Industrial-Driven Big Data as a Self-Service Solution

Introduction

Despina Kopanaki (FORTH)

I-BiDaaS Final Event

December 21, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780787

Housekeeping rules

- All attendees are muted by default during this session.
- To ask a question, use the Zoom Q&A panel or raise your hand.
- The speakers will reply back to you via text or answer live during the Q&A session.
- Contact the session facilitator, Dr. Andreas Miaoudakis, if you have any technical issue.

Now you

- **To which of our stakeholder type do you belong?**
  - *(Big Data Provider, Big Data Technology Provider, Research & Academia, Financial, Telecommunication, Manufacturing, other)*
- **Are you working with Big Data?**
  - *(Yes, No)*
- **Are you interested in Big Data Technologies to optimize your customer experience?**
  - *(Yes, No, Maybe)*
- **What is the main barrier or risk preventing you from implementing Big Data analytical solutions in your organization?**
  - *(Costs, Lack of expertise, Uncertain Value (ROI))*

Speakers

Prof. Sotiris Ioannidis

Associate Professor - Technical University of Crete, Greece

Affiliated Researcher - Institute of Computer Science, Foundation for Research and Technology – Hellas

I-BiDaaS Project Coordinator

Nuria de Lama

European Programs Manager, Atos Research & Innovation –
Member of Board of Directors, BDVA, Spain

Speakers

Prof. Dušan Jakovetić

Associate Professor at the Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Serbia

I-BiDaaS Scientific & Technical Manager.

Dr. Ramon de Pozuelo

Project Manager at Security Innovation & Transformation, CaixaBank, Barcelona, Spain

Dr. Ioannis Arapakis

Researcher at Telefonica, Barcelona, Spain

Speakers

Giuseppe Danilo Spennacchio

Flexible and Adaptive Systems Specialist at Factory Innovation Department, CRF, Italy

Dr. Vassilis Chatzigiannakis

Technical Manager, ITML, Greece

Dr. Hernan Ruiz Ocampo

Project Manager, Ecole Nationale des Ponts et Chaussées, Paris, France

Agenda

- 10:00 Welcome and Introduction – Despina Kopanaki (FORTH)
- 10:05 **Keynote**: A 5-years journey through the European Big Data Landscape – Nuria de Lama (ATOS, BDVA)
- 10:35 I-BiDaaS **Overview** – Prof. Sotiris Ioannidis (TUC, FORTH)
- 10:45 I-BiDaaS **Scientific & Technical view** – Prof. Dusan Jakovetic (UNSPMF)
- 11:05 **Pilot 1**: I-BiDaaS Application to the **banking** sector – Dr. Ramon Martin de Pozuelo (CAIXA)
- 11:35 Break
- 11:50 **Pilot 2**: I-BiDaaS Application to the **telecommunication** sector – Dr. Ioannis Arapakis (TID)
- 12:20 **Pilot 3**: I-BiDaaS Application to the **manufacturing** sector - Giuseppe Danilo Spennacchio (CRF)
- 12:50 I-BiDaaS **Platform Demonstration** – Dr. Vassilis Chatzigiannakis (ITML)
- 13:05 The I-BiDaaS **business & commercial offering** – Hernan Ruiz Ocampo (ENPC)
- 13:25 Q&A, Closing

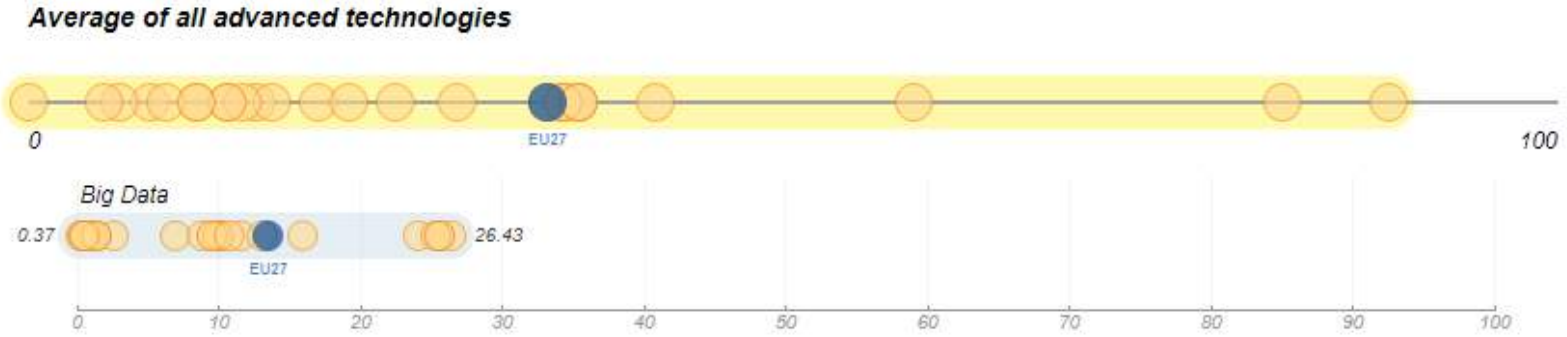
A 5-years journey through the European Big Data Landscape

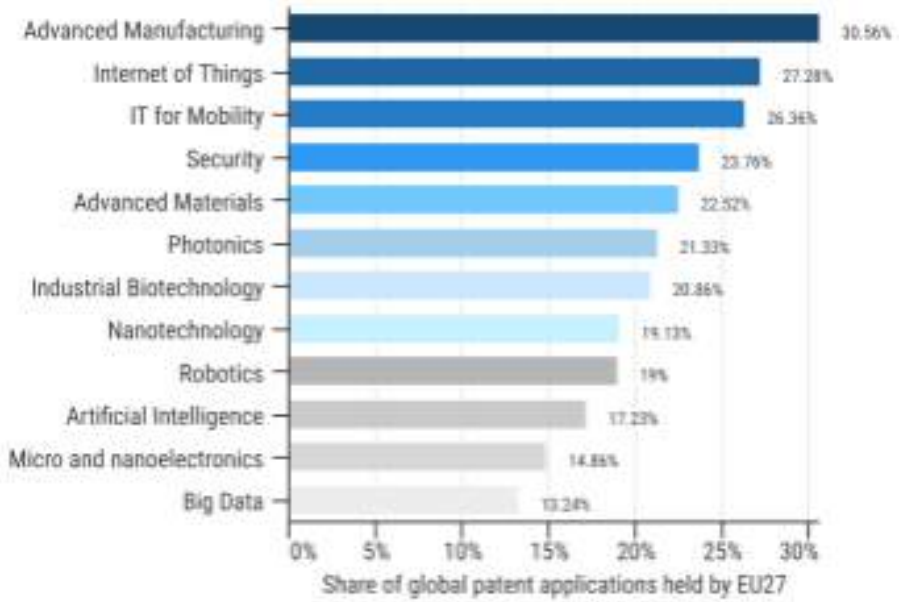
Nuria de Lama
European Programs Manager, Atos
Board of Directors BDVA
Nuria.delama@atos.net

21-12-2020

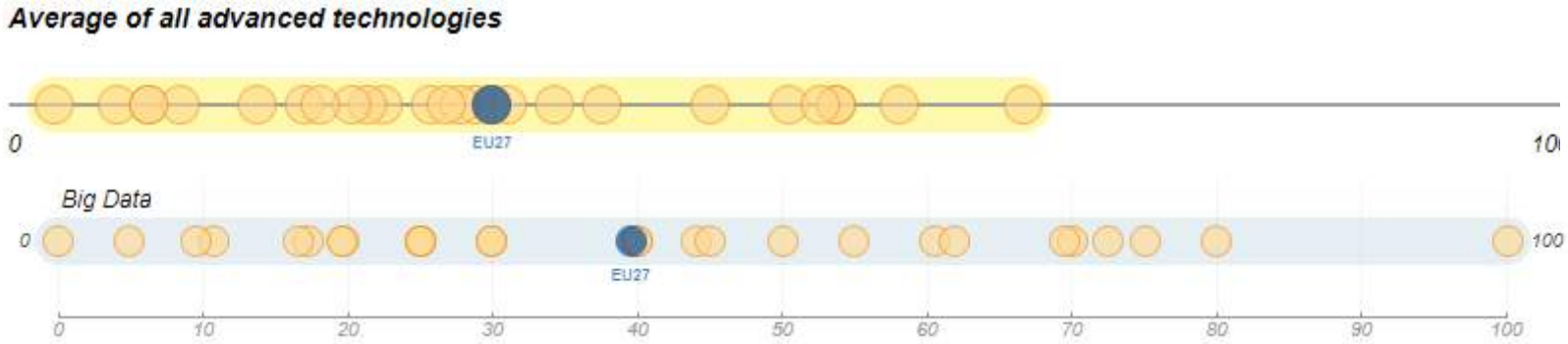
The role of data in Digital Transformation

- Technology Generation





- Technology Uptake



Source: Data Dashboard (Advanced Technologies for Industry) https://ati.ec.europa.eu/

Development paths or altenative scenarios

BDV

CONTRACTUAL ARRANGEMENT

SETTING UP A PUBLIC-PRIVATE PARTNERSHIP IN THE AREA OF DATA

BETWEEN

THE BIG DATA VALUE ASSOCIATION

AND

THE EUROPEAN UNION

The Big Data Value Association (registered offices: rue de Trèves 49/51, 1040 Brussels, Belgium), hereinafter referred to as 'the Private Side', and the European Union, represented by the Commission (jointly hereinafter referred to as 'the Parties'),

CONSIDERING THAT:

- The European Union's Horizon 2020 Framework Programme for research and innovation[1] may be implemented through public-private partnerships taking the form of a contractual arrangement between the partners committed to supporting the development and implementation of research and innovation activities of strategic importance to the Union's competitiveness and industrial leadership.
- The specific programme implementing Horizon 2020[2] has identified potential areas for establishing public-private partnerships.
- The rules for participation and dissemination in Horizon 2020[3] apply to the indirect actions to be financed by the Commission in the context of this arrangement.

[1] Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 — the Framework Programme for Research and Innovation (2014-2020) and repealing Decision No 1982/2006/EC (OJ L 347, 20.12.2013, p. 104).

[2] Council Decision (EU) No 743/2013 establishing the specific programme implementing Horizon 2020 — the Framework Programme for Research and Innovation (2014-2020) and repealing Decisions 2006/971/EC, 2006/972/EC, 2006/973/EC, 2006/974/EC and 2006/975/EC (OJ L 347, 20.12.2013, p. 965).

[3] Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 December 2013 laying

Development paths, dimensions and indicators

Source: datalandscape.eu

POLICY & SOCIETAL

PROGRAMME

TECHNICAL

IMPACT

APPLICATION

COMMUNITY

BUSINESS

COMMUNCIATION

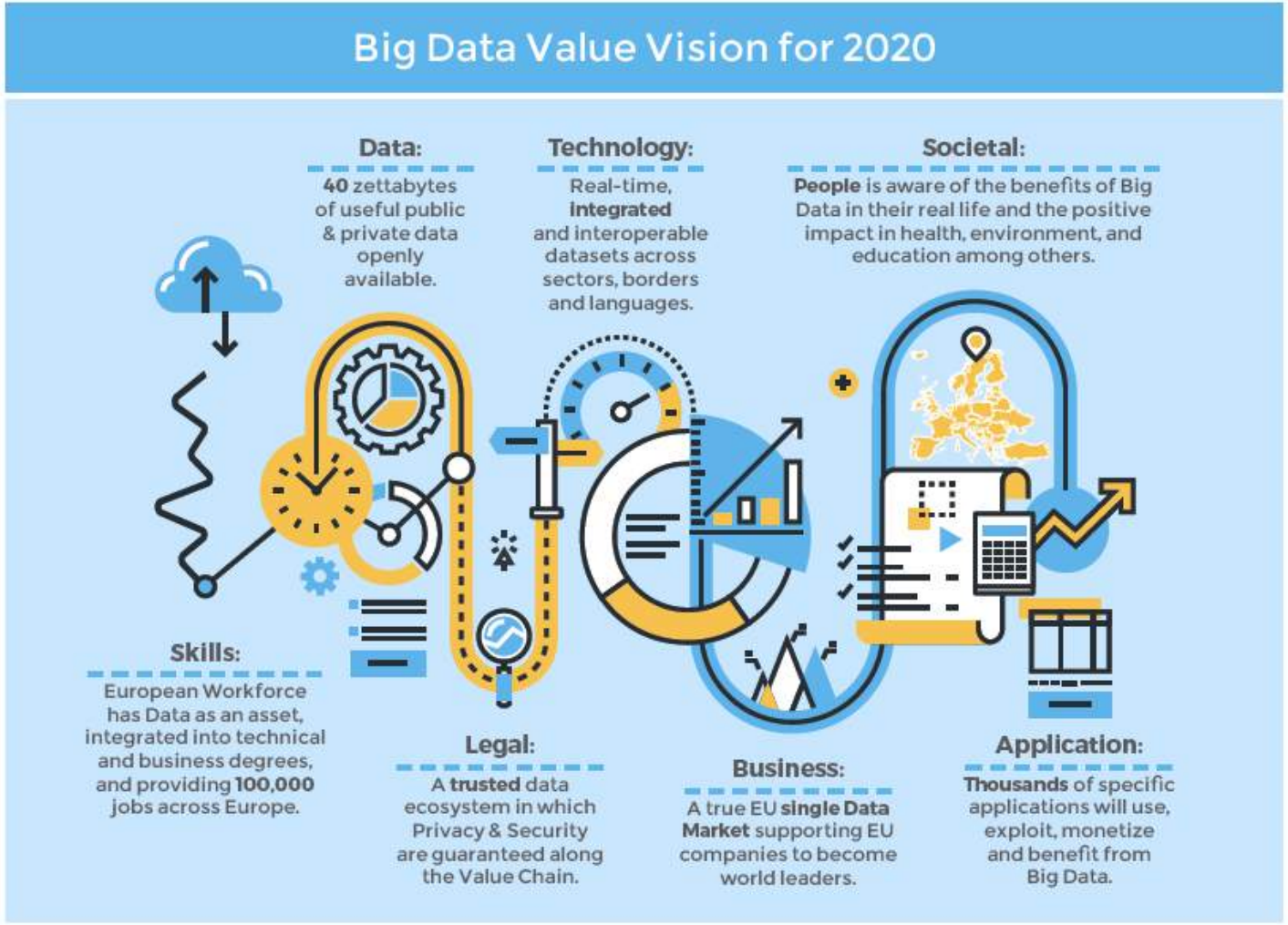
SKILLS AND EDUCATION

STERRING COMMITTEE

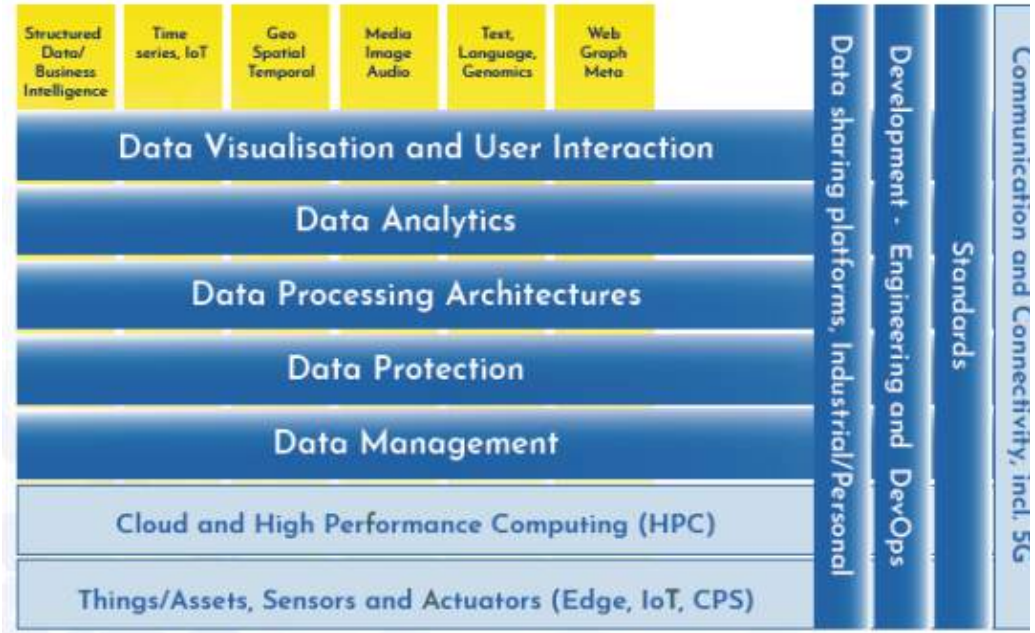
TECHNICAL COMMITTEE

COMMUNICATION TEAM

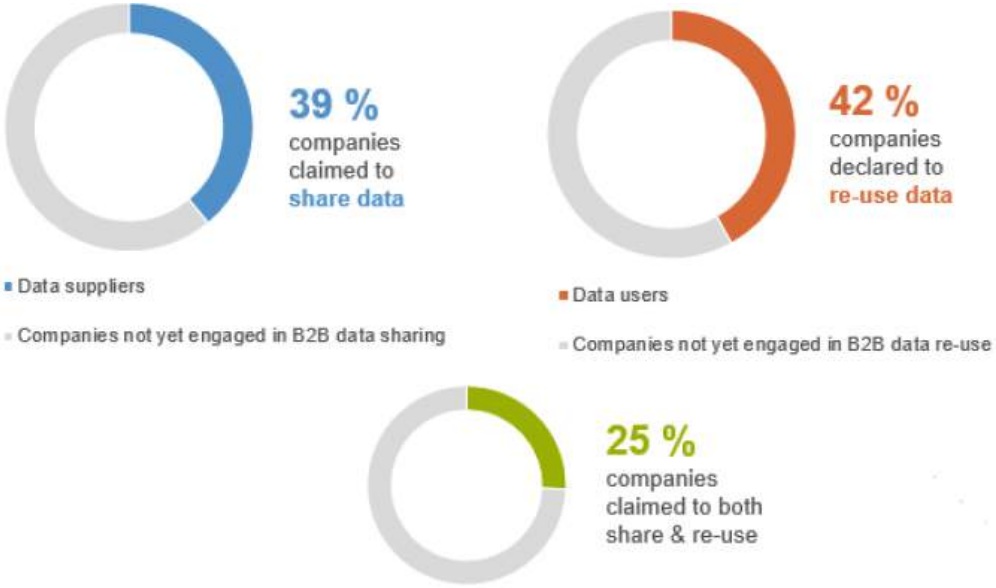
BDVA Reference Framework

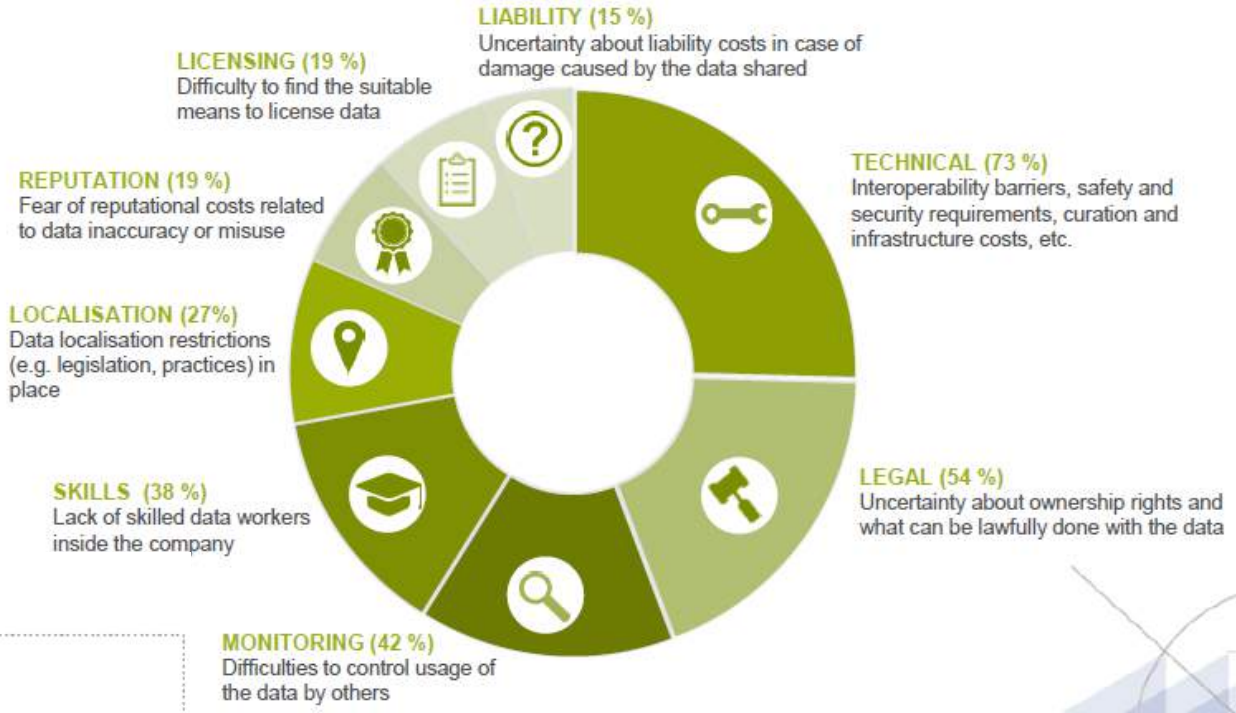


- 50+ Projects are part of the PPP portfolio, contributing to technology generation and uptake



Access to data and data sharing





Profile of data suppliers

Data suppliers

- Mainly large companies
- 90% share data **within their own business sector**
- A majoriy only share a **small percentage** of the data they generate
- 1 in 5 consider data sharing as their **main economic activity**
- 1 in 3 have been sharing data for over 8 years
- Very few adopt an Open Data Policy
- **Developing new business models** and/or services and products as main motivation

1 in 2 experience **obstacles** to make their data available

Profile of data users

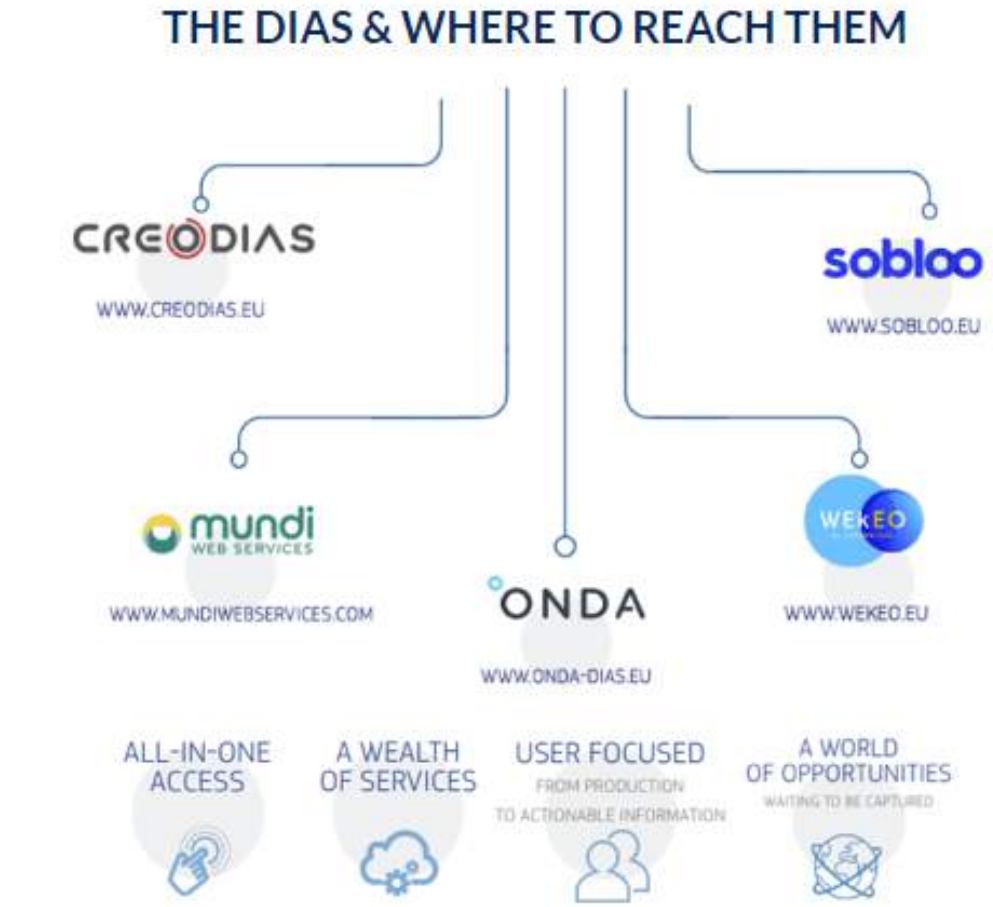
Data users

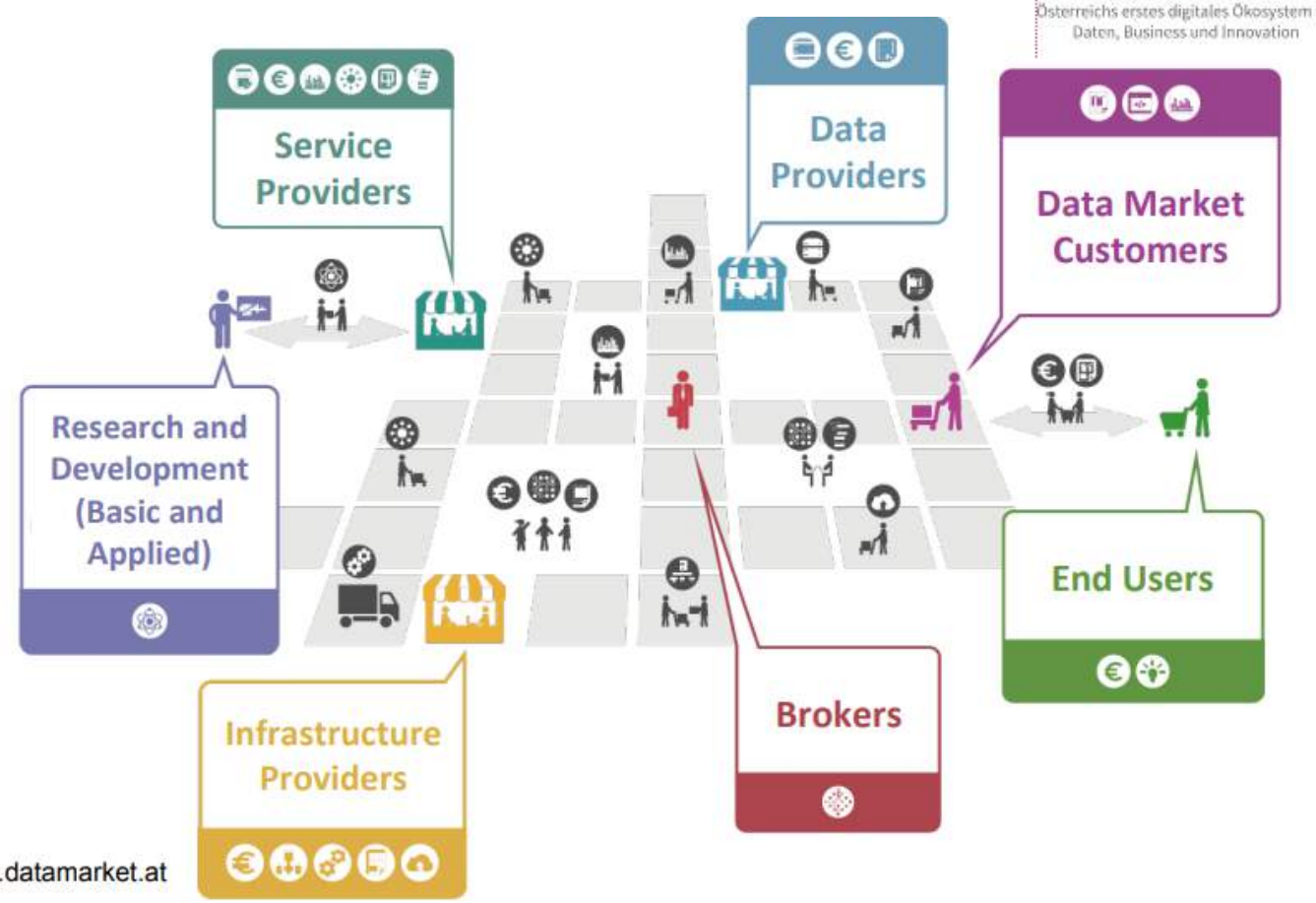
- Mainly large companies
- Mostly re-using data **from their own sector (83%)**
- **7 in 10 strongly depend on data** for their business
- Half of them have been re-using data for over 6 years
- More than 40 % have spent over €50 000/year to access data in the last years
- **Developing/improving** the catalogue of services and products as main motivation

6 in 10 experience **obstacles** to access data from others

Source: Everis (Study on B2B data sharing; 2018)

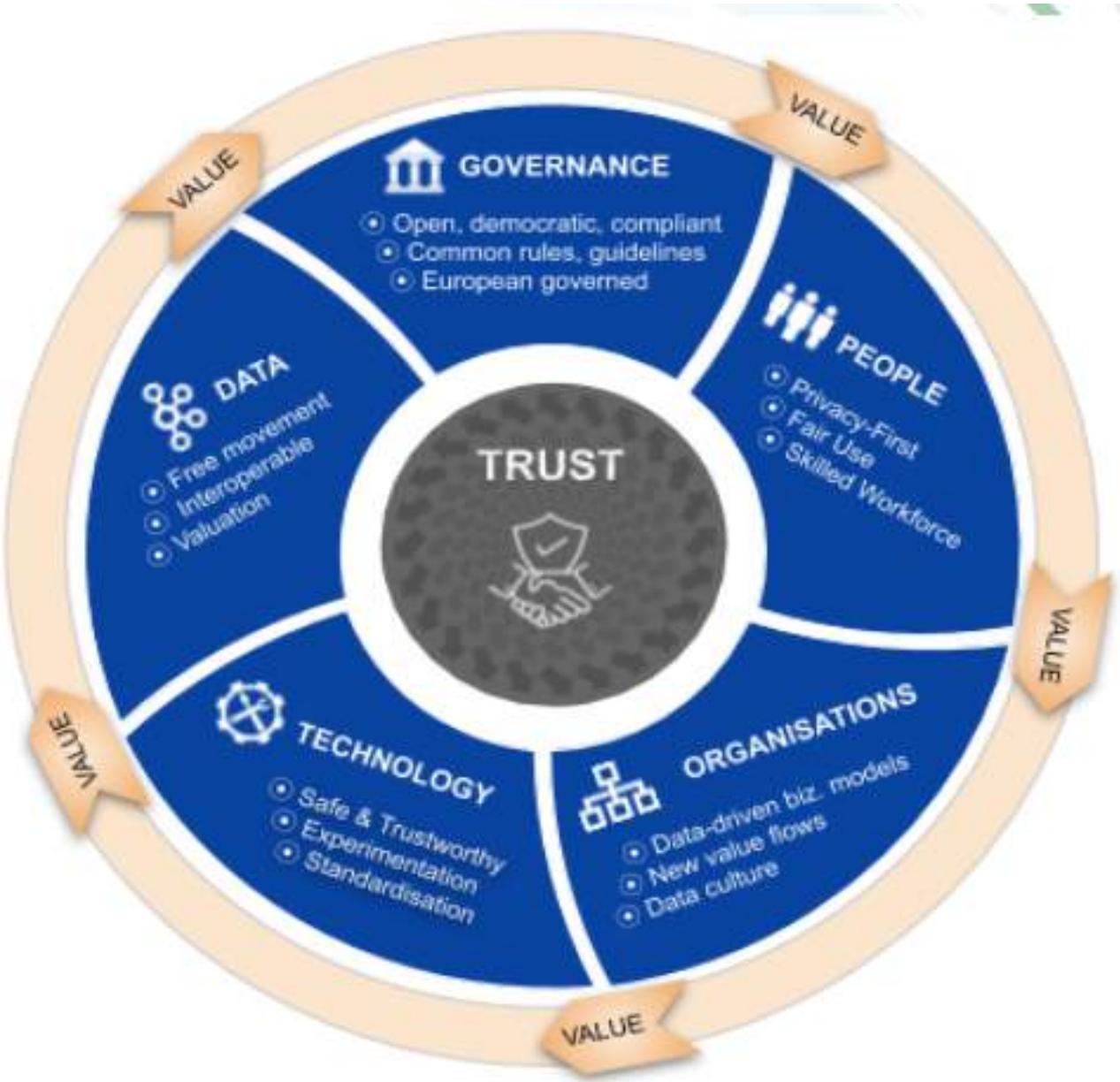
Data sharing: the role of Data Platforms



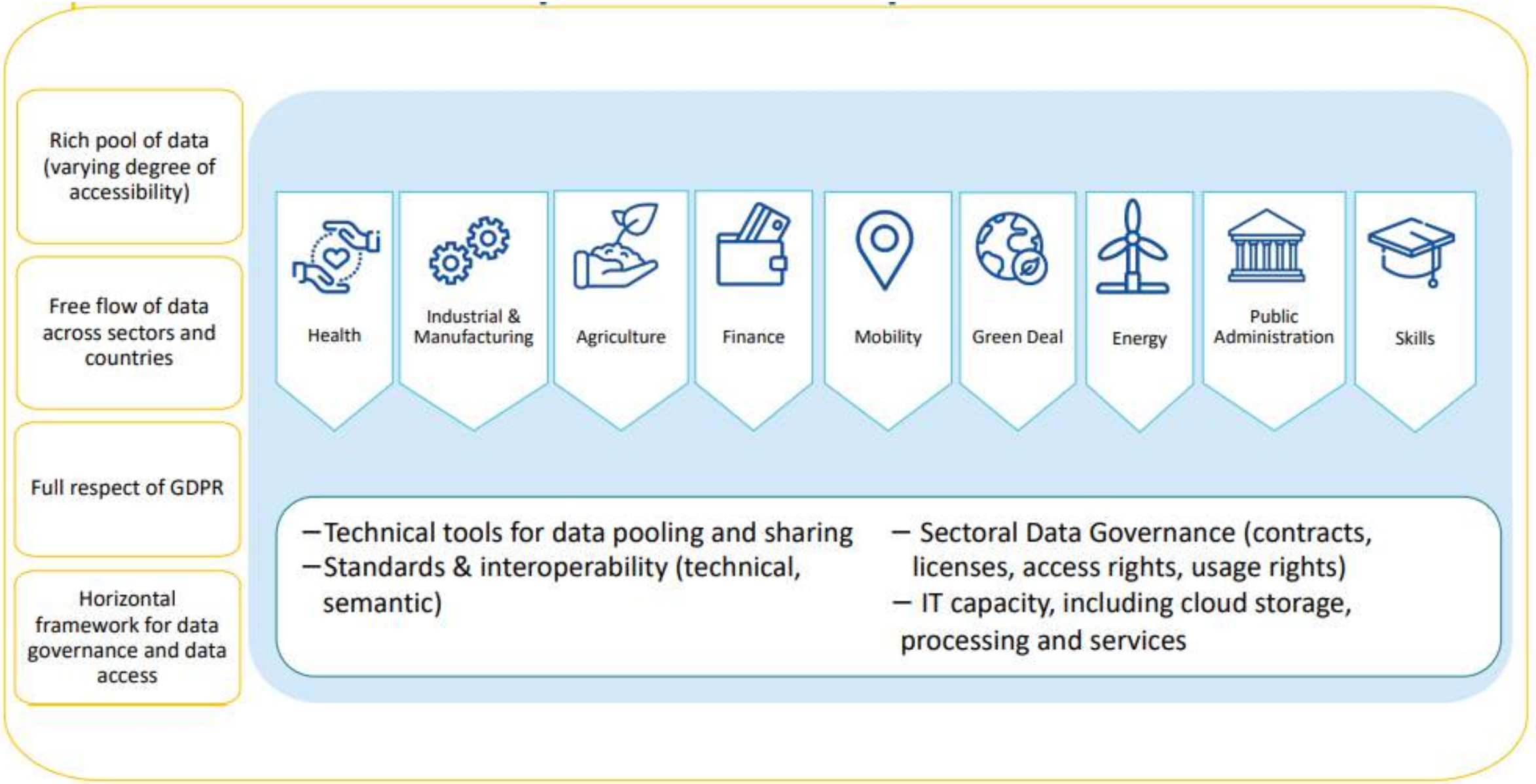


Towards a European-governed Data Sharing Space

BDV BIG DATA VALUE ASSOCIATION

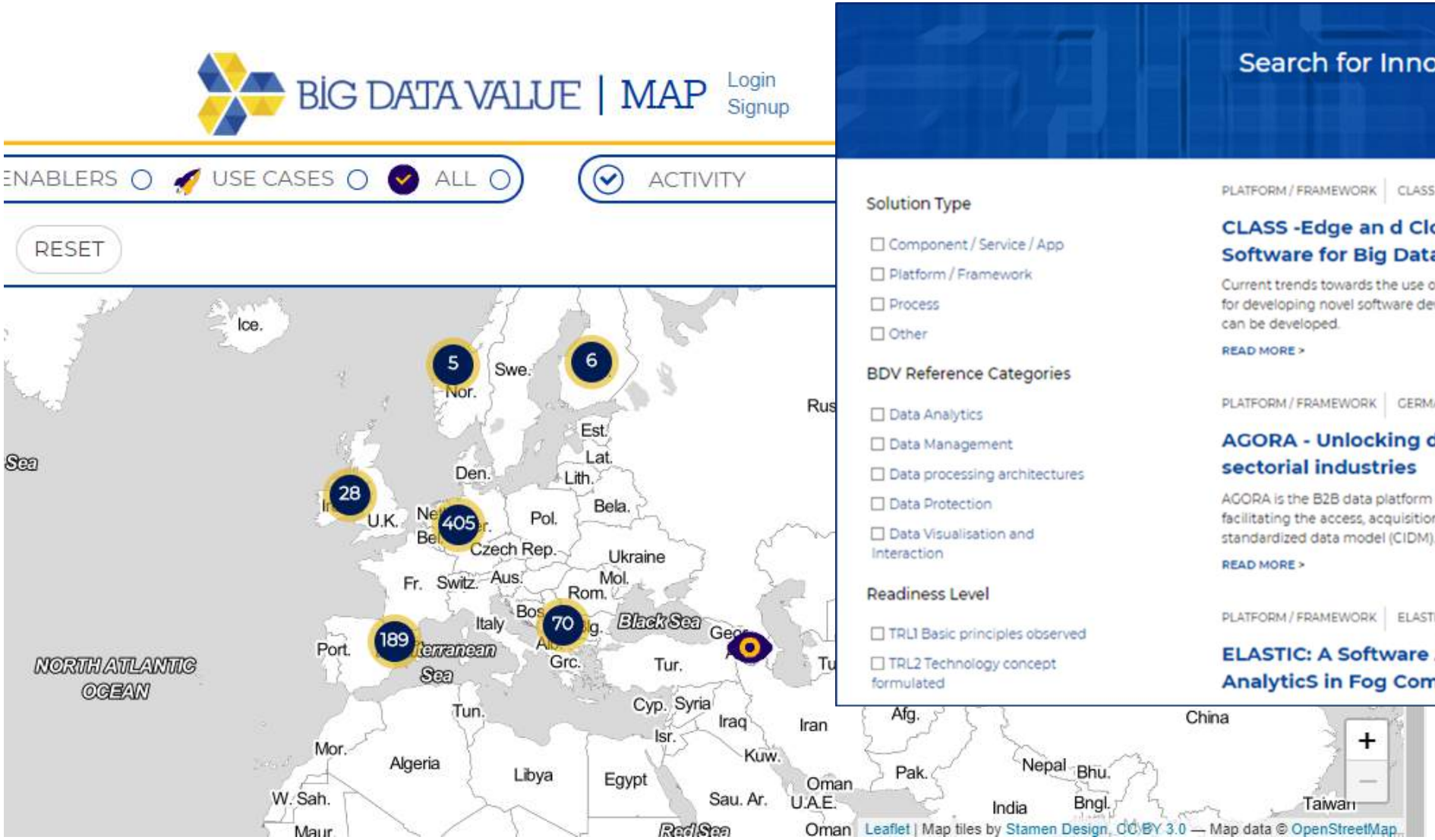


Common European Data Spaces



Lighthouse Projects to illustrate the transformative power of data

Available tools and mechanisms



The value of Data: Data companies, economy and society

Key Numbers 2019 for EU27 + U.K.

Indicator 2: Data Companies

Data suppliers have as their main activity the production and delivery of digital data-related products, services, and technologies.

Data suppliers

Thousand	2018	2019	Growth '19/'18	Companies Share as a % of total companies in ICT and Professional Services, '19
	283	290	2.3%	15%

Data users are organisations that generate, exploit collect and analyse digital data intensively and use what they learn to improve their business.

Data users

Thousand	2018	2019	Growth '19/'18	Companies Share as a % of total companies, '19
	712	716	0.6%	6.7%

Indicator 3: Data suppliers' revenues

The aggregated value of all the data-related products and services generated by EU Data suppliers companies.

€ Bn	2018	2019	Growth '19/'18
	77	84	8%

Indicator 4: Value of the Data Market

The marketplace where digital data is exchanged as "products" or "services" as a result of the elaboration of raw data.

€ Bn	2018	2019	Growth '19/'18
	72	75	5%

Indicator 5: Value of the Data Economy

The Data Economy measures the overall impacts of the data market on the economy as a whole.

€ Bn	2018	2019	Growth '19/'18	Share of EU GDP, '19
	378	406	7.6%	2.8%

Baseline Scenario 2025 for EU27

Indicator 2: Data Companies

Data suppliers have as their main activity the production and delivery of digital data-related products, services, and technologies.

Data suppliers

Thousand	2019	2025	CAGR '25/'19	Companies Share as a % of total companies in ICT and Professional Services, '25
	149	173	2.6%	12.8%

Data users are organisations that generate, exploit collect and analyse digital data intensively and use what they learn to improve their business.

Data users

Thousand	2019	2025	CAGR '25/'19	Companies Share as a % of total companies, '25
	535	583	1.4%	6.3%

Indicator 3: Data Suppliers' revenues

The aggregated value of all the data-related products and services generated by EU Data suppliers companies.

€ Bn	2019	2025	CAGR '25/'19
	64	99	6.8%

Indicator 4: Value of the Data Market

The marketplace where digital data is exchanged as "products" or "services" as a result of the elaboration of raw data.

€ Bn	2019	2025	CAGR '25/'19
	58	83	6%

Indicator 5: Value of the Data Economy

The indicator captures the potential gap between demand and supply of data skills in Europe.

€ Bn	2019	2025	CAGR '25/'19	Share of EU GDP, '25
	325	550	9.2%	4%

Impact of the Big Data Value PPP (I)

- 200 members (+27% SMEs)
- **Private investments mobilized** by end of 2018: 1,57 Billion € (Leverage factor: 7,8)
- Growth of **SMEs in the data economy**:
  - Growth in revenues 60% with respect to 2014 (beginning of H2020) and 17,7 % increase in the last year
  - 74,9% increase on FTEs with respect to 2014
- Contribution to increase the number of Data companies in Europe, market share of the European Union in the global Big Data Market, and the number of Data professionals in Europe.
- Contributions to environmental challenges (Energy efficiency, CO2, Waste reduction)

Impact of the Big Data Value PPP (II)

- **106 innovations** of exploitable value delivered **during 2018**:
  - 39 of significant impact (37%)
  - 67 of medium impact (63%)
- From the 106 innovation of exploitable value delivered during 2018:
  - **53 incremental innovation** (50%)
  - 6 architectural innovation (5,7%)
  - **38 disruptive innovation** (35,8%)
  - 1 radical innovation (0,0%)
  - 8 were not categorised under any of these categories
- 99 over 106 innovations can generate **economic impact** (93%) and 48 (45%) **societal impact**

Data skills

Key Numbers 2019 for EU27 + U.K.

Indicator 1: Data Professionals

Workers who collect, store, manage, analyse, interpret, and visualise data as their primary or as a relevant part of their activity.

Million	2018	2019	Growth '19/'18	Share of data professionals on total EU employment, '19
	7.2	7.6	5.5%	3.6%

Indicator 6: Data Professionals Skills Gap

The indicator captures the potential gap between demand and supply of data skills in Europe.

Thousand	2018	2019	Growth '19/18	Share of total skills demand, '19
	406	459	13%	5.7%

Baseline Scenario 2025 for EU27

Indicator 1: Data Professionals

Workers who collect, store, manage, analyse, interpret, and visualise data as their primary or as a relevant part of their activity.

Million	2019	2025	CAGR '25/'19
	6	9.3	7.5%

Indicator 6: Data Professionals Skills Gap

The indicator captures the potential gap between demand and supply of data skills in Europe.

Thousand	2019	2025	CAGR '25/'19	Share of total skills demand, '25
	399	759	11.3%	8.2%

Source: EDM Monitoring Tool, IDC 2020

Contributions of the Big Data Value PPP

Based on figures from MR2018

- Over 7.500 new jobs created by 2023 linked to project activities
- 48 new job profiles reported in 2018
- 85 training activities during 2018 involving over 9700 participants
- 16 developed interdisciplinary programs during 2018 with around 250 participants,
- 396 equivalent Master and Phd FTEs (260 Master and 136 Phd)
- 323 events outreaching over 630.000 participants

Contributions of the Big Data Value PPP

Education Hub

Mobility Program

Skills Recognition Program

Big Data Centres of Excellence

The future: AI, Data and Robotics

Artificial Intelligence – a real business driver for Europe?

Artificial Intelligence (AI), Data and **Robotics** will create new opportunities, transform many if not all verticals and ultimately shift the balance of power in the shortest possible time.

AI, Data and Robotics combined will be the core driver of innovation, productivity and economic growth. Together they can be used to solve the greatest challenges we face: Environmental sustainability; energy, food and water security; and improving health and quality of life.

Europe can and must be the pacemaker worldwide!

In Europe we must not be shy or afraid. We have our strengths – which we should not neglect. We should use them!

Our strengths are our excellent research networks, our well-established companies that are world market leaders in several major verticals, our growing startup communities and, not to forget, our European values.

Of course, we must be open and accept the challenges and worldwide competition.

To leverage our strength, we brought major European activities for **AI (Claire, Ellis, EurAI), Data (BDVA)** and **Robotics (euRobotics)** into a **Partnership** and setup cooperation with other major European and regional initiatives.

This **Partnership** is the European focal point for AI, Data and Robotics. Europe has all the expertise needed to progress rapidly in the deployment of these technologies, but it needs to direct energy towards building a coherent infrastructure to stimulate deployment and adoption, build up an effective innovation ecosystem and drive excellent research.

This **Partnership** will federate and cohere the communities that underpin European AI, Data and Robotics. It will stimulate private investment and orient public funding to address the key challenges. Collaboration within the Partnership will deliver Europe's vision for a human centric and trustworthy use of AI, Data and Robotics.

It is a pleasure for us to present you this paper – the **Strategic Research, Innovation and Deployment Agenda!**

This paper results from the joint work of BDVA, Claire, Ellis, EurAI and euRobotics colleagues. It includes hundreds of contributions collected in consultations with stakeholders, member states, associations, and individuals.

Many thanks to all contributors!

Thomas Hahn
President Big Data Value Association (BDVA)
Lead entity, main contact

Morten Irgens
Executive board, Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE)

Bernd Liepert
euRobotics President

Barry O'Sullivan
Past President, European Artificial Intelligence Association (EurAI)

Arnold WM Smeulders
Envoy of the European Laboratory for Learning and Intelligent Systems, ELLIS

David Bisset
Executive Director euRobotics

Edward Curry
Vice President, Big Data Value Association (BDVA)

Michela Milano
Past Deputy President, European Artificial Intelligence Association

Sonja Zillner
SRIDA Lead
Big Data Value Association (BDVA)

https://ai-data-robotics-partnership.eu/wp-content/uploads/2020/09/AI-Data-Robotics-Partnership-SRIDA-V3.0.pdf

Data Platforms will Fuel AI-Driven Decision-Making

Data Generation and Analysis
(including IoT)

Data Platforms
(Access and Portability)

AI and Decision Platforms

Thank you!

