Developing a government enterprise architecture framework to support the requirements of big and open linked data with the use of cloud computing

Martin Lněnička and Jitka Komárková

Abstract: *Governmental and local authorities are facing many new information and communication technologies challenges. The amount of data is rapidly increasing. The data sets are published in different formats. New services are based on linking and processing differently structured data from various sources. Users expect openness of public data, fast processing, and intuitive visualisation. The article addresses the challenges and proposes a new government enterprise architecture framework. The following partial architectures are included: big and open linked data storage, processing, and publishing using cloud computing. At first, the key concepts are defined. Next, the basic architectural roles and components are specified. The components result from the decomposition of related frameworks. The main part of the article deals with the detailed proposal of the architecture framework and partial views on architecture (sub-architectures). A methodology, including a proposal of appropriate steps, solutions and responsibilities for them, is described in the next step - after the verification and validation of the new framework with respect to the attributes of quality. The new framework responds to emerging ICT trends in order to evolve government enterprise architecture continually and represent current architectural components and their relationships.*

Keywords: *Government enterprise architecture framework; design science research; big data; open linked data; cloud computing; quality attributes; ATAM; methodology.*

1 Introduction

Traditional Information and Communication Technologies (ICT), infrastructures, architectures and methods for data management and analysis struggle with the rapid growth of data (Tian and Zhao, 2015). The continuous emergence of new technologies and digital channels (Hashem et al., 2015) brings new challenges for enterprises and governments to operate more efficiently and cost-effectively in order to provide better, cheaper and faster services to the public. Big and Open Linked Data (BOLD) concept (Dwivedi et al., 2017; Janssen and Kuk, 2016; Lněnička and Komárková, 2018) has evolved because of production of large amounts of data by the public sector and because of effort to present these data to the public. It represents a new set of steps that are required to create usable insights and value from large amounts of structured and unstructured data coming from many different sources and with differing dynamics (Marton et al., 2013; Chen et al., 2014). This should lead to improving the internal processes of the public sector organizations as well as improving the transparency and accountability processes towards citizens and other stakeholders (Janssen and van den Hoven, 2015).

BOLD result in new opportunities and have the potential to transform government and its interactions with the public (Dwivedi et al., 2017; Janssen and van den Hoven, 2015; Lněnička and Komárková, 2018). The combination of size, multiple sources, and unstructured data then presents the problem of having sufficient computing power to process these data. This problem may be solved using cloud computing (Alshomrani and Qamar, 2013; Hashemi et al., 2013; Sharma, 2016). This led to the intersection of cloud computing and BOLD technologies, mostly because they are complementary technological paradigms with a focus on scalability, agility, on-demand availability, and data-oriented challenges (Hashem et al., 2015, Kobielus, 2014; Lněnička and Komárková, 2018; Sharma, 2016). The practice proved that an organization can reach a better performance and ICT management and services by centralizing available computing resources (Buyya et al., 2013; Jallow et al., 2017; Yu et al., 2014). It started with Service Oriented Architecture (SOA), which provided a conceptual basis for web services development.

It was followed by the rise of computer grids, which migrates to cloud and inter-cloud platforms (Demchenko et al., 2014). According to Sutherland and Chetty (2014), SOA, virtualisation and grids are the major trends in ICT architecture. Cloud computing is not an architectural trend but such a deployment has drawn attention by enterprise architects to ensure that cloud computing aligns with any given enterprise architecture. However, although the applicability and acceptability of cloud computing have been broadly accepted by many organizations in private and public sectors, research on BOLD in the cloud remains in its early stages (Elazhary, 2014; Hashem et al., 2015; Lněnička and Komárková, 2018; Sharma, 2016).

Contemporary organisations are complex systems. Their complexity regularly increases by disrupting, continuing, and recurrent changes in their business. Legal and technological environments take an architectural approach, which can help reduce this complexity (Lagerstrom et al., 2011). This importance was already recognized as a critical factor for the whole-of-government capability and many governments have initiated enterprise architectures programs (Ojo et al., 2012). The relationship between organizations' success with ICT and enterprise architecture management activities was tested by Lagerstrom et al. (2011). They found significant correlations between these variables. Hence, the overall enterprise architecture is critical to be compliant with the constant changing trends in ICT. In similar lines, Janssen et al. (2013) stated that current developments within the government sector focusing on access to open data, shared services, cloud computing and data integration between private and public organizations increase the importance of government enterprise architecture

A choice of an approach for building a system with both cloud computing and BOLD becomes a difficult question for many organizations, not only in the private sector, but also in the public sector. Although the topic of Enterprise Architecture Framework (EAF) provides a rigor background, the understanding of what makes enterprise architecture practice effective in the context of the new trends in ICT is limited (Janssen et al., 2013; Luisi, 2014; Ojo et al., 2012). More precisely, current EAFs are limited in their ability to support revision and extension of existing EAF regarding the requirements resulting from BOLD and cloud computing.

Thus, a new architecture framework is needed to solve this issue in the government enterprise architecture. It should encompass the unique components enabling the use of cloud computing technologies for BOLD and their lifecycle, reuse of ICT assets and the sharing of these technologies across organizations in the public sector. It should ensure cost-effective, faster, and timely delivery of ICT services through a cloud infrastructure, and interoperability principles and guidelines that will assist in storing, processing, and publishing of BOLD. It should facilitate creation of concrete architectures, and increase understanding as an overall picture by containing typical components and partial views on architecture. Last but not least, it should enhance transparency, better communication and coordination, and provide the ability to address critical government-wide issues such as improved security and easier system and network management.

Public sector authorities face the same problem. They have started to produce vast amount of digital data, which are paid by the public budgets. The public sector bodies search for new approaches to better and repeatedly use these data to increase obtained benefits. Since the European Union (EU) requires reuse of data by its regulation (Directive 2003/98/EC, revised by the Directive 2013/37/EU), a new suitable Government Enterprise Architecture Framework (GEAF) is needed in order to efficiently implement new technologies. In addition, despite the achievements in the field of e-government development, the Czech Republic currently faces a number of technical and conceptual shortcomings. As reported in the Strategic Framework for the Development of Public Administration in the Czech Republic for the period 2014 - 2020 (MICR, 2016): "*The implementation of projects in the field of ICT takes place in a largely uncoordinated manner, without ensuring technological compatibility, respecting the four-layer* *architecture and without the free sharing of information about completed projects. The development of e-government is then highly fragmented and its potential is limited (e.g. data cannot be mutually shared and citizens must repeatedly document them).*"

Therefore, the main aim of the article is to propose a new GEAF for storing, processing, and publishing of BOLD using cloud computing including partial views on architecture where the components can cooperate with each other. This solution should be cost-effective to ensure low investments in hardware, software, and data maintenance. For this purpose, a Design Science Research (DSR) approach is followed.

The GEAF is set in the context of the Czech public sector and deals with the requirements of organizations on regional and municipal levels. These are two basic territorial self-government units in the Czech Republic while regions as higher units provide services, resources, and infrastructure to municipalities. This relation is important since it enables to centralize resources and standardize the delivery of services across all municipalities. The case organization in this study is Pardubice region consisting of 451 municipalities.

The organization of the article is as follows. In Section 2, the research methodology is formulated and the research methods are chosen and described. Section 3 introduces the literature review and background on EAFs and commonalities between BOLD and cloud computing for the use in this context. Section 4 presents the results of the decomposition and identification of components for cloud computing and BOLD. Section 5 contains a proposal and verification of the GEAF. Finally, results and discussion are provided in Section 6.

2 Methodology and Methods

The methodology follows the DSR approach described by Hevner et al. (2004) is used. The DSR is a widely applied research approach and is concerned with developing useful artifacts. It is a problem-solving paradigm in which the boundaries of organizational capabilities for creating new artifacts are extended together with the understanding of a problem domain through the building and also application of the design artifact (Hevner et al., 2004). The used methods are derived primarily from the qualitative research, particularly interpretive and critical research (Denzin and Lincoln, 2011). The respective steps are shown in Figure 1.

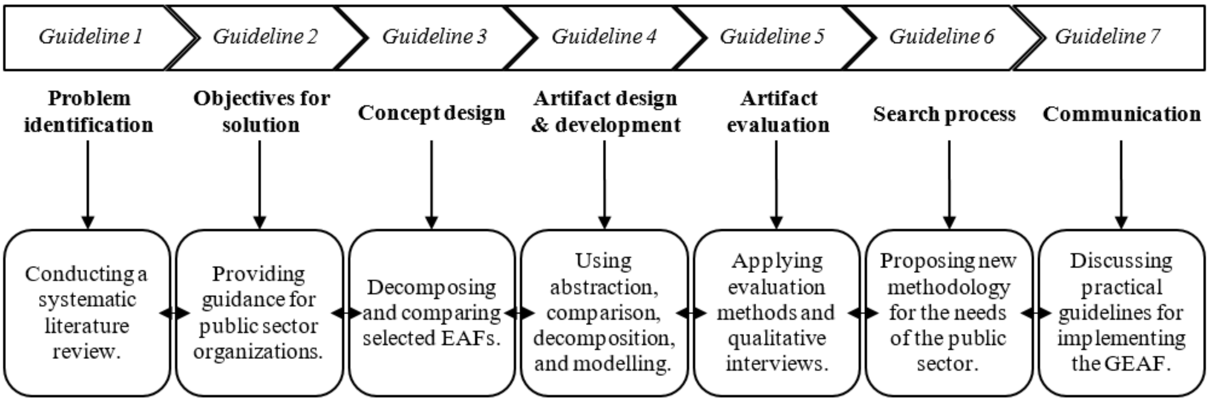

Figure 1: The main steps of conducting the research study

Guideline 1

The first guideline requires the creation of an innovative, purposeful artifact. For this purpose, the systematic literature review as defined by Petticrew and Roberts (2006), i.e.: "*A review that aims to comprehensively identify all relevant studies to answer a particular question, and assesses the validity (or soundness) of each study taking this into account when reaching conclu-*

sions." is used to determine the theoretical background of the research, see Section 3. The approach proposed by Levy and Ellis (2006) is used to conduct the review consisting of the following steps: a) a systematic collection of input papers for the review; b) a systematic processing of relevant papers; c) a systematic synthesis of outputs. The following electronic databases were searched during the collection of input papers: Web of Science, Scopus, and IEEE. The period 2003 – 2018 was covered by the collection of input papers. Both reviewed journals and conference proceedings were included. Various combination of the following terms were used as keywords for search: government enterprise architecture framework, enterprise architecture framework, big data, and open linked data. Table 1 provides the precise list of combinations. Search resulted into 1077 candidate papers and articles.

Table 1: The results of the first step of the systematic literature search

	Web of Science		**Scopus**		**IEEE**	
	proceedings	journals	proceedings	journals	proceedings	journals
"government enterprise architecture framework"	0	0	4	0	1	0
government "enterprise architecture framework"	15	1	37	4	12	0
government enterprise architecture framework	97	24	176	68	95	3
"enterprise architecture framework"	82	24	181	56	60	1
government enterprise architecture framework AND big data	5	1	5	2	4	0
government enterprise architecture framework AND open linked data	1	0	2	0	2	0
"architecture framework" AND big data	13	2	23	8	19	0
"architecture framework" AND open linked data	1	0	0	0	1	0
"software architecture" AND "big data"	47	19	166	32	205	23
"software architecture" AND "open data"	3	2	3	2	11	1
Sum	**264**	**73**	**597**	**172**	**410**	**28**

Based on author keywords, the following number of candidate papers and articles was chosen for the next processing, see Table 2 and Figure 2. This step required manual verification of the authors because search engines in some cases did not include author keywords. The final number of candidate papers and articles is lower because one article is usually described by more included keywords.

Table 2: The results of the second step of the systematic literature search.

