

Sudaryti sąlygas duomenimis grindžiamam viešajam sektoriui

*Nuo šūkių prie veiksmų pasitelkiant
Big Data Test Infrastructure (didžiųjų
duomenų testavimo infrastruktūrą) (BDTI)*

EC-BDTI-PILOTS@ec.europa.eu



CNECT GD

Ryšių tinklų, turinio ir technologijų
generalinis direktoratas

DIGIT GD

Informatikos
generalinis direktoratas



Viešojo sektoriaus informacija ir duomenų analizės (*data analytics*) vaidmuo

Duomenų yra **visur** ir jų kiekis didėja beprecedenčiu tempu.

- Trys **didžiuosius duomenis (*big data*)** nusakantys parametrai: kiekis, įvairovė, sparta



Duomenys – labai svarbus **paslaugų, produktų ir veiksmingo politikos formavimo** elementas.

Siekiant padėti piliečiams, viešajam sektoriui ir įmonėms, užsibrėžta, laikantis ES vertybių ir taisyklių, sukurti bendrąją Europos duomenų rinką ir užtikrinti, kad pasitelkiant galingą bei patikimą infrastruktūrą ir technologijas taptų prieinama daugiau duomenų.





Kas yra *Big Data Test Infrastructure* (BDTI)?

BDTI – ne tik didiesiems duomenims (*big data*), bet ir visai **viešojo sektoriaus informacijai**



Galimybė ES viešojo administravimo institucijoms **paslaugomis šešis mėnesius naudotis nemokamai***



Parengtų naudoti duomenų analizės priemonių rinkinys ir pagalba



Atvirojo kodo (*open-source*) priemonėmis pagrįsta **debesijos (cloud) platforma**



Tikslas – padėti viešajam sektoriui **gauti įžvalgų iš savo duomenų** ir sparčiau pereiti prie **duomenimis grindžiamo sprendimų priėmimo** praktikos.

* Bandomojo projekto išlaidos turi neviršyti BDTI bandomiesiems projektams numatyto biudžeto ribų.

Big Data Test Infrastructure tikslai

Tikslai

- Padėti užtikrinti, kad viešojo sektoriaus informacija pagal **Atvirųjų duomenų (open data) direktyvoje** nustatytą reikalavimą taptų dar lengviau prieinama, sąveikesnė, kokybiškesnė ir tinkamesnė naudoti
- Visoje ES skatinti **atvirųjų viešųjų duomenų pakartotinį naudojimą ir jungimą** informacijos produktams ir paslaugoms, įskaitant DI prietaikus, kurti
- Vykdyti su didelės vertės duomenų rinkiniais nustatytus reikalavimus, nustatytus Atvirųjų duomenų (*open data*) direktyvoje
- Išbandyti bendradarbiavimą, grindžiamą **verslo dalijimusi duomenimis su valdžios sektoriumi (Business-to-Government – B2G) visuotiniam labui**
- Duomenų erdvių (*data space*) paramos centras: tyrinėkite savo duomenis ir su jais eksperimentuokite*
- BDTI suteikia duomenų erdvių (*data space*) naudotojams galimybę saugioje **testavimo aplinkoje eksperimentuoti su didžiais duomenimis (big data)**



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

Apie *Big Data Test Infrastructure* (BDTI)

BDTI yra finansuojama pagal **Skaitmeninės Europos programą**, kurios pagrindinis tikslas – **aprūpinti** įmones, piliečius ir **viešojo administravimo institucijas** skaitmeninėmis technologijomis.

Penkios itin svarbios sritys, kurioms teikiamas strateginis finansavimas pagal Skaitmeninės Europos programą:

Našioji kompiuterija

Kibernetinis saugumas

Dirbtinis intelektas

(debesija, duomenys ir DI)

**Aukšto lygio skaitmeniniai
įgūdžiai**

Skaitmeninių technologijų diegimas ir platus naudojimas



Atvirojo kodo (*open-source*) priemonės, padėsiančios jums dirbti su duomenimis



Orkestravimas
(*Orchestration*)



5. Sprendimų priėmimas

1. Rinkimas

MINIO

Duomenų ežeras
(*Data Lake*)



Duomenų bazės
(*Databases*)

mongoDB



Metabase

4. Vizualizavimas



2. Apdorojimas

elasticsearch

Pažangios
apdorojimo sistemos
(*Advanced
Processing Engines*)



Programavimo
aplinkos
(*Development
Environments*)



3. Analizė



BDTI portalas



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	20.5%
Wednesday, April 19	22.4k	17.5k	21.4%
Thursday, April 20	22.4k	16.9k	24.1%
Friday, April 21	22.4k	16.3k	27.7%
Saturday, April 22	22.4k	19.9k	11.2%
Sunday, April 23	22.4k	19.7k	12.0%
Monday, April 24	22.4k	17.3k	22.3%

Projected savings next week

27.3k Euro


Projected emission reduction next week

3.36 tCO2

Scenario comparison of projected cumulative savings after 1 year *

Traffic level name	Cumulative emission savings (tCO2)	Cumulative savings (Euro)
5 Very high traffic	179k	3,342k
4 High traffic	94.4k	1,762k
3 Medium traffic	726.91	1,382k
2 Low traffic	466.0	8,782k
1 Very low traffic	221.5	1,792k
0 None	0	0

BDTI portalas

100 % 
komponentų –
atvirojo kodo (*open-
source*)



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

Kam skirta BDTI?



