Electronic Government and Electronic Participation
E. Tambouris et al. (Eds.)

doi:10.3233/978-1-61499-570-8-109

Characteristics of business intelligence and big data in e-government: Preliminary findings

Rikke GAARDBOE[a], Tanja SVARRE[a,1] and Anne Marie KANSTRUP[a]
[a] *eLearning Lab, Department of Communication and Psychology, Aalborg University, Rendsburggade 14, DK-9000 Aalborg, Denmark*

Abstract. Business intelligence and big data represent two different technologies within decision support systems. The present paper concerns the two concepts within the context of e-government. Thus, the purpose of the paper is to present the preliminary findings regarding publication patterns and topic coverage within the two technologies by conducting a comparative literature review. A total of 281 papers published in the years 2005–2014 were included in the analysis. A rapid increase of papers regarding big data were identified, the majority being journal papers. As regards business intelligence, researchers publish in conference proceedings to a greater extent. Further, big data journal papers are published within a broader range of journal topics compared to business intelligence journal papers. The paper concludes by pointing to further analyses that will be carried out within the 281 selected papers.

Keywords. Big Data, Business Intelligence, E-government, Literature review

1. Introduction

According to Gartner [1], business intelligence/analytics is the top-priority issue for CIOs and investments. The public sector could increase the effectivity of administration by using big data. An estimate is 250 billion EUR each year, which is equivalent to 0,5% annual productivity growth in the European Union [2]. The two terms applied to describe the software for business analytics are 'business intelligence' and 'big data'. Business intelligence and big data are used for supporting online political participation, e-government service delivery, process transparency and accountability, and is based on opinion mining, social network analysis and data from the accounting system [3]. The term 'decision support system' originated 50 years ago and covers computer-based tools for sense-making and decision-making. The terms used have changed over time, and among others, the term 'business intelligence' is now more commonly used. However, although new terms appear, they cover the same purpose[4]. Some researchers consider big data business intelligence 3.0 [3] as a new generation of business intelligence and analysis, while other researchers consider it as a paradigm shift [5].

The two technologies share the purpose of delivering decision support in a changing world. However, despite terminological overlaps and shared ambitions about

[1] Corresponding Author.

decision support, business intelligence and big data are not identical. Pedersen [6] has identified seven differences between business intelligence and big data which are included in Table 1.

Table 1 Differences between BI and Big Data, based on Pedersen [6]

	Business Intelligence	**Big Data**
Data types	Structured (Mostly)	Unstructured (a.o)
Data sources	Mostly internal	Mostly external
History	Essential	(Often) less relevant
Users	Manager/controller	Data scientist
Precision	Exact results	Approximate results
Privacy	Not critical	Critical
Control over data	Almost full control	Little or no control

The main difference between business intelligence and big data is the data source, as it is crucial to further architecture. Business intelligence uses data structured in relational form, preferably from the organization's internal systems; this means that there is full control of data and that data are structured and validated. In contrast, big data applications often use external data, where there is no control [6]. Consequently, the potential of the two types of technologies in the public sector varies.

Olszak [7] emphasized effectiveness as an important potential of business intelligence in the public sector, especially with respect to the planning and process improvements in an organization. Manyika et al. [2] highlighted the potential of big data, emphasizing the ability to combine structured and unstructured data. In the public sector. That has the following potentials: creating transparency with accessible related data; discovering needs; improving performance; customizing actions for suitable products and services; decision making with automated systems to decrease risks and innovating new products and services. Furthermore, in the health care clinical decision support systems, individual analytics can be applied for patient profiles, personalizing medicine, performance-based pricing for personnel, analyzing disease patterns and improving public health [8].

The idea for this article arose from a PhD study about business intelligence in the public sector. Terminological confusion and important differences between the two types of technologies require researchers to study the literature on business intelligence and big data in e-government. The purpose of the present review is to compare the extent, growth and nature of the technologies in government research. Subsequently, through analysis, we strive to understand and clarify the differences and overlaps between the technologies in research.

