# Interoperability and architecture requirements analysis and metadata standardization for a research data infrastructure in catalysis
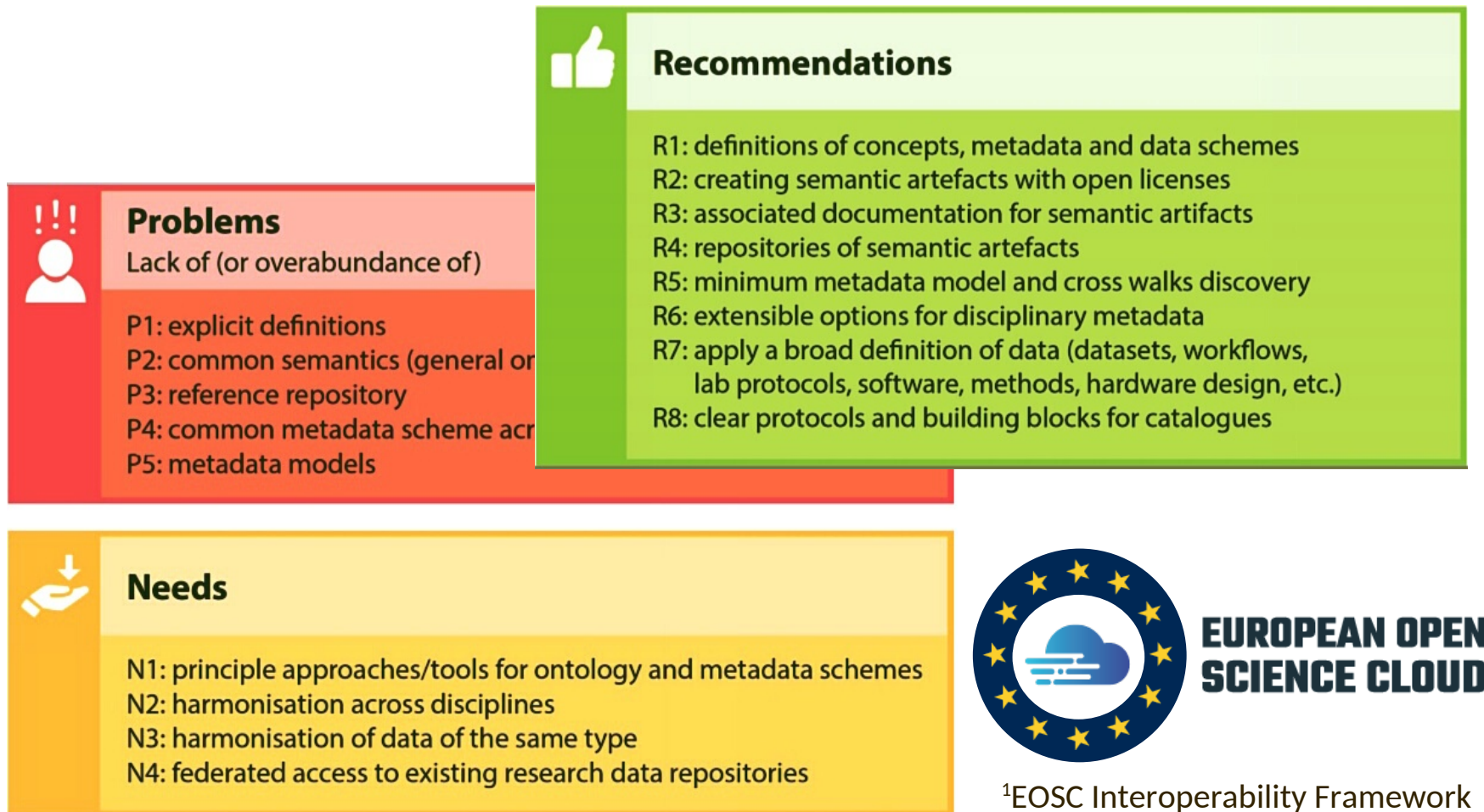
M. T. Horsch,[1,2] T. Petrenko,[1] V. Kushnarenko,[1] B. Schembera,[1]
B. Wentzel,[3] A. Behr,[4] N. Kockmann,[4] S. Schimmler,[3] T. Bönisch[1]

