DOME 4.0

www.dome40.eu

Showcase 1 hack

DOME Hackathon 3

Chung Ting Lao

CMCL

12th Sep 2024

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

List of content

1. Setup
2. Background
3. Tutorial
4. Hack
5. Summary

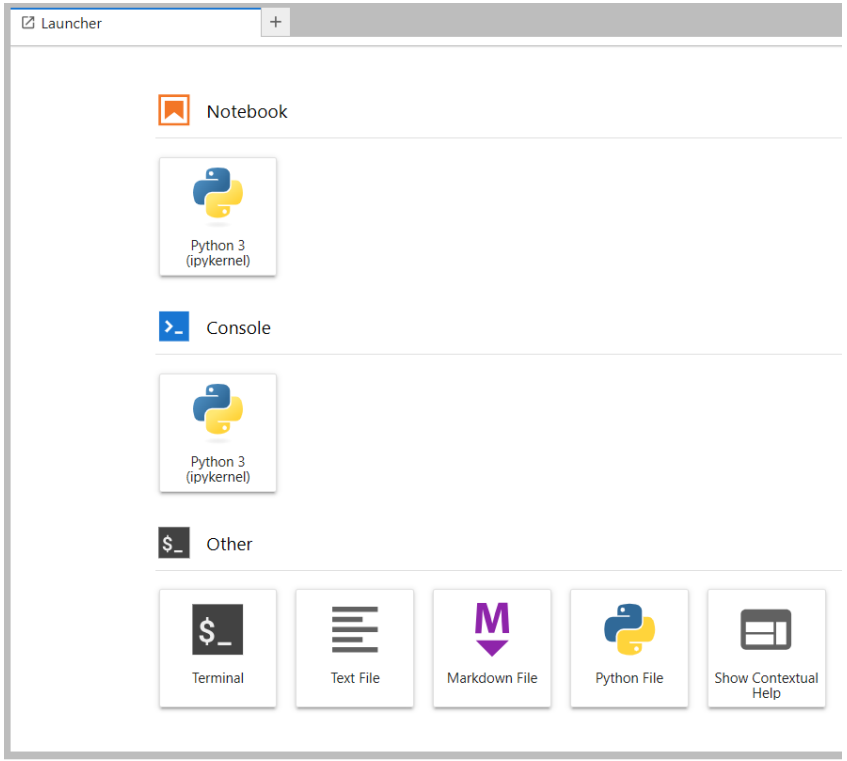
Setup

Set your machine up…

- This exercise is designed to use a docker container of Jupyter Notebook. You are welcome to use other setups.
- Jupyter Notebook + Docker
  - On windows, install WSL2
    - Install WSL | Microsoft Learn
  - Then install docker desktop
    - Get Docker | Docker Docs
  - Start docker

Set your machine up…

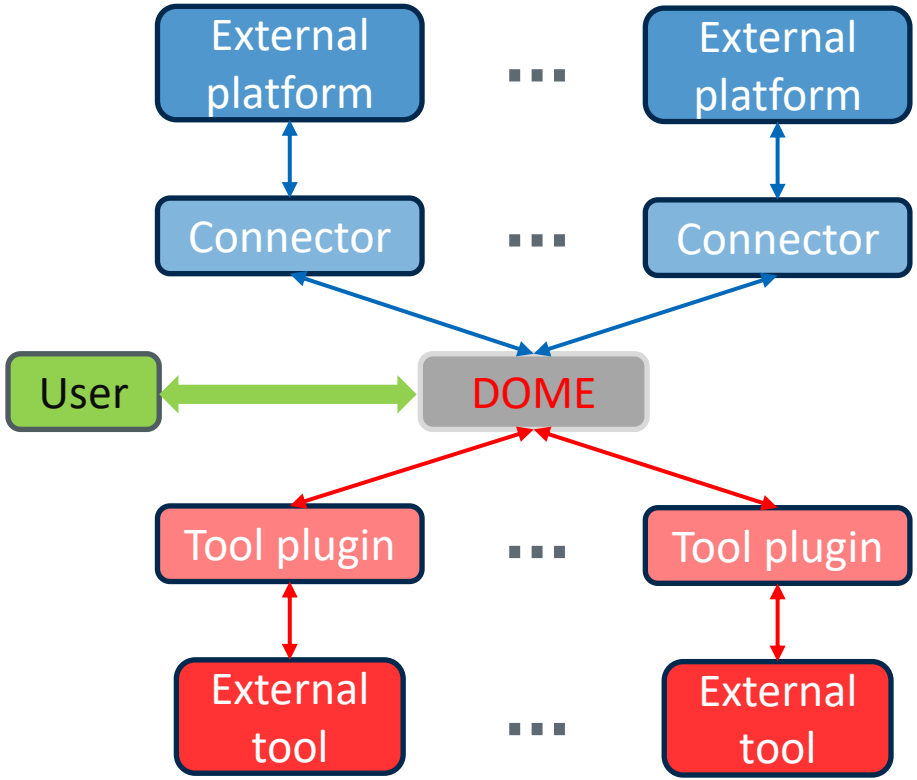
- Download prepared files into a local folder
  - hackathon-3/hack-sc1 at hack-sc1 · DOME-4-0/hackathon-3 (github.com)
  - In the folder, run **docker compose build** in (WSL) terminal to build the docker image
    - This can take a while!
  - After that, run **docker compose up –d** in (WSL) terminal to start the container
  - In internet browser, go to **localhost:8888**
    - Type in password as stored in **password.txt**



Background

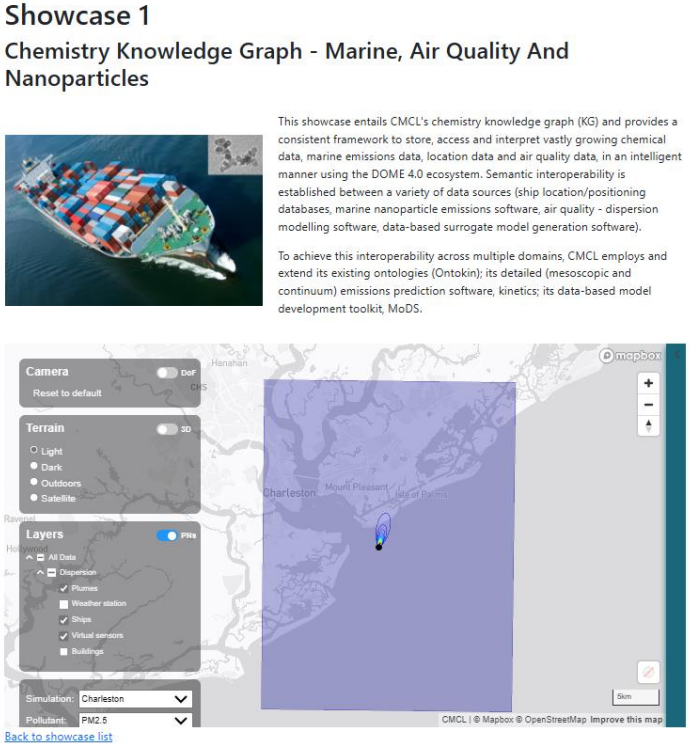
How DOME works

- Connectors enable data from external platforms to be discoverable through DOME
- Tool plugins pass data from DOME to external tool for further processing
- Typical operation
  1. User search on DOME
  2. DOME returns data
  3. Links of relevant tools are provided with individual datasets