This paper presents the preliminary results from our literature study on business intelligence and big data in e-government. These initial results indicate trends and implications for literature studies within these areas. As with all literature reviews, we aim to analyse a given topic and identify research trends and potential gaps that can lead to new studies [9]. Below we present the first steps in this direction.

2. Method

The present paper builds upon a comparative, systematic literature review. Taking this approach, the study is based on informetric methods to quantitatively measure and compare the development of the two subject fields of business intelligence and big data in an e-government context [10], [11]. A common approach in literature reviews is to

limit the sources of papers to core journals within the field in question. However, to be able to test whether research papers on the two technologies in question are published in a variety of channels, we are not limiting the sources of potentially relevant papers to specific journals or conferences. Instead, in collecting relevant references and delimiting the subject field, we searched the following three databases covering computer science (journals, monographs *and* conferences): Web of Science (ISI), Scopus (Elsevier) and ABI/INFORM Complete (Proquest). Another reason for searching these databases was their advanced search interfaces, enabling complex queries for retrieval of documents.

In each database two queries were defined: one for business intelligence and one for big data. To be included in the pool of relevant papers, papers needed to be peer reviewed and written in English. The e-government domain was represented as follows:

(municip OR govern* OR council* OR ministr* OR "public administration" OR "public sector" OR egovern* OR e-govern*)*

The purpose of this representation of the e-government query component was to ensure a broad representation of the domain in question, and to enable the inclusion of papers concerned with government in general. The underlying assumption was that the majority of governments in the Western world will have some level of digitalization within the specific period under investigation. Lastly, the year of publication was defined to include papers within the time window 2005–2014. In sum, the queries carried out consisted of the following search components:

SUBJECT ((BI OR BD) AND e-government) AND DOCUMENT TYPE AND PUBLICATION YEAR AND LANGUAGE

Each query was adjusted to the specific command language of the three databases searched. Subsequently, the retrieved documents were entered into an access database, where duplicates were removed; as a result of this process, the access database now contains 99 papers on business intelligence and 182 papers on big data. No overlap was found between the two pools of papers. The variables characterizing the reviewed papers comprise various bibliographical data, abstracts, Dewey classification codes (for journal papers), and for some papers, full text versions and/or cited references. In the analysis below, we report on and compare the full collection with respect to publication years and document types. In addition, we have added Ulrichsweb's (https://ulrichsweb-serialssolutions-com) Dewey classification codes for journals to the part of our papers covering journal papers. The purpose of adding the Dewey codes where possible is to investigate the extent of topics covered by business intelligence and big data, respectively. However, conferences are not covered by Ulrichsweb, meaning that this particular analysis will be carried out solely for journal papers.

3. Preliminary findings

In this section we present the preliminary findings of the comparative review. The results are presented in two sections; in the first section, the results are analyzed for the whole population based on the 281 collected articles. The second part of the results specifically reports findings regarding journal papers (cf. Section 2 on Dewey codes).

Table 2 Number of e-government research publications concerning Business Intelligence and Big Data in the period of 2005-2014

Year	Business Intelligence	Big Data
2005	3	-
2006	1	-
2007	6	-
2008	12	1
2009	10	2
2010	6	-
2011	17	1
2012	19	19
2013	11	46
2014	14	113
Total	99	182

Within e-government business intelligence, there is a general increase of publications from 2007 to 2012 (see Table 2). After 2012, the number decreased again. The trend in big data publications differs, as the number of publications has grown exponentially since 2011 (see Table 2). From 2013 to 2014, the number of publications more than doubled. Thus, in the period business intelligence represent a minor, but permanent interest within research, whereas big data denotes a newcomer with an increasing focus. With the introduction of NoSQL databases around 2009, an increase in performance and flexibility in handling big data appeared [12]. INMemery technology was also launched for commercial use that year; for example, SAP launched their system HANA, which is an INmemery data management system that improves the performance of analytical and transactional applications [13]. Both technologies support big data by enabling processing of large data volumes. The introduction of new technologies may partially explain the significant increase in the number of research publications on big data and e-government. However, further analyses are needed to fully understand the nature of the growth.

