

Chapter 11

Big Data in the Public Sector

Ricard Munné

11.1 Introduction

The public sector is becoming increasingly aware of the potential value to be gained from big data. Governments generate and collect vast quantities of data through their everyday activities, such as managing pensions and allowance payments, tax collection, national health systems, recording traffic data, and issuing official documents. This chapter takes into account current socio-economic and technological trends, including boosting productivity in an environment with significant budgetary constraints, the increasing demand for medical and social services, and standardization and interoperability as important requirements for public sector technologies and applications. Some examples of potential benefits are as follows:

- **Open government and data sharing:** The free flow of information from organizations to citizens promotes greater trust and transparency between citizens and government, in line with open data initiatives.
- **Citizen sentiment analysis:** Information from both traditional and new social media (websites, blogs, twitter feeds, etc.) can help policy makers to prioritize services and be aware of citizens' interests and opinions.
- **Citizen segmentation and personalization while preserving privacy:** Tailoring government services to individuals can increase effectiveness, efficiency, and citizen satisfaction.
- **Economic analysis:** Correlation of multiple sources of data will help government economists with more accurate financial forecasts.
- **Tax agencies:** Automated algorithms to analyse large datasets and integration of structured and unstructured data from social media and other sources will help them validate information or flag potential frauds.

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

- **Smart city and Internet of things (IoT) applications:** The public sector is increasingly characterized by applications that rely on sensor measurements of physical phenomena such as traffic volumes, environmental pollution, usage levels of waste containers, location of municipal vehicles, or detection of abnormal behaviour. The integrated analysis of these high volume and high velocity IoT data sources has the potential to significantly improve urban management and positively impact the safety and quality of life of its citizens.
- **Cyber security:** Collect, organize, and analyse vast amounts of data from government computer networks with sensitive data or critical services, to give cyber defenders greater ability to detect and counter malicious attacks.

11.1.1 Big Data for the Public Sector

As of today, there are no broad implementations of big data in the public sector. Compared to other sectors, the public sector has not been traditionally using data mining technologies intensively. However, there is a growing interest in the public sector on the potentials of big data for improvement in the current financial environment.

Some examples of the global growing awareness are the Joint Industry/Government Task Force to drive development of big data in Ireland, announced by the Irish Minister for Jobs, Enterprise and Innovation in June 2013 (Government of Ireland 2013), or the announcement made by the Obama administration (The White House 2012), on the “Big Data Research and Development Initiative” where six Federal departments and agencies announce more than \$200 million in new commitments to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data.

11.1.2 Market Impact of Big Data

There is no direct market impact nor competition, as the public sector is not a productive sector, although its expenditure represented 49.3 % of GDP in 2012 of the EU28. The major part of the sector’s income is collected through taxes and social contributions. Hence, the impact of big data technologies is in terms of efficiency: the more efficient the public sector is, the better off are citizens, as less resources (taxes) are needed to provide the same level of service. Therefore, the more effective the public sector is, the more positive the impact on the economy, by transition for the rest of productive sectors, and more positive impact on society. Additionally, the quality of services provided, for example, education, health, social services, active policies, and security, can also be improved by making use of big data technologies.

11.2 Analysis of Industrial Needs in the Public Sector

The benefits of big data in the public sector can be grouped into three major areas, based on a classification of the types of benefits:

Big Data Analytics This area covers applications that can only be performed through automated algorithms for advanced analytics to analyse large datasets for problem solving that can reveal data-driven insights. Such abilities can be used to detect and recognize patterns or to produce forecasts.

Applications in this area include fraud detection (McKinsey Global Institute 2011); supervision of private sector regulated activities; sentiment analysis of Internet content for the prioritization of public services (Oracle 2012); threat detection from external and internal data sources for the prevention of crime, intelligence, and security (Oracle 2012); and prediction for planning purposes of public services (Yiu 2012).

Improvements in Effectiveness Covers the application of big data to provide greater internal transparency. Citizens and businesses can take better decisions and be more effective, and even create new products and services thanks to the information provided. Some examples of applications in this area include data availability across organizational silos (McKinsey Global Institute 2011); sharing information through public sector organizations [e.g. avoiding problems from the lack of a single identity database (e.g. in the UK) (Yiu 2012)]; open government and open data facilitating the free flow of information from public organizations to citizens and businesses, reusing data to provide new and innovative services to citizens (McKinsey Global Institute 2011; Ojo et al. 2015).

