Limits and solutions for ontological representation of chemical reactions

reactions modelling with LARAsuite

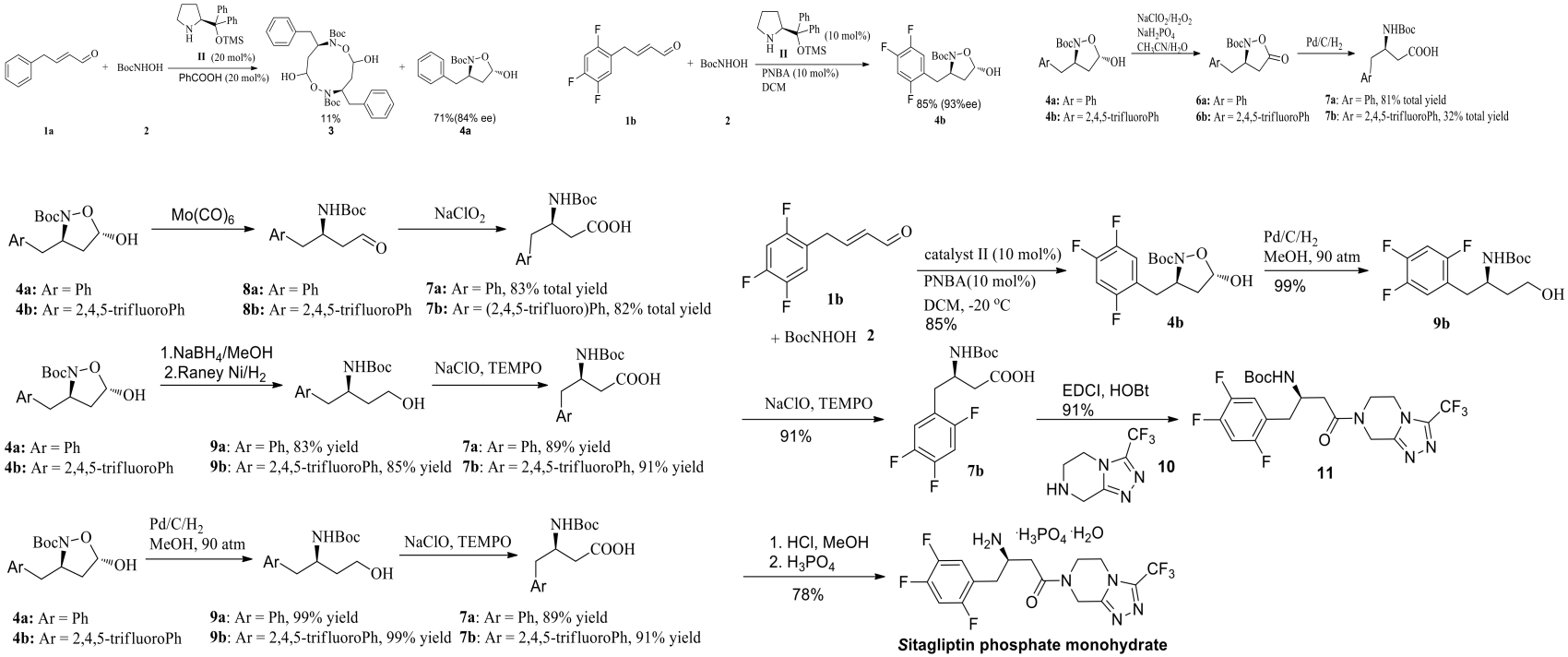
mark doerr and stefan born

university greifswald, tu-berlin

2025-11-11

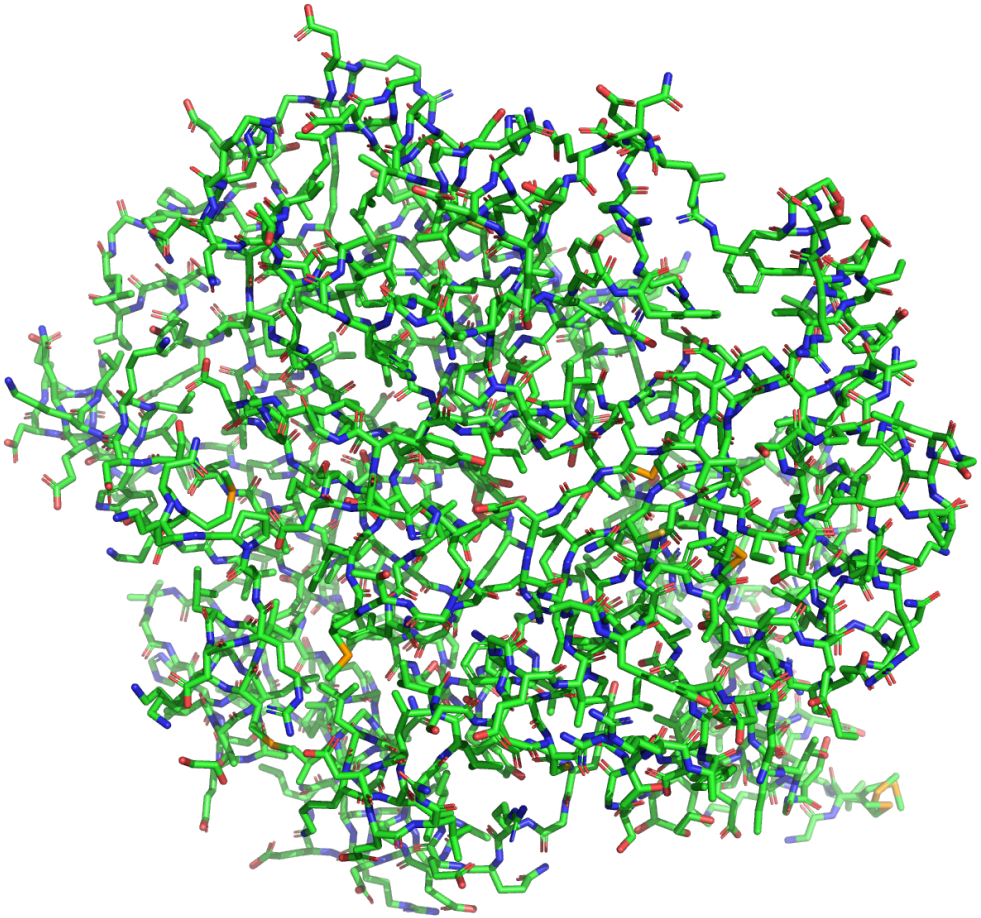
use case 1: drug development

Sitagliptin chemical synthesis

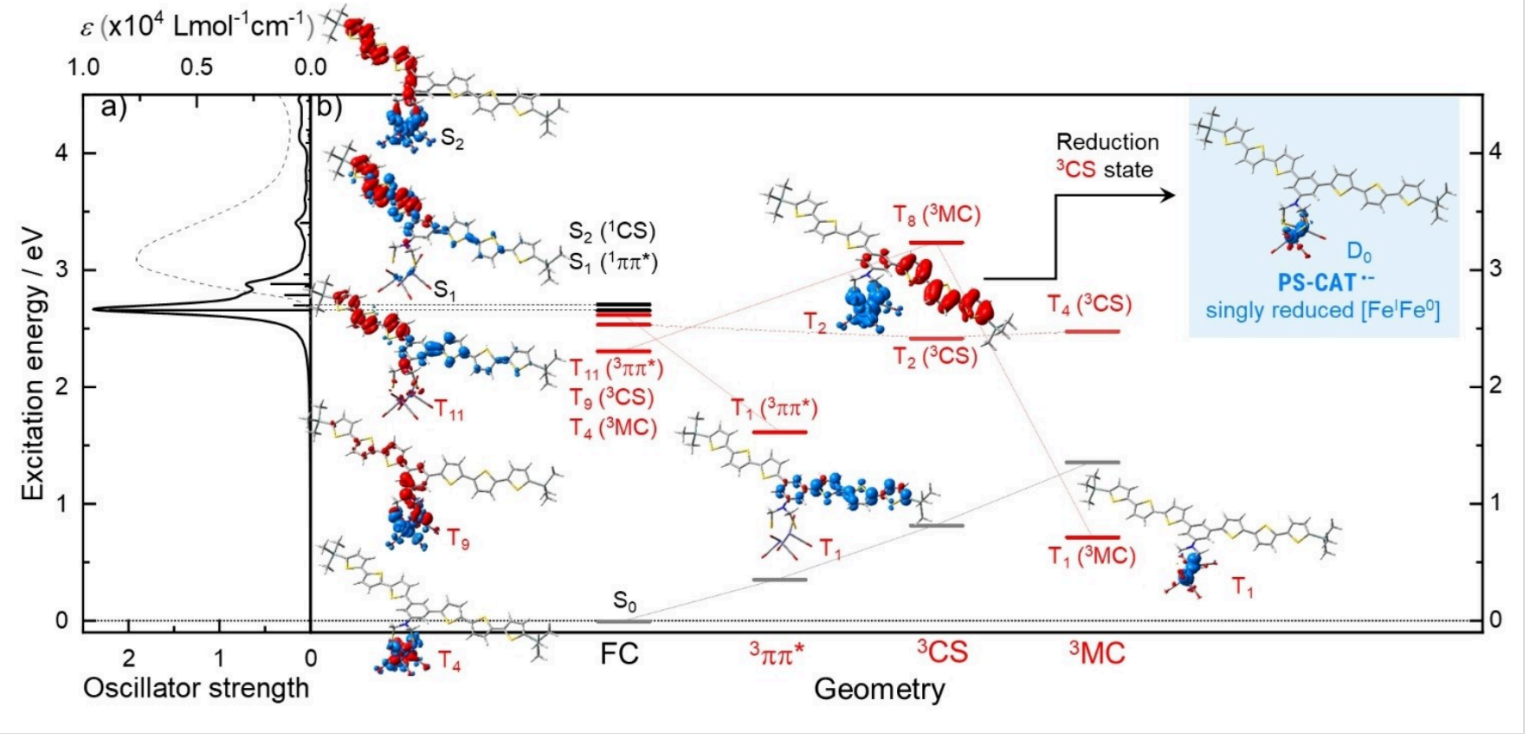


source: Gao et al., Molecules 2018, 23(6), 1440; https://doi.org/10.3390/molecules23061440

case 2: enzyme engineering



case 3: react. mechanism (hydrogenase mimic)



Source: Chizuru Kasahara, ..., **Wolfgang Weigand**, et al ChemCatChem 2024 https://doi.org/10.1002/cctc.202400247

reaction classification and identification

- determine the type of reaction (e.g., redox, acid-base, catalytic, …) based on reactants and products.
- identify all reactions involving a specific molecule as a reactant or product.
- retrieve the stoichiometry of reactants and products for a given reaction.
- what is the catalytic agent for a specific reaction?

reaction mechanism and dynamics

- what are the key steps in the reaction mechanism?
- what are the intermediate species formed during the reaction?
- what are the electronic states of the reactants, intermediates and products during the reaction?
- how do changes in conditions (e.g., temperature, pressure) affect the reaction rate?
- how are charges changed during the reaction?

competency questions 2

- which structural features of the catalyst interact with the reactants to facilitate the reaction?
- what are the thermodynamic properties (e.g., enthalpy, entropy) associated with the reaction?
- what is the activation energy of the reaction?

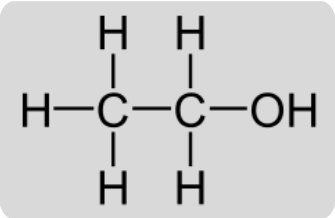
reaction context and environment

- in which solvent does the reaction occur ?
- what is the role of the solvent in the reaction ?
- at which temperature, pressure, and pH does the reaction take place ?

modelling substances

substance class

abstract substance (with general properties, like molar mass, molecular formula , LEWIS-structure ...)



