# Instructions for the live webinar:

This is an interactive session. There will be time for questions and answer throughout the presentation via Slido. We hope you will share your views.

Click on *Connect audio* to hear the presenters but please mute your microphones when you are not speaking.

Please note that this webinar will be recorded.

CEF Digital
Connecting Europe

# Welcome to the live BDTI webinar

**Agenda for today**

**1** Introduction to the CEF (*3 min*)

**2** BDTI in a nutshell (*10 min*)

**3** BDTI Service Offering and how to get started (*10 min*)

**4** Municipality of Milan Pilot Showcase (*10 min*)

**5** City of Florence Pilot Showcase (*10 min*)

**6** BDTI Service Architecture *(5 min)*

**7** Security & Governance (*5 min*)

**8** Q&A section (*10 min*)

**BDTI**

# 01

## Introduction to the CEF

CEF Digital
Connecting
Europe

# How can CEF Digital help you

Our mission is to increase the interconnection between trans-European networks, by helping public administrations, businesses and citizens fully benefit from what digital has to offer.

**We promote the adoption of common digital standards by:**

- Providing **technical support** to Public Administrations and businesses in their digital transition;

- Helping them develop **secure, interoperable digital services**;

- Providing funding to projects that can contribute to a more connected Europe;

This support package comes in the form of **Building Blocks**.

CEF Digital
Connecting
Europe

# Meet the Building Blocks

Find the one you need

**Big Data Test Infrastructure**

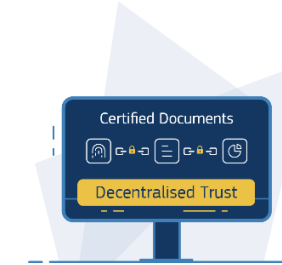Explore and experiment with big data for improved performance and decision making.

**Context Broker**

Gather, manage and share context data, in real-time, throughout Europe

**eArchiving**

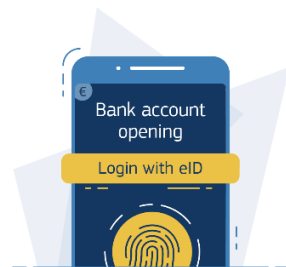Facilitates the preservation, migration, reuse and trust of your data.

**European Blockchain Services Infrastructure**

Harness the power of a Europeanwide network of blockchain services, increasing trust through data security, privacy and transparency..

**eInvoicing**

Promote the implementation of the European standard for electronic invoicing across border

**eID**

Allow citizens to prove who they are across borders, making it easier to access online services in another EU Member State

**eDelivery**

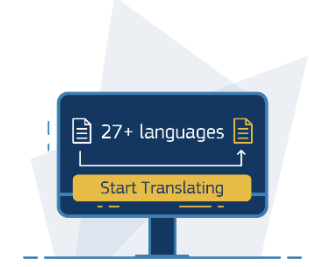Exchange online data and documents reliably and securely

**eSignature**

Create and verify electronic signatures between businesses and EU citizens

**Once Only Principle**

Re-use data held by Public Administrations

**eTranslation**

Use machine translation to translate your documents and web content into any official EU language, and many more

CEF Digital
Connecting
Europe

# How to use a Building Block?

There are 3 options: **buy**, **build** or **reuse** and you can always **co-develop** your solution with other parties.

**Buy**

Buy a compliant, interoperable solution from the market.

**Build**

Build an EU-compliant solution from scratch based on Building Block standards. We'll help you test it for compliance and interoperability.

**Reuse**

Reuse a sample software available through CEF Digital.

Whatever you choose, the relevant CEF Digital team will support you in implementing the Building Block into your project.

CEF Digital
Connecting
Europe

# Why choose CEF Digital

Here are some of the benefits of using a Building Block in your digital project.

## Faster

- The Building Blocks are mature, **ready-to-deploy** digital solutions;
- **This saves you time and money**;
- The CEF Digital website has details on a **range of services available** to help your Building Block implementation.

## Safer

- The Building Blocks are based on open European standards, so you **avoid vendor lockin**;
- Our Building Block team ensures your project is **EU-compliant and interoperable**.

## Future-proof

- Access a **larger market** thanks to interconnectivity
- Building Blocks **can be used in any digital European project**;
- Building Blocks are based on standards and ensure you comply with EU \regulation.

## Funding

- You can visit INEA's website for more information on how to apply, or you can contact the Building Block's onboarding manager.
- Up to 75% of your costs are eligible for funding.

**BDTI**

# 02

# BDTI in a nutshell

Andrea Biancini

Project Manager

DG DIGIT

# What is the Big Data Test Infrastructure

The **Big Data Test Infrastructure** will provide a set of **data and analytics services**, from infrastructure, tools and stakeholder onboarding services, allowing European public organisations to **experiment with Big Data technologies** and move towards **data-driven decision making**

# BDTI Initiative drivers

*Problem - Solution*

**Start**

Adoption of Big Data technologies

**Experiment**

With Big Data sources and datasets

**Share**

Share of Big Data artefacts

**Lack of Big Data technologies**

*Facilitate the prototyping and launching of pilot*

**Lack of Big Data skills**

*Facilitate Big Data knowledge in public sector*

**Data sharing among public organisations is not yet a common practice**

*Provide built-in connectors/APIs and foster the sharing of data sources to better support policy-making*
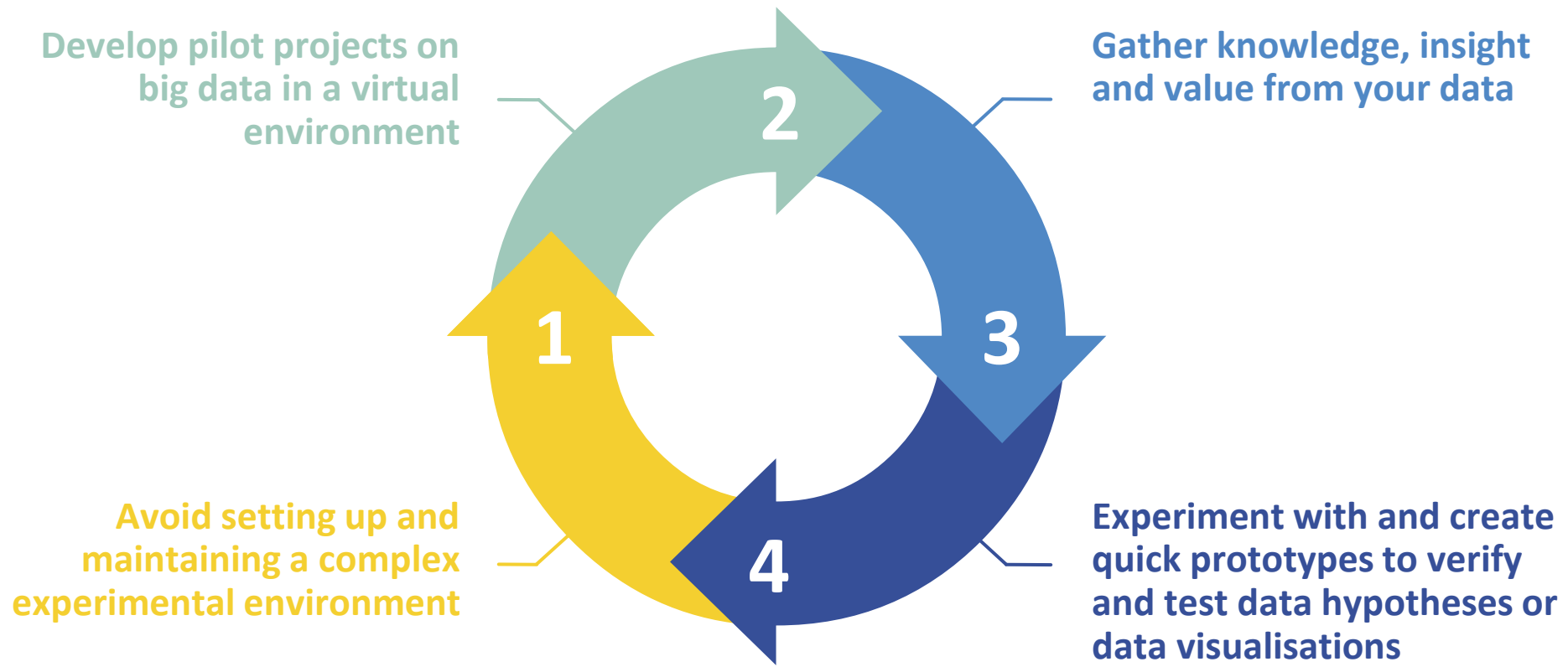
**Risk of replicating the efforts by implementing similar projects**

*Support public organisations through the creation of a Big Data community for the sharing of good practices, pilot outcomes, etc.*

