



Challenges and Drivers for Data Mining in the AEC Sector

Journal:	<i>Engineering, Construction and Architectural Management</i>
Manuscript ID	ECAM-01-2018-0035.R1
Manuscript Type:	Original Article
Keywords:	Decision Support Systems, Information and Communication Technology (ICT) Applications, Technology
Abstract:	

SCHOLARONE™
Manuscripts

Challenges and Drivers for Data Mining in the AEC Sector

Abstract

Comment [u1]: Abstract was reviewed as requested by Reviewer 2

Purpose: This paper explores the current challenges and drivers for data mining in the AEC sector.

Design/methodology/approach: Following a comprehensive literature review, the data mining concept was investigated through a workshop with industry experts and academics.

Findings: The results showed that the key drivers for using data mining within the AEC sector is associated with the sustainability, process improvement, market intelligence, cost certainty and cost reduction, performance certainty and decision support systems agendas in the sector. As for the processes with the greatest potential for data mining application, design, construction, procurement, forensic analysis, sustainability and energy consumption and reuse of digital components were perceived as the main process areas. While the key challenges were perceived as being, data issues due to the fragmented nature of the construction process, the need for a cultural change, IT systems used in silos, skills requirements and having clearly defined business goals.

Originality/value: With the increasing abundance of data, business intelligence and analytics and its related concepts, data mining and big data have captured the attention of practitioners and academics for the last 20 years. On the other hand, and despite the growing amount of data in its business context, the AEC sector still lags behind in utilising those concepts in its end products and daily operations with limited research conducted to explore those issues at the sector level. This paper investigates the main opportunities and barriers for Data Mining in the AEC sector with a practical focus.

Keywords: Business analytics, Data Mining, Data Analytics, AEC, Facilities Management

Paper type: Research paper

Introduction and Background

Business Intelligence and Analytics (BI&A) and the related field of big data analytics have gained great significance over the past two decades (Chen et al., 2012). Top performing organizations are three times more likely to use sophisticated methods in their use of data analytics than lower performers, making analytics a major differentiating factor (Davenport & Harris, 2007; Hopkins et al, 2010). BI&A is often referred to as the techniques, technologies, systems, practices, methodologies, and applications that analyse critical business data to help an enterprise better understand its business and make timely business decision (Chen et al., 2012). According to a report by Gartner (2015), with the mobile and cloud computing technologies, the dynamic BI&A market is undergoing a fundamental shift from highly centralised, IT-led consolidation and standardisation projects to a more mobile, user-centric and interactive styles of analysis and reporting without requiring the users to have static location and IT/data science skills. These business-centric practices and methodologies can be applied to various high-impact applications such as in: e-commerce, market intelligence, e-government, healthcare, security, determining human/asset location, and manufacturing etc. (Turban et al., 2008; Brown et al., 2011).

A closely related term to BI&A is Data-Driven Decision making (DDD), which refers to the practice of making decisions on the analysis of data rather than purely on intuition, and can translate into increased productivity and market value (Provost and Fawcett, 2013). However, the act of gathering and storing large amounts of data for eventual analysis and decision-making is not new. Over the past 20 years, the amount, type and flow of data in many organizations have exponentially increased and is now classified as Big Data (Gantz and Reinsel, 2011; Chen et al., 2014). Moreover, Big Data is accumulating from multiple sources such as systems, sensors and mobile devices etc. at an alarming velocity, volume and variety to an extent that 90% of the data in the world today has been created in the last two years alone (Zikopoulos et al., 2013; IBM, 2017). This influx of data offers insights for making right decisions at the right time based on efficient aggregation and analysis of internal and external data in organisations.

Given the competitive and complex nature of the construction industry, and the constant transformation in its methodologies and processes, more and more businesses need to rely on fact-based decision making and analysis to guarantee their survival (Witten et al., 2016; Shmueli et al., 2017; Tixier et al. 2017). However, the topic is still novel and the volume of research in the AEC domain, particularly in terms of works

investigating the subject at the sector/industry level is limited (Bilal et al., 2016; Ahmed et al., 2017). This paper therefore attempts to develop an understanding of the potential and challenges of data analytics and data mining in the AEC sector and to present a scoping a study, providing a ground for future research in the field.

Business Intelligence, Analytics and Data Mining: What is the difference?

Business intelligence (BI) is an umbrella term that includes the applications, infrastructure and tools, and best practices that help gain access to and analysis of information to improve and optimize decisions and performance (Gartner, 2017). Business Analytics is considered a subset of BI however, there are some differentiating factors which can be understood from Table 1.

Table 1: Distinguishing factors between BI and Analytics

Business Intelligence	Business Analytics	Reference
<ul style="list-style-type: none"> ▪ a broader concept that comprises of technologies and processes used to understand and analyze business performance ▪ encompasses accessing data, reporting and analytics to provide answer questions about an organization's business activities 	<ul style="list-style-type: none"> ▪ a subset of business intelligence ▪ involves rigorous use of data, statistical and quantitative analysis, explanatory and predictive modeling, and fact-based management to drive decisions and actions 	Davenport and Harris (2007); Power et al. (2015)
<ul style="list-style-type: none"> ▪ strong association with conventional, technical reporting 	<ul style="list-style-type: none"> ▪ generates greater value and has a more proactive approach towards data than conventional BI 	Davenport and Harris,(2007); Laursen and Thorlund (2010)
<ul style="list-style-type: none"> ▪ employs a fixed set of metrics for the evaluation of historical performance and for carrying out future planning ▪ only queries and report generation does not guide actions to be taken for the improvement of the organization and so lacks decision support system capabilities 	<ul style="list-style-type: none"> ▪ involves use of more dynamic and developed problem solving tools or methods such as optimization ▪ involves gaining future insights and analyzing data to extract meaningful information and actionable decisions which shows its decision support system capabilities. 	Wang (2014)

▪ develops intelligence through descriptive analytics majorly	▪ inclination towards automation or prescriptive analytics and improved forecasting models	Gangotra and Shankar (2016); Laursen and Thorlund (2010)
▪ uses structured data often stored in relational database management systems	▪ can make use of unstructured data that is widely available through various sources such as the web and the internet	Chen et al. (2012)

Table 1 shows that Business Analytics is BI that relies on statistical modelling, predictive and forecasting techniques. Davenport (2013) suggests that there are three types of analytics; descriptive, which reports on the past to report ‘what has happened?’; predictive, which uses models based on past data to predict the future and to determine ‘what could happen?’; and prescriptive, which uses models to specify optimal behaviours, actions and answer ‘what should we do?’.

Data Mining analyses large quantities of data in order to discover meaningful patterns by using a combination of pattern-recognition rules, statistical rules, as well as rules drawn from machine learning (an area of computer science) (Ledolter, 2013; Larose, 2014). Figure 1 highlights the positioning of data mining on a Business Analytics spectrum. It can be seen that it mainly falls under predictive analytics where the complexity is higher than descriptive analytics. According to Camm et al. (2015), Data Mining can be classified as:

- *Supervised learning*: defined as a set of tools used for prediction (linear model, logistic regression, linear discriminant analysis, classification trees etc.)
- *Unsupervised learning*: an exploratory data analysis technique used for identifying groups (e.g. patterns, clusters, association rules etc.) in the data set of interest.

The ability to embed analytics and optimization in DDD process promises huge benefits for customers and markets (Davenport, 2013). The benefits have also been demonstrated conclusively in a study conducted on how DDD affects organizations’ performance (Brynjolfsson et al., 2011). The research shows statistically that one standard deviation higher on the DDD scale is associated with a 4–6% increase in productivity. The research also shows that DDD is correlated with higher return on assets, return on equity, asset utilization, and market value, and the relationship seems to be causal.

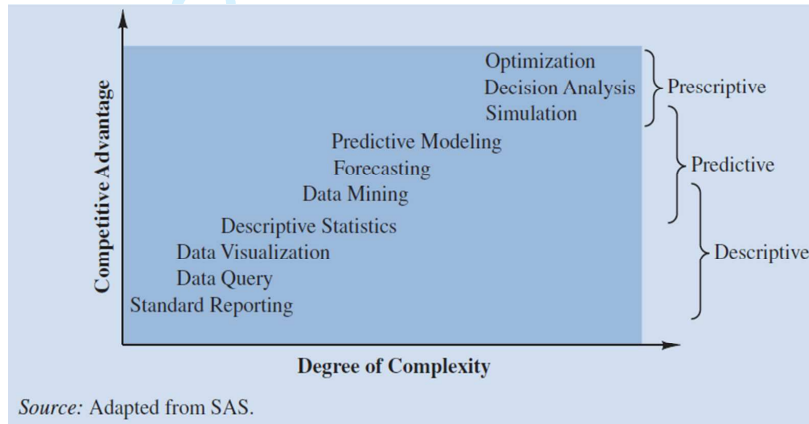


Figure 1: The spectrum of Business Analytics (Source: Camm et al., 2015)

Data and business performance in the AEC sector

The AEC sector, in general, is characterised by its slow pace of change, low productivity, waste, fragmentation and long established processes and ways of doing business that have not changed over the decades (Behera et al., 2015; Arashpour et al., 2017). Whilst there have been incremental improvements in workflows and processes, there has not been a fundamental change in how the business is done. In contrast, other industry sectors (e.g. Aerospace, Automotive, Aviation) have undergone paradigm shifts in the way they deliver value to their customers (Crowley, 1998; Green et al., 2005; Bonev et al., 2015). The increasing scrutiny over construction costs by clients, coupled with the increasing pressure on data transparency has made a strong case for better management of construction data.

