

WEBINAR

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

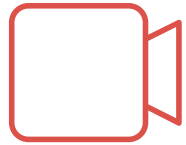
The logo for Data.europa academy is located in the bottom left corner. It features a large red circle with a smaller white circle inside it. The text "data.europa academy" is written in white lowercase letters within the white circle. The word "data" is on the top line, "europa" is on the middle line, and "academy" is on the bottom line. There are small yellow dots above the 'a' in "data" and above the 'o' in "europa".

data.
europa
academy

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.

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



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



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 Informatics

Road Map



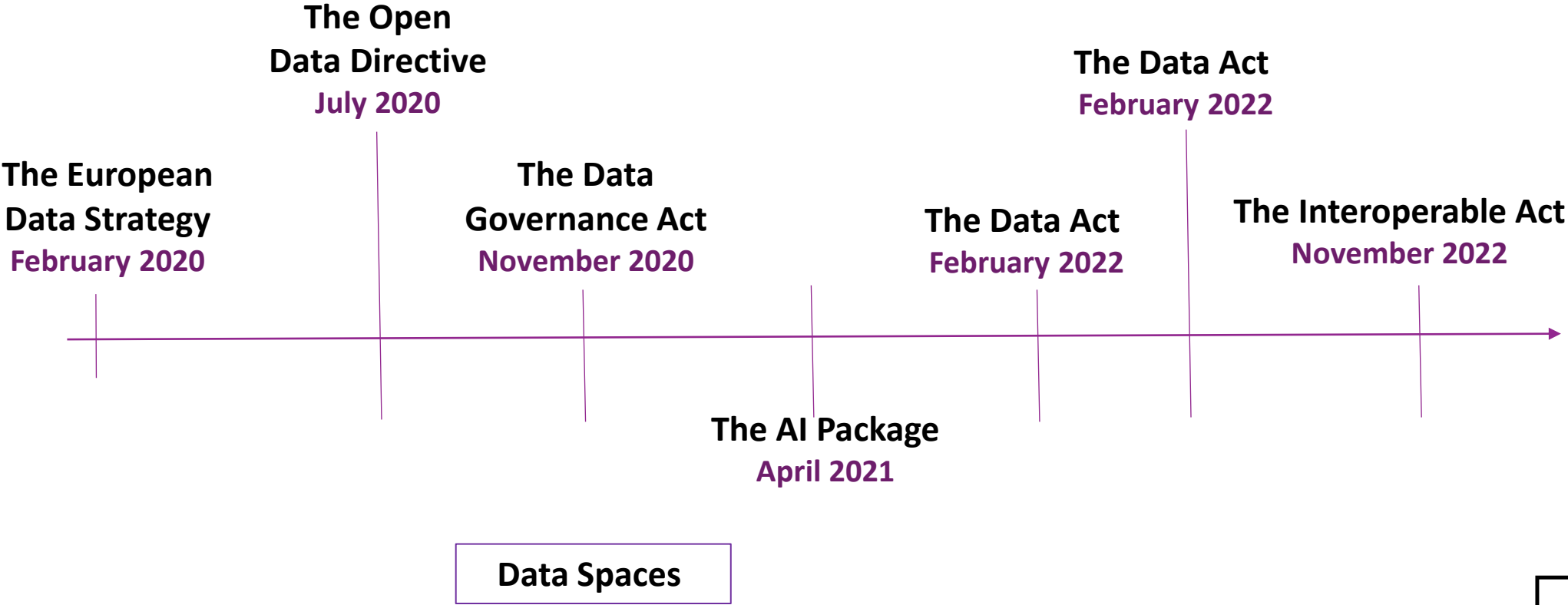
- 1** Policy context
- 2** BDTI in a nutshell
 - Its context and why use it
- 3** BDTI in practice
 - Access and overview of the BDTI portal
 - Concrete application of the BDTI
- 4** BDTI's community
 - Developing the BDTI community and how can you help us



1

Policy context

Policy timeline





2

BDTI in a nutshell

- Its context and why use it

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

2

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



What is the Big Data Test Infrastructure (BDTI) ?



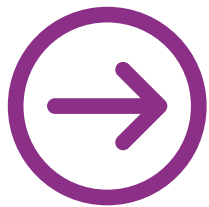
Six months free of charge service
for EU public administrations *



Ready-to-use
data analytics stack and support



Cloud platform based on
open-source tools



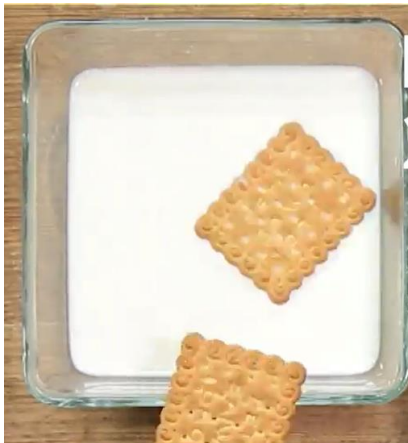
To help the public sector to **derive insights from data**
and accelerate transition towards **data- informed decision making.**

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

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

Challenges

2

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

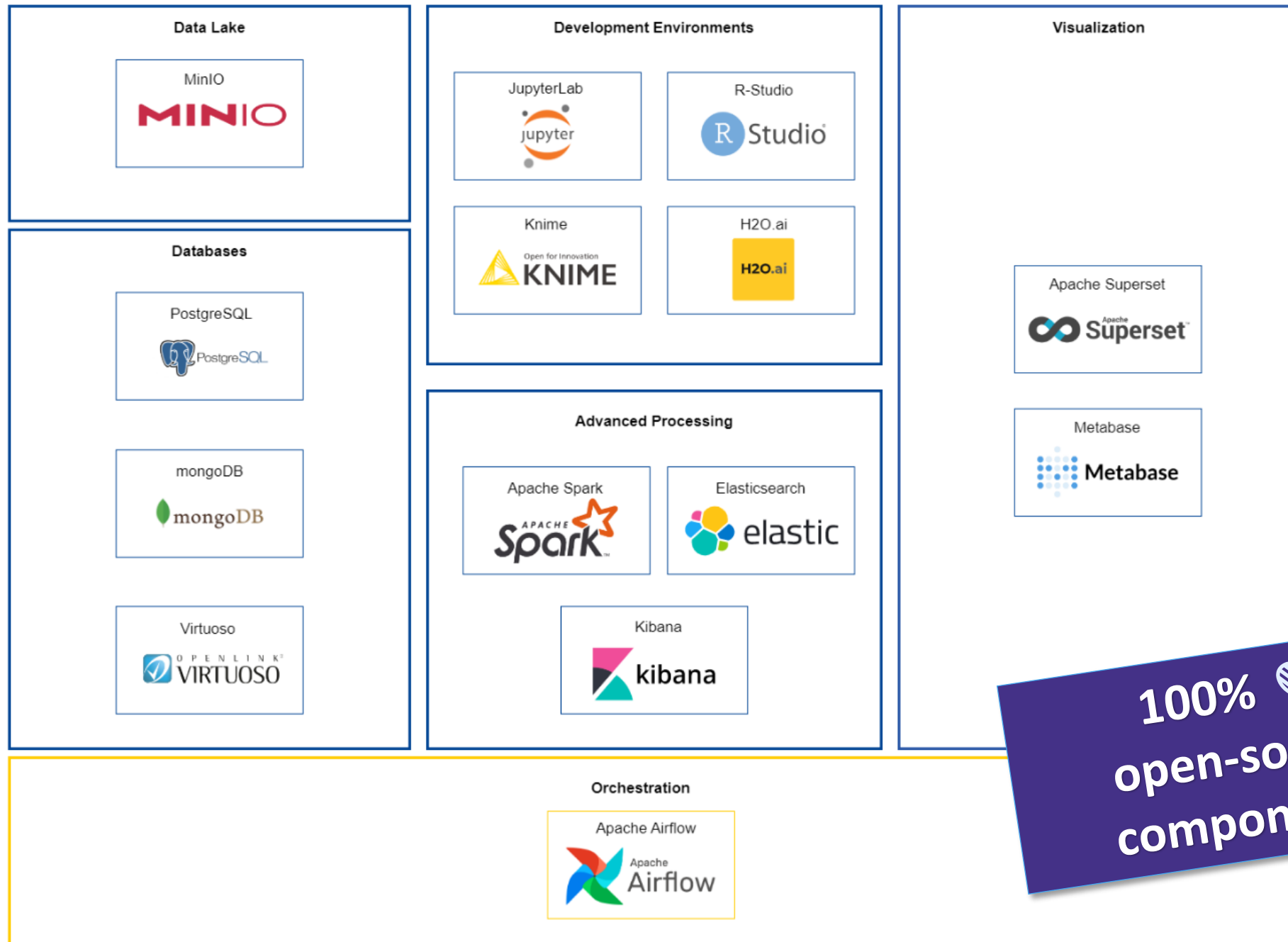
2


- 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.



