



RELATIONSHIP OF EFFORT IN BUILDING INFORMATION MODELLING DATA CREATION TO BIG DATA PROGRESSION: A PILOT STUDY

Faris Hyder Ali^{1*}, Shamsulhadi Bandi²

¹ Faculty of Built Environment & Surveying, Universiti Teknologi Malaysia, Malaysia

Email: faris.ha@graduate.utm.my

² Faculty of Built Environment & Surveying, Universiti Teknologi Malaysia, Malaysia

Email: shamsulhadi@utm.my

* Corresponding Author

Article Info:

Article history:

Received date: 01.10.2021

Revised date: 01.11.2021

Accepted date: 20.11.2021

Published date: 01.12.2021

To cite this document:

Ali, F. H., & Bandi, S. (2021). Relationship of Effort in Building Information Modelling Data Creation to Big Data Progression: A Pilot Study. *Journal of Information System and Technology Management*, 6 (24), 118-124.

DOI: 10.35631/JISTM.624012

This work is licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)



Abstract:

The construction industry is developing steadily from time to time through the infusion of technologies. Though there are several kinds of technologies that have been already adopted, Building Information Modelling (BIM) software has a vital role in the evolution of construction technologies since it is considered as digital working strategy and method capable to integrate various applications. This circumstance gives rise to the explosion of BIM data through the integration where it sparks the relevancy of big BIM data in construction. However, the creation of BIM data in the context of big data still remains uncertain due to the incognizant state of effort being imposed in the generation of data. Therefore, a study to determine the relationship of the current effort in BIM data creation towards big data progression in construction is being carried out. Before carrying out the research main/large scale data collection, this research has been undergone through smaller-scale data collection which is a pilot study. The pilot study was administered via a survey questionnaire to 30 BIM modellers in the Klang Valley to scrutinize the credibility of the instrument to be utilized in the study later. Thus, this paper reports partly the outcome of the pilot study which sought to appraise the reliability of the instrument and the data normality as a prerequisite step to conduct an inferential analysis at a much later stage of the research. The questionnaire for the pilot test was distributed to BIM modellers within Klang Valley. Through the study conducted, it has been found that the instrument developed is reliable and the most suitable analysis method is a non-parametric test which is Spearman's correlation due to non-normal distribution.

Keywords:

Building Information Modelling, Big Data, Effort, Pilot Study

Introduction

Currently, despite being among the least digitized industry, the construction is gradually thriving through the adoption of technologies where backbone of digital strategy which is Building Information Modelling software has favorable impact towards the construction through its integration capability (Kaufmann, Ruaux, & Jacob, 2018). This amalgamation in BIM platform leads to the formation of voluminous data which will cause the incapacity of conventional software to store, process and analyze the data (Bilal, Oyedele, Akinade, et al., 2016; Reyes, Suresh, & Renukappa, 2018). This situation illustrates that BIM data will require another platform to provide data driven insights where some scholars emphasize the relevancy of the concept of big data (Hosey, 2018). Nevertheless, the effort being exerted in the creation of big BIM data is still indecisive due to lack of studies carried out in this aspect. Hence, it is imperative to assess the current state of BIM data creation in the context of big data.

Before conducting the inferential analysis on the complete feedbacks that going to be obtained, it is paramount the instrument developed will be subjected to pilot study in which reliability and normality need to be assessed on the smaller scale data (Kowalski, 1972; Taherdoost, 2016). Therefore, this study has been conducted to appraise the reliability of the instrument and to identify the normality of the data.

Literature Review

The construction is refining by the adoption of several digital innovations which can bring great improvement to the industry. Wyman (2018) mentioned that the gradual enhancement is taking place due to the paradigm shift demanded globally in construction. Therefore, this circumstance leads to the evolution in terms of collaborative working environment, visualization, analytics and others where BIM becomes as a focal point of the development (Hasni, Ismail, & Hashim, 2019; Kaufmann et al., 2018; Motawa, 2017).

The concept of BIM which was introduced a decade ago acts as a game changer that influence the whole lifecycle of building and has become as one of the most applied technologies in construction sector (Gerrish, 2017; Moses, Heesom, & Oloke, 2020). Since BIM software are being used intensely, it leads to the creation of largely structured BIM data such as materials data, design data, cost data and others (Jansen, 2018). Besides, there will be an exponential growth in the BIM data through integration with various relatable data (Arslan, Riaz, & Munawar, 2017; Solihin & Eastman, 2015). As a consequence, the conventional method's storage and processing systems cannot withstand the data explosion (Ismail, Bandi, & Maaz, 2018). Hence, BIM data requires another platform to ensure appropriate data management and utilization by restructuring the way of making use digital data in BIM environment (Correa, 2015; Farghaly, 2019).