[1]*High Performance Computing Center Stuttgart (HLRS)*, [2]*University of Central Lancashire*, [3]*Fraunhofer Institute for Open Communication Systems (FOKUS)*, [4]*TU Dortmund*
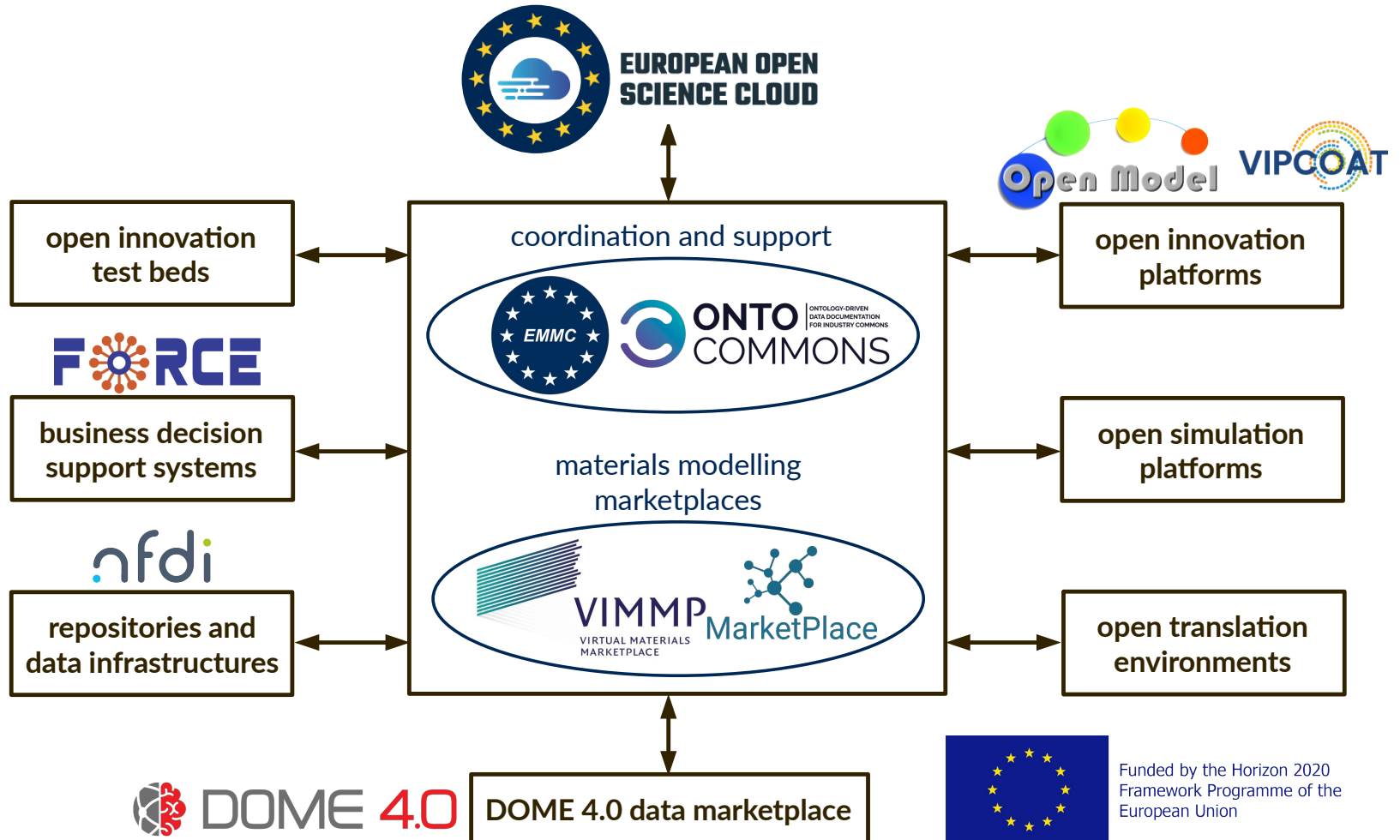
# Semantic interoperability: The European perspective[1]

## Problems
Lack of (or overabundance of)

P1: explicit definitions
P2: common semantics (general or
P3: reference repository
P4: common metadata scheme acr
P5: metadata models

## Recommendations

R1: definitions of concepts, metadata and data schemes
R2: creating semantic artefacts with open licenses
R3: associated documentation for semantic artifacts
R4: repositories of semantic artefacts
R5: minimum metadata model and cross walks discovery
R6: extensible options for disciplinary metadata
R7: apply a broad definition of data (datasets, workflows, lab protocols, software, methods, hardware design, etc.)
R8: clear protocols and building blocks for catalogues

## Needs

N1: principle approaches/tools for ontology and metadata schemes
N2: harmonisation across disciplines
N3: harmonisation of data of the same type
N4: federated access to existing research data repositories

**EUROPEAN OPEN SCIENCE CLOUD**

[1]EOSC Interoperability Framework

# Semantic interoperability: The European perspective

# Semantic interoperability: The European perspective

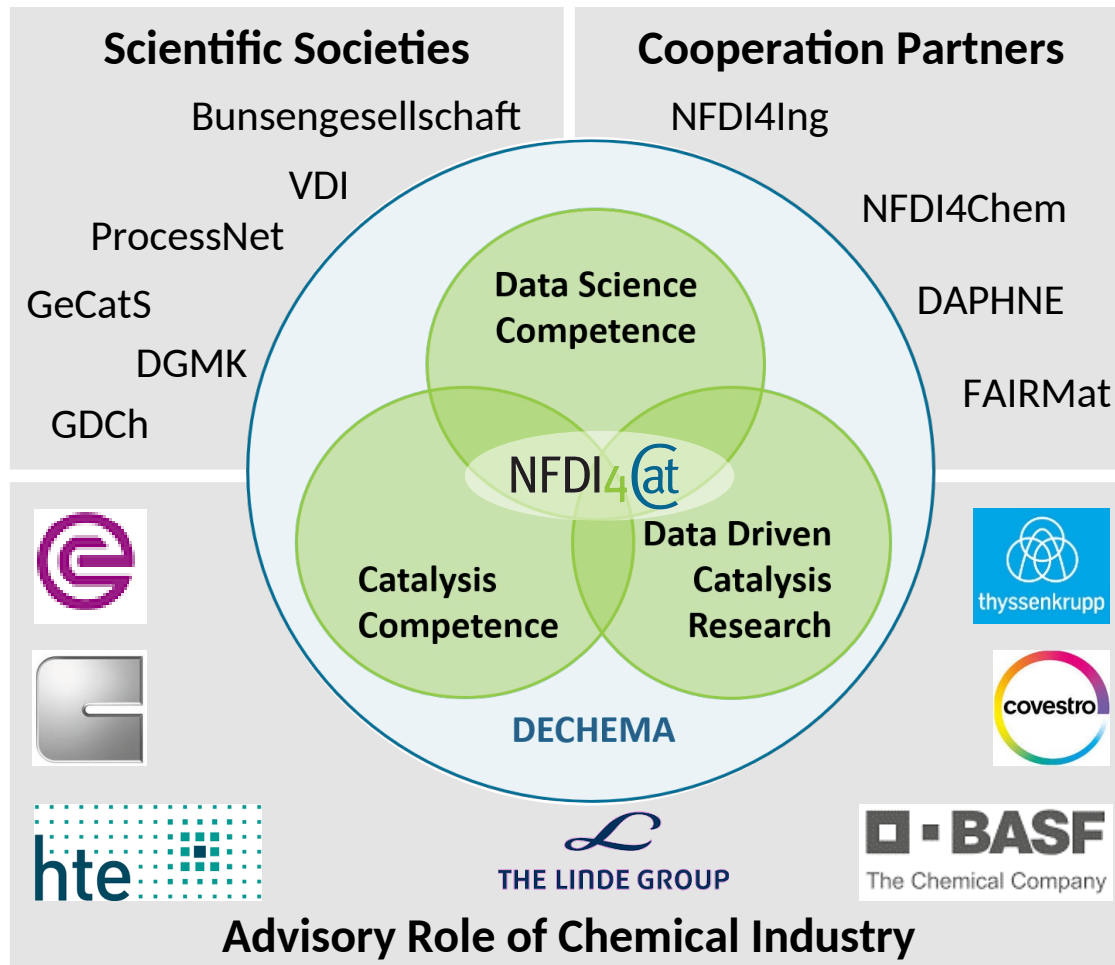**European Materials Modelling Council (EMMC ASBL)**

