A Roadmap to Drive Adoption of Data Ecosystems

Sonja Zillner, Laure Le Bars, Nuria de Lama, Simon Scerri, Ana García Robles, Marie Claire Tonna, Jim Kenneally, Dirk Mayer, Thomas Hahn, Södergård Caj, Robert Seidl, Davide Dalle Carbonare, and Edward Curry

The original version of this chapter was revised with an update in the affiliation of the author Davide Dalle Carbonare. A correction to this chapter can be found at https://doi.org/10.1007/978-3-030-68176-0_17

S. Zillner (✉)
Siemens AG, Munich, Germany
e-mail: sonja.zillner@siemens.com

L. Le Bars
SAP, Paris, France

N. de Lama
Atos, Madrid, Spain

S. Scerri
Fraunhofer IAIS, Sankt Augustin, Germany

A. García Robles
Big Data Value Association, Bruxelles, Belgium

M. C. Tonna
Digital Catapult, London, UK

J. Kenneally
Intel, Leixlip, Ireland

D. Mayer
Software AG, Saarbrücken, Germany

T. Hahn
Siemens AG, Erlangen, Germany

S. Caj
VTT, Espoo, Finland

R. Seidl
Nokia Bell Labs, Munich, Germany

D. D. Carbonare
Engineering Ingegneria Informatica, Rome, Italy

E. Curry
Insight SFI Research Centre for Data Analytics, NUI Galway, Galway, Ireland

Abstract To support the adoption of big data value, it is essential to foster, strengthen, and support the development of big data value technologies, successful use cases and data-driven business models. At the same time, it is necessary to deal with many different aspects of an increasingly complex data ecosystem. Creating a productive ecosystem for big data and driving accelerated adoption requires an interdisciplinary approach addressing a wide range of challenges from access to data and infrastructure, to technical barriers, skills, and policy and regulation. In order to overcome the adoption challenges, collective action from all stakeholders in an effective, holistic and coherent manner is required. To this end, the Big Data Value Public-Private Partnership (BDV PPP) was established to develop the European data ecosystem and enable data-driven digital transformation, delivering maximum economic and societal benefit, and achieving and sustaining Europe's leadership in the fields of big data value creation and Artificial Intelligence. This chapter describes the different steps that have been taken to address the big data value adoption challenges: first, the *establishment of the BDV PPP* to mobilise and create coherence with all stakeholders in the European data ecosystem; second, the introduction of *five strategic mechanisms* to encourage cooperation and coordination in the data ecosystem; third, a *three-phase roadmap* to guide the development of a healthy European data ecosystem; and fourth, a systematic and strategic approach towards actively *engaging the key communities* in the European Data Value Ecosystem.

Keywords Big data value · Public Private Partnership · European data ecosystem · Adoption of big data

1 Introduction

To support the adoption of big data value, it is essential to foster, strengthen and support the development of big data value technologies, successful use cases and data-driven business models. At the same time, it is necessary to deal with many different aspects of an increasingly complex data ecosystem. Creating a productive ecosystem for big data and driving accelerated adoption was possible by relying on an interdisciplinary approach addressing a wide range of central challenges from access to data and infrastructure, to technical barriers, skills, and policy and regulation. Given the broad range of challenges and opportunities with big data value, new instruments, an aligned implementation roadmap and a strategic approach towards cooperation were needed. In this chapter, we set out such a strategy, the formulation of which is the result of an inclusive discussion process involving a large number of relevant European Big Data Value (BDV) stakeholders. The result is an interdisciplinary approach that integrates expertise from the different fields necessary to tackle both the strategic and specific objectives. To this end, the Big Data Value Public-Private Partnership was established to develop the European data ecosystem and enable data-driven digital transformation, delivering maximum economic and societal benefit, and achieving and sustaining Europe's leadership in the fields of big data value creation and Artificial Intelligence.

This chapter starts by detailing the adoption challenges of big data value and all the different steps that were taken to overcome the adoption challenges: first, the *establishment of the Big Data Value Public-Private Partnership (BDV PPP)* to mobilise and create coherence with all stakeholders in the European data ecosystem; second, the introduction of *five strategic mechanisms* to encourage cooperation and coordination in the data ecosystem; third, a *three-phase roadmap* to guide the development of a healthy European data ecosystem; and fourth, a systematic and strategic approach towards actively *engaging the key communities* in the European Data Value Ecosystem.

2 Challenges for the Adoption of Big Data Value

To support the adoption of big data value, it was important to foster, strengthen and support the development of big data value technologies, successful use cases and data-driven business models. At the same time, it was necessary to deal with many different aspects of an increasingly complex data ecosystem. Building on the analysis provided in the literature (Cavanillas et al. 2016; Zillner et al. 2017, 2020), the main challenges that needed to be tackled to create and sustain a robust big data ecosystem have been as follows:

- **Access to Data and Infrastructures:** Availability of data sources and access to data infrastructures is paramount. There is a broad range of data types and data sources: structured and unstructured data, multi-lingual data sources, data generated from machines and sensors, data-at-rest and data-in-motion. Value is created by acquiring data, combining data from different sources and providing access to it with low latency, while ensuring data integrity and preserving privacy. Pre-processing, validating, augmenting data, and ensuring data integrity and accuracy add value. Both academics and innovators (SMEs and start-ups in particular) need proper access to world-class innovation infrastructures, including to data and infrastructure resources such as High Performance Computing (HPC) and test environments.
- **Higher Complexity of Data-driven Applications in Industry and Public Domain:** Novel applications and solutions must be developed and validated in ecosystems to deliver value creation from the data ecosystem. However, implementing data value and data-driven AI in industrial and public environments relies on incorporating the domain knowledge of underlying processes. Handling these challenges requires combining domain-specific process knowledge with knowledge on data-driven approaches.
- **Lack of Skills and Know-How:** To leverage the potential of big data value, a key challenge is to ensure the availability of highly and appropriately skilled people who have an excellent grasp of the best practices and technologies for delivering big data value within applications and solutions. Data experts need to be connected to other experts with strong domain knowledge and the ability to apply this know-how within organisations for value creation. Many European

organisations lack the skills to manage or deploy data-driven solutions with global competition for talent under way.

