



DOME 4.0

www.dome40.eu

iDMT

DOME Hackathon 3

Chung Ting Lao

CMCL

11th Sep 2024



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 953163

- The Innovation Centre in **D**igital **M**olecular **T**echnologies
- An open innovation platform for collaborative R&D in the areas of:
 - Artificial intelligences in molecular technologies
 - Robotic equipment for chemical synthesis
 - Algorithms and tools for digital process development
- Co-fund by
 - European Regional Development Fund (ERDF)
 - AstraZeneca
 - Shionogi
 - University of Cambridge



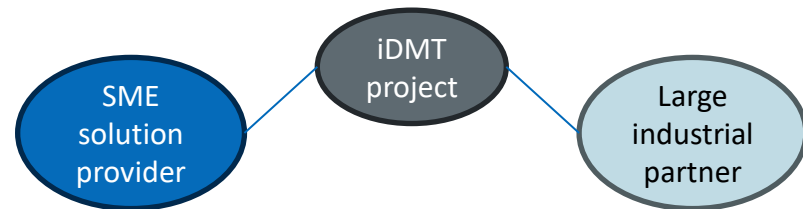
iDMT

Innovation Centre
in Digital Molecular
Technologies

idmt.online

About iDMT

- Offers a platform for rapid development and testing of new products
 - In the sectors of chemical synthesis of molecules and materials
- Set up projects that link
 - large industrial partner with need
 - SME with potential solution
 - via a mixed industry-academic R&D team
- Two laboratories: High throughput lab and **Robotic lab**



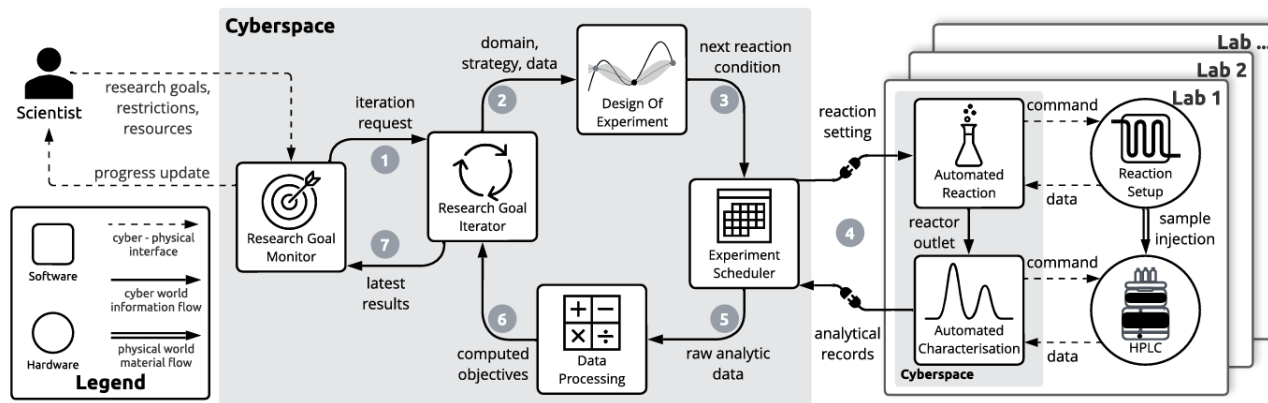
Robotic lab



High throughput lab

Use case: self-driving laboratory

- Needs
 - Accelerated scientific discovery
 - Decentralisation – enables global collaboration
- Challenges
 - Heterogeneous hardware and software
 - Sharing data across organisations

