

Big data in building information modeling research: survey and exploratory text mining

Abstract

It has been argued that Building Information Modeling (BIM) can transform the landscape of Architecture, Engineering, and Construction (AEC) and Facility Management (FM) industries with its potential to reduce cost, project delivery time and increase productivity. Going beyond traditional Computer Aided Design (CAD), BIM has emerged as a data-rich, object-oriented, shared digital representation of a facility that can serve as a reliable basis for decision making across the entire life cycle of a construction project. Simultaneous advancements in big data analytics, storage, as well as information visualization seem to hold the potential to enable truly n-dimensional BIMs integrated with other data-intensive sources such as Geographic Information Systems (GIS), Building Automation Systems (BAS), Energy Management Systems (EMS), etc. A number of recently published articles in the literature seem to indicate that the AEC industry can significantly benefit from big data analytics and architecture and make data-driven decisions considering the volume and variety of information resulting from BIM integration approaches. This paper presents a concise survey of recently published articles highlighting the big data based BIM challenges as well as some exploratory text analysis using bag-of-words text mining, a Natural Language Processing (NLP) technique.

Keywords: big data, analytics, smart infrastructure, resulting, big data analytics, storage

Volume 3 Issue 6 - 2017

Kasthurirangan Gopalakrishnan, Ankit Agrawal, Alok Choudhary

Department of Electrical Engineering & Computer Science, Northwestern University, USA

Correspondence: Kasthurirangan Gopalakrishnan, Senior Research Scientist, Department of Electrical Engineering & Computer Science, Northwestern University, Evanston, IL 60208, USA, Tel 151-545-183-95, Email rangan@northwestern.edu

Received: November 03, 2017 | **Published:** December 19, 2017

Abbreviations: BIM, building information modeling; NLP, natural language processing; BAS, building automation systems; GIS, geographic information systems; CAD, computer aided design; AEC, architecture, engineering and construction

Introduction

Building Information Modeling (BIM) has rapidly grown from merely being a three-dimensional (3D) model of a facility to serving as “a shared knowledge resource for information about a facility, forming a reliable basis for decisions during its life cycle from inception onward”.¹ BIM with three primary spatial dimensions (width, height, and depth) becomes 4D BIM when time (construction scheduling information) is added, and 5D BIM when cost information is added to it. Although the sixth dimension of the 6D BIM is often attributed to asset information useful for Facility Management (FM) processes, there is no agreement in the research literature on what each dimension represents beyond the fifth dimension.² BIM ultimately seeks to digitize the different stages of a building life-cycle such as planning, design, construction, and operation such that consistent digital information of a building project can be used by stakeholders throughout the building life-cycle.³ The United States National Building Information Model Standard (NBIMS) initially characterized BIMs as digital representations of physical and functional aspects of a facility. But, in the most recent version released in July 2015, the NBIMS’ definition of BIM includes three separate but linked functions, namely business process, digital representation, and organization and control.⁴ A number of national-level initiatives are underway in various countries to formally encourage the adoption of BIM technologies in the Architecture, Engineering, and Construction (AEC) and FM industries. Building SMART, with 18 chapters across the globe, including USA, UK, Australasia, etc., was established in

1995 with the aim of developing and driving the active use of open internationally-recognized standards to support the wider adoption of BIM across the building and infrastructure sectors.⁵ The UK BIM Task Group, with experts from industry, government, public sector, institutes, and academia, is committed to facilitate the implementation of ‘collaborative 3D BIM’, a UK Government Construction Strategy initiative.⁶ Similarly, the EUBIM Task Group was started with a vision to foster the common use of BIM in public works and produce a handbook containing the common BIM principles, guidance and practices for public contracting entities and policy makers.⁷

Since data (prices, performance ratings, etc.) is centric to BIM, it is natural that the AEC industry needs to soon come to grips with big data storage, processing, and visualization challenges considering the volume and variety of information managed around BIM and other information sources such as Geographic Information Systems (GIS), Energy Management Systems (EMS), etc. that can be integrated with BIM. A number of previous studies have tried to provide a comprehensive review of the BIM literature from different perspectives.⁸⁻¹¹ In the last few years, a relatively small section of BIM research literature has tried to identify the specific big data challenges in BIM workflows and some have tried to propose solutions. This paper provides a concise survey of this rather narrow range of publications (with the focus on highlighting big-data-BIM issues and challenges) as well as a framework for automating the text analysis of the published literature in this field.