CEF Digital
Connecting
Europe

# Is BDTI for me?

**Yes**, if you need to experiment with big data in a safe environment.

## What can we help you achieve?

**Develop pilot projects on big data in a virtual environment**

**2**

**Gather knowledge, insight and value from your data**

**1**

**3**

**Avoid setting up and maintaining a complex experimental environment**

**4**

**Experiment with and create quick prototypes to verify and test data hypotheses or data visualisations**

CEF Digital
Connecting
Europe

# BDTI Use Cases (1/2)

**Descriptive Analysis**

Use of statistics to quantitatively describe features of a collection of information

**Social Media Analysis**

Gather and analyse data from social media to improve business decisions

**Time-series Analysis**

Analyse time series data in order to extract meaningful statistics and other data characteristics

**Covid 19 Analysis**

Analyse Covid 19 information extracted from a dataset collecting data taken from medical papers

**Predictive Analysis**

Use statistical techniques that analyse current and historical facts to make predictions about future or unknown events
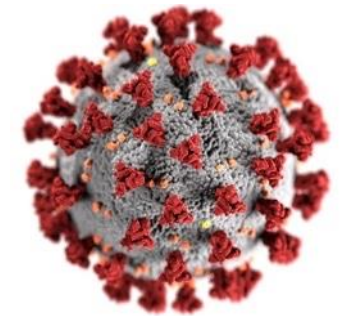
**Network Analysis**

Investigate any structures through the use of network and graph theories

**Text Analysis**

Use natural language processing to analyse unstructured text data, to derive pattern and trends

13

# BDTI Use Cases (2/2)

**Web Analysis (Scraping/Monitoring)**
Gathering information from websites, involving data scraping (using both or web-crawler) and data parsing to extract un-organised web data, as well as data from API's, into manageable format.

**Population/Customer segmentation**
Dividing a broad population into sub-groups of consumers based on shared characteristics such as common needs, interests, similar lifestyles or even similar demographic profiles.

**Route-Traceability/Flow monitoring**
Tracking and detection of objects through the use of sensors (e.g. GPS, mobile phone signals, road cameras) or any other types of data usable for this purpose.

**Image Processing**
Computational operations using any form of signal processing for which the input is an image, a series of images, or frames of a video; output of image processing may be either an image or a set of characteristics / parameters related to the image.
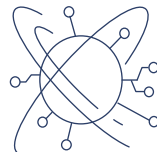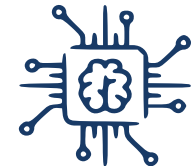
**IoT & Smart City**
Gathering relevant information in IoT environments, in a Smart City context.

**IoT Security**
Safeguarding connected IoT devices and networks, since security often has not been considered in IoT product design.

**Applying Bioinformatics to genetic data**
The use of computational biology, applying data science techniques to understanding/organising biological information and analysing genetic data

CEF Digital
Connecting
Europe

# Main Benefits

**Interoperable** environments and tools that use open source technologies, ensuring their integration with other systems

High **performance** due to an environment architecture that easily scales resources needed for dealing with big data.

**Scalability** due to an environment architecture tailored to the required storage size and computing resources

**Reliability** and **availability** during data transfer and data storage

**Modifiability** for using big data evolving tools and technologies

Necessary **security** implementations for safe data experimentation

**Share and re-use data** across policy domains and organisations

Access to a **knowledge base** and **advisory** for the implementation of pilot project

Access to **insights on best practices** with big data projects and other pilots

Implement a **free of charge** pilot project

15

CEF Digital
Connecting
Europe

# 03

## BDTI Service Offering and how to get started

Francesca Vella

Pilot request Manager

BDTI Team

CEF Digital
Connecting
Europe

# How to get started with BDTI

**Get familiar with BDTI services**

Big Data Test Infrastructure offers a safe environment to experiment with Big Data technologies and to prototype Big Data solutions before deploying them in your own production environment

**Define your data analytics use case**

Build your BDTI use case. The support team is available to guide the users through the process of defining their organisation's use case, as well as to clarify any preliminary questions

**Submit your "BDTI PILOT request"**

Request to use BDTI by submitting an **online form.** The user provides information on the online form in order to clarify the pilot objective (general summary, short description and any supporting evidence)

**Elaborate your business and technical need**

The functional team will work with you on the elaboration of your business case to ensure that it fits within the CEF requirements, while our technical team will design your BDTI test environment which is tailored for your specific technical needs

**Pilot project approval**

The European Commission will give a final validation to run the pilot project.
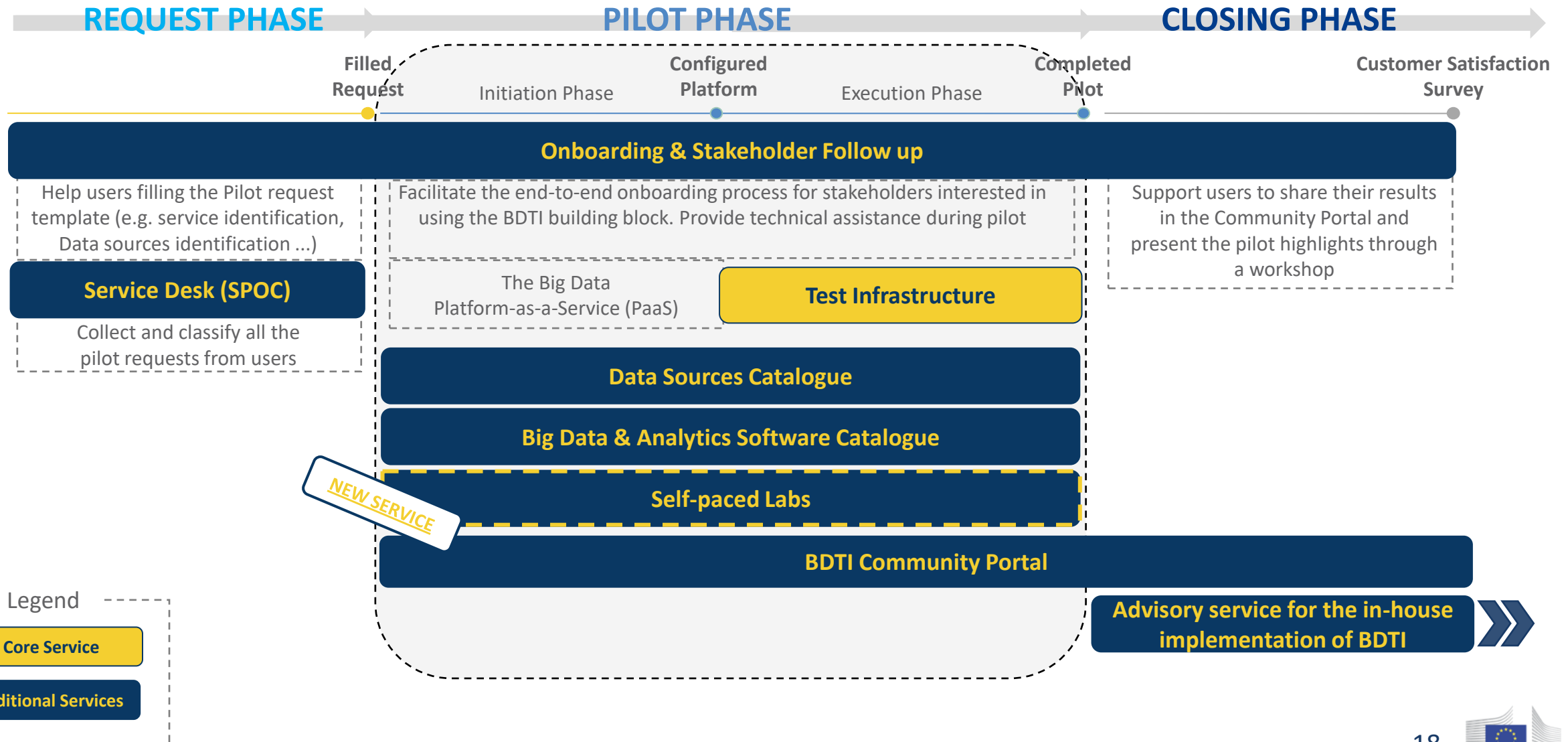
**Test environment set up**

The BDTI technical team will set-up your tailored cloud-based data analytics environment so that you can start with your big data experiments.