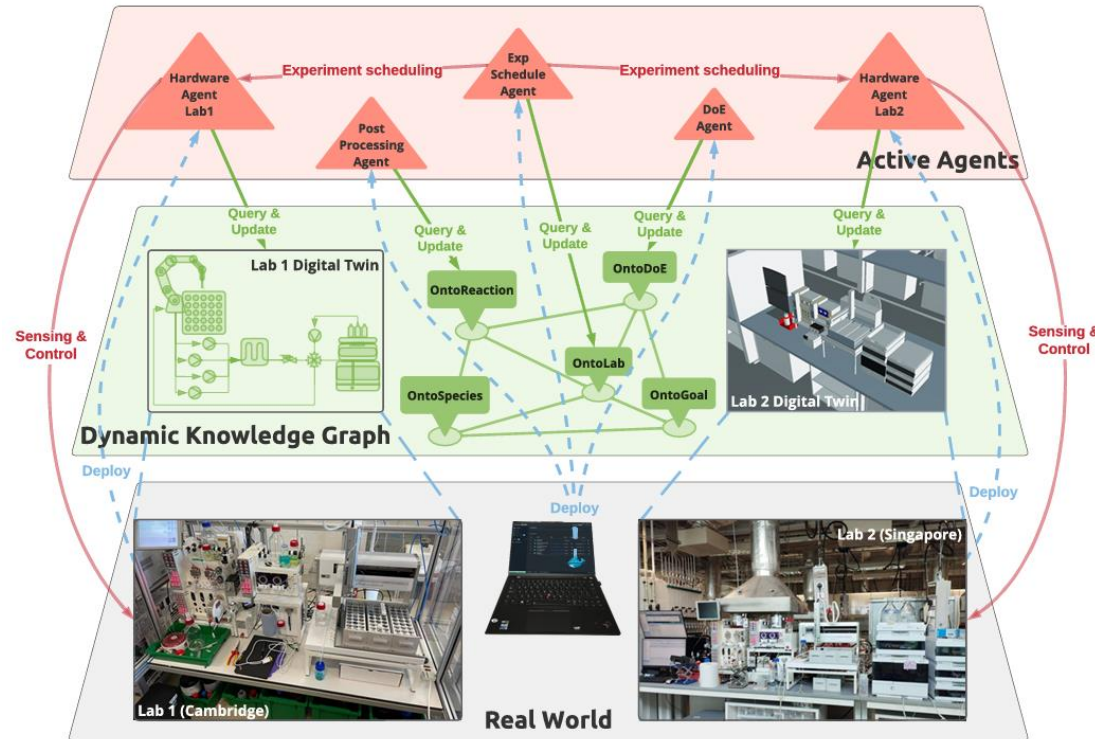


Closed-loop optimisation workflow

Bai, J., Mosbach, S., Taylor, C.J. *et al.* A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat Commun* 15, 462 (2024). <https://doi.org/10.1038/s41467-023-44599-9>

Use case: self-driving laboratory

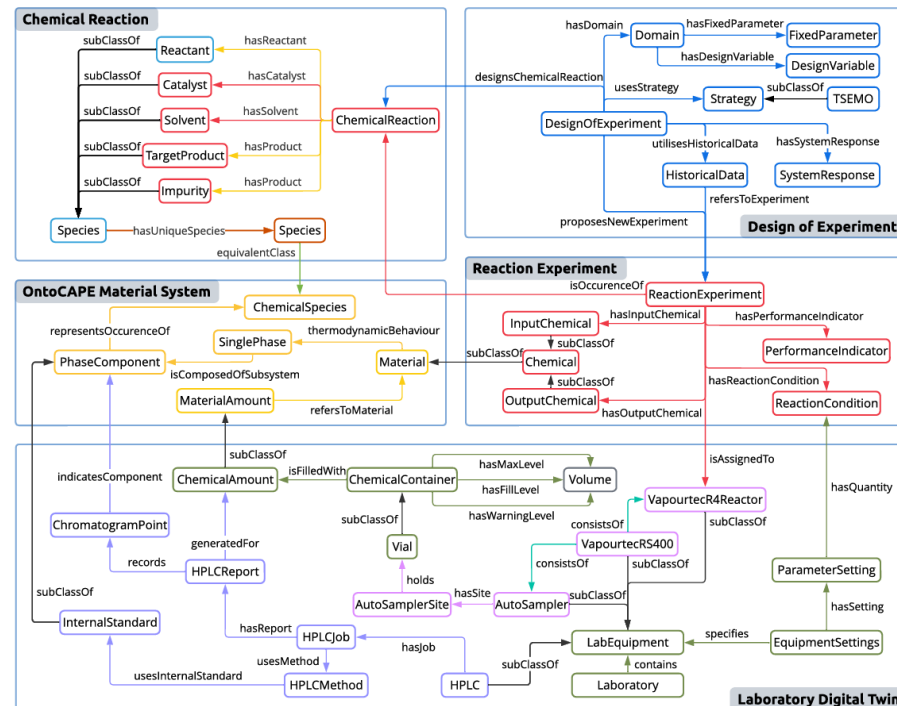
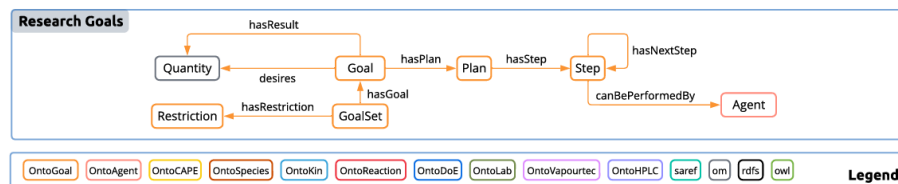
- Dynamic knowledge graph (dKG) approach
 - Hardware in real world
 - dKG hosts all information in cyberspace
 - Store triples: subject-predicate-object
 - Agents manage the dKG



Bai, J., Mosbach, S., Taylor, C.J. *et al.* A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat Commun* 15, 462 (2024). <https://doi.org/10.1038/s41467-023-44599-9>

Use case: self-driving laboratory

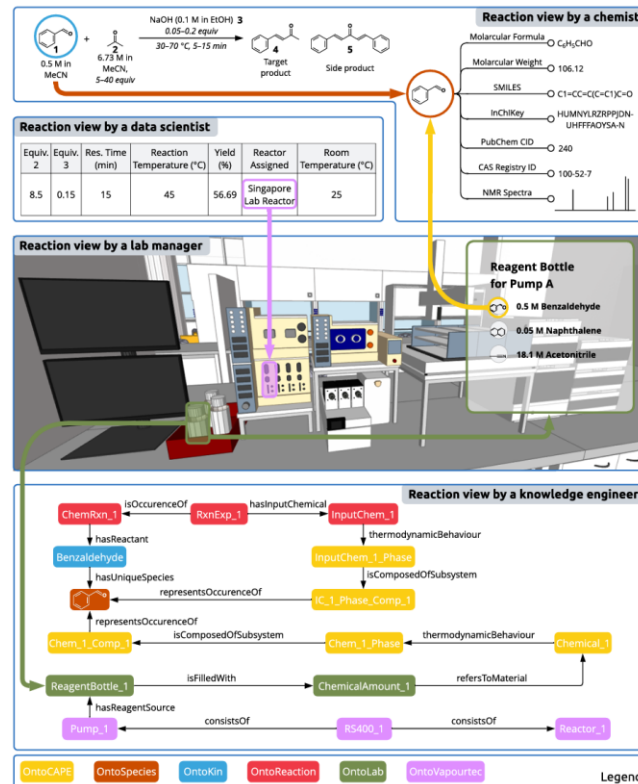
- Ontologies
 - A set of connected ontologies is developed to describe
 - research goals
 - abstract chemistry knowledge
 - hardware (for execution)



