



Linking data automatically from their origin

a demonstration of the LARAsuite open data platform

mark doerr, stefan maak, stefan born & uwe bornscheuer

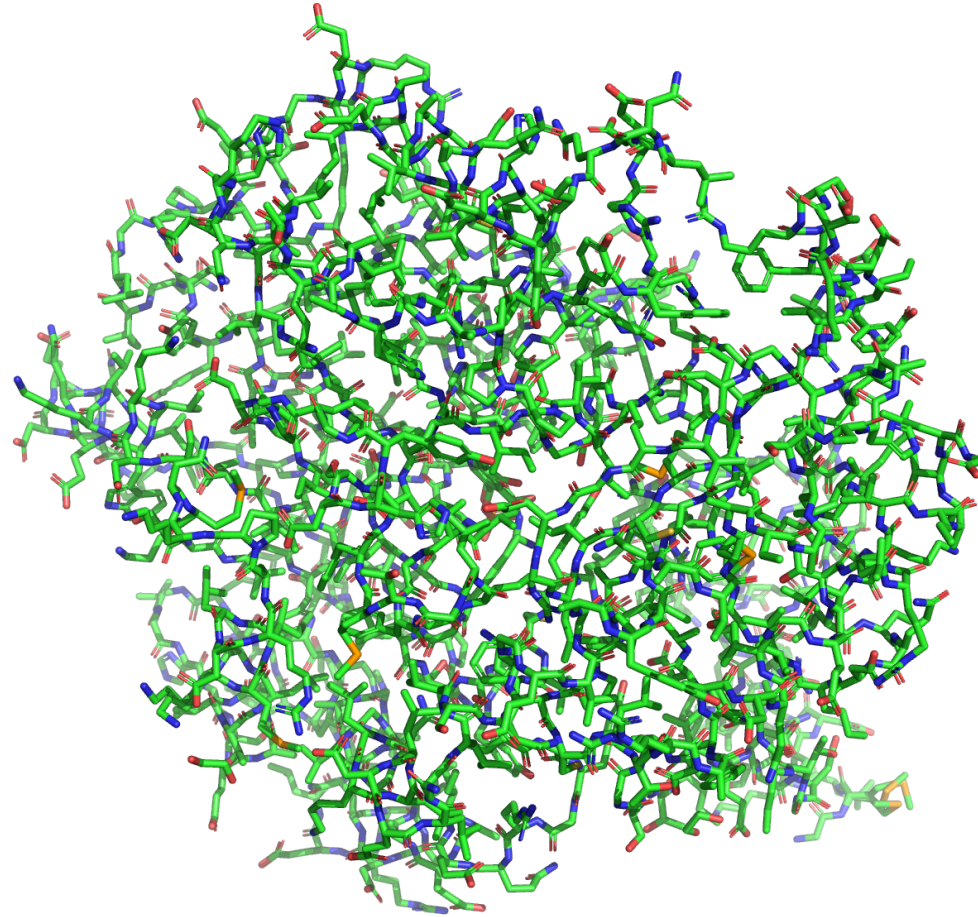
university greifswald / tu-berlin

2024-12-12

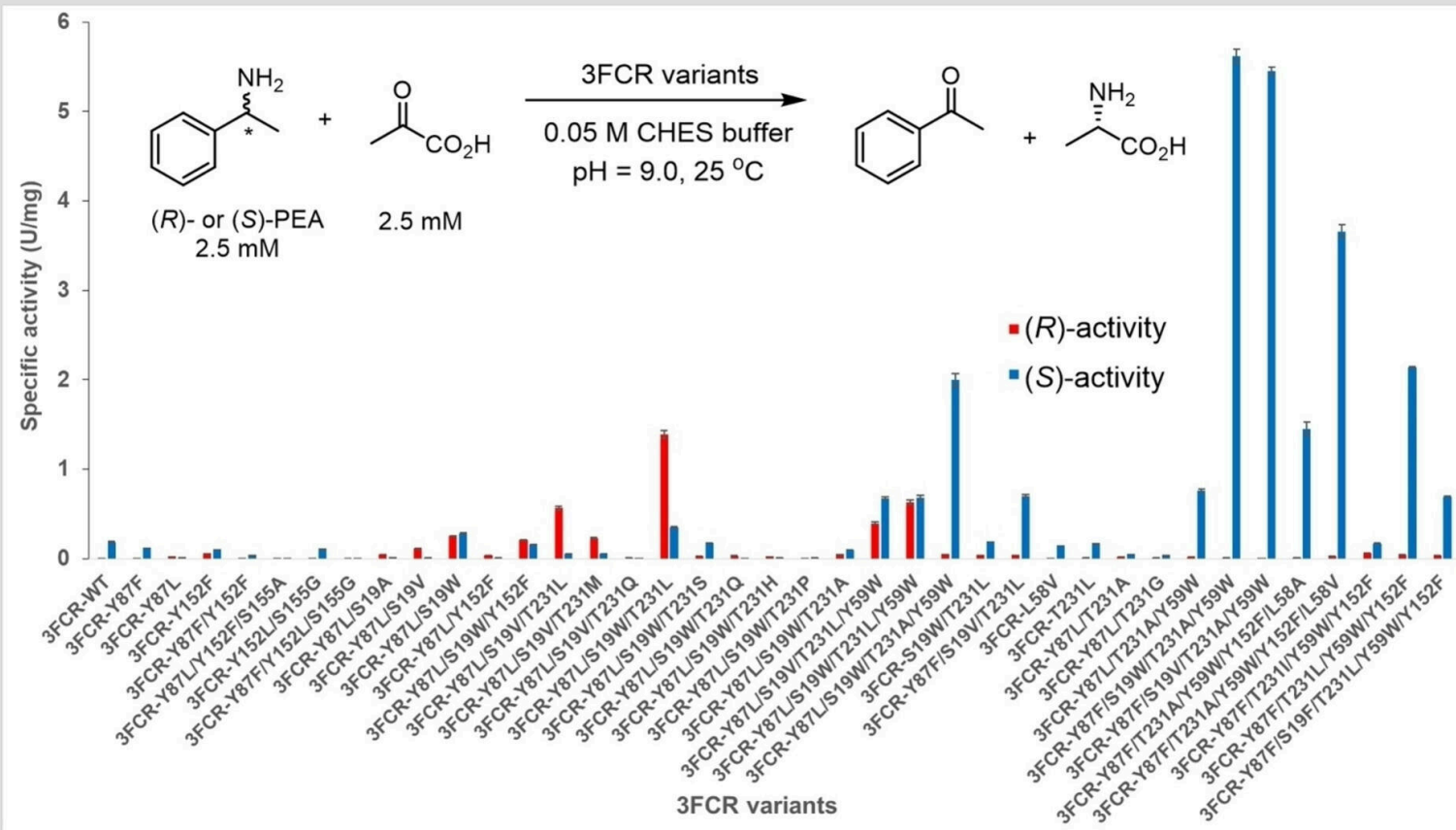