Industrial-Driven Big Data as a Self-Service Solution

Project Overview

Prof. Sotiris Ioannidis (TUC, FORTH)

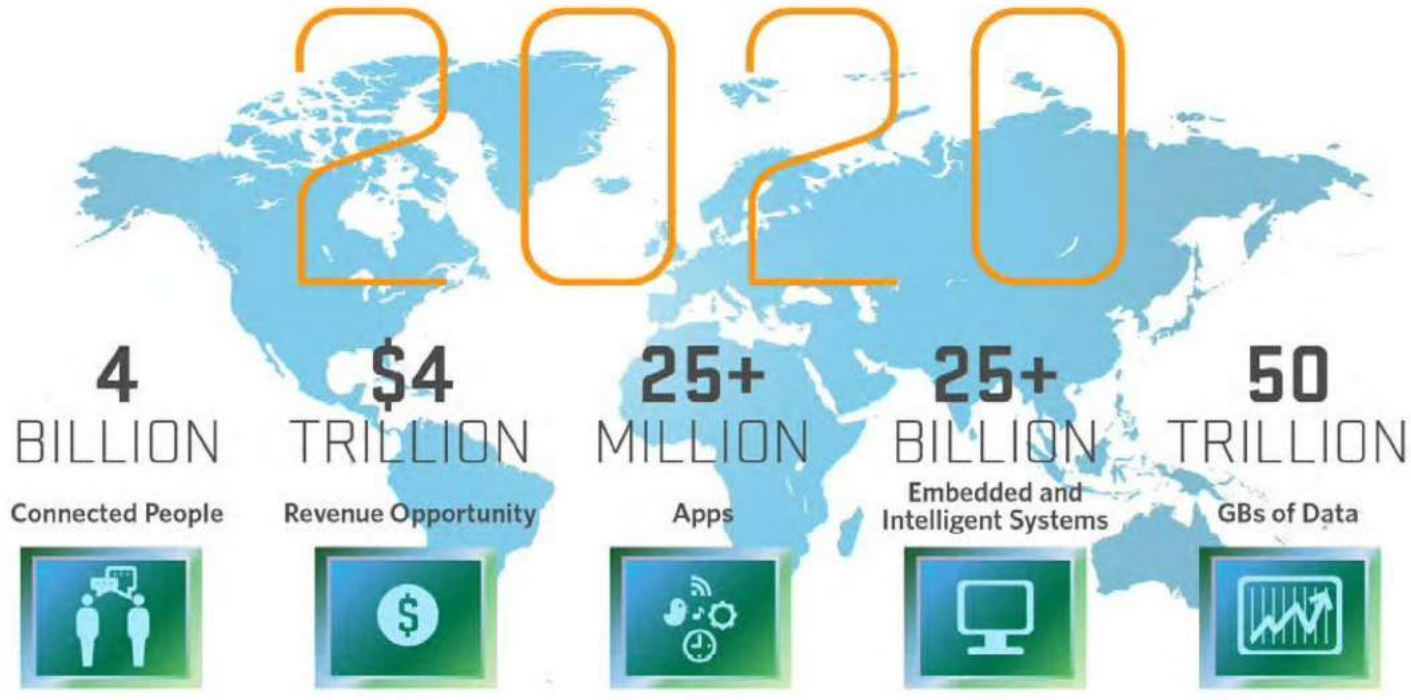
I-BiDaaS Final Event

December 21, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780787

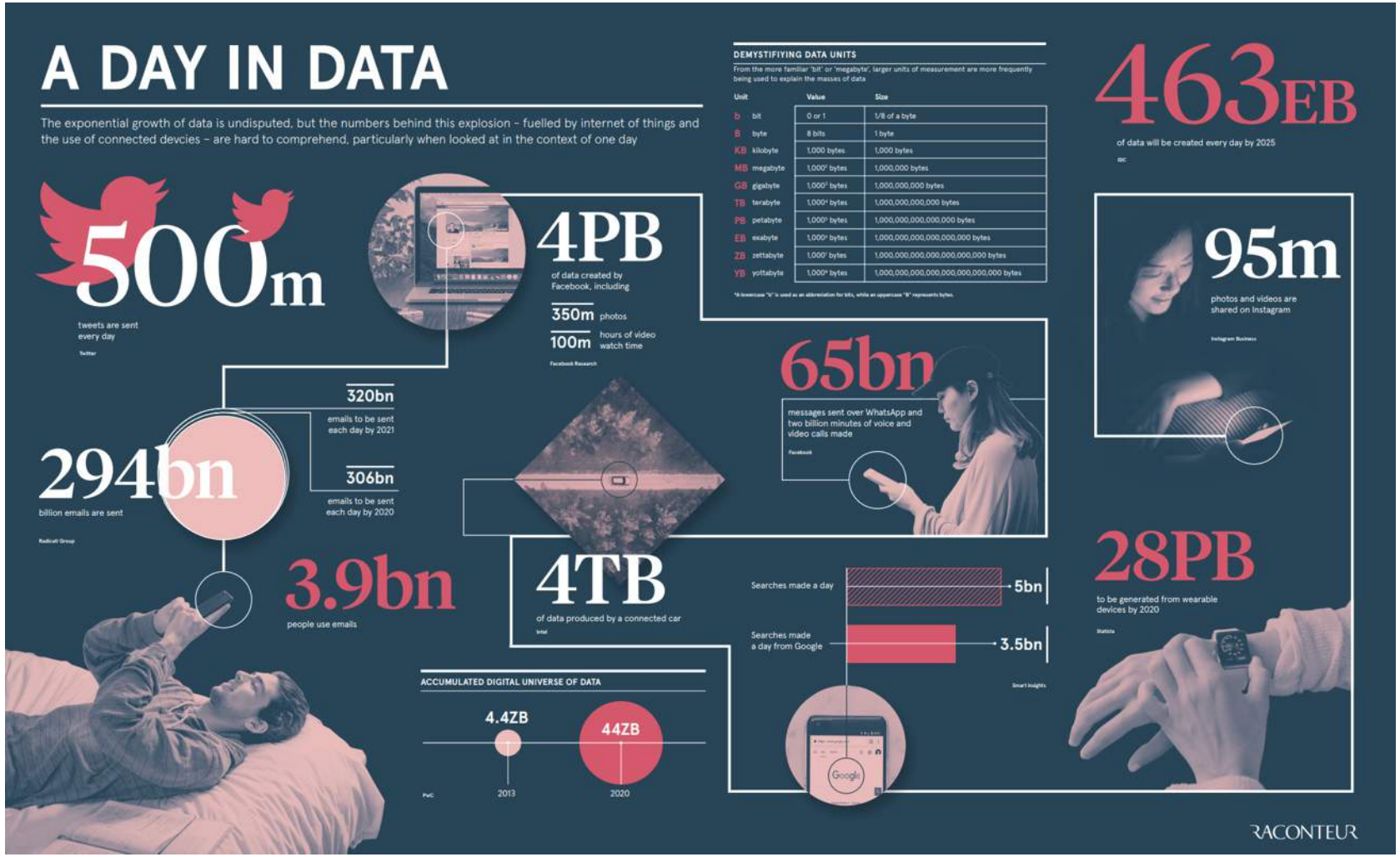
Big Data Era

- Many new **sources of data** become available
  - Most data is produced continuously at high rates



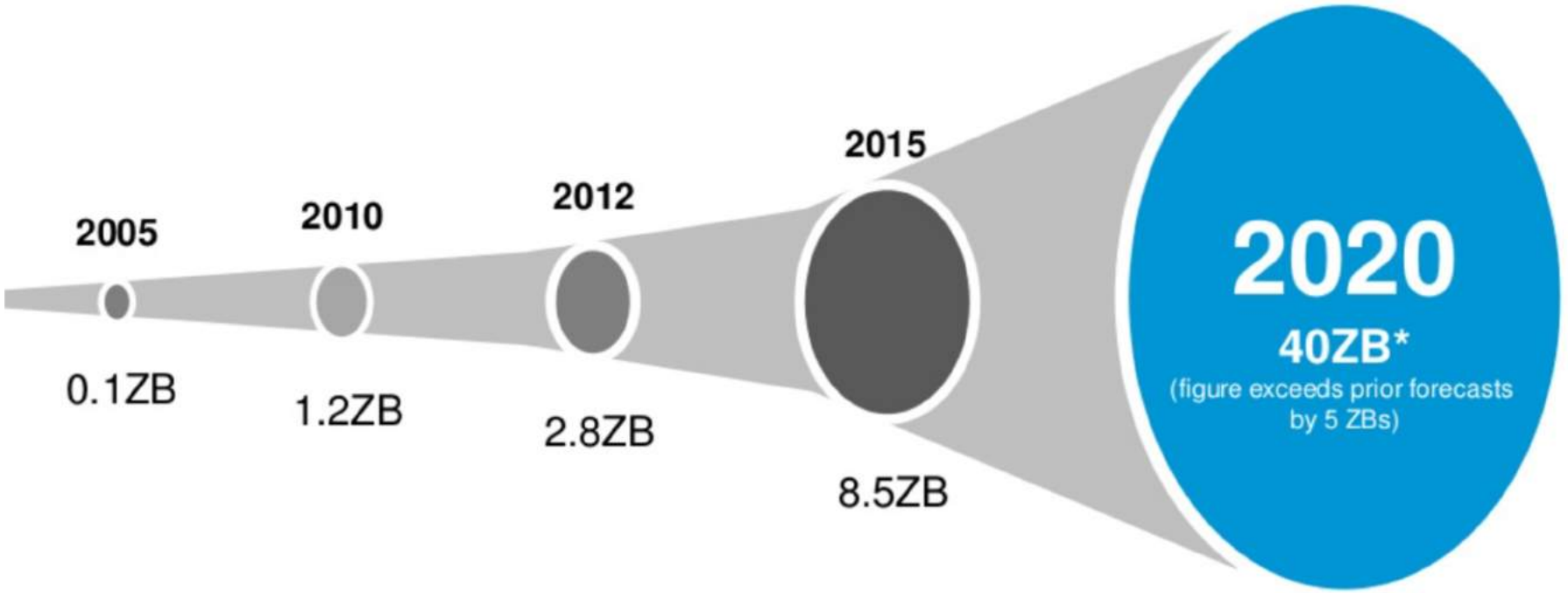
- The variety of data can drive **big-data investments**

How much data is generated each day?



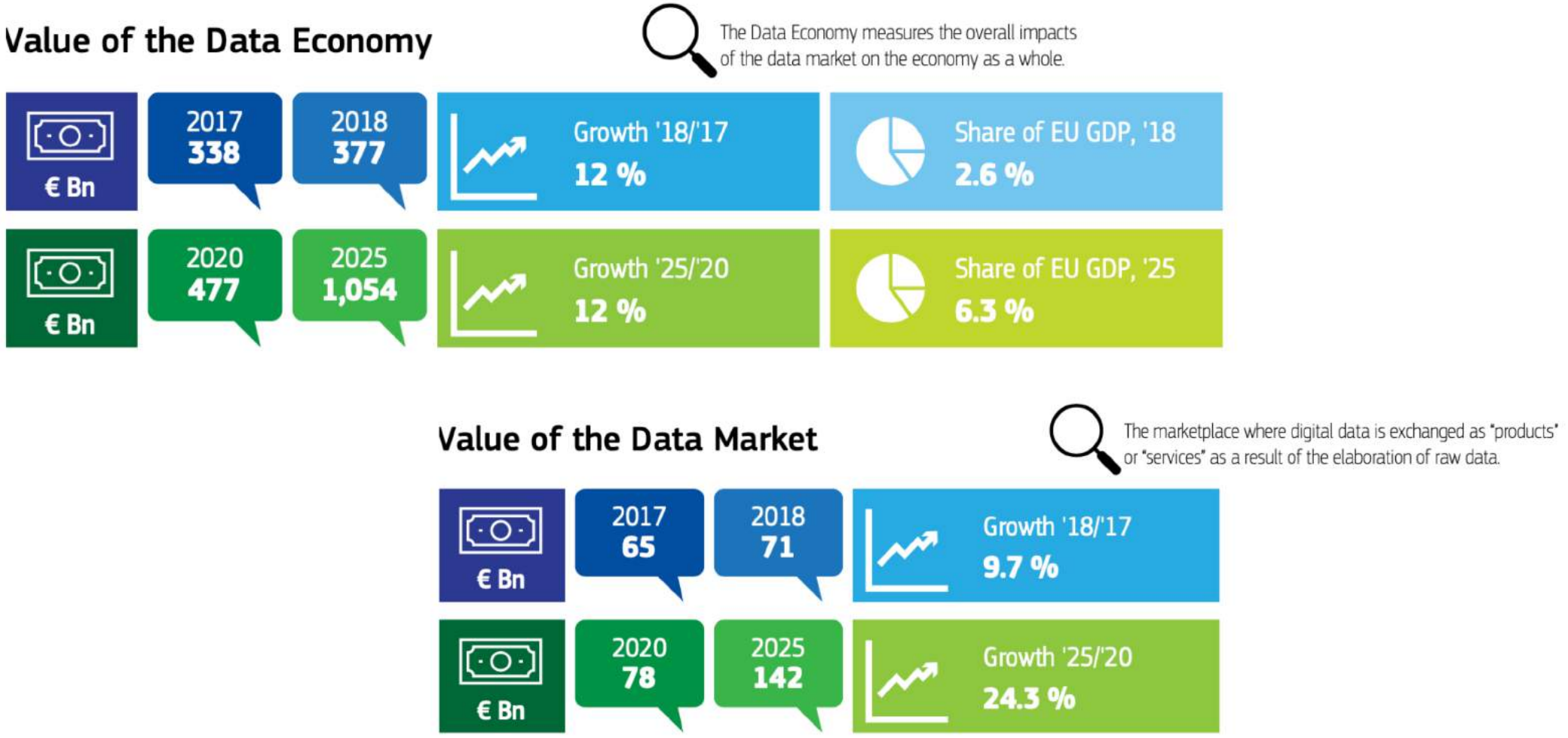
Source: RACONTEUR

The Data Deluge



*Source: IDC

A Data-Powered European Economy that Drives Innovation



**Source: IDC, European Data Market*

A Data-Powered European Economy that Drives Innovation

**Source: IDC, European Data Market*

Identity card

Project Consortium
13 partners

PROJECT NAME
Industrial-Driven Big Data as a Self-Service Solution

TOTAL BUDGET / TOTAL EC FUNDING
€ 4 997 035

PROJECT TYPE
RIA

START DATE
1 January 2018

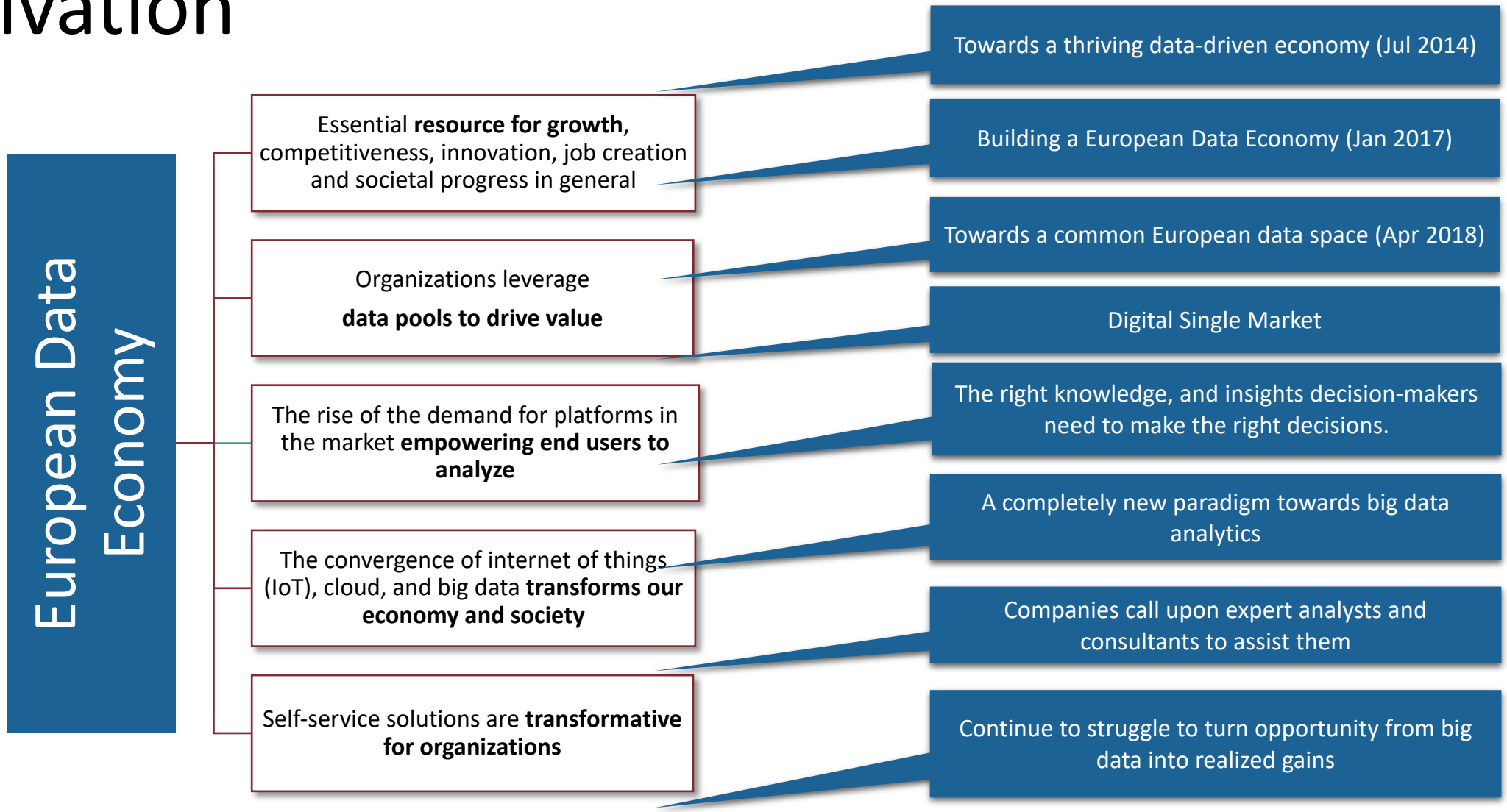
DURATION
36 months

http://www.ibidaas.eu/ @Ibidaas https://www.linkedin.com/in/i-bidaas/

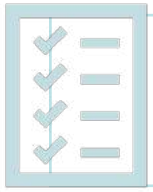
I-BiDaaS Consortium

1. FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS (**FORTH**)
2. BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION (**BSC**)
3. IBM ISRAEL - SCIENCE AND TECHNOLOGY LTD (**IBM**)
4. CENTRO RICERCHE FIAT SCPA (**CRF**)
5. SOFTWARE AG (**SAG**)
6. CAIXABANK, S.A (**CAIXA**)
7. THE UNIVERSITY OF MANCHESTER (**UNIMAN**)
8. ECOLE NATIONALE DES PONTS ET CHAUSSEES (**ENPC**)
9. ATOS SPAIN SA (**ATOS**)
10. AEGIS IT RESEARCH LTD (**AEGIS**)
11. INFORMATION TECHNOLOGY FOR MARKET LEADERSHIP (**ITML**)
12. University of Novi Sad Faculty of Sciences Serbia (**UNSPMF**)
13. TELEFONICA INVESTIGACION Y DESARROLLO SA (**TID**)

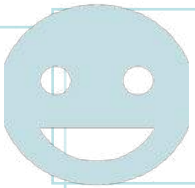
Motivation


European Data Economy
Essential **resource for growth**, competitiveness, innovation, job creation and societal progress in general
Organizations leverage **data pools to drive value**
The rise of the demand for platforms in the market **empowering end users to analyze**
The convergence of internet of things (IoT), cloud, and big data **transforms our economy and society**
Self-service solutions are **transformative for organizations**
Towards a thriving data-driven economy (Jul 2014)
Building a European Data Economy (Jan 2017)
Towards a common European data space (Apr 2018)
Digital Single Market
The right knowledge, and insights decision-makers need to make the right decisions.
A completely new paradigm towards big data analytics
Companies call upon expert analysts and consultants to assist them
Continue to struggle to turn opportunity from big data into realized gains

Our Vision



A **complete** and **safe environment** for methodological **big data experimentation**

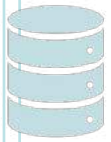


Tool and services to **increase the quality** of data analytics

A Big Data as a **Self-Service solution** that boosts EU's data-driven economy

Tools and services for **fast ingestion and consolidation** of both realistic and fabricated data

Increases impact in research community and contributes to industrial innovation capacity



Tools and services for the management of **heterogeneous infrastructures** including elasticity

Project Statement

***I-BiDaaS** aims to **empower** users to easily **utilize and interact** with **big data technologies**, by designing, building, and demonstrating, a **unified framework** that:*

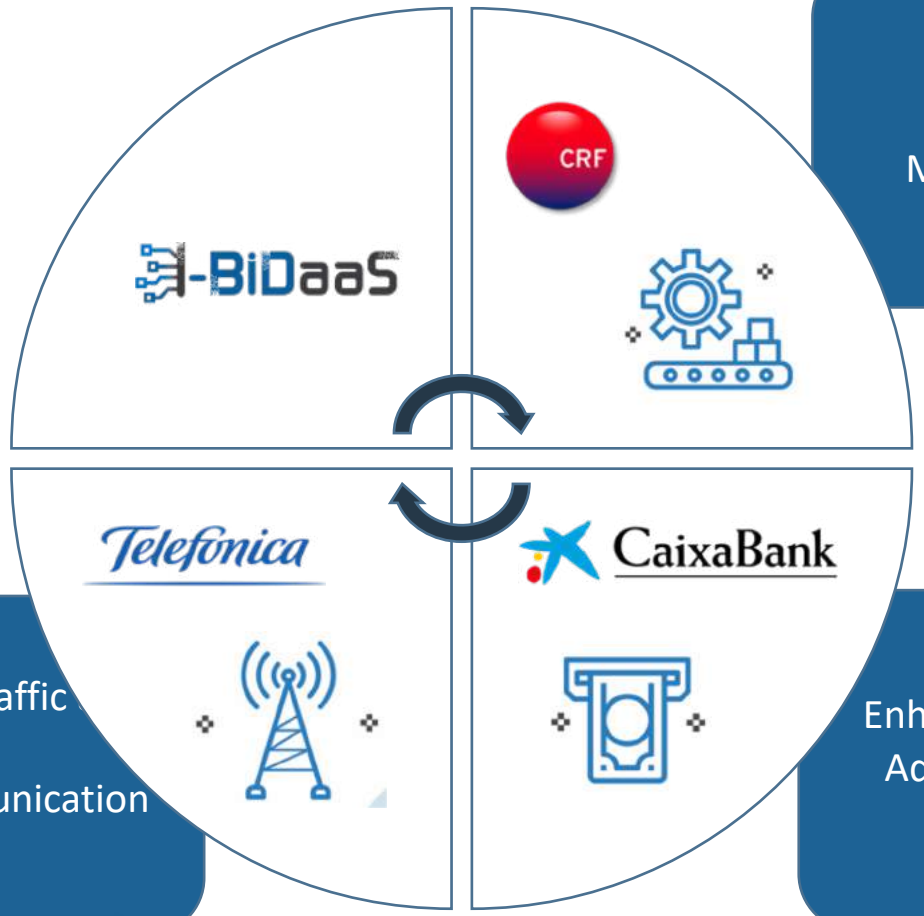
*significantly increases the **speed of data analysis** while coping with the rate of **data asset growth**, and facilitates **cross-domain data-flow** towards a thriving **data-driven EU economy**.*

*I-BiDaaS will be tangibly validated by three real-world, **industry-lead** experiments.*

Objectives

1. Release a **complete and solid Big Data-as-a-Self-Service solution** that can be easily configured and adopted by practitioners
2. **Break** inter- and intra-sectorial **data-silos**, create a data market and offer new business opportunities (support **data sharing**, **exchange**, and **interoperability**)
3. Construct **a safe environment** for methodological **big data experimentation**, for the development of **new products**, **services**, and **tools**
4. Develop **data processing tools and techniques** applicable in **real-world** settings, demonstrate significant **increase** of **speed of data** throughput
5. Develop technologies that will increase the efficiency and competitiveness of all EU companies and organisations that need to manage vast and complex amounts of data

Application / Experimentation



CRF

Production process of aluminium casting

Maintenance and monitoring of production assets

Manufacturing Industry

TID

Accurate location prediction with high traffic visibility

Optimization of placement of telecommunication equipment

Quality of Service in Call Centers

Telecommunications Industry

CAIXA

Enhance control of customers to online banking

Advanced analysis of bank transfer payment in financial terminal

Analysis of relationships through IP address

Banking/Finance Industry

I-BiDaaS Target Domains

I-BiDaaS

Industrial-Driven Big Data as a Self-Service Solution

I-BiDaaS aims to empower IT and non-IT big data experts to easily utilize and interact with big data technologies.

Check out Website & Social

- www.ibidaas.eu
- twitter.com/ibidaas
- linkedin.com/in/i-bidaas
- zenodo.org/communities/i-bidaas

This project has received funding from the European Union's Horizon 2020 Research and Innovation program under grant agreement **No 780787**.

3 We made applications for **3 SECTORS**
- Financial
- Telecommunications
- Manufacturing

13 **PARTNERS**

10 **USE CASES**

11 **SOFTWARE COMPONENTS**
5 Open Source
6 Proprietary

Industrial-Driven Big Data as a Self-Service Solution

Scientific-technical view

Dusan Jakovetic

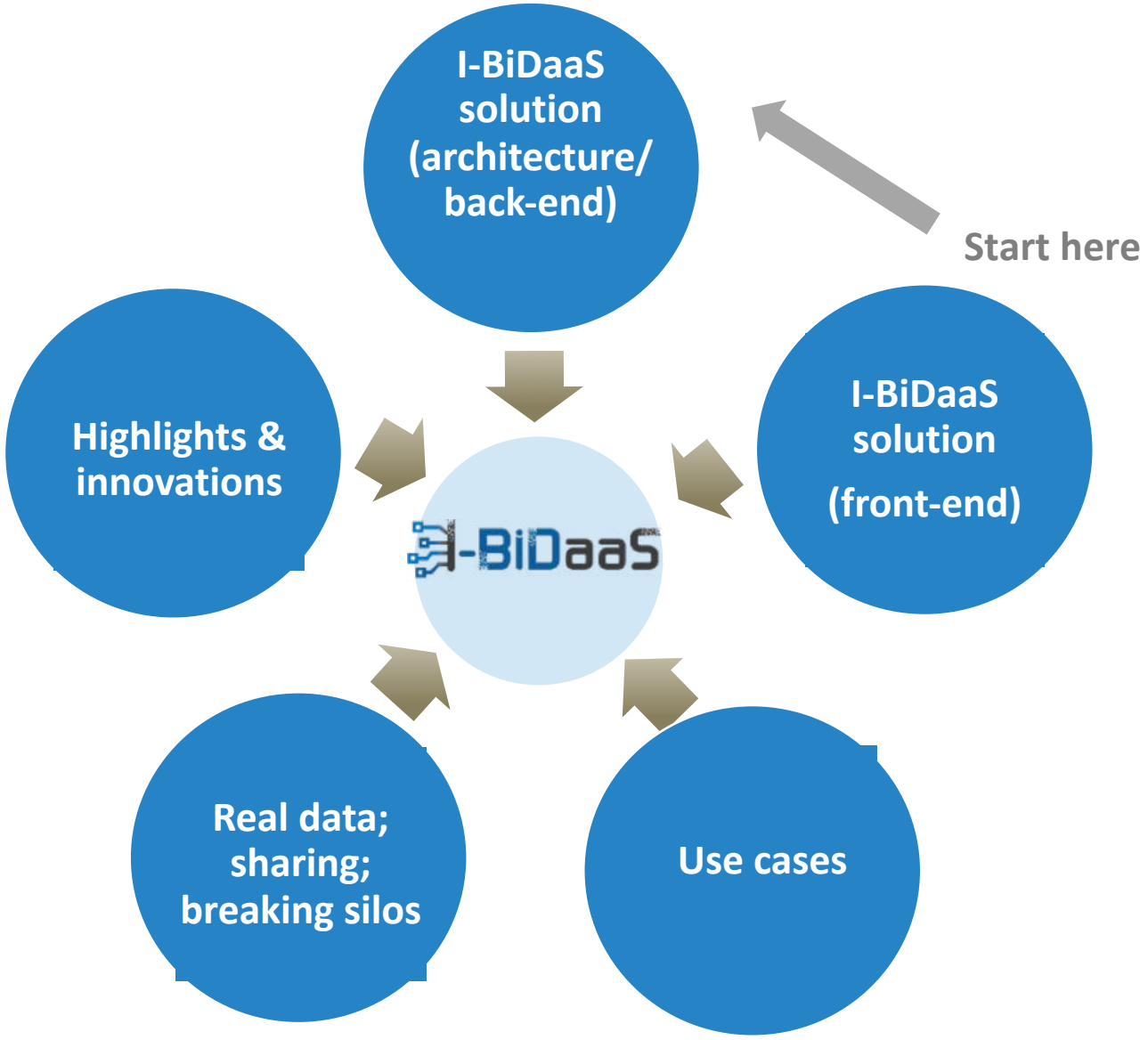
University of Novi Sad, Faculty of Sciences

I-BiDaaS Final Event

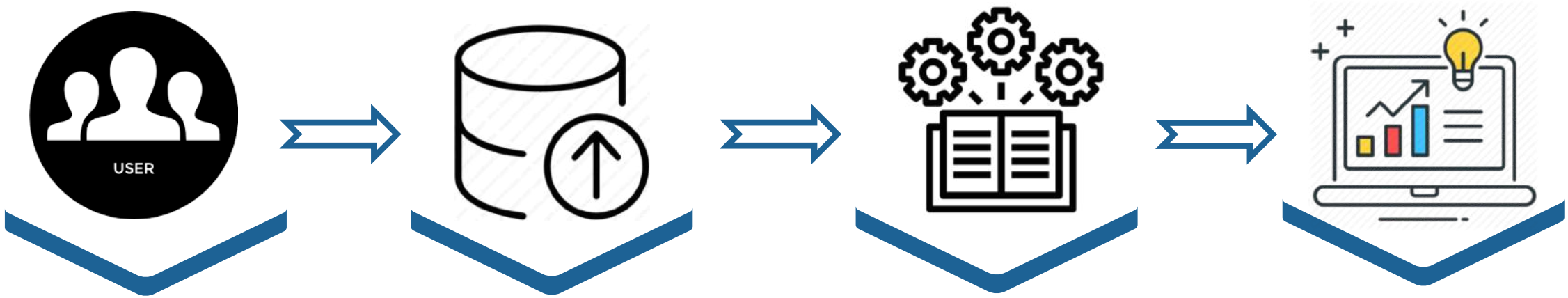
December 21, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780787

The I-BiDaaS project



The I-BiDaaS solution: Front-end



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- Fabricate Data
- Tokenize data

Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code
- Self-service: Select an algorithm from the pool
- Co-develop: custom end-to-end application

Results

- Visualize the results
- Share models

Benefits of using I-BiDaaS

Do it yourself In a flexible manner

Break data silos

Safe environment

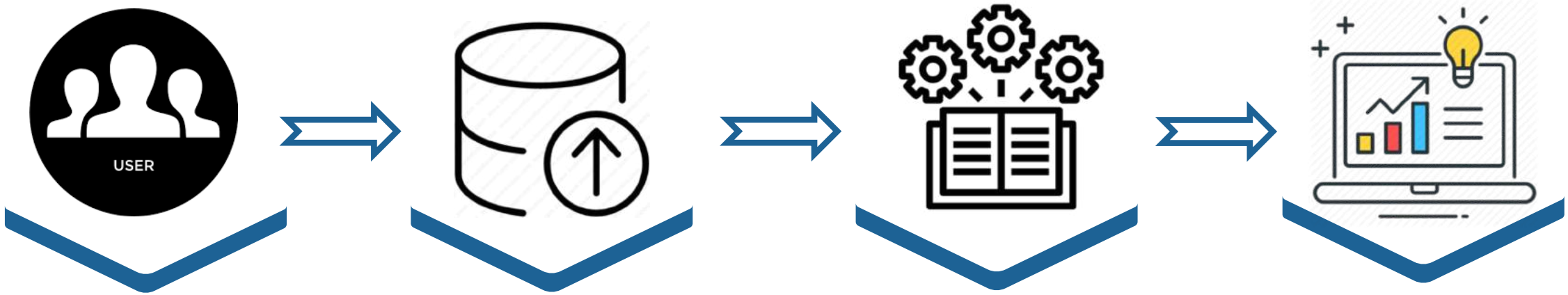
Interact with Big Data technologies

Increase speed of data analysis

Intra- and inter-domain data-flow

Cope with the rate of data asset growth

The I-BiDaaS solution: Front-end



Users

- **Expert mode**
- **Self-service mode**
- **Co-develop mode**

Data

- Import your data

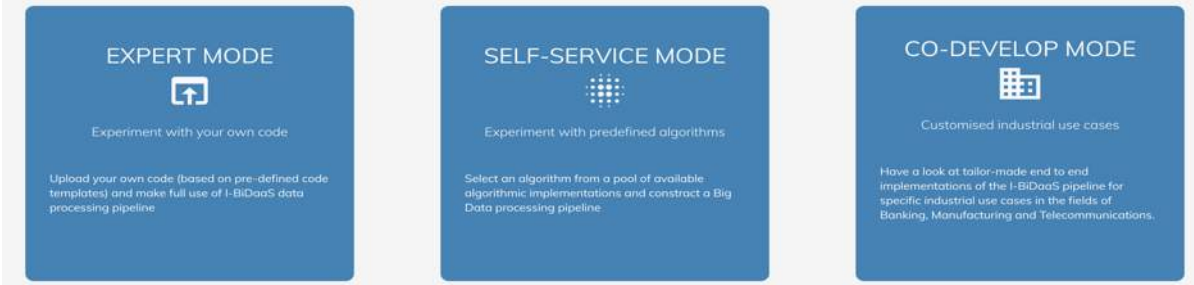
Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code

Results

- Visualize the results
- Share models

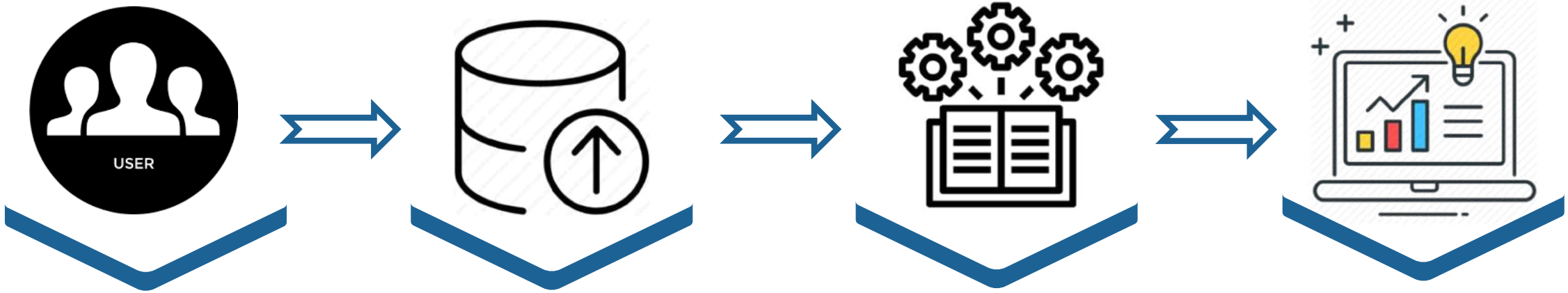
Flexible solution



Benefits of using I-BiDaaS

- Do it yourself In a flexible manner
- Break data silos
- Safe environment
- Interact with Big Data technologies
- Increase speed of data analysis
- Intra- and inter-domain data-flow
- Cope with the rate of data asset growth

The I-BiDaaS solution: Front-end



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- **Fabricate Data**
- Tokenize data

Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code
- Self-service: Select an

Results

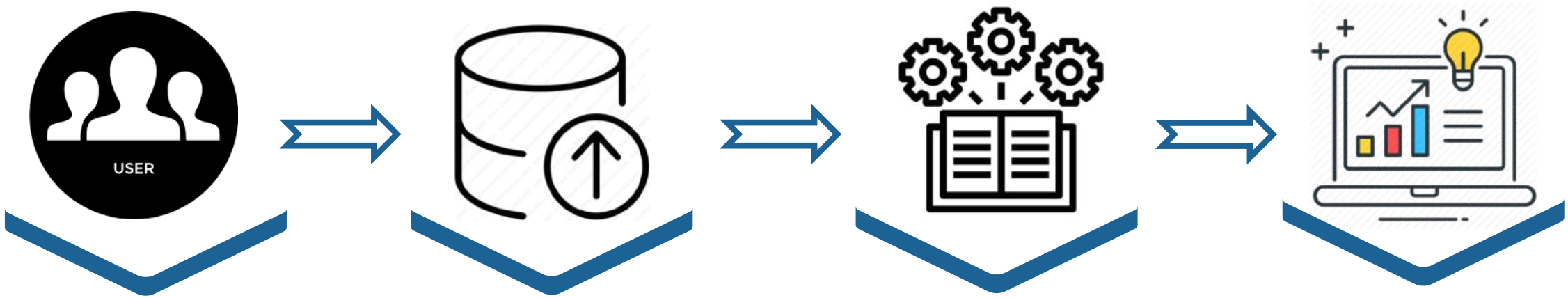
- Visualize the results
- Share models

Data sharing & breaking silos

Benefits of using I-BiDaaS

- Do it yourself In a flexible manner
- Break data silos
- Safe environment
- Interact with Big Data technologies
- Increase speed of data analysis
- Intra- and inter-domain data-flow
- Cope with the rate of data asset growth

The I-BiDaaS solution: Front-end



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- Fabricate Data
- Tokenize data

Analyze your Data

- **Stream & Batch Analytics**
- Expert: Upload your code
- Self-service: Select an algorithm from the pool
- Co-develop: custom end-to-end application

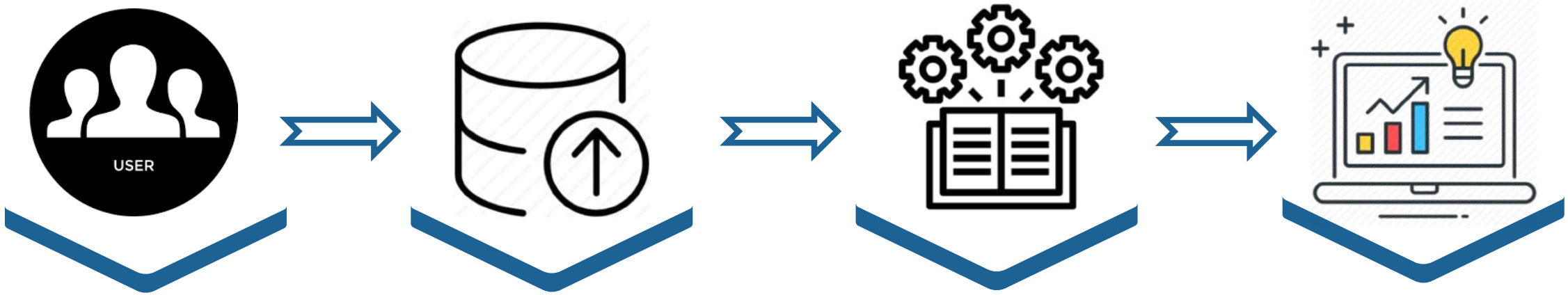
Results

- Visualize the results
- Share models

Benefits of using I-BiDaaS

- Do it yourself In a flexible manner
- Break data silos
- Safe environment
- Interact with Big Data technologies
- Increase speed of data analysis
- Intra- and inter-domain data-flow
- Cope with the rate of data asset growth

The I-BiDaaS solution: Front-end



Users

- Expert mode
- Self-service mode
- Co-develop mode

Data

- Import your data
- Fabricate Data
- Tokenize data

Analyze your Data

- Stream & Batch Analytics
- Expert: Upload your code
- Self-service: Select an algorithm from the pool
- Co-develop: custom end-to-end application

Results

- **Visualize the results**
- **Share models**

Benefits of using I-BiDaaS

Do it yourself In a flexible manner

Break data silos

Safe environment

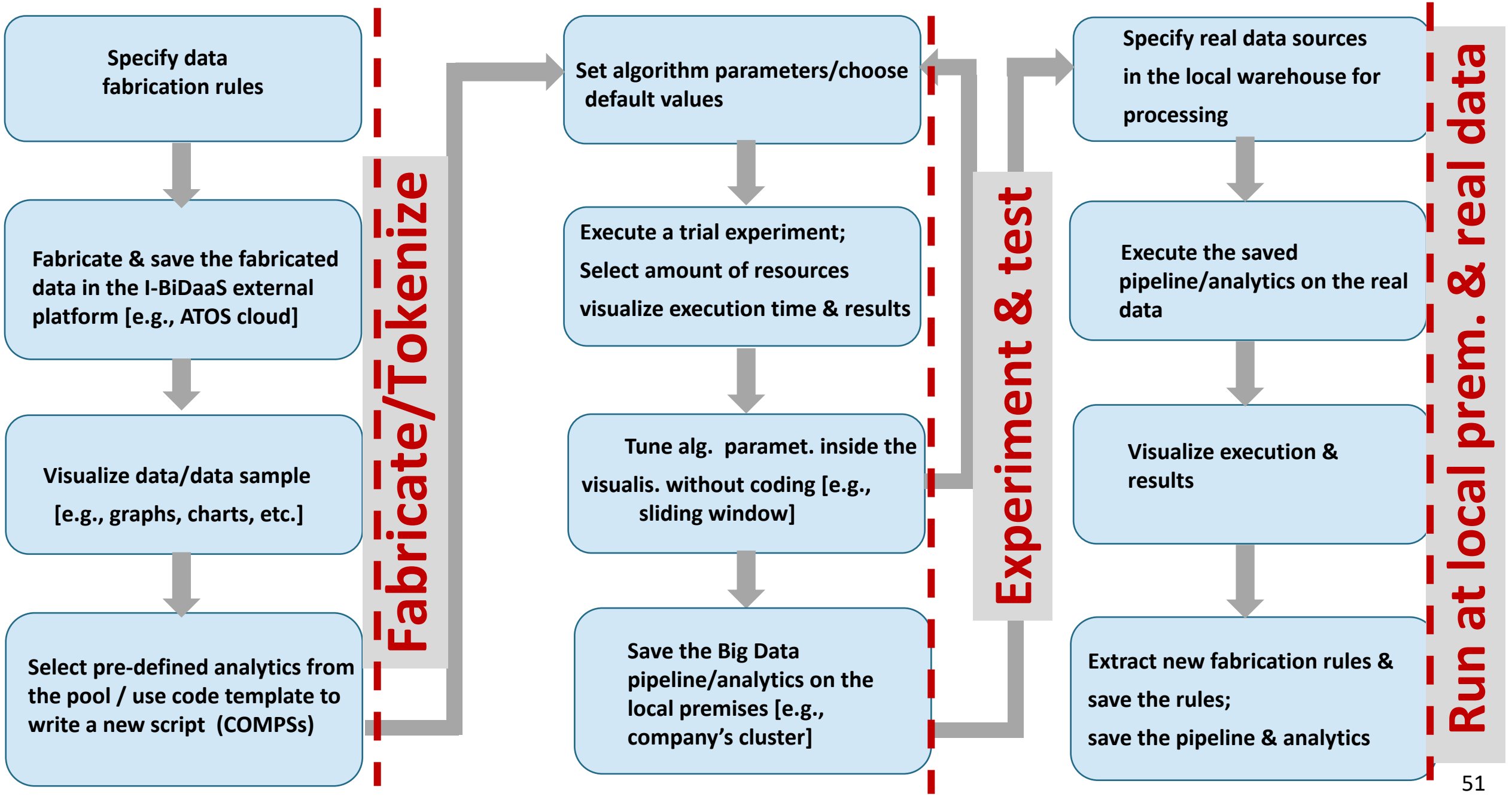
Interact with Big Data technologies

Increase speed of data analysis

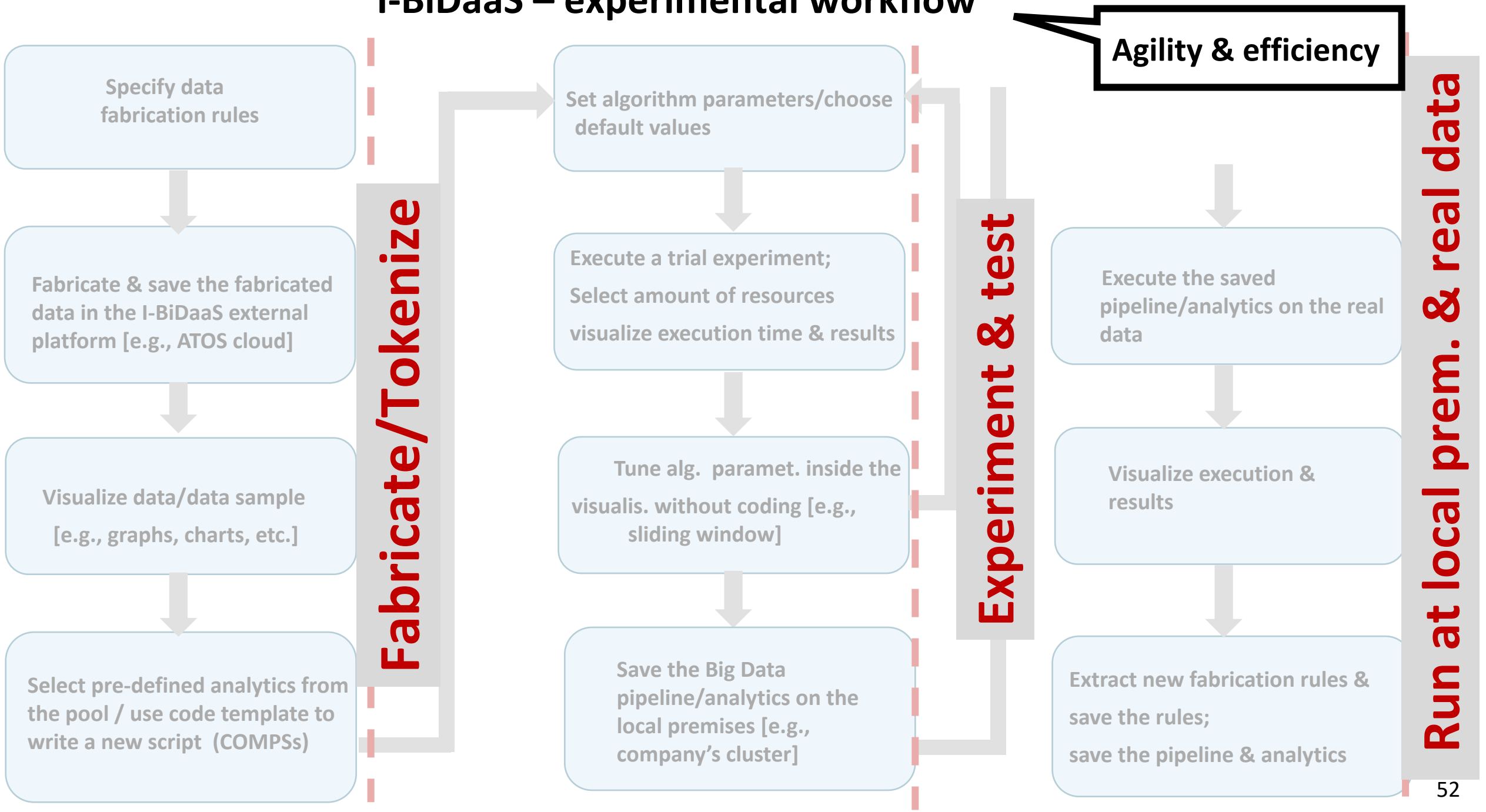
Intra- and inter-domain data-flow

Cope with the rate of data asset growth

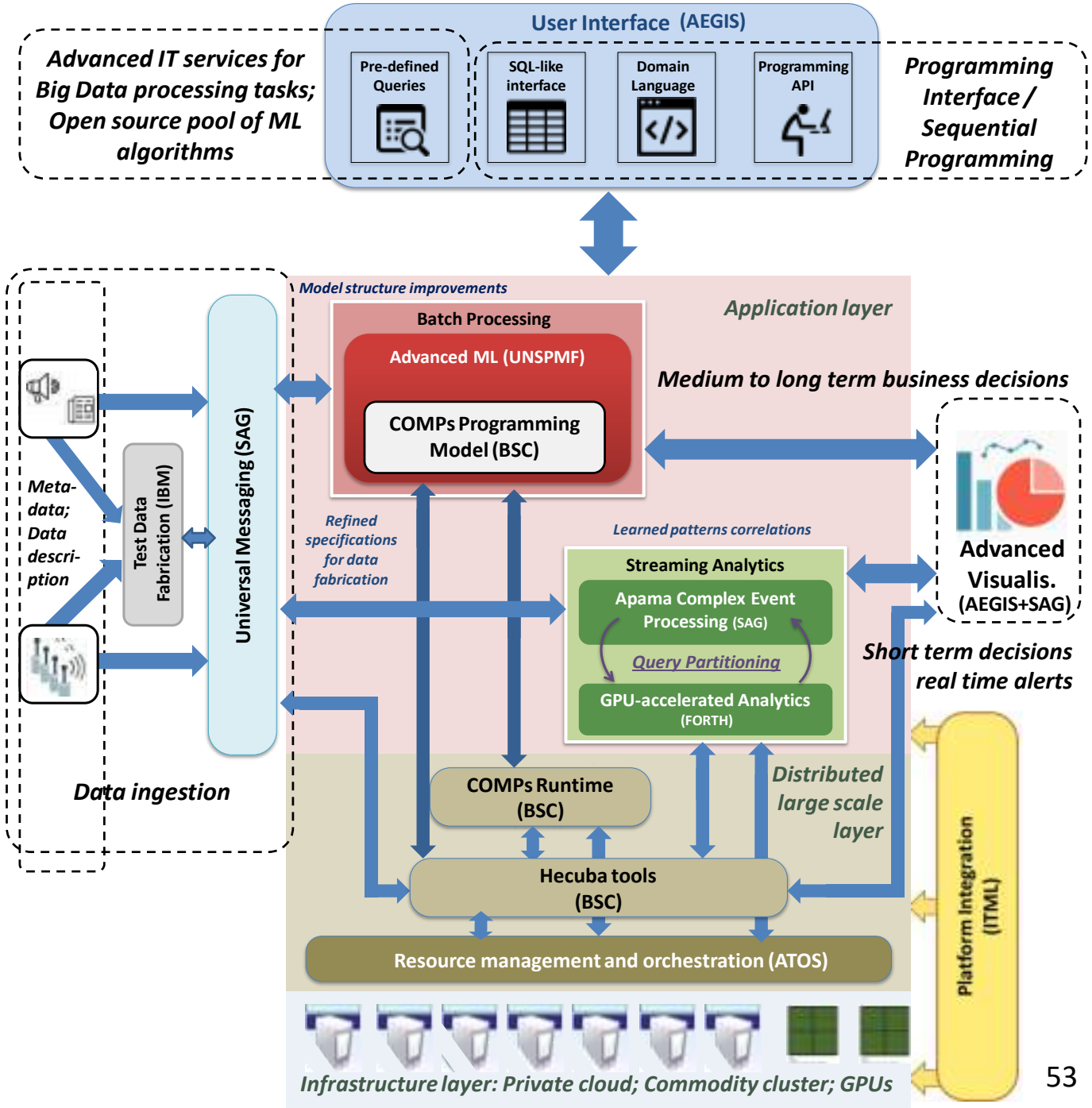
I-BiDaaS – Prototypical experimental workflow



