

# Chapter 14

## Big Data in the Media and Entertainment Sectors

Helen Lippell

### 14.1 Introduction

The media and entertainment industries have frequently been at the forefront of adopting new technologies. The key business problems that are driving media companies to look at big data capabilities are the need to reduce the costs of operating in an increasingly competitive landscape and, at the same time, the need to generate revenue from delivering content and data through diverse platforms and products.

It is no longer sufficient merely to publish a daily newspaper or broadcast a television programme. Contemporary operators must drive value from their assets at every stage of the data lifecycle. The most nimble media operators nowadays may not even create original content themselves. Two of the biggest international video streaming services, Netflix and Amazon, are largely aggregators of others' content, though also offering originally commissioned content to entice new and existing subscribers.

Media industry players are more connected with their customers and competitors than ever before. Thanks to the impact of disintermediation, content can be generated, shared, curated, and republished by literally anyone with an Internet-enabled device. Global revenues from such devices, including smartphones, tablets, desktop PCs, TVs, games consoles, e-readers, wearable gadgets, and even drones were expected to be around \$750 billion in 2014 (Deloitte 2014). This means that the ability of big data technology to ingest, store, and process many different data sources, and in real-time, is a valuable asset to the companies who are prepared to invest in it.

The Media Sector is in many respects an early adopter of big data technologies, but much more evolution has to happen for the full potential to be realized. Better integration between solutions along the data value chain will be essential in order to convince decision-makers to invest in innovation, especially in times of economic

---

H. Lippell (✉)  
Press Association, London, UK  
e-mail: [helenlippell@gmail.com](mailto:helenlippell@gmail.com)

uncertainty. Also, the solutions market is dominated by US, and, increasingly, Asian firms. Therefore, there is an economic imperative for Europe to both develop and use big data technologies more extensively. Media and entertainment content and platforms have a global reach that many companies in other sectors, even retail and manufacturing, would be envious of.

Case studies of successful big data projects in media have tended to come from the left-hand end of the data value chain (i.e. data acquisition and analysis). However, there is a need to identify both exemplars and gaps in the curation and usage of big data, as these are significant areas of competitive advantage for media organizations. Big data contributes to the bottom line by enabling organizations to pursue digital transformation. According to PWC (2014), this forges the trust of consumers, creates the confidence to innovate with speed and agility, and empowers innovation.

Unlike some other sectors, the vast majority of actionable data in the media sector is already in digital form (and analogue products such as newspapers have been created through digital technologies for some years now). However, this does not mean that organizations are deriving the fullest possible financial benefit or cost efficiencies from both their existing data and new sources of data. There is a growing body of evidence that there is much work to do at research and policy levels to support the burgeoning ecosystem of diverse businesses engaged in analysing, enhancing, and delivering content and data.

## 14.2 Analysis of Industrial Needs in the Media and Entertainment Sectors

The media sector has always generated data, whether from research, sales, customer databases, log files, and so on. Equally, the vast majority of publishers and broadcasters have always faced the need to compete right from the earliest days of newspapers in the eighteenth century. Even government or publicly funded media bodies have to continually prove their relevance to their audiences, in order to stay relevant in a world of extensive choice and to secure future funding. But the big data mind-set, technical solutions, and strategies offer the ability to manage and disseminate data at speeds and scales that have never been seen before.

There are three main areas where big data has the potential to disrupt the status quo and stimulate economic growth within the media and entertainment sectors:

1. **Products and Services:** Big data-driven media businesses have the ability to *publish* content in more sophisticated ways. Human expertise in, e.g., curation, editorial nous, and psychology can be complemented with quantitative insights derived from analysing large and heterogeneous datasets. But this is predicated on big data analysis tools being easy to use for data scientists and business users alike.
2. **Customers and Suppliers:** Ambitious media companies will use big data to find out more about their customers—their preferences, profile, attitudes—and they

will use that information to build more engaged relationships. With the tools of social media and data capture now widely available to more or less anyone, individuals are also suppliers of content back to media companies. Many organizations now back social media analysis into their orthodox journalism processes, so that consumers have a richer, more interactive relationship with news stories. Without big data applications, there will be a wasteful and random approach to finding the most interesting content.

