The Use of Big Data in the Public Policy Process: Paving the Way for Evidence-Based Governance

Julia Studinka[a] and Ali Asker Guenduez[a]*

[a] *Smart Government Lab, Institute for Public Management and Governance, University of St.Gallen, St.Gallen, Switzerland*

**corresponding author*

The Use of Big Data in the Public Policy Process: Paving the Way for Evidence-Based Governance

Abstract. Big data holds vast potential for improving decision-making processes, policymaking, and services. This thesis examines the question how big data can be used in different phases of the public policy process (planning, design, delivery, and evaluation of public policies). Based on a systematic literature review, this thesis evaluates how the use of big data in different phases of the policy process has been described in the academic literature. The findings first provide a taxonomy of the different types of big data according to their characteristics (structure, source, access, and size) and an overview over various big data techniques. The core findings of the analysis show how big data can be used in the different phases of the public policy process, including detailed examples. In the planning phase, the use of big data has been described in the areas of agenda-setting, problem definition, policy discussion, and citizen participation, with a focus on social media data. In the design phase, big data can contribute to policy formulation and information-based policy instruments – the techniques employed in this phase often have a predictive element. In the delivery phase, the focus lies on the real-time production of data and immediate feedback regarding the effectiveness of policies in order to improve future implementation processes. Furthermore, big data can be used for continuous evaluation of policies as part of all policy process phases. This thesis show the different ways in which big data influences the policy process and paw the way to evidence-based governance.

Keywords. Big data, Public Policy Process, Evidence-Based Governance, public administration

1 Introduction

The amount of electronic data being generated in the world is increasing as people's lives become more and more digital. Devices that are permanently connected to the Internet (e.g. smartphones, cameras with image recognition and analytics software, sensors, IoT systems etc.) produce vast amounts of different types of electronic data. This development is accompanied by new possibilities of information collection, storage, and processing, which increase the data's availability and usability, as well as new methods of data analysis. These phenomena create new opportunities for accessing and using data in ways that would not have been impossible with traditional methods (Maciejewski, 2017, p. 121). In the digital age, almost any action leaves a digital trail – governments now have access to extraordinary quantities of data about citizens (and vice versa), which offer the potential for vast improvements of public policies and services (Clarke & Margetts, 2014, p. 393).

Governments can use new forms of big data analysis to understand citizens' behavior and to improve public programs (Mergel, Rethemeyer & Isett, 2016, p. 928). The current literature on big data suggests many ways in which big data can be used to improve public sector outcomes. These include improving the efficiency, effectiveness, and transparency of governments, enabling the provision of better services based on enhanced insight into citizens' needs and demands, and more informed policymaking (Klievink et al., 2017, p. 268). Scholars have pointed to the vast potential of the use of big data in different policy areas, such as health care, economy, environment or transportation (Maciejewski, 2017; Jarmin & O'Hara, 2016; Schintler & Kulkarni, 2014). However, although big data solutions are promoted as a way to address public issues, many questions regarding how, where and when they are likely to be successful as well as the limits and potential of big data are still unanswered – according to Desouza and Jacob (2017, p. 1045) there is some tension between the promise and reality of big data. Also, the actual use of big data in the public sector is still very limited. Most big data projects are still being planned for future implementation or are in an early stage of development (Kim et al., 2014, p. 85). Nevertheless, the big data movement has moved past the question *if* towards the *how*, e.g. how can big data be incorporated into policymaking, how can it be regulated and how can it be utilized – "there is no turning back from the expectation that more data can lead to more information and eventually a more efficient and effective government" (Giest, 2017, p. 379).

The use of big data in the policy process builds on the research stream of evidence-based policymaking. The use of big data to enable evidence-based policymaking is increasingly studied – particularly in relation with the policy process or "policy cycle" (De Marchi et al., 2016, p. 20; Giest, 2017, p. 370; Höchtl, Parycek & Schöllhammer, 2016, pp. 156-157). According to Sanderson (2002, p. 5), evidence-based policy fits well with the theoretical model of the policy process. However, where and how big data can be used in the different phases of the policy process and thereby contributing to evidence-based policymaking has not been elaborated upon in detail. Only few scholars have assessed where and how big data can be used in the specific phases of the policy process (Höchtl et al., 2016; Maciejewski, 2017). This paper contributes towards filling this research gap by examining how big data is used in the different phases of the public

policy process. Thus, the overall research question that motivate and guide this study is: *How can big data be used in different phases of public policy processes (planning, design, delivery, and evaluation of public policies)?*

The research question is answered with a systematic literature review, which results in an overview of how the use of big data along the policy process has been described in the academic literature. The paper builds on a theoretical framework, consisting of data-based theories in government as well as theories in the field of policy science. The former include the concepts of Digital-era Governance (DEG) (Clarke & Margetts, 2014), Big Data Readiness (Klievink et al., 2017), and Big and Open Linked Data (BOLD) (Janssen et al., 2017). In the field of policy science, this paper contributes to the research streams of evidence-based policymaking (Head, 2008; Sanderson, 2002) and policy analytics (De Marchi et al., 2016; Tsoukias et al., 2013).

First, the paper provides some theoretical background on the concept and definition of big data, as well as the above-mentioned theoretical concepts. Then, the method used to answer the research question – a systematic literature review – will be outlined in detail, before presenting and discussing the results. Finally, the paper concludes with its limitations, conclusion and an outlook.

2 Background

There is no agreed-upon academic or industry definition on big data. However, there is some consensus among scholars that it is characterized by a variety of factors, or "Vs", ranging from three to seven or more. This characterization was first introduced by Laney (2001, pp. 1-3) who described the three dimensions *volume, velocity* and *variety*. It has been expanded by various scholars to include further dimensions, such as veracity, volatility, complexity, etc. (see Desouza and Jacob, 2017; Malomo & Sena, 2017; Mayer-Schönberger & Cukier, 2017; Kim et al., 2014; Boyd & Crawford, 2012). Big data has many different definitions in different contexts. Some definitions emphasize the size of the data sets, which exceeds the ability of typical database software tools to capture, store, manage, and analyze them (see Desouza and Jacob, 2017; Klievink et al., 2017; Höchtl et al., 2016). In the context of policymaking, Bright and Margetts (2016) define big data as "the creative application of large transactional data sets generated by the Internet (such as comments on social media) to the processes of policymaking" (p. 221).

What exactly can be considered big data is constantly changing due to the high speed of technology advances, making it challenging to express it in specific and measurable terms (Klievink et al., 2017, p. 269). As there is no single generally accepted definition of big data, it is more insightful to take a closer look at the various concepts that are nested in the term big data. Therefore, the next section will provide an overview over the different characteristics of big data described in the current literature.

Big Data Characteristics

Various authors in the literature on big data in public policymaking have described different types of big data according to their characteristics. This section provides an overview over how the different types of big data have been conceptualized, along the characteristics of *structure, source, access,* and *size.*

A first characteristic discussed in the literature is the data structure. Here a distinction is made between structured and unstructured date. The difference lies in the format of the data – the format of unstructured data varies widely and cannot be stored in traditional databases without complex data transformations (Daniell et al., 2016, p. 6). Desouza and Jacob (2017) describe structured, semi-structured and unstructured data as follows (p. 1046):

- *Structured data* have an organized structure and are clearly identifiable, such as a database with specific information stored in columns and rows.
- *Semi-structured data* does not conform to a formal structure, but contains "tags" that enable the separation of data records or fields. An example is data in bibliographical software programs.
- *Unstructured data* have no identifiable structure, for example texts, photos, videos and audio files.

The effort for a computer system to automatically analyze and derive meaningful insights from unstructured data types is much higher and requires a framework to manage computations over large data quantities. However, unstructured data is much better suited to store knowledge (Höchtl et al., 2016, p. 153). The vast majority of available data is unstructured, as much of it originates from sensors or is contained in videos, images or textual information in social networks (Daniell et al., 2016, p. 6; Höchtl et al., 2016, p. 153).