Better utilisation of data and analytics is particularly important in early project decision making, which has a huge influence on the downstream life-cycle stages (Bilal et al., 2016). It is expected that in the near future, decision making will become more complex because of increasing demands and expectations, coupled with paucity of funds and fast paced technology growth. In early project stages, the ability to influence the key project parameters is maximum and the cost of any proposed design changes are minimal.

1
2
3
4
5
6
7
8 Thus, time and effort spent upfront in the planning phase could provide huge dividends
9 throughout the project lifecycle.

10
11 With many data formats flowing from incompatible software, such as Autodesk's
12 DWG (abbreviation of "drawing"), Bentley's DGN (abbreviation of "design") for
13 engineering activities, Microsoft's DOC/XLS/PPT (document format), Adobe's PDF
14 (document format), RM/MPEG (video format), and JPEG (image format), and constantly
15 increasing data volumes distributed over fragmented ICT systems (i.e. ERP systems) and
16 relational databases of different organisations, data in the AEC sector are typically of the
17 Big Data category (Jiao et al., 2013). Adding to the existing data volumes and
18 complexity, automatically data generating technologies such as laser scanning, advanced
19 photogrammetry, mobile computing and sensors have been fast penetrating into the
20 sector's daily operations (Chen and Kamara, 2011; Razavi and Haas, 2011; Kim et al,
21 2013). Alongside the technical data interoperability issues, due to the project centric,
22 dynamic and fragmented nature of the sector, the problem of data integration, and
23 subsequently data analytics, throughout the lifecycle of a project among multiple
24 collaborative organizations remain unsolved (Bakis et al., 2007; Cerovsek, 2011).

25
26 Recently, the Building Information Modeling (BIM) concept has been widely
27 discussed as a single project-life cycle data repository (Howard and Bjork, 2008; Babic
28 et al., 2010). Nonetheless, it is still a challenge to completely integrate both design (e.g.
29 geometric and parametric descriptions) and management (e.g. scheduling, monitoring,
30 work assignments) related data onto a singular data representation through the project-life
31 cycle with additional standardization, interoperability and data sharing issues (Jiao et al.,
32 2013). Ironically, with the improved capabilities of BIM with systems like seismic,
33 lighting and energy simulation in project design , the extension of the use of BIM models
34 from design to construction and maintenance , and the increasing integration of BIM
35 systems with mobile and cloud computing, even bigger data sets to manage have been
36 generated for the sector (Popov et al., 2010; Kim et al., 2011; Redmond et al., 2012).
37 Robust analytics that holistically infer from diverse data resources and those singular
38 project data representations are yet to come.

39
40 Given bigger data sets being generated within almost each project life-cycle and fast
41 developing Big Data analytics opportunities, the sector lags behind in obtaining value
42 from those data as an asset and DDD as a managerial competence (Bilal et al., 2016;
43 Ahmed et al., 2017). Organizations in the AEC sector can succeed based on their ability
44 to achieve constant yield from each project. There is tremendous value in turning
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

operational and transactional data 'into tactical information' in an industry where profit can be so severely influenced by exceptions (Kamara et al., 2002). BI&A can be used to analyse past performance for planning, identify exceptions for action and trend current performance for forecasting. Many construction executives often make decisions based on their intuitions and subjective judgements, justified by years of experience and expertise (Mohemad et al., 2010). However, today's fierce competition requires precision and process improvements to avoid any, if not all, risks associated with construction financial and project management, which requires BI&A to be an integral part of the sector's business strategy. The rest of this section looks into the potential of Data Mining in the AEC sector.

Data Mining in the AEC sector

Data mining is the application of specific algorithms or techniques for extracting knowledge patterns from data sets (Fayyad et al., 1996). The patterns discovered must be meaningful in that they lead to some advantage in decision making. Although the traditional way of making decisions in construction organisations is based on chief executives' experience and gut feelings, the fierce market conditions in construction today requires precision and process improvement to avoid any risks associated with financial and project management issues (Mehta, 2010). Being a relatively recent phenomenon for construction, data mining techniques were mostly introduced to the AEC sector at the beginning of the 2000s (Hammad et al., 2013). In the AEC sector, BI&A can be used in (i) analysing data from the generic, firm-wide level to the more specific, individual level, (ii) performing "what if" analyses for cash flow, revenue and profit, (iii) eliminating cumbersome Excel spread sheets, and (iv) integrating information from different financial software resources such as SAP or Tally for a complete view of the business (Mehta, 2010). Specific data mining applications exploiting Artificial Neural Networks, Genetic Algorithms, Neuro-Fuzzy Systems, Bayesian Networks, Case Based Reasoning, text mining, decision trees, clustering and regression on smaller, structured construction data warehouses have been applied and brought a number of benefits to the AEC sector;

- Historical project data can assist AEC professionals in answering specific questions about the business, the performance of interested operations, business trends, and what can be done to improve the business and operations in general

(Rujirayanyon and Shi, 2006). The practice of inference from that historic data is broadly called data mining, which is the act of extraction of implicit, previously unknown, and potentially useful information and knowledge from recorded data sets (Frawley et al, 1992).

- Data mining is also used for optimizing structural engineering design (Leu et al., 2001; Tinoco et al., 2011; Gandomi et al., 2016), energy efficient and sustainable building design (Kim et al., 2011; Naganathan et al., 2016), supporting collaborative design processes (Chiu and Lan, 2005), construction operation simulation (Akhavian and Behzadan, 2013), investment feasibility analysis (Yun and Caldas, 2009), future cost forecasting (Wilcot and Cheng, 2003), analysing building performance (Ahmed et al., 2009; Fan et al., 2015), construction accident analysis (Abdelhamid and Everett, 2000; Huang and Hinze, 2003; Ding et al., 2018), safety clash detection (Tixier et al., 2017), optimizing bid selection policies (Art Chaovalitwongse et al., 2011), hierarchical construction document classification (Caldas and Soibelman, 2003), asphalt paving pattern prediction and analysis for transportation projects (Nassar, 2007) and predictive cost overrun and mitigation analysis (Ahiaga-Dagbui and Smith, 2014; Williams and Gong, 2014).
- Data mining also provides useful tools that help to explain how building systems and facilities that were once thought to be completely chaotic have predictable patterns (Peitgen et al, 199; Fan et al., 2015). A report by Liew and Rosenblatt (2003) revealed that applying data mining techniques could greatly help facility managers to identify crucial maintenance cost patterns and to form preventive maintenance plans. Reffat et al. (2006) gave a classification of possible Facilities Management benefits with their corresponding data mining techniques: unveiling the correlations between facilities service parameters and malfunctions, and some other independent parameters (i.e. seasons or weather conditions), dynamic adjustments of lighting, temperature and air conditioning, prioritizing

1
2
3
4
5
6
7
8 maintenance works, calculating benefits from successful maintenance practices,
9 maintenance resource allocations and better planning in maintenance works.

10
11 Having outlined those data mining implementations and possibilities, it should be
12 noted most of the data mining discussions in construction are coming from specific
13 academic research efforts, which do not illustrate the general extent of the use of data
14 mining in the sector with its drivers and challenges. Previous studies of data mining in the
15 construction industry have mainly focused on introducing new data mining techniques,
16 comparing between techniques, and comparing between data mining techniques and
17 traditional methods (Yun and Caldas, 2009). Also, the amount of data within the sector
18 has become increasingly overwhelming, as well as the growth of many business,
19 government, and scientific databases coupled with the data generated by individuals and
20 even inanimate objects (i.e. sensors, mobile and wearable computing) has begun to far
21 outpace our ability to interpret and digest those data holistically with conventional
22 approaches.

23
24 Nassar (2007) illustrated some key issues relating to the implementation of data
25 mining in construction;

- 26
27
28
29 ● Lack of standardisation in the construction industry as to collecting and storing
30 project and company specific data. This industry fragmentation significantly
31 hinders the uptake of data mining techniques in practice,
- 32
33 ● The way in which the information is stored in the construction industry is
34 generally not very organised and patchy. In the research presented by Nassar
35 (2007), for example, various pieces of information had to be collected from
36 bulletins, reports, as well as electronic databases and then re- structured into one
37 database in order to facilitate data mining,
- 38
39 ● There is a need for a unified data model for construction data. This unified data
40 model would be similar to the current building product model utilized in the
41 Industry Foundation Classes (IFC) but would focus primarily on construction
42 specific data. One layer would capture project-specific data such as cost, estimate,
43 schedule and productivity. The second layer would capture company-specific
44 data, such as profitability, bids and bonding capacity. These two layers would be
45 obviously interrelated so that information can be indexed from one layer to
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9 another. The third layer would capture industry-specific data such as employment
10 rates, industry wide productivity rates and financial ratios.