use case/motivation



machine learning

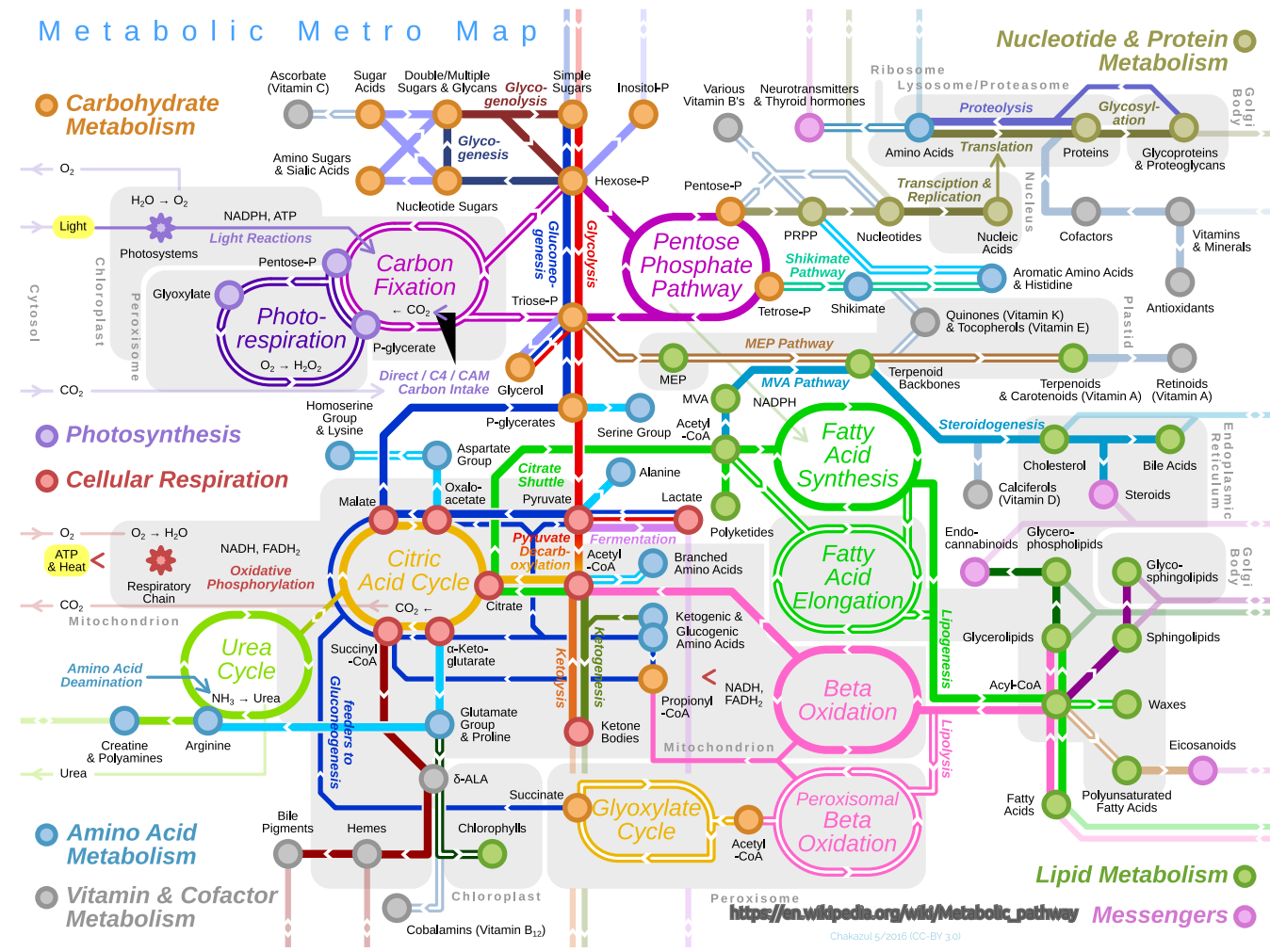


machine learning



Source: Structure- and Data-Driven Protein Engineering of Transaminases for Improving Activity and Stereoselectivity Yu-Fei Ao et. al, Angewandte Chemie 2023. <https://doi.org/10.1002/anie.202301660>