A second characteristic discussed in the literature is the data source. Big data extends the sources of data traditionally used in policymaking. Traditional data sources of massive pre-designed data collection systems (e.g. census, tax collection or governmental surveys) coexist with other sources of large amounts of data that can be useful for various government departments (e.g. electronic medical records, meteorological data, data from surveillance cameras, social media, GPS tracking, etc.) (Daniell et al., 2016, p. 7; Ruggeri et al., 2017, p. 2). Dunleavy (2016, pp. 5-6) distinguishes between *administrative data* and the *digital residues* as the two main sources of new information for policy-makers. *Administrative data* is collected for transactional purposes, rather than being designed as a dataset for analysis or as part of the national statistics reporting. It typically records objective behaviors, not opinions. *Digital residues* allow government agencies to collect digital data series that resemble administrative data but contain a great deal of potentially useful text, image or sound information if decoded (Dunleavy, 2016, p. 10). Similarly, Severo et al. (2016, pp. 358-359) propose the term *soft data* to differentiate the diverse data sources on the Internet from traditionally collected administrative statistics. *Soft data* are defined as sources of information that are "freely available on the Internet, are not controlled by a public administration but are subject to the property rights of public or private actors" (Severo et al., 2016, p. 358).

Data access is a third characteristic of big data found in the literature. Heitmueller et al. (2014, pp. 1524-1525) distinguish among different data types in terms of who controls access to those types:

- *Personal and proprietary data* is controlled by individual or commercial entities, which typically have the right to restrict access to the data, e.g. personal health records or credit card information.
- *Government-controlled data* is data to which a government can restrict access, e.g. census data or personal tax or health records.
- *Open data commons* are data available to all. The data may be private, commercial or government controlled. Other than open data, open data commons are usually kept up-to-date and provided in accessible format, e.g. for geographic, climate, census or financial data.

Finally, data size is a fourth characteristic of big data. Desouza and Jacob (2017, p. 1047) developed a data continuum whereby "big data" reflects one extreme and "small data" the other (Table 1). Size in this sense does not refer to *volume*, which is only one out of four big data characteristics reflected in the continuum, but also includes *variety, velocity*, and *complexity*. While *volume* is a function of the capacity of an organization to collect, store and analyze its data, *velocity* is the speed at which data are created, stored and retrieved. *Variety* refers to the various structures of big data outlined above and *complexity* is the degree to which data are interconnected – many insights that emerge from big data applications are the result of connecting previously unrelated datasets (Desouza & Jacob, 2017, pp. 1046-1047).

Table 1: Big data continuum

"Small Data"				**"Big Data"**
Low volume Low velocity Low variety Low complexity	High volume Low velocity Low variety Low complexity	High volume High velocity Low variety Low complexity	High volume High velocity High variety Low complexity	High volume High velocity High variety High complexity
E.g. land use data for a small city	E.g. census data	E.g. Twitter data or video feeds	E.g. data-sets with different structures	E.g. linked data-sets with different structures

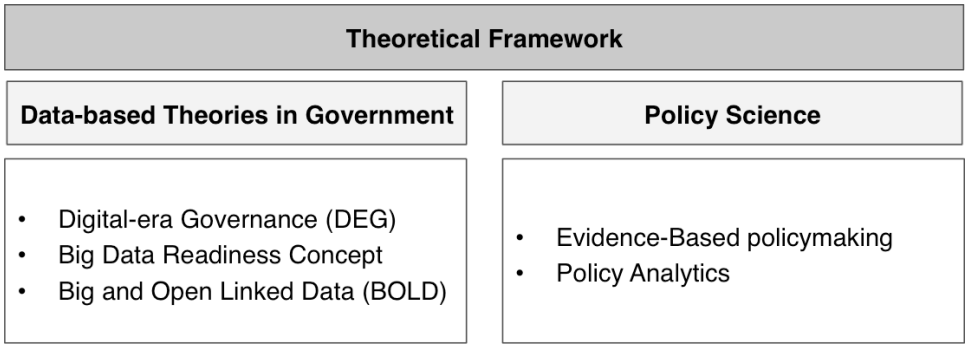
Source: Desouza & Jacob (2017, p. 1047)

To summarize, various authors in the literature on big data in the public sector have conceptualized big data types along the characteristics of *data structure, data source, data access,* and *data size*.

2.2 Big Data and Evidence-based Policymaking

Based on the above-explained features, issues, and research gaps regarding big data in the public sector, this section outlines the theoretical context of this paper. On the one hand, the research question of this paper is related to current data-based theories in government, on the other hand, it also concerns theories in the area of policy science. While the former are rather recent trends in the literature, they tie in with the broader concept of evidence-based policymaking, as explained below (Giest, 2017, p. 370). The combination of these two broader theories serves as a sound theoretical basis for this paper, as they cover newer data-related concepts as well as long-established and widely used theories. The theoretical framework of this paper is illustrated in Figure 1.

Figure 1: The theoretical framework of the paper



Data-based Theories in Government

According to Giest (2017, p. 368) big data in public policy can be analyzed using the current data-based theories of Digital-era Governance (DEG) (Clarke & Margetts, 2014), the Big Data Readiness concept (Klievink, et al., 2017), and the concept of Big and Open Linked Data (BOLD) (Janssen, et al., 2017). These three theoretical concepts build on e-government and New Public Management (NPM) research and tie in with the broader concept of evidence-based policymaking (Giest, 2017, pp. 368-370).

The *Digital-era Governance (DEG) concept* is a successor of the NPM concept. According to the DEG research stream, governments lag behind the private sector in the development of technology and digitization in public services, which leads to low levels of literacy regarding new technologies. Therefore, governments need to acquire new skills and develop the capacity to process information and realize desired outcomes (Giest, 2017, p. 369). DEG is characterized by the complete digitalization of paper and phone-based systems and places digital technologies at the center of bureaucracy. The use of big data in the public policy process contributes to the ideal type of DEG, by making the process more transparent, efficient, and citizen-focused (p. 396).

The *Big Data Readiness concept* raises complementary points to DEG. It assesses public capacities by looking the big data readiness of public organizations (Giest, 2017, p. 369). In order to evaluate big data readiness, Klievink et al. (2017, pp. 272-273) introduce an assessment framework, which rests on three component parts:

- *Organizational alignment* concerns an organization's current structure, main activities and strategy and whether big data use can be reconciled with them (p. 272).
- *Organizational maturity* indicates how far an organization has developed towards better collaboration with other public organizations and the provision of more citizen-oriented services and demand-driven policies (p. 273).
- *Organizational capabilities* address whether an organization possesses the required capacities to use and create value from big data and to avoid negative consequences from its use (p. 273).

According to Klievink et al. (2017), much work is still to be done to unlock the full potential of big data in the public sector. Therefore, public organizations should assess what the use of big data will require and what specific added value it could bring (pp. 275-277). In order to successfully incorporate the use of big data into the public policy process, all three components are important. However, the element of *organizational maturity* is closest to the research question at hand, as it concerns the provision of demand-driven policies and citizen orientation.

Janssen, et al. (2017, p. 189) identify *Big and Open Linked Data (BOLD)* as a driver of innovation in government. The concept of BOLD can be described as the integration of three major developments that affect our society (Janssen & van den Hoven, 2015, p. 363):

- Big data involves large volumes of data from various sources that need to be processed – value is created by combining different data sources.
- Open data enables access to data without any restrictions or usage conditions. The opening of data can lead to efficiency improvements, innovation, and more transparency.
- Linked data involves connecting structured and machine-readable data.

According to Janssen, et al. (2017, p. 189), *policymaking innovation* is the idea that government and the public can use big data to model and understand policy implications and to support policy decisions. The use of BOLD is often linked to enhancing evidence-based policymaking – it subjects policies and models to rigorous testing of their underlying assumptions and predictions and is less dependent on subjective assessment and opinions.

After the data-based pillar of the theoretical framework has been presented in this section, the following section concerns the pillar of policy science.

Policy Science

This section elaborates on the research field of policy science, with a focus on evidence-based policymaking and policy analytics, both of which are particularly relevant for the research endeavor of this paper.

Evidence-based policymaking is a relatively new topic, which has come into use over the recent years and has pervaded the last decade of debate in the social sciences (De Marchi et al., 2016, p. 23). It is consistent with NPM and the public sector's increased interest in efficiency and effectiveness (Head, 2008, p. 2).

According to Howlett et al. (2009), evidence-based policymaking "represents an effort to reform or restructure policy processes by prioritizing data-based evidentiary decision-making criteria over less formal or more 'intuitive' or experiential policy assessments in order to avoid or minimize policy failures caused by a mismatch between government expectations and actual, on-the-ground conditions" (p. 181).

