Data Innovation Spaces

Daniel Alonso

Abstract Within the European Big Data Ecosystem, cross-organisational and cross-sectorial experimentation and innovation environments play a central role. European Innovation Spaces (or i-Spaces for short) are the main elements to ensure that research on big data value technologies and novel applications can be quickly tested, piloted and exploited for the benefit of all stakeholders. In particular, i-Spaces enable stakeholders to develop new businesses facilitated by advanced Big Data Value (BDV) technologies, applications and business models, bringing together all blocks, actors and functionalities expected to provide IT infrastructure, support and assistance, data protection, privacy and governance, community building and linkages with other innovation spaces, as well as incubation and accelerator services. Thereby, i-Spaces contribute to building a community, providing a catalyst for engagement and acting as incubators and accelerators of data-driven innovation, with cross-border collaborations as a key aspect to fully unleash the potential of data to support the uptake of European AI and related technologies.

Keywords Data-driven innovation (DDI) · Data experimentation environment · Data space · Data platform · Data sharing · Data ecosystems and community · Digital Innovation Hub (DIH) · Federation of DIHs

1 Introduction

The term Data Innovation Space (in short i-Space) was initially coined by the Big Data Value Association (BDVA) and included in the first version of its Strategic Research and Innovation Agenda (SRIA) (Zillner et al. 2017) as one of the mechanisms identified to implement its research and innovation strategy, together with (i) lighthouse projects (large-scale demonstrators aimed to showcase the applications

---

D. Alonso (✉)
ITI, Valencia, Spain
e-mail: dalonso@iti.es

of data-driven solutions to different sectors), (ii) technical projects (addressing specific data issues and technical aspects) and (iii) cooperation and coordination projects (to enable international cooperation for efficient information exchange and coordination of activities).

This chapter presents Data Innovation Spaces as environments to test, experiment and deploy new data-driven innovations. More specifically, Sect. 2 introduces the concept of Data Innovation Spaces and their main characteristics. The key elements of Data Innovation Spaces, as well as basic expected services, are presented in Sect. 3. Section 4 presents the role of i-Spaces in the European landscape and their alignment with other initiatives. Section 5 explains the specific certification process implemented by the Big Data Value Association (BDVA) to recognise relevant initiatives in Europe. The impact of the BDVA-recognised i-Spaces in their respective ecosystems is presented in Sect. 6. General collaboration between Data Innovation Spaces and a specific example of creating a European federation are explained in Sect. 7. Finally, the chapter ends with learnt stories and success stories as part of Sect. 8.

2 Introduction to the European Data Innovation Spaces

European Data Innovation Spaces are the main elements to ensure that research on BDV technologies and novel BDV applications can be quickly tested, piloted and thus exploited in a context with the maximum involvement of all the stakeholders of BDV ecosystems. The objective is to facilitate large and small companies, public administration, and European and national projects and society, in general, in easily accessing economic opportunities offered by the BDV and developing working prototypes to test the viability of actual business deployments. As such, i-Spaces enable stakeholders to develop new businesses facilitated by advanced BDV technologies, applications and business models. i-Spaces bring together not only technical and application developments but also all aspects needed to foster skills, competencies and best practices. i-Spaces usually rely on national and regional initiatives, federating, complementing and leveraging activities of similar national incubators/environments, existing Public—Private Partnerships and other national or European initiatives.

The main characteristics of a Data Innovation Space are as follows (as shown in Fig. 1):

- Forming **hubs** to bring technology and application developments together while catering for the development of skills, competencies and best practices. These environments provide new and existing technologies and tools from industry and open-source software initiatives as a basic service to tackle the big data value challenges.
- Ensuring that **data** is at the centre of big data value activities. i-Spaces make data assets based on industrial, private and open data sources accessible. They are

Fig. 1 Data Innovation Spaces concept

secure and safe environments that ensure the availability, integrity and confidentiality of data sources.

- Serving as **incubators** for the testing and benchmarking of technologies, applications and business models. This provides early insights into potential issues and helps to avoid failure in the later stages of commercial deployments. In addition, it is expected that this activity will provide input for standardisation and regulation.
- Developing **skills** and sharing best practices is an important task of i-Spaces and their federation. They will also link with other existing initiatives at both the European and national level.
- New **business models** and ecosystems will emerge from exposing new technologies and tools to industrial and open data. i-Spaces are a playground for testing new business model concepts and the emerging ecosystems of existing and new BDV “players”.
- Gaining early insights into the **social impact** of new technologies and data-driven applications and how they will change the behaviour of individuals and the characteristics of data ecosystems.
- Acting as a **catalyst** to foster data-driven communities in the ecosystem and accelerate value creation.

The establishment of European Data Innovation Spaces and their evolution is reflected in the roadmap of the implementation of the Big Data Value Public-Private Partnership (BDV PPP), as detailed in Chap. “A Roadmap to Drive Adoption of Data Ecosystems”. Phase 1 of this roadmap (2016–2017) is devoted to the establishment of the ecosystem (including i-Spaces and their collaboration towards a federation or network of i-Spaces), phase 2 (2018–2019) proposed disruptive forms of big data solutions, and phase 3 (2020) considers the sustainability and the benefits of the carried-out actions.

3 Key Elements of an i-Space

As mentioned in the previous section, i-Spaces are conceived as interdisciplinary hubs to target BDV challenges encountered by SMEs and small regional actors in the following different dimensions (see Fig. 2).

- **Technical**, providing infrastructure for testing, giving advice on architecture and security of the workspace and tool implementation, and offering help-desk support
- **Application**, supporting the building of precompetitive application, developing (visual) analytics tools and settings for specific domains
- **Business**, creating new data-driven business models, identifying new business opportunities with already existing data, and providing proof of impact and ROI
- **Social**, supporting SME uptake in digitisation, offering services for cultural heritage and local governments, and providing digital solutions for policy development

Fig. 2 Dimensions of i-Spaces (BDVA SRIA)

- **Skills**, training and educating employees to make use of big data technologies and build on data expertise, and providing master's-level students with industrial problems and specific data
- **Legal**, providing templates for data provision contracts and consulting on internal secure data management process and architecture

In terms of services, i-Spaces are supposed to provide to SMEs and industry, society and other European initiatives (including projects) a set of basic tools to allow the demonstration, experimentation and training, testing, showcasing and benchmarking of their data-driven solutions and products, before going to the market. This set of basic services includes:

- **Community Building:** Contributing to the identification and management of stakeholder ecosystem communities along thematic and/or regional dimensions.
- **Asset Support:** Supporting data providers in integrating datasets in a quality-secured way while maintaining a catalogue of available data assets.
- **ICT Support:** Providing basic ICT assistance as well as focused support from big data scientists and data specialists, and business development during research and innovation projects. This includes assistance in benchmarking datasets, technologies, applications, services and business models.
- **On-boarding:** Running an induction process for new project teams.
- **Resourcing:** Allocating the resources (computing, storage, networking, tools and applications) to individual research and innovation projects and scheduling these resources among different projects.
- **Data Protection and Privacy:** Data protection, including ensuring compliance with laws and regulations such as the EU GDPR (General Data Protection Regulation), and the deployment of cutting-edge, state-of-the-art security technologies in protecting data and controlling data access, privacy and anonymisation in terms of handling and deleting personally identifiable information (PII).
- **Data Governance:** Taking into account privacy and protection issues, defining the rules for accessing and sharing data. This includes the standardisation of procedures for sharing metadata, defining the (smart) contract between stakeholders, assessing technologies such as encryption and blockchain, and formulating the necessary solutions to orchestrate the agreed governance.
- **Federation:** Supporting linkages to other innovation spaces and facilitating experiments across multiple innovation spaces. An effective federation will help to support research and innovation activities through accessing and processing data assets across national borders (data spaces).
- **Business Support:** Facilitating start-ups and SME inclusion in the value creation process by leveraging community engagement.
- **Incubation and Acceleration:** Delivering all forms of suitable support to data-driven value creation projects by liaising with existing thematic, national or regional initiatives.

