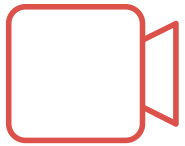
WEBINAR

Stories of use cases: Open data to make Europe fit for the digital age

30 June 2023

10.00 — 11.30 CET

Rules of the game



The webinar will be recorded

Please reserve 3 min after the webinar to help us improve by filling in our feedback form

For questions, please use the ClickMeeting chat.

data europa academy

Our speakers

Inmaculada Farfan Velasco
Project Manager, Data.europa.eu, Publications Office of the EU

Maria Claudia Bodino
IT Strategy and Planning Officer - Data Scientist, Big Data Test Infrastructure, Directorate for Informatics, European Commission

Tim Werkhoven
CEO, Lobium

Dimitris Michailidis
Co-founder, 100 Europeans

data europa academy

Agenda

10:00 – 10:05 CET	Introduction to the political priority ‘A Europe fit for the digital age' and the role of open data
10:05 – 10:20 CET	Get to know: *Big Data Test Infrastructure*
10:20 – 10:35 CET	Get to know: *Lobium*
10:35 – 10:50 CET	Get to know: *100 Europeans*
10:50 – 11:25 CET	Panel discussion
11:25 – 11:30 CET	Short recap of the previous editions and feedback

Introduction to the political priority ‘A Europe fit for the digital age' and the role of open data

A Europe fit for the digital age and open data

- **Empowering people** with a new generation of technologies, while achieving its target of a climate-neutral Europe by 2050
- **Transformation focusing on data, infrastructure and technology.**

Open data:

- Provides **tools and resources** to people and empower them with available free to access information.
- Boosts **economic growth** by providing more competitive advantage and build innovative products.
- Facilitates **innovation** in public administrations to provide better services to citizens

Get to know: Big Data Test Infrastructure

Maria Claudia Bodino

data europa academy

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

European Commission

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

Service Provider:
DG DIGIT
Directorate-General for Informatics

Road Map

1. **Policy context**

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

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

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

1 Policy context

1

Policy timeline

The European Data Strategy
February 2020

The Open Data Directive
July 2020

The Data Governance Act
November 2020

The AI Package
April 2021

The Data Act
February 2022

The Data Act
February 2022

The Interoperable Act
November 2022

Data Spaces

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

2

BDTI in a nutshell

- Its context and why use it

Big Data Test Infrastructure (BDTI) in a nutshell: its context

The BDTI is funded by the the Digital Europe Program (DEP), an **EU funding programme** (€7.5 bn) focused on **bringing digital technology** to businesses, citizens and **public administrations**.

The DEP provides strategic funding in **five crucial areas**:

- **High performance computing**
- **Cybersecurity**
- **Artificial intelligence** *(Cloud, data and AI)*
- **Advanced digital skills**
- **Deployment and wide use of digital technologies**

Learn more about the Digital Europe Programme: https://digital-strategy.ec.europa.eu/en/activities/digital-programme

2

What is the Big Data Test Infrastructure (BDTI) ?

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

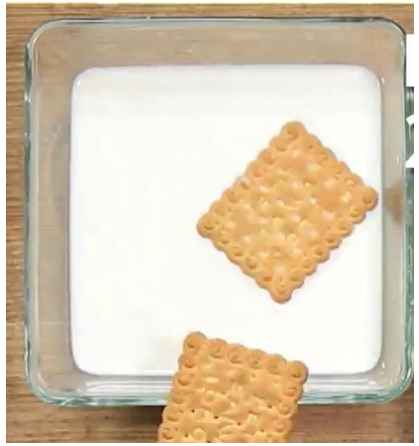
To help the public sector **to derive insights from data**
and accelerate transition towards **data- informed decision making.**

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

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

Why use the BDTI ?

Data → Information → Presentation → Knowledge





You have the key ingredients (datasets),
we provide you the best tool to generate amazing recipes.

data.europa.eu

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

Challenges

Legal, technological, organisational, cultural, ethical, behavioural and institutional challenges.

To mention some of them:

- Lack of **data skills** – limited understanding of data's potential and its value proposition
- Data sharing and **SILOS**
  - **PPP – smart cities…**
- Lack of high-**quality data –poor quality**
- Lack of effective **data governance**
  - Data stewards
- Data discovery and re-use for human and **machines**
  - **FAIR principles**
    - Findable
    - Accessible
    - **Interoperable –** Cross border and cross domain dimensions
    - Reusable - Licenses

Why use the BDTI ?

Benefit of six months free of charge service, including **advisory and technical** support during 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

Big Data Test Infrastructure 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.
- High Value Datasets – Open Data Directive
- Testing **Business-to-Government** (B2G) data sharing collaborations for the **public good**
- Data Space Support Centre: explore and experiment with Big Data
- BDTI provides a safe **testing environment to run big data experiments** for data space customers.