Proponents of evidence-based policymaking believe that it offers decision-makers the opportunity for continuous improvement in policy settings and performance, based on rational evaluation and a well-informed debate of options. The analysis of evidence helps to answer the questions "what works?" and "what happens if we change these settings?" (Head, 2008, p. 1). Furthermore, governments can better learn from experience, avoid repeating past errors, and apply new techniques to old and new problems (Howlett et al., 2009, p. 182). Critics emphasize different forms of information competing in the policy process, which "further require the capacity of decision-makers to comprehend it" (Giest, 2017, p. 368). De Marchi et al. (2016, p. 35) suggest that evidence-based policymaking fails to address certain challenges: as evidence does not exist independently from policies, it does not "objectively" drive the policy process. Furthermore, the same "evidence" may carry multiple interpretations. Nevertheless, there is a demand for using analytic information to support policymaking. Therefore, De Marchi et al. (2016, p. 34) propose the concept of *policy analytics* as a new frame to address this demand differently.

Over the last decade, advances in big data and the growth of computational power have led to the development of the concept of "analytics" (Daniell, Morton, & Insua, 2016, p. 5). Analytics is an umbrella term that covers many different methods and approaches, such as statistics, data mining, decision support systems, etc. (De Marchi et al., 2016, p. 33). In recent years, analytics has been associated with big data, in order to take into account the availability of large databases (Tsoukias et al., 2013, p. 123). In the private sector, the problem of how evidence is constructed to design a business policy has led to the establishment of *business analytics* (De Marchi et al., 2016, p. 33). The same approach can also be applied to the public policy context, leading to the concept of *policy analytics* (Daniell et al., 2016, p. 7). The term was coined in the scientific literature papers by De Marchi et al. (2016) and Tsoukias et al. (2013). Longo et al. (2017, p. 82) equates the term policy analytics with "big data in public affairs" (Mergel et al., 2016, p. 931) and "policy informatics" (Johnston, 2015, p. 3). According De Marchi et al. (2016) *policy analytics* is the development and application of "skills, methodologies, methods and technologies, which aim to support relevant stakeholders engaged at any stage of a policy cycle, with the aim of facilitating meaningful and informative hindsight, insight and foresight" (p. 34). They therefore suggest the concept policy analytics to support policy makers in a way that is:

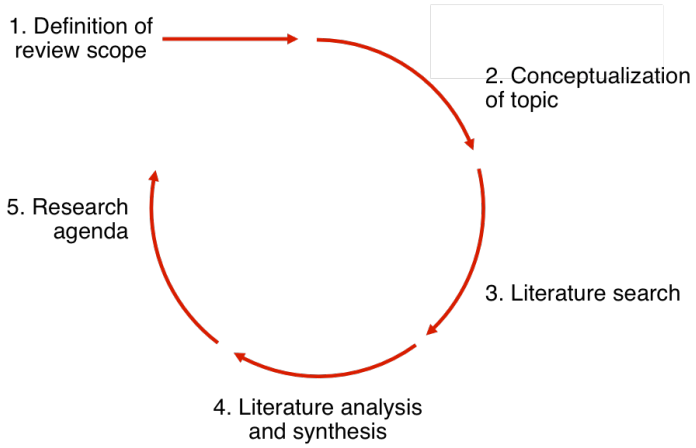
- Meaningful – relevant and adding value to the process
- Operational –practically feasible
- Legitimating – ensuring transparency and accountability (De Marchi et al., 2016, p. 34).

The question how big data can be used in the different phases of the policy process concerns both research streams, evidence-based policymaking and policy analytics. Many of the works analyzed in the literature review of this paper explicitly contribute to these two research fields.

3 Method

In order to answer the research question of this paper, a systematic literature was carried out, following a framework for conducting literature reviews proposed by vom Brocke et al. (2009, pp. 7-8), which includes five phases (see Figure 2).

Figure 2: Framework for literature reviewing



Source: own illustration based on vom Brocke et al. (2009, p. 7)

Phase 1- definition of review scope: Defining an appropriate scope of the review is a first major challenge, as the review can serve many different purposes, such as gaining new research outcomes or identifying research techniques. Also, it may be critical, historical, interpretative etc. (vom Brocke et al., 2009, pp. 6-7). Cooper's (1988) taxonomy for literature reviews can be a helpful tool to clearly define the scope of a literature review (Table 4). It consists of six characteristics, each of which contains various categories[1] (Cooper, 1988, pp. 108-112):

Table 4: Taxonomy of literature reviews and categorization of this literature review

Characteristic	**Categories**			
1. Focus	Research outcomes	Research methods	Theories	Applications
2. Goal	Integration and synthesis	Criticism	Identification of central issues	
3. Perspective	Neutral representation	Espousal of position		
4. Coverage	Exhaustive	Exhaustive with selective citation	Representative	Central or pivotal
5. Organization	Historical	Conceptual	Methodological	
6. Audience	Specialized scholars	General scholars	Practitioners or policy makers	General public

Source: own illustration based on Brocke et al. (2009, p. 7) and Cooper (1988, p. 109)

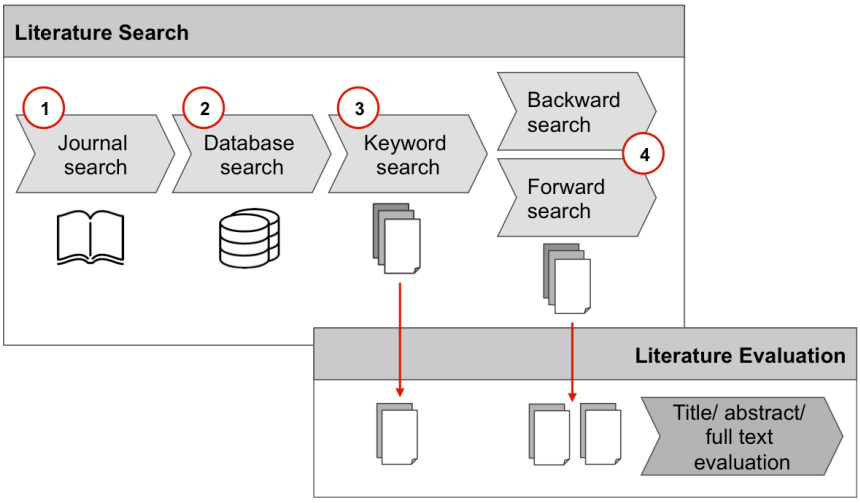
[1] *Perspective* and *coverage* are mutually exclusive, while *audience, organization, goal,* and *focus* can be combined.

As shown in Table 4, this literature review focuses (1) on material concerning research outcomes, theories and applications. The goal (2) of the review is the integration and synthesis of literature and the perspective (3) is neutral, as the author does not take a particular point of view. While the analysis of relevant literature on the use big data in the policy process is exhaustive, only a selected sample of works is described in this paper, in order to provide a detailed overview (4). The review is organized (5) in a conceptual manner – papers relating to the same phase of the policy process are presented together. Finally, the findings of the paper may be relevant for specialized scholars in the field of policy science, particularly regarding evidence-based policymaking and policy analytics, but also for practitioners and policy makers themselves (6).

Phase 2 – conceptualization of topic: According to vom Brocke et al. (2009, p. 8), a broad conception of what is known about the topic is essential at the beginning of a literature review – one should firstly consult the sources that provide a summary or an overview of the key issues relevant for the subject and provide working definitions of key terms. Therefore, relevant literature on big data, with a special focus on the public sector, as well as relevant works in the field of policy science were consulted before beginning the review.

Phase 3 – literature search: The goal of a systematic literature search is to accumulate a relatively complete census of relevant literature (Webster & Watson, 2013, p. 16). The literature search framework proposed by vom Brocke et al. (2009, p. 8) involves journal and database, keyword, backwards, and forward search, as well as an ongoing evaluation of sources (Figure 3).

Figure 3: Literature search process



Source: own illustration based on vom Brocke et al. (2009, p. 9)

It is recommended to focus on articles in scholarly journals or proceedings or renowned conferences (Webster & Watson, 2013, p. 16; vom Brocke et al., 2009, p. 8). For the next steps, database and keyword search, the *Web of Science* was used. Vom Brocke et al. (2009, p. 9) recommend using a precise set of search phrases in order to exclude contributions that cover topics or research questions irrelevant for the litera-

ture review (see 5). The next step is conducting *backward* and *forward search*[2]. In order to limit the amount of literature identified, the articles' contents need to be evaluated by analyzing their titles, abstracts or even full texts (vom Brocke et al., 2009, p. 9). Table 5 provides an overview over the terms used and the number of relevant articles identified for the literature review. The search has been limited by document type and includes only peer-reviewed journal articles and conference proceedings. In total, this literature search resulted in 22 relevant documents. Due to the novelty of the big data concept, the timeframe of the articles analyzed is from 2014-2017.

