Open Government Data for Better Public Services

Data-Driven Smart Governance

Open Data Alliance
Dongpo Deng
2024/03/26

Can Open Government Data (OGD) be used to make better public services?

Data-Driven Smart Governance - the Five I

Interaction

Influence



Integration

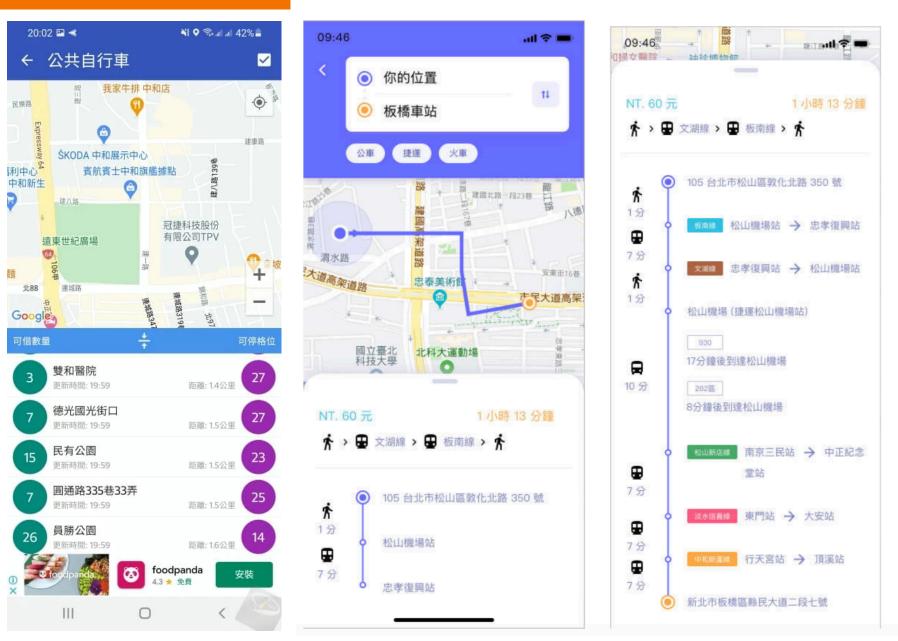
Innovation

Intelligence



- Open Government Data (OGD) is a way allowing citizens to understand the affairs of governments.
 Also, the OGD provides citizens to improve the public services.
 - Taipei City Gov. released the real-time bus tracking data. Citizens developed the Bus Tracker Apps. By their requirements
 - Ministry of Health and Welfare Taiwan openly provided the daily amount of masks in each pharmacy. Citizens can made Mask Maps allowing to know where to get masks during COVID-19
- OGD offers citizens to brain-storming for better public services or problem solutions for citizens
 - The brain storming activities, e.g. Hackathon,
 Data jam, Datapalooza

Bus Tracker APP



Mask Map



全家便利商店 台中京華店 回報目前口罩數量







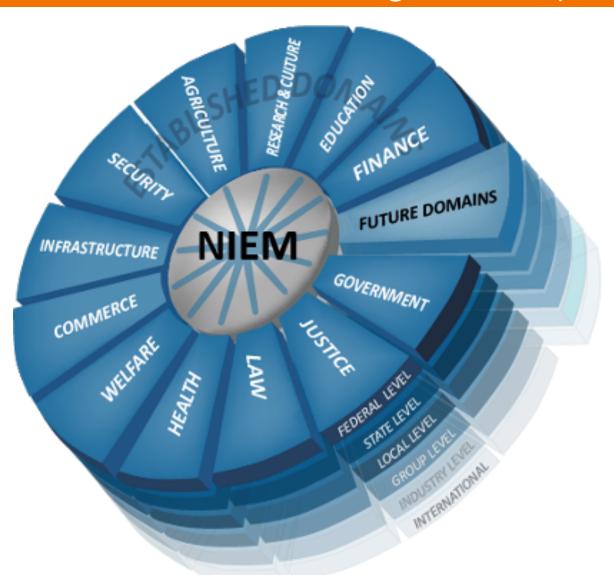




Integration

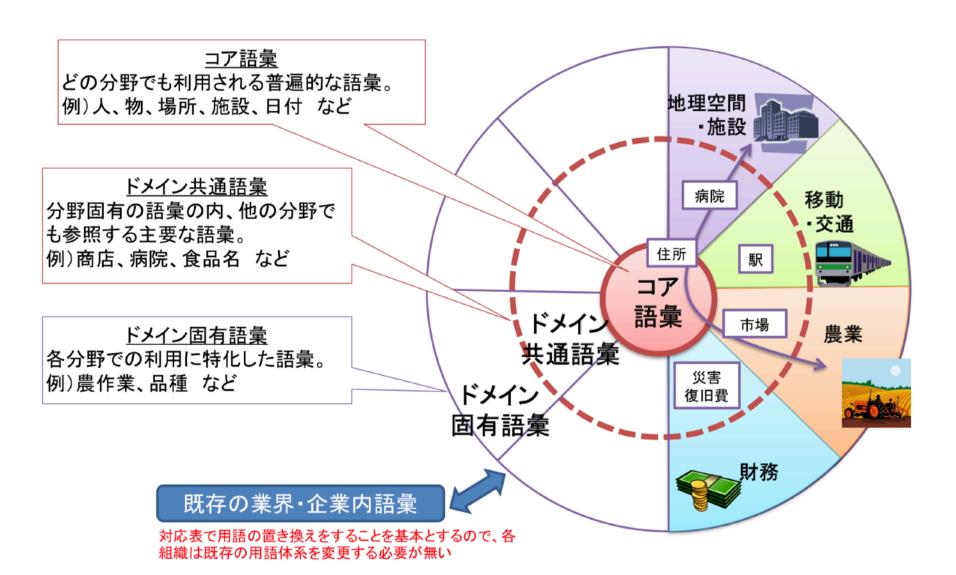
- To make better public services via using OGD, data interoperability is a crucial cornerstone
- The development of common vocabularies and data standards is a way to make OGD interoperability
- In Tim Berners-Lee's 5-stars Open Data, the top rank of open data is linked open data, which actually means the data interoperability.

National Information Exchange Model (NIEM) in US



Picture source: https:// www.sparxsystems.com.au/ domains/niem/nationalinformation-exchange-modelniem-solution-withenterprise-architect.html

Infrastructure for Multi-layer Interoperability (IMI)

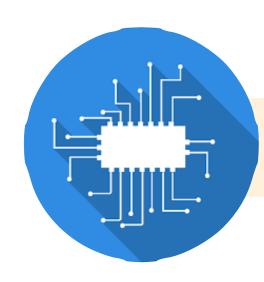




Integration

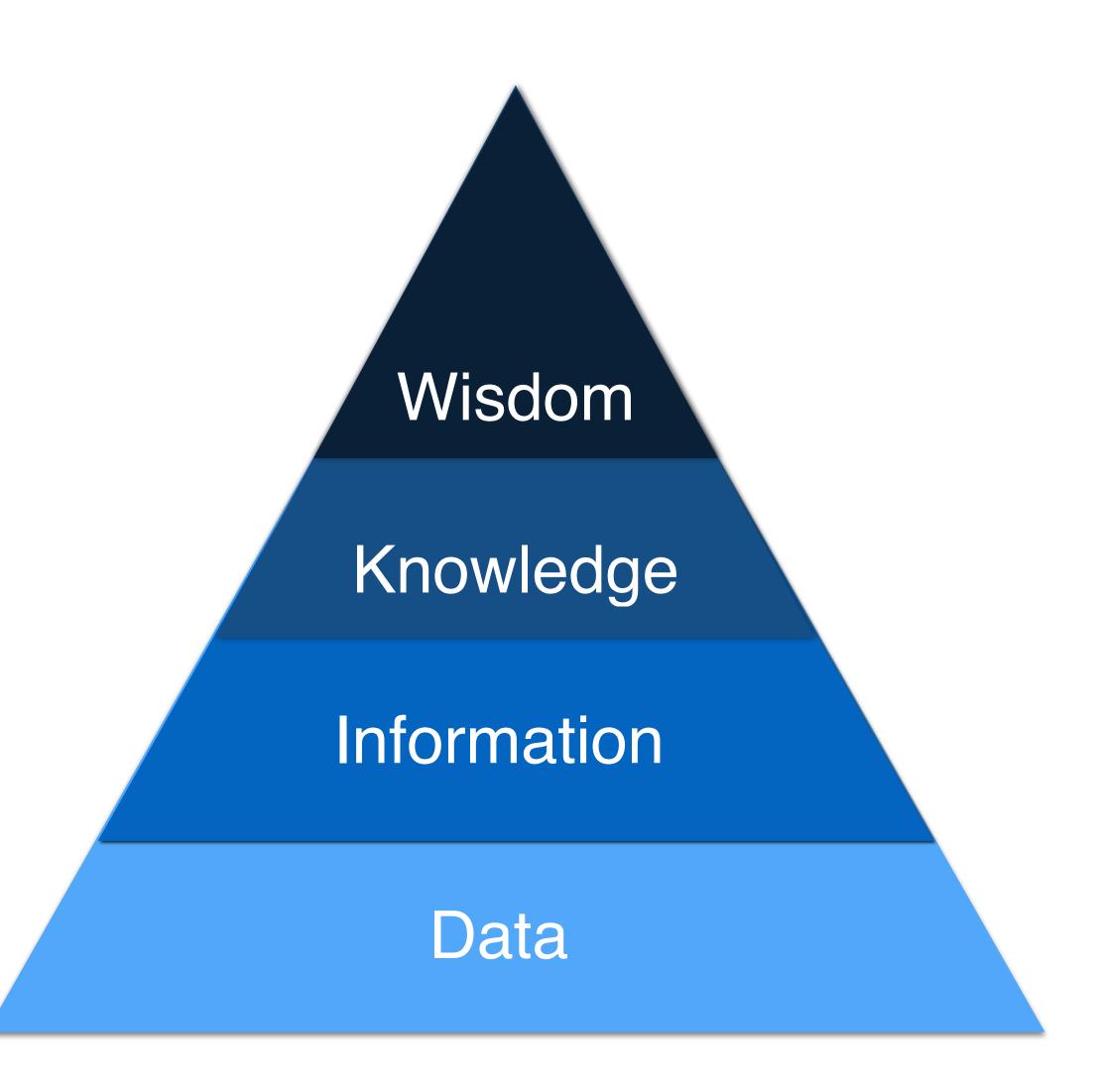
- To make better public services via using OGD, data interoperability is a crucial cornerstone
- The development of common vocabularies and data standards is a way to make OGD interoperability
- In Tim Berners-Lee's 5-stars Open Data, the top rank of open data is linked open data, which actually means the data interoperability.





Intelligence

- Data is one of the sources for intelligence
- Public sectors applied data science for public good
 - Statistics Netherlands (CBS) is partnered with Vodafone for analyzing mobile call records. From the mobility patterns, they can make better traffic management and urban planning.
- The appearance of Chat-GTP revealed the coming of AI era. How to use generative AI for open data analysis to make better public services is imperative



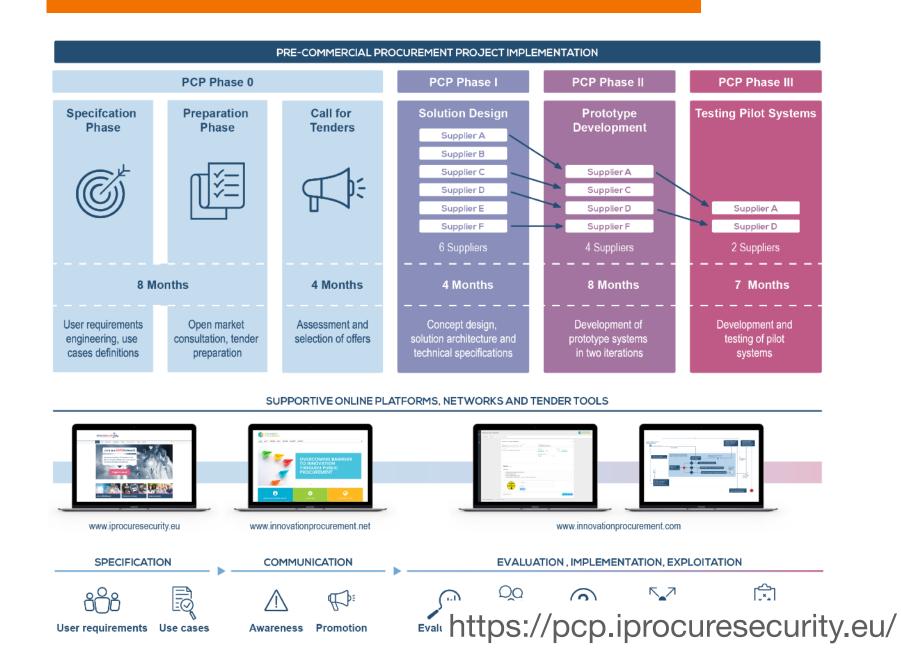


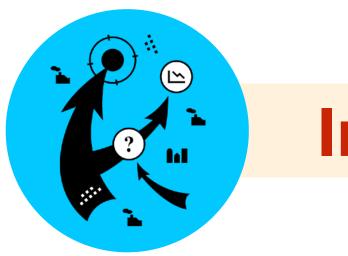
- To match the demands from public sectors, for example,
 - Procurement Innovation Network, EU
 - The Open Regulation Platform, UK
- To match the demands from citizens,
 - Living Lab, using OGD for better quality of life