**Start the pilot and gain insight**

The users can experiment with data on the test infrastructure provided by BDTI and share any lessons learned on our Community portal

# BDTI Service Offering



**REQUEST PHASE** → **PILOT PHASE** → **CLOSING PHASE**

| Filled Request | Configured Platform | Completed Pilot | Customer Satisfaction Survey |
|---|---|---|---|
| | Initiation Phase | Execution Phase | |

**Onboarding & Stakeholder Follow up**

Help users filling the Pilot request template (e.g. service identification, Data sources identification …)

Facilitate the end-to-end onboarding process for stakeholders interested in using the BDTI building block. Provide technical assistance during pilot

Support users to share their results in the Community Portal and present the pilot highlights through a workshop

**Service Desk (SPOC)**

Collect and classify all the pilot requests from users

The Big Data Platform-as-a-Service (PaaS)

**Test Infrastructure**

**Data Sources Catalogue**

**Big Data & Analytics Software Catalogue**

NEW SERVICE

**Self-paced Labs**

**BDTI Community Portal**

**Advisory service for the in-house implementation of BDTI**

Legend

Core Service

Additional Services

CEF Digital Connecting Europe
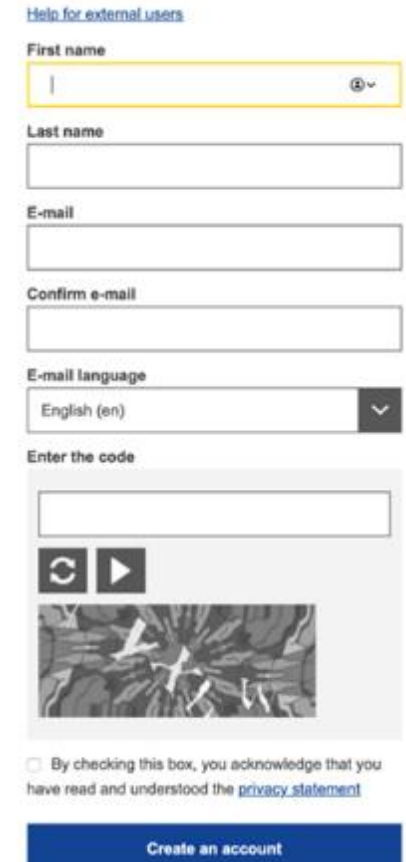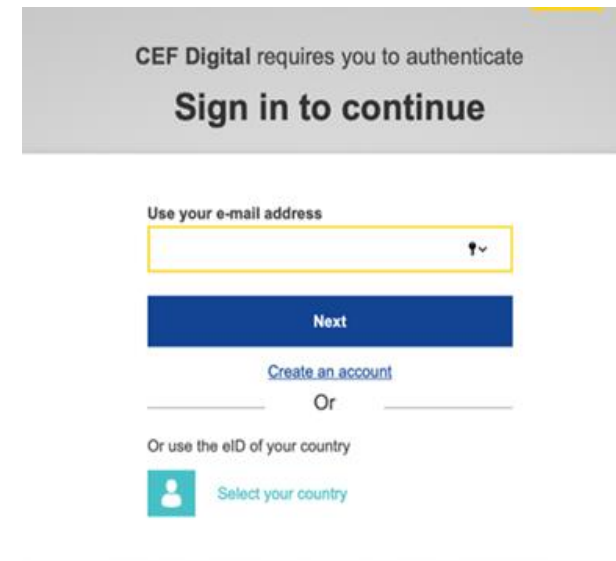
# How to submit a BDTI pilot request (1/2)

**1**

Access to the **BDTI homepage** (https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Big+Data+Test+Infrastructure) and click on the "**REQUEST BDTI PILOT**" section.

**2**

To access the BDTI service desk, you will need to create an **EU login account**, entering your e-mail address and filling in the form with the required information.

# How to submit a BDTI pilot request (2/2)

**3** Once you have filled in the form, you will receive an email to set your EU **account password** and complete your registration.

**4** Now you will be able to access to the BDTI service desk, inserting your password, putting **"Password"** as verification method and clicking on **Sign in**.

**5** Once registration is complete, click on **BDTI PILOT REQUEST**, and fill in the form to submit your request.

# Use cases acceptance criteria

## Business criteria

## Functional criteria

- **Potential users**: Member State or public administration at **national** level, **regional** or **local** level

- **Pilot duration**: 6 months

- **Pilot use cases**: (only use case in scope)

- **Resource usage limit**: based on CEF budget

- **Clear value added**: Business and technical

- **Skills/Maturity level**: adequate skilled resources and/or level of maturity on the big data subject

- Pilot BDTI **geographical distribution/ resource allocation**

- **Clear contact point** for the entire pilot

CEF Digital
Connecting
Europe

# Case studies of ongoing pilots

## Conselleria de Sanitat (CS)

Conselleria de Sanitat (CS) is the Health Public Administration, belonging to the Comunidad Valenciana (CV) Regional Government and it provides health services for all **5.2 million people** in the region. They needed a tool capable of analysing and synthetising the huge quantity of scientific clinical articles coming from different sources: PubMed.gov (more than 30M, and 1M coming every year) and the 100.000 + clinical articles Covid-19 related generated in the first 6 months of pandemic.

## Municipality of Milan

The Municipality of Milan wants to analyse the **movements of people before and during the lock-down period**, in order to predict the future flow of the citizens and assess which areas will be more impacted by the full release of the current restrictions related to Covid-19.The available data collected from Telco operators were used to create a dashboard showing the most crowded areas of the city in the period before, during and after the lockdown.

.
## City of Florence

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 the most and the less crowded neighborhoods during and after the lockdown and about **how services can be relaunched to foster cultural activities and events.**

## How BDTI is helping

BDTI is supporting Concelleria de Sanitat with advanced **data visualization** and **text mining** tools to help **extracting the knowledge contained in the documents**, supporting clinicians and managers in their clinical practices, management process and day-to-day work in fighting the virus.

BDTI is supporting the Municipality of Milan with a scalable virtual environment capable to perform **descriptive, predictive and time-series analysis**. Inside the platform, Telco data can be correlated with on premises data about commercial activities, schools and universities to analyse the future flows of the citizens.

BDTI is supporting the City of Florence with predictive, descriptive and time-series analysis on multiple datasets collected **before, during and after the Covid-19 pandemic** such as: public wifi sensors, parking and geo-referenced data of people movements (i.e. tourists).

# Want to learn more?

If you want to learn more about BDTI pilots, please visit **BDTI Community Portal**

https://ec.europa.eu/cefdigital/wiki/display/BDUC/Big+Data+User+Community+Home



Building a Renewable Energetic Community using data analytics

The increasing decentralisation of energy production is bringing both general changes and challenges to the conventional electricity system. The Municipality of Miglierina needs to optimise energy..

Improving eProcurement through open source software solution

The aim of the eProcurement pilot is to develop an open source software solution to help Member States comply with the einvoice and the public procurement strategic objectives..

Experiments with big data and new ways of collecting statistics

The Eurostat together with the NSIs promoted a challenge on how to modernise statistics with automated data collection and more accurate indicators to better support policy decisions..

Building a reference frame of maritime ships through AIS

The aim of this pilot is to use big data on geo-positioning of ships in order to firstly improve the quality and internal comparability of existing statistics and secondly to produce new..

Producing statistical estimates of Online Job Vacancies

The aim of the project was to set up a pan-European tool for gathering, analysing and presenting the data for all 28 countries. OJV are considered as the source of valuable info..

Improving public procurement processes with data analysis

The objective of the pilot is to develop a value chain from data collection and archiving to data preparation and submission for different statistics and analyses. The value chain will..

# Q&A time

# 04

# Municipality of Milan Pilot Showcase

Milano

Elisa Bellistri

SIAD Management Instructor,
Analytics Data Science Unit

Municipality of Milan

CEF Digital
Connecting
Europe

# Agenda

- Data governance in the city of Milan

- Beyond the Covid 19 emergency

- Use of BDTI Platform
    - First part results
    - Second part aims and preliminary data