Author Keywords	**WoS**	**Scopus**	**IEEE**
Architectural frameworks	0	0	1
Architecture framework	2	7	1
Architecture framework selection scheme	0	1	1
Artifact	1	2	2
Big and open linked data	0	1	0

Author Keywords	**WoS**	**Scopus**	**IEEE**
Big data	7	8	7
Big data analytics	1	2	2
Big data architecture framework	3	3	3
Big data ecosystem	1	1	1
Big data infrastructure	2	2	2
Big data lifecycle management	1	1	1
Big data management	0	1	1
Big data technology	1	1	1
Cloud based big data infrastructure services	1	1	1
Component architecture framework	0	1	0
Conceptual framework	1	1	1
EA framework	2	3	1
EA frameworks	1	1	3
Enterprise architectural framework	1	1	0
Enterprise architecture framework	28	43	17
Enterprise architecture framework adoption	0	0	1
Enterprise architecture framework based on commonality	1	1	0
Enterprise architecture frameworks	1	16	1
Federal enterprise architecture framework	2	3	1
Framework	12	18	13
Frameworks	1	3	1
Frameworks and methodologies for collaborations	0	1	1
Governance enterprise architecture framework	1	1	1
Governance framework	1	1	1
Government enterprise architecture	3	7	0
Governmental enterprise architecture	0	1	0
Linked open data	0	0	1
Open (big) data	1	1	0
Open data	0	0	2
Open government data	0	0	1
Open group architecture framework architecture development method	0	0	1
Public agencies architecture	0	1	0
Software architecture/-s	4	5	13
Software architecture based development	0	1	0
Software architecture development	0	2	2
The enterprise architecture framework of Zachman	0	1	0
The open group architecture framework	16	19	12
Zachman framework	7	13	1
Zachman framework for enterprise architecture	1	1	0
Sum	**104**	**177**	**99**

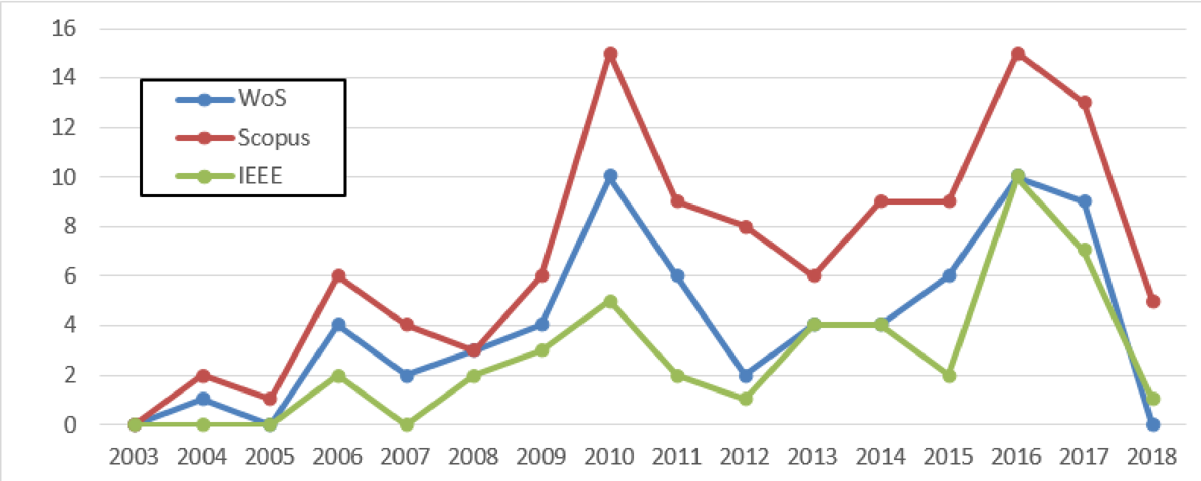

Figure 2: Number of papers and articles in particular databases by year.

Next, duplicate papers were excluded. This step resulted in 143 candidate papers. Authors read abstracts and all author keywords of the candidate papers and excluded not relevant papers. A high number of not relevant papers was caused by utilization of the keywords “framework” and “frameworks”, which included various different frameworks, e.g. e-government, e-governance decentralization, design of national metrology smart system, etc. Next, search engines several times included title pages of conference proceedings, they were excluded too. After the described steps, 93 papers remained. Standards and strategic documents were additionally included because of their importance although they are not indexed by any database. Additionally, several papers and articles were included because of recommendations of experts on various conferences. Authors’ previous works were included as well as far as this article describes continuation of our previous work.

Table 1 and Table 2 reveal that a higher attention is paid to general EAFs then to GEAFs. Contrary, the existence of three GEAFs in the USA confirms the importance of GEAFs (Mary and Rodrigues, 2011; Urbaczewski and Mrdalj, 2006): the Federal Enterprise Architecture Framework (FEAF), the Department of Defense Architecture Framework (DoDAF), and the Treasury Enterprise Architecture Framework (TEAF). Contributions describing existing EAFs and GEAFs were naturally included as the base for the next steps.

The EAFs provide a base for enterprise and software architectures through which an implementation of new technologies should be supported. As it was discussed in the Introduction, contemporary EAFs are limited in their abilities to efficiently support the adoption of emerging technologies and approaches. Authors included articles describing new technologies with a particular focus on big data, open linked data, and cloud computing in the literature review to address these technologies in the proposed framework because existing EAFs/GEAFs do not properly take into account all the relevant components and their relationships.

Contributions proposing and discussing enterprise and software architectures with the focus on the public administration were included in the literature review to deeper address governmental aspects in the newly proposed framework. For example, Ojo et al. (2012) focused on critical success factors and revealed importance of the management commitment and participation of business units as the critical factors.

Analysis of all the included papers and articles represented the next step. It was followed by their decomposition, conceptualisation and linking of the components and dimensions to the newly proposed GEAF.

Guideline 2

The primary objective for the domain of the artifact is to provide guidance to public sector organizations on how to propose, deploy, and implement new technologies related to BOLD with the use of cloud computing in the enterprise architecture context. It is formed by a combination of technology-based artifacts (GEAF conceptualizations and representations, technical capabilities, interfaces, etc.), organization-based artifacts (public sector structures and organisational models, relationships, etc.), and people-based artifacts (involved stakeholders and their training, consensus building, etc.) that address such issues.

The objectives for this solution are drawn from the strategic documents relevant for this domain. The first one is dealing with the cloud infrastructure for services and it is a successor of the concept of regional technological centres. A National Cloud Computing Strategy aims to create a state cloud project, including the data storage and other necessary documents (financial, security, organizational and technical requirements). The importance of BOLD is discussed in an Action Plan of the Czech Republic Open Government Partnership for 2018 to 2020. Together with the Information Concept of the Czech Republic, perceived as a significant element of the National Architectural Plan, and the Strategic Framework for the Development of Public Administration in the Czech Republic for the period 2014 - 2020 they provide a background for designing a new solution.

Guideline 3

The problem space is shaped by decomposing and comparing selected EAFs in order to provide their overview and components based on the key requirements for BOLD and cloud computing, see Section 4.

Guideline 4

The business needs of the public sector are represented by providing cost-effective and transparent services at a good level of quality (MICR, 2016). This also represents a business view. Attention is paid to the performance of public administration, both at the central level and in delegated powers but it will influence territorial self-governing units too. The core strategic objectives are (MICR, 2016):

- Modernization of public administration – namely to optimize the performance.
- Revision and optimization of the performance of public administration in the territory – namely to increase the transparency.
- Completion of the functional framework of e-government – an essential prerequisite and tool for the development and optimization of the functioning of public administration.
- Human resources development – an essential prerequisite and tool for the development and optimization of the functioning of public administration.

In this regard, a new GEAF was developed including the partial views on architecture (process, function, data, software, and hardware sub-architectures) to support phases of the BOLD lifecycle and cover the specific aspects of BOLD and cloud computing in the public sector, see Sections 5.1 and 5.2. For artifact design and development, the methods of analysis, synthesis, abstraction, comparison, decomposition, and modelling are used. The Unified Modeling Language (UML) is used to develop the GEAF. More precisely, the UML component diagram is used to model the logical components that make up the framework.

Guideline 5

Then, chosen architecture evaluation methods and qualitative interviews are used to validate and verify the GEAF, ensuring that it is formally represented, coherent, and internally consistent, see Section 5.3. Namely, the Pattern-Based Architecture Reviews (PBAR) (Harrison and Avgeriou, 2011) and the Architecture Trade-off Analysis Method (ATAM) (Kazman et al., 1999) including its extensions in Bellomo et al. (2015). Because these methods are focused on the architecture evaluation and in this case we do not have a complete architecture to consider because specific functional requirements do not exist, the outputs are generic and not particular for a specific application. In addition, some steps of these methods have to be adapted to this purpose. These adaptations are explained in Section 5.3.

As both these methods, as well as guidelines developed by Hevner et al. (2004), require selecting relevant quality attributes, the quality model of ISO/IEC 25010:2011 (ISO, 2011) is used. The GEAF is evaluated against the following attributes of the current product quality model: functional suitability, reliability, performance efficiency, usability, security, compatibility, maintainability, and portability (ISO, 2011). These quality attributes were also chosen due to their current relevance for several system types in the public sector that are being deployed in cloud environments. The quality attributes are important for all stakeholders as they can assure saving of time and money for them. Interview is an important data-gathering method in qualitative research because of its reliance on direct, usually immediate, interaction between the researcher and participant (Salmons, 2011). Academic people with a good knowledge of IS, system architectures, and quality requirements in the public sector were successfully interviewed to validate and verify the new architecture framework by Svahnberg and Wohlin (2005). Therefore, this method is used in this study to interpret research participants' views of their experience regarding the quality attributes of the new framework.

Guideline 6

Since the artifact should incorporate search process whereby a mechanism enacted to find an effective solution, a new methodology for the needs of the public sector consisting of the description of required steps, related stakeholders and their responsibilities for these steps is proposed, see Section 5.4.

Guideline 7

In addition, the new methodology provides practical guidelines for implementing and using the GEAF and communicates the results of the DSR effectively to the intended audience.

3 Current State of Scientific Knowledge in the Research Area

This section is focused on enterprise architectures, EAFs, BOLD, and cloud computing in government to provide context and background for the next steps.

3.1 Government Enterprise Architecture and its Frameworks

Well-structured and effective processes together with the ICT and enterprise architecture are the essential components of a successful and profitable organization, so it represents a top priority of Information Systems (IS) management (Luisi, 2014). Minoli (2008) defines an enterprise architecture as "*a method and organizing principle that aligns functional business mission with the ICT strategy and execution plans*." ISO/IEC/IEEE 42010:2011 standard defines it as "*fundamental concepts or properties of an enterprise in its environment embodied in its elements, relationships, and in the principles of its design and evolution.*" (IEEE, 2011).

A reference architecture then provides "*an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions.*" It generally serves as a reference foundation for solution architectures, and may be used for comparison and alignment purposes (NIST, 2015). The enterprise architecture is usually divided into several sub-architectures and typically described using views that represents all combined components and their relationships throughout all different layers (Luisi, 2014; Minoli, 2008). For example, the Multidimensional Management and Development of Information Systems (MMDIS) (Bruckner et al., 2012; Nálevka, 2009) deals with these partial views on architecture (dimensions): process and function; application; data; technology; security aspects; organisational and legislative aspects; and labour, social and ethical aspects.

An EAF provides "*logical mechanisms for developing architectures in a uniform and consistent manner*" (Minoli, 2008). The framework also provides a tool for accessing, organizing, and displaying that information. Comprehensive EAFs such as The Open Group Architecture Framework (TOGAF) or the Zachman Framework have been developed to address this issue. There can be found a lot of definitions of these frameworks, see Bruckner et al. (2012), Luisi (2014) or Minoli (2008). These frameworks differ by the stakeholders they address and the issues that concern their focus (Luisi, 2014; Minoli, 2008). Thus, choosing an appropriate EAF including a structure of what the enterprise architecture should contain and how to create it requires a comprehensive analysis of key requirements (Mohamed et al., 2012).

The government enterprise architecture is considered to be an efficient tool to overcome the challenges and problems related to the development of e-government systems including interoperability, integration, complexity, and lack of data standards (Gong and Janssen, 2017; Guijarro, 2007; Janssen et al., 2013; Ojo et al., 2012). Research on this topic includes a conceptual model illustrating the relationships between government enterprise architecture concepts, its use, benefits and public value drivers (Janssen et al., 2013), the interoperability frameworks and the enterprise architectures as tools to achieving the interoperability in the public sector (Guijarro, 2007), criteria to evaluate selected EAFs (Mohamed et al., 2012) or key maturity factors affecting the government enterprise architecture practice (Ojo et al., 2012). Birkmeier et al. (2013) focused on the alignment of business and IT architectures in the German federal government. Gill et al. (2014) proposed the use of an agile EAF to developing and implementing the adaptive cloud technology-enabled enterprise architecture in the Australian government context. However, none of these frameworks provide all of the functionality required for BOLD and cloud computing components in the GEAF.