The non-profit association EMMC ASBL was created in 2019 to ensure the continuity, growth, and sustainability of community activities for modellers, materials data scientists, software owners, materials modelling translators, and manufacturers in Europe. The EMMC regards the **integration of materials modelling and digitalization** as critical for an advancement of industrial process and product design.

https://emmc.eu/

**EMMC Focus Area Digitalization**

In computational engineering, digitalization encompasses aspects of representing, managing, accessing, and utilizing digital information about products, components, materials, their behaviour, and their processing.

# National Research Data Infrastructure (NFDI) in Germany

**Scientific Societies**

Bunsengesellschaft

VDI

ProcessNet

GeCatS

DGMK

GDCh

**Cooperation Partners**

NFDI4Ing

NFDI4Chem

DAPHNE

FAIRMat

Data Science Competence

NFDI4Cat

Data Driven Catalysis Research

Catalysis Competence

DECHEMA

thyssenkrupp

covestro

THE LINDE GROUP

BASF
The Chemical Company

hte

**Advisory Role of Chemical Industry**

**National Research Data Infrastructure (NFDI)** funding programme and association (NFDI e.V.) comprising 19 disciplinary consortia and infrastructures

**National Research Data Infrastructure for Catalysis-Related Sciences**

# Metadata standards and the digital value chain in catalysis
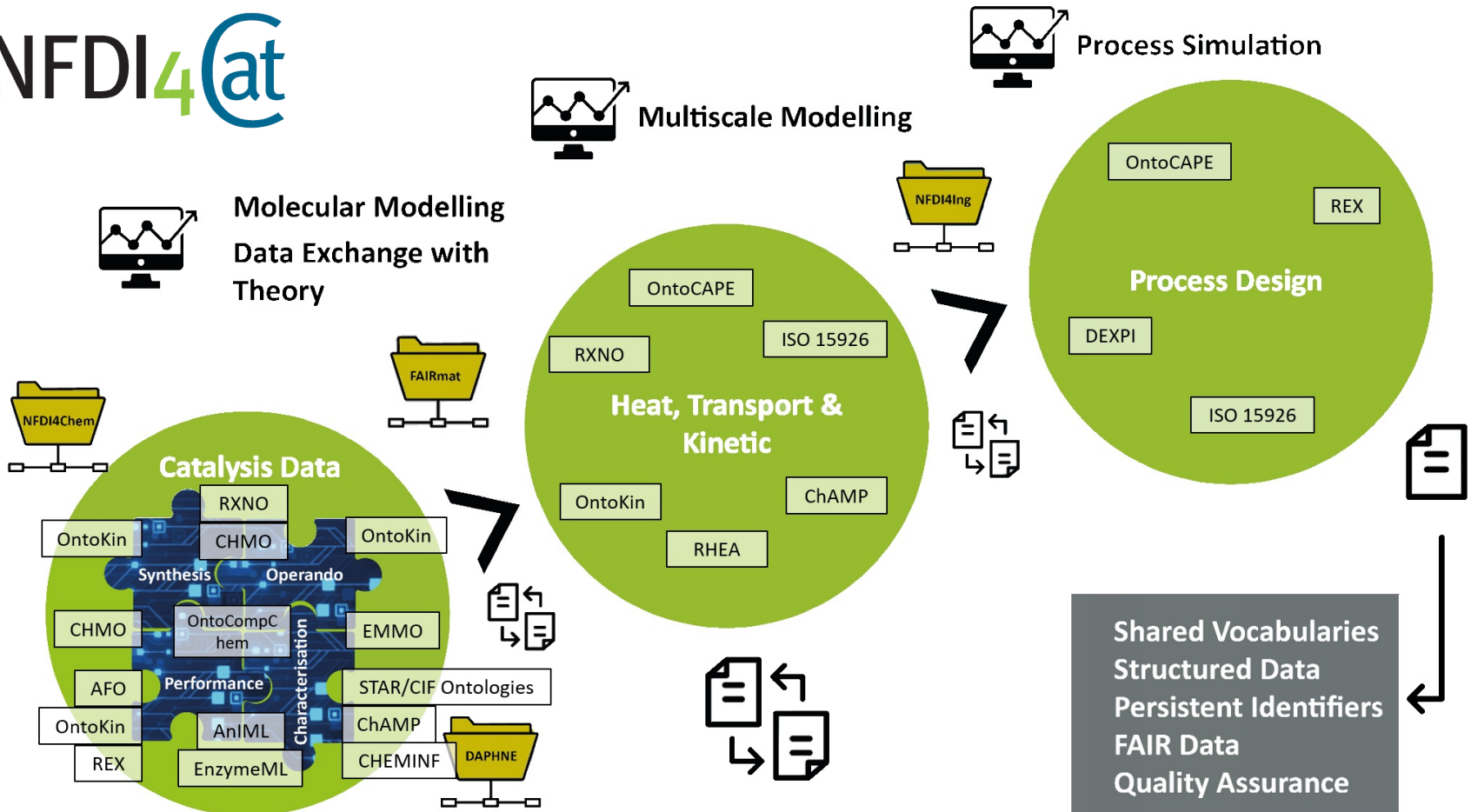
# NFDI4Cat architecture
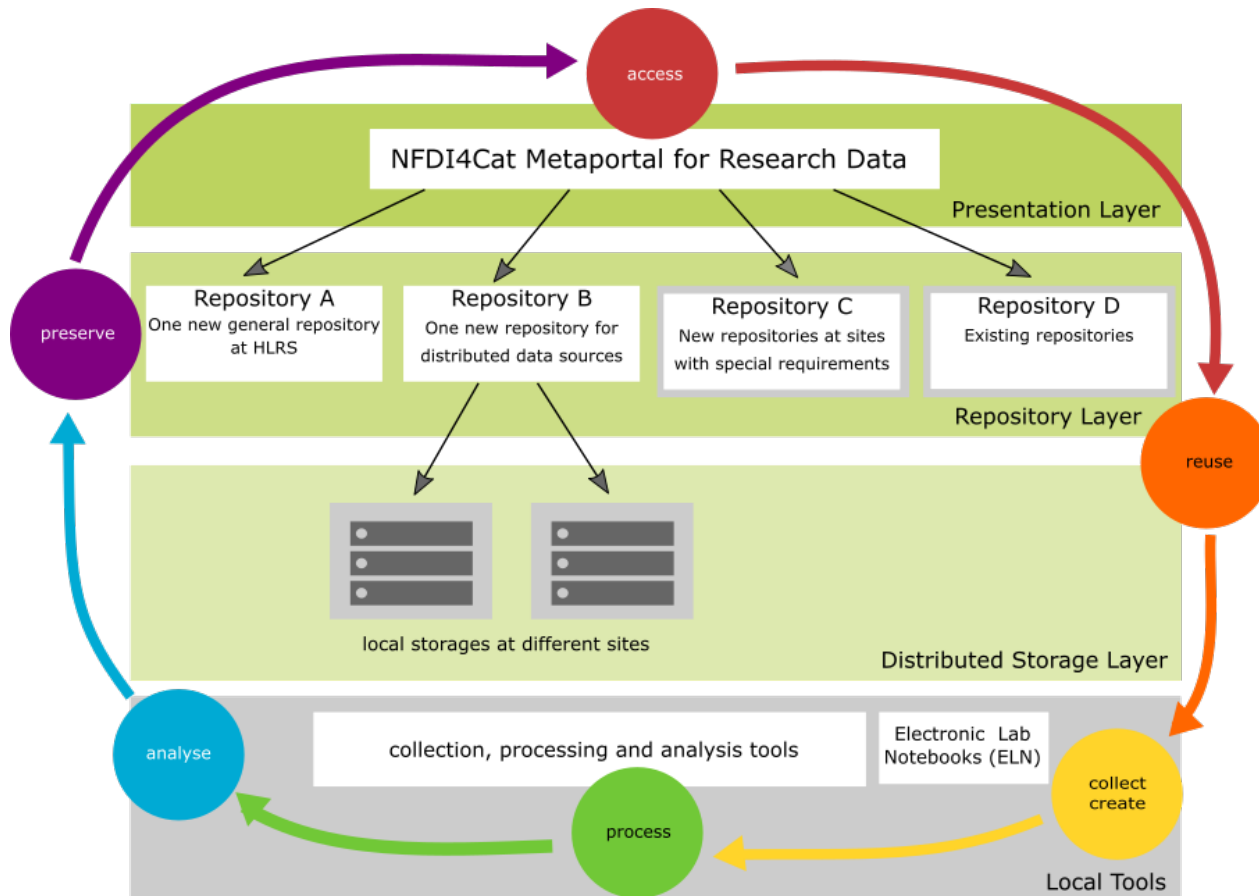