complex expression system



complex instrumentation



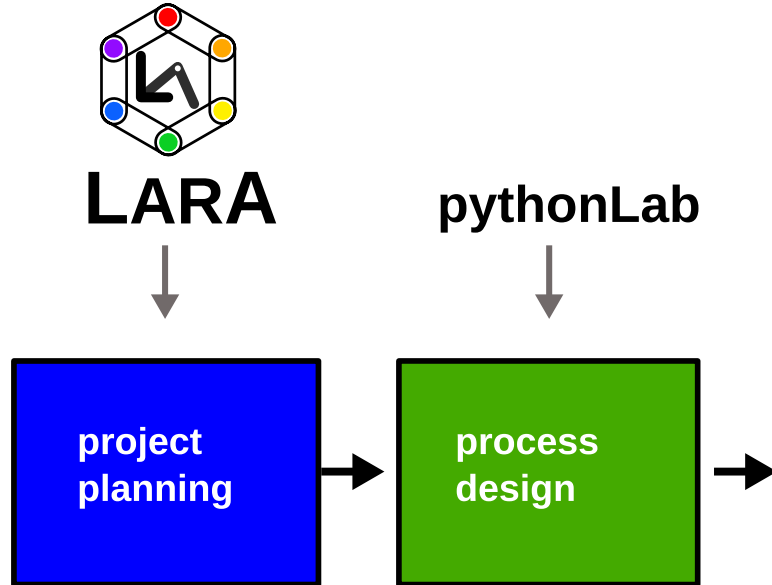


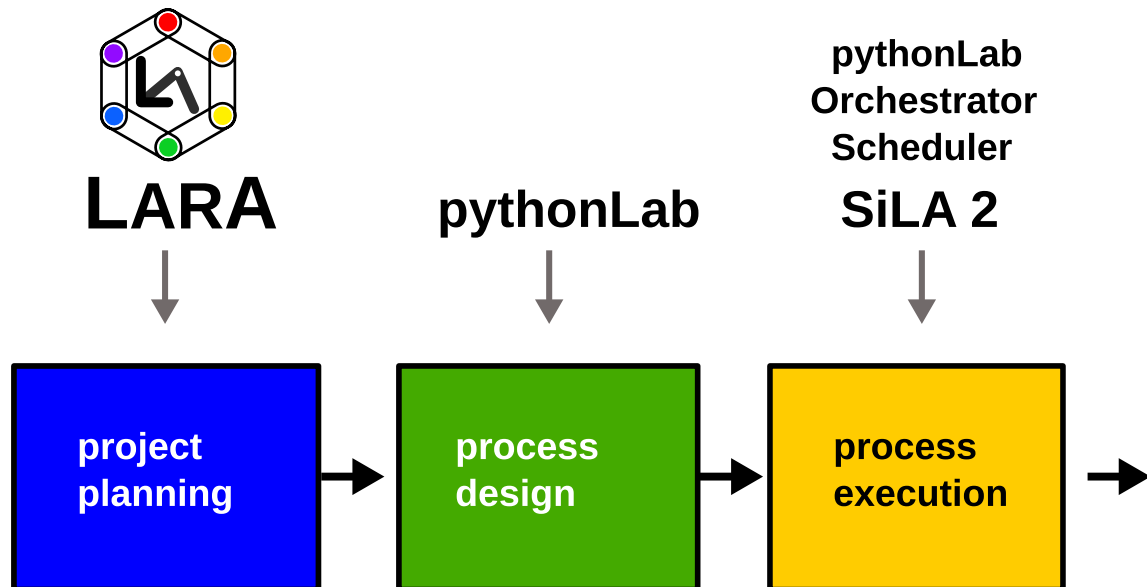
LARA

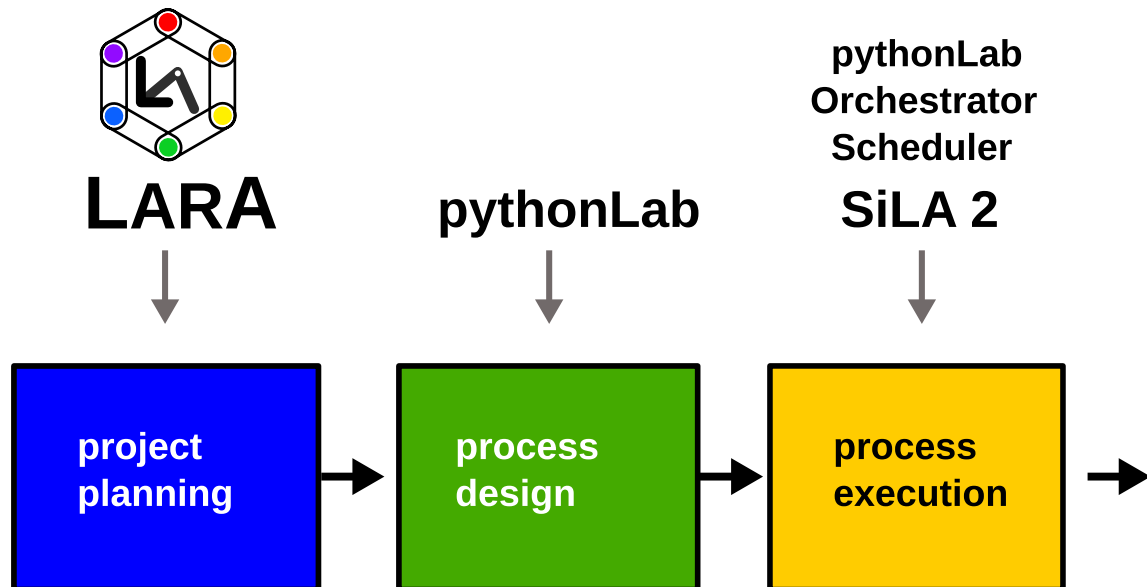