- Next steps
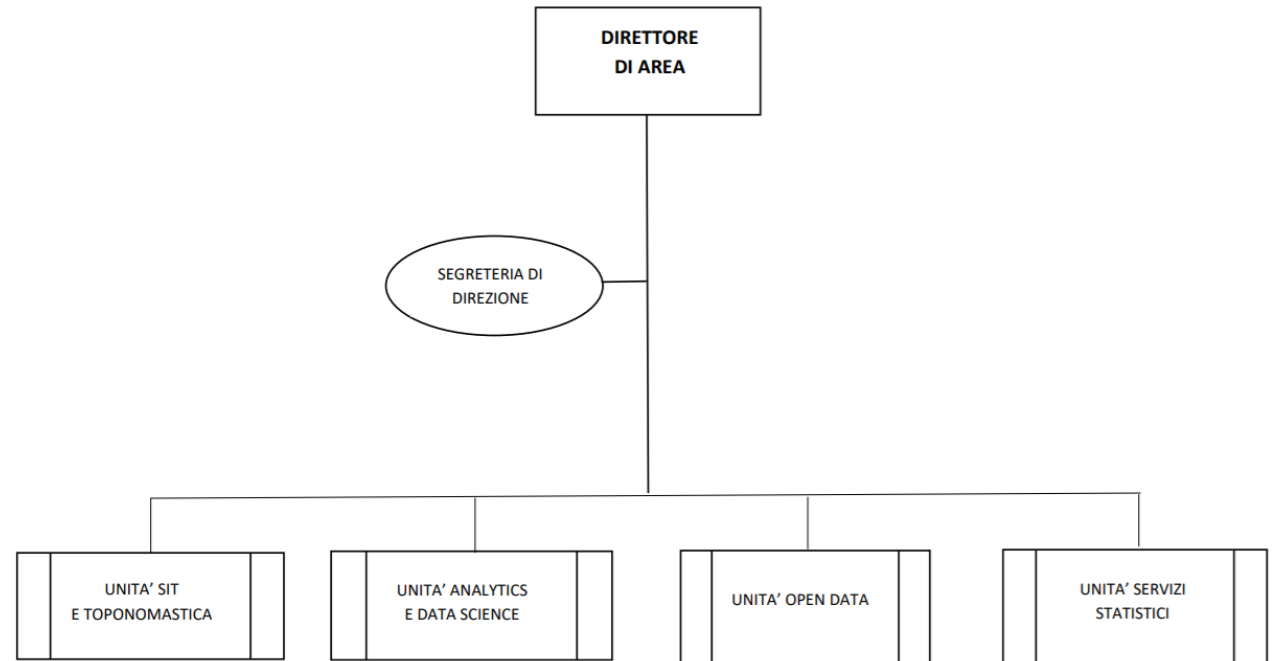
# Who we are

**Data Management and Integration Division**

- Four units composing the division

**Analytics and data science unit**
Main responsibilities

- Managing data and metadata, data governance and support to data driven decisional process, oriented to citizens and city users

- Support to the Chief Technology Officer for the management and analysis of big data

AREA GESTIONE E INTEGRAZIONE DATI

DIRETTORE DI AREA

SEGRETERIA DI DIREZIONE

UNITA' SIT E TOPONOMASTICA

UNITA' ANALYTICS E DATA SCIENCE

UNITA' OPEN DATA

UNITA' SERVIZI STATISTICI

# Beyond Covid-19 emergency

Measures to contrast the pandemic emergency considerably changed Milan city configuration and habits of the people.

The pandemic-period **organization of work, education and commercial activities** suggested a new way to redefine the use of roads and public spaces, to increase non-polluting mobility (walking, cycling, soft mobility) and to develop areas that will allow commercial, recreational, cultural, and sporting developments, while respecting the appropriate physical (but not social!) distances.

Prospecting a closer and closer resolution of the pandemic period, what remains is a renovated idea of the city, where probably some habits definitely changed.

A huge impact of the pandemic period was in particular related to:

- **Traffic and mobility**

- **Open spaces, parks, open air activities**

# New green habits and projects

With bars, restaurants and other spaces closed, a lot of people poured on public parks and gardens.

On the other hand, new squares and public places are planned to increase the green space amount of the city.

# Flow of people during the ongoing pandemic period

Through the collaboration with two of the main Telco companies in Italy, we were able to visualize the people movements within and outside the city during the first lockdown period.
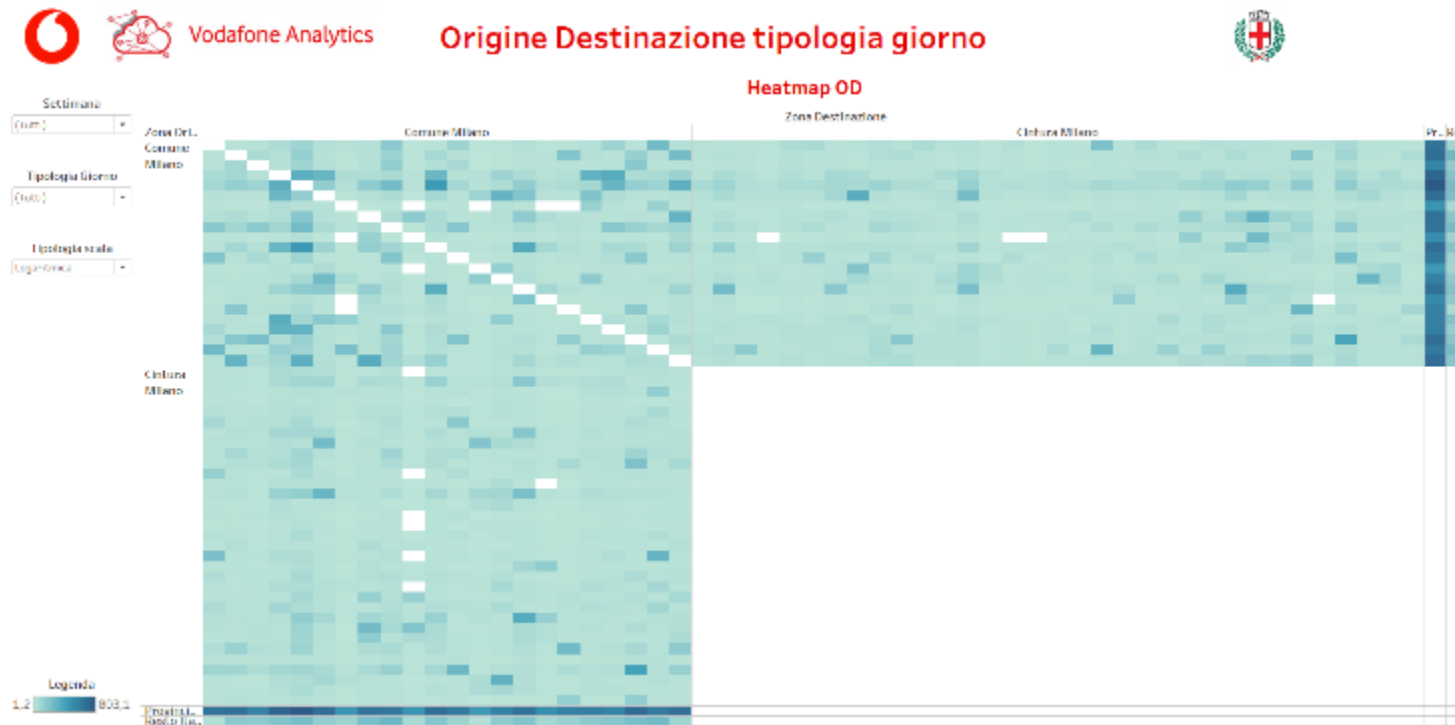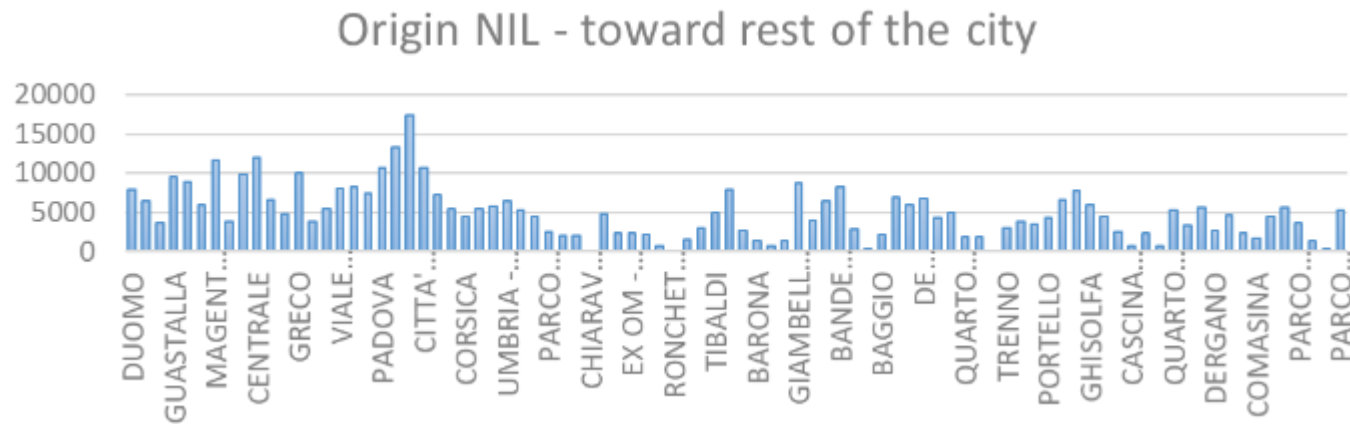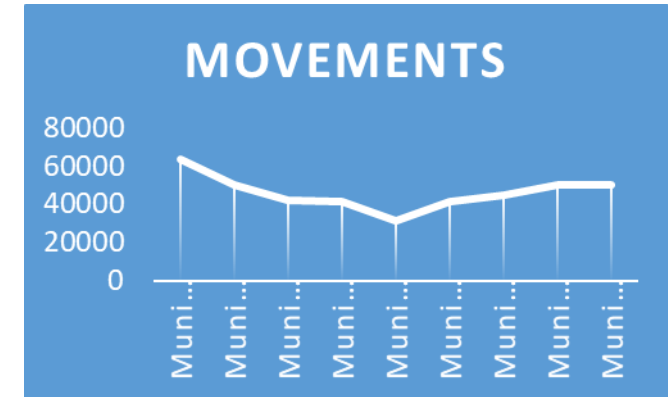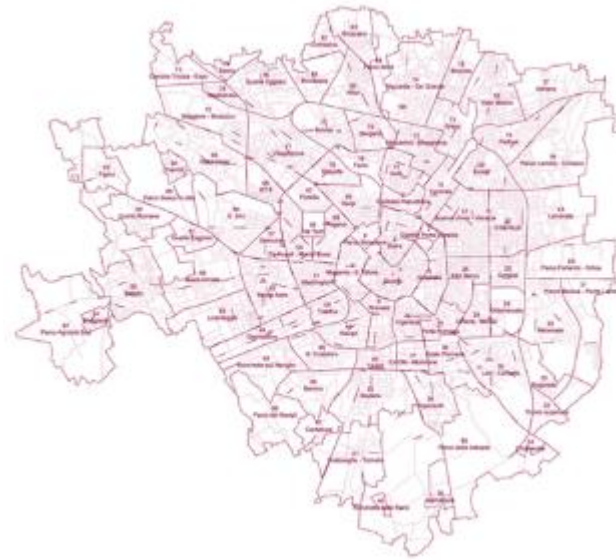
# First part of BDTI project