Fig. 3 Competence Centres and Digital Innovation Hubs (Source: European Commission) (by European Commission licensed under CC BY 4.0)

4 Role of an i-Space and its Alignment with Other Initiatives

As mentioned above, the concept of Data Innovation Space was initially coined in 2014 by the BDVA and identified as a key instrument to foster data-driven innovation based on experimentation, testing and benchmarking. Since then, many other instruments have appeared in Europe, aimed at bringing innovation closer to industry and society, and more specifically to those actors with no capacity to benefit from the latest European digital innovations.

In this way, and considering that only about 1 out of 5 companies across the EU is highly digitalised, and around 60% of large industries and more than 90% of SMEs lag in digital innovation, the European Commission introduced in 2017 the concept of the Digital Innovation Hub (DIH),¹ to ensure that every company, small or large, high-tech or not, can take advantage of digital opportunities. DIHs are one-stop shops that help companies become more competitive with regard to their business/production processes, products or services using digital technologies. DIHs provide access to technical expertise and experimentation so that companies can “test before invest”. They also provide innovation services, such as financing advice, training and skills development, that are needed for a successful digital transformation.

A Digital Innovation Hub brings many actors together, to develop a coherent and coordinated set of services that are needed to help companies (especially SMEs or enterprises from low-tech sectors) that have difficulties with their digitisation through a one-stop shop. However, the core of a DIH is the Competence Centre, which provides technical expertise and access to advanced facilities (see Fig. 3).

¹https://ec.europa.eu/futurium/en/system/files/ged/dei_working_group1_report_june2017_0.pdf

Fig. 4 Data Innovation Space vs. DIH and Competence Centre

The European Commission has developed an online catalogue² to provide a comprehensive picture of DIHs in the EU across varying competences structures and service offerings. It is a repository with more than 400 DIHs, over 200 of which are fully operational, including information on the technology and application specialisation, geographical coverage, markets addressed and general digitisation support available. According to this catalogue, there are around 190 DIHs in Europe specialised in data mining, big data and database management, meaning that these data-driven DIHs are ready, based on the expertise provided by their Competence Centres, to support companies in their respective ecosystems in the development, adoption and testing of data-driven solutions.

In this way, the concept of Data Innovation Space is aligned with that of a Competence Centre on Big Data, in the sense that it provides access to infrastructure, expertise, support to experimentation and production of new services, and best practices regarding data-driven solutions and products. On the other hand, it can also offer advanced services such as brokerage, access to finance, training, and incubation and acceleration. In this case, it would act as a Data-Driven Innovation Hub (actually, all BDVA i-Spaces are recognised DIHs on big data), bringing together not only technical competencies but all tools and aspects needed to allow SMEs to put their data-driven services and products into the market. Taking all of the above into consideration, and depending on the offered services, a Data Innovation Space would range between a Competence Centre on Big Data and a Data-Driven Innovation Hub (see Fig. 4).

Other important instruments developed to mobilise data and foster data sharing and reuse are data platforms and data spaces. According to a BDVA position paper on data sharing and data spaces,³ a data space is an ecosystem of data models, datasets, ontologies, data sharing contracts and specialised management services (e.g. as often provided by data centres, stores and repositories, individually or within “data lakes”), together with soft competencies around it (i.e. governance, social interactions, business processes). These competencies follow a data engineering approach to optimise data storage and exchange mechanisms, in this way preserving, generating and sharing new knowledge. On the other hand, data platforms refer to architectures and repositories of interoperable hardware/software components, which follow a software engineering approach to enable the creation, transformation, evolution, curation and exploitation of both static and dynamic data in data spaces.

²<https://s3platform.jrc.ec.europa.eu/digital-innovation-hubs-catalogue>

³<https://www.bdva.eu/node/1277>

Specific examples of data space and data platforms are mentioned in this BDVA paper, and it is also worth mentioning the nine innovation actions funded by the European Commission under the topic “Supporting the emergence of data markets and the data economy”, especially aimed to address the necessary technical, organisational, legal and commercial aspects of data sharing/brokerage/trading, both for personal and industrial data.

These instruments incorporate in Data Innovation Spaces (and Data-Driven Innovation Hubs) the dimension of data sharing, data trading and data reuse, allowing Data Innovation Spaces to share datasets and data sources with other Data Innovation Spaces, and providing interoperability and scalability in terms of data.

The new Digital Europe Programme will reinforce the role of Digital Innovation Hubs and European Data Spaces as the main instruments to increase the competencies and bring innovation to the European industry and society in terms of data. This programme also includes technology infrastructures with specific expertise and experience of testing mature technology in a given sector, under real or close to real conditions (e.g. smart hospital, smart city, experimental farm, corridor for connected and automated driving), which are the Testing and Experimentation Facilities (TEFs) on AI.

These TEFs will exploit, test and validate data spaces to test AI-powered solutions, also enriching them by providing user feedback. TEFs will contribute to data spaces by collecting and providing data from experimentation. On the other hand, the Digital Innovation Hubs will act as a distribution channel for AI to empower all local companies and users.

Figure 5 shows the different dimensions provided by different European instruments.

According to the European Commission, a Digital Innovation Hub relies on four pillars to increase the competitiveness of companies with regard to their business/production processes, products or services using digital technologies. These pillars are: (i) access to an innovation ecosystem with connection and networking with multiple stakeholders, (ii) test before invest, with access to technical expertise and experimentation, (iii) support to find investments and (iv) skills and trainings. With respect to this last aspect, to find alignments and synergies with the so-called centres of excellence, organisational units within a national system of research and education that provides leadership in research, innovation and training in digital technologies are of utmost importance, given the regional/national scope of both types of initiatives and their complementarities. In the case of big data, the connection between Data-Driven Innovation Hubs and the network of Big Data Centres of Excellence is valuable in identifying gaps in the industry demand side (workforce) at regional level and jointly planning a training programme to fill those gaps. Further details on big data and AI Centres of Excellence are available in Chap. “A Best Practice Framework for Centres of Excellence in Big Data and Artificial Intelligence”.

Fig. 5 European instruments to foster data-driven innovation and experimentation

5 BDVA i-Spaces Certification Process

With the objective of identifying relevant and qualified initiatives in Europe aligned with the concept of i-Spaces, the BDVA launches yearly public calls that are open to any innovation hub on big data⁴ in Europe. The candidates are evaluated in terms of infrastructure and technologies provided, the services that are offered, projects and applications where the DIH is involved, the impact on the local/regional and national/European ecosystem, and the business strategy and sustainability. After the review process, those initiatives that meet specific criteria are qualified as BDVA i-Spaces. This call has been launched over the last 5 years, and during the several editions, new i-Spaces have been incorporated, composing the current group of 15 BDVA i-Spaces (see Fig. 6).

- The different steps of the labelling process are as follows:
- Launch of the open call, aimed at any data-driven competence centre, DIH on big data and AI, etc. in Europe, interested in having the recognition of BDVA as a qualified Data Innovation Space. This recognition guarantees that the innovation environments provided meet the requirements to boost data-driven and AI-based innovation at a local level, and the collaboration with similar initiatives to foster adoption at European level.

