

# Big data in construction industry: systematic literature overview

Atis Verdenhofs<sup>1</sup>, Ineta Geipele<sup>2</sup>, Tatjana Tambovceva<sup>3</sup>

*Institute of Civil Engineering and Real Estate Economics, Riga Technical University, Riga, Latvia*

*E-mail: <sup>1</sup>atis@enkurs.org (corresponding author)*

**Abstract.** Technological advancement has led to tremendous increase of data. Many industries utilize big data to become more efficient or even to create new products or services. Applications of big data in construction industry has been extensively researched in Asia that can be explained with huge construction volumes in the area. This study is aimed at identifying big data applications in construction industry in time period beyond 2016.

Research object is construction industry, research subject is big data applications. Research methods used in this research are systematic literature overview and meta-analysis. Novelty of the research is classification of big data applications based on systematic literature overview.

Authors conclude that existing categorization (Bilal et al., 2016b) can be applied to researches about big data applications in construction industry published in 2016 and later. However, potential for new applications is identified in category of emerging trends triggered by big data and authors propose to perform cross-industry analysis to identify solutions that can be adopted to construction industry.

**Keywords:** big data, systematic literature overview, construction industry, big data applications

## Introduction

Applications of big data in construction industry has been extensively described in research (Bilal et al., 2016b). Authors of the research indicates five categories of opportunities for big data to be applied in construction industry. This research is aimed to perform systematic literature review to verify defined categories and update categories of big data applications in construction industry if such need is identified.

Authors are using well known academic literature databases and creating selection criteria to identify appropriate researches that can be used in further analysis. Two types of analysis are performed to categorize selected researches. During the process authors identified that more academic researches are coming from Asia - region in which construction industry and technological advancement is booming.

## Methodology

Systematic literature review is based on two databases of scientific publications – Web of Science (WoS; now owned by Clarivate Analytics) and Scopus. Search of literature was performed in January 2019. Set of two keywords were used in both searches – “big data” and “construction” to identify baselist of researches for further selection and analysis. Additional limitation for searching publications – keywords search done only in publication titles. This limitation allows to identify publications that have selected keywords as main topic.

Selection criteria for publications is shown in Table 1. It contains number of results in both databases, duplicating researches that are available in both databases and how many of them are published in 2016 or later.

Table 1. Selection criteria for publications

Database	Search criteria	Number of results	Publications identified in both databases	Published in 2016 or later
Scopus	TITLE (“big data” AND “construction”)	120	51	41
Web of Science	TITLE: (“big data” AND “construction”)	113	51	41

For selection to be more precise and representing highest quality of selected publications all identified publications are merged and only duplicated publications indexed in both databases are selected for further analysis. 51 distinct publications are identified that are available in both scientific publication databases. However, since there has been already extensive study performed on big data in construction industry (Bilal et al., 2016b), that was submitted for publishing in beginning of 2016 and published in middle of 2016 then all selected publications with date of publications before 2016 are eliminated from further analysis. 10 publications are excluded from the list due to time restriction and 41 remains for final analysis.

After deep analysis of identified 41 articles authors excluded 14 articles based on content of the article that was only partly or not connected with construction industry and keyword combination matched different area of research. Further analysis of publications is based on remaining 27 researches.

## Description of publications

Authors identified publication (Bilal et al., 2016b) containing categories of big data usage in construction. There are five categories of applications indicated:

1. Resource and waste optimization,
2. Value added services,
3. Facility management,
4. Energy management and analytics,
5. Other emerging trends triggered by big data.

Following categories will be used to identify areas of applications described in selected articles published in 2016 or later. If selected research will not fit into any of existing category, new category will be indicated.

### Characteristics of publications

There are three types of publications that are selected for further analysis – 11 articles, 14 conference papers and 2 reviews. Out of 27 publications all publications are cited in other publications at least once based on information available in Scopus database. Web of Science database indicates that out of 27 publications 16 are not cited in any other publications. Five most highly cited publications in Scopus database and Web of Science database are the same, only ranking and number of citations are different. Detailed list of most cited publications and number of citations in each database is shown in Table 2.

Table 2. Top 5 highest cited publications in WoS and Scopus

Publication	Times cited in Web of Science	Times cited in Scopus
Bilal et al. (2016b)	32	50
Lu et al. (2016)	27	30
Bilal et al. (2016a)	15	22
Liu et al. (2018)	18	20
Chen et al. (2018)	3	7

Publication that authors used as reference for further analysis of big data applications in construction (Bilal et al., 2016b) is most cited publication out of all selected publications. Therefore, it supports the decision that this research is using additional criteria of publication date 2016 and later to identify most up to date applications of big data.