Bai, J., Mosbach, S., Taylor, C.J. et al. A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat Commun* 15, 462 (2024). <https://doi.org/10.1038/s41467-023-44599-9>

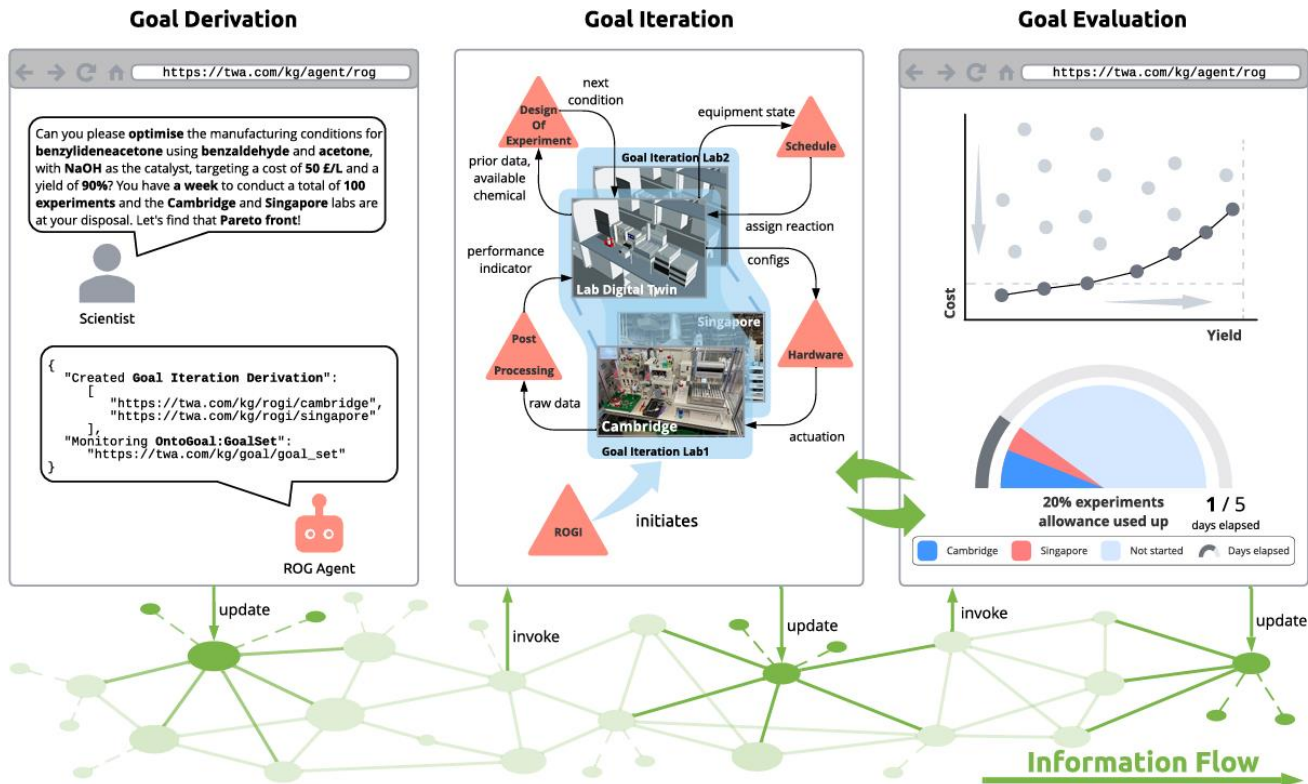
Use case: self-driving laboratory

- dKG approach enables queries of different level of abstractions
 - A chemist may be more interested in the properties of the chemical
 - A data scientist may be more interested in the condition and yield of an experiment
 - A lab manager may be more interested in the hardware and inventory



Bai, J., Mosbach, S., Taylor, C.J. *et al.* A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat Commun* 15, 462 (2024). <https://doi.org/10.1038/s41467-023-44599-9>