Improvements in Efficiency This area covers the applications that provide better services and continuous improvement based on the personalization of services and learnings from the performance of such services. Some examples of applications in this area are personalization of public services to adapt to citizen needs and improving public services through internal analytics based on the analysis of performance indicators.

11.3 Potential Big Data Applications for the Public Sector

Four potential applications for the public sector were described and developed in Zillner et al. (2013, 2014) for demonstrating the use of big data technologies in the public sector (Table 11.1).

Table 11.1 Summary of application scenarios for the public sector

Name	Monitoring and supervision of regulated activities for online gambling operators
Summary	Large volumes of data available make it difficult to effectively regulate and supervise activities
Synopsis	To monitor the online gambling operators for the control of regulated activities and detection of fraud. The user of this application is the public body in charge of the supervisory activity. This procedure is a regulatory obligation from the public administration; the online gambling operators must provide the information to the regulatory public through a specific communication channel. Real-time data is received from gambling operators every 5 min.
Business objectives	Ensure compliance with regulations, fraud prevention and detection, and criminal investigation.
Name	Operative efficiency in labour agency
Summary	Extract value from available large volumes of unused data
Synopsis	Enable a new range of personalized services, improve customer services and cut operation costs in German Federal Labour agency. All unemployed workers were receiving the same standard services despite having different profiles. Historical data on their customers was analysed, including profiles, interventions, and the time it took to find a job. Based on this analysis, customer segmentation was developed.
Business objectives	Reduce the cost and improve the quality of the service: now they are able to find a new job in a shorter period of time.
Name	Public Safety in Smart Cities
Summary	Large volumes of data available from sensors, social media, and emergency calls can be combined to provide effective public safety.
Synopsis	Smart cities equipped with sensors and communication infrastructures help the public sector keep cities and their citizens safe. Having accurate and up-to-date information allows better and faster responses during emergencies and results in less damage and casualties. Typical sources for obtaining such information can come from emergency response calls, surveillance cameras, and mobile forces (such as a police patrol car) that arrived at a site. In recent years social media have shown interesting potential for gathering information that aids in obtaining an accurate situational awareness picture (van Kasteren et al. 2014). All gathered information is collected in a command and control centre where an operator can decide how to steer available mobile forces.
Business objectives	Quick response to emergencies, prevention of damages, and less casualties.
Name	Predictive policing using open data
Summary	Reuse of public open data to provide predictive policing
Synopsis	Governments around the world have started open data initiatives to make public sector data available to the public for the sake of transparency and to allow third parties to offer services based on the data. One such service can be described as predictive policing where historical crime data is used to automatically discover trends and patterns. The identified patterns help in gaining insights into crime-related problems a city is facing and allow a more effective and efficient deployment of police forces (Wang et al. 2013; PredPol 2013).
Business objectives	Significant decrease in crime, efficient use of mobile forces.

11.4 Drivers and Constraints for Big Data in the Public Sector

The key drivers and constraints of big data technologies in the public sector are:

11.4.1 Drivers

The following **drivers** were identified for big data in the public sector:

- **Governments can act as catalysts** in the development of a data ecosystem through the opening of their own datasets, and actively managing their dissemination and use (World Economic Forum 2012).
- **Open data initiatives** are a starting point for boosting a data market that can take advantage from open information (content) and the big data technologies. Therefore active policies in the area of open data can benefit the private sector, and in return facilitate the growth of this industry in Europe. In the end this will benefit public budgets with an increase of tax incomes from a growing European data industry.

11.4.2 Constraints

The **constraints** for big data in the public sector can be summarized as follows:

- **Lack of political willingness** to make the public sector take advantage of these technologies. It will require a change in mind-set of senior officials in the public sector.
- **Lack of skilled business-oriented people** aware of where and how big data can help to solve public sector challenges, and who may help to prepare the regulatory framework for the successful development of big data solutions.
- **New General Data Protection Regulation and the PSI directives display some uncertainties** about the impact on the implementation of big data and open data initiatives in the public sector. Specifically, open data is set to be a catalyst from the public sector to the private sector to establish a powerful data industry.
- **Gaining adoption momentum.** Today, there is more marketing around big data in the public sector than real experiences from which to learn which applications are more profitable, and how it should be deployed. This requires the development of a standard set of big data solutions for the sector.
- **Numerous bodies in public administration** (especially in those which are widely decentralized), so much energy is lost and will remain so until a common strategy is realized for the reuse of cross technology platforms.

