Enabling a data-informed public sector:

From hype to action using

*the **Big Data Test Infrastructure (BDTI)***

Maria Claudia BODINO, BDTI project officer – European Commission

mariaclaudia.bodino@ec.europa.eu

Business Owner:
DG CNECT
Directorate-General for Communications Networks, Content and Technology

Service Provider:
DG DIGIT
Directorate-General for Digital Services

Road Map

1 **BDTI in a nutshell**
- Its context and why use it

2 **BDTI in practice**
- Access and overview of the BDTI portal
- Concrete application of the BDTI

3 **BDTI's community**
- Developing the BDTI community and how can you help us

1

BDTI in a nutshell
- Its context and why use it

What is the Big Data Test Infrastructure (BDTI) ?

Not <u>only</u> for big data, for public sector in general (open data)

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

→ To help the public sector **to derive insights from data**
and accelerate transition towards **data-informed decision making.**

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

2 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 as long as there there is a **clear collaboration** with a Public Administration which will be the main point of contact for the project (Master/PhD, GovTech startups)
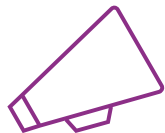
We need YOUR support in promoting BDTI between your public administrations

Contact us:

EC-BDTI-PILOTS@ec.europa.eu

Big Data Test Infrastructure Objectives

Objectives

- Increase the easy accessibility, interoperability, quality and usability of public sector information in compliance with the requirement of the **Open Data Directive**
- Boost the **re-use and combination of open public data** across the EU for the development of information products and services, including AI applications
- Testing **Business-to-Government** (B2G) data sharing collaborations for the **public good**
- Data Space Support Centre: explore and experiment with your data*
    - BDTI provides a safe **testing environment to run big data experiment**s for data space customers

* https://joinup.ec.europa.eu/collection/semic-support-centre/data-spaces

Why use the BDTI ?

Benefit of six months free of charge service, including **advisory and technical** support for the duration of the pilot

Experiment with data analytics using high **performance infrastructure** that leverages the power of the **elastic cloud**

Receive guidance to move from a pilot to a **production-ready** process – **EXIT package**

→ **Test your idea → Extract value → Create knowledge**

Why use the BDTI ?

Data → Information → Presentation → Knowledge
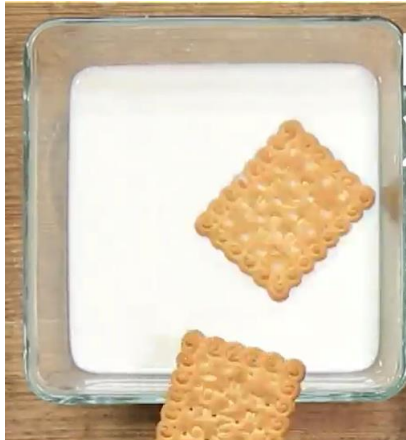
You have the key ingredients (datasets),
we equip you with the best open source tool
to generate amazing recipes.