In the first part of the project, we explored the possibility to integrate different dataset shaped on different granularity of the city zones.

The aim of this part was to perform a quality assessment on the available data and test the possibility of integration between different dataset.





Origin NIL - toward rest of the city

# Evolution of BDTI project
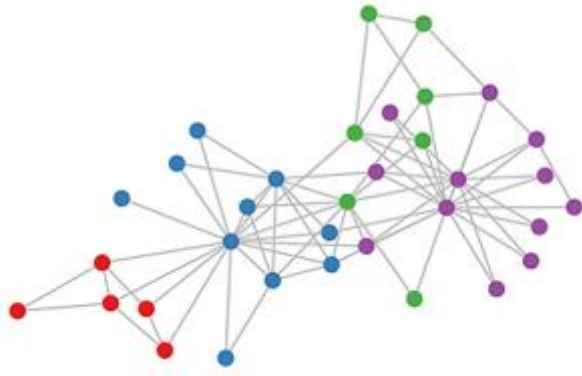
In the following of the project, the aim is to identify specific use cases of the city, in order to populate the integration algorithm with both static parameters and dynamic dataset.

The final goal is to develop a machine learning algorithm that not only calculates the correlation between the different variables, but also would be able to make forecast of the dynamic part of the dataset.
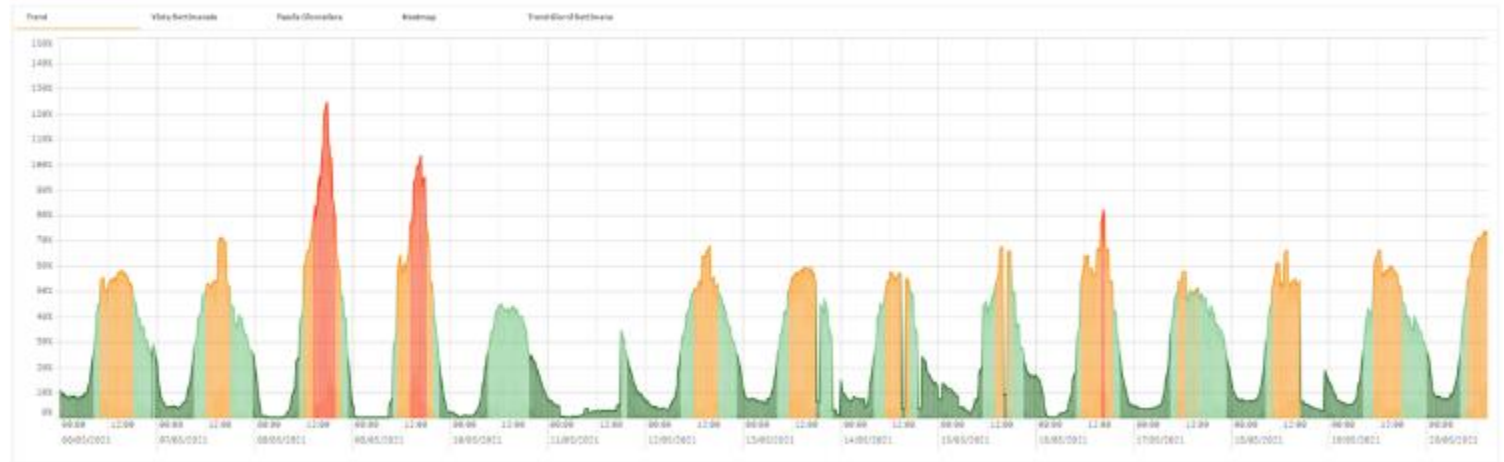
# BDTI opportunities

- Experimenting new ways to approach data, in particular experiment ML algorithm.

- Knowing how other municipalities or groups use the platform. Develop a synergic approach also through partnership with other municipalities (collaboration with the municipality of Florence)

- Sharing knowledge and ideas with other BDTI partners, in order to improve new methodologies of data analysis, using the state of art in the big data analysis tools.

# Next Steps

- Construction of a ML predictive model, based on integration of different dataset

- Developing of advanced tools for data integration and management

- Analysis results sharing with other groups and municipalities

**BDTI**

# 05

# City of Florence Pilot Showcase



Fabio Pazzaglia

Ph. Data Engineer/Scientist

Municipality of Florence

Emanuele Geri

Information Systems Directorate,
Open Data, SIT, Chief Data

Municipality of Florence

# Agenda

- Introduction

- Motivation

- Use of BDTI Platform
  - o Analysis
  - o Results
  - o Data-driven decision

- Next steps

Over the  years, the Municipality of Florence has been very active in collecting data coming from the City and during the last years (since 2015),  from the Smart City.

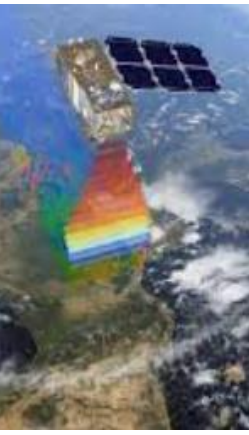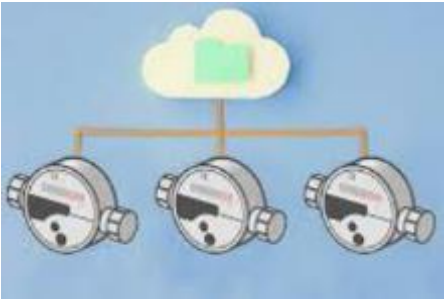GOAL: make the most from the various types of data, extract value from them

Build/share/compare use cases with other cities

Have a powerful platform, scalable and up to date

A team of experts to support and discuss ideas

… that is BDTI

# Introduction



… and more…

The Pandemic has been a period for further specific analyzes that *found interesting work material in the data of the smart city* that has allowed us conduct advanced analysis on the impact of restrictions on people, and more…

# Motivation

The goal of the pilot was to use some significant city data to monitor, analyze and make assumptions over the pandemic evolution.

## FOCUS ON

### MAIN TOPICS

- PEOPLE
- VEHICLE
- TOURISM

### DATA SOURCES

- Outdoor **WiFi** Connections
- **Parking** lots occupancy trend
- Vehicle **traffic**
- Nights in Florence

This is an ongoing project yet. The project aimed at describing the city at all stages of the pandemic event, including a recovery phase that has not been reached yet.