BDTI's Data Analytics Stack

2



100%  open-source components

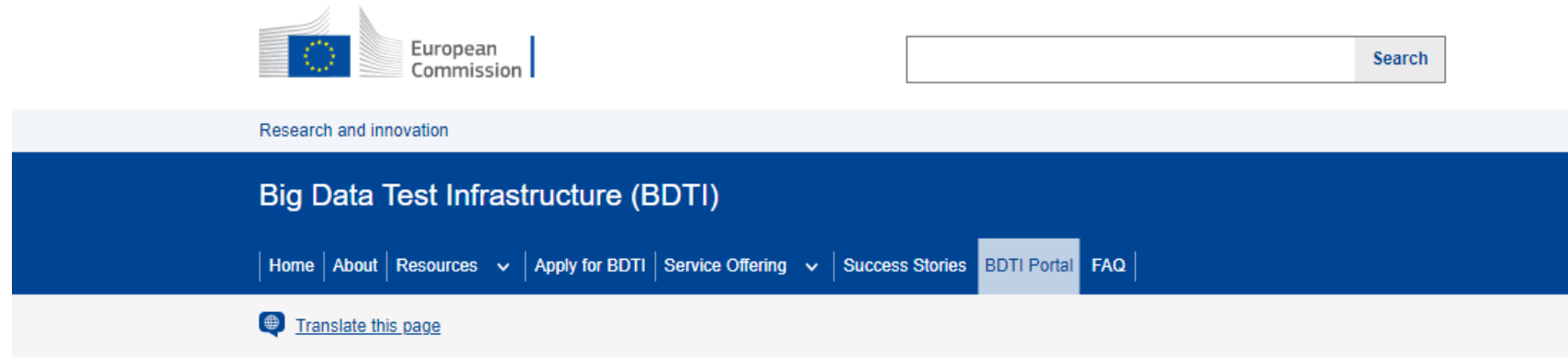


3

BDTI in practice

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

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



Home > BDTI Portal

BDTI Portal

The BDTI portal is a web application which allows users to easily deploy and manage containerized data science workloads. In this section, you can access the portal and find documentation about the portal.

Access the BDTI Portal

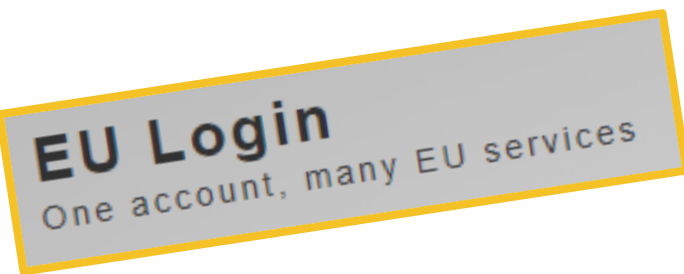
Disclaimer: The BDTI portal is only available to users who have a BDTI pilot.

The user documentation for the BDTI portal can be found [here](#).

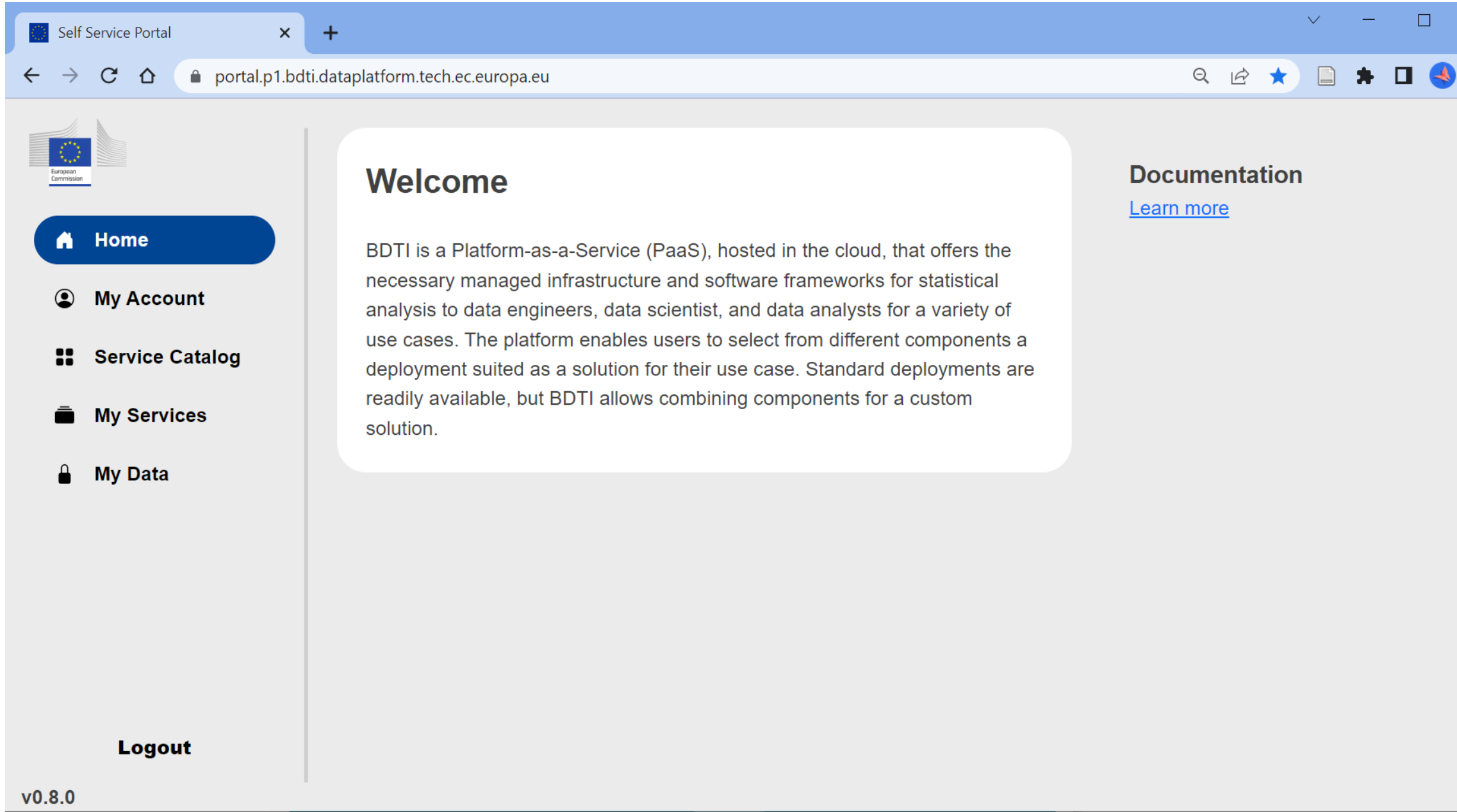
Access the BDTI Portal



For teams part of BDTI pilots



The BDTI portal



The screenshot shows a web browser window with the following elements:

- Browser Tab:** Self Service Portal
- Address Bar:** portal.p1.bdti.dataplatfrom.tech.ec.europa.eu
- Header:** European Commission logo
- Navigation Menu (Left):**
 - Home (highlighted)
 - My Account
 - Service Catalog
 - My Services
 - My Data
- Main Content Area:**

Welcome

BDTI is a Platform-as-a-Service (PaaS), hosted in the cloud, that offers the necessary managed infrastructure and software frameworks for statistical analysis to data engineers, data scientist, and data analysts for a variety of use cases. The platform enables users to select from different components a deployment suited as a solution for their use case. Standard deployments are readily available, but BDTI allows combining components for a custom solution.
- Right Sidebar:**

Documentation

[Learn more](#)
- Footer (Left):** Logout
- Footer (Bottom Left):** v0.8.0

The BDTI portal: My Services

Self Service Portal

portal.p1.bdti.dataplatform.tech.ec.europa.eu/my-services

European Commission

- Home
- My Account
- Service Catalog
- My Services**
- My Data

Logout

Service Deployments

Name	Group	Status	Type	Date	Sharing	
LeonJupyter_6	DSL0003	ACTIVE	JUPYTERLAB	Tue Nov 15 2022	SHARED	Terminate Open
Knime_demo	DSL0003	ACTIVE	KNIME	Fri Apr 28 2023	SHARED	Terminate Open
SharedSuperset_5	DSL0003	ACTIVE	SUPERSET	Wed Jan 11 2023	SHARED	Terminate Open
SharedPost_1	DSL0003	ACTIVE	POSTGRESQL	Tue Nov 29 2022	SHARED	Terminate Copy

v0.8.0

The BDTI portal: service catalogue

The screenshot displays the 'Self Service Portal' for the BDTI data platform. The browser address bar shows the URL: <https://portal.p1.bdti.dataplatform.tech.ec.europa.eu/service-catalog>. The page features a sidebar with navigation options: Home, My Account, Service Catalog (selected), My Services, and My Data. The main content area is titled 'Service Catalog' and contains a grid of 16 service cards, each with an icon, title, version, description, and a 'Launch' button. The 'Knime v4.5.3' card is highlighted with a yellow border.

