

# Enabling a Data-informed Public Sector: An Introductory Course to BDTI Essentials



Session 5: Gathering Data from the Web and  
Geo Visualisation



**Welcome!**  
**The webinar will begin shortly.**

**DG CNECT**

Directorate-General for Communications  
Networks, Content and Technology

**DG DIGIT**

Directorate-General  
for Digital Services



# Some housekeeping



- The Chat and Q&A widgets are located to the right of the presenter screen
- Post questions in Q&A
- The session will be recorded and available shortly after

## Do

- Participate in the chat during the session
- Ask questions
- Reserve 2 minutes after the webinar to share your feedback

## Don't

- Self-promote
- Disrespect anyone



# Enabling a Data-informed Public Sector: An Introductory Course to BDTI Essentials



## Session 5: Gathering Data from the Web and Geo Visualisation



## Agenda

- Welcome [11:00-11:05]
- Session 4 Recap [11.05-11.15]
- Advanced Data Analytics [11:15-12:00]
- Q&A [12:00-12:10]
- Wrap up [12:10 – 12:15]

# What is the Big Data Test Infrastructure (BDTI) ?



Not **only** for big data, for **public sector in general (open data)**

You have the key ingredients (datasets),  
we equip you with the best **open-source tool**  
to create amazing recipes for **public good**.



**Six months free of charge service**  
for EU public administrations \*



**Ready-to-use**  
**data analytics stack** and  
support



Cloud platform based on  
open-source tools

\* The cost of the pilot project must fit within the funding boundaries of the BDTI pilot budget



# Who is the Big Data Test Infrastructure (BDTI) for?



## European Public Administrations

All European Public Administrations at **local, regional and national level** can independently apply for a BDTI pilot project



## Ecosystem with **academia** and **private sector**

Academia, spin-off, startups can apply for pilot projects once there is a **clear collaboration** with a Public Administration which will be the main point of contact for the project (**Master/PhD, GovTech startups**)



**Are you working for a public administration in need of infrastructure for data analytics?**

<https://big-data-test-infrastructure.ec.europa.eu/>



# Use case:

## "Identify Innovative Green Energy Projects from EU Universities"

Zoi and her team, work in the **Educational Department of a highly polluted EU region.**

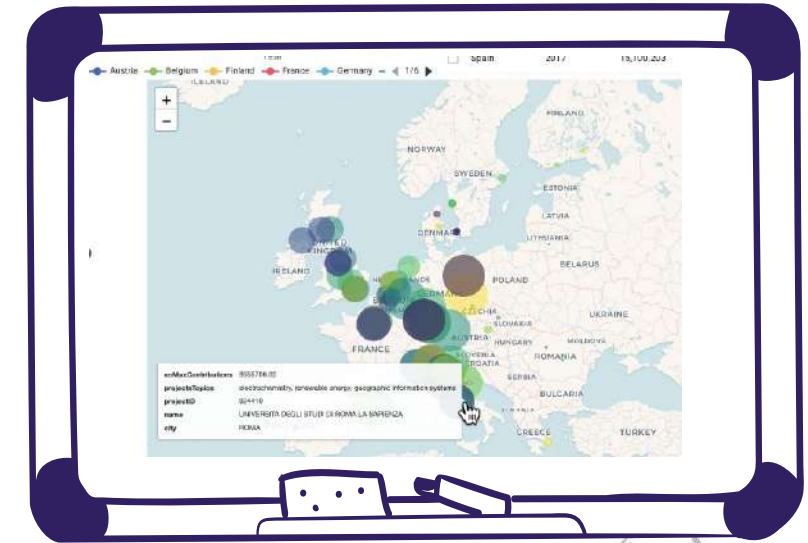
Their **purpose** is to identify potential **"green energy partnerships"** between EU universities to address pollution issues.

Zoi reached out the BDTI team to learn how to **connect the dots**, use the free BDTI playground to **create insights** and produce a **report**, which can be **easily updated** when new data is available.



They found some interesting **open data**:

- **List of innovation green energy projects** across EU countries
- **CO2 emissions** of these countries



# Recap: Zoi Data Journey

<https://code.europa.eu/bdti/bdti-essentials-course>



Access the data from multiple sources

Merge the data and store it in the desired format (end of ETL process)

**SESSION 1:**  
Data Access and Exploration

**SESSION 2:**  
Data Cleaning and Transformation

**SESSION 3:**  
Data Blending and Storage

**SESSION 4:**  
Analytics: aggregation, visualisation, reporting

**SESSION 5:**  
Gathering data from the Web and Geo Viz

EU Open data portal:

[data.europa.eu](https://data.europa.eu)

Formats:  
csv, Excel, JSON

Tools:



Clean and transform the data to the correct shape







Make sense of data, extract insights and visualise them in a report



Enhance analytics by gathering data from the internet





## B BDTI Essentials Course




  Unstar 1  Fork 0 



 98 Commits  1 Branch  0 Tags  2.2 MiB Project Storage

 **Update README.md**  
 Maria Claudia BODINO authored 18 hours ago 85da3666 

master  bdti-essentials-course /  History Find file Edit  Code 









-  README
-  BSD 3-Clause "New" or "Revised" License
-  Auto DevOps enabled
-  Add CHANGELOG
-  Add CONTRIBUTING
-  Add Kubernetes cluster
-  Add Wiki
-  Configure Integrations

Name	Last
 Session 1: Data Access and Exploration	Upda
 Session 2: Data Cleaning and Transformati...	Upda
 img	Uplo
 LICENSE	Add
 README.md	Upda

 **Update README.md**  
 Maria Claudia BODINO authored 18 hours ago 85da3666 

 **Code owners** Assign users and groups as approvers for specific file changes. [Learn more.](#) Manage branch rules

master  bdti-essentials-course / Session 1: Data Access and Exploration /  Lock History Find file Edit  Code 

Name	Last commit	Last update
..		
 img	Delete session1_4.png	1 week ago
 .gitkeep	Add new directory for session 1	2 weeks ago
 BDTI_Essentials_Session1_DataAccess_Exploration.pdf	Upload New File	1 week ago
 Exercise_S1.knwf	Upload New File	1 week ago
 README.md	Update README.md	18 hours ago
 Session1.R	Upload New File	1 week ago
 Session1.ipynb	Upload New File	1 week ago
 Solution_S1.knwf	Upload New File	1 week ago







# Recap Session 4 – Analytics, Visualisation, and Reporting

# How do we get insights?

## 1) Aggregation

Create a summarised version of the data

## 2) Pivoting

Rearrange data to analyse it from different perspectives.

## 3) Visualisation

Transform data tables to communicate their results visually



# Analytics: from objectives to insights



Objective	Insights
Identify countries with the most university involvement in green projects	Aggregate number of projects and plot data with a stacked bar chart
Identify the most active EU-based universities involved in green projects	Aggregate number of projects and plot data with a bar chart
Identify trends in started projects per country per year	Plot data with a line plot
Explore the relation between number of projects per university, energy project funding per university and a country's CO2 emission	Filter plot data with a parallel coordinates plot



# Exercise Review



# **Session 5: Advanced Analytics - Gathering data from the Web and Geo Visualisations**

---

# Session 5: Advanced module



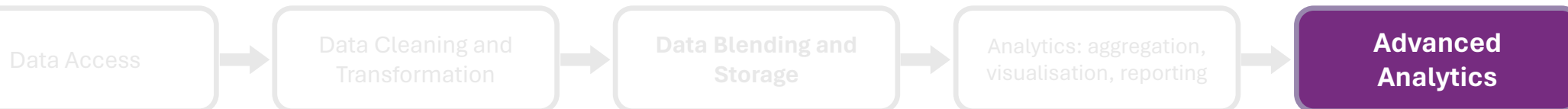
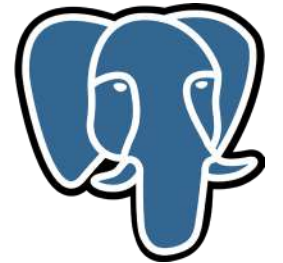
Session	Topic	Duration
Webinar 1	Data Access and Exploration	45 min
Webinar 2	Data Cleaning and Transformation	75 min
Webinar 3	Data Blending and Storage	75 min
Webinar 4	Analytics	75 min
<b>Webinar 5</b>	<b>Advanced Module: Gathering Data from the Web and Geo Visualisation</b>	<b>75 min</b>

- Slides, hands-on exercises, and solutions will be provided on the webinar
- Gitlab repository <https://code.europa.eu/bdti/bdti-essentials-course>

# Learning Outcomes

**At the end of this session, you should be able to:**

1. Sign in to the BDTI framework initialise KNIME, PgAdmin and PostgreSQL
2. Understand the fundamentals of REST Services to gather data from the internet
3. Retrieve Geospatial data using an API
4. Append Geospatial data to the data table and visualise it on a map