- **Policy and Regulation Uncertainty:** The increased importance of data will intensify the debate on data ownership and usage, data protection and privacy, security, liability, cybercrime, Intellectual Property Rights (IPR), and the impact of insolvencies on data rights. These issues have to be resolved to remove the adoption barriers. In the area of data-driven AI, policy and regulation are still unclear in areas including liability, right to explain and data access. Many organisations have concerns about compliance.
- **Technical Barriers:** There is considerable complexity and cost in creating systems with the ability to collect, process, and analyse large quantities of data to make robust and trustworthy decisions and implement autonomy. Key aspects such as real-time analytics, low latency and scalability in processing data, new and rich user interfaces, and interacting with and linking data, information and content all have to be advanced to open up new opportunities and to sustain or develop competitive advantages. Interoperability of data sets and data-driven solutions, as well as agreed approaches, is essential for a wide adoption within and across sectors.
- **Digitalisation of Business:** Businesses have to increase their digitalisation effort to maintain their competitive advantage within a Digital Single Market. A more efficient use of big data, and understanding data as an economic asset, carries great potential for the economy and society. The setup of big data value ecosystems and the development of appropriate business models on top of a strong big data value chain must be supported to generate the desired impact on the economy and employment.
- **Societal Trust in Data:** Big data will provide solutions for major societal challenges in Europe, such as improved efficiency in healthcare information processing and reduced CO₂ emissions through climate impact analysis. However, there are many misconceptions and much misinformation about data-driven systems in societal debates, and the technology seems not to be fully accepted by society in all application areas. It is critical for accelerated adoption of big data to increase awareness of the benefits and the value that big data can create for business, the public sector, the citizen, and the environment.
- **EU Private Investment Environment:** Still lagging behind other parts of the world within its investments in digitalisation, Europe needs to create a competitive, forward-looking private investments ecosystem to boost innovation in data and data-driven AI in a fast and focused way.

Creating a productive ecosystem for big data and driving accelerated adoption requires an interdisciplinary approach addressing all of the challenges above in collective action from all stakeholders working together in an effective, holistic and coherent manner.

3 Big Data Value Public-Private Partnership

Europe must aim high and mobilise stakeholders in society, industry, academia and research to enable a European big data value economy, supporting and boosting agile business actors, delivering products, services and technology, and providing highly skilled data engineers, scientists and practitioners along the entire big data value chain. This will result in an innovation ecosystem in which value creation from big data flourishes.

To achieve these goals, the **European contractual Public-Private Partnership on Big Data Value (BDV PPP)** was signed on 13 October 2014. This signature marks the commitment by the European Commission, industry and academia partners to build a data-driven economy across Europe, mastering the generation of value from big data and creating a significant competitive advantage for European industry, boosting economic growth and employment. The **Big Data Value Association (BDVA)** is the private counterpart to the EU Commission in implementing the BDV PPP programme. BDVA has a well-balanced composition of large, small and medium-sized industries and enterprises as well as research organisations to support the development and deployment of the PPP work programme and to achieve the Key Performance Indicators (KPI) committed in the PPP contract. The BDV PPP commenced in 2015 and was operationalised with the launch of the LEIT work programme 2016/2017. The BDV PPP activities address technology and applications development, business model discovery, ecosystem validation, skills profiling, regulatory and IPR environment, and social aspects. The BDV PPP did lead to a comprehensive innovation ecosystem fostering and sustaining European leadership on big data and delivering maximum economic and societal benefit to Europe – its business and its citizens (see Chap. “Achievements and Impact of the Big Data Value Public-Private Partnership: The Story so Far” for more details).

3.1 *The Big Data Value Ecosystem*

A data ecosystem is a socio-technical system enabling value to be extracted from data value chains supported by interacting organisations and individuals (Curry 2016). Within an ecosystem, data value chains are oriented to business and societal purposes. The ecosystem can create the conditions for marketplace competition between participants or can enable collaboration among diverse, interconnected participants that depend on each other for their mutual benefit.

The clear goal of the BDV PPP was to develop a European data ecosystem that enables data-driven digital transformation in Europe, delivers maximum economic and societal benefit, and fosters and sustains Europe’s leadership in the fields of big data value creation and Artificial Intelligence. The ecosystem is established on a set of principles to ensure openness, inclusion and incubation (see Table 1).

Table 1 The principles of the big data value ecosystem

Focus	BDV ecosystem principle
<i>Openness</i>	The ecosystem should embrace and contribute to openness in terms of data, technology standards, interoperability, best practice, education and innovation.
<i>Connectedness</i>	The ecosystem should prioritise connectedness and synergies between actors, industrial sectors, and languages, and across borders.
<i>Cross-sectorial</i>	The ecosystem should involve all relevant sectors of society and scientific disciplines.
<i>Sustainable</i>	The ecosystem should strive for self-sustainability in structure and function, in order to maintain long-term ecosystem services.
<i>Co-evolution</i>	The ecosystem should support competition, cooperation and co-evolution between actors.
<i>Incubation</i>	The ecosystem should incubate start-ups and entrepreneurs.