3.2 Commonalities between BOLD and Cloud Computing in the Government Enterprise Architecture

Cloud computing is "*a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be provisioned and released with minimal management effort or service provider interaction.*" (NIST, 2011). Enabled by advances in fast and low-cost networks, commoditized faster hardware, high performance virtualization technologies, and maturing interactive web technologies, cloud computing has become a powerful architecture to perform complex computing problems for tackling large amounts of data (Buyya et al., 2013, Shroff, 2010). Although, the concept of cloud computing is mainly available in three varieties of services, namely, Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS), but in this data science age, should be equally expandable to Database-as-a-Service (DBaaS) (Sharma, 2016). Several researchers have also attempted to define cloud frameworks by adding other essential components, such as management, quality, security, communication, storage, information or process as additional services extending the basic structure consisting of SaaS, PaaS and IaaS (Mahmood, 2011). In this regard, cloud computing can provide a useful extension to existing architectures, on demand and without the additional capital investment (Mahmood, 2011). According to Ramachandran (2011), the cloud architecture, its layers and its composition of components and services have to be designed for scalability and re-configurability.

A cloud implementation perspective of the existing EAFs and their usability was analysed by Zota and Fratila (2013). They claimed that EAFs using cloud computing have to be broad enough to use them regardless of the domain or the regulations implied, but specific enough to include all the deliverables needed in most of the existing business domains. Tang et al. (2010) further proposed a way to extend the enterprise service-oriented architecture style to a new hybrid architectural style based on cloud computing. In addition, many cloud vendors such as Amazon, Google, IBM, Oracle and Microsoft also proposed their own architectural styles to combine enterprise architectures and the power of cloud computing powers (Mahmood, 2011; Sharma, 2016; Shroff, 2010). A survey conducted by Sutherland and Chetty (2014) showed that up to 75% of the respondents see the need to incorporate cloud services into their respective enterprise architecture. Therefore, clear understanding of requirements and their relationships with respect to architectural choices is a critical for any enterprise system using the cloud services (Rimal, 2011).

Big data are characterized by very large data volume, velocity of data flows, and highly variety in data types and sources. Open data are about standards on how to make data machine-readable and linkable (Chen et al., 2014; Janssen and van den Hoven, 2015; Marton et al., 2013). The recent hardware and software advances and the shift towards openness of the public sector (Buchholtz et al., 2014) have enabled the emergence of BOLD (Dwivedi et al., 2017). The combination of these data with other data, e.g. organization's own data, could enable creating models and services that were previously hard to develop (Janssen and Kuk, 2016; Marton et al., 2013). The opportunities to collect and integrate these data in the enterprise data architecture offers potential for organizations to improve on the provision of quality services through better management of data (Dwivedi et al., 2017; Jallow et al., 2017; Kamal et al., 2013) and increase their competitiveness (Buchholtz et al., 2014; Luisi, 2014; Marton et al., 2013). In this regard, Gong and Janssen (2017) explored the role of the enterprise architecture as an instrument to integrate big data in the existing business processes and ICT-landscape. Their findings suggested that it provides an incremental approach for adapting the infrastructure step-by-step, before the benefits of big data can be gained.

BOLD are classified into different categories to better understand their characteristics. Figure 3 shows the numerous characteristics defining BOLD. The classification is based on five aspects: data sources, content format, data stores, data staging, and data processing. This classification was firstly presented by Hashem et al. (2015). However, due to the rapid growth of open and linked data, their mashups, integration, and new data stores, it was updated by the authors. Data stores classification is based on the National Institute of Standards and Technology (NIST) Big Data Interoperability Framework (NIST, 2015).

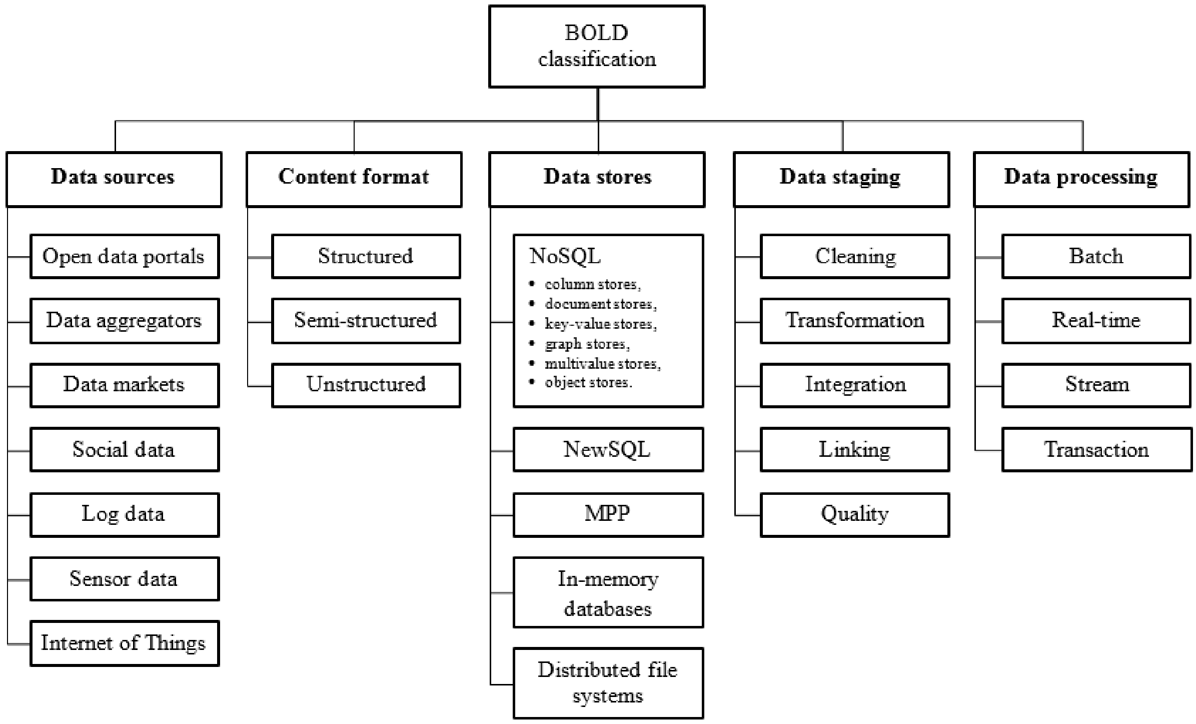

Figure 3: The classification of BOLD

Cloud technologies support data-intensive computing by providing a large amount of compute instances on demand and a storage system optimized for keeping large blobs of data and other distributed data store architectures (Bahrami and Singhal, 2015; Buyya et al., 2013; Kobielus, 2014; Rimal, 2011). Generally, cloud computing technologies simplify the building of related infrastructure and provision it on-demand (Demchenko et al., 2014). Considering that the public sector organizations of the same type or on the same level have very similar needs of ICT resources and data processing tools, cloud computing technologies can provide many benefits to governments (Alshomrani and Qamar, 2013; Hashemi et al., 2013).

According to Hashem et al. (2015) the use of cloud computing in BOLD may be described as follows: "*Large data sources from the cloud and Web are stored in a distributed fault-tolerant database and processed through a programing model for large data sets with a parallel distributed algorithm in a cluster*." BOLD provide stakeholders the ability to use commodity computing to process distributed queries across multiple data sets and return results in a timely manner and cost-effective way. Thus, cloud computing not only provides facilities for the computation and processing of these data but also serves as a service model (Hashem et al., 2015). Sá et al. (2015) then stated that the use of BOLD in a cloud environment requires "*rethinking of data architectures, new architectures must take into account characteristics such as robustness, availability and easy access of data for analysis, allowing data structured, semi-structured, and unstructured, while preserving crucial aspects such as security and data integrity*." Bahrami and Singhal (2015) addressed the role of cloud computing architecture as a solution for important issues related with BOLD.

Kobielus (2014) claimed that the practical overlap between the BOLD and cloud paradigms is so extensive that cloud-based BOLD can be done with an existing Apache Hadoop on premise, NoSQL, or enterprise data warehousing environment. Liu (2013) gave an overview of computing infrastructure for big data processing, focusing on architectural, storage, and networking challenges of supporting these data analysis. Further, many SaaS platforms have been developed specifically for processing BOLD such as sequence analysis, alignment and mapping (Elazhary, 2014). Sharma (2016) investigated the growing role of cloud computing in the big data ecosystem and proposed a novel classification of the cloud-assisted NoSQL big data models. The author found out that each data model is developed with different storage technology and computing mechanism underneath. Di Martino et al. (2014) classified services offered by cloud providers with respect to the functionalities related to big data manipulation and provided a taxonomy of cloud platforms offering big data oriented services. A review of the rise of big data in cloud computing can be also found in Hashem et al. (2015).

On the other side, Miller (2013) argued that too many emphasis on BOLD analytical methods and incorrect use of the methods may produce wrong and costly decisions. The pressure for organizations to quickly adopt and implement these technologies also involves unexpected risks and consequences (Elazhary, 2014; Miller, 2013). There are also many other challenges such as security and privacy issues, transfer issues, the performance in case of extremely large data sets, the complicated pricing models, Qualify of Service (QoS) assurance, quality dimensions, etc. (Bahrami and Singhal, 2015; Elazhary, 2014; Hashem et al., 2015). Janssen and van den Hoven (2015) stated that information architectures should reflect the societal value of privacy protection and mechanisms for access control, security, privacy protection and the inability to combine information should be in place.

Luisi (2014) recommended organizing data in the enterprise architecture into a lifecycle. Generally, a lifecycle of big data can be divided in four phases (Chen et al., 2014): data generation, data acquisition, data storage, and data analysis. Similarly, Di Martino et al. (2014) defined these phases of the big data value chain: data collection, data curation, data integration and aggregation, data storage, data analysis and interpretation. The big data lifecycle management model that reflects new challenges and specifics in the big data management was introduced by Demchenko et al. (2013). Zuiderwijk et al. (2014) emphasized data publication phase in the context of open government.

Therefore, the BOLD lifecycle newly proposed by the authors and described in the Figure 4 provides an example of activities that are performed when working with BOLD. It also provides more detailed components that should be integrated in the architecture framework.

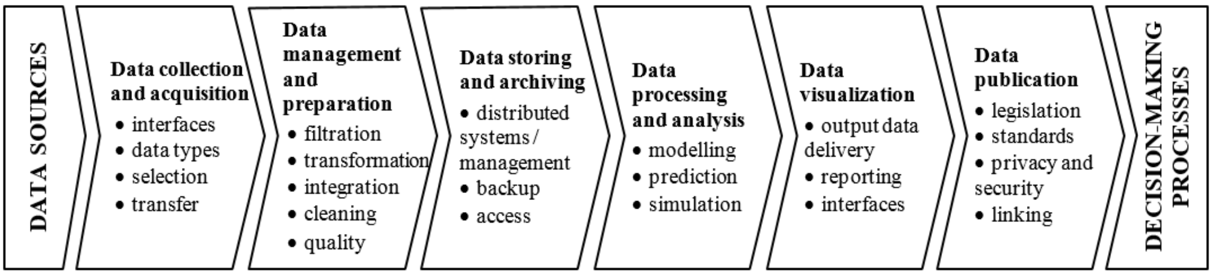

Figure 4: The phases and components of the BOLD lifecycle

4 Architectural Components for Cloud Computing and BOLD

Currently, there are numerous definitions of architecture, emphasising its various aspects and components such as ISO/IEC/IEEE 42010:2011 (IEEE, 2011). An architectural component describes the fundamental aspects of service deployment and service orchestration (NIST, 2011; NIST, 2015). In the public sector context a component is meant as "*a functional sub-system, which consists of a set of steps and activities using platforms, tools, and services defined by the regulations and other documents, usually covering one or more phases of the BOLD lifecycle.*" The specifics of the public sector ensure that each relevant component can be used by agencies and institutions on the same administrative level and with the same scope. Since the functional boundaries of components are shaped by legislation and policy, it is easier to reflect the requirements in the EAF. This issue was explored and described by Kamal et al. (2013) for the context of delivering integrated services in the local government authorities. Although some of the components may seem not to be related to the public sector, they are required by BOLD analytics and cloud computing.

In order to understand the relationships and the interactions between the key components for cloud computing and BOLD, several architecture and reference frameworks are decomposed – see Table 3 for the list. The decomposition follows two approaches: multidimensionality based on the division into layers, and the BOLD lifecycle. The architectural components of cloud computing are classified according to the requirements of cloud providers and cloud consumers into layers. The architectural components of BOLD are classified according to the phases of these data. Each component encompasses a portion of the system's functionality and may consist of one or more sub-components. The allocation of this functionality is a crucial aspect of the decomposition onto a set of components.

The NIST and its reference architectures (NIST, 2011; NIST, 2015) are consensus-based rather than empirical ones. They are too general to be used in the context of enterprise architecture. The other category is comprised of commercial companies and platform-specific vendors' proposals. Their main aim is to build a wider ecosystem around their own products. Although these platforms may cause a vendor lock-in situation, important information about orchestration of the components can be gained from these frameworks. However, there is no work dealing with both BOLD and cloud computing requirements in a single model although there are frameworks that tried to combine cloud computing and big data. Finally, it should be noted that not all the papers and articles identified through the systematic literature review are included in Table 3. It is due to the fact that several components are overlapping.