# Gather data from the web



# Exploring the Power of Web Data



## 1 - Why (purpose)

- Access *external* data to **enrich** with it more information

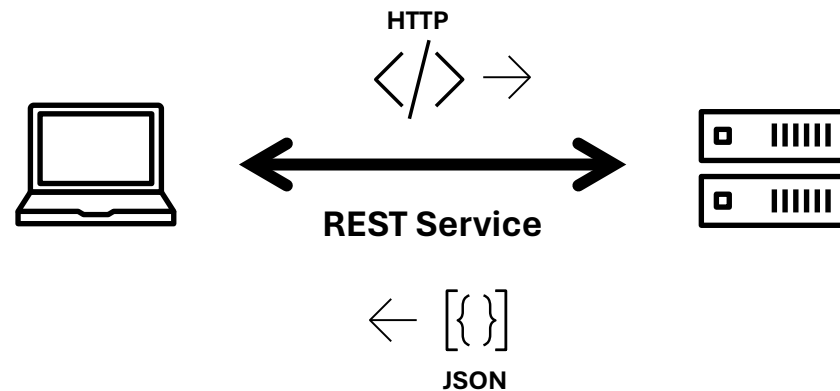
## 2 - How

- Extract web data with a REST Service Protocol
- Make use of API (Application Programming Interface)

# What Is REST Service

- **REST = RE**presentational **State T**ransfer

- A protocol for two computers to communicate over the Internet
- Offers web services for clients to communicate with a server
- Uses API (application programming interface) with the HTTP protocol
- Responses are often in JSON or XML objects



# Elements of REST Service Request



- Each request from a client to a server must contain
  - **Path:** Location of the web service the client is trying to access
  - **HTTP verb:** Type of operation to be performed
  - **Header:** Data for the operation
- HTTP verbs:
  - **GET:** To retrieve information only
  - **POST:** To generate resource or information
  - **PUT:** To update the existing resource or information
  - **DELETE:** To delete resource or information
  - **PATCH:** To partially update resource or information

# What is an API?



## Application Programming Interfaces

- set of rules and protocols that allow different software applications to communicate with each other.

## How does it work?

- defines the methods and data formats to request and exchange information.

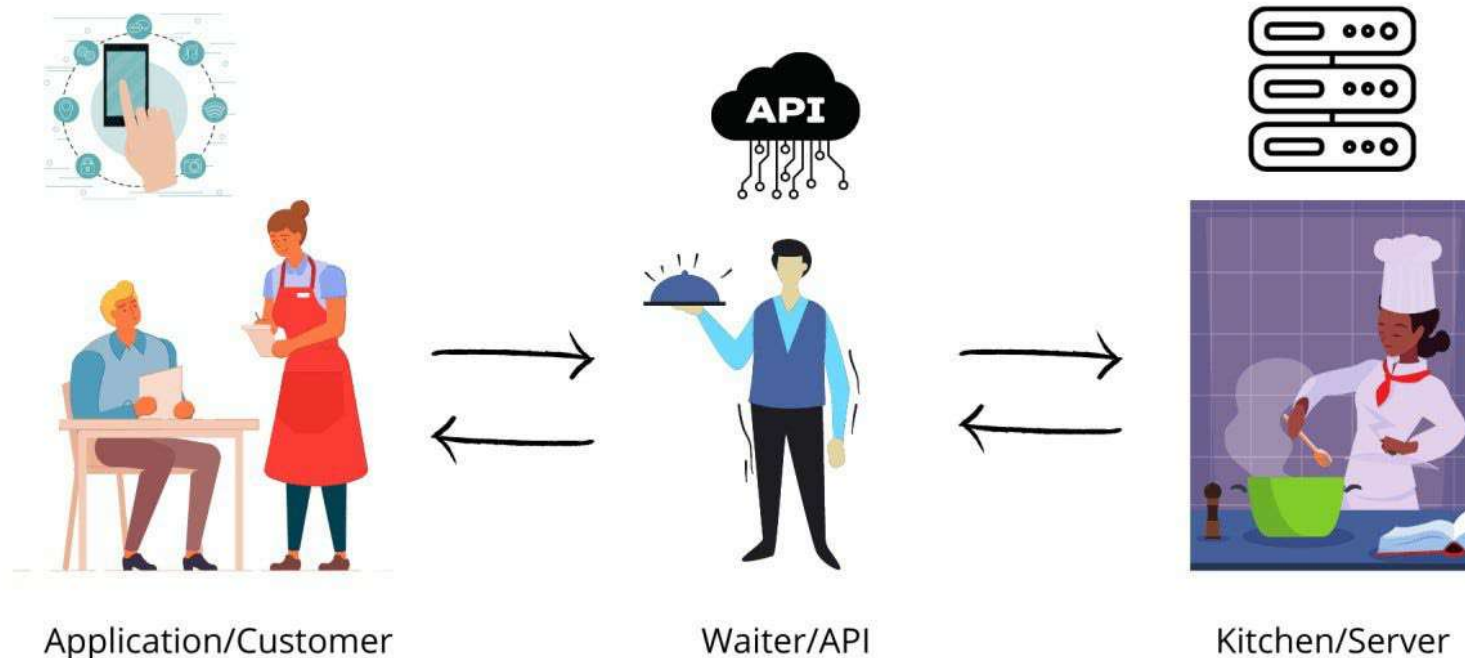
## Why are APIs important?