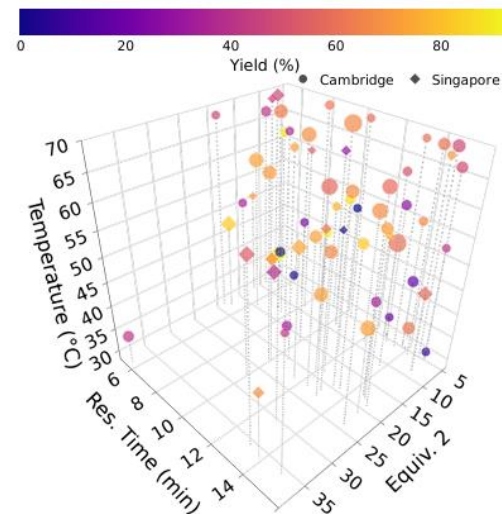
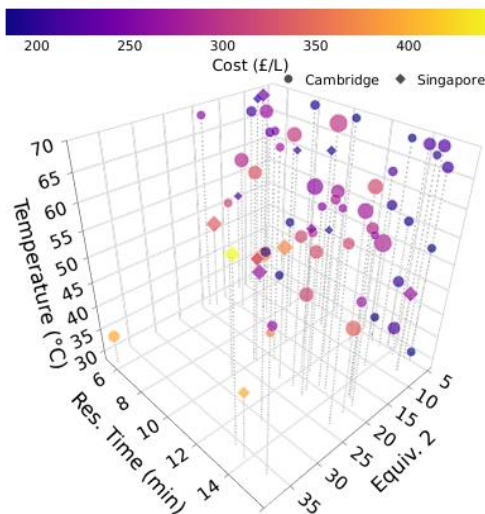
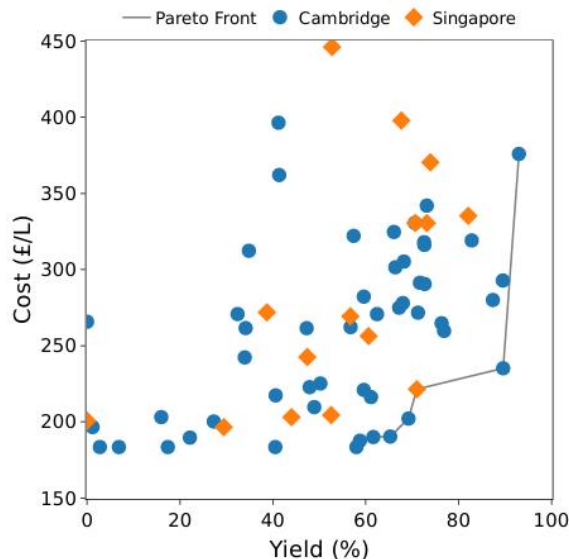
Use case: self-driving laboratory



Bai, J., Mosbach, S., Taylor, C.J. *et al.* A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat Commun* 15, 462 (2024). <https://doi.org/10.1038/s41467-023-44599-9>

Use case: self-driving laboratory

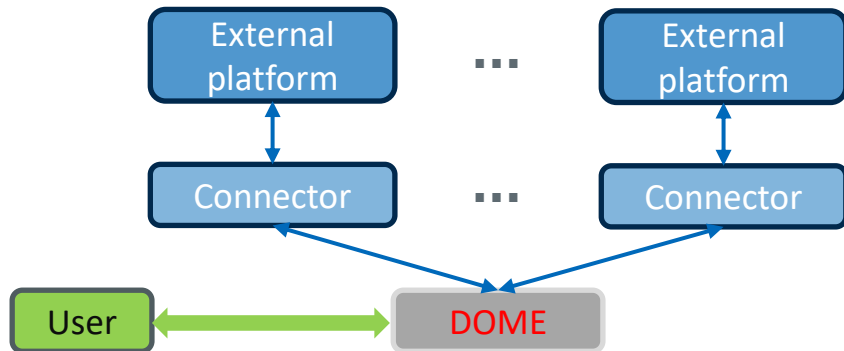
Real-time collaborative distributed SDLs



Bai, J., Mosbach, S., Taylor, C.J. *et al.* A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat Commun* 15, 462 (2024). <https://doi.org/10.1038/s41467-023-44599-9>

Connecting to DOME

- For data on external platform to be discoverable on DOME, we need to
 - Develop and deploy a connector
 - Register connector on DOME
- iDMT onboarding achieved in Hackathon 2



iDMT

Innovation Centre in Digital Molecular Technologies (iDMT) aims to create a digital transformation in the chemical industry by helping companies make the transition to automated discovery processes and machine learning.

Free Platform

true

Domain

ENGINEERING_AND_TECHNOLOGY

Offers

EXPERIMENTAL_DATA

Home Page

<https://www.idmt.online/>

Query URL

<https://nextgen.dome40.io/api/discover/results/iDMT>

Connecting to DOME

[GitHub - DOME-4-0/reference-connector](https://github.com/DOME-4-0/reference-connector)

Reference-Connector-Template

This is a template for a wrapper that will allow third-party connectors to provide data on the DOME 4.0 platform.

The template has been created using [cookiecutter](#).

Generate your project from the template

First, install cookiecutter according to [the documentation](#).

