HAL
open science

Challenges of Big Data in the age of Building Information Modeling: A high-level conceptual pipeline

Conrad Boton, Gilles Halin, Sylvain Kubicki, Daniel Forgues

To cite this version:

Conrad Boton, Gilles Halin, Sylvain Kubicki, Daniel Forgues. Challenges of Big Data in the age of Building Information Modeling: A high-level conceptual pipeline. CDVE 2015, Sep 2015, Mallorca, Spain. hal-02890096

HAL Id: hal-02890096

https://hal.science/hal-02890096

Submitted on 6 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Challenges of Big Data in the age of Building Information Modeling: A high-level conceptual pipeline

Conrad Boton[1 2], Gilles Halin[2], Sylvain Kubicki[3] and Daniel Forgues[1]

[1] École de Technologie Supérieure (ETS), 1100, rue Notre-Dame Ouest, Montréal (Qc) Canada.
[2] CRAI - Research Centre in Architecture and Engineering, 2, rue Bastien Lepage. 54001 Nancy, France
[3] Luxembourg Institute of Science and Technology (LIST), 5, avenue des Hauts-Fourneaux, L-4362 Esch/Alzette, Luxembourg.
conrad.boton.1@ens.etsmtl.ca, gilles.halin@crai.archi.fr, sylvain.kubicki@list.lu, daniel.forgues@etsmtl.ca

Abstract. N-dimensional BIM models integrates many aspects of construction projects information. Construction project managers are facing an important challenge related to making meaningful deduction from this data. Because it is interesting to associate BIM with other information sources and because of the increasing growth of cloud computing, the current data mining tools are showing their limitations. Big data is then gradually getting a reality in the construction industry. This paper tries to define what it means and proposes a high-level conceptual pipeline in order to bridge the gap between BIM-based related visualization works and information visualization domain.

Keywords: BIM, Big Data, Information Visualization, Information pipeline, Architecture, Engineering and Construction.

1 Introduction

With the advent of Building Information Modeling (BIM) in the construction industry, the perspective of information management is changing. Reddy compared the transformative aspect of BIM to the Internet and stated that it is becoming necessary to manage building data on projects of any size [1]. According to him, "the most interesting aspect of BIM in the long term is the application of data mining and analyzing" [1]. Moreover, building information models are increasingly associated with other sources, including urban cartographies, Geographic Information Systems (GIS) [2, 3], Building Automation Systems (BAS) , Facility Management (FM) [4, 5], Electronic Document Management Systems (EDMS) [6] and Computerized Maintenance Management Systems (CMMS) [7]. Combining such different sources in order to extend the BIM-based decision-support can quickly become very complex for current data mining tools due to the variety and the amount of the information.

Thus, Big Data has the potential to become a common reality in the industry. In 2011, the McKinsley Global Institute proposed a map showing how it is easy to capture the value potential of Big Data cross sectors. On a scale of five, construction

sector is positioned on the third quintile for IT intensity, for data-driven mind-set, and for data availability, and for the overall ease of capture index [8].

Because of the peculiarity of the industry, data integration and management is a challenging issue in the development of Architecture, Engineering and Construction (AEC) projects. It is worth to remind that fragmentation, non-standardized collaboration and heterogeneity of actors and skills characterize collaborative processes in AEC projects management. Moreover it is not trivial to give a precise definition of what Big Data currently means in construction projects. Moreover, with the growing implementation of Building Information Modeling approach in the AEC industry, one can state that no conceptual formalization has been provided so far to identify construction projects' related information challenges. This paper proposes to explore the literature in order to define what Big Data really means in construction industry. It then proposes a conceptual model of Big Data based on previous seminal frameworks and information visualization theories.

2 Big Data in construction management in an era of BIM: what does it mean?

Provost and Fawcett defined Big Data as "datasets that are too large for traditional data-processing systems and that therefore require new technologies" [9]. According to the McKinsley Global Institute, it "refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" [8]. McAfee and Brynjolfsson (2012) identified three key differences between Big Data and analytics used before. These differences are related to the variety, the volume, and the velocity of the data [10].