Table 5: Result of keyword, backward, and forward search

Keywords used	Articles obtained	Relevant articles
Big Data (title) AND Public Policy (title)	10	2
Big Data (title) AND Policy (title)	62	7
Big Data (title) AND Policymaking (title)	2	1
Big Data (topic) AND Policy analytics (title)	11	3
Big Data (title) AND Policy making (topic) AND Public (topic)	28	2
Big Data (topic) AND evidence-based policy (title)	1	1
Total keyword search	114	16
Backward and forward search	6	6
Total literature search	**120**	**22**

Phase 4 – literature analysis and synthesis: after sufficient literature has been collected, it has to be analyzed and synthesized. As this literature review is conceptual (concepts determine its organizing framework), it can be structured using a concept matrix (Webster & Watson, 2013, p. 17), as illustrated in Table 6.

Table 6: Concept matrix

Concept Matrix

Articles	**Concepts**				
	A	B	C	D	…
1		x	x		x
2	x	x			
…			x	x	

Source: Webster and Watson (2013, p. 17)

In the context of this literature review, the analysis was structured along the analysis of big data use in the different phases of the policy process. The relevant literature on the use of big data was analyzed along the planning, design, delivery, and evaluation phases of the policy process. The phases of the policy process –

[2] *Backward search* involves reviewing the citations in the articles from the keyword search, to determine older literature, *forward search* means reviewing sources that cite the key articles identified in the previous steps (Webster & Watson, 2013, p. 16).

planning, design, delivery, and evaluation – are elements found in many versions of the policy cycle described in the literature. Limiting the policy process to four phases enables a simplified and more comprehensive literature analysis. As the authors analyzed describe the use of big data in many distinct policy process phases, a certain grouping was necessary in order to provide a clear, yet distinct picture of the use of big data in the policy process.

Phase 5 – research agenda: finally, the synthesis of literature is expected to result in a research agenda, including more acute and insightful questions for future research (vom Brocke et al., 2009, p. 9).

4 The Use of Big Data in the Public Policy Process

Overall, many of the analyzed works outline the potential of using big data to improve the entire policy process. According to Maciejewski (2017), big data supports better policy development and execution *"by strengthening the information input for evidence-based decision-making and provides more immediate feedback on policy and its impacts"* (p. 127). According to Schintler and Kulkarni (2014), big data has great potential as a resource for helping to inform different points in the policy analysis process *"from problem conceptualization to ongoing evaluation of existing policies, and even empowering and engaging citizens and stakeholders in the process"* (p. 343). This section analyzes how big data can be used in the four phases of the policy process, according to the reviewed literature. Table 7 provides a concept matrix of the different policy process phases and demonstrates which author has described the use of big data in which phase.

Table 7: Concept matrix of big data descriptions in the different policy process phases

	Policy Process Phase			
Literature	Planning	Design	Delivery	Evaluation
Lavertu (2014)				X
Schintler & Kulkarni (2014)	X			X
Burnap & Williams (2015)	X			
Whitman Cobb (2015)	X			
Alfaro et al. (2016)	X			
Bright & Margetts (2016)	X			
Ceron & Negri (2016)	X	X	X	X
Daniell et al. (2016)	X	X	X	X
De Gennaro et al. (2016)		X		
Dunleavy (2016)			X	X
Höchtl et al. (2016)	X	X	X	X
Lee et al. (2016)	X			
Severo et al. (2016)	X			
Williamson (2016)		X		
Semanjski et al. (2016)		X		

Brayne (2017)		X	
Giest (2017)	X		
Guerrero & Lopez (2017)	X		
Longo et al. (2017)	X	X	X
Maciejewski (2017)		X	
Panagiotopoulos et al. (2017)	X	X	X
Ruggeri et al. (2017)			X

While the use of big data can be pinpointed to one distinct phase of the policy process in most of the analyzed works, a small number of authors describe the use of big data in multiple or even all phases of the policy process. The following sections shall analyze the use of big data in four distinct phases of the policy process in more detail, summarized in Tables 8-11. The tables contain the categories of techniques, data types, and goals and examples of big data in the respective phase, as described by the authors. These reflect the focus points of the literature analysis, in order to provide a comprehensive overview for each phase, while being able to compare the different authors along certain criteria.

Planning Phase

The discussions and examples described in the planning phase evolve around agenda-setting, problem definition, policy discussion, and participation. Agenda-setting is concerned with the way problems are recognized as requiring government attention, i.e. the identification and specification of the problems that may become the target of public policies (Anderson, 2014, p. 4; Howlett, Ramesh, & Perl, 2009, p. 92). Various scholars have described the use of big data in the problem definition and agenda-setting stage of policy process. According to Longo et al. (2017), big data can serve as an input for "framing a policy problem before it is apprehended as such, indicating where a need is being unmet or where an emerging problem might be countered early" (p. 83). It has long been recognized that media play a central role in agenda-setting by framing issues and spreading relevant information (McCombs & Shaw, 1972). According to Höchtl et al. (2016, p. 159), digital media add additional complexity to the dynamics of agenda-setting. Through the use of social media, any audience member can easily initiate new discussions, and responses to existing discussions can take various forms, such as text, audio, video or images. Therefore, one way for governments to identify emergent topics early and to create relevant agenda points is "to collect data from social networks with high degrees of participation and try to identify citizens' policy preferences, which can then be taken into account by the government in setting the agenda" (Höchtl et al., 2016, p. 159).

Policy discussion focuses on debating the different policy options for the issue agreed upon in the agenda-setting stage. Big data can play a significant role regarding the details of pressing policy problems – it can set policy priorities, such as infrastructure, security, education etc. An example is Boston's *Street Bump Application*, which measures the smoothness of car rides based on movements of cell phones, thereby identifying the areas that should be prioritized for infrastructure improvements (Höchtl et al., 2016, p.

160). This information can be used in open policy discussions by helping to find the most efficient starting point for implementation. Sentiment analysis and opinion mining can also be used to identify opinion streams linked to any topic of interest in public policy, mentioned in textual messages (Alfaro et al., 2016, p. 198). The following two examples show in more detail how Internet-based big data can inform agenda-setting and policy discussion.

Whitman Cobb (2015, pp. 11-12) analyzes new measures for tracking public opinion of U.S. space policy that are enabled by big data, such as Google Trends and social media sources. Public opinion plays an important role in setting the direction for U.S. space exploration – however, the tools that have been used to measure it have suffered from limitations related to time and a lack of available data (p. 11). Google Trends offers a wide range of potential data sources, broken down by country, state, region, and time. While Google Trends offers a longer-term view, Twitter provides policymakers with information on what people are interested in at particular points in time (pp. 12-13). Together, they provide a flexible tool, through which policy analysts can measure public interest. "Policy entrepreneurs in both the space community and political community could find this type of data valuable as they endeavor to lobby Congress and the executive branch to support further activities or spending for NASA" (Whitman Cobb, 2015, p. 16).

Panagiotopoulos et al. (2017, p. 604) examine the value of social media data as part of the policymaking cycle and evidence-based policymaking. They conducted an exploratory study with the UK Department for Environment, Food and Rural Affairs (DEFRA), focusing on farming and agricultural policy. The goal was to explore how collective input by farmers on Twitter could be suitable as input in policy activities. Communities have formed on Twitter around influential accounts and hashtags (e.g. #AgriChatUK, @FarmersGuardian or @NFUTweets) (p. 606). The cluster mapping technique was used to summarize and visualize the large exploratory Twitter dataset and discover how conversations evolve (pp. 604-609). Two separate farming-relevant branches were identified, dairy farming and arable farming. Terms clustering around dairy farming were often connected to topics concerning renewable energy, showing a mutually connected relevance, which provided interesting insights for policymakers. Terms clustering around arable farming were often associated to terms like "economy", "government" or "support", showing that the issues of government funding is interconnected with arable faming – this indicates farmers' perceptions of how government funding is directed (pp. 605-607). The connected topics to both arable and dairy farming thus provided important input for policy activities of the DEFRA.