**Platforms to be designed:**

- Overarching central infrastructure, including the NFDI4Cat metaportal

- Bespoke local repositories

- Generic solution for local repositories

# Competency questions for metadata standardization

| | | |
|---|---|---|
| | ...) (TUDO-NK). | |
| 25 | Gib mir die in der Literatur bekannten Konzentration der Substrate (Katalysatoren, Liganden, Additive, ...) zur Homogenkatalyse mit Platin Pt (Rhodium Rh, Ruthenium Ru, ...) (TUDO-NK). | TUDO-NK |
| 26 | Gib mir die in der Literatur bekannten Ausbeuten (Selektivitäten) der Produkte (Katalysatoren, Liganden, Additive, ...) zur Homogenkatalyse mit Platin Pt (Rhodium Rh, Ruthenium Ru, ...) (TUDO-NK). | TUDO-NK |
| 26 | Gib mir die in der Literatur bekannten Aktivitäten und Standzeiten (Recycle-Strategie, Recycling-Raten, ...) der Homogenkatalysatoren Platin Pt (Rhodium Rh, Ruthenium Ru, ...) (TUDO-NK). | TUDO-NK |
| 27 | Welche Rührreaktoren werden in der Homogenkatalyse eingesetzt (Inhalt, Durchmesser, Druck, Temperatur, Rührerart, Aufbau, ...)? (TUDO-NK) | TUDO-NK |
| 28 | Gib mir die Prozessschritte zur Produktaufbereitung aus der Homogenkatalyse (TUDO-NK). | TUDO-NK |
| 29 | Nenne mir die Messtechniken zur Analyse der Konzentration (Phasenverteilung, Tropfengrößenverteilung, ...) im Rührreaktor / Autoklav bei der Homogenkatalyse (TUDO-NK). | TUDO-NK |
| 30 | Gib mir die Katalysatorträgermaterialien für Platin Pt (Rhodium Rh, Ruthenium Ru, ...) (TUDO-NK). | TUDO-NK |
| 31 | Gib mir die Zusammensetzung der Katalysatorträgermaterialien mit Aluminiumoxid (Ceroxid, Siliziumoxid, ...) (TUDO-NK). | TUDO-NK |
| 32 | Gib mir die Porenstruktur und Porosität (Kristallformation, ...) der | TUDO-NK |

