



DG DIGIT
Directorate-General
for Informatics

DG CNECT
Directorate-General for Communications
Networks, Content and Technology

Big Data Test Infrastructure

Discover how BDTI can support Public Administrations in gathering valuable insights from public sector information

From Hype to Action

March 2nd, starts at 3:00 PM CET



Instructions for the live webinar:



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



Click on *Connect audio* to hear the presenters but please note that your microphones will be muted during the entire session.



Please note that this webinar will be recorded.

Welcome to the live Webinar

Agenda for today

- 01** Introduction to the Digital Europe Programme *(5 min)*
- 02** BDTI in a nutshell *(10 min)*
- 03** BDTI Service Offering and how to get started *(15 min)*
- 04** Pilot Showcase: BDTI for Norwegian Public Procurement *(20 min)*
- 05** BDTI Service Architecture *(20 min)*
- 06** Q&A section *(15 min)*

01

Introduction to the Digital Europe Programme



MARIA CLAUDIA BODINO

BDTI Project Officer

DG DIGIT – EUROPEAN COMMISSION

Introduction to the Digital Europe Programme (DEP)

The **Digital Europe Programme** is a new EU funding programme focused on **bringing digital technology** to businesses, citizens and public administrations to **accelerate the economic recovery and shape the digital transformation of Europe's society and economy.**

The Digital Europe Programme will provide strategic funding in five crucial areas:

- **high performance computing;**
- **artificial intelligence** (cloud, data and AI);
- **cybersecurity;**
- **advanced digital skills;**
- **deployment and wide use of digital technologies.**

- planned overall budget of €7.5 billion



[DIGITAL EUROPE Work Programme 2021-2022](#)

The Digital Europe Programme (DEP) and BDTI

DATA for AI - Public Sector Open Data for AI and Open Data Platform

Objective:

- increase the easy availability, 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.

Work Strand:

- Maintenance and further expansion of the **Big Data Test Infrastructure (BDTI)** Building Block, including the possibility for the public sector to use it for testing **Business-to-Government (B2G)** data sharing collaborations for the **public good**



Context and background

OPEN DATA DIRECTIVE HIGH VALUE DATASETS (July 2019)

The Directive on open data and the re-use of public sector information, also known as the Open Data Directive, entered into force on 16 July 2019, replacing the Public Sector Information (PSI) Directive. It provides common rules for a European market for government-held data and identifies a list of High Value Dataset to be published for **free** and available via **API**

EUROPEAN DATA STRATEGY (February 2020)

The **European strategy for data** aims at creating a **single market for data** that will ensure **Europe's global competitiveness** and **data sovereignty**. Common European data spaces will ensure that more data becomes available for use in the economy and society, while keeping the companies and individuals who generate the data in control.

DATA GOVERNANCE ACT (November 2020)

The European Commission proposed the **Data Governance Act (DGA)** in response to the public consultation on the European Strategy for Data. The EU Data Governance Act is intended to **foster the availability of data** by increasing trust in data intermediaries and strengthening data sharing across the EU and between sectors.

DATA ACT (February 2022)

The new measures complement the Data Governance Regulation proposed in November 2020, the first deliverable of the **European strategy for data**. While the Data Governance Regulation creates the processes and structures to facilitate data, the **Data Act** clarifies **who can create value from data** and **under which conditions**.

02

BDTI in a nutshell



MARIA CLAUDIA BODINO

BDTI Project Officer

DG DIGIT – EUROPEAN COMMISSION

What is the Big Data Test Infrastructure (BDTI)



The Big Data Test Infrastructure:

- **free of charge** service provided by the EC as part of the Digital Europe Programme
- **data analytics stack and services**, based on **open-source tools**
- stakeholder onboarding services allowing EU public organisations to **experiment with Big Data technologies** and move towards **data-driven decision making**.