4 Five Mechanism to Drive Adoption

In order to implement the research and innovation strategy, and to align technical issues with aspects of cooperation and coordination, five major types of mechanisms were identified:

- **Innovation Spaces (i-Spaces):** Cross-organisational and cross-sectorial environments that allow challenges to be addressed in an interdisciplinary way and serve as a hub for other research and innovation activities
- **Lighthouse projects:** To raise awareness of the opportunities offered by big data and the value of data-driven applications for different sectors, acting as incubators for data-driven ecosystems
- **Technical projects:** To tackle specific big data issues, addressing targeted aspects of the technical priorities
- **Data platforms:** To support the sharing and trading of industrial and personal data (free flow of data) as a key enabler of the data economy
- **Cooperation and coordination projects:** To foster international cooperation for efficient information exchange and coordination of activities within the ecosystem

4.1 European Innovation Spaces (i-Spaces)

Extensive consultation with many stakeholders from areas related to big data value (BDV) had confirmed that in addition to technology and applications, several key issues required consideration. First, infrastructural, economic, social and legal issues have to be addressed. Second, the private and public sectors need to be made aware of the benefits that BDV can provide, thereby motivating them to be innovative and to adopt BDV solutions.

To address all of these aspects, European cross-organisational and cross-sectorial environments, which rely and build upon existing national and European initiatives, play a central role in a European big data ecosystem. These so-called European Innovation Spaces (or i-Spaces for short) 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. As such, i-Spaces enable stakeholders to develop new businesses facilitated by advanced BDV technologies, applications and business models. They contribute to the building of communities, providing a catalyst for community engagement and acting as incubators and accelerators of data-driven innovation.

In this sense, i-Spaces are hubs for uniting technical and non-technical activities, for instance, by bringing technology and application development together and by fostering skills, competence and best practices. To this end, i-Spaces offer both state-of-the-art and emerging technologies and tools from industry, as well as open-source software initiatives; they also provide access to data assets. In this way, i-Spaces foster community building and an interdisciplinary approach to solving BDV challenges along the core dimensions of technology, applications, legal, social and business issues, data assets, and skills.

The creation of i-Spaces is driven by the needs of large and small companies alike to ensure that they can easily access the economic opportunities offered by BDV and develop working prototypes to test the viability of actual business deployments. This does not necessarily require moving data assets across borders; rather, data analytic tools and computation activities are brought to the data. In this way, valuable data assets are made available in environments that simultaneously support the legitimate ownership, privacy and security policies of corporate data owners and their customers, while facilitating ease of experimentation for researchers, entrepreneurs and small and large IT providers.

Concerning the discovery of value creation, i-Spaces support various models: at one end, corporate entities with valuable data assets can specify business-relevant data challenges for researchers or software developers to tackle; at the other end, entrepreneurs and companies with business ideas to be evaluated can solicit the addition and integration of desired data assets from corporate or public sources. i-Spaces also contribute to filling the skills gap Europe is facing in providing (controlled) access to real use cases and data assets for education and skills improvement initiatives.

i-Spaces themselves are data-driven, both at the planning and the reporting stage. At the planning stage, they prioritise the inclusion of data assets that, in conjunction with existing assets, present the greatest promise for European economic development (while taking full account of the international competitive landscape); at the reporting stage, they provide methodologically sound quantitative evidence on important issues such as increases in performance for core technologies or reductions in costs for business processes. These reports have been an important basis to foster learning and continuous improvement for the next cycle of technology and applications.

The particular value addition of i-Spaces in the European context is that they federate, complement and leverage activities of similar national incubators and environments, existing PPPs, and other national or European initiatives. With the aim of not duplicating existing efforts, complementary activities considered for inclusion have to stand the test of expected economic development: new data assets and technologies are considered for inclusion to the extent that they can be expected to open new economic opportunities when added to and interfaced with the assets maintained by regional or national data incubators or existing PPPs.

Over recent years, the successive inclusion of data assets into i-Spaces, in turn, has driven and prioritised the agenda for addressing data integration or data processing technologies. One example is the existence of data assets with homogeneous qualities (e.g. geospatial factors, time series, graphs and imagery), which called for optimising the performance of existing core technology (e.g. querying, indexing, feature extraction, predictive analytics and visualisation). This required methodologically sound benchmarking practices to be carried out in appropriate facilities. Similarly, business applications exploiting BDV technologies have been evaluated for usability and fitness for purpose, thereby leading to the continuous improvement of these applications.

Due to the richness of data that i-Spaces offer, as well as the access they afford to a large variety of integrated software tools and expert community interactions, the data environments provide the perfect setting for the effective training of data scientists and domain practitioners. They encourage a broader group of interested parties to engage in data activities. These activities are designed to complement the educational offerings of established European institutions.

4.2 *Lighthouse Projects*

Lighthouse projects¹ are projects with a high degree of innovation that run large-scale data-driven demonstrations whose main objectives are to create high-level impact and to promote visibility and awareness, leading to faster uptake of big data value applications and solutions.

They form the major mechanism to demonstrate big data value ecosystems and sustainable data marketplaces, and thus promote increased competitiveness of established sectors as well as the creation of new sectors in Europe. Furthermore, they propose replicable solutions by using existing technologies or very near-to-market technologies that show evidence of data value and could be integrated in an innovative way.

Lighthouse projects lead to explicit business growth and job creation, which is measured by the clear indicators and success factors that had been defined by all projects in both a qualitative and quantitative manner beforehand.

¹Sometimes also labelled as large-scale demonstrations or pilots.

Increased competitiveness is not only a result of the application of advanced technologies; it also stems from a combination of changes that expand the technological level, as well as political and legal decisions, among others. Thus, Lighthouse projects were expected to involve a combination of decisions centred on data, including the use of advanced big data-related technologies, but also other dimensions. Their main purpose has been to render results visible to a widespread and high-level audience to accelerate change, thus allowing the explicit impact of big data to be made in a specific sector, and a particular economic or societal ecosystem.