# Procurement of architecture requirements

**Hierarchical structure of requirements in agile software engineering**
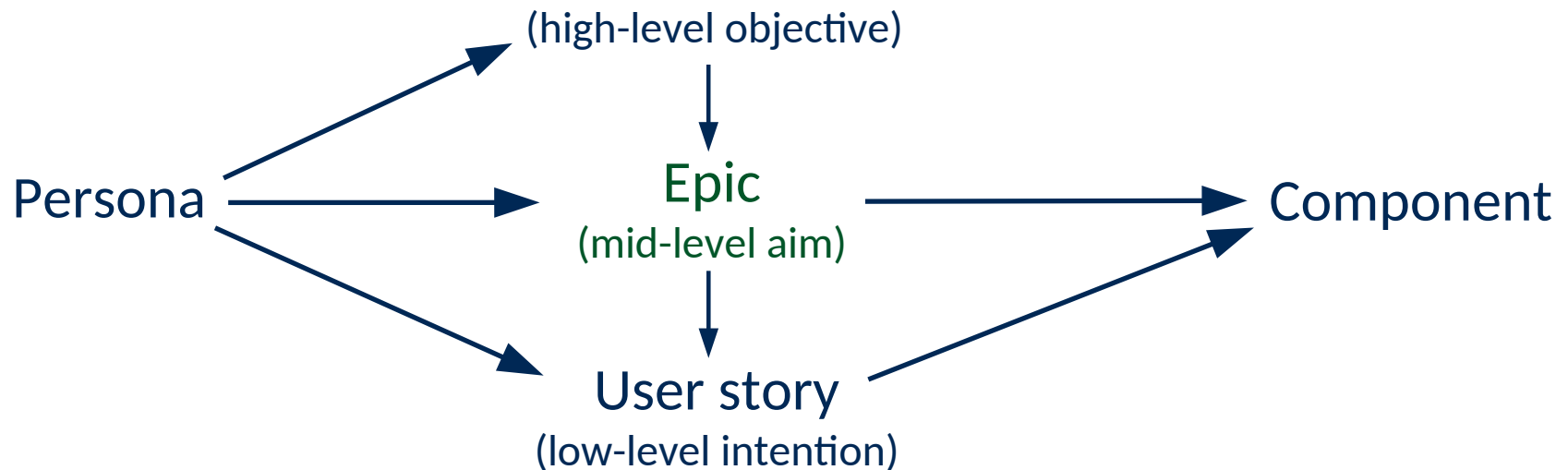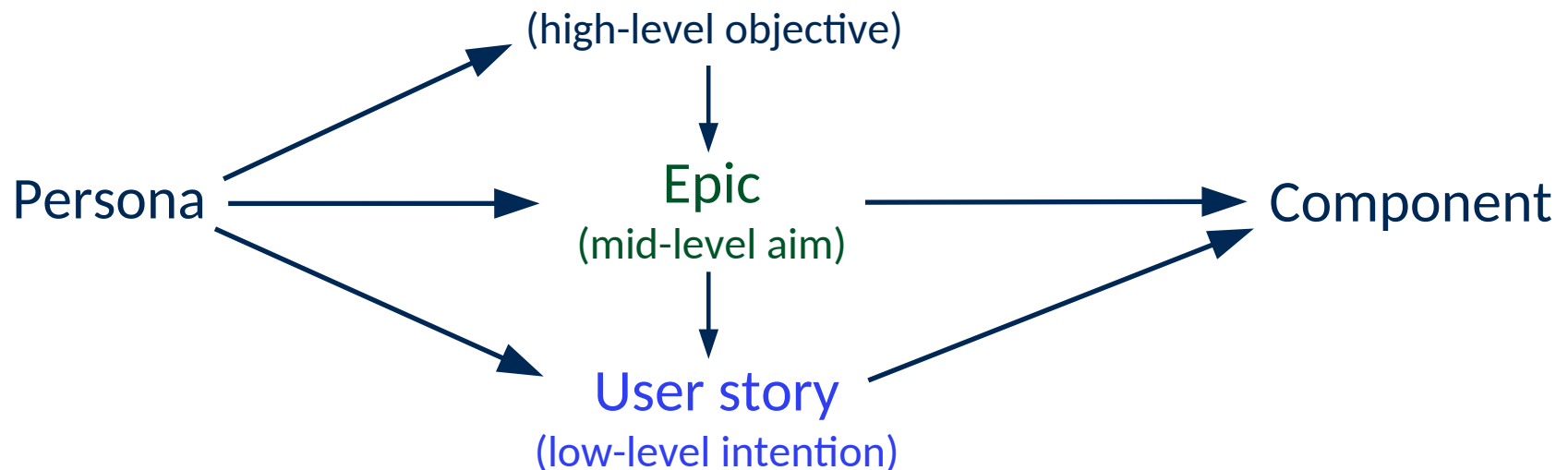


**Epic ui??:epic?**
**As** Domain Scientist,
**I aim at** employing the local XXX knowledge base, which is presently being constructed on the basis of XXX, as a local repository within a federated NFDI4Cat architecture,
**subordinate to my objective** to create synergies between NFDI4Cat and other ongoing work.

# Procurement of architecture requirements

**Hierarchical structure of requirements in agile software engineering**

(high-level objective)

Persona → Epic (mid-level aim) → Component

User story (low-level intention)

**User story** ui??:epic?-us?
**As** Domain Scientist
**I intend to** have an agreement on minimum requirements for query languages, if possible including property-graph based approaches, to be implemented by NFDI4Cat Components.
**subordinate to my aim** of building and operating a local repository within NFDI4Cat.