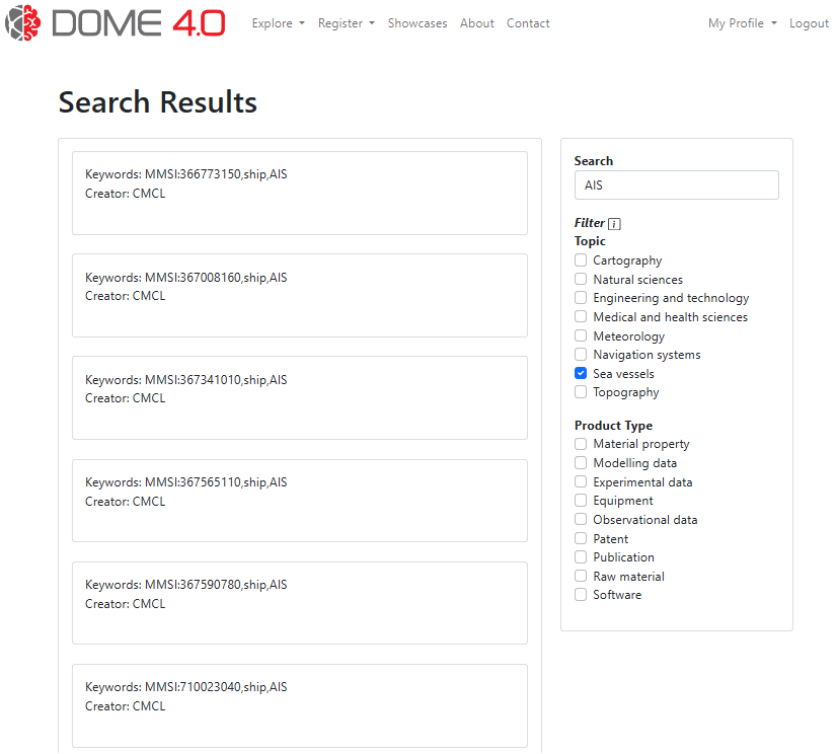


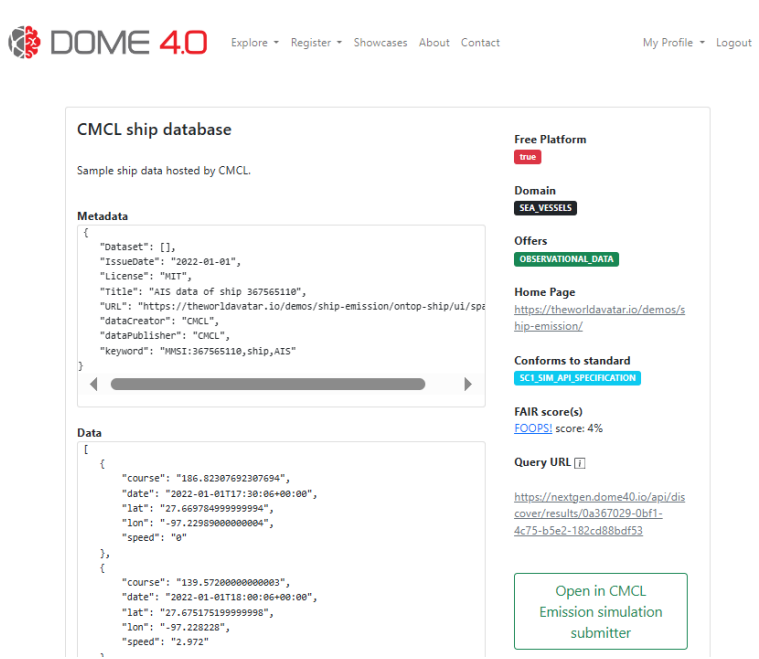
Showcase 1

- nextgen.dome40.io/showcase?show_case_id=1
  - Login with Github account
- **Ship connector** provides ship location and speed data through DOME
- **Emission simulation submitter** consumes ship data and produce emission dispersion data
- **Emission connector** provides emission dispersion data through DOME
- **Emission simulation viewer** visualises emission dispersion data