# Analysis

In particular:

## People behaviour

People presence and preferred city centre
places through WiFi data analysis
Trend and PY comparison

## Vehicle trend

Variations about Indoor car parking lots
occupancy trend and vehicle  traffic flow
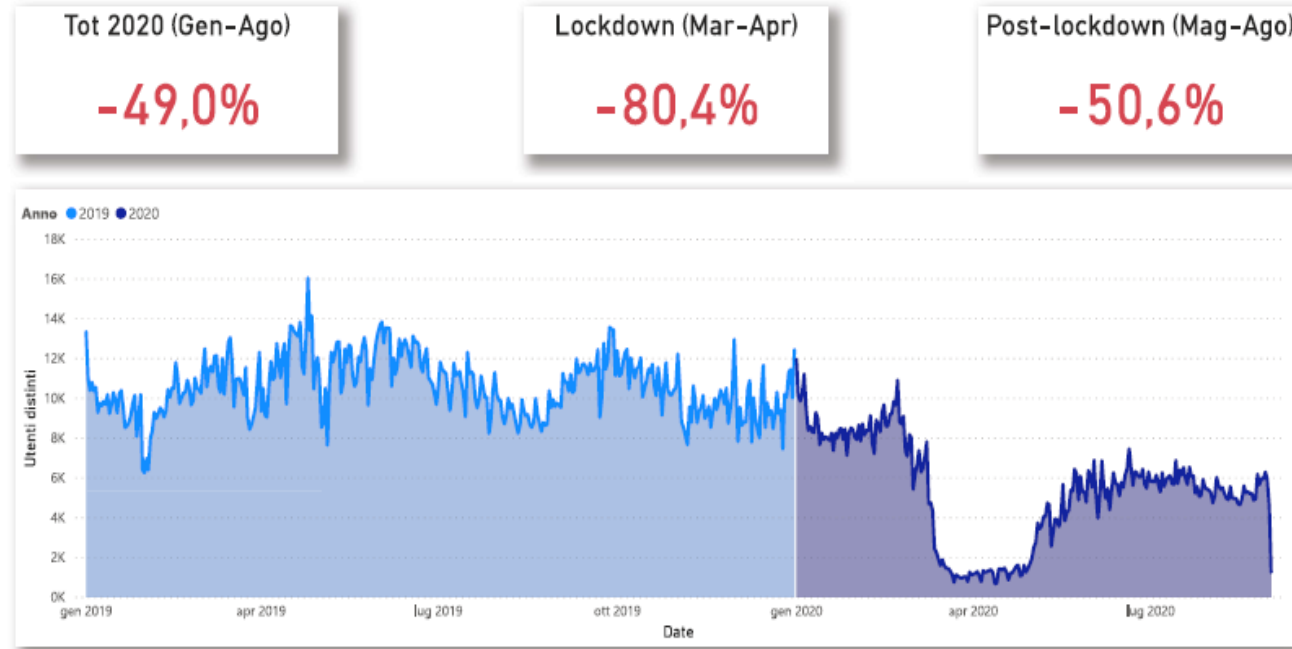inside the LTZ

## Tourists and venues

Nights spent in Florence, by structure and
type of accomodation.
2020 vs 2019 comparison.

# Results

## PEOPLE

| Tot 2020 (Gen-Ago) | Lockdown (Mar-Apr) | Post-lockdown (Mag-Ago) |
|---|---|---|
| -49,0% | -80,4% | -50,6% |



## VEHICLES

| park_name | gennaio | febbraio | marzo | aprile | maggio | giugno | luglio | agosto | settembre | ottobre | novembre | dicembre |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Alberti | -14,87% | 18,94% | 25,37% | 37,88% | 13,43% | -1,23% | 0,75% | -1,87% | -13,42% | -13,91% | -7,90% | -18,21% |
| Parterre | 11,08% | 6,48% | -1,70% | -11,34% | -12,82% | -5,37% | -2,46% | 4,79% | 6,95% | 9,72% | -15,92% | -0,04% |
| Oltrarno | -2,61% | 24,52% | -0,39% | -7,29% | -18,67% | -10,86% | -10,21% | 2,50% | 3,50% | -4,53% | -19,96% | -14,81% |
| SMN | 2,89% | -1,40% | -16,08% | -24,05% | -17,52% | -13,80% | -13,94% | -1,89% | -0,76% | 1,85% | -14,22% | -18,78% |
| Meyer | -8,85% | -1,33% | -27,72% | -31,07% | -21,69% | -23,44% | -14,74% | -77,40% | 0,04% | -4,97% | -8,42% | 51,30% |
| Binario16 | 42,28% | 55,49% | -20,03% | -31,88% | -28,93% | -17,03% | -19,24% | 1,33% | -14,25% | -30,11% | -69,38% | -57,82% |
| PortaalPrato | 6,09% | 19,15% | -36,55% | -31,62% | -40,35% | -38,80% | -28,93% | -9,04% | -11,51% | -21,88% | -31,39% | -31,83% |
| Beccaria | 9,36% | 21,07% | -38,47% | -59,42% | -48,09% | -20,26% | -17,99% | -23,01% | -4,34% | -18,84% | -41,59% | -42,95% |
| Careggi | 29,45% | 23,31% | -50,11% | -62,04% | -51,87% | -22,71% | -15,50% | -31,82% | 14,99% | -55,75% | -46,59% | -47,46% |
| PalGiustizia | -32,28% | -25,87% | 7,80% | 17,58% | 4,07% | -37,88% | -37,80% | -59,00% | -49,97% | -31,17% | -21,65% | -23,23% |
| SAmbrogio | 2,37% | 9,73% | -50,61% | -67,24% | -52,23% | -18,97% | -33,56% | -3,84% | -11,41% | -14,97% | -48,10% | -43,48% |
| SLorenzo | -27,33% | -3,31% | -73,64% | -89,63% | -75,33% | -54,49% | -52,40% | -35,60% | -31,05% | -24,33% | -69,32% | -52,67% |
| FortezzaFiera | -8,22% | 3,53% | -61,10% | -74,11% | -70,61% | -81,88% | -69,10% | -39,58% | -58,90% | -48,94% | -59,52% | -48,62% |

## VENUES

| Typology | Nights (%var over 2019) |
|---|---|
| Apartments | -78,8% |
| Others | -59,7% |
| Hotels | -79,7% |
| B&Bs | -75,7% |
| Hostels | -77,5% |
| Tot | -78,7% |

# Data-driven decision

Starting from the people presence analysis we setup a real-time crowding level warning system, to proactively monitor some crucial city squares and place.

*Area: Piazza Duomo      Period:December 2020*
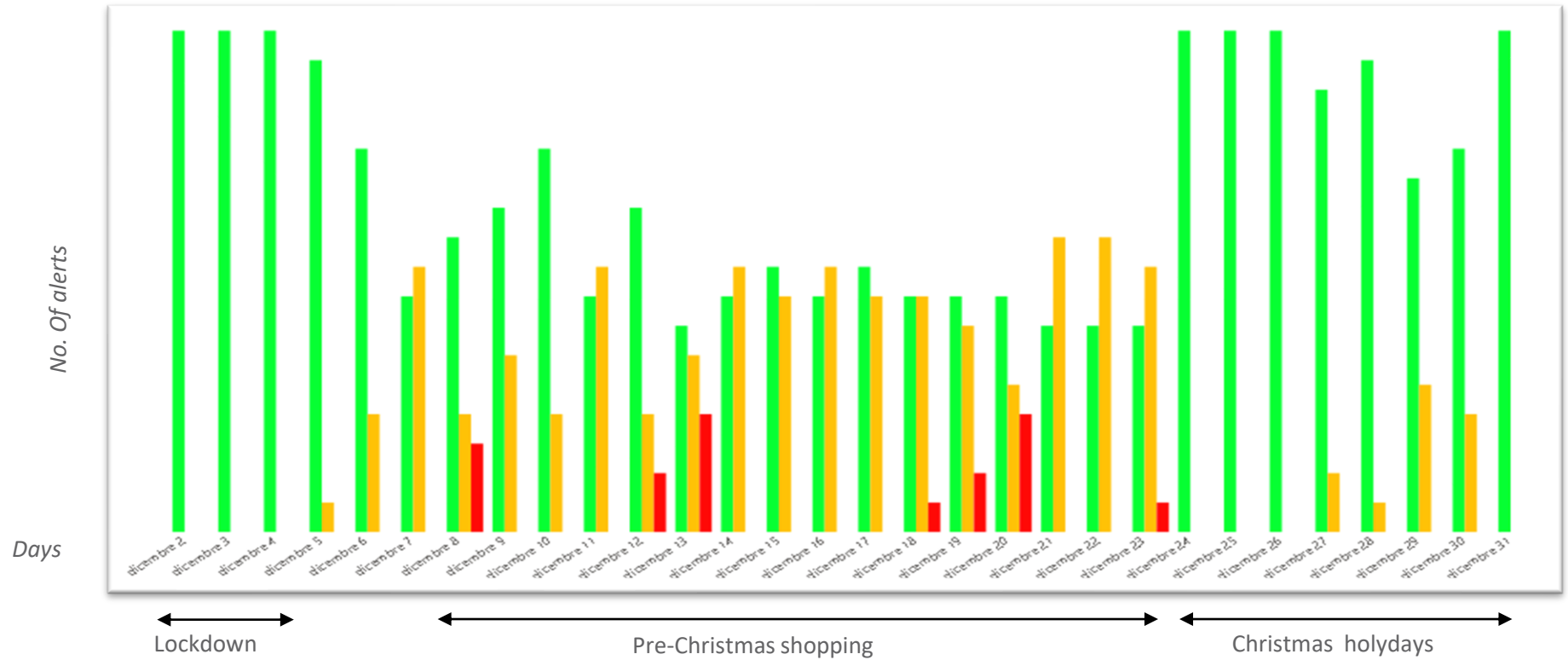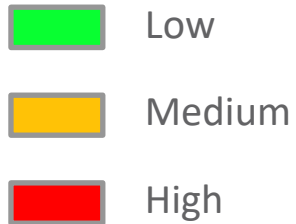
# Next steps