Service Name	Version	Description
Airflow	v1.6.0	Airflow is a platform created by the community to programmatically author, schedule and monitor workflows.
Apache Superset	v1.0	Apache Superset is a modern data exploration and visualization platform. It is fast, lightweight, intuitive, and loaded with options that make it easy for users of all skill sets to explore and visualize their data, from simple line charts to highly detailed geospatial charts.
ElasticSearch	v7.17.3	Elasticsearch is the distributed, RESTful search and analytics engine at the heart of the Elastic Stack.
H2o-3	v36.1.1	H2O is an in-memory platform for distributed, scalable machine learning. H2O uses familiar interfaces like R, Python, Scala, Java, JSON and the Flow notebook/web interface, and works seamlessly with big data technologies like Hadoop and Spark.
Jupyterlab (all-spark-notebook)	v.3.4.2	The Jupyter Notebook is a web application for creating and sharing documents that contain code, visualizations, and text. It can be used for data science, statistical modeling, machine learning, and much more. Used for spark.
Jupyterlab	v3.2.8	The Jupyter Notebook is a web application for creating and sharing documents that contain code, visualizations, and text. It can be used for data science, statistical modeling, machine learning, and much more.
Kibana	v7.17.3	Kibana is your window into the Elastic Stack. Specifically, it is a browser-based analytics and search dashboard for Elasticsearch.
Knime	v4.5.3	KNIME Analytics Platform is the open source software for creating data science. Intuitive, open, and continuously integrating new developments, KNIME makes understanding data and designing data science workflows and reusable components accessible to everyone.
Metabase	v0.43.3	Metabase sets up in five minutes, connecting to your database, and bringing its data to life in beautiful visualizations. An intuitive interface makes data exploration feel like second nature—opening data up for everyone, not just analysts and developers.
MinIO		MinIO offers high-performance, S3 compatible object storage. Native to Kubernetes, MinIO is the only object storage suite available on every public cloud, every Kubernetes distribution, the private cloud and the edge. MinIO is software-defined and is 100% open source under GNU AGPL v3.
MongoDB	v4.4.13	MongoDB® is a relational open source NoSQL database. Easy to use, it stores data in JSON-like documents. Automated scalability and high-performance. Ideal for developing cloud native applications.
PgAdmin4	v6.8	PgAdmin is the most popular and feature rich Open Source administration and development platform for PostgreSQL, the most advanced Open Source database in the world.
Postgresql	v14.2.0	PostgreSQL is a powerful, open source object-relational database system with over 30 years of active development that has earned it a strong reputation for reliability, feature robustness, and performance.
RStudio	v4.1.2	An integrated development environment for R and Python, with a console, syntax-highlighting editor that supports direct code execution, and tools for plotting, history, debugging and workspace management.
Spark	.3.2.1	Apache Spark is an open-source unified analytics engine for large-scale data processing. Spark provides an interface for programming clusters with implicit data parallelism and fault tolerance.
Virtuoso	v7.2.7	OpenLink Virtuoso is a next-generation Universal Server that facilitates the development and deployment of a new generation of Enterprise-wide, Internet, Intranet, and Extranet-based solutions, transcending prevalent enterprise challenge areas such as Disparate Databases and Data Sources, Web Service Composition, and Business Process Management.

Logout

v0.8.0

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

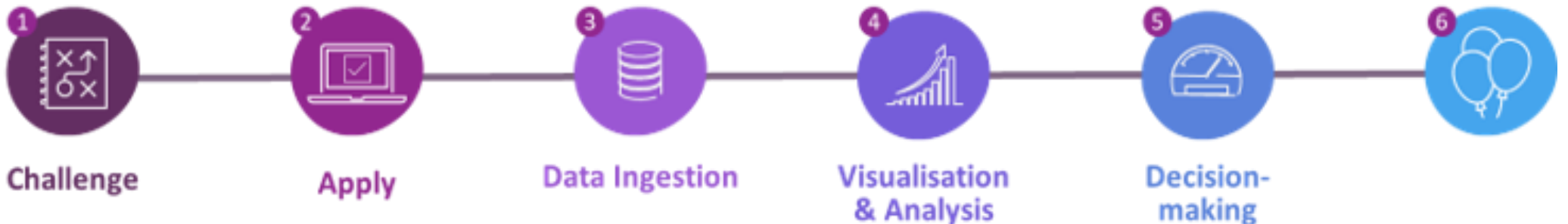
BDTI Demonstrator and KNIME: Data-Informed Government Spending



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.

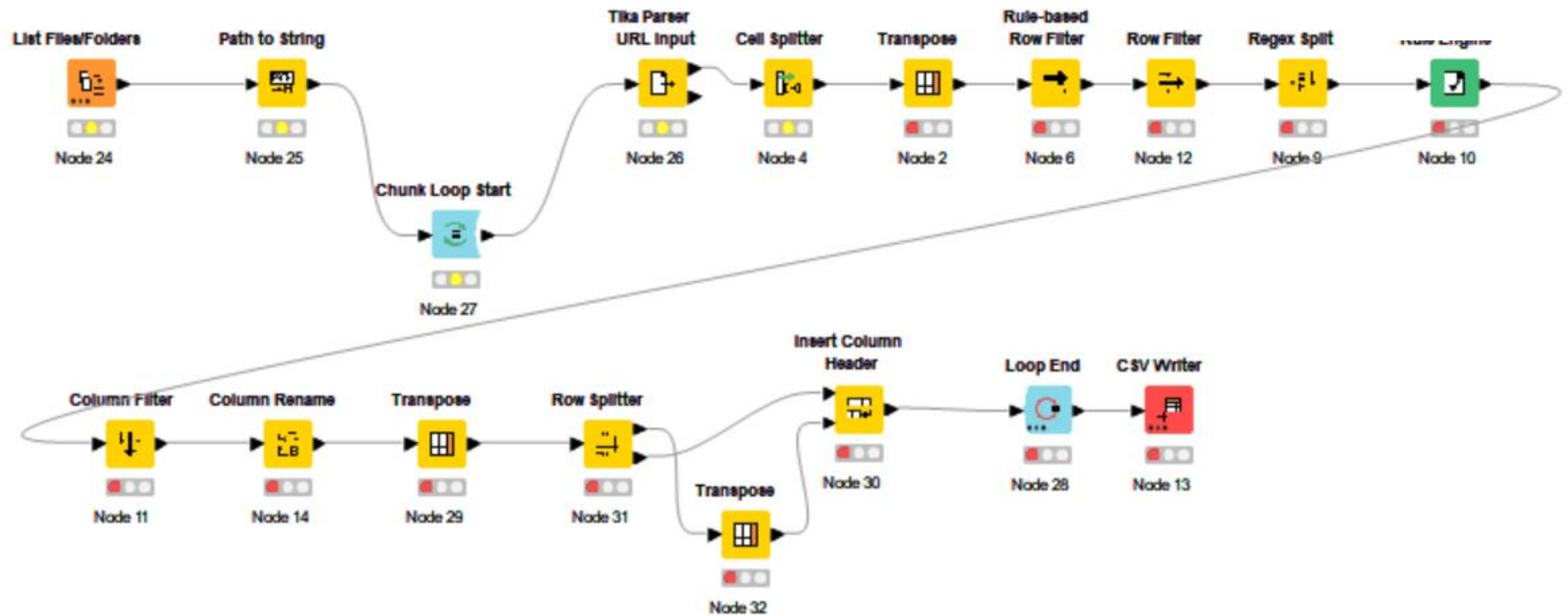
(Services used: KNIME, PostgreSQL, Apache Superset)



BDTI Demonstrator: KNIME Workflow

*3: KNIME_OCR_multiple_pdf X

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 Tika Parser URL Input is the node that performs the actual optical character recognition.



Dún Looghair-Rathdown Country Co.
2 Marine Rd
A96 K6C9 Dublin
Tel: +353 1 205 4700
VAT No: IR82898105810

Invoice

Ref No.: 1003002
Invoice Date: 01-03-2022
Expiry Date: 01-06-2022

PERIOD	DESCRIPTION
Feb 2022	Public Lighting
Feb 2022	Facilities
Feb 2022	Transport

Invoice

Ref No.: 080008
Invoice Date: 01-03-2022
Expiry Date: 01-06-2022

MONTH	DESCRIPTION	QUANTITY	UNIT PRICE	TOTAL
Aug 2021	Public Lighting	2000	€ 2.75	€ 5,500.00
Aug 2021	Facilities	500	€ 2.75	€ 1,375.00
Aug 2021	Transport	500	€ 2.75	€ 1,375.00
TOTAL EXCL. VAT				€ 8,250.00
TOTAL VAT				€ 1,650.00
TOTAL INCL. VAT				€ 9,900.00

Printed on: 08/03/2022
Printed by: Admin User

BDTI Demonstrator: Dashboard

Self Service Portal

Government Spending - Dublin

superset-1673442350.p1.bdti.dataplatform.tech.ec.europa.eu/superset/dashboard/9/?native_filters_key=DekayluHneULOi6S49E...

