European Big Data Value Strategic Research & Innovation Agenda

VERSION 0.99

Executive Summary

This Strategic Research and Innovation Agenda (SRIA) defines the overall goals, main technical and non-technical priorities, and a research and innovation roadmap for the European contractual Public Private Partnership (cPPP) on Big Data Value which has been proposed by a partnership of European Big Data stakeholders led by NESSI, the European Technology Platform (ETP) for software, services and data .

The SRIA explains the strategic importance of Big Data, describes the Data Value Chain and the central role of Ecosystems, details a vision for Big Data Value in Europe in 2020, analyses the associated strengths, weaknesses, opportunities and threats, and sets out the objectives and goals to be accomplished by the cPPP within the European research and innovation landscape of Horizon 2020 and national and regional initiatives.

The multiple dimensions of Big Data Value are described, and the overarching strategic objectives for the cPPP are set out. These embrace data, skills, legal and policy issues, technology leadership through research and innovation, transforming applications into new business opportunities, acceleration of business ecosystems and business models, with particular focus on SMEs, and successful solutions for major societal challenges Europe is facing such as Health, Energy, Transport and the Environment. They are broken out into specific competitiveness objectives, innovation and technology objectives, societal objectives and operational objectives.

The implementation strategy for addressing the goals of the SRIA involves four mechanisms: i-Spaces, Lighthouse projects, technical projects, and cooperation and coordination projects. i-Spaces are cross-organisation cross-sector interdisciplinary Innovation Spaces to anchor targeted research and innovation projects. Offering secure environments as accelerators for experiments with both private data and open data, bringing technology and application developments together. They will act as incubators for new businesses and for the development of skills, competence and best practices. Lighthouse projects are large-scale data-driven innovation and demonstration projects which will create high-level visibility, awareness and impact.

The strategic and specific goals, which together will ensure Europe's leading role in the data-driven world, are supported by key specific technical and non-technical priorities. Five technical priority areas have been identified for research and innovation: deep analysis, to improve data understanding; optimized architectures for analytics of data-at-rest and data-in-motion; mechanisms for managing privacy and anonymisation, to enable the vast amounts of data which are not open data (and never can be open data) to be part of the Data Value Chain; advanced visualization and user experience; and, underpinning these, data management engineering. The complementary non-technical priorities are skills development, business models and ecosystems; policy, regulation and standardization; and social perceptions and societal implications.

Finally, the expected impact of achieving the objectives is summarised, together with KPIs to frame and assess that impact. The activities set out in this SRIA will deliver solutions, architectures, technologies and standards for the data value chain over the next decade, leading to a comprehensive eco-system for achieving and sustaining Europe's role, for delivering economic and societal benefits, and enabling a future in which Europe is the world-leader in the creation of Big Data Value.

Contents

1 Introduction – the strategic importance of Big Data

The economic potential of Big Data

Economic and social activities have long relied on data. But today the increased volume, velocity, variety, and social and economic value of data signals **a paradigm shift towards a data-driven socioeconomic model**.

In parallel with the continuous and significant growth of data has come better data access, availability of powerful ICT[1] systems, and ubiquitous connectivity of both systems and people. This has led to intensified activities around Big Data and Big Data Value. **Powerful tools** have been developed to collect, store, analyse, process, and visualize huge amounts of data. **Open data initiatives** have been launched to provide broad access to data from the public sector, business and science.

The volume of data is rapidly growing: it is expected that **by 2020 there will be more than 16 zettabytes** of **useful data** (16 Trillion GB)[2], which implies growth of 236% per year from 2013 to 2020. This data explosion is a reality that Europe must both face and exploit in a structured, aggressive and ambitious way to create value for society, its citizens, and its businesses in all sectors.

It is clear that Data is now an asset that can create a significant competitive advantage and drive innovation, increase competitiveness and create social impact. As EU Commissioner Kroes has stated on several occasions: "**Big Data is the new oil**". Big Data therefore has to be regarded as a **primary asset** for all sectors, organizations, countries and regions.

The following table provides some examples how Big Data will impact different sectors:

Sectors/Domains	Big Data Value	Source
Public administration	EUR 150 billion to EUR 300 billion in new value (Considering EU 23 larger governments)	OECD[3], 2013
Healthcare & Social Care	EUR 90 billion considering only the reduction of national healthcare expenditure in the EU	McKinsey Global Institute[4], 2011
Utilities	Reduce CO2 emissions by more than 2 gigatonnes, equivalent to EUR 79 billion (Global figure)	OECD[6], 2013
Transport and logistics	USD 500 billion in value worldwide in the form of time and fuel savings, or 380 megatonnes of CO2 emissions saved	OECD[6], 2013
Retail & Trade	60% potential increase in retailers' operating margins possible with Big Data	McKinsey Global Institute[2], 2011
Geospatial	USD 800 billion in revenue to service providers and value to consumer and business end users	McKinsey Global Institute[2], 2011
Applications & Services	USD 51 billion worldwide directly associated to Big Data market (Services and applications)	Various[5,6]

[1] A full list of acronyms and terms is presented in 6.1
[2] "*The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*" Vernon Turner, John F. Gantz, David Reinsel, and Stephen Minton, Report from IDC for EMC April 2014.
[3] "*Exploring Data-Driven Innovation as a New Source of Growth – mapping the policy issues raised by "Big Data""*, Report from OECD 18 June 2013.
[4] Applying assumptions from the McKinsey Global Institute report "Big Data: The next frontier for innovation, competition, and productivity", June 2011, to the European healthcare sector
[5] Big Data Market by Types (Hardware; Software; Services; BDaaS - HaaS; Analytics; Visualization as Service); By Software (Hadoop, Big Data Analytics and Databases, System Software (IMDB, IMC): Worldwide Forecasts & Analysis (2013 – 2018), available online at: www.marketsandmarkets.com, August 2013.
[6] "*Big Data Vendor Revenue and Market Forecast 2013-2017*", article, Wikibon, February 2014.

The Big Data Value market measured by the revenue that vendors earn from sales of related hardware, software and ICT services is a fast growing multibillion-euro business. According to IDC[7] the Big Data market is growing six times faster than the overall ICT market. The compound annual growth rate (CAGR) of the Big Data market over the period 2013 – 2017 will be around 27%, reaching an overall total of $50 billion.

The exploitation of Big Data in various sectors has socio-economic potential far beyond the specific Big Data market. Therefore, it is essential to embrace new technology, applications, use cases, and business models within and across various sectors and domains. This will ensure rapid adoption by organizations and individuals, and provide major returns in growth and competitiveness.

A significant contribution to the European economy

As identified by demosEUROPA, "Overall, by 2020, big & open data can improve the European GDP by 1.9%, an equivalent of one full year of economic growth in the EU"[8]. The increased adoption of Big Data will have positive impact on employment,. and is expected to result in 3.75 million jobs in the EU by 2017[9].

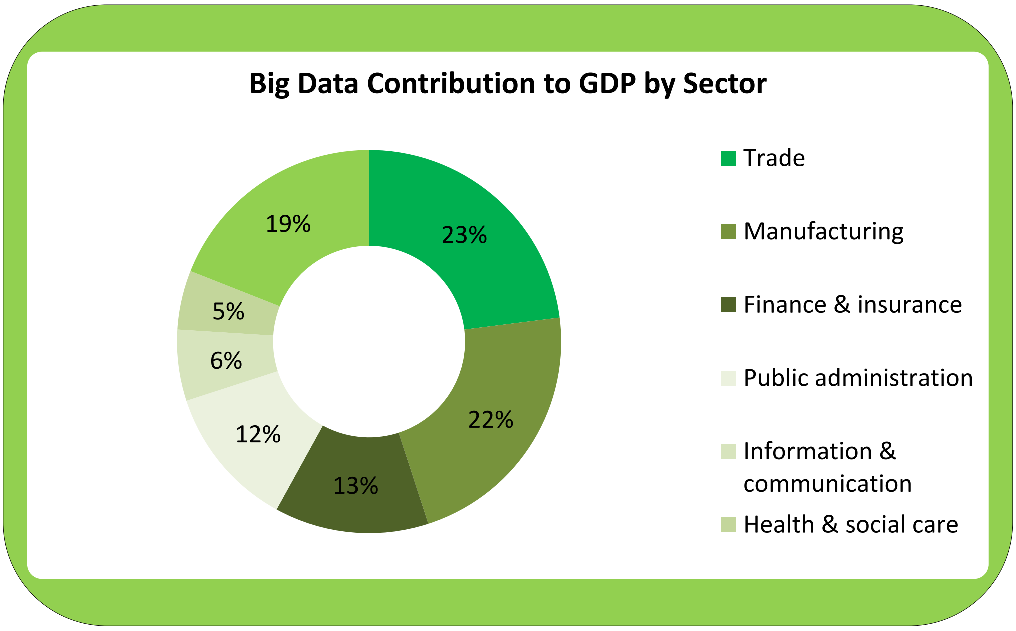

Figure 1: Economic potential of big and open data – source: demosEUROPA

Large companies and SMEs in Europe are clearly seeing the fundamental potential of Big Data for disruptive change in markets and business models, and are beginning to explore the opportunities. IDC confirms that the Big Data adoption in Europe is accelerating[10]. According to IDC[11], 30% of West European companies will adopt Big Data by the end of 2015. For the other 70% of business actors, it is crucial to provide new tools and assets to propel them into the data-driven economy.

---

[7] "*Worldwide Big Data Technology and Services 2013–2017 Forecast*", report, IDC, December 2013
[8] "*Big and open data in Europe - A growth engine or a missed opportunity?",* Sonia Buchholtz, Maciej Bukowski, Aleksander Śniegocki (Warsaw Institute for Economic Studies), report commissioned by demosEUROPA, 2014.
[9] Big Data Value calculation based on http://www.eskillslandscape.eu/ict-workforce-in-europe/ (also re footnote '7')
[10] "*The European Data Market*", Gabriella Catteneo, IDC, presentation given at the NESSI summit in Brussels on 27 May 2014, available online at: http://www.nessi-europe.eu/?Page=nessi_summit_2014
[11] IDC European Vertical Markets Survey, October 2013

However, Europe is still at an early stage of adopting Big Data technologies and services; and is lagging behind North America[12], with the picture in third countries being less well determined. Companies intending to build and to rely on data-driven solutions will face challenges that go well beyond technology. Successful adoption of Big Data will require changes in business orientation and strategy, processes, procedures and the organizational setup. European enterprises will create new knowledge and hire new experts, enhancing a new ecosystem.

The multiple dimensions of Big Data Value

In order to reduce the gap with other countries and regions and drive innovation and competitiveness, Europe needs to foster the development and wide adoption 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 landscape. The main issues that Europe must tackle for the creation of a strong Big Data ecosystem concern the following dimensions:

- **Data**: Availability of data and the access to data sources 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 generated by acquiring those data, combining data from different sources, and providing access to it with low latency while ensuring data integrity and preserving privacy. Value is added by pre-processing, validating, augmenting data and ensuring data integrity and accuracy.
- **Skills**: In order to leverage the potential of Big Data Value, a key challenge for Europe is to ensure the availability of highly and rightly skilled people who have an excellent grasp of the best practices and technologies for delivering Big Data Value within applications and solutions. There will be the need for data scientists and engineers who have expertise in analytics, statistics, machine learning, data mining and data management. These experts will need to be combined with other experts having strong domain knowledge and the ability to apply this know-how within organisations for value creation.
- **Legal:** 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. Those issues have to be resolved in order to remove the adoption barriers and favourable European regulatory environments are needed to facilitate the development of a true pan-European Big Data market.
- **Technical:** Key aspects such as real-time analytics, low latency and scalability in processing data, new and rich user interfaces, 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 are essential for a wide adoption within and across sectors
- **Application:** Business and market ready applications should be the target. Novel applications and solutions must be developed and validated in ecosystems providing the basis for Europe to become world-leader in the creation of Big Data Value.
- **Business:** A more efficient use of Big Data and understanding data as an economic asset carries great potential for the EU economy and society. The setup of Big Data Value ecosystems and the development of appropriate

---

[12] "*The European Data Market*", Gabriella Catteneo, IDC, presentation given at the NESSI summit in Brussels on 27 May 2014, available online at: http://www.nessi-europe.eu/?Page=nessi_summit_2014

business models on top of a strong Big Data Value chain must be supported in order to generate the desired impact on economy and employment.

- **Social:** Big Data will provide solutions for major societal challenges in Europe, such as the improved efficiency in healthcare information processing or reduced CO2 emissions through climate impact analysis. In parallel it is critical for an accelerated adoption of Big Data to increase awareness on the benefits and the Value that Big Data can create for business, the public sector, and the citizen.

Creating a favorable business environment for Big Data and pushing for its accelerated adoption requires an interdisciplinary approach addressing the dimensions of Big Data Value as described above.

The Data Value Chain and the central role of Ecosystems

Figure 2: The Big Data Value chain

In addition, Europe needs strong players along the Big Data Value Chain[13] (Figure 2) ranging from data generation and acquisition through to data processing and analysis, then to curation, usage, service creation and provisioning. All the links of the entire value chain have to be strong so that a vibrant Big Data Value ecosystem based on such a value chain can evolve.

There are already companies in Europe which provide services and solutions along the Big Data Value chain. Some of them generate and provide access to huge amounts of data including structured and unstructured data. They acquire or combine real-time data streams from different sources or add value by pre-processing, validating, augmenting data and ensuring data integrity. There are companies specialized in analyzing data and recognizing correlations and patterns. Furthermore, there are companies which use these insights for predictions and decisions in various application domains.

Despite the growing number of companies active in the data business, an economic community supported by interacting organisations does not yet really exist for the Big Data Value Chain at the European level. Data usage is growing, but both in businesses and in science it is treated and handled in a fragmented way. In order to ensure a coherent use of data, a wide range of stakeholders along the Data Value Chain should be brought together to facilitate cooperation.

The stakeholders that will form the basis for interoperable data-driven ecosystems as a source for new businesses and innovations on Big Data are:

- Vendors of the ICT industry (Large and SME)
- Users across different industrial sectors (private and public)
- Big Data Value companies that do not exist yet and will emerge
- Researchers and academics who can provide knowledge and thought leadership

The cross-fertilisation involving these many stakeholder and many datasets is a key element for the advancing Big Data economy.

[13] "*Competitive Advantage –Creating and Sustaining Superior Performance*", Michael E. Porter, New York, 1998

Finally, it is vital that SMEs and web entrepreneurs participate in this ecosystem and become part of the Big Data Value chain. They are an essential part of the process to create value based on their specific and strong niche competences at the technical, application and business level.