The limitation of BIM software due to proliferation of data, leads to the realization of big data in construction (Boton, Halin, Kubicki, & Forgues, 2015). Though the concept of big data has come to an existence in 90s, but the effort to confabulate this notion in construction only get into consideration recently (Bilal, Oyedele, Qadir, et al., 2016). To understand the concept, there is a need to grasp about the term data itself. Data refers to any unprocessed text, symbols and numbers which require storage and processing through available tools to become beneficial to users (Cambridge, 2017). Meanwhile, big data is incapability of conventional tools to store and process due to massive data creation in the database (Olsson & Bull-Berg, 2015). Thus, it shows that normally data can be processed through conventional method, but once it becomes

big data, the data will become much complex where it emphasizes the need to make use other sophisticated tool which is big data technology.

The concept of big data can be understand further through its three main characteristics which are volume, velocity and variety where scholars concurred that a datasets can be considered as big data when it fulfils those attributes (Chen, Mao, & Liu, 2014; Maaz, Bandi, & Amirudin, 2018). Although big data is still at infancy stage despite its potentiality in construction, the big data technology has already proved its worth in various sectors such as manufacturing, banking, tourism and so on where it is capable to solve underutilization issue by deriving value from the accumulated data (Institute, 2011; Jingjing Li, 2018; Kumar, Shankar, & Thakur, 2018; Rabhi, Falih, Afraites, & Bouikhalene, 2019).

Besides the opportunity attainable through the big data technological adoption, there is a concern whether the current state of BIM data is really leading towards the big data. Though there are several studies had reviewed about the potential relationship between BIM and big data through sub-domains of construction management such as waste, facilities, energy and others, there are still contradiction regarding this technological convergence (Bilal, Oyedele, Qadir, et al., 2016; Burger, 2019). This is because of several scholars emphasized that BIM data would not be large enough to be recognized as big data (Anunson, 2018; Correa, 2015). This situation triggers the need to clarify the uncertain state of effort in BIM data formation in construction. Thus, a study to determine the relationship of the current effort in BIM data creation towards big data progression in construction needs to be carried out. In order to perform the study, a pilot test is required to examine the viability of the method that is going to be employed in larger scale study. This small-scale preliminary study has been carried out by appraising the normality and reliability of obtained data.

Research Methodology

Before carrying out the main data collection, it is imperative to conduct pilot study. It is being done to ensure the reliability and validity of the research instrument (Taherdoost, 2016). The number of pilot sample was finalized based on scholars' view where they mentioned the need to obtain at least 30 respondents (Johanson & Brooks, 2010; Whitehead, Julious, Cooper, & Campbell, 2016). Therefore, feedbacks from 30 BIM modellers which comprises of consultants, contractors and property developers were procured within Klang Valley. Then, the reliability of data was tested by using Cronbach's Alpha to identify the consistency in measurement (Laerd, 2018a; Price, Jhangiani, Chiang, & Leighton, 2017). On the other hand, normality test was done to examine whether the data are normally or non-normally distributed which can be assessed in numerical or graphical format (Laerd, 2018b). Though there are various test available to test the normality, Shapiro-Wilk test has been chosen for the research because it is being perceived as the best option to test the normality (Ghasemi & Zahediasl, 2012).

Main Results

The main concern of the reliability is the value of the Alpha (α) which reflect the consistency. Therefore, scholars have outlined the reliable and acceptable Alpha which is 0.6-0.7 while 0.8 or greater is a very good consistency (Hulin, Netemeyer, & Cudeck, 2001; Ursachi, Zait, & Horodnic, 2015). Based on the Table 1, it has been found that the α is 0.838 for intensity of BIM data creation and 0.825 for big data in BIM platform where α for both items depict very

good internal consistency in the gathered data. Hence, the research instrument can be considered as a reliable one.