example: ethanol, glucose, benzaldehyde

substance instance

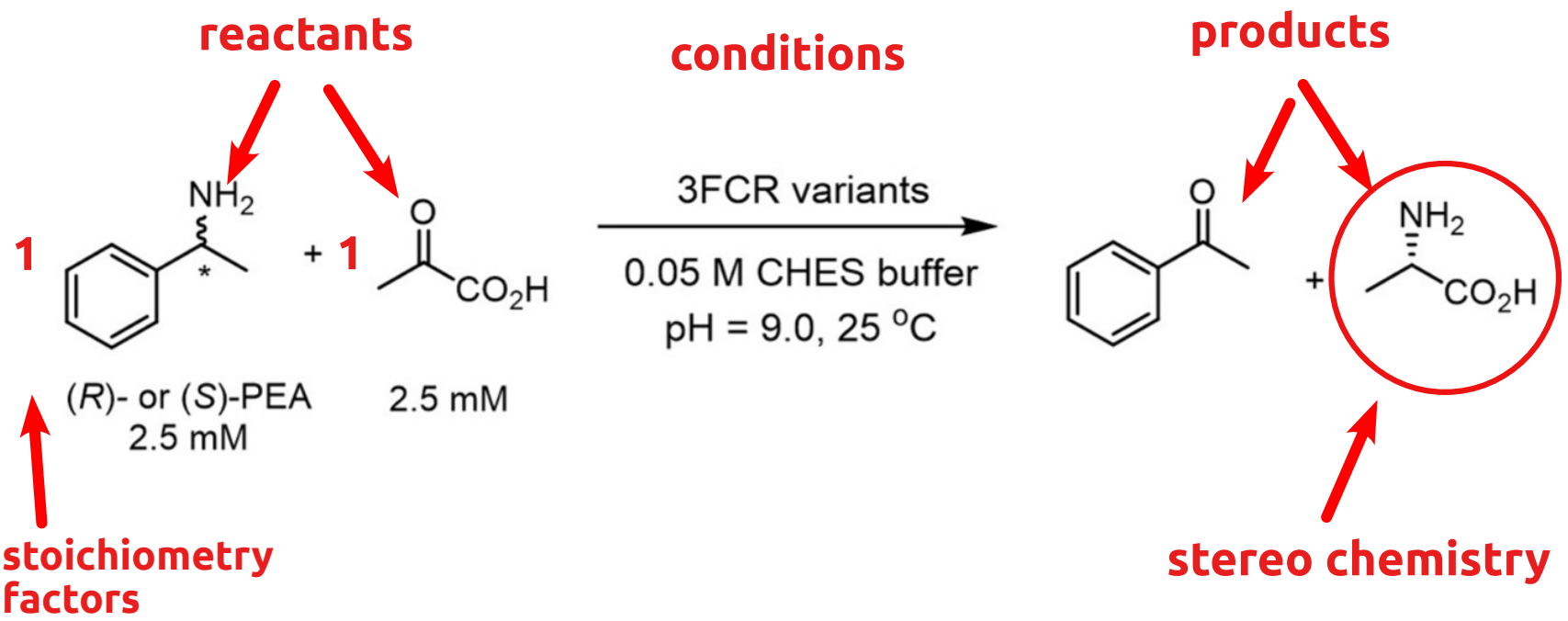
specific substance in a bottle in a lab (with lot number, concentration, purity, ...)

example: ethanol solution 70% v/v,
supplier: Merck
lot #12345

modelling reactions

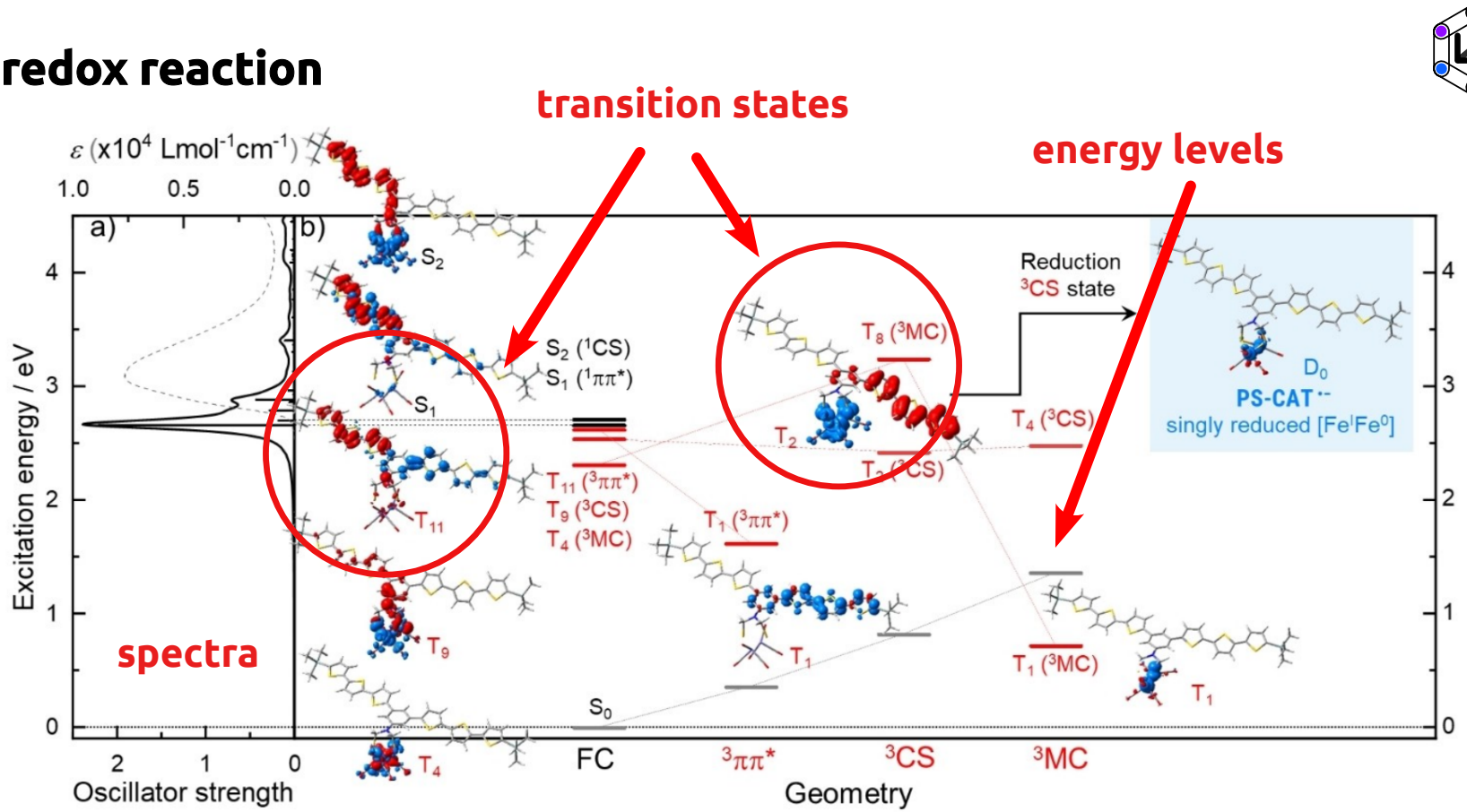
transaminase reaction

class - instance relation



source: Yu-Fei Ao et. al, Angewandte Chemie 2023.
https://doi.org/10.1002/anie.202301660