Need for action

Big Data is one of the key economic assets of the future. Mastering the creation of Value from Big Data will impact the competitiveness of companies and will result into economic growth and jobs in Europe. Strategic investments by industry, public sector and governments accompanied by forward-looking policies will enable Europe to take the lead in the global data-driven digital economy and to reap societal benefits from the unique opportunities offered by Big Data Value. The European Council highlighted Big Data in its conclusion of 24/25 October 2013[14] as a strategic technology and important enabler for productivity and better services. However, immediate action is required not to miss these opportunities. Therefore, **Commissioner Kroes has called for a European Public Private Partnership in Big Data** in her speech at the ICT 2013 event in Vilnius on 7 November 2013.

Europe must aim high and mobilise stakeholders in society, industry, academia and research to enable a European Big Data 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. The result will be an innovation environment in which value creation from Big Data flourishes.

In order to achieve these goals, a **European contractual Public Private Partnership (cPPP)** on Big Data Value has been proposed by a partnership of European Big Data stakeholders led by NESSI, the European Technology Platform (ETP) for software, services and data.

This **Strategic Research and Innovation Agenda (SRIA)** defines the overall goals, main technical and non-technical priorities, as well as a research and innovation roadmap for the cPPP.

A wide range of stakeholders have contributed to this SRIA. It is built upon inputs and analysis from SMEs and Large Enterprises, public organisations, and research and academic institutions. They include suppliers and service providers, data owners, and early adopters of Big Data in many sectors.

The value that the intelligent use of Big Data can generate is already being recognised by some private and public organisations. There are relevant national initiatives in Germany[15], France[16], Ireland[17] and the UK[18]. It is essential that these be connected at the European level, establishing knowledge sharing, and collaborating to advance the technology.

Discussions and workshops have clearly shown that, alongside vital research and innovation in technologies and applications, many economic, societal and legal challenges will have to be addressed in an interdisciplinary fashion. Underpinning successful exploitation will be the availability of skills and access to investment and capital. Citizens must also be involved, and provision has been made to emancipate them as stakeholders so that they can be the ethical integration point of their own data.

---

[14] European Council Conclusion – 24/25 October 2013 – EUCO 169/13, available online at http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/139197.pdf
[15] http://www.sdil.de/
[16] http://www.teralab-datascience.fr/en/home
[17] http://insight-centre.org/content/launch-insight
[18] http://theodi.org/

1.1 A Vision for Big Data

The vision for Big Data Value in Europe is that in 2020:

- **Data:** Zettabytes of useful public and private data will be widely and openly available
- **Skills:** Millions of jobs established for data engineers and scientists, and the Big Data discipline is integrated in technical and business degrees. The European workforce is more and more data-savvy seeing data as an asset and in the domains many rather than a few are data specialists
- **Legal:** Privacy & Security can be guaranteed through the lifecycle of BDV exploitation. Data sharing and data privacy will be fully managed by end-users themselves in a trustworthy society – citizen emancipation
- **Technology:** Real-time integration and interoperability among different multilingual, sensorial, and non-structured datasets and where content is automatically managed and visualised in real-time
- **Application**: Applications using the BDV technologies can be built which will allow anyone to create, use, exploit and benefit from Big Data
- **Business:** A true EU single data market will be established allowing EU companies to increase their competitiveness and become world leaders
- **Social:** Societal challenges are addressed through BDV systems addressing high data volume, high motion of data, high variety of data, etc.

These will impact the European Union's priority areas as follows:

- **Economy:** Competitiveness of European enterprises will be significantly higher compared to its worldwide competitors with improved products and services and higher efficiency based on Big Data value. A true EU single data market will be established allowing EU companies to increase their competitiveness and become world leaders
- **Growth:** There is a blossoming sector of growing new small and large businesses with a significant number of new jobs that create value out of data
- **Society:** Citizen benefits from better and more economical services in trustful economy where data can be shared with confidence. Privacy & security will be guaranteed throughout the lifecycle of BDV exploitation

By 2020 thousands of specific applications and solutions will address Data-in-motion and Data-at-rest. There will be a highly secure and traceable environment supporting organisations and citizens and having the capacity to support various monetization models.

By 2020 Value creation from Big Data will have a disruptive influence on many sectors.. From manufacturing to tourism, from healthcare to education, from energy to telecommunications services, from entertainment to mobility, Big Data Value will be a key success factor in fuelling innovation, driving new business models, and supporting increased productivity and competitiveness.

By 2020, smart applications such as smart grids, smart logistics, smart factories, and smart cities will be widely deployed across the continent and beyond. Ubiquitous broadband access, mobile technology, social media, services, and IoT on billions of devices will have contributed to the explosion of generated data to a global total of 40 zettabytes[19]. Much of this data will yield valuable information. Extracting this information and using it in intelligent ways will revolutionize decision making in businesses, science, and society, enhancing companies' competitiveness and leading to new industries, jobs and services.

By 2020, European research and innovation efforts will have led to advanced technologies which make it significantly easier to use Big Data across sectors, borders and languages.

This foreseen evolution demands rethinking technologies around Big Data. Data collection, storage and processing must be improved in order to allow much more efficient access to data. Data visualisation and data analytics are also areas where new technologies will be needed. These technologies have different innovation cycles (in the range of months for services and applications, and years for ICT infrastructure) implying that architectures, technologies and standards cannot be designed based on pre-defined requirements. It is necessary to make challenging working assumptions on major basic technical requirements based on today's best knowledge in order to meet the needs expected in 2020.

Software based systems provide the flexibility to adapt to new requirements introducing innovation into deployed systems, but the overall architecture and ICT infrastructures for storing and managing data do not offer this flexibility at present. Therefore, from the medium-long-term vision perspective, future systems have to offer high flexibility and have to allow for high adaptability to new schemes.

1.2 Strengths, Weaknesses, Opportunities and Threats

The priorities identified in this Strategic Research and Innovation Agenda reflect the views of industry, research organizations and academia, representing providers and users of technologies and data assets in many sectors. A number of workshops[20] were organised in order to ensure that the objectives set out in this SRIA are based on the real needs of both public and private entities in Europe.

The main task of each workshop was to identify main priorities and a SWOT analysis for each of the sectors, including consideration of the benefits derived from cross-sector fertilization. The workshops addressed different industrial sectors, including energy, manufacturing, environment and geo-spatial, health, public sector, content and media. In addition to the sector workshops, additional workshops were organized to gather feedback on cross-sector aspects and the views from SMEs. A compilation of the workshop results is provided in the following pages as an integrated SWOT analysis for the European market.

These views form the basis for the strategic and specific objectives for the SRIA, set out in Section 1.3.

Strengths

European Aspects:

- Compared to the rest of the world, Europe has a strong medium-sized sector with regard to Big Data
- Europe offers a stable environment in terms of life standards, currency, etc.

---

[19] "*THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*", IDC report, December 2012
[20] For more details about the workshops, please see Section 3.1

Market and Business:

- There is a specific European capacity that allows for companies to start in niches and then grow their business potential
- There are many SMEs which are dynamic and flexible and can react quickly to market changes
- There is an existing and strong content/data market in Europe
- There are established cooperation networks between content providers in several domains

Technical:

- Computer clusters and cloud resources are readily available
- There is a growing interest in archiving, sensing, behavioural data, and personal data

Data and Content:

- There is a large amount of content and data available – the issue is making use of it
- There are already a number of existing ecosystems and portals (for example INSPIRE[21], Copernicus[22] and GEOSS[23])
- Geospatial and environmental data sets and supporting infrastructure data are available

Education and Skills:

- There is a broad and detailed domain know-how as well as process know-how available
- Many domains have innovative technology and skilled people
- There are many universities with high capacity where skills can be developed
- Good engineering /domain specific education can be obtained

Policy, Legal and Security:

- The European Union promotes free and open processes

Weaknesses

European Aspects:

- Europe is decentralized which can lead to disparate policies
- Some domains are characterized by conservatism and long innovation cycles
- There is a lack of a solid start-up culture because of risk aversion and intolerance of failure
- There are few European data analytics solutions providers
- There are few larger companies to lead the market, and many small sized companies that need nurturing

Market and Business:

- There is a lack of access to Big Data facilities which make data more easily accessible
- There is no visibility of ecosystem service offerings
- Is unclear what data should be preserved, and for how long, in all the different sectors and markets

[21] http://inspire.ec.europa.eu/
[22] http://www.copernicus.eu/
[23] https://www.earthobservations.org/geoss.shtml

Technical:
- Lack of processable linked data, and of aggregated/combined data
- Lack of seamless data access and interconnectivity, and low levels of interoperability: data is often in silos and data sharing is difficult due to a lack of standards for e.g. formats and semantics
- Migration of data between systems, versions or partners is challenging
- Access and processing of data sets that are too big to be given to the end user

Data and Content:
- Public data in EU is not available to the extent it should be
- The quality of data in open data portals is often very low
- The different languages within Europe creates a barrier (multilingualism) during data processing
- Structural data sources often lack precise semantics, for instance through labels from ontologies
- Poor and inconsistent use or management of metadata

Education and Skills:
- There is a lack of specialised education programs for data analysts
- There are not enough skilled people to participate in training programmes

Policy, Legal and Security:
- Legislative restrictions on data sharing decreases availability across Europe and makes European-focused initiatives which address these issues more difficult
- Rules and regulations are fragmented across Europe
- There are high security demands which can be difficult to address

Usage:
- Europe is not good at analysing and changing consumer behaviour
- Citizen science – how to qualify and use data from citizens
- Providing Big Data (Value) for SME use

Opportunities

European Aspects:
- Various cultures and various strengths can result in creative thinking if they are mixed
- The existence of BDV topics and best practice examples in other initiatives can lead to synergies
- Strengthening the European market, e.g. by fusing the emerging start-up nucleus
- Create lots of SMEs for the low hanging fruits of Big Data for which agility is required
- Investment in the entire innovation chain, beyond basic research
- Investment support mechanisms for SMEs (e.g. European loans).

Market and Business:
- Opening up of private content to extend and complement existing assets
- Increasing the use of analytics
- Many opportunities exist for particular sectors, for example, environmental monitoring, social media, industrial processing
- There is a potential for extending INSPIRE and COPERNICUS on a global scale
- Improve creativity to create cost-effective solutions

- There is the opportunity to open up completely new and different business areas and services
- New applications can be created throughout the Big Data ecosystem, ranging over acquisition, data extraction, analysis, visualization and utilisation

Technical
- Easier syndication of data and content
- Micropayments for processed data or analytic information
- Wearable sensors and sensor technologies become mainstream generating more data
- The explosion of device types opens up access to any data from any device for greater and more varied usage
- Development of APIs for access becoming standardised and available
- Interoperability tools and standardised APIs to facilitate data exchange
- Greater visibility and increased use of directory services for data sources

Data and Content
- Making use of European cultural and data assets
- Use semantics to align content from various data sources
- Providing facilities to better navigate and curate data
- Contextualisation and personalisation of data
- The evolution of different sectors and the increased volume of data enables more and different applications to be developed

Education and Skills:
- Exploring new research areas
- Training focussed on innovation in BDV
- Use and exploration of Big Data to be ubiquitous in education and training

Policy, Legal and Security:
- Address the safe and secure storage of data on a European basis
- Develop uniform policies for data access in Europe to help build competitive capabilities

Usage:
- User generated and crowd source content increasingly available
- Data as a service can significantly lower the market entry barriers (in particular to new markets)
- Shift from technology push to end-user engagement
- Create rich and complex data value chains

Threats

European Aspects:
- Europe is lagging behind the US in the Big Data market
- US players and their bottom-up ideas are dominating the market
- Europe does not have a Big Data and data-sharing culture

Market and Business:
- Dominant large corporations own important data
- Consolidation of stakeholders and marketplaces are reducing competition
- There are several barriers to market entry for SMEs, for example for owning data

Education and Skills:

- Many skilled professionals leave Europe to work in other countries, there is a risk of “Brain Drain” in Europe
- Continuous lack of skilled professionals and graduates

Policy, Legal and Security:

- Policies are often too connected to the ‘old data’ world
- Complete analysis of ethical and privacy issues are needed
- There is a risk of overregulation and protectionism in Europe; privacy regulations elsewhere are less restrictive
- Policies of data availability; for example companies are not willing to make data available ’just-in-case’ it may cause harm in another territory

Usage:

- Cross-border data flows
- Data-driven services are not tied to a particular location but are subject to different legislation in different countries.

1.3 Strategic and Specific Objectives

The BDV cPPP has its roots in the Industrial Leadership Priority in Horizon 2020, specifically the ICT industrial and technological leadership challenge “Content technologies and information management”. This aims to strengthen Europe’s position as provider of innovative multilingual products and services based on digital content and data.

Cross-cutting initiatives with other strategic areas, particularly societal challenges, will be facilitated in later Work Programmes of Horizon 2020. This will invariably widen the range of objectives of the cPPP and contribute to EU societal challenges through the implementation of BDV applications and solutions.

The following H2020 societal challenges have particular importance for the BDV cPPP due to Europe’s current concerns and challenges in these areas (although the cPPP is domain neutral):

- **Health**: Demographic change and wellbeing
- **Energy**: Secure, clean and efficient energy
- **Transport**: Smart, green and integrated transport
- **Environment**: Climate Action, Resource Efficiency and Raw Materials

The Big Data Value cPPP is driven by the conviction that research and innovation focusing on a combination of business and usage needs is the best long term strategy. This will bring many benefits and stimulate the creation of value from Big Data to reach the level that is needed to create jobs and prosperity.