3. **Infrastructure and Process:** While start-ups and SMEs can operate efficiently with open source and cloud infrastructure, for larger, older players, updating legacy IT infrastructure is a challenge. Legacy products and standards still need to be supported in the transition to big data ways of thinking and working. Process and organizational culture may also need to keep pace with the expectations of what big data offers. Failure to transform the culture and skillset of staff could impact companies who are profitable today but cannot adapt to data-driven business models.

### 14.3 Potential Big Data Applications for the Media and Entertainment Sectors

Six application scenarios for the media sector were described and further developed in Zillner et al. (2013, 2014a). All of these scenarios represent tangible business models for organizations; however, without support from big data technologies, companies will not be able to mature their existing pilots or small-scale projects into future revenue opportunities (Table 14.1).

### 14.4 Drivers and Constraints for Big Data in Media and Entertainment Sectors

Like all businesses, media companies aim to maximize revenue, minimize costs, and improve decision-making and business processes.

#### 14.4.1 Drivers

Specific to the media and entertainment sectors though are the following drivers:

- Aim to **understand customers** on a very detailed level, often by analysing many different types of interaction (e.g. product usage, customer service interactions, social media, etc.).

**Table 14.1** Summary of six application big data scenarios for the media sector

<i>Name</i>	<i>Data journalism</i>
Summary	Large volumes of data become available to a media organization.
Synopsis	Single or multiple datasets require analysis to derive insight, find interesting stories, and generate material. This can then be enhanced and ultimately monetized by selling to customers.
Business objectives	<ul style="list-style-type: none"> <li>– Improve quality of journalism and therefore enhance the brand</li> <li>– Analyse data more thoroughly for less cost</li> <li>– Enable data analysis to be performed by a wider range of users</li> </ul>
<i>Name</i>	<i>Dynamic semantic publishing</i>
Summary	Scalable processing of content for efficient targeting
Synopsis	Using semantic technologies to both produce and target content more efficiently
Business objectives	<ul style="list-style-type: none"> <li>– Manage content and scarce staff resources more efficiently</li> <li>– Add value to data to differentiate services from competitors</li> </ul>
<i>Name</i>	<i>Social media analysis</i>
Summary	Processing of large user-generated content datasets.
Synopsis	Batch and real-time analysis of millions of tweets, images, status updates to identify trends and content that can be packaged in value-added services.
Business objectives	<ul style="list-style-type: none"> <li>– Create value-added services for clients</li> <li>– Perform large-scale data processing in a cost-effective manner</li> </ul>
<i>Name</i>	<i>Cross-sell of related products</i>
Summary	Developing recommendation engines using multiple data sources.
Synopsis	Applications that exploit collaborative filtering, content-based filtering, and hybrids of both approaches.
Business objectives	<ul style="list-style-type: none"> <li>– Generate more revenue from customers</li> </ul>
<i>Name</i>	<i>Product development</i>
Summary	Using predictive analytics to commission new services
Synopsis	Data mining to support development of new and enhanced products for the marketplace
Business objectives	<ul style="list-style-type: none"> <li>– Offer innovative new products and services</li> <li>– Enable development in a more quantitative way than is currently possible</li> </ul>
<i>Name</i>	<i>Audience insight</i>
Summary	Using data from multiple sources to build up a comprehensive 360° view of a customer
Synopsis	Extension of scenario “Product Development”—mining of data external to the organization for information about customer habits and preferences
Business objectives	<ul style="list-style-type: none"> <li>– Reduce costs of customer retention and acquisition</li> <li>– Use insights to aid commissioning of new products and services</li> <li>– Maximize revenue from customers</li> </ul>

- **Operate in crowded sub-sectors** such as digital marketing or book publishing, where very few players have dominance, and consumer preferences and fashions can change very rapidly.

- **Diversify service offerings** wherever possible. Most significant European media companies operate in many areas, for example, newspaper publishers, websites, and commercial apps; or broadcasters may also sell broadband access.
- **Communicate to build influence within society**, e.g. politically. This is less tangible than just selling products but seen as equally important by media owners or governments.