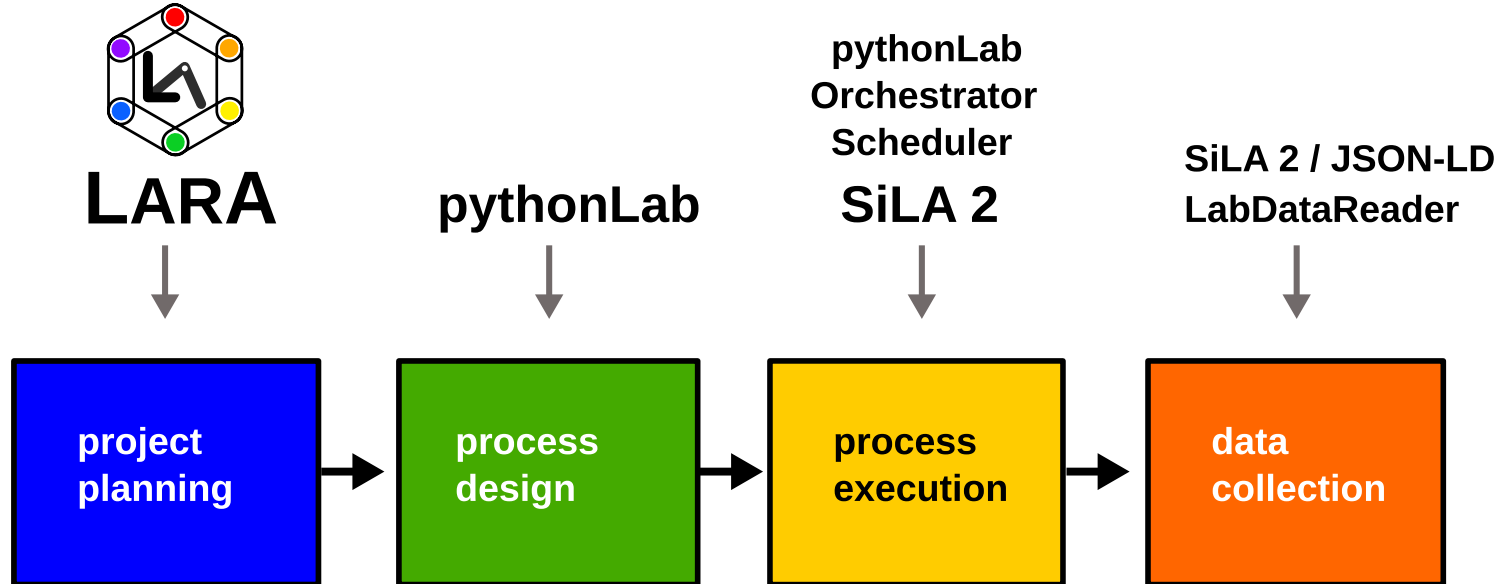
**project
planning**

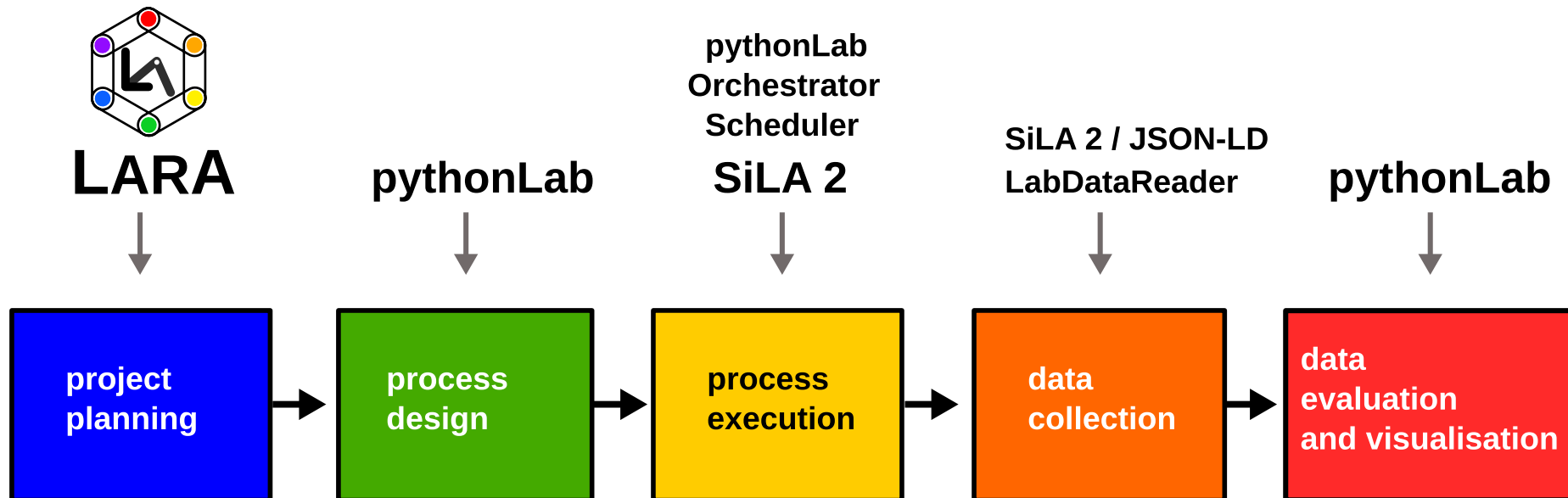


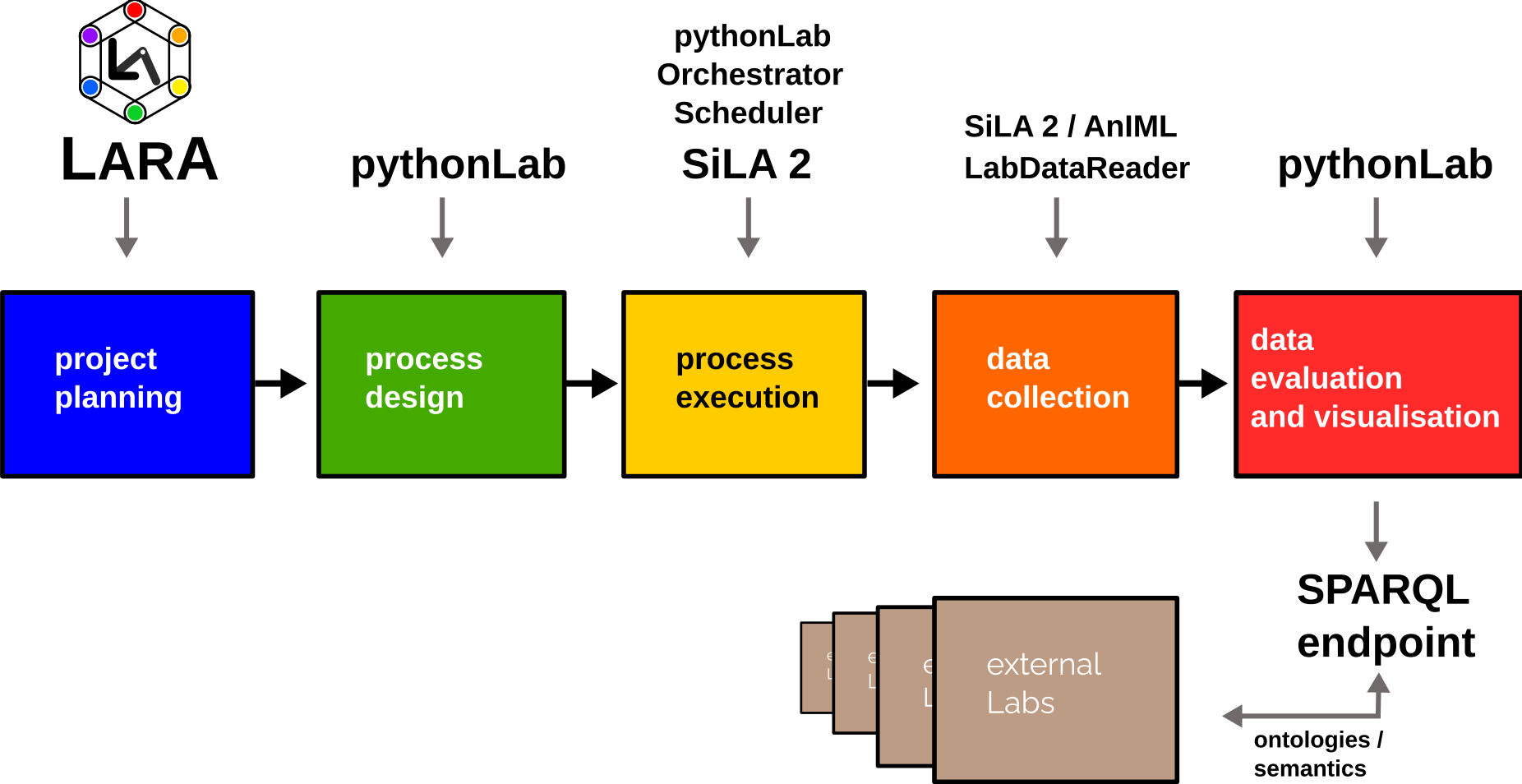


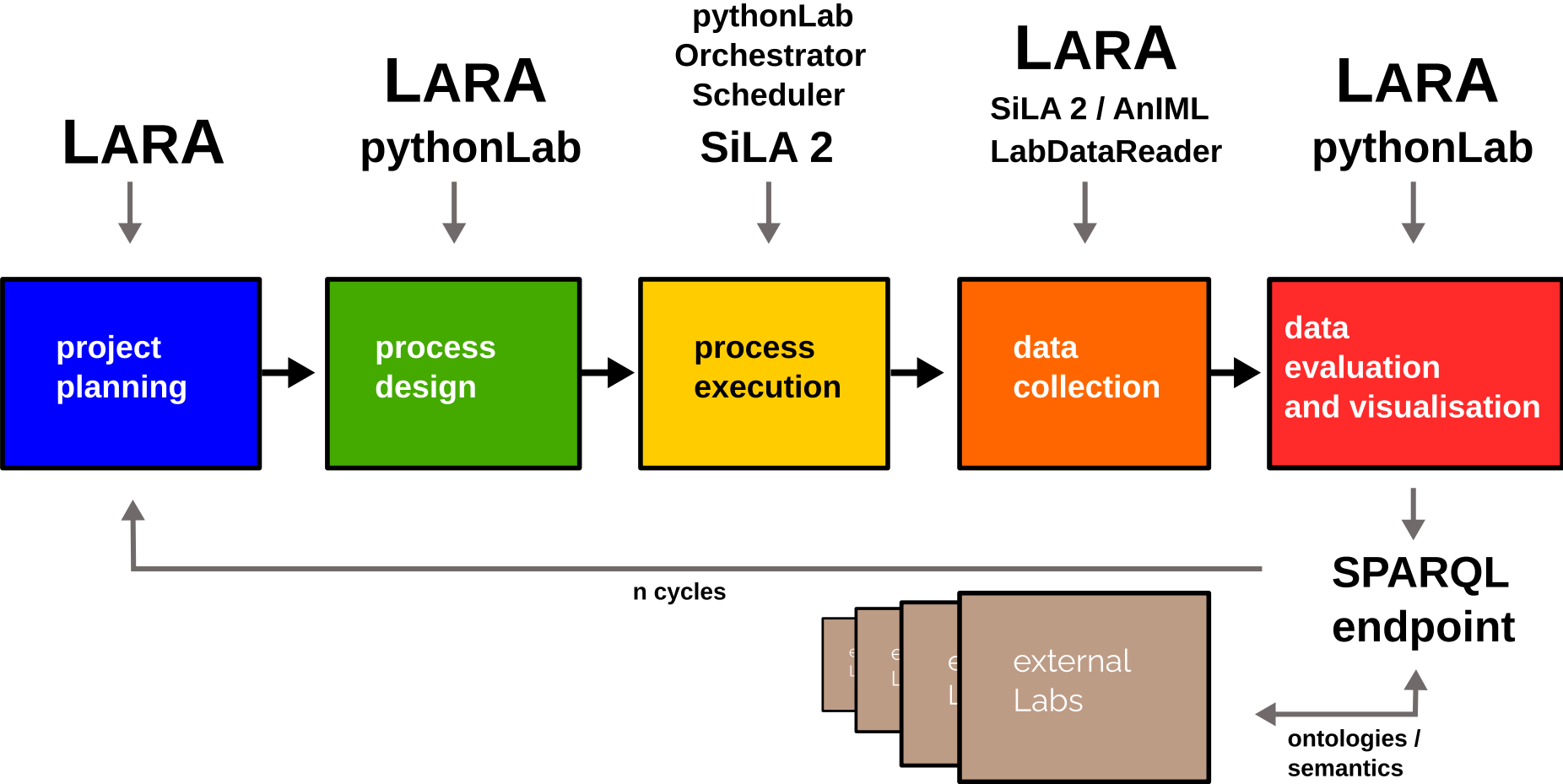












motivation: why we need full