I-BiDaaS – experimental workflow



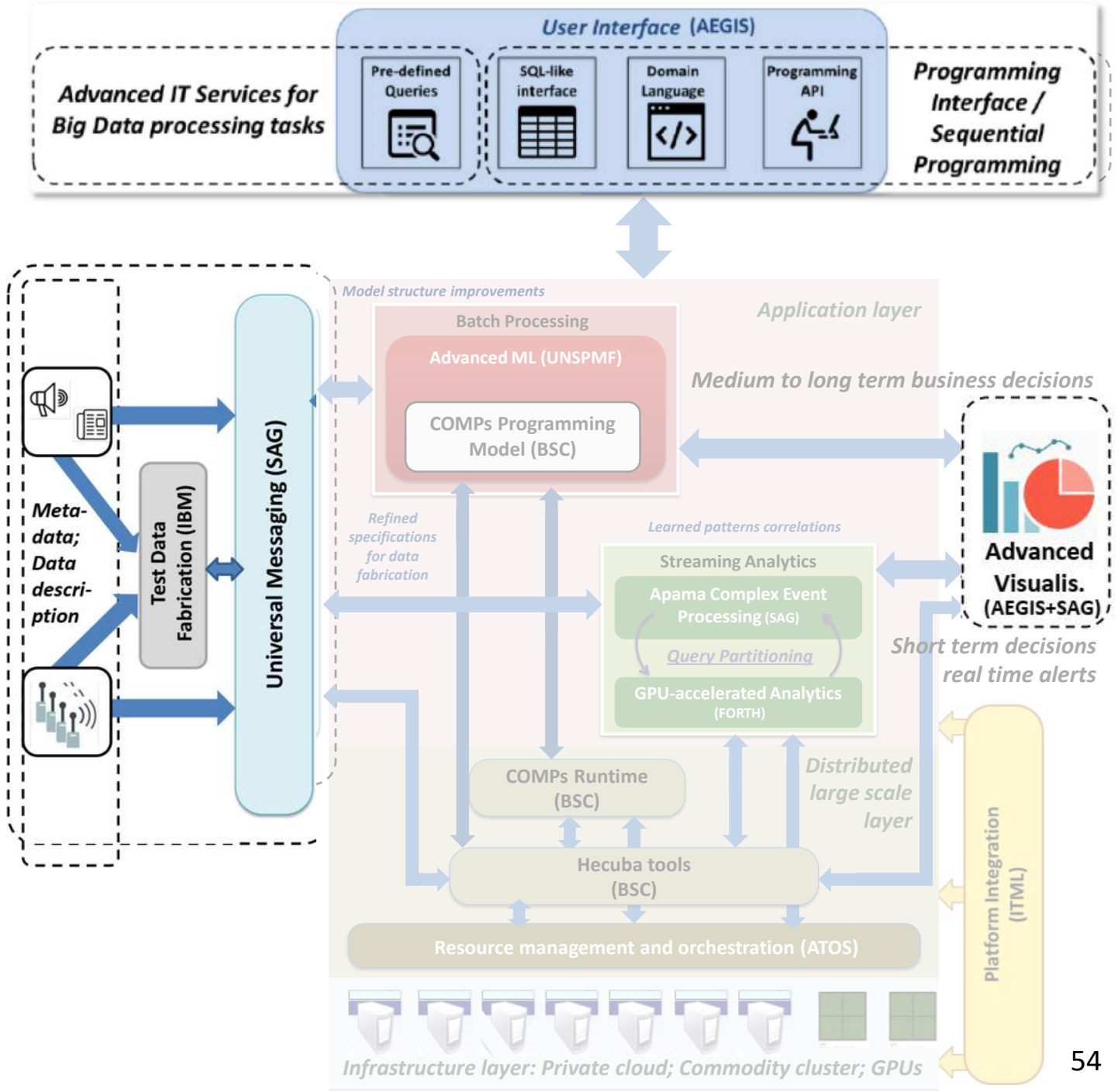
The I-BiDaaS solution: Architecture/back-end



Data, user interface, visualization

Technologies:

- IBM TDF
- SAG UM
- AEGIS AVT

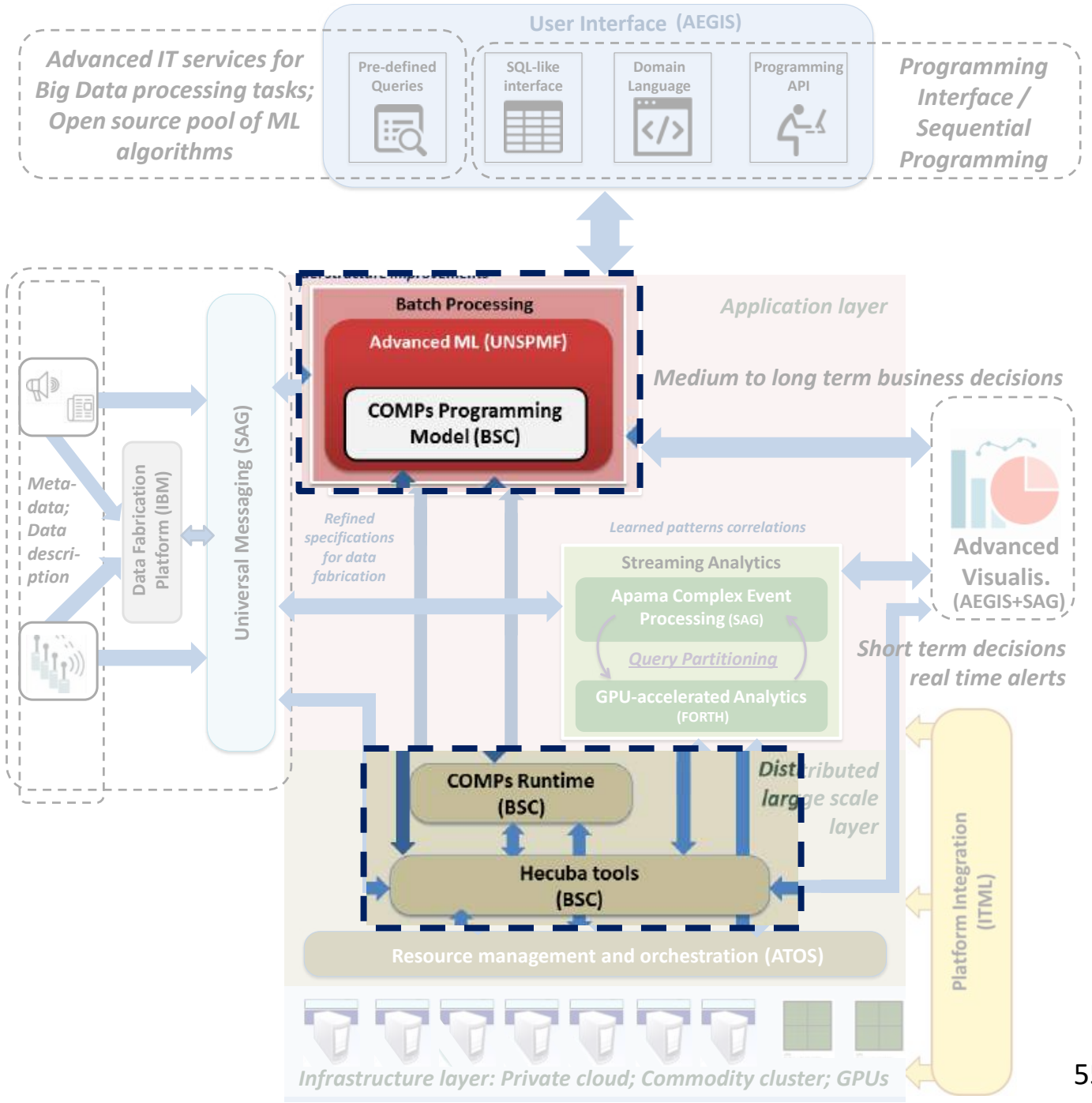


http://ibidaas.eu/tools

Batch analytics

Technologies:

- BSC COMPSs
- BSC Hecuba
- BSC Qbeast
- Advanced ML (UNSPMF)

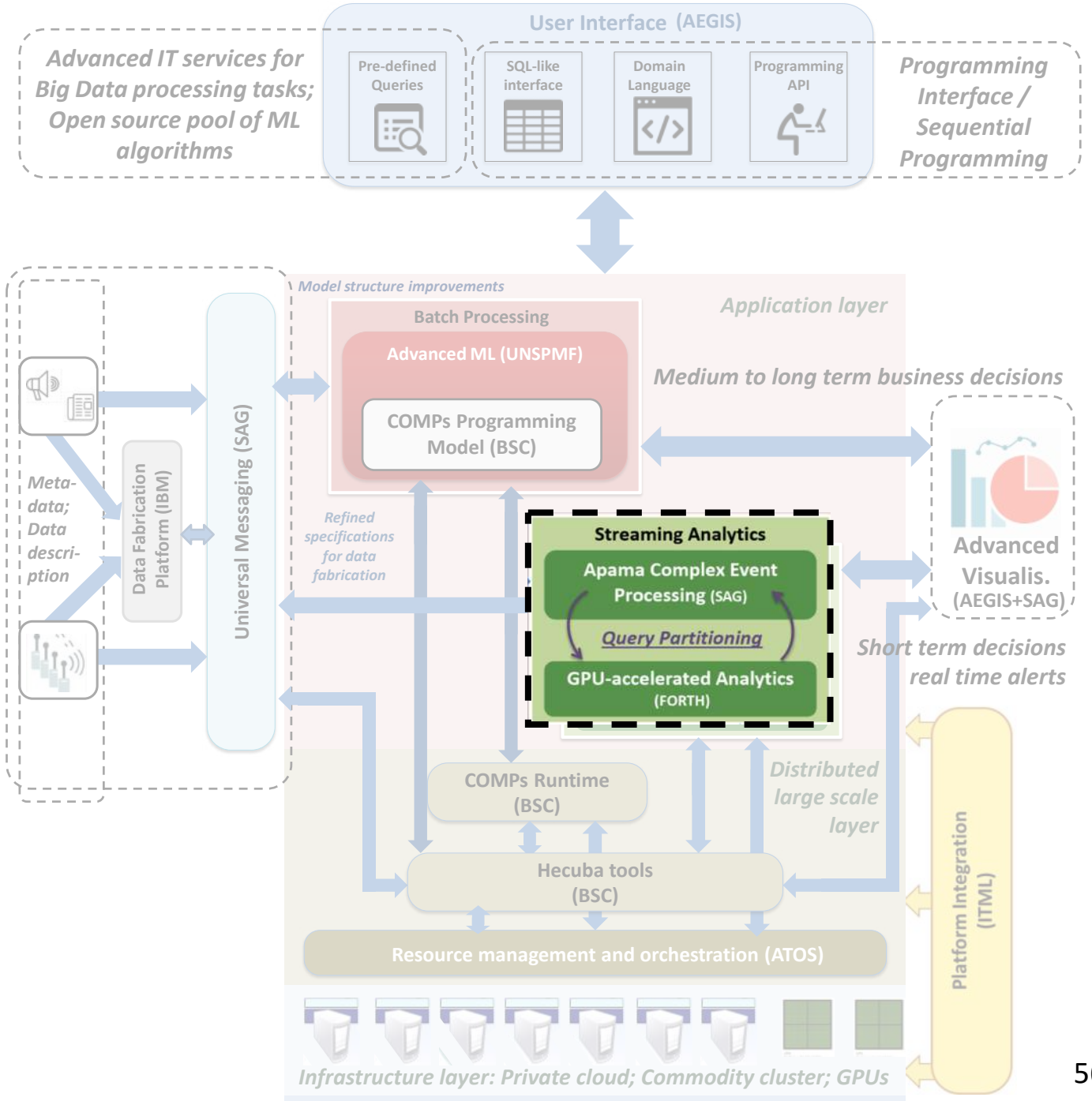


http://ibidaas.eu/tools

Streaming analytics

Technologies:

- SAG Apama CEP
- FORTH GPU-accel. analytics

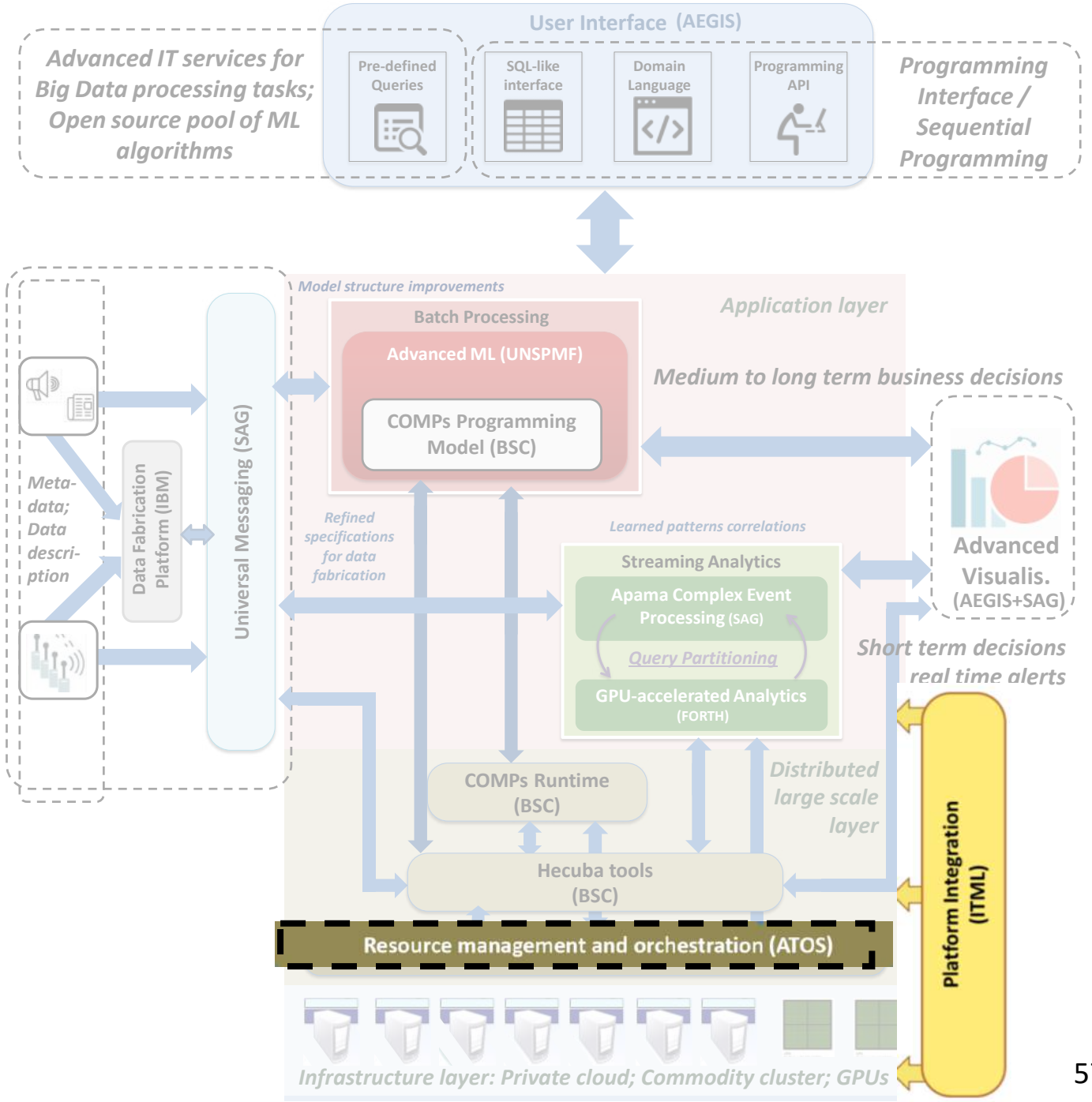


http://ibidaas.eu/tools

Resource mgmt & integration

Technologies:

- ATOS Resource mgmt
- ITML integration services



http://ibidaas.eu/tools

Key features & innovations

- **Data fabrication capabilities**
- Solution flexibility
- Easy to code programming paradigm
- High code reusability

CAIXA: From 3 months to 1-2 weeks for a new proof-of-concept Big Data technology

Key features & innovations

- Data fabrication capabilities
- **Solution flexibility**
- Easy to code programming parad
- High code reusability

CAIXA: Bank transfers use case: ~3-4 less time to analysis; found more useful anomalies than before the project.

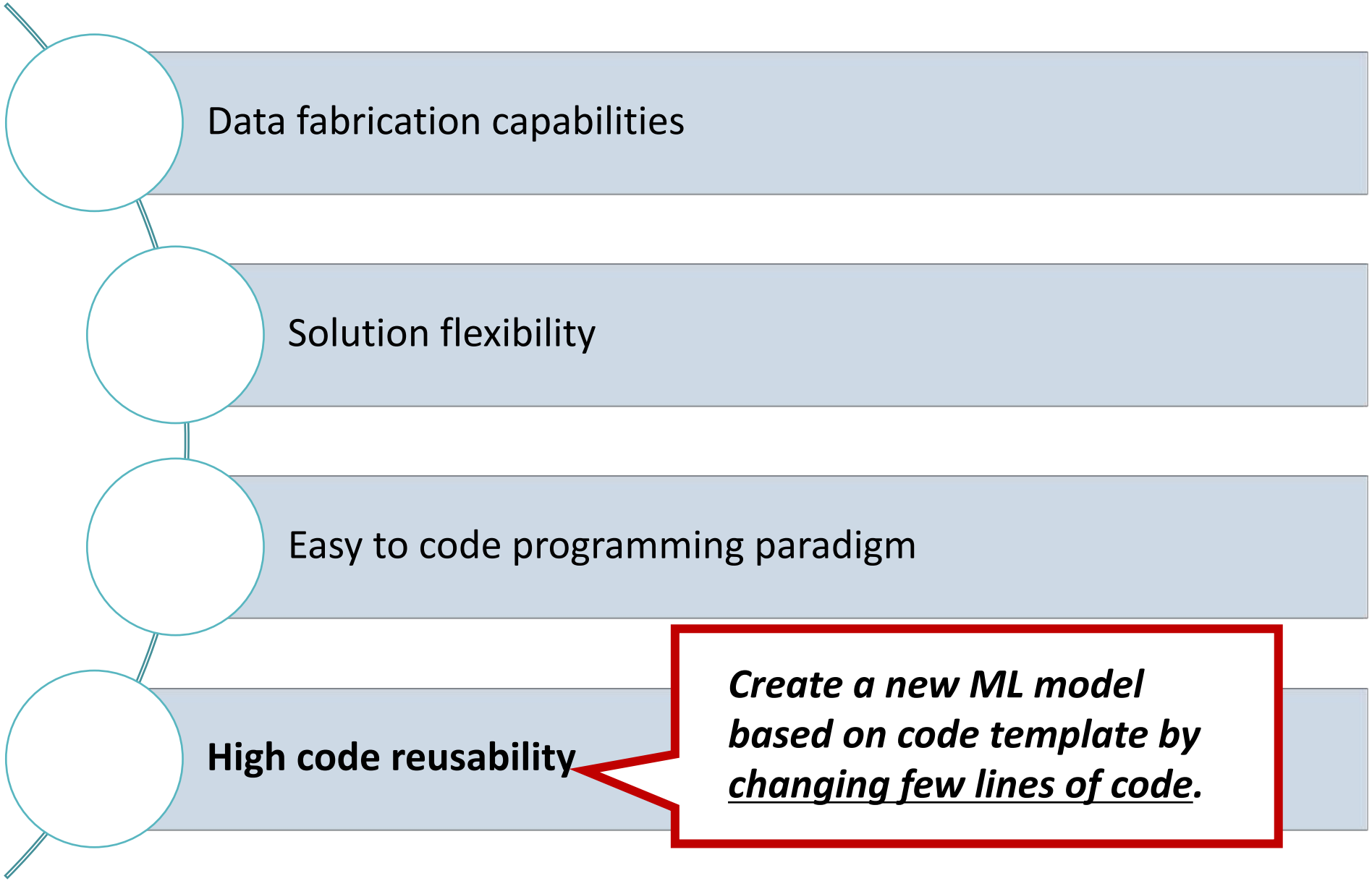
Key features & innovations

- Data fabrication capabilities
- Solution flexibility
- **Easy to code programming paradigm**
- High code reusability

40% reduction of code length thanks to COMPSs.

Example: ADMM-Lasso in COMPSs.

Key features & innovations



Key features & innovations (Cont'd)

- **Synergy of CEP and GPU-accelerated analytics for streaming data**
- Feedback from analytics to data fabric
- Feedback from analytics to problem modelling
- Demonstrated on 8 use cases across 3 different data providers and 3 different industries

30%-60% better throughput compared to the vanilla implementation (CPU only)

Key features & innovations (Cont'd)

- Synergy of CEP and GPU-accelerated analytics for streaming data
- **Feedback from analytics to data fabrication**
- Feedback from analytics to problem modelling
- Demonstrated on 8 use cases across 3 different data providers and 3 different industries

Automatic rule generation based on descriptive analysis with Hecuba

Key features & innovations (Cont'd)

- Synergy of CEP and GPU-accelerated analytics for streaming data
- Feedback from analytics to data fabrication
- **Feedback from analytics to problem modelling**
- Demonstrated on 8 use cases across 3 differe
  3 different industries

Feedback from users collected with Qviz/Qbeast is used to guide ML algorithms

Key features & innovations (Cont'd)

- Synergy of CEP and GPU-accelerated analytics for streaming data
- Feedback from analytics to data fabrication
- Feedback from analytics to problem modelling
- **Demonstrated on 8 use cases across 3 different data providers and 3 different industries**

Key features & innovations (Cont'd)

- Synergy of CEP and GPU-accelerated analytics for streaming data
- Feedback from analytics to data fabrication
- Feedb
- Dem
  different i

5 I-BiDaaS innovations accepted and promoted by the European Commission as Excellent Innovations.

5 I-BiDaaS innovations published at the BDVA innovation marketplace

Excellent Innovations: EU Innovation Radar

Specification of an end-to-end Big Data as-a-self-service platform

Market maturity: Exploring; ATOS, UNIMAN, UNSPMF

Advanced Visualization Toolkit (AVT) supporting scalable data visualisation

Market maturity: Market ready; AEGIS

ADMM Machine Learning Algorithms

Market maturity: Exploring; BSC, UNSPMF

Multidimensional Storage with Efficient Sampling (MuSES)

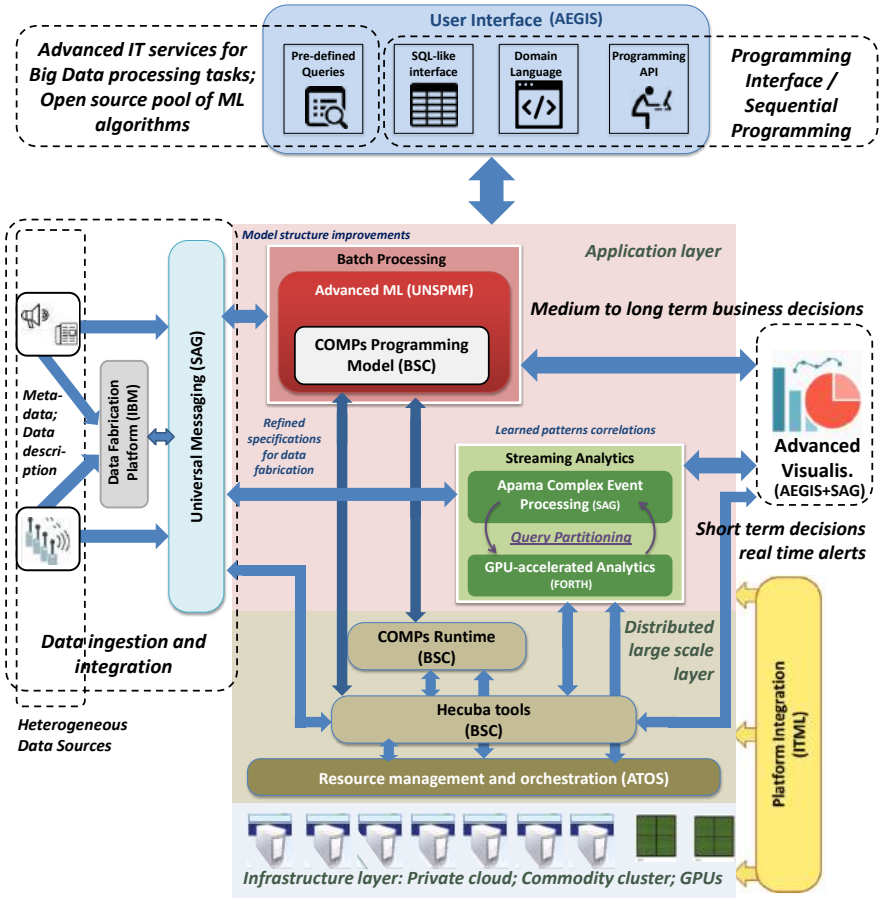
Market maturity: Market ready; BSC

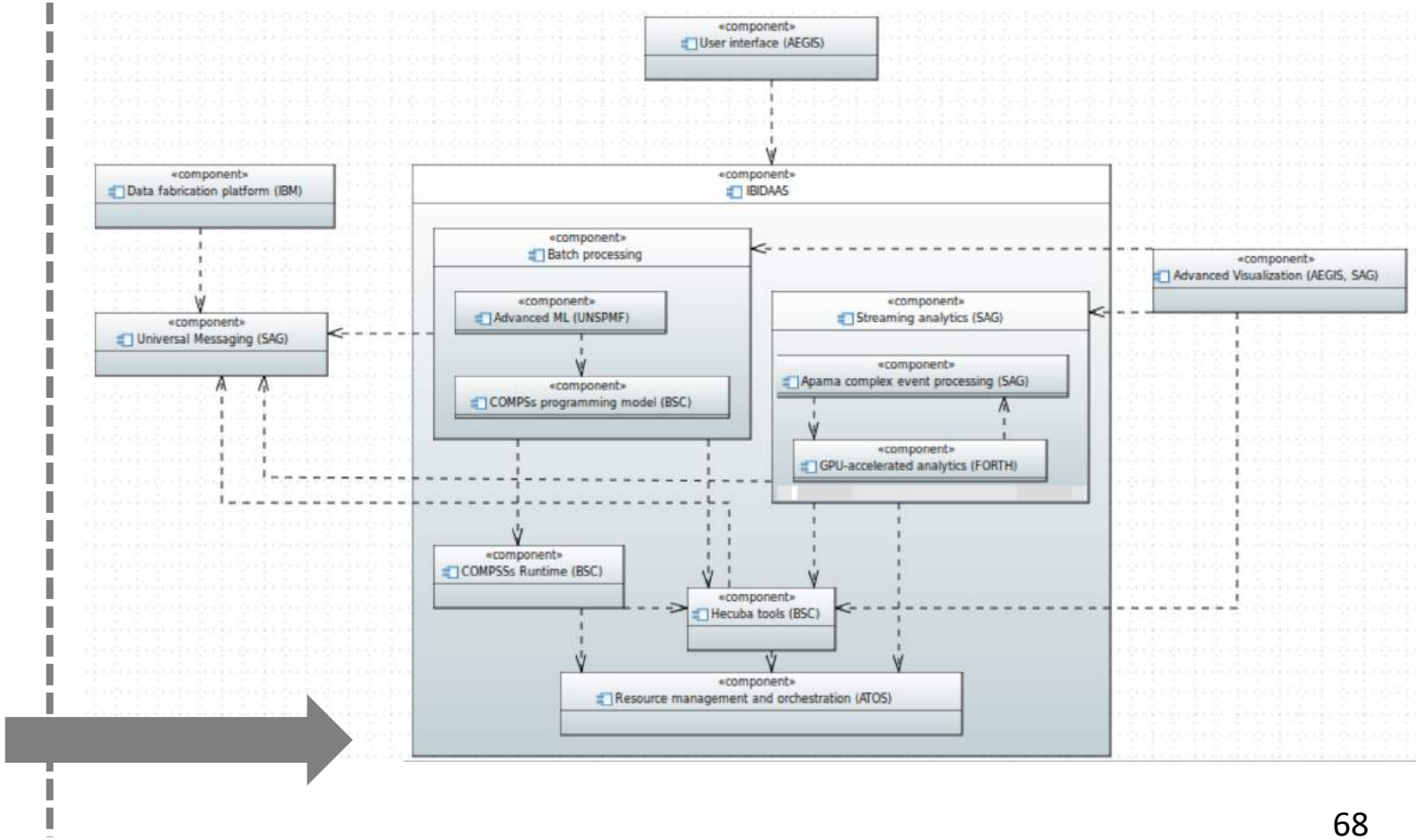
Parallelization of a constraint satisfaction solver within TDF

Market maturity: Market ready; IBM

Excellent Innovations: EU Innovation Radar

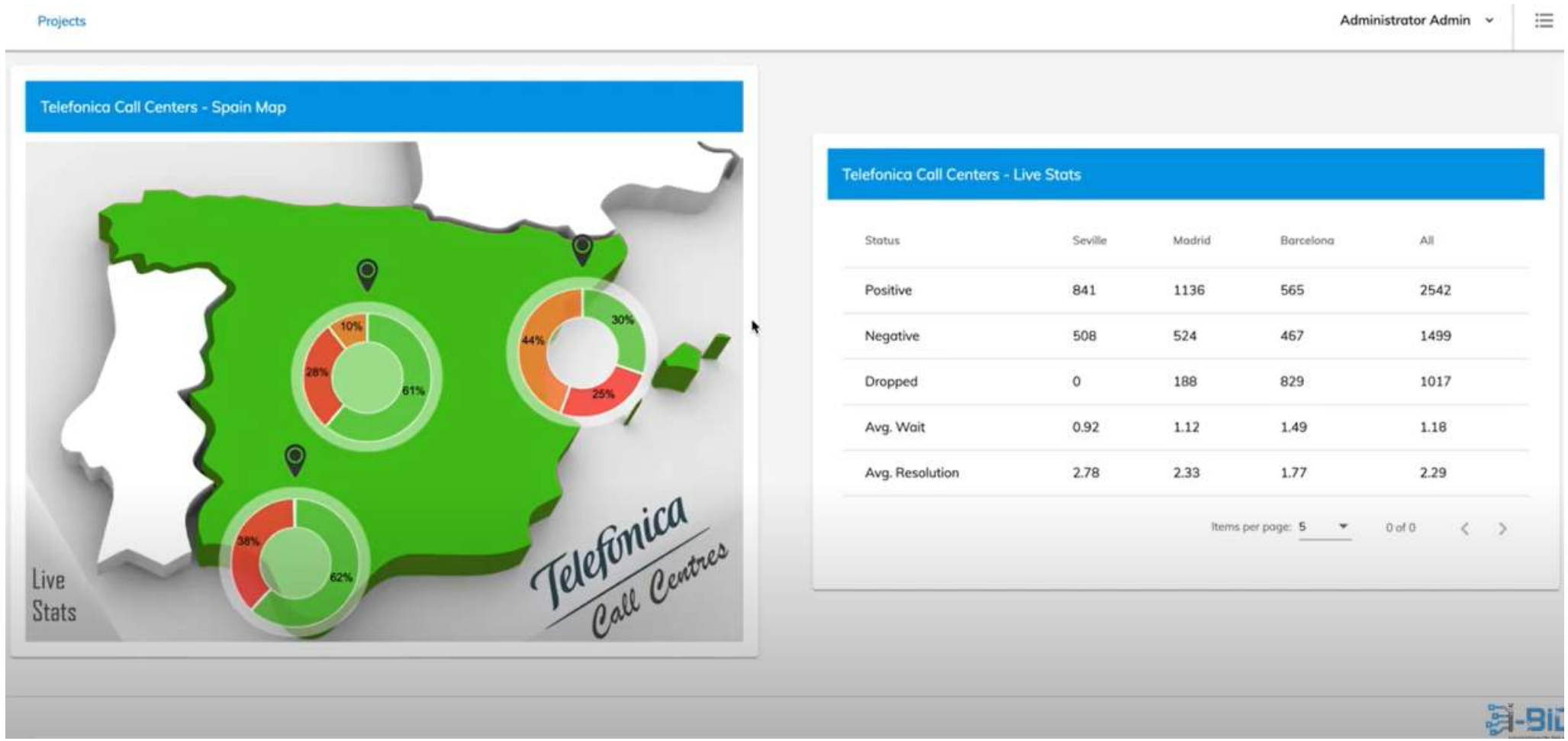
Specification of an end-to-end Big Data as-a-self-service platform





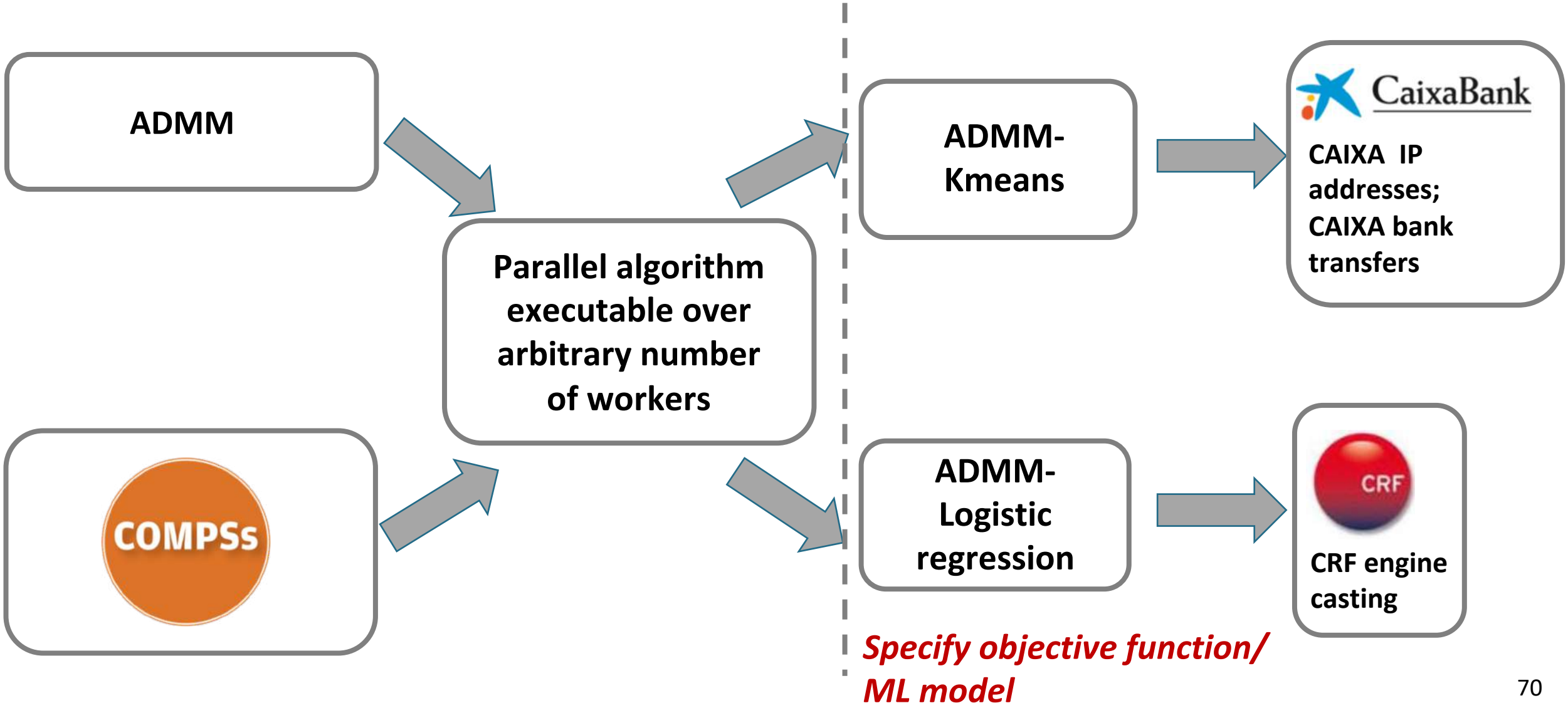
Excellent Innovations: EU Innovation Radar

Advanced Visualization Toolkit (AVT) supporting scalable data visualisation



Excellent Innovations: EU Innovation Radar

ADMM Machine Learning algorithms



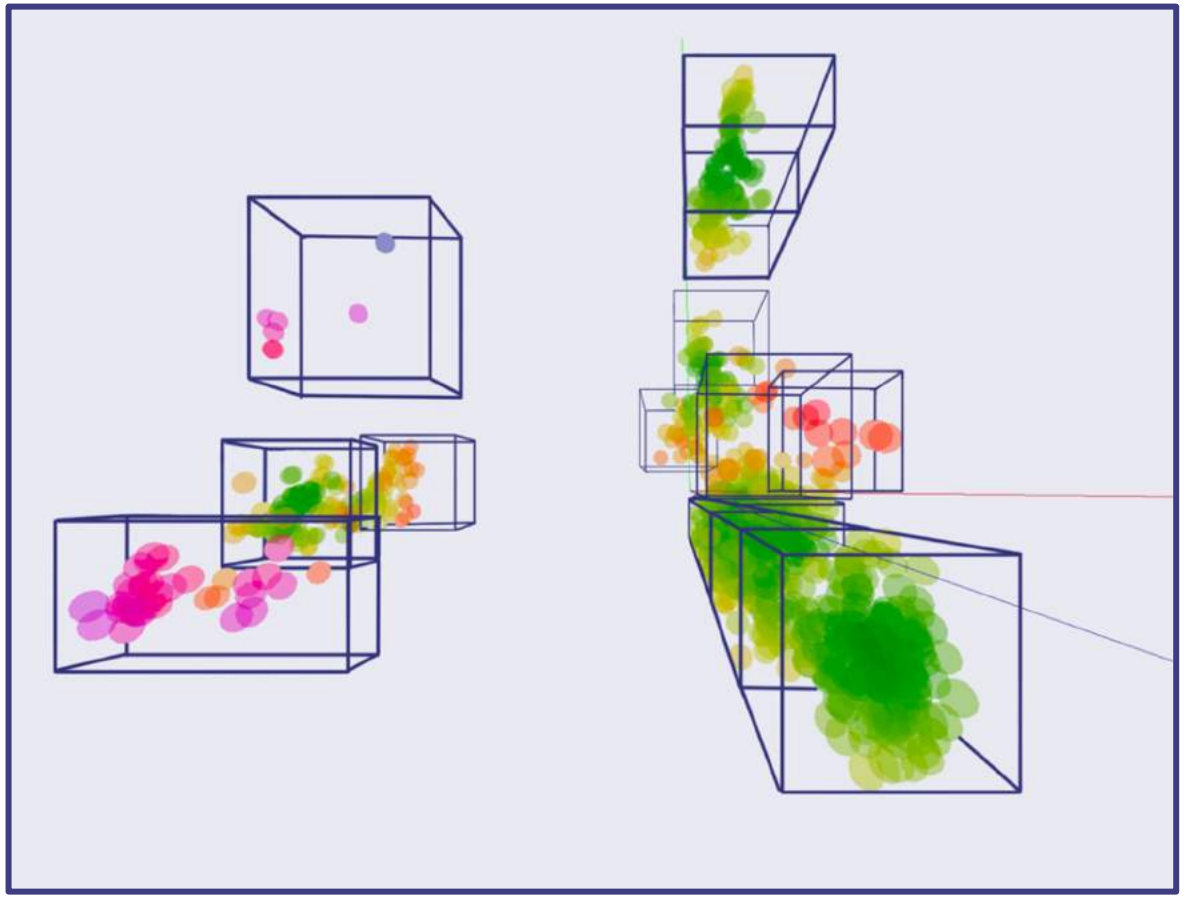
Excellent Innovations: EU Innovation Radar

Multidimensional Storage with Efficient Sampling (MuSES)

Using **MuSES** improves:

1. **initialization**
   initial solution on a sample
2. **convergence**
   better data partitioning
3. **scalability**
   convergence on sample, less data to analyse

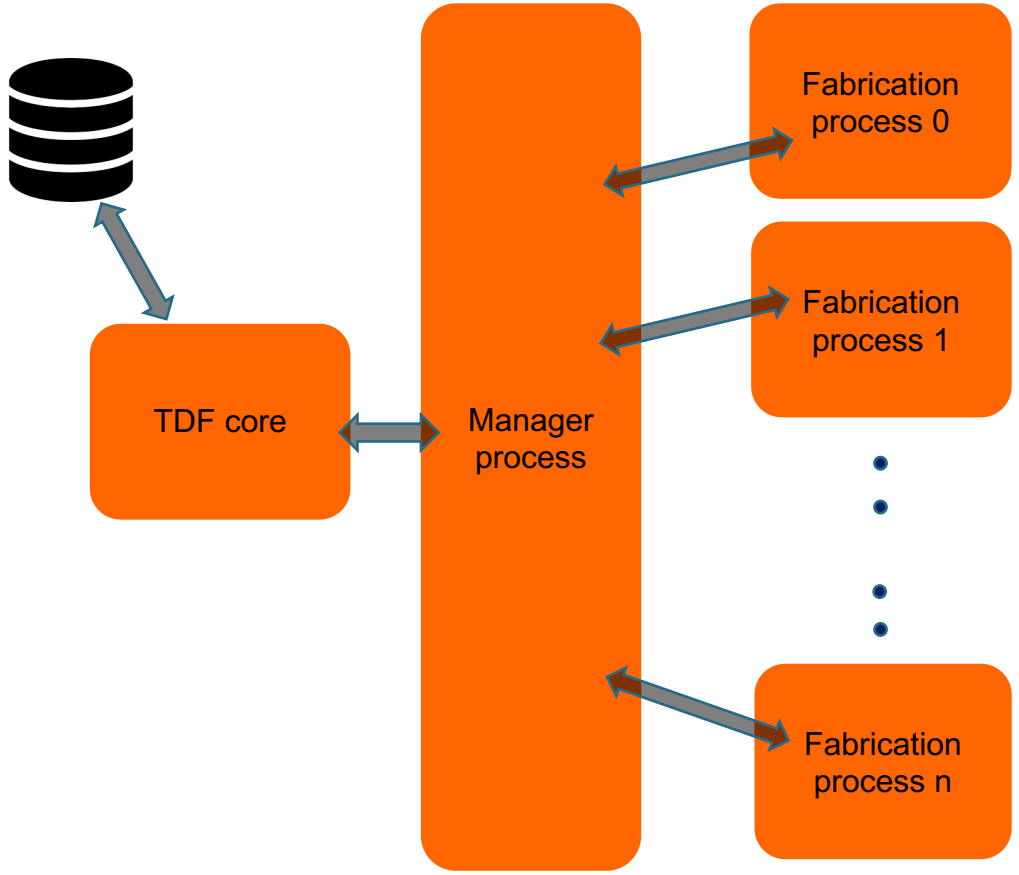
Plus: Allows Interactive visualization



Excellent Innovations: EU Innovation Radar

Parallelization of constraint satisfaction solver within TDF (IBM)

- Simultaneously solve CSP problems to concurrently create values for DB table rows.
- A single manager creates several fabrication processes
- Manager responsibilities
  - Handling data dependencies between different CSP problems
  - Communication with TDF core
  - Sharing DB connections
- Child responsibilities
  - Solving a single CSP problem
  - Fabricating data for a singe table row



Scientific advances – papers

- I-BiDaaS architecture specification & evaluation (4 conference papers) — D1.2; D1.3; D7.7
- Scientific advances on batch analytics (3 journal papers; 6 conference papers) — D3.1-D3.3
- Studies on data sharing & privacy risks (3 conference papers) — D7.3, D7.5, D7.7
- 2 book chapters on lessons learnt w.r.t. Big Data analytics in banking & manufacturing — D6.5
- ...

- 3 Journals
- 18 conf. papers
- 3 posters
- 1 PhD thesis
- 2 book chapters

Real data; sharing; breaking silos

- CRF hackathon; TID hackathon
- CAIXA: 3 real tokenized data sets (CAIXA IP; Bank transfers; Online banking)
- CRF: 4 real anonymized data sets (SCADA, MES, Thermal Images, Process and quality data)
- TID: 3 real anonymized data sets (Mobility, Antenna KPIs, Call centers)
- CAIXA workshop

I-BiDaaS use cases

I-BiDaaS Dataset	Use Case	Partner owner (Domain)	Accessibility
Mobility Data (Real & Synthetic)	Accurate location prediction with high traffic and visibility Optimization of placement of telecommunication equipment	**TID** (Telecom)	Synthetic: Open Access Real: Confidential (In-house)
Call Center Data (Real & Synthetic)	Quality of Service in Call Centers	**TID** (Telecom)	Synthetic: Open Access Real: Confidential (In-house)
Online Banking Control (Real Tokenized)	Enhance control of customers to online banking	**CAIXA** (Banking)	Open Access
Bank Transfer (Real Tokenized)	Advanced analysis of bank transfer payment in financial terminal	**CAIXA** (Banking)	Open Access
IP Address (Synthetic & Real Tokenized)	Analysis of relationships through IP addresses	**CAIXA** (Banking)	Synthetic: Open Access Real: Open Access
SCADA (Real Anonymized) **MES** (Real Anonymized)	Maintenance and monitoring of production assets	**CRF** (Manufacturing)	Open Access
Aluminum Die-casting *Thermal Images, Process and quality data* (Real Anonymized & Synthetic)	Production process of aluminium Die-casting	**CRF** (Manufacturing)	Synthetic: Open Access Anonymized: Confidential (consortium)

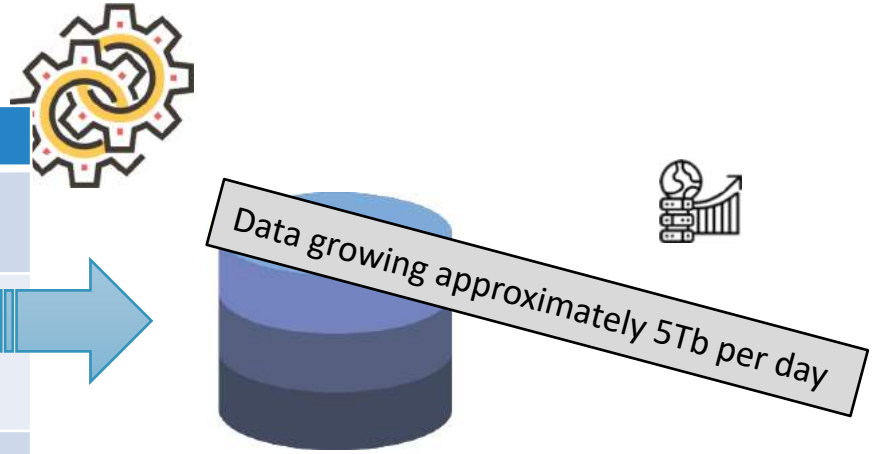
CAIXA: Analysis of relationships through IP addresses

CAIXA 1st case

MVP use case

Find relationships between different customers analysing their bank movements in order to prevent fraud scenarios and bad practices among employees.