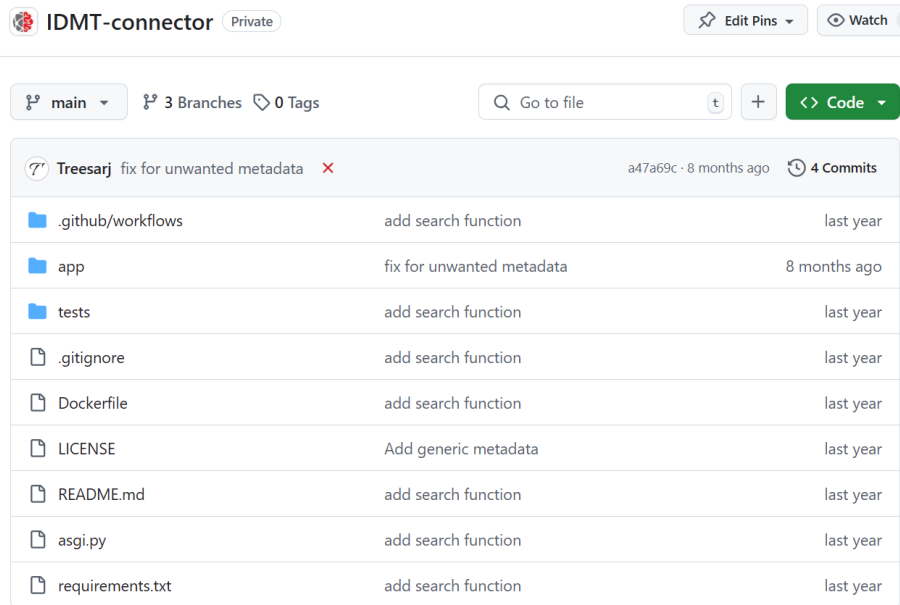
Then run the following command to generate your project:

```
cookiecutter gh:DOME-4-0/reference-connector
```

Now type in the required inputs to customize your repository.

Alternatively, you can pre-define the inputs in a JSON file and pass it to the cookiecutter command using the `--config-file` option. The list of input keys and default values can be found in `cookiecutter.json`. An overview is also provided in the following table:

- Start with a cookiecutter template
- Will create a basic connector as flask app



The screenshot shows the GitHub repository page for 'IDMT-connector'. The repository is private and has 3 branches and 0 tags. The main branch is selected. The file list shows the following files and their commit history:

File	Commit Message	Commit Date
.github/workflows	add search function	last year
app	fix for unwanted metadata	8 months ago
tests	add search function	last year
.gitignore	add search function	last year
Dockerfile	add search function	last year
LICENSE	Add generic metadata	last year
README.md	add search function	last year
asgi.py	add search function	last year
requirements.txt	add search function	last year

- CatalogData

- Confirms to DOME dataset model

```
3 """
4 import json
5
6 from fastapi import APIRouter
7 from pydantic import BaseModel
8
9 from ..utils.utils import fetchTriples
10
11
12 ✓ class CatalogData(BaseModel):
13     """
14     Pydantic data model for DOME 4.0 catalog datasets.
15     """
16     Dataset: list
17     IssueDate: str
18     License: str
19     Title: str
20     URL: str
21     dataCreator: str
22     dataPublisher: str
23     keyword: str
24
25     router = APIRouter()
```

```
27 ✓ def search(search_string: str):
28     """search function
29
30     Args:
31         search_string (str): _description_
32
33     Returns:
34         data and metadata
35     """
36     metadata = []
37     data = []
38     sparql_query = f"""
39     SELECT ?s ?p ?o
40     WHERE {{
41         ?s ?p ?o .
42         FILTER (regex(str(?s), "{search_string}", "i") ||
43                 regex(str(?p), "{search_string}", "i") ||
44                 regex(str(?o), "{search_string}", "i"))
45     }}
46     """
```

- Search function

- Take a search string as input
 - Translate that into API call depending on platform
 - SPARQL query
 - Return **metadata** and **data**

```
48     # Check if data is not empty before fetching and processing
49     if data := json.loads(fetchTriples(sparql_query)):
50         metadata.append(CatalogData(
51             Dataset=["Dataset of Experimental setup and results"],
52             IssueDate="2021-07-11",
53             License="All Rights Reserved",
54             Title="IDMT lab data",
55             URL=f"https://idmt.dome40.io/KB/?query={sparql_query}",
56             dataCreator="IDMT",
57             dataPublisher="IDMT_LABS",
58             keyword=search_string
59         ))
60
61     return metadata, data
```

○ Register DOME 4.0 Connector (dome40.io)

Register DOME 4.0 Connector

Note:

This form is used for registering DOME 4.0 connectors adhering to the Connector API (<https://github.com/DOME-4-0/reference-connector>). This allows external data platform to be discoverable on DOME.

Name of your connector as it will show up in the list of providers* [i](#)

Human readable description of your connector [i](#)

Conforms to a Data Standard: [i](#)

- ☐ Optimade api specification
- ☐ Ids api specification
- ☐ Sc1 api specification
- ☐ Sc4 api specification api specification
- ☐ Sc 4 api specification
- ☐ Sc9 api specification
- ☐ Sc1 sim api specification
- ☒ None

IRI for API (e.g. Query)* [i](#)

Web Page [i](#)

Offers data free of charge* [i](#)

- ☒ True
- ☐ False

Application domain (Type of data)* [i](#)

Application domain (Type of data) ▾

Data Topics (multiple selection)* [i](#)

Data Topics (multiple selection) ▾

- ☐ Cartography
- ☐ Natural sciences
- ☒ Engineering and technology
- ☐ Medical and health sciences
- ☐ Meteorology
- ☐ Navigation systems
- ☐ Sea vessels
- ☐ Topography

A development platform has been created as part of the DOME 4.0 project, European Union's Horizon 2020 Research and Innovation (grant agreement No 953163). The demonstrator is for research and innovation at maturity level (~TRL 4). The project consortium is not responsible for any omissions or errors of any information posted on the development platform and shall not be responsible for any decisions made based on such information.

