An approach to big data analytics in construction industry

Miglena STOYANOVA[1]

[1] University of Economics, Varna, Bulgaria
m_stoyanova@ue-varna.bg

Abstract. Construction is often considered to be a slow-changing industry. However, construction companies are already collecting vast amounts of data that they can use in a meaningful way to keep up with growing customer demands for complex, fast projects. Now more than ever, companies need to find new ways to structure and analyze data to improve productivity and overall performance, as well as differentiate themselves in the marketplace. The study aims to present the different types of big data, specific to the field of construction, and to develop an approach for performing big data analytics within the construction organizations. It is based on the research of existing specific approaches to the analysis of different types of big data in construction. The proposed approach can be used as a reliable analytical process that can be improved and adapted to the specifics of each construction company. The study is part of Project BG05M2OP001-1.002-0002-C02 "Digitalization of Economy in a Big Data Environment".

Key words: big data, analytics, construction, approach.
Received: 04.11.2022
Revised: 07.11.2022
Accepted: 08.11.2022
Published: 30.12.2022

1. Introduction

Technology is rapidly changing the world we live and work in. Robotic systems, software and new technologies replace manual processes, human labour and outdated machines, leading to the transformation of entire industries (Sulova, 2021; Aleksandrova, 2021; Mileva, Petrov, Yankov, Vasilev, Petrova, 2021), one of which is construction. More and more companies in the sector are looking to adopt a comprehensive digital vision and strategy to increase productivity and achieve the goals set. One of the approaches that companies can use to meet changing customer requirements and solve many of their problems is data and its analysis – using computers and algorithms to extract patterns and trends from big data sets (Nacheva and Sulova, 2021). These practices assist companies in making faster and more informed strategic decisions that stimulate their better overall performance and successful positioning in an environment of increasing demand and competition.

An essential aspect of the use of big data is the good knowledge and understanding of the business processes and specific needs of construction companies that can support and improve them. This data needs to be adequately and effectively managed and analyzed to achieve a positive effect for organizations.

The purpose of the current research is to present the different types of big data, specific to the field of construction, and to develop an approach for performing big data analytics within the construction organizations.

2. Application of big data in the construction sector

The construction sector usually creates products that are used to create value in other sectors of the national economies around the world. Although construction can be described primarily as a sector focused on tangible assets, the appropriate use of big data and information is essential for the proper delivery of these tangible assets in a safe and environmentally friendly way at the right price, time and quality. Proper storage and use of data and information can provide many construction benefits.

Data and information in the construction sector are often generated in several ways. One approach to considering this process is to use a timeline (e.g., the life cycle of a typical construction product). Most often, the construction of products begins with an idea/concept and develops through the phases of development, implementation, operation and maintenance until their final removal.

Each of these identified phases provides the opportunity to generate, collect and use big data, which can be a good basis for further research using big data technologies. Unfortunately, much of the generated data/information in construction exists on paper and is sometimes discarded for economic and practical reasons, such as the need for huge storage spaces.

Typical construction projects generate many and varied data (such as drawings, documents (text), images, videos, etc.). The number of documents that can be generated in different projects (from small to very large projects) ranges from 8000 to about 100000 (Snyder et al., 2018). In addition, big data may be generated during the operation and maintenance phases of the assets. The durable nature of the products supplied by the construction sector determines the importance of these aspects in terms of data collection.

The widespread use of digital tools for the generation, storage and use of construction data can save a lot of physical space and provide significant economic, technical and legal benefits. In addition, this big data can form a good basis for improvements in the sector. There are currently a number of opportunities for construction to extract useful knowledge and ideas that facilitate decision-making on current and future projects, operational product maintenance and asset release, and to thrive through the application of relevant technologies.

These arguments explain why the construction sector should pay special attention to the recent technological advances in the field of big data. Visualization, analytical techniques and other appropriate tools used in this field have the potential to bring many benefits to construction participants.

Concerning the use of big data technologies in the construction sector, two categories of benefits can be identified. These include general business benefits and specific construction benefits. Examples of general business benefits are, for example, overall cost savings, overall operational efficiency, speed of informed decision-making and potential competitive advantage.