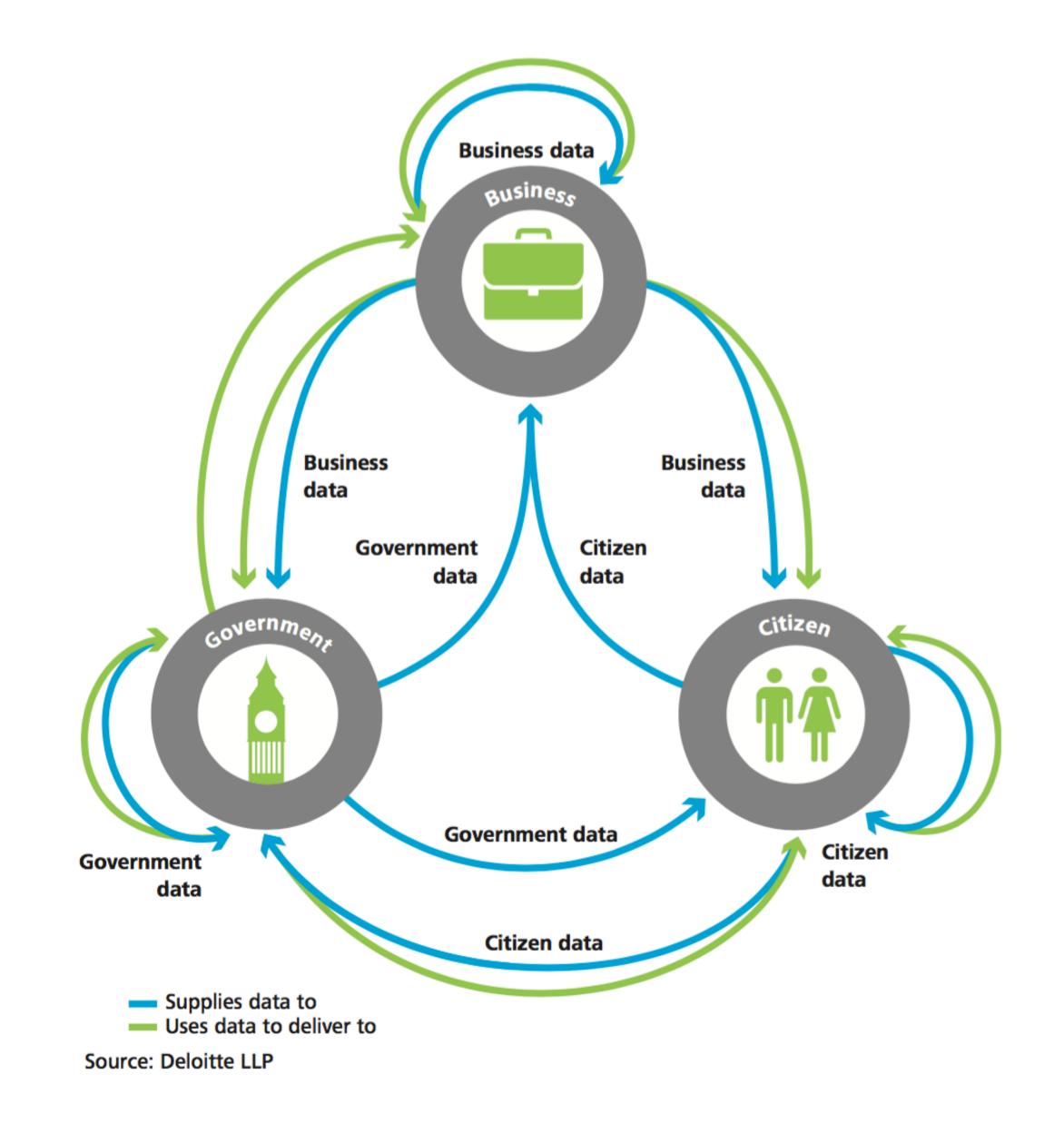
The Innovation Procurement Network





Influence

- To open more data in different domains, for example
 - NGO, open data for social good
 - Personal data, makes better public service experience (MyData policy, Taiwan)
- To change in terms of
 - Data governance
 - Decision culture
 - Data economic
- To make Public Private Partnership (PPP)



The case of Traceable Agriculture Product for better food choice

Traceable Agriculture Product (TAP) dataset

產銷履歷資料

- Providing the agriculture product information to customers for building trust
- A rare dataset around the world
- High value dataset
- JSON format but data structure is tabular
- An good example why the data cannot present its value



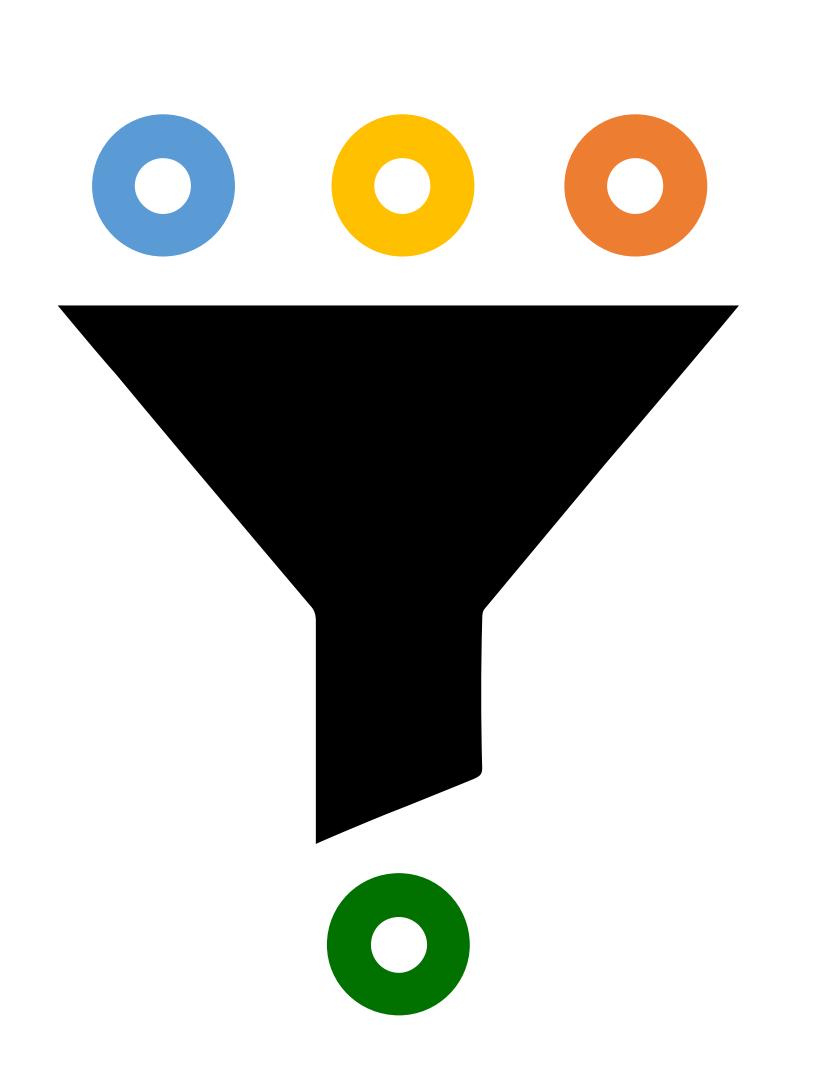
Traceable Agriculture Product (TAP) dataset

產銷履歷資料

- 1. Product Name (產品名稱)
- 2. Organization ID (組織代碼)
- 3. Producer (農產品經營業者)
- 4. Place (產地)
- 5. Farmer Name (生產者姓名)
- 6. Pack date (包裝日期/出貨日期)
- 7. Certification Name (驗證單位)
- 8. Valid Date (驗證有效日期)
- 9. Store Information (通路商資訊)
- 10. Tracecode (追蹤碼)
- 11.Cadastral Number (農產品產地地段地號)
- 12.Parent trace code (原料追溯碼網址)
- 13.Trace code list (一籤一碼追溯碼)
- 14.Update time (更新時間)
- 15.Operation detail (詳細栽種流程)
- 16.Resume detail (詳細履歷資料)
- 17.Process detail (詳細加工流程)
- 18.Certificate detail (其他驗證資訊)

```
ProductName: "大蒜-蒜米(蒜米)",
OrgID: "138090",
Producer: "陳美玲",
Place: "雲林縣四湖鄉東光段、雲林縣四湖鄉和平段、雲林縣四湖鄉保安段",
FarmerName: "陳美玲",
PackDate: "2017/09/15",
CertificationName: "國立中興大學",
ValidDate: "2020/04/19",
StoreInfo: "",
Tracecode: "02138090760777",
LandSecNO: "",
ParentTraceCode: "",
TraceCodelist:
"1060915013400008,1060915013400175,1060915013400271,106091501
Log_UpdateTime: "2017/09/16",
OperationDetail: "http://data.coa.gov.tw:80/Service/OpenData/
ResumeDetail: "http://data.coa.gov.tw:80/Service/OpenData/Res
ProcessDetail: "http://data.coa.gov.tw:80/Service/OpenData/Re
CertificateDetail: "http://data.coa.gov.tw:80/Service/OpenDat
```

Data Process of Traceable Agriculture Product



Data Clean

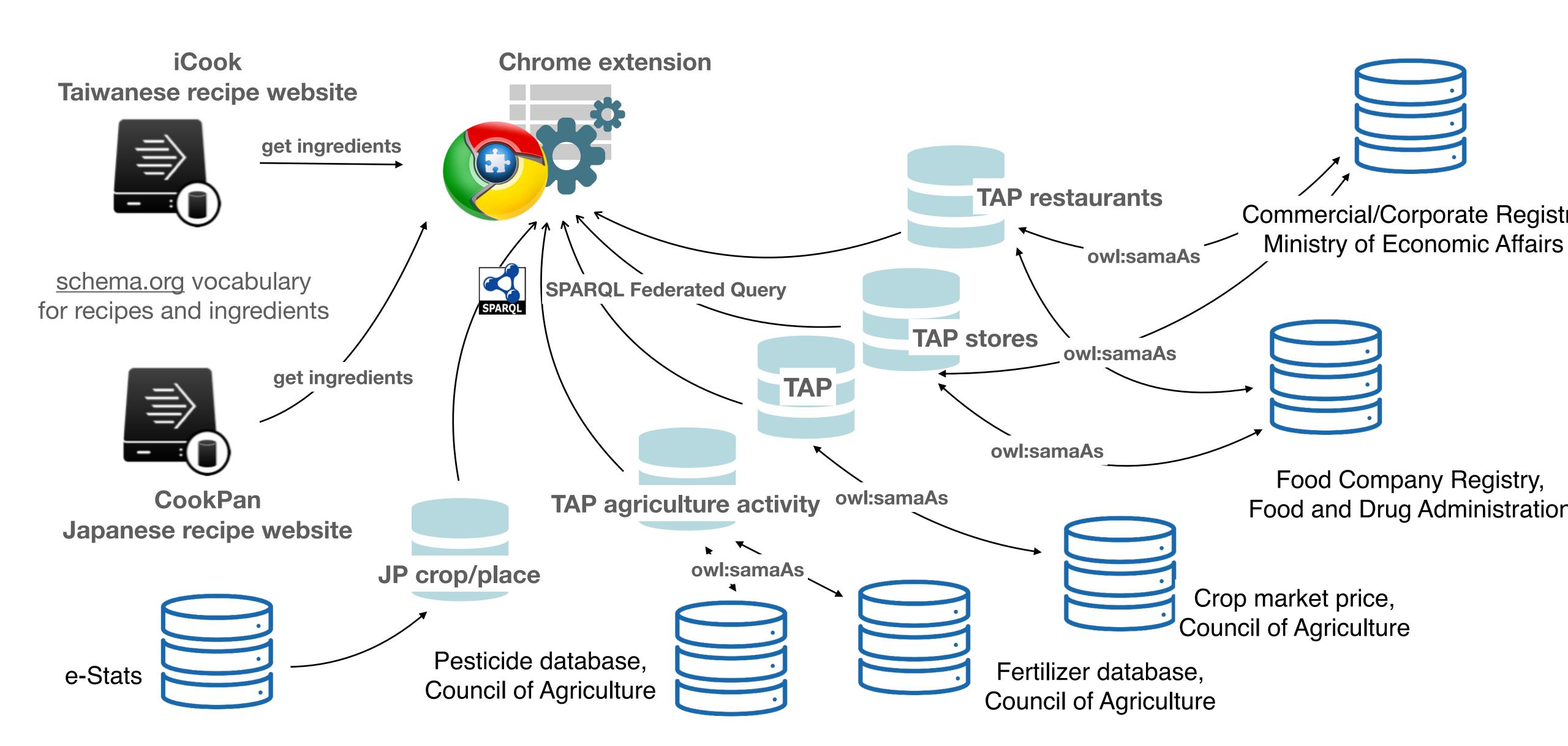
- Clean crop names of TAP to match standard crop names and codes published by Ministry of Agriculture
- Clean store/supermarket names of TAP
- Clean restaurant names of TAP
- Clean agricultural activities of TAP to match Agriculture Activity Ontology published by NII and NARO, and AGROVOC by FAO