### 14.4.2 Constraints

The **constraints** for big data in the media and entertainment sectors can be summarized as follows:

- **Increased consumer awareness** and concern about how personal data is being used. There is regulatory uncertainty for European businesses that handle personal data, which potentially puts them at a disadvantage compared to, say, US companies who operate within a much more relaxed legal landscape.
- **Insufficient access to finance** for media start-ups and SMEs. While it is relatively easy to start a new company producing apps, games, or social networks, it is much harder to scale up without committed investors.
- The **labour market** across Europe is not providing enough data professionals able to manipulate big data applications, e.g. for data journalism and product management.
- Fear of **piracy and consumer disregard for copyright** may disincentive creative people and companies from taking risks to launch new media and cultural products and services.
- **Large US players** dominate the content and data industry. Companies such as Apple, Amazon, and Google between them have huge dominance in many sub-sectors including music, advertising, publishing, and consumer media electronics.
- Differences in **penetration of high-speed broadband** provision across member countries, in cities, and in rural areas. This is a disincentive for companies looking to deliver content that requires high bandwidth, e.g. streaming movies, as it reduces the potential customer base.

## 14.5 Available Media and Entertainment Data Resources

Table 14.2 is intended to give a flavour of the data sources that most media companies routinely handle. One table lists some categories of data that are generated by the companies themselves, while the second shows third-party sources that are or can be processed by those in the media sector, depending on their particular line of business.

Each type of data source is matched to a key characteristic of big data. Customarily, the technology industry has talked of “the three Vs of big data”, that is, volume, variety, and velocity. Kobielus (2013) also discusses a fourth characteristic—veracity. This is important for the media sector because consumer products and services can quickly fail if the content lacks authoritativeness, or it is of poor quality, or it has uncertain provenance. According to IBM (2014), 27 % of respondents to a US survey were unsure even *how much* of their data was inaccurate—suggesting the scale of the problem is underestimated.

**Table 14.2** Media data resources mapped to “V” characteristics of big data

<i>Internally generated data</i>	<i>Key “V” characteristic</i>
Consumer profile details including customer service interactions.	Volume—Large amounts of data to be stored and potentially mined. Variety applies when considering the different ways customers may interact with a media service provider—and hence the opportunity for the business to “join up the dots” and better understand them.
Network logging (e.g. for web or entertainment companies operating their own networks).	Velocity—Network issues must be identified in real-time in order to resolve problems and retain consumer trust.
Organizations own data services to end users.	Characteristic(s) will depend on business objective of the data, e.g., a news agency will prioritize speed of delivery to customers, a broadcaster will be focused on streaming content in multiple formats to multiple types of device.
Consumer preferences inferred from sources including click stream data, product usage behaviour, purchase history, etc.	Volume—Large amounts of data can be gathered. Velocity will become pertinent where the service needs to be responsive to user action, e.g., online gaming networks which upsell extra features to players.
<i>Third-party data</i>	<i>Key “V” characteristic</i>
Commercial data feeds, e.g., sports data, press agency newswires.	Velocity—Being first to use data such as sports or news events builds competitive advantage.
Network information (where external networks are being used, e.g., messaging apps that piggyback on mobile networks).	Velocity—Network issues must be identified in real-time in order to ensure continuity of service.
Public sector open datasets.	Veracity—Open data may have quality, provenance, and completeness issues.
Free structured and/or linked data, e.g., Wikidata/DBpedia	Veracity—crowdsourced data may have quality, provenance, and completeness issues.
Social media data, e.g., updates, videos, images, links, and signals such as “likes”.	Volume, variety, velocity, and veracity—Media companies must prioritize processing based on expected use cases. As one example, data journalism requires a large volume of data to be prepared for analysis and interpretation. On the other hand, a media marketing business might be more concerned with the variety of social data across many channels.

## 14.6 Media and Entertainment Sector Requirements

The Media and Entertainment Sectorial Forum were able to identify and name several requirements, which need to be addressed by big data application in the domain. The requirements are distinguish between non-technical and technical requirements.

### 14.6.1 *Non-technical Requirements*

It is important to note that the widespread uptake of big data within the media industry is not solely dependent on successful implementation of specific technologies and solutions. In Zillner et al. (2014b), a survey was undertaken among European middle and senior managers from the media sector (and also the telecoms sector, where large players are increasingly moving into areas that were once considered purely the realm of broadcasters, publishers, etc.). Respondents were asked to rank several big data priorities based on how important they would be to their own organizations.