2

BDTI's Data Analytics Stack

Data Lake

MinIO

Databases

PostgreSQL

mongoDB

Virtuoso

Development Environments

JupyterLab

R-Studio

Knime

H2O.ai

Advanced Processing

Apache Spark

Elasticsearch

Kibana

Visualization

Apache Superset

Metabase

Orchestration

Apache Airflow

100% ♡ open-source components

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

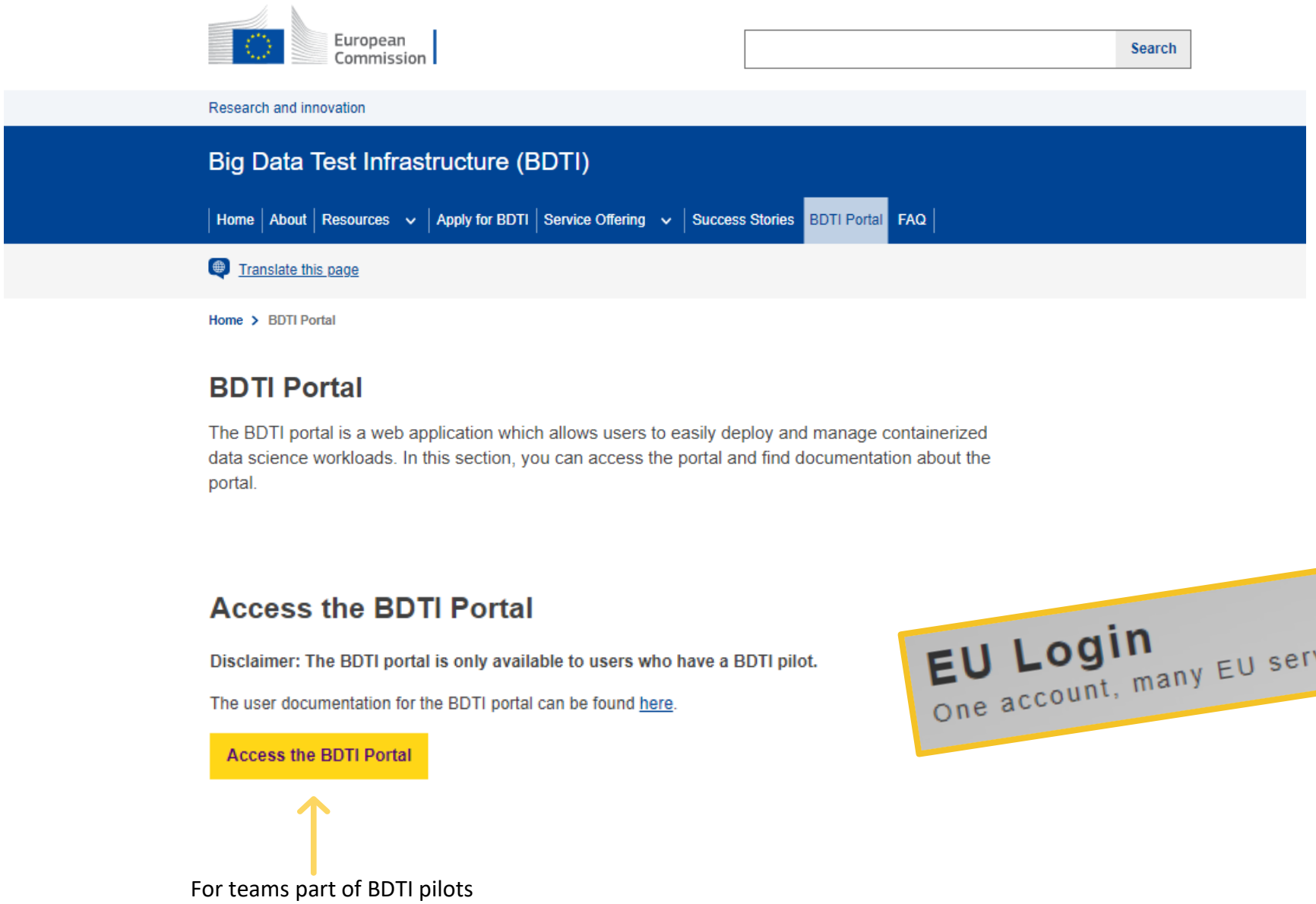
3

BDTI in practice

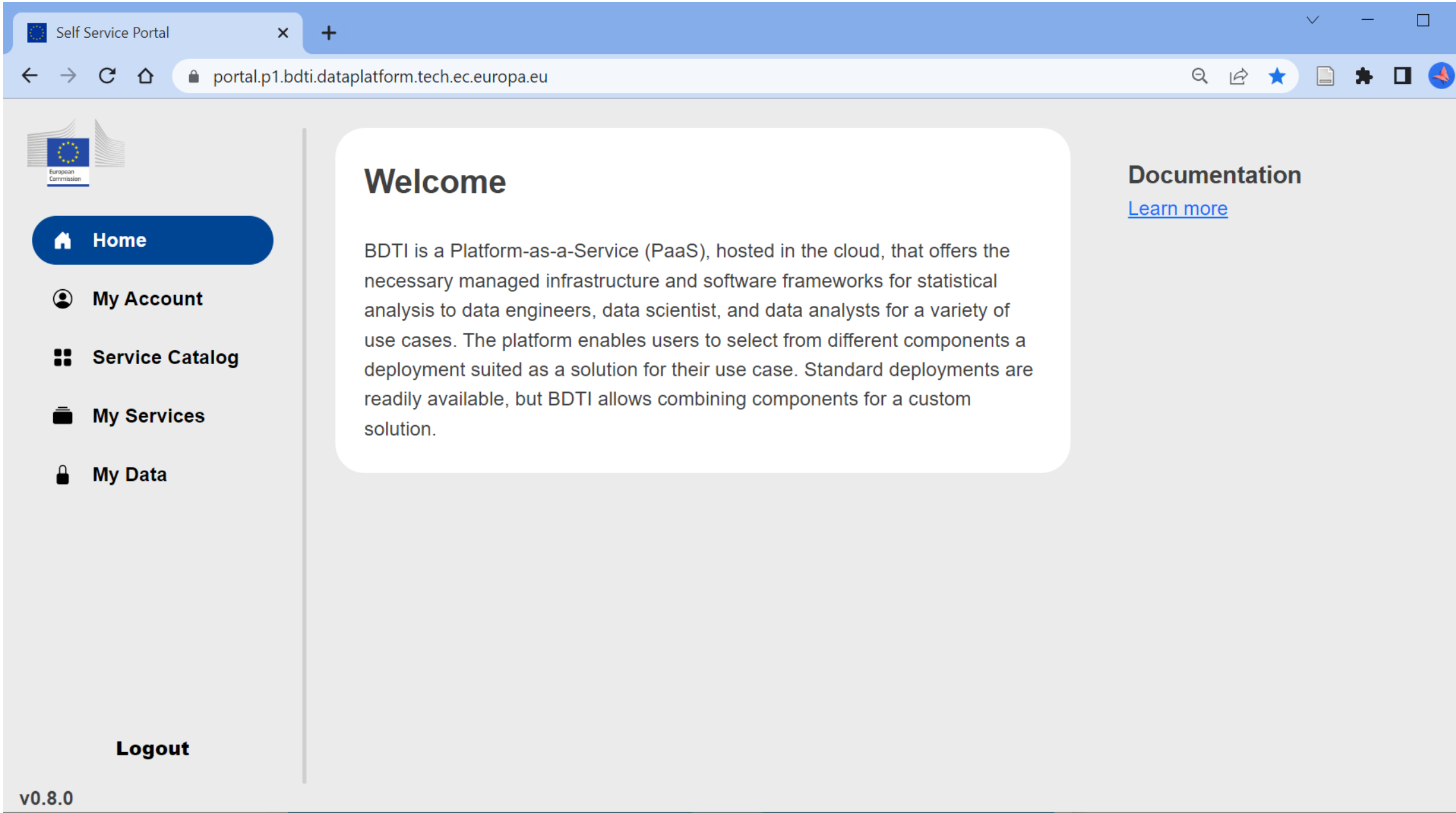
- Access and overview of the BDTI portal
- Concrete application of the BDTI

3

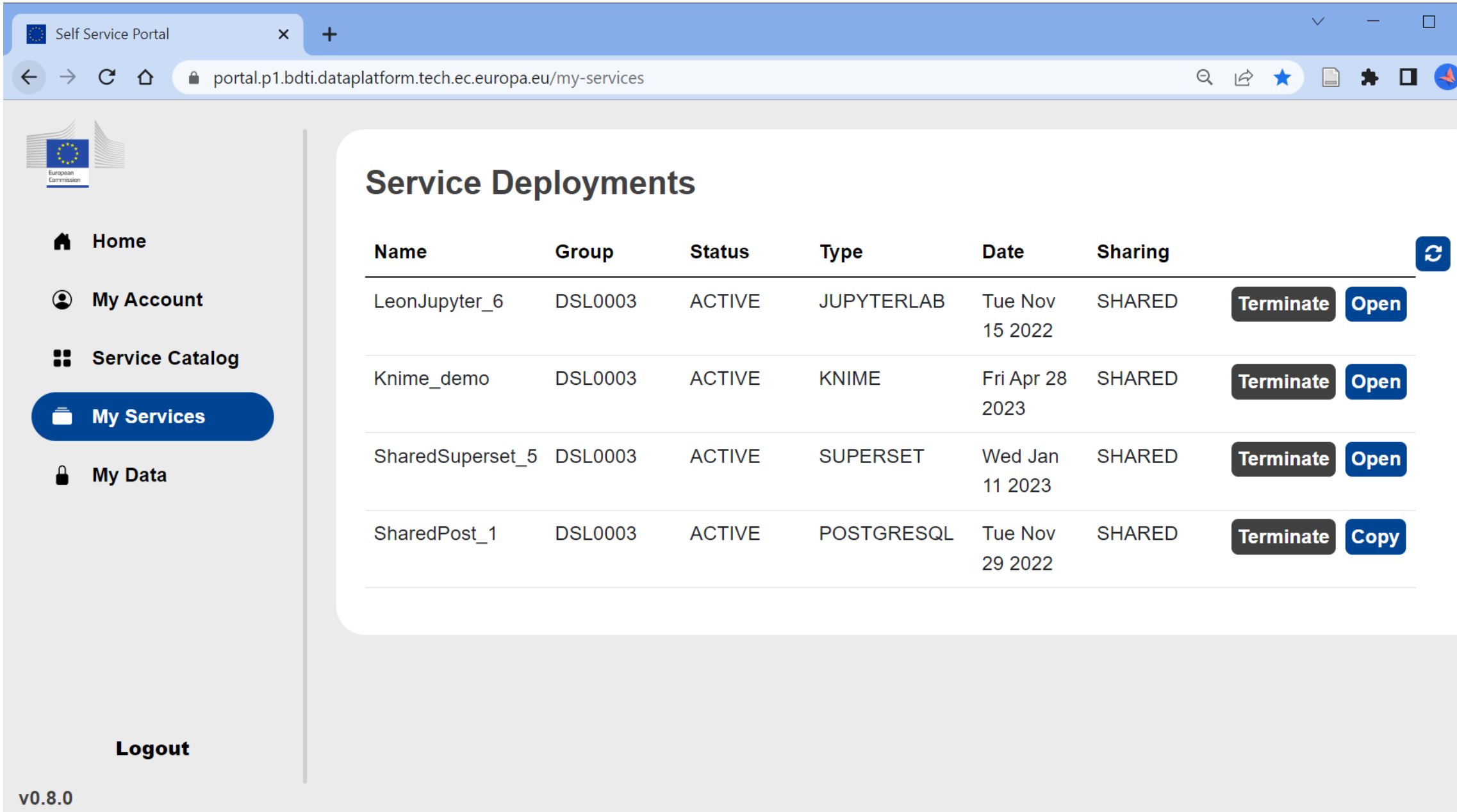
Access to BDTI portal directly from your browser (EU Login integration)



The BDTI portal

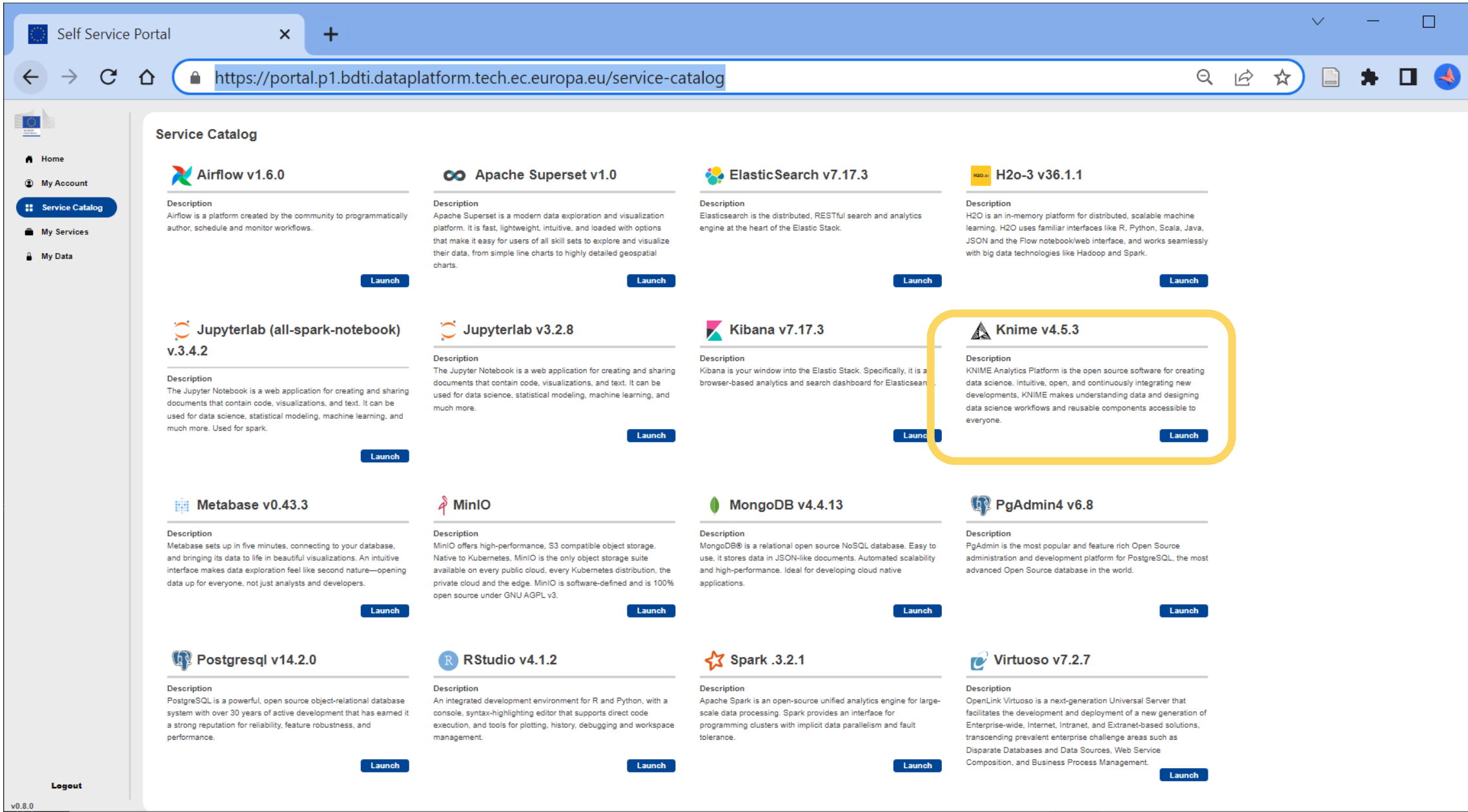


The BDTI portal: My Services



3

The BDTI portal: service catalogue



3

BDTI Demonstrator:
Towards a data-Informed Government Spending

Goal:

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 user-centered approach:

- Elena and Daniel, public servants
- Low data literacy skills
- **Problem**: high government spending in public lighting
- **Solution**: how to optimize public lighting to reduce government spending

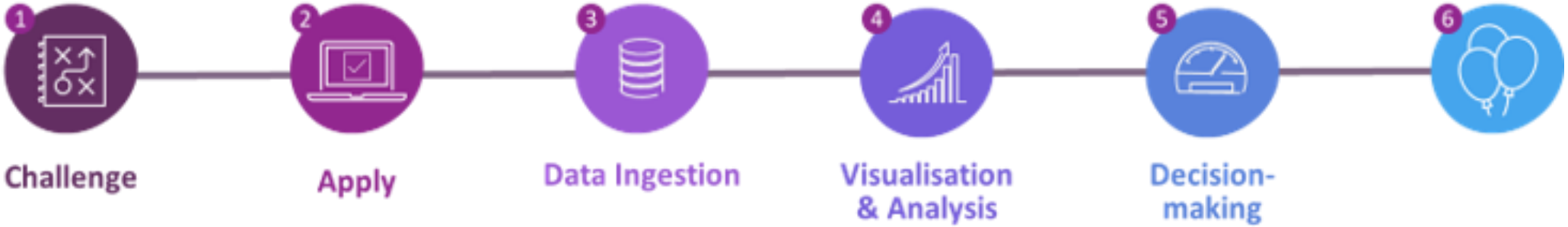
3

BDTI Demonstrator and KNIME: Data-Informed Government Spending

Elena

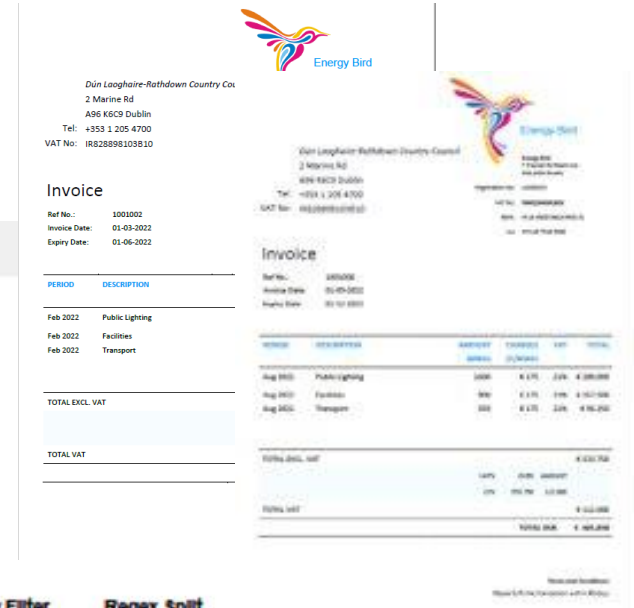
- Observes high government spending reported in the news
- Uses Optical Character Recognition (OCR) to extract relevant information from a folder of PDF invoices from the Energy Supplier.
- She then combines this output with other data (.csv, .xlsx) on her government's spending.
- She feeds the consolidated dataset to a relational database that she can access with her dashboarding service.
- Elena visualizes the enriched government spending data in a Dashboard.
- She analyses the charts and discovers that her government is spending **more on public lighting** than other comparable municipalities.

(**Services used:** KNIME, PostgreSQL, Apache Superset)



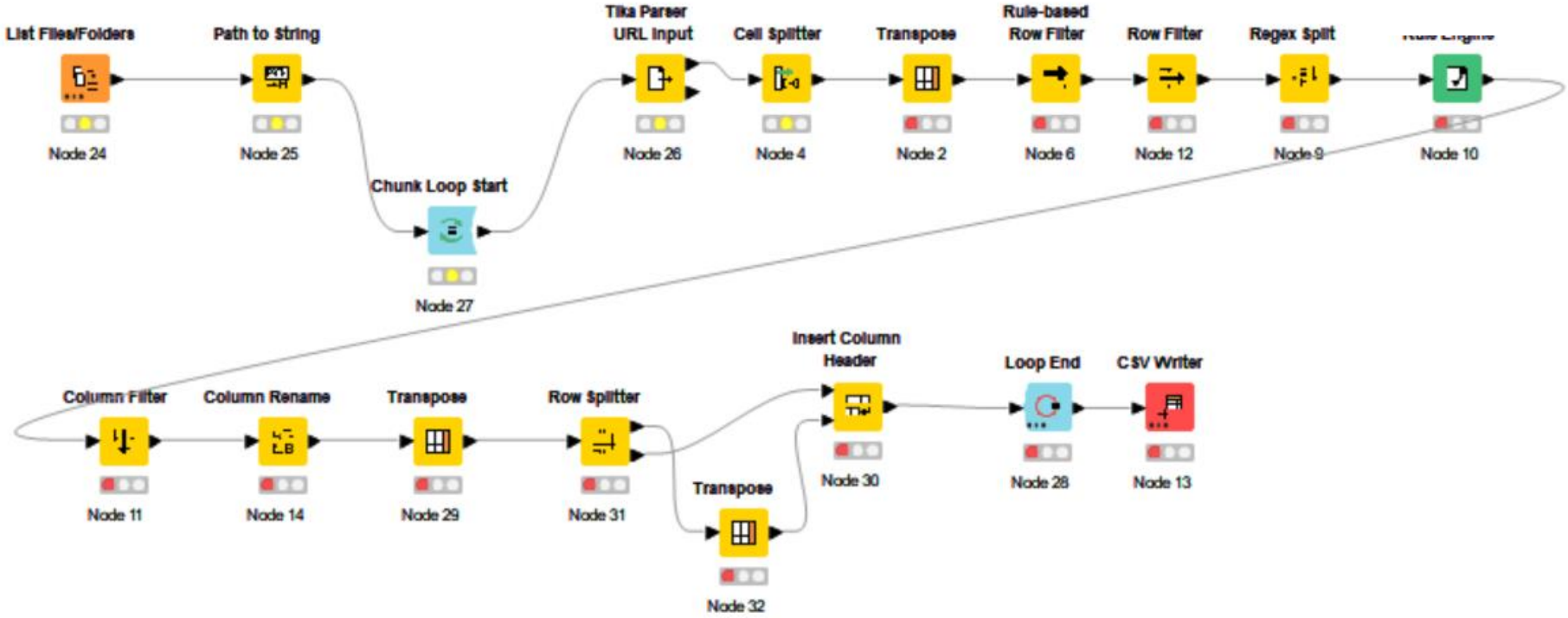
3

BDTI Demonstrator: KNIME Workflow



*3: KNIME_OCR_multiple_pdf

This Knime workflow connects to a folder containing PDF invoices. It loops over the invoices one at a time to apply OCR. When all invoices are read, the retrieved parameters are stored into a single csv file. Right-click a node and select configure to see what the node does. The node called Tike Parser URL Input is the node that performs the actual optical character recognition.



3

BDTI Demonstrator: Dashboard

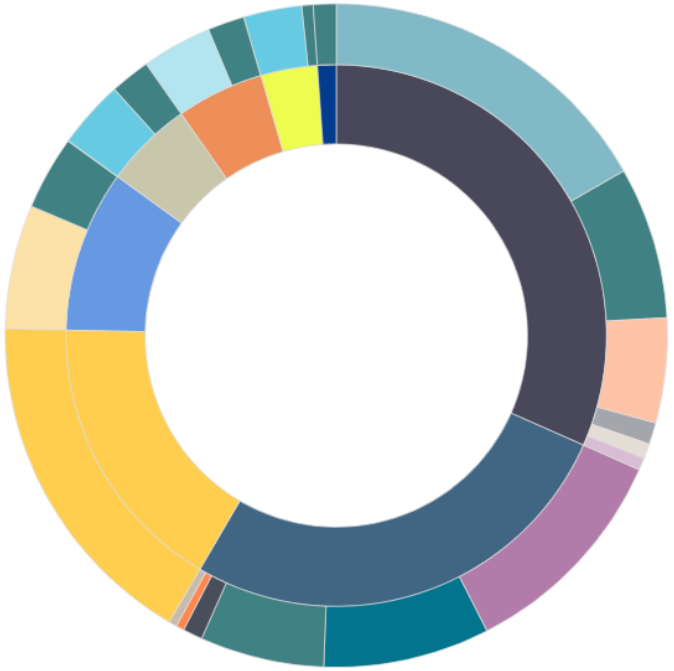
Self Service Portal | Government Spending - Dublin

superset-1673442350.p1.bdti.dataplatform.tech.ec.europa.eu/superset/dashboard/9/?native_filters_key=DekayIuHneULOi6S49E...