Table 3: Comparison of the selected architecture and reference frameworks

Focus	Author / source	Title	Published	Category
Cloud computing	Al-Hmoud	Developing a new enterprise architecture framework based on information technology service delivery	2011	Scientific research
	Bernal et al.	Enterprise architecture framework oriented to cloud computing services	2016	Scientific research
	NIST	Cloud computing reference architecture	2011	(Inter)national organizations
	IBM	Cloud computing reference architecture	2010, 2011, 2013, 2014	Commercial companies
	ORACLE	Cloud reference architecture	2011, 2012	Commercial companies
	Demchenko et al.	Intercloud architecture framework	2013	Scientific research
	Ramachandran	Component-based development for cloud computing architectures	2011	Scientific research
	Rimal et al.	Architectural requirements for cloud computing systems	2011	Scientific research
Big data	NIST	Big data interoperability framework: volume 5, volume 6	2015	(Inter)national organizations
	IBM	Big data and analytics reference architecture	2014	Commercial companies
	ORACLE	Big data and analytics reference architecture	2013	Commercial companies
	Demchenko et al.	Big data architecture framework	2014	Scientific research
	Gong and Janssen	Enterprise architectures for supporting the adoption of big data	2017	Scientific research

	Maier	Towards a big data reference architecture	2013	Scientific research
	Pääkkönen a Pakkala	Reference architecture and classification of technologies, products and services for big data systems	2015	Scientific research
Open linked data	Lee	A life-cycle workflow architecture for linked data	2017	Scientific research
	ODCA	ODCA master usage model: Information as a service	2013	(Inter)national organizations
	Oliveira et al.	A recommendation approach for consuming linked open data	2017	Scientific research
	Zuiderwijk et al.	Innovation with open data: Essential elements of open data ecosystems	2014	Scientific research
	Van Schalkwyk et al.	Viscous open data: The roles of intermediaries in an open data ecosystem	2016	Scientific research
Big data, cloud computing	Liu et al.	How big data ecosystem changes cloud services	2015	Scientific research
	Lněnička et al.	Components of big data analytics for strategic management of enterprise architecture	2017	Scientific research
	Sá et al.	Big data in cloud: A data architecture	2015	Scientific research
	Zimmermann et al.	Towards service-oriented enterprise architectures for big data applications in the cloud	2013	Scientific research
	Schmidt and Möhring	Strategic alignment of cloud-based architectures for big data	2013	Scientific research
	Tian a Zhao	Big data technologies and cloud computing	2015	Scientific research

For the purpose of the new GEAF, the related stakeholders are described to easily define their roles and responsibilities assigned to them. Each stakeholder is an entity (a person or an organization) that participates in a transaction or process and/or performs tasks related to cloud computing and/or BOLD. Stakeholders in these roles interact with each other to ensure that the goals of BOLD analytics will succeed. Each stakeholder may perform one or more roles. These roles of stakeholders focusing on the public sector can be seen in Table 4, including examples of literature sources.

Table 4: Roles of stakeholders for cloud computing and BOLD

Stakeholders for cloud computing		
Role	**Description**	**Literature sources**
Cloud consumer	Maintaining a business relationship with, and uses services (resources) from provider. There are classifications by organization type for easier selection of services.	Demchenko et al. (2013); NIST (2011); Ramachandran (2011); Rimal (2011)
Cloud provider	Ensuring that a service is available to stakeholders. Private or hybrid clouds are most suitable for the public sector agencies and institutions.	Demchenko et al. (2013); NIST (2011); Ramachandran (2011); Rimal (2011)
Cloud carrier	Providing connectivity and transport of services between stakeholders at different levels.	NIST (2011)

Cloud auditor	Assessment of cloud services, cloud performance and their security. This is a typical role for the public sector.	NIST (2011)
Cloud broker	Managing the performance, use and delivery of services, and negotiation of relationships between providers and consumers.	Demchenko et al. (2013); NIST (2011); Rimal (2011)
Cloud service creator	Using service creation tools to offer new services with new technologies or delivery models for the public sector.	Demchenko et al. (2013); Lněnička and Komárková (2015a)
Stakeholders for BOLD		
Role	**Description**	**Literature sources**
System orchestrator	Defining the required data application activities in the context of organization's goals, policy and resources, and monitoring them (performance indicators).	NIST (2015); Zuiderwijk et al. (2014)
Data provider	Creating, maintaining and distribution of data and their sources. Providing various interfaces and other services.	NIST (2015); Van Schalkwyk et al. (2016); Zuiderwijk et al. (2014)
Data application provider	Using applications and other platforms to execute tasks and activities required by the BOLD lifecycle.	Lněnička and Komárková (2015a); NIST (2015)
Data framework provider	Establishing a computing framework in which certain applications are executed. Offering resources for the application provider.	Lněnička and Komárková (2015a); NIST (2015)
Data consumer (user)	Using the results provided by the data application provider. They can also search in data sets, discuss and evaluate them, or request and suggest data to be opened.	NIST (2015); Van Schalkwyk et al. (2016); Zuiderwijk et al. (2014)
Data carrier	Guiding and controlling the data distribution between providers and consumers in the BOLD ecosystem.	Van Schalkwyk et al. (2016)

The most important characteristic of these technologies is the focus on openness, transparency and accountability of public sector agencies and institutions. While the private sector organizations primarily utilizes BOLD to discover hidden insights in their data and gain a competitive advantage, the public sector requires components that improve efficiency, effectiveness, and performance regarding the resources utilization and cost transparency. Furthermore, cloud computing delivery models are mostly deployed as SaaS in the public sector and provide access to applications that all the respective agencies and institutions use. A high demand is there for stream and real-time data transmission and processing. More emphasis is also placed on security and components needed to develop critical infrastructures for providing ICT services across the public sector. Regarding the platforms, tools, and services required to perform tasks defined by the BOLD lifecycle, components for their support in the GEAF have to be taken into account. Finally, while interactions between private sector organizations are mostly competitive, public sector agencies and institutions should cooperate and work together in order to provide quality public services. These characteristics are unique for these technologies in the public sector and are different or missing in past research.

The architectural components for a cloud consumer, identified from the literature review, are: cloud deployment models; cloud services delivery models; cloud service integration tools; data sources management; cloud economy and costs; and consumer in-house IT. These components then consist of sub-components. The support components for both consumer and provider are: security and privacy; architecture management and control; operations support in architecture; architecture performance and elasticity; and coordination and cooperation of components. The other identified architectural components for a cloud provider are distinguished into six layers:

cloud services delivery models; access / service delivery at different levels and access infrastructure; service search and management; service orchestration and cloud management; resources abstraction (virtualization); physical resources for cloud services and equipment. As an example of related sub-components for the cloud provider on the second view level, the service orchestration and cloud management component is chosen. These sub-components are: cloud management and organization's operations; services, abstraction and control; organization services and/or third party services; and performance and quality indicators.

The identified architectural components for BOLD on the first view level are: data sources and their acquisition; data manipulation, pre-processing and management; stream and real-time analytical data processing; data analysis, analytical sources and tools; analytics and database interfaces; visualization, reporting and user interfaces; data usage, analytical views and decision-making; data publication and consumption; and data users. Further, there are components focusing on the infrastructure (mostly cloud), however, also grids or clusters may be used. These components are: data storage, specialized databases and repositories; and highly scalable infrastructure, resources and platforms. The support components are: design, develop and deploy tools; jobs and models identification and specification; data security and privacy management; regulations, policies and risks management; and system and data sources management.

As an example of related sub-components on the second view level, the data publication and consumption component is chosen. These are: data request; data discussion; evaluation of data set (rating); data publication on the web/portal; registration/log in; data set visualization; search data sets; and data licenses search. However, boundaries between these components are not sharply defined as in the case of cloud computing, which is obvious especially on the second view level. Therefore, in the new GEAF, these components are clearly defined for the needs of the public sector.

5 Proposal and Verification of the Government Enterprise Architecture Framework

Although there were identified a lot of common features and functionalities between compared frameworks, some key features were omitted or cannot be used in the public sector. This may result partly from a slightly different focus of these frameworks or developments in ICT are already offering new or modified platforms and tools that can be used in the new GEAF. The main limitations and drawbacks found were: some BOLD phases are omitted, especially data sources, data collection and acquisition, and data publication; the missing link between relational databases and NoSQL databases; missing or vague classification of data storage technologies and their functionalities; inadequate definition of requirements for the phase of data management and preparation, which is essential for data analysis, but some activities are redundantly moved to the phase of data analysis, e.g. data integration and linking; the missing link to open and linked data, their acquisition and publication in the context of related licenses; missing classification of related platforms, tools and services, in particular in connection to individual components or phases of the data lifecycle; more detailed definition and description of the external environment influences on the framework for the needs of the public sector; etc.

5.1 Proposal of the Government Enterprise Architecture Framework

The need for GEAF derives from practices of modern ICT that are implemented and used in the public sector agencies and institutions, but without an underlying concept relating them. Although sufficient quality ICT resources are available, these are not fully utilized to provide delivery of quality and flexible services to citizens and other stakeholders. This pressure for innovation and quality in providing services is also forming the GEAF. Its evolution towards the current form can be described as a shift from back-end to front-end services. The MMDIS was initially focused on information system development and data management. With increasing access to Internet and data services, the public sector had to change the way it worked in order to compete in those circumstances. More dimensions and stakeholders had to be taken into account to develop a product or service. Building of technological centres of regional authorities in the Czech Republic in the beginning of 2010s represented one of the attempts how to address changing situation and support e-government development. They were funded by the European Regional Development Fund (ERDF) within Integrated Operational Program. The centres (focused on infrastructure services) should host, among others, various digital agendas for municipalities within particular regions. Nowadays, the authorities face the problem with next maintenance and development of the centres because they cannot apply for subsequent ERDF funding. In this situation, a framework that enables to systematically relate in a long time period relevant architectural components needed to centralize available computing resources and manage data in order to provide better, cheaper and faster services to the public is required.

A system view on the general concept of the GEAF in the public sector, including the impacts of external conditions and processes on the system, is shown in Figure 5. It also highlights a sub-system in which the new GEAF is to be delivered, i.e. organizations on the regional level should provide services to organizations on the municipal level, respecting the relationship and responsibilities between a cloud provider and a cloud consumer. BOLD flow through the system and respective components and roles are needed to deal with them in order to meet public sector goals.

The new GEAF was proposed based on the decomposition of published architecture and reference frameworks and their components, see Table 3. The theoretical background for the decomposition process, including instructions for reflecting requirements in components properly, is derived from the ISO/IEC/IEEE 42010:2011 standard (IEEE, 2011), the MMDIS (Bruckner et al., 2012) methodology, and the TOGAF (The Open Group, 2009). The design of the structure and interrelationships between the components of the GEAF follows two approaches: multidimensionality based on the division into layers for cloud computing components, and the phases of the BOLD lifecycle for BOLD components. It was developed as a stakeholder/role based model that lays out the key components of cloud computing and BOLD. It was proposed for the use in the public sector focusing on the storing, processing, and publishing of BOLD using cloud computing. Finally, it should be noted that the proposal of the new GEAF in this article does not include developing multiple domains / layers covering from business, software application to infrastructure, because these are already covered by the MMDIS methodology, see Section 5.4.