It is striking that all survey participants identified the need for a European framework for shared standards, a clear regulatory landscape, and a collaborative ecosystem—implying that businesses are suffering from a lack of confidence in their ability to see through the hype and really get to grips with big data in their enterprises. Another area ranked as very important by a notable proportion of respondents was making solutions usable and attractive for business users (i.e. not just data scientists).

### 14.6.2 *Technical Requirements*

Table 14.3 lists 37 requirements that were distilled from the work of the Media Sector Forum. Each requirement is matched to a business objective (although of course in practice some requirements could meet more than one objective). The five columns at the right-hand side of the table place each requirement in its appropriate place(s) along the big data value chain. Media, as a mostly customer-facing, revenue-generating economic sector, has many critical needs in data curation and usage.

**Table 14.3** Big data technical requirements of the media sector

Big data requirement	Business objective	Acquisition	Analysis	Curation	Storage	Usage
Curate heterogeneous data sources in a content and origin agnostic manner	Improve business processes	X		X		
Programmatically interrogate data for trends	Improve business processes		X			
Quickly start processing new data types as they become needed	Improve business processes	X	X	X	X	
Analyse unstructured data with regard to sentiment, topic, and other intangible aspects of text	Improve business processes		X	X		
Transform and augment open data from the public sector with regard to format, semantics, and quality	Improve business processes	X	X	X	X	
Scalable tools for search and discovery applications	Improve business processes				X	X
Visualize data for analytics and metrics (especially for business-technical users)	Improve business processes		X			X
Automatically create and apply metadata to datasets	Improve business processes	X	X	X		
Quickly and accurately process data in near real-time	Improve decision-making	X	X		X	
Apply models and ontologies to data to extract relationships	Improve decision-making	X	X		X	
Transform streams from sensors into actionable views	Improve decision-making	X			X	
Analytics tools which enable powerful querying and manipulation by non-programmers or statisticians	Improve decision-making		X		X	
Inference engines to analyse semantic graph data	Improve decision-making		X	X		

(continued)

**Table 14.3** (continued)

Big data requirement	Business objective	Acquisition	Analysis	Curation	Storage	Usage
Derive value from proprietary datasets	Increase revenue	X	X	X	X	X
Derive value from public open datasets	Increase revenue		X	X		X
Deliver tailored data and content to customers	Increase revenue			X		X
Human-centred editorializing of curated data streams	Increase revenue		X	X		X
Algorithms to crunch data to produce more interesting recommendations than “more of the same”	Increase revenue					X
Algorithm management tools for non-technical users	Increase revenue			X		X
Enrich multimedia content such as images and videos with semantic metadata	Increase revenue		X	X		X
Blend user-generated content with commercially produced media to create new digital products	Increase revenue	X		X		X
Generate insights from data to enable new business models (e.g. cross-selling based on viewing habits)	Increase revenue		X			X
Increase conversions from offline marketing activities (e.g. direct mail) by analysing online data	Increase revenue		X			X
Predictive analytics solutions that can identify trends, segments, and patterns without these explicitly being modelled	Increase revenue		X			
Return more relevant search results in consumer-facing	Increase revenue					X

(continued)

**Table 14.3** (continued)

Big data requirement	Business objective	Acquisition	Analysis	Curation	Storage	Usage
applications using semantic analysis						
Database solutions that can be set-up more quickly than with traditional applications	Reduce costs				X	
Capability to use crowdsourced data curation to complement internal subject matter expertise	Reduce costs			X		
Manage large-scale data in graph databases	Reduce costs				X	
Translate unstructured data (e.g. text or voice) to one or many languages	Reduce costs	X	X		X	
High-volume data scraping and crawling tools	Reduce costs	X			X	
Identify patterns in data to drive insights about consumer behaviour	Understand customers		X			
Take account of many factors (e.g. location, device, user profile, usage context) to better target content delivery	Understand customers	X				
Connect data from all customer interactions to form a 360° view	Understand customers	X	X		X	
Ingest data from new classes of device (e.g. wearables)	Understand customers	X				
Drill down into consumer behaviour in more granular detail	Understand customers		X			X
Foster a more engaged relationship with audiences and customers through unstructured social data analysis	Understand customers					X
Clear policy direction on use of personal data within the EU	Understand customers				X	X

## 14.7 Technology Roadmap for Big Data in the Media and Entertainment Sectors

Of all the sectors discussed in this book, media is arguably the one that changes most suddenly and most often. New paradigms can emerge extremely quickly and become commercially vital in a short space of time (e.g. Twitter was founded only in 2006 and now has a market capitalization of many billions of dollars). The year 2015 onwards will see many media players and consumers alike experimenting with drones (more strictly, “unmanned aerial vehicles”, or UAVs) to see if captured footage can be monetized either directly as content or indirectly to attract advertising.