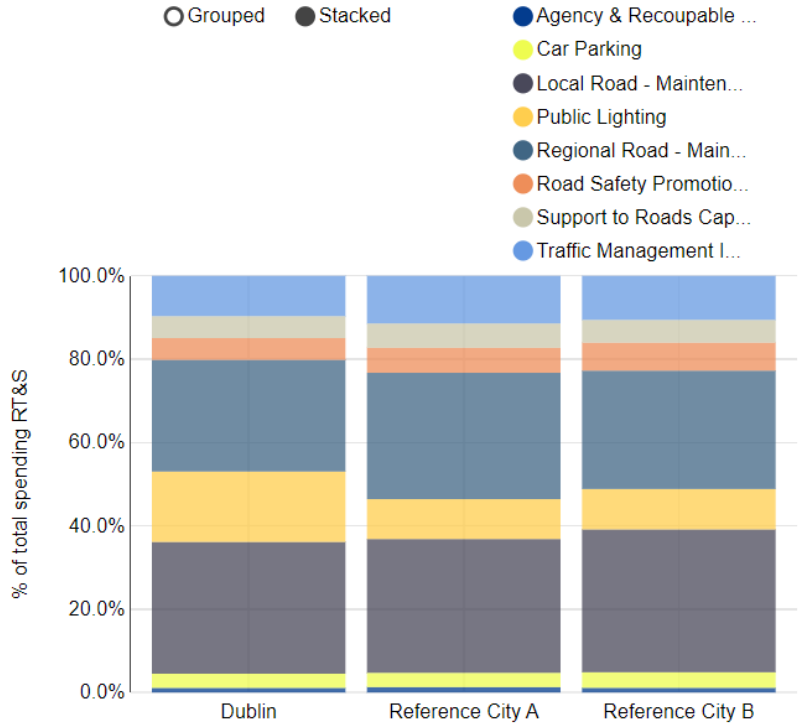
Government Spending - Dublin

2022 Overview | Road Transport & Safety | Road Transport & Safety (yearly) | Additional Information

2022 Road Transport & Safety spending by category

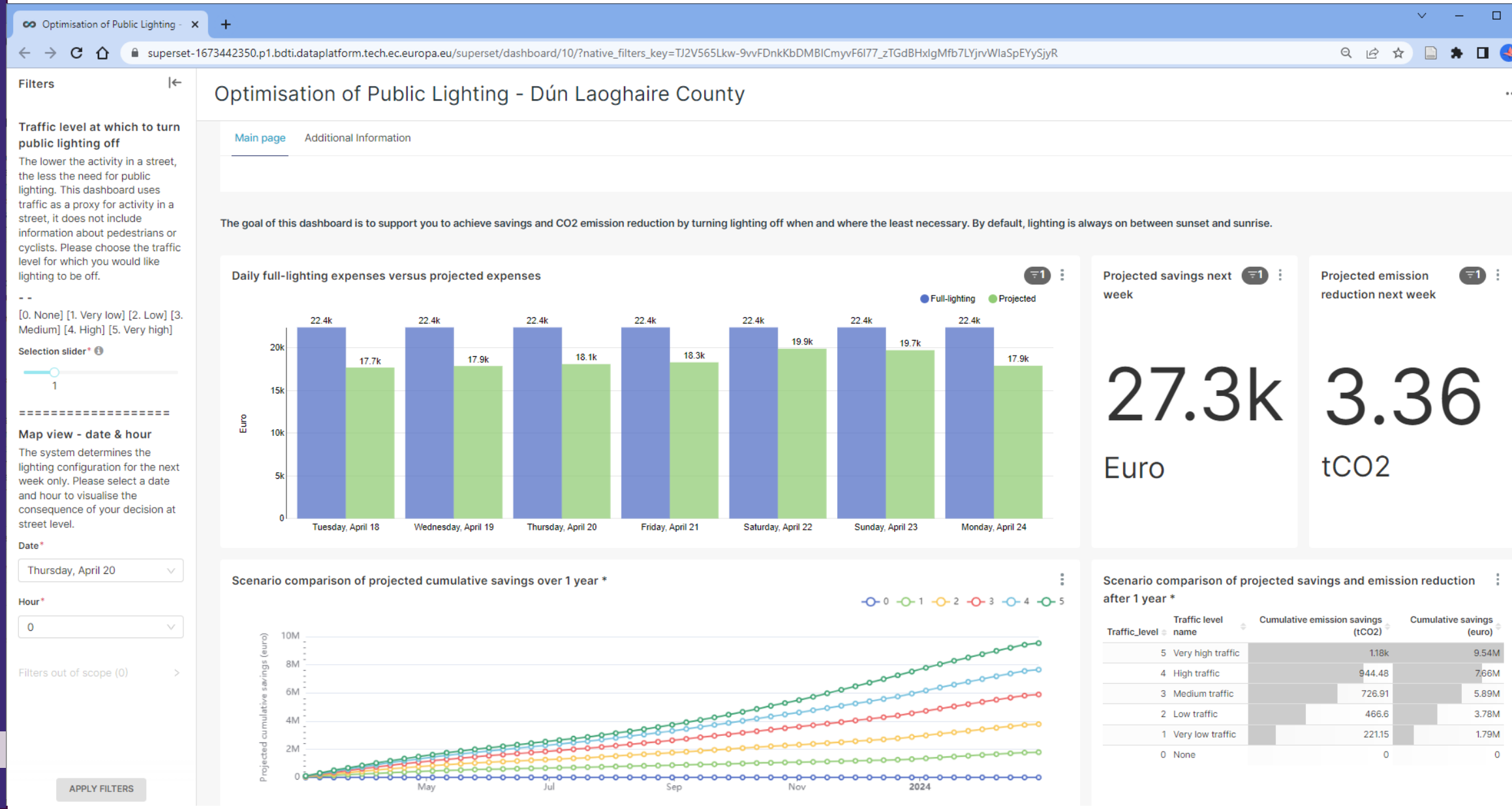


Road Transport & Safety spending - City comparison



3

BDTI Demonstrator: Dashboard



BDTI in practice

4

BDTI's community

- Developing the BDTI community and how can you help us

4

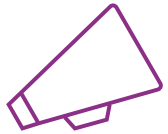
Developing BDTI's Community

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**)



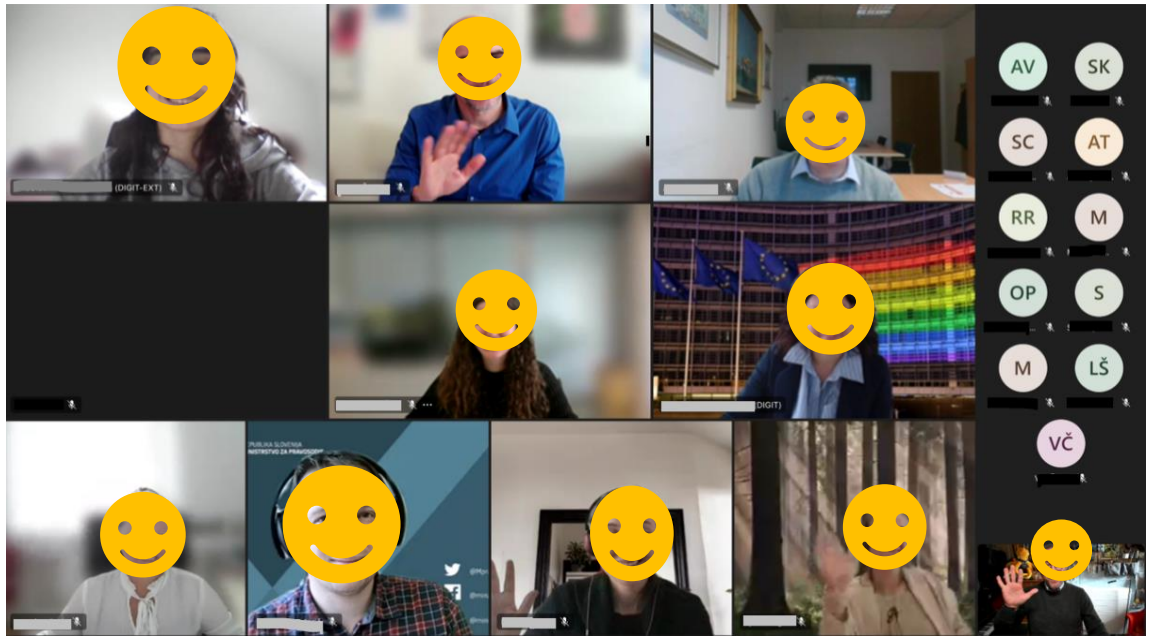
Are you working for a public administration in need of infrastructure for data analytics?

Contact us:

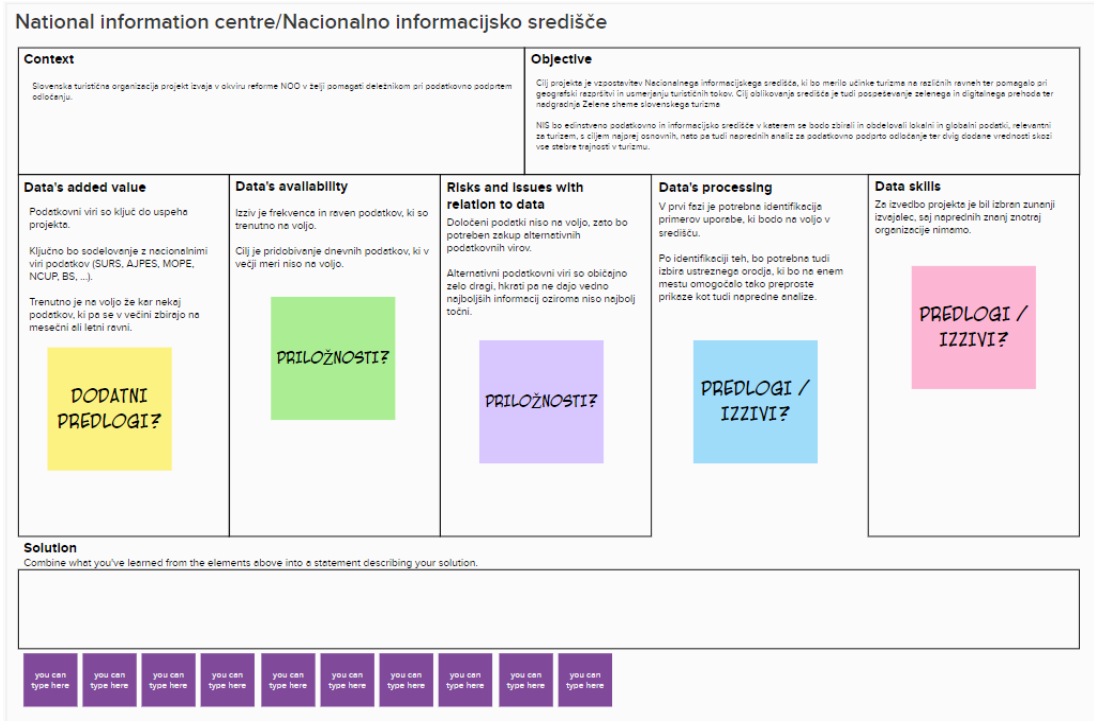
EC-BDTI-PILOTS@ec.europa.eu

4 BDTI National Information Sessions

Goal: introduce BDTI, learn about data analytics projects, develop your data analytics community!