Semantified data

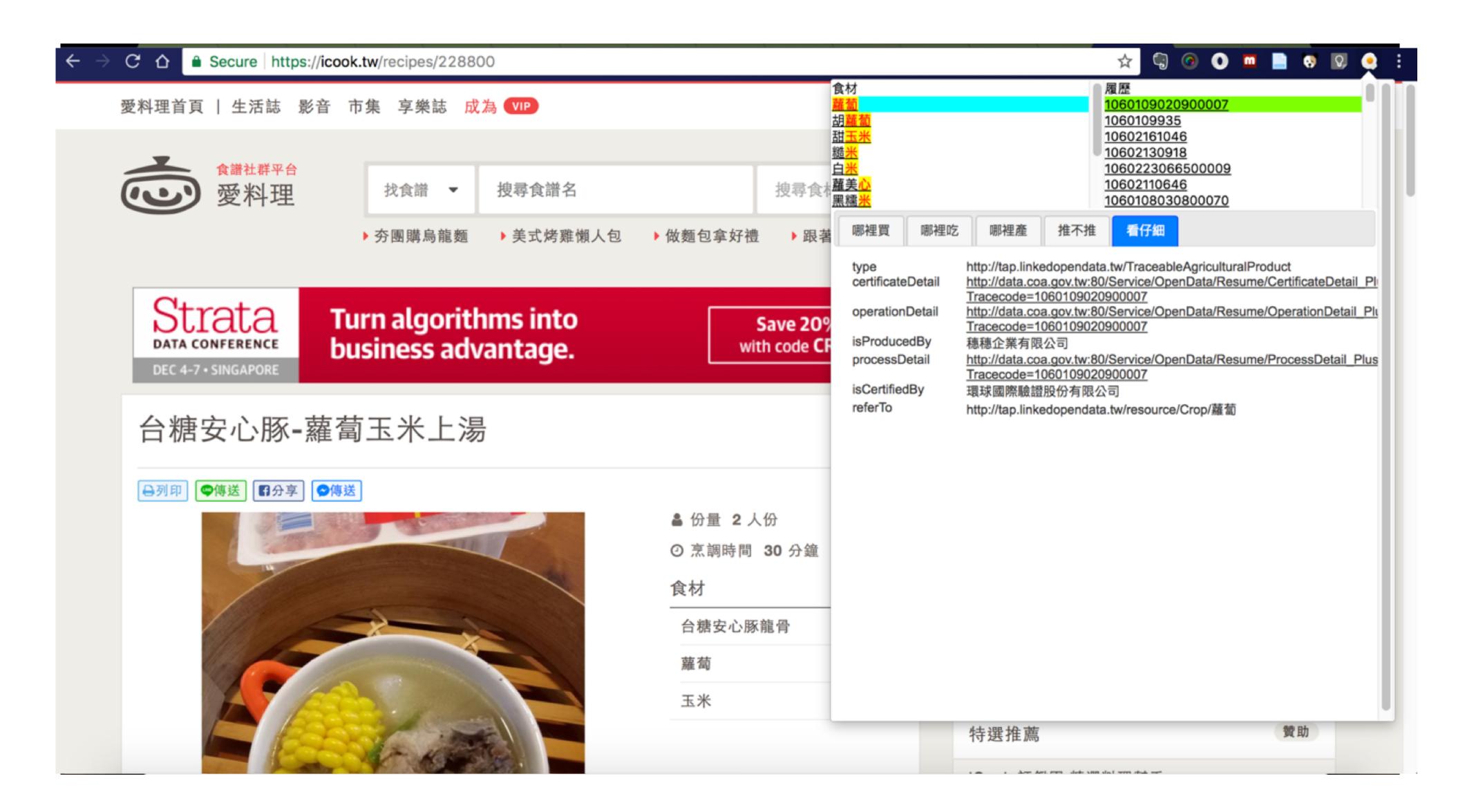
• Develop an TAP ontology defining the properties and relationships among TAP entities such as agriculture products, farmers, agriculture activities, stores,..... etc.

- Linked Data
- Use owl:sameAs to interlink same entities in different domains such as fertilizer, pesticide, food company registry

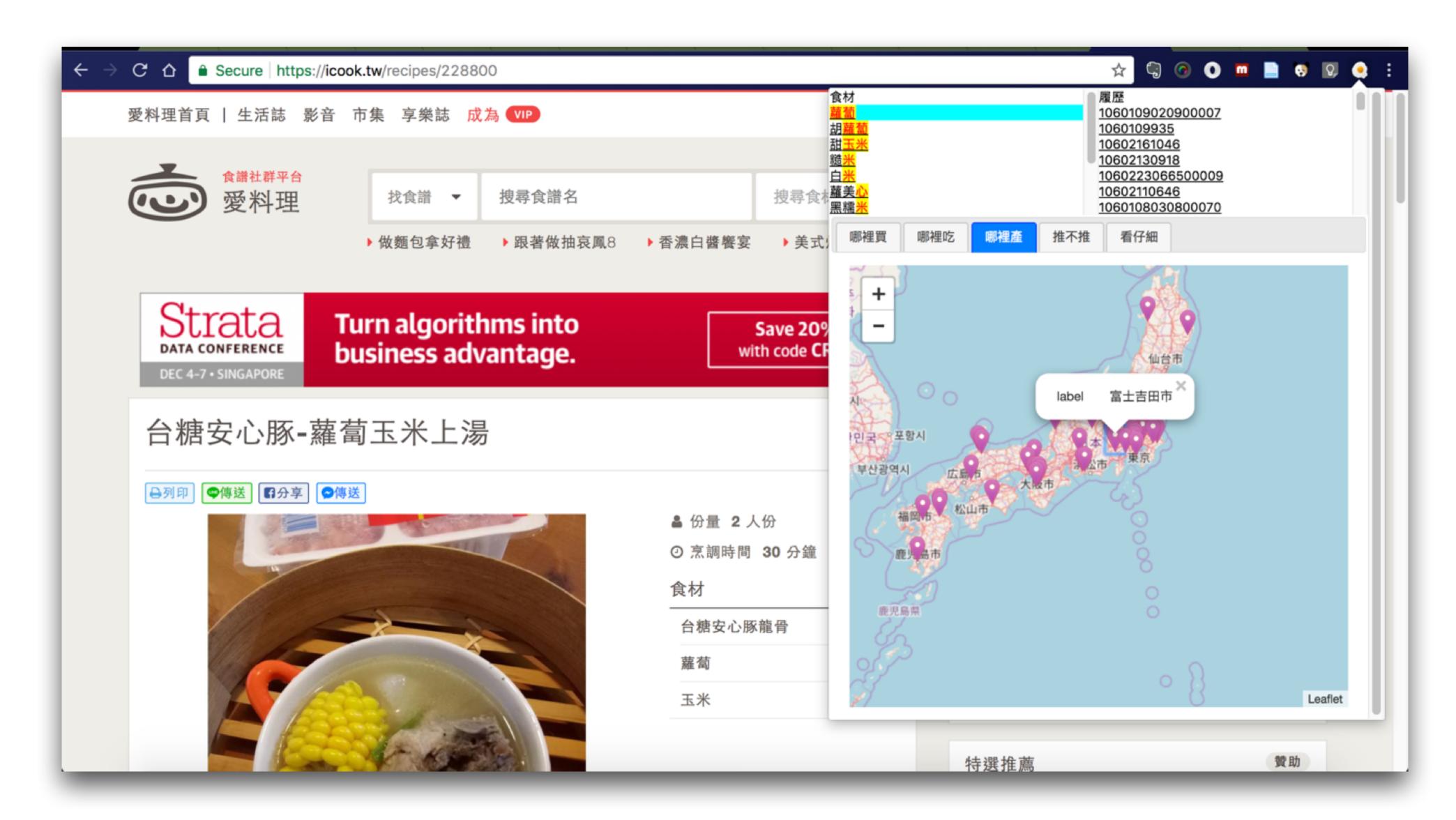
Cross-sectional data integration for TAP queries



LinkedFood: Chrome extension for recipe website



Across language: query Japanese crops



Query Japanese crops on JP website



Query Taiwanese shops on JP recipe web

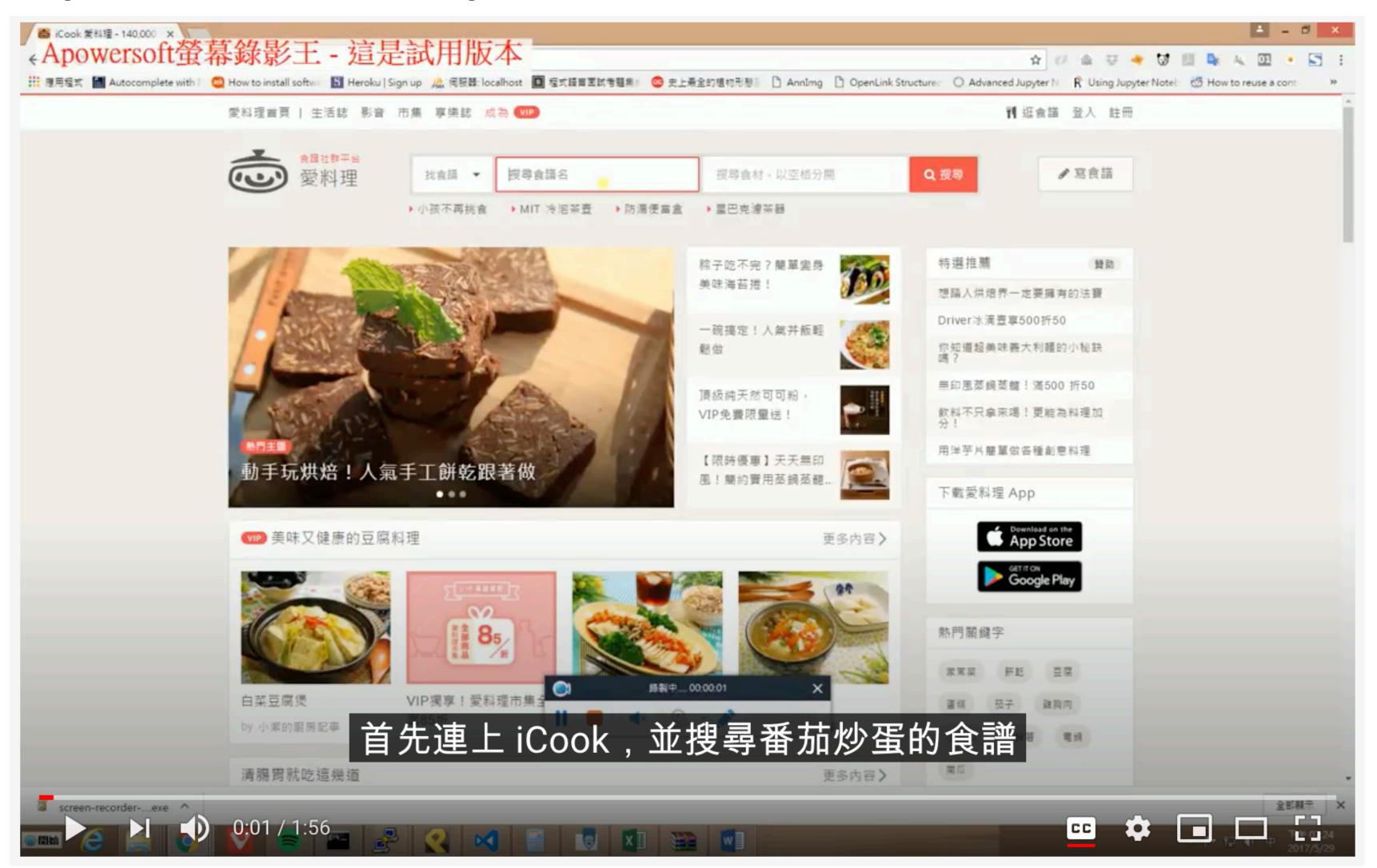


Query Taiwanese restaurants on JP recipe web



Demo

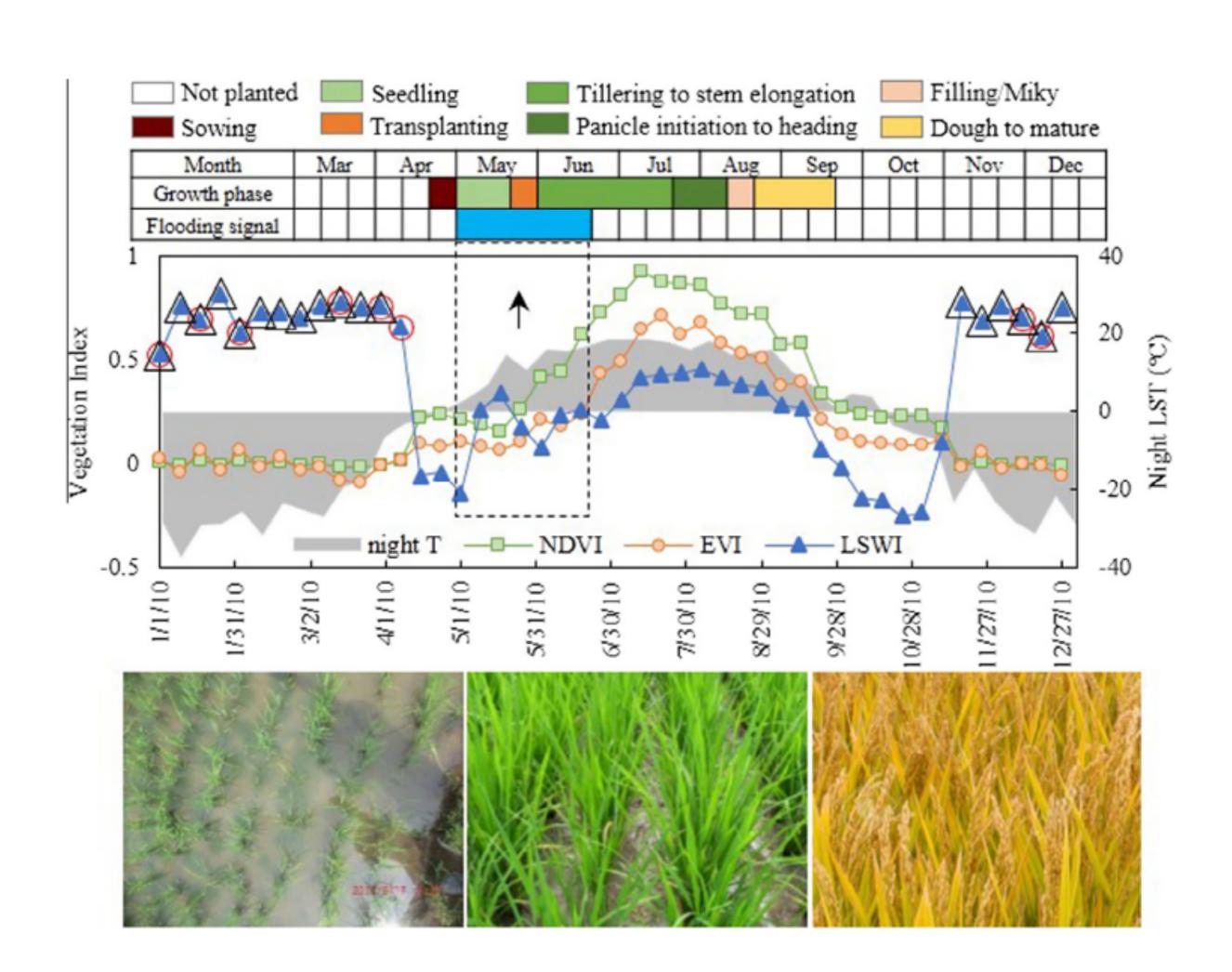
https://www.youtube.com/watch?v=XytMotQwbA4&t



