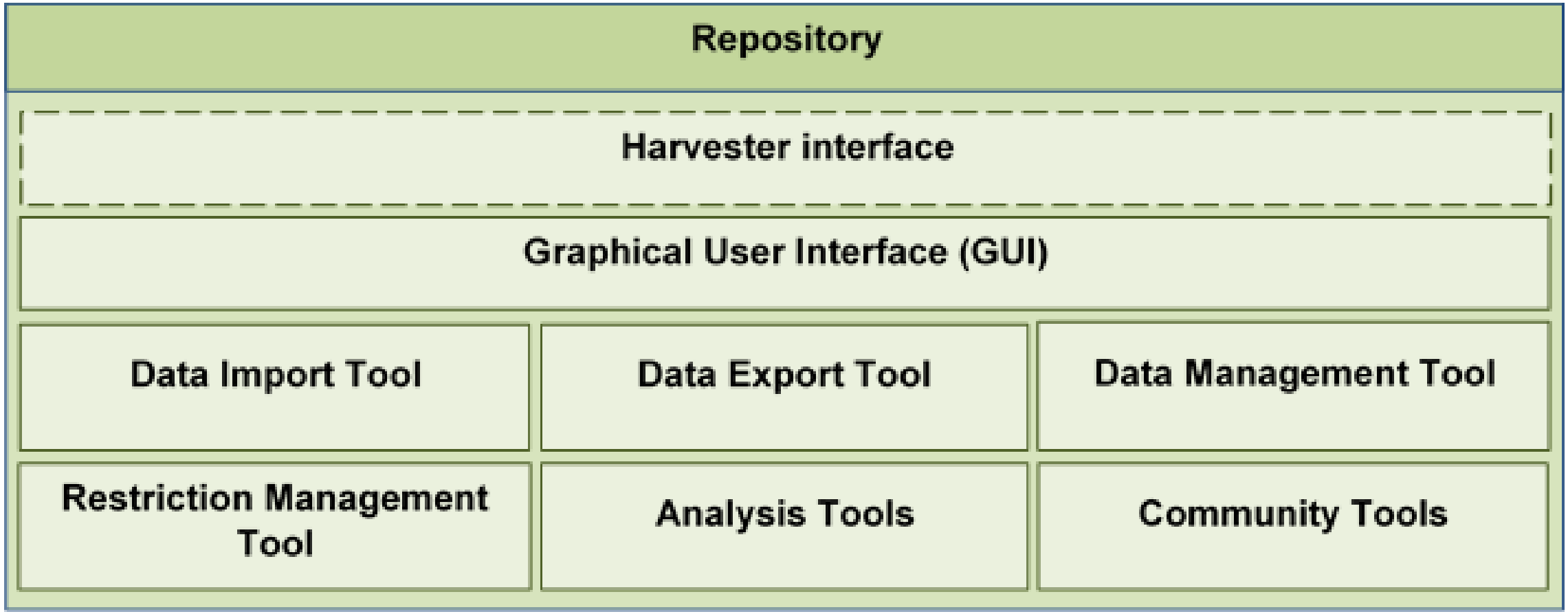
DOI: 10.5281/zenodo.5882747

NFDI4Cat Local Data Storage



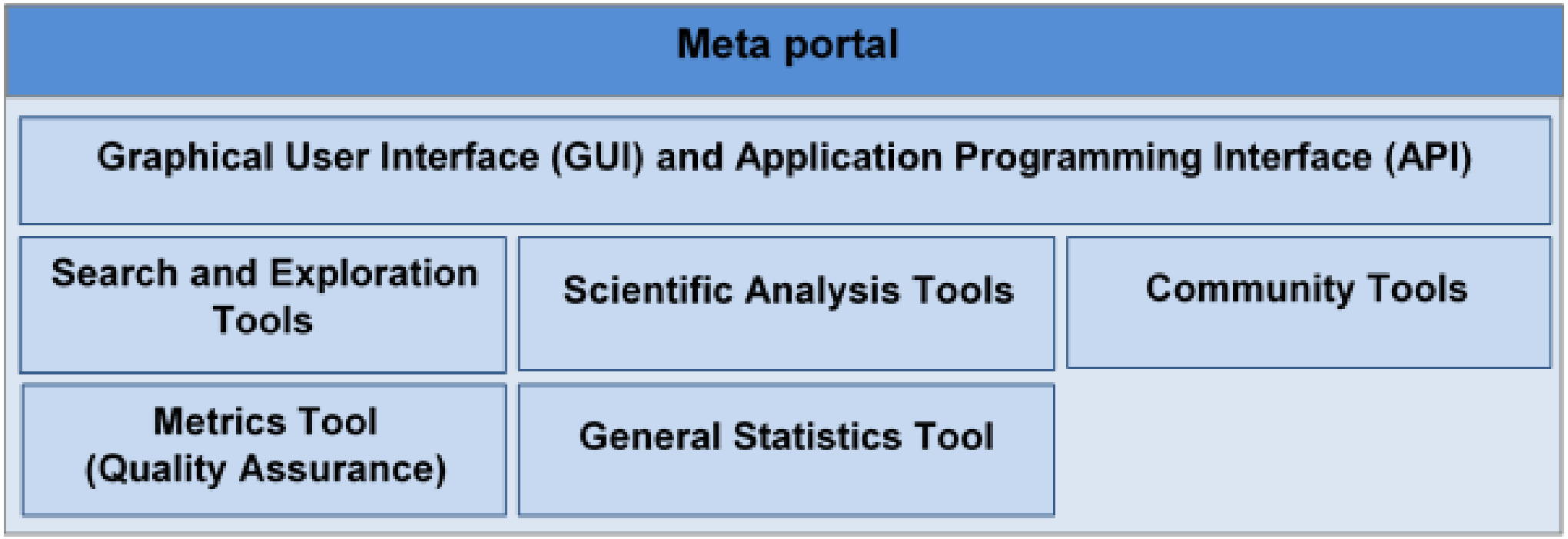
DOI: 10.5281/zenodo.5882747

NFDI4Cat Repository



DOI: 10.5281/zenodo.5882747

NFDI4Cat Portal



DOI: 10.5281/zenodo.5882747

Architecture Document

- concretization of planned architecture
- detailed description of architecture‘s components and interfaces based on the collected requirements from the requirements document and open questions which arised during the feedback round of the requirements analysis

- also agile approach
- similar to requirements document interation of document through partners
- -> extension and improvement of document

