

# BIM BIG DATA SYSTEM ARCHITECTURE FOR ASSET MANAGEMENT: A CONCEPTUAL FRAMEWORK

Karim Farghaly<sup>1</sup>, Henry Abanda<sup>2</sup>, Christos Vidalakis<sup>3</sup> and Graham Wood<sup>4</sup>

**Abstract:** Effective decision making in the AEC/FM industry has been based increasingly on an exponential growth of data extracted from different sources and technologies. It has been argued that Building Information Modelling (BIM) can handle this information efficiently, acting as a data pool where data can be stored, managed and integrated. Indeed, a BIM platform based on cloud computing and Big Data can manage the storage and flow of data, as well as extract knowledge from Geographical Information Systems (GIS), Internet of Things (IoT), asset management, energy management and materials and resources databases. Furthermore, it can also provide an opportunity for multiple users to view, access and edit the data in 3D environment.

This paper describes the requirements and different components of a BIM Big Data platform for facilitating management of building assets. This is achieved by firstly, conducting a critical peer review to ascertain Big Data definitions and stages, and also to define the critical BIM requirements for the Big Data platform. At the crux, this paper presents a conceptual framework for developing a Big Data platform for BIM which incorporates suitable tools and techniques needed to export, store, analyse and visualise BIM data.

**Keywords:** Building Information Modelling, Big Data, Asset Management.

## 1 INTRODUCTION

The era of Big Data Analytics (BDA) has emerged mainly due to the need to deal with the huge volume, complex and growing data generated by social media, sensors, instruments and a plethora of digital sources. Currently, nearly every sector of the economy including accounting, tourism, transportation, education and construction is affected by Big Data. This has mainly been due to the noticeable Big Data capabilities in storing, processing and analysing big volume of diverse data.

At the same time, Building Information Modelling (BIM) has become the new international benchmark for better efficiency and collaboration in design, construction and building operation and maintenance. BIM can act as a data pool during all phases for all data of the building, even from other technologies such as Geographical Information Systems (GIS), Radio Frequency Identifications (RFID), Internet of Things (IoT) and Augmented Reality (AR). Furthermore, Data from different sources such as asset management, energy management, and materials and resources databases can be

---

<sup>1</sup> PhD Student, School of the Built Environment, Faculty of Technology, Design and Environment, Oxford Brookes University, Oxford, UK [karim.ibrahim-2016@brookes.ac.uk](mailto:karim.ibrahim-2016@brookes.ac.uk)

<sup>2</sup> Senior Lecturer, School of the Built Environment, Faculty of Technology, Design and Environment, Oxford Brookes University, Oxford, UK [fabanda@brookes.ac.uk](mailto:fabanda@brookes.ac.uk)

<sup>3</sup> Senior Lecturer, School of the Built Environment, Faculty of Technology, Design and Environment, Oxford Brookes University, Oxford, UK [christos.vidalakis@brookes.ac.uk](mailto:christos.vidalakis@brookes.ac.uk)

<sup>4</sup> Reader in Environmental Assessment and Management, School of the Built Environment, Faculty of Technology, Design and Environment, Oxford Brookes University, Oxford, UK [gjwood@brookes.ac.uk](mailto:gjwood@brookes.ac.uk)

integrated and linked within a BIM environment. As a result, BIM can deal with large volumes of heterogeneous data which would more often than not, result in an interoperability barrier. This massive accumulation of data in BIM platforms has pushed the implementation of Big Data in the construction industry, especially in BIM applications and platforms. However, although Cloud computing and Big Data analytics can improve interoperability, the adoption of Big Data in BIM for asset management remains at a nascent stage lacking clear frameworks and guidelines. The benefits of asset management database based on BIM and BIG Data can include improved financial performance, informed asset investment decisions, improved services and outputs, managed risks, and demonstrated compliance (Spilling 2016).

To address this, the paper firstly presents the Big Data definitions and the different stages the data passes through, as well as the critical BIM requirements for the Big Data platform. Then, taking into consideration BIM requirements and armed with available researches and tools dealing with the problem, a conceptual framework of the Big Data platform for BIM in asset management is proposed.