https://funtip.giallozafferano.it/Torta-mattonella.html

2

With its open-source tools, BDTI supports you throughout your data journey

Orchestration

Apache Airflow

5. **Decision-Making**

1. **Collection**

MINIO — Data Lake

OPENLINK VIRTUOSO UNIVERSAL SERVER

mongoDB — Database

Metabase

Apache Superset

4. **Visualisation**

Development Environments

KNIME — Open for Innovation

R Studio

H2O.ai   jupyter

3. **Analysis**

2. **Processing**

elasticsearch

Apache Spark — Advanced Processing Engines

kibana

BDTI's Data Analytics Stack

https://big-data-test-infrastructure.ec.europa.eu/service-offering_en
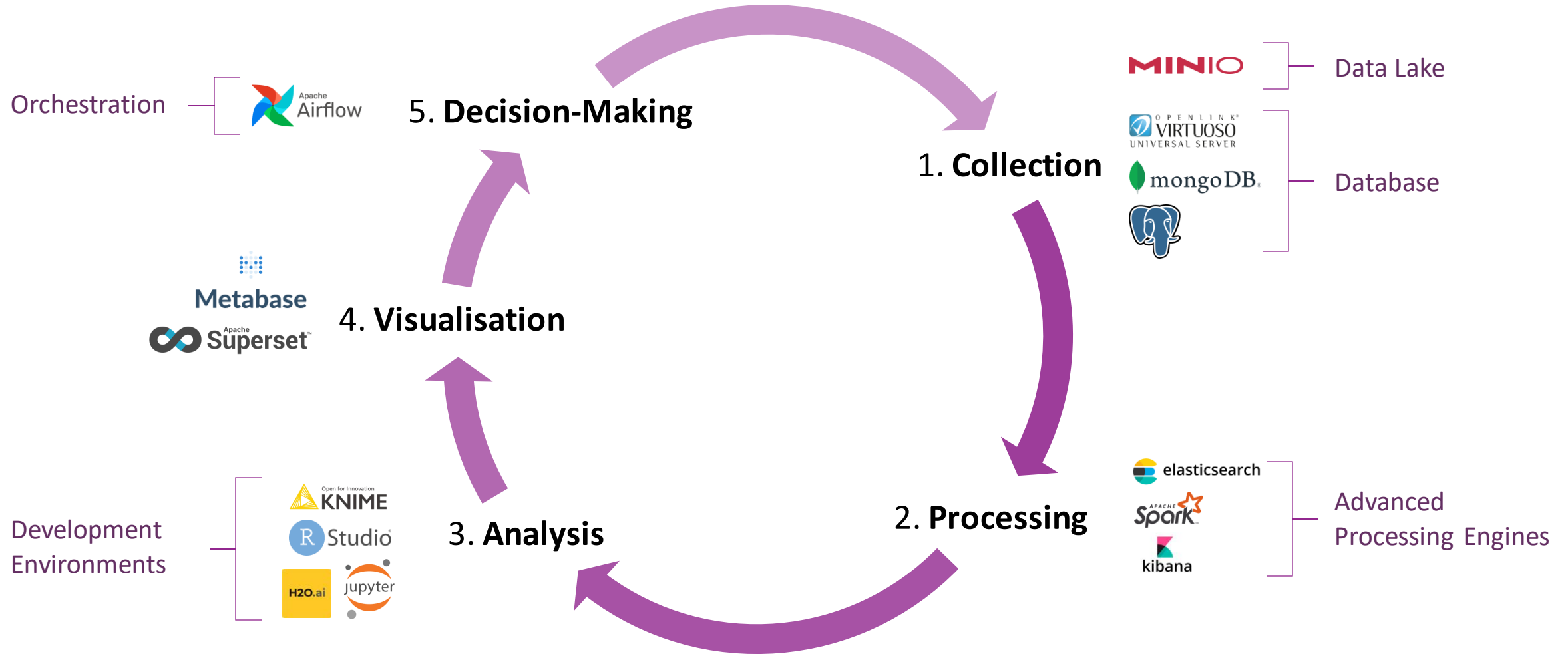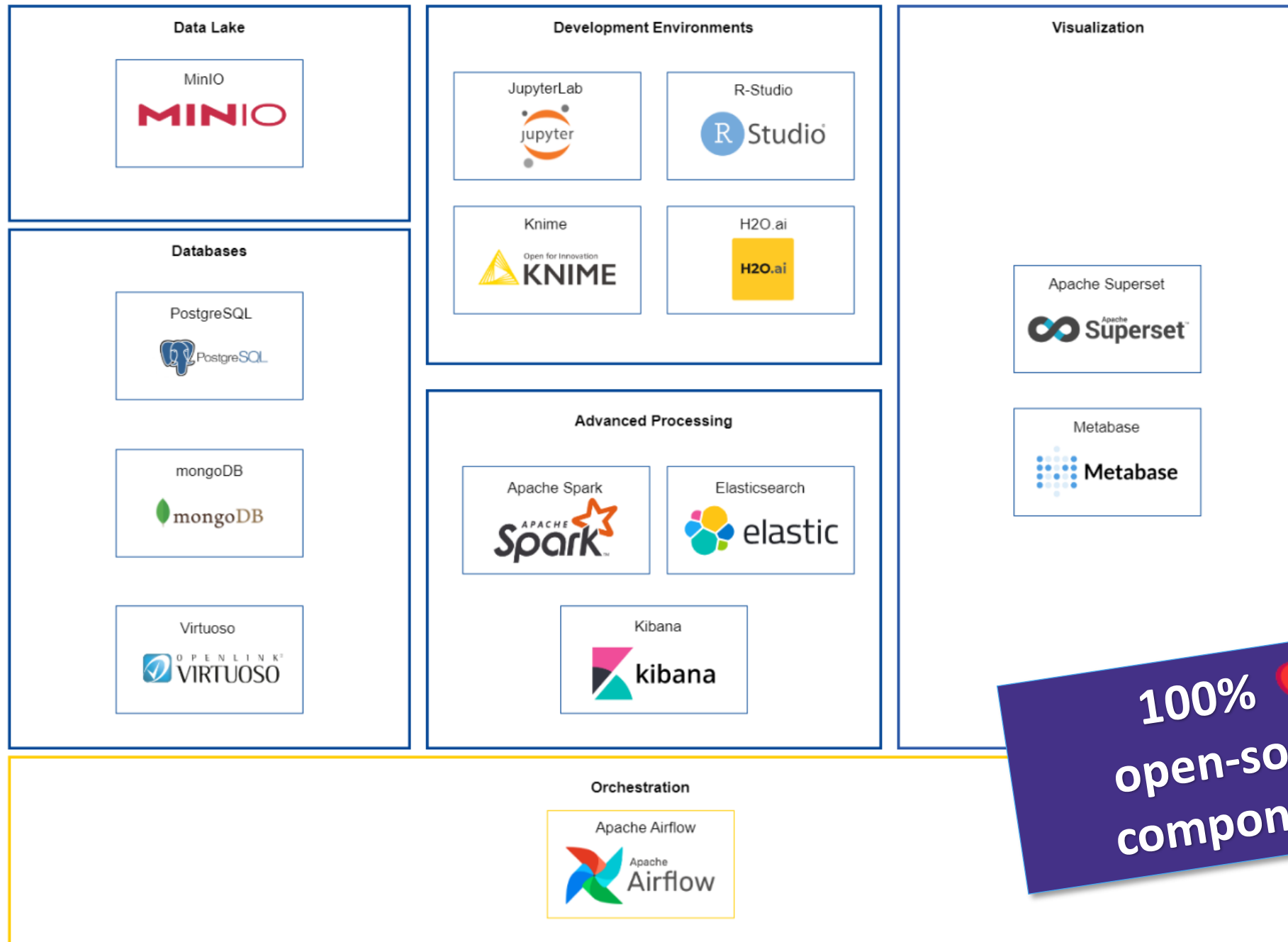
3 BDTI in practice
- Access and overview of the BDTI portal
- Concrete application of the BDTI