Dynamic Cultivation Calendar

動態栽培曆

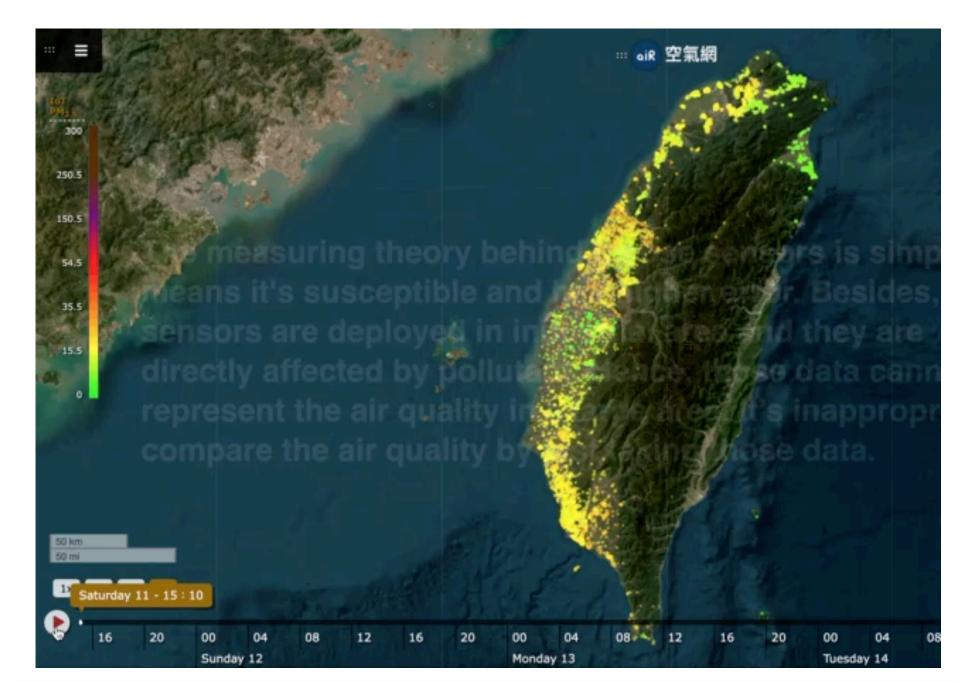
- Use paddy rice as en example
 - Cultivation calendar is mainly the relationship of temperature and water in the growth of paddy rice.
 - Cultivation calendar offers the timings to fertilization, watering, pesticide,...etc.
- The relationships of temperature and water can be observed via using IoT technologies
- The timing of agriculture activities could be extracted from traceable agriculture product datasets.



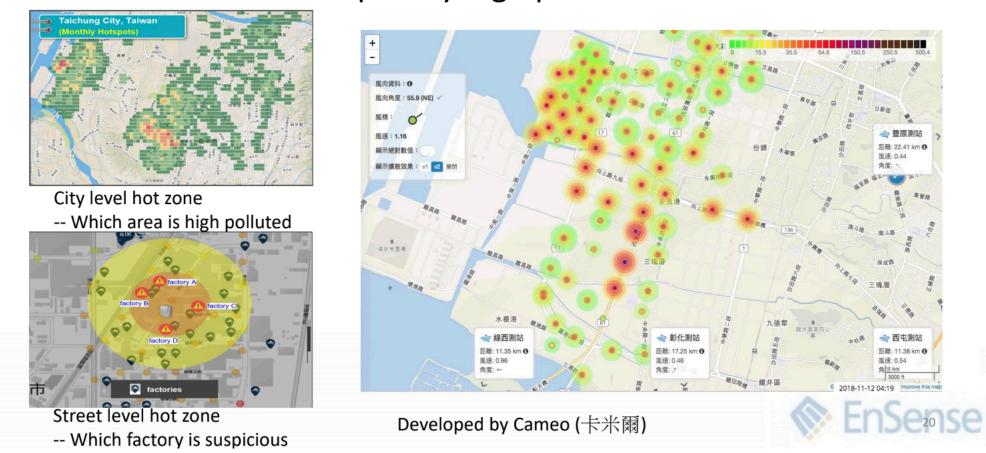
Case of Integration of Civil IoT (Air quality) and open data for better quality of life

Air Quality Micro-Stations in Taiwan

- Started at civic technologies
 - People would like to know air quality surrounding area
 - Airbox developed by LASS (Location-Aware Sensing System)
 - Much cheaper than the national air quality stations, but not so precise
- EPA incorporated with private sectors to develop micro air quality sensors
- More precisely detect air pollution sources and track
- Make machine learning enable to be applied



Find out the most frequently high pollution area



The Civil Internet of Things (IoT) Project of Taiwan

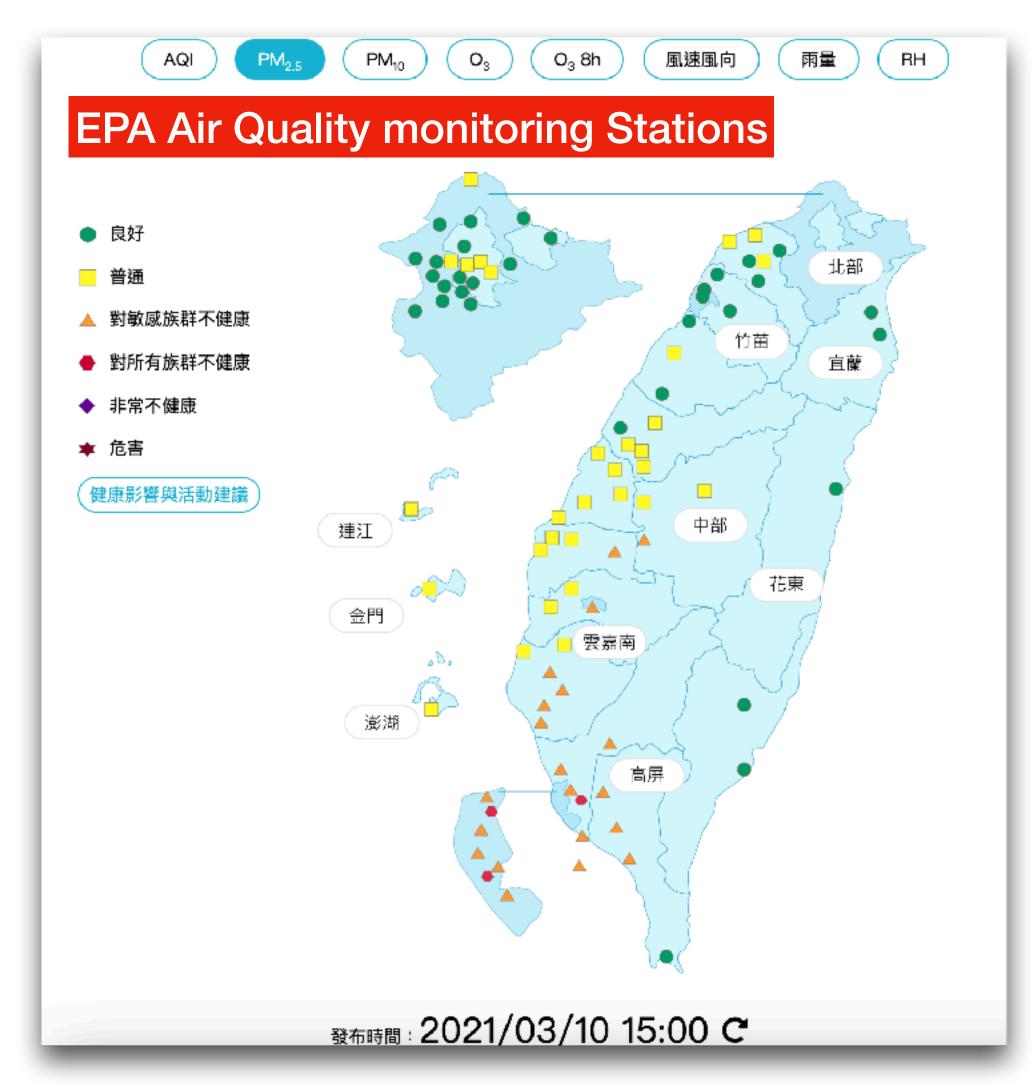
A national-wide project

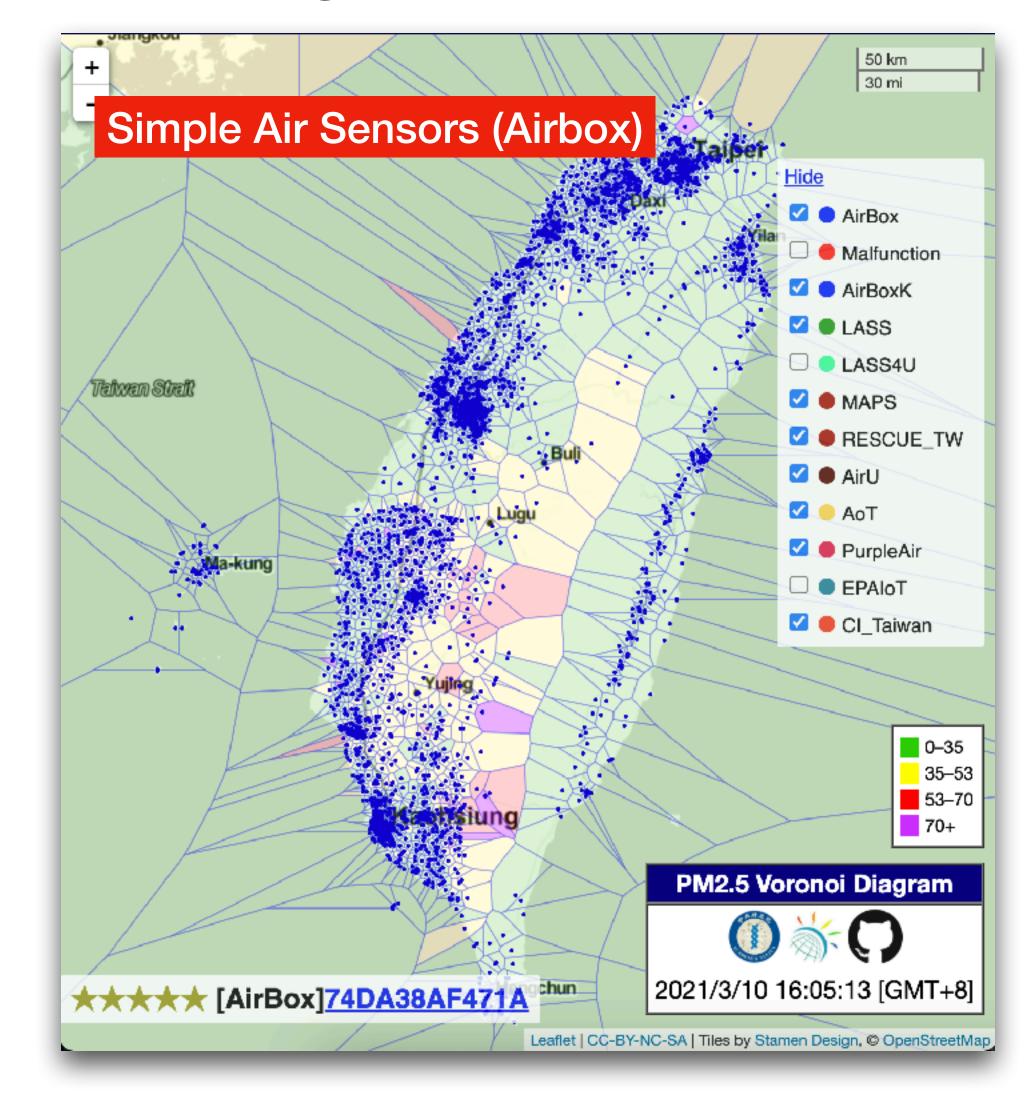
- Providing real-time and historical Civil IoT sensing data by OGC SensorThings API
- Data visualization and search service
- Massive storage and computing power for simulation analysis and Al applications

Platform: https://ci.taiwan.gov.tw/dsp/en/environmental en.aspx Air quality Earthquake Alert Water resource ~~~ Data from Civil IoT Data supply **Data Service** Platform Value-added Government data services Computing Data environment management Non-government data Research institute Industry Analytic models/algorithms Analytic models/algorithms

Why needs civil loT?