11 A study by Silwattananusarn and Tuamsuk (2012) looked into the available literature
12 on data mining applications for Knowledge Management (KM). The study divided
13 knowledge resources into eight groups as to which knowledge object to be stored and
14 manipulated in KM and how data mining aids for health care, retailing, financial banking,
15 small and middle business, entrepreneurial science, collaboration and teamwork, general
16 business and construction organisations. The study revealed that only 19 articles
17 containing both “Data Mining” and “Knowledge Management” in their title, keyword or
18 abstract fields had been published between 2007 and 2012. For construction, the authors
19 stated that a large part of enterprise information was available in the form of textual data.
20 This leads to the influence of text mining techniques to handle textual information source
21 for industrial knowledge discovery and management solutions, which have been applied
22 in a very small number of conceptual research efforts in construction so far.

23 In light of this context, it can be inferred that there is a lack of empirical research
24 exploring the data mining concept for construction (Soibelman and Kim, 2002; Gajzler,
25 2010; Bilal et al., 2016; Ahmet et al., 2017). Therefore, this paper aims at understanding
26 the current challenges and drivers for data mining in the AEC sector.

27 **Research methodology**

28 This paper shares the outcome of an industry workshop that was carried out to develop
29 a scoping study for an SME IT group (XpertRule) in order to explore the potential and
30 challenges for utilising Data Mining in the AEC sector and whether these align with what
31 has already been identified from the literature. The focus group consisted of a mixture of
32 65 participants from the UK construction sector (AEC) and academia with a few
33 participants from IT service providers to the AEC sector. When researchers seek depth or
34 execute scoping studies at industry/sector level on a subject, focus groups are frequently
35 preferred to quickly capture rich insights from relevant individuals, enabling the
36 exploration of novel or complex phenomena (Love et al., 2010; Anvuur et al., 2011;
37 Liamputtong, 2011; Krueger and Casey, 2014; Loosemore, 2014; Loudoun and
38 Townsend, 2017).

39 Details of the workshop attendees can be seen in Table 2. The attendees were
40 identified and invited selectively. Senior and middle level managers from prominent

Comment [u2]: This section was expanded as per the request by the reviewers. Highlighted.

organisations in the industry were contacted initially for the workshop. Also, the academic participants were invited by relevance of their research interests in the built environment to the subject area. Contacts details of the participants were obtained from professional organisations, the industry partner's network, as well as from the authors' and their institutes' contact records.

Table 2: Details of the workshop attendees

Sector	Years of Experience in Sector			Grand Total
	<10	10-20	>20	
AEC		22	14	36
Client		3	2	5
Contractor		10	9	19
Designer		4	3	7
Facilities Manager		5		5
Higher Education	10	4	7	21
Lecturer	2	2	1	5
PhD Researcher	6			6
Professor		1	3	4
Research Fellow	2			2
Senior Lecturer		1	3	4
IT	2	4	2	8
Database		2		2
Software	2		2	4
Consultant		2		2
Grand Total	12	30	23	65

It was anticipated that the level of understanding of the subject among the AEC sector professionals and some Higher Education sector participants might have been limited. Therefore, four presentations on issues associated with construction data types, data management and data driven decision making in the AEC sector, the data mining and BI&A concepts, their current use in the AEC sector compared to other sectors, their possible future use across the project life-cycle, and opportunities facilitated by emerging IT solutions were given to the attendees by subject experts.

After the presentations, it was assumed that the participants had gained a satisfactory understanding of Business Intelligence, Data Mining and their applications across different sectors for the scoping research presented in the paper. Following the presentations and initial discussions, the participants' feedback was collated in two steps

Comment [u3]: For Reviewer 1 comment, to give more detail about the workshop attendees.

Comment [u4]: For Reviewer 1.

1
2
3
4
5
6
7
8
9 as explained below. At the end of the workshop, initial findings were presented to the
10 attendees and also, the attendees were contacted within 2 weeks after the workshop to
11 inform them of the collated findings and to obtain their further comments and suggestions
12 for validation. Some minor comments were received from especially senior managers and
13 academics that were incorporated into the final discussion.
14

15 The following steps were taken to analyse the focus group findings (Liamputtong,
16 2011; Krueger and Casey, 2014); (i) transcribing the written feedback and the discussion
17 voice records of the participants, (ii) noting the main ideas emerging for each question,
18 (iii) reviewing, classifying and grouping the recurring ideas, (iv) identifying the main
19 themes associated with those ideas, (v) reporting the main findings, and (vi) seeking
20 validation and reviewing the findings.
21

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Comment [u5]: For reviewer 2 comments.

Step 1 - Panel discussion

A panel formed of a senior construction manager, a senior academic in the built environment and a director of an IT company led the discussion about the potential for the use of Data Mining for Business Intelligence in the AEC sector.

Step 2 – Participants' feedback

The participants were asked to give written feedback to include three main points in response to the following four questions.

Q1: What are the key drivers to use Data Mining within construction?

Q2: Which processes within construction hold the greatest potential for Data Mining applications?

Q3: What are the key challenges to application of Data Mining within construction?

Q4: What steps should be taken by the construction industry to drive the uptake of Data Mining?

The rest of the paper shares the results generated and captured from the above steps.

Results and analysis

Step 1 - Panel Discussion

The panel discussion evolved around the attendees' feedback for the four main questions.

Based on the discussions, a few generic conclusions were made:

- Data Mining and BI&A is an area that is not well explored in construction and it requires academic research and investigation.
- There is a great potential in using Data Mining in order to improve the sustainability aspects of projects, reducing project waste, improving processes, improving market intelligence, supporting the decision making process and improving business certainty.
- Data Mining principles and concepts can be applied throughout the construction life cycle from conception to operation.
- The main challenges that face the application of data mining in construction have to do with obtaining the right data, understanding its relevance and importance, the intellectual rights, cultural and trust issues, skills and cultural changes that need leadership to drive this concept.
- There is a potential for construction companies and academics to work together on Knowledge Transfer Schemes.

Step 2 - Participants feedback

The participants were asked about the key drivers for using Data Mining in construction. Figure 2 shows the summary of the participants' feedback, whereby six key areas the participants identified are as such;

- Sustainability - Where Data Mining can play an important role in waste management and recycling on construction sites, which helps in responding to the sustainability agenda and managing the percentages of recyclable material and the source of waste. This in return plays an important role in improving the environmental factors around a project such as the percentages of pollutant materials and the level of pollution generated by the construction process. Project generated data can also help predict the time and costs wasted in the life cycle of a project and the pattern of behaviours around these processes to reduce waste.
- Process improvement - The participants' response under this category covered a wide range of activities to do with the different stages of a project from inception

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

to handover and beyond, whereby project data can help with increasing the operational effectiveness within a building, whether it related to the management of the occupancy of the building or energy consumption, identifying the causes of waste during the design and construction process, predicting risk and identifying health and safety measures to improve the project performance.

- Market Intelligence – The participants considered Data Mining and Business Analytics to play an important role in gaining a competitive advantage through a better understanding of the construction market and specific demand and supply patterns, as well as in serving the client needs to provide a product with a competitive advantage in terms of cost, time and quality.
- Cost Certainty and Cost Reduction - The participants' responses also aligned with the literature, which puts much emphasis on the use of Data Mining to predict cost certainty in the business world, and to work out patterns of behaviour for cost reduction, which in return helps to improve the project efficiency and the project profit margins.
- Performance Certainty – The participants felt that Data Mining techniques would be useful to improve the cost, time and quality of construction projects through gaining a better understanding of the history of project performance, and gaining a prediction of improved performance certainty for future projects.
- Decision Support Systems – The participants perceived Data Mining as a tool to help with decision support systems through the sharing of good practice, in order to understand the roots of problems and to enable the project team to create 'what-if-scenarios' to improve the decision making process, making sense of existing project strategies with an attempt for continuous improvement that is driven by data.