3 Access to BDTI portal directly from your browser (EU Login integration)

For teams part of BDTI pilots

3
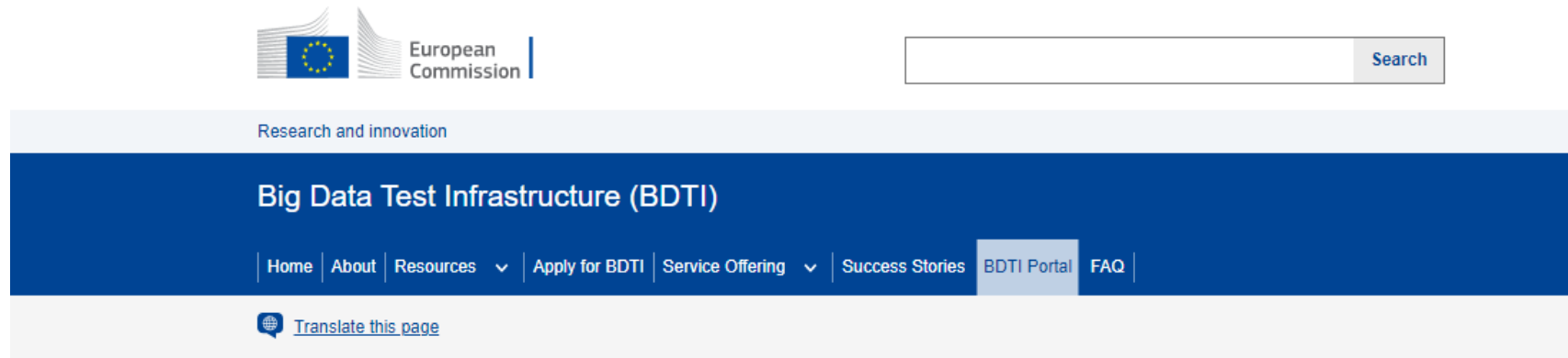
The BDTI portal

The BDTI portal: service catalogue

BDTI Demonstrator:
Towards Data-Informed Government Spending

<u>Goal:</u>
Show how the BDTI can be used by different users (at different levels of complexity) to **derive insights from government spendings to take data-informed actions**

A <u>**user-centered**</u> approach:
- Elena and Daniel, public servants
- Low data literacy skills
- **Problem**: high government spending in public lighting
- **Solution**: how to optimise public lighting to reduce government spending

Data Ingestion: Automating Invoice Extraction

KNIME — Open for Innovation

ETL - Data extraction from non-machine readable **PDF files**

PostgreSQL

Storage & structuring of collected data

3 Data visualisation and analysis

Apache **Superset**™

Interactive overview of expenditures per category.