Specific benefits, associated with construction, include improvement of the final product, design, procurement, physical and maintenance processes in construction, development of new materials, technical skills, etc. Empirical studies on the application of big data technologies/tools can provide adequate evidence to support the identified benefits (Yan et al., 2020).

Construction is one of the main sectors responsible for the development of a country. The construction activities carried out in a project are dynamic and involve a large volume of data exchange from various stakeholders, which should be collected and processed (Wood, 2016). Data is generated during the various phases of construction projects from planning to completion. The data flow may include project and financial data, sensor and equipment data, images and videos, etc. (see Table 1). This data is often large in volume, very diverse in format and dynamic. Multilateral data reflects the many characteristics of data, derived from construction activities, thus in line with the 3V concept of big data (Bilal et al., 2016).

Table 1.

Context of big data in the construction sector

Characteristics	Participants	Examples
Volume	A large amount of data from different sources	Design data, cost data, financial data, contractual data, ERP system data, etc.
Variety	Diversity in the format of the content	DWG (drawing), DXF (drawing exchange format), DGN (project), ifcXML, ifcOWL, DOC/XLS/PPT (Microsoft format), RM/MPG (videos), JPEG (images)
Velocity	Dynamic nature of data sources	Sensors, RFID, Building Management Systems (BMS)

Source: (Bilal et al., 2016)

Table 1 shows that the development of construction processes through the widespread use of this data is expected to be the next frontier of innovation and productivity in the construction sector. This is also confirmed by Harenberg, who considers real-time data processing as a future accelerator of productivity in construction (Harenberg, 2019).

The digital revolution has a significant impact on the construction sector, as a large amount of heterogeneous data is used in construction (Bilal et al., 2016). The main prerequisites for the application of big data in the sector are the following:

- **Building Information Modeling (BIM)**

BIM covers multidimensional CAD data and aims to support the multidisciplinary and coordinated work environment among the stakeholders involved in a project. Since BIM involves collecting additional layers of information throughout the life cycle of buildings, BIM is perceived as transforming the construction sector in different perspectives. Although data volume is an essential feature of BIM, according to some authors (Humphreys, 2016), this data is not exactly big data. This follows from the fact that huge BIM files with a combination of multiple models are still only processed by BIM applications. Also, the use of embedded devices and sensors increases the amount of data generated, which ultimately leads to big BIM data sources (Bilal et al., 2016). Thus, construction enters the era of big data.

- **Cloud computing**

Cloud computing is an Internet trend that gives access to the pooling of configurable resources on demand. The main goal is to provide multiple users with access to data storage and computing without the need for an individual license. Accelerating cloud technology is contributing to the development of big data. As cloud computing supports the coordination of orders in BIM-based applications, they are widely used in the construction sector, and the performance of big data in this revolution is astounding (Bilal et al., 2016). In addition, cloud computing and big data are considered to be an ideal combination that contributes to cost-effectiveness and infrastructure expansion in support of big data and business analysis (Bello et al., 2021).

- **Internet of Things (IoT)**

The Internet of Things (IoT) is the main pillar of the Big Data 3.0 era. In essence, IoT is a system of Internet-connected devices that collect and transfer data through installed sensors (Meola, 2018). IoT applications often involve a significant number of sensor devices for data accumulation. The construction sector offers a large number of cases of using big IoT data. Among the more important areas of IoT applications are logistics, transportation, asset recording, smart homes and buildings, energy and agriculture. According to some researchers (Bilal et al., 2016), IoT and big data are interdependent trends that create a huge amount of data, accessible and analyzed in real time in construction applications. In addition, some researchers (Williams et al., 2019) suggest that during the choice of big data processing technology, the huge flow of information, produced by the IoT, triggers big data on a reciprocal basis, following the choice of big data processing technology.

- **Smart buildings**

Smart buildings adapt modern technologies with existing building systems to achieve an economic compromise between maximizing comfort and reducing energy. These systems generate a huge amount of data and most of this information often remains undisclosed and possibly discarded. These data should be interpreted to truly reflect smart buildings, which is evidence of the important role of big data analysis. Integration and development systems, based on ICT, especially the IoT, are an important catalyst for various applications, both for the sector and for the population, in the implementation of smart buildings. In this sense, some authors (Daissaoui et al., 2020) think that big data and IoT are the perfect combination to increase the energy efficiency of smart buildings.