BDTI Drivers for a data driven public sector

Problem – *Solution: from hype to action*



Lack of Data analytics and Big Data technologies

Facilitate the prototyping and launching of pilot using open-source tools

Lack of Data analytics skills

Facilitate Big Data knowledge in public sector



Data sharing and interoperability 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 Data analytics and Big Data community for the sharing of good practices, pilot outcomes, etc.

Is BDTI for me?

Yes, if you want to move from hype to action, test your ideas and experiment with data analytics open-source tools in a safe environment.

What can we help you achieve?

01

Benefit from **support to manage a complex experimental environment**



02

Develop pilot projects on **public sector information** in a virtual environment with **open-source tools**



03

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



04

Experiment with **open-source tools** and create quick prototypes to **verify** and **test data analytics initiatives**

BDTI some ideas for use cases



DESCRIPTIVE ANALYSIS

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



IOT & SMART CITY

Gathering relevant information in IoT environments, in a Smart City context.



TIME-SERIES ANALYSIS

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



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.

Main benefits



Implement a **free of charge** pilot project for 6 months



Technical support during the entire duration of the pilot



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



Necessary **security** implementations for data experimentation in a safe infrastructure



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

03

BDTI service offering and how to get started



FRANCESCA VELLA
Pilot Request Manager
BDTI team

How BDTI will support you throughout your journey

DISCOVER

The services that BDTI offers and how they leverage insights from data to support the policy-making process and support the onboarding process of stakeholders.

START

To support the set up of the big data pilot and enable users to benefit from advisory, knowledge base and community, and to support the integration of the stakeholder's data. A set of transparent tutorials, named Self-paced-lab, are made available to help non-technical Pilot members to start experimenting with their data in BDTI.

FINISH

To support in implementing in-house BDTI test infrastructure at the end of their pilot projects.



EVALUATE

To help to identify the specific business needs and translate these into requirements; to help to design the BDTI pilot taking into account the techniques, supporting software packages, infrastructure requirements and reporting tools needed; and to help to choose the BDTI offering to use from among open source and commercial solutions (e.g. Data Source Catalogue, Big Data & Analytics Software Catalogue...).

USE

To support the implementation of the big data pilot and to give technical assistance during the pilot.

SHARE

To support users in sharing their results and in presenting the pilot highlights through a workshop (Feedback collection).

How to get started with BDTI



GET FAMILIAR WITH BDTI SERVICES

BDTI offers a safe environment to experiment with public sector information and open-source tools 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 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

DROP US A LINE TO REQUEST A BDTI PILOT

BDTI Functional Mailbox: EC-BDTI-PILOTS@ec.europa.eu

Use cases acceptance criteria



BUSINESS CRITERIA	FUNCTIONAL CRITERIA
<ul style="list-style-type: none">• Potential users: Every EU public administration at national, regional or local level• Clear value added: Business and technical• Clear contact point for the entire pilot	<ul style="list-style-type: none">• Pilot duration: 6 months• Pilot use cases: only use case in scope• Resource usage limit: based on EC budget• Skills/Maturity level: adequate skilled resources in reusing public sector information

DROP US A LINE TO REQUEST A BDTI PILOT

BDTI Functional Mailbox: EC-BDTI-PILOTS@ec.europa.eu

Case studies of completed 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 synthesising 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.



HOW BDTI HELPED

BDTI supported the Conselleria 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.

CONVALESCENT PLASMA DATABASE

The European Blood Alliance (EBA) is working together with the European Commission (DG SANTE, DG CNECT and DG DIGIT) to create and manage an **EU-wide open-access platform** that collects data to support a study on **Covid-19 convalescent plasma therapy**. The aim of the study is to assess in which conditions the convalescent plasma treatment is most effective, in order to take data driven decisions on the therapy and focus the efforts of the research in the most promising directions.



BDTI supported EBA and DG SANTE with a ready-to-use, virtual environment in which **data collected through a custom-built website**, are ingested and anonymized, to be then analysed with advanced data visualization 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**.

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**.



BDTI supported 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).