BDTI Information Session (April 2022) in Slovenia in collaboration with the **Slovenian Ministry of Digital Transformation**



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

From Hype to Action

The BDTI Canva

Created 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

4 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 extract knowledge from the huge quantity of scientific clinical articles coming from different sources (i.e. PubMed.gov, Covid-19 related clinical articles).

GENERALITAT VALENCIANA
Conselleria de Sanitat Universal i Salut Pública

Advanced **data visualization** and **text mining** tools to help **extracting knowledge contained in the documents**, supporting clinicians and managers in their clinical practices andd 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.

eba EUROPEAN BLOOD ALLIANCE

A ready-to-use, virtual environment in which **data collected through a custom-built website** are ingested and anonymized, to be then analyzed with advanced data visualization and analytical tools. Initially, only donation data were 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 the most and about **how services can be redesigned to foster cultural activities and events.**

Città di FIRENZE

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).

4 Who used it already?

The vision: Public Procurement Data Spaces

Every year in the EU, over **250 000 public authorities spend around €2 trillion (around 13.6% of GDP)** on the purchase of services, goods, and supplies. EU directives govern procurement contracts above certain thresholds to ensure the transparency of the procedure.

The Public Procurement Data Space (PPDS) will:

- connect European databases, including TED data on public procurement, and national procurement data sets available in national portals
- facilitate access for companies and SMEs to public procurement procedures across the EU.
- increase transparency, integrity, and accountability of public spending while fighting corruption and collusion.
- generate key insights for policy-making

https://single-market-economy.ec.europa.eu/single-market/public-procurement/digital-procurement/public-procurement-data-space-ppds_en

4 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/

Developing data skills

iTalks Series:

EUROPEAN YEAR OF SKILLS

'Exploring data skills initiatives to foster public sector innovation'

Digital innovationLab

From the 14th of June to the 5th of July,
join us for six iTalks with 10 data literacy experts
to learn about data skills initiatives for the public sector

European Commission

How to apply:

Get familiar with the BDTI service on our website

Brainstorm on 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 done in 6 months, sufficient resources available (skills, team, data)

Your test environment is set up

You can start piloting and create value!

Thank you for your attention!

BDTI website

BDTI's Joinup page (subscribe ;-)

mariaclaudia.bodino@ec.europa.eu

EC-BDTI-PILOTS@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

Get to know: Lobium

Tim Werkhoven

data europa academy

Lobium/Gavagai

Winner EU Datahon, challenge 4: Europe fit for a Digital Age

Our journey

EUDATATHON'22

The teams behind the apps

Video series

Lobium/Gavagai

Tim Werkhoven, Ivan Sukhletcov, Fredrik Olsson, Anyla Pula

Finalist team for

Challenge 4

A Europe fit for the digital age

#eudatathon | op.europa.eu/eudatathon | 20 OCTOBER 2022

Lobium; Vision

1. ***Data must work harder for Public Affairs Management.***

Better quality data and smarter data analysis are key to modern Public Affairs Management.

1. ***Data must work harder to preserve Democracy***

Deep commitment to support and maintain transparency in government and fighting for democracy

1. ***Data must work harder to get to net zero***

Climate change is an emergency. We must go to net zero quickly; unlocking data and combining data sets is a key enabler

Our Journey to date - how do we make EU Open Data actionable?

Open data

Transformation

Application

European Commission

Have your say

Transparency Register

europa.eu

Open Data Portal

European Parliament

Home | Datasets | Developers' corner | Discover | Contact us | About

European Parliament's Open Data Portal

Data Science

NLP

Etc.

Stakeholder-Issue Analysis

EU Parliament Amendment Analyser

Where do we go from here?

Connect to other open (EU-) data sets

joinup

data europa eu
The official portal for European data

EUR-Lex

ELI ontology

Legislative Observatory
European Parliament

Law European data
EU Vocabularies

New use cases

- *Legislative timeline builder*
- *Stakeholder salience*
- *Issue salience*
- *Advocacy benchmarking*
- *Legislation predictor*

New users

- *EU Commission officials*
- *MEP assistants*
- *EU citizens*
- *NGOs*
- *Trade Associations*

New NLP models

- *Sentiment analysis*
- *Legislation summarisation*
- *Legislation prediction*
- *Text to speech / speech analysis*

Unlock open data from national jurisdictions

EU Open Data, what can we improve?

Users

- Better understand users and user needs
  - Citizens
  - Journalists
  - Academia / Think tanks
- Define use cases

EU Institutions

- EU Council
  - Make more data available
- EU Parliament
  - Unlock data (e.g Amendments), in XML format
  - Extend API
- EU Commission
  - Improve EU Have Your Say data: consistency in publication process and simplify consultation formats
  - Create consistency across DGs

Horizontal

- Clarify file hierarchy for better data association (Basic Act, COM Proposal, EP Procedure, Committee docs, Plenary docs, etc.
- Provide more examples and resources how to work with RDF format (knowing the tree, depth, etc.)
- Building Machine Learning models: help with data pre-processing
- Ensure consistent format across data from all institutions

Investigative Journalists: Ready for the Digital Age

Tim Werkhoven

CEO

Lobium.ai

tim@lobium.ai

Dr. Fredrik Olsson

Product Owner and Head of Data Science

Gavagai.io

fredriko@gavagai.io

Get to know: 100 Europeans

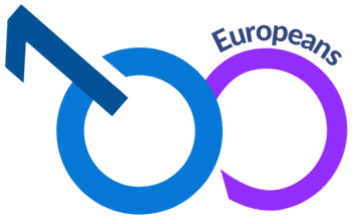
Dimitris Michailidis

data europa academy

Beyond Open Data

Empowering citizens through engaging storytelling

GreeceInFigures.com



100Europeans.org

“What do we mean by unemployment?”

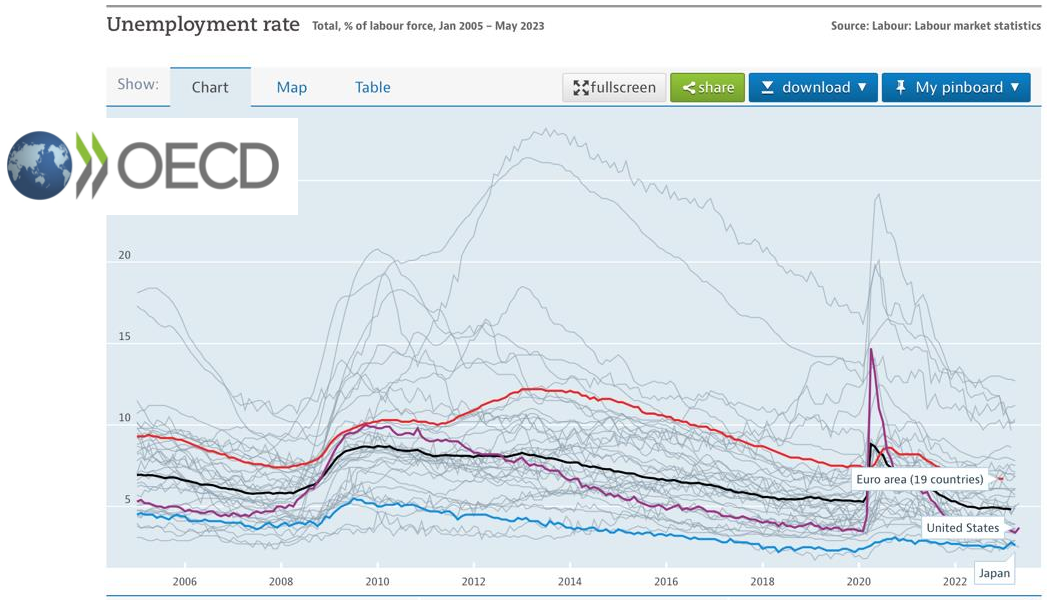
Myself, circa 2019

Unemployment

Article Talk

From Wikipedia, the free encyclopedia

Unemployment, according to the OECD (Organisation for Economic Co-operation and Development), is people above a specified age (usually 15)[2] not being in paid employment or self-employment but currently available for work during the reference period.[3]



What Is the Unemployment Rate?

The unemployment rate is the percentage of the labor force without a job. It is a lagging indicator, meaning that it generally rises or falls in the wake of changing economic conditions, rather than anticipating them. When the economy is in poor shape and jobs are scarce, the unemployment rate can be expected to rise. When the economy grows at a healthy rate and jobs are relatively plentiful, it can be expected to fall.