Professional v.s. civic air quality monitoring





Different levels of air quality sensors

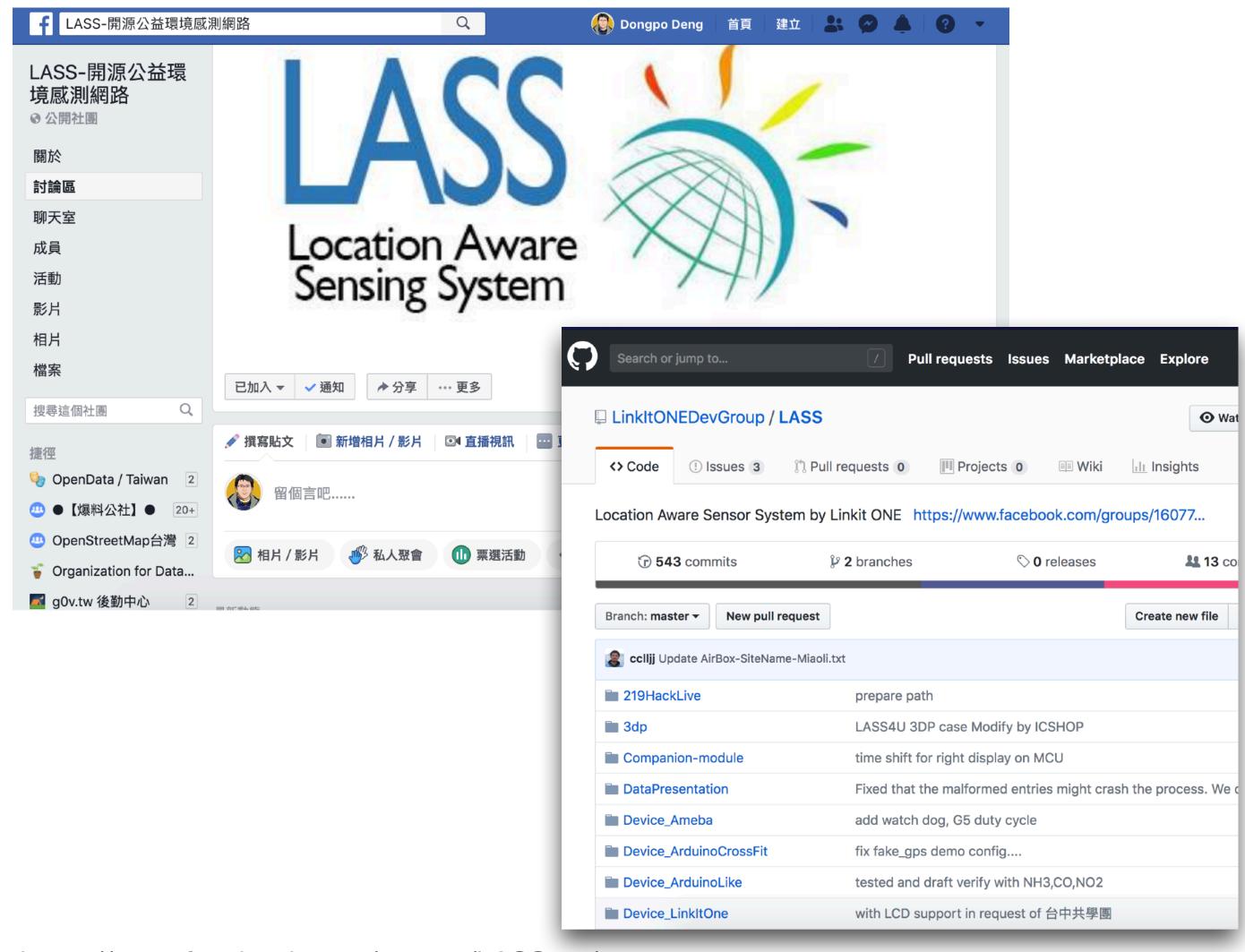
Level	National Monitoring Stations	Smart City Sensors	Educational Sensors	Civil Sensors
Figure				
Purpose	Regulations	Hot zone determination	Envir. Education	Personal
Apply	Evaluate air quality in large area according to regulations	Detect air quality in small area for tracking pollution sources	Know the variety of air quality in micro environment	
Theory	Air-Beta Ray Attenuation Method or Air-Inertial Mass Method	Light scatter		
Particle definition	Aerodynamic Diameter: a sphere of particle which settles in still air at the same velocity	Optical Diameter: The particle diameter is measured by laser diffraction. The surface rough of particle, water, and absorbance would affect the measure		
Health risk relationship	Current health risk research use aerodynamic diameter	Not much health risk research use optical diameter currently		

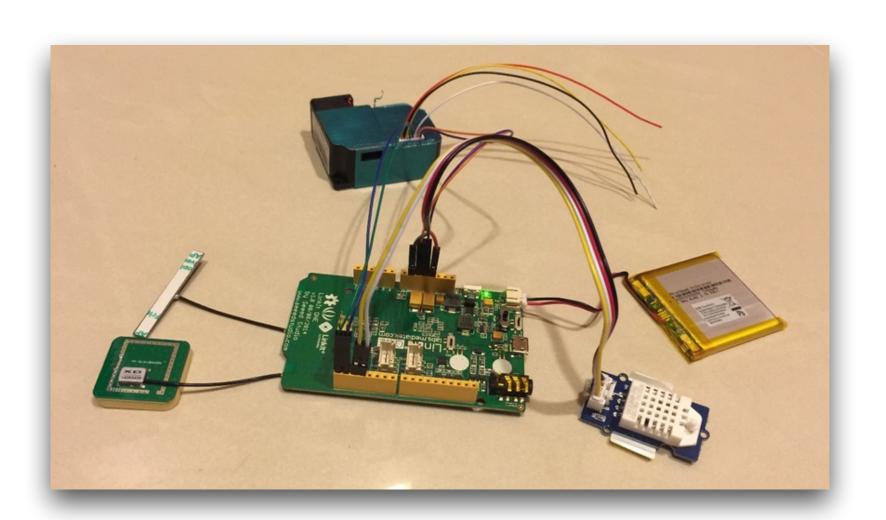
Smart City Sensors

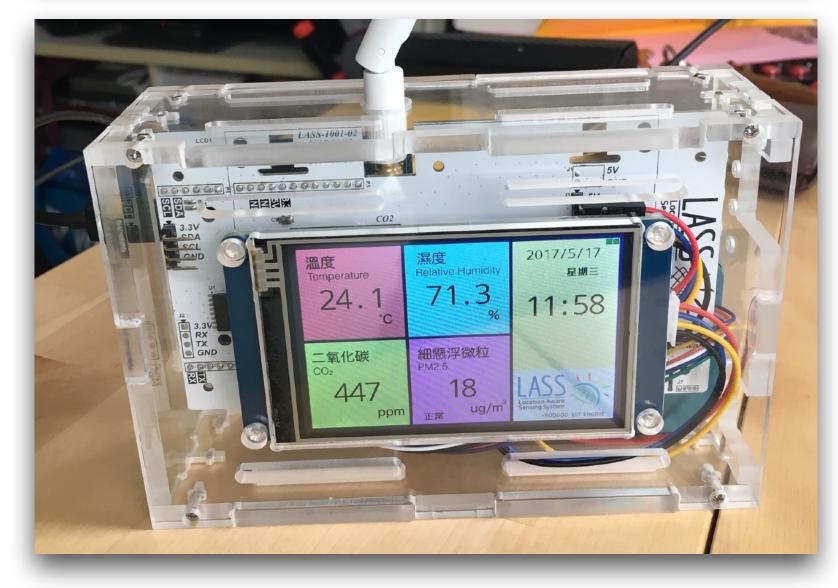
Air quality Water quality Mobile Air quality



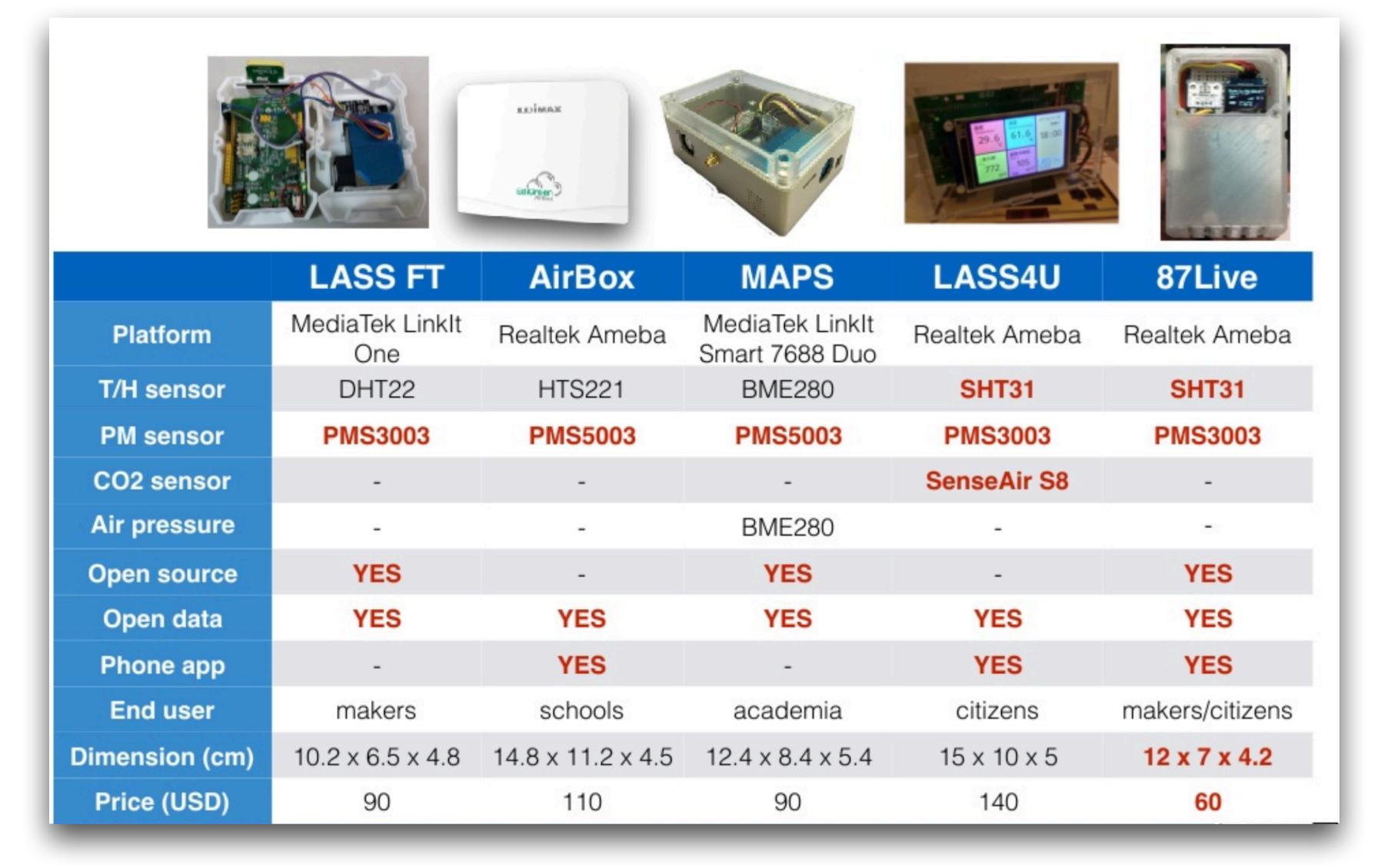
Collaboration with CivicTech communities