Lighthouse projects are defined through a set of well-specified goals that materialise through large-scale demonstrations deploying existing and near-to-market technologies. Projects may include a limited set of research activities if that is needed to achieve their goals, but it is expected that the major focus will be on data integration and solution deployment.

Lighthouse projects are different from Proof of Concepts (which are more related to technology or process) or pilots (which are usually an intermediate step on the way to full production): they need to pave the way for a faster market roll-out of technologies (big data with Cloud and HPC or the IoT), they need to be conducted on a large scale, and they need to use their successes to rapidly transform the way an organisation thinks or the way processes are run.

Sectors or environments that were included were not pre-determined but had been in line with the goal mentioned above of creating a high-level impact.

The first call for Lighthouse projects made by the BDV PPP resulted in two actions in the domains of bioeconomy (including agriculture, fisheries and forestry) and transport and logistics. The second call resulted in two actions for health and smart manufacturing.

Lighthouse projects operate primarily in a single domain, where a meaningful (as evidenced by total market share) group of EU industries from the same sector can jointly provide a safe environment in which they make available a proportion of their data (or data streams) and demonstrate, on a large scale, the impact of big data technologies. Lighthouse projects used data sources other than those of the specific sector addressed, thereby contributing to breaking silos. In all cases, projects did enable access to appropriately large, complex and realistic datasets.

Projects needed to show sustainable impact beyond the specific large-scale demonstrators running through the project duration. Whenever possible, this was addressed by projects through solutions that could be replicated by other companies in the sector or by other application domains.

All Lighthouse projects were requested to involve all relevant stakeholders to reach their goals. This again did lead to the development of complete data ecosystems of the addressed domain or sector. Whenever this was appropriate, Lighthouse projects did rely on the infrastructure and ecosystems facilitated by one or more i-Spaces.

Some of the indicators that were used to assess the impact of Lighthouse projects have been the number and size of datasets processed (integrated), the number of data sources made available for use and analysis by third parties, and the number of

services provided for integrating data across sectors. Market indicators are obviously of utmost importance.

Key elements for the implementation of Lighthouse projects include at least the following areas.

The Use of Existing or Close-to-Market Technologies Lighthouses have not been expected to develop entirely new solutions; instead, they have been requested to make use of existing or close-to-market technologies and services by adding and/or adapting current relevant technologies, as well as accelerating the roll-out of big data value solutions using the Cloud and the IoT or HPC. Solutions should provide answers for real needs and requirements, showing an explicit knowledge of the demand side. Even though projects were asked to concentrate on solving concrete problems which again might easily lead to specific deployment challenges, the replicability of concepts was always a high priority to ensure impact beyond the particular deployments of the project. Lighthouse projects have been requested to address frameworks and tools from a holistic perspective, considering, for example, not only analytics but also the complete data value chain (data generation, the extension of data storing and analysis).

Interoperability and Openness All projects did take advantage of both closed and open data; during the project, they could determine if open source or proprietary solutions were the most suitable to address their challenges. However, it was always requested that projects promote the interoperability of solutions to avoid locking in customers.

The involvement of smaller actors (e.g. through opportunities for start-ups and entrepreneurs) who can compete in the same ecosystem in a fair way was always a must. For instance, open Application Programming Interfaces (APIs) had been identified as an important way forward (e.g. third-party innovation through data sharing). In addition, projects have been requested to focus on re-usability and ways to reduce possible barriers or gaps resulting from big data methods impacting end-users (break the ‘big data for data analysts only’ paradigm).

Performance All projects have been requested to contribute to common data collection systems and to have a measurement methodology in place. Performance monitoring was accomplished over at least two-thirds of the duration of the project.

The Setting Up of Ecosystems Lighthouse projects have a transformational power, that is, they had never been restricted to any type of narrow-minded experiments with limited impact. All projects demonstrated that they could improve (sometimes changing associated processes) the competitiveness of the selected industrial sector in a relevant way. To achieve this, the active involvement of different stakeholders is mandatory. For that reason, the supporting role of the ecosystem that enabled such changes is an important factor to keep in mind: All Lighthouse projects had been connected to communities of stakeholders from the design phase. Ecosystems evolved, extended or connected with existing networks of stakeholders and hubs, whenever this was possible.

As is well known, the European industry is characterised by a considerable number of small and medium-sized enterprises. Therefore, the adequate consideration of SME integration in the projects was always a central requirement to create a healthy environment.

Even though all projects had been requested to primarily focus on one particular sector, the use of data from different sources and industrial fields had always been encouraged, with priority given to avoiding the ‘silo’ effect.

Long-Term Commitment and Sustainability The budgets assigned to the projects have been envisioned as seeds for more widely implemented plans. All funded activities had been integrated into more ambitious strategies that allowed for the involvement of additional stakeholders and further funding (preferably private but also possibly a combination of public and private).

After the launch of the four initial Lighthouse projects, all learnings related to the concept of Lighthouse projects could be consolidated. As a result, a more advanced concept had been proposed including more concrete requirements for the upcoming large-scale pilots, in some cases further specifying aspects that had already been worked out. The following list served as guidance without the claim of completeness:

- It is important to reuse technologies and frameworks by combining and adapting relevant existing technologies (big data with the Cloud, HPC or IoT) that are already in the market or close to it (i.e. those with a high technology readiness level) to avoid the development of new platforms where a reasonable basis already exists (e.g. as part of the Open Source community). In addition, projects are especially encouraged to build on the technologies created by the ongoing projects of the Big Data PPP that fit their requirements (e.g. in the area of privacy-preserving technologies).
- Particular attention should be paid to interoperability. This applies to all layers of the solution, including data (here, some of the results of the projects funded under the Big Data PPP with a focus on data integration could be particularly useful), and to relevant efforts within the HPC, Cloud and IoT communities.
- It is expected that projects will combine the use of open and closed data. While it is understandable that some closed data will remain as such, we also expect these projects to contribute to the increasing availability of datasets that could be used by other stakeholders, such as SMEs and start-ups. This could happen under different regimes (not necessarily for free). Projects should declare how they will contribute to this objective by quantifying and qualifying datasets (when possible) and by including potential contributions to the ongoing data incubators/accelerators and Innovation Spaces.
- Lighthouse projects have to contribute to the horizontal activities of the Big Data PPP as a way of helping in the assessment of the PPP implementation and increasing its potential impact. Some of the targeted activities include contributing to the standardisation of activities, the measurement of KPIs, and coordination with the PPP branding, or active participation in training and educational activities proposed by the PPP.