Figure 14.2 and Table 14.4 consolidate the outcomes of the research completed in Zillner et al. (2013, 2014a), along with additional background research. Figure 14.1 maps out the methodology used to derive the sector roadmap, showing how iterative engagement with industry supported at every stage the definition of the needs and technologies around big data for the media sector.

Any roadmap must be cognisant of the risk that it will be out of date before it is even published. Nevertheless, the key headings shown in the figures in this section are strongly predicted to remain highly relevant to the sector for the following reasons.

### 14.7.1 *Semantic Data Enrichment*

Semantics is a long-established and now fast-developing field that is finally fulfilling its academic promise. Major media applications such as “intelligent personal assistants”, e.g. Siri and Cortana, are underpinned by “artificial intelligence” and semantic analysis technology. More development is needed to help commercial organizations in Europe exploit the potential of ontologies, graph databases, and curation platforms.

### 14.7.2 *Data Quality*

The key technological developments in this area include open data and data standards generally to aid interoperability. Also key are capabilities for processing unstructured (especially natural language) data streams. Finally, there is a need for back-end systems that can absorb different types of data with as little friction as possible, by minimizing the need to define data schemas upfront.

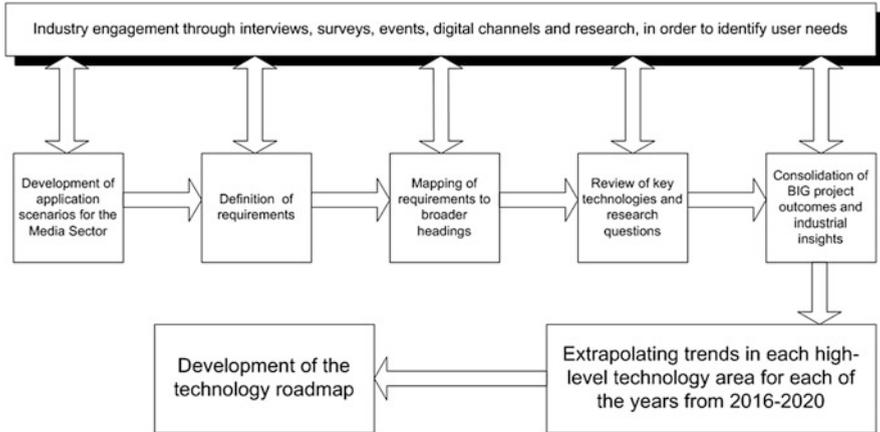


Fig. 14.1 Methodology for deriving the media sector roadmap

Technical Requirement	Technology	Research Question
Semantic Data Enrichment	Common and open ontologies	Relation extraction
	Graph databases	Scalability of non-relational databases
	Crowdsourced curation platform	Blended algorithmic & manual curation at scale
Data Quality	Open data platforms	Data standardisation & interoperability
	Unstructured data processing	Natural language processing at scale
	Heterogeneous data storage	Data-agnostic storage architectures
Data-driven Innovation	Machine learning (ML)	Integrating ML approaches into databases
	Networks, sensors, wearable tech	Commercialisation of auto-generated data
	Customer recommendation tools	Improving algorithmic recommendations
Data Analysis	Descriptive analytics	Data mining for subjective factors e.g. sentiment
	Data visualisation solutions	Business-user friendly applications
	Customer relationship platforms	Understand contexts to enhance data delivery

Fig. 14.2 Mapping requirements to research questions in the media sector

### 14.7.3 Data-Driven Innovation

Three key technologies underpinning the drive for high-quality innovation are machine learning at enterprise scale; the Internet of Things (IoT), which will exponentially increase the volume and diversity of data streams available to anyone