Government Spending - Dublin

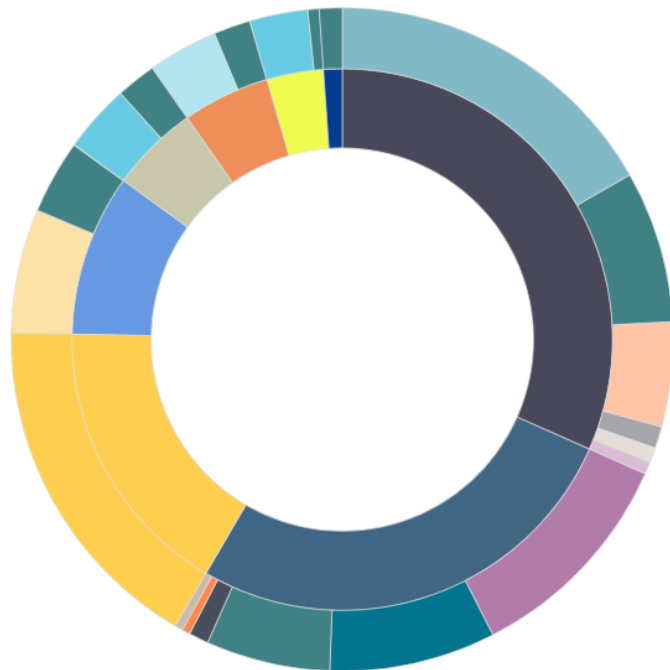
2022 Overview

Road Transport & Safety

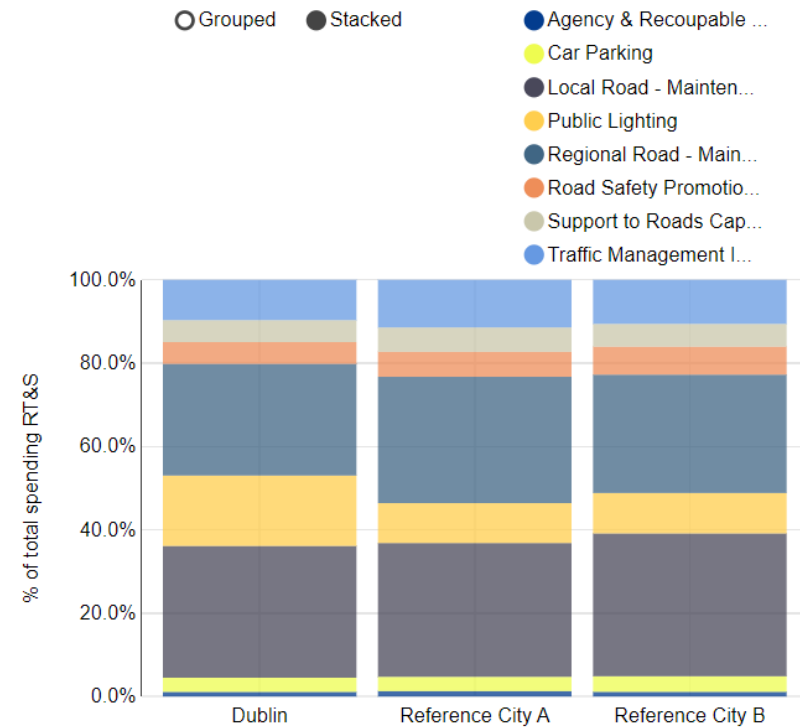
Road Transport & Safety (yearly)

Additional Information

2022 Road Transport & Safety spending by category



Road Transport & Safety spending - City comparison



BDTI Demonstrator: Dashboard

Optimisation of Public Lighting - x
+
superset-1673442350.p1.bdti.dataplatform.tech.ec.europa.eu/superset/dashboard/10/?native_filters_key=TJ2V565Lkw-9vvFDnKbDMBICmyvF6I77_zTGdBHxlgMfb7LYjrvWlaSpEYySjyR

Filters

Traffic level at which to turn public lighting off

The lower the activity in a street, the less the need for public lighting. This dashboard uses traffic as a proxy for activity in a street, it does not include information about pedestrians or cyclists. Please choose the traffic level for which you would like lighting to be off.

--

[0. None] [1. Very low] [2. Low] [3. Medium] [4. High] [5. Very high]

Selection slider ?

1

Map view - date & hour

The system determines the lighting configuration for the next week only. Please select a date and hour to visualise the consequence of your decision at street level.

Date*

Thursday, April 20

Hour*

0

Filters out of scope (0)

APPLY FILTERS

Optimisation of Public Lighting - Dún Laoghaire County

[Main page](#)
Additional Information

The goal of this dashboard is to support you to achieve savings and CO2 emission reduction by turning lighting off when and where the least necessary. By default, lighting is always on between sunset and sunrise.

Daily full-lighting expenses versus projected expenses

Date	Full-lighting (Euro)	Projected (Euro)
Tuesday, April 18	22.4k	17.7k
Wednesday, April 19	22.4k	17.9k
Thursday, April 20	22.4k	18.1k
Friday, April 21	22.4k	18.3k
Saturday, April 22	22.4k	19.9k
Sunday, April 23	22.4k	19.7k
Monday, April 24	22.4k	17.9k

Projected savings next week

27.3k

Euro

Projected emission reduction next week

3.36

tCO2

Scenario comparison of projected cumulative savings over 1 year *

Scenario comparison of projected savings and emission reduction after 1 year *

Traffic_Level	Traffic level name	Cumulative emission savings (tCO2)	Cumulative savings (euro)
5	Very high traffic	1.18k	9.54M
4	High traffic	944.48	7.66M
3	Medium traffic	726.91	5.89M
2	Low traffic	466.6	3.78M
1	Very low traffic	221.15	1.79M
0	None	0	0



4

BDTI's community

- Developing the BDTI community and how can you help us

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 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

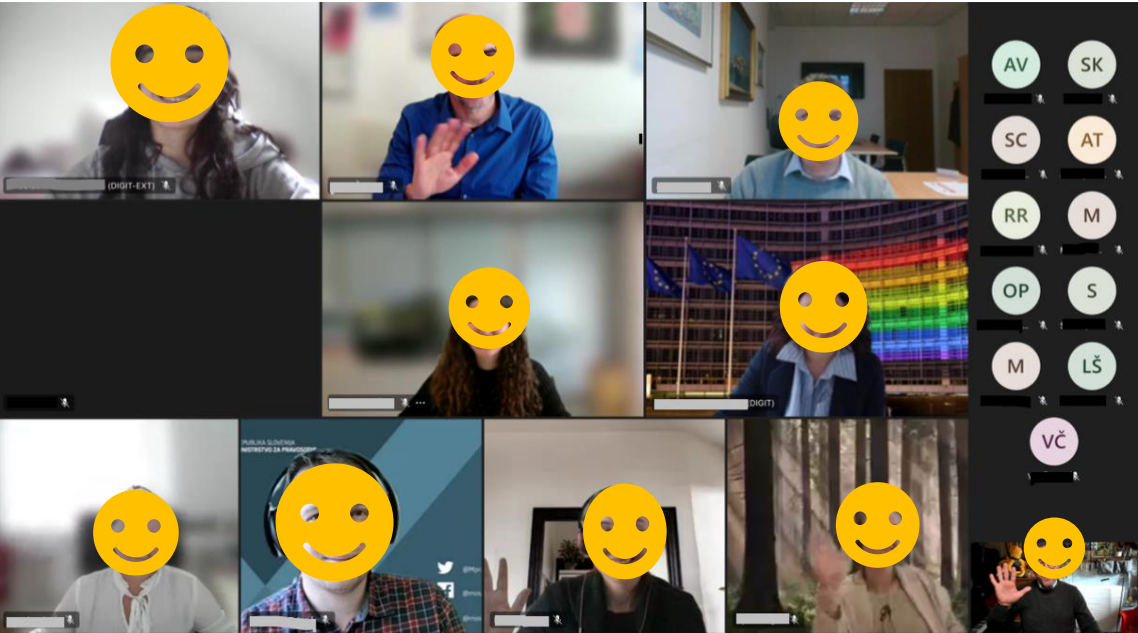


4

BDTI National Information Sessions



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

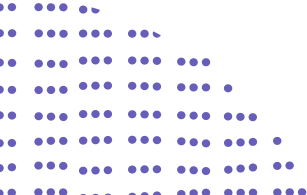


BDTI Information Session (April 2022) in Slovenia in collaboration with the Slovenian Ministry of Digital Transformation

National information centre/Nacionalno informacijsko središče

Context Slovenska turistična organizacija projekt izvede v okviru reforme NOD v želji pomagati deležnikom pri podetovno podpirnem odločitju.		Objective Cilj projekta je vzpostavitev Nacionalnega informacijskega središča, ki bo merilo učinke turizma na različnih ravneh ter pomagal pri geografski razširitvi in usmerjanju turističnih tokov. Cilj oblikovanja središča je tudi opozrevanje zelenega in digitalnega prehoda ter nadgradnje Zeleno sname slovenskega turizma. NIŠ bo edinstveno podatkovno in informacijsko središče v katerem se bodo zbirali in obdelovali lokalni in globalni podatki, relevantni za turizem, s ciljem najprej osnovnih, nato se tudi naprednih analiz za podetovno podortu odločitje ter ovig sodene vrednosti oziro ose stebre trajnosti v turizmu.		
Data's added value Podetkovni viri so ključ do uspeha projekta. Ključno bo sodelovanje z necloninimi viri podetkov (SURS, AJPES, MOPE, NCUP, BS, ...). Trenutno je ne voljo že kar nekaj podetkov, ki po se v večini zbirajo na mesečni ali letni ravni.	Data's availability Izziv je frekvence in raven podetkov, ki so trenutno ne voljo. Cilj je pridobivanje dnevnih podetkov, ki v večji meri niso ne voljo.	Risks and issues with relation to data Določeni podatki niso ne voljo, zato bo potreben zekup alternativnih podetkovnih virov. Alternativni podetkovni viri so običejno zelo dragi, hkrati pa ne dajo vedno najboljših informacij oziroma niso najbolj točni.	Data's processing V prvi fazi je potrebne identifikacije primerov uporabe, ki bodo ne voljo v središču. Po identifikaciji teh, bo potrebne tudi izbire ustreznega orodja, ki bo ne enem mestu omogočelo teko preproste prikaze kot tudi napredne analize.	Data skills Za izvedbo projekta je bil izbrani zunanji izvajalec, saj naprednih znanj znotraj organizacije nimamo.
<div style="background-color: yellow; padding: 10px; border: 1px solid black;">DODATNI PREDLOGI?</div>	<div style="background-color: lightgreen; padding: 10px; border: 1px solid black;">PRILOŽNOSTI?</div>	<div style="background-color: lightpurple; padding: 10px; border: 1px solid black;">PRILOŽNOSTI?</div>	<div style="background-color: lightblue; padding: 10px; border: 1px solid black;">PREDLOGI / IZZIVI?</div>	<div style="background-color: pink; padding: 10px; border: 1px solid black;">PREDLOGI / IZZIVI?</div>
Solution Combine what you've learned from the elements above into a statement describing your solution.				
<div style="display: flex; justify-content: space-around;"> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> <div style="background-color: purple; color: white; padding: 2px 5px; font-size: 8px;">you can type here</div> </div>				

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





The BDTI Canva






Created by the BDTI 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 email: EC-BDTI-PILOTS@ec.europa.eu

<p>Context: Who are you? Who are your stakeholders?</p>	<p>Objective(s): What is the problem you are trying to address? What is your timeframe?</p>
--	--

<p>Data's added value:  Which information helps you address the problem? From which sector and or domain?</p>	<p>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?</p>	<p>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?</p>	<p>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)?</p>	<p>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?</p>
--	---	--	--	---

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).

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.

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**.



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



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**.



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?



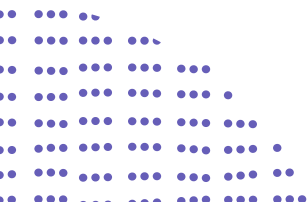
The vision: Public Procurement Data Spaces

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






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 Guasch¹ , Giorgia Lodi² , and Sander Van Dooren¹ 

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 Research 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/>

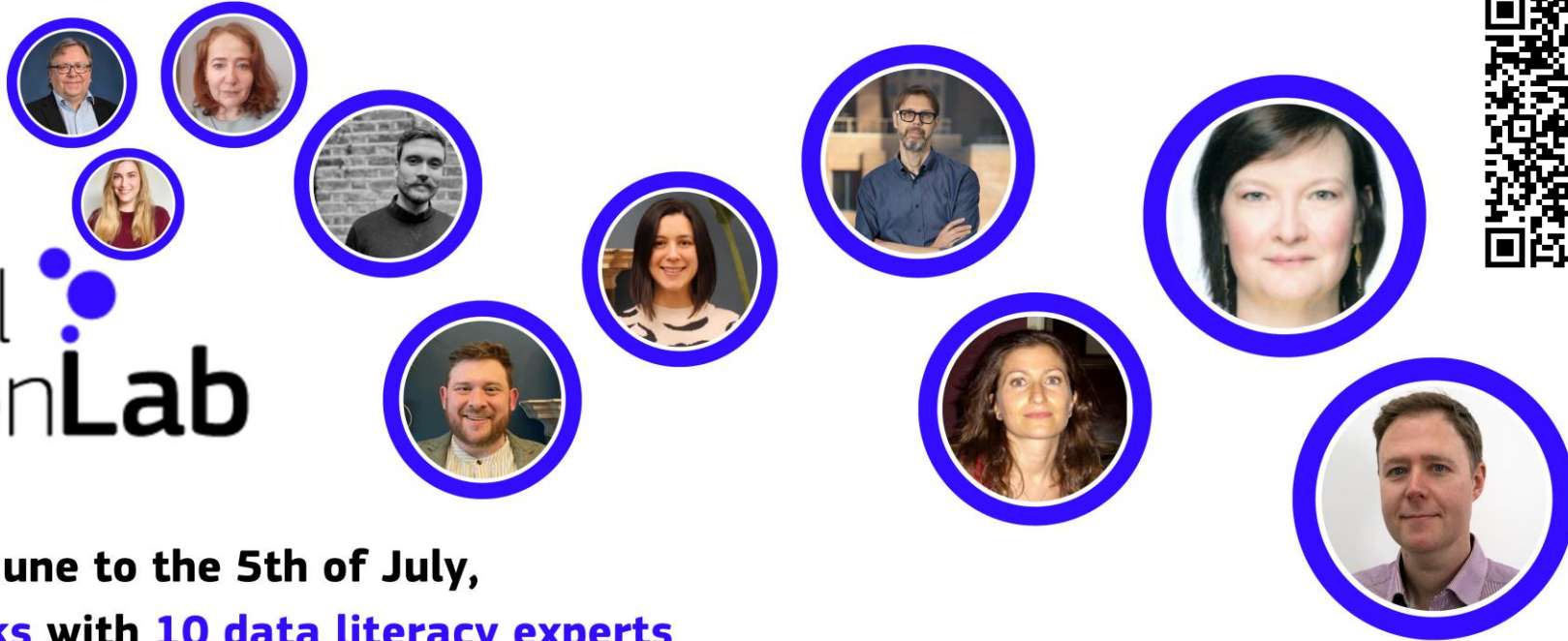
Developing data skills

iTalks Series:

'Exploring **data skills** initiatives to foster public sector innovation'



Digital
innovation**Lab**



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



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., & Cromptvoets, 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>

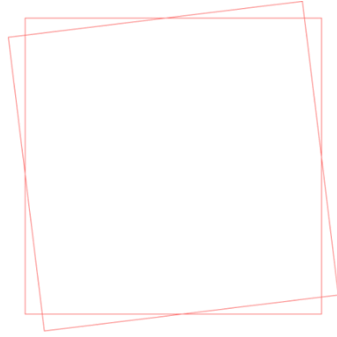
<https://dssc.eu/wp-content/uploads/2023/03/DSSC-Data-Spaces-Glossary-v1.0.pdf>

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

Get to know: Lobium

Tim Werkhoven



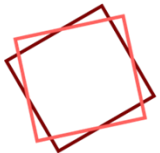


Lobium/Gavagai



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

Our journey



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

Where do we go from here?

Connect to other open (EU-) data sets



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



Beyond Open Data

Empowering citizens through engaging storytelling



GreeceInFigures.com



100Europeans.org



Unemployment

Article [Talk](#)

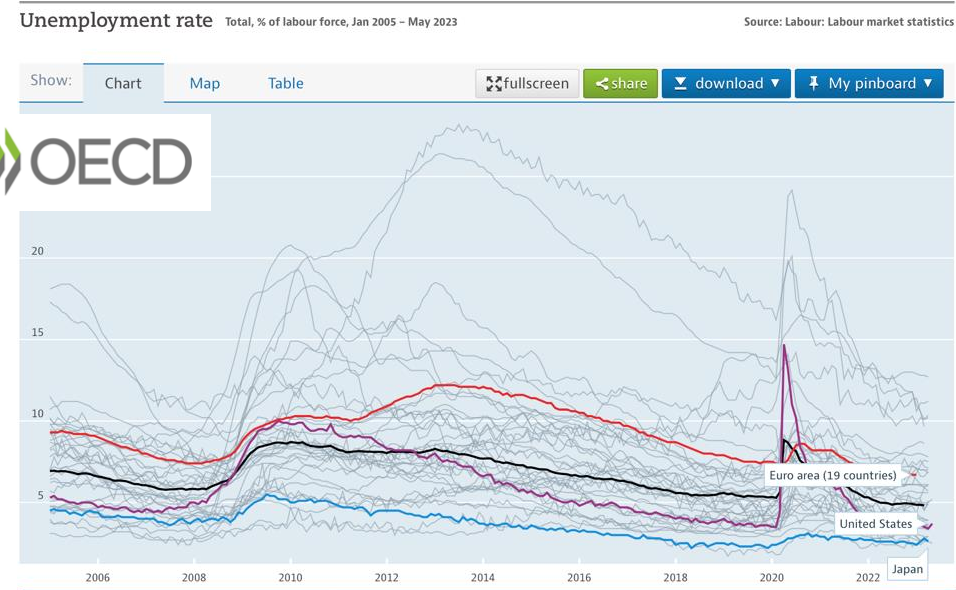
From Wikipedia, the free encyclopedia

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](#)

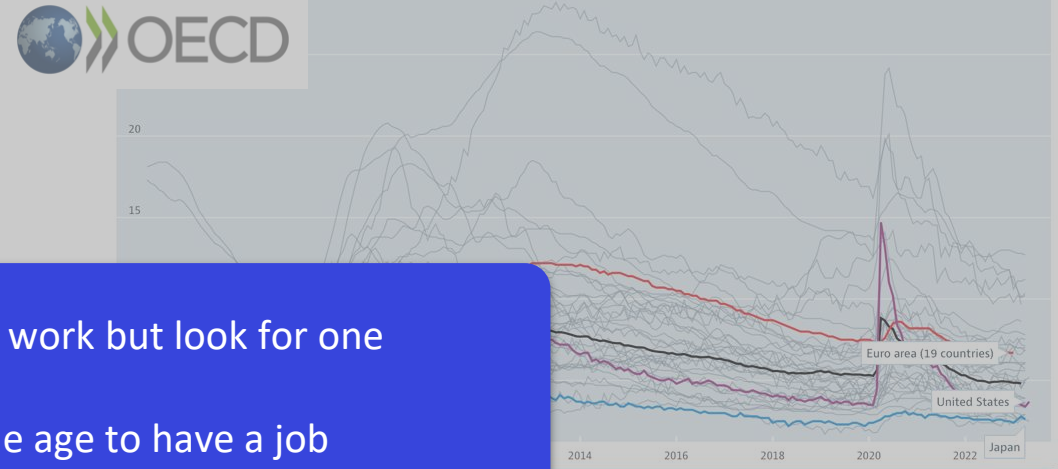
From Wikipedia, the free encyclopedia

Unemployment, according to the **OECD** (Organisation for Economic Co-operation and Development), is the number of people above a specified age who are not in paid **employment** or **self-employment** but are available for **work** during the reference period.

$$\frac{\text{\# People who cannot find work but look for one}}{\text{\# People who are at the age to have a job}}$$

Unemployment rate Total, % of labour force, Jan 2005 – May 2023 Source: Labour: Labour market statistics

Show: [Chart](#) [Map](#) [Table](#) [Fullscreen](#) [Share](#) [Download](#) [My pinboard](#)



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

Greece in Figures

Because Who We Are Data   

Unemployment in Greece

Dionysis Koulloli, Panagiotis Michaelides, Dimitris Michaelides
Last updated: 3/10/2023

Import

Unemployment is one of the most important economic indicators, as it is directly related to the general state of the economy and by extension of the society of a country.

A low unemployment rate means that the economy is capable of creating jobs for those who can and want to work. The more people who work, the more economic activity, incomes and spending increase – factors that directly boost the economy.

Low unemployment also works well for society, as it gives workers more choice about what type and place they work. At the same time, it incentivises employers to raise wages and create better working conditions in order to attract more talent.

i Unemployment is a "lagging indicator". In times of economic recession, it usually takes several months for unemployment to start rising. Correspondingly, a reduction is achieved a few months after economic growth begins.

Ανεργία - 🇬🇷 Ελλάδα και 🇪🇺 Ευρωπαϊκή Ένωση

Ποσοστό ανέργων επί του εργατικού δυναμικού.



Γράφημα: GreeceInFigures.com • Πηγή: ΕΛΣΤΑΤ/Eurostat • Κατεβάστε την εικόνα • Δημιουργήθηκε με το Datawrapper

greeceinfigures.com

It went well.

“Simple words”
“Concise descriptions”
“Nice looking graphs”



Website traffic:

<u>UNIQUE VISITORS</u>	TOTAL VISITS	TOTAL PAGEVIEWS	VIEWS PER VISIT
245k	278k	423k	1.52



Twitter analytics snapshots:

APR 2023 SUMMARY	
Tweet impressions	New followers
85.1K	205

MAR 2023 SUMMARY	
Tweet impressions	New followers
90.4K	128

“A Europe fit for the digital age”

But where does the digital age
take place?



Instagram

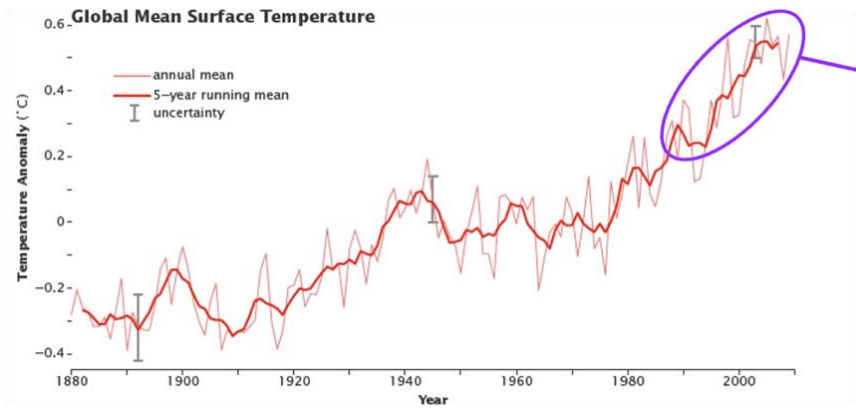
340+ million users in
Europe



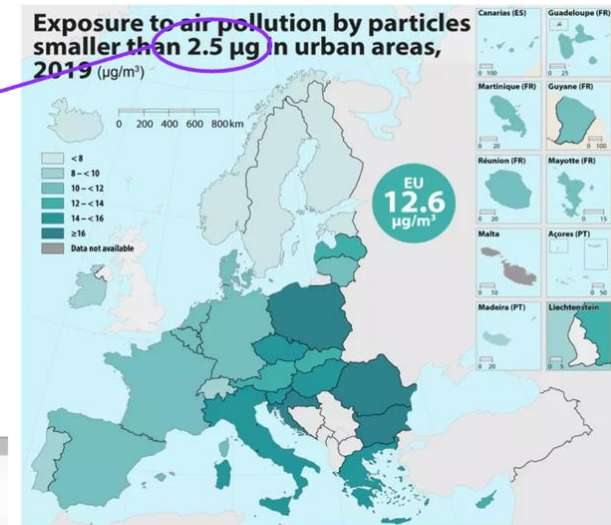
150+ million users in
Europe

Where does communication of big challenges take place?

Climate change:



What does this mean?



Hard to use
Too much information

eurostat
Your key to European statistics

Sign In | Register
Cookies | Privacy policy | Legal notice | My alerts | Contact | English | Translate

Search the Eurostat website + all products

News | Data | Publications | About Eurostat | Help

European Commission > Eurostat > Health > Data > Database

HEALTH | DATABASE

- Overview
- Data
 - Main tables
 - DATABASE
- Publications
- Methodology
- Legislation
- Links

Health (hlth)

- Health status (hlth_state)
- Health determinants (hlth_det)
 - Body mass index (BMI) (hlth_bmi)
 - Physical activity (hlth_ph)
 - Consumption of fruits and vegetables (hlth_cf)
 - Tobacco consumption (hlth_smok)
 - Alcohol consumption (hlth_alc)
 - Social environment (hlth_serv)
 - Health determinants - historical data (2008) (hlth_det_h)
- Health care (hlth_care)
- Disability (hlth_dsb)
- Causes of death (hlth_cdeath)
- Health and safety at work (hsw)

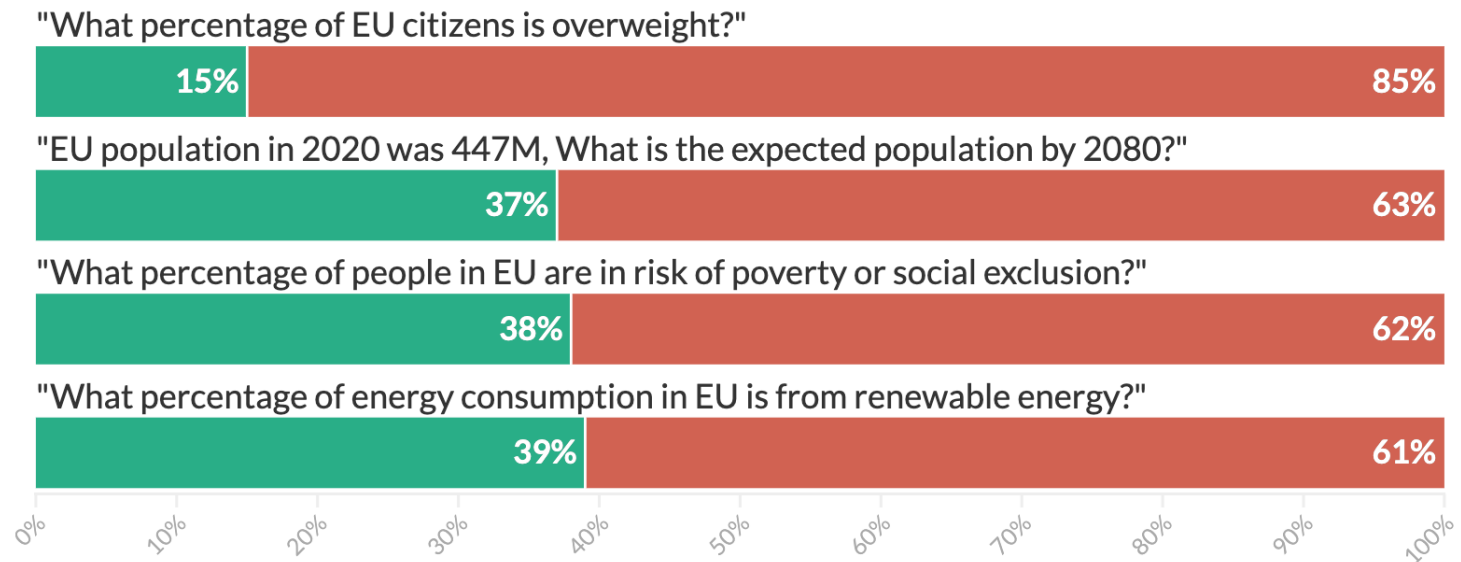
data.europa.eu
The official portal for European data



Is this communication effective?

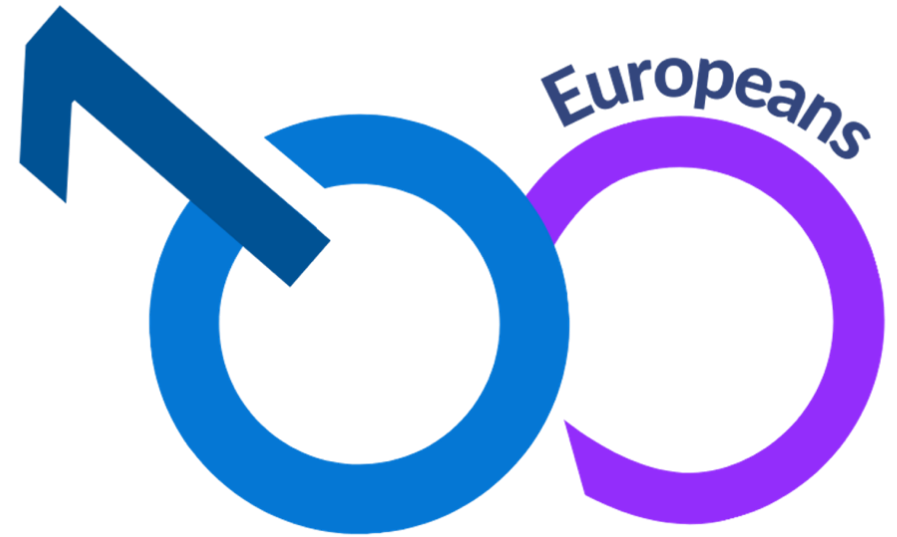
We ran a quiz to find out:

■ Answered Correct ■ Answered Wrong



“Perhaps
we can
change
that.”

Myself & Denis
Circa 2022

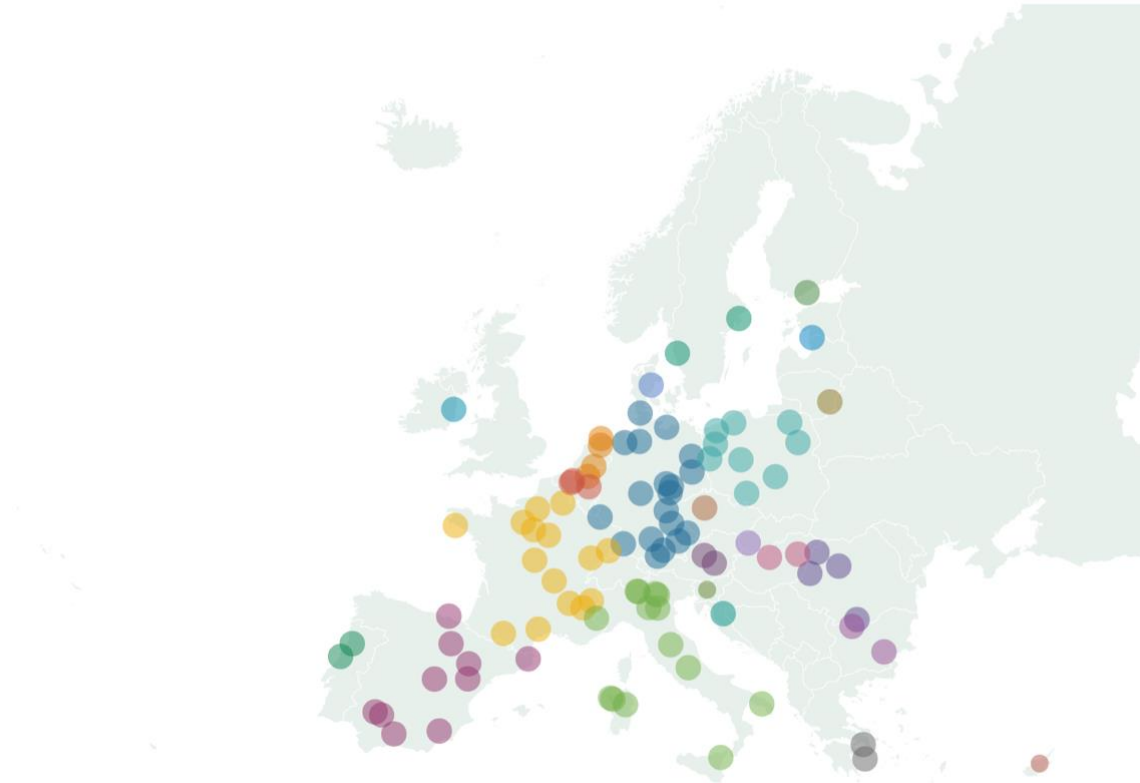


Scrollytelling Europe's biggest challenges

100europeans.org

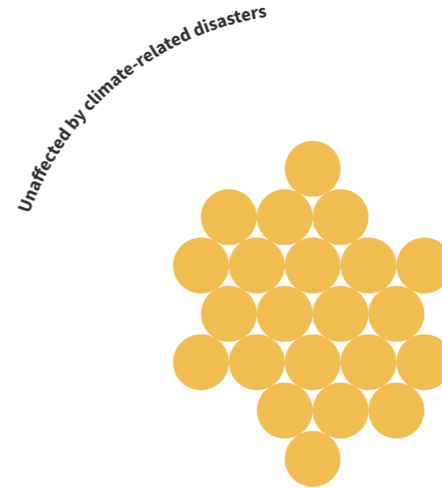
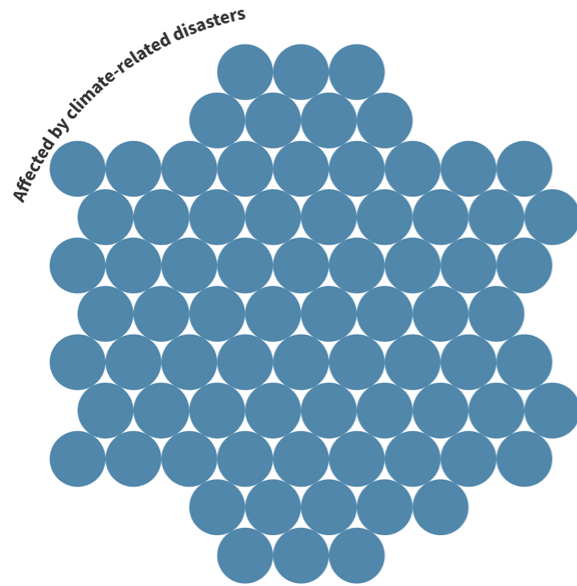
Let's reimagine the EU as a community of just 100 people...

If the EU was 100 people, where would they live?



Source: Eurostat • Note: dots are randomly distributed within countries

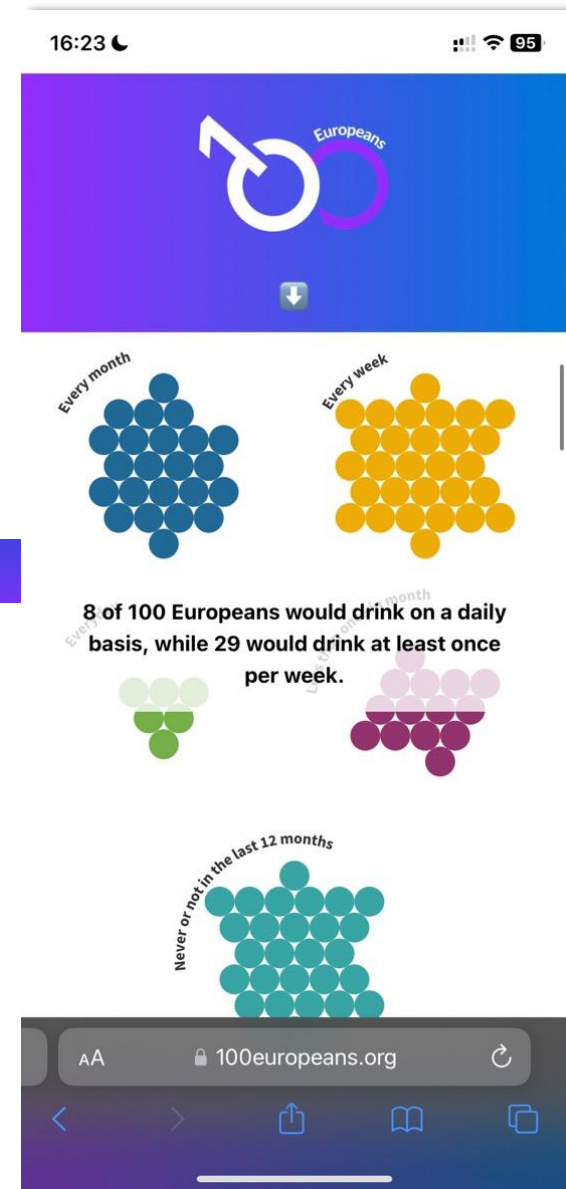
If Europe was 100 people ... 78 of them would be severely affected by climate-related disasters



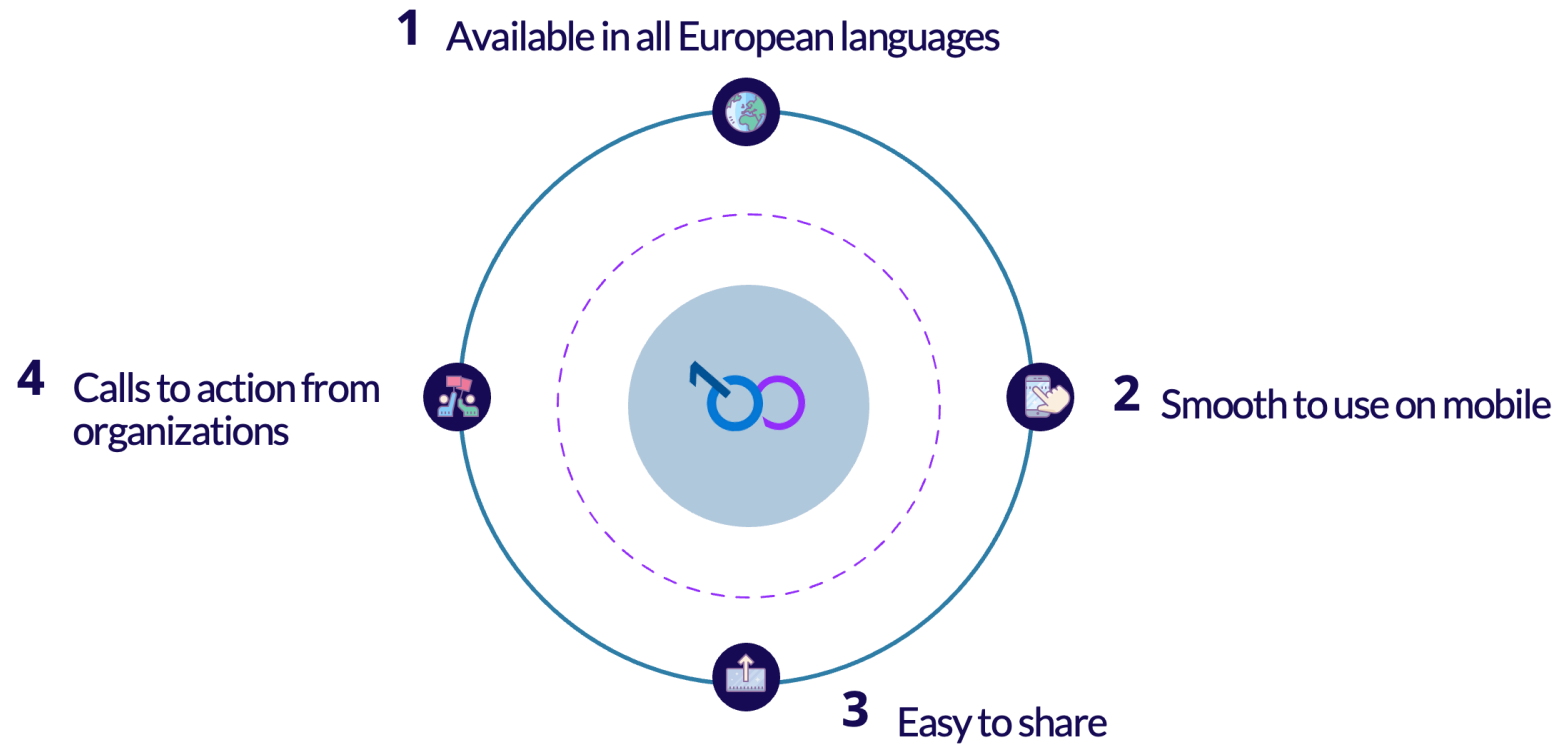
Online content is consumed via endless scrolling ...

Science communication can be done similarly!

Effortless
scrollytelling
learning, like



But how can we do it the right way?



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

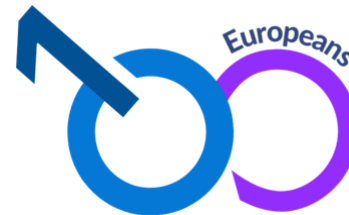
* by telling good stories.

Thank you!

Check the projects out 



GreeceInFigures.com



100Europeans.org

Panel discussion



Inmaculada Farfan Velasco



Maria Claudia Bodino



Tim Werkhoven



Dimitris Michailidis

Stay up-to-date on our
2023 activities!

The logo for Data Europa Academy is located in the bottom-left corner. It consists of a large red circle that partially overlaps a smaller dark blue circle. Inside the dark blue circle, the text "data.europa academy" is written in white, lowercase letters. The word "data" is on the top line, "europa" is on the middle line, and "academy" is on the bottom line. Small yellow dots are placed above the 'a' in "data" and above the 'o' in "europa".

data.
europa
academy

Join our next webinar!

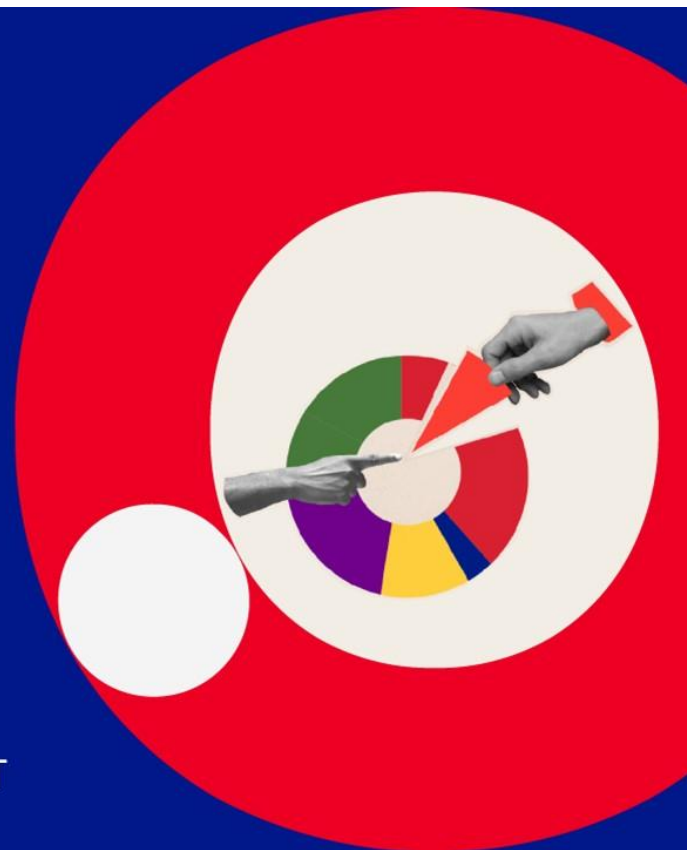
WEBINAR

Open Data Maturity
2022: Diving deeper
into the portal
dimension

14 July 2023

10.00 — 11.30 CET

data.
europa
academy



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

Please provide
your feedback!



Thank you