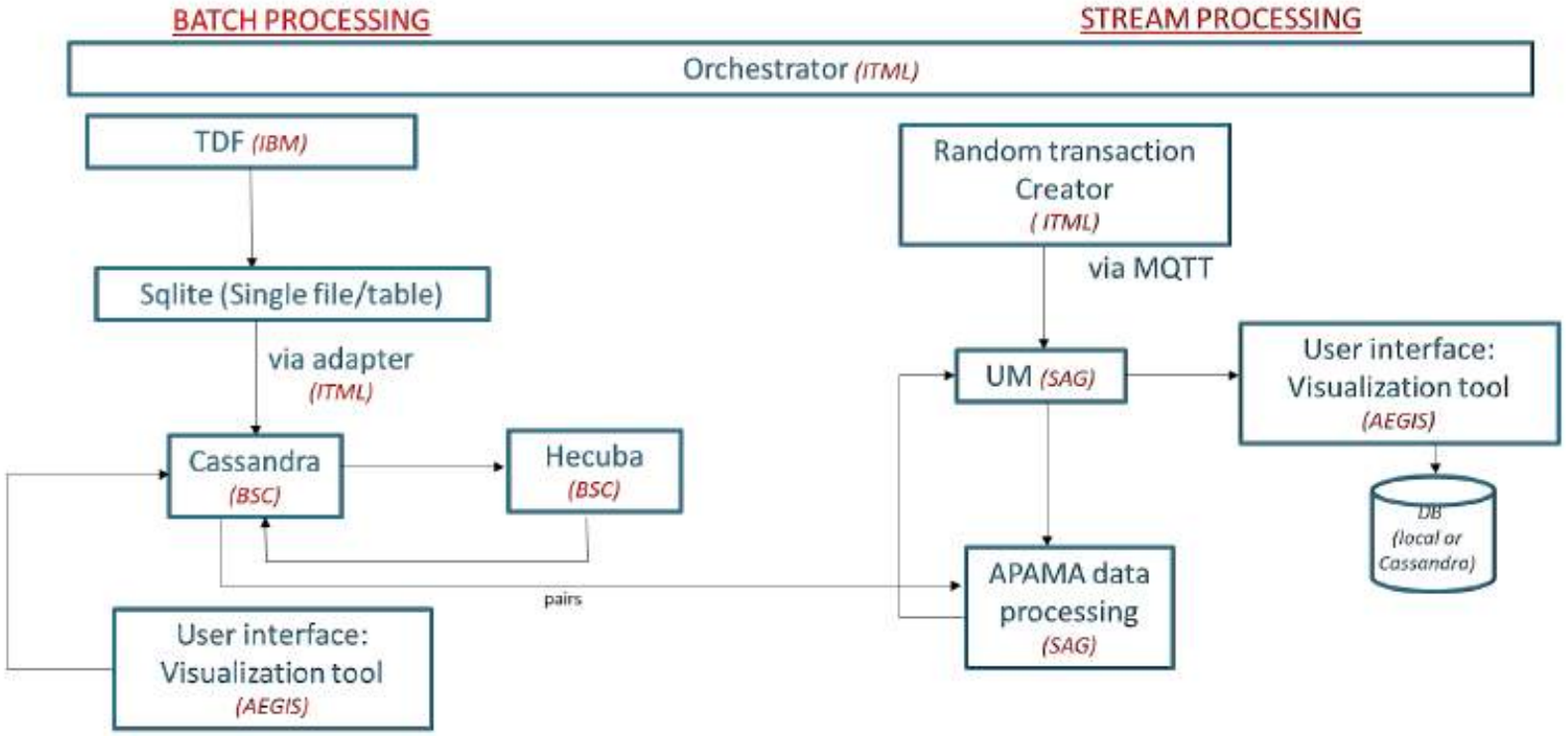
Attribute	Description	Format	Example
FK_NUMPERSO	Identifier of the Person. (Foreign Key).	NUMBER	34523454
PK_ANYOMESDIA	Date (YYYYMMDD) when the connection of the user. (Primary Key).	NUMBER	20180823
IP_TERMINAL	IP Address of the connection of the user.	VARCHAR2	10.8.2.22



CAIXA: Analysis of relationships through IP addresses

- **I-BiDaaS approach:**
  - Synthetic & real tokenized data generated
  - COMPSs script for finding relationships
  - **Batch & streaming integrated solution**
  - **Kmeans & DBSCAN for clustering of users into groups**
  - Validation of the synthetic data by repeating the same analysis on synthetic & real data

D3.2, D5.2, D6.4

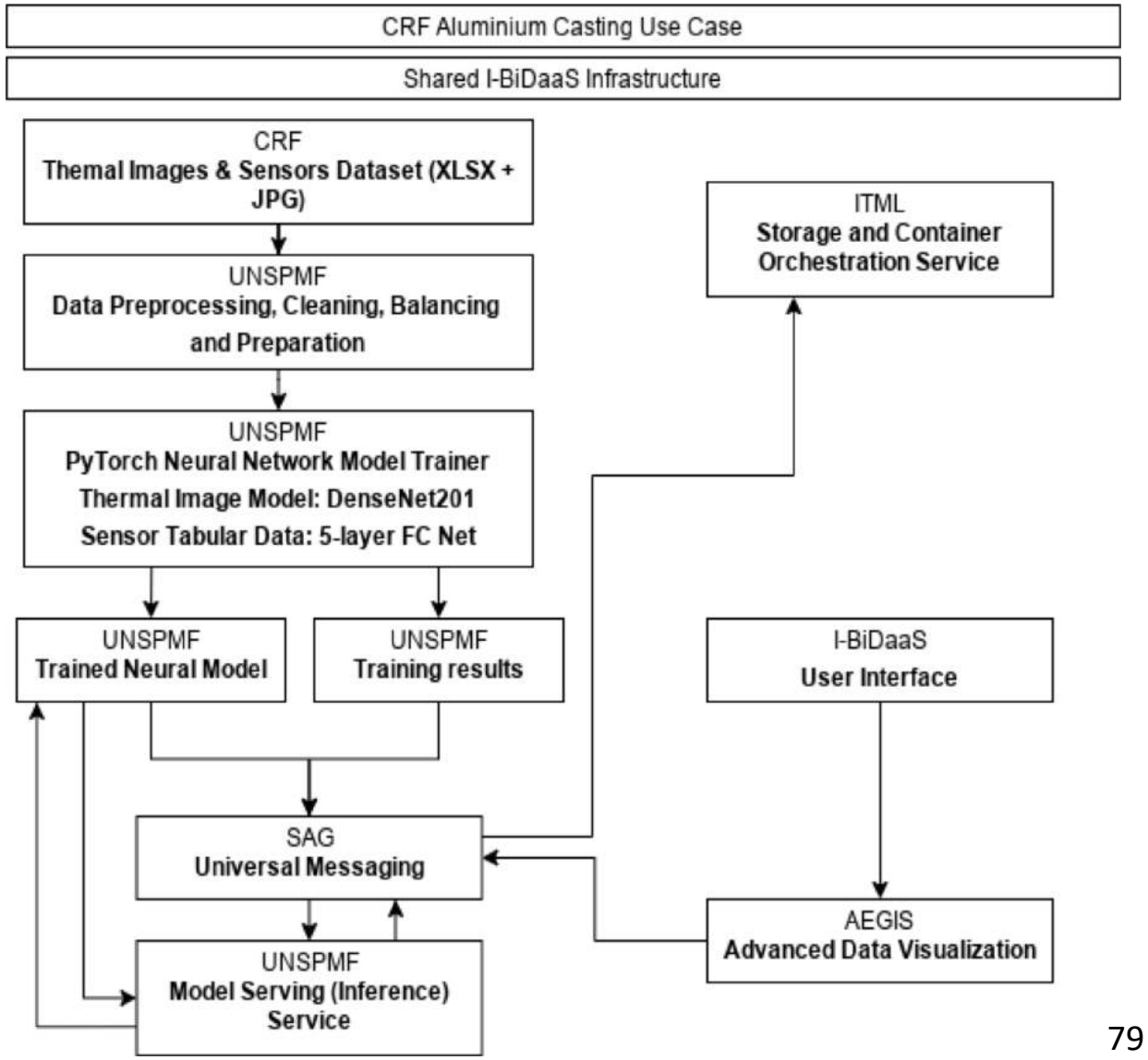


- **Production Process of Aluminium Die Casting**
  - During the casting process of the engine blocks, molten aluminium is injected into the die cavity
  - Can we classify the quality levels of engine blocks correctly at the time they are produced?
  - Can we improve the production process, by increasing the number of high quality engine blocks and decreasing the number of lower quality engine blocks?

CRF: Production process of aluminium die-casting

D3.3, D6.4

- **I-BiDaaS approach:**
  - **Joint neural network model with CNN (thermal images) and FCNN (sensor data)**
    - CNN (DenseNet201) for images – pretrained & non-trained model
    - Fully Connected NN for sensor data
  - **Random under-sampling** to produce a balanced dataset
  - **Good accuracy** – up to 73%



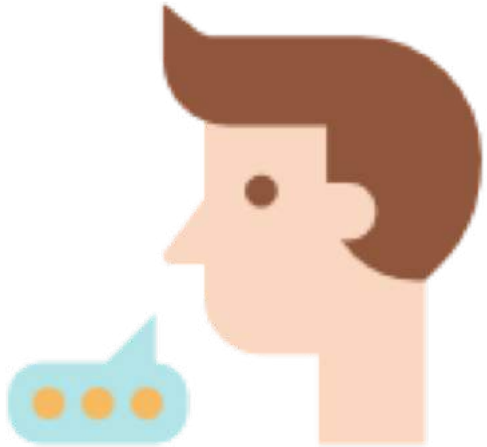
TID: Quality of service in Call Centers

- **Call centers, example of use cases and KPIs**

  - Wide variety of the nature of phone calls: to ask for service and product information, report technical problems, to follow-up with a purchase, to provide feedback, etc.

  - Quickly get familiar and understand customer's perspective and main interests to provide fast response and improve customer service by using big data speech and language analytics.

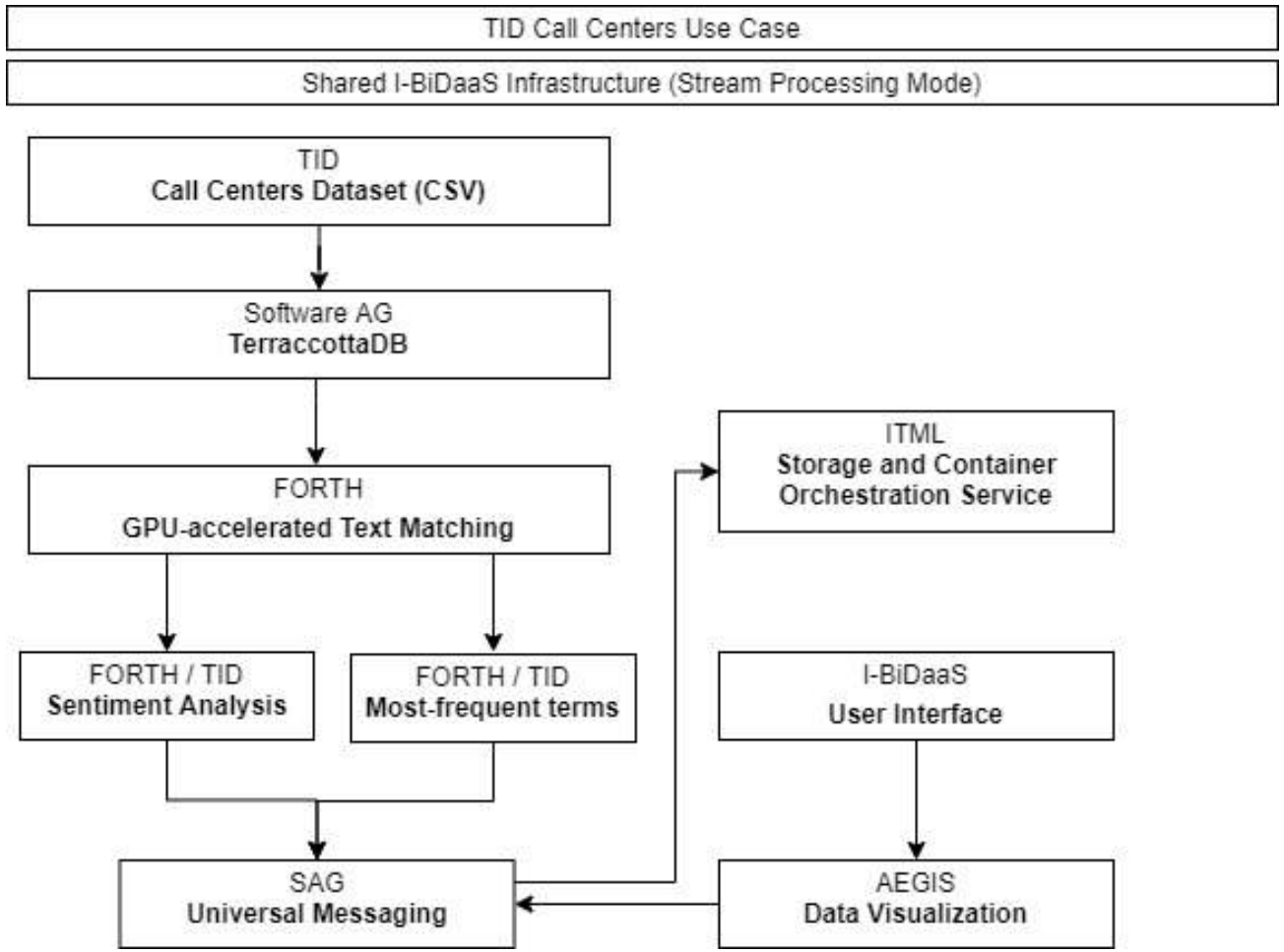
  - Shorten call durations, waiting time and First Call Resolution (FCR) time by anticipating customer's situation based on previous insights, e.g., using the aggregation of previous analytics by Call Centers or regions.



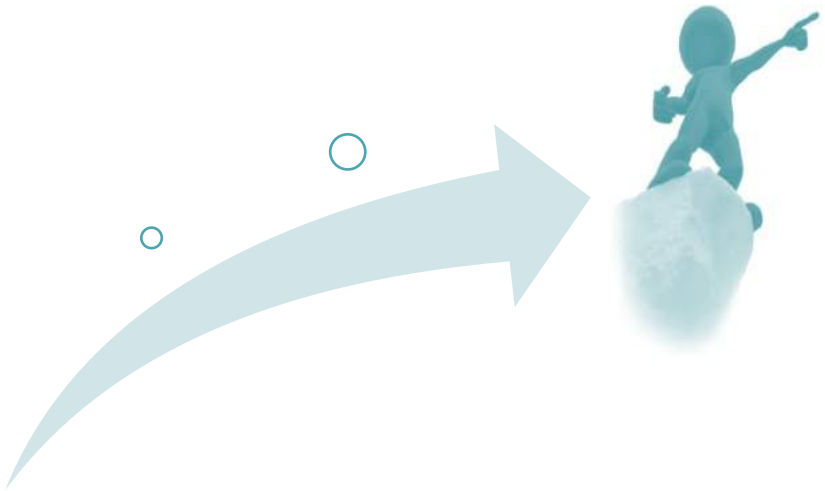
TID: Quality of service in Call Centers

D4.3, D6.2, D6.3

- **I-BiDaaS approach**
  - **GPU-accelerated keywords matching** that estimates a sentiment score in the incoming data stream, aggregating them by call centers/regions and by a given time window
  - Correlation between main sentiment from the calls with Customer Satisfaction Index
  - Top-K frequent words or 2-grams as quick semantic overview of the CC current operation
  - **High efficiency:** 42-48K calls per second



Summary & outlook



- Open-source I-BiDaaS solution variant + User guide
- Open source scalable ML implementations (I-BiDaaS knowledge repository)
- Commercial platform offerings devised
- Qbeast spin-off
- Garage lab experimental platform @ CaixaBank
- Experimental training environment @ CRF
- Internal innovation call @ TID

Industrial-Driven Big Data as a Self-Service Solution

I-BiDaaS application to the Banking Sector

Ramon Martín de Pozuelo (CaixaBank)

I-BiDaaS Final Event

December 21, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780787

CaixaBank and the Use of Data

CaixaBank is the **leading financial group in Spain**, both in banking and insurance and it is developing a strategy of diversification with stakes in international banks and also within leading service companies.

+14M Customers

+7M On-line banking

+5M Mobile Banking

32K Employees

+9K ATMs

4.2K Branches

In January 2014 CaixaBank created the Big Data Department.

- We currently manage more than 4PB of information only in our Big data.
- Department of more than 100 internal people providing data analytics services to all the organization.
- Due to regulation constraints all the infrastructure and the analysis is done internally.

Why do we need I-BiDaaS?

Current data sharing situation in CaixaBank and I-BiDaaS approach

We have **tons of data** but **confidential**.
We are the data managers but the **real owners of the data are our customers**.

Lack of agility in our 'Datapool' for extracting data externally.
Security procedures and constraints are necessary but hinder and slow down data sharing processes.

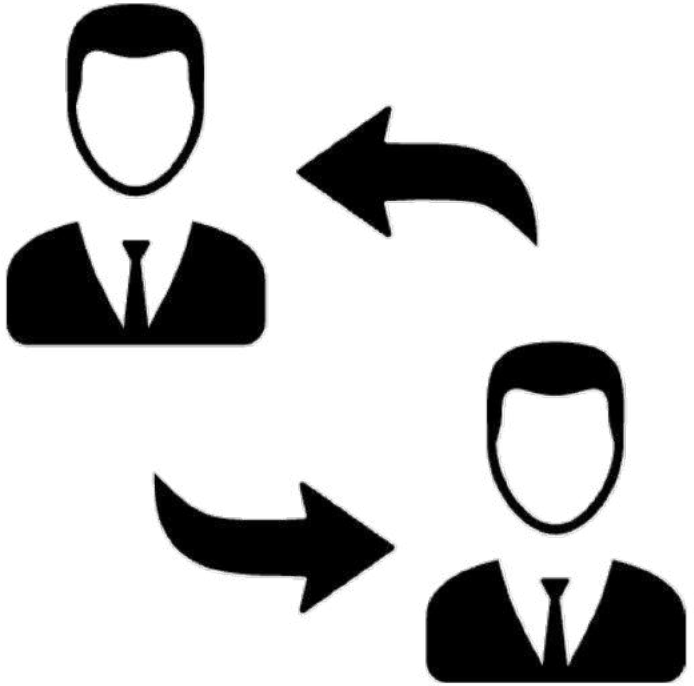
Requirements

Requirement	Control
Data privacy	Data of customers (e.g., social graph) and external suppliers (e.g., SIEM) can not be accessed or shared with other partners
Regulation Compliance	All activities related to CaixaBank within the project must comply with the relevant regulations (e.g., ISO 27001, GDPR)
Fraud and Security Analytics	Use cases presented will be related to the improvement of security and the prevention of fraud.

Objective: Exploit the I-BiDaaS platform to gain agility, efficiency and flexibility in our analytics for security.

Why do we need I-BiDaaS?

- **Breaking external and cross-sectorial data silos while complying Regulation**
  - *Sharing Data models with other FI*
  - *Sharing Data models with other sectors*
  - *Following ECB & Banco de España constraints, we already have probed with I-BiDaas that this is possible*

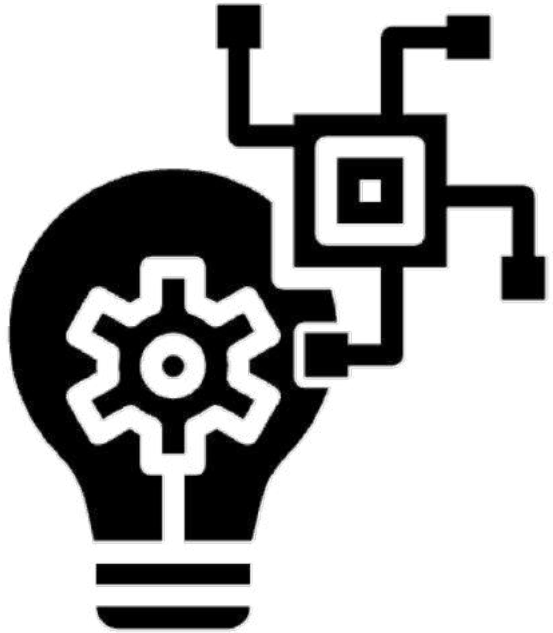


Why do we need I-BiDaaS?

- **Secure Self- Service Infrastructure**
  - *Being able to outsource Big Data Infrastructure preserving privacy & security*
  - *To grow in a dedicated and specialized infrastructure*
  - *To count on dedicated and specialized specialists*

Why do we need I-BiDaaS?

- **Competitiveness & Innovation**
  - *Fast, agile and specialized adaptation to new technology.*
  - *Vs Current proprietary infrastructure*
- **Efficiency**
  - *Reducing the costs and time of analyzing large datasets*

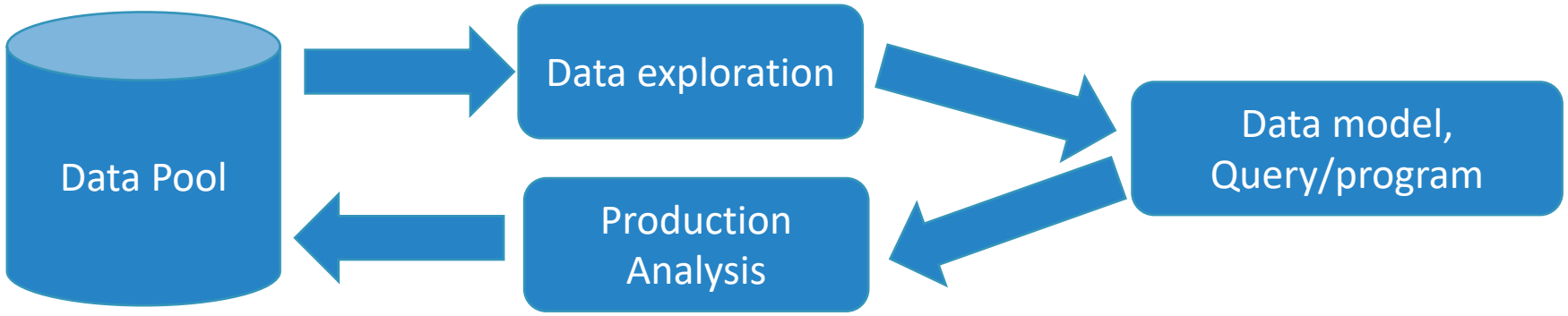


Current situation

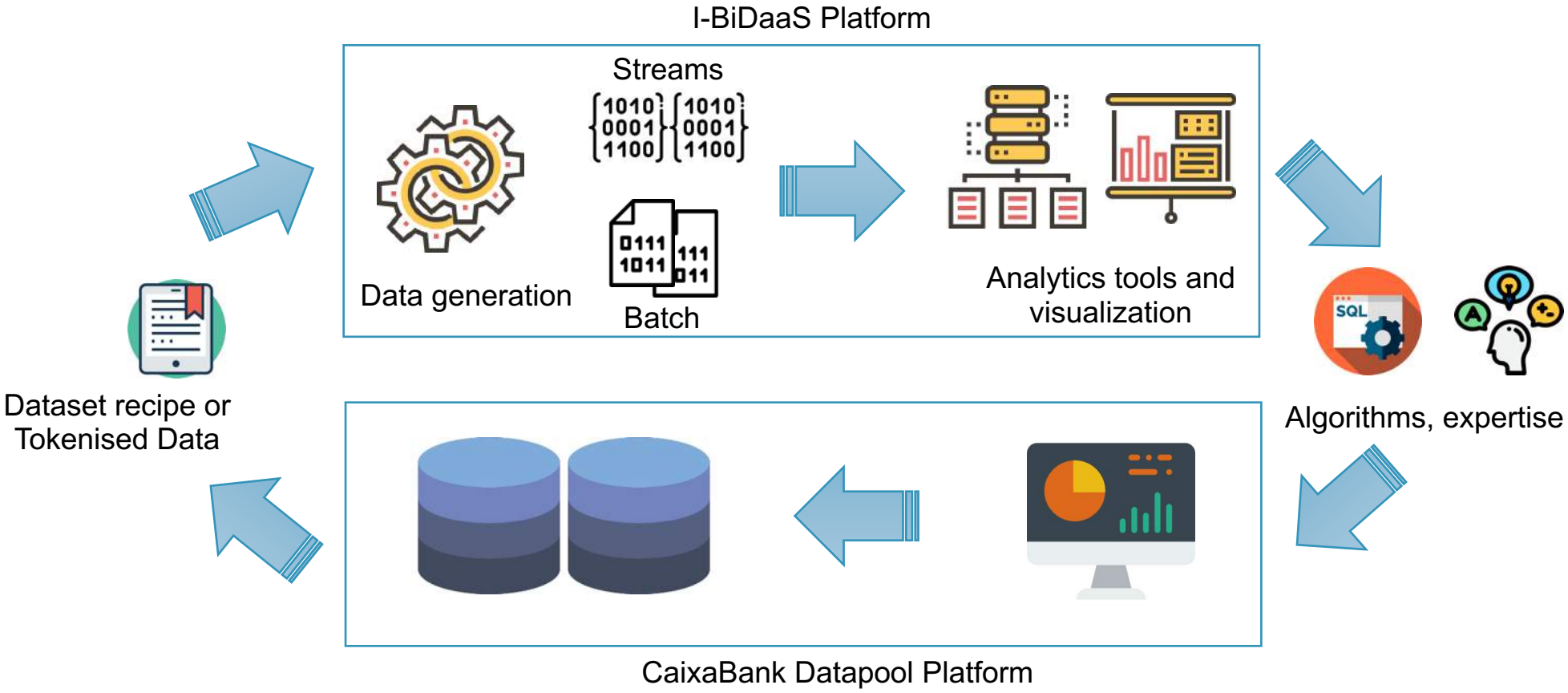
Fraud Detection Analytics in CaixaBank

- **Currently,** CaixaBank data analytics are **executed in-house**.
- **Analytics lifecycle**:
  - *Security analysis data storage: DataPool (Oracle), SIEM (Devo).*
  - *Data exploration phase to build a model/query (expensive to run).*
  - *Execution of model/query on data in production mode.*

•This process is executed **periodically**.

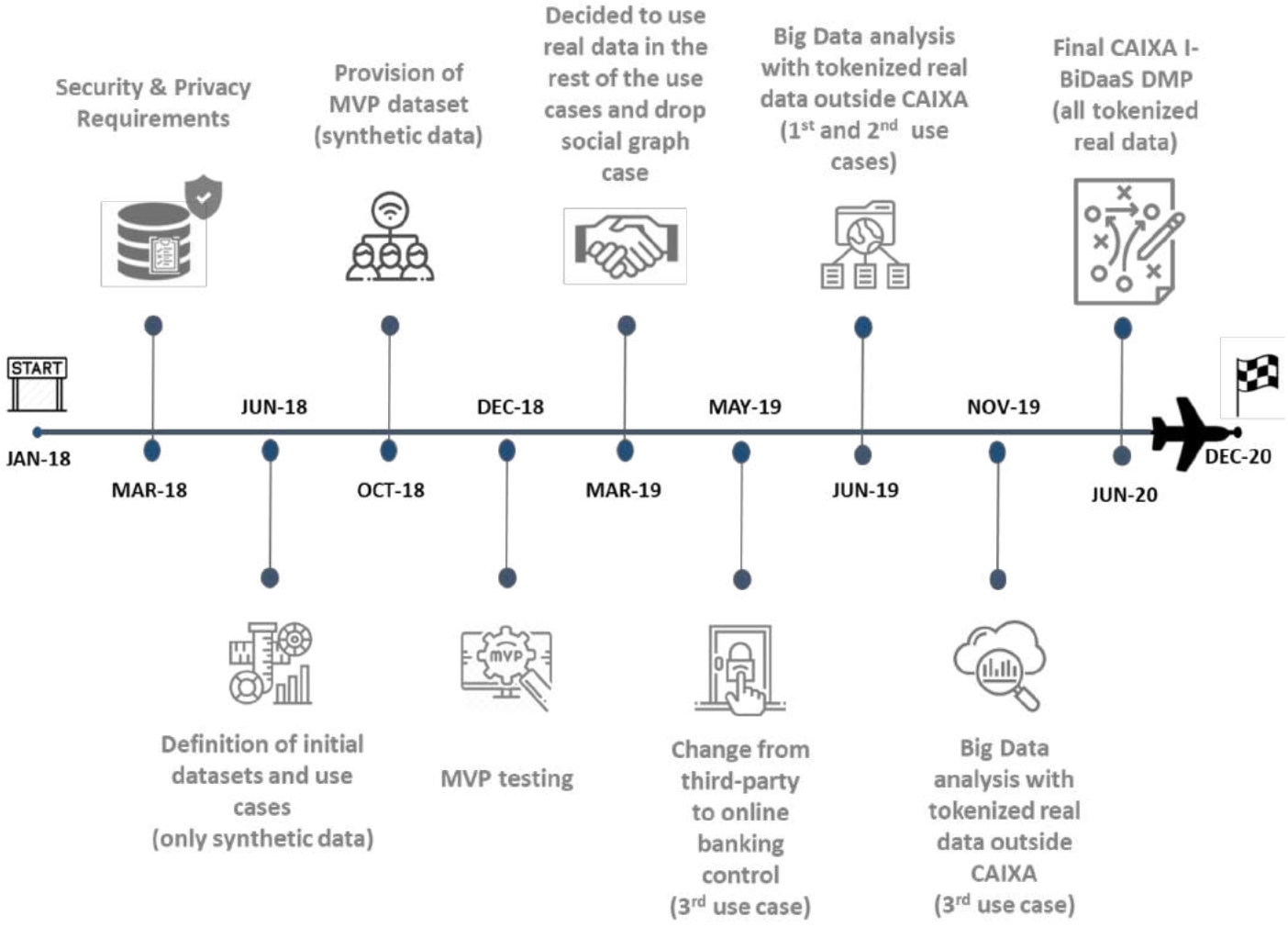


I-BiDaaS solution

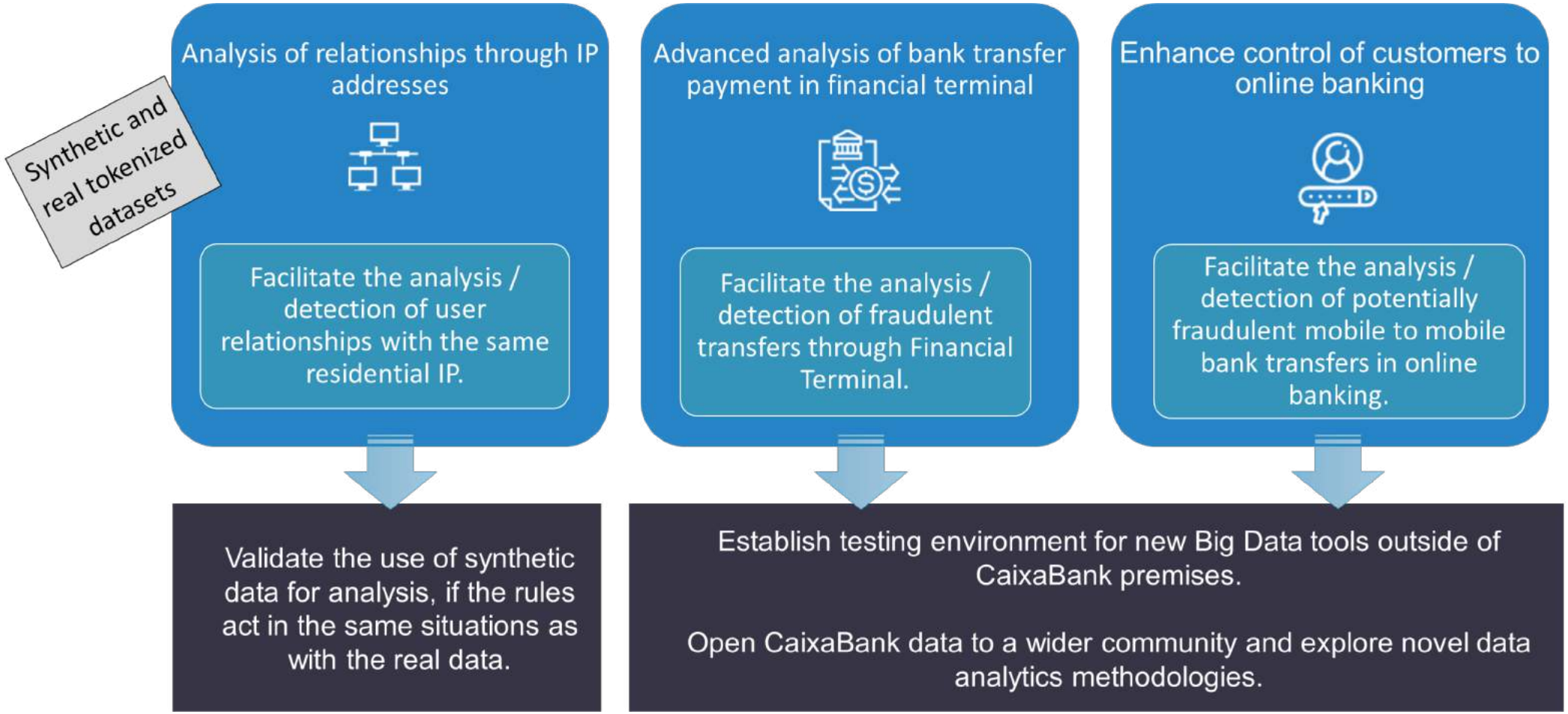


CaixaBank Data Roadmap

- **Synthetic data usage:**
  - *We explored the synthetic data solution in the MVP.*
- **New opportunities**:
  - *The possibility to work with real data outside CaixaBank in a secure way.*
  - *We moved from totally synthetic approach to tokenized real datasets.*
  - *We include a comparison between synthetic and real data to know better the differences*



Final Data & use cases



CAIXA 1st case

MVP use case

Use Case	I-BiDaaS dataset	Data
Enhance control of customers to online banking	Online banking connections	Real tokenized
Advanced analysis of bank transfer payment in financial terminal	Bank transfers	Real tokenized
Analysis of relationships through IP addresses	IP address	**Synthetic** / real tokenized

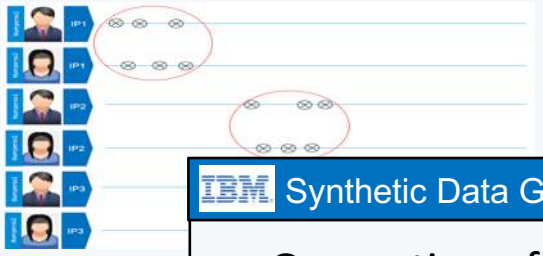
- **Business Goal:**

The transaction between two people related by an IP is not Fraud.

- **Use Case Goal:**

Validation of Synthetic Data usage.

Definition of the relationship



IBM Synthetic Data Generation

Generation of non-sensible data via rule definition

Non-sensitive data

Analytics

Results

CAIXA 1st case

MVP use case: Analysis of relationships through IP addresses

I-BiDaaS Home | Expert Mode | Self-Service Mode | Co-Develop Mode | Administrator Admin

I-BiDaaS Use Cases

IP Address Relations - Batch processsing

The CAIXA IP Addresses Use Case is an application of the I-BiDaaS ideas and tools extracting relations between IP addresses. In this use case batch processing of big data emerges as an innovative tool that can be instrumental in fraud detection.

Status: open Upd: 2018-12-12 10:33:38

IP Address Relations - Stream processsing

The CAIXA IP Addresses Use Case is an application of the I-BiDaaS ideas and tools extracting relations between IP addresses. In this use case batch processing of big data emerges as an innovative tool that can be instrumental in fraud detection.

Status: open Upd: 2018-12-12 12:37:52

Centro Ricerche FIAT Aluminium Casting (Training)

The use case is concerned with the production process of aluminium casting. The goal is to use Big Data for improving the quality of the production process and operational efficiency, in particular, the quality issues on the automotive component

Status: open Upd: 2019-02-26 09:23:03

Telefonica CC Sentiment - Stream Processing

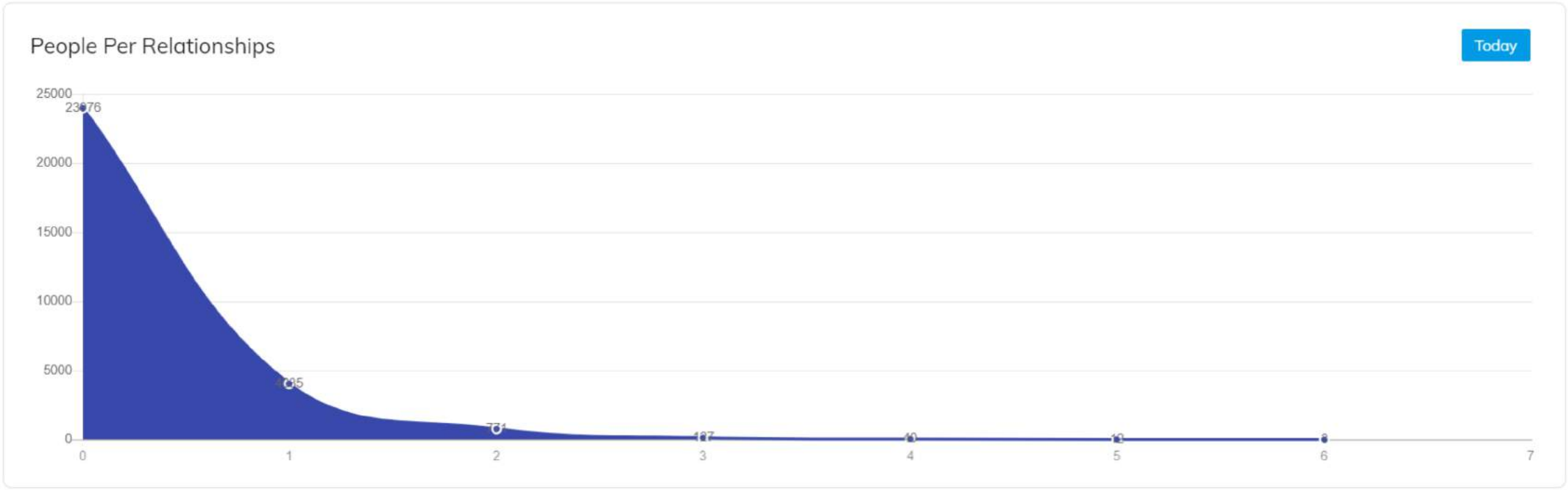
The use case is concerned with the sentiment of callers to Telefonica's call centres in Madrid, Barcellona, and Seville.

Status: open Upd: 2019-02-26 09:23:03

Which algorithm should you select?

CAIXA 1st case

MVP use case: Analysis of relationships through IP addresses



Description:
Download URL:

BACK DOWNLOAD FILE

CAIXA 1st case

MVP use case: Analysis of relationships through IP addresses

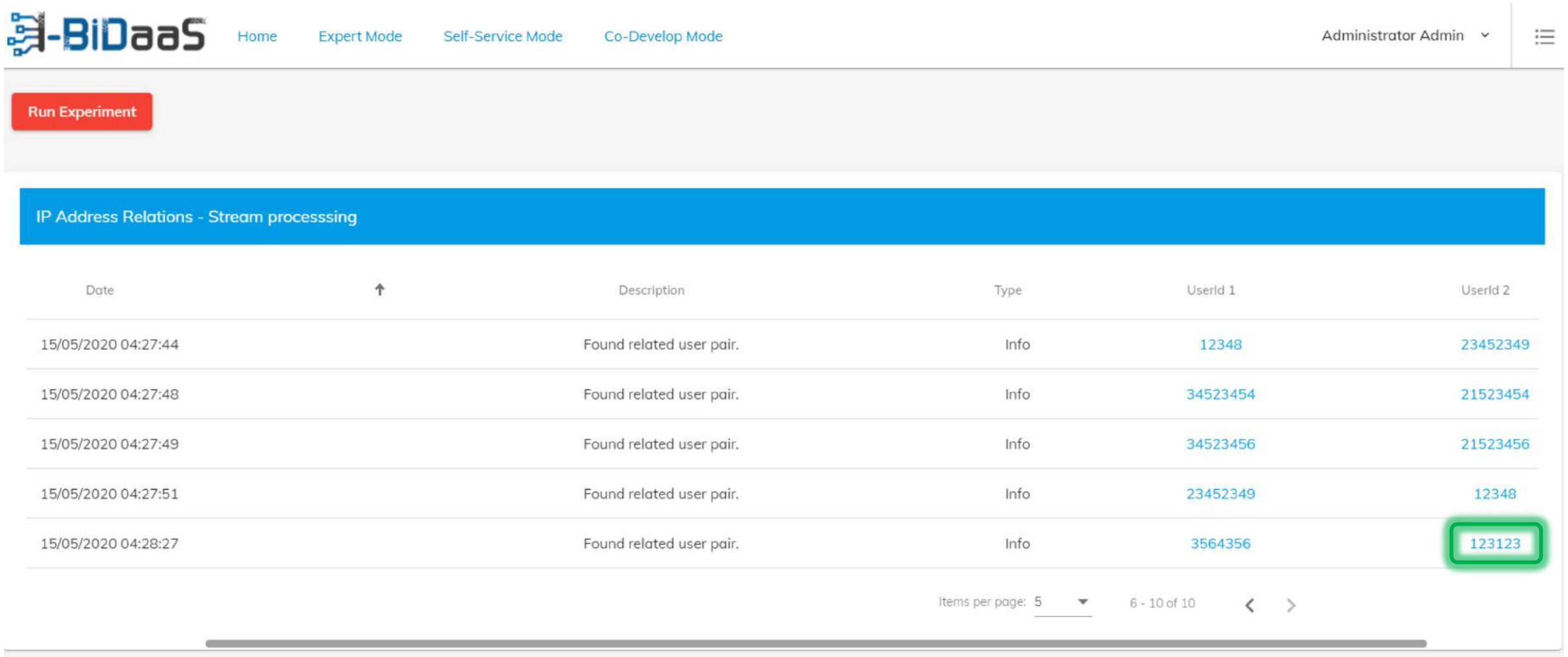
A1 | Client_ID_1

	A	B	C	D
1	Client_ID_1	Client_ID_2		
2	19440	8810		
3	19440	81427		
4	99560	4423		
5	86814	77216		
6	4897	69743		
7	93297	56347		
8	4659	69543		
9	50790	5677		
10	50790	20044		
11	50790	20548		
12	50790	28930		
13	50790	34807		
14	73358	29298		
15	65647	36884		
16	47305	29446		
17	90833	15211		
18	35403	74935		
19	76443	97593		
20	42385	55955		
21	94502	44981		
22	8959	49501		
23	26040	8404		
24	44281	48842		
25	93004	1841		

relationships

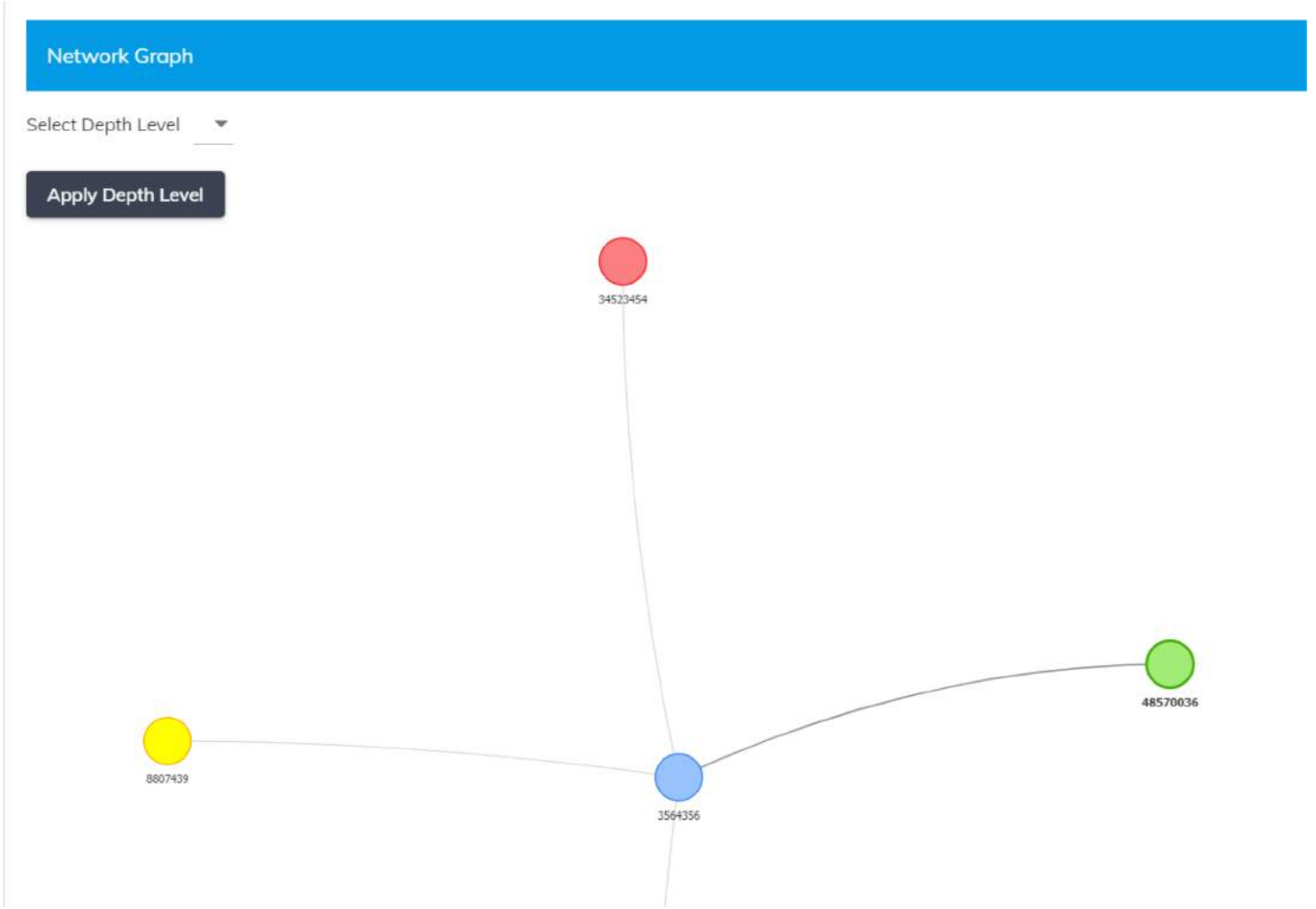
CAIXA 1st case

MVP use case: Analysis of relationships through IP addresses

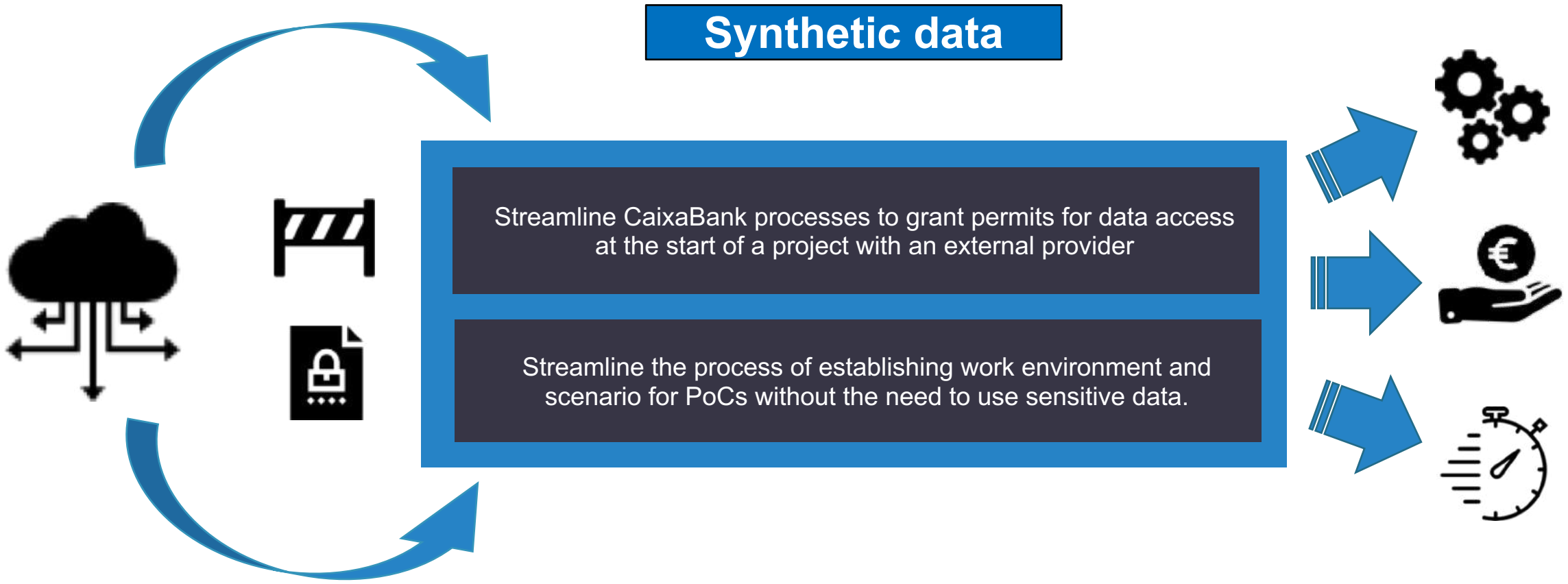


CAIXA 1st case

MVP use case: Analysis of relationships through IP addresses

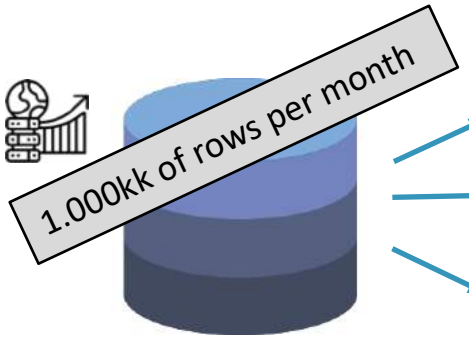


How did I-BiDaaS help us (in this use case)?



CAIXA 2nd case

Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)



TIME OF THE TRANSFER

IDENTITY NUMBER

ADDITIONAL NOTES

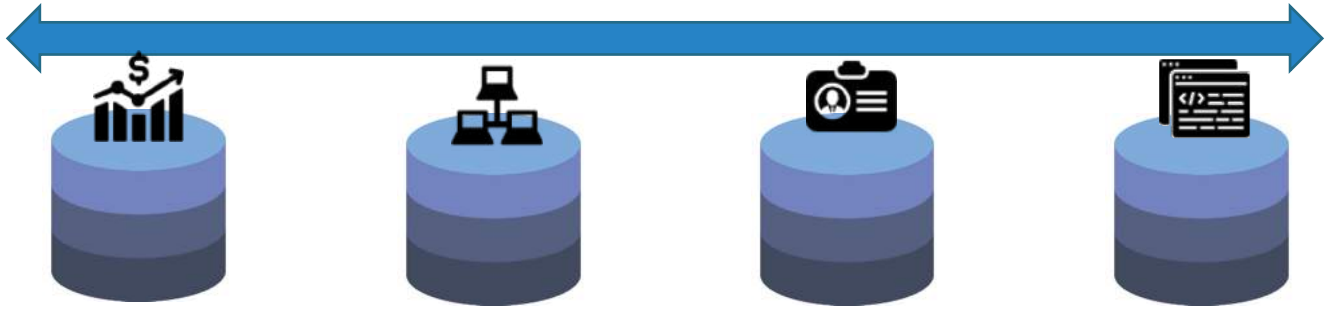
Ensure the security of our data: Decide what are we going to share **clear, tokenized or encrypted.**

The challenge relies on finding the limit of **what and how real data can be shared** to comply to regulation and not lose additional and valuable information for analytics.

Big Data Analytics objective: Discover fraudulent scenarios from our data by analysing the presence or not of the client.

Use case high-level objective: Breaking internal and external silos

- Financial Operation Data, Security Management (SIEM), etc.