11.5 Available Public Sector Data Resources

In Directive 2003/98/EC (The European Parliament and the Council of The European Union 2003), on the re-use of public sector information, public sector information (PSI) is defined as follows: “It covers any representation of acts, facts or information – and any compilation of such acts, facts or information – whatever its medium (written on paper, or stored in electronic form or as a sound, visual or audio-visual recording), held by public bodies. A document held by a public sector body is a document where the public sector body has the right to authorise re-use.”

According to Correia (2004), concerning the availability of the information produced by those public bodies, and in the absence of specific guidelines, the producing body is free to decide how to make it available: directly to the end users, establishing a public/private partnership, or outsourcing the commercial exploitation of that information to private operators. The Directive 2003/98/EC clarifies that activities falling outside the public task: “will typically include supply of documents that are produced and charged for exclusively on a commercial basis and in competition with others in the market”.

On the nature of the PSI available, there are several approaches. The Green paper on PSI (European Commission 1998) proposes some classifications such as:

- PSI distinction between administrative and non-administrative
- PSI distinction regarding its relevance for the public

Additionally it can be distinguished according to its potential market value, and in some cases according to the content of personal data:

- PSI distinction according to its anonymity

The most important amount of data produced by public sector is textual or numerical, versus other sectors like healthcare that produces a large amount of electronic images. As a result of e-government initiatives of the past 15 years, a great part of this data is created in digital form, 90 % according to McKinsey (McKinsey Global Institute 2011).

According to the survey performed for the formulation of the European Big Data Value Partnership to public sector representatives (Zillner et al. 2014), the key data asset is the whole system of public sector, registries, databases, and information systems, of which the most significant are:

- Citizens, business, and properties (e.g. base registries, transactions)
- Fiscal data
- Security data
- Document management especially as the electronic transactions are growing
- Public procurement and expenses
- Public bodies and employees
- Geographical data mainly related to cadastral
- Content related to culture, education, and tourism
- Legislative documents

- Statistical data (socio-economic data that could be used by private sector)
- Geospatial data

11.6 Public Sector Requirements

The requirements of the public sector were broken down into non-technical and technical requirements.

11.6.1 *Non-technical Requirements*

Privacy and Security Issues The aggregation of data across administrative boundaries on a non-request-based manner is a real challenge, since this information may reveal highly sensitive personal and security information when combined with various other data sources, not only compromising individual privacy but also civil security. Access rights to the required datasets for an operation must be justified and obtained. When a new operation is performed over existing data, a notification or a license must be obtained from the Data Privacy Agency. Anonymity must be preserved in these cases, so data dissociation is required. Individual privacy and public security concerns must be addressed before governments can be convinced to share data more openly, not only publicly but sharing in a restricted manner with other governments or international entities. Another dimension is the regulation for the use of cloud computing in a way that public sector can trust cloud providers. Furthermore, the lack of European big data cloud computing providers within the European market is also a barrier for adoption.

Big Data Skills There's a lack of skilled data scientists and technologists who can capture and process these new data sources. When big data technologies become increasingly adopted in business, skilled big data professionals will become harder to find. Public body agencies could go a fair distance with the skills they already have, but then they will need to make sure those skills advance ([1105 Government Information Group n.d.](#)). Besides the technical oriented people, there is a lack of knowledge in business-oriented people who are aware of what big data can do to help them solve public sector challenges.

Other Requirements Other non-technical requirements include:

- Willingness to supply and to adopt big data technologies, and also to know how to use it.
- Need for common national or European approaches (policies)—like the European policies for interoperability and open data. Lack of leadership in this field.

- A general mismatch between business intelligence in general and big data in particular in the public sector.

11.6.2 Technical Requirements

Below is a detailed description of each of the eight technical requirements that were distilled from the four big data applications selected for the Public Sector Forum.

