

# Pour un secteur public guidé par les données

*Des beaux discours aux actes grâce à  
Big Data Test Infrastructure (BDTI)*

[EC-BDTI-PILOTS@ec.europa.eu](mailto:EC-BDTI-PILOTS@ec.europa.eu)



## **DG CNECT**

Direction générale des  
réseaux de communication, du contenu  
et des technologies

## **DG DIGIT**

Direction générale  
Services Numériques



# Les informations du secteur public et le rôle de l'analyse des données (Data analytics)

Les données sont **omniprésentes** et leur volume progresse à un rythme sans précédent.

- **Mégadonnées** (Big Data): les 3 V - Volume, Variété, Vitesse



Les données sont un ingrédient essentiel pour **les services, les produits et l'élaboration efficace des politiques.**

L'ambition est de créer un **marché européen unique des données** et de rendre davantage de données disponibles grâce à des infrastructures et des technologies puissantes et fiables, **dans le respect des valeurs et des dispositions réglementaires de l'Union**, au service des citoyens, du secteur public et des entreprises.



# En quoi consiste l'infrastructure de test de mégadonnées (Big Data Test Infrastructure - BDTI)?



BDTI: ne concerne pas seulement les mégadonnées (big data), mais l'ensemble des **informations du secteur public**



Six mois de service gratuit\*  
pour les administrations publiques de l'UE



Pile d'analyse de données prête à  
l'emploi et assistance



Plateforme en nuage (cloud)  
fondée sur des outils **open source**



Aider le secteur public à **tirer des enseignements de ses données** et à accélérer la transition vers une **prise de décision fondée sur les données**

\* Le coût du projet pilote doit s'inscrire dans les limites de financement du budget du projet pilote BDTI

# Objectifs de l'infrastructure de test de mégadonnées (Big Data Test Infrastructure)

## Objectifs

- Améliorer l'accessibilité, l'interopérabilité, la qualité et l'utilisation des informations du secteur public conformément aux exigences de la **directive sur les données ouvertes (Open Data Directive)**
- Favoriser la **réutilisation et la combinaison des données publiques ouvertes** partout dans l'Union pour la conception de produits et de services d'information, y compris les applications d'intelligence artificielle
- Ensembles de données de forte valeur – Directive sur les données ouvertes (Open Data Directive)
- Mettre à l'épreuve des collaborations de partage de données **entre entreprises et administrations publiques** (Business-to-Government - B2G) dans l'**intérêt général**
- Centre d'appui aux espaces de données (Data Space Support Centre): exploration et expérimentation avec vos données\*
- BDTI constitue un **environnement de test sûr pour mener des expériences avec des mégadonnées** (big data) pour la clientèle des espaces de données (data space)

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



# À propos de l'infrastructure de test de mégadonnées (Big Data Test Infrastructure - BDTI)

BDTI est financée par le **programme pour une Europe numérique**, qui vise à mettre la technologie numérique au service des entreprises, des citoyens et des administrations publiques.

Le programme pour une Europe numérique prévoit un financement stratégique dans cinq domaines primordiaux:

**Calcul à haute performance**

**Cybersécurité**

**Intelligence artificielle**

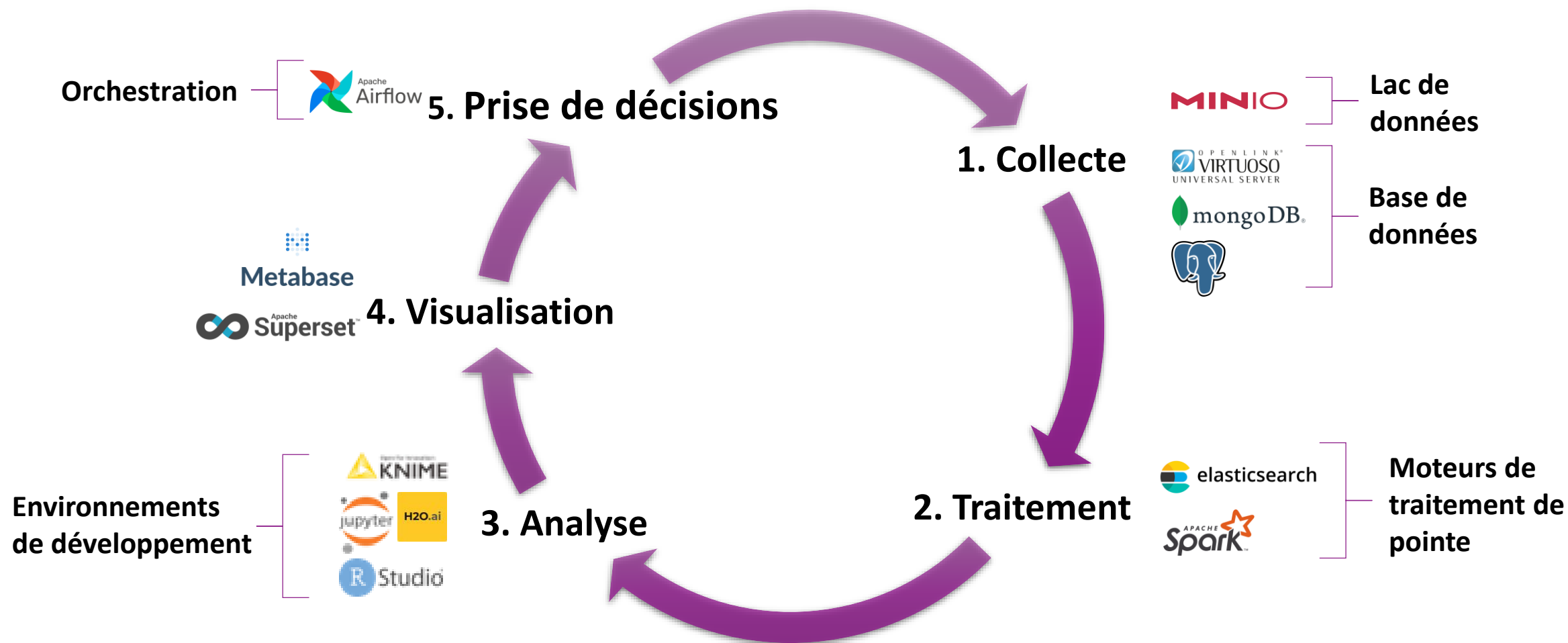
*(Nuage, données et intelligence artificielle)*

**Compétences numériques avancées**

**Déploiement et utilisation à grande échelle des technologies numériques**



# Outils open source pour accompagner le parcours de vos données





# Le portail 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)

Watch on YouTube

<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 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	16.5k	16.5%
Friday, April 21	22.4k	16.3k	16.3%
Saturday, April 22	22.4k	19.9k	19.9%
Sunday, April 23	22.4k	16.7k	16.7%
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 \*

Traffic level name	Cumulative emission savings (tCO2)	Cumulative savings (Euro)
5 Very high traffic	179k	3,342k
4 High traffic	84.4k	1,604k
3 Medium traffic	726.91	5,809k
2 Low traffic	466.0	3,784k
1 Very low traffic	221.5	1,794k
0 None	0	0

# Le portail BDTI

100 %   
de composants  
open source



## Service Catalog

### Airflow - v2.3.0

#### Description

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

[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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### ElasticSearch - v7.17.3

#### Description

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

[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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)

### 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](#)



Home

My Account

Service Catalog

My Services

My Data



# Pour qui BDTI est-elle conçue?



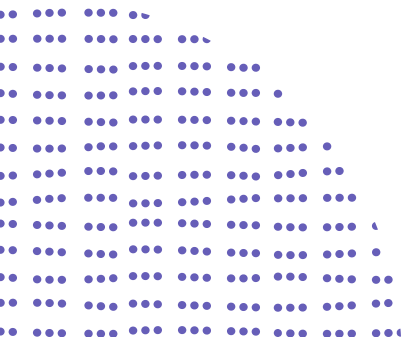
## Administrations publiques européennes

Toutes les administrations publiques européennes au **niveau local, régional et national** peuvent présenter individuellement une demande pour un projet pilote BDTI.



## Partenariats avec les universités et le secteur privé

Les universités (étudiants en master, en doctorat/PhD), les start-ups et les entreprises (secteur de la GovTech) peuvent présenter une demande si elles **collaborent** avec une administration publique.



# Pourquoi faire appel à BDTI?



Bénéficiez de **six mois gratuits**, comprenant **le conseil et l'assistance technique**.



Menez des expériences d'analyse de **données** en utilisant une **infrastructure** de haute **performance** qui tire parti de la puissance d'**Elastic Cloud**.