CAIXA 2nd case

Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)

CONTEXT

Point of view:

- Final User

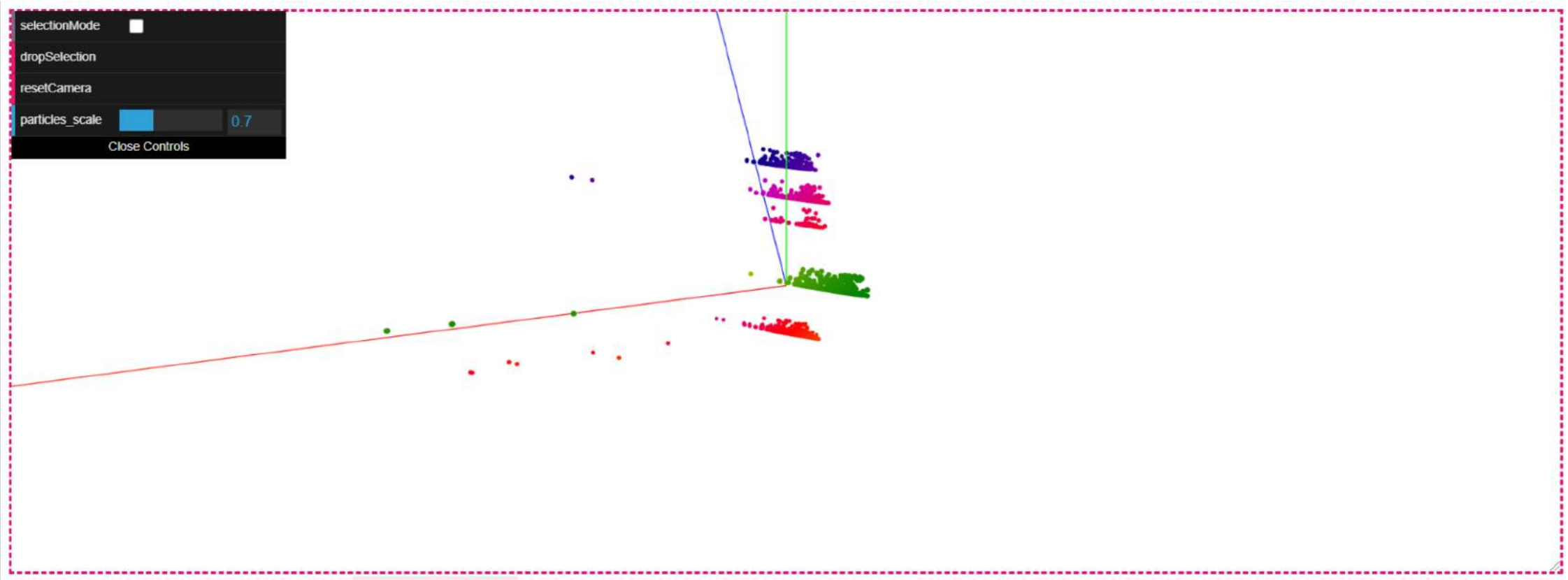
- Programmer

OBJECTIVES

- Identify and glue events, followed by enriching the transfer payment dataset
- Encrypt the data without lossing value
- Advanced analytics

CAIXA 2nd case

Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)



CAIXA 2nd case

Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)

Preview | Sample 1

	pk_anyomes	pk_anyomesdia	pk_ide_usuclo_orig	pk_idenumsesion	pk_referencia	autenticacion_esperada	cod_tipo_terminal	empresa	estado_op_dec	estado_op_dec_ltx	estado_op_trf	fech
0	202001	20200114	1502612200	UGLJ321 1135130	M74XVPEECZ	-2	LA	1	NOK	-1	-2	09/03/
1	202001	20200114	1863674100	I1LG132	M7476K21QD	-2	LA	1	NOK	-1	-2	09/03/
2	202001	20200114	8145655600	cPMd731	M74XV11UTZ	-2	LA	1	NOK	-1	ANUL	09/03/
3	202001	20200114	8147861200	yVMd331	M746XQPTF8	-2	LA	1	OK	ACCP	EMIT	09/03/
4	202001	20200114	8004254000	lkL4921	M74XDDHQCT	-2	LA	1	OK	ACCP	EMIT	09/03/
5	202001	20200114	8069553800	yuLu922	M74XT7W9YV	-2	LA	1	OK	ACCP	EMIT	09/03/
6	202001	20200114	978495900	rKKl822 1718359	M748GUDCRM	-2	LA	1	OK	ACCP	EMIT	09/03/
7	202001	20200114	1599047900	QtLP021 2216429	M74BZ8YWH3	-2	LA	1	OK	ACCP	EMIT	09/03/
8	202001	20200114	8021525800	yQL5731	M7494ZU8PF	-2	LA	1	OK	ACCP	EMIT	09/03/
9	202001	20200114	1904984200	ViMa311	M74EGFRXUR	-2	LA	1	OK	ACCP	EMIT	09/03/
10	202001	20200114	107025400	pbMa241	M74B3CAHAF	-2	LA	1	OK	ACCP	EMIT	09/03/
11	202001	20200114	8020272400	lDL5341 1742436	M748MGJ8VF	-2	LA	1	OK	ACCP	EMIT	09/03/
12	202001	20200114	8118703100	AxLD822	M74A9TGPBD	-2	LA	1	OK	ACCP	EMIT	09/03/
13	202001	20200114	1810458300	0vLs112 0913027	M744GP4TPZ	-2	LA	1	OK	ACCP	EMIT	09/03/
14	202001	20200114	1957767400	5lMp611	M7413J1Z3H	-2	LA	1	OK	ACCP	EMIT	09/03/
15	202001	20200114	8138361700	DVMc831	M74B7ZZ681	-2	LA	1	OK	ACCP	EMIT	09/03/
16	202001	20200114	8024502200	otL6111	M7486FQ6WH	-2	LA	1	OK	ACCP	EMIT	09/03/
17	202001	20200114	1861141700	YcLG222 0922092	M744YQJW4J	-2	LA	1	OK	ACCP	EMIT	09/03/
18	202001	20200114	1898110600	ssLN512 0743142	M743UPHNU3	-2	LA	1	OK	ACCP	EMIT	09/03/
19	202001	20200114	8010656200	pnL4141	M744JTJNJU	-2	LA	1	OK	ACCP	EMIT	09/03/

File format: csv | Generate file

CAIXA 2nd case

Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)

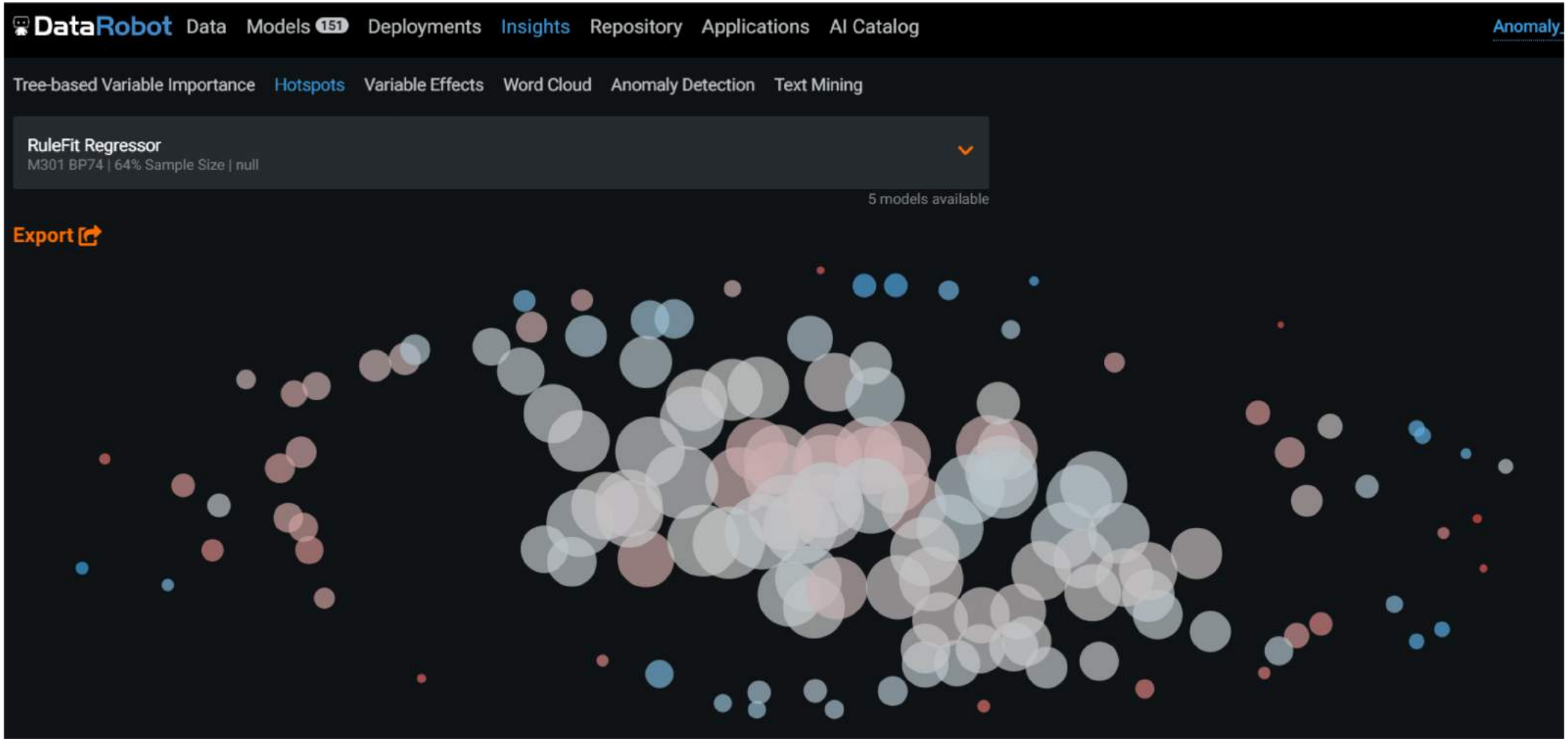
Comparing solutions and processes to analyse data outside CaixaBank premises

Data Analytics commercial products

DataRobot

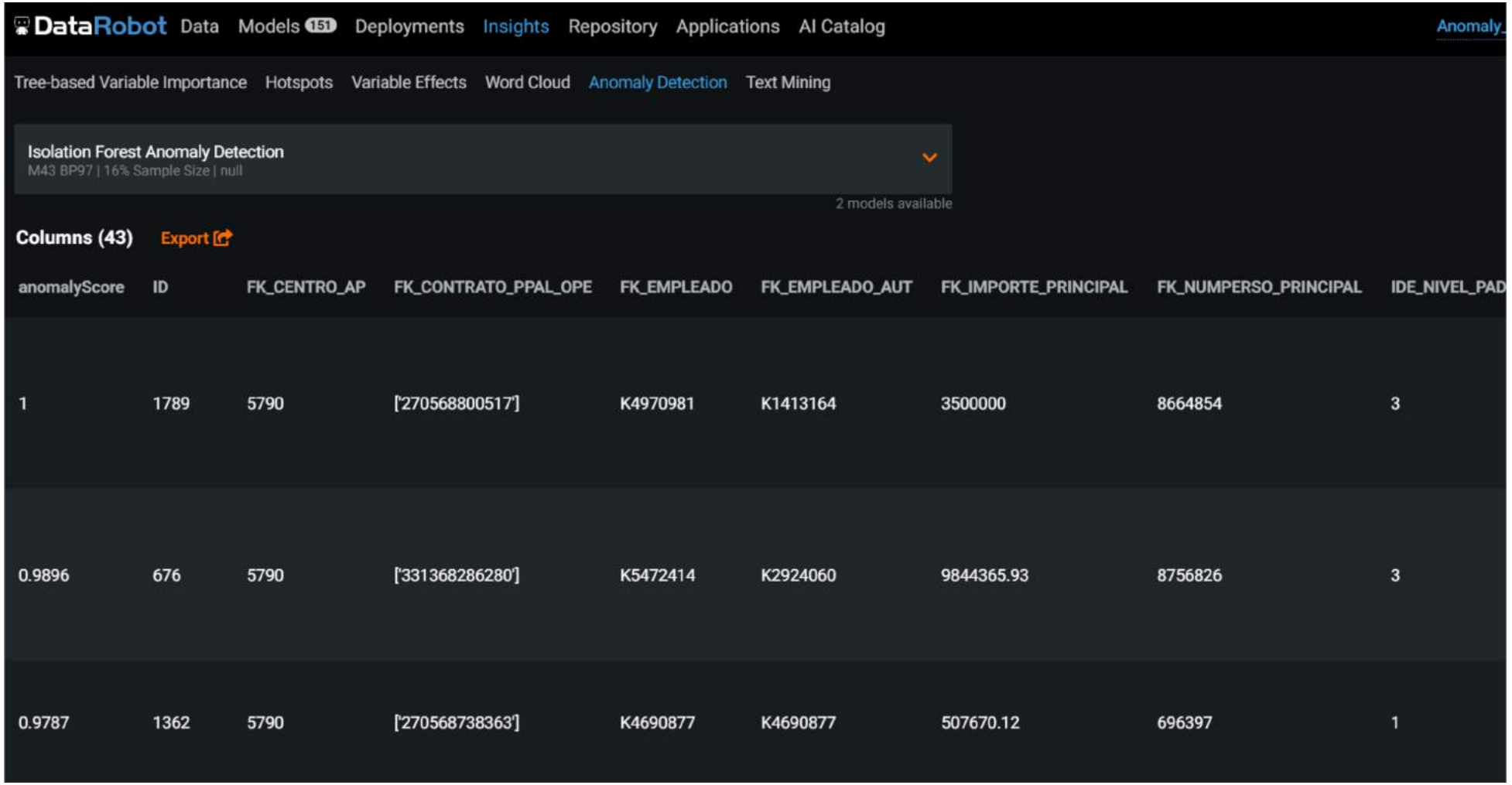
CAIXA 2nd case

Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)



CAIXA 2nd case

Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)



anomalyScore	ID	FK_CENTRO_AP	FK_CONTRATO_PPAL_OPE	FK_EMPLEADO	FK_EMPLEADO_AUT	FK_IMPORTE_PRINCIPAL	FK_NUMPERSO_PRINCIPAL	IDE_NIVEL_PAD
1	1789	5790	['270568800517']	K4970981	K1413164	3500000	8664854	3
0.9896	676	5790	['331368286280']	K5472414	K2924060	9844365.93	8756826	3
0.9787	1362	5790	['270568738363']	K4690877	K4690877	507670.12	696397	1

CAIXA 2nd case

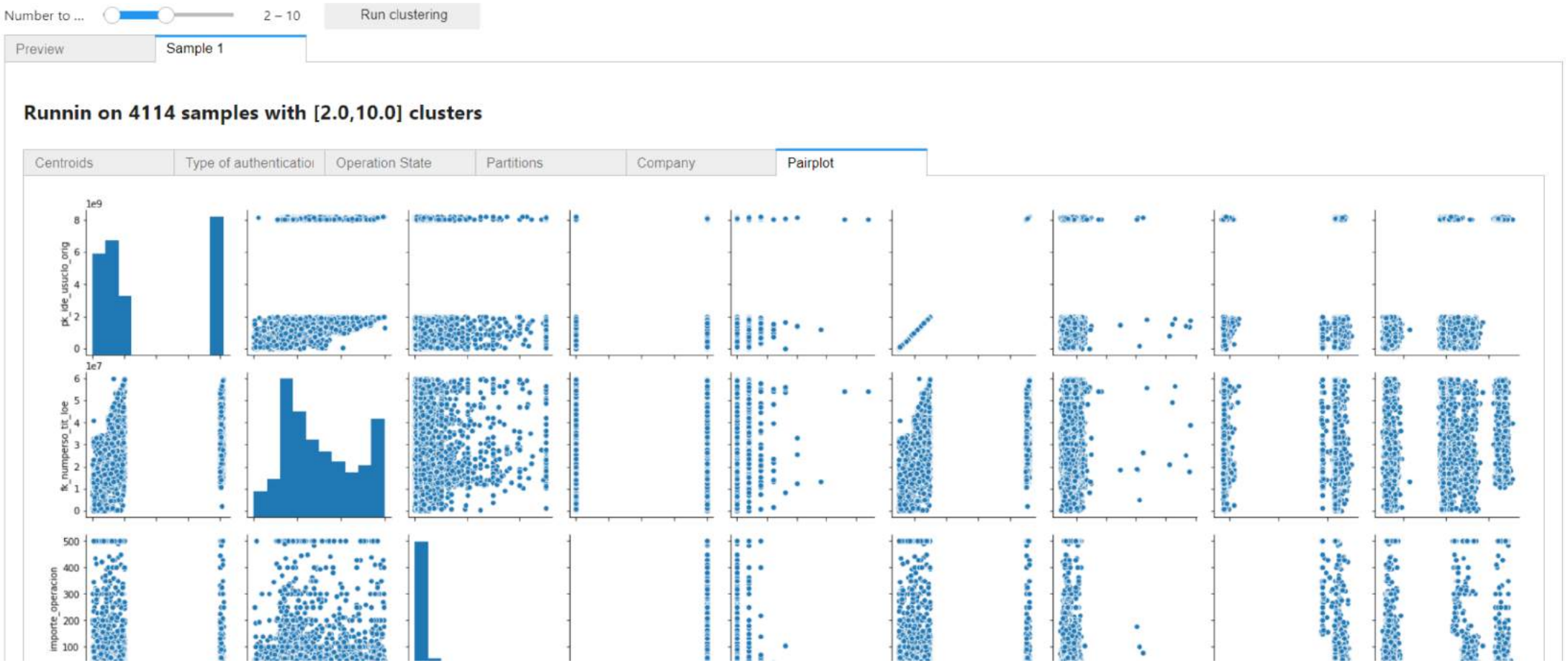
Analyse bank transfers executed from employees financial terminal in the name of a customer.
Potential fraudulent transfer or bad practices (e.g. check that the client was present in the time of the movement.)

DataRobot comparison results

- ***Custom solutions:***
  - I-BiDaaS provide more flexibility in the definition of your own code, scoring metrics, etc.
- ***Unsupervised learning:***
  - DataRobot has very limited number of unsupervised learning models. I-BiDaaS can provide much more detailed results on unsupervised learning use cases.

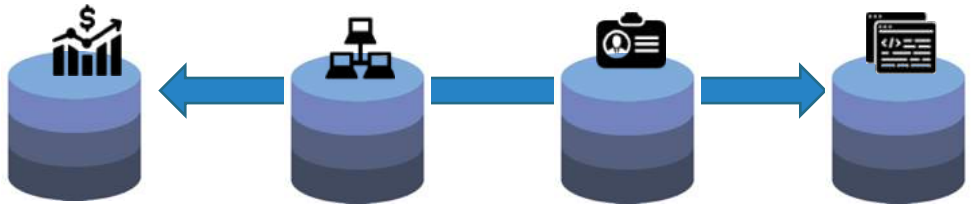
3nd use case

Analyse the patterns on mobile to mobile bank transfers and identify clusters of transfers with higher risk.



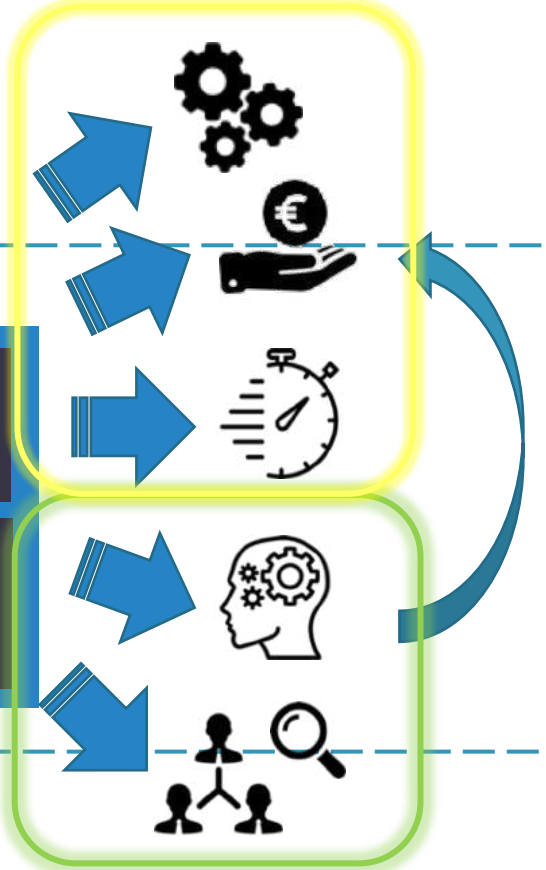
Lessons Learned: How did I-BiDaaS help us?

Main objective → Break information silos



Establish a methodology to extract and analyse sensitive data outside of our premises

Test solutions in a more efficient way, before integrating them into the CaixaBank environment



Synthetic data

- Streamline access to data to new providers or other entities for first data analysis.
- Improve performance evaluation and integration with new tools.

Tokenized data

- Allow the safe extraction of data, and data analytics outside the CaixaBank environment.

Lessons Learned: CaixaBank benefits from I-BiDaaS

Benefits	KPIs
To increase the efficiency and competitiveness in the management of its vast and complex amounts of data.	75% time reduction data access from external stakeholders using synthetic data (From 6 to 1.5 days).
To break data silos not only internally, but also fostering and triggering internal procedures to open data to external stakeholders.	Real data accessed by at least 6 different external entities skipping long-time data access procedures.
To evaluate Big Data analytics tools with real-life use cases of CaixaBank in a much more agile way.	I-BiDaaS overall solution and tools experimentation with 3 different industrial use cases with real data.

Telefonica Research
I-BiDaaS Final Event

Ioannis Arapakis
(PhD)

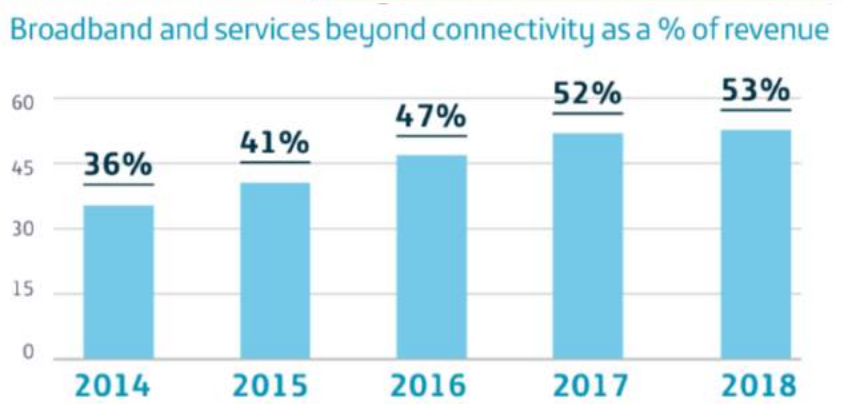
Jordi Luque Serrano
(PhD)

Telefonica

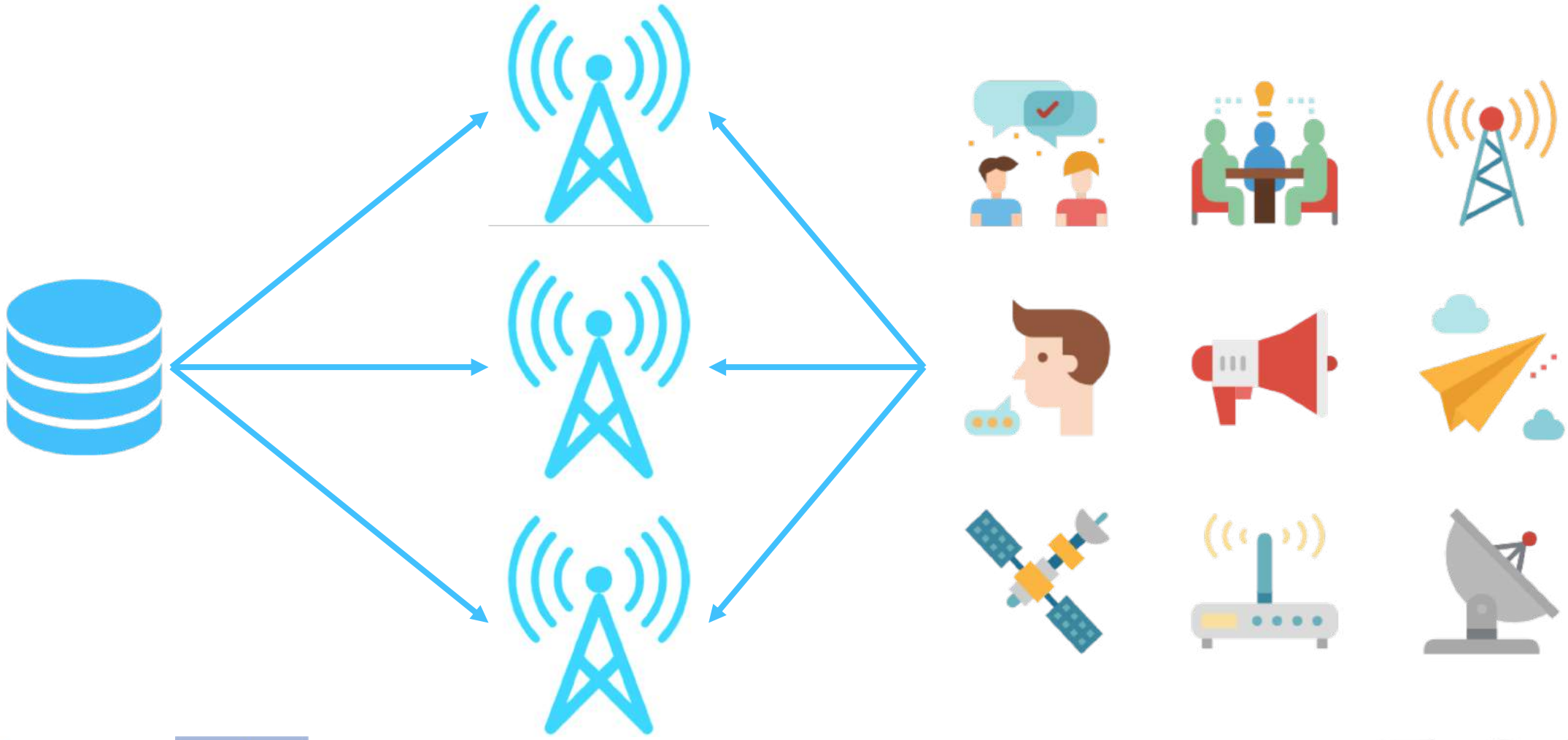
Overview of TID

Telefonica in Numbers

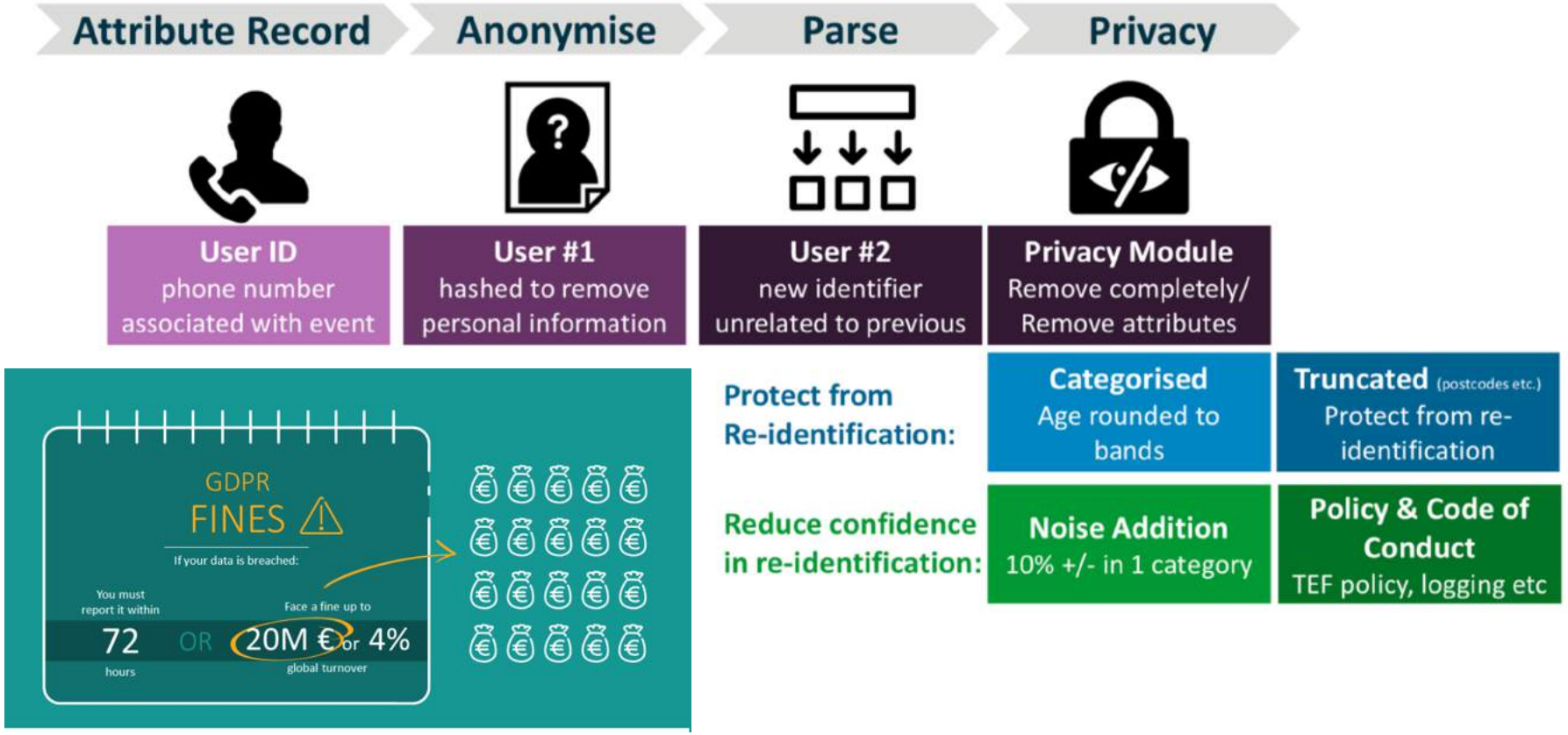
- Telefonica is one of the largest telecommunications companies in the world by market capitalization and number of customers
- It provides quality of connectivity that is delivered over world class fixed, mobile and broadband networks



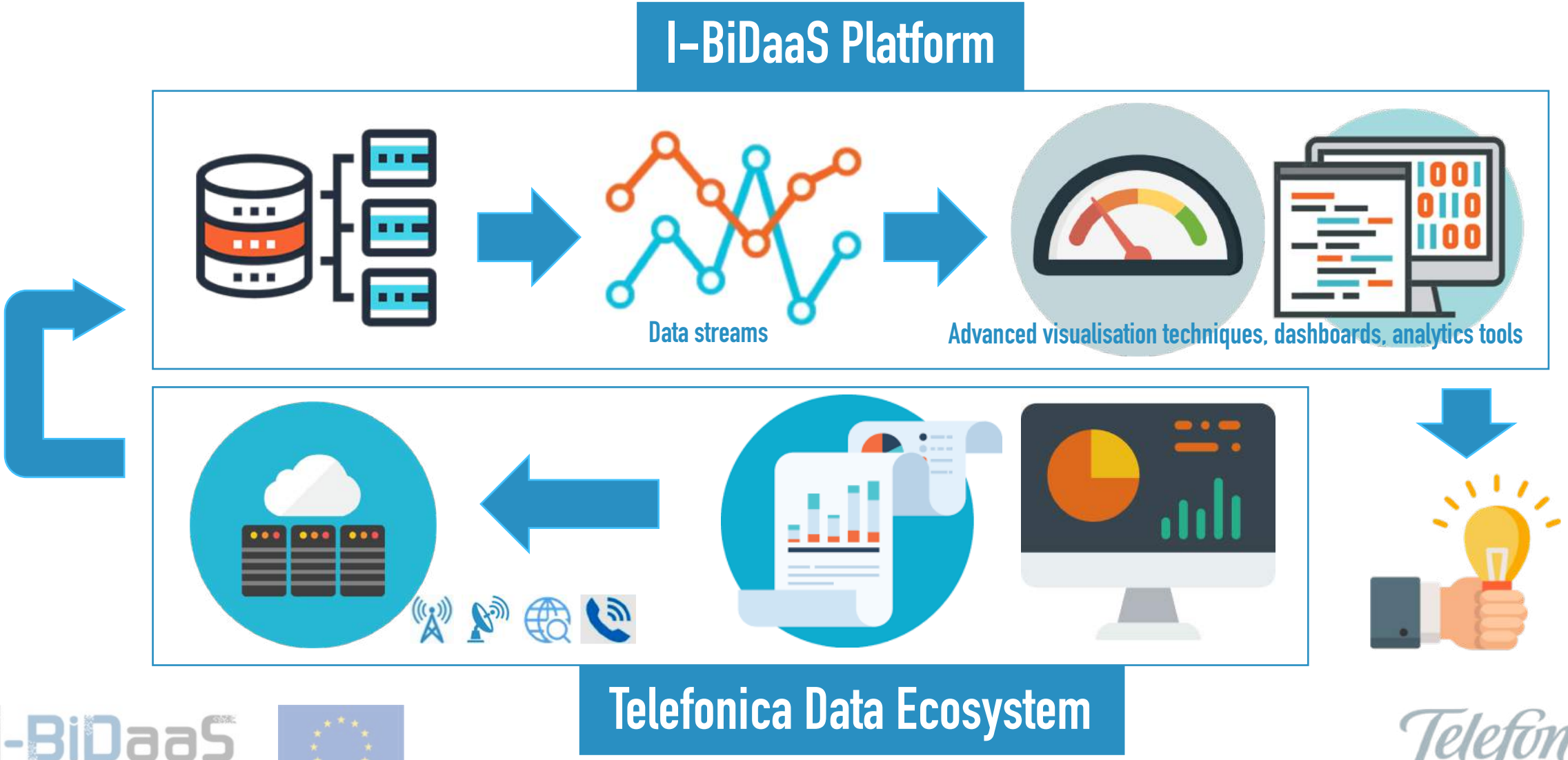
Data Landscape



Industrial Challenges



The I-BiDaaS Solution



I-BiDaaS
Industrial-Driven Big Data as a Self-Service Solution

Telefónica

TID Use Cases

Quality of Service in Call Centers

Improve performance of CC operations by automatically processing audio calls and predicting customer satisfaction index

Accurate location prediction with high traffic and visibility

Enable the automatic extraction of behavioural patterns of customers

Optimization of placement of telecommunication equipment

Improve routing and placement of antennas

Quality of Service in Call Centres (CC)

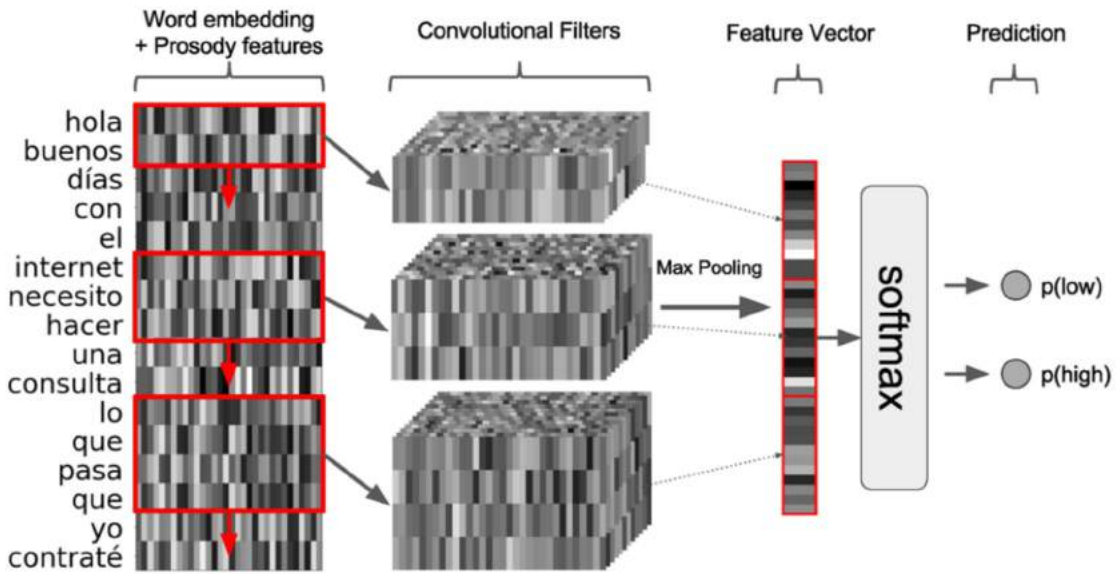
Use Case Requirements

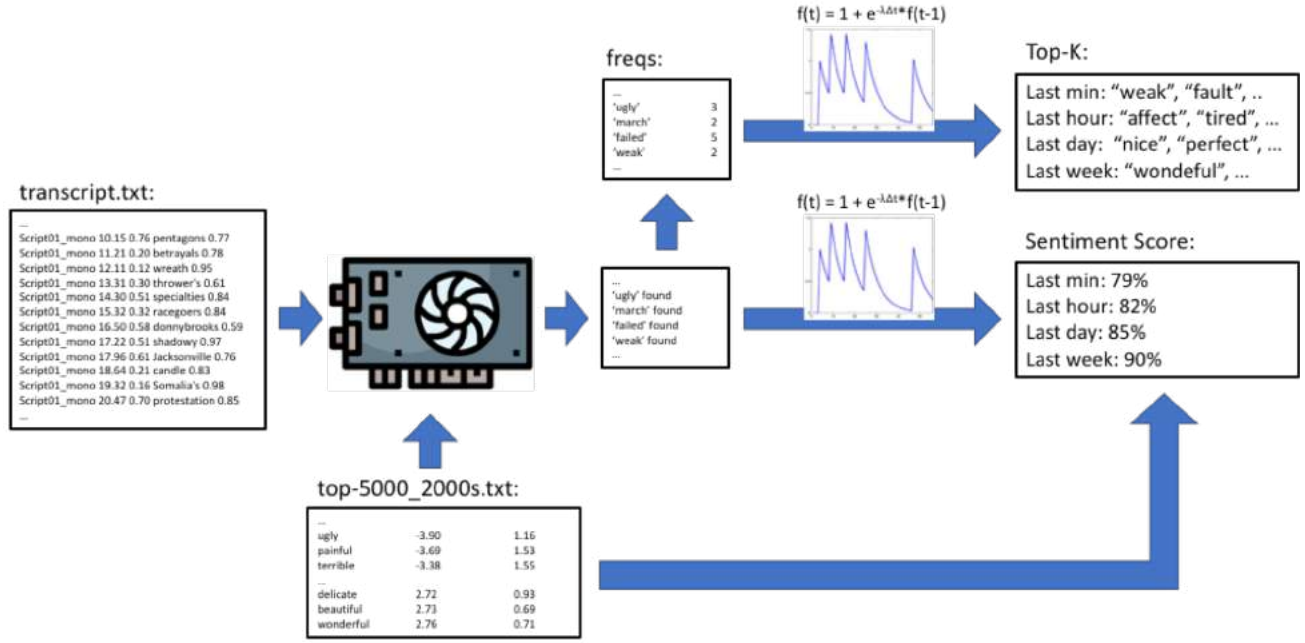
- Nature of phone calls varies greatly
- Quickly get familiar and understand customer's perspective to provide fast response and improve customer service by using big data speech and language analytics
- Shorten call durations, waiting time and First Call Resolution (FCR) time by anticipating customer's situation based on previous insights, e.g., using the aggregation of previous analytics by Call Centers or regions
- Detect customer satisfaction based on the automatic transcript analysis

Data

- Anonymised dataset with 1.3M transcripts, that is, around 12.3 years of continuous speech recordings
- Analyzed the significance of various acoustic features extracted from customer-agents' spoken interaction in predicting self-reported satisfaction
- Investigated whether speech prosodic features can deliver complementary information to speech transcriptions provided by an ASR

Predictive Modeling





	Dataset AUC, (F-score)	
	Spain	Latam
BoW-PCA	0.716 (0.324)	0.689 (0.262)
XGBoost (prosody)	0.58 (0.185)	0.610 (0.2398)
BoW-PCA + prosody (MW)	0.7309 (0.3420)	0.701 (0.269)
CNN	0.605 (0.212)	0.759 (0.410)
CNN + prosody	**0.733 (0.242)**	**0.772 (0.427)**

KPIs

- No automated solution in place; customer audio calls are processed by human agents (Business Units inspect <1% of total calls per year) which is a costly and time-consuming solution

Existing Solution	I-BiDaaS Solution
~11,520 calls processed (per year) by human agents	**~3.5B** calls processed per year in a single GPU Max. real-time throughput: **40K** transcripts/second (I-BiDaaS platform)
This manual process allows to identify about **2,300 low customer satisfaction audio calls**	Increase the number of detected low customer satisfaction calls by human agents to **7,000 (200% increase*)** by ranking the calls based on sentiment

Intelligent Dashboard

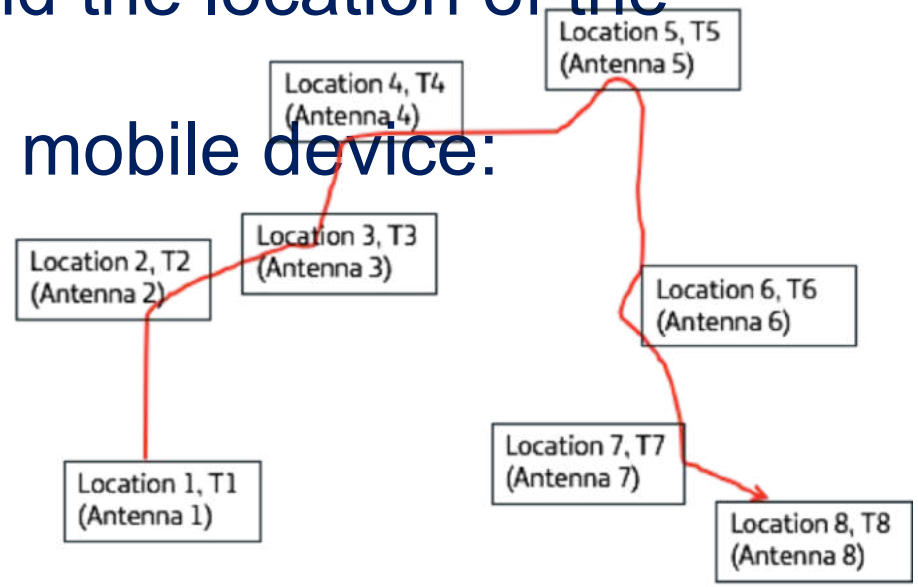
Accurate Location Prediction with High Traffic and Visibility

Use Case Requirements

- Predict places with high traffic and congestion events in order to optimise their resource distribution
- Important challenges that stem from this use case:
  - Interpolate missing events to recover plausible event trajectories
  - Minimize processing time with respect to growing data size
  - Maintain real-time delivery of results

Data

- Our dataset consists of synthesized traces mimicking cellular network events
- Each trace is a time series of mobile events that contain the encrypted user identifier, a timestamp, and the location of the associated base station
- A mobile event is generated every time a mobile device:
  - activates/deactivates in the network
  - makes/receives a call
  - sends/receives an SMS
  - moves from one location area code to another
  - changes from one technology to another
  - requests access to data (2G/3G) or requests a high-speed data channel (4G)

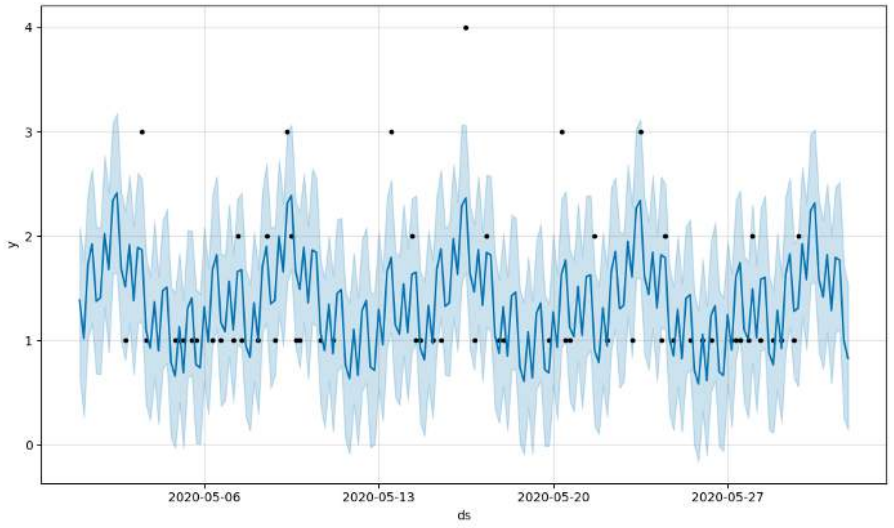


Predictive Modelling

- We want to predict movement from one sector to another
- Estimate the delta (%) in connected users per sector x hours in advance
- The forecasting model can account for:
  - Information about the other sectors' current status and load
  - Contextual data, such as the weather for the location
  - Occurring events (holiday events or other)

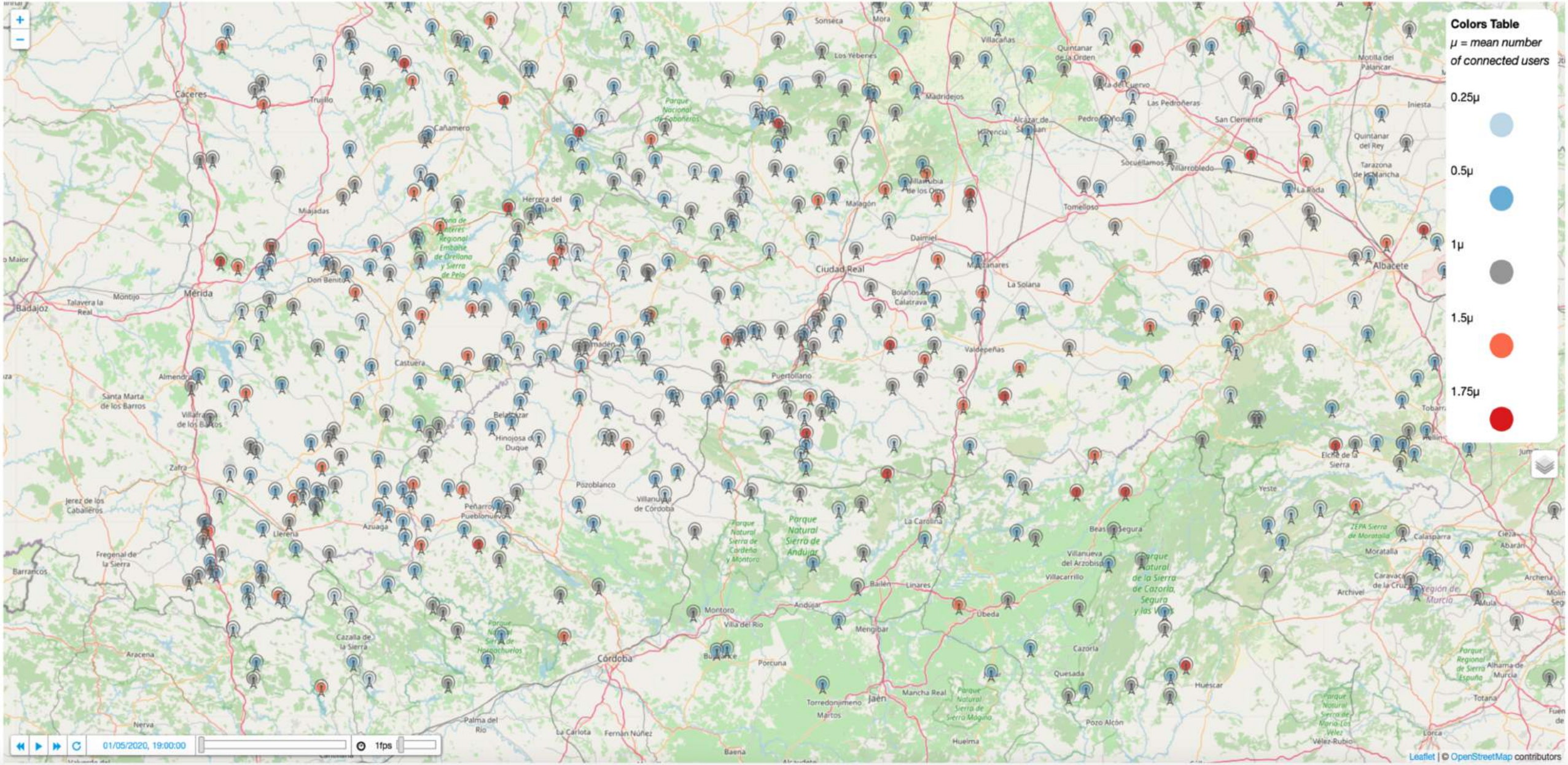
KPIs

- We used process-based parallelization where every CPU core was assigned a process which trained a time series model for one antenna
- Training ~120K time series took three (3) hours on the TID server
- We selected the best 100 models based on the mean average error (we obtained a baseline metric of the average mean absolute error of 1.2



Mobility Data Locations (normalised by the mean) - May 2020

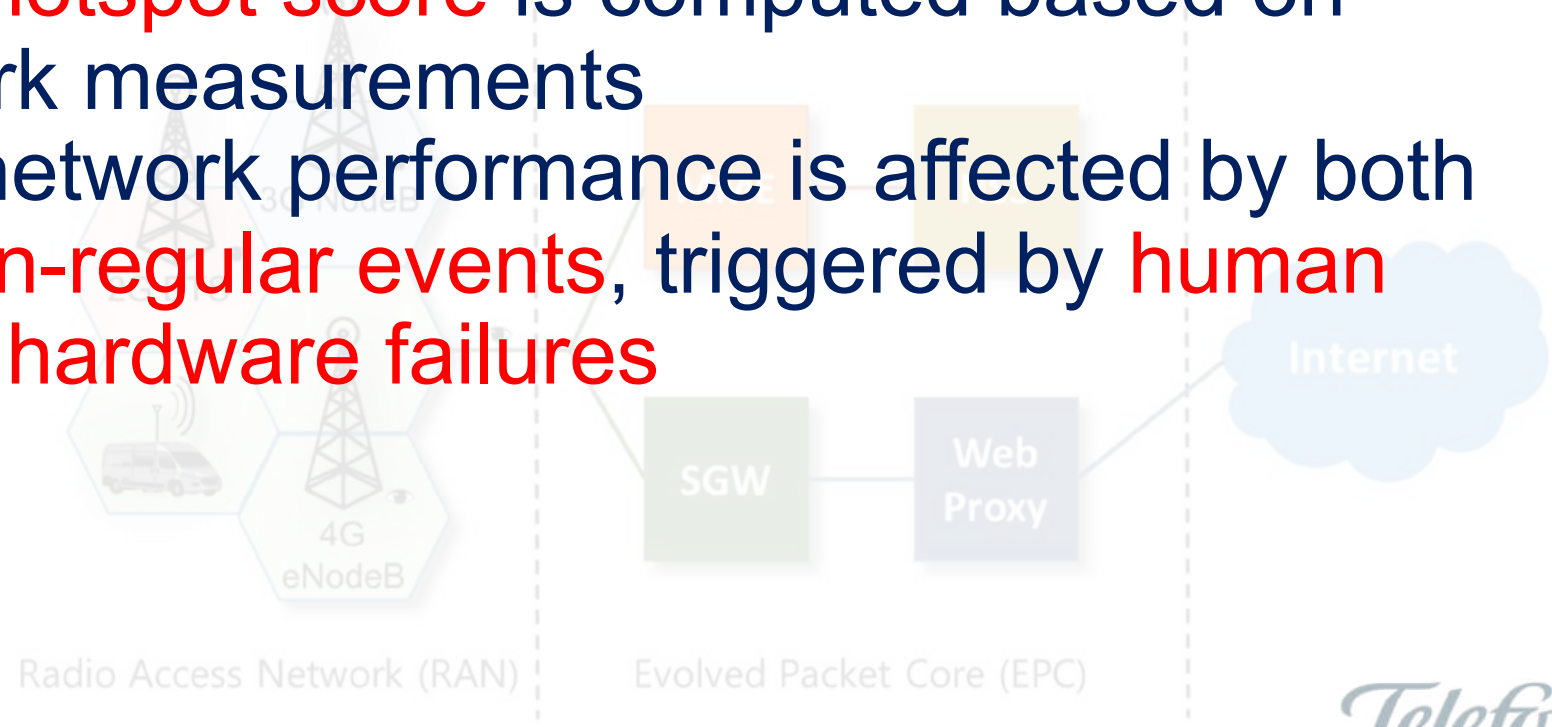
Highlight Antennas Below μ | Highlight Antennas Above μ | Highlight All Antennas | *Adjust Opacity :*