**Table 14.4** Big data technology roadmap for the media sector

Technical requirement	Year 1	Year 2	Year 3	Year 4	Year 5
Semantic data enrichment	Large media firms with resources create and publish open ontologies	Common ontologies for specific use cases in media and entertainment industries	Ontology management and manipulation tools available for a wide range of commercial uses	Relation extraction technology available at scale and affordability	Semantic inference to support predictive analysis of data, e.g., user behaviour, tracking news stories
Data quality	Limited open data available to companies looking to generate new business models	Open data published in machine-readable formats by more public and private sector bodies	Natural language processing tools scalable to large volumes of data (including speech)	Standardization of data acquisition protocols	Data-agnostic architectures enable diverse data streams to be analysed simultaneously and in real-time
Data-driven innovation	Curation platforms to enhance value-add of data products	Scalable recommendation tools for non-technical users	Machine learning frameworks embedded into decision-making tools	Real-time aggregation of streams generated by networks, sensors, body-worn devices	Product development platforms for rapid iteration of data-driven services
Data analysis	More detailed segmentation of customers based on subjective factors	Intuitive data visualization tools for interactive applications	Convergence of business intelligence and product analytics applications	Actionable predictive analytics of events or trends across large, dispersed data streams	Combinable analytics approaches enable deep insights into patterns based on context

involved in media or data-driven storytelling; and finally, tools to better interpret customer interactions with products and services.

#### ***14.7.4 Data Analysis***

Media and entertainment companies need to analyse data not only at the customer and product levels, but also at network and infrastructure levels (e.g. streaming video suppliers, Internet businesses, television broadcasters, and so on). Key technologies in the coming years will be descriptive analytics, more sophisticated customer relationship management solutions, and lastly data visualization solutions that are accessible to a wide range of users in the enterprise. It is only by “humanizing” these tools that big data will be able to deliver the benefits that data-driven businesses increasingly demand (Table 14.4).

### **14.8 Conclusion and Recommendations for the Media and Entertainment Sectors**

Europe has much to offer in culture and content to the global market. European publishers and TV companies are globally renowned, but no EU-based competitor has emerged to the multinational giants of Google, Amazon, Apple, or Facebook. Differences between the European and US economies, such as ease of access to venture capital, would seem to preclude this happening. Therefore, the best way forward for Europe is to build on its strengths of creativity and free movement of people and services, in order to bring together communities of industrial players, researchers, and government to tackle the following priorities:

- Making sense of data streams, whether text, image, video, sensors, and so on. Sophisticated products and services can be developed by extracting value from heterogeneous sources.
- Exploiting big data step changes in the ability to ingest and process raw data, so as to minimize risks in bringing new data-driven offerings to market.
- Curating quality information out of vast data streams, using algorithmic scalable approaches and blending them with human knowledge through curation platforms.
- Accelerating business adoption of big data. Consumer awareness is growing and technical improvements continue to reduce the cost of storage and analytics tools among other things. Therefore, it is more important than ever that businesses have confidence that they understand what they want from big data and that the non-technical aspects such as human resources and regulation are in place.

**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

- Deloitte. (2014). *Technology, media and telecommunications predictions 2014*. Retrieved from <http://www.deloitte.co.uk/tmtpredictions/assets/downloads/Deloitte-TMT-Predictions-2014.pdf>
- IBM. (2014). *Infographic – The four V's of big data*. Retrieved from <http://www.ibmbigdatahub.com/enlarge-infographic/1642>
- Kobielus, J. (2013). *Measuring the business value of big data*. Retrieved from <http://www.ibmbigdatahub.com/blog/measuring-business-value-big-data>
- PWC. (2014). *Global entertainment and media outlook 2014-2018 – key industry themes*. Retrieved from <http://www.pwc.com/gx/en/global-entertainment-media-outlook/insights-and-analysis.jhtml>
- Zillner, S., Neurerer, S., Munné, R., Lippell, H., Vilela, L., Prieto, E. et al. (2013). *D2.4.1 first draft of sectors roadmap*. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4).
- Zillner, S., Neurerer, S., Munné, R., Lippell, H., Vilela, L., Prieto, E. et al. (2014a). *D2.3.2. Final version of the sectorial requisites*. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4).
- Zillner, S., Neurerer, S., Munné, R., Lippell, H., Vilela, L., Prieto, E. et al. (2014b). *D2.4.2 Final version of sectors roadmap*. Public deliverable of the EU-Project BIG (318062; ICT-2011.4.4).