Big data and BIMS: a survey

Most current commercial BIM software are stand-alone systems in the sense that a single computer is used for majority of the computations posing severe restrictions to massive storage, efficient management, sharing, and synchronization of the BIMs

that are growing in size and complexity day by day.¹² To overcome limitations such as computational burden, difficulty in managing BIMs of multiple projects, unfriendly collaborative environment, etc. associated with the use of a stand-alone system for BIM applications, Chen et al.¹² proposed a cloud-based framework (using cloud computing technology) for providing a web-based service for viewing, storing, and processing massive dynamic BIMs. The use of cloud computing technology in the AEC industry is still not mature, although growing fast. Wong et al.¹³ provided a state-of-the-art literature review on the cloud computing technology based BIM research and its implementation in building life cycle management. One important observation was that the building planning/design and construction phases have received more attention in cloud-based BIM research, whereas its application in the operation, maintenance and facility management, energy efficient and demolition stages are quite limited.^{13,14} A similar review on the potential of cloud-based collaboration in the construction industry was provided by Almaatouk et al.¹⁵ which noted that cloud-based BIM is still in the early stages of development mainly due to lack of IT infrastructure and the cost of IT services in handling massive BIMs with big data. Wong & Zhou¹⁶ also noted the need for employing cloud based BIM technology to enable building sustainability management using big data in green BIM. It is clear that the AEC industry is favoring the use of cloud computing technology as the demand for increasing amounts of data in building models continue to grow.¹⁷ The Cloud BIM framework proposed by Chen et al.¹² seems to have addressed the scalability and management issues associated with massive BIMs. The proposed Cloud BIM utilizes the Apache Hadoop Bigtable framework for big data storage using multiple servers in a distributed manner, performs parallel computing and analysis using Map Reduce, and provides real-time online services (such as displaying 3D BIMs in standard web browsers) to multiple users simultaneously.¹² Huang¹⁸ integrated data mining techniques into a cloud-based BIM system to perform online big data analysis on dynamic BIMs. Two kinds of methods have been commonly adopted by the AEC industry for interoperability and collaboration based on BIM, namely domain ontologies and semantic web based methods and methods based on building SMART's Industry Foundation Classes (IFC).¹⁹ However, as the dimensions and consequently the information and data integrated into the BIMs are continuing to increase, the need for automated processing and extraction of desirable and useful information from a massive BIM has become critical for both expert and nonexpert users of the BIM software.¹⁹ Lin et al.¹⁹ proposed a Natural Language Processing (NLP) based approach to intelligent data retrieval and representation for Cloud BIM.¹² In the proposed framework, the user inputs (keyword queries) in natural language are extracted and mapped to IFC entities or attributes through the International Framework for Dictionaries (IFD) for retrieving results from an IFC-structured BIM data model and visualized as per user expectations.¹⁹

Du et al.²⁰ proposed a cloud-based application called BIMCS (Building Information Modeling Cloud Score), using the software as a service (SaaS) model of cloud computing, for benchmarking BIM performance based on the BIM performance big data collected from a wide range of BIM users nationwide. Another emerging technology, namely 3D laser scanning and photogrammetry, capable of capturing huge quantities of 3D data about an object has shown significant potential in creating as-built BIM of a facility.^{21,22} Barazzetti et al.²³ presented an innovative procedure for creating BIM objects with parametric intelligence from 3D laser scan point clouds of complex