Table 1: Reliability Test

Item	ALPHA If Item Deleted Content
Intensity of BIM Data Creation	.838
Big Data in BIM Platform	.825

Meanwhile, the test of normality was conducted to verify the normality of distribution which is the basis to determine whether parametric or non-parametric test to be applied (Frost, 2017). Amid several normality tests available, Shapiro-Wilk test has been used to identify whether the data are normally or non-normally distributed (Ghasemi & Zahediasl, 2012). The sig. level should be more than 0.05 to indicate the normal distribution (Laerd, 2018b). The highest sig. level among the items is design data, which is about 0.044, but this value itself reflects insignificant level. Meanwhile, other items' sig. level is in between 0.000 until 0.030 which obviously depicts much lower significant level. Since the sig. level for all of them are less than 0.05 as shown in the Table 2, it seems the gathered data are non-normally distributed. Therefore, the most appropriate correlation analysis method will be non-parametric test which is Spearman Rho. The analysis is applicable since it fulfilled two assumptions of the test which are both variables are ordinal data and the research cannot make use of Pearson Correlation due to the violation of the test assumptions (Frost, 2017). The Spearman inferential analysis will determine and measure whether there is a direction and strength of correlation between two variables of the research which are the effort of BIM data creation and big data progression. Hence, the contradiction among the scholars in relation to the emergence of big data in construction BIM platform can be demonstrated and validated.

Table 2: Normality Test (Shapiro-Wilk Test)

	BIM Data	Statistic	df	Sig.
Big Data	Design	0.655	30	0.044
	Material	0.710	30	0.015
	Cost	0.716	30	0.028
	Schedule	0.623	30	0.017
	Operation & Maintenance	0.708	30	0.002

Conclusion

There were several constraints faced throughout the pilot study such as unavailability of the respondents, slow progress in acquiring the feedbacks, Movement Control Order restriction and others. Despite that, the study still managed to be completed within several months. The study reflects that the instrument is reliable enough for large scale data collection and the distribution of data are non-normally distributed. Hence, non-parametric which is Spearman's Rho will be utilized for the analysis once large scale data collection has been completed. This study is significant since it has determined the most appropriate inferential analysis for the research.

Acknowledgment

This research is part of ongoing research which seeks to establish the status of construction big data progression using BIM data creation as indicator. Previously, there were two articles

published in relation to the ongoing research in two different journals which are “Associations between BIM data and big data attributes” and “Key dimensions of effort for BIM data creation”. In addition, this paper has been presented in the International Graduate Conference of Built Environment and Surveying which was conducted from 22nd to 23rd June 2021. All these works are being supported/funded by Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2019/SS03/UTM/02/4).

References

- Anunson, J. (2018). BIM: The original Big Data?
- Arslan, M., Riaz, Z., & Munawar, S. (2017, 9-13 July 2017). *Building Information Modeling (BIM) Enabled Facilities Management Using Hadoop Architecture*. Paper presented at the 2017 Portland International Conference on Management of Engineering and Technology (PICMET).
- Bilal, M., Oyedele, L. O., Akinade, O. O., Ajayi, S. O., Alaka, H. A., Owolabi, H. A., . . . Bello, S. A. (2016). Big data architecture for construction waste analytics (CWA): A conceptual framework. *Journal of Building Engineering*, 6, 144-156.
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., . . . Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, 30(3), 500-521. doi: <https://doi.org/10.1016/j.aei.2016.07.001>
- Boton, C., Halin, G., Kubicki, S., & Forgues, D. (2015). *Challenges of Big Data in the Age of Building Information Modeling: A High-Level Conceptual Pipeline*, Cham.
- Burger, R. (2019). *How the Construction Industry is Using Big Data*. Cambridge. (2017). *Data, Information and Knowledge*.
- Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A Survey. *Mobile Networks and Applications*, 19(2), 171-209. doi: 10.1007/s11036-013-0489-0
- Correa, F. (2015). *Is BIM big enough to take advantage of big data analytics?* Paper presented at the ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction.
- Farghaly, K. (2019). BIM-linked data integration for asset management. *Built Environment Project and Asset Management*, 9(4), 489-502. doi: 10.1108/BEPAM-11-2018-0136
- Frost, J. (2017). Nonparametric Tests vs. Parametric Tests. from <https://statisticsbyjim.com/hypothesis-testing/nonparametric-parametric-tests/>
- Gerrish, T. (2017). Using BIM capabilities to improve existing building energy modelling practices. *Engineering, Construction and Architectural Management*, 24(2), 190-208. doi: 10.1108/ECAM-11-2015-0181
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, 10(2), 486-489. doi: 10.5812/ijem.3505
- Hasni, M. I. A. K., Ismail, Z., & Hashim, N. (2019). *Contractual Aspects in the Utilisation of Level Two Building Information Modelling (BIM) within Malaysian Public Construction Projects*. Paper presented at the MATEC Web of Conferences.
- Hosey, T. (2018). Reinventing Construction Industry with Big Data Analytics.
- Hulin, C., Netemeyer, R., & Cudeck, R. (2001). Can a reliability coefficient be too high? *Journal of Consumer Psychology*.
- Institute, M. G. (2011). *Big Data: The Next Frontier for Innovation, Competition and Productivity*.