- enables to access the functionality of other applications or services, saving time (interoperability)



# What is an API?

Imagine an API as a waiter at a restaurant. You, the client (or application), place an order (request) with the waiter (API), who then communicates your request to the kitchen (server) and delivers the food (response) back to you.



# What is an API?

Let's play with an API

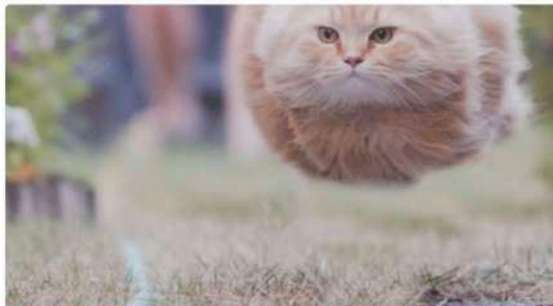


Usage:

```
https://http.cat/[status_code]
```



**Note:** If you need an extension at the end of the URL just add `.jpg`.



100

Continue



101

Switching Protocols



102

Processing



# Extract Open Data for Zoi

# Gather Web Data

Zoi and team wants to extract Geo data from  
“**Open Street Map**” (OSM)

- Free, open geographic database
- Updated and maintained by a community of volunteers
- <https://www.openstreetmap.org/>



## Use Case

Zoi wants to locate in a map the universities that participated in the Horizon program



# Gather Web Data: GET

Open Street Map is our “**Kitchen**”, and to communicate with it we need an API

- Every server has their own rules, is important to read the **documentation** on how to communicate with it
- <https://wiki.openstreetmap.org/wiki/Nominatim>



# Gather Web Data: GET



Nominatim is the API to communicate with OSM, our “**Waiter**”.



Nominatim

- Here we find the rules to extract the information we need.
- <https://nominatim.org/release-docs/develop/api/Search/>

The search API has the following format:

```
https://nominatim.openstreetmap.org/search?<params>
```

# Gather Web Data: GET

Zoi is the “**Customer**” and she wants to communicate with OSM

- Zoi wants to retrieve data from OSM, so it implements a “GET Request” based on OSM rules.
- Zoi communicates with OSM via Nominatim.

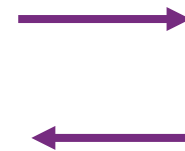


**Nominatim**



# Steps to gather Geo Data

- 1) Send a GET Request to OSM
- 2) If success request, we GET a response
- 3) Read and transform the response (geojson object)
- 4) Transform latitude and longitude to geo objects
- 5) Visualise the location on a Map



# 1) Send a Request to OSM

Zoi has the addresses of the universities and wants the geo-codes (latitude and longitude) to locate them in a map.



```
https://nominatim.openstreetmap.org/search?<params>
```

- In our case we will input the address in the <params> field (parameters)

<https://nominatim.openstreetmap.org/search?q=PIAZZA+LEONARDO+DA+VINCI+32+20133+MILANO+ITALY&format=geojson>



Nominatim



## 2) If success request, we get a response

```
{
  "type": "FeatureCollection",
  "licence": "Data © OpenStreetMap contributors, ODbL 1.0. https://osm.org/copyright",
  "features": [
    {
      "type": "Feature",
      "properties": {
        "place_id": "35811445",
        "osm_type": "node",
        "osm_id": "2846295644",
        "display_name": "17, Strada Pictor Alexandru Romano, Bukarest, Bucharest, S",
        "place_rank": "30",
        "category": "place",
        "type": "house",
        "importance": 0.62025
      },
      "bbox": [
        26.1156689,
        44.4354754,
        26.1157689,
        44.4355754
      ],
      "geometry": {
        "type": "Point",
        "coordinates": [
          26.1157189,
          44.4355254
        ]
      }
    }
  ]
}
```



Nominatim



# 3) Read the response and append columns



```
{
  "type": "FeatureCollection",
  "licence": "Data © OpenStreetMap contributors, ODbL 1.0. https://osm.org/copyrigh",
  "features": [
    {
      "type": "Feature",
      "properties": {
        "place_id": "35811445",
        "osm_type": "node",
        "osm_id": "2846295644",
        "display_name": "17, Strada Pictor Alexandru Romano, Bukarest, Bucharest, S",
        "place_rank": "30",
        "category": "place",
        "type": "house",
        "importance": 0.62025
      },
      "bbox": [
        26.1156689,
        44.4354754,
        26.1157689,
        44.4355254
      ],
      "geometry": {
        "type": "Point",
        "coordinates": [
          26.1157189,
          44.4355254
        ]
      }
    }
  ]
}
```

