Chapter 2
The BIG Project

Edward Curry, Tilman Becker, Ricard Munné, Nuria De Lama, and Sonja Zillner

2.1 Introduction

The Big Data Public Private Forum (BIG) Project (http://www.big-project.eu/) was an EU coordination and support action to provide a roadmap for big data within Europe. The BIG project worked towards the definition and implementation of a clear big data strategy that tackled the necessary activities needed in research and innovation, technology adoption, and the required support from the European Commission necessary for the successful implementation of the big data economy. As part of this strategy, the outcomes of the project were used as input for Horizon 2020.

E. Curry (✉)
Insight Centre for Data Analytics, National University of Ireland Galway, Lower Dangan, Galway, Ireland
e-mail: edward.curry@insight-centre.org

T. Becker
German Research Centre for Artificial Intelligence (DFKI), Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany
e-mail: tilman.becker@dfki.de

R. Munné
Atos Spain, S.A., Av. Diagonal, 200, 08018 Barcelona, Spain
e-mail: ricard.munne@atos.net

N. De Lama
Atos Spain, S.A., Albarracín, 25, 28037 Madrid, Spain
e-mail: nuria.delama@atos.net

S. Zillner
Corporate Technology, Siemens AG, Munich, Germany

School of International Business and Entrepreneurship, Steinbeis University, Berlin, Germany
e-mail: sonja.zillner@siemens.com

J.M. Cavanillas et al. (eds.), *New Horizons for a Data-Driven Economy*,
DOI 10.1007/978-3-319-21569-3_2

Foundational research technologies and innovative sectorial applications were analysed and assessed in the BIG project in order to create technology and strategy roadmaps so that business and operational communities understand the potential of big data technologies and are enabled to implement appropriate strategies and technologies for commercial benefit.

This chapter provides an overview of the BIG project detailing the project's mission and strategic objectives. The chapter describes the partners within the consortium and the overall structure of the project work. The three-phase methodology used in the project is described, including details on the techniques used within the technical working groups, sectorial forms, and road mapping activity. Finally, the project's role in setting up the Horizon 2020 Big Data Value contractual Public Private Partnership and Big Data Value Association is discussed.

2.2 Project Mission

In order to realize the vision of a data-driven society in 2020, Europe has to prepare the right ecosystem around big data. Public and private organizations need to have the necessary infrastructures and technologies to deal with the complexity of big data, but should also be able to use data to maximize their competitiveness and deliver business value.

Building an industrial community around big data in Europe was a key priority of the BIG project, together with setting up the necessary collaboration and dissemination infrastructure to link technology suppliers, integrators, and leading user organizations. The BIG project (from now on referred to as BIG) worked towards the definition and implementation of a strategy that includes research and innovation, but also technology adoption. The establishment of the community together with adequate resources to work at all levels (technical, business, political, etc.) is the basis for a long-term European strategy. Convinced that a strong reaction is needed, BIG defined its mission accordingly:

> The mission of BIG is setting up an ecosystem that will bring together all the relevant stakeholders needed to materialize a data-driven society in 2020. This ecosystem will ensure that Europe plays a leading role in the definition of the new context by building the necessary infrastructures and technologies, generating a suitable innovation space where all organizations benefit from data, and provides a pan-European framework to coherently address policy, regulatory, legal, and security barriers.

The BIG mission was broken down into a number of specific strategic objectives for the project.