Table 3 Publication types applied for research dissemination by researchers

Publication type	Business Intelligence	Big Data
Conference proceedings	58 (59%)	32 (18%)
Generic	1 (1%)	-
Journal article	25 (25%)	139 (76%)
Monograph	15 (15%)	11 (6%)
Total	99	182

Another interesting finding is the type of communication channels chosen by the authors. Table 3 shows the distribution of publication types across the two fields. As illustrated in the table, business intelligence has the largest share of conference papers (59% of all publications), whereas big data has an even larger share of journal papers (76% of all publications). As previously stated, conferences tend to publish earlier research results, whereas scientific journals publish more completed work [14]–[16]. Considering the publication patterns of the two fields, it is surprising that the youngest of these fields has the largest share of journal papers. An expected distribution would be a larger share of conference papers for the younger field and a larger share of journal papers within the older, more established field of research. We will look into possible causes in additional analyses.

Table 4 summarizes the subject distribution of the journals chosen for publication by the authors. For both technologies, the three most frequent categories comprise Computer Science, Information and General Works, Social Sciences and Technology.

Papers concerning big data and e-government are also published in the categories of Science, Arts and Recreation, History and Geography, Language and Philosophy, and Psychology.

Table 4 Dewey main classification of journals (as added by Ulrichsweb) containing the journal papers of the collected papers

Dewey main classification	Business Intelligence	Big Data
Arts & recreation	-	1 (1%)
Computer science, information & general works	10 (40%)	41(29%)
History & Geography	-	2 (1%)
Language	-	1 (1%)
Philosophy & Psychology	-	2 (1%)
Science	-	9 (6%)
Social Sciences	5 (20%)	43(31%)
Technology	10 (40%)	40(29%)
Total	25 (100%)	139 (100%)

Thus, despite overlaps in the most frequent categories, big data research apparently has a more diverse scope of topics compared to business intelligence. The more varied distribution of big data may partially be explained by a considerably larger number of journal papers written about big data (139) compared to business intelligence (25 papers). For the e-government research field, this means that broader perspectives are needed in terms of subject coverage to ensure a sufficient reflection of the field. However, it is nonetheless an interesting finding that we will examine more thoroughly in a follow-up study.

4. Future analysis and perspectives

Above we have presented the preliminary results of the distribution of papers within the two research fields under investigation. We now have an indication of the distribution of publications with respect to publication year, type, volume and journal subject. However, in some cases, we have been puzzled by the distribution. In their review typology, Paré et al. [17] identified various ways of composing a review. To explain the distributions reported in the present paper, we plan to carry out a follow-up study. The follow-up study should be designed as a scoping review based on the comprehensive search strategy also applied here. In a scoping review, the methods applied for analysis include content and thematic analysis. To conduct these analyses, we intend to examine different aspects of the papers.

One aspect of the papers that should be investigated further is the author-supplied keywords added to the papers. As opposed to the controlled Dewey codes reported above, author keywords represent an uncontrolled way of indicating the content of a paper. Uncontrolled elements in a subject description are characterized by not having any delays in incorporating new terms and reflecting the mindset of the author [18]. As an extension of the Dewey code analysis, and to avoid being restricted to journal papers, we intend to conduct an analysis of the author keywords by means of text mining [19]. Using this technique, the most frequent keywords can be identified across authors. We will also carry out citation analyses to identify core authors, papers and journals within the fields.

In addition, we want to carry out more thorough coding of the content of the papers gathered. In the first round of coding, we will identify the use of methods in the papers by means of the typology put forward by Dwivedi [20]. Another interesting coding to perform is to identify the unit of analysis. To do this, we will apply Joseph's [21] categorization.

The method applied for collecting papers presents several benefits, including that we have not limited the scope of the two technologies to a specific journal. Therefore, a broad collection of papers have informed the review. However, we may discuss the number of years selected for the review. Regarding big data, it appears from the number and distribution of papers that we have framed the significant years of publication. However, business intelligence dates further back than 2005. This means that from a methodical perspective, the full range of business intelligence papers is not included in the present comparison of technologies. This condition should be taken into consideration in the assessment of the findings.