Showcase 1



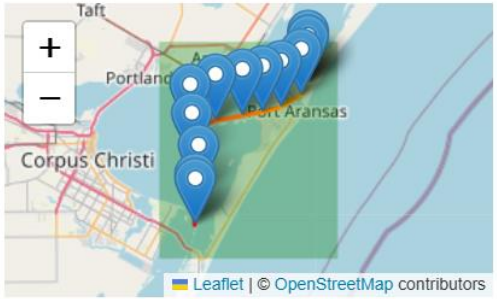


- **Search "AIS"**
- **A button for simulation submitter**

Showcase 1

DOME 4.0

Location data of ship MMSI:367565110

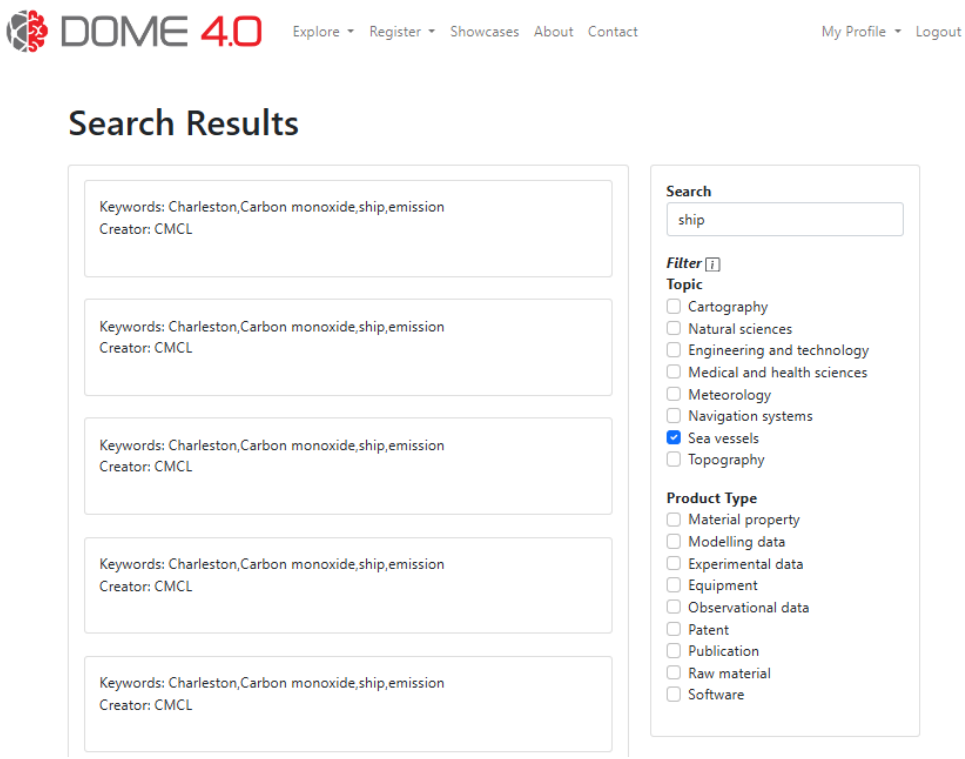


Start new simulation

Name of simulation. This will be MMSI:

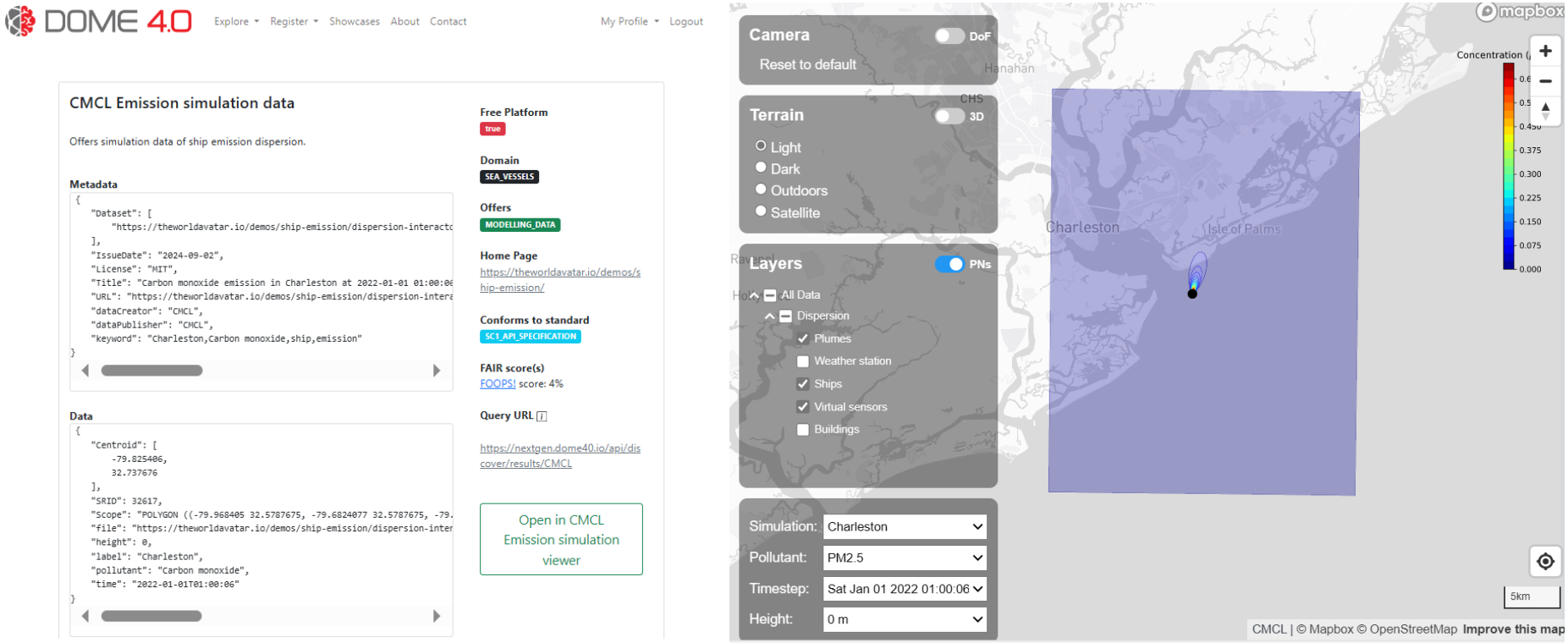
Number of timesteps. All available time

Submit



- simulation submitter page
- Search "ship" to find emission data

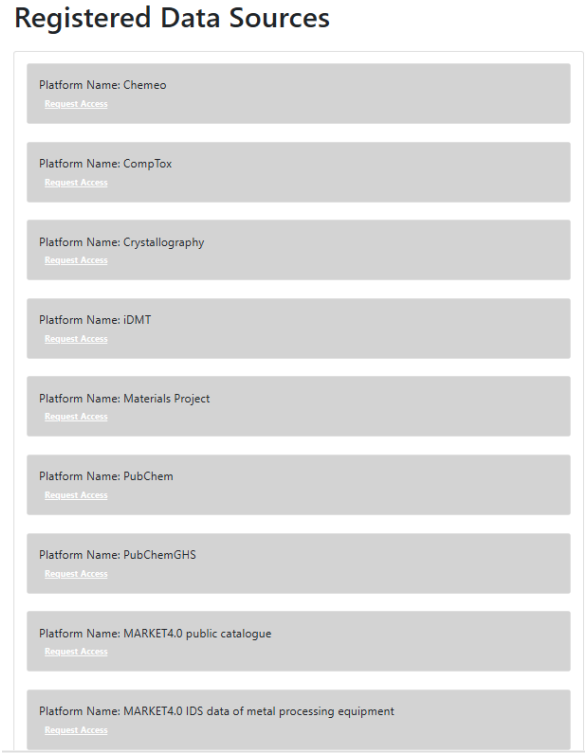
Showcase 1



- **A button for simulation viewer**
- **Visualise emission data**

Next step

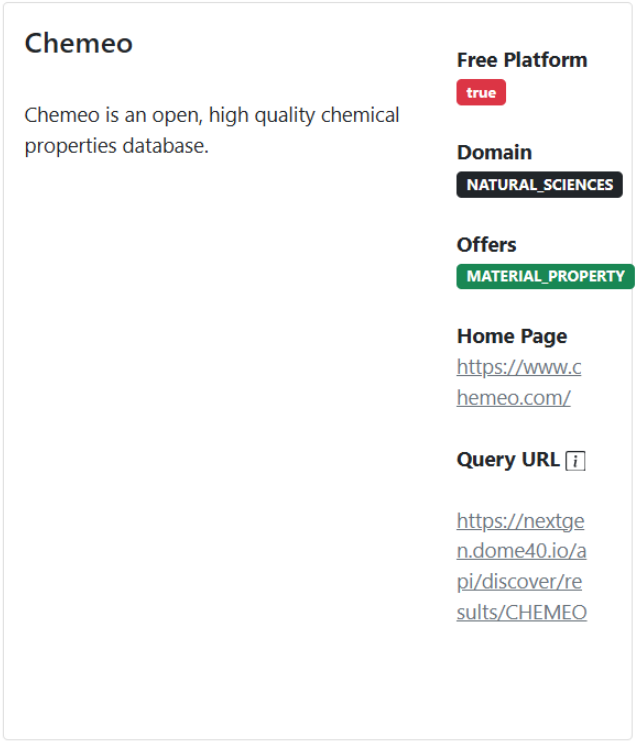
- Can we make our own viewer that goes beyond the current show case?
  - Can we present the data differently?
  - Can we connect to other datasets?
    - Platform providers (dome40.io)
    - Other data sources?



Tutorial

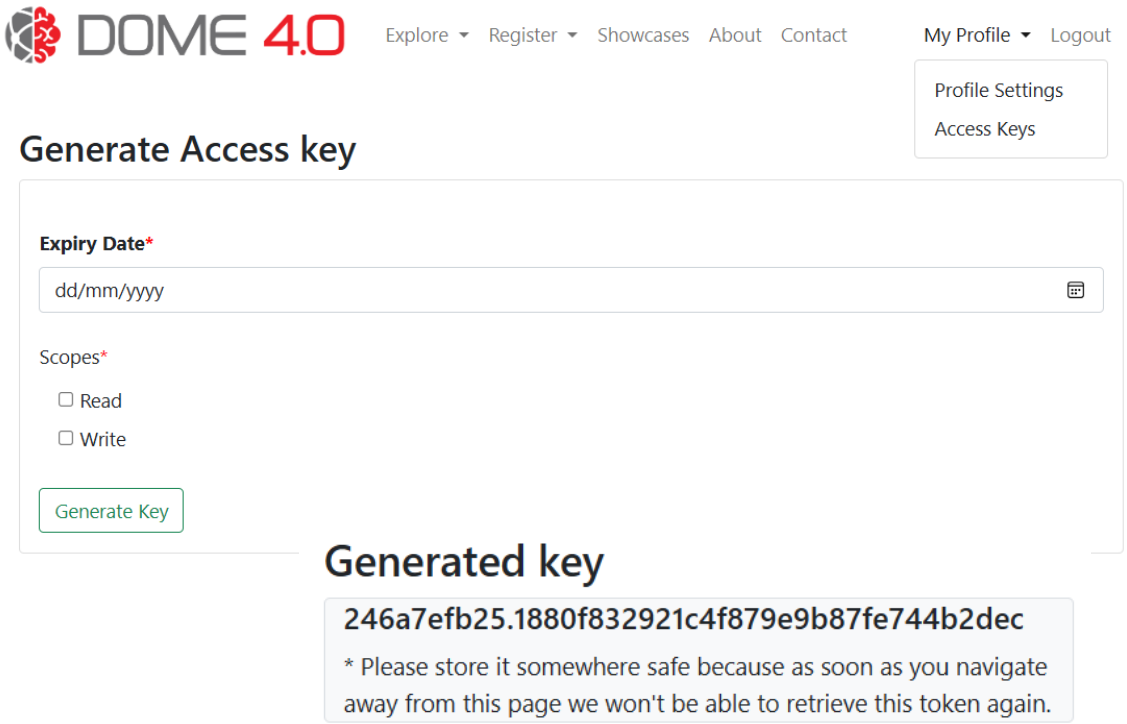
How to query DOME?

1. Find a connector on DOME
2. Find the query URL



How to query DOME?

3. Create an API key for identification
4. Save the generated key



How to query DOME?

5. Send a HTTP request to the query URL with
   - apikey
   - search_string
6. Receive **metadata** and **data**

dome_connector.py

```python
import requests

class Connector:
    def __init__(self, dome_url, api_key, provider_id):
        self.dome_url = dome_url
        self.api_key = api_key
        self.provider_id = provider_id

    def get_data(self, search_string):

        res_conn_query = requests.get(
            f"{self.dome_url}api/discover/results/{self.provider_id}",
            headers={'apikey': self.api_key},
            params={'search_string': search_string},
            timeout=10,
        )

        try:
            return res_conn_query.json()
        except:
            return res_conn_query.content

if __name__ == "__main__":
    API_KEY = "246a7efb25.1880f832921c4f879e9b87fe744b2dec"
    connector = Connector("https://nextgen.dome40.io/", API_KEY, "CHEMEO")
    result = connector.get_data("Carbon monoxide")
    for k, v in result.items():
        print(f'*****{k}*****')
        print(v)
```