# Procurement of architecture requirements

interview 3:
$CO_2$ methanation

interview 2:
synchrotron QEXAFS
experiments

interview 1:
photocatalytic
water splitting
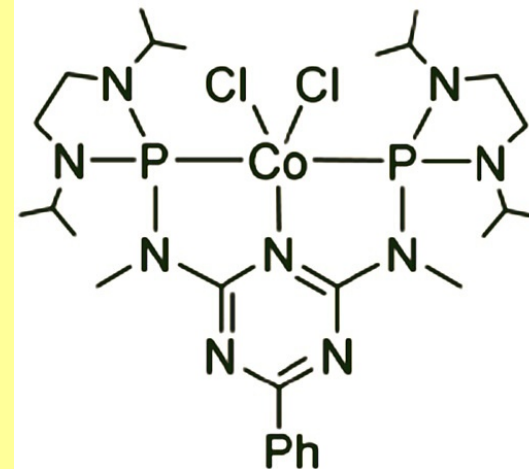
interview 4:
biocatalysis &
multiphase flow

interview 9:
photocatalytic
$CO_2$ reduction

interview 5:
$CO_2$ electrolysis

interview 6:
metal-organic
catalysts

interview 7:
nanoparticle &
SILP catalysts

interview 8:
selective NOx
reduction

# Procurement of architecture requirements



interview 3:
CO$_2$ methanation

interview 2:
synchrotron QEXAFS
experiments

interview 1:
photocatalytic
water splitting

interview 4:
biocatalysis &
multiphase flow

view 9:
atalytic
duction

interview 5:
CO$_2$ electrolysis

interview 6:
metal-organic
catalysts

na
SILP catalysts

w 8:
NOx
tion

# Procurement of architecture requirements



interview 11: Coupled in-situ and offline analytics

interview 10: Fischer-Tropsch process

interview 12: Syn-gas-to-ethanol catalyst characterization

interview 17: FURTHRmind

interview 13: Syngas-to-ethanol reaction

interview 16: Water-gas shift reaction

interview 14: RDM for enzyme activity data

interview 15: Oligomerization reaction

interview 11:

interview 10:
...scher-Tropsch process

interview 17:
FURTHRmind

interview 16:
Water-gas shift reaction
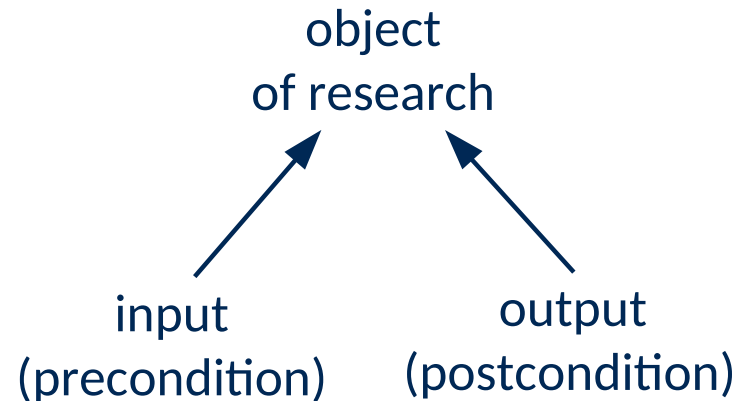
interview 15:
Oligomerization reaction
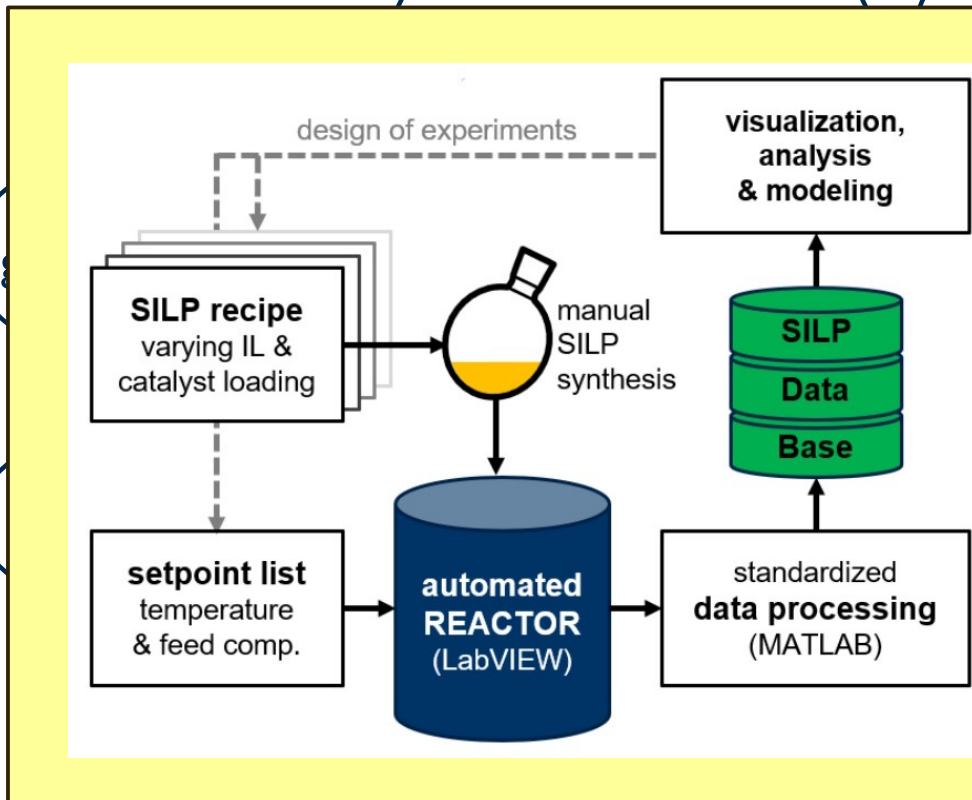
# Requirements for data provenance descriptions

Interviews, 30 minutes each, were conducted with prospective users (including the groups mentioned before).

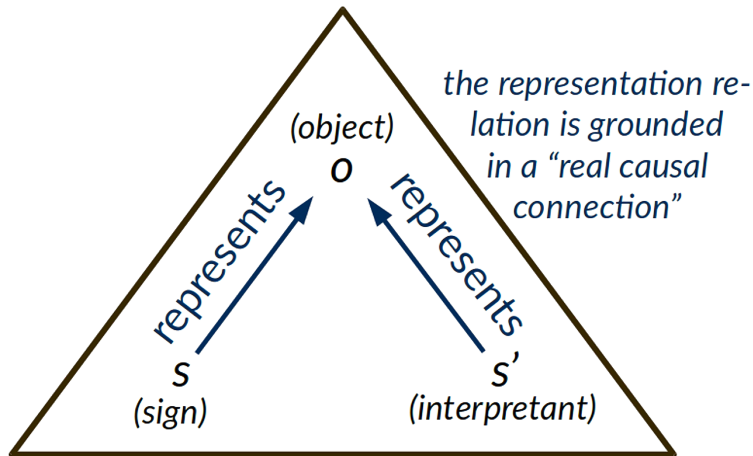For each research step, we jointly identified:

- **input**, *i.e.*, all that needs to be present in advance (incuding equipment);
- **output**, *i.e.*, that which is generated as an outcome of the research step.