## 2 BIG DATA FOR BIM

### 2.1 Big Data Definitions

The term “Big Data” was coined in 2005 by Roger Mougulas of O’Reilly Media (Sangeetha and Sreeja 2015). He used the term to refer to a huge set of data that is almost impossible to manage and process using traditional business intelligence tools. In the same year, the open-source Hadoop - one of the most-used platforms for Big Data- was created by Yahoo based on Google File System and MapReduce. Big Data can be described and defined by several characteristics. The V’s definition started in 2001 in the META group (now Gartner) (Laney 2001). Actually, the Gartner report did not use the term “Big Data” and predates the current trend. However, the report has since been chosen as a key definition for the term ‘Big Data’. Laney (2001) proposed a threefold definition encompassing the “three Vs”: Volume, Velocity, Variety was proposed. Volume focuses on the size of the data set, velocity indicates the data processing and variety describes the range of data types, schemas and sources. IBM (2012) expanded this definition and introduced a fourth V, Veracity. Veracity corresponds to the extent to which data can be trusted. Moreover, Biehn (2013) extended the definition to include a fifth V, Value. Value refers to the monetary worth that an organisation can derive from processing Big Data. Besides, other V’s such as variability, virtual, visualisation and viability are also mentioned in the literature serving complement characteristics of Big Data (Van Rijmenam 2014, Wang et al. 2016).

### 2.2 Big Data Stages

Generally in data processing world, the data goes through the process of capture, store, process, analyse, and produce results (Casado and Younas 2015). Accordingly, Big Data can be carried out in four stages including Big Data capture, Big Data storage and processing, Big Data analytics, and Big Data visualisation and interpretation. However, the data life cycle does not strictly follow this sequence (Casado and Younas 2015).

Big Data capture stage is the stage where data is received from different technologies in heterogeneous formats. This data can be classified into three categories namely, structured, semi-structured and unstructured data. Structured data refers to the tabular data in spreadsheets or relational databases which can be stored in structured query language (SQL) databases. Semi-structured data refers to data that does not reside in a

relational database, nevertheless has associated information. Extensible Mark-up Language (XML), a textual language for exchanging data on the Web, is classified as semi-structured data. Unstructured data is available as audio files, images, videos, presentations, and amorphous texts such as email, and blogs. Depending on the data format and classification the data storage and processing stage differs.

Big Data storage and processing refers to the storage and management of large-scale datasets and accomplishing availability for this data (Chen et al. 2014). A data-storage system consists of two main components, hardware infrastructure and data-storage methods/techniques. The data-storage methods are deployed on top of the hardware infrastructure to handle the data. Traditionally structured relational database management systems (RDBMS) based on SQL cannot handle and maintain the variety of Big Data. Different systems and databases have been developed to meet the demands of Big Data storage such as consistency, availability and partition tolerance (Chen et al. 2014). This Big data storage consists of distributed file systems and NoSQL databases. Distributed file systems is a method of storing and accessing files in one or more central servers. These files can be accessed, with proper authorization rights, by any number of remote clients on the network. In the market there are different competing distributed file systems such as Hadoop distributed file system (HDFS), Parallel virtual file system (PVFS), Google file system (GFS) and ZFS Lustre (Shvachko et al. 2010). While NoSQL database stands for not only structured query language database. NoSQL database systems provide a mechanism to store and manage data in a non-relational data model (semi-structured and unstructured data) without prohibiting relational data. DB-Engines lists 309 different database management systems (DBMS) classified in relation to their database model (Solid 2015). Currently, the most popular NoSQL databases include graph DBMS, document stores, wide column store, search engines, and Key/Value DBMS (Solid 2015).

Big Data Analytics is the process/stage where advanced analytic techniques operate on Big Data to reveal hidden patterns, unknown correlations and other useful information. Like Big Data itself, the analytics evolution has been made possible by a number of key innovations mainly related to advances in capturing, storing, and processing data capabilities (Berson et al. 2004, Marr 2015). This stage is focusing on finding patterns in Big Data. As the world grows in complexity and overwhelming data is generated, data mining becomes the only process for clarifying the patterns that underlie this Big Data (Witten and Frank 2005). Data mining process applies methods from many different areas including computer science, artificial intelligence and mathematics in order to identify unknown patterns in already stored data (Witten and Frank 2005). These methods could include statistical algorithms, machine learning, content analytics, and time-series analysis. In the market, the two main stacks of horizontal scaling are Hadoop Big Data Analytics Stack and Spark Big Data Analytics Stack.