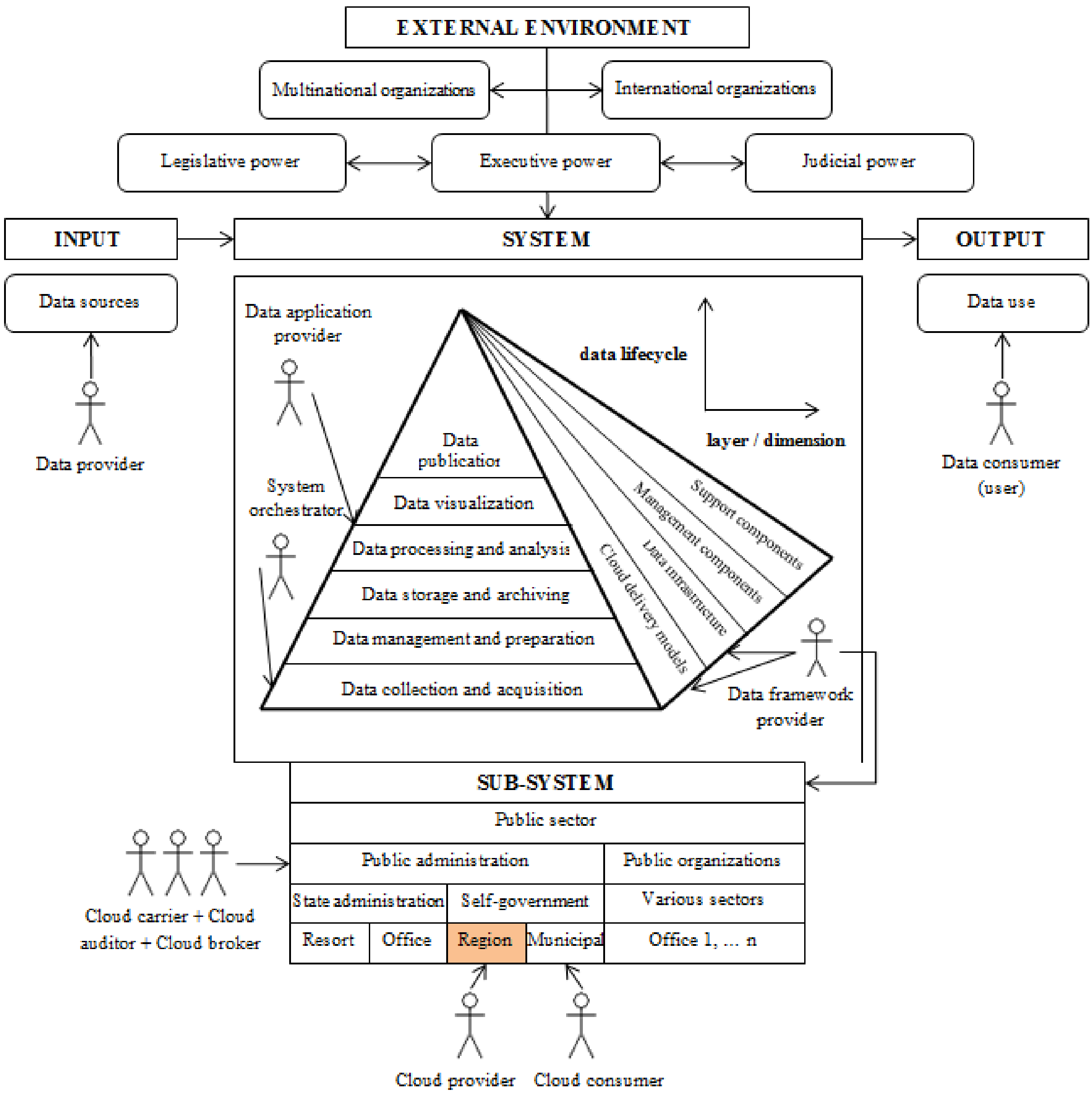

Figure 5: A system view on the general concept of the GEAF in the public sector

The newly proposed GEAF contains two views and descriptions that are the basis for discussing the characteristics and uses for cloud computing and BOLD. The first view level can be seen in Figure 6. There is also the second view level of the framework in which all the components from the first level are further decomposed into activities and functions that need to be addressed in the GEAF. As an example, the second view level for the data availability and accessibility component consists of the technical equipment of the user; work mode (regime) of the user; knowledge and skills of the user; and documentation and manuals.

The multidimensional approach consists of several different dimensions influencing the success of a IS/ICT in the organization. Vertical layers are visualised in Figure 6 in a form of particular components of the framework. Each dimension is solved separately at first. These solutions are later integrated into the overall methodology with the use of requirements defined in the TOGAF, i.e. Architecture Development Method (ADM) and Model Driven Architecture (MDA), and the MMDIS, i.e. methodological and organizational dimension. The models of partial views on architecture focus on organising architectural information according to the stakeholders' needs and concerns, captured by views with the appropriate levels of abstraction. Resources for the data infrastructure are allocated on the regional level, i.e., cloud provider covers (offers cloud services) all the municipalities in the region.

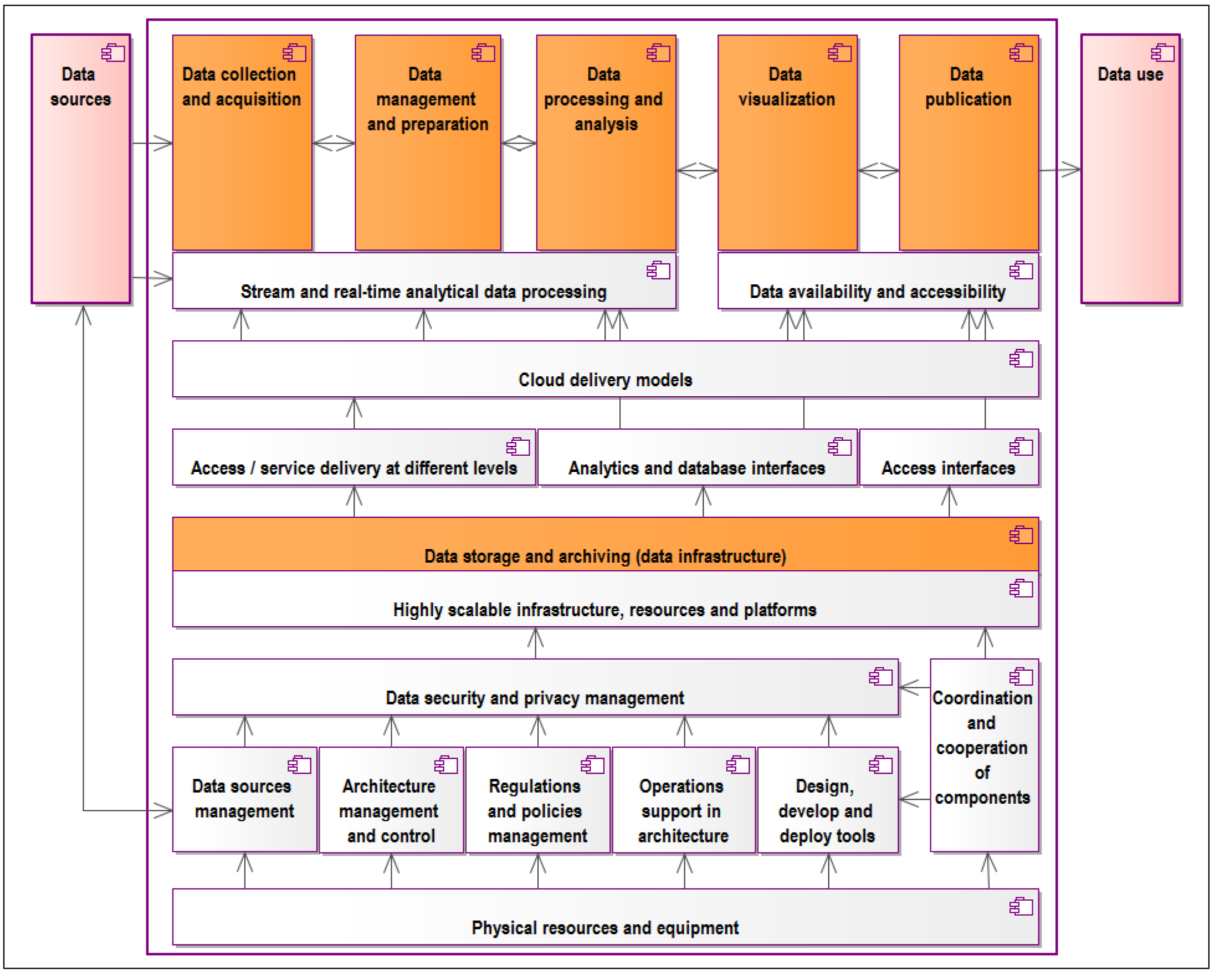

Figure 6: The GEAF at the first view level – visualisation of its components

One of the important characteristics of the framework is the ability to import and use data from a variety of data sources. Consequently, data from different sources may have different security and privacy considerations. Therefore, the connection between these sources and data sources management is critical. The key components on this level are defined according to the BOLD lifecycle, while it is important to distinguish between batch processing and stream and real-time processing. Cloud delivery models consist of deployment (private or hybrid cloud) and service models (IaaS, PaaS and SaaS) with QoS guarantee. A data storing and archiving component covers (supports) all the other phases as their data infrastructure. With the introduction of new storage paradigms analytics techniques should be modified for different types of data access through different interfaces. Also, various users have different demands, access rights and devices, which have to be solved in the new framework, i.e. as a data availability and accessibility component. Regarding the structure and goals of the public sector, various support components and systems / resources for the coordination and cooperation of components are required. These support components focus on the users of the ICT services and ensure that they have access to them. The most important component is dealing with data security and privacy management that determines the implementation of other components. It is followed by a data sources management component, i.e. defined rules on what to process and publish. Finally, a data use component includes end users or other systems that require access to data.

5.2 Key Architecture Views

Within this article, the key architecture views are: a) process; b) function; c) data; d) software; and e) hardware. The last one is security, which covers all these views focusing on the needs of the public sector. The business view is defined by MICR (2016).

As a part of the **process view** the framework proposes to focus on the support of related processes which may be modelled in a hierarchical manner. A process architecture is the means by which the organization establishes a set of rules, principles, guidelines and models for the implementation of the process management across the organization. For this purpose, a definition of external events with respect to the related roles for cloud computing and BOLD was done. Then, management and support processes for the key operational process of the BOLD storing, processing, and publishing were identified and described. Finally, related standards and norms for the public sector such as Common Assessment Framework (CAF) or European Foundation for Quality Management (EFQM) have to be involved in this view.

A **function architecture** is designed based on the requirement analysis. A function view defines hierarchical decomposition of functions of the IS/ICT from the high-level functions to elementary functions and to define inputs, outputs and, control data for each function. Therefore, an overview of these requirements classified according to the BOLD lifecycle phases have to be proposed. A data view on architecture provides an understanding of what data exist and how are transferred throughout the organization and related systems. Therefore, the main goal is to describe the data flows and sources in the new architecture defined by the architecture framework.

A **software view** on architecture focuses on the related platforms, tools, and services, and it is distinguished into three categories: 1) data processing; 2) data storage; and 3) data publication. The platforms, tools, and services can be found in Lněnička and Komárková (2015a). A data processing is the most important phase in the lifecycle of BOLD with the purpose of extracting useful values, providing suggestions or decisions. With the emergence of the MapReduce framework, rise of the open-source Apache Hadoop and related platforms together with the commoditization of large-scale and reliable clusters, the new opportunities to process and analyse these data arise. Thus, the classification of these platforms was created and a multi-criteria decision-making model for the selection of the most suitable platform was proposed. The defined criteria are under three categories based on their feasibility and integrability: technical (hardware and resources configuration requirements) perspective; social (people and their knowledge and skills) perspective; and cost and policy perspective. Based on the literature review of the possible advantages and disadvantages of various data analytics platforms, eleven tools were selected as alternatives to be compared. Further, three use cases were designed to meet the various users' needs: public sector organization; scientist or advanced user from the public sector; and medium-sized business. This model can be found in Lněnička (2015).

A **data storage** requires new types of distributed file systems, i.e. NoSQL, NewSQL, Massively Parallel Processing (MPP) or in-memory database architecture to adapt to large amounts of data and changing structures. However, it is unrealistic to assume that these new types of data stores are a replacement for traditional relational databases such as SQL. Therefore, the new architecture contains the interface for the integration of relational databases and NoSQL databases. Another important part of this phase is a programming language that helps to simplify the programming model and to provide a higher level language, in which are data operations performed. Last, there are also cloud storages, which can be used to store, synchronize or share data. A data publication is focused on open government data and their publishing on open data portals. Thus, the most important portals have to be described and the most suitable platform (open data management system) has to be chosen (Máchová and Lněnička, 2017).

A **hardware view** on architecture refers to the identification of a system's physical components and their interrelationships. It is defined by the physical resources and equipment component in the new architecture framework and is used with the IaaS provisions in the private cloud hosted by the cloud provider. An example focused on the deployment of the infrastructure for the processing of BOLD in the public sector can be found in Lněnička and Komárková (2015b).

Security, privacy and protecting information are integral to government operations, and are part of the architecture. Governments must protect information against unauthorized access, denial of service, and both intentional and accidental unauthorised modification. Data security and audit controls have to be maintained across the data lifecycle management. Therefore, data are available to users who require the information as part of their roles. Data confidentiality has to be maintained for sensitive and critical data that can be used to distinguish or trace an individual's identity through anonymization. Security is a cross-cutting aspect of the architecture that spans across all layers of the GEAF, ranging from physical security to application security (NIST, 2011; NIST, 2015). Therefore, security in the architecture concerns is not solely under the responsibility of the cloud providers, but also cloud consumers and other relevant stakeholders.