Unemployment

Article Talk

From Wikipedia, the free encyclopedia

Unemployment, according to the OECD (Organisation for Economic Co-operation ar
people above a specified age
in paid employment or self-en
available for work during the

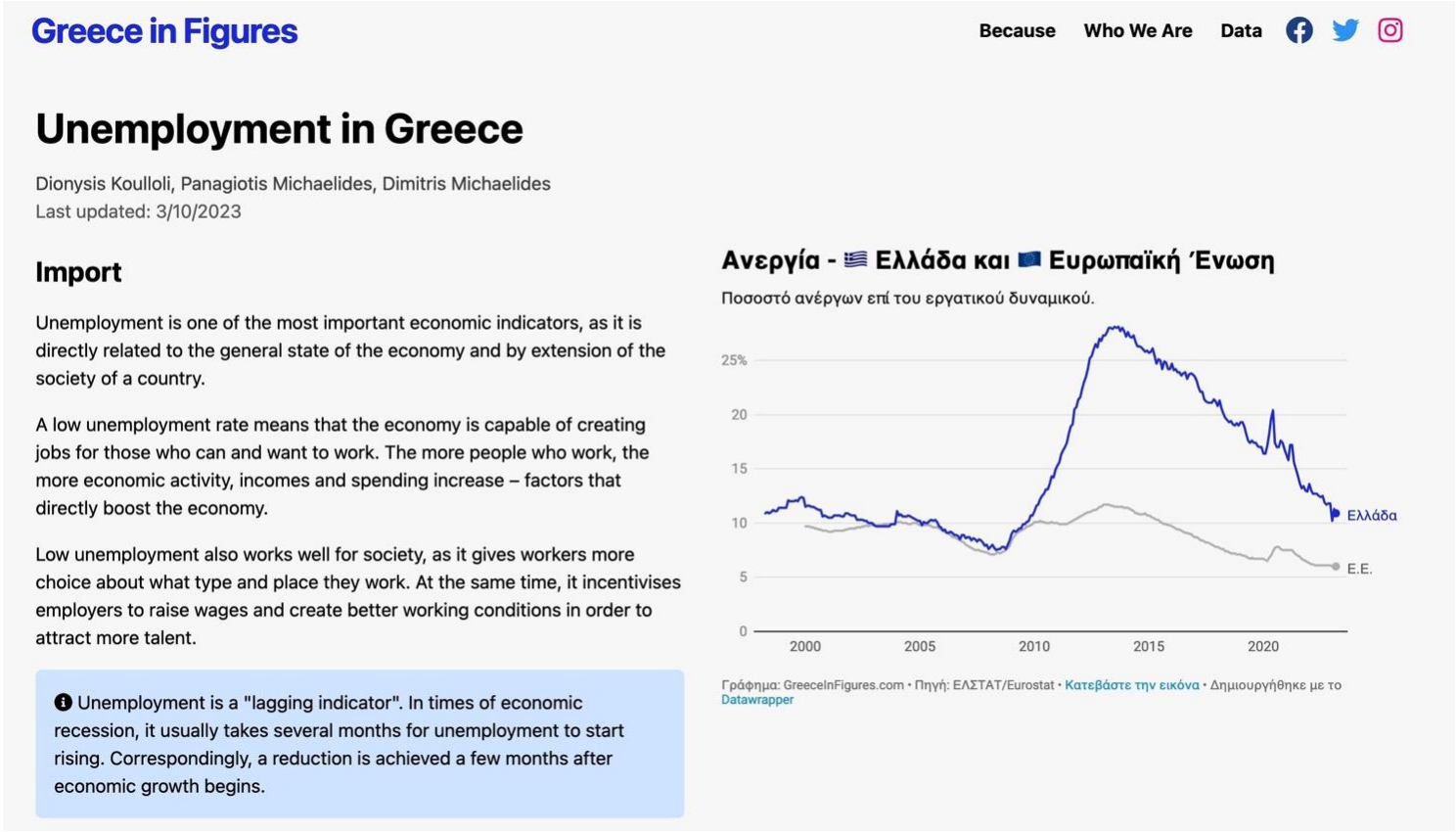
People who cannot find work but look for one
/
People who are at the age to have a job

What Is the Unemployment Rate?

The unemployment rate is the percentage of the labor force without a job. It is a lagging indicator, meaning that it generally rises or falls in the wake of changing economic conditions, rather than anticipating them. When the economy is in poor shape and jobs are scarce, the unemployment rate can be expected to rise. When the economy grows at a healthy rate and jobs are relatively plentiful, it can be expected to fall.

“What else do we not know?”

Myself & Denis & Panagiotis
Circa 2021



greeceinfigures.com

It went well.

"Simple words"
"Concise descriptions"
"Nice looking graphs"

Vebsite traffic:

UNIQUE VISITORS	TOTAL VISITS	TOTAL PAGEVIEWS	VIEWS PER VISIT
245k	278k	423k	1.52

Twitter analytics snapshots:

“A Europe fit for the digital age”

But where does the digital age take place?

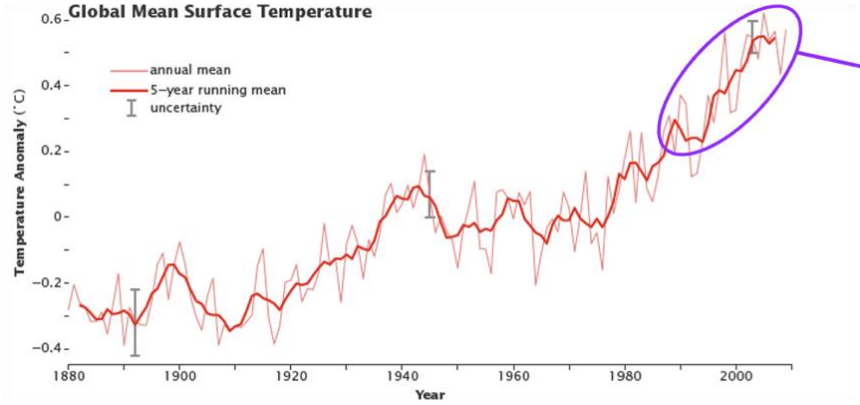
Instagram

340+ million users in Europe

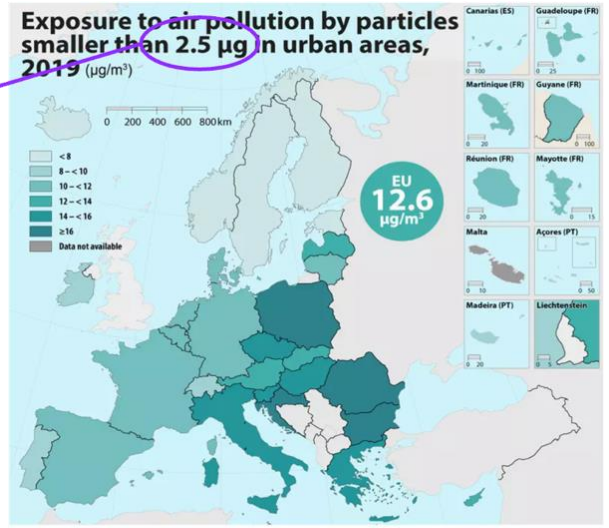
150+ million users in Europe

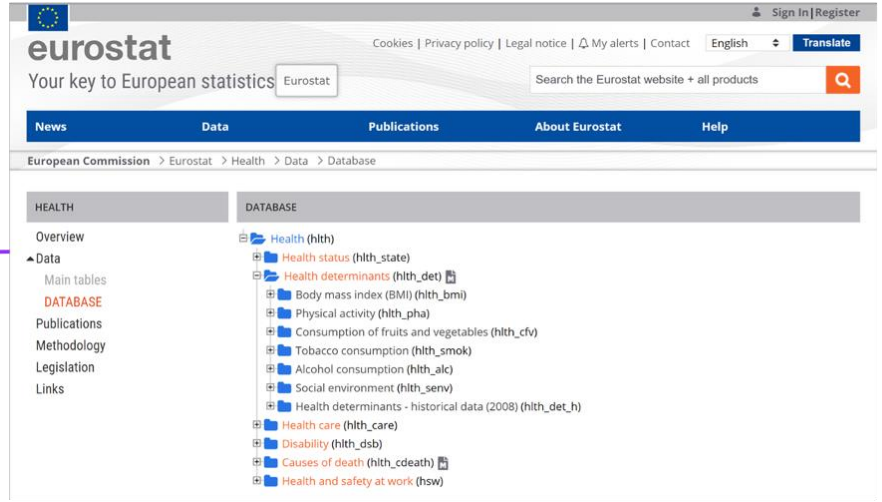
Where does communication of big challenges take place?

Climate change:



What does this mean?