automation of metadata
acquisition and semantic data ?

- meaning of the data is well defined

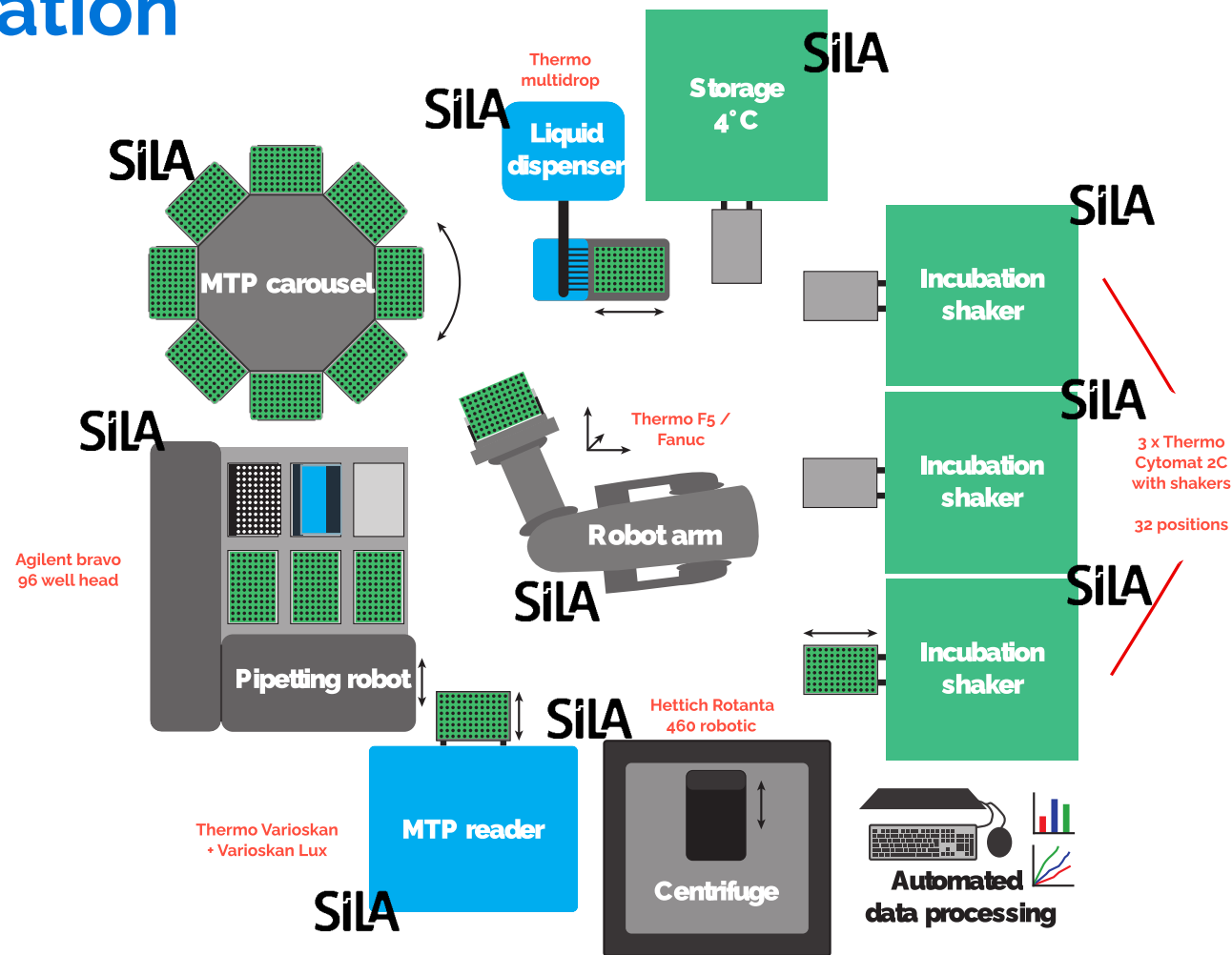
- meaning of the data is well defined
- machine readable