Big Data visualisation and interpretation stage is the stage of human-computer interaction where the data is presented in a pictorial or graphical format. This stage is important because the analysed data/information generated from the Big Data Analytics stage must be standardised, interpreted, and visualised to help users to extract knowledge and make decisions (Zhong et al. 2016). However, Big Data visualisation could make information more expressive. However, due to the complexity of the Big data, data visualisation is more challenging than traditional cases (Wang et al. 2016). Different visualisation approaches can be used to visualise the Big Data such as object-oriented, location-based (interactive maps), and network visualisation. Generally, open-source and web-based data visualization tools are preferred by end-users for free application which could be easily interoperate and integrate with existing systems (Bughin et al. 2010). From

this perspective, cloud computing has been emerged in the Big Data visualisation and interpretation stage and even in further stages like storage, processing and analysis.

### 2.3 BIM Requirements in the Big Data platform

The identification of BIM requirements is the first step in choosing the appropriate Big Data platform. These requirements can be categorised according to the three phases where the raw data is transformed to information, knowledge and wisdom enabling better decision making in the design, construction, operation and maintenance of building facilities. The three phases are data to information, information to knowledge and finally knowledge to wisdom.

**Data to information stage:** A huge volume of input data is present in the BIM platforms during design, construction and operation. These data have to be handled, managed and categorised to generate information. Meanwhile, further data from different technologies such as Geographical Information Systems (GIS), Internet of Things (IoT), energy sensors, and materials and resources databases have to be engaged and integrated with the BIM data. However, this data has to be firstly extracted from their authoring tools in a proper format to integrate with other surrounding data. Industry Foundation Classes (IFC) format is the most common schema for exporting BIM data. Still, different schema has to be taken into consideration as IFC is not a suitable choice for real-time queries (Solihin and Eastman 2016). A suitable schema that allows fast and efficient queries, and deals with graphical information and geographical reference location is required. Once the data is exported, it has to be stored and managed in a database which can handle large volumes of data, stream data, integrated geometry and unstructured data.

**Information to knowledge stage:** The data stored in the database from different sources has to be smoothly and rapidly integrated to reach the knowledge related to all building aspects. The platform has to be able for high-performance read. In other words, the data is written once and read many times. The platform has also to accumulate massive BIM data and other technologies data, and run multiple services concurrently for access by multiple users (Chen et al. 2014). Furthermore, the platform needs to provide techniques and tools to find the pattern/relation between the data from the different sources. Consequently, the data is managed and categorised based on object instead of source criteria.

**Knowledge to wisdom:** Once the data is integrated, efficient visualisation and secured access are required in a user friendly platform for user interpretation. Cloud computing environment is also a requirement to allow better collaboration between stakeholders.

## 3 CONCEPTUAL FRAMEWORK

Conceptual system framework provides an important foundation for the system software development. Based on Zachman's framework (2002), the proposed conceptual framework can be classified as an application architecture framework where the principle perspective is designer's perspective (Row 3) and the abstraction is function (Column 2). The proposed framework is formulated by synthesizing concepts from peer-reviewed papers and integrating four empirically tested and proven research developed prototypes. These models are a tender price evaluation system (Zhang et al. 2015), cloud-based online system for big data of massive BIMs (Chen et al. 2014), construction waste analytics system (Bilal et al. 2016), and a simplified BIM model server on a Big Data platform (Solihin and Eastman 2016). Figure 1 illustrates diagrammatical overview of the BIM Big Data system

framework for asset management in its four levels. This is based on Big Data phases which include extracting, integrating, analysing and interoperating data.