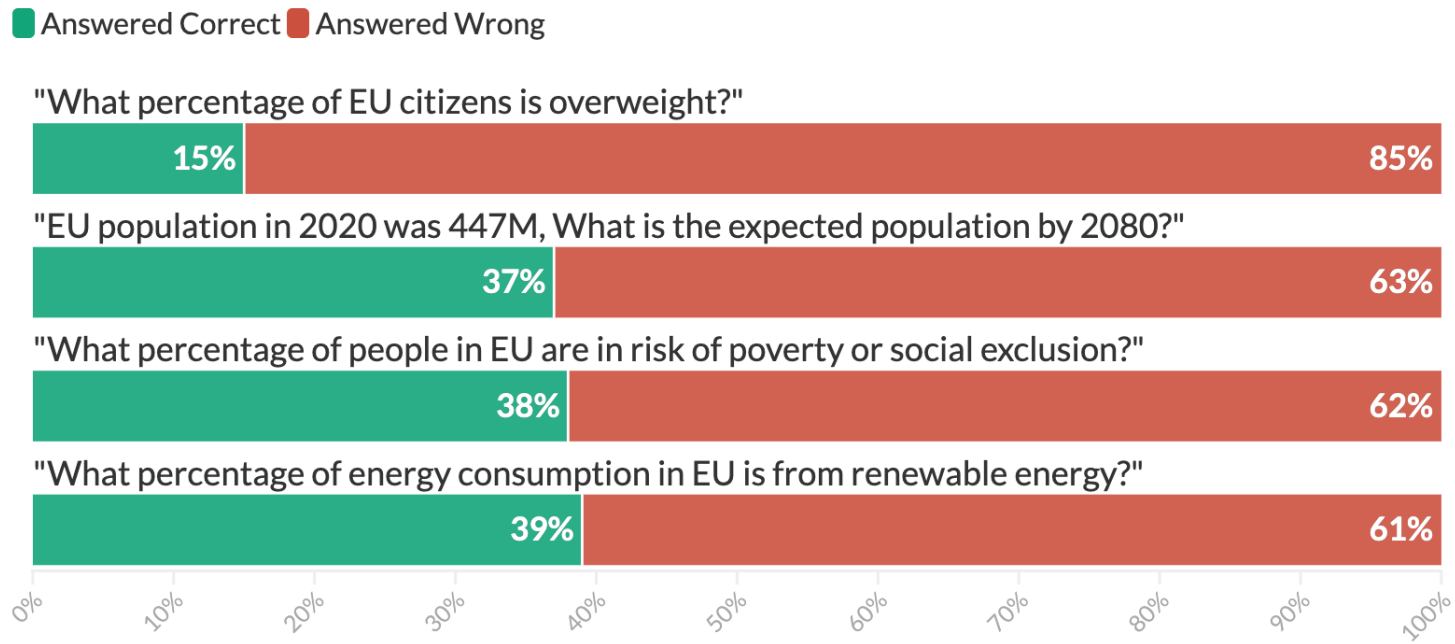




Hard to use
Too much information

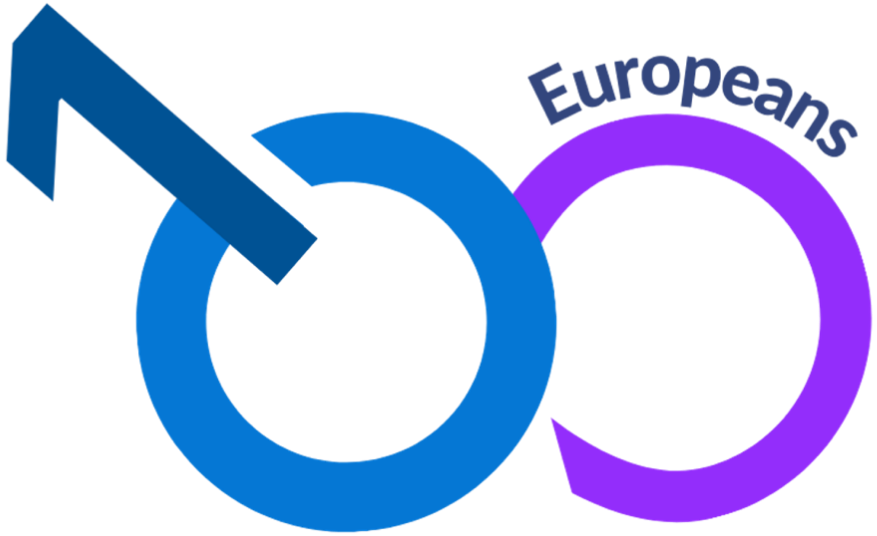
Is this communication effective?

We ran a quiz to find out:



“Perhaps we can change that.”

Myself & Denis
Circa 2022

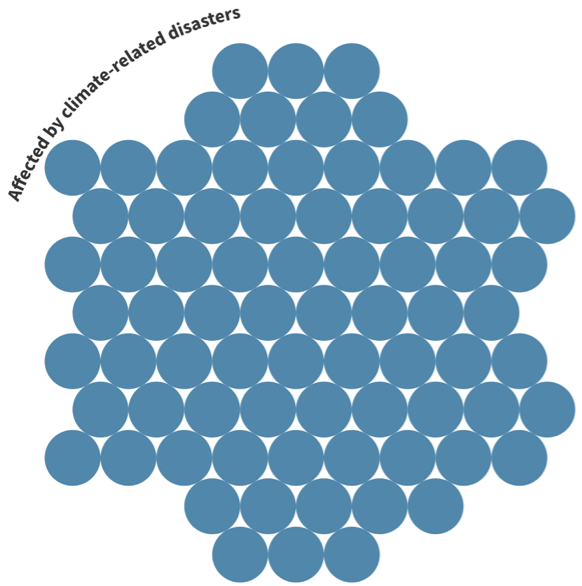


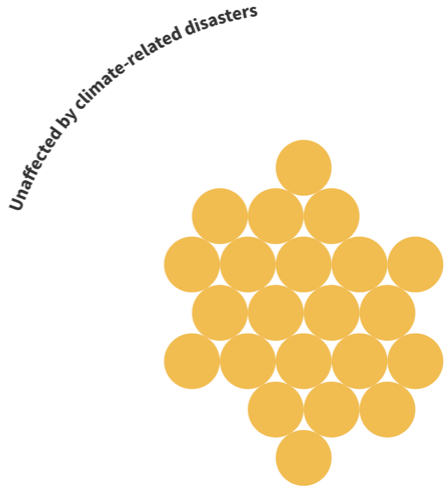
Scrollytelling Europe’s biggest challenges

100europeans.org

Let’s reimagine the EU as a community of just 100 people…

If Europe was 100 people … 78 of them would be severely affected by climate-related disasters

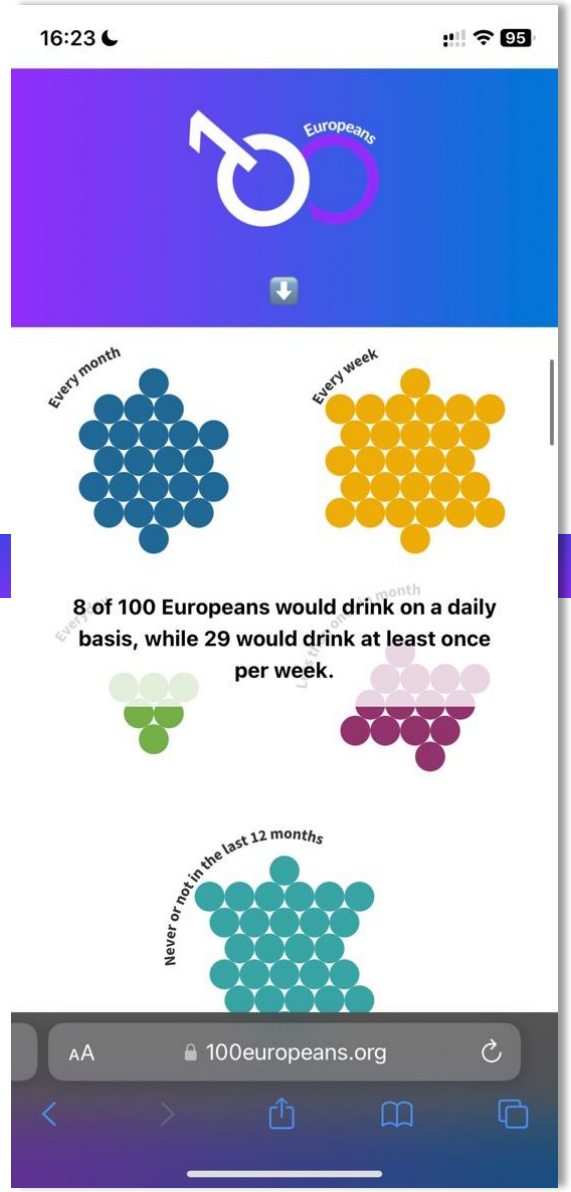




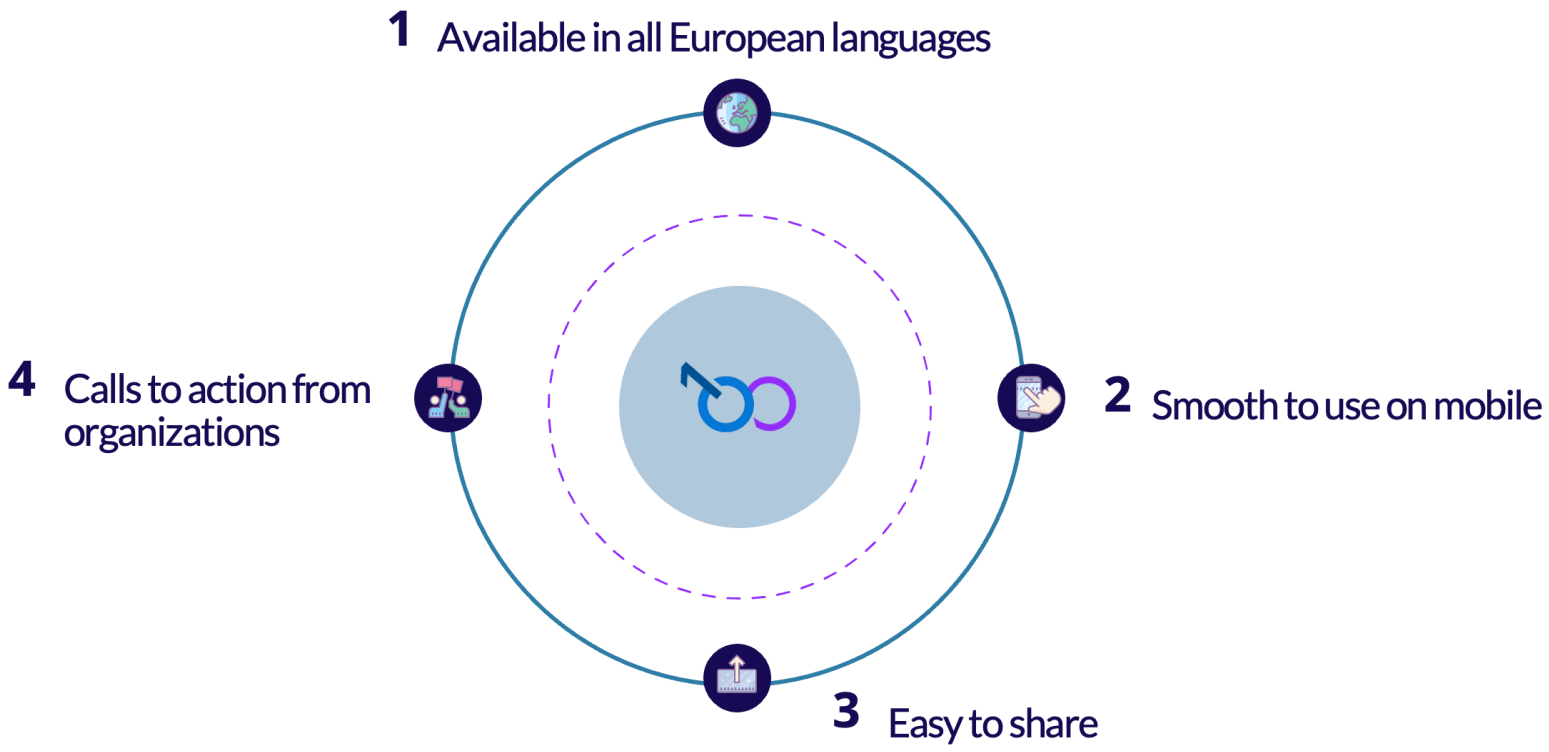
Online content is consumed via endless scrolling …

Science communication can be done similarly!