⁴<http://www.bdva.eu/node/1173>

Fig. 6 Map of recognised BDVA i-Spaces 2019

- Online survey/questionnaire. Candidates are invited to fill in an online questionnaire, to collect information from their initiatives in the following domains:
  - Infrastructure, including computing, storage and communication capacities, allocation of resources, data access methods and tools, policies, standards and certificates
  - Services, including technical support, data management, analysis and visualisation, data governance, privacy and protection, incubation and acceleration, business support, skills and training
  - Projects and sectors, including most relevant projects and aggregated number of experiments per year
  - Ecosystem and collaborations, including actors engaged in the ecosystem, involvement in regional clusters, outreach and collaborations
  - Business strategy, including growth, impact and sustainability models
- Review (including review committee meeting). Received applications are reviewed by a review committee composed of external experts also recruited through an open call. Each of the five domains of the applications is scored

Fig. 7 i-Spaces labelling criteria

between 1 and 5. Final results are agreed in a review committee meeting. Applications are granted either a gold, silver or bronze label according to the criteria shown in Fig. 7.

- Proposal to the BDVA Board of Directors and announcement to i-Spaces. The results from the review committee are submitted to the BDVA Board of Directors for approval and communicated to the candidates.
- Trophy hand-out ceremony, usually co-located with the European Big Data Value Forum (www.ebdvf.eu) and where trophies are handed out to i-Spaces on stage by the BDVA president.

6 Impact of i-Spaces in Their Local Innovation Ecosystems

Digital Innovation Hubs, in general, and BDVA i-Spaces, in particular, are expected to contribute to the digital transformation and development of their respective ecosystems. They should be deeply rooted in innovation ecosystems and offer digital transformation services to companies in their proximity. They are also expected to contribute to the development of the RIS3 (Research and Innovation Strategies for Smart Specialisation) strategy.⁵ To illustrate this, below we sketch several specific actions carried out by the BDVA i-Spaces supporting the emergence of their respective ecosystems.

CeADAR: Ireland's Centre for Applied Artificial Intelligence

The CeADAR centre is a main plank in Ireland's Smart Specialisation Strategy, particularly in applied AI and data analytics. The centre is directly funded by the

⁵<https://s3platform.jrc.ec.europa.eu/s3-guide>

Department of Business, Enterprise and Innovation through its two main industry agencies, Enterprise Ireland (EI) and the Industrial Development Authority (IDA), which are in charge of the S3 R&I strategies and priorities for Ireland. In 2018, CeADAR went through an international review process where it was referred to as a key contributor to the digital transformation of Ireland's industry. As part of this review, the centre has received funding from the State Agencies of €12 million to drive its data analytics and artificial intelligence agenda. CeADAR as the National Technology Centre for Applied Data Analytics and Artificial Intelligence has developed links with some of the other technology centres to combine their domain knowledge in specific areas with their expertise in different fields of AI.

CINECA

Embedded in the Italian national HPC centre, CINECA i-Space operates at the intersection of big data, HPC and deep learning technologies to support research and innovation with the most advanced infrastructure, tools, services and skills. The RIS3 Emilia-Romagna strategy is based on four strategic priorities: (i) to increase Emilia-Romagna enterprise competitiveness, (ii) to sustain the emerging specialisation areas, (iii) to provide orientation to the digital transformation and (iv) to develop services of excellence, in four specialisation areas: (a) building and construction, (b) mechatronics and motoring, (c) health and wellness industries and (d) cultural and creative industries. CINECA developed dozens of projects involving large companies and SMEs of all specialisation areas, providing value-added services rooted in advanced simulation, big data and AI technologies.

EURECAT/Big Data CoE Barcelona

The Barcelona Big Data Centre of Excellence (Big Data CoE) is an initiative led by EURECAT, which was launched in February 2015 with the support of the Barcelona City Council, the Government of Catalonia and Oracle. Its impact in the regional ecosystem includes actions as being:

- A pillar of the SmartCat Strategy led by the Catalan Government to promote key enabling digital technologies which include big data and data analytics.
- An evolution phase to embrace not only data-related technologies but also AI technologies is a core element for the deployment of the Catalan AI strategy.
- Developing projects with local companies aligned with the RIS3CAT strategy in Catalonia, notably in sectors like Digital Health, Industry 4.0 and Tourism.

ITAINNOVA/Aragon DIH

DIH on "HPC-Cloud and Cognitive Systems for Smart Manufacturing processes, Robotics and Logistics" is the Aragonese initiative that, within a framework of European cooperation (DIH), extends the strategy of economic and industrial promotion of Aragon and the intelligent regional strategy of Aragon, forming the technological and innovative action of the Aragonese Innovation System. Within the National Strategy for Industry 4.0, it has developed an advisory action that will identify the degree of digitisation of the Spanish Industry. Only 15 entities have been selected to carry out this advisory task throughout Spain. ITAINNOVA has been selected as a qualified consultancy entity for the development of these actions in its

areas of influence. This will allow Aragon DIH the ability to offer its services fully integrated into the national strategy of digitisation of the industry.

ITI/Data Cycle Hub

The Data Cycle Hub, coordinated by ITI, is a Digital Innovation Hub composed of a consortium of organisations with complementary experience that supports companies and the public sector in the Valencia region in their digital transformation. The Valencian Institute of Business Competitiveness (IVACE) is the coordinator of the RIS3CV (development of the RIS3 strategy specifically for the Valencia region). ITI has been working with IVACE in the RIS3CV strategy since the beginning, carrying out the ICT secretariat and working with all the ICT ecosystems. ITI also developed the Industry 4.0 agenda in the Valencia region. Activities of the Data Cycle Hub are aligned with almost all of the RIS3CV areas, including industry (working directly with the Industry 4.0 Lab with IVACE), Health, Tourism, Agrifood, Habitat and Cities, Transport and Energy (also working in Smart Grid Lab with IVACE) – all of them included in the RIS3CV priorities.

Know-Center

Know-Center Graz was founded in 2000 within the framework of the COMET K1 program, and became Austria's leading research centre for data-driven business innovative information and communication technologies. It actively integrates into national cooperation and networks including Green Tech Cluster, AC Styria, Human. Technology Styria, Styrian Service Cluster, Silicon Alps Cluster and IT Clusters. It has close ties with competence centres such as Pro2Future, Virtual Vehicle, Materials Center Leoben and Large Engines Competence Center.

RISE/ICE by RISE

ICE, the Infrastructure and Cloud datacenter test Environment, is a research data centre inaugurated in January 2016. The facility is open to use primarily for European projects, universities and companies. However, customers and partners from all over the world are welcome to use ICE for their testing and experiments. ICE's mission is to contribute to Sweden being at the absolute forefront regarding competence in sustainable and efficient data centre solutions, cloud applications and data analysis, including links with other regional DIHs such as EIT RawMaterials CLC North, Luleå EIT InnoEnergy. ICE is fully aligned with the regional development plan and is running an S3 pilot for an AI and big data ecosystem in the region.

Smart Data Innovation Lab (SDIL)

The SDIL supports pre-commercial research between academia and industries, especially SMEs, in the areas of smart infrastructure, medicine and Industry 4.0. Its potential analysis service under the programme Smart Data Solution Center Baden-Württemberg (SDSC-BW) aims to facilitate entry into smart data analytics application and Industry 4.0 for SMEs. All of these correspond to the digitisation strategy of Germany as well as the RIS3.

TeraLab

TeraLab provides AI and big data “one-stop shop” support to research organisations, web innovators, start-ups, midcaps and large groups, as well as governmental and educational organisations. TeraLab is actively involved in France’s regional and national initiatives around AI and big data:

- It is a consortium partner of the regional initiative PACK IA.
- It contributed to the national AI mission led by Cédric Villani.
- It participated in the projects ADMIRR, EXPRESSO, GeoLytics, M4P, PULSE and Data&Musée.