- **Augmented reality (AR)**

Augmented reality is a technology that mixes images of virtual objects with images from the real world. It is also widely recognized as a technology that improves the human perspective. Furthermore, the means to improve the mainstream big data visualization techniques are related to AR and virtual reality (VR). Therefore, AR and big data are certainly inevitable when the complexity of big data in construction is enormous and needs to be overcome through modern visualization methods, in particular AR and VR.

- **Social networking services**

Social media is one of the fascinating trends that can help the construction industry improve communication between project teams. However, one of the main challenges is to assess its value and explore ways to analyze it. This follows from the huge amount of heterogeneous data produced by social networks. Therefore, to properly analyze data from social media, data analytics techniques need to be modified and included in the big data for big data processing (Bello-Orgaz et al., 2016). In this regard, big data can be used to develop attractive applications for the field through the high volume, speed and variety of data on social networks to improve the productivity of stakeholders.

The review of the scientific literature on the application of big data in the construction sector provides an opportunity to derive and summarize the more important concepts identified in the analyzed studies. Based on them, the main orientation of the current research related to big data in construction can be determined, as well as the possible directions for future research in the field.

Analyzed studies of big data in construction show that they are focused on the concept of "management", especially "project management", "energy management" and "resource management". In this context, big data in project management includes those construction-related data that provide a broader understanding of complex projects. Studies show that big data leads to better project management, especially in terms of cost-effectiveness, as well as reducing delays. Also, big data generated by IoT devices such as drones, sensors or smartphones helps to record progress in construction activities and monitor performance. Real-time data can be used to take appropriate actions to increase project productivity. In addition, IoT devices generate data with regard to "safety" - for example, data related to worker safety behavior and safety conditions on a construction site through sensors, automated equipment, tracking devices, and visualization technologies.

Big data also contributes to better project management by improving the decision-making process, especially when predicting project orientation, which reduces its risk.

Energy management covers the integration of IoT or BIM with big data analysis when considering the energy consumption of buildings to increase their energy efficiency and performance. Energy analysis further improve decision-making on the "design" process, which could be a determining factor in the creation of integrated building design models. In addition, big data provides an "aerial" view of all aspects of the architectural environment, which facilitates better design decision-making.

Similarly, tracking and monitoring resources through sensors or mobile applications help improve resource management decision-making and optimize resources. Other possible applications of big data found in the scientific literature include construction waste management, as well as efficiency in data exchange to improve communication.

Based on the performed analysis, the most common areas for big data research in construction can be summarized. The results of this study reveal the following five current areas: (1) project management; (2) safety; (3) energy management; (4) design decision making and (5) resource management. Table 2 summarizes the context of the most frequently explored areas related to big data in the construction sector.

Table 2.

Context of the explored areas related to big data in the construction sector

Research area	Important concepts	Details for the study
Construction project management	Monitoring	Monitoring progress/productivity through IoT devices
	Time, cost	Better time and cost management
	Decision-making	Decision-making using estimated data, leading to lower project risk
Safety	Safety of construction sites, safety behavior of workers	Big data generated through IoT devices to track and visualize the safety conditions of construction sites, as well as the safety behavior of workers
Energy management	Consumption, operational characteristics	Improving energy efficiency and performance by estimating energy consumption in buildings
Design decision-making	Decision-making	Big data on timely and informed decision-making
Resource Management	Resource Tracking	Track resources using IoT devices to improve resource usage efficiency

Source: Own elaboration

In conclusion, it can be pointed out that current big data studies in construction cover a variety of research areas. The analysis shows that researchers in the field of construction are intensively conducting research on big data related to monitoring, tracking and decision-making. This fact implies a rapid pace of development of big data and continuing interest in its application in the construction sector. Of the five research areas examined, big data on construction project management can be defined as an area in which research is intensifying. This follows from the fact that construction is a data-dependent industry, which is why it must be managed effectively with the right tools to ensure the success of a project. Big data research in construction offers good prospects for improving the sector. This is a step forward in the current digitalization efforts and provides an opportunity to obtain practical information from the vast amount of data.