2.3 Strategic Objectives

In September 2012, the project identified a set of strategic objectives to ensure it delivered on its mission. The specific objectives were:

- **BIG will set up an industrial-led initiative around Intelligent Information Management and Big Data to contribute to EU competitiveness and position it in Horizon 2020:** Industrial leadership will guide actions towards real business benefits, but will be complemented by the views of academia and research organizations, which will also take part in this endeavour. The project will take a long-term approach to represent the views and interests of IIM stakeholders, with a special focus on big data due to its relevance in the current and future context. Decisions such as establishing it as a legal entity will be considered, and potential mergers with relevant associations at the EU level will also be envisaged for the sake of sustainability and impact.
- **BIG will elaborate an integrated roadmap that takes into consideration technical, business, policy, and society aspects,** focusing not only on pure technical issues, but also establishing priorities based on expected impact. The BIG consortium will engage the necessary expertise to ensure contributions not only from project partners, but also from a wider community comprised of experts in relevant technical domains as well as experts in sectors or application domains where the use of these technologies is expected to produce a high impact.
- **BIG will ensure that technical research areas selected by the project cover the needs expressed by the industry in different application domains:** For this to happen, a sharp understanding is needed of how big data can be applied within industrial sectors. This understanding needs to be transmitted to domain experts to establish a clear path for the adoption of the technology in each of the selected sectors.
- **BIG will promote adoption of earlier waves of big data technology:** Instead of adopting only a futuristic approach, BIG will use as a starting point those technologies that are already in place. The objective is to reach a clear understanding of the level of maturity of different technical solutions as well as the feasibility of their implementation. This will be valuable information with respect to the state of the art and will be used as input for the elaboration of both the sectorial and the integrated roadmaps.
- **BIG will define and promote actions dealing with policy and regulation,** including aspects such as data security, intellectual property, privacy, liability, and data access. BIG will contribute to the entire ecosystem related to big data implementation without restricting its activities to only technical issues.
- **BIG will carry out dissemination actions targeting different stakeholders and players in the value chain:** Dissemination actions will be customized to the different communities (e.g. technical experts, data scientists, technical managers, business managers, and executives in both Multinational Corporation (MNC) and Small and Medium-sized Enterprises (SMEs)). BIG addressed all

the relevant communities with an ambitious strategy including presence in mass media, relevant conferences, organization of workshops and events, and maximization of the use of web channels.

- All this will not have been possible without **providing the right collaboration infrastructures**. Collaboration among projects, but also many discussions between all the relevant stakeholders and actors in the value chain, including major industrial organizations in the EU landscape, will take place. Bearing this in mind BIG set up and maintained a support infrastructure that will enable collaboration, information sharing, and customization of actions toward different targeted audiences.

2.4 Consortium

The participants of the BIG consortium (illustrated in Fig. 2.1) were carefully selected to include key players with complementary skills in both industry and academia. Each of the project partners had experience in cutting-edge European projects and significant connections to key stakeholders in the big data marketplace. The academic partners using their expert knowledge in the field lead the technical investigations of big data technology. The industrial partners were well positioned in their knowledge of large-scale data management products and services and their application within different industrial sectors.

The partners of the BIG consortium were:

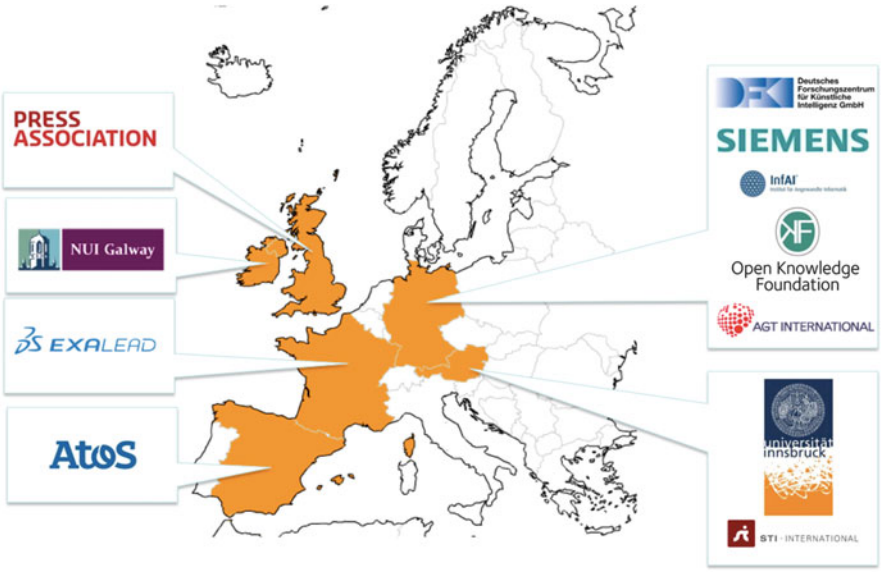

Fig. 2.1 BIG project consortium members

- **Industry:** Atos, Press Association (PA), Siemens, AGT International, Exalead, and the Open Knowledge Foundation (OKF)
- **Academia:** University of Innsbruck (UIBK), National University of Ireland Galway (NUIG), University of Leipzig, German Research Centre for Artificial Intelligence (DFKI), and STI International

2.5 Stakeholder Engagement

Essential for the success of a large-scale, cross-fertilization, and broad road mapping effort is the involvement of a large fraction of the community and industry, not only from the point of view of technology provision but also technology adoption. The project took an inclusive approach to stakeholder engagement and actively solicited inputs from the wider community composed of experts in technical domains as well as experts in business sectors. An open philosophy was applied to all the documents generated by the project, which were made public to the wider community for active contribution and content validation. The project held stakeholder workshops to engage the community within the project. The first workshop was held at the European Data Forum (EDF) 2013 in Dublin to announce the project to the community and gather participants. The second workshop took place at EDF 2014 in Athens to present the interim results of the project for feedback and further validation with stakeholders. Over the duration of the project a number of well-attended sector-specific workshops were held to gather needs and validate findings. At the end of the project a final workshop was convened to present the results of the project in October 2014 in Heidelberg.

2.6 Project Structure

The work of the BIG project was split into groups focusing on industrial sectors and technical areas. The project structure comprised of sectorial forums and technical working groups.

Sectorial forums examined how big data technologies can enable business innovation and transformation within different sectors. The sector forums were led by the industrial partners of the project. Their objective was to gather big data requirements from vertical industrial sectors, including health, public sector, finance, insurance, telecoms, media, entertainment, manufacturing, retail, energy, and transport (see Fig. 2.2).

Technical working groups focused on big data technologies for each activity in the data value chain to examine their capabilities, level of maturity, clarity, understandability, and suitability for implementation. The technical groups (see Fig. 2.3) were led by the academic partners in BIG and examined emerging technological and research trends for coping with big data.