Pre- and postcondition share a referent: The **object of research**.
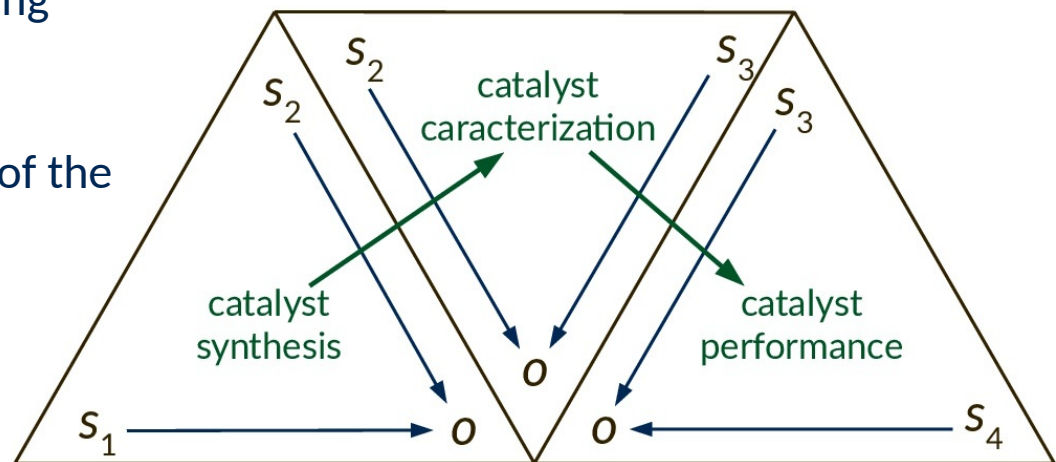
object
of research

input
(precondition)

output
(postcondition)



$S_2$

$S_2$

$S_3$

$S_3$

catalyst
caracterization

catalyst
synthesis

$O$

catalyst
performance

$S_1$ ———→ $O$      $O$ ←——— $S_4$

**Peircean semiotics**



*the representation re-lation is grounded in a "real causal connection"*

(object)
O

represents

represents

S
(sign)

S'
(interpretant)

*the semiosis, a process by which a new representamen, the interpretant, is created*



C. S. Peirce

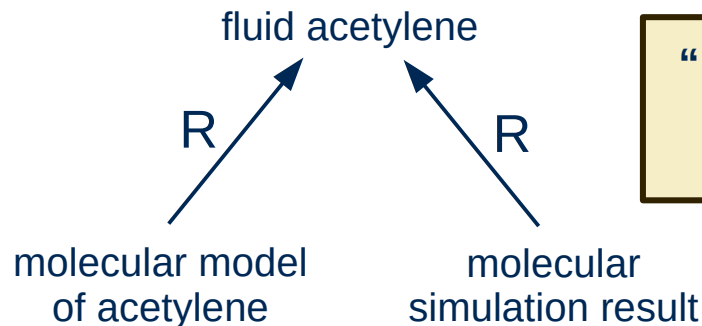**Elementary Multiperspective Material Ontology[1, 2]**