Connecting to DOME

○ Results (dome40.io)



Search Results

Keywords: benzaldehyde
Creator: IDMT

Search

benzaldehyde

Filter

Topic

- ☐ Cartography
- ☐ Natural sciences
- ☒ Engineering and technology
- ☐ Medical and health sciences
- ☐ Meteorology
- ☐ Navigation systems
- ☐ Sea vessels
- ☐ Topography

iDMT

Innovation Centre in Digital Molecular Technologies (iDMT) aims to create a digital transformation in the chemical industry by helping companies make the transition to automated discovery processes and machine learning.

Metadata

```
{
  "Dataset": [
    "Dataset of Experimental setup and results"
  ],
  "IssueDate": "2021-07-11",
  "License": "All Rights Reserved",
  "Title": "IDMT lab data",
  "URL": "https://idmt.dome40.io/KB/?query=\n",
  "dataCreator": "IDMT",
  "dataPublisher": "IDMT_LABS",
  "keyword": "benzaldehyde"
}
```

Data

```
[
  {
    "@id": "http://www.theworldavatar.com/kg/ontos\nhttp://www.w3.org/2004/02/skos/core#altLabel",
    {
      "@value": "benzaldehyde"
    }
  }
]
```

Free Platform

true

Domain

ENGINEERING_AND_TECHNOLOGY

Offers

EXPERIMENTAL_DATA

Home Page

<https://www.idmtonline/>

FAIR score(s)

[FOOFS!](#) score: 4%

Query URL

<https://nextgen.dome40.io/api/discover/results/IDMT>

- Data in knowledge graph may be exported into TTL files which can be uploaded to DOME

```
@prefix ontogoal: <https://www.theworldavatar.com/kg/ontogoal/> .
@prefix placeholder: <https://www.theworldavatar.com/kg/ for species/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix saref: <https://saref.etsi.org/core/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

default1:GoalSet_1 a ontogoal:GoalSet ;
    ontogoal:hasGoal default1:Goal_1,
        default1:Goal_2 ;
    ontogoal:hasRestriction default1:Restriction_1 .

lab1:HPLCMethod_Aldol a OntoHPLC:HPLCMethod ;
    rdfs:comment "HPLCMethod records the operation details of an HPLC job, e.g. HPLC analysis
was performed using a XXXX system equipped with a XXXX pump, XXXX column and XXXX variable
wavelength detector (VWD)... https://doi.org/10.1002/cmtd.202000044"^^xsd:string ;
    OntoHPLC:hasResponseFactor lab1:ResponseFactor_2,
        lab1:ResponseFactor_4,
        lab1:ResponseFactor_5 ;
    OntoHPLC:hasRetentionTime lab1:RetentionTime_1,
        lab1:RetentionTime_2,
        lab1:RetentionTime_4,
        lab1:RetentionTime_5 ;
    OntoHPLC:retentionTimeMatchThreshold 1.18e-01 ;
    OntoHPLC:usesInternalStandard lab1:PhaseComponent_InternalStandard .

lab1:Laboratory_Dummy a OntoLab:Laboratory ;
    OntoLab:contains lab1:AutoSampler_Dummy,
        lab1:AutomatedRxnPlatform_Dummy,
        lab1:BFR_Dummy,
        lab1:HPLC_Dummy,
        lab1:VapourtecR2_1_Dummy,
        lab1:VapourtecR2_2_Dummy,
        lab1:VapourtecR2_3_Dummy,
        lab1:VapourtecR4_Dummy,
        lab1:VapourtecRS400_Dummy .
```



Explore ▾ Register ▾ Showcases About Contact

Search for Data

example keyword: carbon

[Advanced Search](#)

Register Catalog Data
Register DOME 4.0 Connector
Register DOME 4.0 Tool/Service
Register API Specification
Upload Ontology

earch



Upload An Ontology

Note:

This is a pre-release, that only accepts .ttl files, a more general upload will be implemented shortly.

Choose File idmt.ttl

The data will reside in a separate graph (a name space) please provide a valid Graph URI:

https://idmt/example

For example: "https://your_organisation_uri/yourname/dataname", replace organisation etc as appropriate.

Upload

Uploading to DOME

- Uploaded ontologies may be visualised
- nextgen.dome40.io/visualise/graph/https://idmt/example

Uploaded Ontologies

Note:

Successfully Added <https://idmt/example>, from file

http://dome40.io/dataset/data/dome4.0_core_dataset_trial0_reasoned [Visualise](#)

http://dome40.io/dataset/data/platforms_dome_core_reasoned_Hermit [Visualise](#)

<http://dome40.io/dataset/data/dome-all-data> [Visualise](#)

<https://imd.ucl.io/miso> [Visualise](#)

<https://db1> [Visualise](#)

<https://simulation/test> [Visualise](#)