Fig. 2.2 Sectorial forums within the BIG project

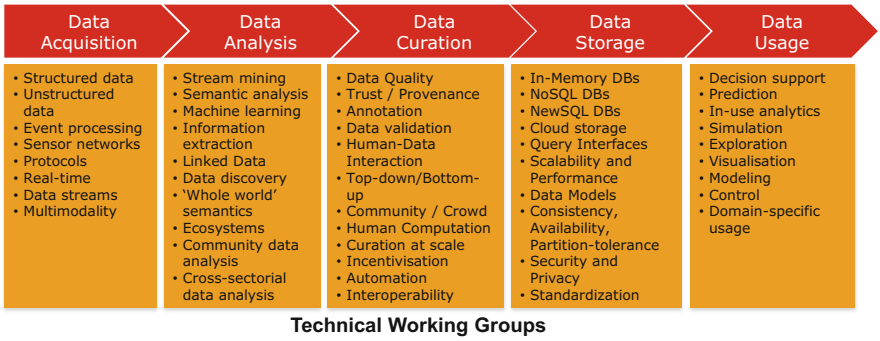

Fig. 2.3 Technical working groups within the BIG project

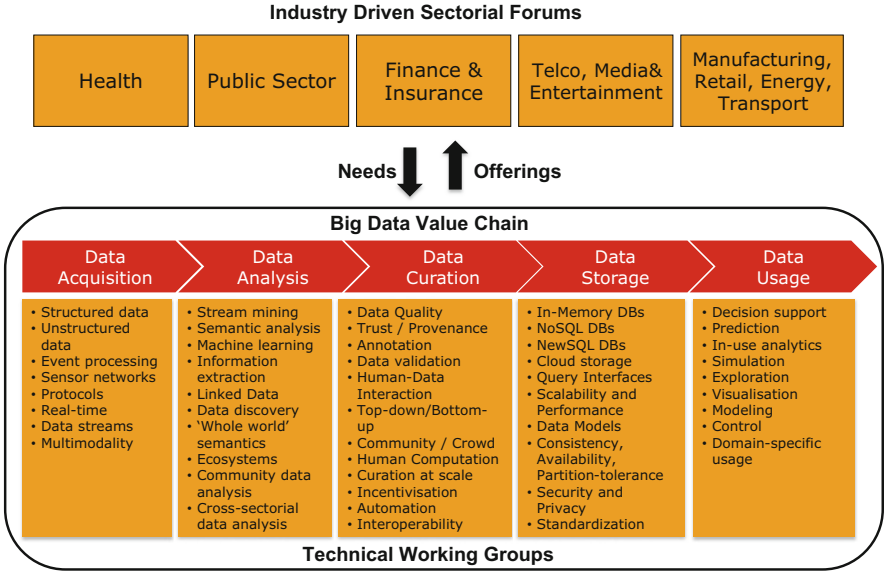

Fig. 2.4 The BIG project structure

As illustrated in Fig. 2.4, the needs identified by sector forums were used to understand the maturity and gaps in the capability offered by current big data technology. This analysis provided a clear picture on the limitations and expectations regarding big data technology deployment. The outputs of the analysis were

used to produce a series of consensus-reflecting roadmaps that defined priorities and actions needed for big data in each sector.

2.7 Methodology

From an operational point of view, BIG defined a set of activities based on a three-phase approach as illustrated in Fig. 2.5. The three phases were:

1. Technology state of the art and sector analysis
2. Roadmapping activity
3. Big data public private partnership

2.7.1 Technology State of the Art and Sector Analysis

In the first phase of the project, the sectorial forums and the technical working groups performed a parallel investigation in order to identify:

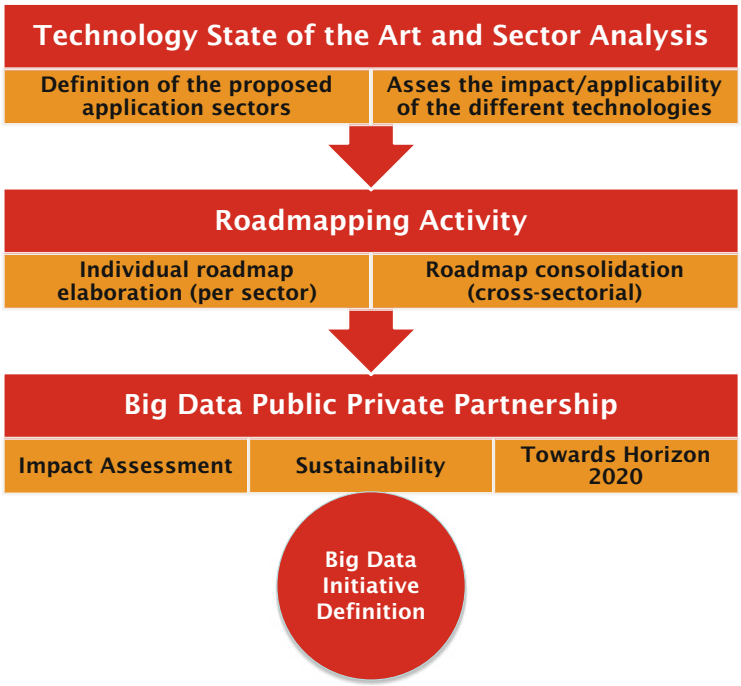

Fig. 2.5 Three-phase methodology of BIG

- Sectorial needs and requirements gathered from different stakeholders
- The state of the art of big data technologies as well as identifying research challenges

As part of the investigation, application sectors expressed their needs with respect to the technology as well as possible limitations and expectations regarding its current and future deployment.