Effortless scrollytelling learning, like TikTok



But how can we do it the right way?



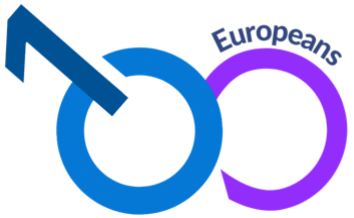
A Europe fit for the digital age...
... with open data and simple, relatable communication.*

* by telling good stories.

Thank you!

Check the projects out 👇

GreeceInFigures.com



100Europeans.org

Panel discussion

Inmaculada Farfan Velasco

Maria Claudia Bodino

Tim Werkhoven

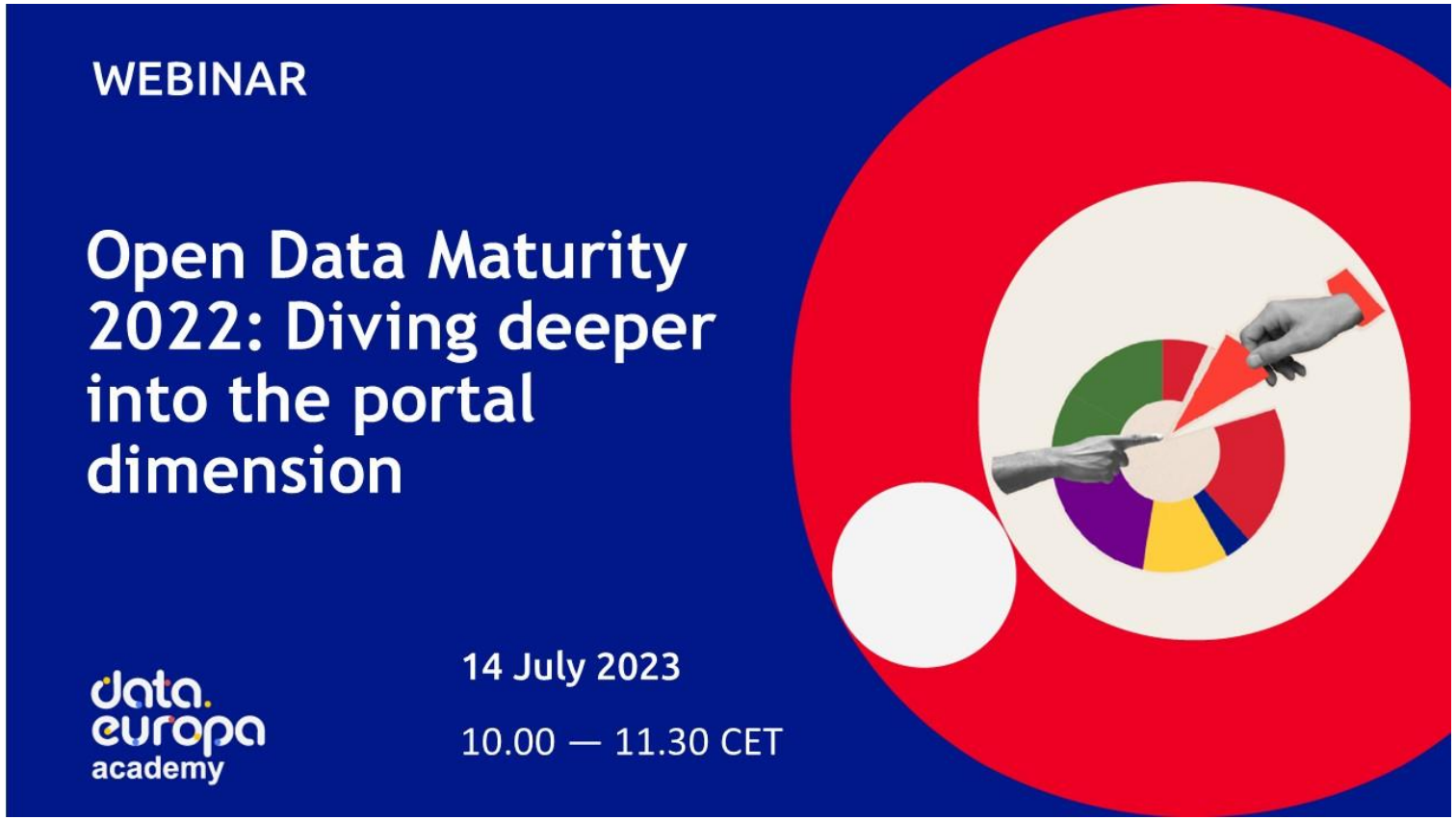
Dimitris Michailidis

data europa academy

Stay up-to-date on our
2023 activities!

data.europa academy

Join our next webinar!



data europa academy

Sign up for the newsletter: data.europa.eu/newsletter

Follow us on social media:

EU_opendata

Publications Office of the European Union

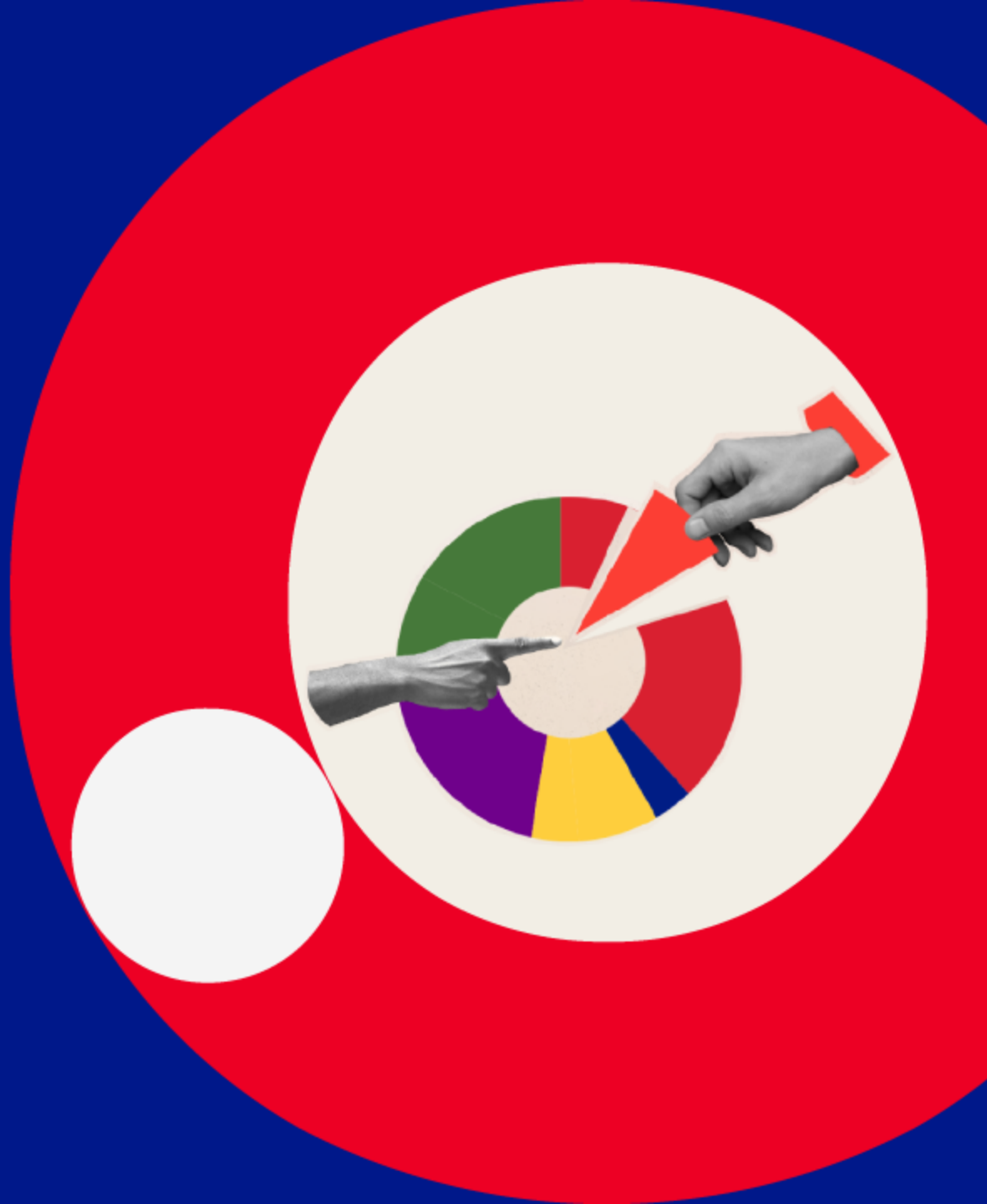
data.europa.eu

data europa academy

Please provide your feedback!

data europa academy

Thank you



data.
europa
academy