3. Sources of big data in construction

Construction is one of the sectors with the largest scope and using the largest amount of heterogeneous data. Thanks to the new software technologies, digitalization of design, documentation and logic of planning is achieved. This is particularly true for BIM, which gives new meaning to data collection in design. Construction projects have a relatively long-life cycle and usually include the following phases: 1) design, 2) preparation for construction, 3) construction, 4) finishing works and 5) operation and maintenance. Each of the phases generates a large amount of data from a variety of sources, including:

- many engineering drawings in the design and construction of buildings (BIM)
- raw materials, main components, prices
- utilities and construction services, measuring instruments, building management systems
- infrastructure and transport systems
- corporate systems such as purchasing systems, performance reporting, work planning, etc.
- maintenance and replacement systems
- monitoring of operating costs
- ICT systems and equipment.

Thus, a comprehensive project involves the collection, storage and management of a large amount of heterogeneous data from its inception to its finalization, which supports all phases of business processes in construction. The data comes in many different types. Each type can have great value for the construction business. For some types the value of the data is easier to extract than for others. Different types of data require different storage solutions and therefore need to be processed in different ways.

The data generated in the construction sector can be divided into three main types - **structured, semi-structured and unstructured**. Structured data is tabular data that is very well defined. Often such data is stored in databases (Oracle, MySQL, MSSQL, DB2, etc.). Queries can easily extract data sets, which can be used in decision making. Unstructured data is the rawest form of data. It can be any type of file, such as text documents (PDF, DOCX, TXT), drawings, images (JPEG, PNG, DWG, DXF), sound files (MP3), videos (RM, MPG), etc. This data is often stored in file repositories. Extracting value from this form of data is often the most difficult, as structured characteristics from the data that describes it must first be determined. Semi-structured data has a certain format, but its structure is not very strict, as parts of the data may be incomplete or of different types. Semi-structured data is often stored as files. However, some types of semi-structured data (such as JSON or XML) can be stored in document-oriented databases that allow the use of queries.

These three types of data can often be found in a construction organization but can also be found in external data sources such as the Internet. All these forms of data from different sources can be combined into one source - a data lake.

The sources of big data in construction can be divided into two main groups - **internal** (within construction enterprises) and **external** (Internet, other organizations).

a) Internal sources of big data in construction

Any information collected from sources such as: design phase, machinery, employees, contracts, planning tools, corporate systems, specifications and manuals can be considered big data within construction companies (Fig. 1). The number of these sources may vary depending on the specific construction project.

Information from the design phase is available using technologies such as BIM, CAD and Document Management Software (DMS). In addition to big data, including construction projects and modelling, environmental data and stakeholders' investments are also used to help decide not only what to build, but also where to build.

Information from machines can be obtained in real time using IoT. In addition, high-level data flow can be achieved with robotic technologies. During the building of a construction site, big data is considered, related to weather conditions, traffic, public and economic activity, which can help determine the optimal performance of construction activities. The information, obtained from sensors embedded into various machines used in the sites, shows the time of their activity and inactivity. After its processing, appropriate conclusions can be drawn, for example, about the best combination of purchasing and renting such equipment, the most efficient use of fuel in relation to cost reduction and environmental impact. The geolocation of the equipment allows improvement of logistics, provision of spare parts if necessary and avoidance of inactivity.

In the operation and maintenance phase, big data from sensors embedded into buildings, bridges and all other structures allows their monitoring at many levels. The data obtained may be re-submitted to BIM systems in order to plan the relevant maintenance activities.

Contracts, specifications and manuals can be managed using search document management software. Planning and ERP are built on structured data systems. The data in these systems can be easily analyzed. Currently, most construction companies use ERP, CRM, workflow management systems, etc. This type of systems often uses a database to store data in a structured way. These databases contain vast amounts of data from which value can be easily extracted. For example, the workflow management system can easily give an idea of bottlenecks in business processes or sales forecasts can be made, using data from an ERP system.

An interesting point is to get to the data generated by employees. Unfortunately, access to this data is very limited. In general, employees generate data related to their experience and expertise, efficiency, issues they face, solutions they find, and coordination problems. There is a flow of data in construction, which consists only of reports and notes. For this reason, the sector loses some of its experience and know-how in each project.

