

Enabling a data-informed public sector

*From hype to action using
the Big Data Test Infrastructure (BDTI)*

EC-BDTI-PILOTS@ec.europa.eu



DG CNECT

Directorate-General for Communications
Networks, Content and Technology

DG DIGIT

Directorate-General
for Informatics

Public Sector Information and the role of Data analytics

Data is **everywhere** and growing at an unprecedented pace.

Big Data: 3V- Volume, Variety, Velocity



Data is a key ingredient for **services, products, and effective policy making.**

There is an ambition to create a **single European market for data** and make more data available through powerful and trustworthy infrastructures and technologies, **in line with EU values and regulations, to support citizens, public sector and companies.**



What is the Big Data Test Infrastructure (BDTI)?



BDTI: Not **only** for **big data**, but for all **public sector information**



Six months free* of charge service
for the 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 its data** and accelerate
transition towards **data-driven decision making**

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

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
- 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 your data*
- BDTI provides a safe **testing environment to run big data experiments** for data space customers



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

About the Big Data Test Infrastructure (BDTI)

The BDTI is funded by the **Digital Europe Programme (DEP)**, which focuses 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

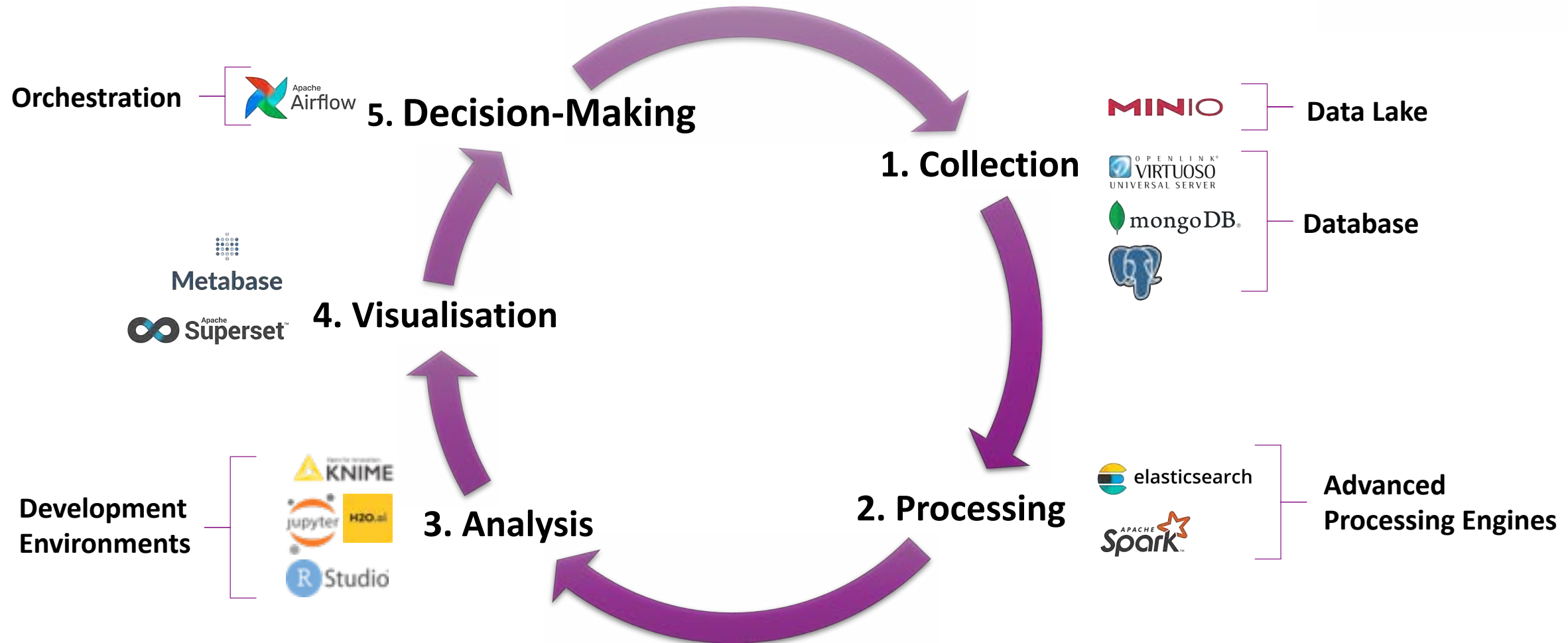
(Cloud, data and AI)

Advanced digital skills

Deployment and wide use of digital technologies



Open-source tools to support your data journey



The BDTI portal



portal.p1.bdti.datapatform.tech.ec.europa.eu

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.

Documentation
[Learn more](#)

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

Logout

v0.9.4



<https://youtu.be/fdzNCB1CVUM>

Optimisation of Public Lighting - Dún Laoghaire County

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 without projected expenses

Day	Full Lighting (€)	Projected (€)
Tuesday, April 10	27.3k	17.6k
Wednesday, April 11	27.3k	17.6k
Thursday, April 12	27.3k	17.6k
Friday, April 13	27.3k	17.6k
Saturday, April 14	27.3k	17.6k
Sunday, April 15	27.3k	17.6k
Monday, April 16	27.3k	17.6k


Projected savings next week: 27.3k Euro

Projected emission reduction next week: 3.36 tCO2

Scenario comparison of projected cumulative savings over 1 year

Traffic level	Cumulative emission savings (tCO2)	Cumulative savings (€)
Highway traffic	1,100	1,100
Urban traffic	1,000	1,000
Suburban traffic	1,000	1,000
Low traffic	1,000	1,000
Medium traffic	1,000	1,000
Other	0	0

The BDTI portal

100% 
open-source
components



The screenshot displays the BDTI Service Catalog interface. On the left is a navigation sidebar with links for Home, My Account, Service Catalog (active), My Services, and My Data. The main content area is titled "Service Catalog" and contains a grid of 20 service cards, each for a different open-source tool. Each card includes the tool's logo, name, version, a brief description, and a "Launch" button.

Service Name	Version	Description
Airflow	v2.3.0	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 geographical charts.
Apache Superset	v2.1	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 geographical 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 - lab-3.2.8 - datascience-notebook		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.
Jupyterlab - lab-3.4.2 - all-spark-notebook		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 - lab-4.0.4 - all-spark-notebook		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 - lab-4.0.4 - datascience-notebook		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.
Knime	v5.1.0	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 fun for the serious nature-loving data set for everyone, not just analysts and developers.
MinIO - RELEASE.2022-06-20T23-13-45Z		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	v3.2.1	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 Extrane-based solutions, transcending prevalent enterprise challenge areas such as: Disparate Databases and Data Sources, Web Service Composition, and Business Process Management.

Who is BDTI for ?



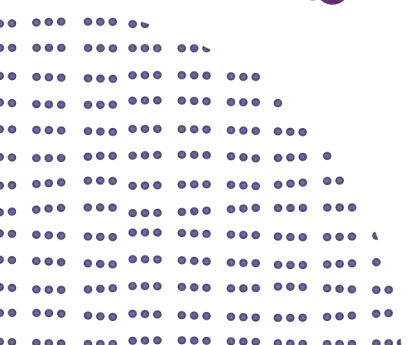
European Public Administrations

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



Partnerships with academia and the private sector

Academia (master, PhD students) and startups and companies (GovTech sector) can apply if they are **collaborating** with a public administration



Why use BDTI?



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



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



Test your idea → Extract value → Create knowledge

Who used it already?



Generalitat Valenciana needed a solution to sift through vast scientific clinical articles from various sources like PubMed.gov and COVID-19-related publications. Their challenge was to extract valuable insights and knowledge from this immense volume of data.



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.



The European Blood Alliance and DG SANTE embarked on a collaborative effort to establish a comprehensive open-access platform in Europe. This initiative aimed to collect crucial data for a focused study on COVID-19 convalescent plasma therapy. The challenge was to identify the most effective conditions for this treatment, utilising data analysis to make informed decisions and streamline research efforts.



A ready-to-use, virtual environment in which **data collected through a custom-built website** are ingested and anonymised, to be then analysed with advanced data visualisation 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**.



The Municipality faced the complex task of correlating diverse datasets within the city. Their objective was to understand the movement patterns of people across different districts over time. This data analysis challenge was essential for gaining insights into mobility trends and for redesigning services to promote cultural activities and events effectively.



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).

How to apply: a fast and simple process



Get familiar with the BDTI service on our [website](#)



Define your data analytics use case using our [BDTI Canva](#) and then fill in the template request form (see [website](#))



Submit your pilot request (template) [by email](#)



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!



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



EC-BDTI-PILOTS@ec.europa.eu

[BDTI website](#)



[BDTI's Joinup page](#)



[BDTI's newsletter](#)



Links and resources



- <https://big-data-test-infrastructure.ec.europa.eu/>
- <https://code.europa.eu/bdti/bdti-demonstrator>
- https://commission.europa.eu/publications/interoperable-europe-act-proposal_en
- 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://digital-strategy.ec.europa.eu/en/policies/data-governance-act>
- <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>
- <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>
- https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113
- <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024>
- <https://joinup.ec.europa.eu/collection/egovernment/solution/big-data-test-infrastructure-bdti>