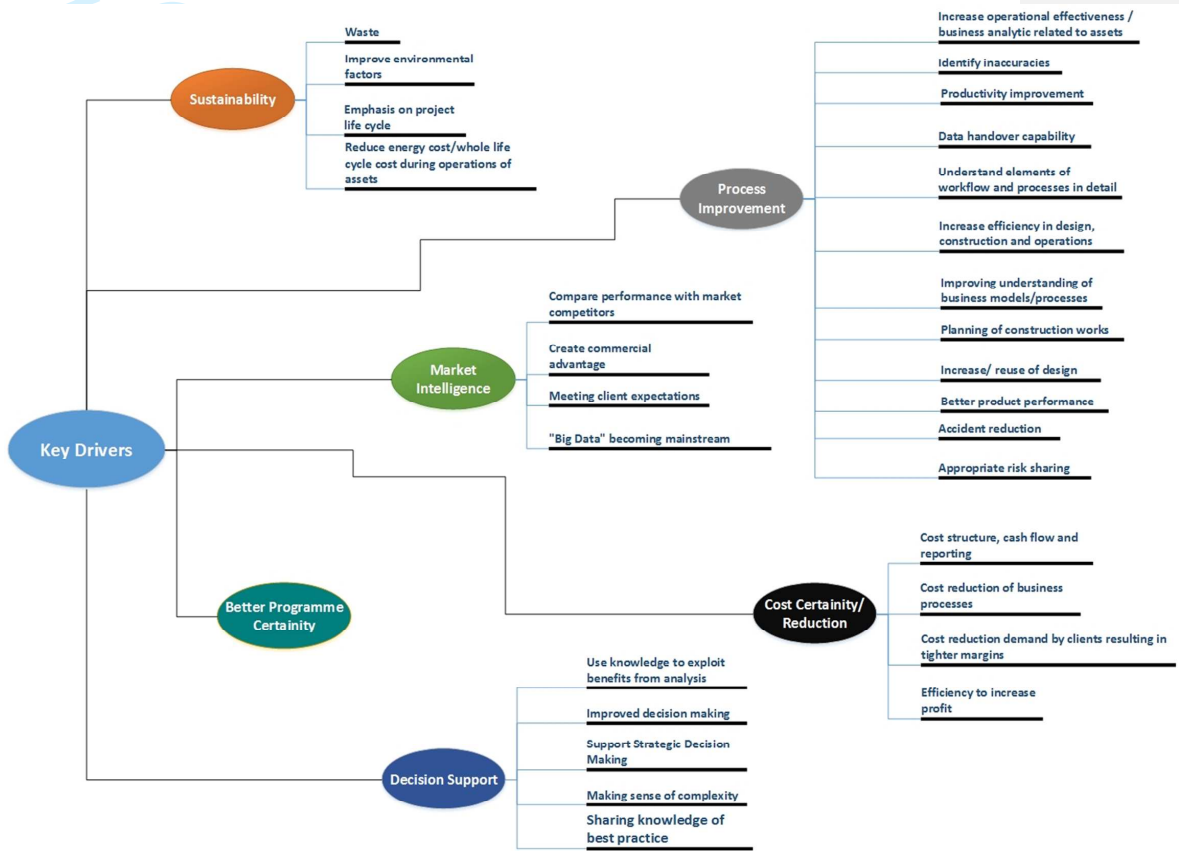


Figure 2: Key drivers to use Data Mining within construction

The participants were also asked to give their views of the 'Processes with the Greatest Potential for Data Mining Applications in Construction'. Figure 3 shows a summary of the participants' feedback, where design, construction, procurement, forensic analysis, sustainability and energy consumption and the reuse of digital components were perceived as the main process areas.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- Design – The Participants indicated that data generated from the design feedback loop gives architects and clients a good insight into the lessons learnt from previous projects at the design stage, and to define patterns of behaviours at this stage.

- Construction - The majority of the participants agreed that there is a lot of data generated from the construction stage particularly at the planning and programming stage, where Data Mining can analyse the time and cost efficiencies and efficiencies of different stages of projects to help enhance the performance of on-going, as well as future projects. In addition, the participants also saw great potential for Data Mining to improve the off-site manufacturing processes, where the same principles apply, as well as data on plant and machinery performance on construction sites.

- Forensic Analysis – The participants gave a broad range of factors that could benefit from Data Analytics and Data Mining techniques in order to develop an understanding of the general trends as to why project are delayed, project defects' causes and effects, the market trends in the AEC sector over the years in relation to the supply chain and the client requirements.

- Sustainability and Energy Consumption - The participants saw real potential for extracting data to detect patterns of behaviour through real time and truly embedded carbon processes and components such as the fabrication, transportation and extraction of construction materials and building components.

- Reuse of Digital Components – The participants recognise that there is a great potential in using Data Mining techniques, if the existing project digital materials, such as documents, images, designs, programmes etc. are made use of to extract the relevant data.

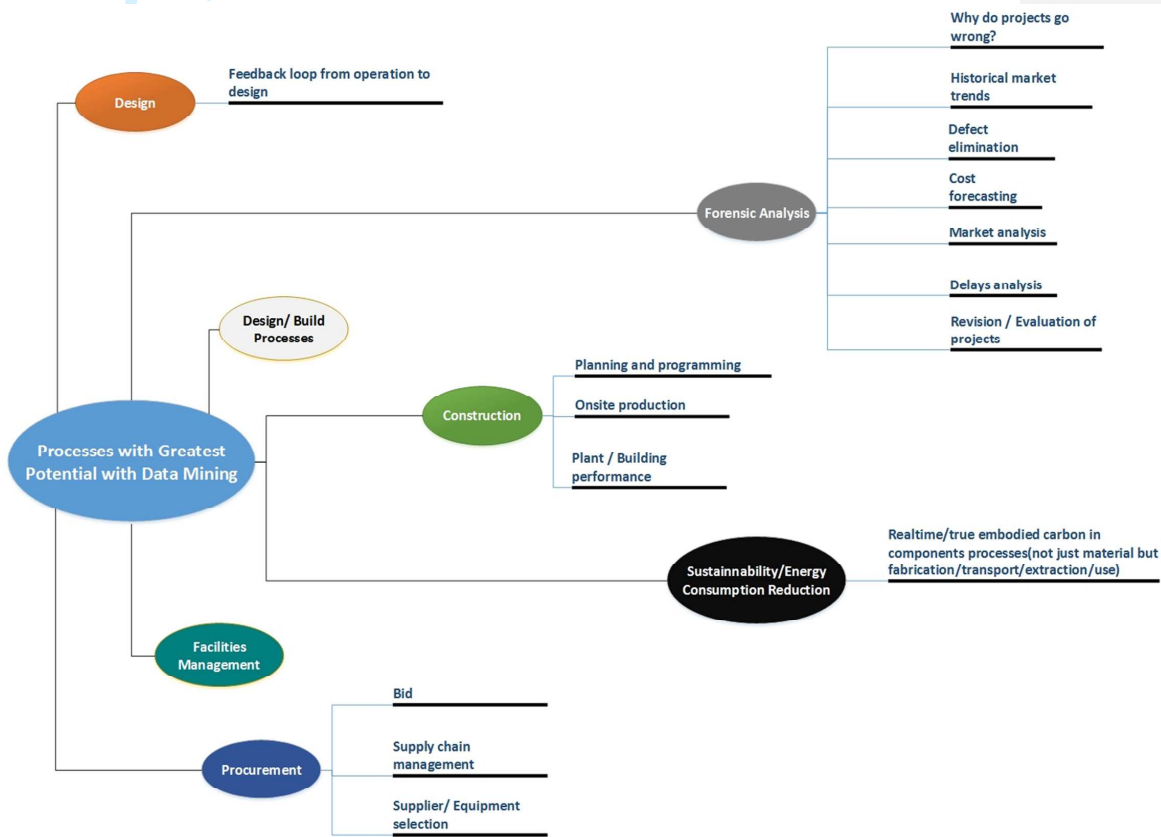


Figure 3: Processes with the greatest potential for Data Mining applications in construction

The participants were asked about the key challenges that face the Application of Data Mining within the AEC sector. The result in Figure 4 groups these challenges as being;

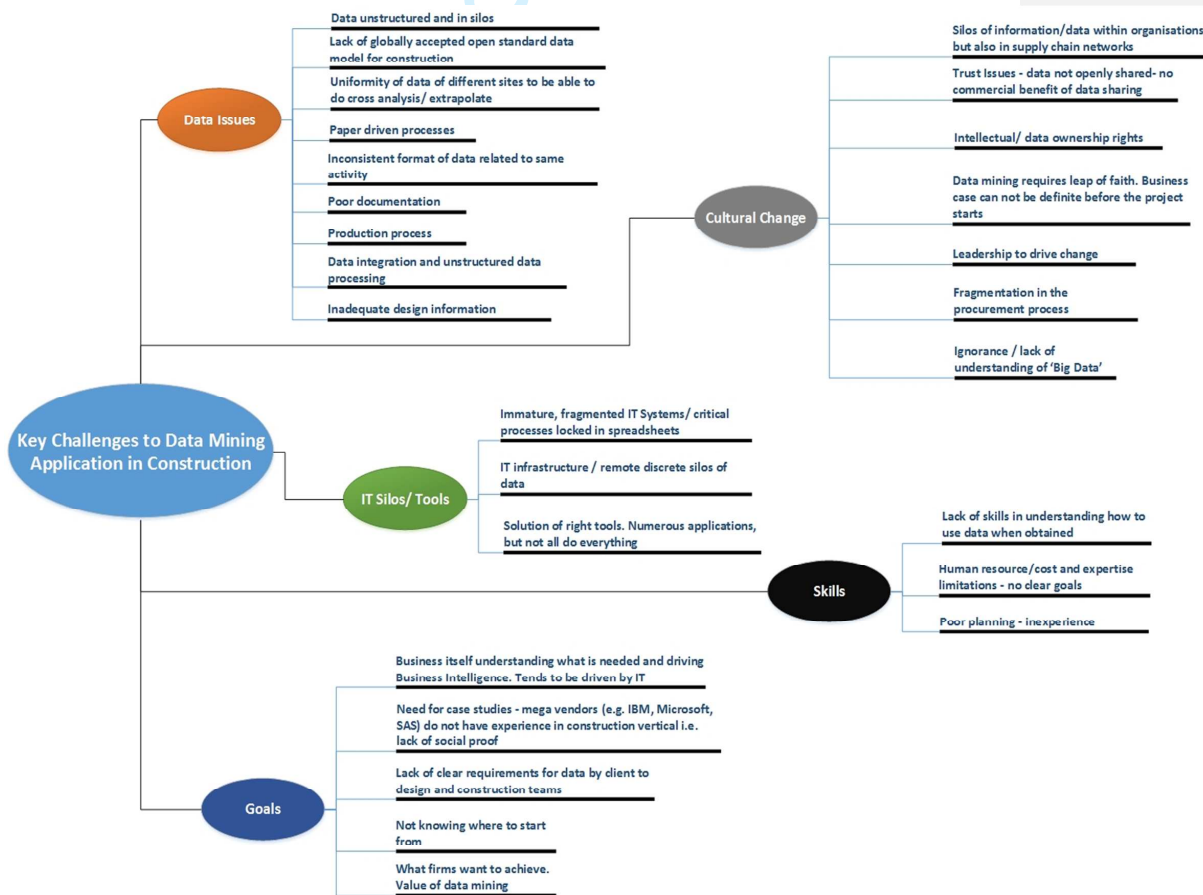
- Data Generation Issues - Much of the data related challenges evolved around the availability of massive data sets generated throughout the life cycle of a

1
2
3
4
5
6
7
8 construction project. The data is often times in a paper based format and/or
9 available in silos and in an unstructured manner.