The overarching **strategic objectives** are:

- **Data:** To access, compose and use data in an easy to use way according to clearly defined model and allowing the transformation of data into information
- **Skills:** To contribute to the conditions for skills development in industry and academia
- **Legal & Policy:** To contribute to policy processes for finding favourable European regulatory environments, and address concerns of privacy and citizen inclusion
- **Technology:** To foster European BDV technology leadership for job creation and prosperity by creating a European wide technology and application base and building up competence. In addition To enable research and innovation work,

including the support of interoperability and standardisation, for the future basis of BDV creation in Europe

- **Application:** To reinforce the European industrial leadership and capability to successfully compete on global level in the data value solution market by advancing applications transformed into new opportunities for business
- **Business:** To facilitate the acceleration of business ecosystems and appropriate business models with particular focus on SMEs, enforced by Europe-wide benchmarking of usage, efficiency and benefits
- **Social:** To provide successful solutions for major societal challenges Europe is facing such as: Health, Energy, Transport and the Environment. And in addition to increase awareness about BDV benefits for businesses and public sector and better acceptance by citizen, to engage them as prosumers and accelerate take-up

Specific objectives are:

Competitiveness Objectives (Business):

- To use BDV technology for increased productivity, optimised production, more efficient logistics (inbound and outbound), effective service provision from public and private organisations
- To create a Big Data Economy including new ecosystems and markets between data providers, knowledge providers and consumers that will profit from sector, organizations and individual collaboration
- To reduce the gap between traditional economy and new digital business models that will smooth changes in the economic value chain and stakeholders including End Users
- To implement European-wide strategic projects (Lighthouse projects – See Section 2.2) for specific reference deployments of existing or near-to-market technologies that demonstrate the impact that can be achieved by BDV creation across sectors

Innovation objectives (Technology):

- To develop and make available to industry and the public sector technology, applications and solutions for the creation of value from Big Data
- To optimize architectures for real-time analytics of both data at rest and in motion enabling data-driven decision-making on the fly with low latency as well as to improve scalability and processing of data validation and information discovery especially in heterogeneous data sets
- To drive the integration of the BDV services into private and public decision making systems such as Enterprise Resource Planning and marketing systems for optimising the functioning of existing industries and potentially establishing entirely new business models
- To validate technologies from a technical and a business perspective within cross-organisational, cross-sector, and cross-lingual innovation environments through early trials
- To enable European industry, business, public sector and citizens to use and take advantage of value creation from Big Data that is validated with user involvement based on open and private data in secure and privacy respecting environments
- To integrate advanced visualization of data and analytics for augmented user experience and prepare platforms, technologies and tools for disruptive changes in management of data

From the above, technical domains for the development of the BDV cPPP are detailed in Section 3 of this document: deep analytics to improve data understanding, optimized architectures of both data at rest and in motion, privacy and anonymisation mechanisms, advanced visualization, and data management engineering.

Societal objectives (Society):

- To support building extensive know-how, education and skills in Europe (e.g. by European curricula and sharing of best practices) for future systems in the research community and industry
- To enable European industry, business, public sector and citizens to use and take advantage of value creation from Big Data; that is validated with user involvement based on open and private data in secure and privacy respecting environments
- To develop and provide validated technology and tools for “deep data analysis” to improve data understanding, deep learning and increased meaningfulness of data, by creating the comprehension of the importance of data definition for optimal information content
- To create new personalized and enhanced product and services adapted to citizens and organizations need that will respect security and privacy of individuals
- To address European framework aspirations such as IPR, liability etc. within the Digital Single Market and the pan European innovation environments
- To support the societal challenges that Europe faces through Lighthouse projects and i-Spaces in these challenge areas

Operational Objectives:

- To create an environment for productive research and innovation activities by utilising proven approaches including clustering actors around key research and innovation areas
- To establish a governance model that provides for efficient decision making at the level of concerned activity and at the same time following the target of openness and transparency
- To maintain an effective flow of information between and amongst the cPPP projects that is overcoming barriers and resentments while respecting interests and rights of individual beneficiaries
- To enable a collaboration amongst the projects to support the common targets to positively impact European society and industry

In order to achieve the strategic and specific objectives, the research and innovation strategy requires dedicated actions and mechanisms along its overarching strategic goals:

In short, some of the main **technical** aspects are:

- **Data:** Placing Data is at the centre of the Big Data Value activities and making data sets and assets accessible. These will typically be based on their domain of operation and include industrial, private and open data sources whilst ensuring their availability, integrity, and confidentiality.
- **Technology:** Fostering research and innovation activities to develop Big Data Value technologies and tools. Focus will be put on those technologies and tools which are needed to support data-driven applications and business opportunities along the data value chain. They will be addressed in technical projects (see Section 3).
- **Application:** Benchmarking and incubating by allowing testing of technologies, applications, and business models. This will provide early insights on potential issues and will help to avoid failures in the later stages of commercial deployments. In addition, it can be expected that these activities will provide inputs for standardization and regulation.

Complementary, further **socio-economic** aspects need to be addressed

- **Skills:** Developing skills and federating best practices through linking with other similar initiatives at the European and national level.
- **Business**: Exploring and stimulating new business models and ecosystems that will emerge from the exposure of new technologies and tools to both closed (industrial) and open data. They will be the playground to test new business model concepts and emerging ecosystems of existing and new entrants
- **Legal, Policy and Privacy:** Providing insight into country, regional and European-wide legislation, regulation, and similar issues which impact the implementation or use of data-intensive technologies.
- **Social**: Acquiring 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 eco-systems.

2 Implementation Strategy

Given this broad range of objectives around many aspects of Big Data a complete implementation strategy is needed. In this section we set out such a strategy, resulting from a very broad discussion process involving a large number of relevant European BDV stakeholders.

The result is an interdisciplinary approach which integrates expertise from all the fields necessary to tackle the strategic and specific objectives. To this end, European cross-organisational and cross-sector environments have to be incubated, such that large enterprises and SMEs alike will find it easy to discover economic opportunities based on data integration and analysis and then develop working prototypes to test the viability of actual business development.

The growing number and complexity of valuable data assets will drive existing and new research challenges. The cross-sectorial and cross-organisational environment will allow research and innovations for new and existing technologies. Business applications that need to be evaluated for usability and fitness of purpose can then be built on top of these technologies. In addition, a further key aspect is ensuring practical applicability of those outcomes. This in turn will require validations, trials and large-scale experiments in existing or emerging business fields, the public sector, industry, and jointly with end-user and individual consumers.

To support validations, trials and large-scale experiments, access to valuable data assets needs to be provided with low obstacles in environments that simultaneously support legitimate ownership, privacy and security related to data owners and their customers. Such environments will ease experimentation for researchers, entrepreneurs, SMEs and large ICT providers.

Four kinds of mechanisms

In order implement the research and innovation strategy, and to align technical with cooperation and coordination aspects, the four major kinds of mechanisms are recommended to be realized:

- **Innovation Spaces** (i-Spaces): Cross-organisational and cross-sector environments – will allow challenges to be addressed in an interdisciplinary way and will serve as a hub for other research and innovation activities.
- **Lighthouse projects:** These will help raise awareness about the opportunities offered by Big Data and the value of data-driven applications for different sectors and they will be an incubator for data-driven ecosystems.
- **Technical projects:** These will take up specific Big Data issues addressing targeted aspects of the technical priorities as defined in Section 3.

- **Cooperation and coordination projects:** These projects will foster international cooperation for efficient information exchange and coordination of activities.

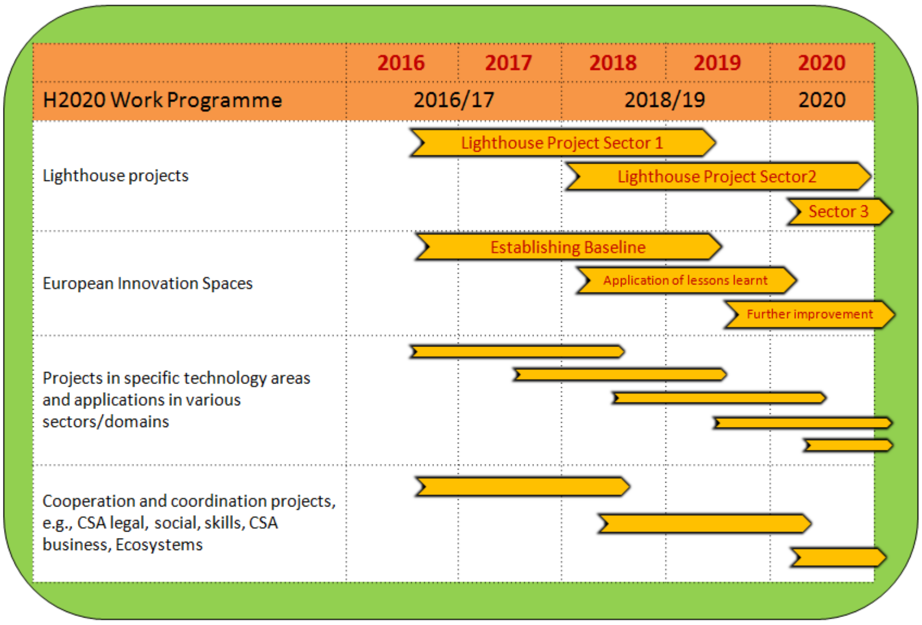

Figure 3: Timeline of activities

The lighthouse projects will start in the most mature sectors to play an early role regarding large scale demonstrations for accelerated take up of BDV creation in industry and public sector.

SRIA Preparation Process with broader community

Within the SRIA preparation process, the proposers have heavily engaged with the wider community. Multiple workshops and consultations took place to ensure the widest representation of views and positions including the full range of public and private sector entities. These have been carried out in order to identify main priorities with approximately 200 organisations and other relevant stakeholders physically participating and contributing. Extensive analysis reports were then produced which helped both formulate and construct this SRIA.

The series of workshops gathered views from different stakeholders in existing value chains of different industrial sectors, including: energy, manufacturing, environment and geo-spatial, health, public sector, content and media; additional workshops have been organized to gather feedback on cross-sectorial aspects, for example, the views from SMEs. The selection of sectors was based on the criteria of their weight in the EU economy and potential impact of their data assets (source: demosEUROPA). The community involved in the Workshops included: Actors such as AGT International (DE), Hospital de la Hierro (ES), Press Association (UK), Reed Elsevier (NL); BIODONOSTIA (ES), Merck8 (ES), Kongsberg Group (NO), and many more.

In addition NESSI, together with partners from the FP7 project BIG, ran an online public consultation on the BDV Strategic Research and Innovation Agenda between 9 April and 15 May 2014. The aim was to validate the main ideas put forward in the SRIA on how to advance Big Data Value in Europe in the next 5 to 10 years. 195 organisations from all over Europe participated in the consultation including companies such as Hitachi Data Systems, OKFN Belgium, TNO Innovation for Life, Euroalert, Tecnalia Research and Innovation, ESTeam AB, and CGI Nederland B.V. In addition, around 20 organisations and companies such as Wolters Kluwer

Germany, Reed Elsevier and LT-Innovate shared in more detail their views on the content of the SRIA.

Although the primary target is to create impact at a European level, cooperation with stakeholders outside Europe will allow the transfer of knowledge and experiences around the globe. For future collaborations, NESSI has already set up links to the following regions through NESSI partners: Mediterranean countries[24], LatAm countries[25], South East Asian countries[26] and the Russian speaking countries[27].

SRIA Update Process

The technical and non-technical priorities reflected in this version of the SRIA reflect the consolidated results from the needs analysis performed by involving all relevant stakeholders of the Big Data environment. We are well aware that in a fast moving area such as Big Data, those priorities need to be regularly reflected and updated if needed. To this end, it is planned to release regular updates of this SRIA document. Instrumental to those SRIA updates will be the BDV Stakeholder Platform.

BDV Stakeholder Platform

The BDV Stakeholder Platform constitutes a permanent platform for BDV Stakeholders to express their views, expectations and requirements. They will be captured in a continuous manner, thereby establishing a series of long-term SRIAs. BDV stakeholders that cannot commit to participate on a regular manner but wish to participate and contribute as they can, will be also gathered by the platform.

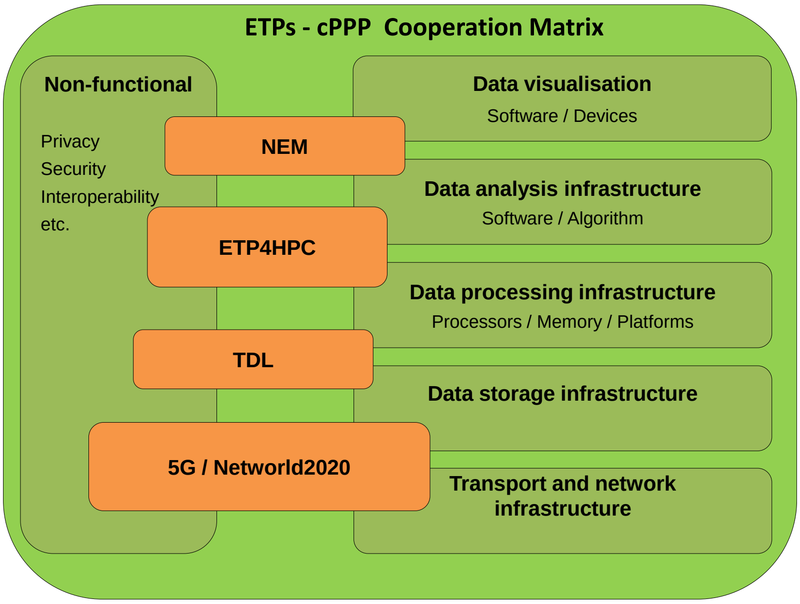

Figure 4: Examples of collaboration with other ETPs and PPPs [28] [29] [30] [31] [32]

---

[24] MOSAIC (http://www.connect2sea.eu/) and MED-Dialogue (www.med-dialogue.eu)
[25] CONECTA 2020 (www.conecta2020.net)
[26] CONNECT2SEA (http://www.connect2sea.eu/)
[27] EAST HORIZON (http://www.eeca-ict.eu/about/new_projects/easthorizon)
[28] NEM: New European Media, http://nem-initiative.org/
[29] ETP4HPC: The European Technology Platform for High Performance Computing, http://www.etp4hpc.eu/
[30] TDL: Trust in the Digital World, http://www.trustindigitallife.eu/
[31] 5G: 5G PPP, http://5g-ppp.eu/
[32] NetWorld2020: The European Technology Platform for communications networks and services, http://networld2020.eu/

The BDV Stakeholder Platform will take advantage of already established stakeholder groups and communities, such as started in BIG and BYTE and will take them into account wherever appropriate. Once set-up it will have the capacity to gather and coordinate BDV stakeholder recommendations along technology, application, skills, eco-system and social dimensions. The stakeholder platform will be open, neutral, independent and representative of the different communities needed to set up a successful data-driven ecosystem in Europe including technology providers, industrial players both large and SMEs, academia, public sector, users and/or user communities, start-ups etc. The BDV Stakeholder Platform will address a number of cross domain and cross sector topics. As an example, collaboration will be sought for with other ETPs such as ETP4 HPC, NEM and partnerships like TDL, 5G or EUROGI (see Figure 4). It should also be open to more dynamic agents that can provoke new innovative usages of the data and business models typified by e.g. web entrepreneurs. The activity of the platform is on a more detailed and typically domain specific level and will form key inputs to the BDVA in the development of its roadmaps and SRIAs.

Regular revision of the BDV Stakeholder Platform will create trust and ensure independency. Advisory sessions and hearing with international experts should be part of the activities. Steering committees for organizational and technology issues will be set up to enable the projects to work as efficiently as possible and identify interdependencies and complementarities. The federation of Big Data initiatives from private and public sector on European, Members States or regional level will make the collaboration with other organizations vital. Examples are the Smart Data Innovation Lab in Germany or the Teralab in France.