Bénéficiez de **conseils** pour passer d'un processus pilote à un processus **prêt pour la production**



Mettez votre idée à l'épreuve → Retirez-en de la valeur → Créez des connaissances

# Qui a déjà fait appel à BDTI?



La Generalitat Valenciana avait besoin d'une solution pour passer au crible un grand nombre d'articles scientifiques cliniques provenant de plusieurs sources telles que PubMed.gov et des publications liées à la COVID-19. Elle cherchait à tirer des enseignements et des connaissances utiles de cet immense volume de données.



Des **outils avancés de visualisation de données et de fouille de textes** permettant d'**extraire les connaissances** contenues dans les documents, afin d'aider les cliniciens et les gestionnaires dans leurs pratiques cliniques et leur travail quotidien.



L'Alliance européenne du sang (EBA) et la DG SANTE ont entrepris de collaborer à la mise en place d'une plateforme globale en libre accès en Europe. Cette initiative visait à collecter des données essentielles pour une étude ciblée sur l'utilisation thérapeutique de plasma prélevé sur des convalescents de la COVID-19. La difficulté consistait à déterminer les conditions optimales pour ce traitement, en utilisant l'analyse des données pour prendre des décisions éclairées et rationaliser le travail de recherche.



Un environnement virtuel prêt à l'emploi dans lequel les **données collectées par l'intermédiaire d'un site web ad hoc** sont importées et anonymisées, pour être ensuite analysées à l'aide d'outils avancés de visualisation et d'analyse des données. Au départ, seules les données relatives aux dons étaient traitées, puis le champ d'application a été élargi pour donner une bonne représentation de **tout le parcours du plasma sanguin, depuis le don jusqu'au patient ou à l'essai clinique.**



La municipalité de Florence devait s'atteler à la tâche complexe de mettre en corrélation différents ensembles de données au sein de la ville. Son objectif était de comprendre les habitudes de déplacement des personnes dans les différents quartiers au fil du temps. Ce délicat travail d'analyse des données était essentiel pour se faire une idée des tendances en matière de mobilité et pour repenser les services afin de promouvoir efficacement les activités et les manifestations culturelles.



Analyse prédictive, descriptive et de séries chronologiques concernant de nombreux ensembles de données collectées **avant, pendant et après la pandémie de COVID-19**, notamment à partir de capteurs Wi-Fi publics, de données de stationnement et de données géoréférencées sur les déplacements des personnes (c'est-à-dire des touristes).

# Présenter une demande: une procédure simple et rapide



Familiarisez-vous avec le service BDTI sur notre [site web](#)



» Définissez votre cas d'utilisation en analyse de données à l'aide de notre [Canva BDTI](#), puis remplissez le formulaire de demande (voir le [site web](#))



» Soumettez votre demande de projet pilote (formulaire) par courriel

[EC-BDTI-PILOTS@ec.europa.eu](mailto:EC-BDTI-PILOTS@ec.europa.eu)



» Rencontrez-nous pour élaborer votre cas d'utilisation



» Le projet pilote est approuvé si:  
il apporte une valeur ajoutée, il peut être réalisé en six mois et des ressources suffisantes sont disponibles (compétences, équipe, données)

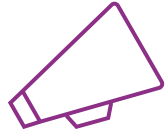


» Votre environnement de test est en place



» Vous pouvez commencer à expérimenter et à créer de la valeur!

# Merci de votre attention!



Vous travaillez pour une administration publique qui a besoin d'une infrastructure pour l'analyse de données? Prenez contact avec nous!



[EC-BDTI-PILOTS@ec.europa.eu](mailto:EC-BDTI-PILOTS@ec.europa.eu)

[Site web de BDTI](#)



[Page Joinup de la BDTI](#)



[Bulletin d'information de BDTI](#)



# Liens et ressources

---

- <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/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://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/fr/policies/data-governance-act>
- <https://digital-strategy.ec.europa.eu/fr/policies/european-approach-artificial-intelligence>
- <https://digital-strategy.ec.europa.eu/fr/activities/digital-programme>
- <https://dssc.eu/wp-content/uploads/2023/03/DSSC-Data-Spaces-Glossary-v1.0.pdf>
- <https://digital-strategy.ec.europa.eu/fr/library/staff-working-document-data-spaces>
- [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_22\\_1113](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>