- 10
11
12 - Data Formatting Issues – The participants perceived the attainment of data as one
13 of the main challenges that face the AEC sector due to the fragmented nature of
14 its processes, which forces some of the stakeholders to the projects to work in
15 silos, while the actual data is either available in paper format (not on a shared
16 drive where everyone can access), in inconsistent formats, or is unstructured,
17 poorly and inconsistently documented and not easy to access. The participants
18 also drew attention to the lack of an existing standard or a uniform model for
19 reporting and storing information, particularly when it comes to work generated
20 on different construction sites.
21
22 - Need for Cultural Changes - The participants clearly recognised the need for
23 cultural changes in the way people work, and share information openly for the
24 benefit of the business itself, where a certain level of trust needs to be achieved
25 with good leadership to drive this change. Leadership would also mean training
26 people to show the value in sharing project data, and investing in IT and
27 infrastructure systems. There also needs to be some clear protocols around data
28 sharing rights and intellectual property rights, while not many construction
29 companies have tackled these issues, particularly where much of the data are
30 generated through the design and procurement process.
31
32 - IT Tools Used in Silos - The participants emphasised the importance of having an
33 appropriate IT infrastructure that is capable of coping with different file formats
34 and sharing, distributing and updating information that sit in silos of spread
35 sheets. Therefore, having an IT structure that helps to deal and interpolate the data
36 is key.
37
38 - Skills Requirements - Another interesting set of challenges that was highlighted
39 by the participants evolves around the skills required to understand how the data
40 could be used when generated, and how to plan ahead to store the data so that it
41 makes sense when it is generated, and for there to be a clear differentiation
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

between automated systems that store and generate data, and systems that produce meaningful data to serve the business intelligence.

- Goals - The participants agreed that one of the main challenges is not knowing where to start from in terms of adopting data mining techniques, and that construction organisations need to have defined goals and business objectives to be achieved from the data mining with a realised value. Additionally, having clear requirement strategies from different stakeholders, such as the design and construction teams, is of critical importance. Therefore, to drive the value of data mining, the participants suggested that sharing good case studies from different software vendors would be useful.



1
2
3
4
5
6
7
8
9
10 **Figure 4: Key challenges to the application of Data Mining within Construction**

11
12 Finally, the participants were asked about the “Steps to be taken by the Construction
13 Industry to Drive the uptake of Data Mining”. The results can be seen in Figure 5. The
14 participant's responses were categorised as follows;
15

- 16 - Organisational Issues - This category mainly focused on the organisational mind-
17 sets, and the need to change the perceptions in terms of using IT and for it not
18 only to be a communication tool but for it to be a mean for Business Intelligence.
19 In alignment with the previous points made, the participants also felt that
20 openness and transparency (which are the resultants of trust) are key to providing
21 and sharing valuable data, and for this to happen, sufficient resources and
22 investments are required to promote good practices that will reap the long term
23 benefits.
24
- 25 - Benefit Realisation – Again, the participants emphasised the need for
26 organisations to demonstrate the benefits that can be realised through different
27 processes, where tendering was given as an example, as that much of the data
28 generated from this process can help with understanding the probability of
29 success or failure of project prospects.
30
- 31 - Technologies - The participants agreed that changing the mind-set to an
32 understanding that the technology is there to enable data sharing through
33 transparent and open environments, is a step forward toward integrating and
34 gathering the relevant data for business intelligence.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

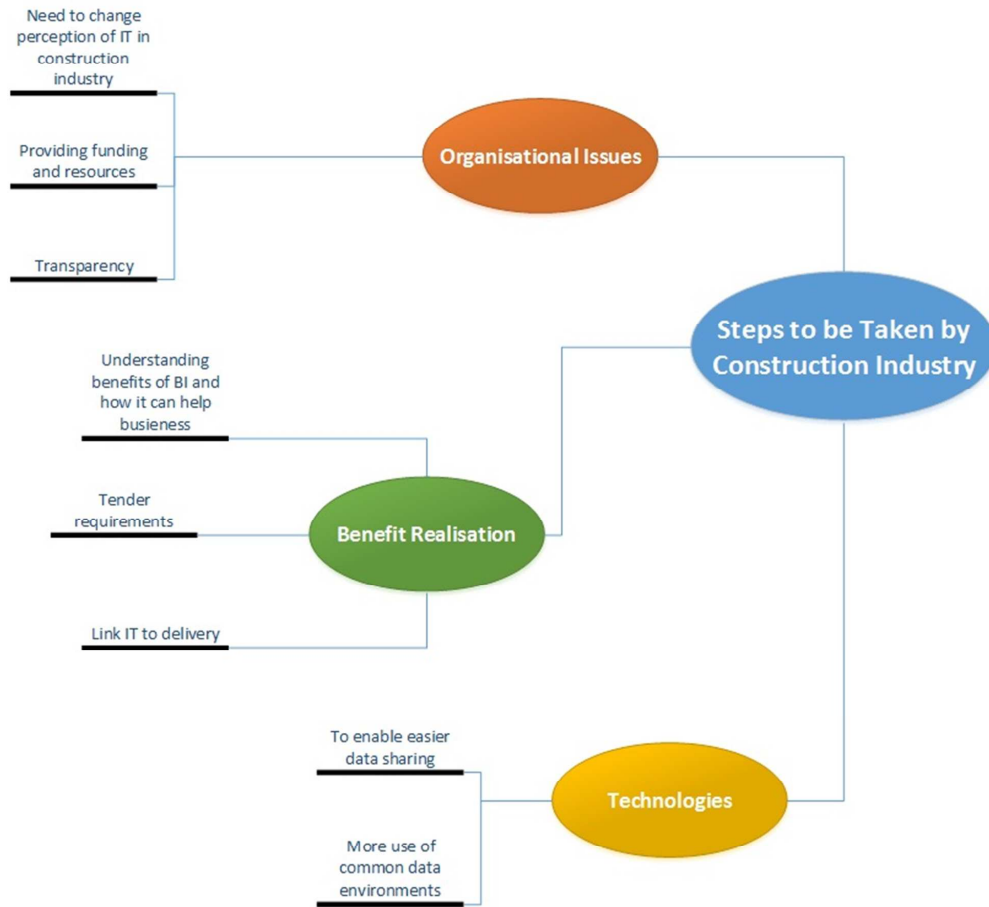


Figure 5: Steps to be taken by the construction industry to drive uptake of Data Mining

Summary and Conclusions

Business Intelligence and Analytics enables Data-Driven Decision making, which refers to the practice of basing decisions on the analysis of data rather than purely on intuition, that can translate into an increase in productivity, higher return on assets, return on equity, better asset and resource utilisation, and higher market value. This paper argued that although Data Driven Decision making has been widely utilised in different sectors and on the rise in the business world, beyond some specific cases or research

Comment [u6]: Reviewed and expanded

1
2
3
4
5
6
7
8 investigations demonstrating the use of some specific data mining techniques, its
9 utilisation and penetration have been somewhat limited in the AEC sector. The findings
10 are in accord with the recent sector level investigations by Bilal et al. (2016) and Ahmed
11 et al. (2017). The wider applications for the sector can lead to data-driven predictions in
12 project KPIs, such as cash flow, and how projects will generate or consume cash,
13 liquidity of project assets, labour productivity, factors that cause project and margin
14 variance, rectifying time and cost defects and improving project deficiencies; and better
15 predictions in asset whole-life performance. Therefore, to further identify challenges and
16 drivers for using Data Mining (Data-Driven Decision making) and the steps that should
17 be taken by the AEC industry to realise this potential, this paper shared the results of an
18 industry workshop which brought together 65 academics and practitioners to explore
19 these issues.