## ML & AI

- switching analysis into prediction
- Using BDTI to experiment Machine Learning techniques applied to our data

## Collaboration and sharing

- Ongoing collaborations with other partners (e.g. Milan ) in order to think and develop new application scenarios

- Join forces to achieve bigger impact results

- Share data and compare results

---

**"THE CITY OF TOMORROW"**

A city where data are turned in useful information in order to:

- Make citizens' life easier

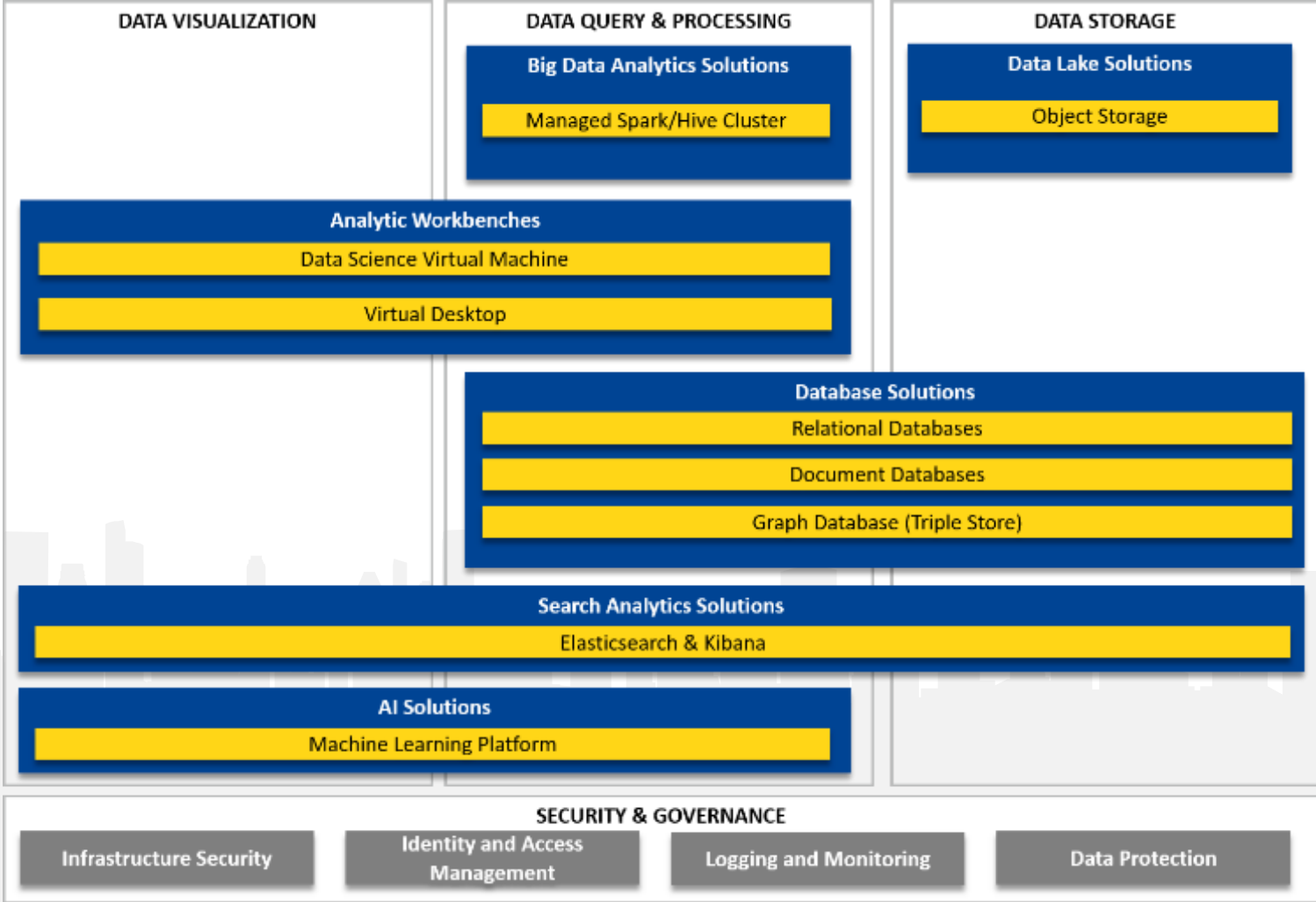- Support industries and the

  growth of local economy

# 06

# BDTI Service Architecture

Ze Wen Wu

Cloud Engineer
BDTI Team

# General BDTI service offering

# Analytic Workbenches

An analytics workbench empowers users with the ability to autonomously produce and publish insights, mainly through self-service data preparation and visual data discovery tools.

## WorkSpaces

Connect to a full-featured Windows Desktop environment through a remote client. Jupyter notebooks, Rstudio, KNIME Analytics, Anaconda are available and can be used for processing, analyzing and visualizing data.

- A Virtual Desktop as your Analytics Workbench, perform ad-hoc analysis with Jupyter notebooks, KNIME Analytics or RStudio.

- Access data from Database & Data Lake solutions

- Store processed data from Database & Data Lake solutions

- Visualize analysis with Jupyter, Rstudio or your favorite plotting library.

## Data Science Virtual Machine

Connect to a full-featured Ubuntu 18.04 LTS Desktop environment through your browser. Jupyter notebooks, RStudio Server or a Terminal session can be accessed securely from your browser.

- A DSVM as your Analytics Workbench, perform ad-hoc analysis with Jupyter notebooks or RStudio.

- Access data from Database & Data Lake solutions

- Store processed data from Database & Data Lake solutions

- Visualize analysis with Jupyter or Rstudio

- Execute jobs on Spark/Hive cluster with Jupyter sparkmagic kernels

# Data Lake Solutions

A Data Lake is a repository of data stored in its raw format. It is the main starting place for self-service analytics.

Typically, Data Lakes store large volumes of information, classified as Big Data.

## Object Storage

Amazon S3 is an object storage solution that stores objects up to 5 TB in buckets. In your pilot you can upload data to your S3 bucket(s) and structure them accordingly in folders and subfolders. You can upload as many objects as you want in Amazon S3, storage capacity will automatically scale.

- Use object storage as your primary storage for raw data (data lake solution)

- Starting point for big data analytics

- Access object storage from other BDTI resources such as a Data Science Virtual Machine or Spark/Hive Cluster
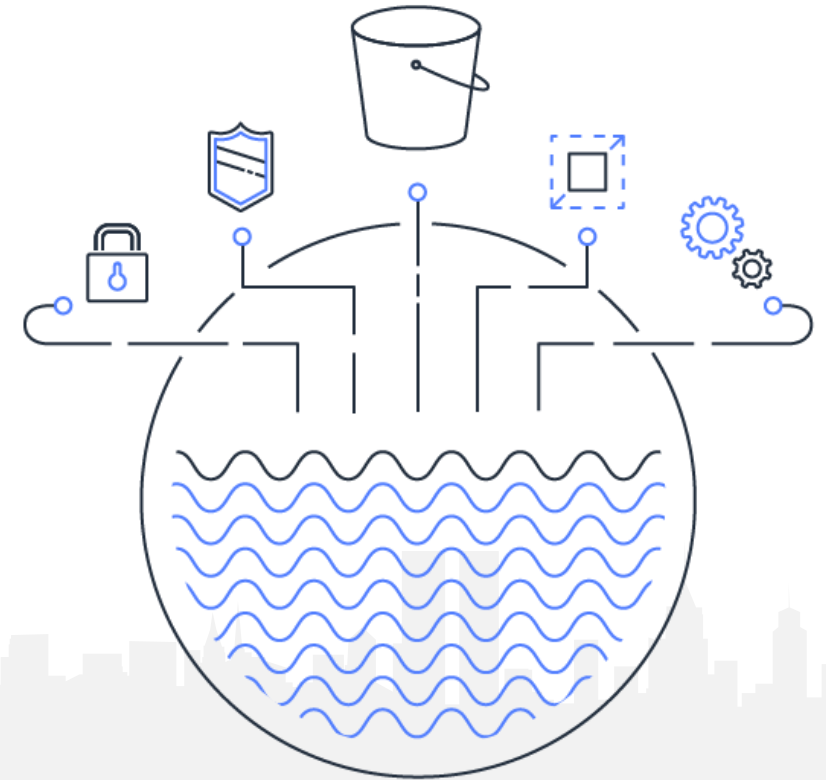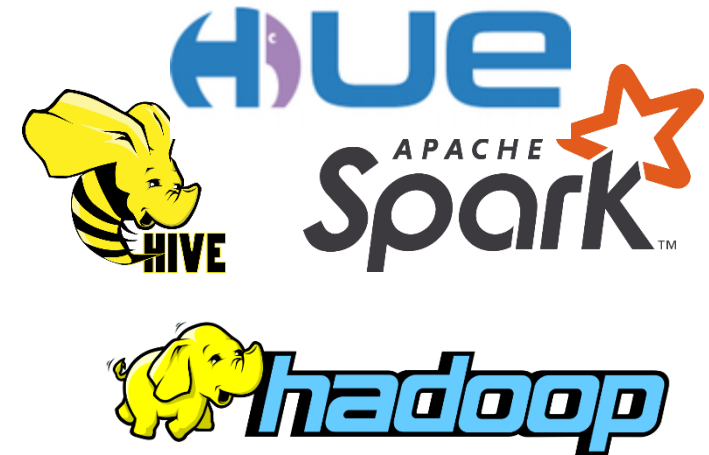
# Big Data Analytics Solutions

Analyzing large data sets requires significant compute capacity that can vary in size based on the amount of input data and the type of analysis.