modelling mechanism



generic (reaction) workflow language

- define reaction steps and their sequence
- specify conditions and parameters for each step
- capture data and results at each stage
- enable reproducibility and automation of reaction processes

workflow language - pythonLAB

pythonLAB (https://opensourcelab/pythonLab)

```python
incubation_duration = 120
cont1 = self.containers[0]
# move to incubator
self.robot_arm.move(cont1, source_loc=cont1.pos, target_loc=self.incubator)
# incubate
self.incubator.incubate(cont1, duration=incubation_duration, temperature=310)
# move to reader
self.robot_arm.move(cont1, source_loc=cont1.pos, target_loc=self.reader)
# make absorbance measurement
self.reader.single_read(cont1, wavelengths=[600, 660], temperature=305)
# move back to hotel
self.robot_arm.move(cont1, source_loc=cont1.pos, target_loc=self.hotel)
```

LARA

simplified syntax for process design

pythonLAB 2.0

simplified syntax for process design

binding to Ontologies (RDF representation)

pythonLAB 2.0

simplified syntax for process design

binding to Ontologies (RDF representation)

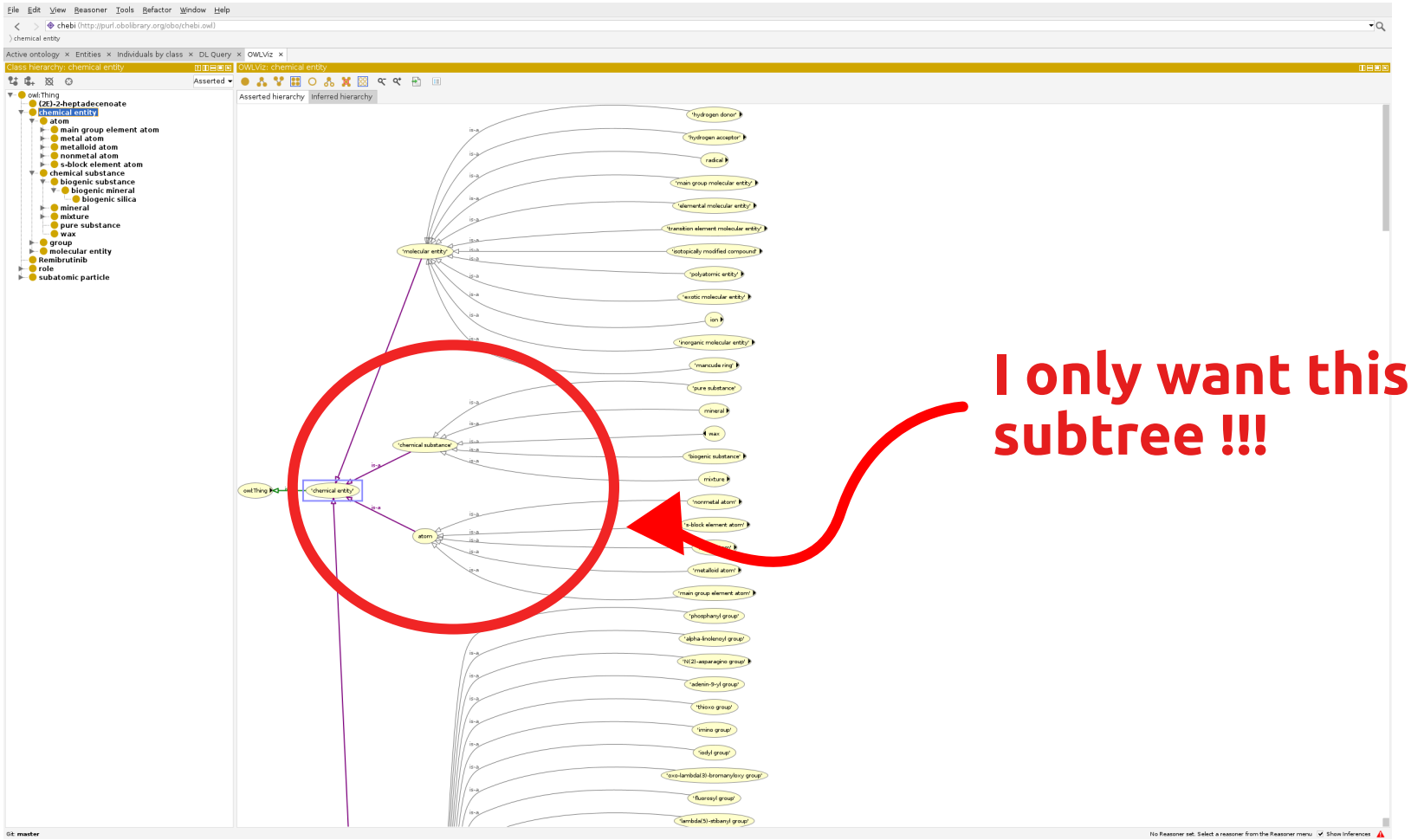
workflow graphs as RDF graphs

Master Student Max Schmidt, TU Darmstadt

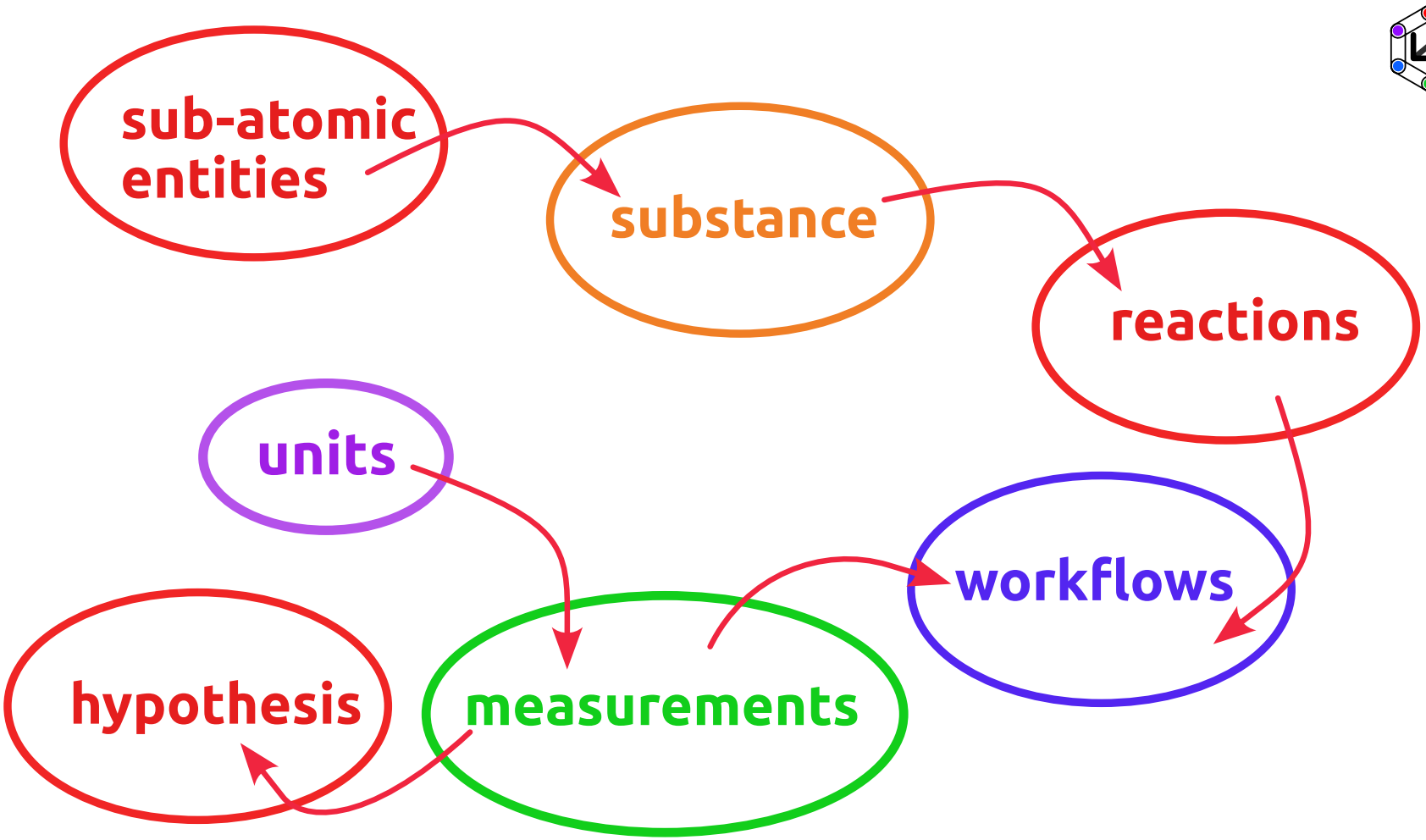
recommendations for ontologies design

- cover mainly T-Box (terminology) aspects
- use **OCCAM**'s razor principle: (minimalistic, based on first principles)
- modular design with clear interfaces
- check for logic consistency (reasoning)
- provide clear documentation and examples

rather than ...