20
21
22
23
24 The results showed that there needs to be a cultural change in the way the sector
25 works and its mind set. Organisation leaders will play an important role in this by a top-
26 down approach, investing in technologies that bring together valuable data generated
27 from different sites and fragmented processes, investing in awareness raising and
28 training, and employing people with necessary skill sets (Ofori, 2008). However, it
29 should be noted that the use of those integrating technologies by different project
30 stakeholders should be facilitated as well with relational project delivery structures and
31 commercial configurations that enable data sharing and transparency between the
32 stakeholders (Porwal and Hewage, 2013).
33
34

35 The results also showed that the key drivers in using data mining within the AEC
36 sector are related to the sustainability, process improvement, market intelligence, cost
37 certainty and cost reduction, performance certainty and decision support system agendas
38 in the sector. As for the processes with the greatest potential for data mining applications,
39 design, construction, procurement, forensic analysis, sustainability and energy
40 consumption and reuse of digital components were perceived as the main process areas.
41 While the key challenges were perceived as being data issues due to the fragmented
42 nature of the construction process, the need for a cultural change, IT systems used in
43 silos, skills requirements and having clearly defined business goals.
44
45
46

47 **Acknowledgement**

48 The authors would like to acknowledge the contribution of support of XpertRule [
49 <http://xpertrule.com/>] and TSB for their support for this study.
50
51
52
53
54
55
56
57
58
59
60

References

- Abdelhamid, T. S. and Everett, J. G. (2000), "Identifying root causes of construction accidents", *Journal of Construction Engineering and Management*, Vol. 126 No. 1, pp. 52-60.
- Ahmed, A., Ploennigs, J., Gao, Y. and Menzel, K. (2009), "Analyze building performance data for energy-efficient building operation", In Proceedings of CIB W 26th W78 Conference on Information Technology in Construction. International Council for Building Research Studies and Documentation, Vol. 78.
- Ahmed, V., Tezel, A., Aziz, Z. and Sibley, M. (2017). "The future of big data in facilities management: opportunities and challenges", *Facilities*, Vol. 35 No. 13-14, pp. 725-745.
- Ahiaga-Dagbui, D. D. and Smith, S. D. (2014), "Dealing with construction cost overruns using data mining", *Construction Management and Economics*, Vol. 32 No. 7-8, pp. 682-694.
- Akhavian, R. and Behzadan, A. H. (2013), "Knowledge-based simulation modeling of construction fleet operations using multimodal-process data mining", *Journal of Construction Engineering and Management*, Vol. 139 No. 11.
- Anvuur, A. M., Kumaraswamy, M. M. and Mahesh, G. (2011), "Building "relationally integrated value networks"(RIVANS)", *Engineering, Construction and Architectural Management*, Vol. 18 No. 1, pp. 102-120.
- Arashpour, M., Abbasi, B., Arashpour, M., Hosseini, M. R. and Yang, R. (2017). "Integrated management of on-site, coordination and off-site uncertainty: theorizing risk analysis within a hybrid project setting", *International Journal of Project Management*, Vol. 35 No. 4, pp. 647-655.
- Art Chaovaitwongse, W., Wang, W., Williams, T. P. and Chaovaitwongse, P. (2011), "Data mining framework to optimize the bid selection policy for competitively bid highway construction projects", *Journal of Construction Engineering and Management*, Vol. 138 No. 2, pp. 277-286.
- Babič, N. Č., Podbreznik, P. and Rebolj, D. (2010), "Integrating resource production and construction using BIM", *Automation in Construction*, Vol. 19 No. 5, pp. 539-543.
- Bakis, N., Aouad, G. and Kagioglou, M. (2007), "Towards distributed product data sharing environments—progress so far and future challenges", *Automation in Construction*, Vol. 16 No. 5, pp. 586-595.
- Behera, P., Mohanty, R. P. and Prakash, A. (2015), "Understanding construction supply chain management", *Production Planning & Control*, Vol. 26 No. 16, pp. 1332-1350.
- Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., ... and Pasha, M. (2016), "Big Data in the construction industry: A review of present status, opportunities, and future trends", *Advanced Engineering Informatics*, Vol. 30 No. 3, pp. 500-521.
- Bonev, M., Wörösch, M. and Hvam, L. (2015), "Utilizing platforms in industrialized construction: A case study of a precast manufacturer", *Construction Innovation*, Vol. 15 No. 1, pp. 84-106.
- Brown, B., Chui, M. and Manyika, J. (2011), Are you Ready for the Era of 'Big Data', McKinsey Quarterly, available at: <http://www.mckinsey.com/business->

- functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data (accessed 29 February 2016).
- Brynjolfsson, E., Hitt, L. M. and Kim, H. H. (2011), "Strength in numbers: how does data-driven decisionmaking affect firm performance", In Proceedings of ICIS 2011, available at <http://dx.doi.org/10.2139/ssrn.1819486> (accessed 29 February 2016).
- Caldas, C. H. and Soibelman, L. (2003), "Automating hierarchical document classification for construction management information systems", *Automation in Construction*, Vol. 12 No. 4, pp. 395-406.
- Camm, J., Cochran, J., Fry, M., Ohlmann, J., Anderson, D., Sweeney, D., Williams, T. (2015) "Essentials of Business Analytics", Cengage.
- Cerovsek, T. (2011), "A review and outlook for a 'Building Information Model'(BIM): A multi-standpoint framework for technological development", *Advanced Engineering Informatics*, Vol. 25 No. 2, pp. 224-244.
- Chen, H., Chiang, R. H. and Storey, V. C. (2012), "Business intelligence and analytics: from big data to big impact", *MIS Quarterly*, Vol. 36 No. 4, pp. 1165-1188.
- Chen, M., Mao, S. and Liu, Y. (2014), "Big data: A survey", *Mobile Networks and Applications*, Vol. 19 No. 2, pp. 171-209.
- Chen, Y. and Kamara, J. M. (2011), "A framework for using mobile computing for information management on construction sites", *Automation in Construction*, Vol. 20 No. 7, pp. 776-788.
- Chiu, M. L. and Lan, J. H. (2005), "Information and IN-formation: information mining for supporting collaborative design", *Automation in Construction*, Vol. 14 No. 2, pp. 197-205.
- Crowley, A. (1998), "Construction as a manufacturing process: Lessons from the automotive industry", *Computers & Structures*, Vol. 67 No. 5, pp. 389-400.
- Davenport, T. (2013) Analytics 3.0. Harvard Business Review. Issue: December 2013.
- Davenport, T., D'Amico, B. and Patil, D. J. (2013). *The Complete Guide to Business Analytics* (Collection). FT Press: Upper Saddle River, NJ.
- Davenport, T. and Harris, J. (2007) *Competing on analytics*. Boston: Harvard Business School Press, 46.
- Ding, L., Fang, W., Luo, H., Love, P. E., Zhong, B. and Ouyang, X. (2018), "A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory", *Automation in Construction*, Vol. 86, pp. 118-124.
- Fan, C., Xiao, F., and Yan, C. (2015), "A framework for knowledge discovery in massive building automation data and its application in building diagnostics", *Automation in Construction*, Vol. 50, pp. 81-90.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, Vol. 17 No. 3, pp. 37.
- Frawley, W., Piatetsky-Shapiro, G. and Matheus, C. (1992), "Knowledge discovery in databases: An overview", *AI Magazine*, Vol. 13 pp. 57-70.
- Gajzler, M. (2010), "Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry", *Technological and Economic Development of Economy*, Vol. 16 No. 2, pp. 219-232.