5. Conclusion

The publication of business intelligence articles in the context of e-government has steadily increased since 2005, whereas the publication of big data articles has grown exponentially since 2011. Despite big data being a new technology, a lot of research findings have been published as journal papers. On the contrary, the more established field of business intelligence were to a larger extend disseminated through scientific conferences. The purpose of a literature review is to identify a research area in depth, meaning that journal articles and conferences must both be involved as common procedure. Therefore, 281 publications were analyzed in this study. However, we have made several interesting findings. To be able to understand the reasons behind, we need to dig deeper into the publications and apply additional analyses in a follow up study.

References

[1] Gartner, *Flipping to Digital Leadership: Insights from the 2015 Gartner CIO Agenda Report*, Gartner, Stamford, US, 2014.
[2] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, *Big Data: The next Frontier for Innovation, Competition, and Productivity*, McKinsey Global Institute, 2011.
[3] H. Chen, R. H. Chiang, and V. C. Storey, Business intelligence and analytics: From big data to big impact., *MIS Quarterly*, **36** (2012), 1165–1188.
[4] S. Alter, A work system view of DSS in its fourth decade, *Decision Support Systems*, **38** (2004), 319–327.
[5] A. McAfee and E. Brynjolfsson, Big data: The management revolution. (cover story), *Harvard Business Review*, **90 (**2012), 60–68.
[6] T. B. Pedersen, Managing big multidimensional data, *EGC 2015. Revue des Nouvelles Technologies de l'Information*, vol. RNTI-E-28, 2015.
[7] C. M. Olszak, Business intelligence and analytics in organizations, in *Advances in ICT for Business, Industry and Public Sector*, vol. 579, M. Mach-Król, C. M. Olszak, and T. Pełech-Pilichowski, Eds. Cham: Springer International Publishing, (89–109), 2015.
[8] S. Sagiroglu and D. Sinanc, Big data: A review, in *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, (2013), 42–47.
[9] Y. Levy and T. J. Ellis, A systems approach to conduct an effective literature review in support of information systems research, *Informing Science: International Journal of an Emerging Transdiscipline*, **9** (2006), 181–212.

[10] J. Tague-Sutcliffe, An introduction to informetrics, *Information Processing & Management*, **28** (1992), 1–3.
[11] L. Egghe, Expansion of the field of informetrics: Origins and consequences, *Information Processing & Management*, **41** (2005), 1311–1316.
[12] M. Stonebraker, SQL databases v. NoSQL databases, *Communications of the ACM*, **53** (2010), 10-11.
[13] F. Färber, S. K. Cha, J. Primsch, C. Bornhövd, S. Sigg, and W. Lehner, SAP HANA database: data management for modern business applications, *ACM Sigmod Record*, **40** (2012), 45–51.
[14] W. D. Garvey and B. C. Griffith, Communication and information processing within scientific disciplines: Empirical findings for Psychology, *Information Storage and Retrieval*, **8** (1972), 123–136.
[15] W. D. Garvey, N. Lin, C. E. Nelson, and K. Tomita, Research studies in patterns of scientific communication: I. General description of research program, *Information Storage and Retrieval*, **8 (**1972), 111–122.
[16] B.-C. Bjork, A model of scientific communication as a global distributed information system, *Information Research*, **12** (2007).
[17] G. Paré, M.-C. Trudel, M. Jaana, and S. Kitsiou, Synthesizing information systems knowledge: A typology of literature reviews, *Information & Management*, **52**, (2015), 183–199.
[18] C. P. R. Dubois, Free text versus controlled vocabulary, *Online Review*, **1 (**1987), 243–253.
[19] W. Fan, L. Wallance, S. Rich, and Z. Zhang, Tapping the power of textmining, *Communications of the ACM*, **49** (2006), 76-82.
[20] Yogesh K. Dwivedi, An analysis of e-Government research published in Transforming Government: People, Process and Policy (TGPPP), *Transforming Government*, **3** (2009), 7–15.
[21] R. C. Joseph, A structured analysis of e-government studies: Trends and opportunities, *Government Information Quarterly,* **30** (2013), 435–440.