4.3 *Technical Projects*

Technical projects focus on addressing one issue or a few specific aspects identified as part of the BDV technical priorities. In this way, technical projects provide the technology foundation for Lighthouse projects and i-Spaces. Technical projects may be implemented as Research and Innovation Actions (RIA) or Innovation Actions (IA), depending on the amount of research work required to address the respective technical priorities.

To identify the most important technical priorities to be addressed within these projects, the stakeholders within the data ecosystem had been engaged within a structured methodology to produce a set of consolidated cross-sectorial technical research requirements. The result of this process was the identification of five key technical research priorities (data management, data processing architectures, deep analytics, data protection and pseudonymisation, advanced visualisation and user experience) together with 28 sub-level challenges to delivering big data value (Zillner et al. 2017). Based on this analysis, the overall, strategic technical goal could be summarised as follows:

Deliver big data technology empowered by deep analytics for data-at-rest and data-in-motion, while providing data protection guarantees and optimised user experience, through sound engineering principles and tools for data-intensive systems.

Further details on the technical priorities and how they were defined are provided in Chap. “Technical Research Priorities for Big Data”. The Big Data Value Reference Model, which structures the technical priorities identified during the requirements analysis, is detailed in Chap. “A Reference Model for Big Data Technologies”.

4.4 *Platforms for Data Sharing*

Platform approaches have proved successful in many areas of technology (Gawer and Cusumano 2014), from supporting transactions among buyers and sellers in marketplaces (e.g. Amazon), to innovation platforms which provide a foundation on top of which to develop complementary products or services (e.g. Windows), to integrated platforms which are a combined transaction and innovation platform (e.g. Android and the Play Store).

The idea of large-scale “data” platforms has been touted as a possible next step to support data ecosystems (Curry and Sheth 2018). An ecosystem data platform would have to support continuous, coordinated data flows, seamlessly moving data among intelligent systems. The design of infrastructure to support data sharing and reuse is still an active area of research (Curry and Ojo 2020).

Data sharing and trading are seen as important ecosystem enablers in the data economy, although closed and personal data present particular challenges for the free

flow of data. The following two conceptual solutions – Industrial Data Platforms (IDP) and Personal Data Platforms (PDP) – introduce new approaches to addressing this particular need to regulate closed proprietary and personal data.

4.4.1 Industrial Data Platforms (IDP)

IDPs have increasingly been touted as potential catalysts for advancing the European Data Economy as a solution for emerging data markets, focusing on the need to offer secure and trusted data sharing to interested parties, primarily from the private sector (industrial implementations). The IDP conceptual solution is oriented towards proprietary (or closed) data, and its realisation should guarantee a trusted, secure environment within which participants can safely, and within a clear legal framework, monetise and exchange their data assets. A functional realisation of a continent-wide IDP promises to significantly reduce the existing barriers to a free flow of data within an advanced European Data Economy. The establishment of a trusted data-sharing environment will have a substantial impact on the data economy by incentivising the marketing and sharing of proprietary data assets (currently widely considered by the private sector as out of bounds) through guarantees for fair and safe financial compensations set out in black and white legal terms and obligations for both data owners and users. The ‘opening up’ of previously guarded private data can thus vastly increase its value by several orders of magnitude, boosting the data economy and enabling cross-sectorial applications that were previously unattainable or only possible following one-off bilateral agreements between parties over specific data assets.

The IDP conceptual solution complements the drive to establish BDVA i-Spaces by offering existing infrastructure and functional technical solutions that can better regulate data sharing within the innovation spaces. This includes better support for the secure sharing of proprietary or ‘closed’ data within the trusted i-Space environment. Moreover, i-Spaces offer a perfect testbed for validating existing implementations of conceptual solutions such as the IDP.

The identified possibilities for action can be categorised into two branches:

- **Standardisation:** Addressing the lack of an existing standard platform (technical solution) that limits stakeholders from participating in the European Digital Single Market, and the availability of clear governance models (reference models, guidelines and best practices) regulating the secure and trusted exchange of proprietary data.
- **Implementation:** Establishing, developing or aligning existing IDP implementations to provide a functional European-wide infrastructure within which industrial participants can safely, and within a clear legal framework, monetise and exchange data assets.

Standardisation activities outlined by the Strategic Research and Innovation Agenda (SRIA) (Zillner et al. 2017) and in Chap. “Recognition of Formal and Non-formal Training in Data Science” have taken into account the need to

accommodate activities related to the evolving IDP solutions. The opportunity to drive forward emerging standards also covers the harmonisation of reference architectures and governance models put forward by the community. Notable advanced contributions in this direction include the highly relevant white paper and the reference architecture² provided by the Industrial Data Space (IDS) Association. The Layered Databus, introduced by the Industrial Internet Consortium,³ is another emerging standard advocating the need for data-centric information-sharing technology that enables data market players to exchange data within a virtual and global data space.