2.1 As regards the variety

A large variety of information is managed and used in construction projects. It generally encompasses functional brief requirements, feasibility studies, architectural drawings, engineering drawings, technical specifications, tender documents, bills of quantities, cost estimations, invoices, reports, lifecycle information, technical data sheets, as-built records. This list is far to be exhaustive and we can add design know-how, building regulations, construction know-now, client brief, schematic drawings, observations, measurements, etc. [11].

Jiao et al. (2013) distinguishes project engineering data (PED) from project management data (PMD). According to them, PED refers to "geometric presentation, parametric descriptions and legal regulations associated with the construction of a building" [12]. PMD are "control and communication information that are generated in and closely related to management activities throughout the construction lifecycle". Björk showed that information is required to control the material activities, to specify the building requirements (design information) and the activities to be performed (management information) [11]. The construction projects information is presented in

Table 36 of the OmniClass[1] classification. Its three broad categories (general references, office resources, project information) and their content well illustrate the variety of information in construction project.

With the development of the BIM approach, both PMD and PED are integrated and coordinated with the aim of managing nD models. nD modeling can be defined as the addition of supplementary information to tridimensional model(s) for analyses and simulation purposes. Thus, it is commonly agreed that a 4D model links scheduling information to 3D model's objects in order to simulate construction process over time. The fifth dimension is related to costs information. It is more and more accepted that the sixth dimension is represented by facility management information but it is important to note that beyond the fifth dimension (cost), there is no consensus about what each dimension represents. Lee et al. identify eight possible dimensions which notably show how nD modelling can increase the variety of information to be managed in BIM models. The dimensions are: maintenance needs, acoustics, process, cost, energy requirements, crime deterrent features, sustainability, people's accessibility [13].

This variety of model-related dimensions of information is highly structured and relies on object-based conceptual modeling techniques, such as BuildingSmart's IFC. However, design and construction processes also rely on huge raw datasets, such as the 3D point clouds delivered in the surveying tasks using laser scanning techniques. Moreover, as mentioned in the introduction many research works have reported other information sources which can be integrated with BIM. Such integration approaches of course dramatically increase the variety of information to be managed, but also the volume.

2.2 As regards the volume

As regards the volume, it is hard to say beyond which size data can be called Big data because it depends on the sector, the advances of the technology over time and the complexity of data [8]. In the AEC industry, "construction projects are associated with voluminous and often unstructured data sets" [15]. According to Jiao et al., "individual file sizes currently range from Megabytes to Gigabytes, however there is an apparent trend of increasing file size" [12]. These authors mention a huge volume of data representing "tens of thousands of files" at a project level. The McKinsey Global Institute (2011) estimated the data storage for the USA construction industry in 2009 at 51 petabytes, with 231 terabytes for each of the 222 firms with more than 1000 employees [8].

Traditionally, the stored data is in the form of images and text/numbers but with BIM and nD modeling, other forms of data are now stored. For example, 4D models are commonly exported and stored as video animations. In the frame of nD modeling, the size of the handled data can grow very rapidly due to the multiple dimensions conveyed, their size, their variety and their complexity. Joyce recently published in a professional journal a very interesting case related to 4D modelling. He stated that when managers associate costs information to "the three-day schedule of a crane, the five-day schedule of its crew, the seven-day schedule of the materials and the cost of

[1] http://omniclass.org

the crew's labor" the size of the data to manage can rapidly get out of control [16]. According to him, with the growth of 4D BIM in the construction sector, "big data arrives each day with an ever-louder thud"[16].

2.3 As regards the velocity

Thanks to new delivery methods including Integrated Project Delivery (IPD), BIM level 3 has already been adopted in a few construction projects. A survey conducted in 2010 by Kent and Becerik-Gerber showed that the adoption of IPD by construction professionals in the United States was still not very significant, but noted a huge interest in the principle and some good perspectives [17]. Both experienced professionals (66.7%) and informed ones (58.3%) see the IPD becoming one day a widely adopted method for project delivery. BIM level 3 is defined as network-based integration, a collaborative use of an integrated object-oriented model, shared over a network in which users can connect to feed and use the model, within their respective areas of responsibility [18, 19]. Beyond "building-related" data, recent policies plan to integrate "infrastructure-related" to the models required for public tendering in UK.