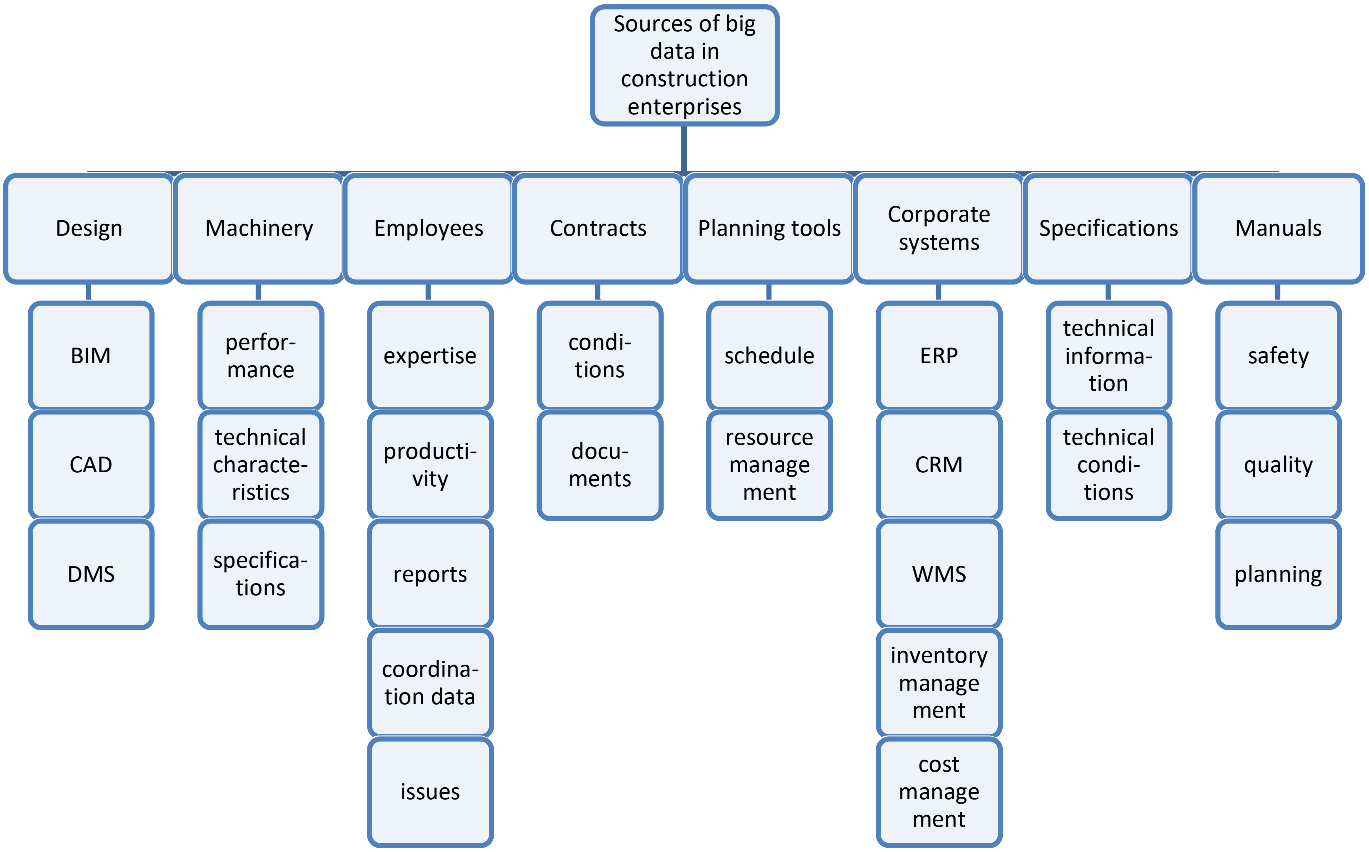

Figure 1. Sources of big data in construction enterprises
Source: Own elaboration

b) External sources of big data in construction

The data of construction companies may be enriched with external data sources, which are also of different nature. Examples of such data are publicly available datasets. Government organizations often publish demographic and economic datasets over a certain period (e.g., population/km2 for a particular region). Such data can be used to improve risk assessment.