URL	Status	Content t...	longitude	latitude
String	Number (inte...	String	Number (dou...	Number (...)
https://nomin...	200	application/js.	12.524	55.786
https://nomin...	200	application/js.	4.376	51.999
https://nomin...	200	application/js.	9.227	45.478
https://nomin...	200	application/js.	8.412	49.009
https://nomin...	200	application/js.	7.662	45.063
https://nomin...	200	application/js.	5.486	51.447
https://nomin...	200	application/js.	9.984	57.014
https://nomin...	200	application/js.	6.077	50.778



Nominatim



## 4) Transform latitude and longitude to geo objects



The data extracted is number, and to visualise it on a map, we need to transform it to the correct Geodata type.

URL	Status	Content t...	longitude	latitude	geometry
https://nomin...	200	application/js...	12.524	55.786	POINT - EPSG...
https://nomin...	200	application/js...	4.376	51.999	POINT - EPSG...
https://nomin...	200	application/js...	9.227	45.478	POINT - EPSG...
https://nomin...	200	application/js...	8.412	49.009	POINT - EPSG...
https://nomin...	200	application/js...	7.662	45.063	POINT - EPSG...
https://nomin...	200	application/js...	5.486	51.447	POINT - EPSG...
https://nomin...	200	application/js...	9.984	57.014	POINT - EPSG...
https://nomin...	200	application/js...	6.077	50.778	POINT - EPSG...



# Geo data types



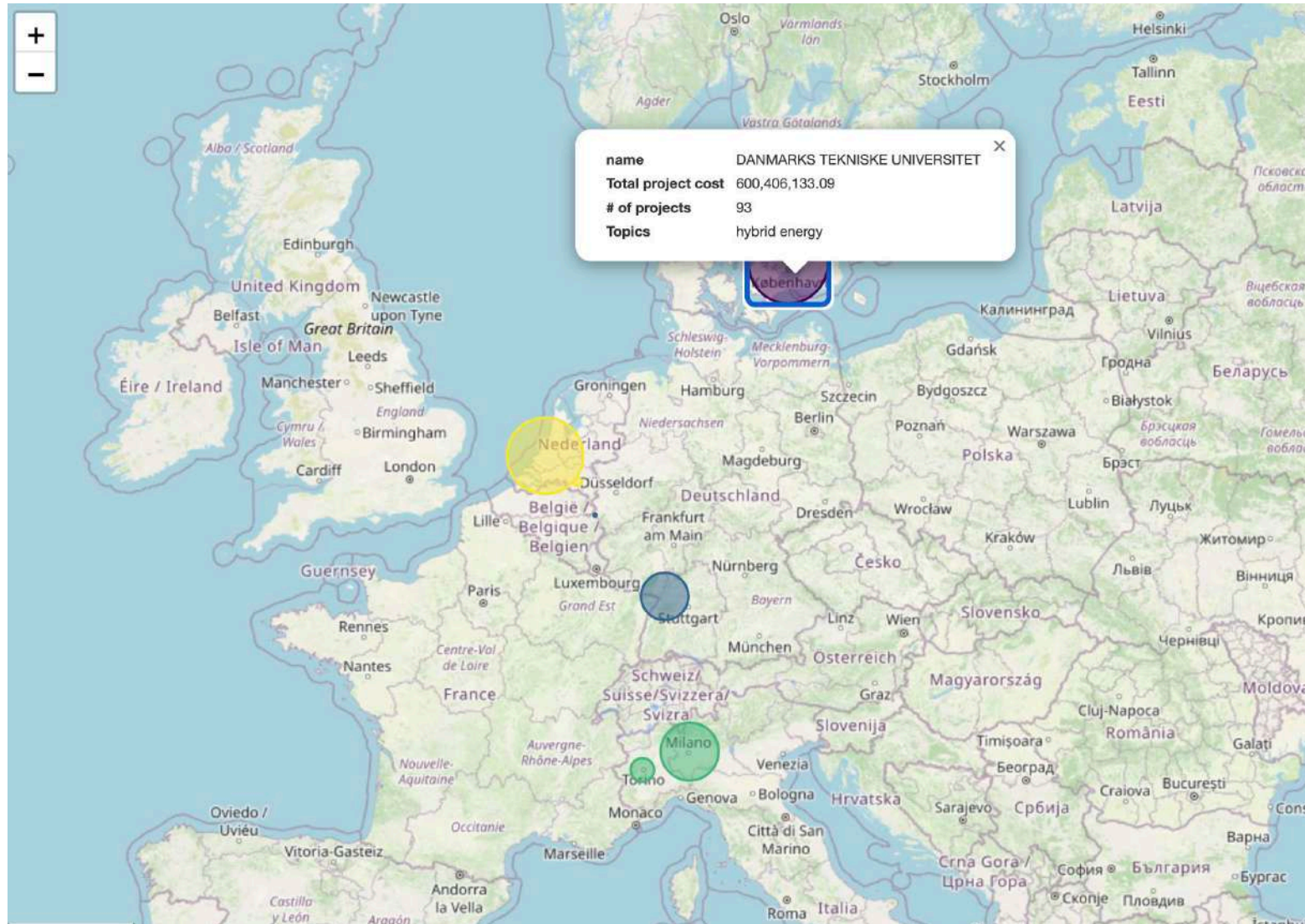
## 1. Vector Data: points, lines and polygons