Ability to benchmark against different data sets.

https://big-data-test-infrastructure.ec.europa.eu/resources/bdti-value-demonstrator_en
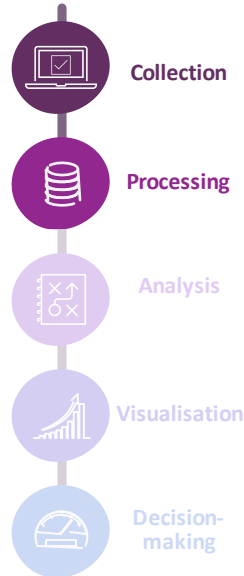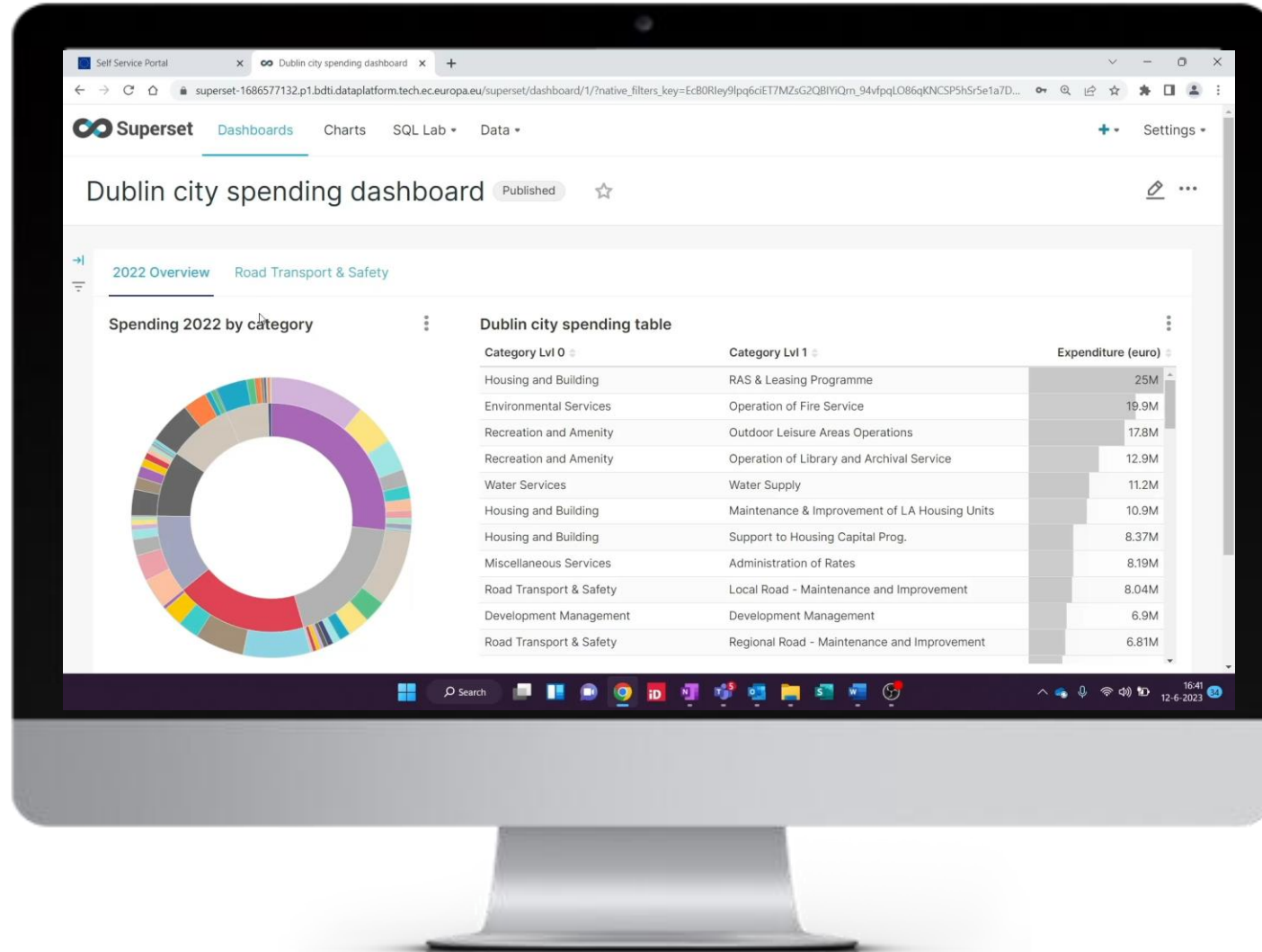
3 Dublin's Data-Driven Energy Expenditure Reduction