The implementation of IDPs needs to be approached on a European level, and existing and planned EU-wide, national and regional platform development activities could contribute to these efforts. The industries behind existing IDP implementations, including the IDS reference architecture and other examples such as the MindSphere Open Industrial Cloud Platform,⁴ can be approached to move towards a functional European Industrial Data Platform. The technical priorities outlined by the SRIA (Zillner et al. 2017), particularly the Data Management priority, need to address data management across a data ecosystem comprising both open and closed data. The broadening of the scope of data management is also reflected in the latest BDVA reference model, which includes an allusion to the establishment of a digital platform whereby marketplaces regulate the exchange of proprietary data.

4.4.2 Personal Data Platforms (PDP)

So far, consumers have trusted companies, including Google, Amazon, Facebook, Apple and Microsoft, to aggregate and use their personal data in return for free services. While EU legislation, through directives such as the Data Protection Directive (1995) and the ePrivacy Directive (1998), has ensured that personal data can only be processed lawfully and for legitimate use, the limited user control offered by such companies and their abuse of a lack of transparency have undermined consumers' trust. In particular consumers experience everyday leakage of their data, traded by large aggregators in the marketing networks for value only returned to consumers in the form of often unwanted digital advertisements. This has recently led to a growth in the number of consumers adopting adblockers to protect

²Reference Architecture Model for the Industrial Data Space, April 2017, https://www.fraunhofer.de/content/dam/zv/de/Forschungsfelder/industrial-data-space/Industrial-Data-Space_Reference-Architecture-Model-2017.pdf

³The Industrial Internet of Things, Volume G1: Reference Architecture, January 2017, https://www.iiconsortium.org/IIC_PUB_G1_V1.80_2017-01-31.pdf

⁴*MindSphere: The cloud-based, open IoT operating system for digital transformation*, Siemens, 2017, https://www.plm.automation.siemens.com/media/global/en/Siemens_MindSphere_Whitepaper_tcm27-9395.pdf

their digital life,⁵ while at the same time they are becoming more conscious of and suspicious about their personal data trail.

In order to address this growing distrust, the concept of Personal Data Platforms (PDP) has emerged as a possible solution that could allow data subjects and data owners to remain in control of their data and its subsequent use.⁶ PDPs leverage ‘the concept of user-controlled cloud-based technologies for storage and use of personal data (“personal data spaces”).’⁷ However, so far consumers have only been able to store and control access to a limited set of personal data, mainly by connecting their social media profiles to a variety of emerging Personal Information Management Systems (PIMS). More successful (but limited in number) uses of PDPs have involved the support of large organisations in agreeing to their customers accumulating data in their own self-controlled spaces. The expectation here is the reduction of their liability in securing such data and the opportunity to access and combine them with other data that individuals will import and accumulate from other aggregators. However, a degree of friction and the lack of a successful business model are still hindering the potential of the PDP approach.

A new driver behind such a self-managed personal data economy has recently started to appear. As a result of consumers’ growing distrust, measures such as the General Data Protection Regulation (GDPR), which has been in force since May 2018, have emerged. The GDPR constitutes the single pan-European law on data protection, and, among other provisions and backed by the risk of incurring high fines, it will force all companies dealing with European consumers to (1) increase transparency and (2) provide users with granular control for data access and sharing and will (3) guarantee consumers a set of fundamental individual digital rights (including the right to rectification, erasure, data portability and to restrict processing). In particular, by representing a threat to the multi-billion euro advertising business, we expect individuals’ data portability right, as enshrined in the GDPR, to be the driver for large data aggregators to explore new business models for personal data access. As a result, this will create new opportunities for PDPs to emerge. The rise of PDPs and the creation of more decentralised personal datasets will also open up new opportunities for SMEs that might benefit from and

⁵Used by 615 million devices at the end of 2016, <http://uk.businessinsider.com/pagefair-2017-ad-blocking-report-2017-1?r=US&IR=T>

⁶See a Commission paper on ‘Personal information management services – current state of service offers and challenges’ analysing feedback from public consultation: <https://ec.europa.eu/digital-single-market/en/news/emerging-offer-personal-information-management-services-current-state-service-offers-and>

⁷A Personal Data Space is a concept, framework and architectural implementation that enables individuals to gather, store, update, correct, analyse and/or share personal data. This is also a marked deviation from the existing environment where distributed data is stored throughout organisations and companies internally, with limited to no access or control from the user that the information concerns. This is a move away from the B2B (business to business) and B2C (business to consumer) models, with a move towards Me2B – when individuals start collecting and using data for their own purposes and sharing data with other parties (including companies) under their control (<https://www.ctrl-shift.co.uk/news/2016/09/19/shifting-from-b2c-to-me2b/>).

investigate new secondary uses of such data, by gaining access to them from user-controlled personal data stores – a privilege so far available only to large data aggregators. However, further debate is required to reach an understanding on the best business models (for demand and supply) to develop a marketplace for personal data donors, and on what mechanisms are required to demonstrate transparency and distribute rewards to personal data donors. Furthermore, the challenges organisations face in accessing expensive data storage, and the difficulties in sharing data with commercial and international partners due to the existence of data platforms which are considered to be unsafe, need to be taken into account. Last but not least, questions around data portability and interoperability also have to be addressed.

4.5 Cooperation and Coordination Projects

Cooperation and coordination projects aimed to work on detailed activities that ensured coordination and coherence in the PPP implementation and provided support to activities. The portfolio of support activities comprised support actions that addressed complementary, non-technical issues alongside the European Innovation Spaces, Lighthouse projects, data platforms, and research and innovation activities. In addition to the activities addressed, the governance of the data ecosystem, cooperation and coordination activities focused on the following.