Two main methods are currently used for models integration in the BIM approach. The first method use object-oriented database named "BIM servers" to manage IFC objects with their versions, their variants, their business views and the associated notifications. The interface with the BIM server is enabled via BIM authoring software systems or intermediary viewers coupled to the server with APIs. The second method consists in the use of Project Lifecycle Management (PLM) collaborative platforms to establish and maintain the link between IFC objects collections corresponding to the responsibility and the production of the different stakeholders within the same construction phase or among different phases of the project. This requires the use of new technologies such as "cloud computing" [20].

Whether for PLM or BIM server, the velocity of data will be an important challenge in the next years. Indeed, the speed needed for data input/output is getting increasingly higher because data must be processed quickly despite its big volume. In fact, the volume of data and the diversity of transactions coming from multiple sources considerably increase the flow of information. Moreover, Jiao et al. (2013) have notably well stated the current data velocity challenges in construction.

3 A Big Data conceptual pipeline for Construction Industry

3.1 Background

The following proposition relies on important previous works dedicated to both 1) the information technology and flow in construction and 2) the (Big) data flow proposed by Information Visualization scientific fields.

In 1999, Bjork defined the definition and the main research issues associated with the emerging domain of Information Technology in construction. The work proposed a global model of construction process relying on two interacting sub-processes: the material process and the information process [11]. It then used the IDEF0

methodology to model the interactions of the four generic information process activities: person-to-person communication, creation of new information, information search and retrieval, making information available. Due to the complexity of processes in construction and their interrelations, Turk proposed to distinguish the glue processes (or integration processes) from the base processes [21]. According to Turk, while base processes are the core processes (including information creation processes and information utilization processes), glue processes "make sure that material items or information flow from creation process to utilization process and that the utilization process can use them" [21]. Information creation includes four steps: create, edit, record, and distribute [21]. The glue process includes five steps: find, retrieve, store, convert, and store (the converted version). According to Turk, information processes have five main features: input, output, performer, customer, method, and time frame.

As developed in previous sections, BIM can generate voluminous and various data set to support the management functions in construction projects. According to Russell et al. (2009), making meaningful deduction from this data is one important challenge construction project managers are facing. He also remarked that information processing and visualization can be very helpful. Mazza defined visualization as a "cognitive activity, facilitated by external visual representations from which people build an internal mental representation of the world" [22]. Geovisualization (about geospatial data), Scientific Visualization (about physical "things") and Information Visualization (about abstract data) are the three main research fields that provide scientific body of knowledge on information processing and visualization. With the advances of IT, many tools exist to assist users in creating visual representations of abstract data. Although these tools are very different from each other, they all follow the same steps and process of generating visual representations. Mazza (2009) presented a conceptual model inspired from the pipeline proposed by Haber and McNabb [23]. Three main steps have been identified: pre-processing, visual mapping, and view creation. One of the reference models in collaborative visualization is also based on this pipeline [24].

In construction industry, Russell et al. identified four main steps necessary for an efficient information processing activity: identifying the purpose, selecting the data representations and transformations, choosing the visual representation and the interaction technology, and presenting and disseminating [15]. More recently, assuming that 4D/nD simulation works are largely collaborative, Boton et al. proposed a method to design adapted visualization in collaborative 4D applications [25]. The method includes four main steps: characterizing collaborative situation, identifying visualization needs, selecting adapted business views, composing coordinated multiple views.

The amount of data in every sectors and function of the global economy has been exploding and the current models and applications seem inappropriate to manage the volume, the variety and the velocity of data. New conceptual models are then appearing. In 2012, Fisher et al. proposed a conceptual big data pipeline, based on interviews with experienced practitioners [26]. Five main steps appeared in the proposed pipeline: acquiring data, choosing architecture, shaping data into architecture, coding/debugging, and reflecting.

3.2 A new (big) data pipeline dedicated to construction industry

In this section, and based on the previous approaches, we propose a conceptual data pipeline for construction industry. The IDEF0 formalism is used, which is well used by the Information Technology in Construction scientific community. IDEF0 is known to be used for the enterprise activities and information flows modeling [27]. In IDEF0 simple syntax, activities are placed in a box and described by verb-based labels. Arrows entering the left side of the box represent inputs. Arrows exiting from the right side of the box show outputs of the activities. The top entering arrows represents the control concepts while the bottom entering arrows show the mechanisms or involved actors.