Using the results of the investigation a gap analysis was performed between what technology capability was ready, with the sectorial expectations of what technological capability was currently required together with future requirements. The analysis produced a series of consensus-reflecting sectorial roadmaps that defined priorities and actions to guide further steps in big data research.

2.7.1.1 Technical Working Groups

The goal of the technical working groups was to investigate the state of the art in big data technologies to determine its level of maturity, clarity, understandability, and suitability for implementation. To allow for an extensive investigation and detailed mapping of developments, the technical working groups deployed a combination of a top-down and bottom-up approach, with a focus on the latter. The approach of the working groups was based on a 4-step approach: (1) literature research, (2) subject matter expert interviews, (3) stakeholder workshops, and (4) technical survey.

In the first step each technical working group performed a systematic literature review based on the following activities:

- Identification of relevant type and sources of information
- Analysis of key information in each source
- Identification of key topics for each technical working group
- Identification of the key subject matter experts for each topic as potential interview candidates
- Synthesizing the key message of each data source into state-of-the art descriptions for each identified topic

The experts within the consortium outlined the initial starting points for each technical area, and the topics were expanded through the literature search and from the subject matter expert interviews.

The following types of data sources were used: scientific papers published in workshops, symposia, conferences, journals and magazines, company white papers, technology vendor websites, open source projects, online magazines, analysts' data, web blogs, other online sources, and interviews conducted by the BIG consortium. The groups focused on sources that mention concrete technologies and analysed them with respect to their values and benefits.

The synthesis step compared the key messages and extracted agreed views that were then summarized in the technical white papers. Topics were prioritized based

on the degree to which they are able to address business needs as identified by the sectorial forum working groups.

The literature survey was complemented with a series of interviews with subject matter experts for relevant topic areas. Subject matter expert interviews are a technique well suited to data collection and particularly for exploratory research because it allows expansive discussions that illuminate factors of importance (Oppenheim 1992; Yin 2009). The information gathered is likely to be more accurate than information collected by other methods since the interviewer can avoid inaccurate or incomplete answers by explaining the questions to the interviewee (Oppenheim 1992).

The interviews followed a semi-structured protocol. The topics of the interview covered different aspects of big data, with a focus on:

- Goals of big data technology
- Beneficiaries of big data technology
- Drivers and barriers for big data technologies
- Technology and standards for big data technologies

An initial set of interviewees was identified from the literature survey, contacts within the consortium, and a wider search of the big data ecosystem. Interviewees were selected to be representative of the different stakeholders within the big data ecosystem. The selection of interviewees covered (1) established providers of big data technology (typically MNCs), (2) innovative sectorial players who are successful at leveraging big data, (3) new and emerging SMEs in the big data space, and (4) world leading academic authorities in technical areas related to the Big Data Value Chain.

2.7.1.2 Sectorial Forums

The overall objective of the sectorial forums was to acquire a deep understanding of how big data technology can be used in the various industrial sectors, such as healthcare, public, finance and insurance, and media.

In order to identify the user needs and industrial requisites of each domain, the sectorial forums followed a research methodology encompassing the following three steps as illustrated in Fig. 2.6. For each industrial sector, the steps were accomplished separately. However, in the case where sectors were related (such as energy and transport) the results have been merged for those sectors in order to highlight differences and similarities.

The aim of the first steps was to identify both stakeholders and use cases for big data applications within the different sectors. Therefore, a survey was conducted including scientific reviews, market studies, and other Internet sources. This knowledge allowed the sectorial forums to identify and select potential interview partners and guided the development of the questionnaire for the domain expert interviews.