- 1
2
3
4
5
6
7
8
9 Gandomi, A. H., Sajedi, S., Kiani, B. and Huang, Q. (2016), "Genetic programming for
10 experimental big data mining: A case study on concrete creep formulation",
11 *Automation in Construction*, Vol. 70, pp. 89-97.
- 12 Gangotra, A. and Shankar, R. (2016). Strategies in managing risks in the adoption of
13 business analytics practices. *Journal of Enterprise Information Management*, 29(3),
14 pp.374-399.
- 15 Gantz, J. and Reinsel, D. (2011), "Extracting value from chaos", *IDC Review*, Vol. 1142,
16 pp. 1-12.
- 17 Gartner (2017) Gartner IT Glossary: Business Intelligence (BI). Available at:
18 <http://www.gartner.com/it-glossary/business-intelligence-bi/>
- 19 Gartner (2015), Magic Quadrant for Business Intelligence and Analytics Platforms,
20 available at: <https://www.gartner.com/doc/reprints?id=1-2AJGAKH&ct=150225&st=sb> (accessed 29 February 2016).
- 21 Green, S. D., Fernie, S. and Weller, S. (2005), "Making sense of supply chain
22 management: a comparative study of aerospace and construction", *Construction
23 Management and Economics*, Vol. 23 No. 6, pp. 579-593.
- 24 Hammad, A., AbouRizk, S., & Mohamed, Y. (2013), "Application of KDD techniques to
25 extract useful knowledge from labor resources data in industrial construction
26 projects", *Journal of Management in Engineering*, Vol. 30, No.6, 05014011.
- 27 Hopkins, Michael S., LaValle, Steve, Balboni, Fred, Kruschwitz, Nina, and Shockley,
28 Rebecca. (2010). 10 data points: Information and analytics at work. *MIT Sloan
29 Management Review*, 52(1), Fall, 27-31
- 30 Howard, R. and Björk, B. C. (2008), "Building information modelling—Experts' views on
31 standardisation and industry deployment", *Advanced Engineering Informatics*, Vol.
32 22 No. 2, pp. 271-280.
- 33 Huang, X. and Hinze, J. (2003), "Analysis of construction worker fall accidents", *Journal
34 of Construction Engineering and Management*, Vol. 129 No. 3, pp. 262-271.
- 35 IBM (2017) What is Big Data? Available at: <https://www.ibm.com/big-data/us/en/>
- 36 Jiao, Y., Wang, Y., Zhang, S., Li, Y., Yang, B. and Yuan, L. (2013), "A cloud approach
37 to unified lifecycle data management in architecture, engineering, construction and
38 facilities management: Integrating BIMs and SN", *Advanced Engineering
39 Informatics*, Vol. 27 No. 2, pp. 173-188.
- 40 Kamara, J. M., Augenbroe, G., Anumba, C. J. and Carrillo, P. M. (2002), "Knowledge
41 management in the architecture, engineering and construction industry",
42 *Construction Innovation*, Vol. 2 No. 1, pp. 53-67.
- 43 Kim, C., Son, H. and Kim, C. (2013), "Automated construction progress measurement
44 using a 4D building information model and 3D data", *Automation in Construction*,
45 Vol. 31, pp. 75-82.
- 46 Kim, H., Stumpf, A. and Kim, W. (2011), "Analysis of an energy efficient building
47 design through data mining approach", *Automation in Construction*, Vol. 20 No. 1,
48 pp. 37-43.
- 49 Krueger, R. A. and Casey, M. A. (2014), *Focus Groups: A Practical Guide for Applied
50 Research*, Sage Publication.
- 51 Larose, D. T. (2014), *Discovering Knowledge in Data: An Introduction to Data Mining*,
52 John Wiley & Sons, Hoboken, USA.
- 53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9 Laursen, G. and Thorlund, J. (2010). *Business Analytics for Managers: Taking Business Intelligence Beyond Reporting*. Hoboken, New Jersey: John Wiley & Sons.
- 10 Ledolter J. (2013) "Data Mining and Business Analytics with R" Wiley.
- 11 Leu, S. S., Chen, C. N. and Chang, S. L. (2001), "Data mining for tunnel support stability: neural network approach", *Automation in Construction*, Vol. 10 No. 4, pp. 429-441.
- 12
13
14 Liamputtong, P. (2011), *Focus Group Methodology: Principle and Practice*, Sage Publications.
- 15
16 Liew, M. P., & Rosenblatt J. (2003), Using Data Mining Techniques for Improving Building Life Cycle.
- 17
18 Loosemore, M. (2014), "Improving construction productivity: a subcontractor's perspective", *Engineering, Construction and Architectural Management*, Vol. 21 No.3, pp. 245-260.
- 19
20 Loudoun, R. and Townsend, K. (2017), "Implementing health promotion programs in the Australian construction industry: Levers and agents for change", *Engineering, Construction and Architectural Management*, Vol. 24 No. 2, pp. 260-274.
- 21
22 Love, P., Davis, P., Ellis, J. and On Cheung, S. (2010), "Dispute causation: identification of pathogenic influences in construction", *Engineering, Construction and Architectural Management*, Vol. 17 No. 4, pp. 404-423.
- 23
24
25 Mehta, S. (2010). *Business Intelligence for Construction Industry*, *The Master Builder*, September, pp. 196-198
- 26
27 Mohemad, R., Hamdan, A. R., Othman, Z. A. and Noor, N. M. M. (2010), "Decision support systems (dss) in construction tendering processes", *IJCSI International Journal of Computer Science Issues*, Vol. 7 No. 2, pp. 35-45.
- 28
29 Nassar, K. (2007). Application of data-mining to state transportation agencies' projects databases. *ITcon*, Vol.12, pp. 139-149.
- 30
31
32 Naganathan, H., Chong, W. O. and Chen, X. (2016), "Building energy modeling (BEM) using clustering algorithms and semi-supervised machine learning approaches", *Automation in Construction*, Vol. 72, pp. 187-194.
- 33
34
35 Ofori, G. (2008), "Leadership for future construction industry: Agenda for authentic leadership", *International Journal of Project Management*, Vol. 26 No. 6, pp. 620-630.
- 36
37
38 Peitgen, H. O., Jurgens, H. and Saupe, D. (1992), *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag, New York, USA.
- 39
40 Popov, V., Juocevicius, V., Migilinskas, D., Ustinovichius, L. and Mikalauskas, S. (2010), "The use of a virtual building design and construction model for developing an effective project concept in 5D environment", *Automation in Construction*, Vol. 19 No. 3, pp. 357-367.
- 41
42
43 Porwal, A. and Hewage, K. N. (2013), "Building Information Modeling (BIM) partnering framework for public construction projects", *Automation in Construction*, 31, pp. 204-214.
- 44
45
46 Power, D. J., Sharda, R. and Burstein, F. (2015), *Decision Support Systems*, John Wiley & Sons, Ltd.
- 47
48 Provost, F. and Fawcett, T. (2013), "Data science and its relationship to big data and data-driven decision making", *Big Data*, Vol. 1 No. 1, pp. 51-59.
- 49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9 Razavi, S. N. and Haas, C. T. (2011), "Using reference RFID tags for calibrating the
10 estimated locations of construction materials", *Automation in Construction*, Vol. 20
11 No. 6, pp. 677-685.
- 12 Redmond, A., Hore, A., Alshawi, M. and West, R. (2012), "Exploring how information
13 exchanges can be enhanced through Cloud BIM", *Automation in Construction*, Vol.
14 24, pp. 175-183.
- 15 Reffat, R. M., Gero, J. S. and Peng, W. (2006), "Improving the management of building
16 life cycle: A data mining approach", CRC Research Conference, Brisbane.
- 17 Rujiranyong, T. and Shi, J. J. (2006), "A project-oriented data warehouse for
18 construction", *Automation in Construction*, Vol. 15 No. 6, pp. 800-807.
- 19 Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R. and Lichtendahl Jr, K. C. (2017), *Data
20 Mining for Business Analytics: Concepts, Techniques, and Applications in R*, John
21 Wiley & Sons.
- 22 Silwattanusarn, T., & Tuamsuk, K. (2012). Data mining and its applications for
23 knowledge management: a literature review from 2007 to 2012, *International
24 Journal of Data Mining & Knowledge Management Process (IJDKP)* Vol.2 No. 5,
25 13-24
- 26 Soibelman, L., & Kim, H. (2002), "Data preparation process for construction knowledge
27 generation through knowledge discovery in databases", *Journal of Computing in
28 Civil Engineering*, Vol. 16 No. 1, pp. 39-48.
- 29 Tinoco, J., Correia, A. G. and Cortez, P. (2011), "Application of data mining techniques
30 in the estimation of the uniaxial compressive strength of jet grouting columns over
31 time", *Construction and Building Materials*, Vol. 25 No. 3, pp. 1257-1262.
- 32 Tixier, A. J. P., Hallowell, M. R., Rajagopalan, B. and Bowman, D. (2017),
33 "Construction safety clash detection: identifying safety incompatibilities among
34 fundamental attributes using data mining", *Automation in Construction*, Vol. 74, pp.
35 39-54.
- 36 Turban, E., Sharda, R., Aronson, J. E. and King, D. (2008), *Business Intelligence: A
37 Managerial Approach*, Prentice Hall.
- 38 Yun, S. and Caldas, C. H. (2009), "Analysing decision variables that influence
39 preliminary feasibility studies using data mining techniques", *Construction
40 Management and Economics*, Vol. 27 No. 1, pp. 73-87.
- 41 Wang, J. (2014). *Encyclopaedia of Business Analytics and Optimization*. Hershey, Pa.:
42 Business Science Reference.
- 43 Williams, T. P. and Gong, J. (2014), "Predicting construction cost overruns using text
44 mining, numerical data and ensemble classifiers", *Automation in Construction*, Vol.
45 43, pp. 23-29.
- 46 Wilmot, C. G. and Cheng, G. (2003), "Estimating future highway construction costs",
47 *Journal of Construction Engineering and Management*, Vol. 129 No. 3, pp. 272-279.
- 48 Witten, I. H., Frank, E., Hall, M. A. and Pal, C. J. (2016), *Data Mining: Practical
49 Machine Learning Tools and Techniques*, Morgan Kaufmann.
- 50 Zikopoulos, P., DeRoos, D., Parasuraman, K., Deutsch, T., Corrigan, D. and Giles, J.
51 (2013), *Harness the Power of Big Data*, McGraw-Hill.
- 52
53
54
55
56
57
58
59
60