Other organizations focus on collecting, evaluating and selling data. Their datasets contain information such as net income at a specific address, house size, etc. This data can be used to enrich the data of construction companies for various purposes – for example, improving the profile of their customers or predicting their credit risk.

Many websites today offer APIs that can be used to collect data from discussions and opinions on social media (e.g., Twitter, Facebook, LinkedIn). All tweets that contain a specific hash tag can be extracted from Twitter. Customer support software often could extract social media feeds using these APIs and to perform sentiment analysis. Sentiment analysis is a method of determining whether text regarding a topic is positive or negative. Using this method, customer support can effectively focus on unsatisfied customers. Moreover, the analysis of such data can be useful for construction companies in deciding what and where to build.

The increasing use of the Internet for access and storage of data, as well as the advancement of cloud computing, are associated with the creation of new data (del Vecchio et al., 2018). Extracting unstructured data from web content and social media and their subsequent analysis can improve various aspects of construction project work.

The ability to use large amounts of data and extract useful knowledge from them allows companies in the construction sector to meet changing customer requirements, to implement more complex projects in less time and at lower cost, and to accelerate and improve the quality of management decisions. The use of computers and algorithms to extract models and trends from big data sets helps construction companies make faster and more informed decisions that lead to better overall performance.

4. Specific types of big data in construction

The construction sector deals with huge amounts of data, resulting from different disciplines during the different phases of construction projects. The interdisciplinary nature of this industry creates big, heterogeneous and dynamic construction data. In addition, primary big data sources in construction generate data in large volumes, multiple formats and near real-time (Ismail et al., 2018).

The specific types of big data in construction can be considered in relation to the life cycle of construction projects, which includes five main stages.

1. Conceptual planning and design

The project life cycle begins with the formation of a project concept and design development. During this stage, geospatial big data can provide planning experts and designers with important information about project location, infrastructure, public spaces and resources (Wu et al., 2020). Loyola (Loyola, 2018) adds that big data can also offer insights from previous projects for future residents, their behavior and preferences, thus facilitating stakeholders' understanding of end-users needs and designing an optimal project.

Big data can be integrated with BIM and online social networks to choose sustainable energy solutions, capable of optimizing the implementation of construction projects. This can improve their design and is especially valuable for green buildings. Big data can also be used to generate simulations to evaluate different design options in terms of space and efficiency.

Stakeholders can use big data to estimate their profits. Some authors (Bilal et al., 2019) point out that big data allows for rapid research of large amounts of project data, identification of key trends and understanding of the relationship between profit margins and project attributes.

2. Pre-construction planning

The effectiveness of this stage depends on the proper use of all available knowledge, needed to develop an implementation plan. The big data collected from similar past projects can be analyzed and used to ensure the stability of the project plan by reducing uncertainty and allowing more accurate forecasts and planning. Model analysis, simulation and trend analysis are three approaches that are often used to analyze data during pre-construction planning and aim to assess the consequences of current problems and solutions, detect early warnings and threats that may affect the implementation of the project, consider the consequences of project assumptions and simulate future scenarios. Another application of big data at this stage focuses on predicting the behavior of stakeholders and analyzing the reliability of their commitments, the level of cooperation and the willingness to share knowledge.

Furthermore, historical and new data collected during this stage can be used to simulate various construction activities and tasks and thus to improve the implementation of the project. These simulations become very critical when automating activities or tasks, as the effect of automation on safety, performance and parallel tasks must be carefully analyzed.

3. Construction and commissioning

At this stage, big data is used in real time or near real time to track project progress and create built-in 3D models. Computer visual techniques are used for monitoring and analysis of activities performed on construction sites. These visual techniques have the ability to analyze static images and visual streams.

Real-time data can be collected using built-in smartphone sensors to collect equipment-related data. The state of the equipment (off, idling or occupied) and the type of work performed can be analyzed to help construction staff make better use of the equipment, make better decisions and have better control over a project. Moreover, the use of laser scanners and video cameras on construction sites allows the collection of new data that can benefit equipment operators by providing them with 3D workspace data, automatic object recognition and fast 3D surface modeling in near real time.