2.1 European Innovation Spaces (i-Spaces)

Extensive consultation with many stakeholders of relevant areas related to Big Data Value (BDV) have confirmed that besides technology and applications additional, key issues have to be considered: Fist, infrastructural, economic, social and legal issues have to be addressed. Second, the privat and the public sector will have 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 these aspects, European cross-organisational and cross-sector environments, which rely on and build upon existing national and European initiatives, will play a central role for a European Big Data ecosystem. These so-called ***European Innovation Spaces*** (***i-Spaces*** for short) are the main elements to assure that research on BDV technologies and novel BDV application will be quickly tested, piloted and thus exploited in a context with maximum involvement of all the stakeholders who would ultimately be part of the respective BDV ecosystems. As such, i-Spaces will enable stakeholders to develop new businesses facilitated by advanced BDV technologies, applications, and business models.

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

The creation of i-Spaces will be driven by the needs of large and small companies alike to ensure they will find it easy to discover economic opportunities based on data integration and analysis and develop working prototypes to test the viability of actual business development. This does not necessarily require moving data assets across borders. Rather data analytics tools and computation activities could be brought to the data. Thereby, 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, and ease of experimentation for researchers, entrepreneurs and small and large IT providers.

Concerning value creation discovery, i-Spaces will support various models: at one end, corporate entities with valuable data assets will be able to 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, will be able to solicit the addition and integration of desired data assets from corporate or public sources.

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

The particular European value-add of i-Spaces is that they will federate, complement and leverage activities of similar national incubators/environments, existing PPPs and other national or European initiatives (unless this is required for methodologically sound benchmarking activities). With the aim of not duplicating these existing activities, complementary activities considered for inclusion will have to stand the test of expected economic activity development: new data assets and technologies will be considered for inclusion to the extent that they can be expected to open new economic activities when added to and interfaced with the assets maintained by regional or national data incubators or existing PPPs.

The successive inclusion of data assets into i-Spaces will in turn drive and prioritise the agenda for addressing data integration or data processing technologies. One is example is the existence of data assets of homogenous qualities (such as geospatial, time series, graphs and imagery), which requires optimising the performance of existing core technology (such as querying, indexing, feature extraction, predictive analytics and visualization). This in turn requires methodologically sound benchmarking practices to be carried out in appropriate facilities. Similarly, business applications exploiting BDV technologies will be evaluated for usability and fitness of purpose, thereby leading to continuous improvement of these applications.

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

While economic development is the principal objective of BDV, this cannot happen without taking into proper account the legislative requirements pertaining to the treatment of data as well as ethical considerations. In addition, BDV will create value for society as a whole by systematically supporting the transfer of sophisticated data management practices to domains of societal interest such as health, environment or sustainable development, among others. Especially when it comes to SMEs, the issues of skills and training, reliable legal frameworks, reference applications and access to an ecosystem become central for a fast take-up of the opportunities offered by BDV. In this interdisciplinary holistic approach, i-Spaces will be one key mechanism that targets BDV challenges along the dimensions depicted in Figure 5 below. The innovation spaces will be instrumental to test, showcase and validate new technology, applications and business models. The central need for availability of open and industrial data assets will be catered for as well as for skills development, best practices identification, requirements for favourable legal, policy and infrastructural frameworks and tools across sectors and borders.

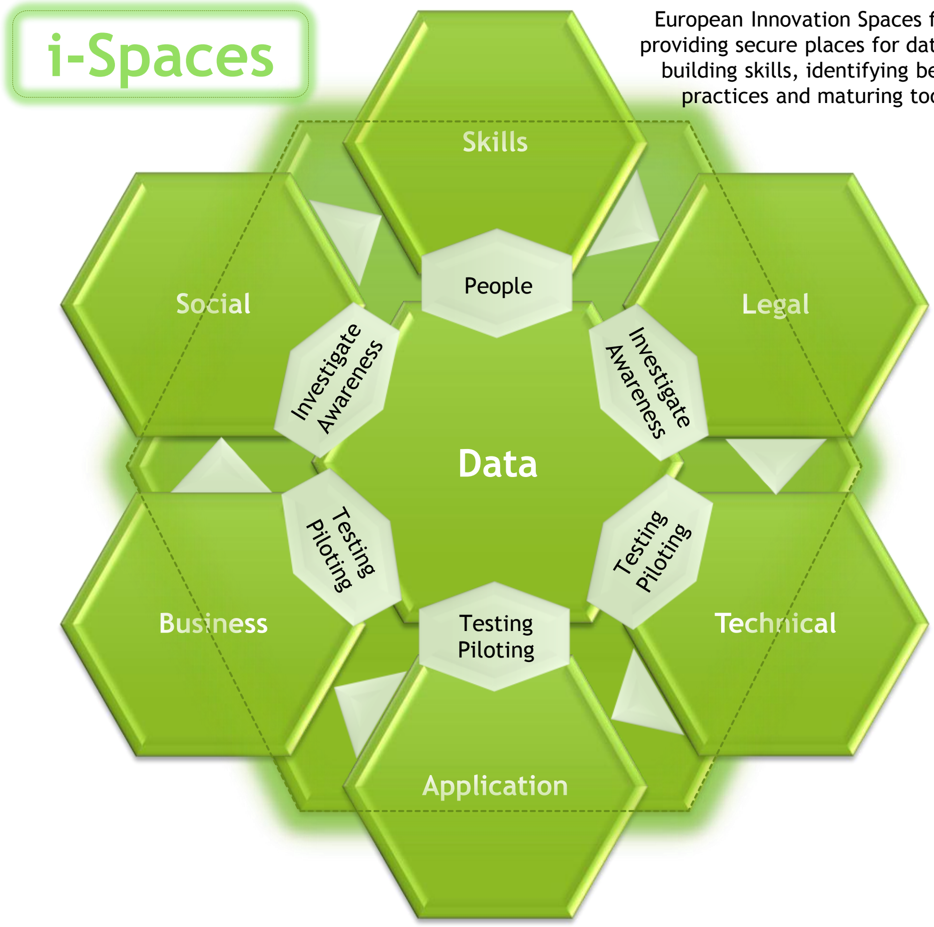

Figure 5: Interconnected Challenges of the cPPP with i-Spaces linking all

All i-Spaces will provide a set of basic services to support Lighthouse Projects, Technical Projects, as well as Collaboration and Coordination Projects. These basic services include:

- **Asset Support:** Supporting data providers in integrating data sets in a quality-secured way and maintaining a catalogue of available data assets
- **ICT Support:** Providing basic ICT assistance as well as focused support by Big Data scientists, Specialists, and business development during the runtime of research and innovation projects. This includes the assistance in benchmarking data sets, technologies, applications, services, and business models
- **On-boarding**: Running an on-boarding 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

- **Privacy:** Data protection and privacy including ensuring the compliance with laws and regulation as well as the deployment of leading-edge state-of-the art security technologies in protecting data and controlling the access to data
- **Federation:** Supporting the federation with other innovation spaces and experiments across multiple innovation spaces. This will especially help to support research and innovation activities accessing and processing data assets across national borders.

The i-Spaces should also be understood as incubator environments, where research outcomes on novel technologies and applications can be quickly tested and piloted in a context with maximum involvement of those who would ultimately be part of the respective ecosystems and develop new business with it and those that use it.

Summarizing, the main characteristics of the i-Spaces are:

- Being the **hubs for bringing the technology and application developments together** and cater for the development of skills, competence, best practices. These environments will offer 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 the Big Data Value activities. The i-Spaces will make accessible those data assets based on industrial, private and open data sources. The i-Spaces will be the secure and safe places that will ensure the availability, integrity, and confidentiality of the data sources.
- Serving as **incubators for testing and benchmarking** of technologies, applications, and business models. This will provide early insights on potential issues and will help to avoid failures in the later stages of commercial deployments. In addition, it can be expected that this activities will provide **input for standardization and regulation**.
- **Developing skills and sharing of best practices** will be an important task of the i-Spaces and their federation and linking with other existing initiatives at European and national level.
- New **Business Models and Ecosystems** will be emerging from exposing new technologies and tools to industrial and open data. The i-Spaces will be the playground to test new business model concepts and emerging ecosystems of existing and new “players”.
- Getting 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 eco-systems.

2.1.1 Setup of i-Spaces

To ensure that i-Space will achieve their ambitious objectives, the following design considerations will be taken into account when setting up i-Spaces:

- A well-managed **IT infrastructure**, including remote access capabilities;
- **Secure and trustworthy data hosting and access**;
- Availability of a team providing basic IT **assistance**, as well as focused **support** by Big Data experts.

Key elements for the implementation of i-Spaces include at least the following:

- **Secure access to data storage** that provides the security mechanisms required for having industry and other data asset owner trust in making their data available to scientists and data specialists for experimentation. At the same time, open data will be made available. . Together, this facilitates running experiments on site or remotely.

- **Offering hybrid computing models**. The Cloud paradigm will be one important computing model for Big Data Value technology and thus i-Spaces. Yet, it will not be the only valid model. For instance, due to the volume and velocity of data, transferring this data from data sources (such as IoT sensors) to Cloud providers might not be feasible. This means that i-Space infrastructures will also consider other computing models, such as "distributed computing", "high-performance computing", as well as "computing at the edge".
- **Delivering platforms and tools** from different sources including open source and proprietary for enabling scientist and data engineers to develop and run new technology and applications. It is envisaged that i-Spaces will start from the "state of the art" and continuously evolve incorporating new technology as it becomes available.

2.1.2 Innovation spaces a tool for continuous benchmarking

An important provision by the i-Spaces is to facilitate benchmarking so that businesses, in particular start-ups and SMEs, are able to evaluate whether their products and services will work in the real world. Hence, benchmarking is about comparing a specific product or service, for instance, with a peer product or service, respectively. Such comparison covers measures such as efficiency, effectiveness, cost, quality and return on investment.

Given the specific characteristics of BDV, i-Spaces focus on four strands of benchmarking:

- **Business (model) benchmarking**: This type of benchmarking may among other aspects focus on process, financial or investor perspective aspects.
- **Technical benchmarking**: This type of benchmarking is about determining how the performance or operational cost of a product or service compares with existing, other products or services.
- **User experience benchmarking**: Besides performance and cost, the customer and user experience of a product and service is key for success. The quality of the user centric approach when it comes to products and services is vital for a product and to become a success.
- **Data set benchmarking**: The data sets are at the core of i-Spaces. Measuring and ensuring data quality, not only on existing data sets, but also on live data streams, is the main concern to this strand.

For all the four strands mentioned above, i-Spaces will facilitate that benchmarking processes. Since benchmarking is not a one size fits all and is highly context sensitive, the organisation planning a benchmark will have a leading role in the process. The following elements therefore will offered by i-Spaces to support such benchmarking:

- Service for assisting in identifying what is to be benchmarked, and how to benchmark it.
- Service for assisting in identifying businesses with, for instance, a peer product or a peer service useful for benchmark measures.
- Service for assisting in identifying best practices and measures.
- Service for assisting in running the benchmark.

2.2 Lighthouse projects

Lighthouse projects run data-driven large scale demonstrations whose main objectives will be to create high level impact, and broadcast visibility and awareness driving towards faster uptake of Big Data Value applications and solutions. Lighthouse projects will propose replicable solutions by using existing technologies or very near to market technologies that could be integrated in an innovative way and show evidence of data value.

Lighthouse projects will be the major, high-impact mechanism for Europe to demonstrate Big Data Value ecosystems and sustainable data marketplaces that lead to increased competitiveness of established sectors as well as the creation of new sectors in Europe. Projects should prepare the path for other domains to replicate the experiences. Such relevance should lead to explicit business growth and job creation and thus, all projects will be required to define clear indicators and success factors that can be measured and assessed in both qualitative and quantitative terms against those goals.

2.3 Technical projects

Technical projects[33] focus on one or few specific aspects of technical priorities, thereby providing the technical foundation for lighthouse projects and i-Spaces.

2.4 Cooperation and coordination projects

Cooperation and coordination projects[34] will work on detailed activities that ensure coordination and coherence along the cPPP implementation and will provide support to activities that fall under the skill, business, policy, regulatory and legal as well as the social domains.

3 Technical Priorities

To identify the key technical priorities that need to be addressed to initiate the development of a European Data Value ecosystem, a two-way analysis was conducted. First, the most important challenges of relevant end users from various economic sectors were identified by performing a needs analysis. Within the already mentioned series of sectorial workshops, the user needs and requirements across a broad range of application domains have been identified. Second, this needs analysis was mapped onto the main roles of the different players in the data value chain. Third, these end users' needs have been cross-checked with available Big Data technical solutions to determine the main technical priorities and challenges to be addressed in order to satisfy those needs.

The technical priorities resulting from this analysis are presented in Section 3.2 to Section 3.6. Before going into the details of these priorities, Section 3.1 provides further background, rationale and a more in-depth description of the approach that was followed to determine these priorities.

[33] for instance Research and Innovation actions (RIAs) and Innovation actions (IAs)

[34]for instance Collaboration and Support Actions (CSAs)

3.1 Analysis and Identification of Technical Priorities

3.1.1 Current Situation and European Assets

Large US IT and Internet companies currently have an unquestionable lead on Big Data infrastructure & storage techniques. It does not seem the most efficient approach to try and overtake or compete with them by simply repeating what they have already achieved, but rather build on top of the commoditized core that they have established.

The fields of Big Analytics & Data Visualization (predictive and decision support systems) is much more open. The EU has an undeniable competitive advantage here, thanks to the very high mathematical and computer literacy level of EU engineers and research scientists as well as the solid base of industries which own most of the underlying data assets, unlike the end consumer data sets dominated by consumer-facing web companies in the US.

This positioning is a major factor of differentiation and a real asset. It is most important that the real added value of Big Data, in terms of scientific and technological innovation, as well as business, lies in application of intelligent decisions driven by the data analysis. Most of the supporting tools and storage architectures are now Open Source (Hadoop, Hive, Spark, Shark, HBase, Riak, Titan, etc.), levelling the playing field for tool vendors in this field.