04

Pilot Showcase: BDTI for Norwegian Public Procurement



PAUL KILLIE

Senior Advisor

Agency for Public and Financial management

Topics covered

- Goal for Public Procurement Datalake
- Datalake Architecture
- POC for Analytics
- BDTI setup and test
- Production setup and development
- Future plan for Concept

Public Procurement



Annually, public procurement amounts to approximately

Euro 50+ billion

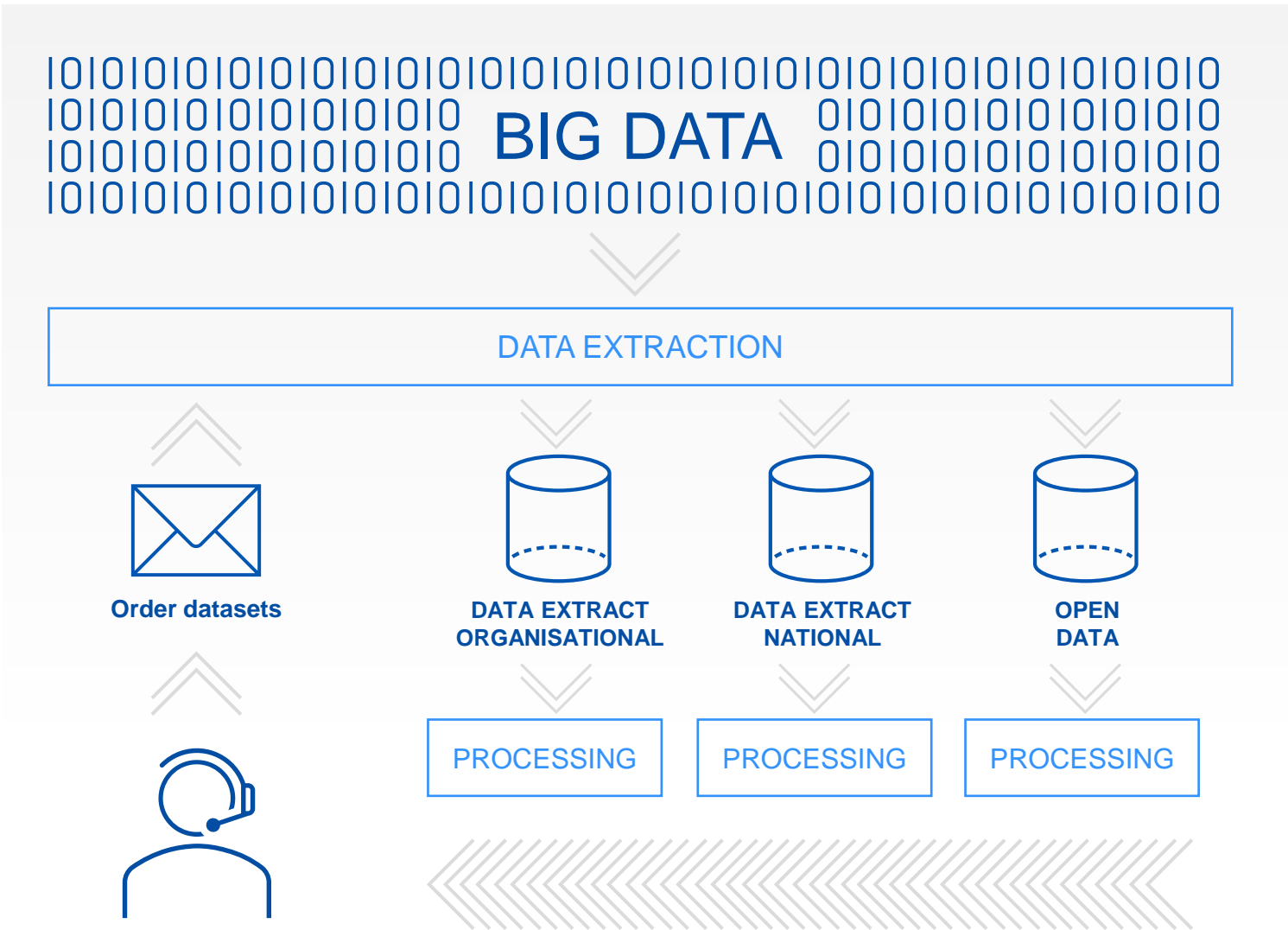
Central government administration spending:

54% Buildings, infrastructure and property

18% Professional services (excl. ICT)

9% ICT (equipment, licenses and services)

Big Data for public sector



ORGANIZATIONAL LEVEL

- Organisational statistics
- Combine with other data sources
- Procurement automation
- Procurement prediction

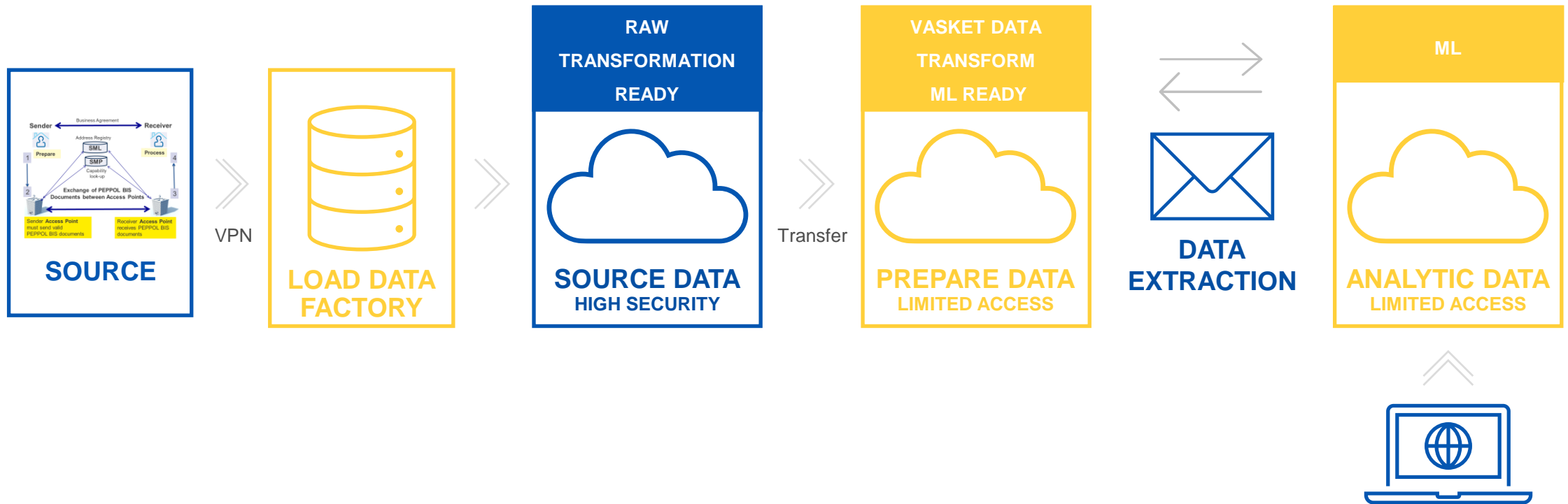
NATIONAL LEVEL

- National statistics
- Benchmarking
- Investigation

OPEN DATA

- Insight in public procurement

Architecture – “whole value chain”



POC 1 – UNSPSC code from Peppol BIS Text line

Developed machine learning models to decide UNSPSC code.

SUCCESS RATE 96%



Codification is done according to product description and «field» code from the vendor.

https://github.com/Fundator/difi_poc1

POC 2 – Environmental data

Hypothesis for algorithm:

- Do we fulfill the environmental requirements in our tenders at a satisfactory level?
- How do we evaluate the environmental requirements in our selection?