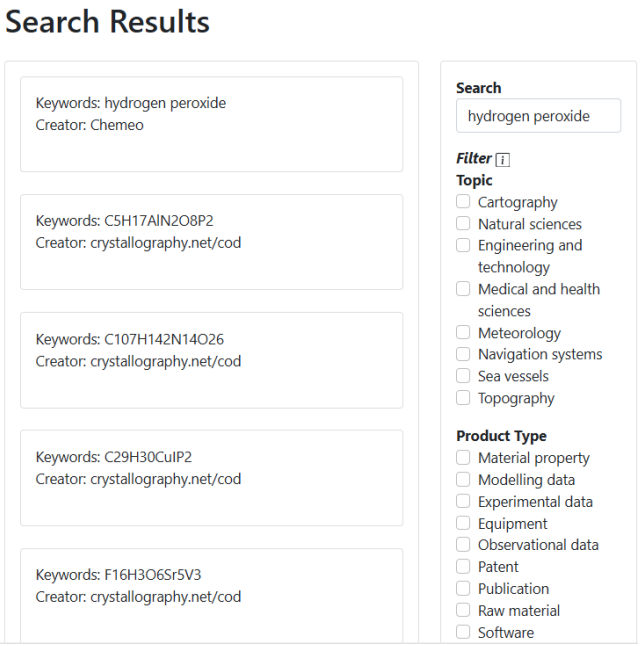
dome_connector.py

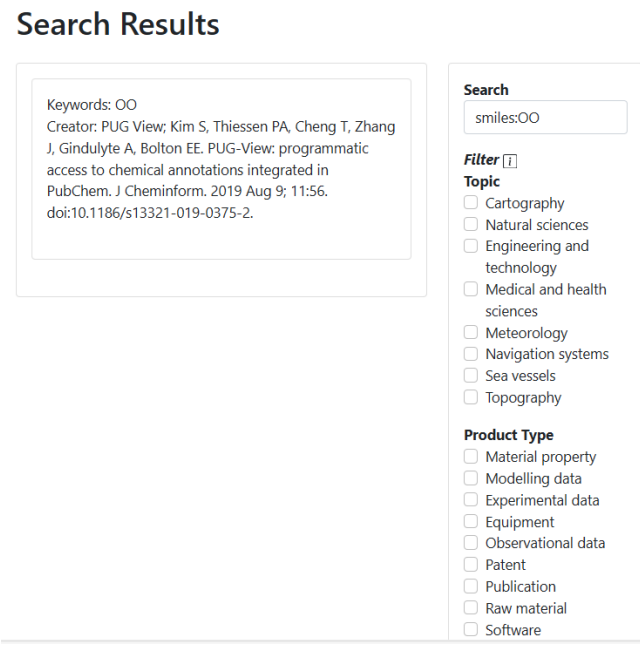
```
(base) jovyan@5ba41a8d74c6:~/work$ python dome_connector.py
*****metadata*****
[{'Dataset': ['dataset'], 'IssueDate': '0000-00-00', 'License': 'Chemeo
'Carbon monoxide'}]
*****data*****
[{'id': '23-012-2', 'rev': 1, 'compound': 'Carbon monoxide', 'other_name
de)', 'Carbonio (ossido di)', 'Exhaust gas', 'Flue gas', 'Kohlenmonoxid'
-08-0', 'inchi': 'InChI=1S/CO/c1-2', 'inchikey': 'UGFAIRIUMAVXCW-UHFFFAO
hEUgAAAPoAAAD6CAIAAAAHjs1qAAAHR0lEQVR4nO3cTUhUaxzH8b+vaFJ0wSQJDcPq2hSWEC
5nhm0hzHMUBD+nQvAEgdcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcoc
cQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcoc
cocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQcocQ
+3FSt8XKGTKvv27fPxr/H72bVsmWOm9RMMTt7BwIBz+LBTWvrLnRQWOnv3Ol1dfkSY5jhOav
Z+npFgpZMGgfPsRsUFZmDx4kcIA0GXKH/xzHNm2ye/fce6qqrL7eqqstJ8e9MxKxtjY7cyZm
bOu8NnFy+6rVdXW5wn6LZssV27omPHscbGpKwlCc/u4XC4u7s7EAhkZ2cnZU2YOb5/t6IiC4
N4dCITMLBoO3bt3yviDMKI8fu+P16xOe/iN3M3vyxPtyvOZePO71pdjzaw1mmvZ2d7xyZcLl
0dj3AiEn76+NEdz56d8PS0NPeamdHRuC62mRC5w0+Z446WR0ensofxl9llZXlcThJyD4fDnZ
jTt+/jzh6Z2dP9/VVHHeHX6qrHTHd+5YU5NlZCQw/fZtd1xV5X05nHeHn8bGrLTUfnxos7XV
fP8vPtwgX35rFj1tAwyXN8c7Pt3OnePH7cysqSs5hkfQoQmMj27TFfJrNqlXPzpvPlS8w2w8
/+TqVSsq+uWfOo6dP28NDXH9enXWLDt0yE6ciLnqxrtkvUxMii/Nm/k/8Xxp3tCQU1/vTPCP
MIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuU
uUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUM
MIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUMIuUM
om\nMore info - https://www.chemeo.com/cid/23-012-2\n  2  1  0  0  0  0
0   -0.0000    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0\n  1  2  2
'}], 'affp': [{'v': 594, 's': 'nist-webbook', 'c': 'at C; HL', 'r': 'Hur
], 'basg': [{'v': 562.8, 's': 'nist-webbook', 'c': 'at C; HL', 'r': 'Hur
], 'dm': [{'v': 0.1, 's': 'kdb-pure'}], 'ea': [{'v': 1.32608, 's': 'nist
d Franklin, 1976'}], 'gf': [{'v': -137.4, 's': 'kdb-pure'}], 'gr': [{'v'
}, {'v': 48.1, 's': 'nist-webbook', 'c': 'gas phase; «DELTA»rH>, DG>; M'
-110.6, 's': 'kdb-pure'}, {'v': -110.53, 'e': 0.17, 's': 'nist-webbook',
ok', 'r': 'N/A'}], 'hfpiz': [{'v': 1238, 's': 'nist-webbook', 'r': 'N/A'
y and entropy were identified with the enthalpy and entropy of activatio
{'v': 159.4, 's': 'nist-webbook', 'c': 'solvent: n-Decane+cyclohexane mi
or the reactions of Cr(CO)6(solution) with a phosphine and an amine. The
6, 'e': 5.9, 's': 'nist-webbook', 'c': 'solvent: Decalin; The reaction e
```

Hack

Task 1: get data from DOME

- Ship emission data from CMCL
- Chemical species data from Chemeo
- Hazard classification from PubChemGHS
  - Search string must be in format smile:XXX where XXX is the SMILES structure
- HINT:
  - Chemo provides SMILES