- Ismail, S. A., Bandi, S., & Maaz, Z. N. (2018). An Appraisal into the Potential Application of Big Data in the Construction Industry. *International Journal of Built Environment and Sustainability*, 5(2).
- Jansen, S. (2018). The "I" In BIM: Which Information Should Be In Your BIM Model?
- Jingjing Li, L. X., Ling Tang, Shouyang Wang & Ling Li (2018). Big data in tourism research: A literature review. *Tourism Management*, 68, 23.
- Johanson, G. A., & Brooks, G. P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Educational and psychological measurement*, 70(3), 394-400. doi: 10.1177/0013164409355692
- Kaufmann, D., Ruaux, X., & Jacob, M. (2018). Digitalization of the Construction Industry: The Revolution is Underway.
- Kowalski, C. J. (1972). On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 21(1), 1-12.
- Kumar, A., Shankar, R., & Thakur, L. S. (2018). A big data driven sustainable manufacturing framework for condition-based maintenance prediction. *Journal of computational science*, 27, 428-439.
- Laerd. (2018a). Cronbach's Alpha Using SPSS Statistics. from <https://statistics.laerd.com/spss-tutorials/cronbachs-alpha-using-spss-statistics.php>
- Laerd. (2018b). Testing for Normality using SPSS Statistics. from <https://statistics.laerd.com/spss-tutorials/testing-for-normality-using-spss-statistics.php>
- Maaz, Z. N., Bandi, S., & Amirudin, R. (2018). Big Data in The Construction Industry: Potential, Opportunities and Way Forward. *The Turkish Online Journal of Design, Art and Communication*.
- Moses, T., Heesom, D., & Oloke, D. (2020). Implementing 5D BIM on construction projects: contractor perspectives from the UK construction sector. *Journal of Engineering, Design and Technology*.
- Motawa, I. (2017). Spoken dialogue BIM systems – an application of big data in construction. *Facilities*, 35(13/14), 787-800. doi: 10.1108/F-01-2016-0001
- Olsson, N. O., & Bull-Berg, H. (2015). Use of big data in project evaluations. *International Journal of Managing Projects in Business*, 8(3), 491-512.
- Price, P. C., Jhangiani, R. S., Chiang, I.-C. A., & Leighton, D. C. (2017). *Research Methods in Psychology* (2nd American Edition ed.).
- Rabhi, L., Falih, N., Afraites, A., & Bouikhalene, B. (2019). Big Data Approach and its applications in Various Fields: Review. *Procedia Computer Science*, 155, 599-605. doi: <https://doi.org/10.1016/j.procs.2019.08.084>
- Reyes, P., Suresh, S., & Renukappa, S. (2018). The Adoption of Big Data Concepts for Sustainable Practices Implementation in the Construction Industry. In A. Sill & J. Spillner (Eds.), *2018 Ieee/Acm International Conference on Utility and Cloud Computing Companion* (pp. 349-352). New York: Ieee.
- Solihin, W., & Eastman, C. (2015). A simplified BIM model server on a big data platform. *CIB W78, 2016*.
- Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *How to Test the Validation of a Questionnaire/Survey in a Research (August 10, 2016)*.

- Ursachi, G., Zait, A., & Horodnic, I. A. (2015). How Reliable are Measurement Scales? External Factors with Indirect Influence on Reliability Estimators. *Procedia Economics and Finance*. doi: 10.1016/S2212-5671(15)00123-9
- Whitehead, A. L., Julious, S. A., Cooper, C. L., & Campbell, M. J. (2016). Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research*, 25(3), 1057-1073. doi: 10.1177/0962280215588241
- Wyman, O. (2018). Digitilization of the construction industry.