To ease the development of Big Analytics and innovative data visualization solutions, it is nonetheless vital to ease the access to Big Data infrastructures (including hardware platforms but also the storage & data processing capabilities) which are pretty costly. But European countries have already understood this issue and some of them have already set up such platforms, precisely to nurture the ecosystem of SMEs in this area and favour the emergence of new business ideas. Such platforms need to be more than just simply technical platforms. They represent innovation spaces (see Section 2.1) where the different players in the Data Value chain, not only Big Data solution providers but also data asset owners, can meet and jointly develop new applications. This is the first step towards the development of a Big Data Value ecosystem. The cPPP will therefore not push for the development of a new technical platform at the European level. It will instead encourage initiatives taking benefits of the existing national innovation spaces and promote the collaboration between these spaces within larger scale projects. Besides, specific partnerships with the existing European innovation spaces will be established within this cPPP.

3.1.2 Needs and Stakeholder Analysis

As already mentioned, in order to systematically elicit the needs towards future Big Data solutions, sectorial workshops have been performed in various fields: geo-spatial/environment, energy, media, mobility, manufacturing, retail, health, public sector. From the analysis of the collected needs it is clear that to address the technical needs of these vertical application markets, a set of cross-sectors technologies are needed. The main technical concerns most often mentioned in these sectorial workshops were:

- Data Integration: Harmonization across different sources: standardized modelling, integration of heterogeneous data sources
- Data curation: handling veracity, life cycle management
- Availability for data in motion: Low latency and real-time data processing
- Advanced analytics: predictive analytics, graph mining, semantic analysis
- Data protection and privacy technologies: to make data owners comfortable about sharing data in an experimental environment

- Advanced visualization, user experience and usability;

To turn Big Data technologies into value, both supply and demand needs to be brought together for a mutual benefit and both of them will be enriched. While this will foster creating a more competitive Big Data “supplier” industry, it will also take care of developing a European market where benefits will be well documented across a wide range of industrial sectors. The impact of such transverse technologies goes well beyond the vertical sectors described as they require an "ecosystem" that will bring together stakeholders from the European Big Data community from both demand and suppliers sides, from legal, societal and technical sides.

Three major stakeholder roles relevant from a technical point of view can be identified in the data ecosystem:

Role	What do they do?	How do they make business?	Main technical concerns, needs
Data provider	Collect, pre-process, transform data into information and sell or distribute the information	Make a margin on the resale of information	Data curation and data integration from heterogeneous sources, Data availability
Data processor and service provider	Buy information, perform deeper analysis to create value and provide services.	Leverages scale effects across multiple clients, service fees	Need for low latency and data analytics with a good benefit/cost ratio, tools; flexibility to serve multiple clients, wide variety of data sources, Predictive analytics
Service consumer	Buy/use services	Applies decisions and insights derived from analysis to optimization of own business	Privacy and anonymisation, Advanced visualization

Table 1: Roles and activities of ecosystem actors

Figure 6 depicts the technical concerns according to the three main roles identified in the data ecosystem (cf. Table 1). There are mainly two types of technical issues:

- Some existing challenges become **more critical** and difficult due to the increasing complexity and need for lower latency.
- **New opportunities** to solve problems that couldn’t be solved previously arise due to the availability of novel technology. For instance, crowd sourcing is generally associated with a high risk for data integrity while it could actually be used as a novel means to perform consistency checks on information from different sources, therefore leading to enhanced data quality, management and integrity solutions.

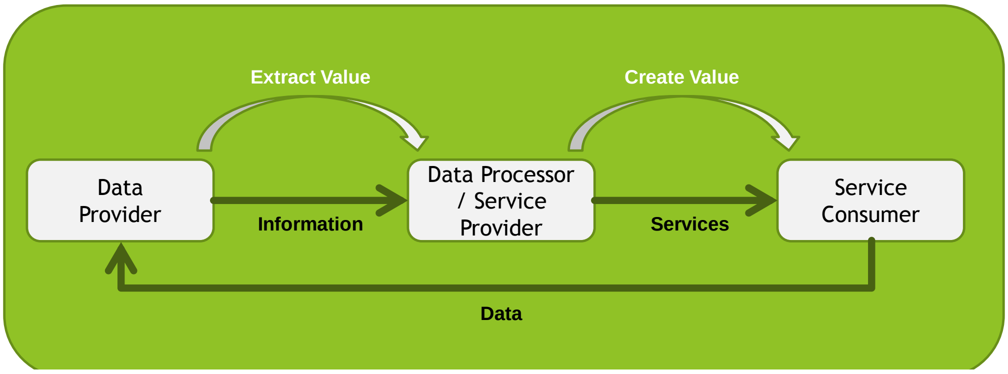

Figure 6: Various technical needs and concerns according to the role in the ecosystems

There are challenges to cope with the volume, velocity, variety, and veracity aspects of data analytics and to integrate novel statistical and mathematical algorithms, as well as prediction techniques into services and applications. This requires new approaches for engineering of data management solutions, advanced technologies for visualisation, veracity checks, anonymisation and deep analytics. All those are considered as strategic priorities. Real value may stem from the capability to deliver shorter and shorter response times, while analysing more and more complex systems and data sources. Organizations able to handle the increasing complexity and dynamicity of data structure and operations will thus gain a clear competitive advantage.

Based on the needs analysis, the overall, strategic technical goal may be stated as:

> Deliver new Big Data technology allowing for deepened analytics capacities on data-at-rest and data-in-motion while providing sufficient privacy guarantees, optimized user experience support and a sound data engineering framework.

Achieving this goal requires addressing at least the following technical priorities, which are elaborated in the remainder of this section:

- Engineering the management of data;
- Optimized and scalable architectures for analytics of both data at rest and in motion with low latency delivering real-time analytics;
- Deep analytics to improve data understanding, deep learning, meaningfulness of data;
- Privacy and anonymisation mechanisms;
- Advanced visualization and user experience.

3.2 Priority “Data Management”

Background

Over many years, several different application sectors have already tried to develop vertical processes for data management including specific data format standards and domain models. The ability to clearly define, interoperate, openly share, access, transform, link, syndicate, and manage data is, however, still missing.

In this perspective, it becomes crucial to have good metadata and semantic techniques to structure the data sets and content, annotate them, document the associated processes, and deliver or syndicate information to recipients. Requirements that need to be addressed include:

- **Annotation:** Data needs to be semantically annotated in digital formats, without imposing extra-effort to data producers
- **Unstructured data** (such as multimedia and other content) has to be pre-processed and enhanced by semantic annotation, overcoming multilingualism issues
- **Semantic Interoperability**: Data silos have to be unlocked by creating interoperability standards and efficient technologies for the storage and exchange of semantic data and tools to allow efficient user-driven or automated transformation
- **Legal Frameworks**: Technical means have to be backed by legal frameworks to ensure the transparent sharing and exchange of data
- **Quality:** Methods for improving and assessing the data quality have to be agreed and curation framework and workflows delivered
- **Syndication:** Data, information and content needs to be syndicated from Data Providers to Data consumers whilst maintaining provenance, control and source information including IPR considerations

Challenges

All main actors involved in the data value chain, from producers to processors and consumers recognize the value and urgency of the definition of an effective and engineered approach to the management of data. Collected data is rapidly increasing, but the methods and tools for data management do not evolve at the same pace. They are hindered by the lack of a broader and common framework according to a clearly defined Data as a Service (DaaS) model. The following research challenges have been identified:

- **Data Tsunami:** Sensors, social-data are exploding the data available: In the current pace of data evolution, important needs will be prominent in the areas of Cyber Physical Systems (CPS) and IoT: New means for data storage with novel data reduction techniques and new applications for their integration, query and analysis
- **Multilingualism**: Resources are often provided in English language only and thus difficult to align with data sources being collected in local languages. In order to join data for integrated approaches of data analysis and business intelligence the language barrier has to be climbed.
- **Data Silos**: In almost every sector isolated and fragmented data pools are found. In order to increase the impact of Big Data applications, technical means for the seamless integration and smart access to the various heterogeneous data sources is needed. In addition, the efficient collaboration between stakeholders (e.g. by relying on the same standards or frameworks or by migration/federation between data pools) is required as the final ROI relies to a high extent on the level of collaborative engagement of the various stakeholders
- **Interoperability:** Different data producers and consumers even in the same sector have different storage, communication and thus access issues. There is an urgent need to build an interoperability layer upon all different systems taking advantage of transformation and semantic integration techniques
- **Curation:** Data cleaning, integration, curation tools and services are required for data users to be able to differentiate noise from valuable data and to be able to integrate them and make them ready for analysis processes. The following will be required:
    - Improving data curation considering crowd-sourcing etc.
    - General-purpose data curation pipelines
    - On-line and Off-line data filtering techniques
    - Improved human-data interaction
    - Standardized data curation models and vocabularies
    - Better integration between data curation tools

- **Data Digitalization and semantic annotation**. Crowd sourced and automated means for semantically describing the content of unstructured data, (such as multimedia, reports, etc.) and facilitating interoperability will be required
- **Knowledge Repositories:** Annotations, Transformation, Ontologies and knowledge need to be stored and sharable to allow industry and citizens to linked their data with others to emancipate the user versus technical experts

Outcome

- **Data management and interoperability mechanisms** available and easily accessible as services. For instance:
    - Novel data management processing algorithms and data quality governance approaches that support the specifics of Big Data
    - Standardized data models and interoperable architectures for different sectors enriched through semantic terminologies
    - Standardized annotation framework for different sectors supporting the technical integration of different annotation technologies and data formats
    - Tackling the whole data management lifecycle, from data curation and cleaning (including pre-processing veracity, integrity, and quality of the data), to long term storage and data access. New models and tools to check integrity and veracity of data, through both machine-based and human-based (crowd-sourcing) techniques
- **Processing:**
    - BPM-oriented process mining integrated with Big Data management algorithms and data quality government approaches
    - Natural language processing for different languages
    - Algorithms for automatic detection of normal and abnormal structures (including automatic measuring)
    - Data integration processes, including APIs for transforming heterogeneous data and connecting heterogeneous data sources
    - Efficient data migration techniques, e.g. supporting data reduction for data in motion
- **Tools:**
    - Harmonization of tools and techniques with the ability to easily re-use, interconnect, syndicate, auto/crowd annotate and bring to real life data management use cases and services across sectors and citizens by diminishing the costs of developing new solutions
    - Tools for pre-processing and analysing sensor, social, geospatial, genomics, proteomics and other domain orientated data
    - User-friendly interfaces for data pre-processing
    - Advanced data access mechanisms to attach and link to other data/content as well as linking to other BDV and facilitating artefacts

3.3 Priority “Optimized Architectures”

Background

In the next 5 years, the Internet of Things (IoT) will become one of the most important drivers for Big Data. European companies, such as Ericsson or Nokia, have already established a leading role in this area and Europe is leading IoT application including automotive (e.g. BMW, Bosch), transport (e.g. Alstom, CAF, Siemens), Smart Cities, health (Siemens, Philips), aerospace (Thales, Airbus, Rolls Royce).

There is little research work that addresses how to integrate IoT technologies and protocols into Big Data platforms. These issues will affect future Big Data infrastructures, dealing with the specificities of IoT data flows, with multitudes of micro-messages to be processed continuously, is a true challenge for Big Data processing architectures. One of the technical solutions is certainly main memory processing, where some European Software Companies (SAP, Software AG) already have a leading role.

Developing such solutions in an ad-hoc fashion is of course possible but only the design of generic architectural solutions will leverage the true potential. Optimized frameworks and toolboxes allowing the best use of both data in motion (e.g. data streams from sensors) and data at rest will leverage the dissemination of reference solutions which are ready and easy to deploy in any economic sector. When such solutions become available to service providers, in a straightforward manner, they will have the opportunity to focus on the development of business models. This opens also an opportunity for Europe to have a leading role in the intersection of these two key fields: IoT and Big Data.

The capabilities of existing systems to process such streaming information and answer queries in real-time and for thousands of concurrent users are limited. Approaches based on traditional solutions like Data Stream Management Systems (DSMS) and Complex Event Processors (CEP), are generally insufficient for the challenges posed by IoT in Big Data scenarios. The problem of effective and efficient processing of data streams (data in motion) in a Big Data context is far from being solved, especially when considering the integration with data at rest and breakthroughs in noSQL databases and parallel processing.

Challenges

There have been advances for Big Data analytics to support new dimensions on Big Data volume (e.g. by NoSQL and Hadoop platforms, or storlets for object stores). Separately stream processing has been enhanced to analytics on the fly to cover the velocity part of Big Data. This is especially important as business needs to know what is happening now. The main challenges to be addressed are:

- **Analytic Techniques:** Being able to apply complex analytics techniques at scale and for data in motion is crucial in order to extract knowledge out of the data and develop decision support applications. For instance predictive systems like recommendation engines must be able to provide real-time predictions while enriching historical databases to continuously train more complex and refined statistical models
- **IoT:** Algorithms for IoT dataflows analytics
- **Performance:** The performance of the algorithms discussed above must be scaled by orders of magnitude while reducing energy consumption with the best effort integration between hardware and software
- **Integrated Processing of data in motion and data at rest:** Real-time Analytics and Stream Processing: which span across the areas of inductive reasoning (machine learning), deductive reasoning (inference), high performance computing (datacenter optimization, efficient resource allocation, quality of service provisioning) and statistical analysis, adapted to allow continuous querying over streams (i.e., on-line processing). New Big Data-specific parallelization techniques and (at least partially) automated distribution of tasks over clusters are crucial elements for effective stream processing

Outcome

- **Architectures for data in motion and data at rest:** Architectures, frameworks and tools for the integration of mostly existing components to new types of platforms, which address the orthogonal challenges in completely new ways, by widening and generalizing known data processing capabilities for data at rest and data in motion

- **Real Time Analytics:** that is capable of analysing large amounts of data-in-motion and data-at-rest by updating the analysis results as the information content changes
- **Software-defined Storage for Big Data:** Novel software-defined storage with built-in functionality for computation near the data (e.g. Storlets) and data reduction to support storing, sharing, and efficient in-place processing of the data

3.4 Priority "Deep Analytics"

Background

The progress of deep analytics of Big Data is expected to have the most significant impact, as it will make Big Data much more usable, and accessible to the wider public. Therefore, it is expected to positively influence all parts of the value-chain and increase not only business opportunities through business intelligence and analytics but also bring societal and citizen benefits.

Deep Analytics is still a field open to competition, in which Europe has strong competitive advantages and which is promising for business development. It was estimated that governments in Europe could save $149 billion[35] by using Big Data analytics to improve operational efficiency. At a smaller scale, big analytics can provide additional value in every sector where it is applied, leading to more efficient and accurate processes. A more recent study, also by the McKinsey Global Institute, has put an even stronger emphasis on this analytics issue, ranking it as the future main driver for the US economic growth, before shale oil and gas production[36].

