BIG DATA APPLICATIONS IN BUILT ENVIRONMENT: TOWARDS A USE FOUNDATION MODEL

Mohamad Kassem[1] and Jennifer Li[2]

Abstract: The term 'big data' is increasingly permeating the current debate over the present and the future of our built environment. There are heightened expectations about the role big data may or can play in enabling new applications and decisions across the whole lifecycle of assets, from early design phase through construction to operation and end-of-life. It is therefore necessary to analyse the state-of-the-art in big data applications in built environment. This paper presents a structured and systematic literature review on this emerging theme. The results are quantitatively analysed using network and clustering techniques.

The research identified distinct clusters within two lenses: (1) the project lifecycle (i.e., design/planning, construction, and operation), and (2) the built environment hierarchy scale (i.e., component/product, building/facility, site, neighbourhood, region and city). Six clusters of big data applications across the two lenses were identified: energy management, traffic management, design & planning, onsite construction activities, city-wide services, and smart cities.

The clusters show the level of research intensity around each of their nodes and the connections between the nodes which indicate the emergence of new avenues of investigations. These results represent the first steps towards developing a descriptive 'Use Foundation model' for big data applications in built environment.

Keywords: Built Environment, Big Data, Internet of Things, and Smart Data.

1 INTRODUCTION

The built environment encapsulates the entire natural and the artificial assets that humans have constructed for their use. It includes villages, towns and cities, infrastructure (i.e. railways, roads, bridges, energy and utility networks), and public spaces such as parks and other urban spaces. The built environment is increasingly pervaded by sensors that track almost everything around its users and constituents enabling the generation of significant amounts of data. This growing generation and availability of data has instigated industry and academia to investigate how such data can be leveraged to simultaneously explore new opportunities and address emergent built environment challenges. Strenuous discussions around 'data revolution' brought a range of new terms and concepts: 'big data', 'smart data', 'smart world', 'Internet of Things' (IoT), and 'linked data'. However, the understanding of the impact this data revolution on our built environment is still limited and there is a dearth of structured reviews on this topic. This paper aims to address this gap by presenting a systematic literature review on the applications or uses of big data in built environment. Frist, the

---

[1]   Associate Professor, Department of Mechanical and Construction Engineering, Faculty of Engineering and Environment, Northumbria University at Newcastle, Newcastle, UK, mohamad.kassem@northumbria.ac.uk

[2]   Masters Student, School of Science and Engineering, Teesside University, Middlesbrough, UK, jenniferli83@gmail.com

paper will clarify the different terms and present a 'visual definition' of Big data. Then, it will illustrate and analyse the results from the systematic literature review using quantitative methods such as network analysis and clustering.

2 KEY TERMS AND DEFINITIONS

Data can be descriptions, counts, or measures, of anything, in any format. Initially, the term 'big data' referred to large data sets that traditional technologies were not sufficient to collect, store and process (Mishra 2016). There is still no universally accepted definition for Big data. According to Boyd and Crawford (2012) Big data is defined as a cultural, technological, and scholarly phenomenon that rests on the interplay of: (1) Technology: maximizing computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets; (2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims, and (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy. Big data is broadly used for two purposes, for exploration and description, to understand trends in current and historical data; and for predictive purposes, using current and historical data to help predict future trends and make decisions in advance (Bilal et al. 2016). In recent years, closely related terms such as 'smart data' and 'Internet of Things' emerged. According to (Pavliscak 2014, pp. 13), smart data is data with a layer of intelligence. Smart data harnesses the challenges posed by volume, velocity, variety and veracity of big data to provide actionable information that improve decision making (Sheth 2013, pp. 17-18, and Lenk et al. 2015, pp. 1749). The IoT focuses on the ability of all devices to communicate with one another or with humans, either in real life or in the virtual world (Isikdag 2015). IoT gives objects the ability to see and hear things that are constantly taking place around it, and at the same time, communicate that data to a processing centre for analysis (Rathore et al. 2016).

A review of the numerous available definitions for big data (i.e., Al Nuaimi et al. 2015; Chandler 2015; Gandomi and Haider 2015; Khan et al. 2014; Kitchin 2014; Marr 2015; Mishra et al. 2016; Nativi et al. 2015; Qi and Wayne 2014) was performed. The only difference between the available definitions was the inclusion of further Vs (e.g. Volume, Velocity, Variety, etc.) as distinct characteristics of big data. Following the mapping of all available definitions against all Vs, this research presents the nine Vs model as a visual definition for big data (Figure 1). The nine Vs are explained in Table 1.

Some further characteristics of big data were identified in other authors' definitions but were not included in the nine Vs model. For example, Kitchin (2014) adds the characteristics of exhaustivity; resolution and indexicality, and relationality and flexibility. However, upon the examination of these characteristics, it was found that these either overlap or derive from the nine Vs. Therefore, it can be argued that the nine Vs model exhaustively represents the 'current' concept of 'Big data' and congregate all possible characteristics. However, it should be noted that the relative importance of the characteristics can vary between different applications and uses of big data.
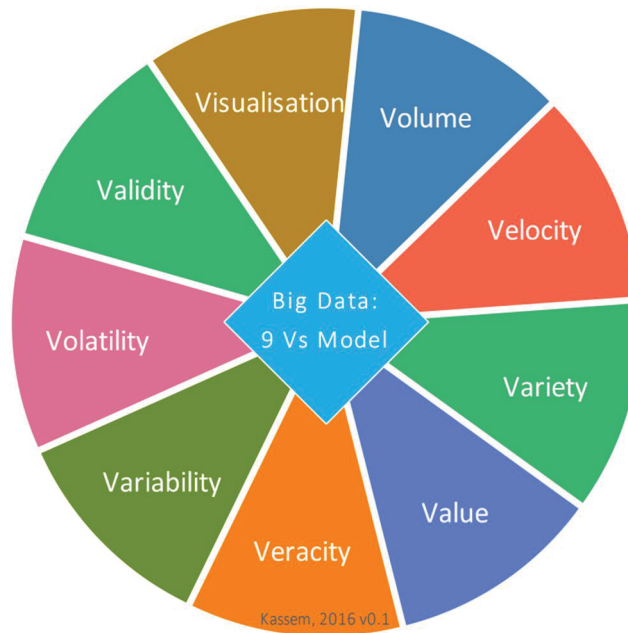
Figure 1: The nine Vs Model: a visual definition of Big data

Table 1: The nine Vs of Big data.

Volume	Size of the big data set.
Velocity	Speed at which data is generated and requires processing.
Variety	Numerous formats in which data exists (e.g. text, images, audio, videos, messages, tweets, sensor data) resulting into heterogeneous data that is often unstructured, uncategorised and unordered.
Veracity	Authenticity of the data as significant amount of data within the big data is often anonymous and unverified hence, affecting the significance of the results in aiding in the decision-making process.
Value	Value (e.g. economic, social, environmental) attainable from collecting, storing and processing the data.
Variability	Temporal and spatial variances in the streams of incoming data (e.g. ebbs and flows) inducing a complexity in connecting, matching, cleansing and transforming the data received from different sources.
Volatility	Policy issues affecting the retention of the data collected from different sources (e.g. retention time, access rights, cyber-security) that could significantly affect the maintenance cost.
Validity	Correctness and accuracy of the data in relation to the anticipated use – it differs from above 'Veracity' that addresses the authenticity of data.
Visualisation	Visual presentation of data that would otherwise be difficult to use in a meaningful and informative way in decision-making.

3   METHODOLOGY

A modular view (Figure 2) of the built environment was adopted to support the objective of exploring the uses of big data at each level within the built environment hierarchy.
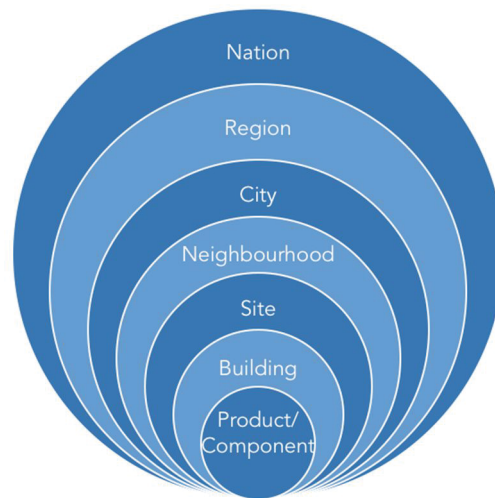
Figure 2: Built environment scale (adapted from BIM Framework, 2016)

A systematic literature review was utilised to identify existing studies on big data uses in built environment. The systematic review helps to reduce the natural tendency of researchers to be swayed by potential biases and provides "a sort of scientific gyroscope with an in-built self-righting mechanism" (Petticrew and Roberts, pp. 5-6). The search strings (Table 2) were applied on the Scopus online database which was selected because it contains the largest number of peer-reviewed academic journals of all online databases in the field of science, technology and engineering. The initial search returned 1,924 papers. After executing the entire process of the systematic literature review (i.e., exclusion of non-journal articles; exclusion of papers published prior to year 2012; exclusion of paper in other than English language; exclusion of non-physical science papers; exclusion of keywords 'computing infrastructure' and 'computing architecture'; exclusion of non-related subject areas; exclusion based on abstract review, and exclusion based on introduction/conclusion review), 71 articles subsisted and passed into the subsequent stage of data extraction and analysis.

The identified studies were analysed using a combination of bibliometric analysis, network analysis and clustering. The bibliometric analysis is generally used to evaluate the qualitative and quantitative interest in a specific field and to describe key patterns. The network analysis provides a quantitative understanding of a series of objects that are interconnected. Data clustering places together sets of articles that share same characteristics. To enhance the interpretation and analysis of results, Gephi was selected as a network analysis tool because: (1) it has the ability to demonstrate links between nodes based on spatial relationships, and (2) it provides flexible visual representations of relationships (Jacomy et al. 2014). The remainder of the paper will focus on profiling big data research in built environment across the different clusters.

Table 2. Search terms for the systematic literature review

Big data	Big data, Big data analytics, Analytics, Smart data, Datafication
Built environment	Built environment, Construction, Sustainability, Civil engineering, Structural engineering, AEC, Architecture, Buildings, Roads, Bridges, Rail, Engineering
Big data applications	Smart cities, Smart government, Smart grid, Energy management, Utilities, Integrated services, Community ownership, Infrastructure planning, Infrastructure, Traffic management, Public safety, Crowd sensing, Crowdsourcing, Building information modelling, Building information modeling, BIM, Building management systems, BMS, Building automation systems, BAS
	Building energy management systems, BEMS, Social networking services, SNS, Early warning systems, Early warning

4   RESULTS AND ANALYSIS

This section addresses the profiling of the identified articles using two knowledge lenses: (1) the phases of the asset lifecycle (i.e., design and planning, construction, and operation); and (2) the built environment hierarchy scale (i.e., product/component, building/facility, site, neighbourhood, city and region). Distinct clusters of big data applications (Figures 3 to 5) across the two lenses are presented and discussed. In these figures, the nodes represent the entities within the dataset. Their size is a measure of the node degree (e.g. the larger a node is, the larger is its degree). The edges represent connections or relationships between the nodes.

4.1   Big data across the asset lifecycle

The identified big data research clusters across the asset lifecycle are illustrated in Figure 3. The results show that big data research is more focussed on the operation phase than on the design and construction phases. However, there is an increasing stream of big data research connecting the design and planning phase to the operation phase. This outcome will be reconfirmed - in section 4.3 - where big data applications are profiled and described.
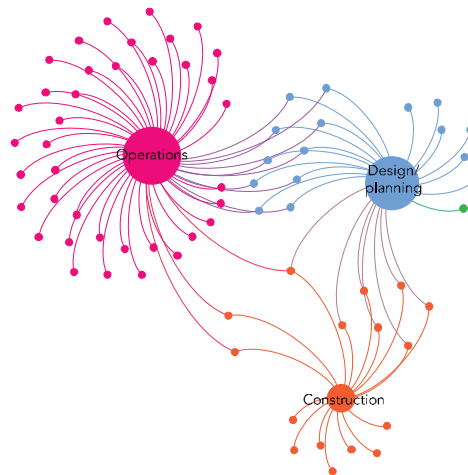

Figure 3: Clusters and spatial distribution of big data research across the project lifecycle

4.2 Big data across the Built Environment hierarchy

The profiling of big data research across the built environment hierarchy is illustrated in Figure 4. There is research across every level of the built environment hierarchy with the highest concentration being at city level and building/facility level. There is also an increase in research effort aimed at integrating the component, building/facility, neighbourhood, and city levels. Site and region have much fewer links with the remaining nodes. Their distance away from the centre demonstrates that they are less integrated with other big data applications in built environment.
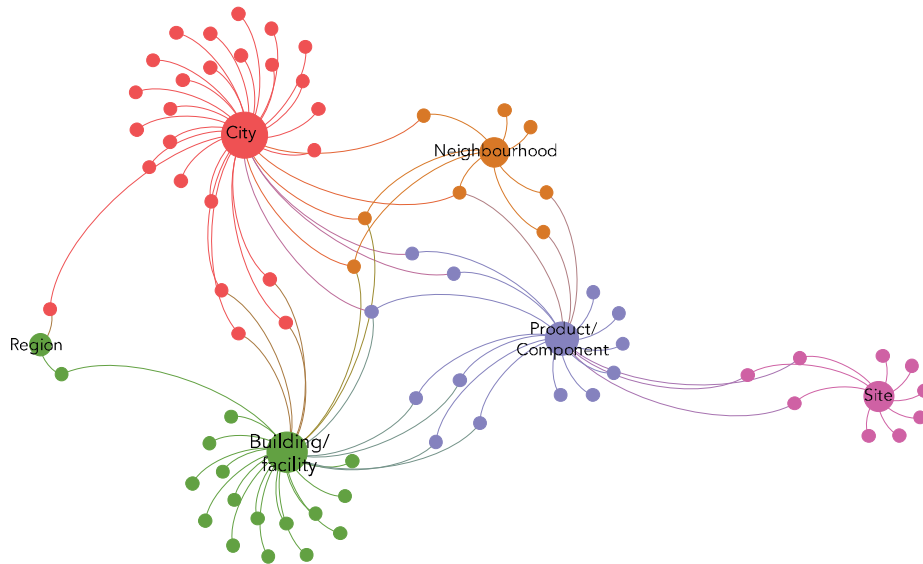

Figure 4: Clusters and spatial distribution of Big data research across the built environment hierarchy

4.3 Big data uses in built environment

The identified six clusters of big data applications in built environment (Figure 5) are: energy management, traffic management, design/planning use, onsite construction, city-wide services, and smart cities. Applications in energy management and traffic management dominate the current research agenda of big data applications in built environment. Onsite construction ranks third however, its corresponding cluster shifts away from the other clusters - a situation referred to as isolation - and have fewer links with nodes of other clusters. The top application area (i.e. energy management) is briefly illustrated in this section.

Big data research in energy management was found in three sub-areas: energy [waste] management, smart grids, and smart buildings. In energy waste management, the common objective between the available studies is the identification of energy wastes and the implementation of ways to achieve energy efficiencies. To achieve this objective, researchers analyse patterns in consumption big data and identify anomalies (Janetzko et al. 2013). Techniques such as non- intrusive load monitoring (Gillman et al., 2014; Paris et al., 2014); non-linear usage patterns (Grolinger et al. 2016), analytics and benchmarking (Pol et al. 2012; Park et al. 2015, Zhu et al. 2016), and predictive analytics (Vaghefi et al. 2015) were used. In smart grids applications, big data, enabled by the ability to take meter readings at short time intervals as opposed to the traditional monthly readings, is allowing improved predictions and planning of power generation needs and daily usages (Aman et al. 2015). Big data related topics for this application

include: smart meters (Balac. 2013), statistical learning tools (Kekatos et al. 2014), integration with renewable energy sources (Balac. 2013) and open service cloud for integration of different 'smart' things (Deindl et al. 2015). In the smart buildings area, big data were found in applications such as: dynamic energy management enabled through real-time data reading, benchmark, visualisation and optimisation (Hong et al. 2015, pp. 703); providing comfort, usability and endurance through the amalgamation of data about energy consumption, environmental conditions and levels of occupancy; improving energy efficiency and reducing Green House Gas by analysing data from building management system (Lee et al. 2013); adapting smart lighting applications and artificial lighting system to occupancy and daylight changes and offering building usage information services (Pandaripande et al. 2014); detecting faults automatically through rule-based systems enabled by energy efficiency experts and data mining (Peña et al. 2016); detecting fault or anomalies at whole building level through integrated portfolio management (Ulickey. 2016), and communicating with utilities at city level of sensors data related to smoke levels, temperature, electricity, gas and water usage (Rathore et al. 2016).
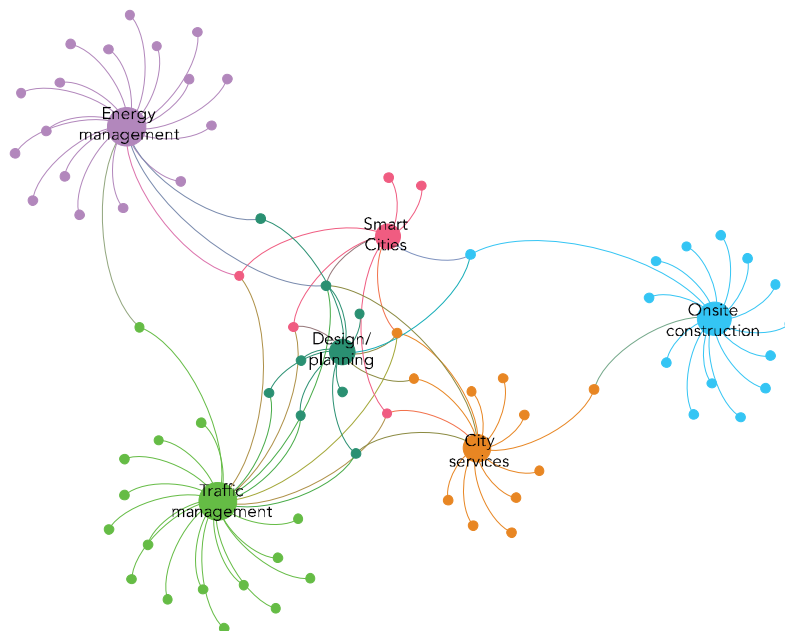

Figure 4: Clusters and spatial distribution of big data applications in built environment

4.4 Benefits, challenges and gaps

The benefits associated with big data applications in built environment can be summarised as: enabling new decisions and improving existing decision making processes; reducing CO_2 emissions; improving the use of urban space; improving health through reducing pollution; making cost savings for producers and consumers; offering better safety and security for citizens; enhancing citizens' access to amenities; increasing sustainability through the introduction of new technological developments; providing better communication networks, and focusing on a more integrated society.

   The identified key challenges facing big data applications in built environment include: security and privacy concerns; handling of non-interoperable unstructured data; availability; cost and resources for implementation, and governance.

Some gaps in the application of big data in built environment were identified. However, these gaps can be short-termed as the whole area of big data application in built environment is still an emerging area of investigation. The identified gaps include: application utilities other than electricity (i.e. water and gas); use of data from, or integration with, Building Information Modelling / Models (BIM); Retrofitting, Regeneration and Refurbishment (RRR); impact of social networking services in cities, and end-of-life activities (i.e., in disposal activities: reconstruction, reuse, recycling and landfill).

5   CONCLUSIONS

This paper identified and analysed big data applications in built environment using a systematic review and network analysis. Although big data applications were found across the entire hierarchy of the built environment and the lifecycle of assets, their distribution across the hierarchy's scale is unbalanced. The six use clusters were identified across the two selected lenses (i.e. built environment hierarchy, and asset lifecycle) include: energy management; traffic management; design and planning; onsite construction; city-wide services, and smart cities. This paper briefly summarised the uses of big data in the energy management cluster and its sub areas. The explanation of the other clusters and the descriptive 'Use Foundation model' are deferred to an extended journal article.

The limitations of this research are those typically associated with a systematic literature review. Different search terms would return different results. However, an extensive and varied set of terms, keywords, their synonyms and spelling variations was used as reasonably as practicable to guide the search strategy.

These early results can be used by domain investigators to: direct attention on identified gaps; exploit existing uses of big data, and envisage new applications across the different levels of the built environment hierarchy and the phases of asset lifecycle.

6   REFERENCES

Al Nuaimi, E., Al Neyadi, H., Mohamed, N. and Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), pp. 1-15.

Aman, S., Simmhan, Y. and Prasanna, V.K. (2015). Holistic measures for evaluating prediction models in smart grids. *IEEE Transactions on Knowledge and Data Engineering*, 27(2), pp. 475-486.

Balac, N. (2013). Green machine' intelligence: Greening and sustaining smart grids', *IEEE Intelligent Systems*. 28(5), pp. 50-55.

Bilal, M., Oyedele, L.O., Qadir, J., Munir, K., Akinade, O.O., Ajayi, S.O., Alaka, H.A. and Owolabi, H.A. (2015). 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*. 6(4), pp. 211-228.

BIM Framework (2016). *Asset Hierarchy*, Available at: http://www.bimframework.info/2014/10/asset-hierarchy.html (Accessed: 26 July 2016).

Boyd, D. and Crawford, K. (2012). Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), pp.662-679.

Chandler, D. (2015). A world without causation: big data and the coming of age of posthumanism. *Millennium: Journal of International Studies*, 43(3), pp. 833-851.

Deindl, M., Roscher, M. and Birkmeier, M. (2015). An architecture vision for an open service cloud for the smart car. in Filho, W.L. and Kotter, R. (ed.) E-Mobility in Europe. Springer International Publishing Switzerland, pp. 281-295.

Gandomi, A. and Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), pp. 137-144.

Gillman, M.D., Donnal, J.S., Paris, J., Leeb, S.B., El Sayed, M.A.H., Wertz, K. and Schertz, S. (2014). Energy accountability using nonintrusive load monitoring. *IEEE Sensors Journal,* 14(6), pp. 1923-1931.

Grolinger, K., L'Heureux, A., Capretz, M.A.M. and Seewald, L. (2016). Energy forecasting for event venues: Big data and prediction accuracy. *Energy and Buildings*, 112, pp. 222-233.

Hong, T., Koo, C., Kim, J., Lee, M. and Jeong, K. (2015). A review on sustainable construction management strategies for monitoring, diagnosing, and retrofitting the building's dynamic energy performance: Focused on the operation and maintenance phase. Applied Energy, 155, pp. 671-707.

Isikdag, U. (2015). BIM and IoT: A Synopsis from GIS Perspective. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 1, pp. 33-38.

Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9(6), pp. e98679.

Janetzko, H., Stoffel, F., Mittelstädt, S. and Keim, D.A. (2014). Anomaly detection for visual analytics of power consumption data. *Computers and Graphics (Pergamon),* 38(1), pp. 27-37.

Kekatos, V., Zhang, Y. and Giannakis, G.B. (2014). Electricity market forecasting via low-rank multi-kernel learning. *IEEE Journal on Selected Topics in Signal Processing*, 8(6), pp. 1182-1193.

Khan, M.A., Uddin, M.F. and Gupta, N. (2014). Seven v's of big data: understanding big data to extract value. Proceedings of 2013 Zone 1 Conference of the American Society for Engineering Education (ASEE Zone 1), Bridgeport, CT, 3-5 April.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. London: Sage Publications Ltd.

Lee, Y.M., An, L., Liu, F., Horesh, R., Chae, Y.T. and Zhang, R. (2013). Applying science and mathematics to big data for smarter buildings. *Annals of the New York Academy of Sciences*, 1295(1), pp.18-25.

Lenk, A., Bonorden, L., Hellmanns, A., Roedder. And Jaehnichen, S. (2015). Towards a taxonomy of standards in smart data. 2015 IEEE International Conference on Big data (IEEE BigData 2015), Santa Clara, CA: October 29 – November 1, pp. 1749-1754.

Marr, B. (2015). Big data: using smart big data analytics and metrics to make better decisions and improve performance. Chichester, West Sussex: John Wiley & Sons, Ltd.

Mishra, D., Gunasekaran, A., Papadopoulos, T. and Childe, S.J. (2016). Big data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research*, Issue: 6, pp. 1-24.

Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M. and Ochiai, O. (2015). Big data challenges in building the Global Earth Observation System of Systems. *Environmental Modelling and Software*, 68, pp. 1-26.

Pandharipande, A., Calabrese, F., Lim, H. and Rajagopal, R. (2014). Guest editorial special issue on sensing technologies for intelligent urban infrastructures. *IEEE Sensors Journal*, 14.

Paris, J., Donnal, J.S., Cox, R. and Leeb, S. (2014). Hunting cyclic energy wasters. *IEEE Transactions on Smart Grid*, 5(6), pp. 2777-2786.

Park, S., Ryu, S., Choi, Y., Kim, J. and Kim, H. (2015). Data-driven baseline estimation of residential buildings for demand response. *Energies*, 8(9), pp. 10239-10259.

Pavliscak, P. (2014). Big data, deep data, smart data, *Information Outlook*. 18(3), pp. 12-14.

Peña, M., Biscarri, F., Guerrero, J.I., Monedero, I. and León, C. (2016). Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Systems with Applications*, 56, pp. 242-255.

Petticrew, M. and Roberts, H. (2006). *Systematic reviews in the social sciences: a practical guide*. Malden, Massachusetts: Blackwell Publishing Ltd.

Pol, O., Palensky, P., Kuh, C., Leutgöb, K., Page, J. and Zucker, G. (2012). Integration of centralized energy monitoring specifications into the planning process of a new urban development area: A step towards smart cities. Elektrotechnik Und Informationstechnik. 129(4), pp. 258-264.

Qi, G. and Wayne, S.F. (2014). A Framework of Data-Enabled Science for Evaluation of Material Damage Based on Acoustic Emission. *Journal of Nondestructive Evaluation*, 33(4), pp. 597-615.

Rathore, M.M., Ahmad, A., Paul, A. and Rho, S. (2016). Urban planning and building smart cities based on the Internet of Things using Big data analytics. *Computer Networks*, 101, pp.63-80.

Sheth, A. (2013). Transforming big data into smart data for smart energy: delivering value, harnessing volume, variety and velocity. Building Research Collaborations: Electricity Systems, Purdue University, West Lafayette, Indiana: August 28-29. Available at: http://www.purdue.edu/discoverypark/energy/assets

Ulickey, J.M. (2016). One Dashboard to Rule Them All: Case Study of an Integrated Facility Management System. *Energy Engineering*, 113(5), pp.52-63.

Vaghefi, A., Farzan, F. and Jafari, M.A. (2015). Modeling industrial loads in non-residential buildings. *Applied Energy*, 158, pp. 378-389.

Zhu, J., Zhuang, E., Fu, J., Baranowski, J., Ford, A. and Shen, J. (2016). A Framework-Based Approach to Utility Big data Analytics. *IEEE Transactions on Power Systems*, 31(3), pp. 2455-2462.