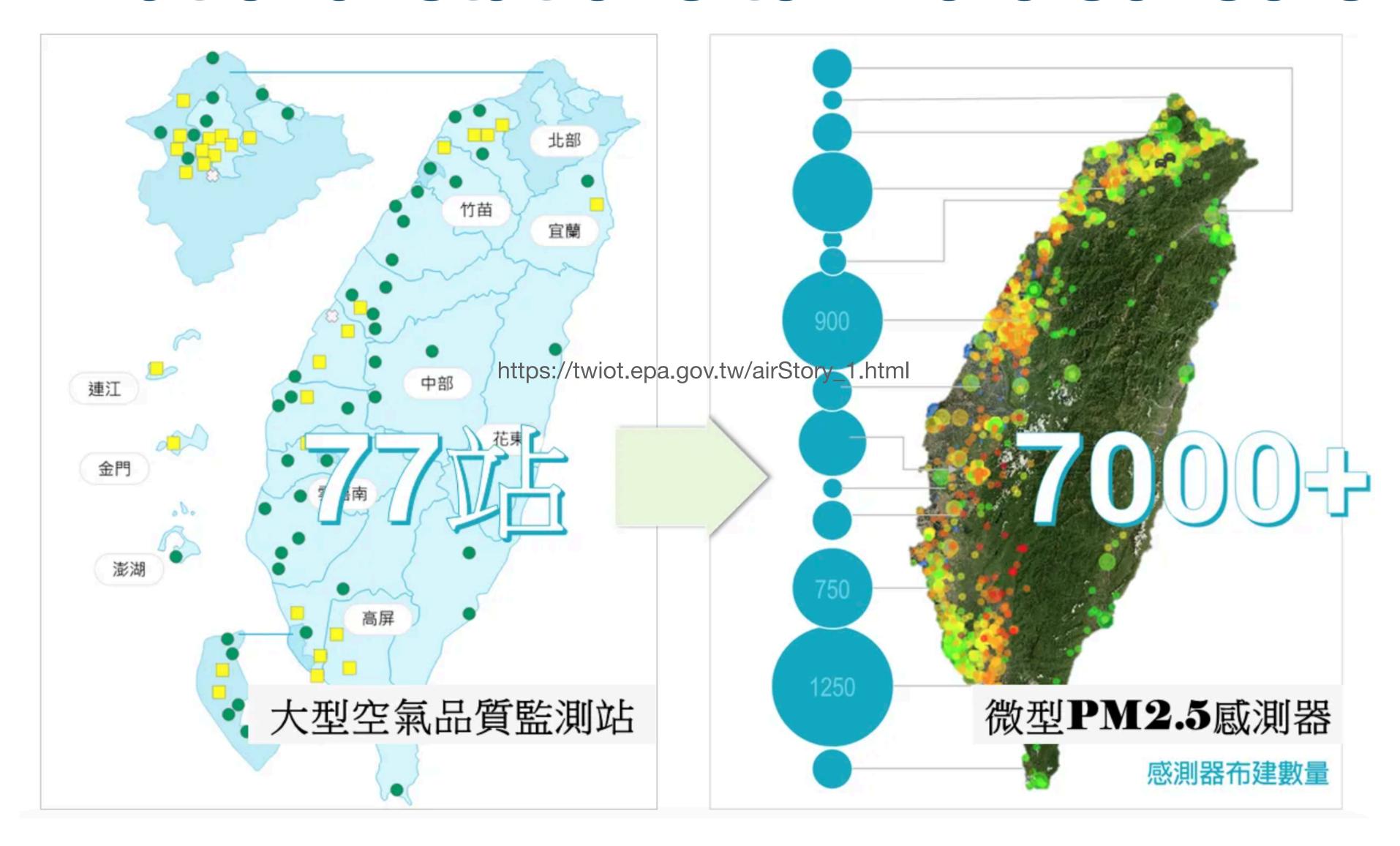




Airbox: Community version for air quality monitoring



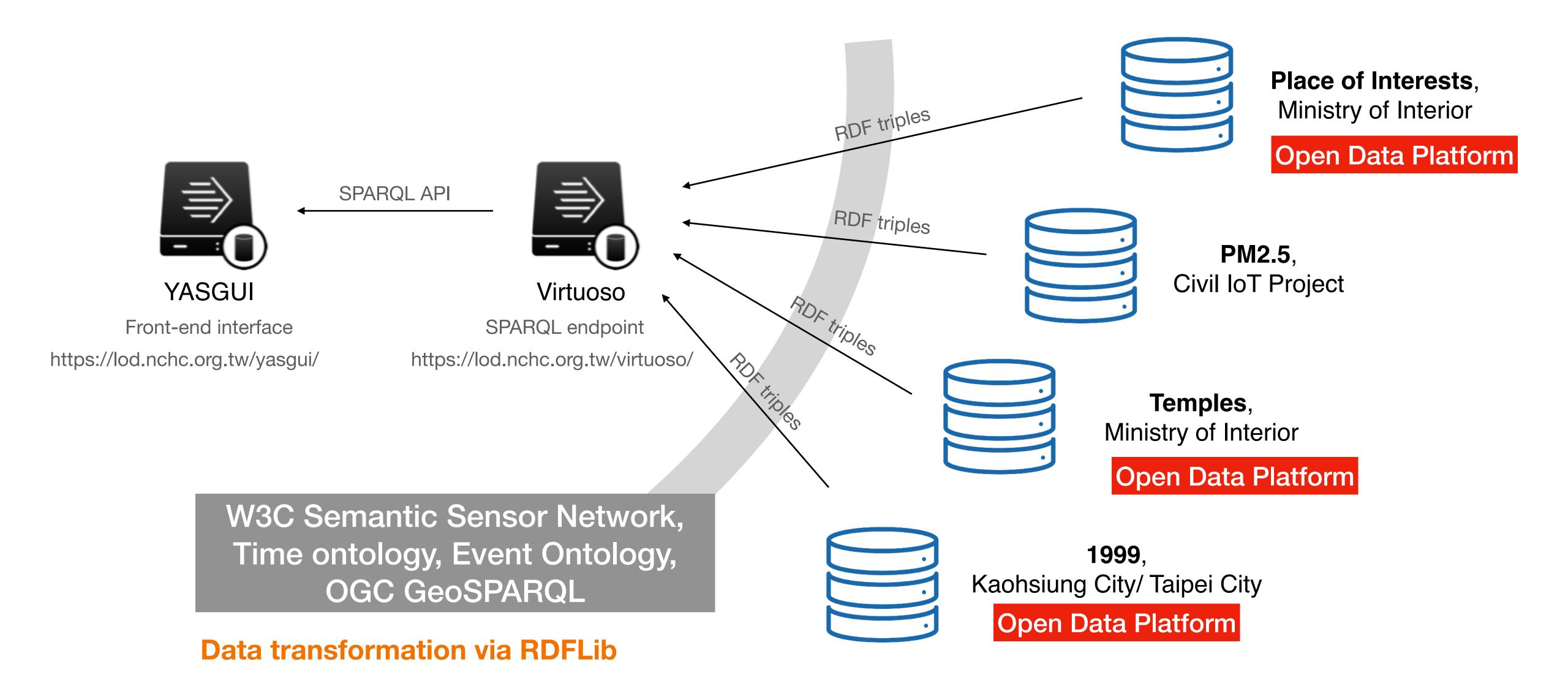
From national stations to micro sensors



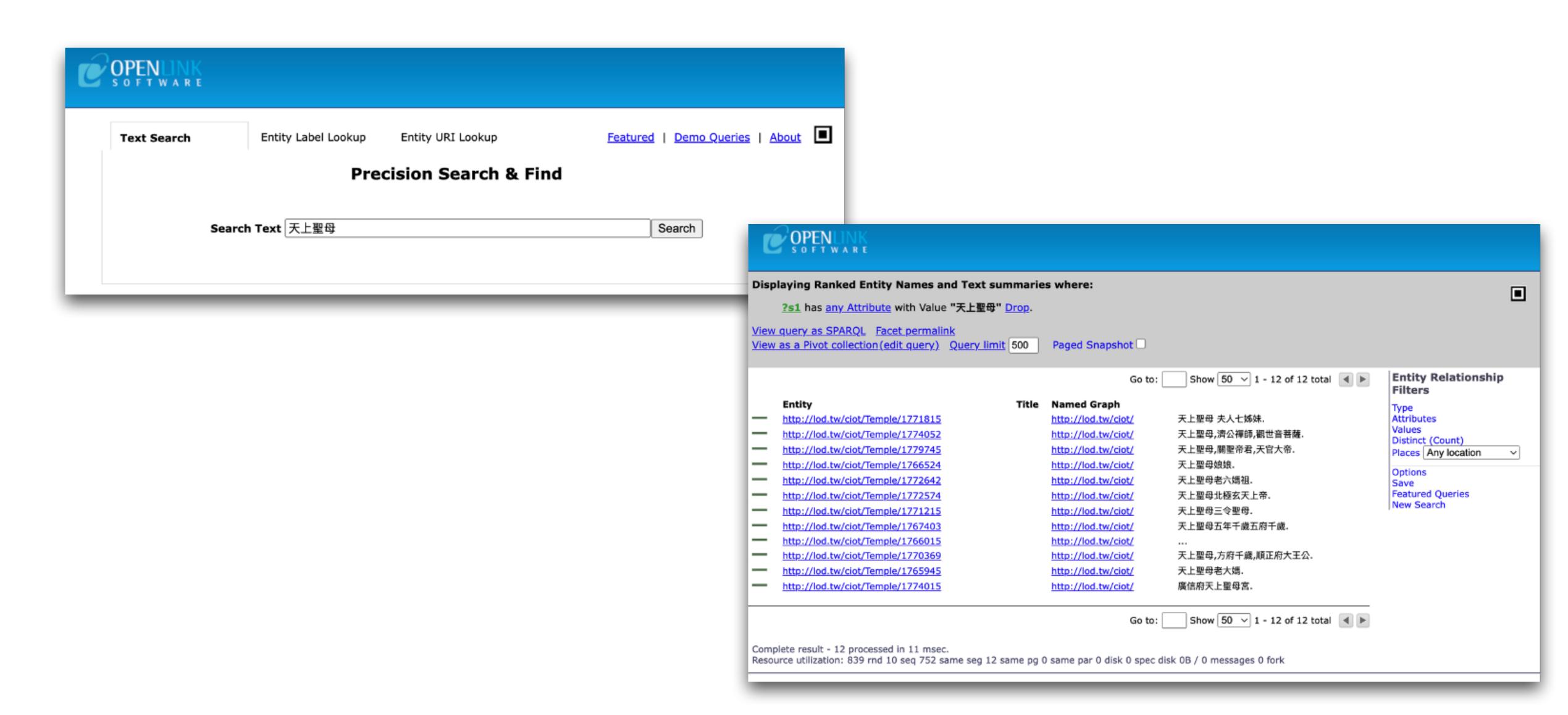
Civil loT Air Quality datasets 民生公共物聯網空氣品質資料