- meaning of the data is well defined
- machine readable
- machine actionable / “understandable”

- meaning of the data is well defined
- machine readable
- machine actionable / “understandable”
- machine reasoning (inference) possible (reasoners)

https://en.wikipedia.org/wiki/FAIR_data

standardisation of communication



localhost:3000/127.0.0.1:50060

Servers > 127.0.0.1:50060

SiLA Service

org.silastandard/core/SiLAService/v1

^

Device Info Provider

de.unigreifswald.biochemie/device/DeviceInfoProvider/v1

^

Light Intensity Controller

de.unigreifswald/instruments/LightIntensityController/v1

^

Logging Service

de.unigreifswald/infrastructure/LoggingService/v0

^

Simulation Controller

org.silastandard/core/SimulationController/v1

^

Temperature Controller

de.unigreifswald/instruments/TemperatureController/v1

v

This is a simple example of a generic Feature for controlling and retrieving the temperature. A new target temperature can be set anytime with the 'Control Temperature' Command. The temperature range has been limited to prevent major damages of a device. In case the first target temperature has not been reached, a ControlInterrupted Error should be thrown.

Current Temperature

^

Target Temperature

^

Current Temperature JSONLD

^

Target Temperature JSONLD

^

Metadata

^

Control Temperature

^

Defined Exceution Error

v

mark doerr

LinkedData/AI Berlin

2024-12-12

18 / 31

Oct. 29th | Prefect - dbt Showcase: Scaling Data Workflows



PRODUCT

SOLUTIONS

RESOURCES

PRICING

16,030

SIGNUP / LOGIN

BOOK A DEMO

flow.py

```
1 from prefect import flow, task
2
3
4 @task(log_prints=True)
5 def say_hello(name: str):
6     print(f"Hello {name}!")
7
8
9 @flow
10 def hello_universe(names: list[str]):
11     for name in names:
12         say_hello(name)
13
14
15 if __name__ == "__main__":
16     # create your first deployment to automate your flow
17     hello_universe.serve(name="your-first-deployment")
```

Acme Co.