Selected 27 publication by the year of publication are split in following way – 10 are published in 2016, 10 are published in 2017, 5 are published in 2018 and 2 are published in 2019. The most represented research areas of publications are computer science and engineering. Only 6 out of 27 publications have only 1 author. 13 publications have 4 or more authors. There are 2 publications that both have 9 authors. Those are (Bilal et al. 2016a) and (Bilal et al., 2016b) – two publications that are among most cited in Scopus and WoS out of all selected publications. Researches that have the greatest number of authors are also the ones that have highest number of citations.

In all 27 publication authors have used 128 keywords as indication for their research. By merging all keywords into one list and calculating most commonly used keywords authors of this research found that most commonly used keyword in selected articles is “Big Data” – 20 different publications have this keyword as author keyword. Only 5 more keywords are used more than once in all selected publications, those are “Big Data Analytics”, “Data Mining”, “Hadoop”, “Ecology” and “Construction Waste Management”. Last two keywords can be associated with one of five categories previously defined – resource and waste optimization. Diversity of keywords indicate wide range of research areas represented in selected publications while still qualifying for research object and subject – big data applications in construction industry.

## Systematic overview

Two approaches are used in this research to categorize selected 27 articles. First method is based on text analytics by automatically analysing abstracts of selected researches to fit each in one or more categories defined by (Bilal et al., 2016b). Second method is based on manual research categorization into same categories or creating new categories if content of research indicates it.

Authors of this research is not analysing if all selected researches for systematic literature overview is using aligned definition of the term “big data”. This limitation is set because there is more than one process, way how to utilize big data and there are also more than one definition of “big data” available in academic environment. (Koseleva & Ropaite, 2017) research indicate at least two options how “big data” can be described – either with three or four main characteristics (3V or 4V models).

### *Categorization with text mining approach*

Authors use simple text mining approach to categorize selected articles into identified 5 categories. Text mining is based on using specific keywords that describe each of 5 categories. Abstracts of selected researches are used for identification of selected keywords to categorize researches. Table 3 show what keywords are used for each category and how many researches can fit into each of 5 categories based on this approach.

Table 3. Text mining keywords and categories

Category	Keywords used in text mining	Researches that fit in category
Resource and waste optimization	Resource, waste, optimization	9
Value added services	Value, service	7
Facility management	Facility, equipment	1
Energy management and analytics	Energy, analytic	5
Other emerging trends triggered by big data	Big data	22

This approach categorizes 85% of selected researches. Only 4 out of 27 selected researches did not fit any of predefined category. 8 researches fit one category, 10 researches fit two categories, 4 researches fit 3 categories and 1 research (Lu, 2019) fit 4 categories. 55% of researches fit category “Other emerging trends triggered by big data” – this result is based on having keyword combination “big data” in abstract of researches. Least represented category is “Facility management” that has only 1 research that fit this category.

Based on text mining approach categories of big data application in construction industry could be decreased from 5 to 4 categories by merging “Facility management” into “Other emerging trends triggered by big data”. This categorization approach also does not cover 4 researches that did not contain any of keywords in research abstracts therefore manual categorization is needed to categorize all researches.

### *Manual categorization of researches*

After getting acquainted with selected 27 articles authors of this research categorize all researches into proposed 5 categories by (Bilal et al., 2016b) based on the content of each research. Results of manual categorization can be seen in Table 4.

Table 4. Manual categorization results

Category	Researches that fit in category
Resource and waste optimization	9
Value added services	9
Facility management	5
Energy management and analytics	4
Other emerging trends triggered by big data	8

One of the differences in two categorization approach is that manual categorization distributes selected researches more evenly between 5 categories than text mining approach. Second categorization approach also categorized all selected articles into at least of given categories. Two categories have the same number of researches that fit into the category. Categories “Facility management” and “Energy management and analytics” had less researches that qualified for the category – 5 and 4 respectively.

During the manual research analysis and categorization authors of this research did not identify new applications of big data in construction industry that did not fit into any of given 5 categories defined by (Bilal et al., 2016b).

### **Categories of big data applications**

Aim of this research is to classify applications of big data in construction industry. Authors identified research (Bilal et al., 2016b) that had extensive analysis done on the topic published in 2016. Since the technological development is happening so fast together with potential applications of new technologies in different areas authors are willing to verify if 5 categories of big data applications in construction industry is still relevant based on researches published in 2016 and onward and if not propose update in categories of big data application in construction industry.