The proposed pipeline relies both on the pipeline proposed by Fisher et al. (2012) and the existing work in the construction industry, including the 4 main steps identified by Russell et al. (2009) and the method proposed by Boton et al. (2013). It identifies four main steps: acquiring data, choosing architecture, Shaping and coding data, reflecting and interacting.

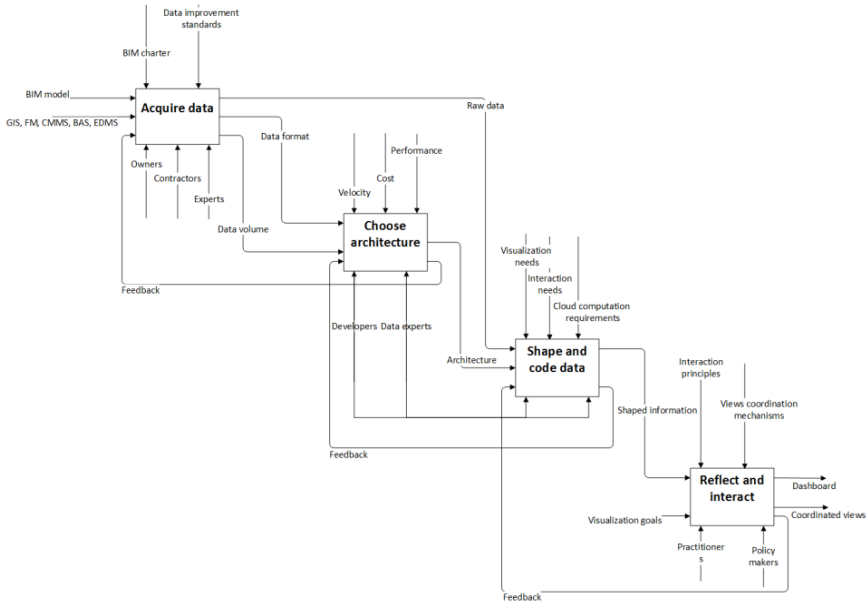

Figure 1: A high-level conceptual pipeline

The first challenge in the Big Data pipeline is related to acquiring data. Data can come from BIM models but also from other sources including GIS, FM systems, CMMS, BAS and EDMS. The structured data from BIM have to be deconstructed in order to be processed, restructured and analyzed. Data come from owners, contractors

and other experts. BIM charters and data improvement standards are used to control the way data are acquired.

The second challenge in the Big Data pipeline is related to choosing architecture. This activity is mainly devoted to developers and data mining experts. The formats and the volume of the data from the previous step will highly determine the choice. The cost, the velocity and other performance criterion are used for architectures comparison in order to choose the most appropriate one.

The third challenge in the Big Data pipeline is related to shaping and coding data. Raw data acquired at the first step are shaped and coded in the frame of the chosen architecture. It is important here to understand users' visualization and interaction needs. It is also interesting to take into account the cloud computing requirements. This activity is also devoted to data mining experts.

The fourth challenge in the Big Data pipeline is related to reflecting information and interacting with it. The aim is to provide users, policy makers and other practitioners with decision support dashboards which present meaningful information and coordinated multiple views. Users' visualization goals are some important input in the achievement of this step. Interactions principles related to visualization tasks and views coordination mechanism are also to be considered.

4 Conclusion

This paper explored the literature in order to define what Big Data can mean for the construction industry at the age of Building Information Modeling. It also proposed a high-level conceptual model dedicated to the industry based on previous seminal frameworks and information visualization theories.

This is a first step of an important work which aim at bridging the gap between BIM-based related visualization works and information visualization domain, usually applied to unstructured data. Future works will apply the proposed pipeline to real construction organization in order to evaluate it.

References

1. Reddy, K.P.: BIM for Building Owners and Developers: Making a Business Case for Using BIM on Projects. JOHN WILEY & SONS, INC., Hoboken, New Jersey (2012).
2. DeLaat, R., VanBerlo, L.: Integration of BIM and GIS: The Development of the CityGML GeoBIM Extension. Lecture Notes in Geoinformation and Cartography. pp. 211–225 (2011).
3. Irizarry, J., Karan, E.P., Jalaei, F.: Integrating BIM and GIS to improve the visual monitoring of construction supply chain management. Autom. Constr. 31, 241–254 (2013).
4. Forns-Samso, F., Laine, T., Hensel, B.: Building information modeling supporting facilities management. Ework Ebus. Archit. Eng. Constr. Proc. ECPPM. 51–57 (2012).