Universidad Politécnica de Madrid/Madrid’s i-Space for Sustainability/AIR4S DIH

This DIH/i-Space, aligned with the RIS3-Madrid priorities, supports the digitisation of industry, especially SMEs but also midcaps, big companies and public administrations, to improve their products, services and processes, by introducing the great advantages of artificial intelligence and robotics into their business. AIR4S provides companies in all disciplines with a multidisciplinary and personalised approach and consequently addresses multisector domains in a confident way. It brings together world-class technological expertise and infrastructure on AI and robotics but also deep knowledge on how to apply these technologies on different market domains, while being aligned with the Sustainable Development Goals and being respectful of the social, legal and ethical aspects of these technologies.

In the context of data spaces and data communities, AIR4S supports the creation of links between different local initiatives related to access to open data and facilitates cooperation among different data holders at the local level. These links can be created and maintained thanks to the permanent collaboration among European DIHs and the connection to local public systems.

7 Cross-Border Collaboration: Towards a European Federation of i-Spaces

To fully exploit the benefits that the different Digital Innovation Hubs (DIHs) are bringing to the industry, one step beyond in the collaboration among those initiatives and towards a network of DIHs is necessary. In the report “Digital Innovation Hubs: Mainstreaming Digital Innovation across All Sectors”,⁶ the creation of a Europe-wide network of DIHs supporting any business at a “working distance” is seen as an ambitious but achievable objective. In this way, the EC has invested EUR 500 million in the Horizon 2020 programme in initiatives for:

---

⁶https://ec.europa.eu/futurium/en/system/files/ged/dei_working_group1_report_june2017_0.pdf

- Networking and collaboration of digital competence centres and cluster partnerships
- Supporting cross-border collaboration of innovative experimentation activities
- Sharing of best practices and developing a catalogue of competencies
- Wide use of public procurement of innovations to improve efficiency and quality of the public sector

As a result, there exist some running initiatives whose objectives are to break silos, find synergies and foster collaboration among DIHs in different technologies and domains (as relevant examples, the AI DIH network (<https://ai-dih-network.eu>) aims at establishing a framework for continuous collaboration and networking between DIHs focusing on artificial intelligence, MIDIH project (<https://www.midih.eu>) aims to create a network of manufacturing DIHs in the area of IoT/Cyber-physical systems (CPS), DIHNET (<https://dihnet.eu>) supports collaboration among DIH networks across Europe, and DIHelp (<https://dihelp.eu>) is a mentoring and coaching programme supporting 30 DIHs to develop and/or scale up their activities).

This role of DIHs is reinforced in the envisioned Digital Europe Programme⁷ (see Fig. 8), as a means to ensure the digital transformation of all businesses as well as public administrations, in a broad roll-out of digital technologies and digital skills to the entire economy. DIHs are supposed to work closely with the relevant specialised centres and make sure that companies and public administrations can experiment with those technologies (test before investing) and develop skills to meet their needs. As part of this programme, the European Commission also envisages the creation of a network of European DIHs including all regions of Europe, to cover activities with a clear European added value and promote the transfer of expertise.

Regarding big data, the creation of a European federation of Data-Driven Innovation Hubs was included as part of the H2020 programme in 2020, under the topic DT-ICT-05,⁸ with the main challenge of breaking “data silos” and stimulating sharing, reusing and trading of data assets, federating data sources and fostering collaborative initiatives with relevant digital innovation hubs, with the ultimate objective of contributing to the creation of the European Common Data Space. The call explicitly mentioned the BDVA i-Spaces among those initiatives to coalesce towards this federation of Data-Driven Innovation Hubs.

The concept is completely aligned with the strategy of the BDVA i-Spaces group, as is reflected in the BDVA SRIA, where supporting linkages to other innovation spaces and facilitating experiments across multiple innovation spaces is seen as a crucial point towards an effective federation that will help to support research and innovation activities through accessing and processing data assets across national borders. The i-Spaces group has been working in recent years with that objective in mind, to foster collaborations and define the processes towards the creation of a

---

⁷<https://ec.europa.eu/futurium/en/digital-innovation-hubs/digital-innovation-hubs-digital-europe-programme>

⁸<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/dt-ict-05-2020>

Fig. 8 Schematic overview of the role of EDIHs in Digital Europe Programme (European Commission) (by European Commission licensed under CC BY 4.0)

network of i-Spaces. Among those activities, it is worth mentioning the organisation of the workshops “Towards a Federation of European Data Spaces” (BDV PPP Meetup, Sofia, May 2018), “Shaping the European Ecosystem: From i-Spaces and Centres of Excellence to Big Data DIHs” (European Big Data Value Forum 2018, Vienna, October 2018) and “Federation of data services to foster the adoption of data-driven AI in Europe” (BDV PPP Summit, Riga, June 2019), and the joint participation in the 5th meeting of the Working Group on DIHs: Big Data and AI,⁹ organised by the EC in Brussels (November 2018), where i-Spaces shared knowledge, experiences, best practices and their views towards a federation of DIHs on big data.

This collaboration crystallised in a successful project proposal under the call DT-ICT-05. This EUHubs4Data project started in September 2020 and will run for 3 years, with the overarching objective of creating the reference federation in Europe for big data cross-border experimentation and innovation, providing a complete pan-European catalogue of data sources and services to foster data-driven innovation at local and regional level. The project also aims to:

⁹<https://ec.europa.eu/digital-single-market/en/news/fifth-meeting-working-group-digital-innovation-hubs-big-data-and-artificial-intelligence>

- Contribute to the creation of the European Common Data Space, by mobilising, sharing and making available all types of data (close/open, personal/industrial, private/public, research, etc.), with the objective of generating data value from them and fostering data-driven innovation in Europe.
- Lay the foundations for the creation of a pan-European federation of initiatives focused on data-driven innovation and experimentation (DIHs on big data) that, based on strong collaboration and value co-creation, will support European business in their development, and launch data-driven products and solutions to the market, assisting them in their whole journey along the data value chain.
- Add value to the ecosystem of existing initiatives in Europe, positioning a one-stop shop for data-driven innovation and experimentation; building community around the data economy; establishing a liaison among data-driven research and innovation, regulatory bodies and policy makers, industry and data service providers; and bringing together and aligning all actors necessary to boost data-driven innovation.

To accomplish its objectives, the EUHubs4Data project will rely on the following pillars:

- A starting point, with an initial ecosystem composed of the BDVA i-Spaces and some relevant players to link with data-driven initiatives in Europe
- The expansion of the ecosystem during and after the lifecycle of the project, defining a model to incorporate new DIHs into the federation of DIHs; access to local, national and European data incubators; and the involvement of more SMEs
- The offer of the federation, with a global catalogue of data-related services, which will configure the offer of the federation of DIHs to end users. This global catalogue will rely on the individual catalogues of the DIHs, will be enriched with outcomes and assets coming from past and existing European actions (projects) and will be accessible at the local level through the regional DIH or local access point.
- The attraction of the demand side, by a cross-border data innovation programme, with the three-fold objective of (i) attracting the demand side to use the federated services in a cross-border basis, (ii) testing the model of service provisioning and (iii) defining the model to be applied once the project is finished.
- The community around the federation, with whom links will be established with the objective of bringing together all European initiatives working around the data economy and data technologies.
- Business and sustainability, to define a model that includes all aspects that guarantee the continuity of all activities of the federation once the project is finished.

The main outcome of the project will be a federated catalogue that will be made available to companies in the different European regions through their respective DIHs, which will provide access to specific federated services following the paradigm “European catalogue, regional offer” (as reflected in Fig. 9). Specificities about the federated catalogue and how the local offer is instantiated by the regional DIH

Fig. 9 EUHubs4Data European catalogue and regional offer

based on the catalogue will remain transparent for local companies, which will have access to an improved offer through its regular point of sale. Hence, DIHs of the federation will act as bridges for European SMEs to a unique catalogue that will include European data-driven innovations coming from multiple stakeholders.