To train Machine Learning (ML) models to analyse big (or small) amount of data

PostgreSQL

Stores the newly analysed data

https://big-data-test-infrastructure.ec.europa.eu/resources/bdti-value-demonstrator_en
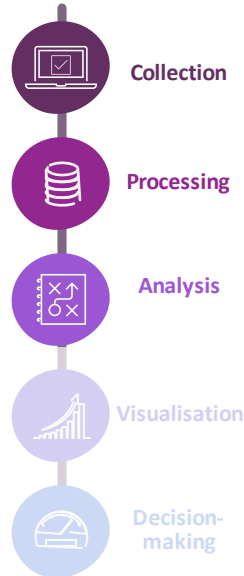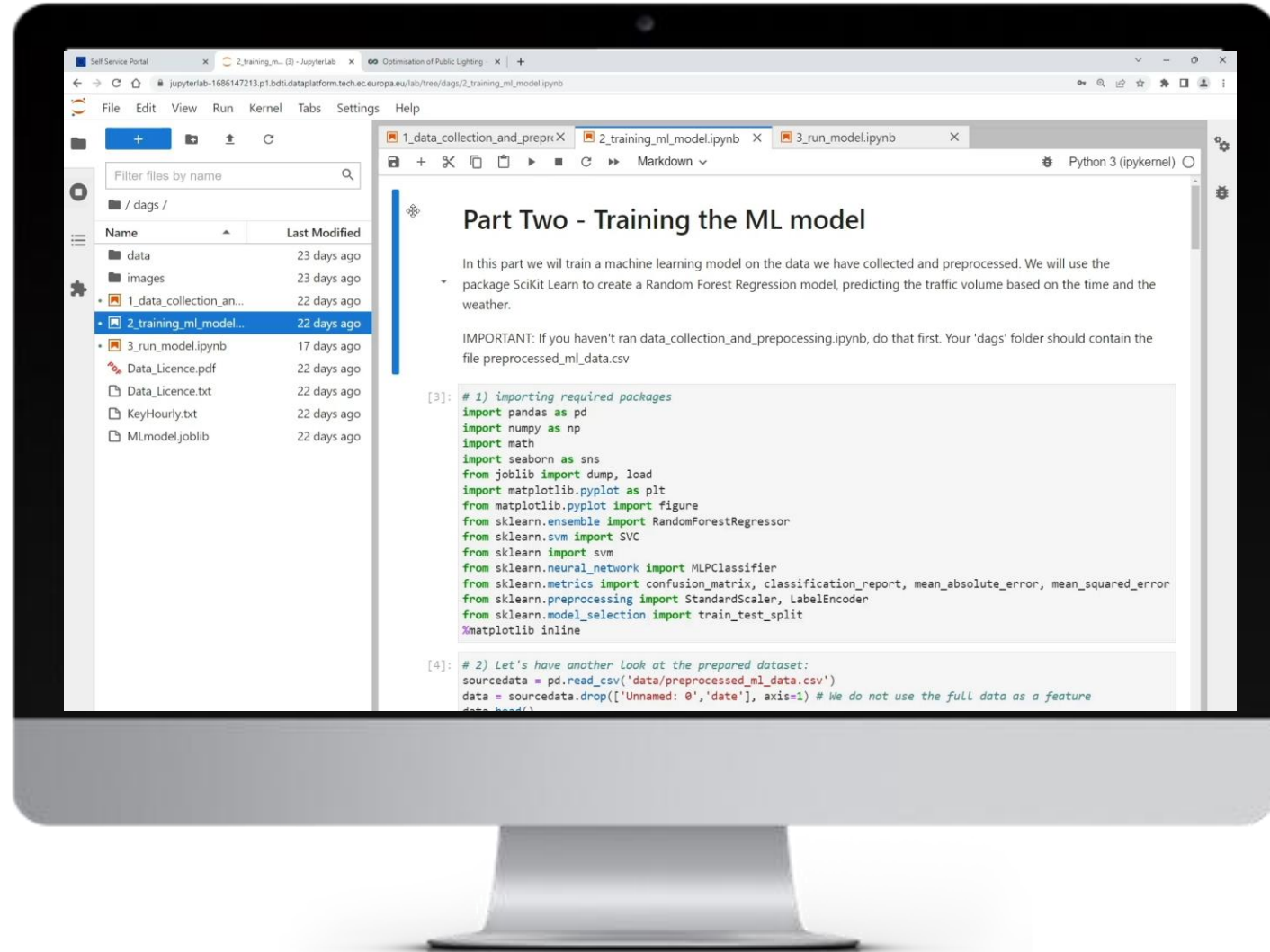
3 **Communication and implementation**