The questionnaire consisted of up to 12 questions that were clustered into three parts:

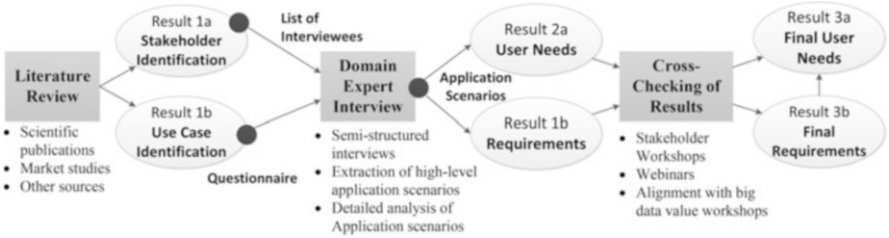

Fig. 2.6 The three steps of the sectorial forums research methodology

- Direct inquiry of specific user needs
- Indirect evaluation of user needs by discussing the relevance of the use cases identified at Step 1 as well as any other big data applications of which they were aware
- Reviewing constraints that need to be addressed in order to foster the implementation of big data applications in each sector

In the second step, semi-structured interviews were conducted using the developed questionnaire. At least one representative of each stakeholder group identified in Step 1 was interviewed. To derive the user needs from the collected material, the most relevant and frequently mentioned use cases were aggregated into high-level *application scenarios*. The data collection and analysis strategy was inspired by the triangulation approach (Flick 2004). Reviewing and quantitatively assessing the high-level application scenarios derived a reliable analysis of user needs. Examinations of the likely constraints of big data applications helped to identify the relevant requirements that needed to be addressed.

The third step involved a crosscheck and validation of the initial results of the first two steps by involving stakeholders of the domain. Some sectors conducted dedicated workshops and webinars with industrial stakeholders to discuss and review the outcomes. The results of the workshops were studied and integrated whenever appropriate.

2.7.2 Cross-Sectorial Roadmapping

Comparison among the different sectors enabled the identification of commonalities and differences at multiple levels, including technical, policy, business, and regulatory. The analysis was used to define an integrated cross-sectorial roadmap that provides a coherent holistic view of the big data domain. The cross-sectorial big data roadmap was defined using the following three steps:

1. **Consolidation** to establish a common understanding of requirements as well as technology descriptions and terms used across domains
2. **Mapping** to identify any technologies needed to address the identified cross-sector requirements

3. **Temporal alignment** to highlight which technologies need to be available at what point in time by incorporating the estimated adoption rate by the involved stakeholders

The remainder of this section describes each of these steps in more detail.

2.7.2.1 Consolidation

Alignment among the technical working groups, and between the technical working groups and the sectorial forums, was important and facilitated through early exchange of drafts, one-on-one meetings, and the collection of consolidated requirements through the SFs. In order to align the sector-specific labelling of requirements, a consolidated description was established. In doing so, each sector provided their requirements with the associated user needs. In dedicated meetings, similar and related requirements were clustered and then merged, aligned, or restructured. Thus, the initial list of 13 high-level requirements and 28 sub-level requirements could be reduced to 8 high-level requirements and 25 sub-level requirements. In summary the consolidation phase reduced the total number of requirements by 20 %.

2.7.2.2 Mapping

For mapping technology to requirements the technical working groups indicated which technology could be used to address the consolidated requirements. Besides providing a mapping between requirements and technologies, the technical working groups also indicated the associated research challenges.

Within a 1-day workshop, the initial mapping of technologies and requirements was consolidated in two steps. First, the indicated technological capabilities were analysed in further detail by describing how the sector-specific aspects of each cross-sector requirement can be handled. Second, for each cross-sector requirement it was investigated whether the technologies from various technical working groups need to be combined in order to address the full scope of the requirement. At the end of the discussion, any technologies that were requested by at least two sectors were included into the cross-sector roadmap.