Optimization of Placement of Telecommunication Equipment

Use Case Requirements

- To manage and maintain large-scale cellular networks, MNOs need to monitor constantly cellular network performance, which is a non-trivial task
- The so-called hotspot score is computed based on multiple network measurements
- However, the network performance is affected by both regular and non-regular events, triggered by human behaviour and hardware failures



Current Limitations

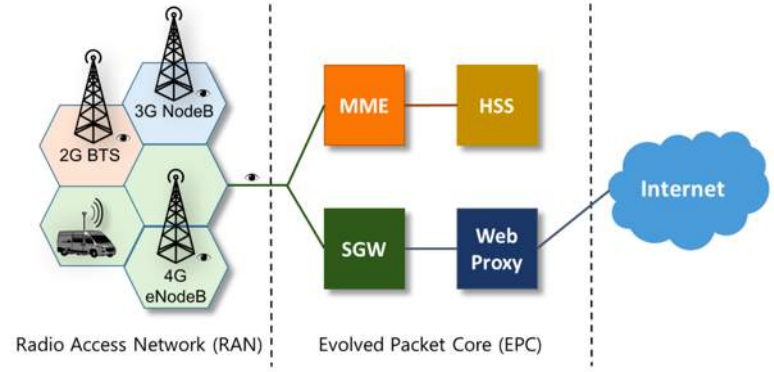
- While operators have a good understanding of the current performance of a network, forecasting the performance of each sector over time is a non-trivial task
- The network performance is affected by both regular and non-regular events, triggered by human behaviour and hardware failures

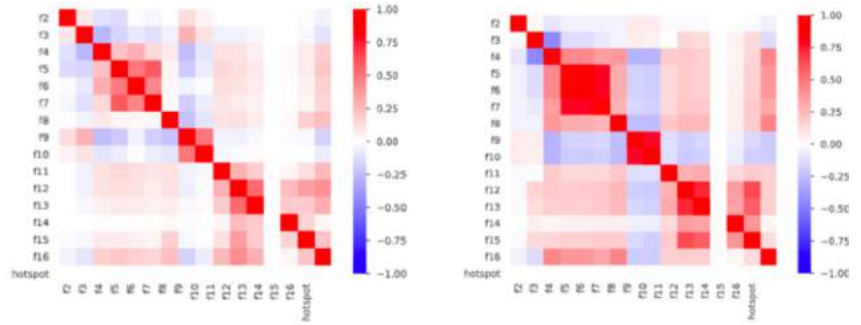
Current Limitations

- A manual assessment cannot scale and it is error-prone
- Heuristic or rule-based models require domain expertise and periodic updates to account for changes temporal and contextual factors (e.g. COVID19 pandemic)
- Conventional forecasting techniques cannot model accurately the dynamics and latent interactions among the process variables

Data

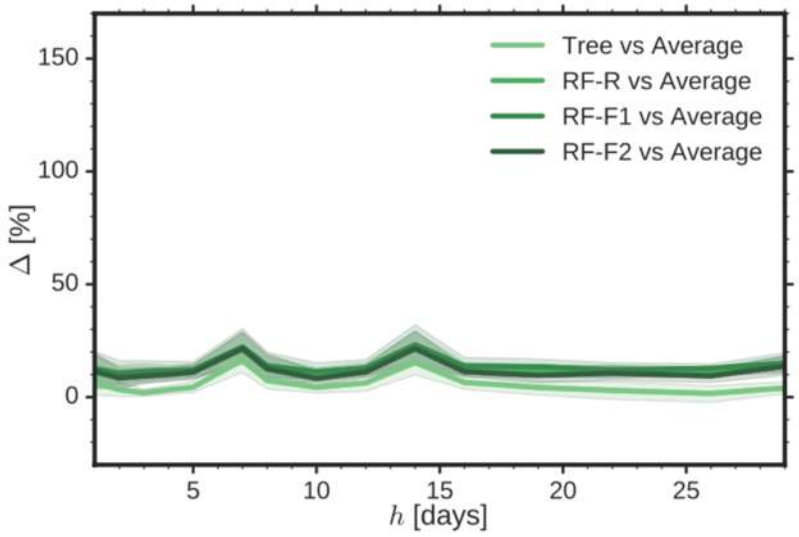
- Synthetic dataset of 999,257 observations with 16 features and ground truth:
  - More than 40K cell sites
  - 24 hours cycle
  - Dataset was unbalanced – 0.0119% of the positive class, mimicking the read data attributes

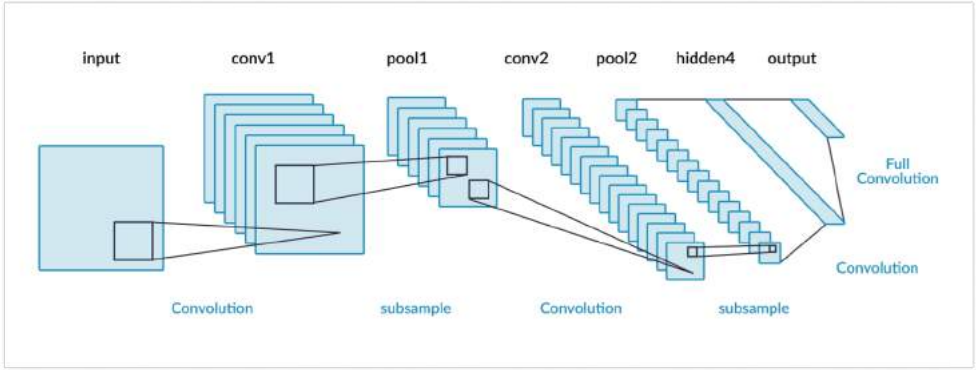


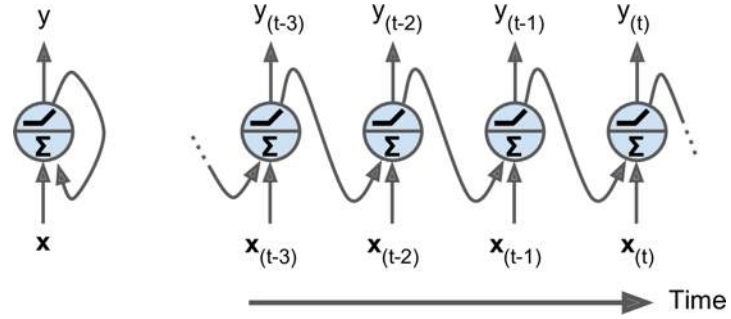


Predictive Modeling

- We address this problem as a binary classification task; hence, our target variable is a binary label that corresponds to the notion of being a “hot spot” at a certain day





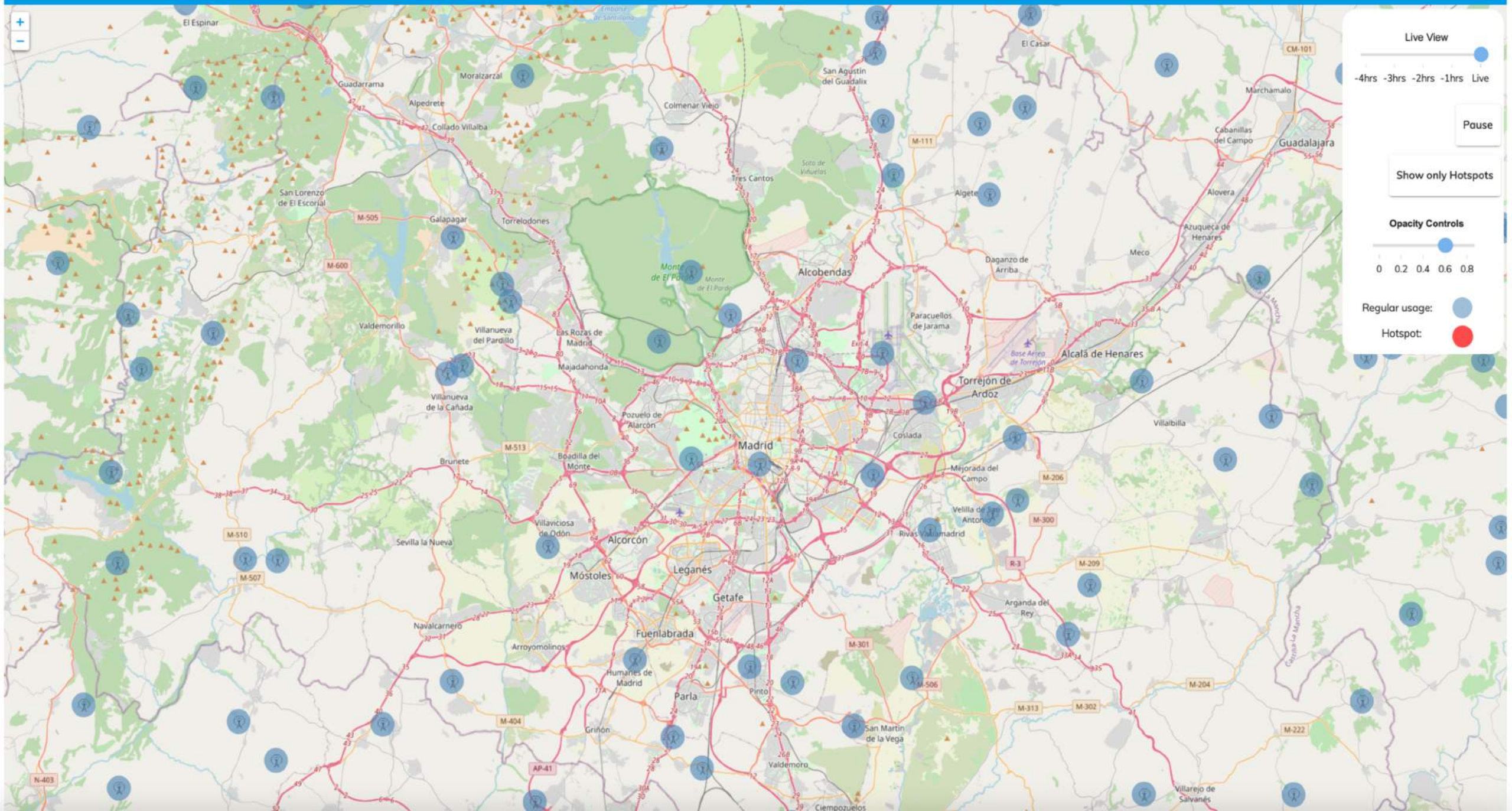


KPIs

- Tested Random Forest (RF), XGBoost (XGB) and CATBoost (CB)
- Used a 80/20 train-test split
- Predict if an antenna will become hotspot in the next hr

Model	Accuracy	Precision	Recall
XGB	0.999	0.998	0.998
CB	0.999	0.977	0.961
RF	0.999	0.998	0.975

Antenna KPIs - Hotspot Prediction (Live Streaming)



Telefonica Research Hackathon

Telefónica

I-BiDaaS

Telefónica Research HACKATHON

THE FUTURE OF SPEECH-TO-TEXT TECHNOLOGIES

REGISTER HERE:

Hosted online

23rd October 2020

wayra

An initiative of: Telefónica

We are looking for you!

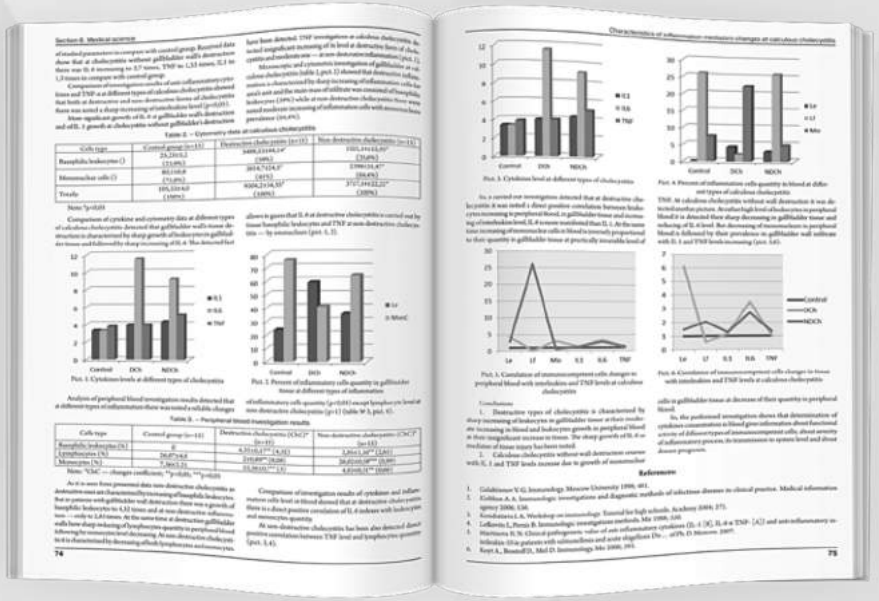
Participate in the Telefónica Research Hackathon and win a fast track to the next Wayra Pitch Day

#WeAreWayra

Telefonica Research Hackathon

- Online event on the “Quality of Service in Call Centers” between the dates 23rd to 25th of October 2020
- Students, scientists and employees from startups and SMEs from all over Europe, teamed up to address the challenge
- 8 teams, 8 mentors, 8 cutting-edge AWS servers
- Broke the inter- and intra-sectorial data-silos
- Involved different business units and external companies
- Raised awareness about the research output produced by the I-BiDaaS project and Telefonica Research

Research Output



I-BiDaaS Publications

- Xin, X., Karatzoglou, A., Arapakis, I., and Jose, J. (2020). Graph Highway Networks. ArXiv, abs/2004.04635.
- Gálvez, A., Luque, J., and Gravano, Agustin. (2020). A unifying framework for modeling acoustic/prosodic entrainment: definition and evaluation on two large corpora. In Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue (pp. 215–224). Association for Computational Linguistics.
- Yuan, F., Karatzoglou, A., Arapakis, I., Jose, J.M. and He, X. (2019). A Simple Convolutional Generative Network for Next Item Recommendation. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19). Association for Computing Machinery, New York, NY, USA, 582–590.
- I. Arapakis et al., "Towards Specification of a Software Architecture for Cross-Sectoral Big Data Applications," 2019 IEEE World Congress on Services (SERVICES), Milan, Italy, 2019, pp. 394-395.

Thank you for your attention!

 iarapakis

 ioannis.arapakis@telefonica.com

 http://iarapakis.github.io

Pilot 3: I-BiDaaS application to the Manufacturing Sector

Giuseppe Danilo Spennacchio
Flexible and Adaptive Systems Specialist
at Factory Innovation Department
CRF

December, 21st 2020

CRF's Participation in I-BiDaaS

Manufacturing production processes are complex in that production lines have several **robots and digital tools**

massive amounts of raw data are gathered; **data** that do not only **help to monitor processes, but can also improve process robustness and efficiency**

Within the I-BiDaaS project, we identified two scenarios, in which **complex and initial structured/unstructured data sets** are retrieved **from real processes**, and defined two use cases:

1. **Production process of Aluminium die-casting**
2. **Maintenance and Monitoring of production assets**

The project focuses on providing a self-service solution that will give us the insights and tools we need **to develop a methodology to implement in production sites** :

- **improve the quality of the processes and products** in a much more agile way, **reducing costs** due to reworks and scraps
- **predict faults** and **unnecessary actions** (preventive or planned maintenance)

Requirements

- Data are strictly **confidential**
- Computerized production lines store large amounts of complex structured/unstructured data

- **Lack of time** to extract and analyse data due to fast rhythms of production
- **Fast internal changes** due to rescheduling production quantities and component variations
- **Slowdown of data sharing** due to security procedures

Requirements:

- **Data privacy:** anonymise data that can not be accessed or shared with external
- **Data cleaning:** identify incomplete, inaccurate and irrelevant parts of the data – dataset generation
- **Data analysis and advanced visualisation tools:** empower data providers decisions

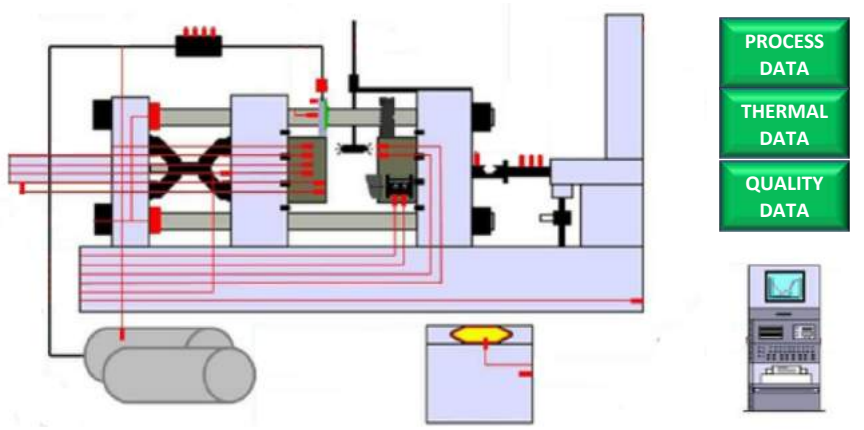
Objective:

- **Quality Improvement and PM:** Exploit the I-BiDaaS platform to easily utilise Big Data analytics and develop a methodology for quality and process improvement and for Predictive Maintenance

Production Process of Aluminium die-casting

Die-casting process:

- heterogeneous data such as piston speed in the first and second phase, piston stroke, intensification pressures, temperatures, cooling capacity, etc
- operator's data: qualitative evaluation of the process, events, etc.(e.g.: defect manually detect)



Business Goal

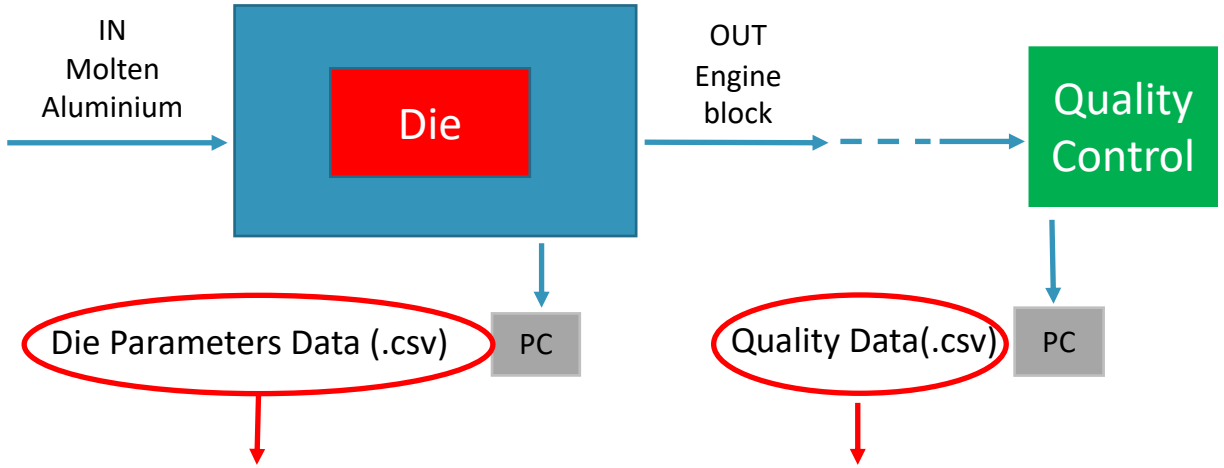
Quality Process Improvement:

- reduce scrap and waste
- prevent repairs and further processing, by avoiding unnecessary actions after casting: impregnation, cooling, storing and management of failed engines

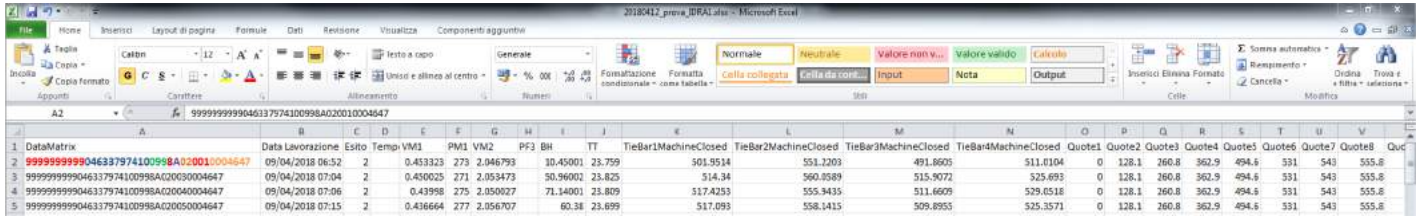
Use Case Goals

- develop high level algorithms
- identify critical parameters
- provide a mode to timely check the status of the process and classify the quality levels that we are using as KPIs

Production Process of Aluminium die-casting



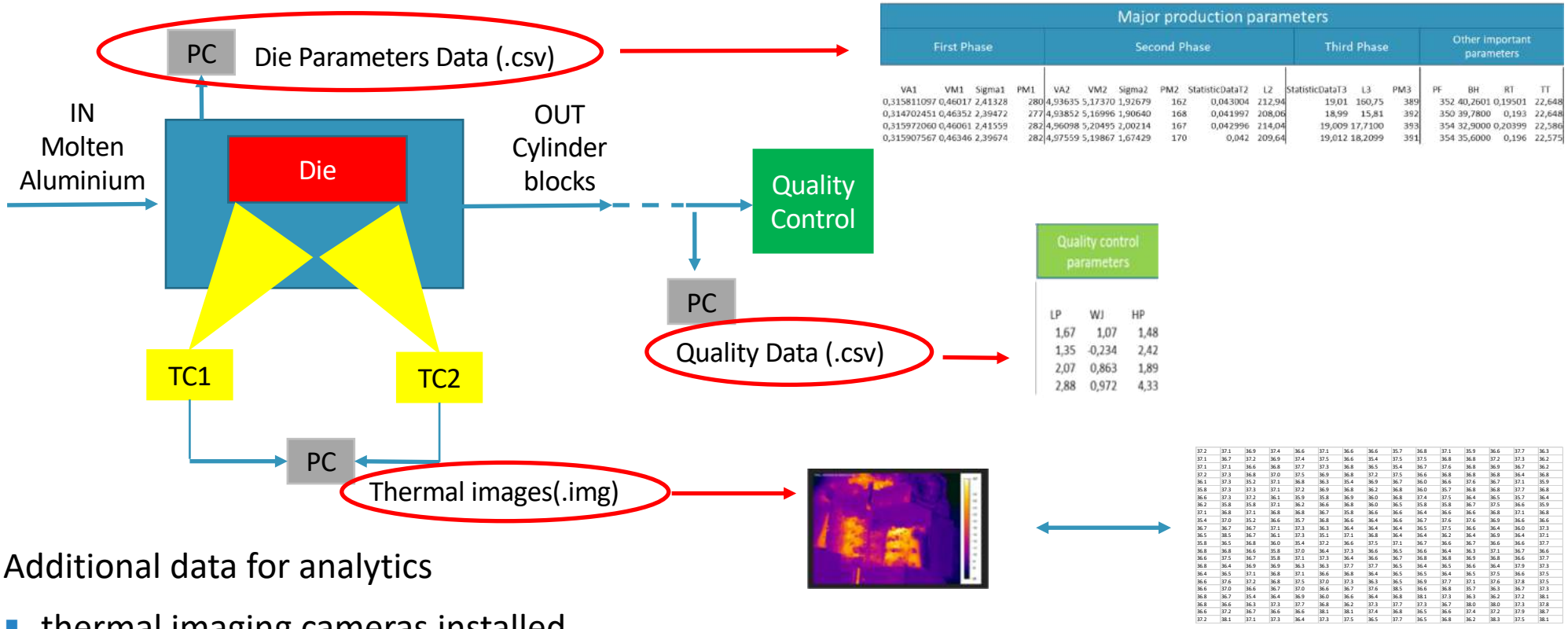
- The initial data set contained: production, process and control parameters of the production of the engine block by die-casting
- Initially, CRF provided 2 datasets: 'Synthetic' and 'Anonymised'



Breaking data silos

- data sharing between company owning the data, company performing the analysis (CRF) and consortium partners
- integrate data from different sources/ levels
- involve external/ internal departments

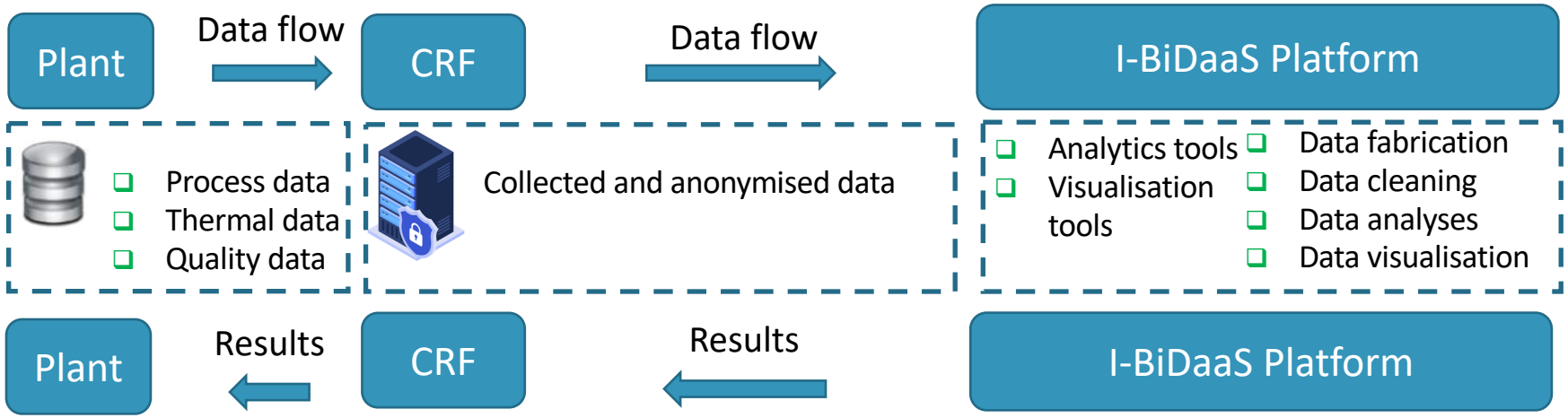
Production Process of Aluminium die-casting



Additional data for analytics

- thermal imaging cameras installed
  - every engine is assessed
  - each frame is discretised
  - an output file (CSV) is generated
- Temperatures values (estimated 400MB per day) to share for further analysis

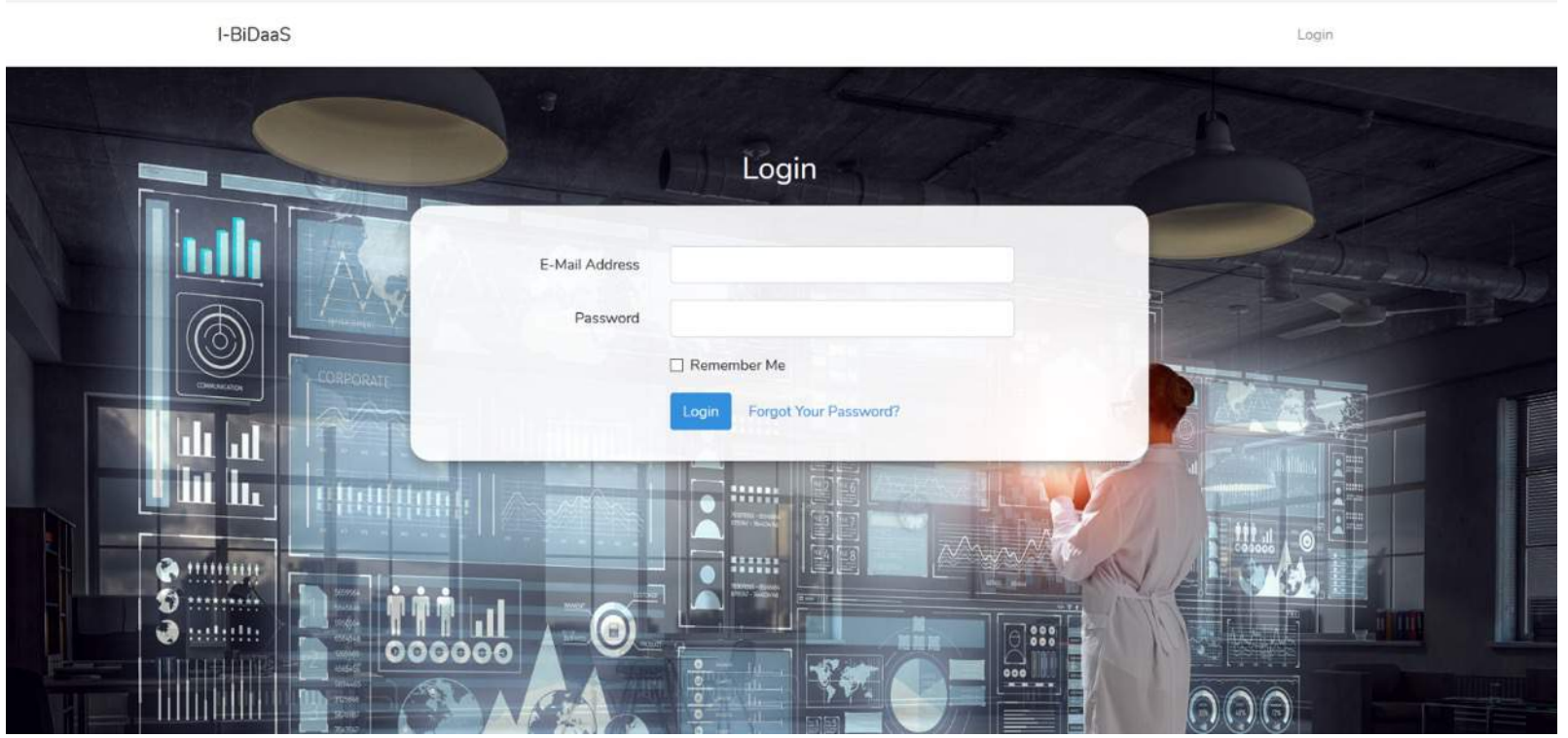
Production Process of Aluminium die-casting



- corporate constraints do not allow to give access to CRF internal server
- data were sent to the Virtual Machine that host I-BiDaaS Platform

Production Process of Aluminium die-casting

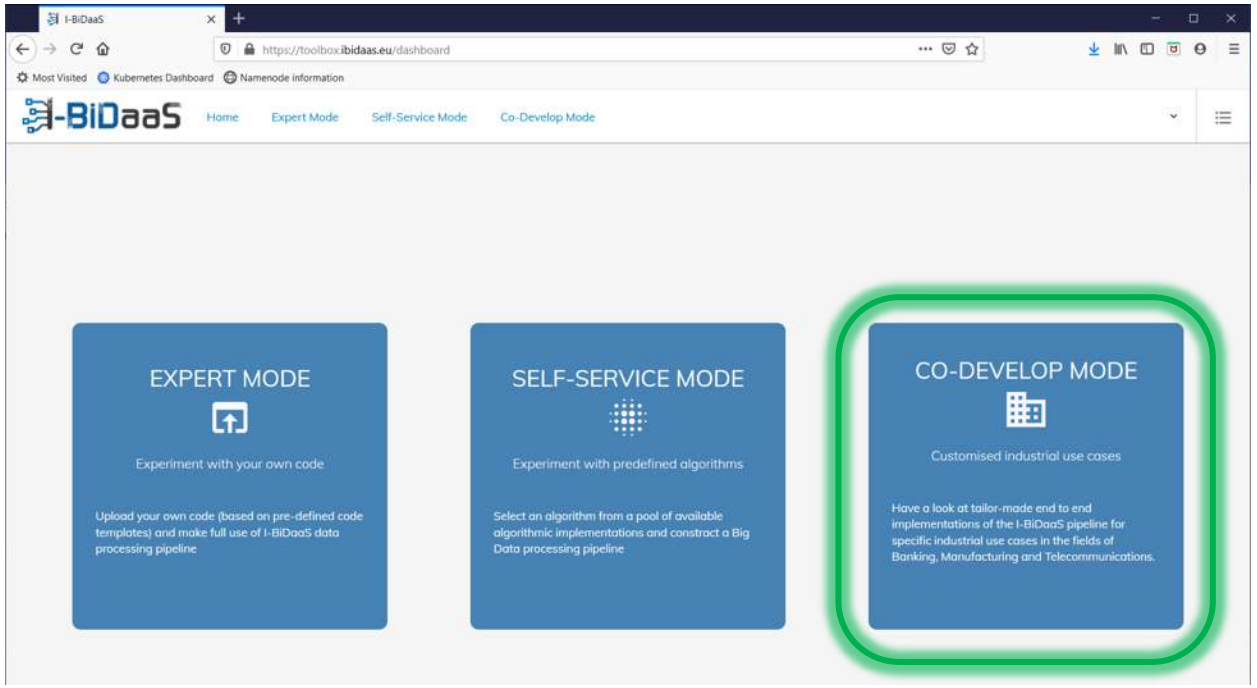
I-BiDaaS platform to easy utilise I-BiDaaS solutions developed by data analysts and technologists



Production Process of Aluminium die-casting

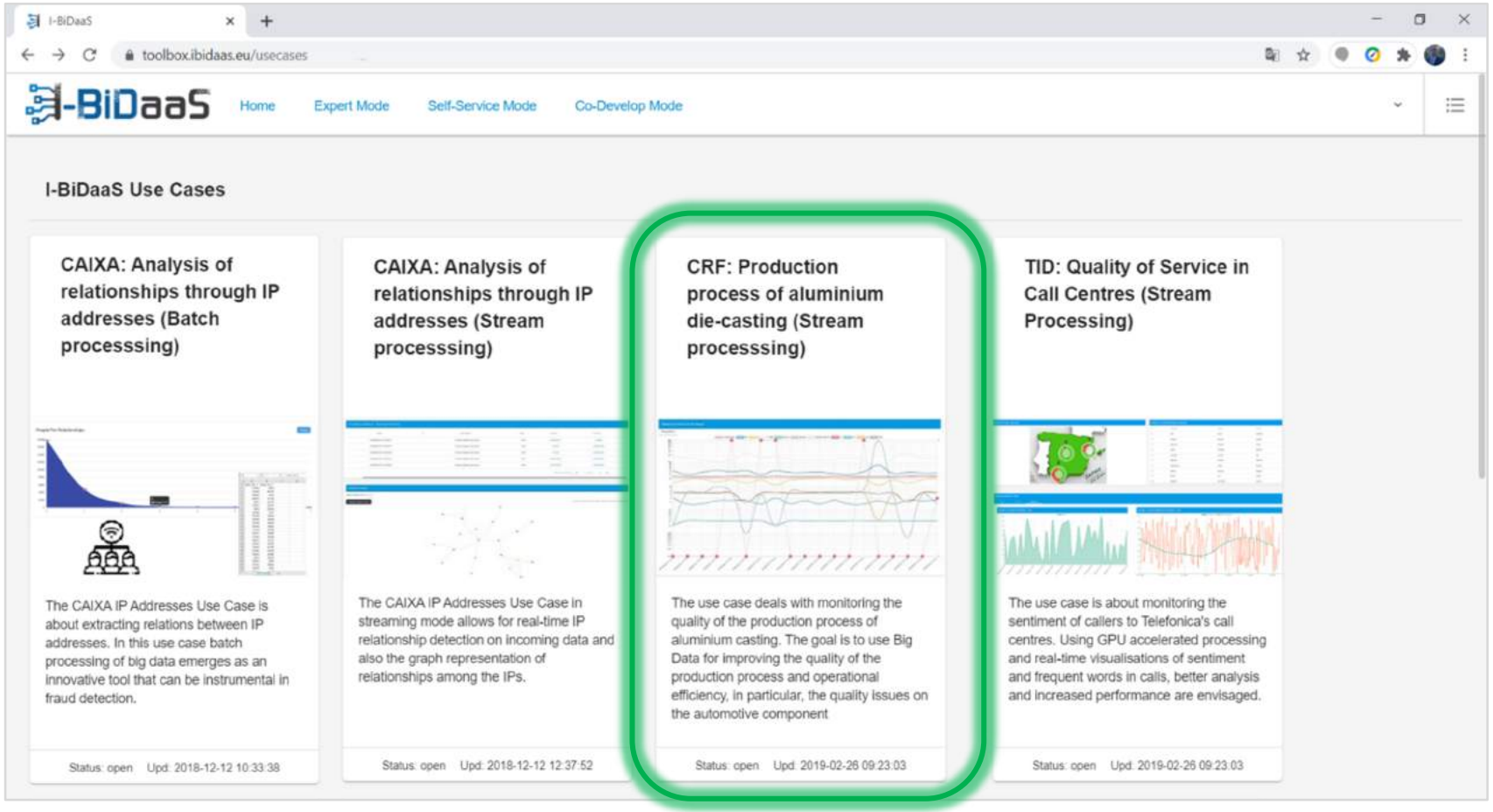
After login, we select the co-develop mode:

- end-to-end solution developed by the I-BiDaaS team for this use case
- dataset formats and analysis setup are predefined and custom-made according to our needs
- examples of how the I-BiDaaS platform and tools can be applied to real-world applications based on real business needs



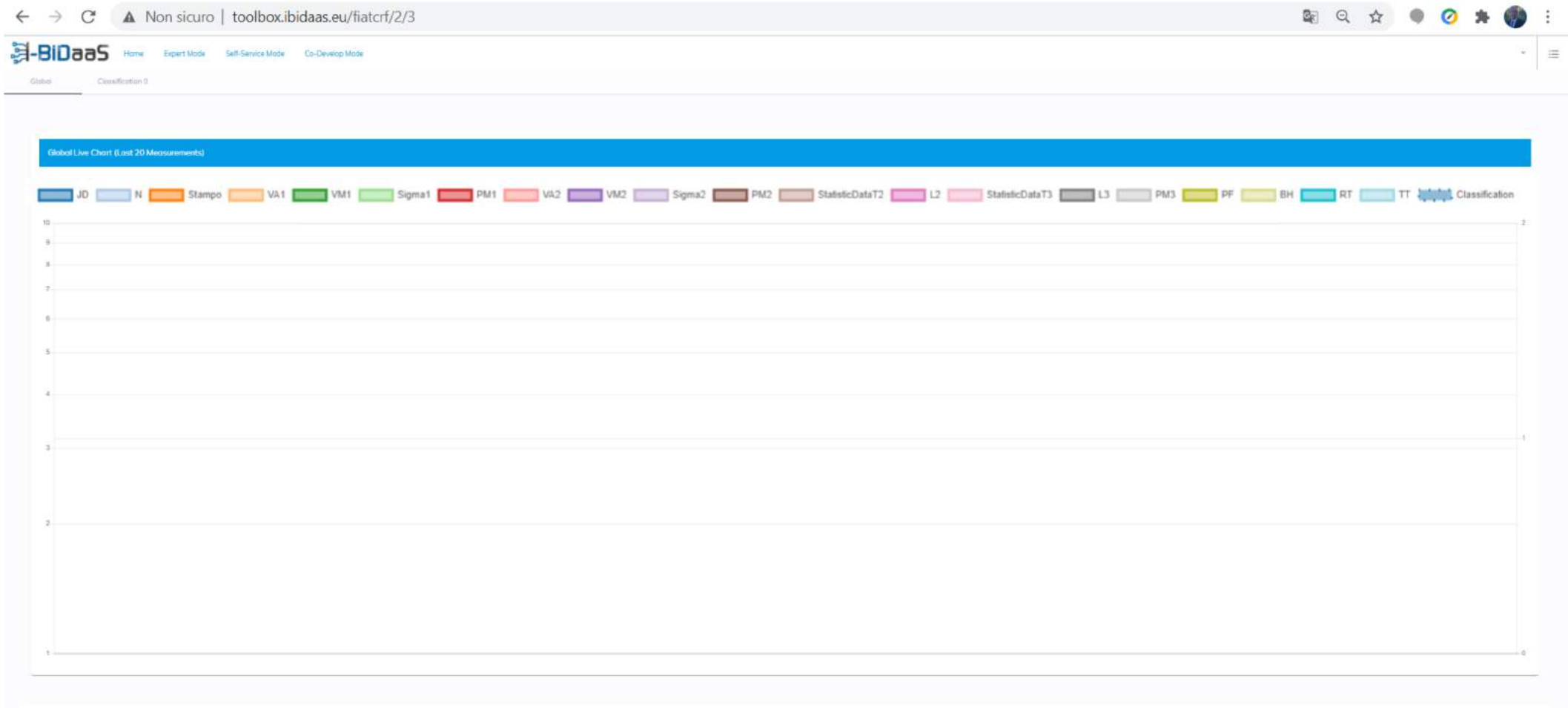
Production Process of Aluminium die-casting

The production process of Aluminium die-casting is a complex industrial process with many heterogeneous parameters. The co-develop mode allows to us to quickly check the quality level classification



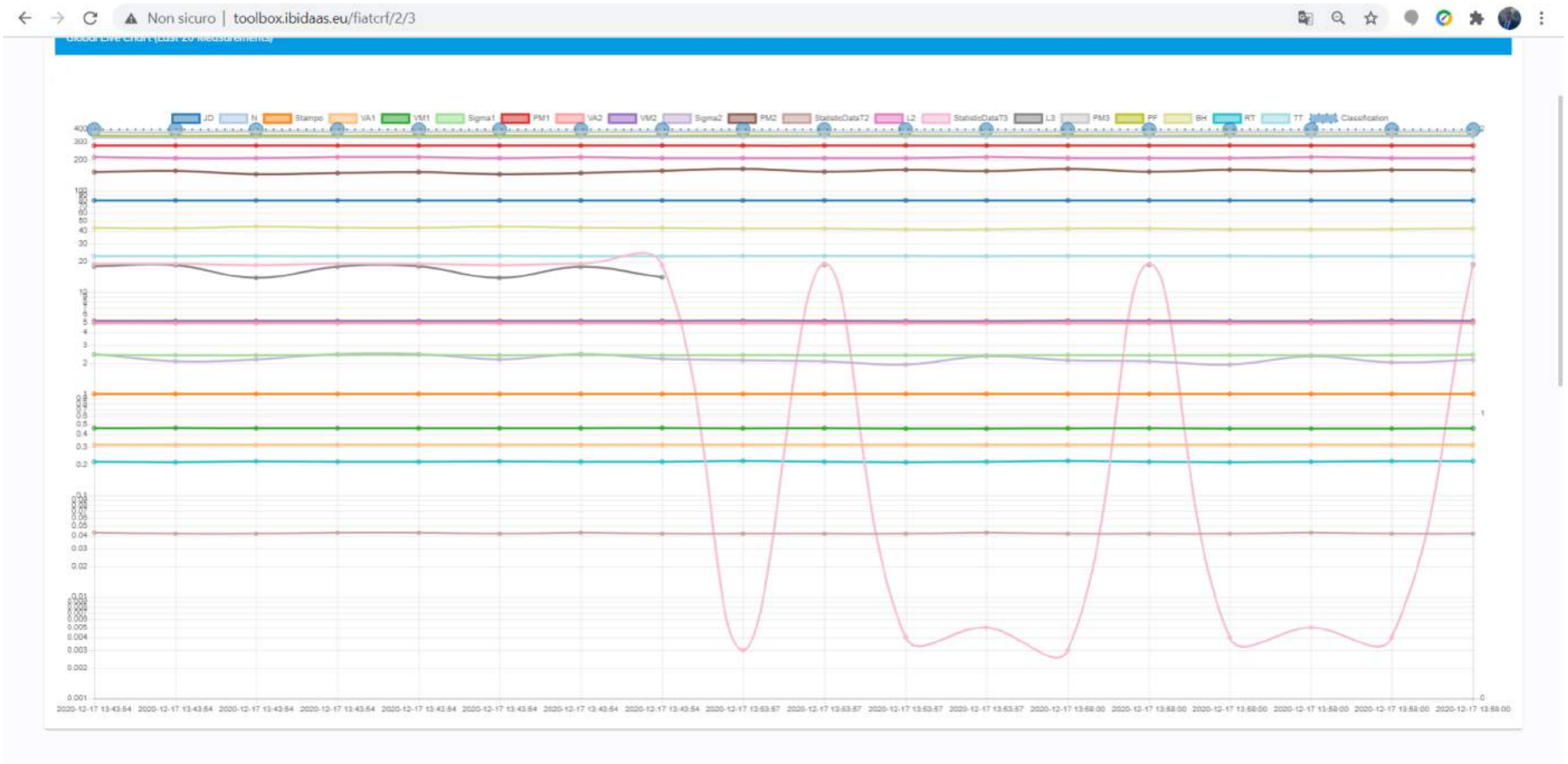
Production Process of Aluminium die-casting

A Global live chart shows the main parameters of the incoming streaming data in real-time



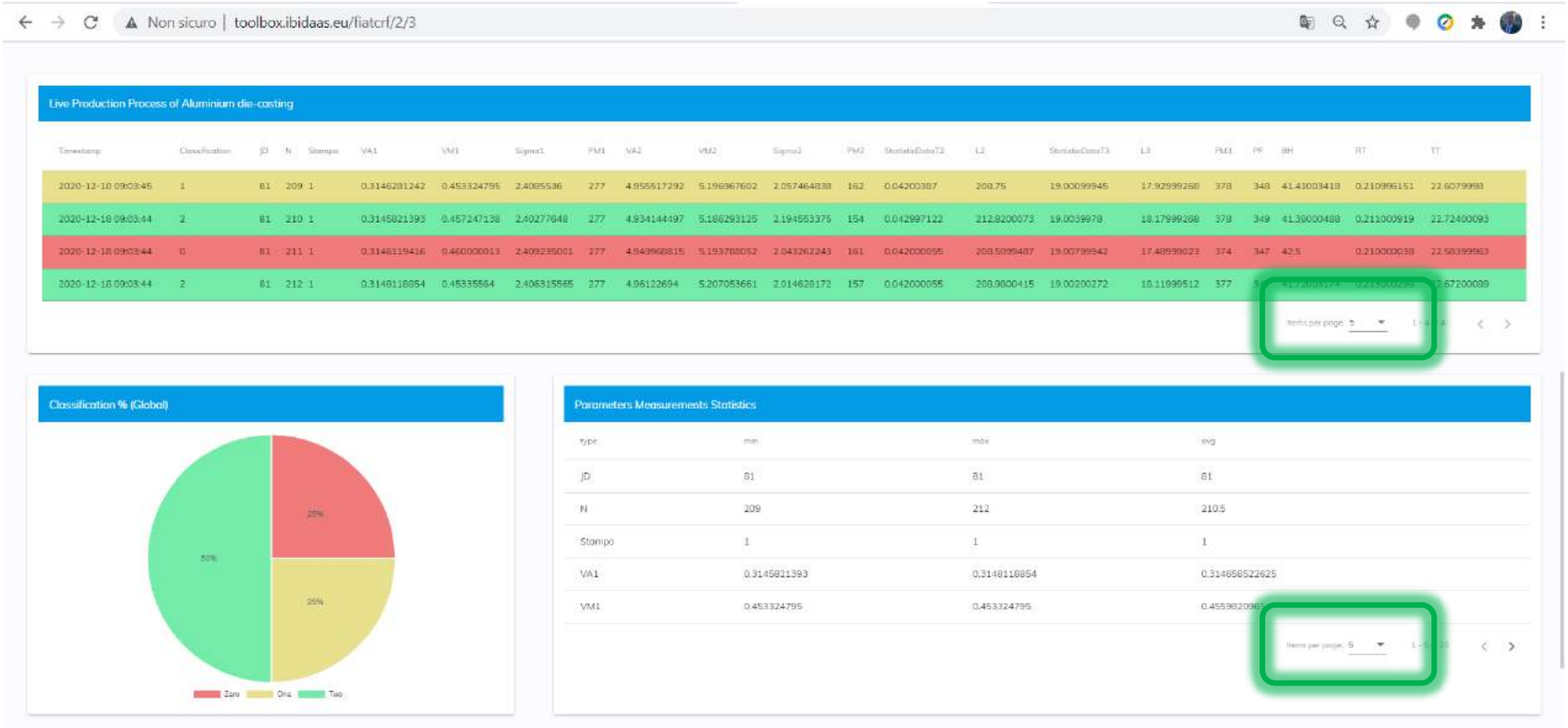
Production Process of Aluminium die-casting

We can visualise the trend of parameters and quickly check if there is something of anomalous compared to the set values of parameters



Production Process of Aluminium die-casting

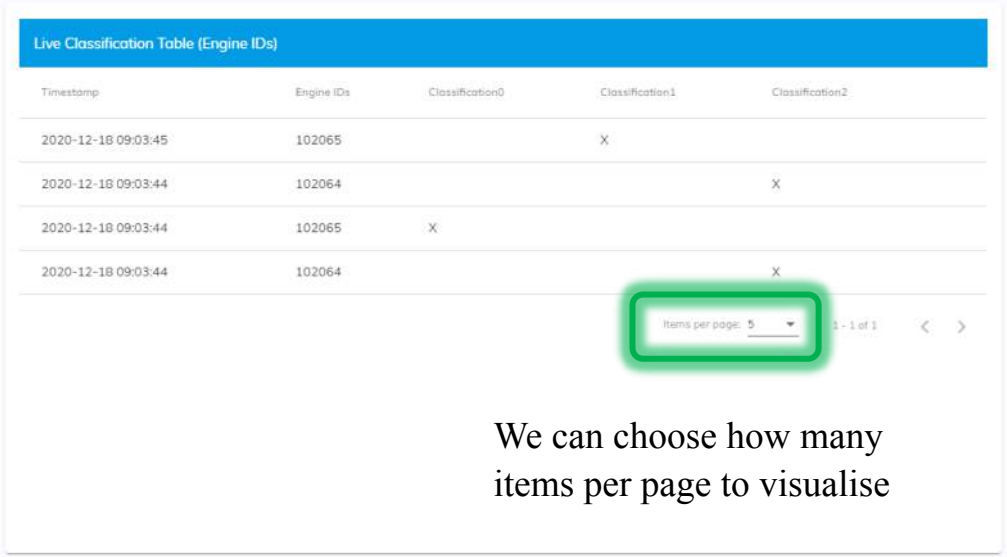
We can visualise the data table, that shows a colour- coded presentation of the results according to a given dimension of data (classification level)



We can choose how many items per page to visualise

Production Process of Aluminium die-casting

We can visualise in real time and quickly the sequence of engine block with their level of classification

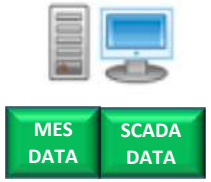


We can choose how many items per page to visualise

Maintenance and Monitoring of production assets

Welding cell data come from:

- sensors mounted on different machines:
  - accelerometers mounted on linear stages, robots, elevators
  - electric consumption, flows, temperature and energy vectors
- production data: model, variant, mix



Business Goal

Predictive Maintenance:

- predict faults before they happen
- predict unnecessary actions (preventive or planned maintenance)
- avoid micro or macro stoppages
- predict retooling and refurbishing

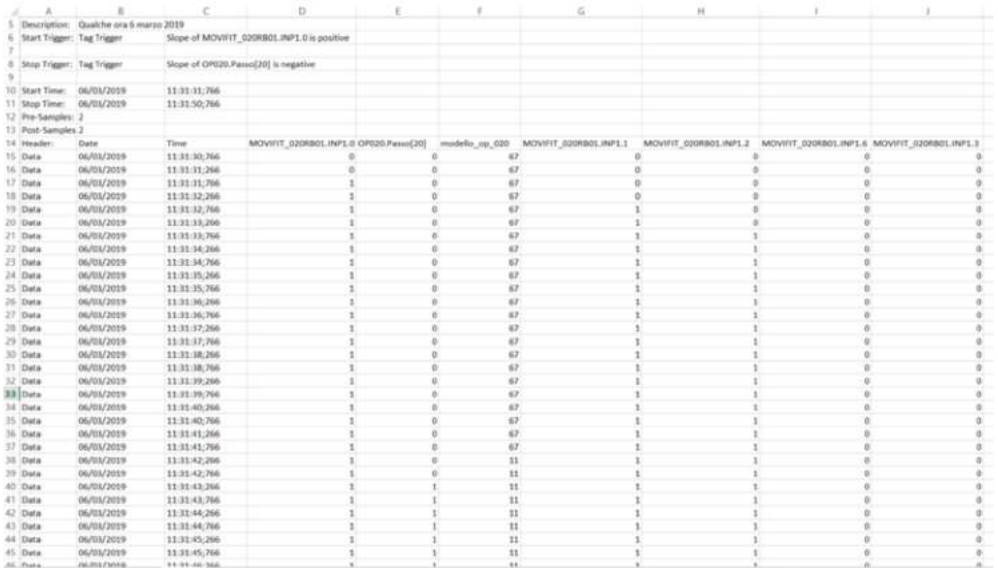
Use Case Goals

- collect, structure and analyse sensor data with high level algorithms
- obtain thresholds for anomalous measurements
- create a structured foundational database to be easily utilised to check outlier detections for a continuous and periodic control of the service conditions (PM)

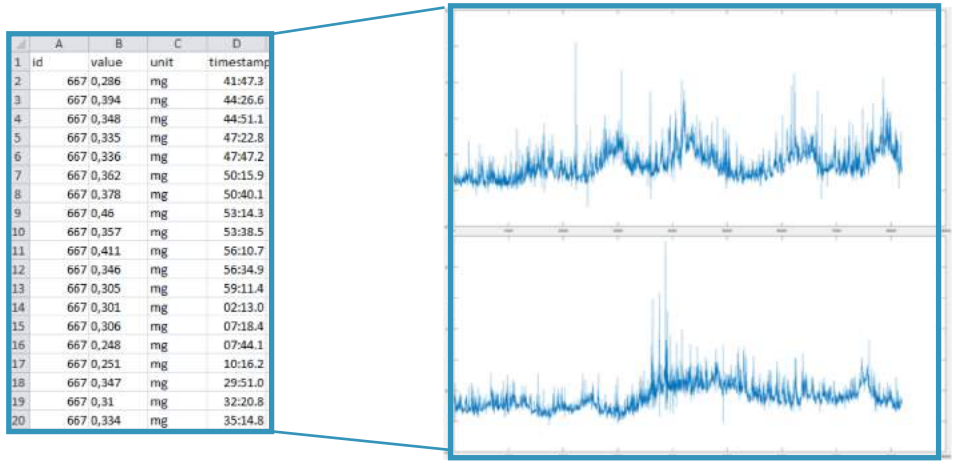
Maintenance and Monitoring of production assets

The data set contains: production, process and control parameters of the production of the Daily vehicle.