**1) Taxonomy:**
   Conceptual hierarchy (subclass relation)

**2) Mereotopology:**
   Spatiotemporal parthood and connectivity

**3) Semiotics:**
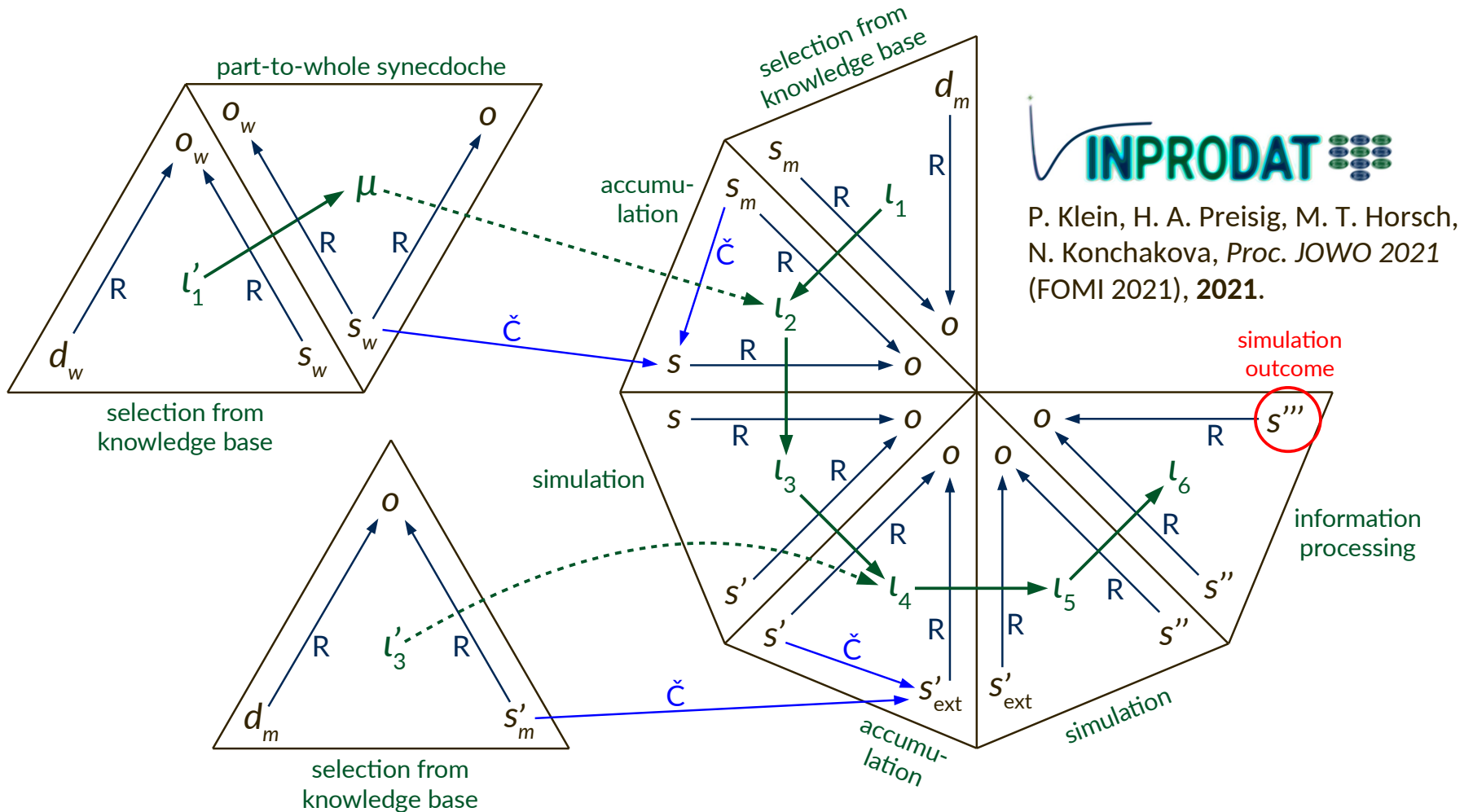   Representation of physical entities by signs

fluid acetylene

R          R

molecular model
of acetylene

molecular
simulation result

**"represents"** or **"is sign for"** is here abbreviated by **R**

[1]J. F. Morgado, E. Ghedini, G. Goldbeck, *et al.*, *Proc. SeDiT 2020*, **2020**.
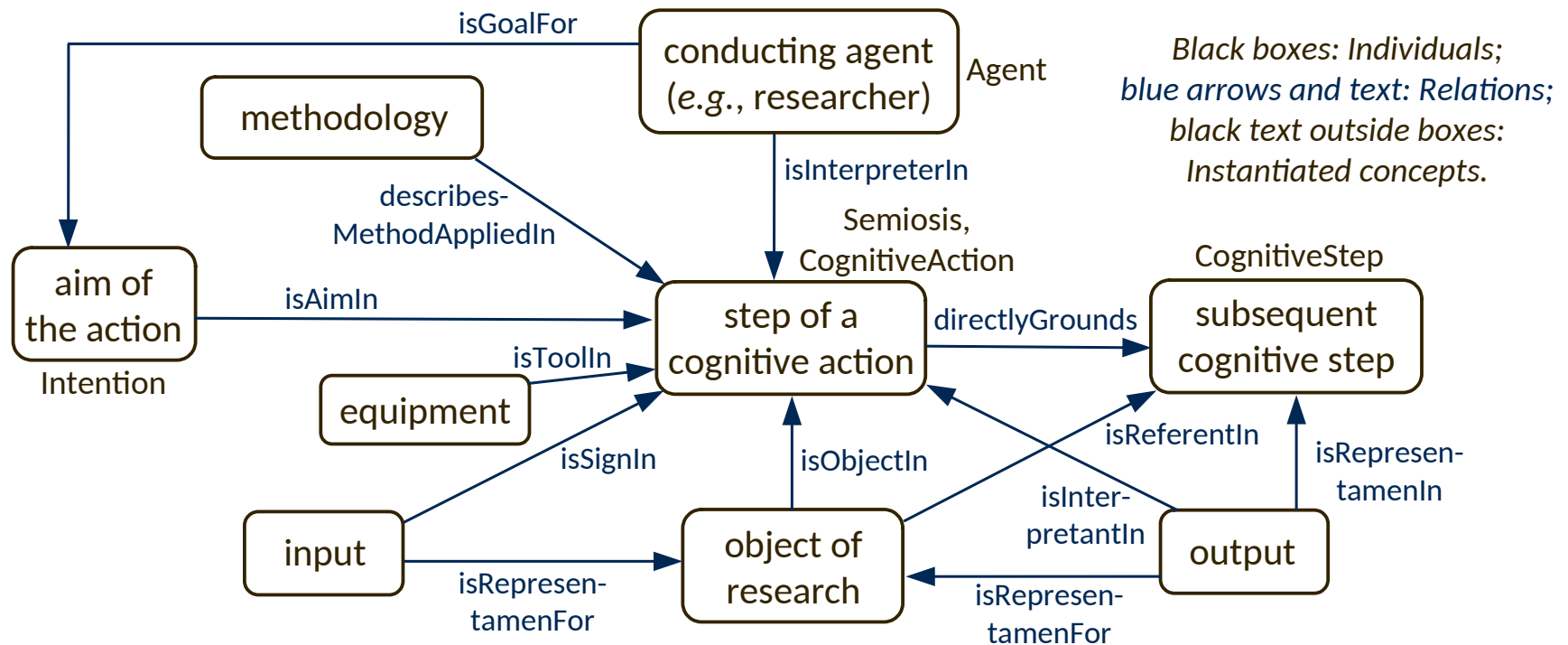[2]H. Preisig, T. Hagelien, J. Friis, *et al.*, *Proc. WCCM-ECCOMAS 2020*, **2021**.

# Application to disciplinary provenance metadata



P. Klein, H. A. Preisig, M. T. Horsch, N. Konchakova, *Proc. JOWO 2021* (FOMI 2021), **2021**.

# Schema for a step in a research workflow

**PIMS interoperability infrastructure[1, 2, 3] (PIMS-II) knowledge graph template**

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

M. T. Horsch,[1, 2] T. Petrenko,[1] V. Kushnarenko,[1] B. Schembera,[1] B. Wentzel,[3] A. Behr,[4] N. Kockmann,[4] S. Schimmler,[3] T. Bönisch[1]

[1]*High Performance Computing Center Stuttgart (HLRS),* [2]*University of Central Lancashire,* [3]*Fraunhofer Institute for Open Communication Systems (FOKUS),* [4]*TU Dortmund*