Specialized tooling is therefore also necessary.

## Managed Spark/Hive Cluster

Using open-source tools such as Apache Spark and Apache Hive coupled with Object Storage enables performing big data analytics on vast amounts of data.

A Spark/Hive cluster is implemented with Amazon EMR which allows leveraging EMRFS, an S3 interface for Hadoop workloads

- Perform big data analytics on large amounts of data

- Leverages the Hadoop ecosystem, an open-source framework for distributed processing and big data analytics

- Execute Spark and Hive jobs on HDFS or Object Storage (Amazon S3)

- Leverage the Spark API for advanced analytics and processing

- Leverage Hive for Big Data warehousing

49

# Database Solutions

Database systems are the most well-known and standardized solutions for data storage and querying. Different flavors are available depending on the use case.

### Relational Database

Uses open-source relational databases such as MySQL and Postgres.

Relational databases are implemented with Amazon RDS.

- Perfect fit for most transactional and analytical processing (OLTP & OLAP cases)
- Perform flexible SQL queries to extract and store data
- Can be scaled depending on the needs

### Document Store

Leverage the NoSQL paradigm to store data as a collection of documents to allow further scaling of your data storage solution.

Document stores are implemented with Amazon DocumentDB.

- NoSQL solution, does not comply with the ACID constraints to allow better performance and scaling for specific use cases.
- Stores data as a colletion of documents.
- Query data with a Mongo-compatible API.

# Search & Analytics Solutions

Search & Analytics solutions provide users with the ability to perform search queries on large amounts of text documents and in addition analyze its content.

## Elasticsearch & Kibana

Use an open-source search engine to perform text search and analytics on large collections of documents. Visualize insights with Kibana.

Elasticsearch & Kibana is implemented with Amazon Elasticsearch Service

- Search and analytics engine for performing real-time search in text-based data.

- Graphical user interface for visualization and querying (Kibana)

- Operational analytics (log analysis).


elasticsearch

# AI Solutions

AI Solutions are specialized tools and frameworks to provide users with the ability to develop and publish artificial intelligence in the form of Natural Language Processing, Machine Learning, Data Mining and Predictive Modeling.

## Machine Learning Platform (h2o.ai)

Use an easy-to-use open-source platform to train machine learning models on your data.

Machine Learning Platform is implemented with h2o.ai

- Supports the most widely used statistical & machine learning algorithms (including gradient boosted machines, generalized linear models, deep learning and more).

- Use H2O Flow, an open-source user interface for H2O. It is a web-based interactive environment that allows you to combine code execution, text, mathematics, plots, and rich media in a single document

# General BDTI service offering
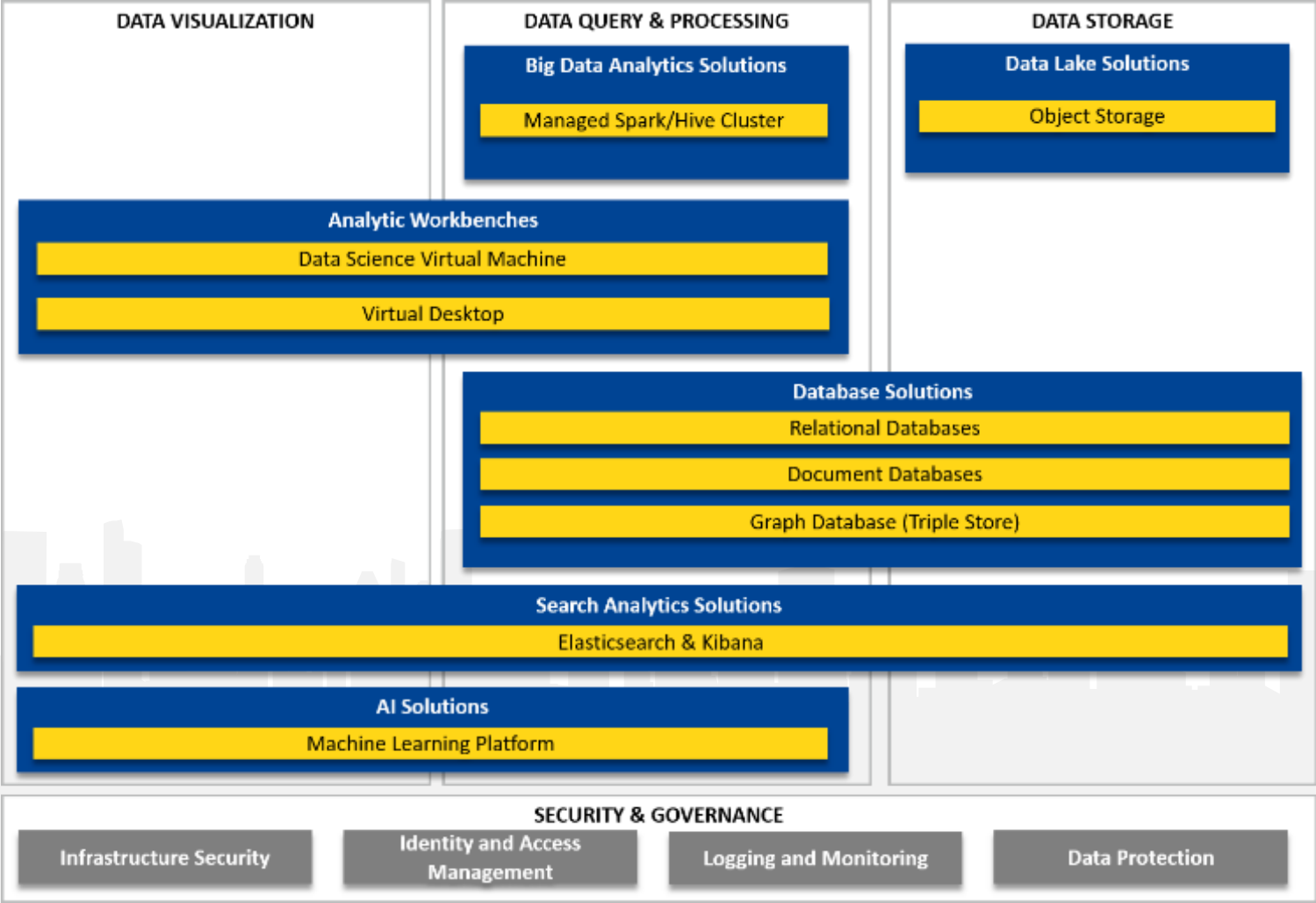
# 07

# Security & Governance

Kasper Rutten

Cloud Engineer
BDTI Team

# Common cloud concerns

## Privacy
How public is public cloud?

## Tech gap
Is it too high-tech for me?

## Security
Is it safe?

### Who else can access my data?
*BDTI keeps all data of your pilot logically separated, in European data centers*

### Can I put my data on the cloud contractually?
*The European Commission has an agreement with cloud vendors that should cover most NDA requirements*

### Is the cloud not difficult to use and understand?
*The BDTI team will abstract most of the technical details away, and provides extensive support during the onboarding process*

### Will I be targeted by cyberattacks?
*The team works closely together with the major cloud vendors to curate a secure environment at all times. Real-time monitoring and clear response protocols allow us to swiftly detect and respond to issues.*

CEF Digital
Connecting
Europe

# BDTI Security & Governance

## Infrastructure Security

BDTI takes care of all operational aspects regarding infrastructure security (network security, disaster recovery, maintenance,…)

## Identity & Access Management

BDTI offers a managed centralized solution for identity and access management.

Users can access all resources of their pilot through one BDTI account.

## Logging & Monitoring

BDTI monitors all pilots for incidents and has logging in place for auditing and traceability.

## Data Protection

BDTI protects all data that is being stored and processed in pilot environments.

All access needs to be authenticated and authorized. Data is encrypted in-transit and at-rest.

CEF Digital
Connecting
Europe

# Q&A time

# Ready to get started?

Reach out to us to learn more!

Visit us at https://ec.europa.eu/cefdigital/bdti

## Follow us on social media

@Connecting_EU

Connecting Europe

Connecting Europe

Drop us a line

CEF-BDTI-SUPPORT@ec.europa.eu

CEF Digital
Connecting
Europe