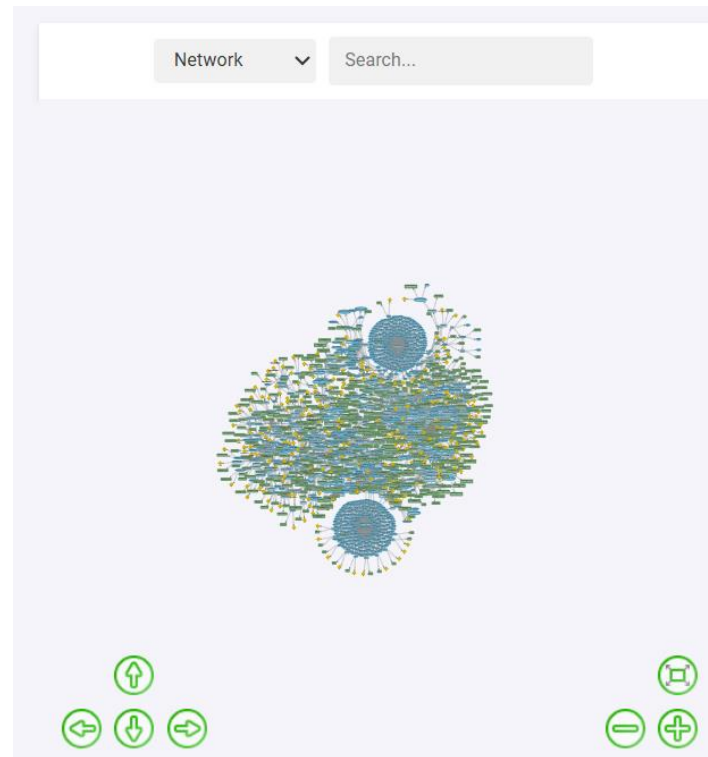
<https://ucl/ob/db1> [Visualise](#)

<https://ddmd.imd.ucl.io/owain/db1> [Visualise](#)

<https://vimmp/semantics/viso/viso-atomistic-mesoscopic> [Visualise](#)

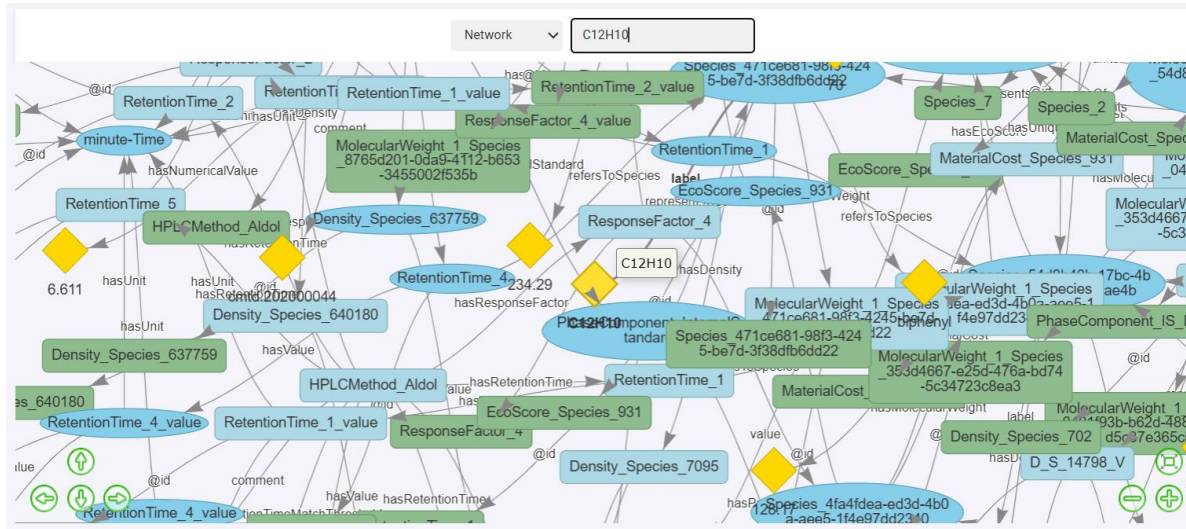
<https://idmt/example> [Visualise](#)

https://testorganisation_uri/testname/testdataname [Visualise](#)