Pattern Discovery Identifying patterns and similarities to detect specific criminal or illegal behaviours in the application scenario of monitoring and supervision of online gambling operators (and also for similar monitoring scenarios within the public sector). This requirement is also applicable in the scenario to improve operative efficiency in the labour agency, and in the predictive policing scenario.

Data Sharing/Data Integration Required to overcome lack of standardization of data schemas and fragmentation of data ownership. Integration of multiple and diverse data sources into a big data platform.

Real-Time Insights Enable analysis of fresh/real-time data for instant decision-making, for obtaining real-time insights from the data.

Data Security and Privacy Legal procedures and technical means that allow the secure and privacy preserving sharing of data. The solutions to this requirement may unlock the widespread use of big data in public sector. Advances in the protection and privacy of data are key for the public sector, as it may allow the analysis of huge amounts of data owned by the public sector without disclosing sensitive information. These privacy and security issues are preventing the use of cloud infrastructures (processing, storage) by many public agencies that deal with sensitive data.

Real-Time Data Transmission Because the capability of placing sensors is increasing in smart city application scenarios, there is a high demand for real-time data transmission. It will be required to provide distributed processing and cleaning capabilities for image sensors so as not to collapse the communication channels and provide just the required information to the real-time analysis, which will be feeding situational awareness systems for decision-makers.

Natural Language Analytics Extract information from unstructured online sources (e.g. social media) to enable sentiment mining. Recognition of data from natural language inputs like text, audio, and video.

Predictive Analytics As described in the application scenario for predictive policing, where the goal is to distribute security forces and resources according to the prediction of incidents, provide predictions based on the learning from previous situations to forecast optimal resource allocation for public services.

Modelling and Simulation Domain-specific tools for modelling and simulation of events according to data from past events to anticipate the results from decisions taken to influence the current conditions in real-time, for example, in scenarios of public safety.

11.7 Technology Roadmap for Big Data in the Public Sector

For each requirement in the sector, this section presents applicable technologies and the research questions to be developed (Fig. 11.1). All references presented here are from Curry et al. (2014).

11.7.1 *Pattern Discovery*

- Data Analysis Technology: Semantic pattern technologies including stream pattern matching.
 - Research Question: Scalable complex pattern matching. Reaching trillions over datasets will take 5 years.
- Data Curation Technology: Validation of pattern analytics outputs with humans via curation.
 - Research Question: Curation at scale depends on the interplay between automated curation platforms and collaborative approaches leveraging large pools of data curators. Commercial application results could be reached in 6–10 years.
- Data Storage Technology: Analytical Databases, Hadoop, Spark, Mahout.
 - Research Question: Standard Array Query Language. Currently there is a lack of standardized query languages but efforts such as ArrayQL are on their way. Currently there is no widespread adoption and existing DBs (SciDB, Rasdaman) are used in the scientific community. This may change in 3–5 years from now.

Technical Requirement	Technology	Research Question
Pattern discovery	Semantic pattern technologies	Scalable complex pattern matching for trillions
	Validation of outputs with humans	Machine learning approaches to the discovery of data curation patterns
	Analytical Databases	Standard Array Query Language
Real Time Insights	Linked Data and machine learning supporting analysis	High performance coping with the 3 Vs
	In-memory databases	Ad-hoc queries with minimal latencies
Data Sharing / Data Integration	Facilitate the integration as well as analysis	Fragment selection for graph-like data through Quantum computing
	Linked Data for sharing and ontologies for integrating data	Scalability, high speed and data variety for trillion records
	Metadata and data provenance frameworks	Integration of provenance-awareness into existing tools
Real Time Data Transmission	Data acquisition: Storm	Distributed processing and cleaning
	Write optimized storage solution	Improving random read/write performance of DB technologies
Predictive Analytics	Analytical DBs	Efficient support of predictive analytics in DBs
Natural Language Analytics	Entity linking and co-reference resolution	Increase scalability and robustness
	Validation of NLA outputs with humans via curation	Software infrastructures integrating NLP pipelines into data curation
Modelling and simulation	Temporal databases	Management of time-series data for effective analysis
	Application of simulation in planning	Making models explicit and/or transparent
Data Security and Privacy	Encrypted storage and DBs	Privacy by design – Queries on encrypted storage

Fig. 11.1 Mapping requirements to research questions in the public sector

11.7.2 Data Sharing/Data Integration

- Data Acquisition Technology: To facilitate the integration as well as analysis.
 - Research Question: Data fragment selection, sampling and scalability. Solutions will be brought about by quantum computers (predicted to be available in 5–10 years, but 15–20 years seems more realistic.)