Validating the vast amount of information potentially available in Big Data scenarios is a major challenge. There are a very large number of different types of sources with differentiated characteristics and levels of trust, such as blogs, social networking platforms, or news sites with social networking functionalities. The same applies for different types of content, such as articles, comments, tweets, etc. from different domains. In addition, this vast amount of data and the frequency of update, render the validation task technically challenging, especially in a cost-effective and economically sustainable way.

On the one hand there is a need to create complex and fine-grained predictive models on heterogeneous and massive datasets such as time series or graph data. On the other hand such models must be applied in real-time on large amounts of streaming data. This ranges from structured to unstructured, numerical to micro-blogs streams of data. The latter is extremely challenging because the data, besides its volume, is very heterogeneous and highly dynamic which also calls for scalability and high throughput. For instance, data collection related to a disaster area can easily occupy terabytes in binary GIS formats and data streams can show bursts of gigabytes per minutes.

Challenges

Understanding data, whether it is numbers, text, or multimedia content, has always been one of the greatest challenges for ICT. Entering into the era of Big Data this challenge has scaled to a degree that makes the development of new methods necessary. In the following we detail the research areas identified:

- **Graph Data:** Predictive modelling, deep learning techniques and graph mining techniques applied on extremely large graphs. As graph data are so widespread and affect almost all sectors this should definitely be considered as a high priority challenge

---

[35] "*Big Data: The next frontier for innovation, competition and productivity*", McKinsey Global Institute, June 2011
[36] "*Game changers: Five opportunities for US growth and renewal*", McKinsey Global Institute, 2013

- **Context:** Context awareness to improve the quality of existing problem solving. In the Big Data context, contextualization will combine heterogeneous data and data streams to improve the quality of mining processes or classifiers. In fact, in the last years, context awareness has demonstrated its crucial role to achieve optimized management of resources, systems, and services in many application domains, from mobile and pervasive computing to dynamically adaptive service provisioning. Context-aware Big Data solutions will exploit context awareness to focus on only some portions of data (at least in first approximation) by keeping high probability of hit for all application-relevant events, with manifest advantages in terms of cost reduction and complexity decrease
- **Content Validation:** Implementation of veracity (source reliability / information credibility) models for validating content and exploiting content recommendations from unknown users
- **Frameworks:** New frameworks and open APIs for the quality-aware distribution of stream processing tasks over datacenter resources, with minimal development effort from Big Data application developers and domain experts
- **Processing:** Improvement of the scalability and processing speed for the aforementioned algorithms in order to tackle linearization and computational optimization issues
- **Business Analytics:** All the above items enable the realisation of real and static business analytics and business intelligence empowering business and other organisations to make accurate and instant decisions to shape their market. The simplification and automation of this is necessary especially for SMEs

Outcome

The main expected advanced analytics innovations are the following:

- **Event Space:** Move beyond limited samples used so far in statistical analytics to samples covering the whole or the largest part of an event space
- **Model Accuracy:** Improve the accuracy of statistical models by enabling fast non-linear approximations in very large datasets
- **Event Discovery:** Discover rare events that are hard to identify since they have a small probability of occurrence, but have a great significance (such as physical disasters, a few costly claims in an insurance portfolio, rare diseases and treatments)
- **Semantic Analysis:** Deep learning, contextualization based on IA, machine learning, semantic analysis in near-real time, graph mining
- **Unstructured Data:** Processing of unstructured data (multi-media, text). Linking and cross-analysis algorithms to deliver cross-domain and cross-sector intelligence
- **Canonical forms:** Provide canonical paths so that data can be aggregated and shared easily without dependency on technicians or domain experts and provide a path for the smart analysis of data across and within domains
- **Business Intelligence:** Coupled with the priorities on visualisation and engineering to provide user-friendly tools which connect to open and other data sets and streams (including a citizens data), provided intelligent data interconnection for business and citizen orientated analytics, and allow visualisation

3.5 Priority "Privacy and Anonymisation Mechanisms"

Background

Privacy and data anonymisation is one of the major concerns in the area of Big Data and data analytics involving all stakeholders in the value chain. Data privacy and security is indeed often a main hurdle, which prevents data owners from joining Big

Data innovation environments. A similar statement can be made for citizens who are more and more seriously taking into account privacy guarantees.

There are a number of current research areas including differential privacy, private information retrieval and homomorphic encryption, which appear to be promising, but are, unfortunately, not yet suitable for commercial, large-scale processing. For instance, the framework CASD[37], which was already successfully implemented at the French Innovation Space TERALAB, can be used as starting point to develop a more generic data protection approach.

Risk based approaches regarding privacy must be considered especially when dealing with combined processing of multiple data sets. It has indeed been shown that when processing combinations of anonymised or pseudonymised data sets there is a risk that identities can be retrieved. Thus providing tools to assess the risk associated with such a processing is an issue of significant importance.

Challenges

Data usage should conform to the current legislation and policies. This is difficult to be assured. On the technical side, mechanisms are needed in order to provide the data owners with the means to control the access and usage of their data throughout its whole lifecycle. Ideally the privacy and anonymity of data should be ensured while Big Data analytics is in operation. Citizens for example should be able to decide on the destruction of their personal data (right to be forgotten).

With the integration of multiple data sources, the opportunities to crack and reverse the anonymisation process greatly increase. Ensuring irreversibility of the anonymisation of Big Data assets is a key Big Data issue. Scalability of the solutions is also a critical feature. Preserving anonymity often implies removing the links between data assets. However, the approach to preserve anonymity also has to be reconciled with the needs for data quality, on which this link removal has a very negative impact. Anonymisation is an important challenge, but it also implies other challenges, such as the need for data analytics to cope with anonymised datasets.

Outcome

- **Complete Data Protection Framework**: All data protection mechanism for innovation spaces by, for instance, generalising the CASD framework
- **Cloud Data Protection**: Protect the cloud infrastructure, analytics applications, and the data from leakage and threats
- **Data minimisation**: Methods for secure deletion of data and data minimization
- **Algorithms**: Robust anonymisation algorithms
- **Reversibility:** Risk assessment tools to evaluate the reversibility of the anonymisation mechanisms
- **Mining Algorithms:** Developed privacy-preserving data mining algorithms
- **Pattern Hiding:** Design of mechanisms for pattern hiding so data is transformed in a way that certain patterns cannot be derived (via mining), while others can
- **Multiparty Mining:** Secure multiparty mining mechanisms over distributed datasets, so data on which mining is to be performed is partitioned, horizontally or vertically, and distributed among several parties. The partitioned data cannot be shared and must remain private but the results of mining on the “union” of the data are shared among the participants

[37] a French acronym which stands for Secure Remote Data Access Centre

3.6 Priority "Advanced Visualisation and User Experience"

Background

Data visualisation is vital if people are to effectively consume Big Data. Reports generated from data analytics can be thought of as complex documents which contain varying forms of media for the end-user, including text, charts and visualisations. In order for users to quickly and correctly interpret these large and complex documents, carefully designed presentation and digital visualisation is required.

When representing complex information on screen(s), the design issues multiply rapidly. Complex information interfaces need to be humane[38], i.e., responsive to human needs and capacity. Knowledge workers need relevant information in a just-in-time manner. Too much information, which cannot be efficiently searched, can hide the information that is most relevant. In fast moving time constrained environments they need to be able to quickly understand the relevance and relatedness of information.

Challenges

In the data visualisation domain, the tools that are currently used to communicate information need to be improved due to the significant changes brought about with the volume and variety of Big Data. Advanced visualisation techniques must consider this variety (i.e. graphs, geospatial, sensor, mobile, etc) of data available from diverse domains. Tools need to support capabilities for the exploration of unknown and unpredictable data within the visualisation layer.

Access to information is at present based on a user-driven paradigm: the user knows what they need and the only issue is to define the right criteria. With the advent of Big Data, this user-driven paradigm no longer proves to be the most efficient. Data driven paradigms will emerge where information is proactively extracted through data discovery techniques and systems are anticipating the user's information needs.

There are significant challenges in visual analytics in the area of multiple scale data. Appropriate scales of analysis are not always clear in advance and single optimal solutions are unlikely to exist. Interactive visual interfaces have a great potential for facilitating the empirical search for the acceptable scales of analysis and the verification of results by modifying the scale and the means of any aggregation.

What is needed is an evolution of visual interfaces towards becoming more intuitive and exploiting the advanced discovery aspects of Big Data analytics. This is required in order to foster effective exploitation of the information and knowledge that Big Data can deliver.

Outcome

The main expected advanced visualisation and user experience are the following:

- **End User Centric:** Adaptation to the needs of end users (user adaptation and personalization but also advanced search capabilities) rather than predefined visualization and analytics. User feedback should be as simple as possible
- **Scale:** In order to handle extremely large volumes of data, interaction must focus on aggregate data at different scales of abstraction rather than on individual objects. There is a need to develop corresponding analysis supporting interaction techniques, which should enable easy transitions from one scale or form of

[38] "*The Humane Interface: New Directions for Designing Interactive Systems"*, Raskin, J. Addison-Wesley, Reading, MA, 2000

aggregation to another (e.g. from neighbourhood-level to city-level) while supporting aggregation and comparisons among different scales
- **Clusters:** Dynamic clustering of information based on similarity or relatedness to the problem rather than on individual categories
- **Geospatial:** New visualisation for data with geo-locations, distances, and space/time correlations (i.e. sensor data, event data)
- **Interrelated Data:** Rather than data islands, visual interfaces must take account of semantic relationships, relying on interactive graph visualization techniques to allow easy exploration of network structures.
- **3-D Visualization:** Real-time and collaborative 3-D visualization techniques
- **Qualitative Analysis:** With large data volumes, qualitative analysis at a high semantic level must augment classical quantitative approaches
- **Time:** Necessity to take into account the specifics of time[39]. In contrast to common data dimensions which are usually "flat", time has an inherent semantic structure and a hierarchical system of granularities which must be addressed
- **Plug and Play:** User adaptable interactive visualization tools that support the combination any visualization asset in a real-time plug-and play manner – for instance: maps, graph visualization, dashboards

3.7 Roadmap and Timeframe

In order to achieve the overall, strategic technical goal laid out in Section 1.3 and to address the aforementioned technical priorities, Table 2 depicts a potential roadmap defining specific tasks and expected outcomes that build on top of each other.

Priority	Year 1	Year 2	Year 3	Year 4	Year 5
Data Manage-ment	Mechanisms for integration of hetero-geneous data sources Semantic based data and content interoperability Generalisation of secure remote data access techniques	Collaborative Tools and techniques for Data Quality (including integrity and veracity check) Harmonized description format for meta-data and for data reduction	Methodology, models and tools for data lifecycle management	Data management as a service	
Optimized Architect-ures	Specialized (software-defined) architectures and infrastructures for large data-sets and analytics	Optimized tools for the integration of existing components to new types of platforms with both data at rest and in motion.	Enhancing integrated batch and stream processing architectures for IoT		

[39] "*Space, Time, and Visual Analytics* ", G.Andrienko, N.Andrienko, U.Demšar, D.Dransch, J.Dykes, S.Fabrikant, M.Jern, M.-J.Kraak, H.Schumann, C.Tominski , International Journal Geographical Information Science, 2010, v.24 (10), pp. 1577-1600

Priority	Year 1	Year 2	Year 3	Year 4	Year 5
Deep analytics	Improved statistical models by enabling fast non-linear approximations in very large datasets	Real-time analytics Predictive modelling and graph mining techniques applied on extremely large graphs	Semantic analysis in near-real-time Algorithms for multimedia data mining	Descriptive language for deep analytics	Deep learning techniques
Privacy and Anonymi-sation	Complete Data Protection framework	Method for deletion of data and data minimization	Robust anonymisation algorithms		
Advance isuali-sation and User Experience		End-user Centric data search and solutions paradigms	Semantic driven data visualisation	Integration of analytics and visualization	Contextuali-sation Collaborative real-time, dynamic 3D solutions

Table 2: Timeframe of the major expected outcomes

4 Non-Technical Priorities

The portfolio of activities of the Big Data Value SRIA needs to comprise support actions that address complementary, non-technical issues alongside the European Innovation Spaces, Lighthouse projects, and the research and innovation activities. Beside the activities addressing the governance of the cPPP[40], the non-technical activities will focus on:

- Skills development
- Business Models and Ecosystems
- Policy, Regulation and Standardization
- Social perceptions and societal implications

4.1 Skills development

In order to leverage the potential of Big Data Value, a key challenge for Europe is to ensure the availability of highly and rightly skilled people who have an excellent grasp of the best practices and technologies for delivering Big Data Value within applications and solutions. In addition to meeting the technical, innovation, and business challenges as laid out in this document, Europe needs to systematically address the need for educating people that are equipped with the right skills and are able to leverage Big Data Value and so enabling best practices. Education and training will play a pivotal role in creating and capitalizing on EU-based Big Data Value technologies and solutions.

At this early stage of the Big Data discipline, we see two sub-disciplines emerging that require two distinct breeds of skills and expertise: **Data Scientists** and **Data**

[40] Which are described in detail in the Big Data Value cPPP proposal.

Engineers (as will be elaborated below). In fact, this is very similar to what happened to the software discipline in the years since the seminal NATO conference on Software Engineering in 1968. In software engineering there are now two principle, complementary types of specialists: (1) computer scientists, who are concerned with theoretical foundations and basic technology for creating software; (2) software engineers, who are concerned with establishing principles, tools, methods and sound engineering principles to efficiently and effectively develop, maintain and evolve software.

Drawing on this similarity with the software field, we currently, see a first trend of two important Big Data Value related skill sets[41]:

Data Scientist*:* Successful Data Scientists will require solid knowledge in statistical foundations and advanced data analysis methods combined with a thorough understanding of scalable data management, with the associated technical and implementation aspects. They will be the specialists that can deliver novel algorithms and approaches for the Big Data Value stack in general, such as advanced learning algorithms, predictive analytics mechanisms, etc. For this, Europe needs new educational programmes in data science as well as ideally a network between scientists (academia) and industry that foster the exchange of ideas and challenges. Hence, innovation spaces could be used to a certain extent to build such networks.

Data Engineers*:* Those are the specialists that develop and exploit techniques, processes, tools and methods for developing applications that actually turn data into value. In addition to technical expertise, Data Engineers need to understand the domain and the business of the organizations. This means they need to bring in domain knowledge and are thus working at the intersection of technology, application domains and business. In a sense they thereby constitute the link between technology experts and the business analysts. Data Engineers will foster the development of Big Data applications from an “art” into a disciplined engineering approach. Data Engineers thereby allow the structured and planned development and delivery of customer-specific Big Data solutions, starting from a clear understanding of the domain, as well as customer and user needs and requirements.