5.3 Validation and Verification of the Proposed Framework

Architecture evaluation methods are applied in order to verify whether the architecture is suitable for a given system (its type, technology, goals, environment, etc.) and does not contain flaws that may affect the achievability of the quality requirements. Two basic paradigms of architectural evaluation have emerged from the research carried out to date (Kazman et al., 2005; Zalewski, 2013):

- Searching for architectural flaws included in the architecture.
- Assessing architecture against a set of relevant quality requirements.

Each of these paradigms then determines the proposal of the concrete method. Each method utilizes various evaluation techniques such as questioning (scenarios, questionnaires and checklists), measurements using metrics, simulations, costs evaluation, prototypes, etc. (Babar et al., 2004; Zalewski, 2013). However, most of the existing methods use more than one technique (Zalewski, 2013). A prerequisite for the use of these methods is the existence of the architecture description (Kazman et al., 2005). There are a lot of methods and also a lot of comparisons and criteria to select the right one. More can be found in Babar et al. (2004), Kazman et al. (2005) or Zalewski, (2013).

All these methods cover all relevant steps for the validation, i.e. the process of ensuring that an architecture and its components are correct, complete, consistent and useful for the purpose intended, and the verification, i.e. the process of ensuring that the architecture models are correctly implemented based on the defined standards and practices for these models (Bruckner et al., 2012; Minoli, 2008). We followed both these processes.

Methods for both these paradigms are used to validate and verify the framework. The first one is the PBAR (Harrison and Avgeriou, 2011) and the second one is the ATAM (Kazman et al., 1999). Regarding design evaluation methods of the DSR approach listed in Hevner et al. (2004), these methods can be classified as experimental and descriptive. More precisely, it is a controlled experiment in which artifacts are studied in controlled environment for qualities. Specific adaptations of the methods used are related to the fact that there is no complete architecture to consider because specific functional requirements do not exist. The outputs are thus limited to the list of issues related to the attributes of quality in the case of the PBAR method and the utility tree in the case of the ATAM method.

The first phase consisted of three cycles of interviews with stakeholders, one of them was realized electronically. The related stakeholders were selected to correspond to the roles for BOLD, i.e. system orchestrator, data provider, data application provider, data framework provider, and data consumer (user). The role of data carrier was not involved because according to Babar and Kitchenham (2007), the ideal group size consists of five stakeholders. One person (the researcher) leads the interview, asking all the major questions and seeking to elicit maximum information from the stakeholders. In total, seven people participated in interviews. The first author directed the interviews. Five experts from the University of Pardubice and one expert from practice were interviewed. Participants from the university are experts in the following fields: public information systems analysis and design, requirements engineering, databases, business process modelling, information systems usability, computer networks, data security, programming, e-shops design, and web-based geographic information systems design. The expert from the practice worked as the head of the Department of information technology of Regional authority of Pardubice region with 15 years of practice at this position, namely with development and implementation of IS in the public sector. Each participant from the university represented one of the role mentioned above. Participant from the Regional authority of Pardubice region represented role of the system orchestrator.

The search for architectural flaws paradigm, i.e. the use the PBAR method, does not demand the requirements to be elicited and prioritised before the actual analysis of the architecture. Therefore, such activities are absent in its workflow. The evaluation is done by walking through an appropriately partial architecture to identify of quality attribute issues. In the first cycle, mostly issues related to the attributes of reliability, security, and maintainability were found. Therefore, the framework was slightly modified on the component level.

The second phase was based on ATAM. Its purpose was to find out whether architecture adequately supports goals driving architectural design. Therefore, the viewpoints of various stakeholders are taken into account during the evaluation. As stakeholders' concerns and goals may conflict with each other, the ATAM method provides techniques for managing (requirements prioritisation based on cumulative voting and utility tree) and resolving such conflicts (identifying trade-offs). After the completion of the steps required by the ATAM method, the top-down approach was applied and the hierarchy for the utility tree was proposed. The second cycle was realized electronically and the prioritization for the attributes and their problem areas was done by using a three-level ranking: High (H), Medium (M) or Low (L). The utility tree is prioritized along two dimensions: (1) by the importance of each attribute to the system's success; and (2) by the degree of risk posed by achieving this attribute (i.e. how easy it will be to achieve). It was found out that each stakeholder has different concerns, goals, and priorities about the new framework. In the third cycle, the outputs, i.e., the final utility tree as the summarized results from all the stakeholders and the priorities of the quality attributes were individually approved. For this purpose, a qualitative interview approach was used. Finally, quality scenarios cannot be created because there are no detailed requirements about the concrete architecture, only the GEAF.

The final utility tree is depicted in Figure 7 and provides a top-down mechanism for directly and efficiently translating stakeholders' preferences into concrete quality attribute scenarios. There can be seen only the most important attributes. However, the number of problem areas identified through the ATAM session was 47. The results suggest that the architecture framework focusing on the storing, processing, and publishing of BOLD using cloud computing should pay attention to the functional completeness and correctness; availability and fault tolerance of the new system proposed with the use of this framework; time behaviour; user error protection; confidentiality and integrity of data; and the overall interoperability. The most important problem area to be solved is data anonymization.

A design artifact is complete and effective when it satisfies the requirements and constraints of the problem it was meant to solve (Hevner et al., 2004).

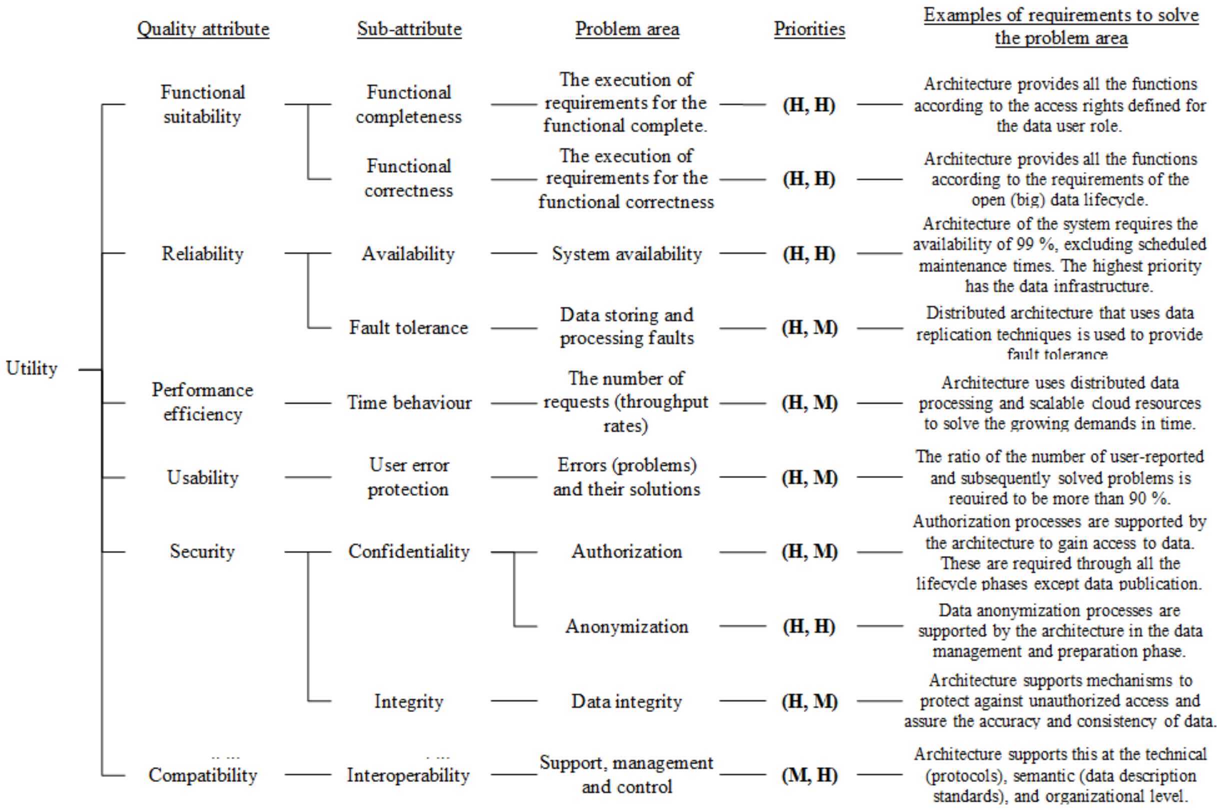

Figure 7: Utility tree for the most important problem areas

5.4 Methodology for the Proposed Framework

The next step was to create a methodology serving as a complex guide for the public sector organizations in the development of their own architectures. According to Zota and Fratila (2013), the EAF should provide a clear architecture development process, related stakeholders and a set of deliverables on each step. As reported by Minoli (2008), if the framework is not described explicitly, it cannot deliver the proposed benefits. The new methodology includes not only the description of required steps, but also related stakeholders and their responsibilities for these steps. It is based on the MMDIS methodology, therefore, it does not already provide the importance of the IS/ICT for the organization, the role of management or related architectures and models. The MMDIS also addresses the complete lifecycle of the system and defines sequential phases as can be seen in Figure 8. However, the new methodology updated only the first two phases on the conceptual level together with the strategic level, which is defined by the new GEAF. It also adapted some agile principles and patterns mentioned in Bruckner et al. (2012) and Nálevka (2009) for the use with the MMDIS.

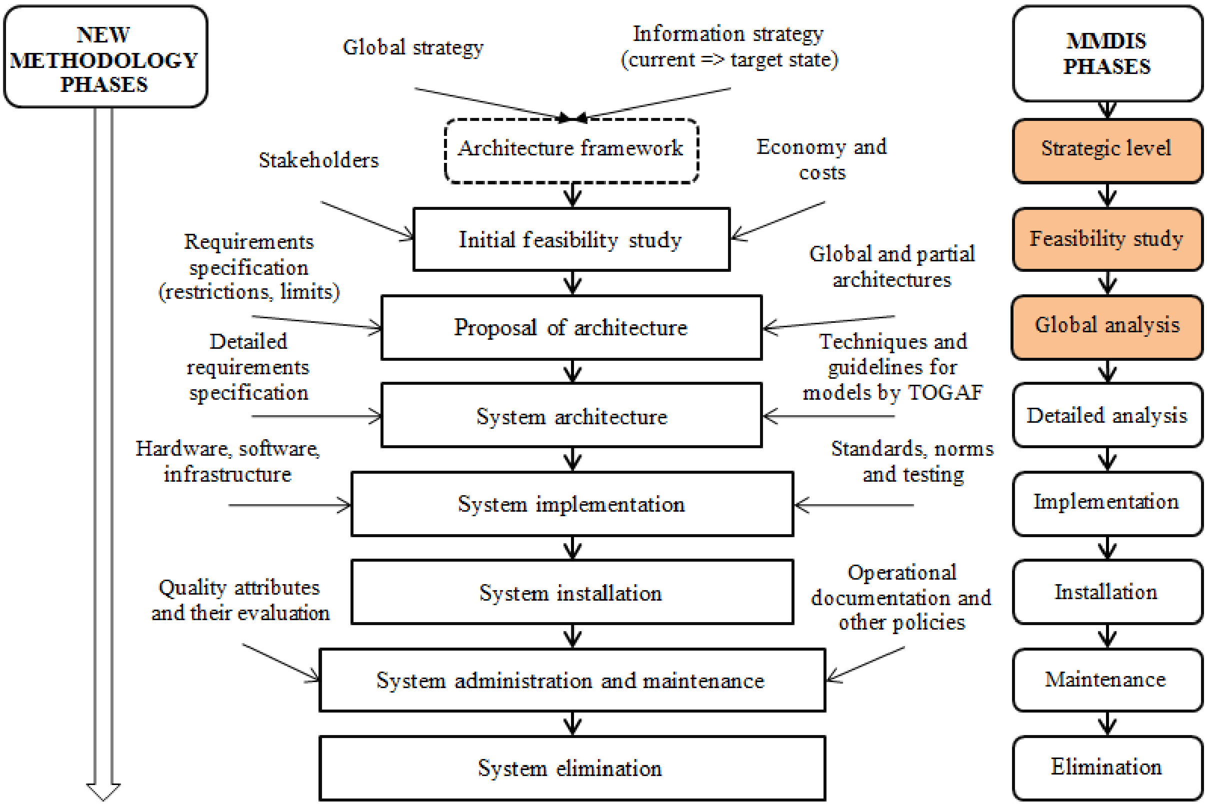

Figure 8: Phases of the new methodology compared to MMDIS phases

The main contribution of the new methodology is therefore the revision of the point of view on the GEAF with regard to the requirements of BOLD and cloud computing for the use in the feasibility study phase and the global analysis phase. Particularly, partial views on architecture are defined more in detail than in the MMDIS methodology. However, this article does not already propose the business architecture and the global architecture, because these are sufficiently described by the MMDIS in Bruckner et al. (2012) and fit in the context of our GEAF. The multidimensional approach is solved in the context of the BOLD lifecycle. For the purpose of the definition of the selected procedures and steps, especially how to develop models, the related of parts of the TOGAF were used.