- **Data Analysis Technology:** Linked data provides the best technology set for sharing data on the Web. Linked data and ontologies provide mechanisms for integrating data (map to same ontology; map between ontologies/schemas/instances).
 - **Research Question:** Scalability, dealing with high speed of data and high variety. Dealing with trillions of nodes will take 3–5 years.
 - **Research Question:** Making semantic systems easy to use by non-semantic (logic) experts. It will take 5 years at least to have a comprehensive tooling support.
- **Data Curation/Storage Technology:** Metadata and data provenance frameworks.
 - **Research Question:** What are standards for common data tracing formats? Provenance on certain storage types, e.g. graph databases, is still computationally expensive. The integration of provenance-awareness into existing tools can be achieved in the short term (2–3 years) once this reaches a critical market demand.

11.7.3 Real-Time Insights

- **Data Analysis Technology:** Linked data and machine learning technologies can support automated analysis, which is required for gaining real-time insights.
 - **Research Question:** High performance while coping with the 3 Vs (volume, variety and velocity). Real-time deep analytics is more than 5 years away.
- **Data Storage Technology:** Google Data Flow, Amazon Kinesis, Spark, Drill, Impala, in-memory databases.
 - **Research Question:** How can ad hoc and streaming queries on large datasets be executed with minimal latencies? This is an active research field and may reach further maturity in a few years' time.

11.7.4 Data Security and Privacy

- **Data Storage Technology:** Encrypted storage and DBs; proxy re-encryption between domains; automatic privacy protection (e.g. differential privacy).
 - **Research Question:** Advances in “privacy by design” to link analytics needs with protective controls in processing and storage. A legal framework, e.g., the General Data Protection Regulation (GDPR), has to be harmonized among EU member states. Beyond legislation, data and social commons are required (Curry et al. 2014). This will require at least a further 3 years of research.

11.7.5 Real-Time Data Transmission

- Data Acquisition Technology: Kafka, Flume, Storm, etc., Curry et al. (2014).
 - Research Question: Distributed processing and cleaning. Current approaches should be able to let the user know the type of resources that they require to perform tasks specified by the user (e.g. process 10 GB/s). First approaches towards these ends are emerging and they should be available on the market within the next 5 years.
- Data Storage Technology: Current best practice: write optimized storage solution (e.g. HDFS), columnar stores.
 - Research Question: How to improve random read/write performance of database technologies. The Lambda Architecture described by Marz and Warren reflects the current best practice standard for persisting high velocity data. Effectively it addresses the shortcoming of insufficient random/read write performances of existing DB technologies. Performance increases will be continuous and incremental and simplify overall development of big data technology stacks. Technologies could reach a level of maturity that leads to simplified architectural blueprints in 3–4 years.

11.7.6 Natural Language Analytics

- Data Analysis Technology: Information extraction, named entity recognition, machine learning, linked data. Entity linking and co-reference resolution.
 - Research Question: Increasing scalability and robustness. Robust scalable solutions are at least 3–5 years away.
- Data Curation Technology: Validation of Natural Language Analytics (NLA) outputs with humans via curation.
 - Research Question: Curation at scale depends on the interplay between automated curation platforms and collaborative approaches leveraging large pools of data curators. Technically, this integration can be achieved in the short term (2–3 years).

11.7.7 Predictive Analytics

- Data Storage Technology: Analytical databases.
 - Research Question: How can databases efficiently support predictive analytics? From a storage point of view, analytical databases address the problem of better performance as the DB itself is able to execute analytical code.

Currently there is a lack of standardized query languages but efforts such as ArrayQL are on their way. This may change in 3–5 years from now.

11.7.8 Modelling and Simulation

- **Data Storage Technology:** Best practices; batch and in-stream processing (Lambda architecture), temporal databases.
 - **Research Question:** How can time-series data be managed in a general way for effective analysis? Spatiotemporal databases are an active research field and results may be beyond a 5-year time scale.
- **Data Usage Technology:** Standards in (semantic) modelling; application of simulation in planning (e.g. plant planning).
 - **Research Question:** Making models explicit and/or transparent. This is a research question with a long timeline (beyond 2020).