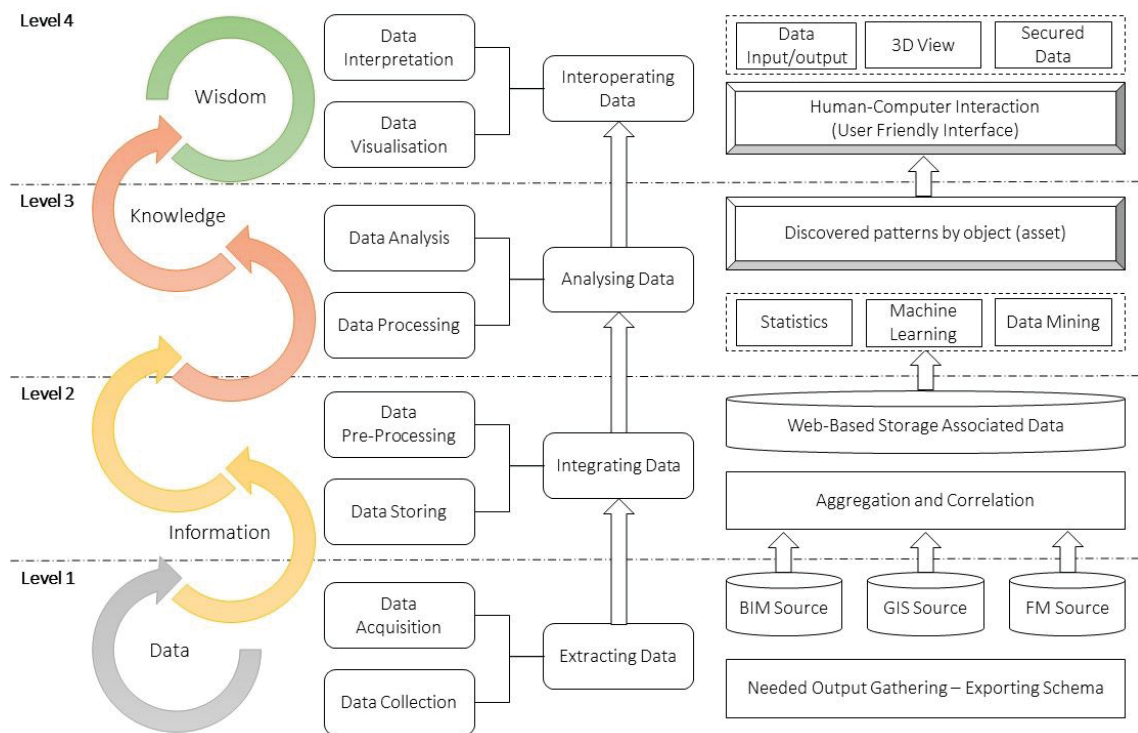


Figure 1: BIM Big Data framework for asset management

### 3.1 Level 1 - Extracting Data

This is the foundation level of the system and the most important stage. It mainly focuses on data collection and acquisition related to building assets. Data from the BIM model, GIS and sensors are included. Structured, semi-structured and unstructured data are extracted from the different sources through automatic data extraction mechanism (suitable schema). This level also includes data cleaning and quality assessment based on the required output.

Generally for BIM data extraction, the first choice of BIM exporting tool is the Industry Foundation Classifications (IFC), mainly because IFC is an open, vendor-neutral BIM data repository for the semantic information of building object and supported by plenty of platforms. However, the IFC language is significantly large and complex and its definition of IFC includes 327 data types, 653 entity definitions, and 317 property sets which would result exporting asset data which is not required (Steel et al. 2012). Each individual data exchange specifies a compulsory dataset which is covered by a small part of the whole IFC model. Implementation of the model view definitions (MVD) concept can overcome the IFC generalisation (Venugopal et al. 2015). MVD is the international standard concept of model (Hietanen and Lehtinen 2014) and referred as Building Compliance Model (Dimyadi et al. 2016). Based on the platform requirements and needs from the BIM model, the exported BCM subset can be defined as Asset Compliance Model (ACM) (Hietanen and Lehtinen 2014). BIMQL (developed operators for query languages) could be a promising solution for MVD (Dimyadi et al. 2016, Ying et al. 2016). Other query languages can be used such as Java, SQL, EDMexpressX, EQL, PMQL, and ProMQL (Ying et al. 2016). Also in this stage, a meaningful list of what data/information is needed to operate the assets

has to be specified. An additional requirement is the development of a taxonomy for the required data in order to facilitate exporting the data (Love et al. 2014, Mayo and Issa 2015, Ibrahim et al. 2016).