Skills Development The educational support for data strategists and data engineers needs to meet industry requirements. The next generation of data professionals needs this wider view to deliver the data-driven organisation of the future. Skill development requirements need to be identified that can be addressed by collaborating with higher education institutes, education providers and industry to support the establishment of:

- New educational programmes based on interdisciplinary curricula with a clear focus on high-impact application domains
- Professional courses to educate and re-skill/up-skill the current workforce with the specialised skillsets needed to be data-intensive engineers, data scientists and data-intensive business experts
- Foundational modules in data science, statistical techniques and data management within related disciplines such as law and the humanities
- A network between scientists (academia) and industry that leverages Innovation Spaces to foster the exchange of ideas and challenges
- Datasets and infrastructure resources provided by industry that enhance the industrial relevance of courses.

Business Models and Ecosystems The big data value ecosystem will comprise many new stakeholders and will require a valid and sustainable business model. Dedicated activities for investigating and evaluating business models will be

connected to the innovation spaces where suppliers and users will meet. These activities include:

- Delivering means for the systematic analysis of data-driven business opportunities
- Establishing a mapping of technology providers and their value contribution
- Identifying mechanisms by which data value is determined and value is established
- Providing a platform for data entrepreneurs and financial actors, including venture capitalists, to identify appropriate levels of value chain understanding
- Describing and validating business models that can be successful and sustainable in the future data-driven economy

Policy and Regulation The stakeholders of the data ecosystem need to contribute to the policy and regulatory debate about non-technical aspects of the future big data value creation as part of the data-driven economy. Dedicated activities addressed the aspects of data governance and usage, data protection and privacy, security, liability, cybercrime, and Intellectual Property Rights (IPR). These activities enabled the exchange between stakeholders from industry, end-users, citizens and society to develop input to ongoing policy debates where appropriate. Of equal importance was the identification of concrete legal problems for actors in the Value Chain, particularly SMEs that have limited legal resources. The established body of knowledge on legal issues was of high value for the wider community.

Social Perceptions and Societal Implications Societal challenges cover a wide range of topics including trust, privacy, ethics, transparency, inclusion efficacy, manageability and acceptability in big data innovations. There needs to be a common understanding in the technical community leading to an operational and validated method that applies to data-driven innovations development. At the same time, it is critical to develop a better understanding of inclusion and collective awareness aspects of big data innovations that enable a clear profile of the social benefits provided by big data value technology. By addressing the listed topics, the PPP ensured that citizens' views and perceptions were taken into account so that technology and applications were developed with a chance to be widely accepted.

5 Roadmap for Adoption of Big Data Value

The roadmap ensured and guided the development of the ecosystem in distinct phases, each with a primary theme. The three phases, as depicted in Fig. 1, are as follows:

- **Phase I:** Establish the ecosystem (governance, i-Spaces, education, enablers) and demonstrate the value of existing technology in high-impact sectors (Light-houses, technical projects)

Fig. 1 Three-phase timeline of the adoption of Big Data Value PPP

- **Phase II:** Pioneer disruptive new forms of big data value solutions (Lighthouses and technical projects) in high-impact domains of importance for EU industry, addressing emerging challenges of the data economy
- **Phase III:** Develop long-term ecosystem enablers to maximise sustainability for economic and societal benefit, including the establishment of data platforms

Phase I: Establish an Innovation Ecosystem

The first phase of the roadmap focused on laying the foundations necessary to establish a sustainable European data innovation ecosystem. The key activities of Phase I included:

- Establishing a European network of i-Spaces for cross-sectorial and cross-lingual data integration, experimentation and incubation
- Demonstrating big data value solutions via large-scale pilot projects in domains of strategic importance for EU industry, using existing technologies or very near-to-market technologies.
- Tackling the main technology challenges of the data economy by improving the technology, methods, standards and processes for big data value
- Advancing state-of-the-art in privacy-preserving big data technologies and exploring the societal and ethical implications
- Establishing key ecosystem enablers, including support and coordination structures for industry skills and benchmarking.

Phase II: Disruptive Big Data Value

Building on the foundations established in Phase I, the second phase had a primary focus on Research and Innovation (R&I) activities to deliver the next generation of big data value solutions. The key activities of Phase II included:

- Supporting the emergence of the data economy with a particular focus on accelerating the progress of SMEs, start-ups and entrepreneurs, as well as best practices and standardisation
- Pioneering disruptive new forms of big data value solutions with the Cloud, HPC or IoT technologies via large-scale pilot projects in emerging domains of importance for EU industry using advanced platforms, tools and testbeds
- Tackling the next generation of big data research and innovation challenges for extreme-scale analytics
- Addressing ecosystem roadblocks and inhibitors to the take-up of big data value platforms for data ecosystem viability, including platforms for personal and industrial data
- Providing (continuing) support, facilitating networking and cooperation among ecosystem actors and projects, and promoting community building among BDV, Cloud, HPC and IoT activities.

Phase III: Long-Term Ecosystem Enablers

While the sustainability of the ecosystem has been considered from the start of the PPP, the third phase had a specific focus on activities that could ensure long-term self-sustainability. The key activities of Phase III included:

- Sowing the seeds for long-term ecosystems enablers to ensure self-sustainability
- Creating innovation projects within a federation of i-Spaces (European Digital Innovation Hubs for Big Data) to validate and incubate innovative big data value solutions and business models
- Ensuring continued support for technology outputs of PPP (Lighthouse projects, R&I, CSA), including non-technical aspects (training and Open Source Community, Technology Foundation)
- Establishing a Foundation for European Innovation Spaces with a charter to continue collaborative innovation activity, in line with the concept of European Digital Innovation Hub for Big Data
- Liaising with private funding (including Venture Capital) to accelerate entry into the market and socio-economic impacts, including the provision of ancillary services to develop investment-ready proposals and support scaling for BDV PPP start-ups and SMEs to reach the market
- Tackling the necessary strategy and planning for the BDV ecosystem until 2030, including the identification of new stakeholders, emerging usage domains, technology, business and policy roadmapping activity.