architectural features which cannot be handled by commercial BIM software. The proposed semi-automated procedure utilized Non-Uniform Rational Basis Splines (NURBS) to generate advanced BIM models of architectural objects with irregular shapes from point clouds.²³ Even for facilities with regular shapes and objects, significant manual labor is invested in converting raw point cloud data sets (PCDs) to BIM descriptions to generate as-built BIMs.²⁴ Zhang et al.²⁵ proposed a novel sparsity-based optimization-based algorithm for automatically extracting planar patches from large-scale, noisy raw PCDs, thus drastically reducing the cost of generating as-built BIMs. In fact, the need for a Scan-to-BIM data standard in the near future was identified as one of the grand challenges in Visualization, Information Modeling, and Simulation (VIMS) for the construction industry.²⁶ As mentioned previously, the application of BIM in building operation and management phases of the life cycle is currently limited despite the big data collected using sensors²⁷ and Building Management Systems (BMS) during the operation phase of a building. Oti et al.¹⁴ proposed a framework for integrating building energy consumption data into BIM for providing feedback and improving design and facility management. Going beyond just the operation phase, Yuan & Jin²⁸ also proposed a framework for carrying out a full life cycle assessment of building energy consumption using BIM, big data, and cloud computing technologies. Pasini et al.²⁹ suggested the adoption and integration of an Internet of Things (IoT) framework into BIM for collecting real-time information during the building use phase that would fill the current gaps in the operation and management of cognitive buildings. Xie et al.³⁰ employed BIM simulations and big data science to understand challenges of low-load homes with respect to high performance ventilation systems and indoor air quality strategies. Stonecipher & Williams³¹ discussed best practices and current technology concerning high-performance building design using BIM, data capture and management and highlighted interoperability issues across multiple data authoring, analysis and management platforms.

Considering the large proportions of waste generated by the construction industry, a number of studies have suggested the idea of waste prevention, i.e., to consider construction waste at the design stage prior to physical construction, rather than to consider it down the pipeline after it has been generated.^{32,33} Bilal et al.³² identified and discussed critical features of BIM that could be leveraged to implement a data-driven construction waste prediction and minimization plug-in at the design stage. Considering the large datasets associated with construction waste, Bilal et al.³² proposed the use of big data technologies for handling massive materials database, and the use of graph-based representation, analysis, and visualization for handling complex spatio-temporal multi-dimensional data. Going further, Bilal et al.³⁴ proposed the first big data architecture (Neo4J, a graph database, combined with Spark, a fast big data processing engine) for construction waste analytics as an extension for BIM and validated it using 200,000 waste disposal records from 900 completed projects. Botton et al.¹² proposed a big data conceptual pipeline for bridging the gap between BIM-based related visualization works and the information visualization domain after identifying the following major big data challenges in the construction industry: acquiring data, choosing the IT architecture, shaping and coding data, reflecting information and interacting with it. The volume of visual data collected at a construction site is ever-increasing, thanks to smart devices, wearables, and advances in camera-equipped Unmanned Aerial Vehicles (UAVs). However, their real value can

be tapped only when they are localized with respect to BIM, leading to accurate documentation of as-built status.³⁵ An exploratory study by Han & Golparvar-Fard³⁵ investigated the current potential of big visual data and BIM in construction performance monitoring. Industry 4.0 technologies, like augmented and virtual reality, have not yet reached the market maturity for widespread adoption by construction companies into their BIM workflows.³⁶ Correa³⁷ argued that the current use of BIM restricted to planning and design of building infrastructure alone does not warrant the use of big data analytics, however, when BIM is applied fully to FM across the entire life cycle or when integrated with Geographic Information System (GIS) to represent Smart Cities, big data analytics would be the right choice. Considering the large scale data that goes into n-dimensional BIMs, the wide range of users, and the challenges in delivering such data at high-performance rates and in a suitable format, Pauwels et al.³⁸ outlined a performance benchmark for querying and reasoning over large scale building data sets. In fact, big data sources shared across a construction project life cycle, including querying and compression, was considered the second-most-important grand information modeling challenge in the construction industry based on a recent survey conducted by the VIMS committee expert task force of the American Society of Civil Engineers (ASCE) Computing and Information Technology Division.

Exploratory text analysis: problem definition

Parallel and significant recent advances in big data and BIM technologies offer an opportunity to mine insights into big data-driven BIM processes and challenges from published literature. Text mining uses NLP and analytical methods to derive high-quality information and insights from text.³⁹⁻⁴³ The goal of this study is to identify and assemble recently published literature focusing on the big-data-BIM issues and challenges, apply text mining techniques to it, and try to derive some meaningful insights. Since the big data-driven BIM research is still in its infancy, it should be expected that the insights derived from the narrow range of publications may be limited in scope. The text mining methodology used in this study followed a three-step process⁴¹ and was implemented in the R programming environment:⁴⁴ establish corpus, preprocess data, and extract knowledge. The same methodology was employed on two different case studies. In the first case study (abstract corpus), the abstracts of the articles were used as the only source of information. Titles and keywords were omitted since the abstracts are likely to include this information. Also, keywords are often set by the authors and they could be terms that the authors would like their articles to be associated with rather than what is actually contained in the articles.⁴¹ In the second case study (full-text corpus), full-text articles (where available) were used as the source of information.