Europos viešojo administravimo institucijoms

Visos Europos **vietos, regioninio ir nacionalinio lygmenų** viešojo administravimo institucijos paraiškas dėl BDTI bandomųjų projektų gali teikti savarankiškai.



Partnerystėms su akademinė bendruomene ir privačiuoju sektoriumi

Akademinė bendruomenė (magistrantūros ir doktorantūros (PhD) studentai) ir startuoliai (*startups*) bei įmonės (*GovTech* sektorius) paraiškas gali teikti, jei **bendradarbiauja** su viešojo administravimo institucijomis.

Kodėl verta naudotis BDTI?



Paslaugomis, įskaitant **konsultacinę ir techninę** pagalbą, **šešis mėnesius** naudokitės **nemokamai**



Eksperimentuokite su duomenų analize naudodamiesi **našiąja infrastruktūra**, kuriai pasitelkiama „Elastic Cloud“ galia



Gaukite patarimų, kaip nuo bandomojo projekto pereiti prie **praktiškai taikyti parengto** proceso



Išbandykite savo idėją → Gaukite naudos →
Generuokite žinias

Kas jau pasinaudojo šia infrastruktūra?



Valensijos generalitetui reikėjo rasti sprendimą, kaip išanalizuoti daugybę mokslinių klinikinių straipsnių, pateikiamų įvairiuose šaltiniuose, kaip antai „PubMed.gov“, ir su COVID-19 susijusių publikacijų. Jų uždavinys buvo iš šio didžiulio duomenų kiekio gauti vertingų įžvalgų ir žinių.



Pažangios duomenų vizualizavimo ir teksto duomenų gavybos priemonės, padedančios iš dokumentų išgauti žinias, pravarčias gydytojų ir vadovų klinikinei praktikai ir kasdieniam darbui.



Europos kraujo aljansas ir Sveikatos ir maisto saugos GD ėmėsi užduoties drauge sukurti visapusišką europinę platformą, grindžiamą atvirąja prieiga. Šia iniciatyva siekta surinkti duomenis, itin svarbius tiksliniam gydymo naudojant imuninę pasveikusiuųjų nuo COVID-19 plazmą tyrimui. Uždavinys buvo, remiantis duomenų analize, nustatyti veiksmingiausias šio gydymo sąlygas, kad būtų galima priimti pagrįstus sprendimus ir racionalizuoti mokslinių tyrimų veiksmus.



Parengta naudoti virtuali aplinka, kurioje per specialiai sukurtą svetainę surenkami duomenys yra integruojami ir anonimizuojami, kad paskui galėtų būti analizuojami naudojant pažangias duomenų vizualizavimo ir analizės priemones. Iš pradžių buvo apdorojami tik donorystės duomenys, bet vėliau analizės mastas buvo išplėstas, kad apimtų visą su kraujo plazma susijusį procesą – nuo donorystės iki paciento (klinikinio tyrimo).



Savivaldybei kilo sudėtingas uždavinys susieti skirtingus miesto duomenų rinkinius. Jos tikslas buvo perprasti žmonių judėjimo po skirtingus rajonus dėsningumus pagal laiką. Tokia duomenų analizė buvo labai svarbi siekiant gauti įžvalgų apie judumo tendencijas ir pertvarkyti paslaugas taip, kad būtų veiksmingai remiama kultūrinė veikla ir renginiai.



Prognozuojamoji, aprašomoji ir laiko eilučių analizė, per kurią analizuoti keli prieš COVID-19 pandemiją, per ją ir po jos surinkti duomenų rinkiniai, apimantys, pavyzdžiui, viešųjų belaidžių vietinių tinklų jutiklių duomenis, automobilių stovėjimo duomenis ir georeferencinius žmonių (turistų) judėjimo duomenis.

Paraiškų teikimo tvarka: greitas ir paprastas procesas



Susipažinkite su BDTI paslauga mūsų [svetainėje](#)



» Pasitelkdami mūsų priemonę „[BDTI Canva](#)“, apibūdinkite savo duomenų analizės naudojimo atvejį ir tada užpildykite šabloninę prašymo formą (žr. [svetainę](#)).



» Pateikite savo (šabloninį) prašymą dėl bandomojo projekto e. pašto adresu

EC-BDTI-PILOTS@ec.europa.eu



» Susitikite su mumis nuodugniau aptarti savo naudojimo atvejo



» Bandomasis projektas patvirtinamas, jei:
jis yra naudingas ir gali būti atliktas per 6 mėnesius ir jei turima pakankamai išteklių (įgūdžių, darbuotojų, duomenų)



» Jums skirta testavimo aplinka sukurta



» Galite pradėti bandymus ir gauti naudos!

Ačiū už dėmesį!



Dirbate viešojo administravimo institucijoje, kuriai reikalinga duomenų analizės infrastruktūra? Susisiekite su mumis adresu



EC-BDTI-PILOTS@ec.europa.eu

[BDTI svetainė](#)



[BDTI puslapis platformoje „Joinup“](#)



[BDTI naujienlaiškis](#)



Nuorodos ir ištekliai

- <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>