Further authors indicate categories of big data application and identified researches that fit into each of the category together with examples of applications that describe the variety of categories.

#### *Resource and waste optimization*

In both categorization methods 9 researches fit into the category while 5 of them fit the category by using both methods. (Lu, Chen, Ho, & Wang, 2016) performed an analysis of the waste management in Hong Kong area comparing public and private sector performance by using big data. Same author only alone performed research on Hong Kong illegal waste dump identification based on big data solutions (Lu, 2019). (Bilal et al., 2016a) however propose a conceptual framework for construction waste analytics. (Xianglan, 2017) is proposing to use big data to analyse and improve performance in the coal mine industry. (X. Liu, Dong, Y. Liu, A. Liu, & Xiong, 2018) research the challenge of delays in communication and reduction of complexity between connected dominating sets that can lead to reduction of energy consumption.

Resource and waste optimization category cover wide range of researches that fit into construction industry and big data applications. Authors indicate that this big data application category is still valid based on identified researches published in 2016 and onward.

All publications that are assigned to this category – Bilal et al. (2016a), Bilal et al. (2016b), Han and Wang (2017), R. H. Liu, Kuo, Yang, Chen, and J. C. Liu (2016), Lu (2019), Lu et al. (2016), A. Konikov and G. Konikov (2017), Weijie, Jinna, Zhengzheng, and Na (2016), Xianglan (2017).

#### *Value added services*

Researches that fit into this category are 9 based on manual categorization. Identified researches that represent value added services also offer wide range of applications of big data in construction industry. (Chen, Wang, Sun, & Lu, 2016) propose to automate decisions in the approval process for construction based on historical approval data. An area that is not directly construction related but which will improve early stage process of construction projects. (Liu, 2016) research an intelligent traffic management system that would improve traditional ETL technology and adopt it to big data gathered from smart traffic technologies. (Chi, Meng, K. Zhai, G. Zhai, & Liu, 2017) propose an algorithm that evaluates historical records of pipe constructions to identify if there are any old pipelines in engineering sites. Solution is based on big data analysis.

Big data applications can be applied and be useful in early stages of construction processes that would make construction process faster or manage early stages of projects more efficiently.

All publications that are assigned to this category – Bilal et al. (2016b), Chen et al. (2016), Chi et al. (2017), Duan, Wang, Li, and Zhang (2018), Fang, Jiaming, Yaohui, and Mei (2016), Guo and Zhai (2017), Liu (2016, 2017), Yun (2016).

#### *Facility management*

This category relates to development of different sources of data, especially combination and analysis of data coming from different sensors. (Zhang et al., 2017) research applications of big data that improve water and fertilizer usage in agriculture. Solution is based on sensor data available in real time and automatic collection and analysis of growth status of crops. Similar area is analysed in research by (Shuo, 2017). Author use the development of big data solutions to create an industrial flow for ecological agriculture.

Facility management category is closely related to the term “Internet of Things” and development of sensors that generate data. And if there are new data sources that generate data then also researches that cover data flows, communication between devices can be fit into this category of big data applications.

All publications that are assigned to this category – Bilal et al. (2016b), Liu et al. (2016), Liu et al. (2018), Shuo (2017), Zhang et al. (2017).

#### *Energy management and analytics*

This category contains several researches that fit also under the category “Resource and waste optimization” since energy can be considered as a resource. One of the researches that is categorize by authors in both categories is (Liu et al., 2016) research that propose a smart solution how energy consumption in a campus can be analysed based on big

data solutions. Authors indicate that analysis of existing situation with energy consumptions will lead to next step of potential savings of energy. This results in both – energy management and also optimization of resources, in this case energy usage optimization in campus. Research by (Liu et al., 2018) also is categorized into this category and “Resource and waste optimization” since it analyses improvements in energy efficiency in network based on big data analytics.

Authors propose to keep this category as separate for big data applications in construction industry since energy as a resource increases in importance together with increasing volumes of construction and overall economic development.

All publications that are assigned to this category – Bilal et al. (2016b), J. S. Kim, and B. S. Kim (2018), Liu et al. (2016), Liu et al. (2018).