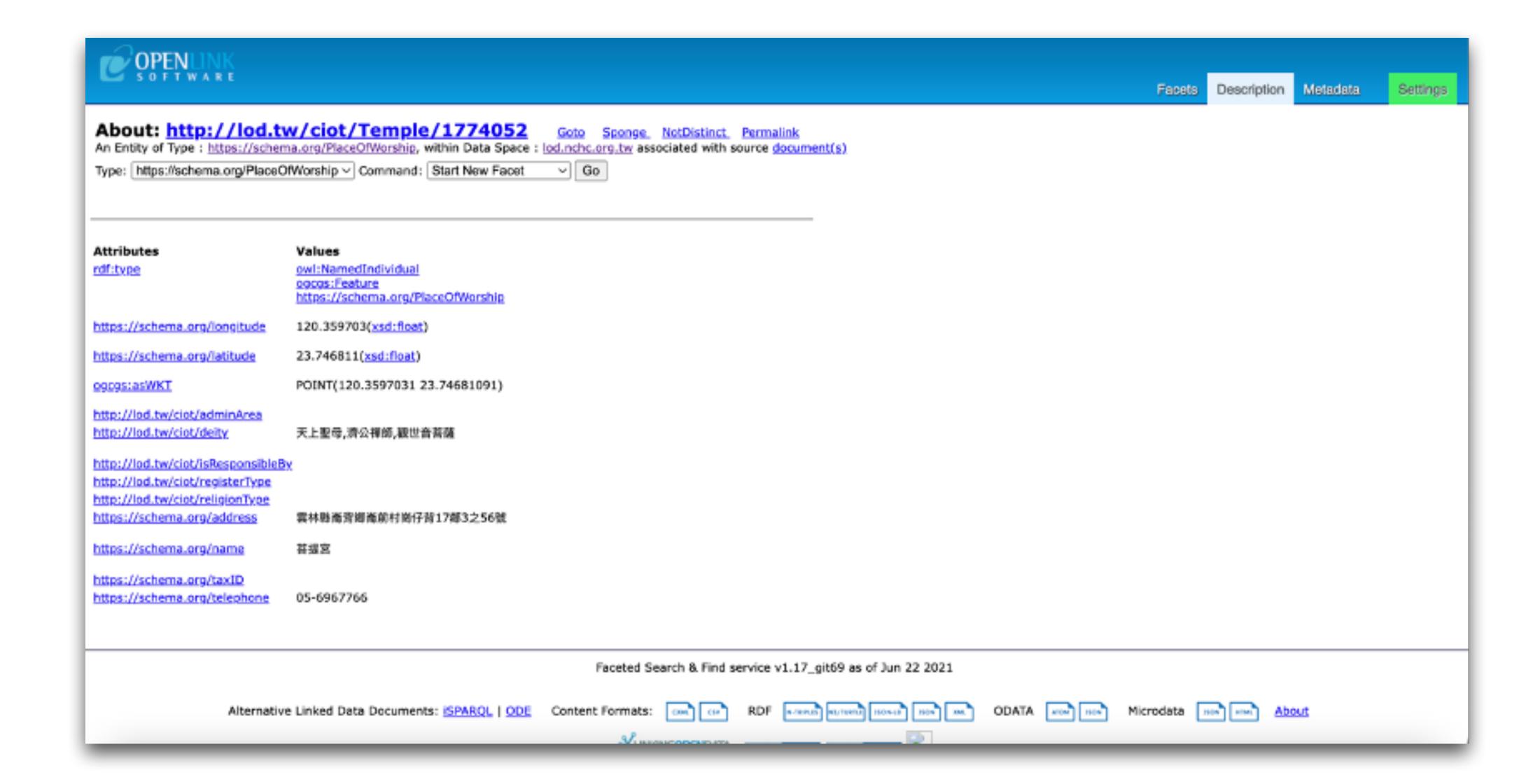
Data integration and process



Virtuoso - text search



Virtuoso - browse a triple



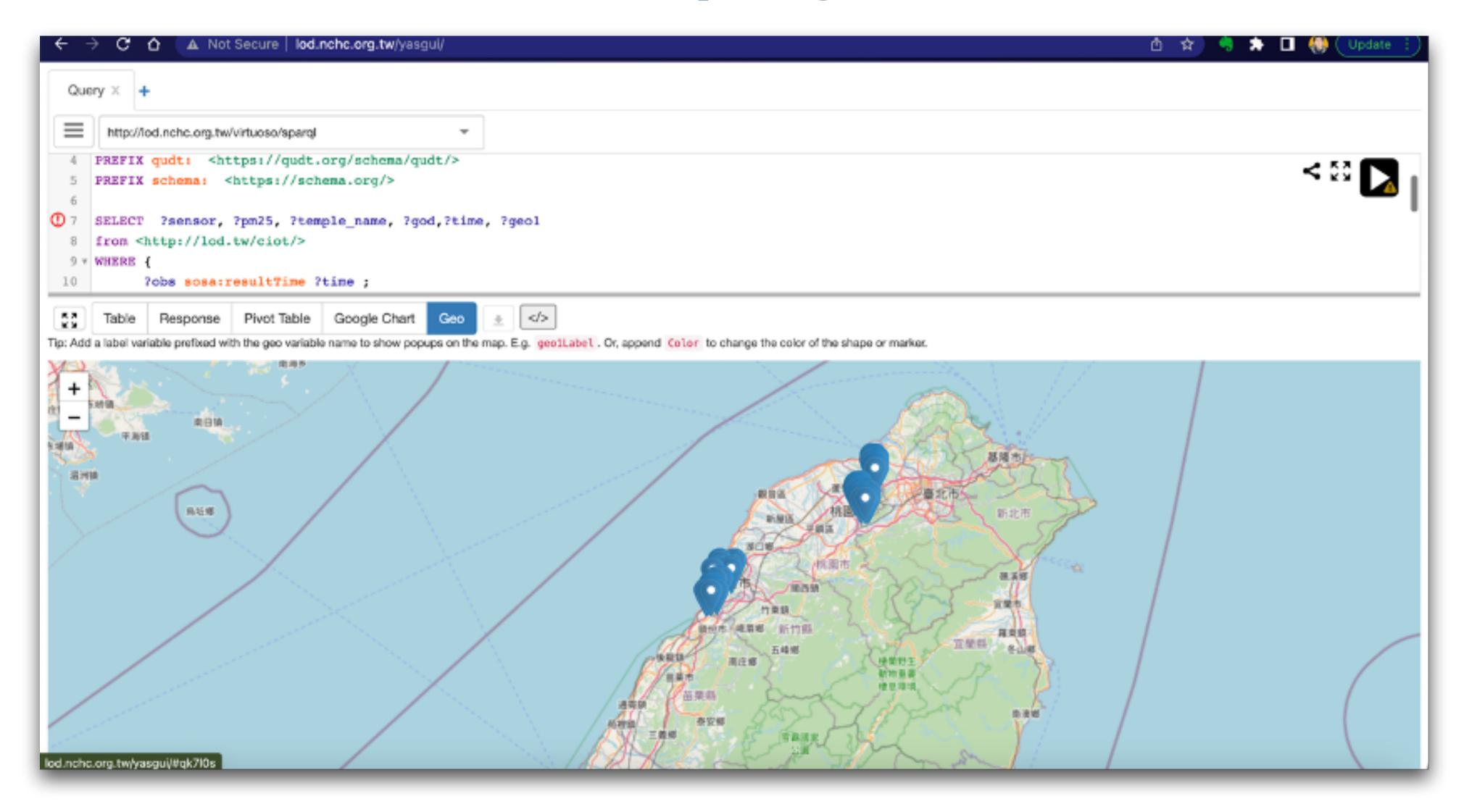
An example of SPARQL query

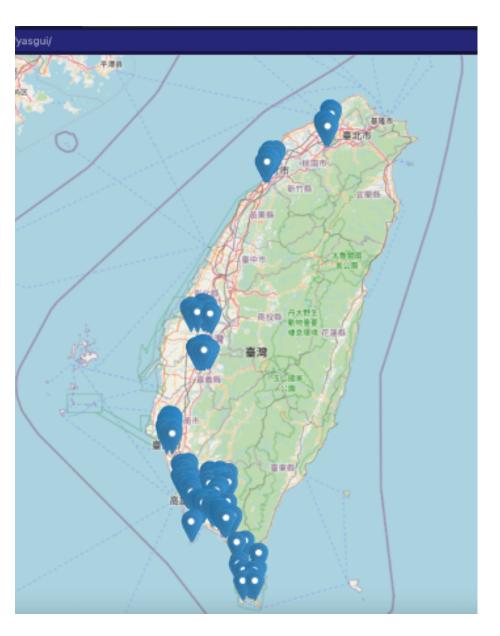
SPARQL query is similar to SQL

Find out the PM2.5 sensors observing over 300 µg/m3 and the distance to temples within 500 meters

```
PREFIX ciot:
                    <http://lod.tw/ciot/>
                    <a href="http://www.opengis.net/ont/geosparql#">http://www.opengis.net/ont/geosparql#</a>
PREFIX geo:
PREFIX sosa: <a href="http://www.w3.org/ns/sosa/">http://www.w3.org/ns/sosa/</a>>
PREFIX qudt: <a href="https://qudt.org/schema/qudt/">https://qudt.org/schema/qudt/>
PREFIX schema: <a href="https://schema.org/">https://schema.org/</a>
SELECT ?sensor, ?pm25, ?temple_name, ?god,?time
from <a href="http://lod.tw/ciot/">http://lod.tw/ciot/</a>
WHERE {
        ?obs sosa:resultTime ?time;
                       sosa:hasResult ?value;
                       sosa:madeBySensor?sensor.
       ?value qudt:numericValue ?pm25.
       ?sensor ciot:hasSite ?site.
       ?site geo:asWKT ?geo1.
         ?temple a schema:PlaceOfWorship;
                    geo:asWKT ?geo2;
                    schema:name?temple_name;
                    ciot:deity?god.
         FILTER (?pm25 > 300)
         FILTER (bif:st within(?geo1,?geo2, 500))
```

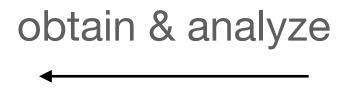
YASGUI - SPARQL query





Apply of Generative Al to civil loT data analysis



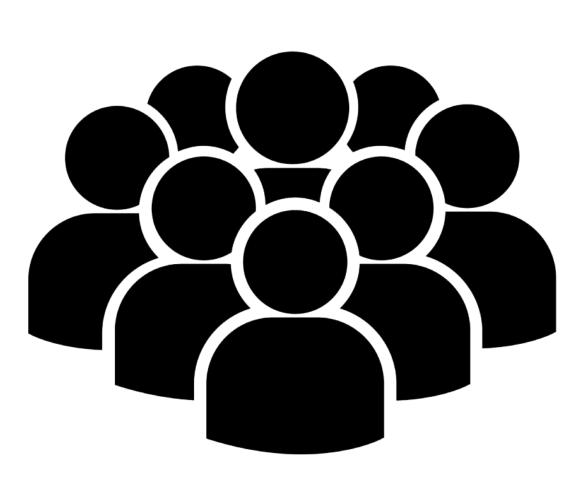




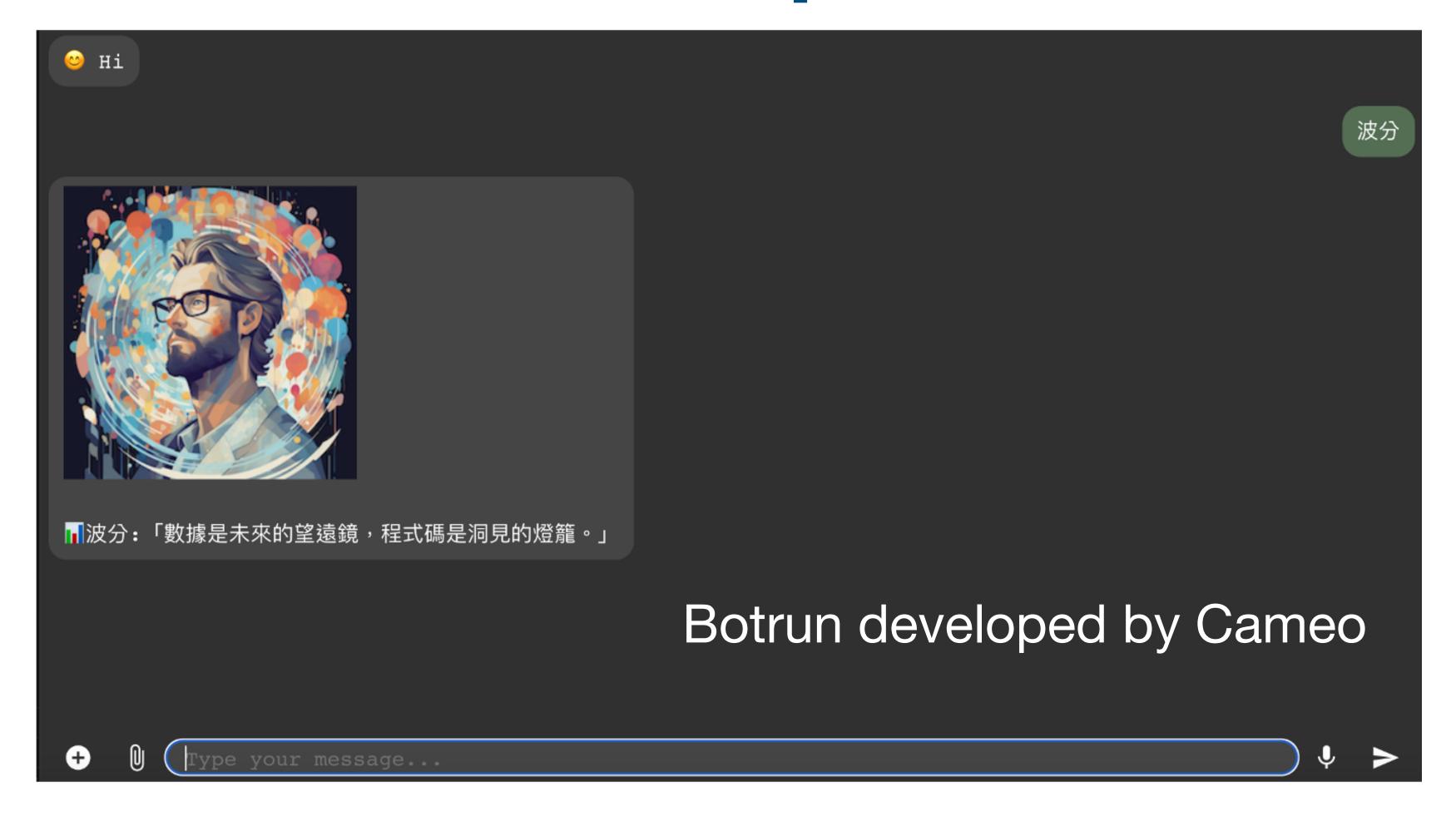
use

A generative Al

If you're not a data scientists or you're even not familiar with how to deal with data, a smart 'assistant' is quit helpful



Users



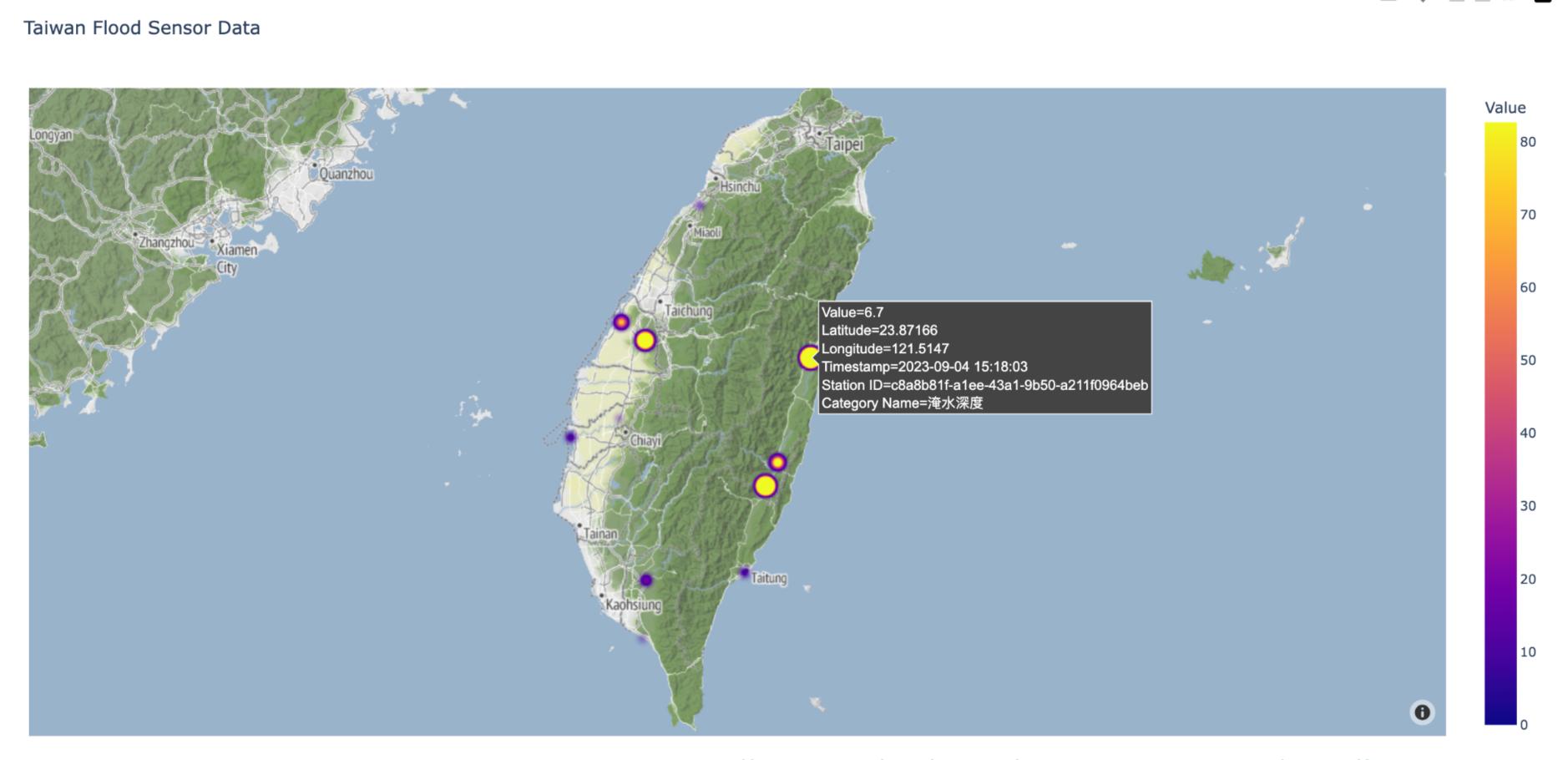
To get more detail

https://www.youtube.com/watch?v=MtCsDPbecyw&t=1s https://www.youtube.com/watch?v=sSCOGKC3SL4&t=5s

Make a map for spatial distribution of flooding sensors

```
001 請用 plotly 的 density_mapbox 繪製台灣地圖
                                            002 將 Plotly.js 庫直接嵌入到 HTML 文件中。可以設定 include_plotlyjs=True
                                            003 將data/flood-sensor_20230903-0905_clean.csv 的資料繪製到台灣地圖上。
                                             'Longitude', 'Latitude' : 為座標欄位
                                             'value': 為主要淹水數值欄位
                                             希望滑鼠點到地圖上資料點位時,顯示: timestamp , station_id, CategoryInfos_Name, value
@begin code("./users/wengroy/code/plot map.py")
import pandas as pd
import plotly.express as px
# Load the data
df = pd.read_csv('data/opendata/flood-sensor_20230903-0905_clean.csv')
# Create the map
fig = px.density_mapbox(df, lat='Latitude', lon='Longitude', z='value', radius=10,
                       center=dict(lat=23.5, lon=121), zoom=7,
                       mapbox_style="stamen-terrain",
                       hover_data=['timestamp', 'station_id', 'CategoryInfos_Name',
'value'])
# Save the map as HTML
fig.write html("users/wengroy/code/flood map.html", include plotlyjs=True)
@end
flood map.html
```