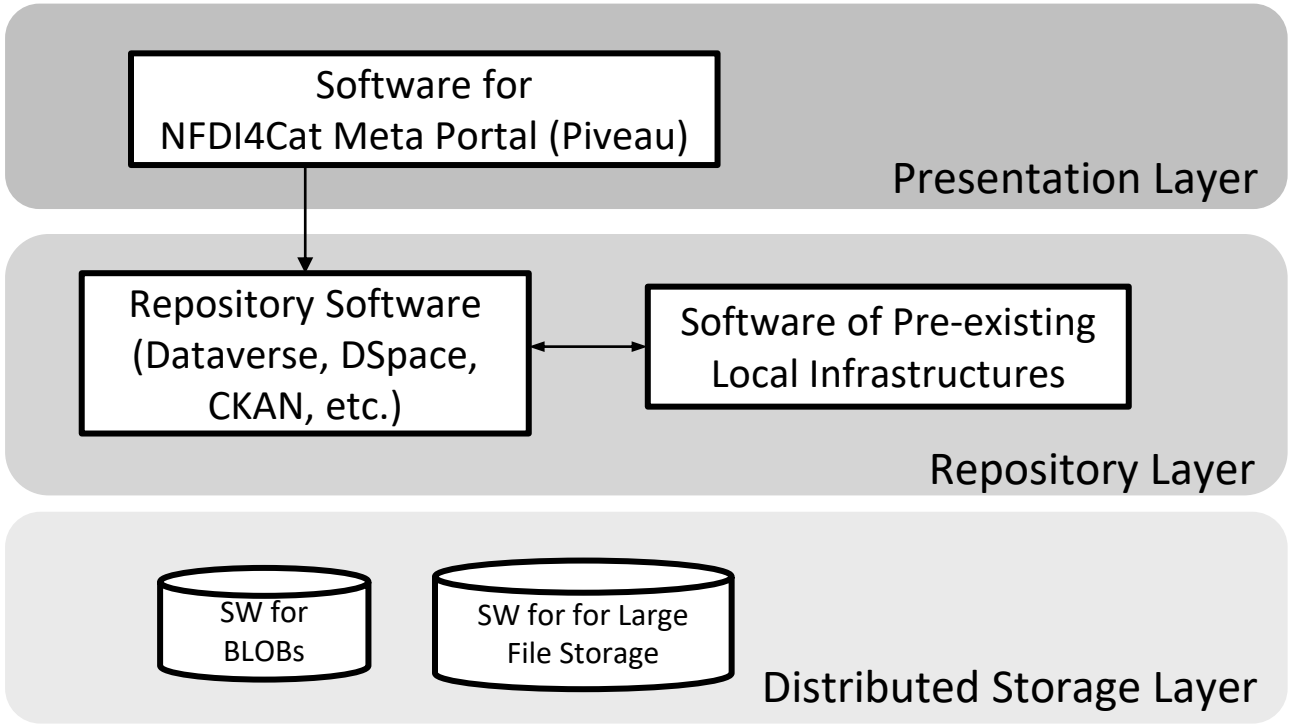
DOI: 10.5281/zenodo.5882747

Measure 2: Software Evaluation

- In the initial phase, HLRS and FOKUS concentrate on evaluating:

Tools for Repositories & Metadata
semantic-logical layer
triple stores
PID generation

Tools for Lab
electronic lab notebooks



DOI: 10.5281/zenodo.5882747

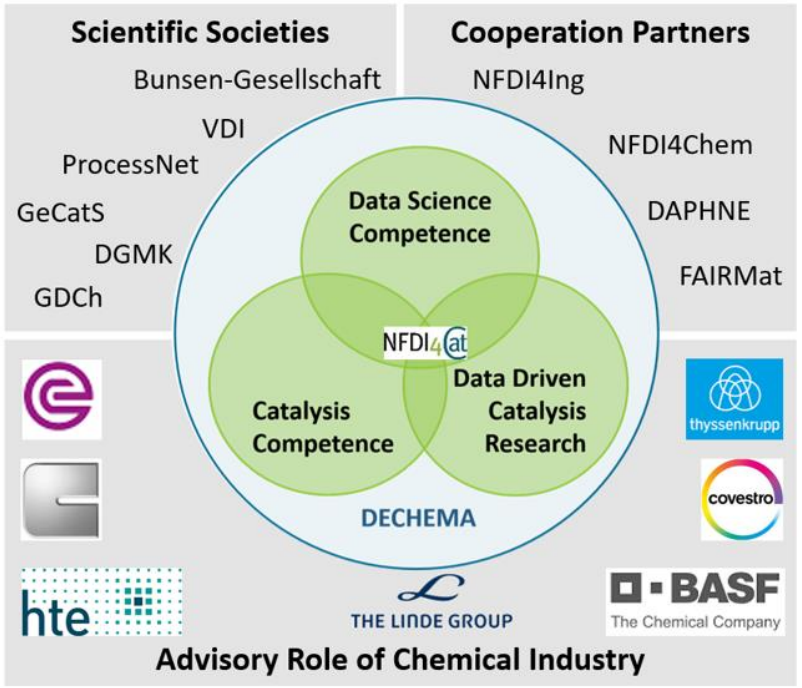
Measure 3: Pilot System (with TA2/3)

- Pilot system at BasCat@TU Berlin
  - Covers a wide range of aspects
  - Will serve as a blueprint
- Further pilot systems at other institutions
  - May (further) develop their own components
  - May reuse components later on

DOI: 10.5281/zenodo.5882747

Finally

- Requirement analysis with User Stories as starting point
- Ontologies are the key for FAIR data principles
- Electronic Laboratory Notebooks are a main data source
- Data Standards are important for research
- Data Analysis needs
  integrated quality management
- Repository Infrastructure and
  Interfaces for Data Access



DOI: 10.5281/zenodo.5882747