Reviewers' Comments	Authors' response
<p>Reviewer 1</p> <p>Comments: It seems a little preciseness who the workshop participants are. Emphasizing the participants (65) divided on academia and construction sector representing number of companies and their characteristics will improve the article. It seems unclear what types of introduction given to the participants, besides they gained good understanding and insight about BI&A, data mining, and data generating process of the AEC companies. It is important to emphasize the level of understanding, especially when the topics are not well-known by AEC. Based on their general answers, it seems they have not understood the topics well enough.</p> <p>Additional Questions: 1. Originality: Does the paper contain new and significant information adequate to justify publication?: The topic is surely relevant justifying publication as the exploration and utilization in AEC is limited even though the topic itself is not new.</p> <p>2. Relationship to Literature: Does the paper demonstrate an adequate understanding of the relevant literature in the field and cite an appropriate range of literature sources? Is any significant work ignored?: The literature citations is fairly well understood and used in the article.</p> <p>3. Methodology: Is the paper's argument built on an appropriate base of theory, concepts or other ideas? Has the research or equivalent intellectual work on which the paper is based been well designed? Are the methods employed appropriate?: It seems a little unprecise who the workshop participants are. Emphasizing the participants (65) divided on academia and construction sector representing number of companies and their characteristics will improve the article. It seems unclear what types of introduction given to the participants, besides they gained good understanding and insight about BI&A, data mining, and data generating process of the AEC companies.</p> <p>4. Results: Are results presented clearly and analysed appropriately? Do the conclusions adequately tie together the other elements of the paper?: Acceptable.</p> <p>5. Implications for research, practice and/or society: Does the paper identify clearly any implications for research, practice and/or society? Does the paper bridge the gap between theory and practice? How can the research be used in practice (economic and commercial impact), in teaching, to influence public policy, in research (contributing to the body of</p>	<p>Authors' response to Reviewer 1 Thank you for your valuable comments. All changes are tracked and highlighted.</p> <p>Comments:</p> <ul style="list-style-type: none"> • A pivot table displaying the attendee profiles was added to the paper under the research methodology section (Table 2). • The level of understanding of the attendees and activities to support that level of understanding were detailed under the research methodology section. The research methodology section was expanded. Highlighted and tracked. • The aim of the paper is to provide a scoping study in the subject area. It is hard to achieve detailed operational suggestions from these kind of scoping studies. This was further highlighted in the paper. <p>Additional questions: Thank you.</p> <p>Relationship to Literature: Thank you.</p> <p>Methodology:</p> <ul style="list-style-type: none"> • A pivot table displaying the attendee profiles was added to the paper under the research methodology section (Table 2). • The level of understanding of the attendees and activities to support that level of understanding were detailed under the research methodology section. The research methodology section was expanded. Highlighted and tracked. <p>Results: Thank you.</p> <p>Implications for research: Thank you.</p> <p>Quality of Communication: Thank you.</p>

<p>knowledge)? What is the impact upon society (influencing public attitudes, affecting quality of life)? Are these implications consistent with the findings and conclusions of the paper?: Steps to be taken by the AEC industry was indeed mentioned, although too general and less operational.</p> <p>6. Quality of Communication: Does the paper clearly express its case, measured against the technical language of the fields and the expected knowledge of the journal's readership? Has attention been paid to the clarity of expression and readability, such as sentence structure, jargon use, acronyms, etc.: Acceptable</p>	
<p>Reviewer 2</p> <p>Referee: 2</p> <p>Comments: Altogether, the paper is quite relevant and does an interesting work and could be useful for theoretical knowledge. It could have a broader amplitude if the author(s) take into account the following:</p> <p>The collection/analysis of the data is weak. There isn't adequate and significant information to justify the findings.</p> <p>Literature review is not extensive. Some important, updated and related references are missing. The paper lacks review of big data application in AEC. It also lacks references e.g. Page 5 of 25, Line 26-56. Some comparative analyses should be conducted. Challenges and Drivers for Data Mining in the AEC Sector have been studied in the paper published in Facilities (Emerald) e.g. "The future of Big Data in facilities management: opportunities and challenges". The authors could check international mainstream journals for an update. More recent citations needed. Only 2 references from 2015, 1 from 2016 and 1 from 2017. No element of criticality analyzed from previous researches relevant to the same research area.</p> <p>The Method appears somewhat 'weak' as it is inter-woven around panel discussion and feedback. There are no supporting references to justify the research method. The paper could have benefited from a more suitable method or statistical analysis for the data and its validation.</p> <p>The Results are questionable and largely unsupported. Not sure about the reliability, e.g. criteria for selecting the panel members, experiences of the participants, the distribution of the feedback, etc. More details are needed. In essence, the results lack validation i.e. by the method and by existing literature on the research area.</p>	<p>Authors' response to Reviewer 2 Thank you for your valuable comments. All changes are tracked and highlighted.</p> <p>Comments:</p> <ul style="list-style-type: none"> • The literature review was significantly expanded to cover more recent publications in the area, including Big Data. However, the focus of the paper is not Big Data analytics per se. It should also be noted that the topic is still novel and volume of research in the AEC domain, particularly in terms of works taking the subject at sector/industry level in comparison to practical applications of data mining in some specific areas or projects, is limited. • Some justification as to the research method was added under the research methodology section. The level of understanding of the attendees and activities to support that level of understanding were detailed under the research methodology section. The research methodology section was expanded. Highlighted and tracked. Statistical analysis is not relevant to this research method and context as the authors sought depth through group discussions as opposed to width that would be better captured through statistically analysing questionnaires and surveys. • More details regarding the analysis and validation were given under the research methodology section.

The paper does not have any clearly defined implication. The findings are not explicitly made: only unsupported drivers, processes and challenges have been listed. Not clear as to how the research findings are able to address the key issues raised in Page 9 of 25, Lines 18-48.

Despite the merits of the paper, the way that the research is formulated and the paper is written lacks organization and rather out of focus. There is a big issue also in the presentation of the paper. The appropriate format of the Journal was not used.

The English writing is very poor. The paper needs a big purification.

The paper should be improved to illustrate a more organized scientific writing, e.g. Abstract, Introduction and Conclusion need to be improved.

Additional Questions:

1. Originality: Does the paper contain new and significant information adequate to justify publication?: Although this paper does an interesting work and could be useful for theoretical knowledge, the collection/analysis of the data is weak. There isn't adequate and significant information to justify the findings.

2. Relationship to Literature: Does the paper demonstrate an adequate understanding of the relevant literature in the field and cite an appropriate range of literature sources? Is any significant work ignored?: Literature review is not extensive. Some important, updated and related references are missing. The paper lacks review of big data application in AEC. It also lacks references e.g. Page 5 of 25, Line 26-56.

Some comparative analyses should be conducted. Challenges and Drivers for Data Mining in the AEC Sector have been studied in the paper published in Facilities (Emerald) e.g. "The future of Big Data in facilities management: opportunities and challenges". The authors could check international mainstream journals for an update.

More recent citations needed. Only 2 references from 2015, 1 from 2016 and 1 from 2017.

No element of criticality analyzed from previous researches relevant to the same research area.

3. Methodology: Is the paper's argument built on an appropriate base of theory, concepts or other ideas? Has the research or equivalent intellectual work on which the paper is based been well designed? Are the methods employed appropriate?: The Method appears somewhat 'weak' as it is inter-woven around panel discussion and feedback. There are no supporting references to justify the research method. The paper could have benefited from a more suitable method or

- The paper is a scoping paper on a scarcely researched topic. We are not aiming to find solutions to BI&A and data mining issues in the AEC sector. The aim of the paper is to highlight those issues and to provide a ground for future research in this emerging field. This aim was added to the introduction section to highlight the aim of the paper better.
- The paper was proof read and reviewed as per your comments. Various errors were corrected.

<p>1</p> <p>2</p> <p>3 statistical analysis for the data and its validation</p> <p>4</p> <p>5 4. Results: Are results presented clearly</p> <p>6 and analysed appropriately? Do the conclusions</p> <p>7 adequately tie together the other elements of the</p> <p>8 paper?: The Results are questionable and largely</p> <p>9 unsupported. Not sure about the reliability, e.g.</p> <p>10 criteria for selecting the panel members,</p> <p>11 experiences of the participants, the distribution of</p> <p>12 the feedback, etc. More details are needed. In</p> <p>13 essence, the results lack validation i.e. by the</p> <p>14 method and by existing literature on the research</p> <p>15 area.</p> <p>16</p> <p>17 5. Implications for research, practice and/or</p> <p>18 society: Does the paper identify clearly any</p> <p>19 implications for research, practice and/or</p> <p>20 society? Does the paper bridge the gap between</p> <p>21 theory and practice? How can the research be</p> <p>22 used in practice (economic and commercial</p> <p>23 impact), in teaching, to influence public policy, in</p> <p>24 research (contributing to the body of</p> <p>25 knowledge)? What is the impact upon society</p> <p>26 (influencing public attitudes, affecting quality of</p> <p>27 life)? Are these implications consistent with the</p> <p>28 findings and conclusions of the paper?: The paper</p> <p>29 does not have any clearly defined implication. The</p> <p>30 findings are not explicitly made: only unsupported</p> <p>31 drivers, processes and challenges have been</p> <p>32 listed. Not clear as to how the research findings</p> <p>33 are able to address the key issues raised in Page</p> <p>34 9 of 25, Lines 18-48.</p> <p>35</p> <p>36 6. Quality of Communication: Does the</p> <p>37 paper clearly express its case, measured against</p> <p>38 the technical language of the fields and the</p> <p>39 expected knowledge of the journal's</p> <p>40 readership? Has attention been paid to the clarity</p> <p>41 of expression and readability, such as sentence</p> <p>42 structure, jargon use, acronyms, etc.: Despite the</p> <p>43 merits of the paper, the way that the research is</p> <p>44 formulated and the paper is written lacks</p> <p>45 organization and rather out of focus. There is a big</p> <p>46 issue also in the presentation of the paper. The</p> <p>47 appropriate format of the Journal was not used.</p> <p>48</p> <p>49 The English writing is very poor. The paper needs</p> <p>50 a big purification.</p> <p>51</p> <p>52 The paper should be improved to illustrate a more</p> <p>53 organized scientific writing, e.g. Abstract,</p> <p>54 Introduction and Conclusion need to be improved.</p> <p>55</p> <p>56</p> <p>57</p> <p>58</p> <p>59</p> <p>60</p>	
--	--