The key steps of the new methodology are defined as follows:

- the first step is focused on the global and information strategy (a change from current to target state), in this case, it is about the solution, which should be cost effective so as to ensure low cost investments in hardware, software, and data maintenance, it is followed by the analysis of current trends in ICT, analysis of the existing IS/ICT infrastructure in the organization, etc., the responsible stakeholder is system orchestrator, the output of this step is the SWOT analysis;
- the second step is the initial feasibility study phase where it is required to select related stakeholders and create a team, which will define and approve the requirements specification, the output is report with an overview of the basic requirements (boundaries), including various alternatives, and a roadmap containing a list of risks, budget and schedule in the context of economy and costs;
- the third step is the global analysis phase, i.e., proposal of the architecture at the conceptual level, which can be described by the related components for each of the partial views on architecture, the output is a revised list of requirements, global and partial architectures, strategy for ensuring the security and quality attributes, etc., more can be found in the content dimension of the MMDIS;

- it is followed by the detailed requirements specification, while this step is defined by the techniques and guidelines for models suggested by the TOGAF, i.e., ADM/MDA, the output of this step are the concrete models at required levels of views ready to implement, respectively to validate and verify them;
- the following steps then are: system implementation, system installation, system administration and maintenance, and system elimination.

More detailed description is focused on the global analysis phase and it is defined in the context of the BOLD lifecycle. The steps for this part of the new methodology including the related stakeholders and their responsibilities for these steps are shown in Table 5.

Table 5: A defined workflow for the BOLD lifecycle phases. Source: author.

Required activities and steps	Stakeholders	Layers and techniques
a) Data collection and acquisition		
Data acquisition request for a specific purpose (goal)	System orchestrator	Data security and privacy management
Identify appropriate data sources for a given request	Data provider Data application provider	Data sources management
Choose an interface for data access, acquisition and load Data validation and transformation	Data application provider	Regulations and policies management
Data import and storage	Data application provider Data framework provider	Data infrastructure Data security and privacy management Data sources management
b) Data management and preparation		
Data quality control Data search, sort and filter Data transformation and formatting Data integration and linking Data compression	Data application provider	Data security and privacy management Data sources management Regulations and policies management
Data storage and access	Data application provider Data framework provider	Data infrastructure Data security and privacy management Data sources management
c) Data storing and archiving		
Choose appropriate components of the data infrastructure Configuration and service delivery support Infrastructure management and control Cloud and operations management Virtualization management	Data framework provider	Data infrastructure Data security and privacy management Data sources management Architecture management and control Operations support in architecture
Performance and quality indicators measurement and monitoring Cloudonomics management Data backup and archiving	Data application provider Data framework provider	Coordination and cooperation of components Design, develop and deploy tools
d) Data processing and analysis		
Choose of appropriate methods and algorithms for data processing	Data application provider	Data security and privacy management

Choose a framework for distributed data processing Choose support tools and other extensions Data processing and analysis according to the specified purpose (goal) Ensure computing resources (scalability)	Data framework provider	Data sources management Regulations and policies management Data infrastructure Architecture management and control Operations support in architecture
e) Data visualization		
Choose a user interface for access to the results Use reporting and presentation Use graphical visualization tools Choose other tools/services for decision-making support	Data application provider	Data security and privacy management Data sources management Regulations and policies management Data availability and accessibility
Results delivery according to the given request	Data application provider System orchestrator	Data infrastructure Data availability and accessibility
f) Data publication		
Ensure data publication on the web/portal Ensure compliance with open data standards and licenses	Data application provider	Data infrastructure Data security and privacy management
Ensure access for web and mobile applications Ensure access for other systems in the organization	Data application provider Data framework provider	Data sources management Regulations and policies management
Registration and/or log in Data request and feedback Data discussion and evaluation	Data consumer (user)	Data infrastructure Data availability and accessibility

6 Results and Discussion

The main contribution of this article is the proposal of the GEAF suitable for storing, processing, and publishing of BOLD with the use of cloud computing. It follows two approaches: multidimensionality based on the division into layers for cloud computing and the BOLD lifecycle phases. The following key architecture views are proposed: a) process; b) function; c) data; d) software; and e) hardware. Security and privacy cover all the views and are discussed as well with focus on the needs of the public sector. The business view is defined by MICR (2016). The relationships between the views are suitably addressed by the MMDIS (Bruckner et al., 2012).

The own approach based on the modification of the ATAM method, developed for the validation and verification of the framework at the conceptual level represents the authors' contribution to the research on the architecture frameworks evaluation. The lower levels, e.g. software and applications, were included to deepen the analytical part. The proposed approach may be also used to validate and verify the other reference frameworks. It is followed by the prioritization of selected quality attributes including the utility tree with the most important problem areas.

Next benefit is represented by the new methodology proposed as a part of the GEAF. It revises the point of view on the architecture framework and derived architectures with regard to the requirements of BOLD and cloud computing for the use in the feasibility study phase and the global analysis phase in the public sector. A final contribution can be found in the generation of new knowledge, skills, and relationships between stakeholders, which was realized by the use of interpretive and critical research.

The proposal can be beneficiary for the e-government development regarding the proposal of a solution for dealing with large amounts of data. The importance of cloud computing and BOLD in the government enterprise architecture was thoroughly described. As stated by Dwivedi et al. (2017) and Janssen and van den Hoven (2015), BOLD can result in new opportunities and have the potential to transform government and its interactions with the public. Since there does not exist a comprehensive summary of knowledge for commonalities between BOLD and cloud computing in the government enterprise architecture, the systematic literature review was conducted to overcome this shortcoming. Then, the decomposition of published architecture and reference frameworks was done and the key architectural components were identified. In addition, related stakeholders and their roles were determined.

The limitations to validity of the study's results and architectural claims in other countries or organizations are related to different types of administrative structures and responsibility, coverage, and liability between them in the public sectors worldwide. Since cloud computing is a centralized solution that requires a high-performance, reliable, and secure infrastructure, it may be an issue for some countries. In addition, regarding the use of the MMDIS methodology developed in the Czech public sector environment together with that fact that most of the related materials and guidelines are available only in Czech language, it may hinder the applicability of our findings. The next limitation of the GEAF may be in its ability to explain dynamic and interrelated processes in the context of cloud computing and BOLD maturity. For this reason, the organizations should apply different maturity models to determine the current state of these technologies in the organization and identify specific capabilities that are lacking or lagging. Several maturity models or readiness metrics are available for each topic. More can be found in Buyya et al. (2013) or Chen et al. (2014). There are also data source risks, which include data quality risks and risks related to how data are defined, stored, and generated prior to transmission to the cloud (Miller, 2013). Furthermore, as stated by Sharma (2016), to construct a suitable, efficient, and cost-effective cloud computing environment is still a major challenge, especially in the public sector.

The limitations of the used methods are: different stakeholders may have different concerns, hence they may be interested in different aspects of the proposed framework and the results may vary based on the selection of concrete stakeholders and individuals. Also, the costs of a fully-blown ATAM-based evaluation may raise high because of these various opinions, concerns and priorities. Some authors, e.g. Zalewski (2013), are critical for the use of these methods based on the quality scenarios, because there is a missing link on measurable outcomes and sustainability of the resulting system throughout its lifecycle. However, after the detailed requirements are available, it can be recommended to use measurements using metrics and simulations. Another limitation can be the question of topicality with regards to the related platforms, tools, and services, especially the compatibility and maintainability aspect and related problem areas. Limitations can be also based on a lack of interest among the organizations or lack of political support.

Future research will be focused on the detailed requirements, especially in the context of the integration into the existing IS/ICT infrastructure in the organization. After a detailed requirements specification, related hierarchy of models may be proposed and these models may be then evaluated with the use of related methods. Also, it is necessary to deal with the procedures and methods for data anonymization, without which it is impossible to publish these data. More detailed research will be focused on other attributes and problem areas that were found during the validation and verification and are represented by the final utility tree. Finally, more case studies should be done on related platforms, tools, and services for the needs of the public sector.

7 Conclusion

Merging BOLD with cloud computing promises to provide new capabilities for creating more values for citizens and other stakeholders at reduced costs. BOLD storing, processing, and publishing for the public sector organizations of all types are empowered by cloud computing with its wide spectrum of service models. Contemporary organizations collect increasing amounts of data but they usually cannot reach their full potential because they cannot fully analyse the data. Public sector bodies belong to these organizations and they need to implement suitable technologies to better use and reuse collected data, which is even required by the EU Directive on the re-use of public sector information (Directive 2003/98/EC, revised by the Directive 2013/37/EU).

Implementations of new ICT require a suitable EAF because contemporary organization are very complex and they already have adopted many different ICT. Public sector bodies should prevent wasting money because of not well controlled procedure of new ICT. Therefore, this article presents the newly proposed GEAF focusing on BOLD and cloud computing in the context of the Czech public sector. The GEAF should help organizations in dealing with these problems. The framework provides guidelines, architecture principles, the component model from process to hardware services perspectives, and the architecture development methodology for better and efficient service delivery to the citizens in a country. This alignment between public sector goals and IS/ICT allows public sector agencies and institutions to apply available resources to the relevant tasks and activities. The business architecture view is adopted from the MMDIS.

The proposed GEAF provides various theoretical and practical benefits for the issue of the BOLD and cloud computing in the government enterprise architectures. It aims at creation of a unified ICT environment across the public organization with tight links to the management side of the organization and its strategy. This framework is characterized by the partial architecture views to understand the relationship and the interaction between defined components and is applicable in the public sector.

References

Al-Hmoud, H. Y. (2011). Developing a new enterprise architecture framework based on information technology service delivery. In K. S. Soliman (Ed.), *Proceedings of the 17th International Business Information Management Association* (pp. 443–456). Norristown: IBIMA.

Alshomrani, S., & Qamar, S. (2013). Cloud based e-government: Benefits and challenges. *International Journal of Multidisciplinary Sciences and Engineering*, *4*(6), 15–19.

Babar, M. A., & Kitchenham, B. (2007). The impact of group size on software architecture evaluation: a controlled experiment. In *Proceedings of the First International Symposium on Empirical Software Engineering and Measurement (ESEM '07)* (pp. 420–429). Washington, DC: IEEE.

Babar, M. A., Zhu, L., & Jeffery, R. (2004). A framework for classifying and comparing software architecture evaluation methods. In *Proceedings of the 2004 Australian Software Engineering Conference (ASWEC '04)* (pp. 309–318). Washington, DC: IEEE.

Bahrami, M., & Singhal, M. (2015). The role of cloud computing architecture in big data. In *Information Granularity, Big Data, and Computational Intelligence* (pp. 275–295). Cham: Springer International Publishing.

Bellomo, S., Gorton, I., & Kazman, R. (2015). Toward agile architecture: Insights from 15 years of ATAM Data. *IEEE Software*, *32*(5), 38–45.

Bernal, W. N., Caballero, G. C., Sánchez, J. O., & Paéz-Logreira, H. (2016). Enterprise architecture framework oriented to cloud computing services. In *Proceedings of the 2016 6th International Conference on Computers Communications and Control (ICCCC)* (pp. 64–69). Los Alamitos: IEEE.

Birkmeier, D. Q., Gehlert, A., Overhage, S., & Schlauderer, S. (2013). Alignment of business and IT architectures in the German Federal Government: A systematic method to identify services from business processes. In *Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS 2013)* (pp. 3848–3857). Los Alamitos: IEEE.

Bruckner, T., Voříšek, J., Buchalcevová, A., et al. (2012). Tvorba informačních systémů: Principy, metodiky, architektury. Prague: Grada.

Buchholtz, S., Bukowski, M., & Śniegocki, A. (2014). *Big and open data in Europe: A growth engine or a missed opportunity*. Warsaw: demosEUROPA.

Buyya, R., Vecchiola, C., & Selvi, S. T. (2013). *Mastering cloud computing: Foundations and applications programming*. Waltham, MA: Morgan Kaufmann.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209.