2.7.2.3 Temporal Alignment

After identifying the key technologies, their temporal alignment needed to be defined. This was achieved by answering two questions:

- How long is the development time of the technology?
- When will the stakeholder involved adopt the technology?

The development time for each technology indicates how much time is needed to solve the associated research challenges. This time frame depends on the technical complexity of the challenge together with the extent to which sector-specific extensions are needed. In order to determine the adoption rate of big data technology (or the associated use case) non-technical requirements such as availability of business cases, suitable incentive structures, legal frameworks, potential benefits, as well as the total cost for all the stakeholders involved (Adner 2012) were considered.

2.8 Big Data Public Private Partnership

The Big Data Public Private Forum, as it was initially called, was intended to create the path towards implementation of the roadmaps. The path required two major elements: (1) a mechanism to include content of the roadmaps into real agendas supported by the necessary resources (economic investment of both public and private stakeholders) and (2) a community interested in the topics and committed to making the investment and collaborating towards the implementation of the agendas.

The BIG consortium was convinced that achieving this result would require creating a broad awareness and commitment outside the project. BIG took the necessary steps to contact major players and to liaise with the NESSI European Technology Platform to jointly work towards this endeavour. The collaboration was set up in the summer of 2013 and allowed the BIG partners to establish the necessary high-level connections at both industrial and political levels. This collaboration led to the following outcomes:

- **The Strategic Research & Innovation Agenda (SRIA) on Big Data Value** that was initially fed by the BIG technical papers and roadmaps and has been extended with the input of a public consultation that included hundreds of additional stakeholders representing both the supply and the demand side.
- **A cPPP (contractual PPP) proposal** as the formal step to set up a PPP on Big Data Value. The cPPP proposal builds on the SRIA by adding additional content elements such as potential instruments that could be used for the implementation of the agenda.
- The formation of a **representative community of stakeholders** that has endorsed the SRIA and expressed an interest and commitment in getting involved in the cPPP. The identification of an industrially led core group ready to commit to the objectives of the cPPP with a willingness to invest money and time.
- The **establishment of a legal entity based** in Belgium: a non-profit organization named Big Data Value Association (BDVA) to represent the private side of the cPPP. The BDVA had 24 founding members, including many partners of the BIG project.

- And finally, the **signature of the Big Data Value cPPP** between the BDVA and the European Commission. The cPPP was signed by Vice President Neelie Kroes, the then EU Commissioner for the Digital Agenda, and Jan Sundelin, the president of the Big Data Value Association (BDVA), on 13 October 2014 in Brussels. The BDV cPPP provides a framework that guarantees the industrial leadership, investment, and commitment of both the private and the public side 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 that will boost economic growth and jobs.

2.9 Summary

The Big Data Public Private Forum (BIG) Project was an EU coordination and support action to provide a roadmap for big data within Europe. The BIG project worked towards the definition and implementation of a clear big data strategy that tackled the necessary activities needed in research and innovation, technology adoption, and the required support from the European Commission necessary for the successful implementation of the big data economy.

The BIG project used a three-phase methodology with technical working groups examining foundational technologies, sectorial forums examining innovative sectorial applications, and a road mapping activity to create technology and strategy roadmaps so that business and operational communities understand the potential of big data technologies and are enabled to implement appropriate strategies and technologies for commercial benefit. The project was a key contributor to setting up the Horizon 2020 Big Data Value Association contractual Public Private Partnership (cPPP) and Big Data Value Association.

References

Adner, R. (2012). *The wide lens: A new strategy for innovation*. London: Penguin.

Flick, U. (2004). Triangulation in qualitative research. In U. Flick, E. V. Kardorff, & I. Steinke (Eds.), *A companion to qualitative research* (p. 432). London: Sage.

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Continuum.

Yin, R. K. (2009). Case study research: Design and methods. In L. Bickman & D. J. Rog (Eds.), *Essential guide to qualitative methods in organization research*. London: Sage. doi:10.1097/FCH.0b013e31822dda9e.