CONCLUSION

Label-model missed only 1 tender with environmental specification out of 924 investigated, Success rate **99,9 %**



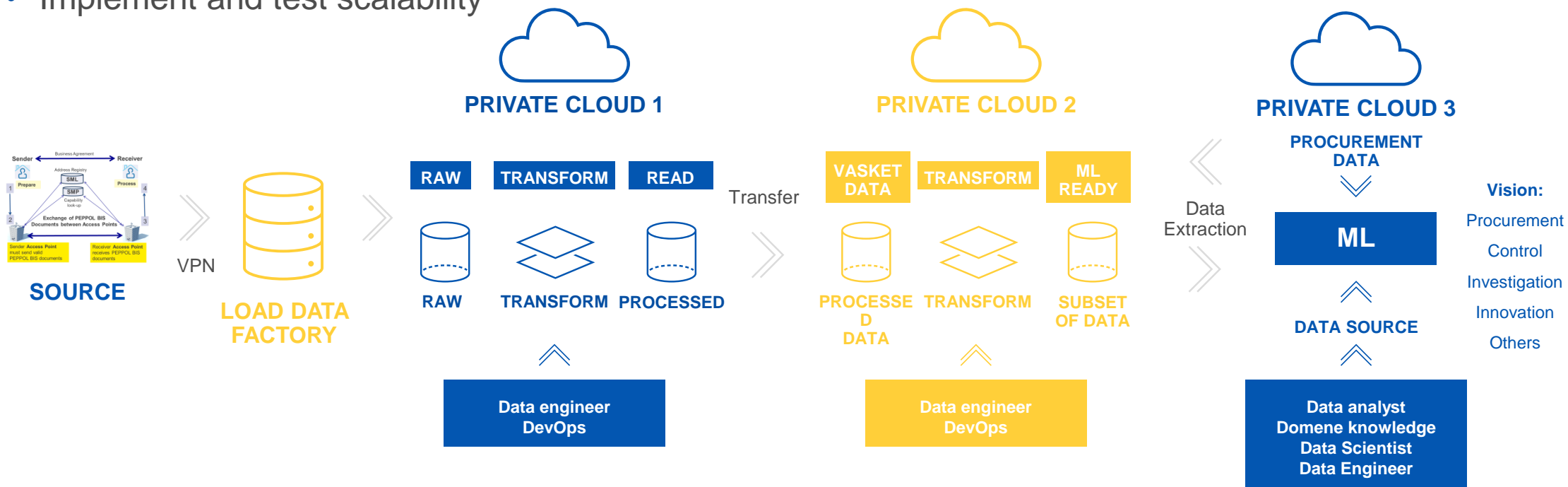
DOMAIN EXPERT



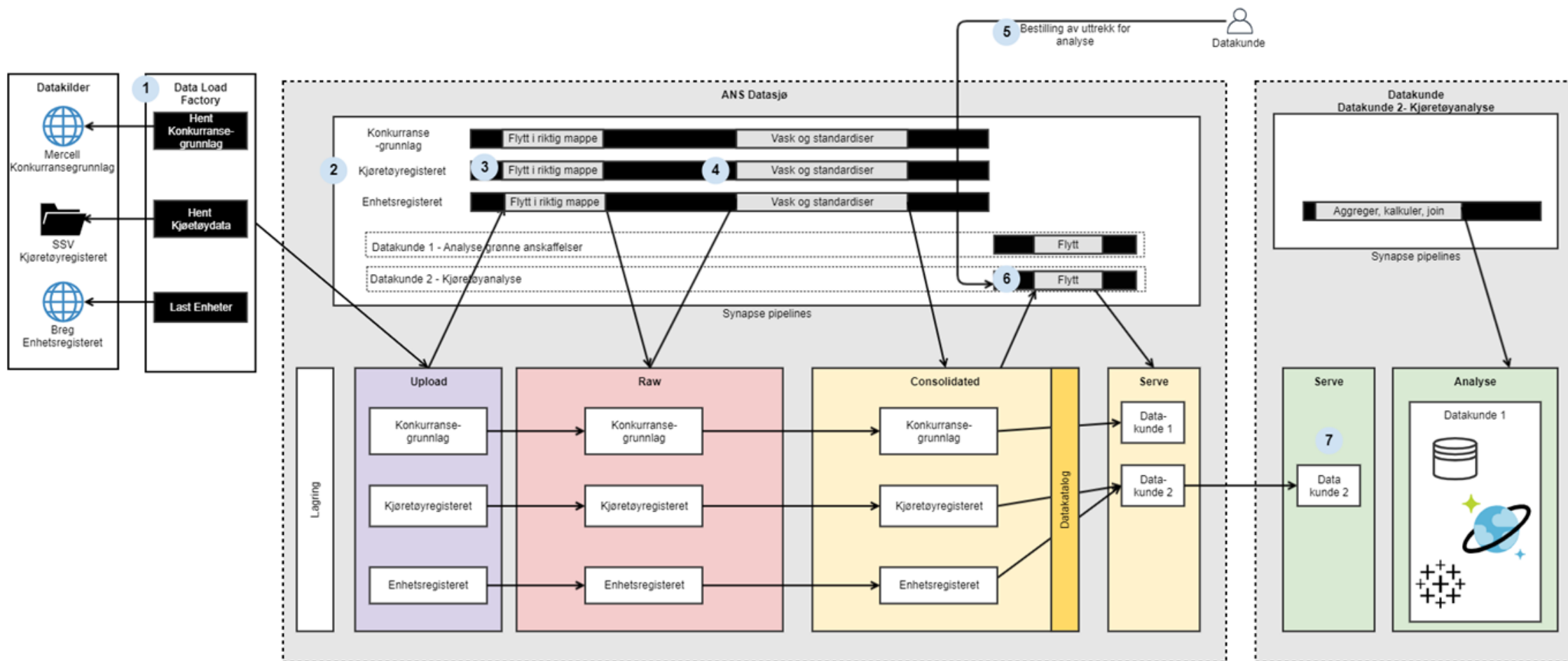
Pilot – from January 2020

EU commission financed through CEF

- Develop and test functionality
- Implement and test scalability



Infrastructure architecture in production



Planning for the future

- Adding more sources for data
- Build a «washing machine» for GDPR
- Use the datalake as hub for original data «permafrost»
- Build dataorder module with authorisation governance
- Build external dCat data catalogue