Another important aspect of the EUHubs4Data project will be to actively contribute to the alignment of existing European initiatives towards the common objective of mobilising, sharing and making available all types of data (close/open, personal/industrial, private/public, research, etc.), in order to get value from them, foster data-driven innovation in Europe, and contribute to the creation of a Common European Data Space. To achieve this, a specific task of the project will be devoted to (i) identifying relevant existing European initiatives on big data and related technologies, (ii) defining a clear value proposition in order to define the guidelines of collaboration with the mentioned objectives in mind, (iii) establishing the necessary links with those initiatives and (iv) specifying a roadmap that defines the work to be done (Fig. 10).

8 Success Stories

Below, we report on the success stories for each of the different BDVA i-Spaces, particularly highlighting their contribution and use in key actions and projects.

Fig. 10 EUHubs4Data community

8.1 *CeADAR: Ireland's Centre for Applied Artificial Intelligence*

Bespoke Innovation and Collaborative Projects CeADAR provides translational research projects to companies for integration in their operational/production systems. As part of this service, companies benefit by (i) starting their data and artificial intelligence journey, (ii) outsourcing key problems to explore new technological avenues, (iii) developing their own in-house data science team and (iv) participating in consortiums to tackle big challenges.

65 Market-Oriented Demonstrators CeADAR delivers approximately eight demonstrator projects per year in two cycles of 6 months, each in collaboration with industry partners. Each project is proposed by the industry members and is focused on a close-to-market challenge. Project development costs are met from the Centre's core budget. The Centre aims to deliver the following for each project: (i) state-of-the-art review, (ii) technical specification, (iii) demonstrators and (iv) assistance with member on-premise demonstrator evaluation. The extensive catalogue of over

65 technology demonstrators from previous platform research is available to all member companies (<https://www.ceadar.ie/outputs/our-demos>). These demonstrators have proven very useful for companies to start tapping into the benefits of data analytics in their organisations.

Data Science Awards CeADAR is a co-founder of the DatSci Awards, the National Data Science Awards (<https://www.datsciawards.com>). This is the major annual event in Ireland showcasing and celebrating data analytics and AI talent.

Industry Impact and Economic Value Add CeADAR has been in existence for over 7 years and in 2018 went through its 5-year term review, achieving the highest marks on each of the evaluation criteria with an international panel of experts from industry and academia. Due to this success, associated government agencies have increased (by 2.5 times) the funding to the centre for the next 5 years.

8.2 CINECA

Anomaly Detection in an HPC System Inside the project “Deriving and Validating Models for the Infrastructure Monitoring”, the anomaly detection project, carried out by the Multithermal Lab of the University of Bologna on CINECA monitoring data, identified a deep learning model able to achieve high accuracy (90–97%) with a semi-supervised learning approach. This use case is peculiar as CINECA’s role is that of data provider and, of course, of data user, and the automation of the anomaly detection would improve its services. These monitoring data are in the orders of TBytes, are currently used for different purposes (deriving thermal models for each core in the system, predicting a specific algorithm computation time, predictive maintenance, etc.) and are undergoing a process of anonymisation in order to be shared with a larger community of researchers.

Risk Management Code Optimisation for a Large Insurance Company The risk assessment in the life insurance field may require considerable computing power. The algorithm that the large insurance company was previously using took many hours and would not allow for calculating the risk measurement with a nested Monte Carlo approach. In fact, nested Monte Carlo involves two stages, scenario generation (outer stage) and portfolio re-valuation (inner stage), that produce millions of Monte Carlo trajectories to be executed for each of the millions of life policies. The simulation becomes an immediate computational challenge. The insurance company asked CINECA to develop a Proof of Concept (PoC) to demonstrate the improved efficiency that could be obtained with efficient code parallelisation and optimisation. The nested Monte Carlo with parameters 100000×100 for all of the 12M policies was achieved. The insurance company then decided to establish a commercial contract with CINECA for the provision of the service.

Sequential patterns of errors from on-board diagnostic devices for TEXA, a European leader company on electronic diagnostic. In the PRESERVE project,

which has been funded within the Fortissimo EU project, sensor data from TEXA on-board diagnostic tools have been analysed in order to identify the driving habits on the one hand and patterns of operating parameters that are predictive of failures and damages on the other. The result is a portfolio of prototypes of services that can predict failures, mechanical problems or damage at the component level, and offer the manufacturer detailed information to better re-design or upgrade their spare parts or vehicle. The return on innovation investment (ROI2) for TEXA from this project has been estimated as 2,72.

LIGA: A Platform for the Game-Content Market LIGA is a project funded within the Fortissimo EU project in partnership with CNR (Consiglio Nazionale delle Ricerche) and Kumo (an SME in the field of 3D technologies and digital asset creation and management).

The current advantage of Kumo is that it is a platform for collecting, sharing, managing and collaborating on 3D content, where consumers of 3D content can access leading museums, gaming and other brands' data. At the end of July 2018, LIGA stored 25 million entries in its database, describing the popularity of game entities among players. Assuming no new game entities will be created in the future, LIGA will add 12 million of entries per month to its database, resulting in 720 million database rows by mid-2023.

Tax Fraud Detection for SOGEI, the Italian Revenue Agency Computing Centre CINECA, with its IOP4HPDA data scientists, developed predictive models of the fraudulent behaviour of companies in the entailment of tax credit and provided methodological solutions for impact and compliance assessment, in particular relating to training sample bias and model estimation and evaluation. The fraudulent behaviour model increased the auditing success rate from 39% to 65% (precision).

Managing Scientific Data for Various Scientific Communities Among the scientific research projects that the HPC department of CINECA supports, many can be reported as being both very successful and data-intensive projects, e.g. EMODnet (European Marine Observation and Data Network; <http://www.emodnet-chemistry.eu/>) and SPHINX (Data Storage and Preservation of High-resolution climate experiments; <http://sansone.to.isac.cnr.it/sphinx/>).

8.3 EGI

EOSC-Hub (www.eosc-hub.eu) EOSC-hub brings together multiple service providers to create the hub: a single contact point for European researchers and innovators to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research. The project mobilises providers from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and other major European research infrastructures to deliver a common catalogue of research data, services and software for research.

EOSC-hub collaborates closely with GÉANT and the EOSCpilot and OpenAIRE-Advance projects to deliver a consistent service offer for research communities across Europe:

- Start: January 2018
- End: December 2020
- 100 partners from 53 countries, including 19 research communities
- 13 work packages
- 49 services ready for use

eXtreme DataCloud (<http://www.extreme-datacloud.eu>) The eXtreme DataCloud (XDC) is an EU H2020-funded project aimed at developing scalable technologies for federating storage resources and managing data in highly distributed computing environments. The services provided will be capable of operating at the unprecedented scale required by the most demanding, data-intensive research experiments in Europe and worldwide. XDC will be based on existing tools, whose technical maturity is proved, and the project will be enriched with new functionalities and plugins already available as prototypes (TRL6+) that will be brought at the production level (TRL8+) at the end of XDC. The targeted platforms are the current and next-generation e-Infrastructures deployed in Europe, such as the European Open Science Cloud (EOSC), EGI and the Worldwide LHC Computing Grid (WLCG), and the computing infrastructures funded by other public and academic initiatives.