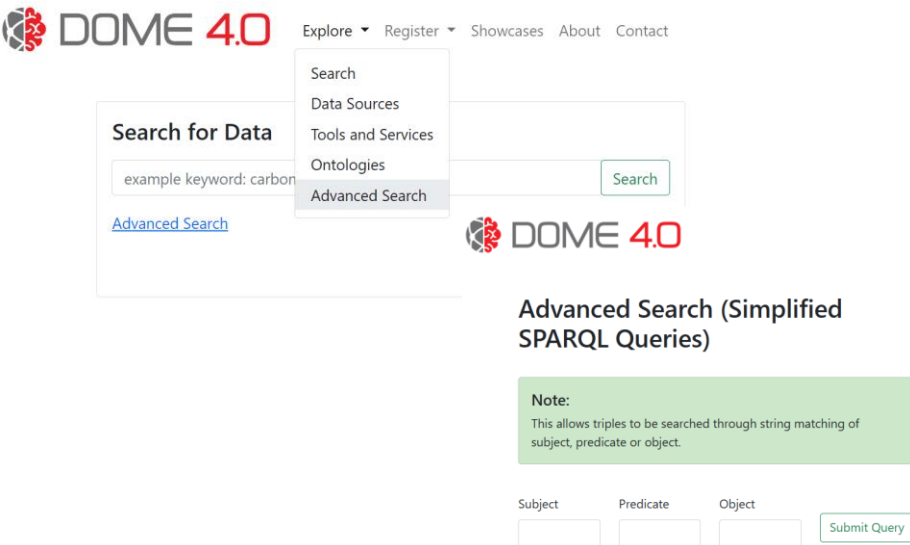
Explore the data

- String search supported in visualisation
 - Auto zoom-in



Explore the data

- Advanced search
 - Allow string matching of subject, predicate or object



Search for Data

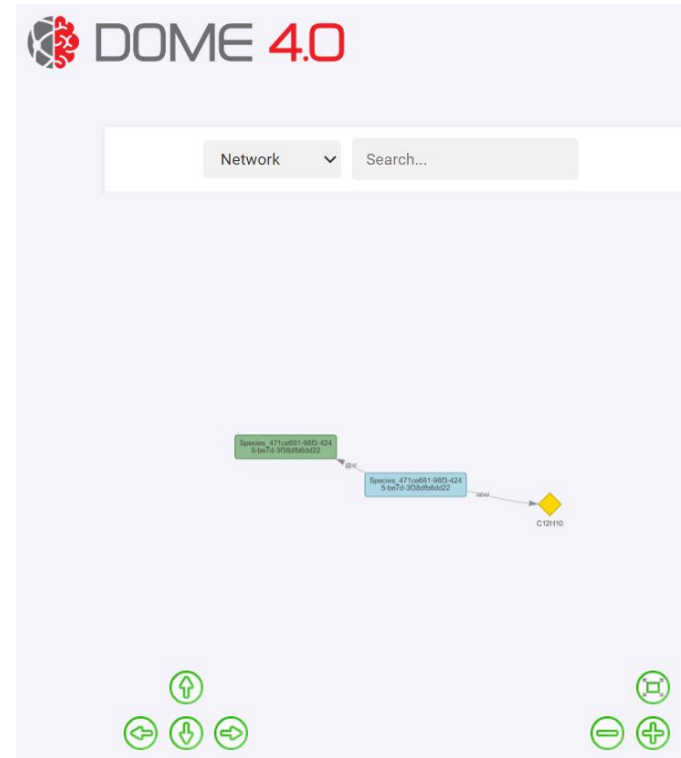
example keyword: carbon

[Advanced Search](#)

Advanced Search (Simplified SPARQL Queries)

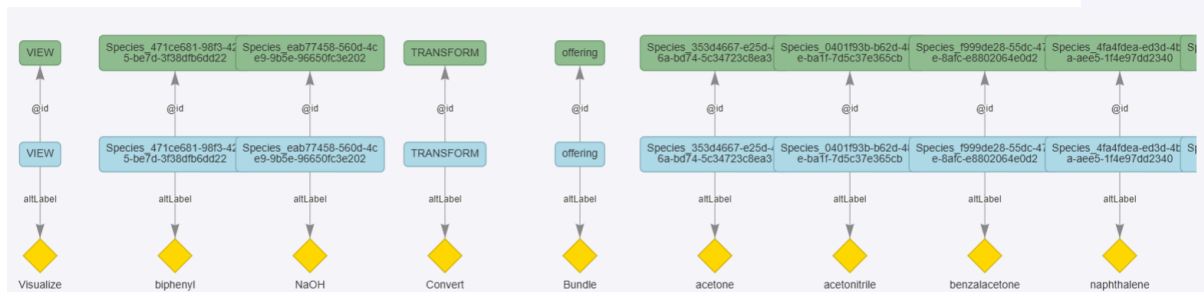
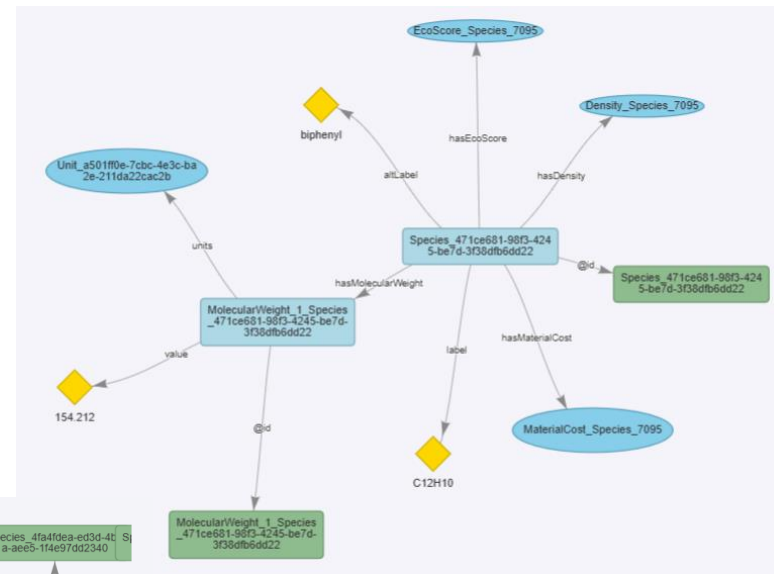
Note:
This allows triples to be searched through string matching of subject, predicate or object.

Subject Predicate Object



Explore the data

- Try it yourself!
 - nextgen.dome40.io/new-advanced-search



- iDMT data were onboarded using a connector in Hackathon 2
- Onboarded iDMT data in the form of TTL files in Hackathon 3
- Data may be explored with visualisation and advanced search



DOME 4.0



@DOME40_H2020



DOME40

www.dome40.eu



SIEMENS



EPFL

netcompany
intrasoft



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 953163