5. Becerik-Gerber, B., Jazizadeh, F., Li, N., Calis, G.: Application Areas and Data Requirements for BIM-Enabled Facilities Management. J. Constr. Eng. Manag. 138, 431–442 (2012).
6. Goedert, J.D., Meadati, P.: into Building Information Modeling. J. Constr. Eng. Manag. 134, 509–516 (2008).
7. Motawa, I., Almarshad, A.: A knowledge-based BIM system for building maintenance. Autom. Constr. 29, 173–182 (2013).
8. McKinsey Global Institute: Big data : The next frontier for innovation , competition , and productivity. (2011).
9. Provost, F., Fawcett, T.: Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data. 1, 51–59 (2013).
10. McAfee, A., Brynjolfsson, E.: Big Data: The Management Revolution. Harv. Bus. Rev. 90, 60–68 (2012).
11. Björk, B.C.: Information Technology in Construction–domain definition and research issues. Int. J. Comput. Integr. Des. Constr. 1, 1–16 (1999).
12. Jiao, Y., Wang, Y., Zhang, S., Li, Y., Yang, B., Yuan, L.: A cloud approach to unified lifecycle data management in architecture, engineering, construction and facilities management: Integrating BIMs and SNS. Adv. Eng. Informatics. 27, 173–188 (2013).
13. Lee, A., Aouad, G., Cooper, R., Fu, C., Marshall-Ponting, A., Tah, J., Wu, S.: nD modelling-a driver or enabler for construction improvement? RICS Res. Pap. Ser. RICS, London. 5, 1–16 (2005).
14. Mignard, C., Nicolle, C.: Computers in Industry Merging BIM and GIS using ontologies application to urban facility management in ACTIVe3D. Comput. Ind. 65, 1276–1290 (2014).
15. Russell, A.D., Chiu, C.-Y., Korde, T.: Visual representation of construction management data. Autom. Constr. 18, 1045–1062 (2009).
16. Joyce, E.: What You Don't Know About Your (Big) Data Can Hurt You, http://enr.construction.com/technology/construction_technology/2012/1203-big-data-little-data-and-aec.asp.
17. Kent, D.C., Becerik-gerber, B.: Understanding Construction Industry Experience and Attitudes toward Integrated Project Delivery. J. Constr. Eng. Manag. 136, 815–825 (2010).
18. Succar, B.: Building information modelling framework: A research and delivery foundation for industry stakeholders. Autom. Constr. 18, 357–375 (2009).
19. Boton, C., Kubicki, S.: Maturité des pratiques BIM: Dimensions de modélisation, pratiques collaboratives et technologies. Interaction des Maquettes Numériques, Actes du 6ème Séminaire de Conception Architecturale Numérique (SCAN'14) (2014).
20. Chong, H.-Y., Wong, J.S., Wang, X.: An explanatory case study on cloud computing applications in the built environment. Autom. Constr. 44, 152–162 (2014).
21. Turk, Z.: Construction IT: Definition, framework and research issues. Fac. Civ. Geod. Eng. doorstep Millenn. Occas. its 80th Anniv. 17–32 (2000).
22. Mazza, R.: Introduction to Information Visualization. Springer London (2009).
23. Haber, R.B., McNabb, D.A.: Visualization Idioms: A Conceptual Model for Scientific Visualization Systems. Visualization in Scientific Computing. pp. 74–93. IEEE (1990).
24. Wood, J., Wright, H., Brodlie, K.: CSCV - Computer Supported Collaborative Visualization. International Conference on Visualization and Modelling. , Leeds (1995).
25. Boton, C., Kubicki, S., Halin, G.: Designing adapted visualization for collaborative 4D applications. Autom. Constr. 36, 152–167 (2013).
26. Fisher, D., DeLine, R., Czerwinski, M., Drucker, S.: Interactions with Big Data Analytics population by running controlled. Interactions. May+June, 50–59 (2012).
27. Dorador, J.M., Young, R.I.M.: Application of IDEF0, IDEF3 and UML methodologies in the creation of information models. Int. J. Comput. Integr. Manuf. 13, 430–445 (2000).