CRF provided 2 data sets: “**Anonymized SCADA**”, “**Anonymized MES**”,



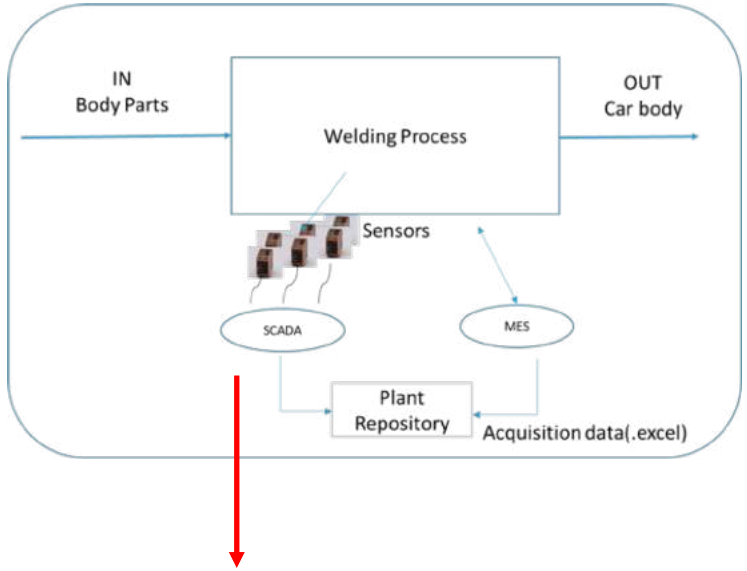
- data related to the car body model processed
- to allow to detect different vibrations due to the different car body model that is passing on the production line



	id	value	unit	timestamp
1				
2	667	0,286	mg	41:47.3
3	667	0,394	mg	44:26.6
4	667	0,348	mg	44:51.1
5	667	0,335	mg	47:22.8
6	667	0,336	mg	47:47.2
7	667	0,362	mg	50:15.9
8	667	0,378	mg	50:40.1
9	667	0,46	mg	53:14.3
10	667	0,357	mg	53:38.5
11	667	0,411	mg	56:10.7
12	667	0,346	mg	56:34.9
13	667	0,305	mg	59:11.4
14	667	0,301	mg	02:13.0
15	667	0,306	mg	07:18.4
16	667	0,248	mg	07:44.1
17	667	0,251	mg	10:16.2
18	667	0,347	mg	29:51.0
19	667	0,31	mg	32:20.8
20	667	0,334	mg	35:14.8

- accelerometers: used for measuring vibration and shock on machines and basically anything that moves
- the monitoring of vibrations is important to check the status of a machine
- the analysis of the trend of vibrations over time allows to predict the onset of deterioration and to intervene in time before the failure

Maintenance and Monitoring of production assets



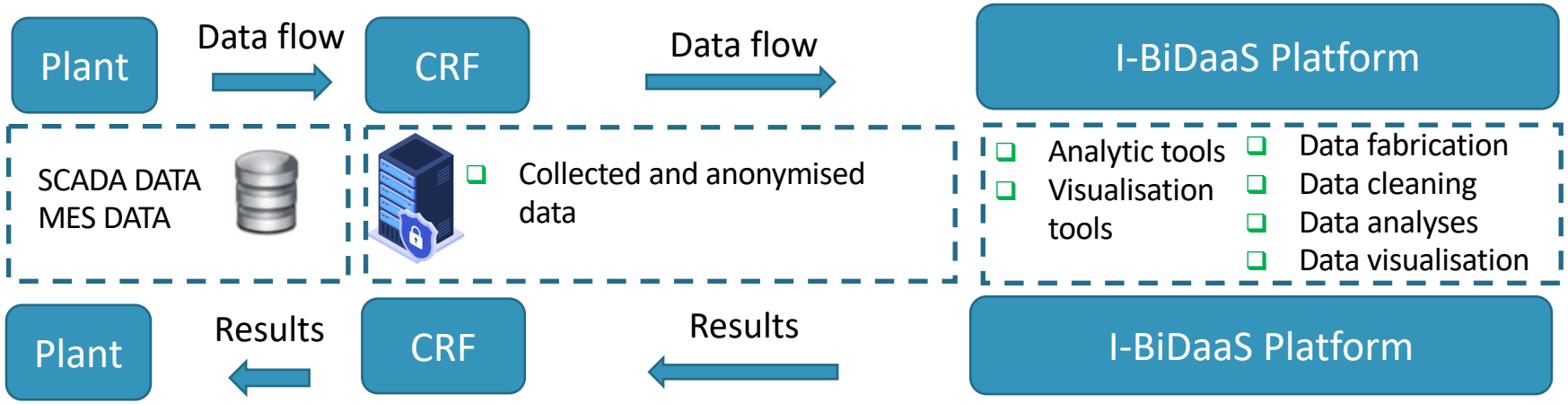
- MES data showed high variance in measurements even for the same vehicle type
- problems to retrieve MES data because of rescheduling activities and changes in the production lines, partially due to the Covid-19 situation
- relevant information are in SCADA data, used to obtain thresholds for anomalous measurements for all sensors

Number of sensors	Physical Quantity Measured	Unit of measurement
87	Acceleration	mg
30	Velocity	mm/s
9	Temperature	°C
8	Pressure	bar
8	Flow	l/min
1	Displacement	mm
2	Energy Vector(water)	l
1	Energy Vector(air)	m^3
1	Energy Vector(air)	m^3/h

Id identifies the sensor

id	value	unit	timestamp
109	122,149	mg	31/1/20 5:07:49
47	188,334	mg	31/1/20 5:07:49
68	153,919	mg	31/1/20 5:07:49
92	166,489	mg	31/1/20 5:07:49
32	3549,391	mg	31/1/20 5:07:49
44	1092,876	mg	31/1/20 5:07:49

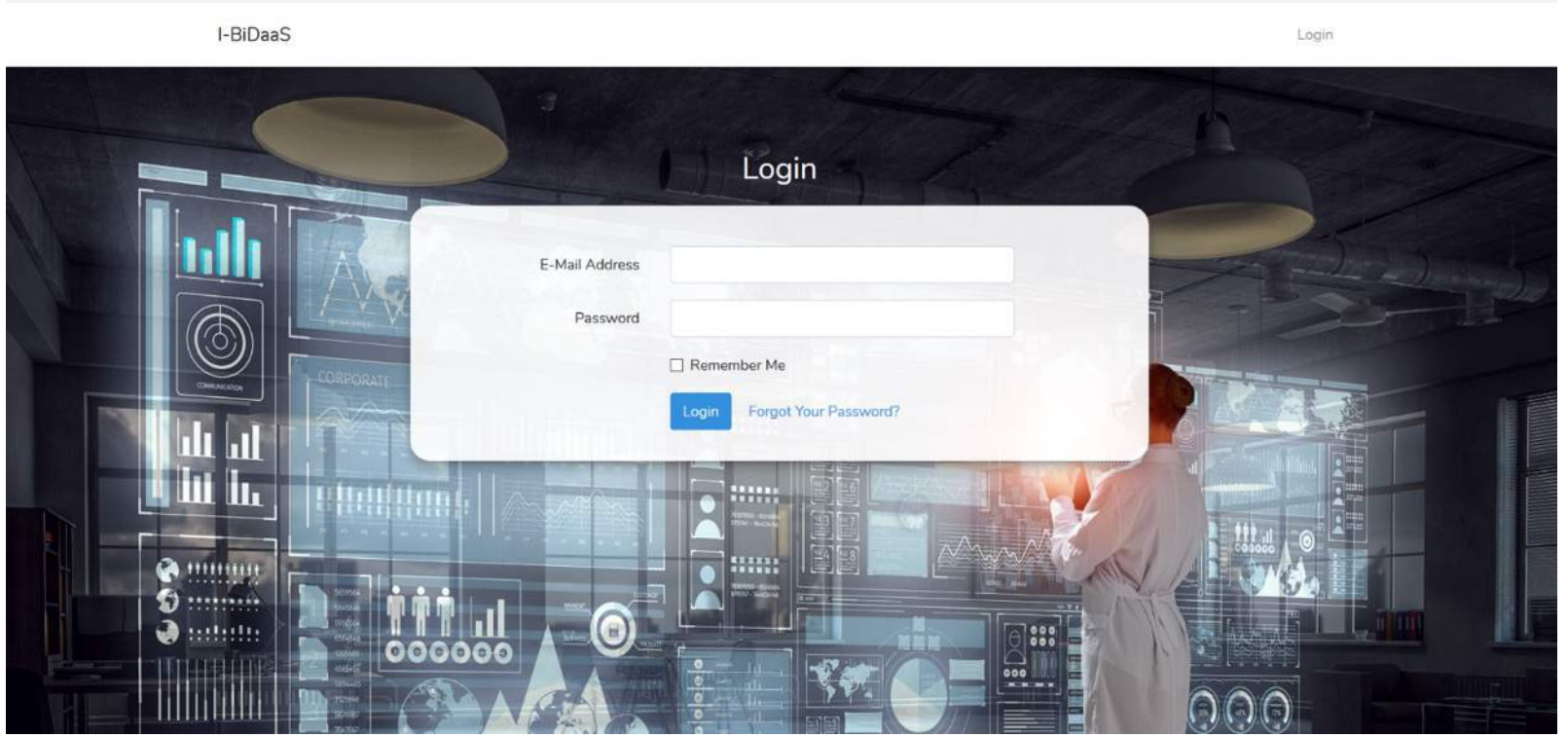
Maintenance and Monitoring of production assets



- corporate constraints do not allow to give access to internal servers
- data were sent to the Virtual Machine that host I-BiDaaS Platform

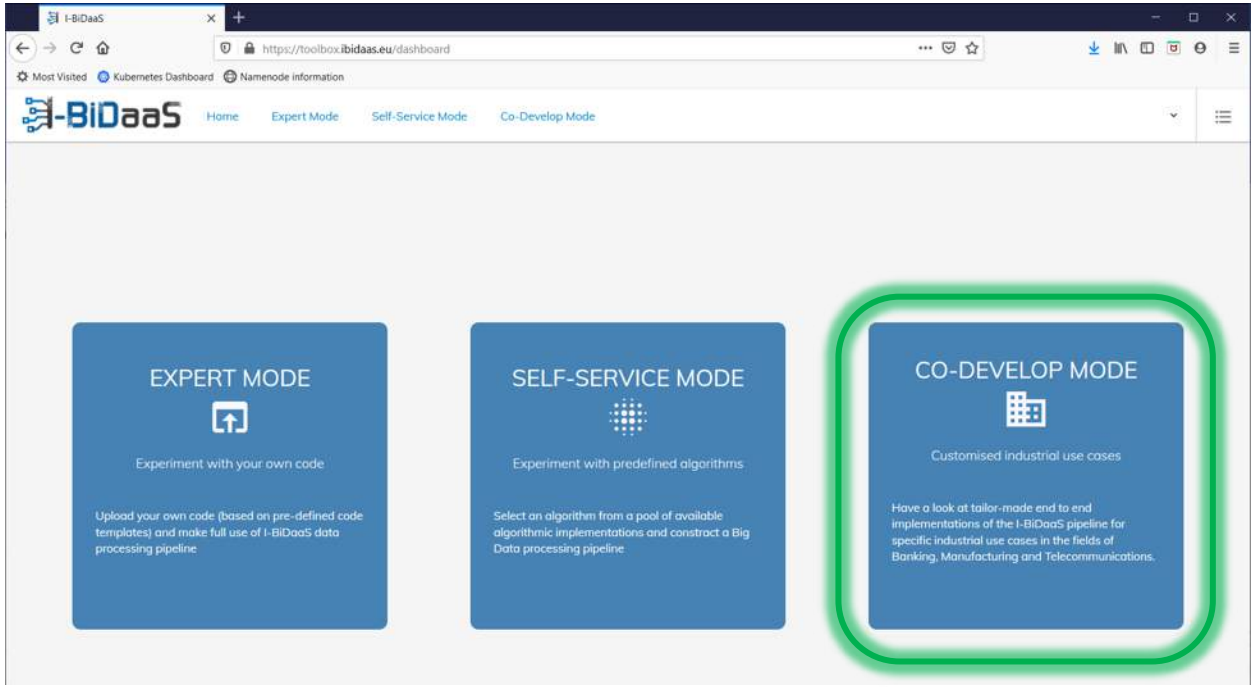
Maintenance and Monitoring of production assets

I-BiDaaS platform to easy utilise I-BiDaaS solutions developed by data analysts and technologists



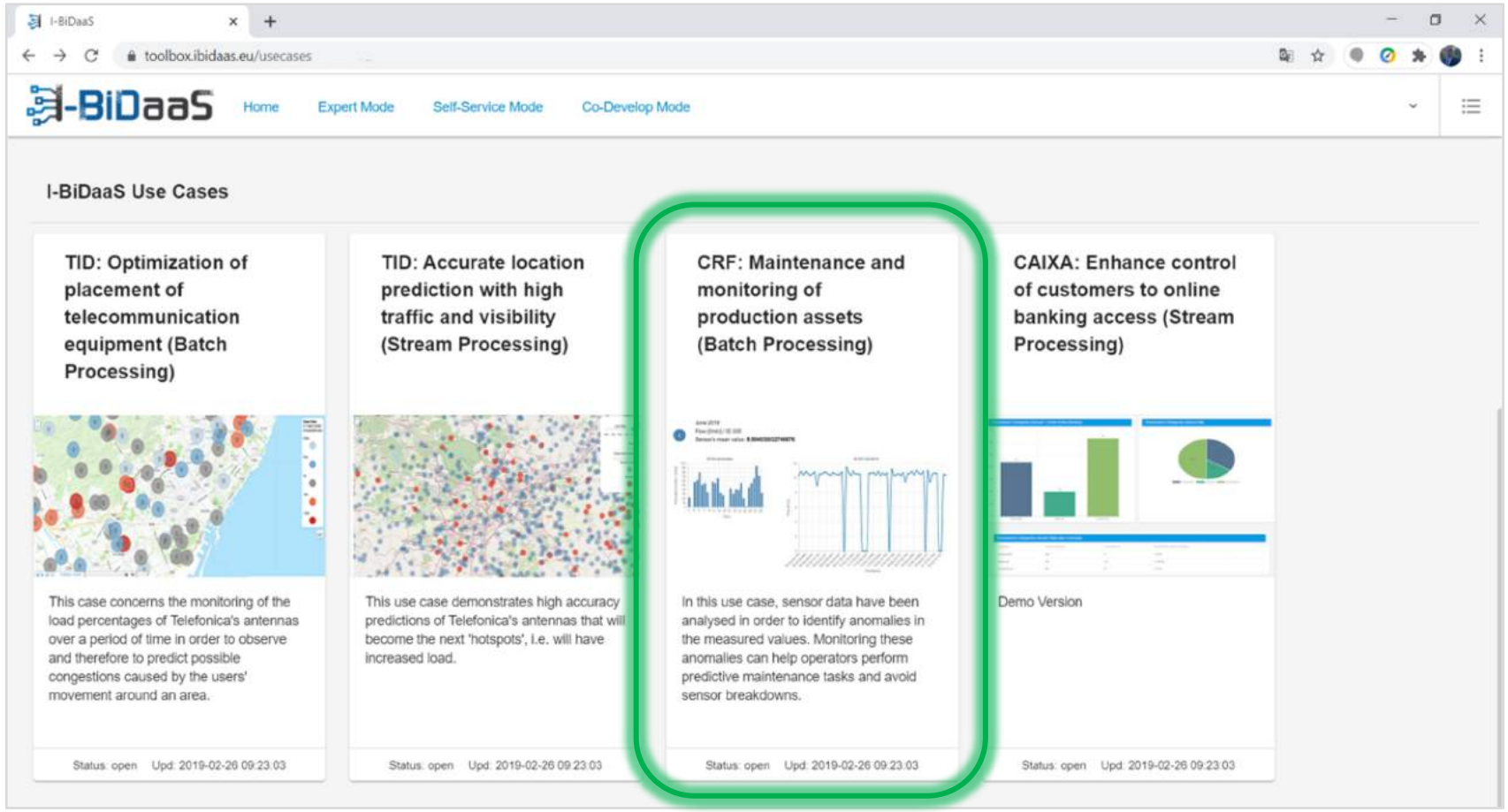
Maintenance and Monitoring of production assets

After login, we selects co-develop mode, in which the data source and analysis algorithm are predefined



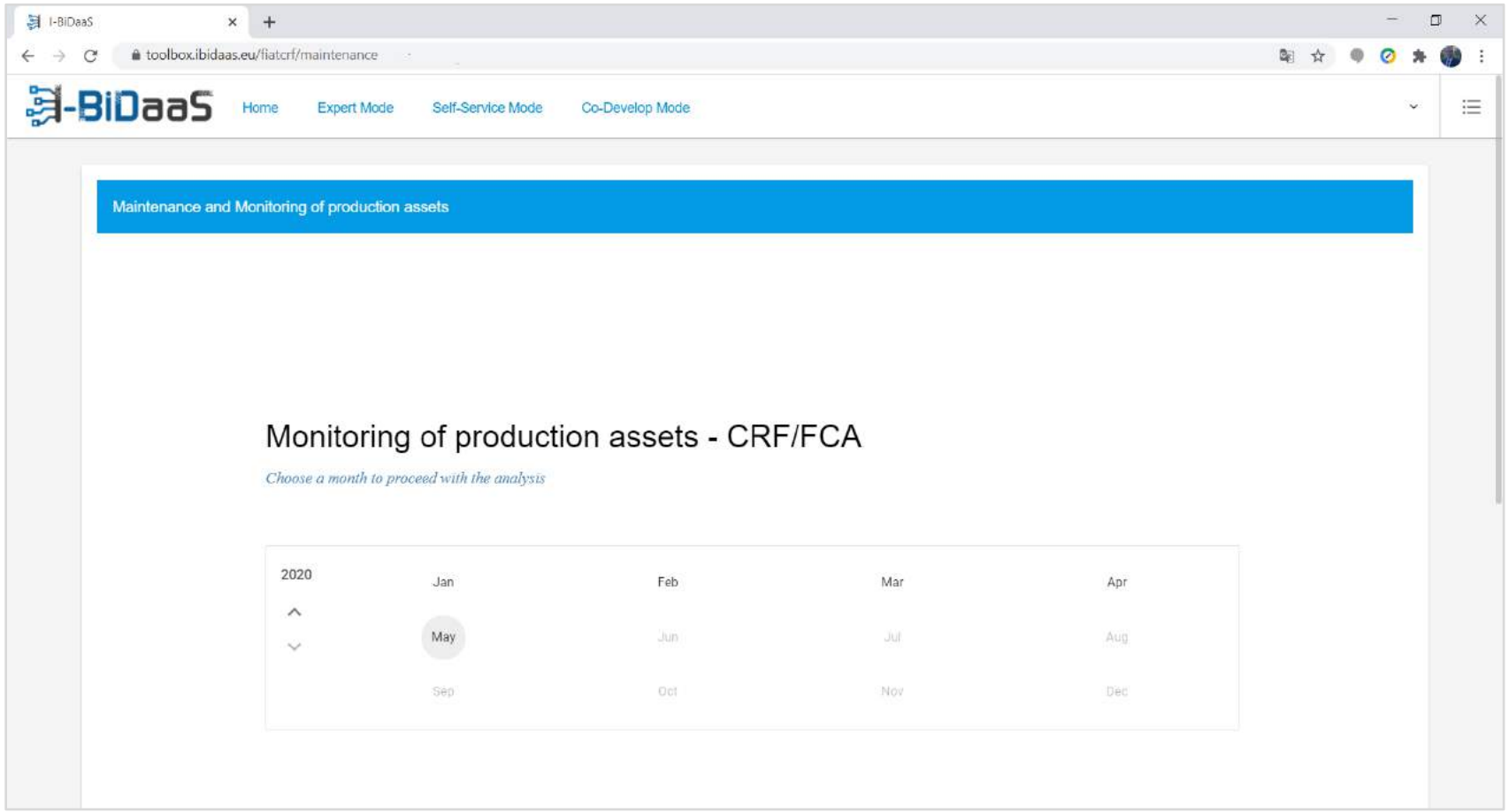
Maintenance and Monitoring of production assets

Maintenance and Monitoring of production assets. The co-develop mode allows us to monitor the outlier time series for each sensor any day



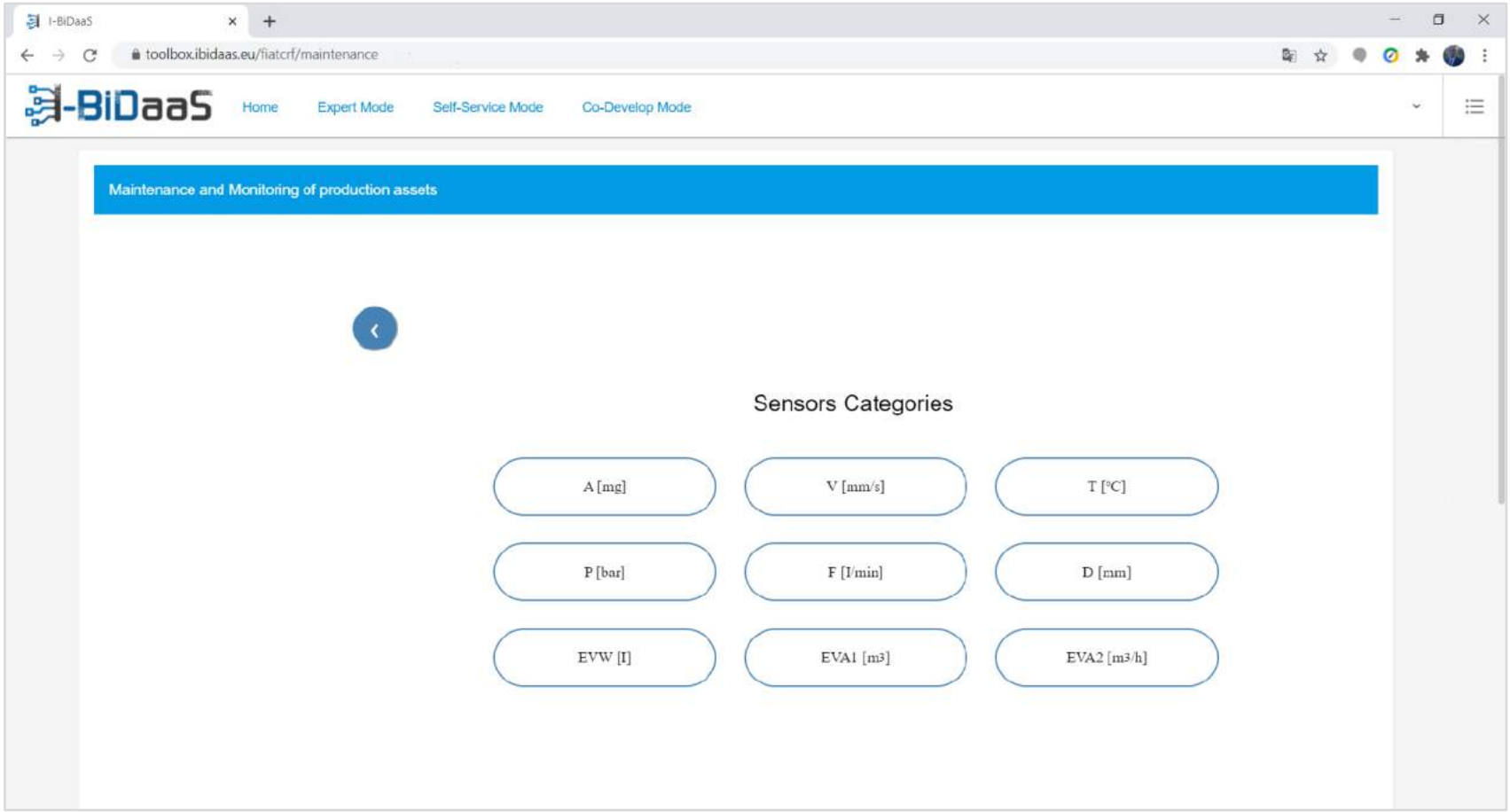
Maintenance and Monitoring of production assets

Structured foundational database to be easily utilised to check outliers for a continuous and periodic control of the service conditions (Predictive Maintenance).



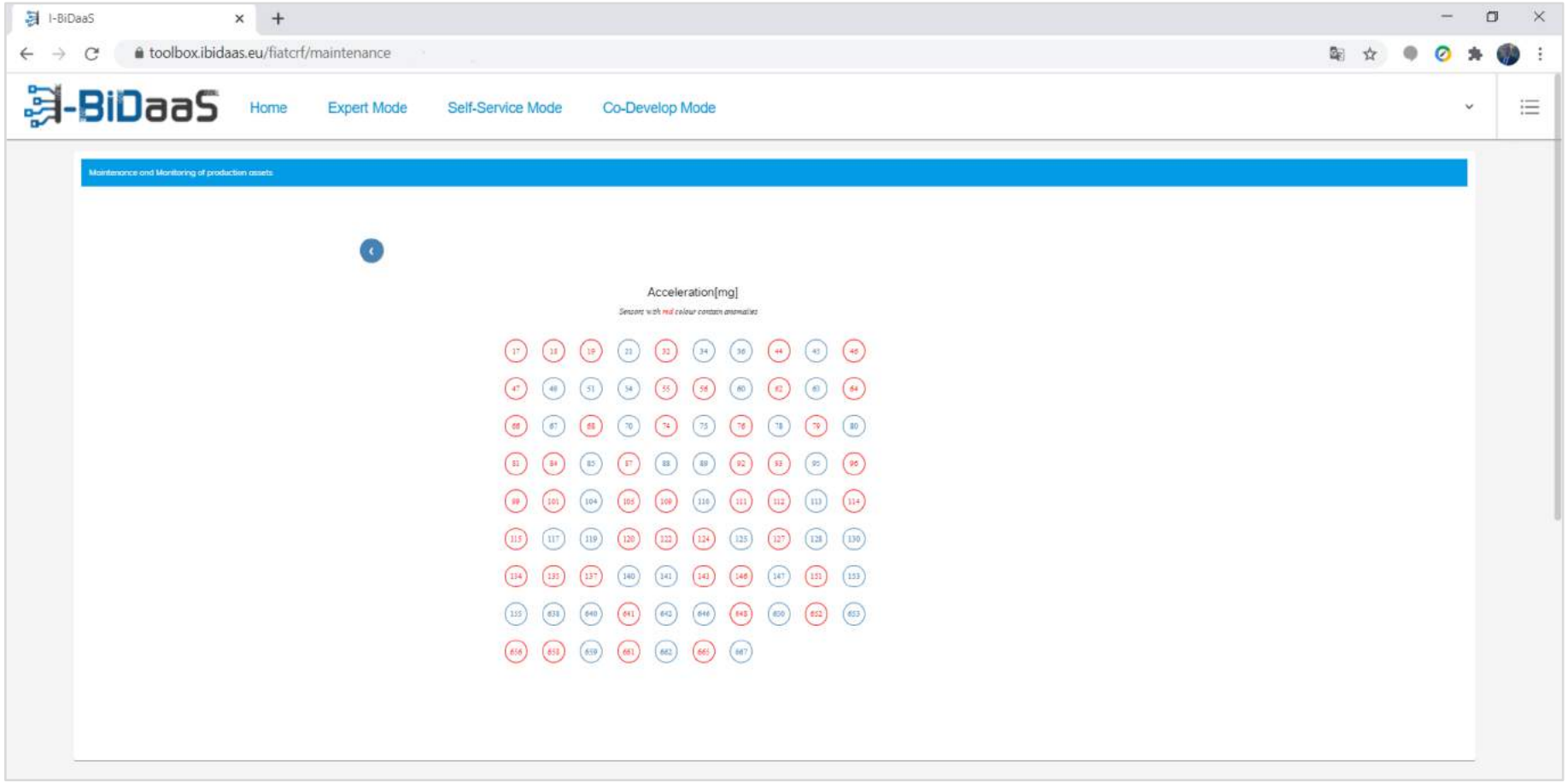
Maintenance and Monitoring of production assets

Sensors - Categories



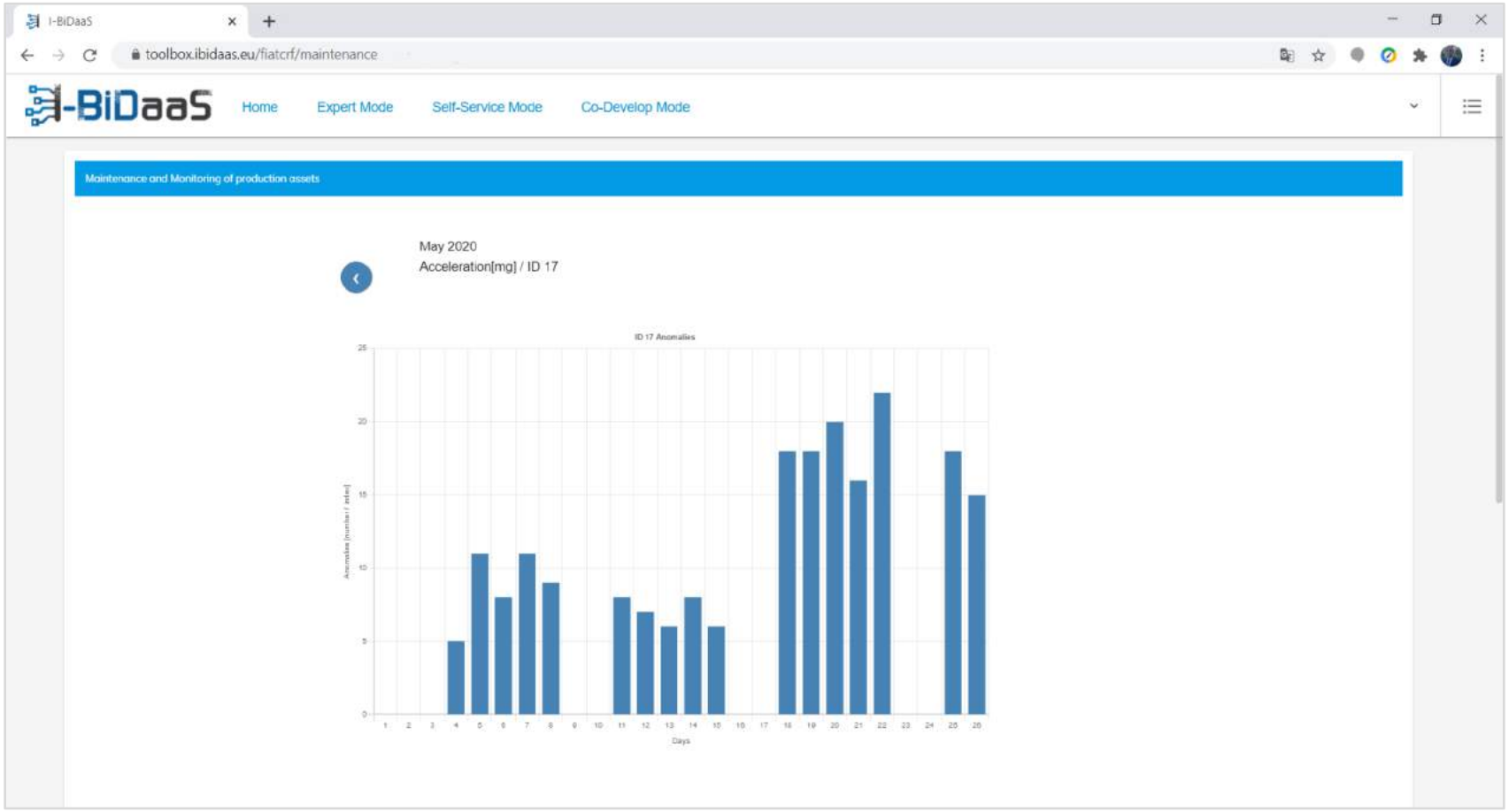
Maintenance and Monitoring of production assets

Accelerometer [mg] / Id sensor



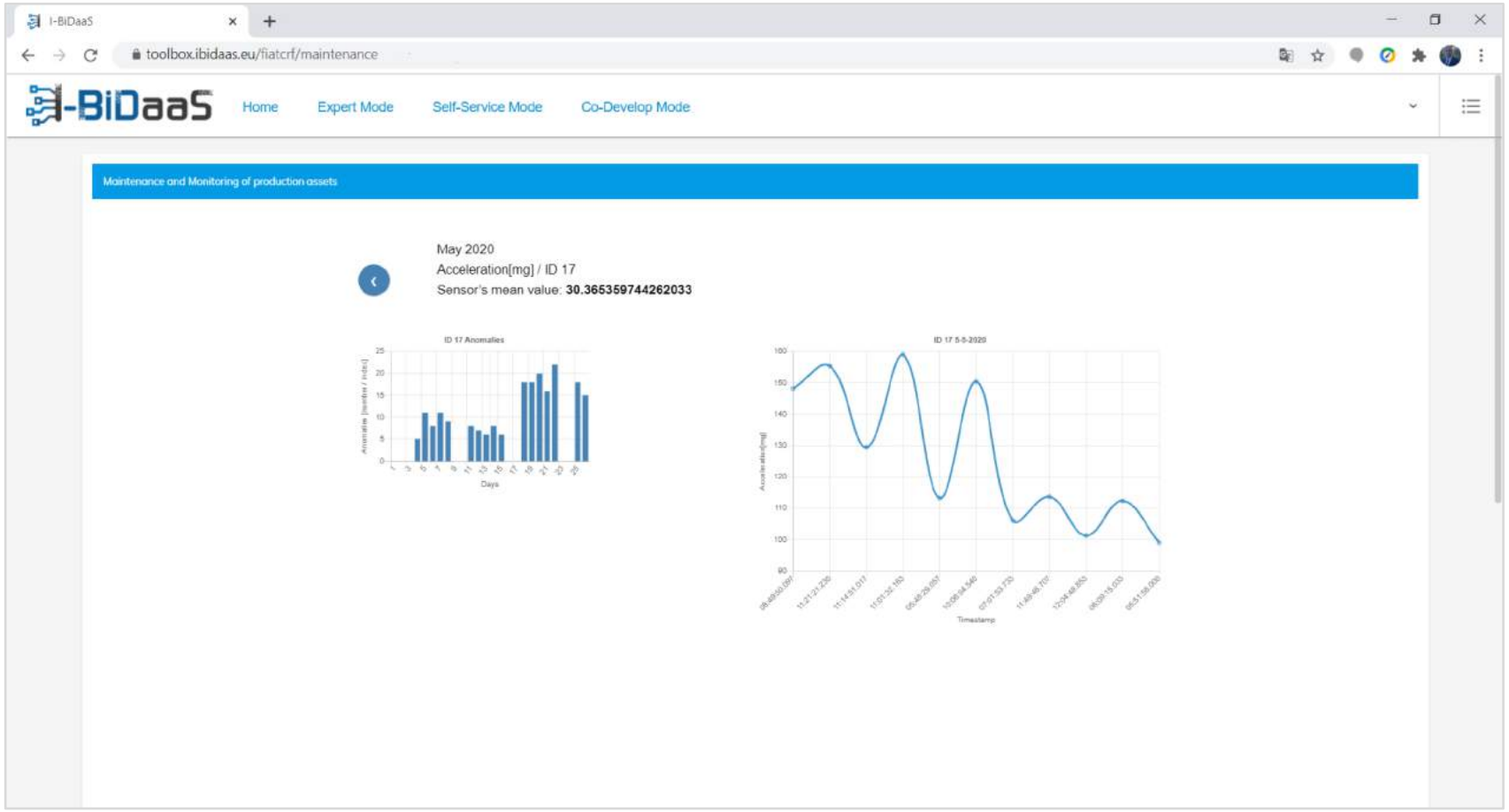
Maintenance and Monitoring of production assets

Anomalies number per day



Maintenance and Monitoring of production assets

- Data has been transformed into separate time series - one per sensor
- anomalous measurements



CRF benefits from I-BiDaaS

Benefits	KPIs
enhance production times, to reduce costs and, consequently, to obtain results that satisfy manufacturers' requests in terms of product quality, machine performance and timing	Increase of 3-7 % of quality control level related to good products and decrease of 1-4% and 0.5-2 % of two quality control levels related to defective products
improve the efficiency of manufacturing plants, by getting the best performance from the machinery to reduce production losses and achieve greater competitiveness of the company	Increase of 1-1.5 % in current Overall Equipment Effectiveness (OEE) and decrease of 20-30 % in maintenance costs
reduce time to produce decisions	From one month to few hours, a turn of job or 1 day
achieve accuracy of new models with respect to internal CRF models in use	Increase of 6 % for the first use case and 20% for the second one

Hackathon in Campus Melfi – June 2019

- students, PhDs, researchers and employees from SMEs have been involved in the CRF's Hackathon
- the data set was previously processed by the CRF which structured and anonymized the large volume of data provided by the plant
- the teams worked by using different softwares (Minitab, Excel, R) and methods: Neural Networks (Multi Layer Perceptron MLP), Decision Trees, Random Forest and Generalized Additive Model (GAM) for Big Data
- each group was able to find valuable results using also different techniques than the ones used in the analysis performed by CRF previously

Hackathon in Campus Melfi – June 2019

- Technical results:

  - ✓ **Benchmark**: With the same data set, the I-BiDaaS team achieved 80% classification accuracy, while best reported by other teams was 73-74%.
  - ✓ **Validation**: Other teams found similar features as the I-BiDaaS team did, while the modelling approach was different. This cross-validates the project approach with others.
  - ✓ **Casting process phases**: Other teams took into account the three phases of the casting process while the project team did not.
  - ✓ **Data cleaning**: Other teams pointed to a problem in Result_2 data, i.e., there was class label 4, while we did not take this into account.

- Other important outcomes/advances:
  - Break silos: I-BidaaS got access to real anonymized data and engine casting process, the data is at least order of magnitude larger than the data previously available to the consortium.
  - Data fabrication: New approaches for fabricating data *with labels* were brainstormed and delivered during the hackathon.

Lessons learned - Best practices

1. **Data ingestion** and **data management** from different data sources and levels over time
2. Extraction of the **value of all data** collected
3. Involvement of **different departments** belonging to the same or different organizations (**breaking data silos)**
4. **Data cleaning** (to identify incomplete, inaccurate and irrelevant parts of the generated dataset)
5. **Fabrication of realistic synthetic data** for experimentation and testing
6. **Batch and stream analytics** for increasing the speed of data analysis for the specific identified business requirements
5. To bridge the **I-BiDaaS infrastructure** with the industrial internal server **for real-time data transfer and near to real-time data analysis**
6. To create an **iterative process** for the data collection design process and data analysis
7. To allow **flexibility and redundancy** to take into account unexpected or severe conditions, including subsystem failure, unavailability of data, etc.
8. **Simple, intuitive and effective visualization of results** and interaction capabilities for the end-users

Conclusion

Specialists and competencies:

- discover new information:

  data are stored, but also they are unstructured and noisy

- identify patterns that enable us to improve processes and identify variables that affect production:

  industries are not sufficiently able to extract the value (sometimes hidden) “inside” data

- improve new digital services in the industrial sector:

  advanced manufacturing processes are increasingly complex and computerized

As a service:

- empower expert/ non expert usage of huge quantities of data
- enable real time or near real time data transfer/ data analysis / internal procedures update
- include new competences in data analysis/ update automatically (SMEs, partners)

THANK YOU

to I-BiDaaS Final Event
Attendees, Speakers and Partners

FORTH Institute of Computer Science
Barcelona Supercomputing Center Centro Nacional de Supercomputación
IBM
CRF
software AG
CaixaBank
MANCHESTER 1824 The University of Manchester
École des Ponts business school
AEGIS IT RESEARCH
Atos
ITML
Telefónica

Industrial-Driven Big Data as a Self-Service Solution

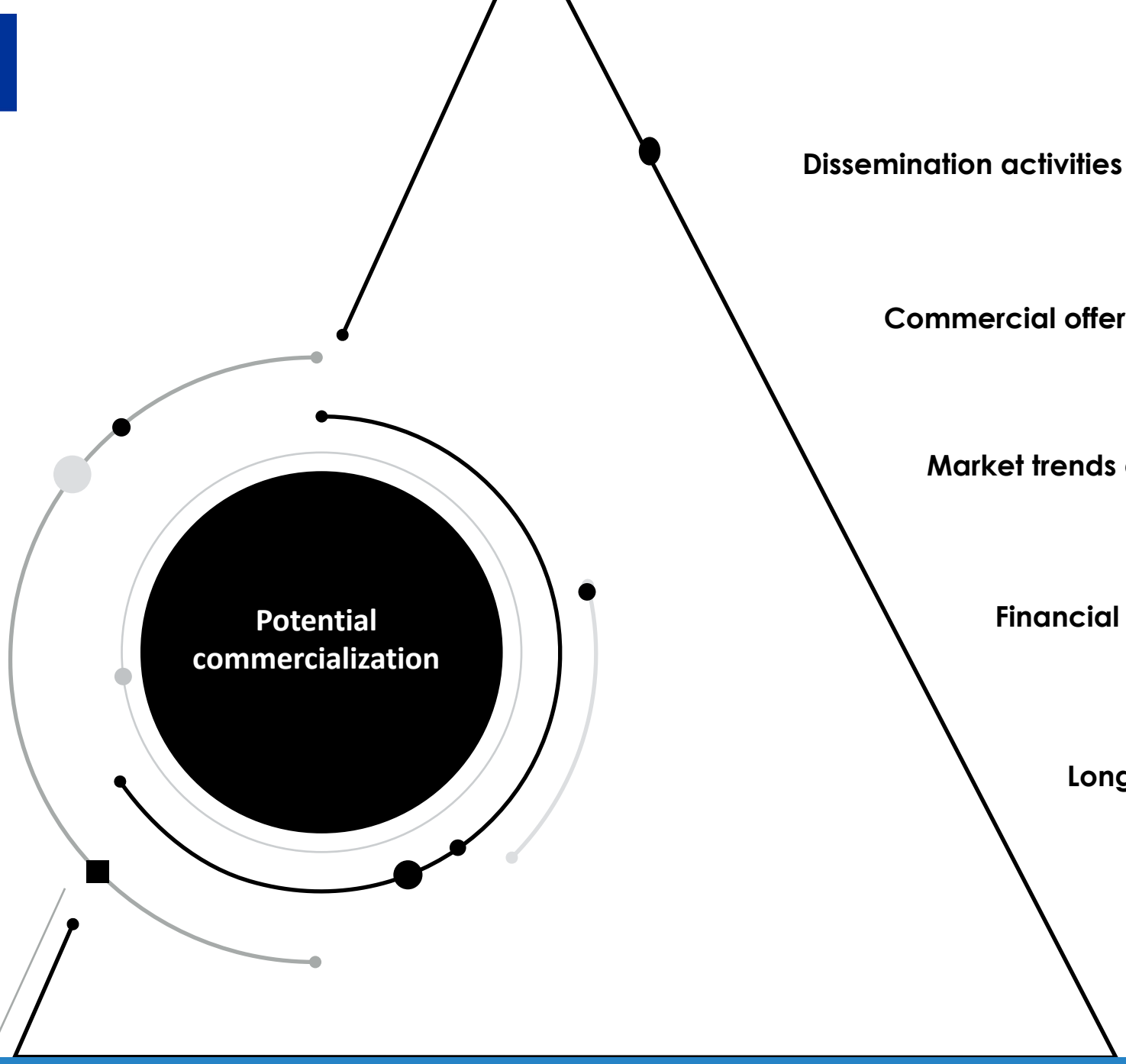
Commercial Offering

Hernan Ruiz Ocampo (ENPC)