1. Points → single locations (e.g., cities, *universities*).
2. Lines → linear features (e.g., roads, rivers).
3. Polygons → areas or regions (e.g., countries, lakes).

## 2. Raster Data: grid of cells or pixels

1. Each cell holds a value representing a certain attribute (e.g., elevation, temperature).
2. Used for continuous data such as elevation models or satellite imagery.

# 5) Visualise the location on a Map



# GET Request in the different BDTI tools



## PYTHON

```
import requests
# Define the API endpoint URL
url =
'https://nominatim.openstreetmap.org/search<params>'
# Send a GET request to the API endpoint
response = requests.get(url)

# Check if the request was successful (status code 200)
if response.status_code == 200:
    # Print the response content (data retrieved from the API)
    print(response.json())
else:
    print('Error:', response.status_code)
```

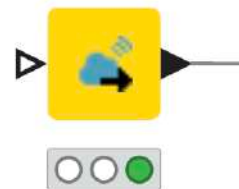
## R

```
library(httr)
# Define the API endpoint URL
url <- "
https://nominatim.openstreetmap.org/search<params>"

# Send a GET request to the API endpoint
response <- GET(url)

# Check if the request was successful (status code 200)
if (status_code(response) == 200) {
    # Print the response content (data retrieved from the API)
    print(content(response))
} else {
    print(paste("Error:", status_code(response)))
}
```

## KNIME GET Request



# Recap: Steps to gather Geo Data

- 1) Send a GET Request to OSM using a BDTI-tool
- 2) If success request, we GET a response
- 3) Read the response and transform (geojson object)
- 4) Transform latitude and longitude to geo objects
- 5) Visualise the location on a Map

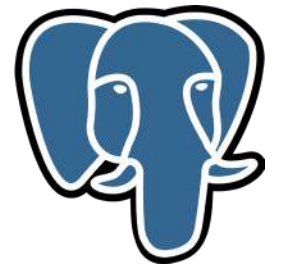


# Summary



## Now you should be able to:

1. Sign in to the BDTI framework initialise KNIME, PgAdmin and PostgreSQL
2. Understand the fundamentals of REST Services to gather data from the internet
3. Retrieve Geospatial data using an API
4. Append Geospatial data to the data table and visualise it on a map



# Bonus

More open data sources to keep enriching ours!  
- [https://www.wikidata.org/wiki/Wikidata:REST\\_API](https://www.wikidata.org/wiki/Wikidata:REST_API)



A screenshot of the Wikidata main page. At the top, there's a navigation bar with 'Main Page' and 'Discussion' tabs, and a search bar. The main content area features a large network diagram with nodes and edges, and a central 'Welcome to Wikidata' message. Below this, there are several informational boxes: 'Welcome!', 'Learn about data', and 'Get Involved'. The 'Learn about data' box includes images of Earth, a document, and Mount Everest, with labels for 'Item: Earth (Q2)', 'Property: highest point (P610)', and 'custom value: Mount Everest (Q513)'. The 'Get Involved' box provides links to the community portal and other resources. The left sidebar contains a list of navigation links and tools.





# Webinars Recap

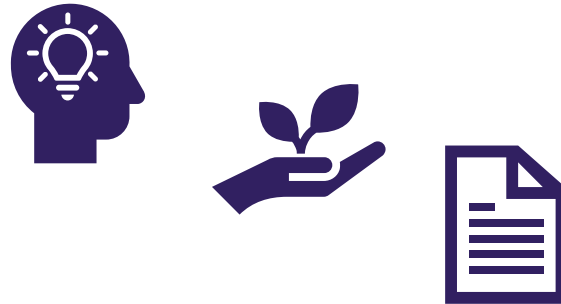
# Use case:

## "Identify Innovative Green Energy Projects from EU Universities"

Zoi and her team, work in the **Educational Department of a highly polluted EU region.**