Establish corpus

The first step involved collecting relevant academic papers for our analysis using academic search engines like Science Direct, Google Scholar, etc. Since the results of text mining is highly dependent on the quality of data collected, great care was taken to collect only papers that were relevant to the main focus of this study, namely BIM-related papers which also discussed big data aspects to different degrees. Search terms included “BIM big data”, “Building Information Modeling big data”, etc. After filtering out irrelevant records, the finalized collection included 26 publications (2 book chapters, 8 conference papers, 14 journal articles, 1 magazine article, and 1 thesis). Among

the 14 journal publications, 4 were from Automation in Construction and 2 from Computer-Aided Civil and Infrastructure Engineering. Others were from Photogrammetric Record, Journal of Construction Engineering and Management, Journal of Building Engineering, Computers in Industry, International Journal of Sustainable Building Technology and Urban Development, Corporate Real Estate Journal, The Journal of Information Technology in Construction, and Journal of Computing in Civil Engineering. Citations (metadata) and full-text articles (where available) were imported using Zotero, an open-source reference management software that “collects all your research in a single, searchable interface.” A quick profiling of the articles by publication year (until 2016) (Figure 1) showed that interest and research in this area is growing exponentially over just the last couple of years. Similar analysis by the first author’s country of origin showed that both USA and UK are at the forefront followed by China (Figure 2). As mentioned previously, two corpora were established: the abstract corpus and the full-text corpus. While the abstract corpus included the abstracts of all 26 publications, the full-text corpus included the full-text of 24 publications (access to a journal article published in Corporate Real Estate Journal³¹ and a thesis document¹⁸ were not available to the authors).

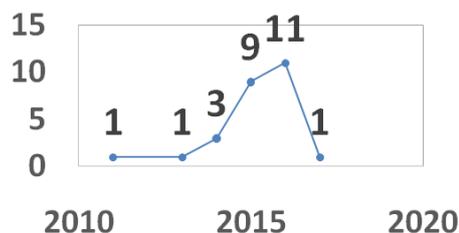


Figure 1 Number of articles by publication year.



Figure 2 Articles by country of origin.

Preprocess text data

After establishing the corpus, it needs to be cleaned up so that meaningful and actionable insights can be distilled from it. The following common preprocessing functions were applied to both the abstract and full-text corpus during text mining:⁴⁴

- Make all text lowercase - helpful for term aggregation
- Remove punctuation like periods and exclamation points – more helpful for preprocessing social media text
- Remove numbers – helpful if the study is not focused on quantitative outcomes
- Remove excess white space

- e. Remove common stop words like “the”, “of”, “was”, “from” etc. which are frequent, but provide little information. In addition to these common stop words, other domain-specific stop words like “BIM”, “Building Information Modeling”, “big data”, “building”, “proposed”, etc. (in this study) were also pulled out since leaving them in the corpus will not add any new insight and lead to over-emphasis of these terms in a frequency analysis.