I-BiDaaS Final Event

December 21, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 780787



Dissemination activities

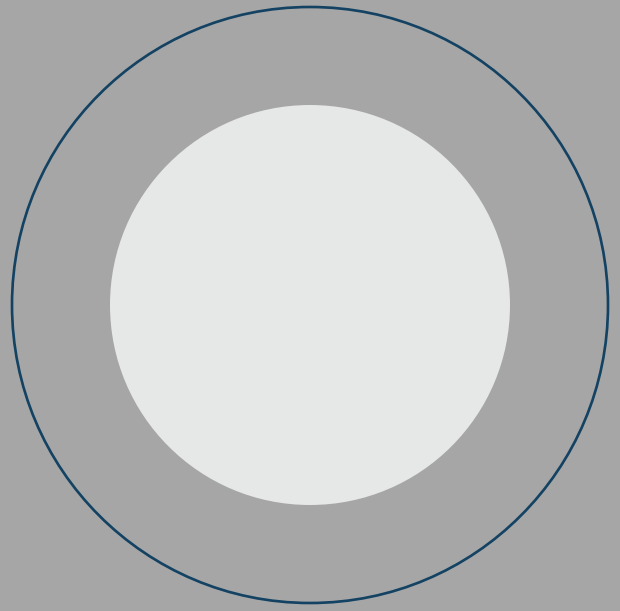
Commercial offering

Market trends and Business Modelling

Financial Forecast Plan

Long-term sustainability

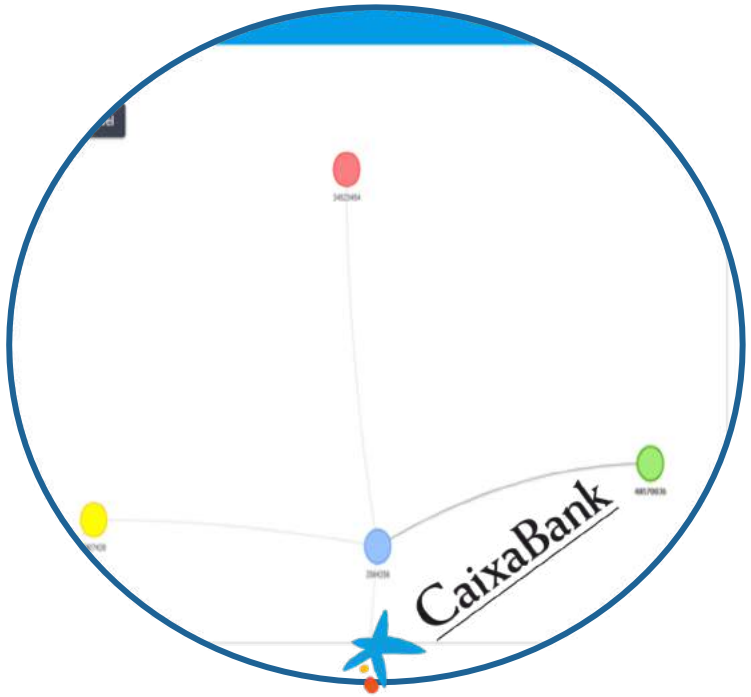
Dissemination activities

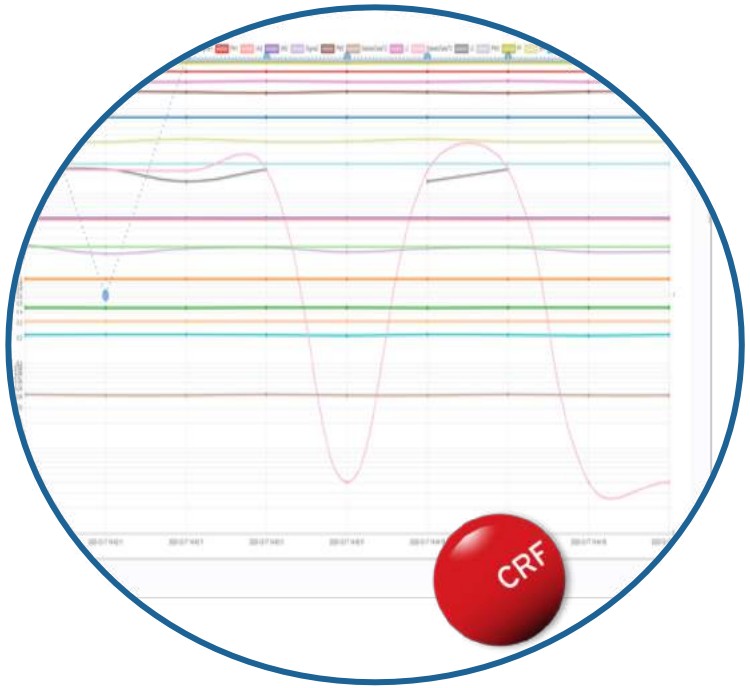


Commercial offering

We believe we are creating a unique product that allows you to

Incorporate BigData Analytics into your Business Decisions





Unleashing the **value of Big Data**

SWOT Analysis

STRENGTH

Flexible and modular platform

Empowers end users, improving the uptake of Big Data technologies

Unlocks data silos

Use of lean business models

THREATS

Competitors developing platforms for big data analytics

Lack of data valuation standards in marketplaces

Legal block to the free-flow of data and uncertainty around data policies

Restrictions in data sharing and data availability

WEAKNESS

Areas of application

Function differently across browsers

OPPORTUNITIES

Increase of global investments in data-driven innovation

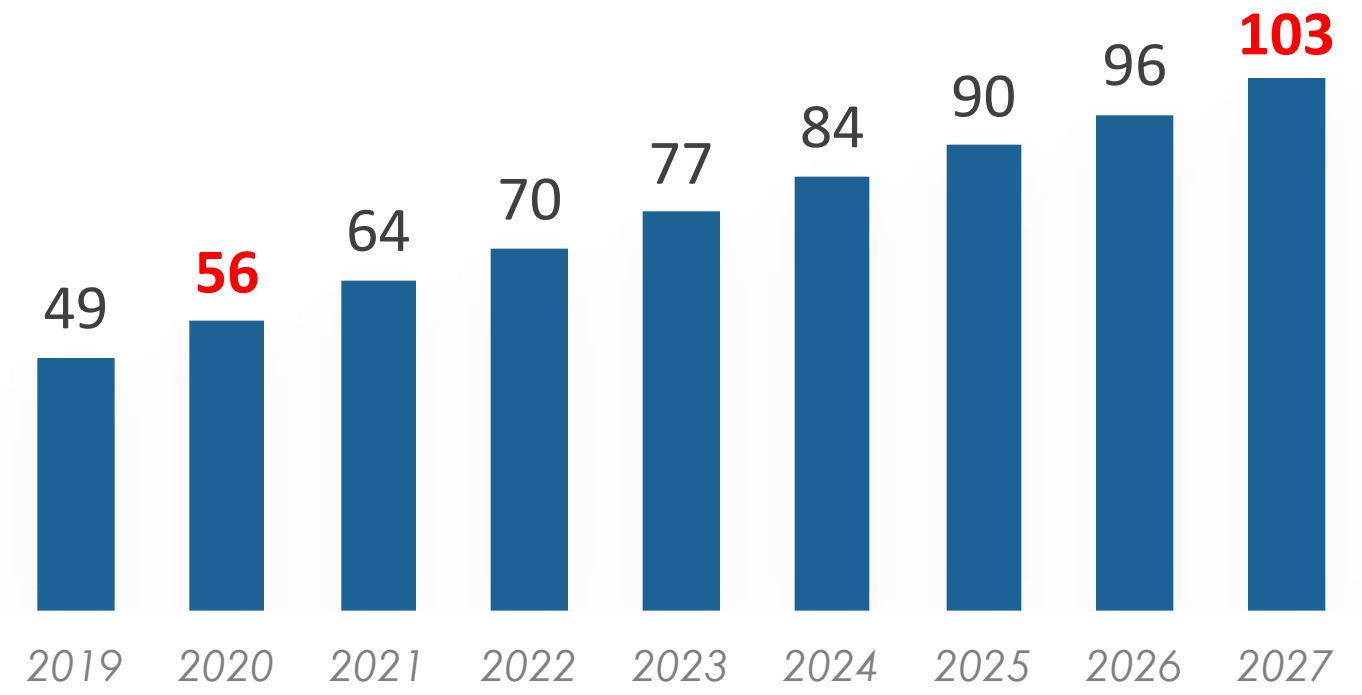
Growth in data volume and velocity

Widely use of big data

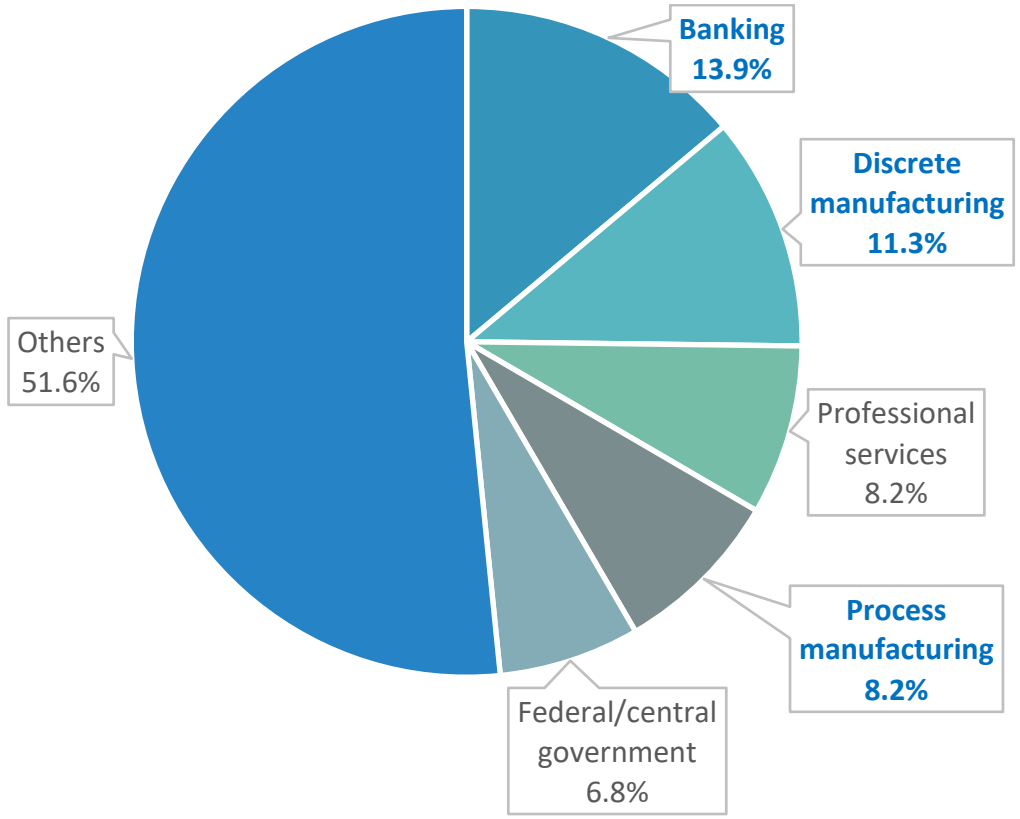
EU data strategy for creating a single market for data

Market trends are in our favor

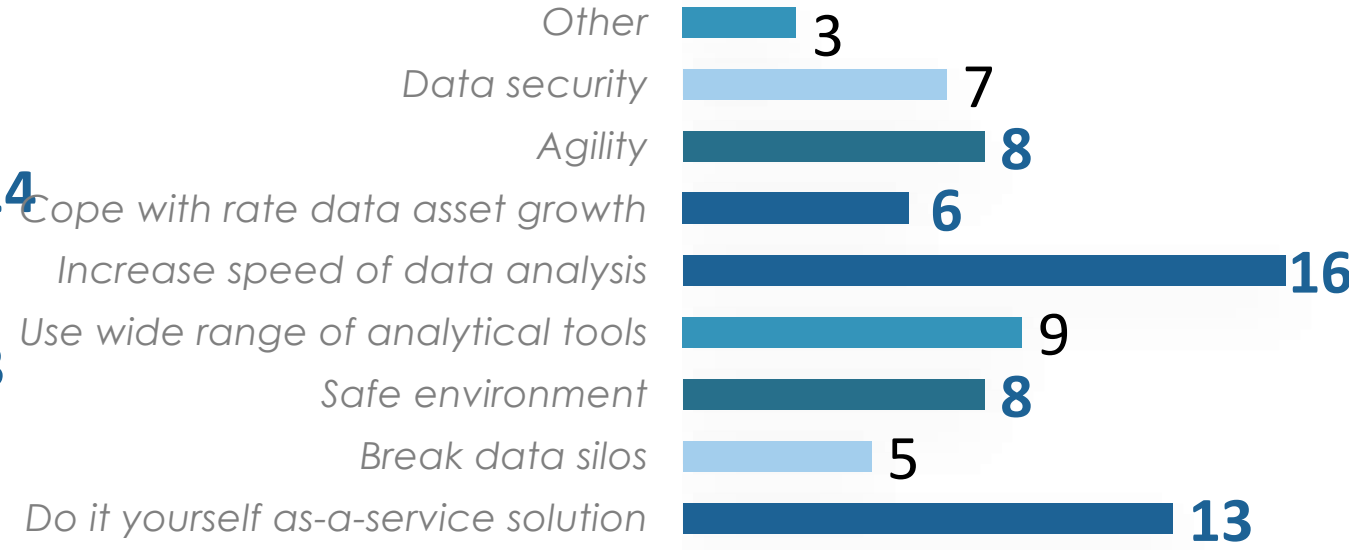
Forecast: Big Data size revenue worldwide from 2019 – 2027 (in billion U.S. dollars)

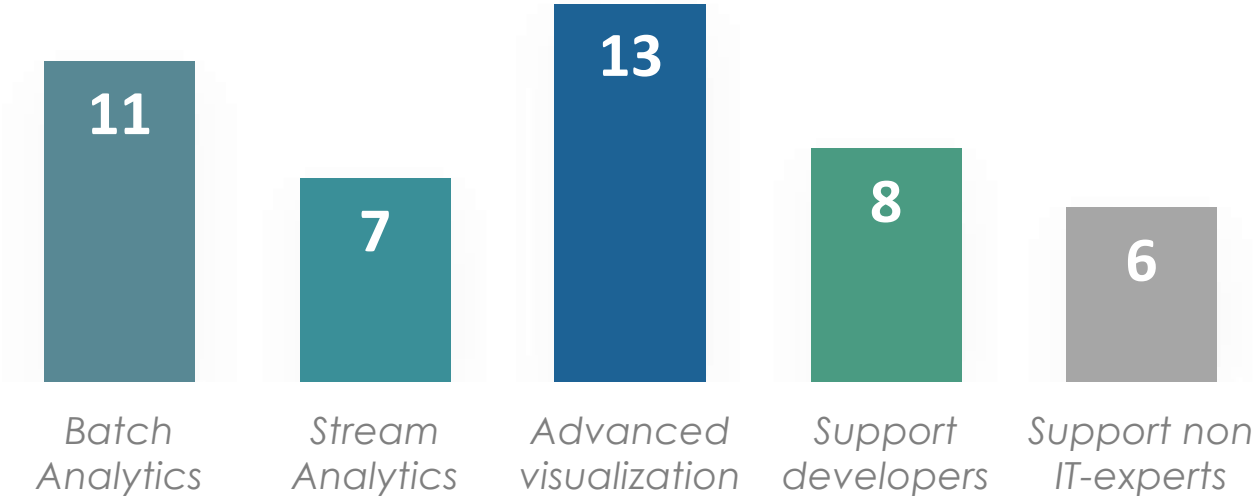


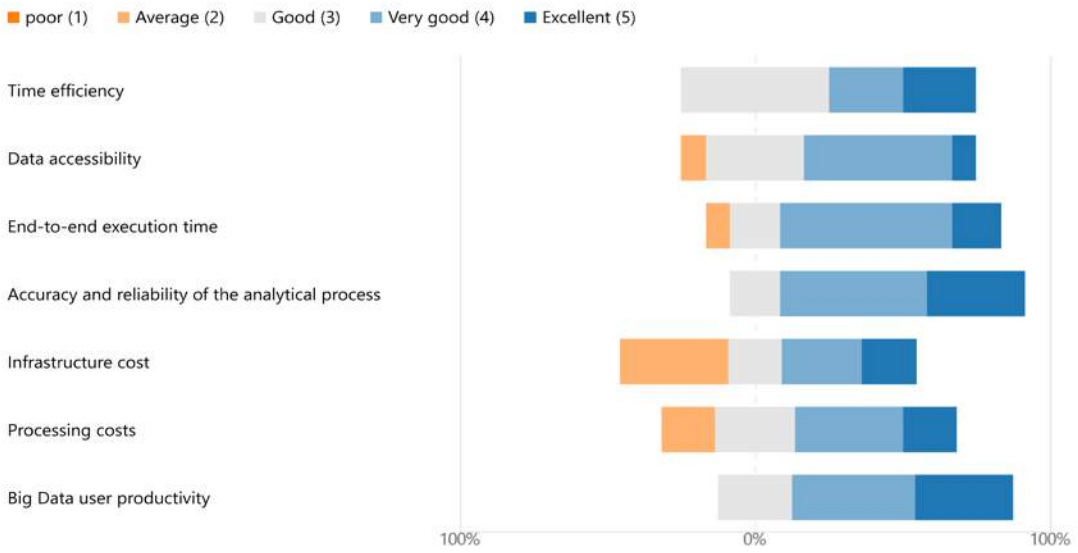
Share of Big Data and business Analytics revenue worldwide in 2019 by Industry



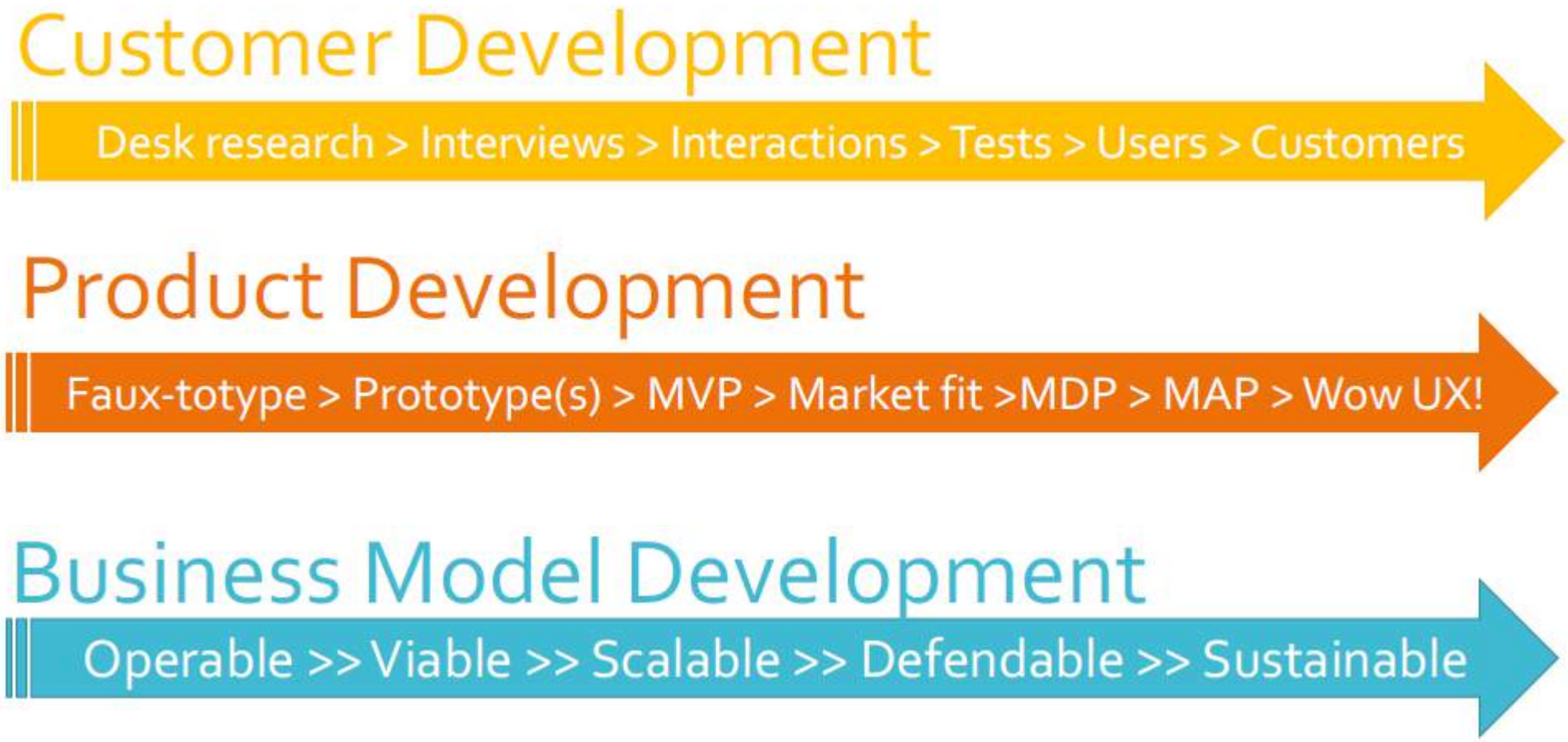
Our results: Feedback from potential users



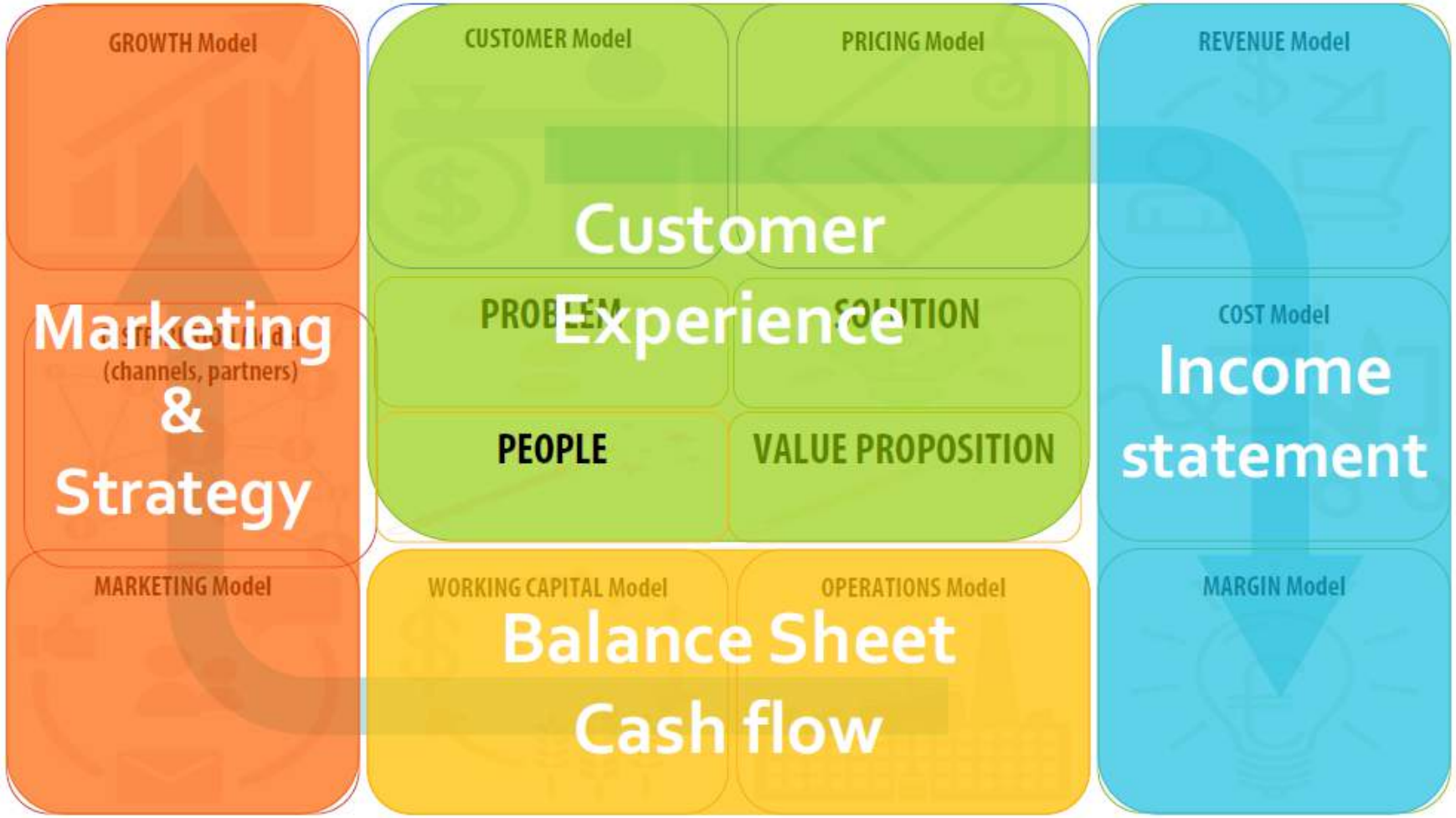




Lean innovation and agility



Dynamic Business Model: DBM



I-BiDaaS DBM

Business Plan for potential commercialization

GROWTH

Scale 10x: Industrial partners investment – partners provide technology

How you will finance growth: Raise of Equity

CUSTOMER

10 first pioneers:

Existing use cases: banking, telecom, automotive industries.

100 early adopters:

Similar industries: Financial institutions, manufacturing companies, utilities and service operators.

New industries: Healthcare, retail, construction

PRICING / PAYMENT

License fee and consultancy: At least 12 – 15 k€ per year – Subscription fee

REVENUE

Revenue model: Open-Source License - Subscription

Revenue streams: batch processing and stream analytics

- License fee
- Consultancy

Expected revenue 1st year: 89.6 k€

SALES

(distribution, channels, partners)

Find the Product: Open-source Apache 2.0

Sales force: Marketing and biz dev team (B2B; B2C)

Partners: I-BiDaaS consortium

SOLUTION

Big Data as-a-service to empower IT experts and non-expert users to easily take advantages of the Big-Data technologies while increasing the speed of data analytics.

VALUE PROP

Experimentation user friendly, modular, flexible, fast, and safe platform that handles significant volumes of sensitive data, and combine data sources.

COSTS

Main drivers:

- Development, maintenance, wages
- Fixed costs
- No changes with scale

Expected cost 1st year: 350 k€

PROBLEM

Companies with significant amount of diverse, distributed, noisy, incomplete, intermittent data, making difficult the extraction of value for non-expert users.

PEOPLE & SKILLS

I-BiDaaS team

MARKETING

- **Industrial Partners Success Stories**
- **Social media (Posts)**
- **Emails**
- **Website**
- **Partners website**
- **Participation in Trade Fairs**

CASH FLOW

	2021	2022	2023	2024
OPEX (€)	-190 562	200 396	732 717	1 364 685
CAPEX (€)	74 800	124 800	174 800	209 800
Financing (€)	309 180	-	-	-

OPERATIONS

What will you do:

- Dev. and maintenance use cases
- Marketing, Biz dev

What will you outsource:

License: Apama; Usage: Qbeast

Operation resources:

Capital: Raise of equity

I-BiDaaS technologies

MARGINS

	2021	2022	2023	2024
Gross margin	-112,3%	68,1%	83,3%	88,4%
EBIT (€)	-239 300	204 800	860 200	1 635 050

Medium margin

Value Proposition

“Improve decision making whenever you want”

User Friendly

- IT expert & non-expert

Save Time

- Increasing the speed of data Analytics

Save money

- Less expensive data analytics

Upskill your team

- Empower employees with crucial knowledge

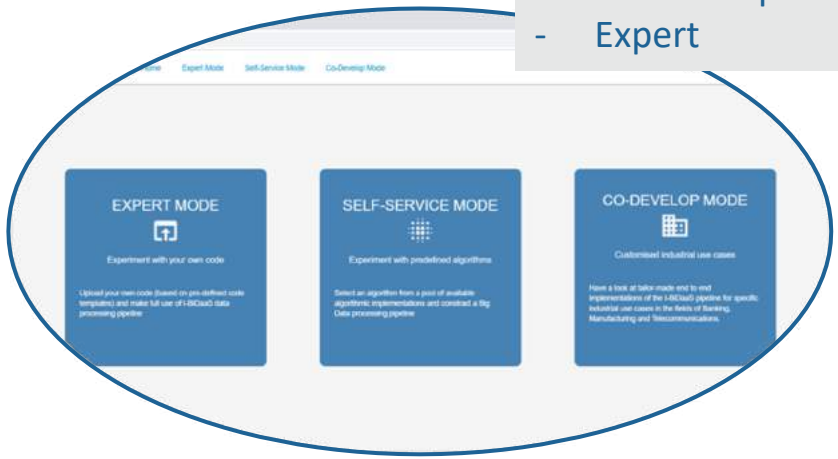
Our business model

- Cloud
- On-premises

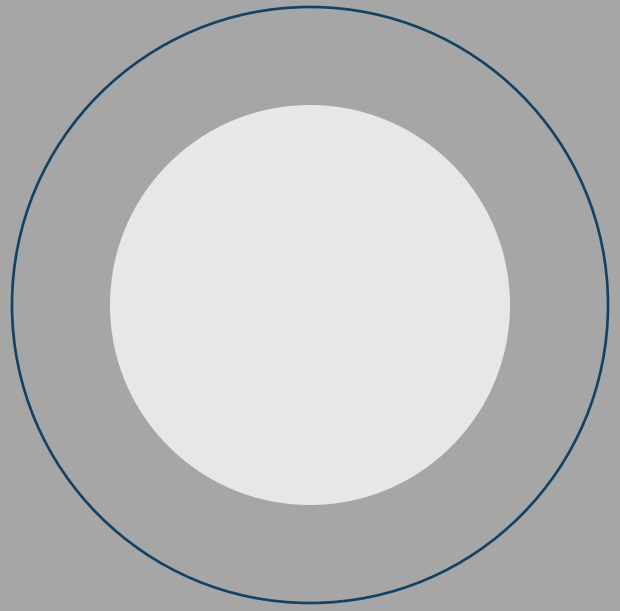
- IT and non-IT experts
- Batch and stream Analytics 3 modes:
  - Self-service
  - Co-develop
  - Expert

License fee
- Subscription: Annual, Monthly

Consultancy services
- Tailor made solutions



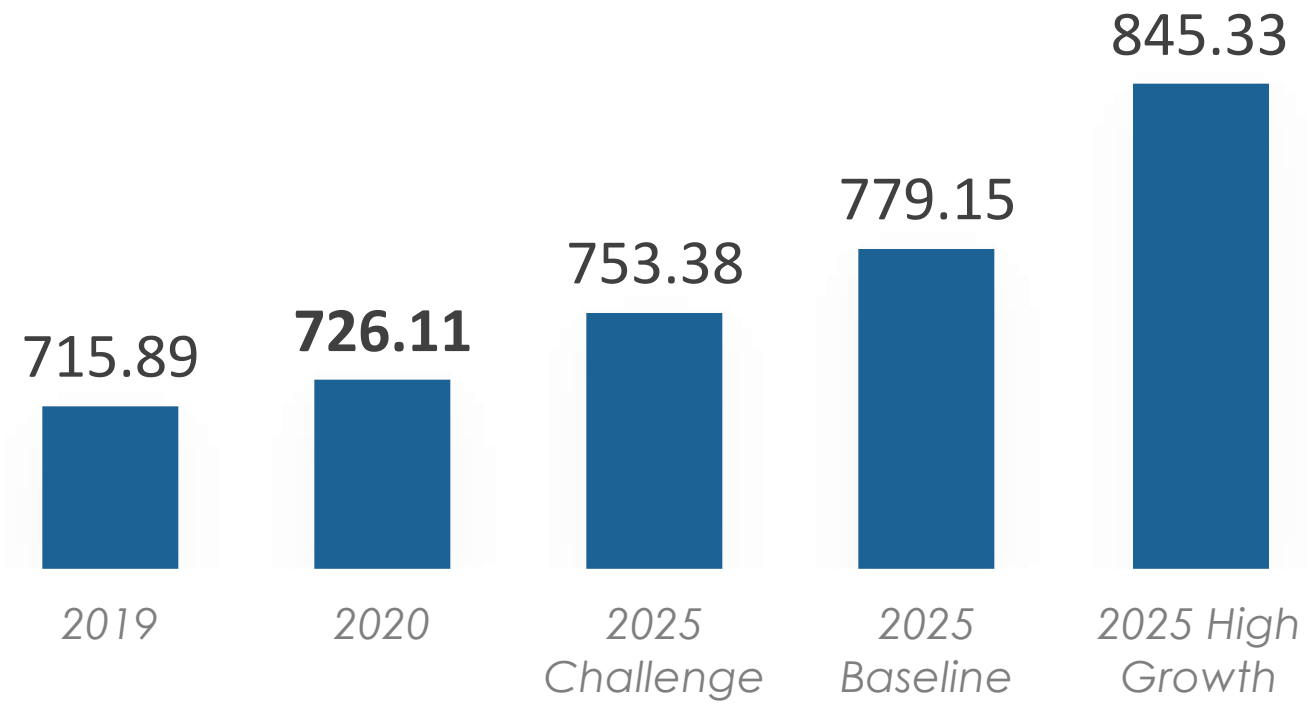
Platform
- Select algorithms
- Create business insights
- Visualizations



Financial Forecast (4-Y)

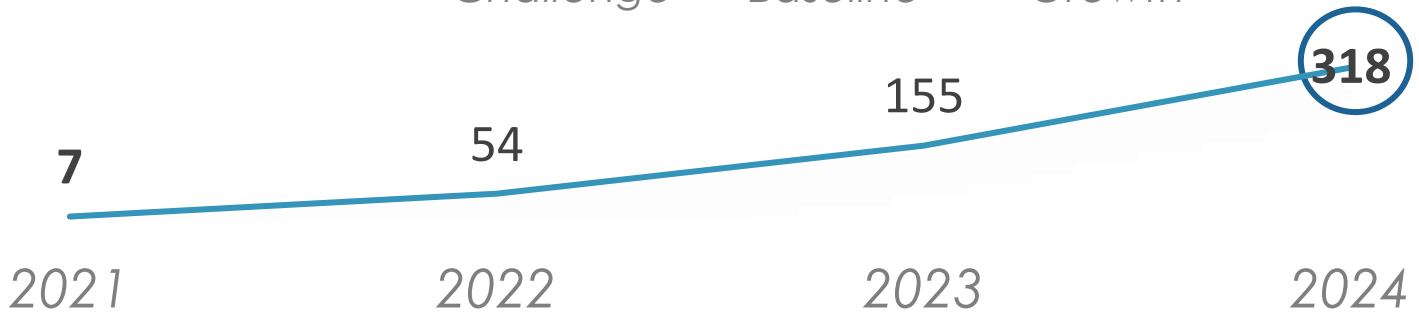
Data Market

Number of company data users in the European Union (EU) and United Kingdom (UK) Forecast 2025 (in thousands)



2020 – 2025: 27 000 - 119 000 companies

Forecast 4-Y



Market share: O,3 – 1,2%

Our customers are clearly defined

Customer Acquisition

2021
1st phase

LAUNCH STAGE (Innovators)

Implementation use cases industrial partners

- Banking
- Manufacturing
- Telecommunication

2022 - 2023
2nd phase

GROWTH STAGE (Innovators – Early adopters)

Similar industries

- Financial institutions
- Fabricated metal industry
- Industrial & commercial machinery
- Services & Infrastructure operators

2024
3rd phase

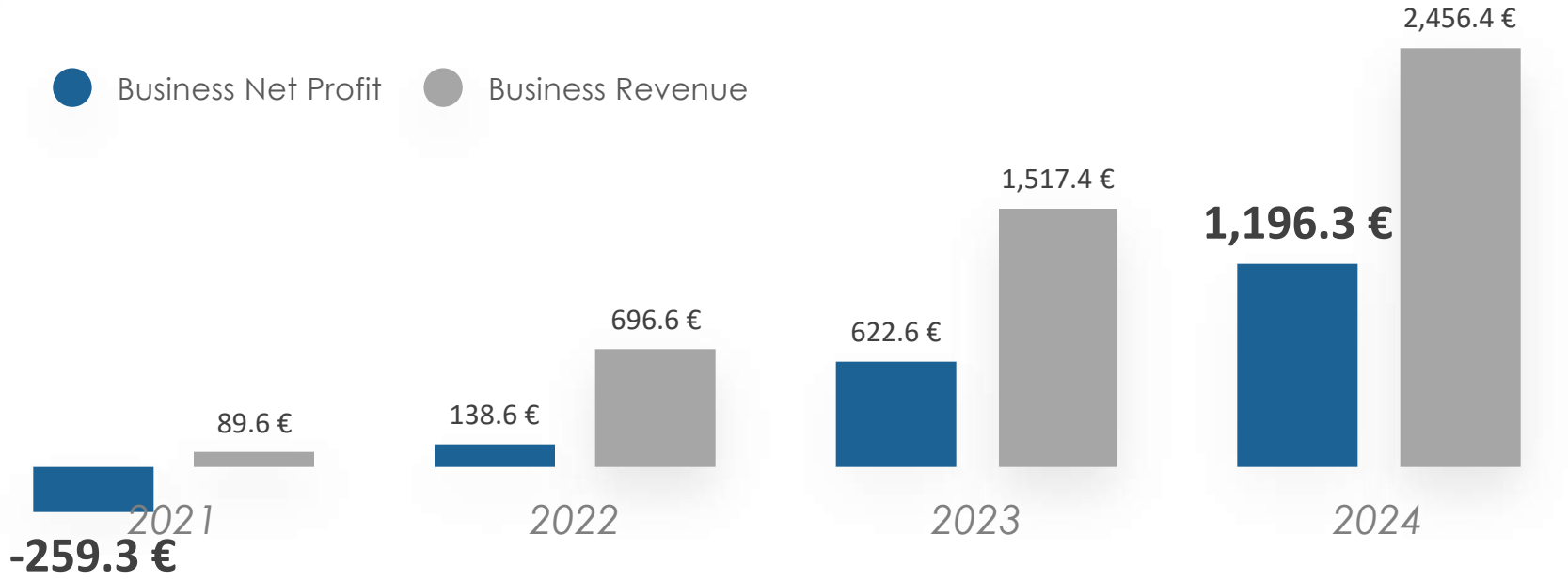
MATURITY STAGE (Early adopters)

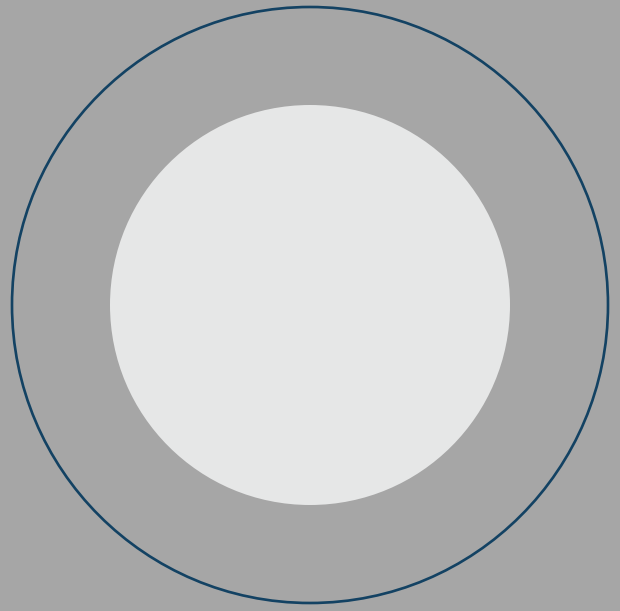
New Industries

- HealthCare
- Retail

Financial Projection (expressed in thousands)

According to our conservative financial projection, our revenue as well as net profit will steadily increase over the next four years.



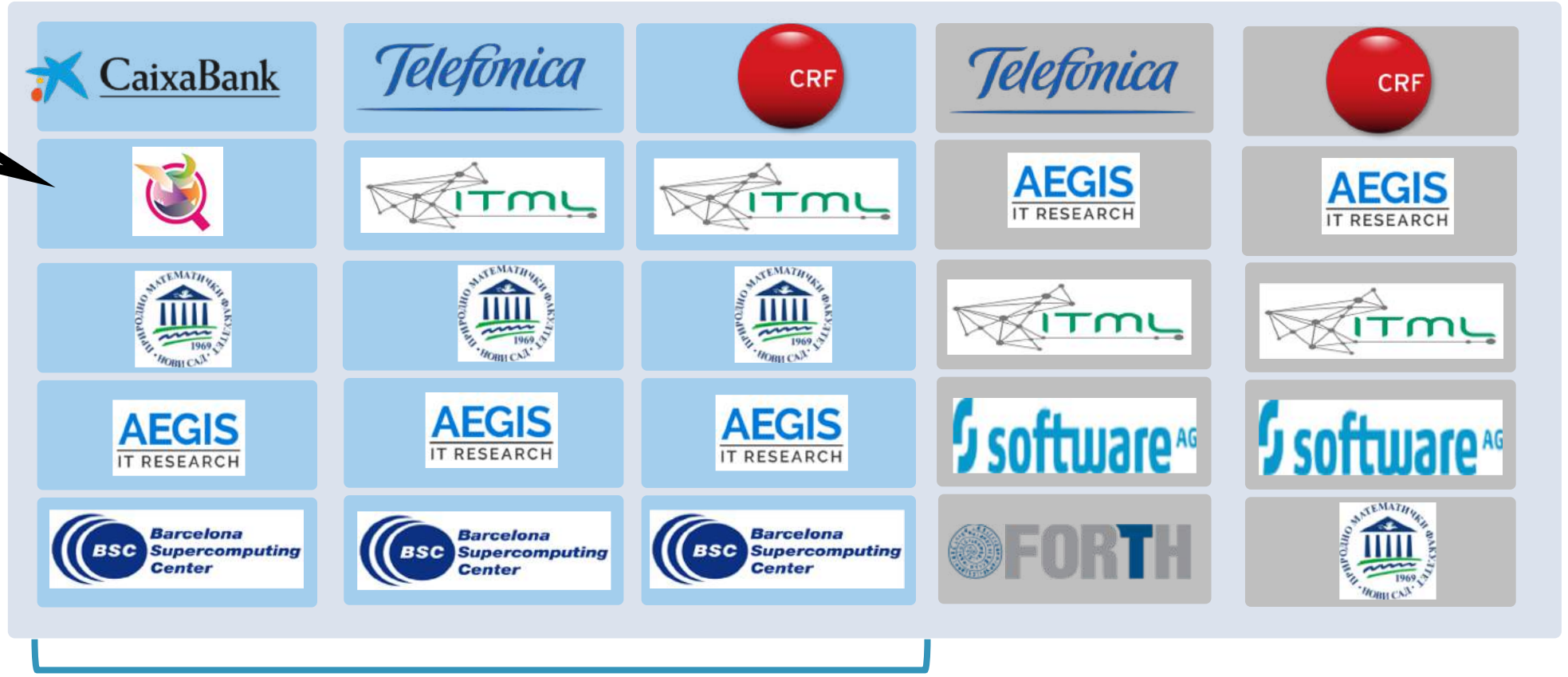


The Path Forward

The path forward: Long-term sustainability

Identification of Synergies

Qbeast (e.g.) evolution of a technology → patented



Qbeast – Big Data free from the unnecessary

Why waste time and money computing data that doesn't make a difference

Get the **results you need,** accessing only a **fraction of the data.**

- Patented technology from the Barcelona Supercomputing Center
- Storage accelerator based on our proprietary **multidimensional indexing and sampling** technology

for

Data Scientists struggling with slow and complex Big Data platforms.

Data Engineers dissatisfied with the complexity of building and maintaining analytics products

3x

more productive data teams

100x

Up to 100X faster data analysis

Value

Advanced ML tools for the whole organization

Cost Savings

Pay only for the data analyzed

Early Adopters

Enterprise Spark users

https://qbeast.io/

Pricing

Qbeast Storage

8€ per Terabyte **accessed**

The path forward: Banking sector

Caixa Garage Lab

Description

- Information Security Garage Lab during the first semester of 2021.
- This infrastructure is a sandbox especially designed to evaluate innovative tools

Strategy

1st

Access I-BiDaaS & extract information from the three use cases.

2nd

Deploy I-BiDaaS in Caixa Garage Lab to evaluate the tool.

3rd

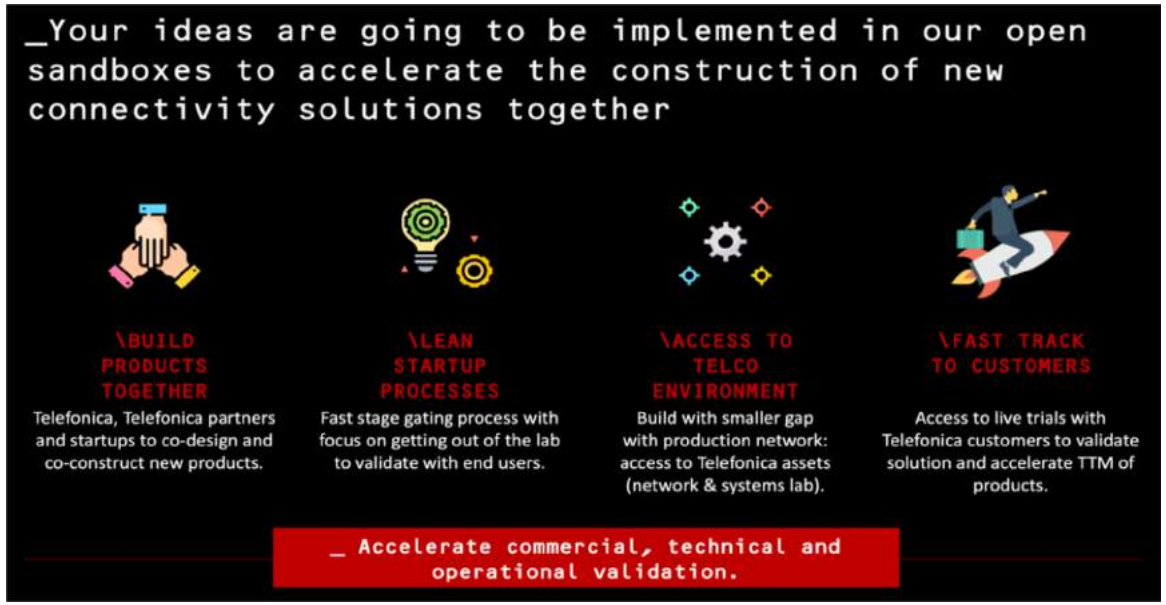
Potential achievements: Integrate I-BiDaaS inside Caixa & experiment with different use cases.

The path forward: Telecommunication sector

Telefonica Internal Innovation Call

Description

Potential achievements



The path forward: Manufacturing sector

CRF Training environment

Description

- Training sessions with end-users in 2021.
- Provide Big Data culture

Strategy

1st

Presentation use cases, type of data collected and how data have been managed to perform the analytics

2nd

To show and explain step-by-step the I-BiDaaS platform's Interactive session to test the different features with the employees

3rd

Integrate I-BiDaaS inside CRF premises & implement Circular Business Model to extend life cycle of products

That ultimately drives long term value creation

Circular Business Models: Product life-extension (CRF)

Product life extension enables companies to extend the lifecycle of products and assets so value that would otherwise be lost through wasted materials is instead maintained or even improved

Overall Equipment Effectiveness (OEE)

Increase of 1-1.5 % in current Overall Equipment Effectiveness (OEE)

Cost reduction

Decrease of 20-30 % in maintenance costs

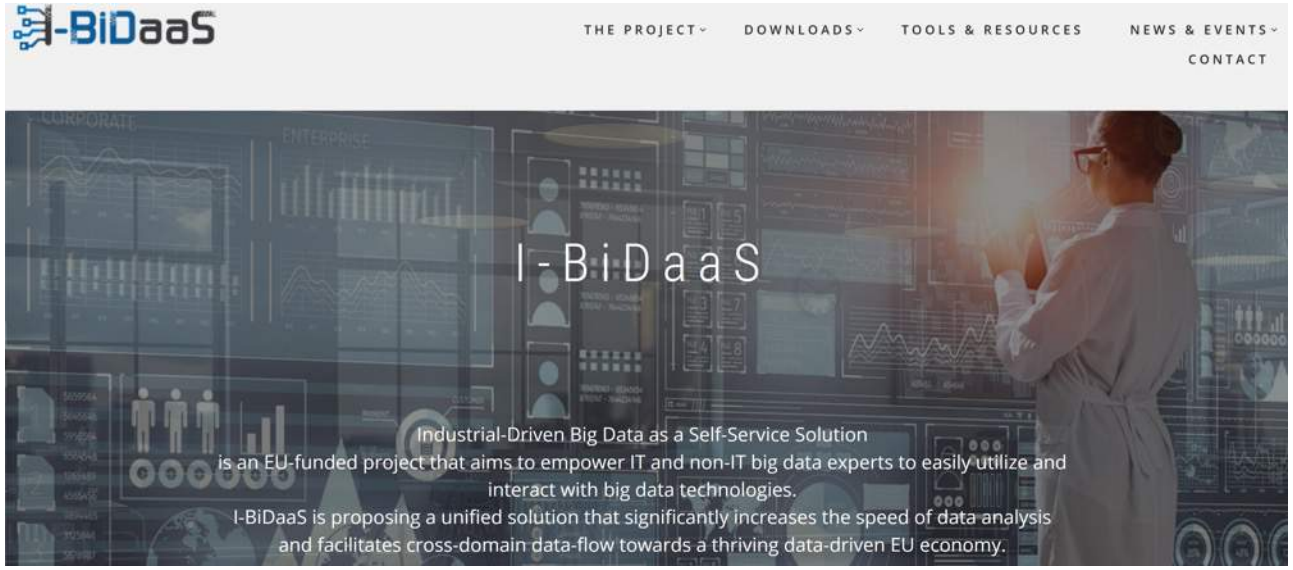
Job-per-Hour (JPH)

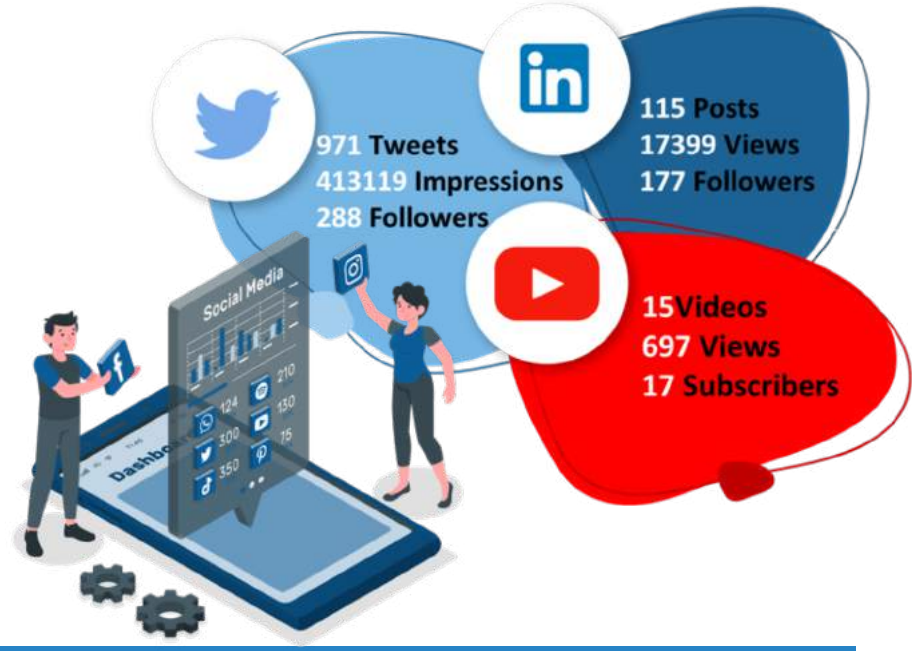
Assess the process performance & measures the throughput of the line per hour.

Quality levels

Increase of 3-7 % good products
Decrease of 1-4% and 0.5-2 % defective products

Contact us





https://www.ibidaas.eu/

Email: sotiris@ics.forth.gr
dkopanaki@ics.forth.gr
h.ruizocampo@pontsbschool.com

I-BiDaaS

Industrial-Driven Big Data as a Self-Service Solution

Thanks for your attention