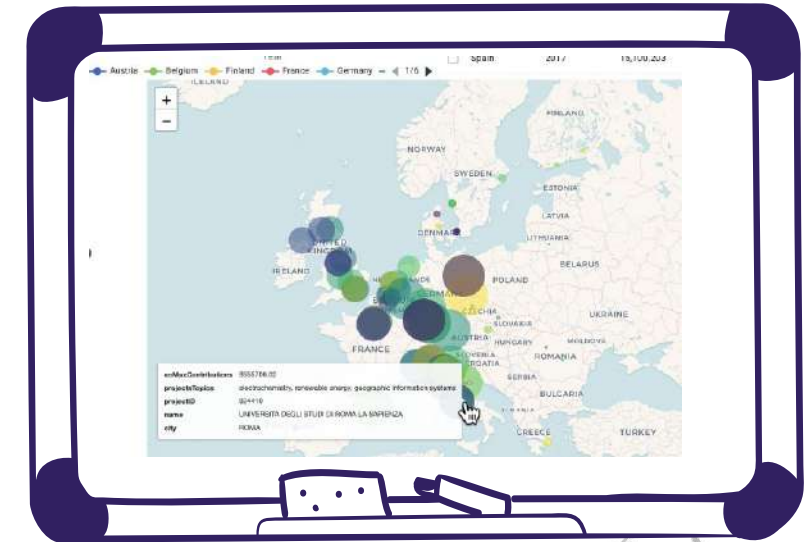
Their **purpose** is to identify potential **"green energy partnerships"** between EU universities to address pollution issues.

Zoi reached out the BDTI team to learn how to **connect the dots**, use the free BDTI playground to **create insights** and produce a **report**, which can be **easily updated** when new data is available.



They found some interesting **open data**:





- **List of innovation green energy projects** across EU countries
- **CO2 emissions** of these countries

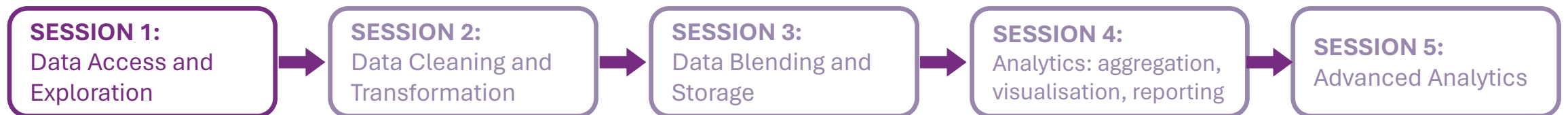




# Zoi Data Journey: Session 1




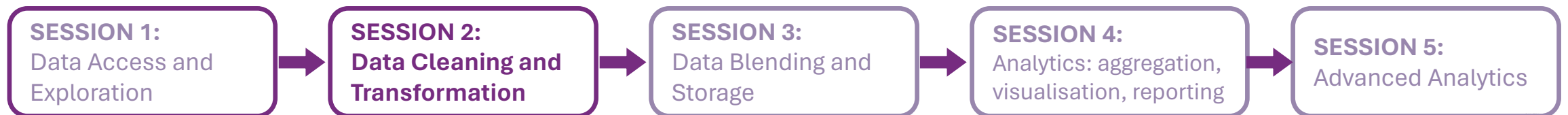
Goal	Data Sources and formats	BDTI tools
<b>Access</b> the data from multiple sources and <b>explore</b> it to <b>make decisions</b>		 <small>Open for Innovation</small> 
	csv, Excel, JSON	



# Zoi Data Journey: Session 2





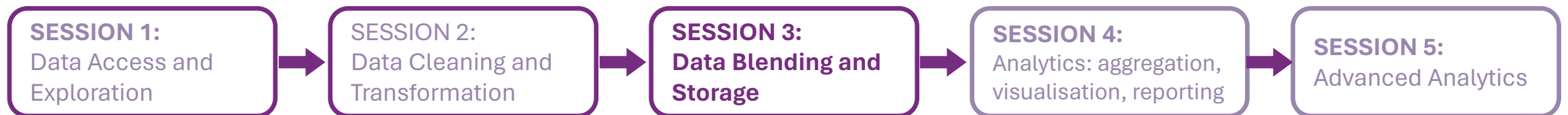
Goal	BDTI tools
<p>Learn skills to clean, transform, and ensure the datasets are in the right format for subsequent analysis</p> <p>From <b>raw</b> data to <b>clean</b> and <b>transformed</b> data</p>	