Extensive experience and skills acquired by working on projects in the specified technical priority areas of the SRIA will guide the identifying of skill development requirements that can be addressed by collaborating with higher education institutes and education providers to support the establishment of:

- New educational programmes in data science and data engineering based on an 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 Engineers and Data Scientists
- Foundational modules in data science, statistical techniques, and data management within related disciplines such as law and 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 enhances the industrial relevance of courses

The regularly updated strategic challenge areas will provide orientation for the development of required data skills to support building extensive know-how (e.g. by

[41] Please note that, as always in novel fields, there are many different, even contrary definitions out there; e.g., some further consider a data analysts being a specific additional type of specialist (in our case we subsumed the competencies in our definition of data engineer); some flip the definitions of data engineer and scientists altogether.

European curricula and sharing of best practices) and skills in Europe for future systems in the industry and research community.

4.2 Ecosystems and Business Models

The Big Data Value ecosystem (see Figure 7) will comprise many new stakeholders. New concepts of data collecting, processing, soring, analysing, handling, visualisation and most importantly the usage will be found and business models created around it. Identifying valid and sustainable business models and eco-systems around sectors or platforms will be challenging. Particularly for the many SMEs involved in specific, if not niche, roles will need support to align and adapt to the new value chain opportunities.

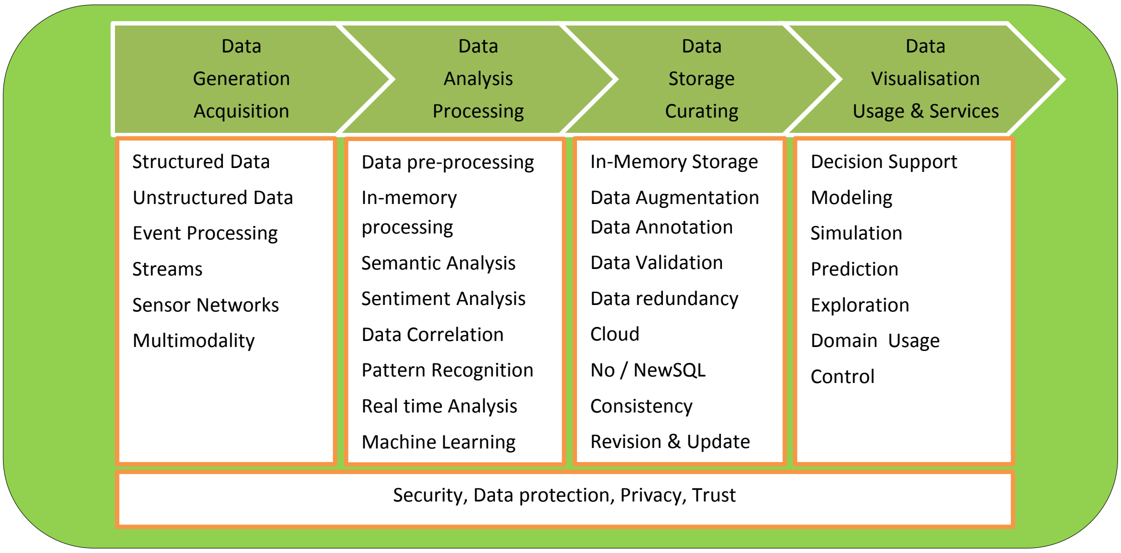

Figure 7: Big Data Ecosystem along the Value Chain

Dedicated projects for investigating and evaluating business models will be connected to the innovation spaces where suppliers and users will meet. Those projects will:

- Establish a mapping of technology providers and their value contribution
- Identify mechanisms by which data value is determined and value is established
- Provide a platform for entrepreneurs and financial actors including venture capitalists to identify appropriate levels of value chain understanding
- Describe and validate business models that can be successful and sustainable in the future data driven economy

The outcomes of these projects will foster the creation of a more stable business environment that enables business, particularly web entrepreneurs and SMEs, to enter the Big Data markets and ecosystems.

Europe needs to establish strong players in order to make the whole Big Data Value ecosystem, and consequently Europe's economy, strong, vibrant and valuable. The following **key stakeholders** are seen as actors along the Big Data Value chain:

- **User Enterprises**, e.g. enterprises in all sectors and of all sizes that want to improve their services and products using Big Data technology, data products and services
- **Data Generators and Providers** that create, collect, aggregate, transform and model raw data from various public and non-public sources and offer it to customers
- **Technology Providers** that provide tools & platforms that offer data management and analytics tools to extract knowledge from data, curate and visualize it

- **Service Providers** that develop Big Data applications on top of the tools and platforms to provide services to the user enterprises

In addition, the following organisations and communities will have impact on data-driven ecosystems that are building on top of the Big Data Value chain:

- **Regulatory bodies** to define privacy and legacy issues related to the data usage
- **International/national dejure and defacto standardisation bodies** in order to promote new concepts, systems and solutions for global adoption in international standards
- **Collaborative networks** where different players in the value chain collaborate to offer value services to their customers based on data value creation

Furthermore, the known stakeholders in H2020 along the phases of research, innovation, exploitation, and usage will play one or more roles of the Big Data Value chain:

- **Industry-Large** as providers of technologies and services who can also become users
- **Industry-SMEs** to provide particular know how and innovative solutions for specific concepts
- **Universities** that research on new algorithms and technologies to be applied in tools & platforms
- **Research centers** which research on new algorithms, methodologies and which define new business cases

4.3 Policy, Regulation and Standardization

4.3.1 Input to policy making and legal support

The cPPP has no mandate or competence to be involved directly in policy making for legal or regulatory framework conditions. However, still the cPPP needs to contribute to the policy and regulatory debate about the non-technical aspects of the future Big Data Value creation as part of the data-driven economy. Dedicated projects have to address the circumstance of new data ownership and usage, data protection and privacy, security, liability, cybercrime, Intellectual Property Rights (IPR), etc.

Those projects will initiate activities that are foreseen for exchange between stakeholders from industry, end users, citizen and society to develop input to ongoing policy debates where appropriate. Equally it will identify the concrete legal problems for actors in the Value Chain particularly SMEs who have no legal resources. This will establish a body of knowledge on the legal issues with a help desk for the project participants and ultimately for the wider community. The mentioned projects will:

- Establish an inventory of roadblocks inhibiting a flourishing data driven economy, e.g. by materializing the value of Big Data collections
- Make and collect observations about the discovery of new legal and regulatory challenges along with the implementation of state-of-the-art technology and the introduction of new technology
- Create a catalogue of legal practice in European Member States and other OECD countries and critical issues for BDV actors
- Establish a Big Data help Desk and prepare for legal clinics to give advice on legal issues

By doing so these projects will contribute from the perspective of developments of novel technology and solutions and will have direct contact with the actors to help legislators and regulators make exhaustive considerations about framework conditions. Furthermore these projects will support the BDV actors particularly SMEs to get around legal barriers to integrate into new eco-systems.

4.3.2 Standardisation

Standardisation is essential to the creation of a Data Economy and the cPPP will support establishing and augmenting both formal and de-facto standards. The cPPP will achieve this by:

- Leveraging existing common standards as the basis for an open and successful Big Data market
- Integrating national efforts on an international (European) level as early as possible
- Ensuring availability of experts for all aspects of Big Data in the standardisation process
- Providing education and education material to promote developing standards

Standards play a pivotal role on any market to provide customers with true choice by being able to choose comparable and compatible goods or services from multiple suppliers. In the Big Data ecosystem this applies to both the **technology** and to the **data**.

Technology Standardisation: Most technology standards for Big Data processing technology are *de facto* standards that are not prescribed (but at best *de*scribed after the fact) by a standards organisation. However, the **lack of standards is a major barrier**. One example are NoSQL databases. The history of NoSQL is based on solving specific technologies challenges that lead to a range of different storage technologies. The large range of choices coupled with the lack of standards for querying the data makes it harder to exchange data stores as it may tie application specific code to a certain storage solution. The NoSQL databases are designed for scalability, often by sacrificing consistency. Compared to relational databases, they often use a low level, non-standardized query interface that makes it harder to integrate in existing applications that expect an SQL interface. The lack of standard interfaces also makes it harder to switch vendors. While it seems plausible to define standards for a certain type of NoSQL databases, creating one language for different NoSQL database types is a hard task with an unclear outcome. The cPPP would take a pragmatic approach to standardisation and would look to influence besides NoSQL databases the standardisation of technologies such as complex event processing for real-time Big Data applications, languages to encode the extracted knowledge bases, computation infrastructure, data curation infrastructure, query interfaces, and data storage technologies.

Data Standardisation: Data “variety” of Big Data makes it very difficult to standardise. Nevertheless, there is a lot of potential for data standardisation in the areas of data exchange and data interoperability.

Big Data is valuable for any organisation across many sectors. Exchange and use of data assets is essential for functioning ecosystems and the data economy. Enabling the seamless flow of data between participants (i.e. companies, institutions, and individuals) is a necessary cornerstone of the ecosystem.

To this end the cPPP would undertake collaborative efforts to support, where possible and pragmatic, the definition of semantic standardized data representation ranging from domain (industry sector) specific solutions, like domain ontologies to general concepts such as Linked Open Data. If such standards for data descriptions and meta-data could be established, it would simplify and reduce the cost of data exchange. Insufficiently described data formats, which are a barrier for global & efficient data exchange and processing, are then eliminated.

4.4 Social perceptions and societal implication

Big Data will provide solutions for major societal challenges in Europe. For an accelerated adoption of Big Data it is critical to increase awareness of the benefits and the Value that Big Data offers and to understand how trust can be built up and privacy concerned built into the solutions and services. While the solution for the societal challenges will be addressed in dedicated projects:

- Look into the lack of trust in the Big Data Value technology, applications and solutions
- Establish a competence about impact on trust that can be made by technology changes
- Address privacy-by-design principles and create a common understanding amongst the technical community
- Identify key privacy concerns and develop answers based on new solutions
- Work towards a clearer profile of the social benefits that Big Data Value technology can provide.

These projects will assure that the citizen’s views and perception is taken into account so that technology and applications are not developed without a chance to be widely accepted.

5 Expected Impact

5.1 Expected Impact of strategic objectives

The expected impact of the cPPP should be recognised in the great enhancement that Big Data analysis techniques will provide to all decision-making processes. From this point of view every sector private or public, industrial or academic, will be touched as will be the society. The cPPP will show that Big Data Value is not just a new buzzword, but shorthand for advancing trends in technology that open the door to a new approach to understanding the world and making decisions.

The general impact of the cPPP is expected in the following lines:

- **Effective service provision** from public and private organisations will be achieved to develop and make available to industry and public sector technology, applications and solutions for creation of value from Big Data for increased productivity, optimised production, more efficient logistics (inbound and outbound)
- **Extensive experience and skills will be acquired** and IPR base will be set up to support building extensive know-how (e.g. by European curricula and sharing of best practices) and skills in Europe for future systems in the industry and research community
- **New Business Models and Optimisation** of existing industries will be established to drive the integration of the BDV services into private and public decision making systems such as Enterprise Resource Planning and marketing systems

Significant impact is expected on society with opportunities for a wide range of applications:

- **Big Data Value technologies** will be a key contributor to solutions on current major societal challenges, in areas such as health, demographic change, climate change, transport, energy, and cities. Novel Big Data technologies will provide insight on the different aspects of the big challenges and unlock new potential to address them. Similarly, BDV is associated other areas such as the Future Internet or the Internet of Things emerging market where the expected integration and huge volume of data need to be supported by solid data-orientated technologies. All these solutions will lead to a transformation of our everyday lives with direct impact on the individual's behaviour and habits. In the future, citizens can expect for example benefits from a more personalized healthcare system, novel decision-support systems for their everyday life or new ways to interact with companies and administrations, based on Big Data Value solutions

- **Availability of public government information and open data** will influence educational and cultural services. Large databases containing information on cultural heritage such as digitalized books and manuscripts, photos and paintings, television and film, sculpture and crafts, diaries and maps, sheet music and recordings will be made available and allow for new ways of educating people and novel forms of interaction between people and across cultural borders
- **Big Data technology will improve societal insight** on individual and society behaviour. Such technologies may allow for greater fact-based decisions in politics and the economy. Fundamental research will be deeply impacted by the availability of such Big Data resources and analysis, providing new insights and new development in many areas such as biology, physics, mathematics, material, energy. These developments themselves will provide new Big Data and enhance further societal developments
- **Collaboration**. Big Data Value will help to improve collaboration by providing access to various data sources such as media content, traffic flow, etc. Better services and collaboration will be possible for instance for emergency and crisis situations. Individuals will be empowered by their new role as co–creator or co-innovator as well as generator and provider of personal data

Industry surveys show that the gains from Big Data Value are expected across all sectors, from industry and production to services and retail. The following are examples of sectors that are especially promising with regard to Big Data Value.