Dashboard

Flow Runs

Artifacts

Event Feed

Flows

Blocks

Variables

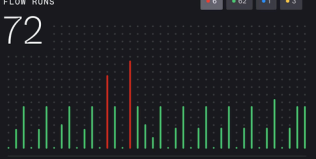
Work Pools

Task Run Concurrency

Automations

FLOW RUNS

72



myUniverse

1h 12m ago

2

farewell-universe

production

etl

2

6s

10m ago

Deployment

galaxy-deployment

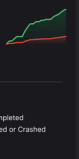
default

Failed

TASK RUNS

1.2k

6.8%



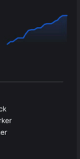
Completed

Failed or Crashed

EVENTS

7.4k

8.2%



Block

Worker

Other

WORK POOLS

azure-push

8s

2.2s

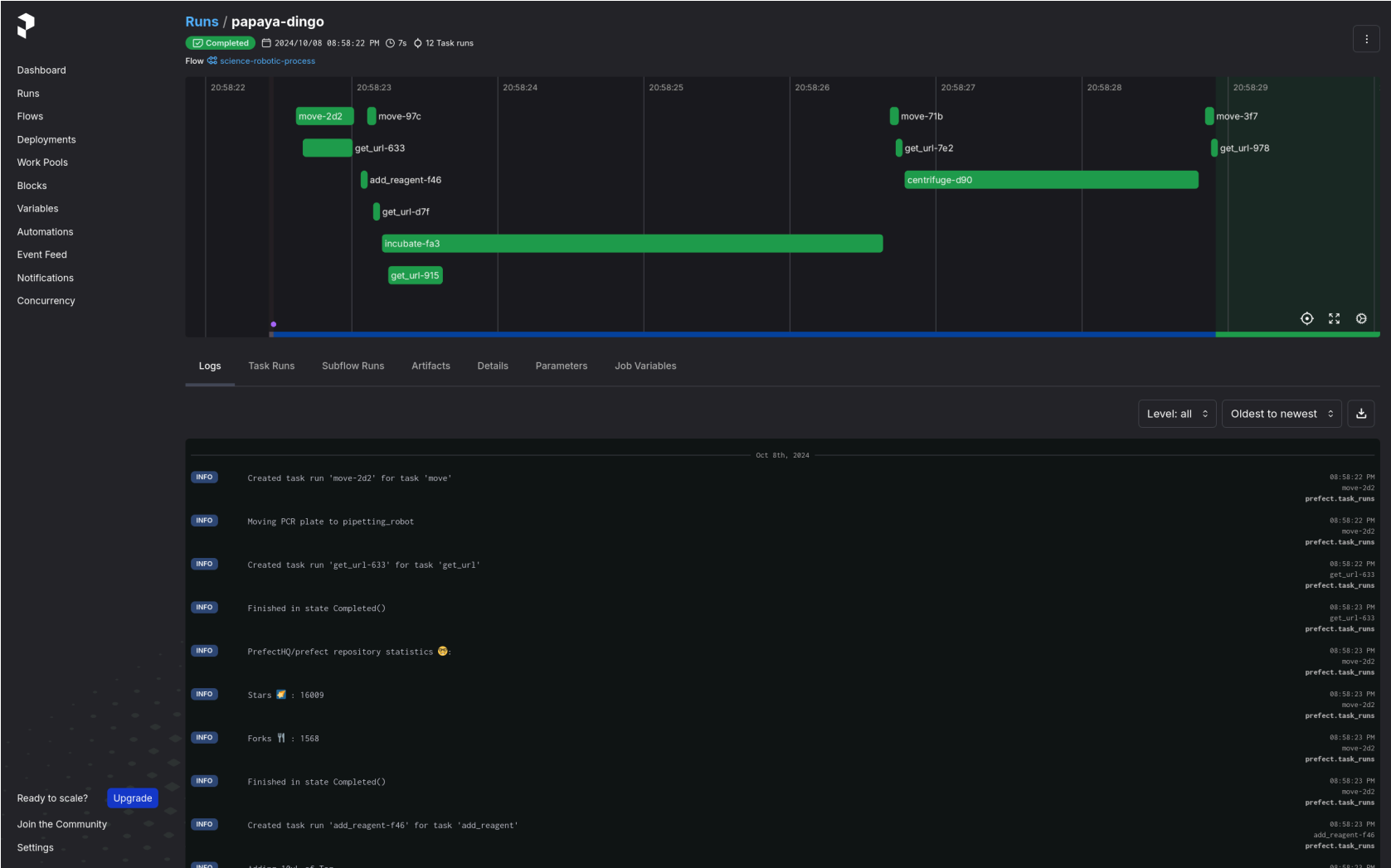
3

89%

5

azure-worker-b

all



LARAsuite

[Home](#)
[Apps](#)
[Tools](#)
[Help & Infos](#)
[Admin](#)

LARA Main Modules - Apps

Projects

Management of Experiments and Projects.

[Projects](#)
[Experiments](#)

Processes

Management of Processes and Procedures.

[Procedures ...](#)
[Processes ...](#)
[Methods ...](#)

[Procedure Inst.](#)
[Process Inst.](#)
[Method Inst.](#)

People

People and Groups database.

[People...](#)

Data

Data database.

[Data](#)

Substances, Reactions, Structures

Substances, Polymers, Mixtures, Reactions and Structures databases.

[Substances ...](#)
[Polymers](#)
[Mixtures](#)
[Reactions](#)

[Substance Inst.](#)
[Polym.I.](#)
[Mixtures I.](#)
[React I.](#)

Sequences

Sequences database.

[Sequences ...](#)

Material, Parts, Devices and Labware

Material, Parts, Devices and Labware databases

[Parts](#)
[Devices](#)
[Labware](#)

[Parts I.](#)
[Devices I.](#)
[Labware I.](#)

Samples

Samples database .

[Samples ...](#)

Organisms

Organism databases

[Organisms ...](#)
[Organisms Inst.](#)

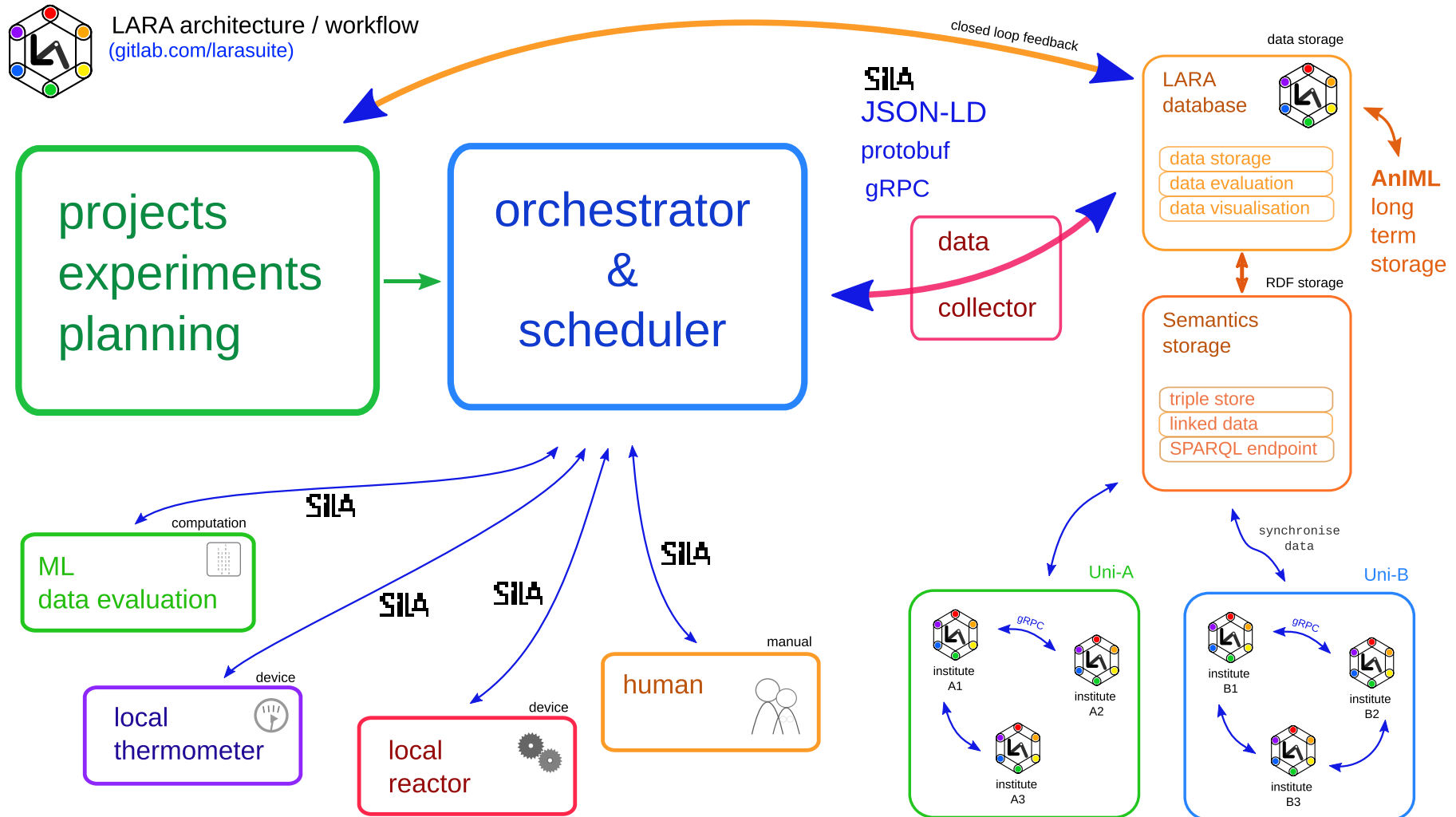
2023 LARAsuite | design by benjamin lear and [mark doerr](#)

mark doerr

LinkedData/AI Berlin

2024-12-12

22 / 31



SPARQL Query Editor

AboutTables

ConductorPermalink

Extensions:

cxmlsave to davsponge

User: SPARQL

Default Data Set Name (Graph IRI)

urn:spargl:lara:data

Query Text

```
SELECT *
WHERE {
    ?s ?p ?o .
}
LIMIT 30
```

Results Format

HTML

Execute Query

Reset

Execution timeout

0

milliseconds

Options

☒ Strict checking of void variables

☐ Log debug info at the end of output (has no effect on some queries and output formats)