Table 8: The Use of Big Data in the Planning Phase

Phase	**Literature**									
Planning Phase	Longo et al. (2017)	Höchtl et al. (2016)	Severo et al. (2016)	Alfaro et al. (2016)	Whitman Cobb (2015)	Lee et al. (2016)	Burnap & Williams (2015)	Schintler & Kulkarni (2014)	Bright & Margetts (2016)	Panagiotopoulos et al. (2017)
Technique			Soft data analysis	Text & sentiment analysis; opinion mining		Machine classification; statistical modeling	Supervised machine learning			Social media analysis; cluster mapping
Data Types		Social media data	Social media data	Social media data	Social media data	Social media data	Social media data		Social media data	Social media data
Goal	Agenda-setting	Agenda-setting; policy discussion	Agenda-setting	Identifying opinion streams and incorporating this feedback into the policy process	Measuring salience	Informing policymaking to formulate policy that is suited to the needs of local people	Monitoring the reaction to large-scale emotive events that may lead to hate crimes	Enhancing stakeholder participation & accountability	Enabling more accessible forms of public (mass) participation in policymaking	Discover conversations and spontaneous reactions in/near real time t
Example		Boston's Street Bump Application			Measuring public opinion of space policy in the U.S.	Measuring the effect of environmental attitudes of citizens on the adoption of green electricity policies in the U.S.	Analyzing the spread of online hate speech immediately after the murder of Drummer Lee Rigby in London			Analyzing collective input by farmers on Twitter for policy activities of the DEFRA

Source: own illustration

Design Phase

The concept of policy design is linked to the idea that governments want to implement goals effectively and efficiently and are interested in using knowledge and experience about policy issues (Giest, 2017, p. 371). Once a public problem has entered onto the formal agenda of government, policy makers can formulate specific courses of action. *Policy formulation* involves the development of alternative courses of action for dealing with (resolving or ameliorating) a public problem (Anderson, 2014, p. 4). According to Giest (2017), most of the design activities come into play at the formulation stage. "The policy design concept looks at these considerations in policy formulation and the outcomes in implementation" (p. 371). This perspective pays special attention to policy instruments. When exploring policy options, policy makers consider not only what to do but also how to do it.

Giest (2017) links these concepts to big data and argues that the increased use of big data is shaping policy instruments. "The vast amount of administrative data collected at various governmental levels and in different domains, such as tax systems, social programs, health records and the like, can— with their digitization—be used for decision-making in areas of education, economics, health and social policy" (Giest, 2017, p. 376). The different information-based policy tools used by governments illustrate the various ways in which big data can be used for pursuing specific policy outcomes (pp. 371-372). *Procedural* informational instruments describe government activities to regulate information. They are "designed to affect policy processes in a way consistent with government aims and ambitions through the control and selective provision of information" (Howlett, 2011, p. 120). Some efforts are aimed at promoting information release (e.g. freedom of information legislation) while others are aimed at preventing it (e.g. censorship). Open data policy frameworks or the release of government data are examples of how big data can be used as procedural policy instruments (Giest, 2017, p. 377). Substantive informational instruments "describe government collecting data to enhance evidence-based policymaking" (Giest, 2017, p. 377), such as judicial inquiries, executive commissions, national statistical agencies, surveys and polling (Howlett, 2011, pp. 118-119). Governments increasingly complement these more traditional data with (real-time) big data based on social media input, cameras and sensors (Giest, 2017, p. 376).

The education sector is an example where real-time big data techniques are increasingly used as policy instruments. According to Giest (2017, p. 377) and Williamson (2016, p. 134), they provide up-to-date information on the education system for policymakers, e.g. by creating digital and interactive data visualizations. Learning analytics platforms are able to capture data from children's educational activities to track and assess their development and attainment and to algorithmically optimize and customize their future educational experience. For policy-makers, this provides fine-grained knowledge, which can be used to formulate policy options (Giest, 2017, pp. 377-378; Williamson, 2016, pp. 136-137).

Table 9: The Use of Big Data in the Design Phase

Phase	**Literature**					
Design Phase	Giest (2017)	Williamson (2016)	Höchtl et al. (2016)	Guerrero & Lopez (2017)	De Gennaro et al. (2016)	Semanjski et al. (2016)
Technique	Data visualization	Learning analytics platforms	Advanced predictive analytics methodologies; scenario techniques	Network science; computational methods	Data collection with navigation systems; data processing platform	Smartphone app based data collection process; machine learning
Data Types	Administrative data, complemented with (real-time) data based on social media input, cameras and sensors	Educational data		Big data from administrative records (administrative data)	Datasets of driving and mobility patterns	Mobility data contributed by citizens
Goal	Shaping information-based policy instruments	Providing fine-grained knowledge and intelligence to formulate policy options	Contributing to evidence-based policymaking in the phase of policy formation	Using highly granular big data sets to create better policymaking tools	Increasing the effectiveness of future policies in the fields of transport and energy	Providing insights on mobility indicators and imminent feedback on implemented measures; shorter data collection and processing phases
Example	Providing real-time information on the education system for by creating digital and interactive data visualizations	Learning analytics platforms capture data from children's educational activities to track and algorithmically optimize their educational experience; predicting the future performance of the system and the student		Using employer-employee microdata in the labor flow network (LFN) model to capture labor mobility patterns and construct new labor market measures and policymaking tools for unemployment policies	Using the Transport Technology and Mobility Assessment (TEMA) processing platform s to support the development of effective transport regulation in the EU	Using the application *Routecoach,* 8300 citizens in Leuven voluntarily contribute their mobility data. Their crowd-sourced behavior was mapped to the wider population using machine learning approach

Source: own illustration

Stakeholder participation can also play an important role when using big data in the design phase, particularly regarding substantive information-based policy tools. Semanjski, et al. (2016, p. 14) show how big data can be used for decision-making in the area of transport systems with an app-based data collection process: the application *Routecoach* enables citizens to voluntarily contribute their mobility data. Using a machine leaning based approach, the results of more than 8,300 participants in the city of Leuven could be used to derive insights on various sustainable mobility indicators, such as CO_2 emissions or cost per trip (pp. 5-6). Due to shorter data collection and processing phases as well as the improved relevance of the data, policymakers receive imminent feedback on implemented measures, e.g. the construction of a new bike line, speeding up the decision-making process (p. 14).

Delivery Phase

When describing the use of big data in the delivery phase, most authors refer to the implementation stage of the policy process – i.e. the application of the policy, which often includes further development or elaboration of policies (Anderson, 2014, p. 4). It comprises "the effort, knowledge, and resources devoted to translating policy decisions into action" (Howlett et al., 2009, p. 160). Although the means to pursue a policy goal are mostly identified in the policy decision, subsequent choices are inevitably required to make a policy work, such as allocating funds, assigning personnel and developing rules of procedure (Howlett et al., 2009, p. 160). One way in which big data can influence the implementation stage of the policy process is the real-time production of data. The execution of new policies immediately produces new data, which can be used to evaluate the effectiveness of policies and improving the future implementation processes. Testing a new policy in real time can provide insights whether it has the desired effect or requires modification. This leads to increased autonomy for public administrations, which are enabled to react quickly to evaluation results (Höchtl et al., 2016, p. 162). Governments can use real-time micro-experimentation to test policies by manipulating input variables in law, markets, architecture, social norms, and information. The impacts that correlate with these changed variables can be measured with great accuracy in order to propose, test, evaluate, and redesign policy intervention (Longo et al., 2017, p. 83).

Dunleavy (2016, pp. 12-13) shows the potential of using big data for behavioral insights. *Online randomized control trials* (online RCTs) enable the evaluation of small-scale effects using the availability of huge datasets – they can often be undertaken at low cost and in real time by government agencies or businesses. For example, the UK is getting 1.9 million people a year to pay court fines – the government is chasing unpaid debts using contractors, which involves great costs. However, people's willingness to pay may be influenced by very small factors, e.g. the design of reminder letters. Using an *online RCT*, several treatments, such as redesigned forms of the reminder letter, can be sent to large, randomly assigned treatment groups that are compared with a control group. Finding out which treatment works best can generate great saving for government finances (p. 13).