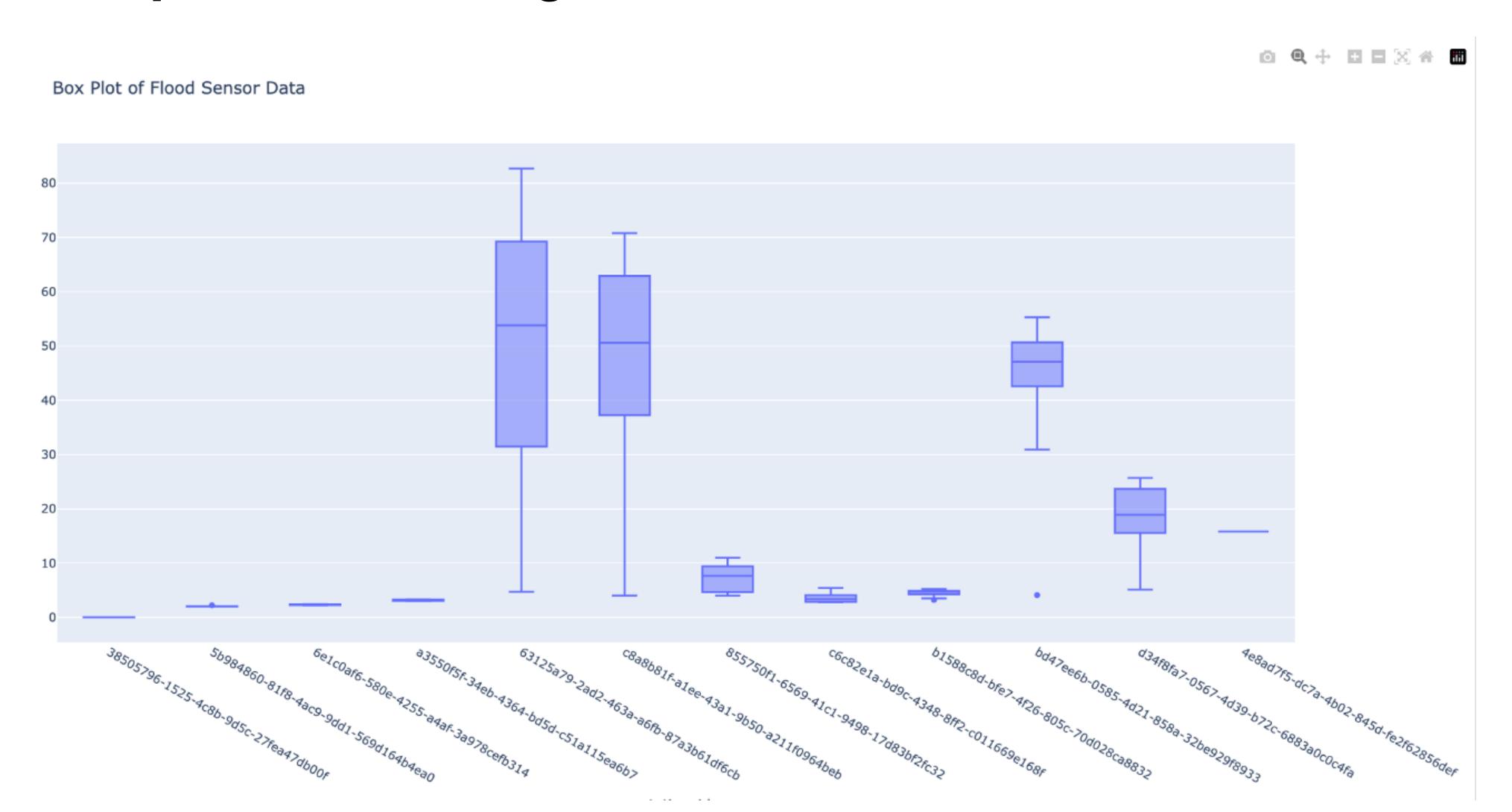
Make a map for spatial distribution of flooding sensors



Make a box plot for flooding condition

```
001 請用 plotly 的 box
                                        002 將 Plotly.js 庫直接嵌入到 HTML 文件中。可以設定 include_plotlyjs=True
                                        003 將data/flood-sensor_20230903-0905_clean.csv 的資料,以 station_id 為 x軸
                                        'value': 為 y 軸主要淹水數值欄位
@begin code("./users/wengroy/code/flood_box.py")
import pandas as pd
import plotly.express as px
# Load data
df = pd.read_csv('data/opendata/flood-sensor_20230903-0905_clean.csv')
# Create box plot
fig = px.box(df, x='station_id', y='value', title="Box Plot of Flood Sensor
Data")
# Save to HTML
fig.write html("users/wengroy/code/flood box.html", include plotlyjs=True)
@end
Box Plot of Flood Sensor Data
```

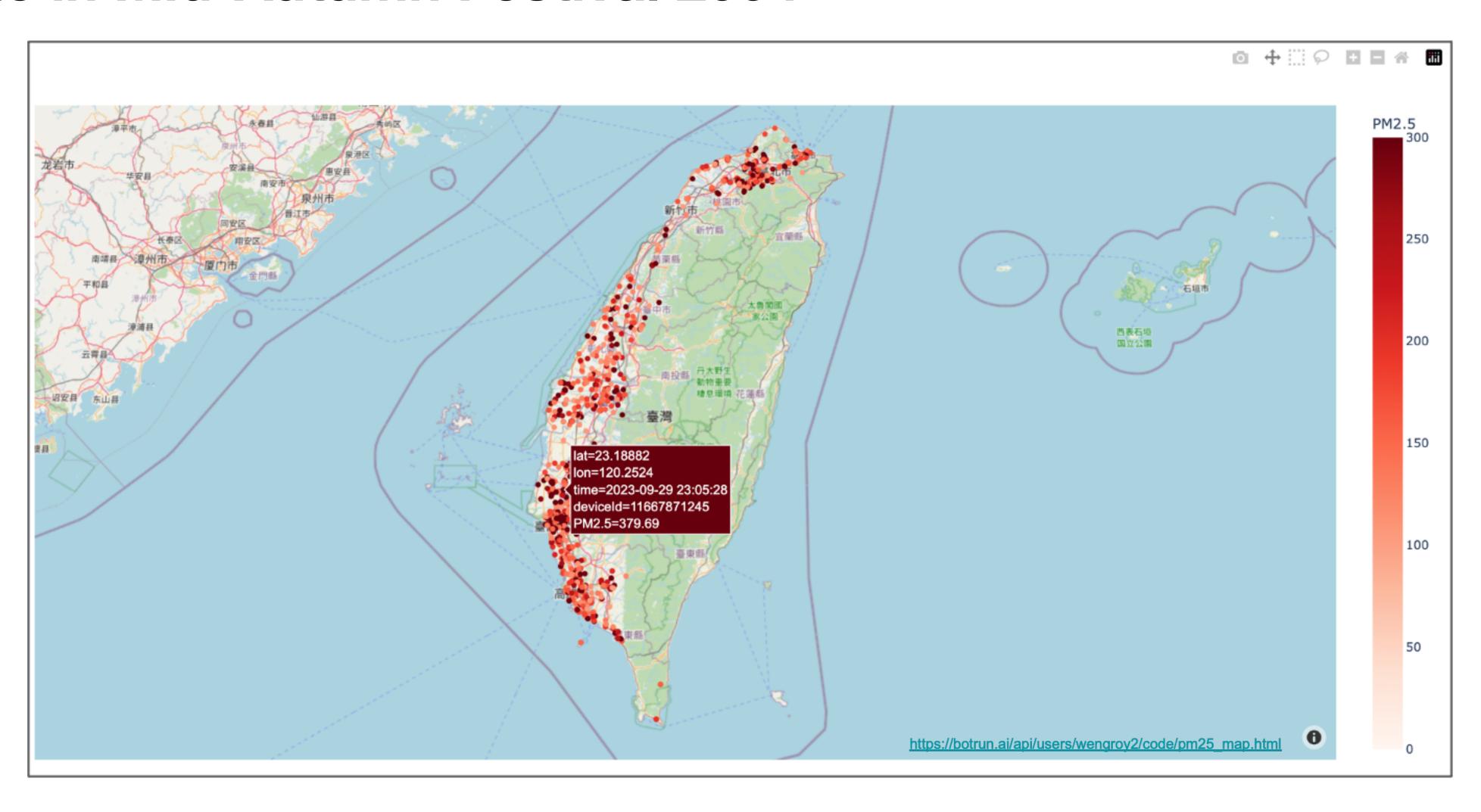
Make a box plot for flooding condition



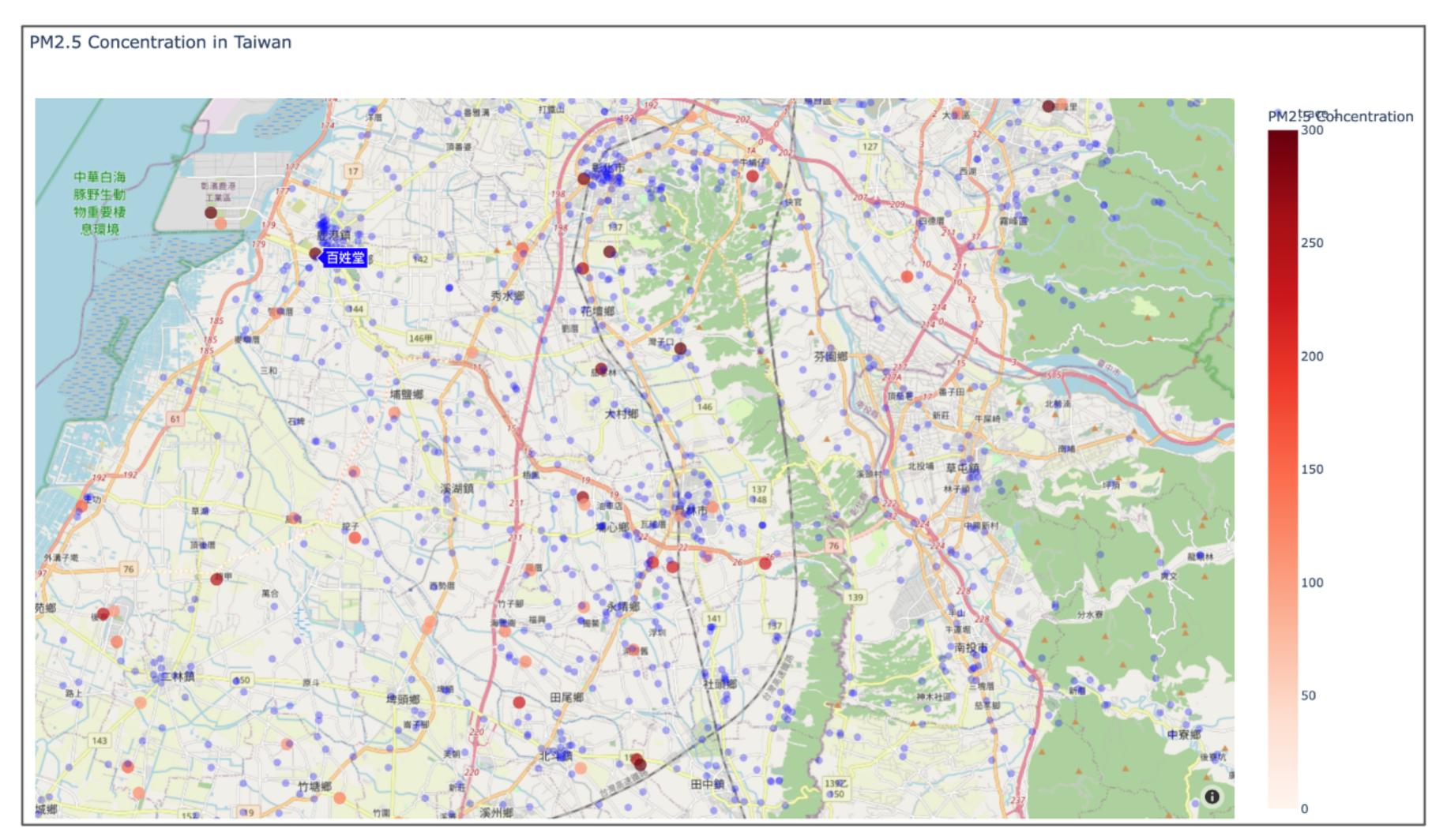
Apply to air quality data (PM2.5)

```
the 'PM2.5' value ranging from 0-300 and use red color.
@begin code("./users/wengroy2/code/plot_map.py")
import pandas as pd
import plotly.express as px
# Load the data
data = pd.read_csv('./users/wengroy2/data/opendata-
air/moenv micro pm25 20230929.csv')
# Filter the data
filtered_data = data.loc[data.groupby('deviceId')
['PM2.5'].idxmax()]
filtered_data = filtered_data[filtered_data['PM2.5'] > 100]
# Create the map
fig = px.scatter_mapbox(filtered_data, lat='lat', lon='lon',
color='PM2.5', hover_data=['time', 'deviceId', 'PM2.5'],
color_continuous_scale='reds', range_color=[0,300])
fig.update_layout(mapbox_style="open-street-map")
fig.write_html("./users/wengroy2/code/pm25_map.html",
include_plotlyjs=True)
@end
pm25 map.html
```

PM2.5 in Mid-Autumn Festival 2004



PM2.5 in Mid-Autumn Festival 2004





Thank for your attentions!

Contact email dongpo.deng@gmail.com