8.4 EURECAT/Big Data CoE Barcelona

Big Data and IoT to Improve Tourism Management in Barcelona With the goal of improving real-time decision making of tourism management in Barcelona as well as in policy definition, Barcelona Big Data CoE conceptualised and executed a big data and IoT-based project in partnership with the Barcelona City Council, the GSM Association Mobile World Capital and Orange. The target was the Sagrada Familia district, the city's hottest tourist attraction which causes severe mobility disruption in this area. We studied the macro-mobility (at district level) using call data records from Orange as well as micro-mobility (at street level) using the dedicated infrastructure of 10 Wi-Fi and GSM sensors around the Sagrada Familia streets as well as 3D cameras at the exits of the closest Metro extensions. We made use of the DATURA platform to perform the analysis of more than 50 TB of data accounting for more than 20 million users (aggregating all sources with national and international tourists) over a year. The main results of the project include seasonal macro- and micro-mobility patterns as well as visitors' profiles (segmented into tourists, excursionists and nightlife visitors) (<https://www.bigdatabcn.com/portfolio-item/bcn-tourism-management-big-data-iot-in-action/>).

Leading eCommerce Company The objectives of the project were to design and develop a new data platform as a critical technology component for a large

e-commerce organisation to become a data-driven company, better support existing core business, and provide new capabilities aimed at a more personalised interaction with the customers. The deployed big data analytics platform scales to support 28 million users' daily interactions around the world, both with batch and real-time use cases.

Advanced Analytics for Cruïlla Cruïlla is a very popular and crowded music festival that takes place every year in Barcelona. Today it is one of the most successful music festivals in Europe. The goal of the project, commissioned to the Big Data CoE by the festival sponsors, was to apply data analytics to improve customer knowledge and develop strategies to boost customer engagement with and loyalty to the festival. User profiling was used to improve customer experience, make better marketing decisions and perform customised campaigns that were monitored through Google analytics and social network data.

Analysis of Wi-Fi Data Sources to Extract Origin—Destination Patterns in a Tram Network TRAM is a company that exploits Barcelona's tram network. The project consisted of analysing data from Wi-Fi sensors installed in trains of the tram lines operated by TRAM. The purpose was to compute O/D (origin and destination) matrices and other indicators and visualise them in a dashboard. In the use case, three trains of two tram lines were equipped with Wi-Fi sensors, which count the aggregated information of MAC id corresponding to passengers' mobile phones with active Wi-Fi. These data are analysed to determine the position of the users and, later, to verify the first and last station of a trip, which is the basic information to compute the O/D matrix. The data are calibrated with IR data sensors (for presence detection), already installed in the trains. The use of accurate data filtering and validation techniques was fundamental to distinguish actual tram passengers from other pedestrians around the train, therefore obtaining realistic O/D matrices.

Data Analysis to Improve Mobility Decisions A Proof of Concept (PoC) was commissioned by AlphaNet Seguretat, an SME which provides a wide range of security services to municipalities. The PoC included the design and deployment of a data analysis solution whose data source was car license plate numbers provided by AlphaNet's infrastructure. The PoC also included the development of algorithms to achieve AlphaNet security objectives and the development of a control dashboard.

8.5 ITAINNOVA/Aragon DIH

The ***Moriarty® platform*** is the result of more than 15 years of research in the field of AI and cognitive systems. Moriarty® is a tool for the design and implementation of advanced artificial intelligence software solutions, developed by ITAINNOVA, that solves various business problems with large volumes of data (big data). With Moriarty® one will be able to understand and structure information, identify hidden

patterns and correlations in data, induce knowledge as well as build learning systems. In an agile, precise and simple way, it will allow one to convert their data into valuable information, facilitating the making of strategic decisions.

A very recent success case using Moriarty is the ***Aragon Tourism Smart Observatory***, which is a dashboard for the regional Tourism Authority in order to let them see the trends of users of social media networks (among other sources) talking about Aragon's tourist places.

This dashboard includes sentiment analysis, tourist places and products, Twitter trends, and semantic searches on relevant tourist websites. Information is updated and analysed in real time in order to provide the latest trends and comments by tourists in the region. This is a technological asset aimed to be used for controlling and developing the regional tourism strategy.

8.6 ITI/Data Cycle Hub

EUHubs4Data The European Federation of Data-Driven Innovation Hubs (Coordinator) (1 September 2020–31 August 2023) (no website yet), with the ambition of becoming a reference instrument for data-driven cross-border experimentation and innovation, supports the growth of European SMEs and start-ups in a global data economy. ITI Data Space is the coordinator and leader of the project and one of the i-Spaces providing support to experiments (42 experiments).

REACH REACH is a European incubator for trusted and secure data value chains (01 September 2020–31 August 2023) (no website yet). It is a second-generation incubator for data-fuelled start-ups and SMEs aiming to develop innovative experiments within data value chains. ITI Data Space is one of the nodes of REACH providing support to experiments and incubation.

TECH4CV TECH4 CV is an alliance of Competence Centres in Enabling Technologies (<https://tech4cv.com/data-hub/>) (1 January 2018–31 December 2020). Especially those based on data, to solve the present and future problems of any company of the Valencian Community. ITI Data Space is leading the alliance and providing the Data Space infrastructure for experiments.

DATAPORTS DATAPORTS is a Data Platform for the Cognitive Ports of the Future (<https://dataports-project.eu/>) (1 January 2020–31 December 2022). It provides a secure environment for the aggregation and integration of data coming from several data sources existing in the digital ports and owned by different stakeholders to improve processes, offer new services and devise new AI-based and data-driven business models. ITI Data Space provides knowledge, tools and methodologies related to big data and AI in digital infrastructures and data-driven business models.

TransformingTransport (TT) TT (<http://transformingtransport.eu/>) (1 January 2017–30 June 2019) demonstrated transformations that big data will bring to the

mobility and logistics market. ITI leads the Ports Domain and the Valencia Port Pilot, providing ITI Data Space for analysing data in ports.

8.7 *Know-Center*

“Mobile Phone Data Analysis” This is an extraordinary example of transferring research results into business. Geospatial data that is continuously generated by cell phones is used to analyse movements of groups of people, thus enabling innovative use cases in intelligent transportation systems (ITS) and in digital marketing. The usage of sensors and embedded technology in vehicles and transportation infrastructure yields new applications in the field of intelligent transportation systems (ITS), such as the prediction of traffic flows and critical transport situations, trip planning in multi-modal transportation and increased traffic. Yet such technology is not pervasively available. Therefore, the application of other location-aware services such as satellite tracking (GPS, Galileo) and cell phone networks is attractive. The latter is of high interest since the technology is available at low cost almost everywhere. Mobile phones regularly generate location-aware (geospatial) events. Other events are cell changes and whenever the user is taking a call or using the data connection. A first study looked at the feasibility of cell phone data in order to detect unusual events such as traffic congestions. The task was to identify congestions, especially on lower-ranking streets, by applying cell phone data without having access to the exact position of individuals, thus satisfying privacy concerns. An algorithmic challenge was how to deal with mobile phone events and their possibly inaccurate data in order to reconstruct trajectories. This resulted in a pool of knowledge, robust tools and scientific publications (Horn et al., 2014KC). Additionally, we addressed topics like transportation mode detection and map matching (Schulze et al. 2015KC). A further challenge was the processing of such data since it arrives as a stream of millions of users simultaneously.

Visual Multi-Perspective Optimisation of Logistic Processes A logistics dashboard is an interactive platform for optimising global logistics processes involving relevant stakeholders in the discussion of strategic alternatives. Logistical processes in production are characterised by a multitude of perspectives with orthogonal optimisation goals. This project addressed the problem of creating a global optimisation strategy for logistical processes through a data-driven visualisation which depicts key parameters and computes models to perspectives. In moderated discussions with the stakeholders different perspectives have been analysed, key parameters identified and interrelations between perspectives established. To inspect the logistical process from all perspectives, an optimum is devised from a dialogue between humans, machines and data. A crucial point that was successfully addressed is that in the optimisation process, human aspects and department interests play as much of a role as data and computational considerations. The interactive visual interface (dashboard) shows information for one or more selected parts. The parameters from

various stakeholders are adjusted to view the impact on relevant key performance indicators. Green bars represent the optimum (i.e. corresponding to lowest costs). The key success factors of the resulting solution are both the model and the simulation, as well as the involvement of all stakeholders in discussing strategic alternatives based on real data.