***Pick carbon monoxide or sulphur dioxide**

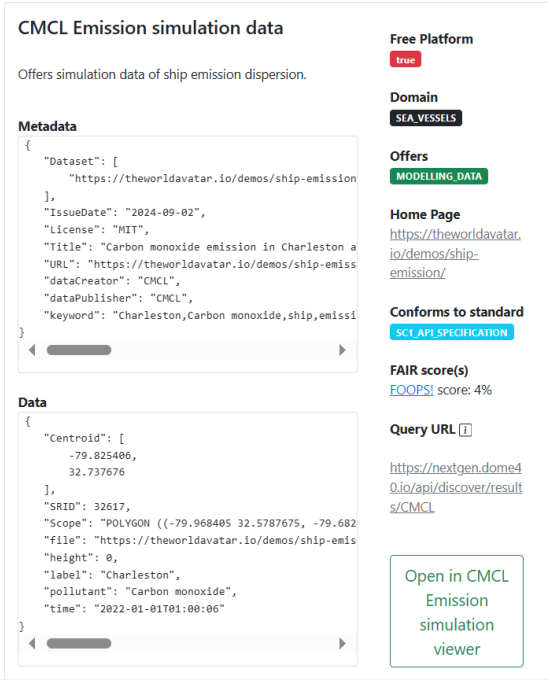
Task 1: get data from DOME

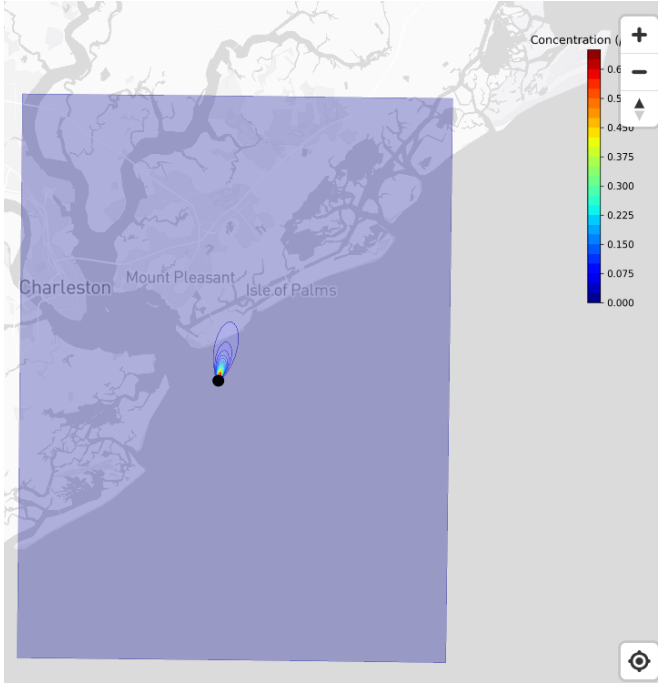
- Reference result
  - Metadata and data in JSON/ python dictionary

```
*****Emission metadata*****
{'Dataset': ['https://theworldavatar.io/demos/ship-emission/dispersion-interactor/GetRaster?filename...
*****Emission data*****
{'file': 'https://theworldavatar.io/demos/ship-emission/dispersion-interactor/GetRaster?filename=325...
*****Chemical metadata*****
[{'Dataset': ['dataset'], 'IssueDate': '0000-00-00', 'License': 'Chemeo EULA', 'Title': 'Title', 'UR...
*****Chemical data*****
[{'id': '23-012-2', 'rev': 1, 'compound': 'Carbon monoxide', 'other_names': ['CARBON OXIDE', 'CARBON...
*****Hazard metadata*****
[{'Dataset': ['smiles:[C]=O'], 'IssueDate': '[]', 'License': 'https://www.nlm.nih.gov/web_policies.h...
*****Hazard data*****
[{'712': {'40': ['H301', 'H311', 'H314', 'H317', 'H318', 'H330', 'H331', 'H341', 'H350', 'H351'], '4...
```

Task 2: process data

- Ship emission data
  - "file" provides hyperlink to raw simulation data
  - Process it for final visualisation
  - Hint:
    - The file is in GeoTIFF format
    - Pre-installed python packages may be useful:
      - **rasterio**
      - **matplotlib**

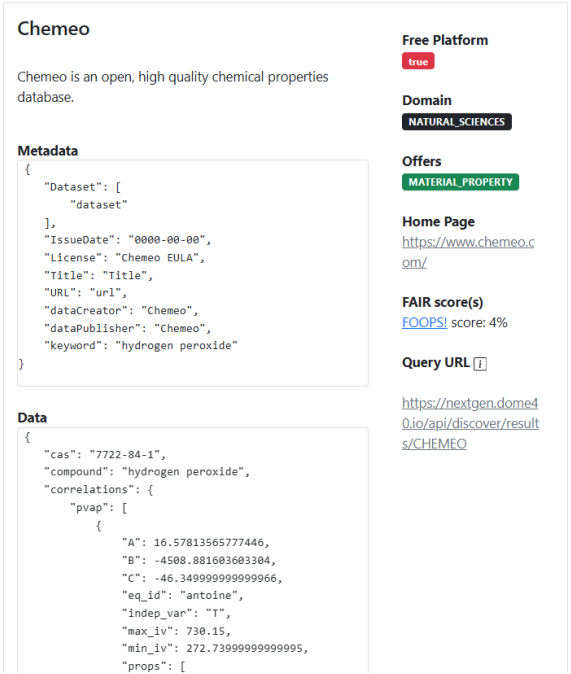




DOME 4.0 — H2020 Grant Agreement No. 953163

Task 2: process data

- Chemeo data
  - Process data for final visualisation
  - HINT:
    - “drawing” is the chemical structure encoded base64 string
    - Can also plot graphs



Chemeo

Chemeo is an open, high quality chemical properties database.

Metadata


```
{
    "Dataset": [
        "dataset"
    ],
    "IssueDate": "0000-00-00",
    "License": "Chemeo EULA",
    "Title": "Title",
    "URL": "url",
    "dataCreator": "Chemeo",
    "dataPublisher": "Chemeo",
    "keyword": "hydrogen peroxide"
}
```


Data


```
{
    "cas": "7722-84-1",
    "compound": "hydrogen peroxide",
    "correlations": {
        "pvap": [
            {
                "A": 16.57813565777446,
                "B": -4508.881603603304,
                "C": -46.349999999999966,
                "eq_id": "antoine",
                "indep_var": "T",
                "max_iv": 730.15,
                "min_iv": 272.73999999999995,
                "props": [
```


Free Platform

true

Domain

NATURAL_SCIENCES

Offers

MATERIAL_PROPERTY

Home Page

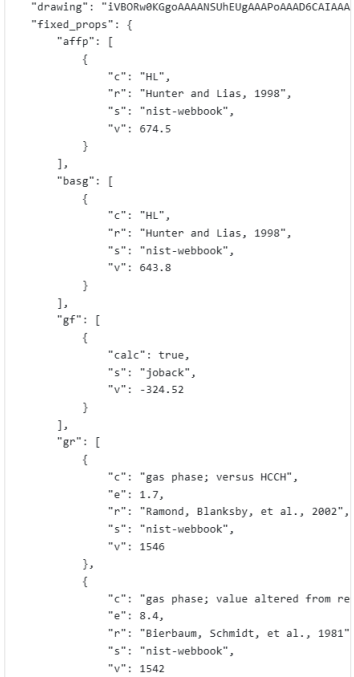
https://www.chemeo.com/

FAIR score(s)