Table 10: The Use of Big Data in the Delivery Phase

Phase	**Literature**									
Delivery Phase	Höchtl et al. (2016)	Höchtl et al. (2016)	Höchtl et al. (2016)	Longo et al. (2017)	Dunleavy (2016)	Dunleavy (2016)	Maciejewski (2017)	Maciejewski (2017)	Maciejewski (2017)	Brayne (2017)
Technique				Real-time micro-experimentation	Online randomized control trials (online RCTs)	Sentiment analysis; machine learning; data mining	Sentiment analysis	Big data analytics	Performance monitoring; network analysis	Predictive analytics; network analysis
Data Types	Real-time data		Budgetary data	Real-time data		Social media data	Social media data		Mobility data	
Goal	Testing new policies by using real-time data produced in the execution	Improving the accuracy of information sources for policy implementation, e.g. census data	Improving decisions on required personnel and financial means for policy implementation	Testing policies by manipulating input variables	Evaluating small-scale effects using the availability of huge datasets	Preventive policing to improve arrest records, crime prevention and deterrence effects	Deriving feedback about policies to make an immediate response	1) Public supervision 2) Public regulation	Optimizing the transport infrastructure and commuting patterns	Using big data for a law enforcement-related activities
Example	Reducing crime rates at their origin by focusing an increase in policing more specifically on problem areas		Analysing the data generated from budgetary processes to detect patterns and design more efficient provision of means for a policy;		Using online RCTs to test the design of reminder letters for court fines (and ist influence on peoples' willingness to pay)	1) Police forces in Manchester monitored would-be rioters' chatter on social media and broadcasted own messages 2) Predictive policing in LA	Using the software Vizie (monitoring and analysis tool for social media) to quickly alert decision-makers of any changes that might require their attention	1) Using big data analytics to detect tax fraud, e.g. the British Connect system 2) Using big data to monitor adverse effects of FDA-approved drugs	The Land Transport Authority in Singapore (LTA) applies big data methods to improve public transport by gathering information of daily commuter rides	The LAPD compiles and analyzes big data for predictive policing. Geo-fences are used to generate real-time notifications and ALPR data is used for investigations

Source: own illustration

Predictive policing programs are another example of using big data in the delivery phase. The Los Angeles Police Department (LAPD) is at the forefront of data analytics and invests heavily in its data collection, analysis, and deployment capacities, in order to harness big data (Dunleavy, 2016, p. 15). It uses big data systems and predictive analytics for a wide variety of law enforcement-related activities, such as algorithms predicting where and when future crimes are most likely to happen or risk models that identify officers most likely to engage in at-risk behavior (Brayne, 2017, p. 981; 984). The LAPD uses *Palantir*, one of the leading analytic platforms for law enforcement and intelligence agencies to compile and analyze massive and disparate data. One of the most fundamental transformations is that the police increasingly utilizes data on individuals who have not had any police contact before. For example, *Automatic License Plate Readers* (ALPRs) take readings on everyone and create data that can be used in several ways. Cameras on police cars and static ALPRs at intersections take two photos of every car and record the time, date, as well as GPS coordinates. The ALPR data can then be compared against a "heat list" of outstanding warrants or stolen cars, a geo-fence can be placed around a location to track cars near the location or the data can be stored for potential use in future investigations (Brayne, 2017, pp. 992-993).

Evaluation Phase

The concept of policy evaluation refers to the stage of the policy process at which it is determined how a public policy has performed in action – i.e. a policy's consequences, whether it was effective, and why or why not. It involves an evaluation of the means being employed and the objectives being served (Anderson, 2014, p. 4; Howlett et al., 2009, p. 178). After a policy has been evaluated, it may be reconceptualized or the status quo may be maintained – reconceptualization can happen at the planning or any other phase of the policy process and may consist of minor changes or fundamental reformulation of the problem, including a termination of the policy (Howlett et al., 2009, p. 178). The use of big data in this phase of the policy process was least frequently described in the literature.

Traditionally, evaluation happened at the end of the policy process. Big data enables fast policy evaluation, which allows the responsible departments of public administrations to find out whether policies have the desired effect in a short time (Höchtl et al., 2016, p. 149). As apparent from the previous sections, it is often pointed out that big data can be used for continuous evaluation of policies, instead of merely as a last phase of the policy cycle – evaluation using big data was mentioned in all phases of the policy process in the examined works. Höchtl et al. (2016, pp. 162-163) suggest a novel approach to evaluation: they describe continuous evaluation is an integral part of every policy process phase, instead of a clearly defined process step at the end of the cycle. They propose a redesigned policy cycle, in which "evaluation does not happen at the end of the process but continuously, opening permanent possibilities of reiteration, reassessment, and consideration" (p. 162). Schintler and Kulkarni (2014, p. 343) also describe ongoing evaluation of existing policies as one of the great potentials of big data to inform the policy analysis process, which can even empower and engage citizens and stakeholders in the process.

Table 11: The Use of Big Data in the Evaluation Phase

Phase	**Literature**				
Evaluation Phase	Ceron & Negri (2016)	Höchtl et al. (2016)	Schintler & Kulkarni (2014)	Ruggeri et al. (2017)	Lavertu (2014)
Technique	Supervised Aggregated Sentiment Analysis (SASA)				
Data Types	Social media data				
Goal	Ex-post evaluation of policies (reaction of online public opinion on policy alternatives and monitoring mobilization of opposition groups)	Continuous evaluation using big data as an integral part of every policy process phase	Ongoing evaluation of existing policies using big data to empower and engage citizens and stakeholders in the process	Smart regulation: using big data when rolling out policies to revise interventions in real time	External political actors are increasingly able to observe and evaluate the administration of public programs, using performance information
Example	Analysing citizens' opinions on two major public policies in Italy (job market reform & school reform) using Twitter data	The *Automated Continuous Evaluation System* of the U.S. Army uses big data analytics and context aware security to analyze government, commercial, and social media data to uncover patterns of applicants			

Source: own illustration

5 Discussion and Conclusion

This section summarizes and discusses the main findings of the systematic literature analysis and provides the answer to the research question of this paper.

Evidence-based policymaking favors data-based decision-making criteria over more 'intuitive' or experiential policy assessments. The goal is to minimize policy failures that result from diverging government expectations and actual conditions (Howlett et al., 2009, p.181). Rational evaluation and a well-informed debate of options offer decision-makers the opportunity for continuous improvement. As the literature review in this paper has shown, the use of big data in each of the different phases of the policy process directly contributes to reaching these goals.

The findings of the *planning phase* show that the authors described the use of big data in the areas of agenda-setting, problem definition, policy discussion, and citizen participation. Social media data was the predominantly described type of big data in this phase. It can be used to identify citizens' policy preferences, in order to take them into account in setting the agenda and debating the different policy options. The techniques of sentiment analysis, opinion mining, machine learning, and clustering were included in the descriptions and examples, which ranged across different policy areas.

In the *design phase*, the authors have described the use of big data for policy formulation and as information-based policy instruments, particularly highlighting the contribution to evidence-based policymaking. The types of big data described include educational data, employer-employee microdata, and mobility data (e.g. GPS-data). The techniques employed in this phase often have a predictive element, such as predictive analytics, scenario techniques, network science, computational methods, and data visualizations.

In the *delivery phase*, the focus was particularly on the real-time production of data and immediate feedback, including public supervision and public regulation. This enables continuous evaluation of the effectiveness of policies, in order to improve future implementation processes. The described types of big data included social media data, mobility data, administrative data (e.g. census and budgetary data), while the employed techniques ranged from machine learning, data mining to sentiment and network analysis, as well as online RCTs.

Finally, most of the authors who mention the *evaluation phase* point out the potential of big data to enable *continuous* evaluation of policies – as outlined above, evaluation using big data was part of all policy process phases in the examined works. The real-time production and processing of data, sentiment analysis and micro-experimentation are the most frequently mentioned techniques in this regard.

The results of systematic literature review therefore provide an answer to the research question of this paper and show how big data can big data be used in different phases of public policy processes (planning, design, delivery, and evaluation of public policies).

Big data can be used in all of the four examined phases of the policy process, as the analysis and synthesis of the numerous descriptions and examples in the analyzed literature have shown. While the different types and techniques of big data outlined above are used in all of the four phases, the described use of big data has different focus areas in each of the phases. While the planning phase is dominated by descriptions and examples of social media used for agenda-setting, the uses described in the design phase often have a predictive element. The focus of big data use in the delivery and evaluation phases was on real-time data, providing immediate feedback. Also, as tables 8-11 show, the use of big data was most frequently described in the planning and delivery phase of the policy process, followed by descriptions in the design phase. As continuous policy evaluation was often mentioned as part of all policy process phases, explicit descriptions concerning big data use evaluation phase were scarcest.