### 3.2 Level 2 - Integrating Data

After data extraction, it is necessary to store, process and manage building assets data extracted from various sources. This data is stored and managed in a web-based storage for massive associated data where aggregation and correlation of data occurs. In other words, this level aims to transform the raw extracted data to data stored according to certain rules and form Big Data centre/pool (Zhang et al. 2015).

The chosen/required database requires dealing with various relationships supported in the BIMRL or other query languages (Exporting query language in level 1). Meanwhile, the database has to deal with other data such as tabular data (from Sensors and GIS) and documents (asset specifications). Nowadays, there is a variety of non-relational sequence query language (NOSQL) databases which are suggested to be multi-mode NOSQL databases (Solihin and Eastman 2016). Graph database stores data in vertices and edges and it is the most suitable NOSQL database for our purpose as it allows the relationship/pattern to be directly linked through its edges. The most popular graph databases used with BIM data are Neo4J (Bilal et al. 2016) and OrientDB (Dimiyadi et al. 2016). Meanwhile, a key-value store database is required to handle the stream simple structured data received from the sensors (Solihin and Eastman 2016).

### 3.3 Level 3 - Analysing Data

Further data processing and analysis are continued by using mathematical and statistical methods, data mining and machine learning algorithms. This stage can provide the correlation law discovered from the asset data (integrated from different sources) for facility managers' decision making. At the same time, simple data visualisation techniques are considered a requirement for presenting the results effectively.

Horizontal scaling stack is required to analyse the data and find a pattern to integrate the data coming from the different sources and stored in the NOSQL database. Horizontal scaling platforms distribute processing across multiple servers and scale out by adding multiple machines to the cluster to increase the speed and performance. The selection is mainly influenced by the requirements for achieving interoperability between BIM, GIS and FM data which would include iterative algorithms, compute intensive tasks and near real-time visualisation. Based on the research finding, the most popular two stacks of horizontal scaling able to achieve the requirements are Hadoop Big Data Analytics Stack and Spark Big Data Analytics Stack.

### 3.4 Level 4 - Interoperating Data

The top level is the human-computer interaction level, including data input, data output, secured data, results of statistical analysis, 3D view, etc. in a user friendly interface. This level is crucial as the analysed data/information generated from level 3 must be standardised, interpreted and visualised to help facility managers and other end-users to extract knowledge and identify new patterns from visualised information (Zhong et al. 2016). A secured web-based software as a service (SaaS) is required, where the users are provided access to enterprise applications using multiuser architecture via the internet. In literature, different technologies were emerged to serve display of BIM in 3D viewer on browser such as WebGL and HTML5 (Chen et al. 2014), Gaming Engines like Unity 3D web player (Lee et al. 2016), and DWF and PHP (Nakama et al. 2015).

## 4 CONCLUSION AND FURTHER WORK

This study proposed a BIM Big Data based system framework for capturing, storing, analysing and visualising data from BIM, GIS, sensors and asset databases for data integration in order to enable improved asset management decisions. The proposed system framework can provide online services (cloud computing) that allow facilities managers, BIM users and asset managers to interact with asset data in a 3D environment. The system, which can be defined as a BIM platform that facilitates integration of data from different systems, consists of four main modules in four different levels. Module 1 is responsible for extracting the needed/required data from the BIM model through a schema such as BIMRL. Module 2 is the NOSQL database (Graph database) where the different data is stored, managed and integrated. Module 3 is the Big Data analytic stack (Hadoop or Spark) where advanced analytic techniques operate on Big Data stored in Module 2 to reveal hidden patterns and categorise the data based on the object (asset) instead of the source. Finally, Module 4 is the Human-Computer interaction where the integrated data is presented as 3D components through a gaming engine (Unity 3D) or web programming language (WebGL and HTML). This study is part of an ongoing research which aims to develop a BIM Big Data system for asset management. Further work will involve developing the system and evaluate it on real-world case studies.