☐ Generate SPARQL compilation report (instead of executing the query)

Copyright © 2024 [OpenLink Software](#)

Virtuoso version 07.20.3240 (a1fd8195b) on Linux (x86_64-ubuntu_focal-linux-gnu) Single Server Edition (47 GB total memory, 117 MB memory in use)

SPARQL | HTML5 table

s	p	o
http://w3id.org/lara/demo_data_1b_v0.0.1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#NamedIndividual
http://w3id.org/lara/demo_data_1c_v0.0.1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2002/07/owl#NamedIndividual
http://w3id.org/lara/demo_data_1b_v0.0.1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://w3id.org/lara/Data
http://w3id.org/lara/demo_data_1c_v0.0.1	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://w3id.org/lara/Data
http://w3id.org/lara/demo_data_1b_v0.0.1	http://w3id.org/lara/datetime_last_modified	2024-11-06T08:24:08.575Z
http://w3id.org/lara/demo_data_1c_v0.0.1	http://w3id.org/lara/datetime_last_modified	2024-11-06T08:24:17.634Z
http://w3id.org/lara/demo_data_1b_v0.0.1	http://w3id.org/lara/version	"v0.0.1"^^<http://www.w3.org/2001/XMLSchema#string>
http://w3id.org/lara/demo_data_1c_v0.0.1	http://w3id.org/lara/version	"v0.0.1"^^<http://www.w3.org/2001/XMLSchema#string>
http://w3id.org/lara/demo_data_1b_v0.0.1	http://w3id.org/lara/datetime_created	2024-11-06T09:24:07.308Z
http://w3id.org/lara/demo_data_1c_v0.0.1	http://w3id.org/lara/datetime_created	2024-11-06T09:24:15.938Z
http://w3id.org/lara/demo_data_1b_v0.0.1	http://w3id.org/lara/name_full	"https://de.unigreifswald/biochem/akb/demo_data_1b_v0.0.1"^^<http://www.w3.org/2001/XMLSchema#string>
http://w3id.org/lara/demo_data_1c_v0.0.1	http://w3id.org/lara/name_full	"https://de.unigreifswald/biochem/akb/demo_data_1c_v0.0.1"^^<http://www.w3.org/2001/XMLSchema#string>
http://w3id.org/lara/demo_data_1b_v0.0.1	http://w3id.org/lara/data_id	"48e9088d-9476-4f0e-90f8-c7923939266e"^^<http://www.w3.org/2001/XMLSchema#string>
http://w3id.org/lara/demo_data_1c_v0.0.1	http://w3id.org/lara/data_id	"627d3639-f41a-428b-87a0-a208fa8150a2"^^<http://www.w3.org/2001/XMLSchema#string>
http://w3id.org/lara/demo_data_1b_v0.0.1	http://w3id.org/lara/title	"Demo Data 1b"^^<http://www.w3.org/2001/XMLSchema#string>
http://w3id.org/lara/demo_data_1c_v0.0.1	http://w3id.org/lara/title	"Demo Data 1c"^^<http://www.w3.org/2001/XMLSchema#string>

SPARQL Query Editor

AboutTables

ConductorPermalink

Extensions: cxmlsave to davspongeUser: SPARQL

Default Data Set Name (Graph IRI)

urn:spargl:lara:data

Query Text

```
PREFIX rdfo: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX emmo: <http://emmo.info/emmo-inferred#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX lara: <http://w3id.org/lara/>

SELECT ?data_title ?date ?data_id
WHERE
{
  ?data lara:datetime_created ?date .
  ?data lara:data_id ?data_id .
  ?data lara:title ?data_title .
  FILTER (?date > "2024-11-01T00:00:00Z"^^xsd:dateTime)
}
LIMIT 30
```

Results Format

HTML

Execute Query

Reset

Execution timeout

0

milliseconds

Options

☒ Strict checking of void variables

☐ Log debug info at the end of output (has no effect on some queries and output formats)

☐ Generate SPARQL compilation report (instead of executing the query)

Copyright © 2024 [OpenLink Software](#)

[Virtuoso](#) version 07.20.3240 (a1fd8195b) on Linux (x86_64-ubuntu_focal-linux-gnu) Single Server Edition (47 GB total memory, 117 MB memory in use)

SPARQL | HTML5 table

data_title	date	data_id
"Demo Data 1b"^^<http://www.w3.org/2001/XMLSchema#string>	2024-11-06T09:24:07.308Z	"48e9088d-9476-4f0e-90f8-c7923939266e"^^<http://www.w3.org/2001/XMLSchema#string>
"Demo Data 1c"^^<http://www.w3.org/2001/XMLSchema#string>	2024-11-06T09:24:15.938Z	"627d3639-f41a-428b-87a0-a208fa8150a2"^^<http://www.w3.org/2001/XMLSchema#string>

labDataReader

- generic reader of proprietary data (e.g. HPLC, plate readers)
- primary output : pandas data frame and *JSON-LD* (metadata)

<https://gitlab.com/opensourcelab/scientificdata/labDataReader>

SciDat

- packing tabular data / data frames into *parquet* files, including *JSON-LD* metadata

<https://gitlab.com/opensourcelab/scientificdata/scidat>

LARA is a tool ...

LARA is a tool ...

- for planning, executing and evaluating experiments as complex as BioCatalytic / Enzymatic processes

LARA is a tool ...

- for planning, executing and evaluating experiments as complex as BioCatalytic / Enzymatic processes
- leverages automation in many steps

LARA is a tool ...

- for planning, executing and evaluating experiments as complex as BioCatalytic / Enzymatic processes
- leverages automation in many steps
- generates a knowledge graph on the fly

LARA is a tool ...

- for planning, executing and evaluating experiments as complex as BioCatalytic / Enzymatic processes
- leverages automation in many steps
- generates a knowledge graph on the fly
- synchronises with external data sources (like Dataverse)

acknowledgements

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

project partners

- Stefan Born (TU Berlin)
- Peter Neubauer's group (TU Berlin)
- Johannes Kabisch's group and associates (Uni Trondheim)
- Egon Heuson (Uni Lille)
- Lukas Bromig and Julian Willand (unitelabs)

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.