11.8 Conclusion and Recommendations for the Public Sector

The findings after analysing the requirements and the technologies currently available show that there are a number of open research questions to be addressed in order to develop the technologies such that competitive and effective solutions can be built. The main developments are required in the fields of scalability of data analysis, pattern discovery, and real-time applications. Also required are improvements in provenance for the sharing and integration of data from the public sector.

It is also extremely important to provide integrated security and privacy mechanisms in big data applications, as the public sector collects vast amounts of sensitive data. In many countries legislation limits the use of the data only for purposes for which it was originally obtained. In any case, respecting the privacy of citizens is a mandatory obligation in the European Union.

Other areas, especially interesting for the safety applications in public sector, are the analysis of natural language, which can be useful as a method to gather unstructured feedback from citizens, e.g. from social media and networks. The development of effective predictive analytics, as well as modelling and simulation tools for the analysis of historical data, are key challenges to be addressed by future research.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-Noncommercial 2.5 License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

The images or other third party material in this book are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt, or reproduce the material.

References

- 1105 Government Information Group. (n.d.). *The chase for Big Data skills*. Retrieved March 26, 2013, from GCN.com: <http://gcn.com/microsites/2012/snapshot-managing-big-data/04--chasing-big-data-skill-sets.aspx>
- Correia, Z. P. (2004). Toward a stakeholder model for the co-production of the public-sector information system. *Information Research*, 10(3), paper 228. Retrieved February 27, 2013, from InformationR.net: <http://InformationR.net/ir/10-3/paper228.html>
- Curry, E., Ngonga, A., Domingue, J., Freitas, A., Strohbach, M., Becker, T. et al. (2014). D2.2.2. Final version of the technical white paper. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4).
- European Commission. (1998). COM(1998)585. *Public sector information: A key resource for Europe. Green paper on public sector information in the information society*. European Commission.
- Government of Ireland. Department of Jobs, Enterprise and Innovation. (2013, June 24). *Joint Industry/Government Task Force to drive development of Big Data in Ireland – Minister Bruton*. Retrieved February 17, 2013, from Working for Jobs, Enterprise and Innovation: <http://www.djei.ie/press/2013/20130624.htm>
- McKinsey Global Institute. (2011, June). *Big Data: The next frontier for innovation, competition, and productivity*. McKinsey & Company.
- Ojo, A., Curry, E., & Sanaz-Ahmadi, F. (2015). A tale of open data innovations in five smart cities. 48th Annual Hawaii International Conference on System Sciences, IEEE (pp. 2326–2335)
- Oracle. (2012). *Big Data: A big deal for public sector organizations*. Oracle.
- PredPol. (n.d.). Retrieved September 08, 2013, from PredPol Web site: <http://www.predpol.com/>
- The European Parliament and the Council of The European Union. (2003, November 17). Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. Official Journal L 345, 31/12/2003 P. 0090 - 0096. Brussels: The European Parliament and the Council of the European Union.
- The White House. (2012, March 29). *Big Data is a Big Deal*. Retrieved January 18, 2013, from The White House: <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>
- van Kasteren, T., Ulrich, B., Srinivasan, V., & Niessen, M. (2014). Analyzing tweets to aid situational awareness. *36th European Conference on Information Retrieval*.
- Wang, T., Rudin, C., Wagner, D., & Sevieri, R. (2013). Detecting patterns of crime with series finder. *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*.
- World Economic Forum. (2012). *Big Data, Big Impact: New possibilities for international development*. Geneva: The World Economic Forum.
- Yiu, C. (2012). *The Big Data Opportunity. Making government faster, smarter and more personal*. London: Policy Exchange.
- Zillner, S., Rusitschka, S., Munné, R., Lippell, H., Lobillo, F., Hussain, K. et al. (2013); D2.3.1. *First draft of the sectorial requisites*. Public Deliverable of the EU-Project BIG (318062; ICT-2011.4.4).
- Zillner, S., Neururer, S., Munné, R., Prieto, E., Strohbach, M., van Kasteren, T. et al. (2014). D2.3.2. *Final version of the sectorial requisites*. Public Deliverable of the EU-Project BIG (318062; ICT-2011.4.4).