Demchenko, Y., De Laat, C., & Membrey, P. (2014). Defining architecture components of the big data ecosystem. In *Proceedings of the 2014 International Conference on Collaboration Technologies and Systems (CTS 2014)* (pp. 104–112). Washington, DC: IEEE.

Demchenko, Y., Grosso, P., De Laat, C., & Membrey, P. (2013). Addressing big data issues in scientific data infrastructure. In *2013 International Conference on Collaboration Technologies and Systems (CTS 2013)* (pp. 48–55). Washington, DC: IEEE.

Denzin, N. K., & Lincoln, Y. S. (2011). *The SAGE handbook of qualitative research*. Los Angeles: SAGE Publications.

Di Martino, B., Aversa, R., Cretella, G., Esposito, A., & Kołodziej, J. (2014). Big data (lost) in the cloud. *International Journal of Big Data Intelligence*, *1*(1-2), 3–17.

Dwivedi, Y. K., Janssen, M., Slade, E. L., Rana, N. P., Weerakkody, V., Millard, J., ... & Snijders, D. (2017). Driving innovation through big open linked data (BOLD): Exploring antecedents using interpretive structural modelling. *Information Systems Frontiers*, *19*(2), 197–212.

Elazhary, H. (2014). Cloud computing for big data. *MAGNT Research Report*, *2*(4), 135–144.

Gill, A. Q., Smith, S., Beydoun, G., & Sugumaran, V. (2014). Agile enterprise architecture: A case of a cloud technology-enabled government enterprise transformation. In *Proceeding of the 18th Pacific Asia Conference on Information Systems (PACIS 2014)* (pp. 1–11). AISeL.

Gong, Y., & Janssen, M. (2017). Enterprise architectures for supporting the adoption of big data. In *Proceedings of the 18th Annual International Conference on Digital Government Research* (pp. 505–510). New York: ACM.

Guijarro, L. (2007). Interoperability frameworks and enterprise architectures in e-government initiatives in Europe and the United States. *Government Information Quarterly*, *24*(1), 89–101.

Harrison, N., & Avgeriou, P. (2011). Pattern-based architecture reviews. *IEEE Software*, *28*(6), 66–71.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98–115.

Hashemi, S., Monfaredi, K., & Masdari, M. (2013). Using cloud computing for e-government: Challenges and benefits. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, *7*(9), 1240–1247.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

IEEE. (2011). *42010-2011 - ISO/IEC/IEEE Systems and software engineering -- Architecture description*. Geneva: International Organization for Standardization.

ISO. (2011). *ISO/IEC 25010:2011 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- System and software quality models*. Geneva: International Organization for Standardization.

Jallow, A. K., Demian, P., Anumba, C. J., & Baldwin, A. N. (2017). An enterprise architecture framework for electronic requirements information management. *International Journal of Information Management*, *37*(5), 455–472.

Janssen, M., Flak, L. S., & Sæbø, Ø. (2013). Government architecture: Concepts, use and impact. In *Proceedings of the 12th IFIP WG 8.5 International Conference (EGOV 2013)* (pp. 135–147). Berlin Heidelberg: Springer.

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.

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.

Kamal, M. M., Hackney, R., & Ali, M. (2013). Facilitating enterprise application integration adoption: An empirical analysis of UK local government authorities. *International Journal of Information Management*, *33*(1), 61–75.

Kazman, R., Barbacci, M., Klein, M., Carrière, S. J., & Woods, S. G. (1999). Experience with performing architecture tradeoff analysis. In *Proceedings of the 21st International Conference on Software Engineering (ICSE '99)* (pp. 54–63). New York: ACM.

Kazman, R., Bass, L., Klein, M., Lattanze, T., & Northrop, L. (2005). A basis for analyzing software architecture analysis methods. *Software Quality Journal*, *13*(4), 329–355.

Kobielus, J. (2014). *Deploying big data analytics applications to the cloud: roadmap for success* [Online]. http://www.thesupplychaincloud.com/wp-content/uploads/2014/07/Deploying-Big-Data-Analytics-Applications-to-the-Cloud.pdf

Lagerstrom, R., Sommestad, T., Buschle, M., & Ekstedt, M. (2011). Enterprise architecture management's impact on information technology success. In *Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS 2011)* (pp. 1–10). Los Alamitos: IEEE.

Liu, L. (2013). Computing infrastructure for big data processing. *Frontiers of Computer Science*, *7*(2), 165–170.

Liu, C. H., Chen, S. C., & Hsieh, P. H. (2015). How big data ecosystem changes cloud services: A design science perspective. *Open Journal of Social Sciences*, *3*(7), 74–78.

Lee, Y. (2017). A life-cycle workflow architecture for linked data. In *Proceedings of the 2017 International Conference on Machine Learning and Soft Computing* (pp. 117–121). New York: ACM.

Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: International Journal of an Emerging Transdiscipline*, *9*(1), 181–212.

Lněnička, M. (2015). AHP model for the big data analytics platform selection. *Acta Informatica Pragensia*, *4*(2), 108–121.

Lněnička, M., & Komárková, J. (2015a). The impact of cloud computing and BOLD on the enterprise architecture framework. In K. S. Soliman (Ed.), *Proceedings of the 26th International Business Information Management Association Conference* (pp. 1679–1683). Norristown: IBIMA.

Lněnička, M., & Komárková, J. (2015b). The performance efficiency of the virtual Hadoop using open big data. *Scientific Papers of the University of Pardubice, Series D, Faculty of Economics and Administration*, *22*(33), 88–100.

Lněnička, M., & Komárková, J. (2018). Big and open linked data analytics ecosystem: Theoretical background and essential elements. *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2018.11.004

Lněnička, M., Máchová, R., Komárková, J., & Čermáková, I. (2017). Components of big data analytics for strategic management of enterprise architecture. In *SMSIS 2017 - Proceedings of the 12th International Conference on Strategic Management and its Support by Information Systems 2017* (pp. 398–406). Ostrava: Technical University of Ostrava.

Luisi, J. (2014). *Pragmatic enterprise architecture: Strategies to transform information systems in the era of big data*. Waltham: Morgan Kaufmann.

Maier, M. (2013). *Towards a big data reference architecture*. Master's thesis, Eindhoven: University of Technology.

Máchová, R., & Lněnička, M. (2017). Evaluating the quality of open data portals on the national level. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(1), 21–41.

Mahmood, Z. (2011). Cloud computing for enterprise architectures: Concepts, principles and approaches. In *Cloud Computing for Enterprise Architectures* (pp. 3–19). London: Springer.

Marton, A., Avital, M., & Jensen, T. B. (2013). Reframing open big data. In J. van Hillegersberg, E. van Heck, & R. Connolly (Eds.), *Proceedings of the 21st European Conference on Information Systems (ECIS 2013)* (Paper 169). Utrecht: Utrecht University.

Mary, S. R., & Rodrigues, P. (2011). Survey and comparison of frameworks in software architecture. In *International Conference on Advances in Computing and Communications* (pp. 9–18). Berlin Heidelberg: Springer.

MICR – Ministry of the Interior of the Czech Republic. (2016). *Strategic framework for the development of public administration in the Czech Republic for the period 2014 – 2020*. Prague. [Online]. http://www.mvcr.cz/soubor/strategic-framework-for-the-development-of-public-administration-in-the-czech-republic-for-the-period-2014-2020.aspx

Miller, H. E. (2013). Big-data in cloud computing: A taxonomy of risks. *Information Research, 18*(1), paper 571.

Minoli, D. (2008). *Enterprise architecture A to Z: Frameworks, business process modeling, SOA, and infrastructure technology*. Boca Raton: Auerbach Publications.

Mohamed, M. A., Galal-Edeen, G. H., Hassan, H. A., & Hasanien, E. E. (2012). An evaluation of enterprise architecture frameworks for e-government. In *Proceedings of the Seventh International Conference on Computer Engineering & Systems (ICCES 2012)* (pp. 255–260). Los Alamitos: IEEE.

Nálevka, P. (2009). Critical analysis of MMDIS concepts: Comparing agile with MMDIS approaches to information system development. *Systémová integrace*, *2009*(4), 119–142.

NIST. (2011). NIST cloud computing reference architecture: Recommendations of the national institute of standards and technology. *NIST Special Publication 500-292* [Online]. http://www.nist.gov/customcf/get_pdf.cfm?pub_id=909505

NIST. (2015). NIST big data interoperability framework: Volume 6, reference architecture. *NIST Special Publication 1500-6* [Online]. http://bigdatawg.nist.gov/_uploadfiles/NIST.SP.1500-6.pdf

Ojo, A., Janowski, T., & Estevez, E. (2012,). Improving government enterprise architecture practice – maturity factor analysis. In *Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS 2012)* (pp. 4260–4269). Los Alamitos: IEEE.

Oliveira, J., Delgado, C., & Assaife, A. C. (2017). A recommendation approach for consuming linked open data. *Expert Systems with Applications*, *72*, 407–420.

Pääkkönen, P., & Pakkala, D. (2015). Reference architecture and classification of technologies, products and services for big data systems. *Big Data Research*, *2*(4), 166–186.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Malden: Blackwell Publishing.

Ramachandran, M. (2011). Component-based development for cloud computing architectures. In *Cloud Computing for Enterprise Architectures* (pp. 91–114). London: Springer.

Rimal, B. P., Jukan, A., Katsaros, D., & Goeleven, Y. (2011). Architectural requirements for cloud computing systems: An enterprise cloud approach. *Journal of Grid Computing*, *9*(1), 3–26.

Sá, J. O., Martins, C., & Simões, P. (2015). Big data in cloud: A data architecture. *Advances in Intelligent Systems and Computing*, *353*, 723–732.

Salmons, J. (2011). Designing and conducting research with online interviews. In *Cases in Online Interview Research* (pp. 1–30). Thousand Oaks: SAGE Publications.

Schmidt, R., & Möhring, M. (2013). Strategic alignment of cloud-based architectures for big data. In *Proceedings of the 17th IEEE International Enterprise Distributed Object Computing Conference Workshops (EDOC 2013)* (pp. 136–143). Los Alamitos: IEEE.

Sharma, S. (2016). Expanded cloud plumes hiding big data ecosystem. *Future Generation Computer Systems*, *59*, 63–92.

Shroff, G. (2010). *Enterprise cloud computing: Technology, architecture, applications*. New York: Cambridge University Press.

Sutherland, S., & Chetty, G. (2014). Migration to cloud computing: A sample survey based on a research in progress on the investigation of standard based interoperability protocols for the convergence of cloud computing, servcie oriented architecture and enterprise architecture. *International Journal of Information Processing and Management*, *5*(1), 50–61.

Svahnberg, M., & Wohlin, C. (2005). An investigation of a method for identifying a software architecture candidate with respect to quality attributes. *Empirical Software Engineering*, *10*(2), 149–181.

Tang, L., Dong, J., Zhao, Y., & Zhang, L. J. (2010). Enterprise cloud service architecture. In *Proceedings of 2010 IEEE 3rd International Conference on Cloud Computing* (pp. 27–34). Los Alamitos: IEEE.

Tian, W., & Zhao, Y. (2015). Big data technologies and cloud computing. In *Optimized Cloud Resource Management and Scheduling: Theory and Practice* (pp. 17–49). Waltham: Morgan Kaufmann.

The Open Group. (2009). *TOGAF version 9: The open group architecture framework (TOGAF)*. The Open Group.

Urbaczewski, L., & Mrdalj, S. (2006). A comparison of enterprise architecture frameworks. *Issues in Information Systems*, *7*(2), 18–23.

Van Schalkwyk, F., Willmers, M., & McNaughton, M. (2016). Viscous open data: The roles of intermediaries in an open data ecosystem. *Information Technology for Development*, *22*(sup1), 68–83.

Yu, F., Stella, C., & Schueller, K. A. (2014). A design of heterogeneous cloud infrastructure for big data and cloud computing services. *Open Journal of Mobile Computing and Cloud Computing*, *1*(2), 1–16.

Zalewski, A. (2013). *Modelling and evaluation of software architectures*. Warsaw: Oficyna Wydawnicza Politechniki Warszawskiej.

Zimmermann, A., Pretz, M., Zimmermann, G., Firesmith, D. G., Petrov, I., & El-Sheikh, E. (2013). Towards service-oriented enterprise architectures for big data applications in the cloud. In *Proceedings of the 2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops* (pp. 130–135). Los Alamitos: IEEE.

Zota, R. D., & Fratila, L. A. (2013). Preliminary aspects in order to develop an efficient method for cloud implementation standardization. In *Proceedings of the 2013 RoEduNet International Conference 12th Edition: Networking in Education and Research* (pp. 1–6). Los Alamitos: IEEE.

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, *19*(1, 2), 17–33.