Direktoratet
for forvaltning og
økonomistyring

THANKS

05

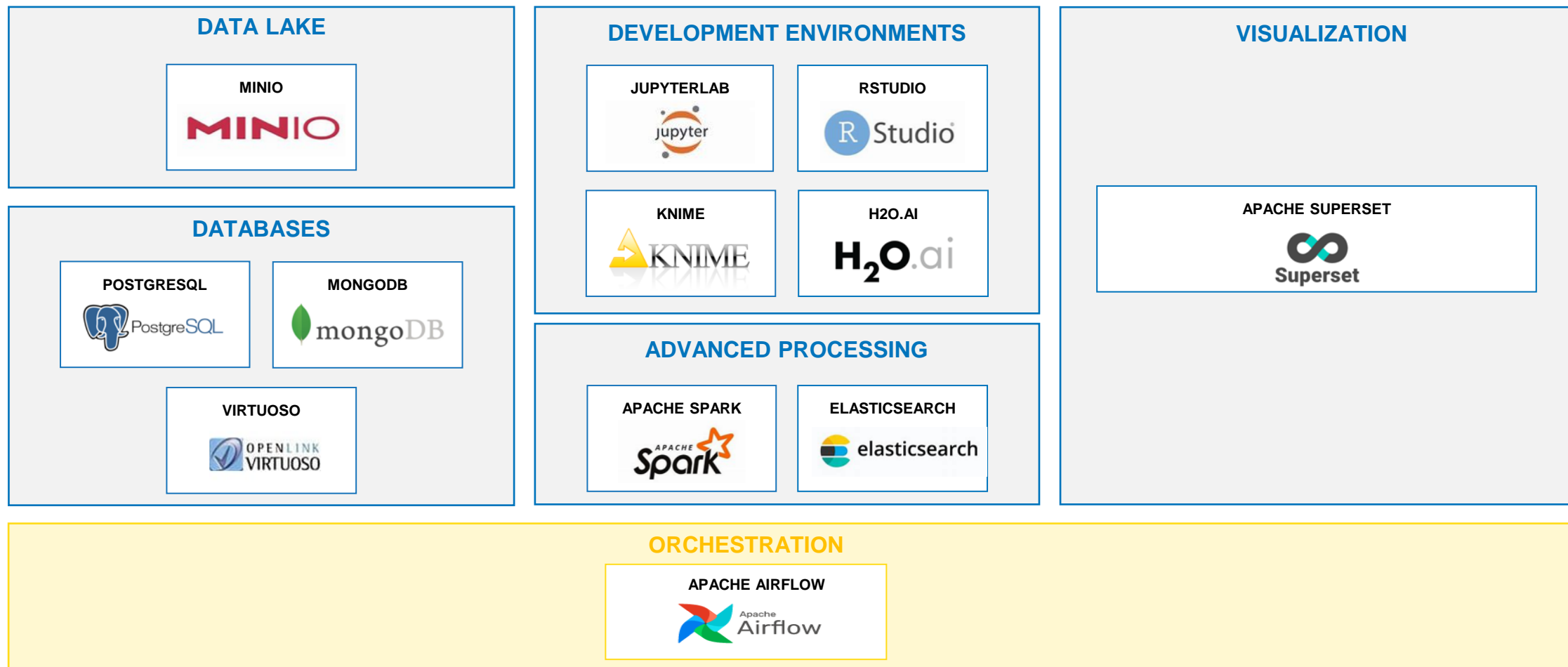
BDTI Service Architecture



KASPER RUTTEN

Cloud Solution Architect
BDTI team

BDTI open-source components



Discover the entire catalogue of BDTI available components at this link: <https://ec-europa.github.io/bdti-infrastructure/>

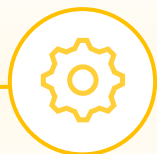
The Big Data Test infrastructure in use

COLLECT & AGGREGATE DATA

Databases (such as [Postgres](#) and [MongoDB](#)) and data lakes are available for storing and combining multiple data sets.

VISUALIZE DATA

Analysed data can be visualized in dashboards and graphs to clearly display insights, by using [Apache Superset](#), which is a modern data exploration and visualization platform.



GENERATE DATA

Public administrations generate a lot of data which can be used for decision-making.



ANALYSE DATA

Data can be analysed using the available development environments such as [JupyterLab](#), combined with for example [Apache Spark](#) to carry out data analytics.



PRESERVE DATA

The results of an analysis can be stored in databases or in a data lake. Multiple database types are supported such as [MongoDB](#), [Postgres](#), etc.

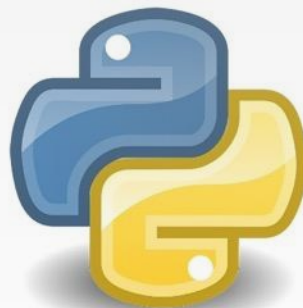


Or use [Apache Airflow](#) to orchestrate the process.

Let's do a demo



**JUPYTER
NOTEBOOK**



**PYTHON
FRAMEWORK**



**OPEN-SOURCE
DATA**

DATA EXPLORATION

A demo of a self-paced lab will be shown using EU open data related to COVID19 in Belgium.
During the self-paced lab tools such as Jupyter Notebook and Python will be used.



Q&A time

Thank you for your participation!



DROP US A LINE

BDTI Functional Mailbox: EC-BDTI-PILOTS@ec.europa.eu

Learn more about the Big Data Test Infrastructure

<https://digital-strategy.ec.europa.eu/en/policies/bdti>

Learn more about the BDTI success stories

<https://digital-strategy.ec.europa.eu/en/related-content?topic=190&type=185>

Big Data Test Infrastructure FAQ

<https://digital-strategy.ec.europa.eu/en/faqs/big-data-test-infrastructure>