#### *Other emerging trends triggered by big data*

Any other research connected with big data applications in construction industry can be fit into this category. Therefore, variety of researches that fit in this category is very wide. (Valpeters, Kireev& Ivanov, 2018) research machine learning solution (specific type of big data application) to determine probability of successful contract execution in construction project. This type of solution could be adjusted in any other industry as well if historical data of contracts and actual outcomes is available. (Pradeepa & Manjula, 2016) describe solution of geographical data extraction from different web documents and media that can be visualized and further used in early stages of potential construction projects. (Taylan, Kabli, Porcel, & Herrera-Viedma, 2018) research similar challenge to (Valpeters et al., 2018) – how to build an automatic solution for the best contractor selection in each specific construction situation to have best possible outcome.

This category should be observed very closely to potentially adopt new ground-breaking solutions based on big data since identified solutions from researches indicate very innovative and creative solutions of big data in construction industry.

All publications that are assigned to this category – Bilal et al. (2016b), Lede, Chunyan, and Meiyang (2016), Omran, and Chen (2016), Pradeepa, and Manjula (2016), Taylan et al. (2018), Tian, Wang, Zhou, and Peng (2018), Liu (2016), Valpeters et al. (2018).

## **Conclusions**

Systematic literature overview performed by authors support 5 category usage proposed by (Bilal et al., 2016b) for the big data usage in construction industry. Research show that academic publications selected by the authors that are published in 2016 and onward can be categorized in the same 5 categories and no new categories are identified. Rapid technological development has not caused new big data applications in construction industry in time period from 2016 untill beginning of 2019 that could not be categorized in one of 5 categories.

Authors have identified potential to perform systematic literature overview on larger scale that could result in more granular classification of big data applications in construction industry. Another direction of further analysis is identified in cross-industry big data application research. Such research would allow to identify potential applications from other industries that could be adopted to construction industry.

Existing classification and identified researches indicate innovative applications of big data solutions that are not directly connected with construction process but rather complementing and improving processes around main construction project.

## **References**

- Bilal, M., Oyedele, L. O., Akinade, O. O., Ajayi, S. O., Alaka, H. A., Owolabi, H. A., ..., & Bello, S. A. (2016a). Big Data architecture for construction waste analytics (CWA): A conceptual framework. *Journal of Building Engineering*, 6, 144-156. <https://doi.org/10.1016/j.jobe.2016.03.002>
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., ..., & Pasha, M. (2016b). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, 30(3), 500-521. <https://doi.org/10.1016/j.aei.2016.07.001>
- Chen, G., Wang, E., Sun, X., & Lu, Y. (2016). An intelligent approval system for city construction based on cloud computing and Big Data. *International Journal of Grid and High Performance Computing (IJGHPC)*, 8(3), 57-69. <https://doi.org/10.4018/IJGHPC.2016070104>
- Chi, B., Meng, H., Zhai, K., Zhai, G., & Liu, Y. (2017, August). Big Data in urban construction archives and urban management: Detected underground pipes alignment with urban construction records. In *2017 IEEE International Conference on Big Knowledge (ICBK)* (pp. 173-178). IEEE. <https://doi.org/10.1109/ICBK.2017.53>
- Duan, Y. Y., Wang, X. A., Li, F., & Zhang, J. D. (2018, July). Coupling relationship construction of key elements in emergency intelligence with Big Data. In *Conference on Complex, Intelligent, and Software Intensive Systems* (pp. 1069-1076). Springer, Cham. [https://doi.org/10.1007/978-3-319-93659-8\\_99](https://doi.org/10.1007/978-3-319-93659-8_99)