linked micro-ontologies



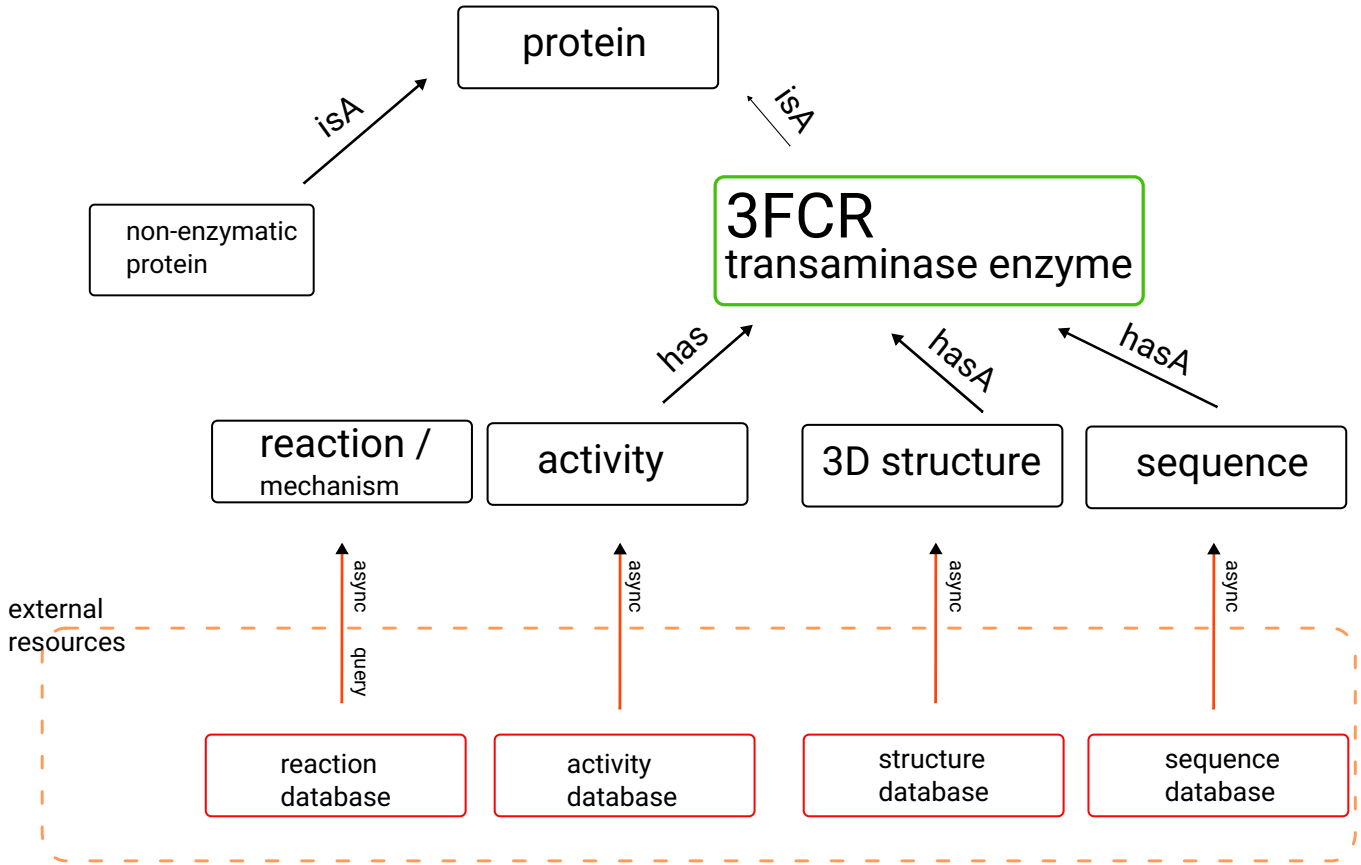
LARA's hybrid approach

combine ontologies with relational databases and file-based data storage

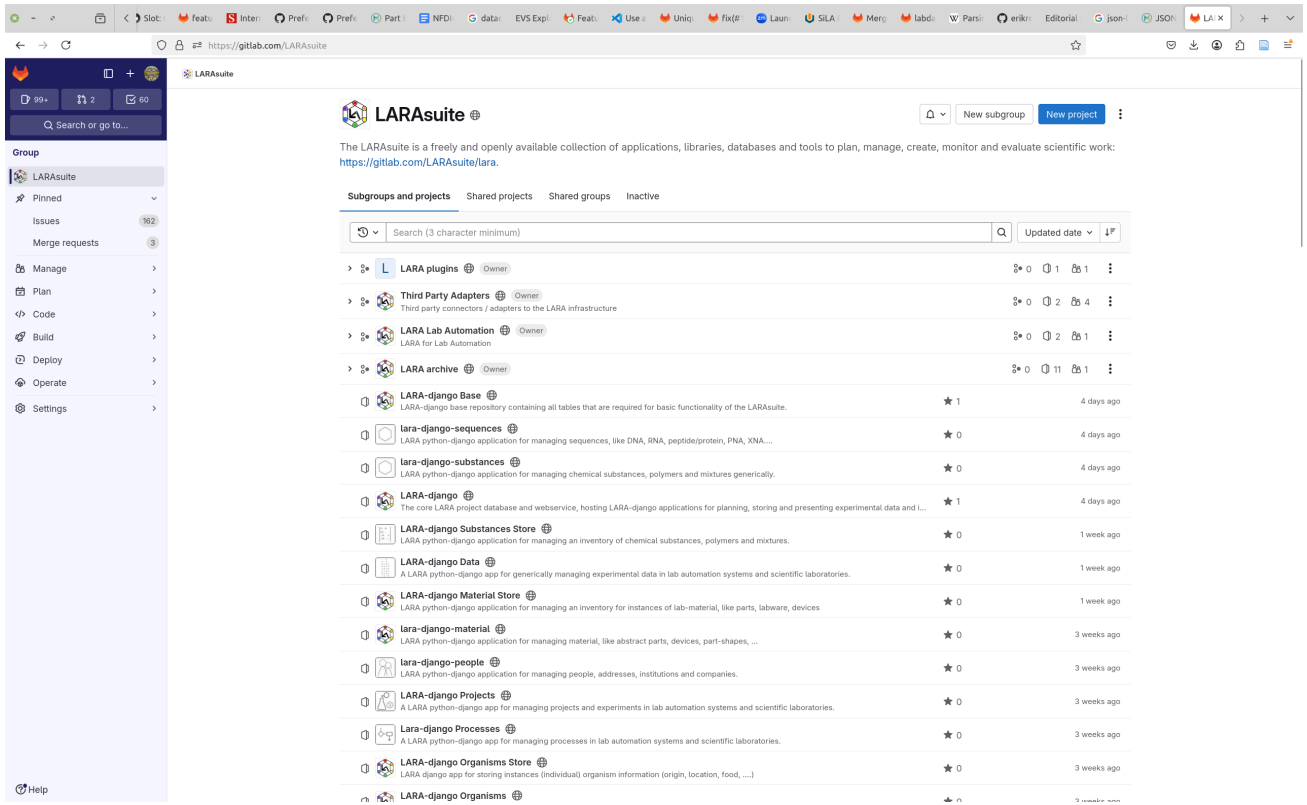
- use ontologies for high-level concepts and relationships
- use relational databases for structured data storage and querying
- use file-based storage for large datasets and unstructured data
- integrate the three layers for comprehensive data management