An important step in preprocessing the text in bag of words text mining is to aggregate similar terms. For instance, “complicatedly”, “complicated”, and “complicate” should be considered as one term rather than three different terms. This is accomplished by applying the stemming function which reduces the words to its root (for instance, “complicatedly”, “complicated”, and “complicate” will be reduced to “complic” and thus be recognized as the same word). After cleaning the corpus, a term-document matrix (TDM) or its transpose document-term matrix (DTM) is created to conduct word frequency analysis and for further processing.⁴⁵ In a TDM, the rows represent all the unique terms identified in the corpus, the columns are the documents, and the cells represent the frequency of each word for each document. DTM is obtained when TDM is transposed. A DTM is more useful when comparing authors (of the documents) within rows or when you want to preserve the time series when the documents are listed chronologically. On the other hand, a TDM is more useful when there are more terms than documents or authors as in language analysis. The first form of TDM, especially when there is a large number of a document, will be quite sparse and thus it will be computationally intensive to process such a huge matrix. After applying some transformations to remove sparse terms, a bar plot of the frequently used words (that appear at least 15 times) in the abstract corpus is displayed in Figure 3. Similar operations on the full-text corpus yields Figure 4 with words appearing at least 400 times in the full-text corpus. Similar terms are captured as being frequently occurring in both the abstract and full-text corpus. In addition, terms like “cloud”, “comput”, and “analysi” appear more frequently in the full-text corpus. After the initial word frequencies are computed, additional transformations such as the following can also be applied to reduce the dimensionality of TDM to a manageable size as well as to aggregate the extracted information:⁴⁵ log frequencies, binary frequencies, and inverse document frequencies.

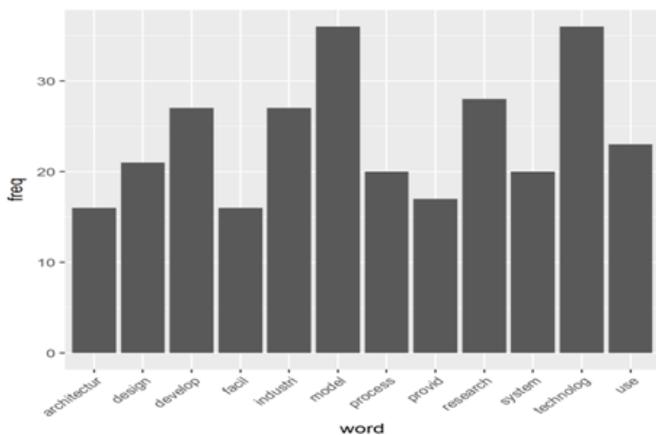


Figure 3 Frequency of words that occur at least 15 times in the abstract corpus.

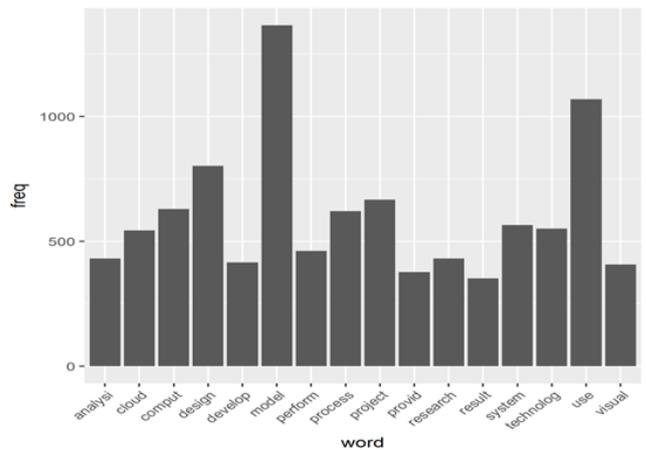


Figure 4 Frequency of words that occur at least 400 times in the full-text corpus.

Extract knowledge

A more popular form of the frequency plot that is more visually engaging is a word cloud where size is typically relative to frequency. In more specialized word clouds, colors may indicate another measurement. The choice of stop words in the text preprocessing stage will have impact on the word frequency plot as well as the word cloud. For instance, in this study, if we did not choose “BIM” and “big data” as stop words, they would have shown up as the largest words in the word cloud and thus mask underlying insights. Figure 5 displays a word cloud of the abstract corpus with words that occur at least 10 times. It is seen that words like “technolog”, “model”, “perform”, “research”, and “develop” appear more frequently than other terms. A word cloud of the frequently occurring words (at least 100 times) in the full-text corpus is displayed in Figure 6. So far, we focused on creating and analyzing TDM or DTM using single words. We could also create tokens containing two or more words which can help us extract useful phrases from the text leading to some additional insights. For example, “proposed” and “framework” as separate words can have different meaning compared to “proposed framework”. This is called n-gram to kenization, where n can be uni, bi, tri, etc. It should be noted that increasing the tokenization or n-gram length will increase the TDM or DTM size.



Figure 5 A word cloud of words that occur at least 10 times in the abstract corpus.

Table 1 Comparison of RAKE extracted keywords to manually assigned keywords for selected abstracts

Reference	Extracted by RAKE	Manually assigned
Barazzeti et al. ²³	Parametric building information modeling, uniform rational basis splines, geometry involving spatial relationships, advanced parametric representation, complex architectural features	BIM, building information modeling, interoperability, NURBS, point cloud
Bilal et al. ³²	Waste intelligence based waste management software, big data based waste analytics architecture, big data based simulation tools, highly resilient graph processing system, construction waste analytics, sophisticated big data technologies	Construction waste, big data analytics, Building information modeling (BIM), design optimization, construction waste analytics, Waste prediction and minimization
Wong et al. ¹³	Building life cycle management, refereed journal articles, building information management, cloud computing technology, facility management, building planning	Cloud computing, BIM, construction sector, building life cycle
Leite et al. ²⁶	VIMS grand challenges, investigate current practices, facility management industries, expert task force, survey results, future research directions, effective decision making	Visualization, information modeling, simulation, architectural engineering, construction engineering, facility management

Summary and conclusion

A number of recently published articles in the literature seem to indicate that the AEC industry can significantly benefit from big data analytics and architecture and make data-driven decisions considering the volume and variety of information resulting from BIM integration approaches. In the last few years, a relatively small section of BIM research literature has tried to identify the specific big data challenges in BIM workflows and some have tried to propose solutions. This paper explored this rather narrow range of publications through a concise survey and exploratory text analysis to highlight key big-data-BIM issues, challenges, and solutions. This is a highly emerging area of research and the number of peer-reviewed journal articles focusing on the various aspects of this research topic considered in this paper was relatively fewer in number compared to most survey papers where the number of articles reviewed runs in hundreds. Consequently, the results of the exploratory text mining do not appear to provide very useful insights regarding the evolving research themes and future challenges.

The following are some significant conclusions from both the state-of-the-art survey and exploratory text mining:

- i. Big data sources shared across a construction project life cycle, including querying and compression, is considered the second-most-important grand information modeling challenge in the construction industry based on industry survey.
- ii. The use of cloud computing technology in the AEC industry, as one solution to address the big-data-BIM challenge (ex., for performing online processing of massive dynamic BIMs), is still not mature, although growing fast. In general, building planning/design and construction phases have received more attention in cloud-based BIM research rather than the operation, maintenance and facility management phases.
- iii. NLP-based techniques have been proposed for automated processing and extraction of desirable and useful information from a massive BIM.
- iv. One source of big data in BIM is the use of acquired 3D laser scan point clouds of a facility to generate as-built BIMs. Semi-automated and automated methods have been proposed to reduce the cost of generating as-built BIMs. There is a critical need for a Scan-to-BIM data standard.

- v. Considering the potential for collecting massive data sets using sensors, BMS, etc. during the building in-use phase, studies have proposed computational big data science based frameworks on how to make the best use of this big data in conjunction with BIM, especially in performing a full life cycle building energy consumption analysis.
- vi. Advances have been made in BIM big data visualization domain, with respect to construction monitoring and waste analytics.
- vii. Model, industry, technology, development, research, and performance are some key terms that seem to characterize the recent literature concerning the big-data-BIM nexus based on the 26 publications analyzed.
- viii. The bag-of-words text mining applied on the abstract corpus produced results almost similar to that applied on full-text corpus. Thus, it may be sufficient to analyze just the research abstracts (without keywords) to identify major themes and the overall direction of the research field without having to analyze the full-text documents. With this approach, a significant number of abstracts can be analyzed to identify the evolution of themes with time and within specific journals of interest.
- ix. As more number of publications becomes available, future research can apply advanced text mining techniques like Latent Dirichlet Allocation (LDA) to extract topics underlying the text.

Funding

None.

Acknowledgments

This work is supported in part by the following grants: NSF award CCF-1409601; DOE awards DE-SC0007456, DE-SC0014330; AFOSR award FA9550-12-1-0458; NIST award 70NANB14H012.

Conflicts of interest

None.

References

1. National Institute of Building Sciences. National BIM Standard-United States(R) Fact Sheet. Washington, USA; 2015.

2. Boton C, Halin G, Kubicki S, et al. Challenges of Big Data in the Age of Building Information Modeling: A High-Level Conceptual Pipeline. In: Cooperative Design, Visualization, and Engineering. *Springer International Publishing*. 2015:48–56.
3. Eastman C, Teicholz P, Sacks R, et al. Wiley: BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors. 2nd ed. USA: John Wiley & Sons; 2011. p. 648.
4. Building SMART alliance. National BIM Standard - United States(R) Version 3 - Transforming the Building Supply Chain Through Open and Interoperable Information Exchanges [Internet]. Washington, USA; 2017.
5. Building SMART. 2016.
6. BIM Task Group - A UK Government Initiative, UK; 2013.
7. EU BIM Task Group. 2016.
8. Bradley A, Li H, Lark R, et al. BIM for infrastructure: An overall review and constructor perspective. *Automation in Construction*. 2016;71(2):139–152.
9. Chen K, Lu W, Peng Y, et al. Bridging BIM and building: From a literature review to an integrated conceptual framework. *International Journal of Project Management*. 2015;33(6):1405–1416.
10. Volk R, Stengel J, Schultmann F. Building Information Modeling (BIM) for existing buildings-Literature review and future needs. *Automation in Construction*. 2014;38:109–127.
11. Yalcinkaya M, Singh V. Patterns and trends in Building Information Modeling (BIM) research: A Latent Semantic Analysis. *Automation in Construction*. 2015;59:68–80.
12. Chen HM, Chang KC, Lin TH. A cloud-based system framework for performing online viewing, storage, and analysis on big data of massive BIMs. *Automation in Construction*. 2016;71:34–48.
13. Wong J, Wang X, Li H, et al. A review of cloud-based BIM technology in the construction sector. *The Journal of Information Technology in Construction*. 2014;19:281–291.
14. Oti AH, Kurul E, Cheung F, et al. A framework for the utilization of Building Management System data in building information models for building design and operation. *Automation in Construction*. 2016;72(2):195–210.
15. Almaatouk Q, Othman MSB, Al-khazraji A. A review on the potential of cloud-based collaboration in construction industry. In: 3rd MEC International Conference on Big Data and Smart City (ICBDSC); 2016. p. 1–5.
16. Wong JKW, Zhou J. Enhancing environmental sustainability over building life cycles through green BIM: A review. *Automation in Construction*. 2015;57:156–165.
17. Robins S. The Future of BIM. *Heating/Piping/Air Conditioning Engineering*. 2013;85(5):14–19.
18. Huang TW. Application Case of Big Data Analysis on Dynamic BIMs with Parameters for Use Scenarios. National Taiwan University of Science and Technology, Taiwan. 2016.
19. Lin JR, Hu ZZ, Zhang JP, et al. A Natural-Language-Based Approach to Intelligent Data Retrieval and Representation for Cloud BIM. *Computer-Aided Civil and Infrastructure Engineering*. 2016;31(1):18–33.
20. Du J, Liu R, Issa R. BIM Cloud Score: Benchmarking BIM Performance. *Journal of Construction Engineering and Management*. 2016;140(11).
21. Fan Y, Shi Y. BIM combined with the 3D laser scanning technology to lead the changes of the design industry in the future. In: Yang S, editor. Proceedings of the 4th International Conference on Information Technology and Management Innovation. Paris: Atlantis Press; 2015. p. 514–517.
22. Tang P, Huber D, Akinci B, et al. Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in Construction*. 2010;19(7):829–843.
23. Barazzetti L, Banfi F, Brumana R, et al. Creation of Parametric BIM Objects from Point Clouds Using Nurbs. *Photogrammetric Record*. 2015;30(152):339–362.
24. Backes D, Thomson C, Malki-Epshtein L, et al. Chadwick Green BIM: Advancing Operational Understanding of Historical Buildings with BIM to Support Sustainable Use. In: Proceedings of the 2014 Building Simulation and Optimization Conference, USA; 2014.
25. Zhang G, Vela PA, Karasev P, et al. A Sparsity-Inducing Optimization-Based Algorithm for Planar Patches Extraction from Noisy Point-Cloud Data. *Computer-Aided Civil and Infrastructure Engineering*. 2015;30(2):85–102.
26. Leite F, Cho Y, Behzadan A, et al. Visualization, Information Modeling, and Simulation: Grand Challenges in the Construction Industry. *Journal of Computing in Civil Engineering*. 2016;30(6).
27. Khan A, Hornbaek K. Big Data from the Built Environment. In: Proceedings of the 2nd International Workshop on Research in the Large. New York, NY, USA; 2011. p. 29–32.
28. Yuan Y, Jin Z. Life Cycle Assessment of Building Energy in Big-Data Era: Theory and Framework. In: 2015 International Conference on Network and Information Systems for Computers; 2015. P. 601–605.
29. Pasini D, Ventura SM, Rinaldi S, et al. Exploiting Internet of Things and building information modeling framework for management of cognitive buildings. In: 2016 IEEE International Smart Cities Conference (ISC2); 2016. p. 1–6.
30. Xie H, Liang T, Li H, et al. Understanding Air Quality Challenges Through Simulation and Big Data Science for Low-Load Homes. In: Intelligence Science and Big Data Engineering Big Data and Machine Learning Techniques. *Springer International Publishing*. 2015:595–602.
31. Stonecipher D, Williams G. High-performance design, big data and energy management: Three critical aspects of smart buildings. *Corporate Real Estate Journal*. 2016;5(3):276–285.
32. Bilal M, Oyedele LO, Qadir J, et al. Analysis of critical features and evaluation of BIM software: towards a plug-in for construction waste minimization using big data. *International Journal of Sustainable Building Technology and Urban Development*. 2014;6(4):211–228.
33. Cheng JCP, Ma LYH. A BIM-based system for demolition and renovation waste estimation and planning. *Waste Management*. 2013;33(6):1539–1551.
34. Bilal M, Oyedele LO, Akinade OO, et al. Big data architecture for construction waste analytics (CWA): A conceptual framework. *Journal of Building Engineering*. 2016;6:144–156.
35. Han KK, Golparvar-Fard M. Potential of big visual data and building information modeling for construction performance analytics: An exploratory study. *Automation in Construction*. 2017;73:184–198.
36. Oesterreich TD, Teuteberg F. Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. *Computers in Industry*. 2016;83:121–139.

37. Correa F. Is BIM Big Enough to Take Advantage of Big Data Analytics? In: Proceedings of the 32nd ISARC Conference. Oulu, Finland, 2015. p. 1–8.
38. Pauwels P, de Farias TM, Zhang C, et al. Querying and Reasoning over Large Scale Building Data Sets: An Outline of a Performance Benchmark. In: Proceedings of the International Workshop on Semantic Big Data. New York, NY, USA; 2016.
39. Ishikiriya CS, Miro D, Gomes CFS. Text Mining Business Intelligence: A small sample of what words can say. *Procedia Computer Science*. 2015;55:261–267.
40. Kumar BS, Ravi V. A survey of the applications of text mining in financial domain. *Knowledge-Based Systems*. 2016;114:128–147.
41. Miner GD, Delen D, Elder J, et al. Practical Text Mining and Statistical Analysis for Non-structured Text Data Application. Boston: Academic Press; 2012.
42. Paul A, Agrawal A, Liao W, et al. Anony Mine: Mining anonymous social media posts using psycho-lingual and crowd-sourced dictionaries. In: 22nd ACM SIGKDD Workshop on Sentiment Discovery and Opinion Mining; 2016.
43. Polepeddi L, Agrawal A, Choudhary A. Poll: A Citation Text Based System for Identifying High-Impact Contributions of an Article. In: 2011 IEEE 11th International Conference on Data Mining Workshops. 2011. p. 965–958.
44. Williams G. Hands-on Data Science with R Text Mining. 2017.
45. Delen D, Crossland MD. Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications*. 2008;34(3):1707–1720.
46. Hu YH, Chen YL, Chou HL. Opinion mining from online hotel reviews – A text summarization approach. *Information Processing & Management*. 2017;53(2):436–449.
47. Rose S, Engel D, Cramer N, et al. Automatic Keyword Extraction from Individual Documents. In: Berry MW, Kogan J, editors. Text Mining. USA: John Wiley & Sons, Ltd; 2010. p. 1–20.