- Fang, Y., Jiaming, Z., Yaohui, L., & Mei, G. (2016, July). Semantic description and link construction of smart tourism linked data based on Big Data. In *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 32-36). IEEE. <https://doi.org/10.1109/ICCCBDA.2016.7529530>
- Guo, S., & Zhai, L. (2017). Analysis on the construction of computer data processing mode in the era of Big Data. *Agro Food Industry Hi-Tech*, 28(1), 798-802.
- Han, Z., & Wang, Y. (2017). The applied exploration of Big Data technology in prefabricated construction project management. In *ICCREM 2017* (pp. 71-78). <https://doi.org/10.1061/9780784481059.007>
- Kim, J. S., & Kim, B. S. (2018). Analysis of fire-accident factors using Big-Data analysis method for construction areas. *KSCE Journal of Civil Engineering*, 22(5), 1535-1543. <https://doi.org/10.1007/s12205-017-0767-7>
- Konikov, A., & Konikov, G. (2017, October). Big Data is a powerful tool for environmental improvements in the construction business. In *IOP Conference Series: Earth and Environmental Science* (vol. 90, No. 1, p. 012184). IOP Publishing. <https://doi.org/10.1088/1755-1315/90/1/012184>
- Koseleva, N., & Ropaite, G. (2017). Big Data in building energy efficiency: Understanding of Big Data and main challenges. *Procedia Engineering*, 172, 544-549. <https://doi.org/10.1016/j.proeng.2017.02.064>
- Lede, N., Chunyan, L., & Meiying, J. (2016, December). Prediction of construction land in Kunming based on big data. In *2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)* (pp. 147-150). IEEE. <https://doi.org/10.1109/ICITBS.2016.46>
- Liu, Q. A. (2016). A Design of ETL for the Construction of Traffic Network Based on Big Data. *Chemical Engineering Transactions*, 51, 451-456. <https://doi.org/10.3303/CET1651076>
- Liu, R. H., Kuo, C. F., Yang, C. T., Chen, S. T., & Liu, J. C. (2016, November). On construction of an energy monitoring service using big data technology for smart campus. In *2016 7th International Conference on Cloud Computing and Big Data (CCBD)* (pp. 81-86). IEEE. <https://doi.org/10.1109/CCBD.2016.026>
- Liu, X. (2017). The construction of tourism model based on big data statistics. *Agro Food Industry Hi-Tech*, 28(3), 2208-2211.
- Liu, X., Dong, M., Liu, Y., Liu, A., & Xiong, N. N. (2018). Construction low complexity and low delay CDS for big data code dissemination. *Complexity*, 2018, ID 5429546. <https://doi.org/10.1155/2018/5429546>
- Lu, W. (2019). Big data analytics to identify illegal construction waste dumping: A Hong Kong study. *Resources, Conservation and Recycling*, 141, 264-272. <https://doi.org/10.1016/j.resconrec.2018.10.039>
- Lu, W., Chen, X., Ho, D. C., & Wang, H. (2016). Analysis of the construction waste management performance in Hong Kong: The public and private sectors compared using big data. *Journal of Cleaner Production*, 112(Part 1), 521-531. <https://doi.org/10.1016/j.jclepro.2015.06.106>
- Omran, B. A., & Chen, Q. (2016). Trend on the implementation of analytical techniques for big data in construction research (2000–2014). In *Construction Research Congress 2016* (pp. 990-999). <https://doi.org/10.1061/9780784479827.100>
- Pradeepa, S., & Manjula, K. R. (2016, March). Construction of gazetteers from geo big data using machine learning technique on Hadoop. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1619-1622). IEEE.
- Shuo, Q. (2017). Construction of the industry chain of ecological agriculture combined with big data. *Agro Food Industry Hi-Tech*, 28(1), 2499-2503.
- Taylan, O., Kabli, M. R., Porcel, C., & Herrera-Viedma, E. (2018). Contractor selection for construction projects using consensus tools and big data. *International Journal of Fuzzy Systems*, 20(4), 1267-1281. <https://doi.org/10.1007/s40815-017-0312-3>
- Tian, L., Wang, H., Zhou, Y., & Peng, C. (2018). Video big data in smart city: Background construction and optimization for surveillance video processing. *Future Generation Computer Systems*, 86, 1371-1382. <https://doi.org/10.1016/j.future.2017.12.065>
- Valpeters, M., Kireev, I., & Ivanov, N. (2018). Application of machine learning methods in big data analytics at management of contracts in the construction industry. In *MATEC Web of Conferences* (vol. 170, p. 01106). EDP Sciences. <https://doi.org/10.1051/mateconf/201817001106>
- Wei jie, L., Jinna, J., Zhengzheng, L., & Na, C. (2016, July). Research on the big data construction of equipment support. In *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)* (pp. 26-31). IEEE. <https://doi.org/10.1109/ICCCBDA.2016.7529529>
- Xianglan, L. (2017, March). Digital construction of coal mine big data for different platforms based on life cycle. In *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)* (pp. 456-459). IEEE. <https://doi.org/10.1109/ICBDA.2017.8078862>
- Yun, Y. (2016). Construction solutions on data center of PV power systems of Qinghai Province based on big data technologies. *Advances in Energy, Environment and Materials Science*, 53-56.
- Zhang, P., Zhang, Q., Liu, F., Li, J., Cao, N., & Song, C. (2017, July). The construction of the integration of water and fertilizer smart water saving irrigation system based on big data. In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)* (vol. 2, pp. 392-397). IEEE. <https://doi.org/10.1109/CSE-EUC.2017.258>