The advances in information-based policy tools with a predictive element, the real-time production of data providing immediate feedback, as well as the possibility of continuous evaluation of policies particularly contribute to evidence-based policymaking. Furthermore, the literature analysis has shown that the use of big data in the different phases of the policy cycle enables policy analytics – supporting relevant stakeholders engaged at any stage of the policy cycle, aiming at facilitating informative hindsight insight and foresight (De Marchi et al., 2016, p. 34). Considering the three criteria of policy analytics introduced by De Marchi et al. (2016, p. 34) – meaningful, operational, and legitimating – the above-described results show that the three criteria apply to the use of big data in all phases of the policy process. However, while the use of big data is meaningful in all four phases (relevant and adding value to the process), the real-time production of data in the delivery phase, which enables continuous evaluation, may be the most meaningful contribution of big data, compared to policymaking in a pre-big data setting. Furthermore, the use of big data is operational and legitimating particularly in the planning phase, as social media data are easily accessible and stakeholder participation is improved.

The findings therefore contribute to closing the research gap identified at the beginning of this paper and tie in with the theoretical framework of data-based theories in government and policy science.

Limitations and further research

As mentioned in the beginning, the concept of big data is still in its infancy. Many questions regarding the true potential of big data, as well as where, when and how it is likely to be successful in the area of policymaking are still unanswered. As the analyzed literature is very new, it is often based on assumptions or sets a strong focus on the potential of big data. The analysis and synthesis of this literature therefore entails the risk of providing a hypothetical rather than actual picture of how big data can be used in the policy process.

Although the literature analyzed in this paper contains various real-world examples, most big data projects in policymaking are still in a planning or early implementation phase. The validity of the results would be further improved by including own case studies of real-world examples. Also, instead of focusing solely on

academic journals and conference proceedings, future research endeavors might also take into account reports, textbooks, etc. as the study of big data in policymaking advances. Furthermore, future research should use the analysis of case studies to provide more detailed insights. Also, it could focus on a particular level of government (local, regional, national etc.), policy area or specific data types and techniques. The paper at hand has not taken a focused perspective on any of these elements, as the literature on big data in the policy process is still too scarce and would therefore not have provided enough documents for analysis.

Finally, as mentioned above, "surrounding conditions", such as privacy implications, differences among different political systems, or the ability and skills of governments to use big data, were not considered in the analysis, as this would have gone beyond the scope of this paper. Future research should take these factors into account, as they may provide important limitations, threats, and requirements that need to be considered by practitioners and scholars when assessing the use of big data in policymaking.

Conclusion and outlook

The use of big data in the different phases of the policy process is of paramount significance for the goal of evidence-based policy-making or "the attempt to ground policy making in more reliable knowledge of 'what works'", as Sanderson (2002, p. 1) describes it. There is a high demand for using analytic information to support policymaking – policy analytics relies on big data to "support relevant stakeholders engaged at any stage of a policy cycle, with the aim of facilitating meaningful and informative hindsight, insight and foresight" (De Marchi et al., 2016, p. 34). But big data alone are not enough to provide the insights needed for evidence-based policymaking and policy analytics. The theories, techniques and best practices required to enable the successful use of big data need to be investigated with rigor, in order to harness the potential of big data in policymaking and to counter its threats. This paper contributes to closing the research gap associated with these demands. However, much work remains to be done in this area, in order to ease the tension between the promise and reality of big data in the public sector.

References

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications, 6*(25), 1-15. doi: 10.1186/s13174-015-0041-5

Alfaro, C., Cano-Montero, J., Gómez, J., Moguerza, J. M., & Ortega, F. (2016). A multi-stage method for content classification and opinion mining on weblog comments. *Annals of Operations Research, 236*(1), 197-213. doi: 10.1007/s10479-013-1449-6

Amankwah-Amoah, J. (2015). Safety or no safety in numbers? Governments, big data and public policy formulation. *Industrial Management & Data Systems, 115*(9), 1596-1603. doi: 10.1108/IMDS-04-2015-0158

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired.* Retrieved May 11, 2018, from Wired: https://www.wired.com/2008/06/pb-theory/

Anderson, J. E. (2014). *Public Policymaking: An Introduction, Eighth Edition.* Stamford: Cengage Learning.

Anderson, J. E. (1984). *Public Policy-Making: An Introduction, Third Edition.* Boston: Houghton Mifflin.

Australian Government Information Management Office. (2013). *The Australian Public Services Big Data Strategy.* Commonwealth of Australia. Retrieved March 12, 2018, from https://www.finance.gov.au/sites/default/files/Big-Data-Strategy.pdf

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography, 3*(3), 274-279. doi: 10.1177/2043820613513390

Barbero, M., Coutuer, J., Jackers, R., Moueddene, K., Renders, E., Stevens, W., et al. (2016). *Big data analytics for policy making.* European Commission, Directorate-General for Informatics. Retrieved April 21, 2018, from https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf

Birkland, T. A. (2015). *An Introduction to the Policy Process: Theories, Concepts, and Models of Public Policy Making.* New York: Routledge.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15*(5), 662-679. doi: 10.1080/1369118X.2012.678878

Brayne, S. (2017). Big Data Surveillance: The Case of Policing. *American Sociological Review, 82*(5), 977-1008. doi: 10.1177/0003122417725865

Bright, J., & Margetts, H. (2016). Big Data and Public Policy: Can It Succeed Where E-Participation Has Failed? *Policy & Internet, 8*(3), 218-224. doi: 10.1002/poi3.130

Burnap, P., & Williams, M. L. (2015). Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making. *Policy & Internet, 7*(2), 223-242. http://10.1002/poi3.85

Butler, D. (2013). *When Google got flu wrong. US outbreak foxes a leading web-based method for tracking seasonal flu.* Retrieved February 10, 2018, from https://www.nature.com/news/when-google-got-flu-wrong-1.12413

Ceron, A., & Negri, F. (2016). The "Social Side" of Public Policy: Monitoring Online Public Opinion and Its Mobilization During the Policy Cycle. *Policy & Internet, 8*(2), 131-147. doi: 10.1002/poi3.117

Clarke, A., & Margetts, H. (2014). Governments and citizens getting to know each other? Open, closed, and big data in public management reform. *Policy & Internet, 6*(4), 393-417. doi: 10.1002/1944-2866.POI377

Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society, 1*, 104-126. doi: 10.1007/BF03177550

Daniell, K. A., Morton, A., & Insua, D. R. (2016). Policy analysis and policy analytics. *Annals of Operations Research, 236*(1), 1-13. doi: 10.1007/s10479-015-1902-9

De Gennaro, M., Paffumi, E., & Martini, G. (2016). Big data for supporting low-carbon road transport policies in europe: Applications, challenges and opportunities. *Big Data Research* (6), 11-25. doi: http://10.1016/j.bdr.2016.04.003

De Marchi, G., Lucertini, G., & Tso, A. (2016). From evidence-based policy making to policy analytics. *Annals of Operations Research, 236*(1), 15-38. doi: 10.1007/s10479-014-1578-6

Desouza, K. C., & Jacob, B. (2017). Big Data in the Public Sector: Lessons for Practitioners and Scholars. *Administration & Society, 49*(7), 1043-1064. doi: 10.1177/0095399714555751

Dunleavy, P. (2016). 'Big data' and policy learning. In G. Stoker, & M. Evans (Eds.), *Evidence-based Policy Making in the Social Sciences: Methods That Matter* (pp. 143-151). Bristol: Policy Press. Retrieved 12 March, 2018, from https://www.researchgate.net/profile/Patrick_Dunleavy/publication/299467976_%27Big_data%27_and_policy_learning/links/56fa2cef08ae81582bf4435e/Big-data-and-policy-learning.pdf?origin=publication_detail

Giest, S. (2017). Big data for policymaking: fad or fasttrack? *Policy Sciences, 50*(3), 367-382. doi: 10.1007/s11077-017-9293-1

Guerrero, O. A., & Lopez, E. (2017). Understanding Unemployment in the Era of Big Data: Policy Informed by Data-Driven Theory. *Policy & Internet, 9*(1), 28-54. doi: 10.1002/poi3.136

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce, 26*(1-2), 147-169. doi: 10.1080/10919392.2015.1125187

Hart, C. (1998). *Doing a literature review: releasing the social science research imagination.* London: Sage Publications.