- **Environment:** Better understanding and managing of environmental and geospatial data is of crucial importance. Environmental data helps to understand how our planet and its climate are changing and also addresses the role humans play in these changes. For example, the European Earth observation programme, Copernicus, aims to provide reliable and up-to-date information on how our planet's climate is changing to provide a foundation, which will support the creation of sustainable environmental policies. In addition, the EU project Galileo will offer a global network of satellites providing precise timing and location information to users on the ground and in the air. The overall intention is to improve the accuracy and availability of location data to the benefit of for example the transport and industry sectors as well as Europe's new air-traffic control system
- **Energy**: The digitization of the energy system from production, to distribution, to smart meters at the consumer, enables the acquisition of real-time, high-resolution data. Coupled with other data sources, such as weather data, usage patterns and market data, accompanied with advanced analytics, efficiency levels can be increased immensely. Existing grid capacities could be better utilized and renewable energy resources could be better integrated
- **Mobility, transport and logistics**: Urban multimodal transportation is one of the most complex and rewarding Big Data settings in the logistics sector. In addition to sensor data from infrastructure, vast amounts of mobility and social data are generated by smart phones, C2x technology (communication among and between vehicles) and end users with location-based services and maps. Big Data will open up opportunities for innovative ways of monitoring, controlling and managing logistics business processes. Deliveries could be adapted based on predictive monitoring, using data from stores, semantic product memories, internet forums, and weather forecasts, leading to both economic and environmental savings

- **Manufacturing and production**: With industry's growing investments into smart factories with sensor-equipped machinery that is both intelligent and networked (Internet of Things, Cyber-Physical Systems), the production sectors in 2020 will be one of the major producers of (real-time) data. The application of Big Data into this sector will bring efficiency gains and predictive maintenance. Entirely new business models are expected since the mass production of individualized products becomes possible and for which consumers may have direct access to influence and control
- **Public Sector**: Big Data Value will contribute to increase efficiency in public administrations processes. The continuous collection and exploitation of real-time data from people, devices and objects will be the basis for smart cities, where people, places and administrations get connected through novel ICT services and networks. In the physical and the cyber-domain, security will be significantly enhanced with Big Data techniques; visual analytics approaches will be used to allow algorithms and humans to cooperate. From financial fraud to public security, Big Data will contribute to establish a framework that enables a safe and secure digital economy
- **Healthcare**: Applications range from comparative effectiveness research to the next generation of clinical decision support systems, which make use of comprehensive heterogeneous health data sets as well as advanced analytics of clinical operations. Of particular importance are aspects such as patient involvement, privacy and ethics
- **Media and Content**: By employing Big Data analysis and visualisation techniques, it will be possible to allow users to interact with the data, and have dynamic access to new data as they appear in the relevant repositories. Users would also be able to register and provide their own data or annotations to existing data. The environment will move from few state-orientated broadcasters to a prosumer approach and where data and content is linked together blurring the lines between data sources and modes of viewing. Content and information will find organisations and consumers rather than vice versa and overall the impact will be a seamless content experience
- **Financial services**: Huge amounts of data are processed to detect fraud and risk, to analyse customer behaviour, segmentation, trading, etc. Big Data analysis and visualization will open up new use cases and permit new techniques to be realised. Possibilities include managing regulation, reporting, audits and compliance, and automatic detection of behaviour patterns and cyber-attacks. Open sources of information can be combined with proprietary knowledge to analyse competitive positions, and recommendation engines will be able to identify potential customers for products
- **Telecommunications services**: Big Data enables improved competitiveness by transforming data into customer knowledge. Possible use cases can include improvement of service levels, churn reduction, services based on combining location with data about personal context, and better analysis of product and service demand
- **Retail**: Digital services for customers provided by smart systems will be essential for the success of future retail business. The retail domain will especially be focused on highly efficient and personalized customer assistance services. Retailers are currently confronted with the challenge to meet the demand of a new generation of customers who expect information to be available anytime and anywhere. New intelligent services that make use of Big Data will allow a new level of personalized and high-quality Efficient Consumer Response (ECR)
- **Tourism**: Personalized services to tourists are essential for creating real experiences and consolidate a powerful European Market. The analysis of real-time and context-aware data with the help of historic data will provide customized information to each person and contributes to a better and more efficient

management of the whole tourism value chain. The application of Big Data in this sector will enable new business models and services

5.2 Monitoring of objectives

Big Data Value generation and the technology for it will have a tremendous impact on industry and economy as a whole. In terms of measuring this impact there are two basic types of measurements and related indicators:

- **Indirect Monitoring**: The monitoring is done using indicators which cannot directly be influenced or monitored by activities resulting from this Big Data Value SRIA. Typically, the monitoring is based on tracking the progress of some developments and is using comparison rather than specific numbers or targets. For example, the proposed Big Data Value activities can provide a research and innovation ecosystem but ultimately jobs, sales information and business progress will be under the control of individual organisations. Indirect indicators include for instance economic and usage information. The success of the SRIA strategy will be mainly measured based on indirect indicators
- **Direct KPIs:** Key Performance Indicators (KPIs) which are directly related to the performance of the SRIA activities themselves and which are clearly measurable – for example providing solutions to the technical priorities or the stimulation of SME participation in research and innovation activities and i-Spaces

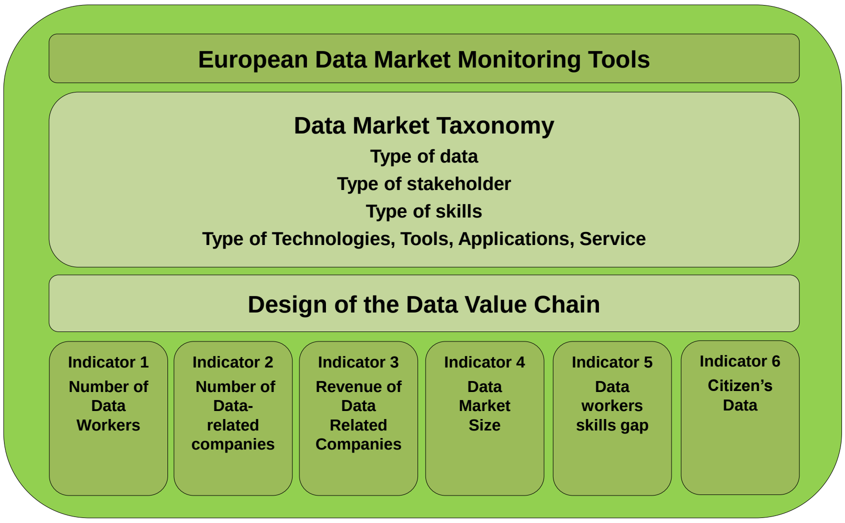

Figure 8: Preliminary European Data Market Indicators; IDC, 2014

According to the strategic and specific objectives of the Big Data Value SRIA described in Section 1.3, an interdisciplinary and holistic approach will be followed. Consequently, the indicators to be used for measuring the impact of the SRIA have to address strategic, social, competitiveness, and innovation aspects.

Indicators to measure the achievement of the strategic objectives

The development of the BDV market will be pushed by new, innovative and novel products. However, the success of those developments depends heavily on various market conditions and the overall economic climate. IDC[42] has proposed some preliminary European Data Market Indicators which are shown in Figure 8. The SRIA proposes to use this kind of market metrics as indirect indicators for monitoring its strategic impact. This monitoring will need to be adapted in alignment with currently ongoing work such as that in IDC on establishing metrics for the BDV market.

Strategic Indicators		Societal	Competitiveness	Innovation	Operational
KPI.S.1	**Number of Data Workers in different sectors, domains and sub-professions**				
KPI.S.2	**Economic impact on productivity within the EU and comparison to other geographical areas**				
KPI.S.3	**Development of the positioning of European companies in the ranking of leading global BDV companies**				
KPI.S.4	**Number of SMEs and web entrepreneurs dealing with Big Data services and products**				
KPI.S.5	**Number of SMEs and web entrepreneurs in the cPPP and BDVA**				
KPI.S.6	**Year-on-Year increase of number and % SMEs and web entrepreneurs in the BDV Implementation at R&I as well as user level**				
KPI.S.7	**Big Data market revenues in Europe and globally including market share of EU industry**				
KPI.S.8	**Deployment of Big Data technology in industry, public sector and its use by citizen**				
KPI.S.9	**Skill gap of workers, graduates and scientists including comparisons with other geographies**				
KPI.S.10	**Inclusion of Citizens in the BDV ecosystem as well as Citizens' contribution of data (where they approve)**				

Direct KPIs to measure the achievement of the specific objectives

The SRIA activities will deliver solutions, architectures, technologies and standards for the data value chain over the next decade. The following KPIs are proposed to frame and assess the impact of those SRIA activities.

Direct KPIs			Societal	Competitiveness	Innovation	Operational
Business	KPI.D.1	**At least 50 large-scale experiments are conducted in i-Spaces involving closed data.** Multiple SMEs should be encouraged to perform experiments by using i-Spaces. This will foster their growth from small companies into larger ones and/or their expansion from national markets to the EU (or even global) market. The iSpace and the residing experiments will provide a unique opportunity for exploitation.				

[42] "*The European Data Market*", Gabriella Catteneo, IDC, presentation given at the NESSI summit in Brussels on 27 May 2014, available online at: http://www.nessi-europe.eu/?Page=nessi_summit_2014

	KPI.D.2	**30% year-on-year increase in Big Data Value use cases supported in i-Spaces.** The number of use cases within the large-scale experiments will be an indicator of acceptance and will also prove the innovative capacity of the BDV partnership. An ever-expanding increase will guarantee a continuous value creation out of Big Data and will speed up the innovation process, thus also addressing the time to market. It will support market development in existing industries and potentially in establishing entirely new business models.				
Skills	KPI.D.3	**At least 50 training programs are established with participation of at least 100 participants per training session arising from cPPP.** Continuous development of skills and competences on the basis of the Big Data Value cPPP will be supported by training and education activities. An appropriate environment (e.g. e-learning platform, contribution to University curricula) should be created to attract potential participants. This broadens the number of skilled people and serves as a unique opportunity to create new jobs and start-ups as a result of the cPPP activities.				
	KPI.D.4	**At least 10 European training programs involving 3 different disciplines with the participation of at least 100 participants**. These inter-disciplinary programs will highly contribute to knowledge and skills of the huge complexity of Big Data. To broaden the number of people Massive Open Online Courses (MOOC) would be proposed building on the diversity of skills and European multiculturalism				
Applications	KPI.D.5	**At least 10 major sectors and major domains are supported by Big Data technologies and applications developed in the cPPP**. The usage of BDV technologies and applications developed in the cPPP in different sectors will lead to increased value generation and finally to job growth in all the addressed sectors. The broad take-up of those technologies and applications across a number of sectors is also an indicator for efficient sharing of best practices and expertise leading to a build-up of a broad skills basis. Furthermore, cross-sector activities should prove the domain independent and cross-domain deployment leading to standards.				
Data	KPI.D.6	**Total amount of data that has been made available to i-Spaces - including in particular closed data – is in the Zettabyte range.** Experiments conducted in i-Spaces benefit from their scale, amount of different but integrated data sources, and especially on the value of data. This is a key to perform deep analytics to improve data understanding, deep learning and meaningfulness of data. Combined with advanced visualization techniques it will guide to a unique user experience. In order to assure privacy protection and to respect user's privacy relevant measures and anonymisation mechanisms need to be applied.				
	KPI.D.7	**Availability of metrics for measuring the quality, diversity and value of data assets.** It is not only the amount of data made available to perform data analysis; of utmost importance are the quality, diversity and value of these data. Ultimate goal to create value out of Big Data is to derive analytical findings on a minimal, yet most significant data set, thus allowing faster data processing and management of data for deep analytics. During the cPPP relevant metrics will be derived.				
Technical	KPI.D.8	**The speed of data throughput is increased by 100 times compared to 2014.** One of the main problems regarding today's data storage and processing techniques is the time required for accessing large set of datasets in order to analyse them. Techniques to be implemented in the scope of the Data Management priority will make data access for analysis much more efficient.				
	KPI.D.9	**The energy required to process the same amount of data is reduced by 10% per year.** One of the main problems today is the energy consumed processing data due to the huge amount of data and lack of algorithms coupled with new hardware designed devices that will make energy required to process data highly reduced. Beyond hardware optimization, new tools and algorithms will require less resources and time to provide the same quality of analytics				
	KPI.D.10	**Enabling advanced privacy and security respecting mechanisms (including anonymisation) for data access, process and analysis** 10% year-to-year increase of closed data sets available in i-Spaces				

6 Annexes

6.1 Acronyms and Terminology

Acronym/Term	Name/Description
General	
API	Application Programming Interface
BDV	Big Data Value
BPM	Business Process Management
CASD	Secure Remote Data Access Centre
cPPP	(Contractual) Public Private Partnership
CSA	Coordination and Support Action
CEP	Complex Event Processing
DSMS	Data Stream Management Systems
EIP	European Innovation Partnership
EU	European Union
ETP	European Technology Platform
FI	Future Internet
FIRE	Future Internet Research & Experimentation
GDP	Gross Domestic Product
ICT	Information Communication Technologies
IoT	Internet of Things
IPR	Intellectual Property Rights
i-Space	(European) Innovation Space
KPI	Key Performance Indicators
MOU	Memorandum of Understanding
MPP	Massively Parallel architectures
NoSQL	Not only SQL (referred to databases)
SME	Small and Medium sized Enterprise
SRIA	Strategic Research & Innovation Agenda
SWOT	Strengths, Weaknesses, Opportunities and Threats
Data Orientated	
Open Data	Data available to everyone to use and republish
Private Data	Data which is generated by organisations, typically companies and in particular users, which and has not been made “open” and often is kept internally or has restricted conditions around it (eg NDAs)
Public Data	Freely reusable datasets from local, regional and national public bodies. Public Data is generally also Open Data
Closed Data	Data that has restrictions on its access or reuse (i.e. charges, technology, memberships, etc.). Typically Closed Data include Private Data
Free Data	Data that can be accesses or reused without a charge
Non-Free Data	Data which has a charge associated with use or reuse

6.2 Contributors

The following individuals and organisations are thanked for their direct involvement it creating this specific proposal or other documents to which it heavily relates

SRIA Core Team		
Nuria de Lama	Co-Editor	ATOS
Julie Marguerite	Co-Editor	Thales
Klaus-Dieter Platte	Co-Editor	SAP
Josef Urban	Co-Editor	Nokia
Sonja Zillner	Co-Editor	Siemens
Primary Editing Team		
Antonio Alfaro	Contributor	Answare
Edward Curry	Contributor	Insight @ NUI Galway
Ernestina Menasalves	Contributor	UPM
Andreas Metzger	Contributor	Paluno, Univ. Duisburg-Essen
Robert Seidl	Contributor	Nokia
Colin Upstill	Contributor	IT Innovation
Walter Waterfeld	Contributor	Software AG
Stefan Wrobel	Contributor	Fraunhofer IAIS
Contribution		
Paolo Bellavista	Contributor	CINI
Stuart Campbell	Contributor	TIE Kinetix
Thomas Delavallade, Yves Mabiala	Contributor	Thales
Nuria Gomez, Paolo Gonzales, Jesus Angel	Contributor	INDRA
Thierry Nagellen	Contributor	Orange
Dalit Naor, Elisa Molino	Contributor	IBM Research
Stefano de Panfilis, Stefano Scamuzzo	Contributor	Engineering
Nikos Sarris	Contributor	ATC
Bjørn Skjellaug, Arne Berre, Titi Roman	Contributor	SINTEF
Tonny Velin	Contributor	Answare
Alexandra Rosén, Francois Troussier	Contributor	NESSI Office
Other Participants involved in SRIA, Proposal or other related documents and discussions		
Volker Markl (TU Berlin), Burkhard Neidecker-Lutz (SAP), José María Cavanillas (ATOS), Roberto Martínez (UPM), Michael May (Siemens), Wolfgang Wahlster (DFKI) and Brigitte Cardinael (Orange).		
The 200 active participants of the NESSI Organised Big Data Value Workshops, organised between February and March 2014.		
All participants of the Public Consultation on the SRIA that was available at www.bigdatavalue.eu from 9 April to 15 May 2014.		
Numerous NESSI Partners and members as well as other contributors and their personnel		
The partners of the EU project BIG.		
Other European Technology Platforms, which through discussions contributed to defining the content of this SRIA, for example the ETPs NEM, ETP4HPC and Net!works.		