“Participation in Global Scientific Challenges” Participating in global scientific challenges is our method of choice to benchmark ourselves with research teams worldwide, to test our skills and boost our motivation. We participate in global scientific challenges and compete with research teams worldwide to boost motivation and test our skills. Examples include SemEval, INEX, PAN, SciSumm and SemPub hosted at conference series like JCDL and ESWC, or at venues like CLEF. We won the Book Search shared task at INEX. We were awarded Most Innovative Approach at SemPub and we achieved Second Best Performance at SciSumm, with results having been presented at the SIGIR’17 and being an integral part of a master’s thesis finished in 2018.

“Magna Painting Finishing Optimization” Based on the parameters of the paint job, MagnaPaint predicts the types of paint imperfections and informs the operator on which parameters have the strongest influence. Our industrial partner Magna is continuously trying to improve its processes and products via innovative technologies and methods. One focus area is the paint finishing process, where vehicles are coated with a protective lacquer. Due to external and internal influences, the coating may contain imperfections, which need to be manually removed, which is a costly process. By applying data science methods, we analysed the data and identified a number of root causes for various types of imperfections, which help the operator to increase the overall quality. The data consists of a large number of parameters, ranging from chemical measurements to process information. Together with the domain experts of our industrial partner, we developed a machine learning model, in order to forecast the expected quality of the processes. In cooperation with the Knowledge Visualisation Area, we developed a tool allowing the operator to visually interact with the learnt model. With this tool, the operator can experiment with different parameter sets and observe the predicted results, without the need to actually test these parameters in the production environment. This again saves time and costs and also avoids potential disruptions in the production process.

8.8 NCSR Demokritos/Attica Hub for the Economy of Data and Devices (ahedd)

National Network for Precision Medicine in Oncology Demokritos operates one of the four national units that are providing next-generation sequencing genetic diagnostics (solid tumours and peripheral blood) to the oncology clinics of Greece as well as management and big data analytics of the genetic archives.

National Network for the Environment and Climate Change: Demokritos operates a cluster of analytical laboratories evaluating toxic (particle, chemical and radioactive) pollution in the atmosphere, soil, water, the food chain and biological tissues.

NanoNOSE A recently initiated action with impact on both the agricultural and health sectors, NanoNOSE, will develop AI methodologies that will be used to combine expert input and advanced sensory data for identifying and predicting risk related to the existence of harmful microorganisms in crop silos.

Marie Curie fellowship for the design of material for gas separation membranes: The research will be based on the incorporation of machine learning techniques in a smart screening methodology that will illustrate the missing correlation between structural modification of the materials and their separation performance.

AI4EU The EU's landmark AI project (€20 million project, Jan. 2019–2022) seeks to develop an EU AI ecosystem, integrating the knowledge, algorithms, tools and resources available, and making it a compelling solution for users. Involving 80 partners across 21 countries, AI4EU will unify the EU's AI community.

IASIS Its aims are to seize the opportunity provided by a wave of data heading our way and turn this into actionable information that would match the right treatment with the right type of patient.

8.9 *RISE/ICE by RISE*

The aim of the D-ICE project is to establish an arena for data-driven innovation. The objective is to improve the conditions for value creation based on advanced data analytics in the industry and society.

The project is financed by national funding (Vinnova) over 21 months, and the partners are Ericsson, RISE SICS and the start-up Logical Clocks. The objective was to strengthen the Swedish competence in data handling, analysis and processing. The project built a collaboration (meeting and tools) platform for data owners and data analysis providers. The basis for the project is the national data centre initiative ICE with all server capacities; analytic tools, for example Flink and HOPS; and the data analytics and industry knowledge that exists within all parts of RISE.

The first pilot case in the project was done together with Scania, a supplier of heavy trucks to a global market. The number of connected Scania vehicles exhibits exponential growth, resulting in large amounts of streaming telematics data. In their own project FUMA, Scania's objective is to develop a big automotive data analytics framework that utilises its collected geolocation data to analyse the behaviour of vehicles from both an individual vehicle perspective and a fleet perspective.

When connecting FUMA to the D-ICE project, new possibilities were created for Scania, to be able to use our collaboration platform for testing new big data platforms

and meet and work together with other organisations in our neutral third-party development environment.

The second pilot, for Mobilaris, was done to improve their product and service for positioning of mobiles and other connected equipment. Mobilaris's market is mobile operators, mining industries and public safety. The positioning system of users or equipment data has an operation user dashboard with analytics capabilities. The large dataset requires a distributed data management and analytics system to achieve low response times.

The services provided were Hadoop-as-a-service and analytics tools for the development of algorithms and queries, expert service for consultancy, and two racks of servers for comparison of different types of Hadoop distributions by different vendors.

The problem was solved with a Hadoop-based big data distributed file and analytics system. The i-Space provided a low-hurdle Hadoop as-a-service to get started with distributed data management and analytics, and an expert service as learning support and query analysis, as well as infrastructure in the form of two racks with 20 servers each for comparison operation of different types of Hadoop distributions for an understanding of product implementations.

The ICE i-Space can deliver a system and service not available in a smaller company that does not have the initial skills for operating a data centre and a Hadoop system and does not implement big data-based analytics. Smaller companies do not have the financial muscle to either do this by themselves to get started or to carry out a pre-study for decision making.

8.10 *Smart Data Innovation Lab (SDIL)*

Smarte Techniker-Einsatzplanung (STEP) The research project *Smarte Techniker-Einsatzplanung*, or “Smart Technician Mission Planning” (STEP), aims to simultaneously increase the efficiency of technician assignments and the availability of machinery. Information from and about machines generated by emerging technologies, such as predicted service demand, will be used. STEP is funded by the Federal Ministry for Economic Affairs and Energy (BMWi) in the context of the programme “Smart Service Welt I”. Several project partners will work on the simulation model with real dispatching operation data. This requires a safe and cooperative setting which is offered by SDIL (<http://www.sdil.de/en/projects/smart-technician-mission-planning-step/>).

BigGIS: Fusion of Geospatially Distributed Heterogeneous Sensor Data BigGIS is a joint project between the regional office for environmental protection and various universities and firms in Baden-Württemberg. The project deals with big data and the fusion of uncertain geographic data. Increasing data volumes and increasingly complex calculation models require fast and robust procedures. Together with the SDIL, suitable algorithms are implemented, tested and further

developed on the basis of temperature data. It aims at a scalable system that takes into account the peculiarities of spatial and temporal relationships. Therefore, the system must be able to merge the geospatial data as well as a model of its uncertainty, taking into account the heterogeneity of the data sources. The computing resources of the SDIL offer considerable added value for BigGIS, since data volumes in the gigabyte to terabyte range are processed (<http://www.sdil.de/en/projects/biggis-fusion-of-geospatially-distributed-heterogeneous-sensor-data/>).

Smart Data Solution Center Baden-Württemberg Project Networking Knowledge. Building a technology referral service is a complex venture. The demands on smart technologies and continuous evaluation are very high and require a well-established methodology. Coral Innovation, a young start-up of the University of Stuttgart, implemented just such a service and was supported by experts from SDSC-BW. The free-of-charge potential analysis with more than 8000 binary test classification questions was carried out on the SDIL platform and showed possible optimisation of the classification values (<http://www.sdil.de/en/projects/sdsc-bw-networking-knowledge/>).