Head, B. W. (2008). Three Lenses of Evidence-Based Policy. *Australian Journal of Public Administration, 67*(1), 1-11. doi: http://10.1111/j.1467-8500.2007.00564.x

Heitmueller, A., Henderson, S., Warburton, W., Elmagarmid, A., Pentland, A. S., & Darzi, A. (2014). Developing public policy to advance the use of big data in health care. *Health Affairs, 33*(9), 1523-1530. doi: 10.1377/hlthaff.2014.0771

Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying public policy: Policy cycles and policy subsystems* (Vol. 3). Ontario: Oxford University Press.

Jann, W., & Wegrich, K. (2007). Theories of the policy cycle. In F. Fischer, G. J. Miller, & M. S. Sidney (Eds.), *Handbook of Public Policy Analysis. Theory, Politics, and Methods.* Boca Raton: CRC Press.

Janssen, M., & Kuk, G. (2016). Big and open linked data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce, 26*(1-2), 3-13. doi: 10.1080/10919392.2015.1124005

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly, 32*(4), 363-368. doi: http://10.1016/j.giq.2015.11.007

Janssen, M., Konopnicki, D., Snowdon, J. L., & Ojo, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information Systems Frontiers, 19*(2), 189-195. doi: 10.1007/s10796-017-9746-2

Jarmin, R. S., & O'Hara, A. B. (2016). Big data and the transformation of public policy analysis. *Journal of Policy Analysis and Management, 35*(3), 715-721. doi: http://10.1002/pam.21925

Johnston, E. W. (2015). *Governance in the Information Era: Theory and Practice of Policy Informatics.* New York: Routledge.

Kauffman, R. J., Kim, K., Lee, S.-Y. T., Hoang, A. P., & Ren, J. (2017). Combining machine-based and econometrics methods for policy analytics insights. *Electronic Commerce Research and Applications, 25*, 115-140. doi: http://10.1016/j.elerap.2017.04.004

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. *Communications of the ACM, 57*(3), 78-85. doi: 10.1145/2500873

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1-14. doi: 10.1007/s10708-013-9516-8

Klievink, B., Romijn, B.-J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers, 19*(2), 267-283. doi: 10.1007/s10796-016-9686-2

Lane, J. (2016). Big data for public policy: The quadruple helix. *Journal of Policy Analysis and Management, 35*(3), 708-715. doi: http://10.1002/pam.21921

Laney, D. (2001). *3D Data management: Controlling data volume, velocity and variety*. Meta Group. Retrieved April 28, 2018, from: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lasswell, H. D. (1971). *A Pre-View of Policy Sciences*. New York: American Elsevier.

Lasswell, H. D. (1958). *Politics: Who Gets What, When, How*. New York: Meridian Books.

Lasswell, H. D. (1956). *The Decision Process: Seven Categories of Functional Analysis*. College Park: University of Maryland Press.

Lasswell, H. D. (1951). The Policy Orientation. In S. Braman (Ed.), *Communication researchers and policy-making.* Cambridge: The MIT Press.

Lavertu, S. (2014). We All Need Help: "Big Data" and the Mismeasure of Public Administration. *Public Administration Review, 76*(6), 864-872. doi: http://10.1111/puar.12436

Lee, D., Kim, M., & Lee, J. (2016). Adoption of green electricity policies: Investigating the role of environmental attitudes via big data-driven search-queries. *Energy Policy, 90*, 187-201. doi: http://10.1016/j.enpol.2015.12.021

Longo, J., Kuras, E., Smith, H., Hondula, D. M., & Johnston, E. (2017). Technology use, exposure to natural hazards, and being digitally invisible: implications for policy analytics. *Policy & Internet, 9*(1), 76-108. doi: http://10.1002/poi3.144

Maciejewski, M. (2017). To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences, 83*(1S), 120-135. doi: 10.1177/0020852316640058

Malomo, F., & Sena, V. (2017). Data Intelligence for Local Government? Assessing the Benefits and Barriers to Use of Big Data in the Public Sector. *Policy & Internet, 9*(1), 7-27. doi: 10.1002/poi3.141

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute. Retrieved January 28, 2018, from https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx

Mayer-Schönberger, V., & Cukier, K. (2017). *Big Data: The Essential Guide to Work, Life and Learning in the Age of Insight.* London: John Murray.

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *Public opinion quarterly, 36*(2), 176-187. doi: 10.1086/267990

Mergel, I., Rethemeyer, R. K., & Isett, K. (2016). Big data in public affairs. *Public Administration Review, 76*(6), 928-937. doi: http://10.1111/puar.12625

Panagiotopoulos, P., Bowen, F., & Brooker, P. (2017). The value of social media data: Integrating crowd capabilities in evidence-based policy. *Government Information Quarterly, 34*(4), 601-612. doi: http://10.1016/j.giq.2017.10.009

Poel, M., Schroeder, R., Treperman, J., Rubinstein, M., Meyer, E., Mahieu, B., et al. (2015). *Data for Policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking. Report about the State-of-the-Art.* Technopolis Group. Retrieved April 21, 2018, from https://ofti.org/wp-content/uploads/2015/05/dataforpolicy.pdf

Richards, N. M., & King, J. H. (2014). Big Data Ethics. *Wake Forest Law Review, 49*, 393–432.

Ruggeri, K., Yoon, H., Kácha, O., van der Linden, S., & Muennig, P. (2017). Policy and population behavior in the age of Big Data. *Current Opinion in Behavioral Sciences, 18*, 1-6. doi: 10.1016/j.cobeha.2017.05.010

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration, 80*(1), 1-22. doi: 10.1111/1467-9299.00292

Schintler, L. A., & Kulkarni, R. (2014). Big data for policy analysis: The good, the bad, and the ugly. *Review of Policy Research, 31*(4), 343-348. doi: 10.1111/ropr.12079

Semanjski, I., Bellens, R., Gautama, S., & Witlox, F. (2016). Integrating Big Data into a Sustainable Mobility Policy 2.0 Planning Support System. *Sustainability, 8*(11), 1-19. doi: http://10.3390/su8111142

Severo, M., Feredj, A., & Romele, A. (2016). Data and Public Policy: Can Social Media Offer Alternatives to Official Statistics in Urban Policymaking? *Policy & Internet, 8*(3), 354-372. doi: http://10.1002/poi3.127

Singapore Economic Development Board. (2018). *How Singapore plans to become Asia's big data hub in 2018*. Retrieved May 11, 2018, from https://www.edb.gov.sg/en/news-and-resources/insights/talent/how-singapore-plans-to-become-asias-big-data-hub-in-2018.html

Sivarajah, U., Weerakkody, V., Waller, P., Habin, L., Irani, Z., Choi, Y., et al. (2016). The role of e-participation and open data in evidence-based policy decision making in local government. *Journal of Organizational Computing and Electronic Commerce, 26*(1-2), 64-79. doi: 10.1080/10919392.2015.1125171

Sutton, R. (1999). *The Policy Process: An Overview*. London: Overseas Development Institute.

Tsoukias, A., Montibeller, G., Lucertini, G., & Belton, V. (2013). Policy analytics: an agenda for research and practice. *EURO Journal on Decision Processes, 1*(1-2), 115-134.

UK Department for Business Innovation and Skills. (2013). *Seizing the Data Opportunity: A Strategy for UK Data Capability*. Retrieved March 12, 2018, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/254136/bis-13-1250-strategy-for-uk-data-capability-v4.pdf

United States Executive Office of the President. (2014). *Big data: Seizing opportunities, preserving values*. Retrieved March 12, 2018, from https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. *ECIS 2009 Proceedings*, *9*.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, xiii-xxiii.

Whitman Cobb, W. N. (2015). Trending now: Using big data to examine public opinion of space policy. *Space Policy, 32*, 11-16. doi: http://10.1016/j.spacepol.2015.02.008

Williamson, B. (2016). Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments. *Journal of Education Policy, 31*(2), 123-141. doi: 10.1080/02680939.2015.1035758

Yiu, C. (2012). *The Big Data Opportunity. Making government faster, smarter and more personal*. London: Policy Exchange. Retrieved February 8, 2018, from https://policyexchange.org.uk/publication/the-big-data-opportunity-making-government-faster-smarter-and-more-personal/

Zickuhr, K. (2012). *Three-quarters of smartphone owners use location-based services*. Washington, D.C.: Pew Research Center's Internet & American Life Project. Retrieved March 7, 2018, from http://www.ris.org/uploadi/editor/1344244410PIP_Location_based_services_2012_Report.pdf