# Zoi Data Journey: Session 3






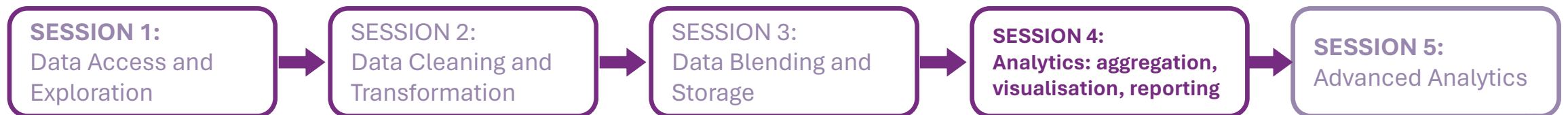
Goal	BDTI tools
<p>Learn to <b>blend</b> all the different data sources by <b>stacking</b> rows and <b>joining</b> columns.</p> <p>End of ETL process and <b>storing</b> the data in different formats such as CSV and in a <b>database</b></p>	 



# Zoi Data Journey: Session 4





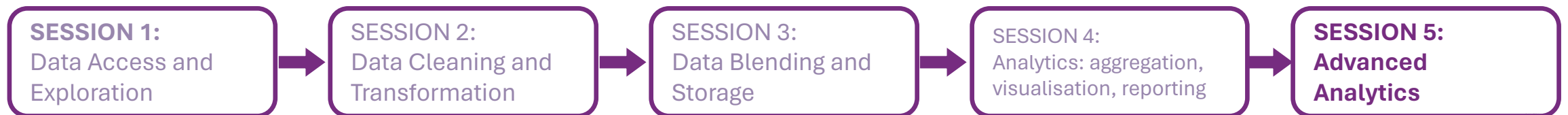
Goal	BDTI tools
<p>Learn to <b>extract insights</b> from the data by <b>aggregation, pivoting</b> and <b>visualisation</b>.</p> <p>Create a <b>Report</b> with this information.</p>	  



# Zoi Data Journey: Session 5



Goal	BDTI tools
Learn how to <b>enrich</b> the data with <b>external</b> data sources using REST protocol.	 



# Practice, practice and practice!



- Slides, hands-on exercises, and solutions will be provided on the webinar Gitlab repository
- <https://code.europa.eu/bdti/bdti-essentials-course>

A screenshot of the GitLab repository page for 'BDTI Essentials Course'. The page shows the repository name, project ID (635), and a 'Request Access' button. It indicates 12 commits, 1 branch, 0 tags, and 2 MiB of project storage. A recent commit by Maria Claudia BODINO is highlighted, titled 'Upload banner session 1', with a commit hash of 9cce1d4d. Below this, a table lists the repository's files and their last commit details.

Name	Last commit	Last update
Session 1: Data Access and Exploration	Upload banner session 1	4 days ago
BDTI_Banner_generic.png	Upload New File	5 days ago
LICENSE	Add LICENSE	5 days ago
README.md	Update README.md	5 days ago

Below the table, the README.md content is visible, including copyright information for the European Union and a license notice. At the bottom of the README, there is a banner image with the European Commission logo and the text 'From hype to action:' followed by a graphic with the 'DIGITAL EUROPE PROGRAMME' logo.

# Practice, practice and practice!

- Slides, hands-on exercises, and solutions will be provided on the webinar Gitlab repository
- <https://code.europa.eu/bdti/bdti-essentials-course>

## Session 5

### External data sources

[https://www.youtube.com/watch?v=Snj3\\_XPcJtU](https://www.youtube.com/watch?v=Snj3_XPcJtU)  
<https://nominatim.org/release/docs/develop/api/Search/#geojson>

### Geo data visualizations

Create nicer visualisation merging all the data

### Retrieve Data from the Web

Static Example with only one address

### Geo Visualizations:

### Preparing the data for input

Dynamic Example with addresses from the data





**Q&A**

---



# Well done!



**Attended all 5 sessions? You deserve a **reward!****

Our kind partners at KNIME are offering free access to all learning resources, including online courses and ebooks.

The BDTI team will send those who qualify a redemption code and instructions for use via email in the coming days.



# Congratulations on completing the course!



•  
**BDTI  
Essentials  
Course**



**You are here**



•  
**Brainstorm your data project**



•  
**Apply for BDTI Pilot-  
light process**

Ready to discuss your data project?



# What's next?



You are here



Information session with the BDTI team



## Am I eligible?



**European Public Administrations**  
at local, regional and national level



**Academia** and **private sector** can apply for once there is a clear collaboration with a Public Administration



**Not sure?** Check our BDTI CANVAS and let's discuss your idea

**EC-BDTI-PILOTS@ec.europa.eu**



# Learn more about BDTI



Email the team: [EC-BDTI-PILOTS@ec.europa.eu](mailto:EC-BDTI-PILOTS@ec.europa.eu)

Visit the BDTI  
website



Stay up to date with  
news & activities



<https://big-data-test-infrastructure.ec.europa.eu/>





**Thank you!**

---