TransformingTransport: Ports as Intelligent Logistics Hubs This project is part of the TransformingTransport EU lighthouse project that aims to demonstrate, in a realistic, measurable and replicable way, the transformative effects that big data will have on the mobility and logistics market. TransformingTransport brings together knowledge, solutions and impact potential of major European ICT and big data technology providers with the competence and experience of key European industry players and public bodies in the mobility and logistics domain. This project should demonstrate how solutions for objectives of a seaport pilot can be replicated and reused for the more challenging setting of an inland port. Compared to seaports, the added complexity in an inland port stems, for example, from the fact that the port is situated in the middle of a large city and at the centre of a large metropolitan area. This means that it has a multitude of roads, tracks and waterways that serve as entry and exit points for containers to and from the actual terminals and ports. In addition, roads need to be shared with many other cars within the metropolitan area. This task will extend the results of a large national innovation project on logistics control towers and enhance them with advanced big data analytics and visualisation capabilities that integrate the various relevant data sources from the port and terminals (<http://www.sdil.de/en/projects/ports-as-intelligent-logistics-hubs>).

8.11 TeraLab

MIDIH (“Manufacturing Industry Digital Innovation Hub”, H2020, I4MS) (fully operational since October 2017): (www.midih.eu). MIDIH is a “one-stop shop” of services, providing industry with access to the most advanced digital solutions and industrial experiments, pools of human and industrial competencies, and access to “ICT for manufacturing” market and financial opportunities.

BOOST4.0, operational, started in January 2018 (www.boost.eu). *BOOST 4.0* “Big Data Value Spaces for Competitiveness of European Connected Smart Factories 4.0” will demonstrate, in a realistic, measurable and replicable way, an open, certifiable and highly standardised and transformative shared data-driven Factory 4.0. BOOST 4.0 will also demonstrate how European industry can build unique strategies and competitive advantages through big data across all phases of the product and process lifecycle (engineering, planning, operation, production and after-market services) building upon the BOOST 4.0 connected smart Factory 4.0 model to meet the Industry 4.0 challenges.

AI4EU will efficiently build a comprehensive European AI-on-demand platform to lower barriers to innovation, to boost technology transfer, and to catalyse the growth of start-ups and SMEs in all sectors through open calls and other actions. The platform will act as a broker, developer and one-stop shop providing and showcasing services, expertise, algorithms, software frameworks, development tools, components, modules, data, computing resources, prototyping functions and access to funding. Training will enable different user communities (engineers, civic leaders, etc.) to obtain skills and certifications.

Proof of ROI (Insurance) Client Profile: Mutual health insurance company (confidential). Client Needs: Early stage data experiment prototype scenario: A large French mutual health insurance company is considering an important strategic move towards novel big data techniques to improve knowledge of their subscriber behaviour. The business lines had identified several use cases, involving heavy machine learning algorithms. They requested support from the IT division, which evaluated the necessary investment. At this stage, the business lines were unable to provide ROI evaluation without concrete experimentation to allow authorisation of such an investment.

Access to research and technology (logistic).

Client Profile: La Poste, Mail Division.

Client Needs: Real value of the data collected by the mail sorting machines. Quality of data and then extraction of useful conclusions about the processes with the focus on two aspects: the fraud of the franking marks and data visualisation of the real process inside a sorting centre to be compared with the theoretical process.

Provided Solution to Meet the Needs: TeraLab provided a workspace and worked closely with La Poste to be able to get 15 Tbytes of data on TeraLab. A research team worked on anonymisation. An innovative company worked on the two use cases described previously. It was the first time La Poste was able to work on the entire dataset.

8.12 *Universidad Politécnica de Madrid/Madrid's i-Space for Sustainability/AIR4S DIH*

TransformingTransport (H2020-731932). This project demonstrated how big data can be used in the context of mobility and logistics (1 January 2017–31 July 2019) (<https://transformingtransport.eu/>). Our role in this project has been the creation of a data portal for all the open and closed data used by pilots in this project. The data portal is available at <https://data.transformingtransport.eu/>.

BigStorage (H2020-642963). This Marie Curie ITN project focused on training data scientists in order to enable them to apply holistic and interdisciplinary approaches, taking advantage of a data-overwhelmed world, which requires HPC and cloud infrastructures (1 January 2015–31 December 2018) (<http://bigstorage-project.eu/>). Our role in this project was in the development of efficient I/O techniques for big data management.

BigDataStack (H2020-779747). A project focused on delivering a completely open-source stack of high-performance technologies (1 January 2018–31 December 2020) (<https://bigdatastack.eu/>). Our role in this project is in the development of part of the open-source technology stack.

BigMedilytics (H2020-780495). A project focused on the application of big data technologies in the health sector (1 January 2018–28 February 2021) (<https://www.bigmedilytics.eu/>). In this project, our role is focused on the application of data mining and text mining techniques to health-related documents.

Ciudades Abiertas. This project is funded by the Spanish Government institution red.es, for the provision of Open Government solutions to cities in Spain, piloted in Madrid, Zaragoza, Santiago de Compostela and A Coruña (30 May 2018–31 December 2020) (<https://ciudadesabiertas.es/>). Our role in this project is the creation of ontologies to guide the publication of open data for these cities.

9 Summary

Despite the increasing relevance of the data economy in Europe, and the importance of data-driven innovation in fostering the digitalisation of companies and society, there are still many actors (small and medium) at national and regional level that do not have access to the benefits of data. There have been many efforts in recent years to solve this issue, from the European Commission, with the Digital Innovation Hubs as main instruments, and also from others, like the Big Data Value Association that is focused more on data, with the Data Innovation Spaces. This chapter presented these and other instruments, introducing their main aspects and characteristics and presenting alignments among them. It also focused on the certification process followed by the Big Data Value Association to recognise relevant initiatives in this field across Europe, and highlighted the importance of collaboration, with the project EUHubs4Data aimed at creating a European federation of Data-Driven

Innovation Hubs, as a meaningful practical example. Finally, the chapter presented some best practices and success stories that could be seen as experiences and lessons for the future.

Acknowledgements We are grateful for the contributions of the following individuals: Claudio Arlandini (Project Manager HPC for industry presso CINECA), Roberta Turra (Team Lead CINECA), Anne-Sophie Taillandier (Director of TeraLab), Natalie Cernecka (Head of Business Development of TeraLab), Maria Eugenia Fuenmayor Garcia (Scientific Director of ICT and Media Areas Eurecat), Professor Stefanie Lindstaedt (CEO Know-Center and Director of Institute of Interactive Systems and Data Science at Graz University of Technology), Tor Björn Minde (Head of Lab at RISE ICE Datacenter), Dr. Ricardo Simon Carbajo (Head of Innovation and Development CeADAR – Ireland’s Centre for Applied AI), Dr. Edward McDonnell (Centre Director, CeADAR – Ireland’s Centre for Applied AI), Sergio Mayo Macías (Information Systems Project Manager ITAINNOVA), Michael Beigl (Head of national competence centre for Big Data AI, the Smart Data Innovation Lab SDIL), Sy Holsinger (Strategy and Innovation Lead/Business Development Manager at EGI Foundation), Periklis Terlixidis (Executive Officer of Attica Hub for the Economy of Data and Devices (ahedd) at NCSR “DEMOKRITOS”), Cristina Sandoval (International Projects Office at Universidad Politécnica de Madrid), Oscar Corcho (Professor at Ontology Engineering Group UPM). The research leading to these results received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 951771 (EUHubs4Data) and under grant agreement no. 732630 (BDVe).

Reference

Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.). (2017). *European big data value strategic research & innovation agenda*. Retrieved from Big Data Value Association website https://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.