Apache **Superset**™

Can display the created data through an interactive dashboard.
Thanks to the ML trained model, it is possible to create different simulations and visualise the outcome through different graphics
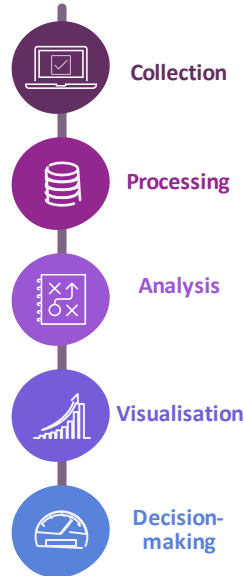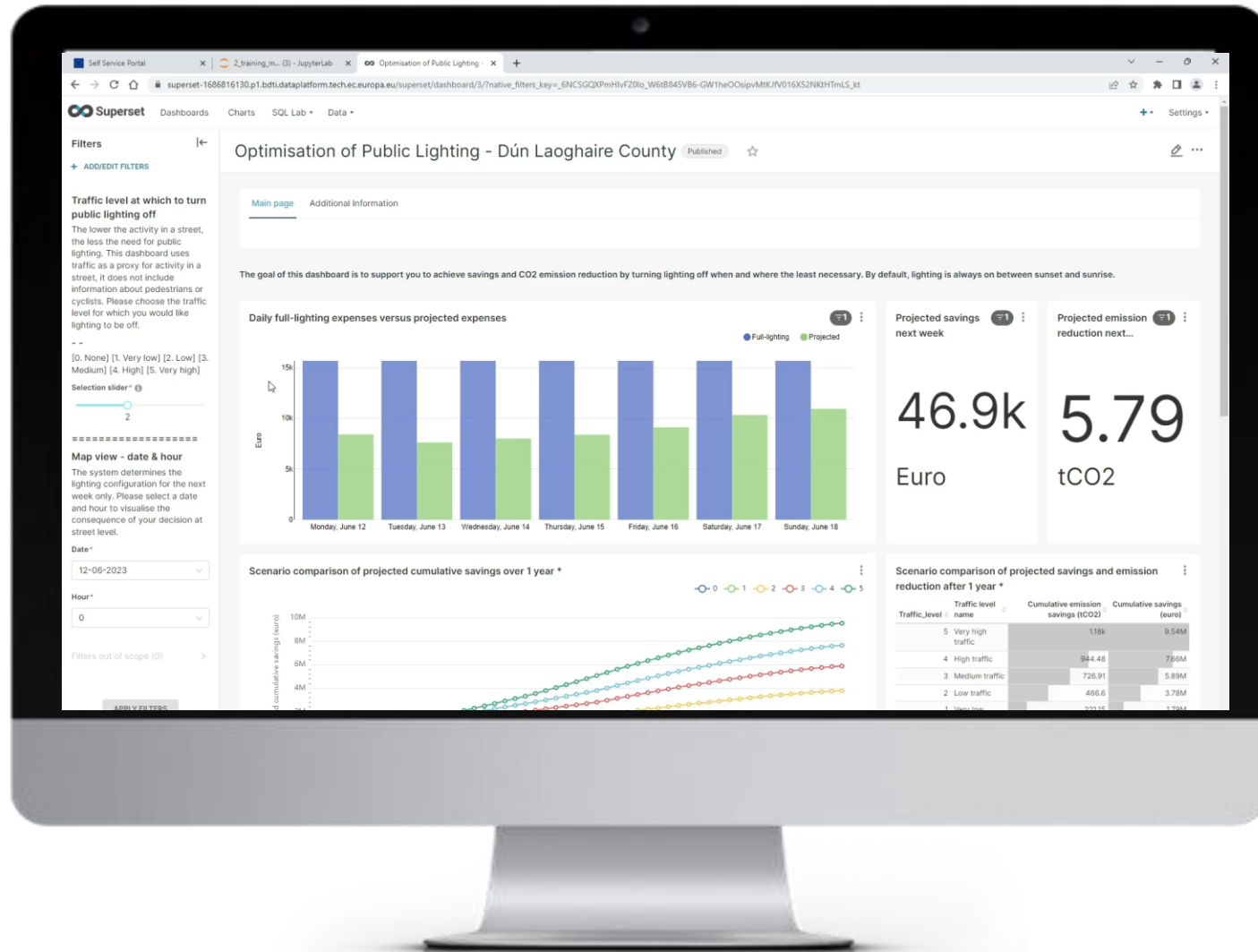
https://big-data-test-infrastructure.ec.europa.eu/resources/bdti-value-demonstrator_en

OPEN SOURCE CODE: https://code.europa.eu/bdti/bdti-demonstrator

3

4

BDTI's community
- Developing the BDTI community and how can you help us

Who used it already?

CONSELLERIA DE SANITAT (CS) - Text Mining

Conselleria de Sanitat, the Health Public Administration of the Comunidad Valenciana Regional Government, needed a tool capable of analysing and extracting knowledge from the huge quantity of scientific clinical articles coming from different sources (i.e. PubMed.gov, Covid-19 related clinical articles).

Advanced **data visualisation** and **text mining** tools to help **extract knowledge contained in the documents**, supporting clinicians and managers in their clinical practices and day-to-day work.

EU CONVALESCENT PLASMA DATABASE – Data sharing

The European Blood Alliance is working together with the European Commission (DG SANTE) to create and manage an **EU-wide open-access platform** that collects data to support a study on **Covid-19 convalescent plasma therapy**. The aim of the study is to assess in which conditions the convalescent plasma treatment is most effective, in order to take data-driven decisions on the therapy and focus the efforts of the research in the most promising directions.