## REFERENCES

- Berson, A., Smith, S. and Thearling, K. (2004). *An overview of data mining techniques, Building Data Mining Application for CRM*.
- Biehn, N. (2013). The Missing V's in Big Data: Viability and Value, *Wired. Innovation insights community*, 5.
- Bilal, M., Oyedele, L. O., Akinade, O. O., Ajayi, S. O., Alaka, H. A., Owolabi, H. A., Qadir, J., Pasha, M. and Bello, S. A. (2016). Big data architecture for construction waste analytics (CWA): A conceptual framework, *Journal of Building Engineering*, 6, 144-156.
- Bughin, J., Chui, M. and Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch, *McKinsey Quarterly*, 56(1), 75-86.
- Casado, R. and Younas, M. (2015). Emerging trends and technologies in big data processing, *Concurrency and Computation: Practice and Experience*, 27(8), 2078-2091.
- Chen, M., Mao, S., Zhang, Y. and Leung, V. C. (2014). *Big data: related technologies, challenges and future prospects*, Springer.
- Dimyadi, J., Solihin, W., Eastman, C. and Amor, R. (2016). Integrating the BIM Rule Language into Compliant Design Audit Processes, translated by Brisbane, Australia.
- Hietanen, J. and Lehtinen, S. (2014). *MVD AND SIMPLEBIM*, Datacubist Oy.
- IBM (2012). The Four V's of Big Data, [online], available: <http://www.ibmbigdatahub.com/infographic/four-vsbig-data> [accessed
- Ibrahim, K. F., Abanda, F. H., Vidalakis, C. and Woods, G. (2016). BIM for FM: Input versus Output data, translated by Brisbane, Australia.
- Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety, META Group Research Note, 6, 70.
- Lee, W.-L., Tsai, M.-H., Yang, C.-H., Juang, J.-R. and Su, J.-Y. (2016). V3DM+: BIM interactive collaboration system for facility management, *Visualization in Engineering*, 4(1), 1.

- Love, P. E., Matthews, J., Simpson, I., Hill, A. and Olatunji, O. A. (2014). A benefits realization management building information modeling framework for asset owners, *Automation in construction*, 37, 1-10.
- Marr, B. (2015). *Big Data: Using SMART big data, analytics and metrics to make better decisions and improve performance*, John Wiley & Sons.
- Mayo, G. and Issa, R. R. (2015). Nongeometric Building Information Needs Assessment for Facilities Management, *Journal of Management in Engineering*, 32(3), 04015054.
- Nakama, Y., Onishi, Y. and Iki, K. (2015). Development of building information management system using BIM toward strategic building operation and maintenance, translated by.
- Sangeetha, S. and Sreeja, A. (2015). No Science No Humans, No New Technologies No changes. Big Data a Great Revolution, *International Journal of Computer Science and Information Technologies*, 6, 3269-3274.
- Shvachko, K., Kuang, H., Radia, S. and Chansler, R. (2010). The hadoop distributed file system, translated by IEEE, 1-10.
- Solid, I. (2015). DBMS popularity broken down by database model, [online], available: [http://db-engines.com/en/ranking\\_categories](http://db-engines.com/en/ranking_categories) [accessed
- Solihin, W. and Eastman, C. (2016). A Simplified BIM Model Server on a Big Data Platform, translated by Brisbane, Australia.
- Spilling, M. (2016). *Asset Management Surveying Practice*. The British Institute of Facilities Management (BIFM).
- Steel, J., Drogemuller, R. and Toth, B. (2012). Model interoperability in building information modelling, *Software & Systems Modeling*, 11(1), 99-109.
- Van Rijmenam, M. (2014). *Think Bigger: Developing a Successful Big Data Strategy for Your Business*, AMACOM Div American Mgmt Assn.
- Venugopal, M., Eastman, C. M. and Teizer, J. (2015). An ontology-based analysis of the industry foundation class schema for building information model exchanges, *Advanced Engineering Informatics*, 29(4), 940-957.
- Wang, H., Xu, Z., Fujita, H. and Liu, S. (2016). Towards felicitous decision making: An overview on challenges and trends of Big Data, *Information Sciences*, 367, 747-765.
- Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann.
- Ying, H., Lee, S. and Lu, Q. (2016). Comparative analysis of the applicability of BIM query languages for energy analysis, translated by Brisbane, Australia.
- Zachman, J. (2002). The zachman framework for enterprise architecture, *Zachman International*, 79.
- Zhang, Y., Luo, H. and He, Y. (2015). A system for tender price evaluation of construction project based on big data, *Procedia Engineering*, 123, 606-614.
- Zhong, R. Y., Newman, S. T., Huang, G. Q. and Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives, *Computers & Industrial Engineering*.