hybrid reasoning

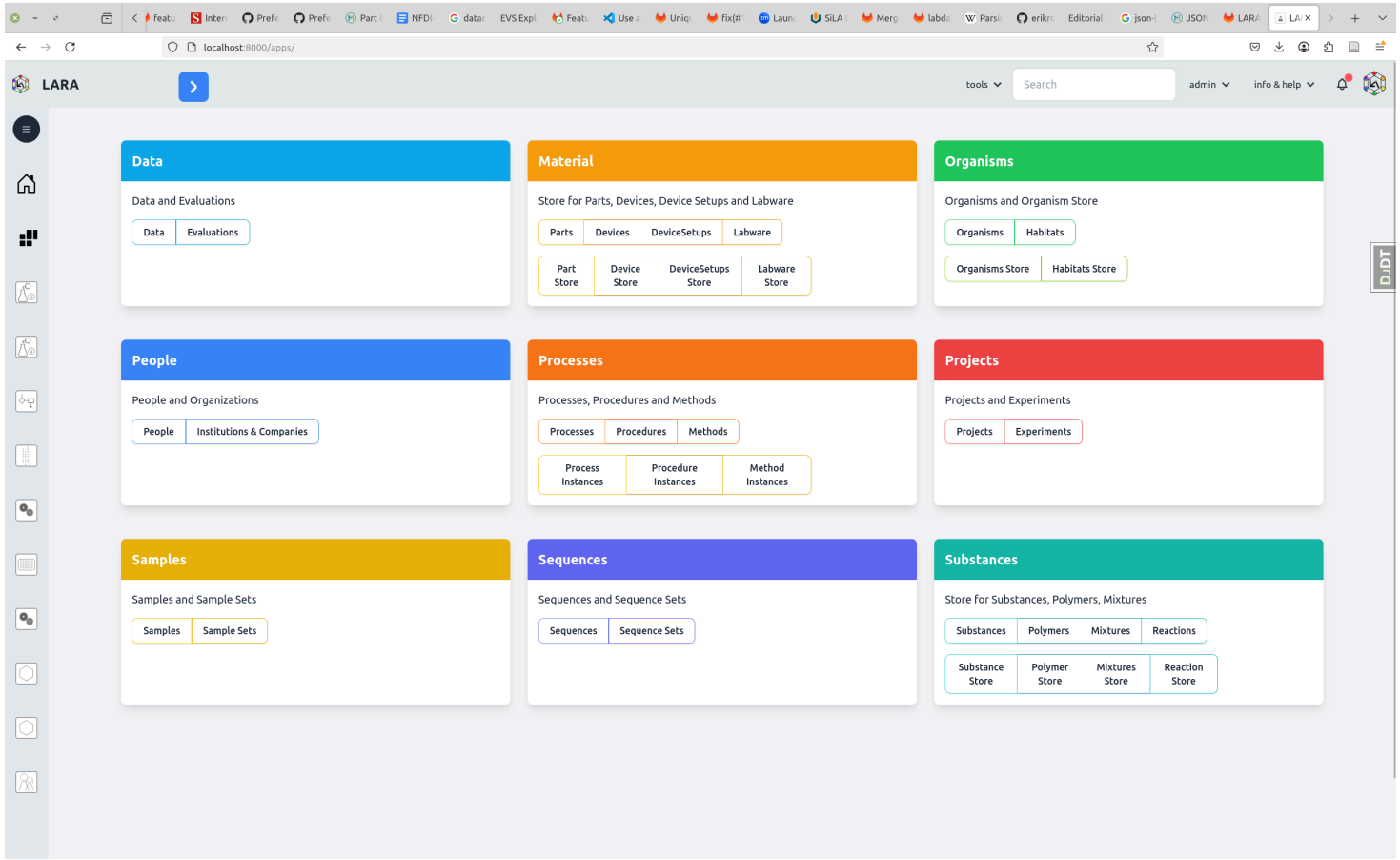
knowldedge graph



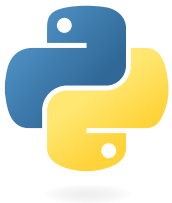
gitlab.com/LARAsuite



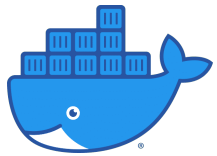
gitlab.com/LARAsuite



build with ...

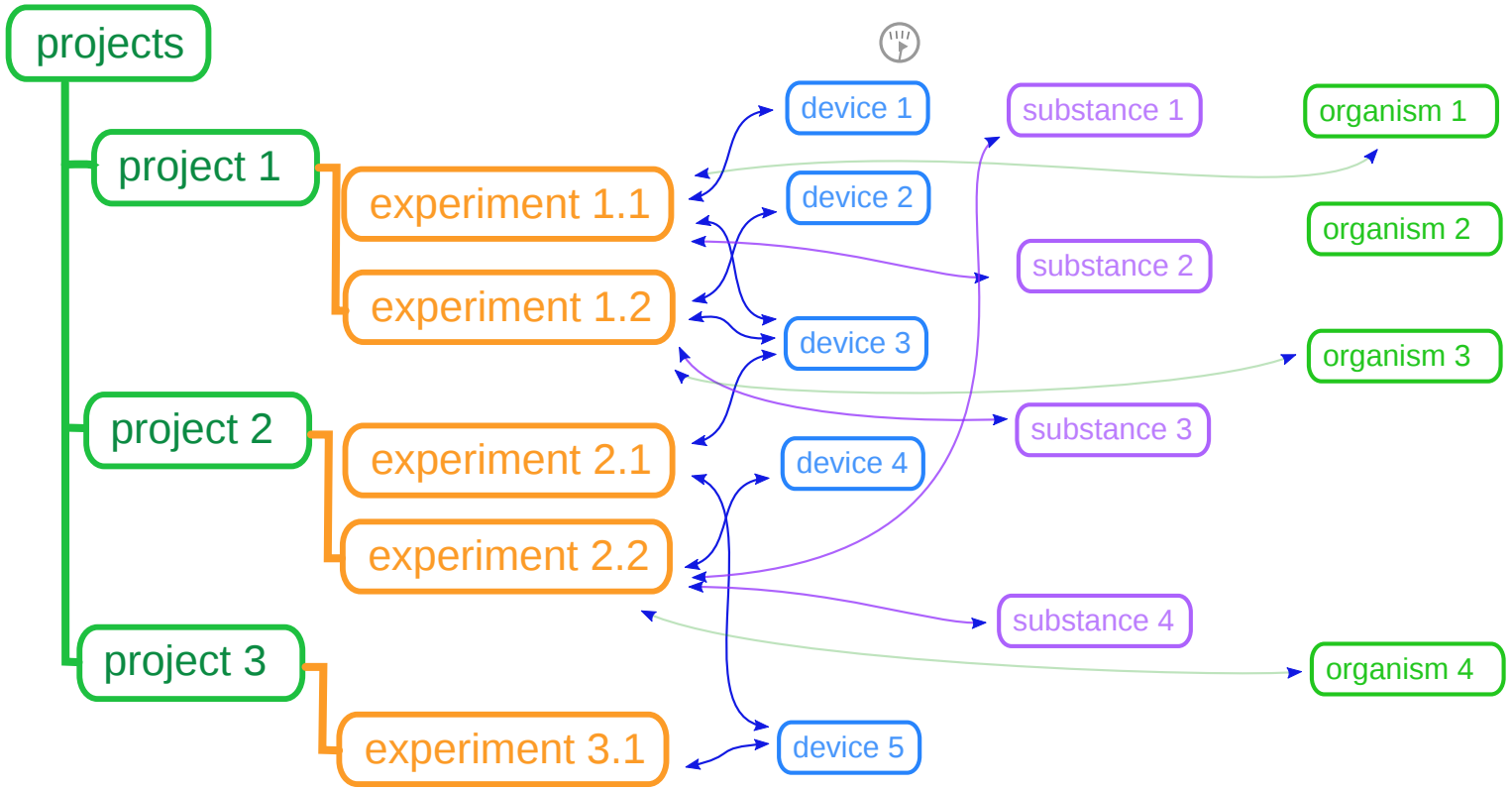



- standards based on: python, SiLA, OWL, AnIML, docker/kubernates,...
- starting at the point of data generation (LabDataReader)
- application ontology design (in collaborations, e.g. NFDI4Cat, EMMO)
- orchstrator, scheduler, aggregator, database
- all open source