A ready-to-use, virtual environment in which **data collected through a custom-built website** is ingested and anonymised, to be then analysed with advanced data visualisation and analytical tools. Initially, only donation data was processed, then the scope was increased to capture the **end-to-end of blood plasma, from donation to patient/clinical trial.**

CITY OF FLORENCE – Mobility data

The main goal of the Municipality is to perform a **cross correlation between the multiple datasets** available within the city to understand how people were and are moving between the different districts, to then derive precious insights about mobility and about **how services can be redesigned to foster cultural activities and events.**

Predictive, descriptive and time-series analysis on multiple datasets collected **before, during and after the Covid-19 pandemic** such as: public Wi-Fi sensors, parking and geo-referenced data of people movements (i.e. tourists).
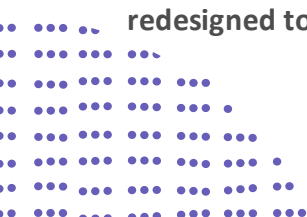
Who used it already?
Semantic Knowledge Graphs for Distributed Data Spaces

The Public Procurement Pilot Experience

Semantic Knowledge Graphs for Distributed Data Spaces: The Public Procurement Pilot Experience

Cecile Guasch[1], Giorgia Lodi[2]([✉]), and Sander Van Dooren[1]

[1] European Commission, DG DIGIT, Brussels, Belgium
{cecile.guasch,Sander.VAN-DOOREN}@ext.ec.europa.eu
[2] Institute of Cognitive Sciences and Technologies of the Italian National Resea
Council (ISTC-CNR), Rome, Italy
giorgia.lodi@cnr.it

Abstract. This paper presents the experience gained in the context of a European pilot project funded by the ISA2 programme. It aims at constructing a semantic knowledge graph that establishes a distributed data space for public procurement. We describe the results obtained, the follow up actions and the main lessons learnt from the construction of the knowledge graph. This latter requires to support different data governance scenarios: some partners control, with their own tools, the building process of their portion of the knowledge graph. Other partners participate in the pilot by providing only their open CSV/XML/JSON datasets, in which case transformations are required. These are performed on the infrastructure made available by the European Big Data Test Infrastructure (BDTI). The paper introduces the design and implementation of the knowledge graph construction process within such a BDTI infrastructure. By instantiating an OWL ontology created for this purpose, we are able to provide a declarative description of the whole workflow required to transform input data into RDF output data, which form the knowledge graph. The declarative description is therefore used to provide instructions to a workflow engine we use (Apache Airflow) for knowledge graph construction purposes.

Guasch, C., Lodi, G., & Dooren, S. V. (2022, October). Semantic Knowledge Graphs for Distributed Data Spaces: The Public Procurement Pilot Experience. In *The Semantic Web–ISWC 2022: 21st International Semantic Web Conference, Virtual Event, October 23–27, 2022, Proceedings* (pp. 753-769). Cham: Springer International Publishing. https://iswc2022.semanticweb.org/index.php/accepted-papers/

How to apply:

Get familiar with the BDTI service on our website

Brainstorm your data analytics project using our BDTI Canva and then fill in the BDTI template request form

Submit your pilot request (template) by email: EC-BDTI-PILOTS@ec.europa.eu

Meet with us to elaborate on your use case

Pilot Project is approved if:
- Brings value
- Can be completed in 6 months
- Sufficient resources available (skills, team, data)

Your test environment is set up

You can start piloting and create value!

DIGITAL EUROPE PROGRAMME

The BDTI Canva

by the BTDI Team

The BDTI Canva aims to help you build a strong data use case through a series of questions.

For more information, visit the BDTI website

Contact us by emai: EC-BDTI-PILOTS@ec.europa.eu

Context:
Who are you? Who are your stakeholders?

Objective(s):
What is the problem you are trying to address?
What is your timeframe?

Data's added value:
Which information helps you address the problem? From which sector and or domain?

Data's availability:
Does the data you need exist?If it doesn't exist, can you collect it? From whom can you get the data you need? Can you reuse the data? What license applies to the data you'd like to use? How is the quality of the data you'd like to use? Are the different datasets interoperable? Do you know how to connect the dots?

Data's risk(s):
What could go wrong when using data to address this objective? Are there legal and ethical considerations to make? Are you dealing with personal data?

Data's processing:
What do you need to gather, process and analyze the data (i.e., tools, software, computing power, ...)? Do you already have them? If you do not, where can you get them (e.g., applying to the BDTI)?

Data skills:
What data literacy and skills do you need (i.e., data engineering, data analysis, data science, data visualization)? Do you already have these available within your team/organization?

Your solution
Combine what you've learned from the elements above into a statement describing your solution

Register now to the BDTI Essentials Course
February 2, 2024

Unlocking the Potential of Open-Source Analytics: Introduction to BDTI Essentials

A series of five e-learning courses to discover the Big Data Test Infrastructure's set of open-source tools.