FOOPS! score: 4%

Query URL

https://nextgen.dome40.io/api/discover/results/CHEMEO



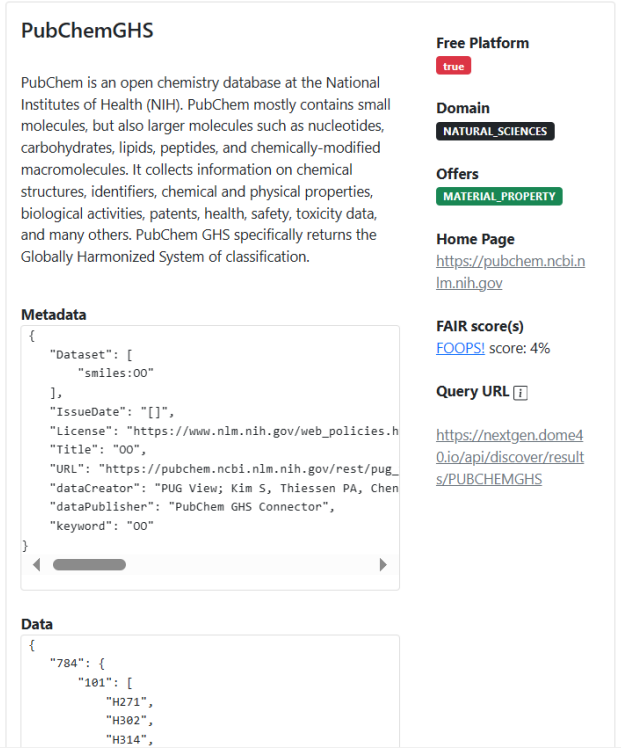

```
"drawing": "iVBORw0KGgoAAAANSUhEUgAAAPoAAAD6CAIAAA
"fixed_props": {
    "affp": [
        {
            "c": "HL",
            "r": "Hunter and Lias, 1998",
            "s": "nist-webbook",
            "v": 674.5
        }
    ],
    "basg": [
        {
            "c": "HL",
            "r": "Hunter and Lias, 1998",
            "s": "nist-webbook",
            "v": 643.8
        }
    ],
    "gf": [
        {
            "calc": true,
            "s": "joback",
            "v": -324.52
        }
    ],
    "gr": [
        {
            "c": "gas phase; versus HCCH",
            "e": 1.7,
            "r": "Ramond, Blanksby, et al., 2002",
            "s": "nist-webbook",
            "v": 1546
        },
        {
            "c": "gas phase; value altered from re
            "e": 8.4,
            "r": "Bierbaum, Schmidt, et al., 1981"
            "s": "nist-webbook",
            "v": 1542
```



```
            "s": "nist-webbook",
            "v": 272.26
        }
    ],
    "vc": [
        {
            "calc": true,
            "s": "joback",
            "v": 0.0735
        }
    ]
},
"formula": "H2O2",
"id": "67-803-6",
"inchi": "InChI=1S/H2O2/c1-2/h1-2H",
"inchikey": "MHAJPDPJQMAIIY-UHFFFAOYSA-N",
"mol2d": "hydrogen peroxide\nChemeo Renderer - htt
"mol3d": "  mol1915\r\n\r\n  created with ArgusLab
"mw": 34.0147,
"n_mixtures": 1,
"other_names": [
    "ALBONE",
    "DIHYDROGEN DIOXIDE"
],
"rev": 1,
"smiles": "OO",
"src": {
    "KDB_vapor_pressure": {
        "db": "KDB-pvap",
        "title": "KDB Vapor Pressure Data",
        "url": "https://www.cheric.org/research/kd
    },
    "crippen": {
        "db": "Crippen",
        "desc": "<p class=\"authors\">S.A. Wildman
        "title": "Prediction of Physicochemical Pa
        "url": "http://pubs.acs.org/doi/abs/10.102
    },
    "crippen_log10ws": {
        "db": "Crippen/Chem\u00e9o",
```

Task 2: process data

- PubChemGHS data
  - Process data for final visualisation
  - HINT:
    - Not all data are shown on the web interface
    - Check “URL” in Metadata



PubChemGHS

PubChem is an open chemistry database at the National Institutes of Health (NIH). PubChem mostly contains small molecules, but also larger molecules such as nucleotides, carbohydrates, lipids, peptides, and chemically-modified macromolecules. It collects information on chemical structures, identifiers, chemical and physical properties, biological activities, patents, health, safety, toxicity data, and many others. PubChem GHS specifically returns the Globally Harmonized System of classification.

Metadata


```
{
    "Dataset": [
        "smiles:OO"
    ],
    "IssueDate": "[]",
    "License": "https://www.nlm.nih.gov/web_policies.h
    "Title": "OO",
    "URL": "https://pubchem.ncbi.nlm.nih.gov/rest/pug_
    "dataCreator": "PUG View; Kim S, Thiessen PA, Chen
    "dataPublisher": "PubChem GHS Connector",
    "keyword": "OO"
}
```


Data


```
{
    "784": {
        "101": [
            "H271",
            "H302",
            "H314",
```


Free Platform: true

Domain: NATURAL_SCIENCES

Offers: MATERIAL_PROPERTY

Home Page: https://pubchem.ncbi.nlm.nih.gov

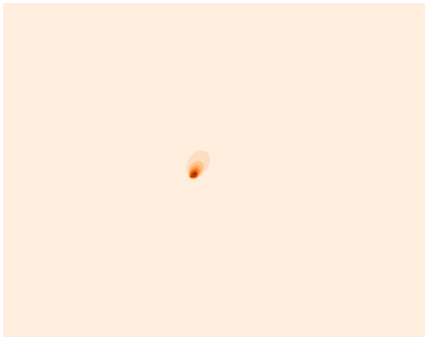
FAIR score(s): FOOPS! score: 4%

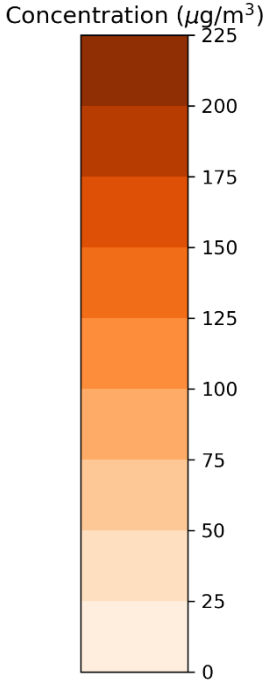
Query URL: https://nextgen.dome40.io/api/discover/results/PUBCHEMGHS


```
            "H314",
            "H332",
            "H335"
        ],
        "67": [
            "H314",
            "H318",
            "H335",
            "H402",
            "H412"
        ],
        "83": [
            "H271",
            "H302",
            "H311",
            "H314",
            "H318",
            "H330",
            "H331",
            "H351",
            "H370",
            "H372",
            "H400"
        ],
        "84": [
            "H271",
            "H302",
            "H313",
            "H314",
            "H318",
            "H331",
            "H361",
            "H370",
            "H372",
            "H373",
            "H401"
        ]
    }
}
```

Task 2: process data

- Expected result
  - Emission data
    - Raw data was in greyscale
    - Present it in colour
    - A colourbar to show values





DOME 4.0 — H2020 Grant Agreement No. 953163

Task 2: process data

- Expected result
  - Chemeo data
    - Save the drawing into a file
    - Alternatives:
      - Plot a graph of its properties
  - GHS data
    - Hazard code and description

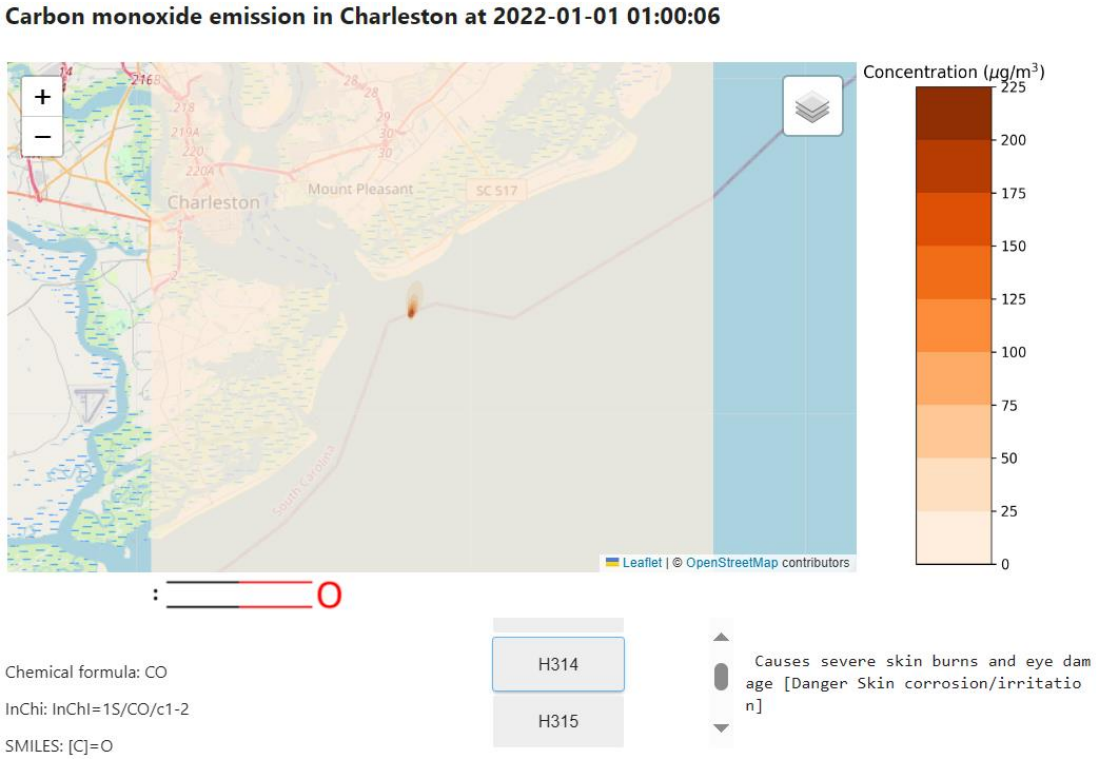
```
H220: Extremely flammable gas [Danger Flammable ga...
H227: Combustible liquid [Warning Flammable liquid...
H280: Contains gas under pressure; may explode if ...
H301: Toxic if swallowed [Danger Acute toxicity, o...
H302: Harmful if swallowed [Warning Acute toxicity...
H311: Toxic in contact with skin [Danger Acute tox...
H314: Causes severe skin burns and eye damage [Dan...
H315: Causes skin irritation [Warning Skin corrosi...
H317: May cause an allergic skin reaction [Warning...
H318: Causes serious eye damage [Danger Serious ey...
H319: Causes serious eye irritation [Warning Serio...
H330: Fatal if inhaled [Danger Acute toxicity, inh...
H331: Toxic if inhaled [Danger Acute toxicity, inh...
H334: May cause allergy or asthma symptoms or brea...
H341: Suspected of causing genetic defects [Warnin...
H350: May cause cancer by inhalation [Danger Carci...
H351: Suspected of causing cancer [Warning Carcino...
H370: Causes damage to organs [Danger Specific tar...
H372: Causes damage to organs through prolonged or...
H401: Toxic to aquatic life [Hazardous to the aqua...
H402: Harmful to aquatic life [Hazardous to the aq...
H412: Harmful to aquatic life with long lasting ef...
```

Task 3: make it interactive

- Make it as impressive as possible!
- Ideas
  - Show data on a map, like the original viewer
  - Allow user to control showing of different data
- HINT
  - To show images on a map, **folium** may be useful
    - needs latitude-longitude → need to reproject to EPSG:4326
      - Can be done with **rasterio**
  - To enable interaction, use **IPython** and **ipywidgets**
    - ipywidgets.Button can response to user's click
    - ipywidgets.Output can display output of functions

Task 3: make it interactive

- Reference result
  - Contour shown on map using folium
  - Image from Chemeo
  - Buttons for each PubChemGHS label
    - Update description on click



Summary

Summary

- Achievements
  - Learn to get data from DOME using API
  - Created a tool that combines data from different sources from DOME
- Next steps
  - How could connectors be improved?
    - Currently, still rely on a lot of prior knowledge
    - “Links” between datasets are rigid

DOME 4.0

@DOME40_H2020 DOME40 www.dome40.eu

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