6 European Data Value Ecosystem Development

Developing the European Data Value Ecosystem is at the core of the mission and strategic priorities of the Big Data Value Association and the Big Data Value PPP. The European Data Value Ecosystem brings together communities (all the different stakeholders who are involved, affected or stand to benefit), technology, solutions

and data platforms, experimentation, incubation and know-how resources, and the business models and framework conditions for the data economy. In this section, we refer to the ‘community’ and stakeholder aspect of the European big data value ecosystem (see Fig. 2).

A dimension to emphasise in the European Data Value Ecosystem is its twofold nature of vertical versus horizontal in respect to the different sector or application domains (transport health, energy, etc.). While specific data value ecosystems are needed per sector (concerning targeted markets, stakeholders, regulations, type of users, data types, challenges, etc.), one of the main values identified for the Big Data Value Association and the PPP is its horizontal nature, allowing cross-sector value creation, considering both the reuse of value from one sector to another, and the creation of innovations based on cross-sector solutions and consequently new value chains.

Establishing collaborations with other European, international and local organisations is crucial for the development of the ecosystem, to generate synergies between communities and to impact research and innovation, standards, regulations, markets and society.

Collaborations, in particular with other PPPs, European and international standardisation bodies, industrial technology platforms, data-driven research and innovation initiatives, user organisations and policymakers, had been identified and developed at national, European and international level since the launch of the PPP and the creation of the Association, influencing the level of maturity of these collaborations.

A key part of ensuring the sustainability of the BDV ecosystem was to develop collaborations with complementary ecosystems with an impact on technology integration and the digitisation of industry challenges. These collaborations, detailed in Fig. 2, include the ETP4HPC (European Technology Platform for HPC) (for HPC),

Fig. 2 Map of collaboration for BDV ecosystem

ECSO (for cybersecurity), AIOTI (for IoT), 5G (through 5G PPP), the European Open Science Cloud (EOSC) (for the Cloud) and the European Factories of the Future Research Association (EFFRA) (for factories of the future).

7 Summary

Creating a productive ecosystem for big data and driving accelerated adoption requires an interdisciplinary approach addressing a wide range of challenges from access to data and infrastructure, to technical barriers, skills, and policy and regulation. To overcome these challenges, collective action is needed from all stakeholders working together in an effective, holistic and coherent manner. To this end, the Big Data Value Public-Private Partnership was established to develop the European data ecosystem and enable data-driven digital transformation, delivering maximum economic and societal benefit, and achieving and sustaining Europe's leadership in the fields of big data value creation and Artificial Intelligence. The BDV PPP follows a phased roadmap with the use of five strategic mechanisms to drive the adoption of big data value and to encourage cooperation and coordination in the data ecosystem. The PPP proactively engaged with the key communities, which helped to enhance the development of the European Data Value Ecosystem.

Acknowledgements We greatly acknowledge the collective effort of the SRIA teams: Carlos A. Iglesias, Antonio Alfaro, Jesus Angel, Sören Auer, Paolo Bellavista, Arne Berre, Freek Bomhof, Stuart Campbell, Geraud Canet, Giuseppa Caruso, Paul Czech, Stefano de Panfilis, Thomas Delavallade, Marija Despenic, Wolfgang Gerteis, Aris Gkoulalas-Divanis, Nuria Gomez, Paolo Gonzales, Thomas Hahn, Souleiman Hasan, Bjarne Kjær Ersbøll, Bas Kotterink, Yannick Legré, Yves Mabiala, Julie Marguerite, Ernestina Menasalves, Andreas Metzger, Elisa Molino, Thierry Nagellen, Dalit Naor, Maria Perez, Milan Petkovic, Roberta Piscitelli, Klaus-Dieter Platte, Pierre Pleven, Dumitru Roman, Titi Roman, Alexandra Rosén, Nikos Sarris, Stefano Scamuzzo, Simon Scerri, Corinna Schulze, Bjørn Skjellaug, Francois Troussier, Colin Upstill, Josef Urban, Meilof Veeningen, Tonny Velin, Ray Walshe, Walter Waterfeld and Stefan Wrobel.

References

- Cavanillas, J. M., Curry, E., & Wahlster, W. (Eds.). (2016). *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*. New York: Springer. <https://doi.org/10.1007/978-3-319-21569-3>
- Curry, E. (2016). The big data value chain: Definitions, concepts, and theoretical approaches. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New horizons for a data-driven economy: A roadmap for usage and exploitation of big data in Europe*. New York: Springer. https://doi.org/10.1007/978-3-319-21569-3_3
- Curry, E., & Ojo, A. (2020). Enabling knowledge flows in an intelligent systems data ecosystem. In *Real-time linked dataspace* (pp. 15–43). https://doi.org/10.1007/978-3-030-29665-0_2

- Curry, E., & Sheth, A. (2018). Next-generation smart environments: From system of systems to data ecosystems. *IEEE Intelligent Systems*, 33(3), 69–76. <https://doi.org/10.1109/MIS.2018.033001418>
- Gawer, A., & Cusumano, M. A. (2014). Industry platforms and ecosystem innovation. *Journal of Product Innovation Management*, 31(3), 417–433. <https://doi.org/10.1111/jpim.12105>
- Zillner, Sonja, Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.). (2017). *European big data value strategic research & innovation agenda*. Retrieved from Big Data Value Associate website www.bdva.eu
- Zillner, S., Bisset, D., Milano, M., Curry, E., Hahn, T., Lafrenz, R., et al. (2020). *Strategic research, innovation and deployment agenda – AI, data and robotics partnership. Third Release (3rd)*. Brussels: BDVA, euRobotics, ELLIS, EurAI and CLAIRE.

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.