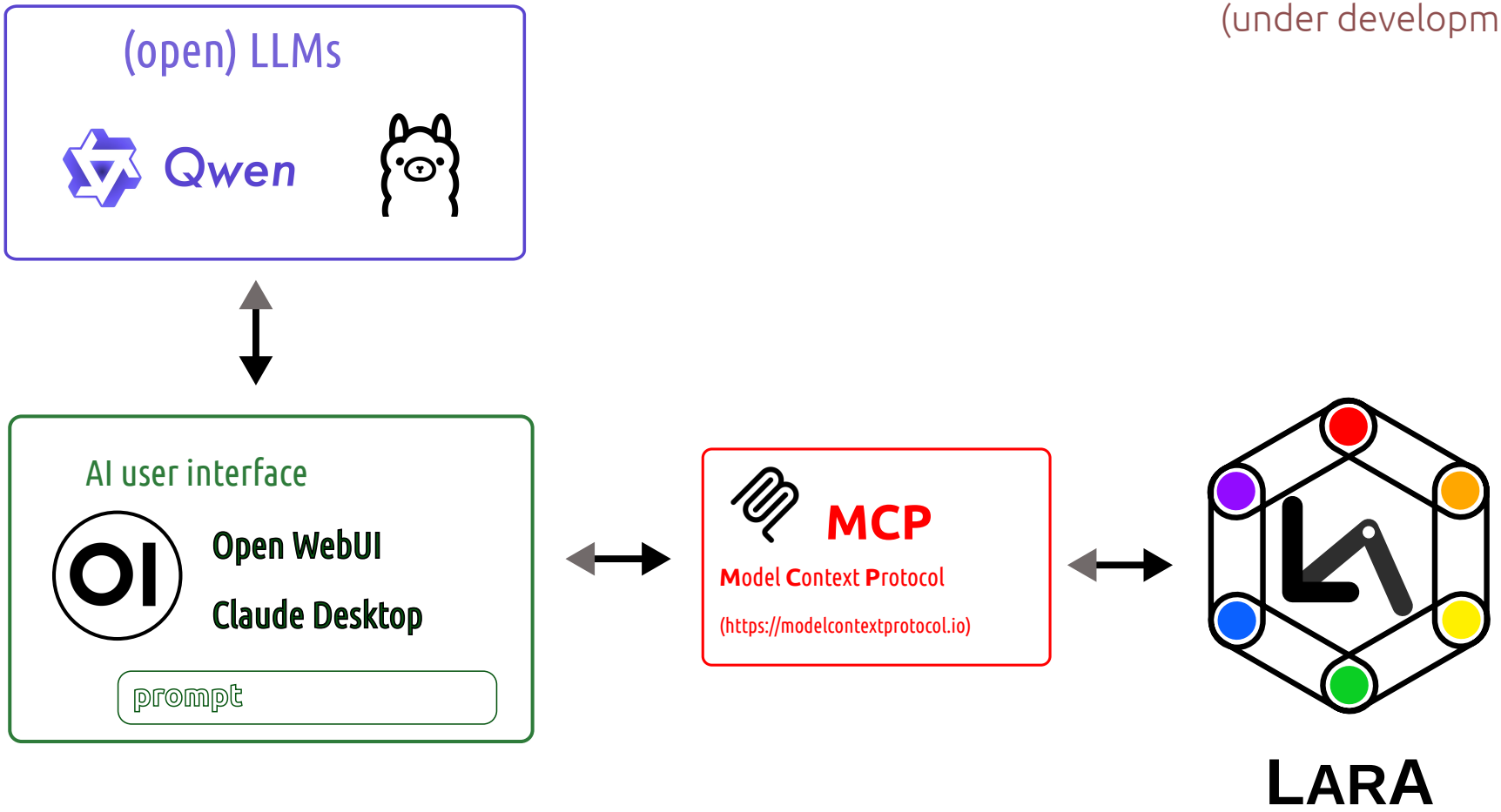
Linked Data

LARA architecture / linked data
(gitlab.com/larasuite)



process planning - MCP

(under development)



summary

- remove semantic overhead, use OCCAM's razor
- use derived knowledge whenever possible
- modular micro-ontologies
- build interlinked network of ontologies
- hybrid relational / graph approach ?

collaborations & partners

scientific partners

- Stefan Born (TU Berlin)
- Peter Neubauer's group (TU Berlin) / KIWI biolab
- Johannes Kabisch's group and associates (Uni Trondheim)
- Egon Heuson (Uni Lille)
- Max Schmidt - (Dr. Sudo's Team) TU-Darmstadt
- Hana Faldynova, David Lacko, Jan Velecky, Zbynek Prokop, Stanislav Mazurenku, Jiri Damborski (University Brno)
- Torsten Veit, Jürgen Deinlein, Jie Peng (Herrenhauszentrum Greifswald)

collaborations & partners

software development / devops

- Lukas Bromig and Julian Willand (unitelabs)
- Erik Trygg, Christian Hejesen (Novo Nordisk, Danmark)
- SiLA labautomation workgroup
- Bryam Nuez Flores, Stefan Kemnitz, Florian Wagner, Ralf Schneider (Rechenzentrum Greifswald)

semantics / knowledge graphs

- Hendrik Borgelt (TU Dortmund), Phlip Strömert (TIB)
- NFDI4Cat / NFDI4Chem (and related NFDIs)
- NFDIBase
- EOSC - Semantic Interoperability Task Force

acknowledgements

- Stefan Maak
- Uwe Bornscheuer's group (Univ. Greifswald)

KIWI-biolab team, NFDI4Cat teams, SiLA team & AnIML team

This work was supported by the German Federal Ministry of Education and Research through the Program "International Future Labs for Artificial Intelligence" (Grant number 01DD20002A) as well as the NFDI4Cat grant

We are grateful to the Deutsche Forschungsgemeinschaft (DFG, INST 292/118-1 FUGG) and the federal state Mecklenburg-Vorpommern for financing the robotic platform.

lara.uni-greifswald.de/presentations

thank you for your attention !