Starts February 2, 2024. Register now!

==Become familiar with the BDTI open-source data analytics tools==

✓ A **free course** helping public administrations explore BDTI delivered through a **practical use case**:

> Analysing H2020 funding allocated for research and innovation to universities across EU nations with high carbon emissions

✓ **Use open-data sources for public sector innovation:** Learning how to harness open data sources to address a real-world application by leveraging the resources offered by data.europa.eu

✓ **Prepare to build your own data use case:** After this course, you will be ready to apply for BDTI and build a public sector data use case using the platform.

https://big-data-test-infrastructure.ec.europa.eu/resources/courses-and-training/bdti-essentials-course_en

The BDTI Kitchen: Baking data

BDTI National Information Sessions

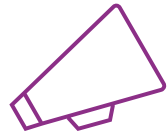**Goal**: introduce BDTI, learn about data analytics projects, develop your data analytics community!
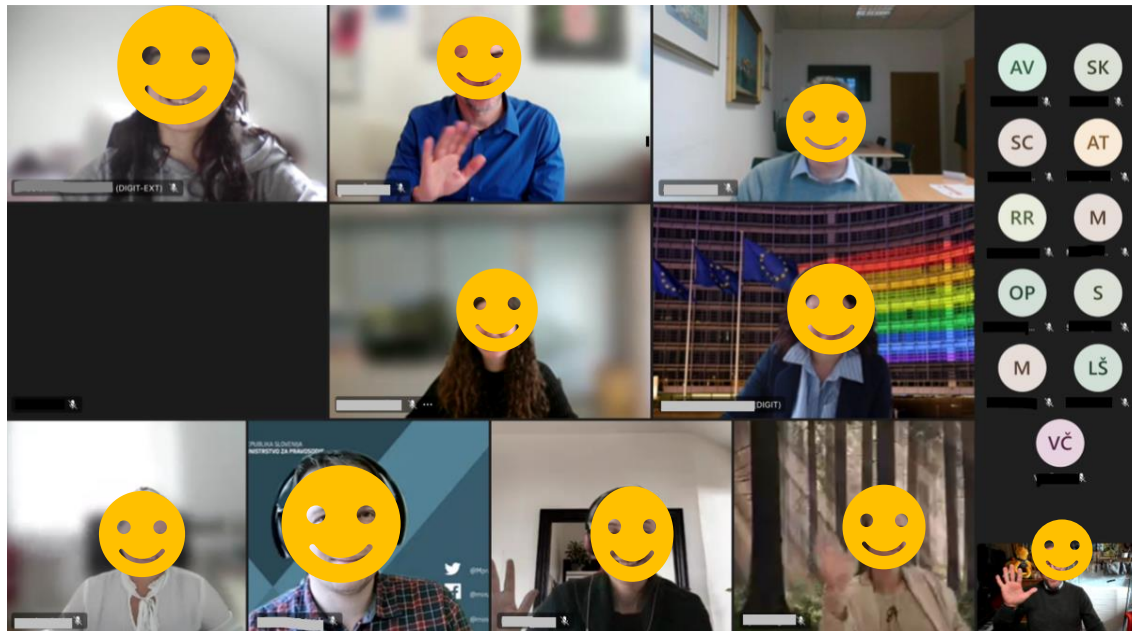
BDTI Information Session in Slovenia in collaboration with the **Slovenian Ministry of Digital Transformation**

BDTI Canva used in Mural during the BDTI Information Session in Slovenia

Get in touch and follow the BDTI activities

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

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

Visit BDTI's website

Subscribe to BDTI's newsletter

Subscribe to BDTI's Joinup

Register for BDTI Essentials online course

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

References

Academic references:

Guasch, C., Lodi, G., & Dooren, S. V. (2022, October). Semantic Knowledge Graphs for Distributed Data Spaces: The Public Procurement Pilot Experience. In The Semantic Web–ISWC 2022: 21st International Semantic Web Conference, Virtual Event, October 23–27, 2022, Proceedings (pp. 753-769). Cham: Springer International Publishing. https://iswc2022.semanticweb.org/index.php/accepted-papers/

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review*, *76*(6), 928-937.

Pirog, M. A. (2014). Data will drive innovation in public policy and management research in the next decade. *Journal of Policy Analysis and Management*, 537-543.

Tan, E., & Crompvoets, J. (Eds.). (2022). *The new digital era governance: How new digital technologies are shaping public governance*. Wageningen Academic Publishers.

European Commission websites:

https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data

https://commission.europa.eu/publications/interoperable-europe-act-proposal_en

https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

https://digital-strategy.ec.europa.eu/en/activities/digital-programme

https://dssc.eu/wp-content/uploads/2023/03/DSSC-Data-Spaces-Glossary-v1.0.pdf

https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces