

Para un sector público basado en datos

*De la palabra a la acción gracias a la **Big Data Test Infrastructure (BDTI)**
[Infraestructura de Prueba de Big Data]*

EC-BDTI-PILOTS@ec.europa.eu



DG CNECT

Dirección General de Redes de Comunicación
Contenido y Tecnología

DG DIGIT

Dirección General
de Informática



La información del sector público y el papel del análisis de datos

Los datos están por todas partes y crecen a un ritmo que no tiene precedentes

- **Macrodatos (Big Data):** 3V - Volumen, Variedad, Velocidad



Los datos son un ingrediente clave de **los servicios, los productos y la elaboración eficaz de políticas.**

Existe la ambición de crear un **mercado único de datos** y de hacer accesibles más datos gracias a infraestructuras y tecnologías potentes y fiables, **acordes con los valores y normas de la UE** y que ayuden a los ciudadanos, el sector público y las empresas.



¿Qué es la Infraestructura de Prueba de Macrodatos (*Big Data Test Infrastructure*, BDTI)?



BDTI: no solo para los macrodatos, sino para **toda la información del sector público**



Seis meses de servicio gratuito*
para las administraciones públicas de
la UE



Instrumentos y soporte de
análisis de datos listos para su
uso



Plataforma en la nube basada en
herramientas de código abierto



Ayudar al sector público a **extraer conocimiento de sus datos** y acelerar la transición hacia una **toma de decisiones basada en datos**.

* El coste del proyecto piloto debe ajustarse a los límites de financiación del presupuesto piloto de la BDTI

Objetivos de la BDTI

Objetivos

- Mayor facilidad de acceso, interoperabilidad, calidad y facilidad de uso de la información del sector público, acorde con los requisitos de la **Directiva sobre datos abiertos [open data]**
- Impulsar la **reutilización y combinación de datos públicos abiertos** en toda la UE con vistas a desarrollar productos y servicios de información, incluidas aplicaciones de inteligencia artificial
- Conjuntos de datos de alto valor - Directiva sobre datos abiertos [open data]
- Poner a prueba colaboraciones de intercambio de datos **de empresa a administración** (*Business to Government, B2G*) de **interés público**
- Centro de apoyo al espacio de datos: explorar y experimentar con tus datos*
- La BDTI ofrece un **entorno de pruebas** seguro para realizar **experimentos con macrodatos** para clientes del espacio de datos

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



Acerca de la Infraestructura de Prueba de Macrodatos (BDTI)

La BDTI se financia a través del **programa Europa Digital**, centrado en llevar la tecnología digital a las empresas, los ciudadanos y las administraciones públicas.

El Programa Europa Digital proporciona financiación en cinco áreas cruciales:

Informática de alto rendimiento

Inteligencia artificial

(Nube, datos e inteligencia artificial)

Ciberseguridad

Competencias digitales avanzadas

Despliegue y uso generalizado de tecnologías digitales



Herramientas de código abierto (open source) para ayudarte en tu recorrido por los datos



**Orchestration
(Orquestación)**



5. Toma de decisiones

1. Recogida

MINIO

**Data Lake
(Lago de
datos)**



**Data base
(Base de
datos)**

mongoDB



Metabase

4. Visualización

2. Procesamiento

elasticsearch

**Advanced
Processing Engines
(Motores
avanzados de
procesamiento)**



**Apache
Superset**

3. Análisis

**Entornos de
desarrollo**



El portal BDTI



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

European Commission

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

Logout

v0.9.4

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](#)

Big Data Test Infrastructure - BDTI Introduction Video

BIG DATA TEST INFRASTRUCTURE (BDTI)

Watch on YouTube

<https://youtu.be/fdzNCB1CVUM>

Optimisation of Public Lighting - Dún Laoghaire County

Filters: Traffic level at which to turn public lighting off

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)	Savings (%)
Tuesday, April 18	22.4k	17.7k	17.7%
Wednesday, April 19	22.4k	17.5k	17.5%
Thursday, April 20	22.4k	16.9k	16.9%
Friday, April 21	22.4k	16.3k	16.3%
Saturday, April 22	22.4k	19.9k	19.9%
Sunday, April 23	22.4k	15.7k	15.7%
Monday, April 24	22.4k	17.3k	17.3%

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 name	Cumulative emission savings (tCO2)	Cumulative savings (Euro)
5 Very high traffic	179k	3,342k
4 High traffic	84.4k	2,024k
3 Medium traffic	726.91	5,809k
2 Low traffic	466.0	3,789k
1 Very low traffic	221.5	1,795k
0 None	0	0

El portal BDTI

100% 
de componentes de
código abierto



Service Catalog

Home

My Account

Service Catalog

My Services

My Data

Airflow - v2.3.0

Description

Airflow is a platform created by the community to programmatically author, schedule and monitor workflows.

Launch

Apache Superset - v1.0

Description

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.

Launch

Apache Superset v2.1

Description

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.

Launch

ElasticSearch - v7.17.3

Description

Elasticsearch is the distributed, RESTful search and analytics engine at the heart of the Elastic Stack.

Launch

H2O-3 - v36.1.1

Description

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.

Launch

Jupyterlab - lab-3.2.8 - datascience-notebook

Description

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.

Launch

Jupyterlab - lab-3.4.2 - all-spark-notebook

Description

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.

Launch

Jupyterlab - lab-4.0.4 - all-spark-notebook

Description

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.

Launch

Jupyterlab - lab-4.0.4 - datascience-notebook

Description

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.

Launch

Kibana - v7.17.3

Description

Kibana is your window into the Elastic Stack. Specifically, it is a browser-based analytics and search dashboard for Elasticsearch.

Launch

Knime - v4.5.3

Description

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.

Launch

Knime - v5.1.0

Description

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.

Launch

Metabase - v0.43.3

Description

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.

Launch

MinIO - RELEASE.2022-06-20T23-13-45Z

Description

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.

Launch

MongoDB - v4.4.13

Description

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.

Launch

PgAdmin4 - v6.8

Description

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.

Launch

Postgresql - v14.2.0

Description

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.

Launch

RStudio - v4.1.2

Description

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.

Launch

Spark - v3.2.1

Description

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.

Launch

Virtuoso - v7.2.7

Description

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.

Launch

¿Para quién es la BDTI?



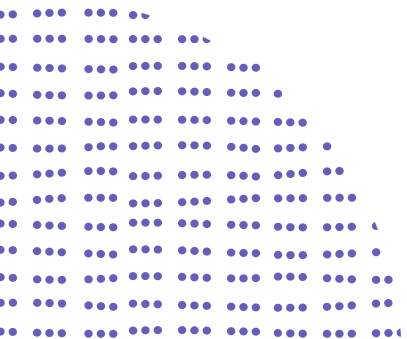
Administraciones públicas europeas

Todas las administraciones públicas europeas de ámbito **local, regional y nacional** pueden presentar una candidatura independiente a un proyecto de BDTI



Asociaciones con el mundo académico y el sector privado

Las universidades (estudiantes de máster y doctorado) y las empresas tanto emergentes como ya existentes (del sector GovTech) pueden presentar su candidatura si **colaboran** con una administración pública



¿Por qué utilizar la BDTI?



Seis meses gratuitos, incluidos **asesoramiento y asistencia técnica**



Experimentar con el análisis de datos utilizando una **infraestructura de alto rendimiento** que aprovecha las posibilidades de la **nube elástica (elastic cloud)**



Recibir orientaciones para pasar de un proceso piloto a un proceso **listo para la producción**



Poner a prueba tu idea → Extraer valor → Crear conocimiento

¿Quién ya la ha utilizado?



La Generalitat Valenciana necesitaba una solución para analizar un gran número de artículos científicos clínicos de diversas fuentes tales como PubMed.gov y publicaciones relacionadas con la COVID-19. El reto consistía en extraer información y conocimientos útiles de este inmenso volumen de datos.



Herramientas avanzadas de **visualización de datos y extracción de textos** que permiten **extraer los conocimientos** contenidos en los documentos, apoyando a los médicos y los gestores en sus prácticas clínicas y su trabajo diario.



La European Blood Alliance y DG SANTE hicieron un esfuerzo de colaboración para establecer una plataforma global de acceso abierto en Europa. La iniciativa tenía por objetivo recopilar datos fundamentales para un estudio centrado en la terapia con plasma de convalecientes de COVID-19. El reto consistía en determinar las condiciones más eficaces para este tratamiento utilizando el análisis de datos a fin de tomar decisiones con conocimiento de causa y racionalizar los esfuerzos de investigación.



Un entorno virtual listo para su uso en el que los **datos recogidos a través de una web hecha a medida** se integran y anonimizan para después analizarse con herramientas analíticas y de visualización de datos avanzadas. En un principio, solo se procesaron datos de donaciones, posteriormente ampliados al **análisis de extremo a extremo del plasma sanguíneo, desde la donación hasta el paciente o el ensayo clínico.**



El Ayuntamiento se encontró ante la compleja tarea de correlacionar distintos conjuntos de datos dentro de la ciudad. Su objetivo era comprender los patrones de movimiento de personas de diferentes distritos a lo largo del tiempo. Este ejercicio de análisis de datos fue fundamental para obtener información sobre las tendencias de movilidad y rediseñar los servicios a fin de promover actividades y eventos culturales con eficacia.



Análisis predictivo, descriptivo y de series temporales sobre múltiples conjuntos de datos recogidos **antes, durante y después de la pandemia de COVID-19**, tales como sensores de wifi pública, datos de aparcamientos y datos georreferenciados de movimientos de personas (turistas).

Cómo presentar una candidatura: un proceso rápido y sencillo



Familiarízate con el servicio BDTI en nuestra [web](#)



» Define tu caso de uso de análisis de datos utilizando nuestra herramienta [BDTI Canva](#) y completa el formulario de solicitud (consulta la [web](#))



» Presenta tu solicitud (formulario) por correo electrónico

EC-BDTI-PILOTS@ec.europa.eu



» Reúnete con nosotros para profundizar en tu caso de uso



» El proyecto piloto se aprueba si: aporta valor, puede realizarse en 6 meses y tiene recursos suficientes (capacidades, equipo, datos)

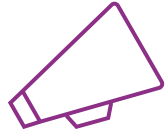


» Crea tu entorno de prueba



» ¡Empieza a experimentar y crear valor!

¡Gracias por tu atención!



¿Trabajas para una administración pública que necesita infraestructura para el análisis de datos? Contáctanos



EC-BDTI-PILOTS@ec.europa.eu

[Web de la
BDTI](#)



[Página de
inscripción en la
BDTI](#)



[Boletín de la BDTI](#)



Enlaces y recursos

- <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>