Big data is used to monitor the quality of construction in real time, in order to ensure timely collection and analysis of data from ongoing activities. GPS, global satellite navigation techniques, sensor technologies and network transmission technologies are used for this purpose.

Big data also applies to human resources and the workforce. A group of researchers (Guo et al., 2015) use big data to analyze the behavior of subway workers in China. Their research forms a knowledge base for behavioral risk, which is used to detect dangerous behavior and analyze the factors, influencing this behavior. Mobile applications and cameras are used for monitoring purposes. All information is stored in a cloud platform for big data and is sorted by Hadoop Distributed File System (HDFS).

4. Operation and maintenance

In the operation and maintenance phase of the construction project lifecycle, with the help of installed technologies such as RFID and sensors, facility managers can obtain information on the exact location and details of the various components of the building to simplify its monitoring, inspection and maintenance. Besides, the use of BIM models and the Internet of Things (IoT) makes it possible to generate big data on buildings that provide geometric and semantic information, as well as information on the condition of building elements. All this data can be used to present buildings in a virtual GIS environment for urban monitoring and management.

Big data, which includes information on construction and especially on energy efficiency, is becoming a major interest in a sustainable society. Big data analysis can be the solution to understanding behavior in terms of energy consumption and improving energy efficiency in the construction sector (Koseleva & Ropaite, 2017). Successful examples include energy consumption analysis, environmental measures and employment information using big data analysis techniques to study building performance.

5. Demolition

The last stage of the construction project is demolition. The big data collected from construction waste management indicators, especially in the demolition of a project, can help manage the disposal of deconstructed materials and reduce waste generation by the contractor (Lu et al., 2018). Examples of performance indicators include levels of waste generation, costs related to the collection, storage, transportation and recycling of waste, as well as revenues and savings from the sale of waste. This data can also be useful to the public by being used for construction waste management and monitoring of air pollution and construction noise (Chen & Lu, 2018; Lu, 2019).

5. Existing approaches to big data analytics in construction

With the advent of big data era, the construction sector is focused on processing large amounts of engineering data and extracting its value. However, inaccurate manual inputs and delayed data collection make it difficult to fully use the information. Meanwhile, difficult data sharing and poor interoperability of data between different business information systems lead to a lack of integration of resources in construction companies, which can facilitate decision-making (Özemre & Kabadurmus, 2020). To overcome these challenges, it is necessary to develop and adapt new approaches to big data analytics in the field of construction. To this end, research on this topic should be studied and analyzed.

A group of authors (You & Wu, 2019) offers a big data infrastructure, called Enterprise Integrated Data Platform (EIDP), to be used by construction companies. In their work, the researchers propose a framework for future business improvement that contributes to the management of the closed-loop construction supply chain, cost management and control, knowledge discovery and decision-making. The proposed solution provides a theoretical basis for the implementation of data exchange and interoperability between business management and project management. On this basis, construction companies are expected to improve their efficiency both in terms of their operations and project implementation by optimizing their business processes and improving decision-making.

Another group of scientists (Le et al., 2019) offers an innovative BIM-based framework for multi-purpose and dynamic temporary design of construction sites. It uses a hybrid approach to systematic design planning and mathematical modeling. The hybrid approach, which follows the construction site planning process, is designed to facilitate the collection and processing of qualitative and quantitative data. A BIM platform is used to determine the required quantitative data, while qualitative data is generated through knowledge-based rules. The proposed framework is expected to serve as a practical application that takes advantage of data collection and processing technologies.

A framework for integrating the construction supply chain in order to solve the problems related to heterogeneity and data exchange in the construction sector is presented by Das, Cheng and Law (Das et al., 2015). The standardized web services technology is used for specification, data transfer and integration in the proposed framework. The open standard SAWSDL is used for descriptions of web services with pointers to concepts, defined in ontologies. The NoSQL database Cassandra is used for distributed data storage between stakeholders in the construction supply chain. The results of the study show that ontologies can be used to support the transfer and integration of heterogeneous data through web services. Distributed data storage facilitates its sharing and improves its control. To demonstrate the proposed framework, the authors also present an example scenario for the material delivery process, involving three parties, namely a project manager, a contractor and a material supplier. The presented web services framework facilitates the storage and sharing of information in a distributed way through ontology-based web services. Security is improved with access control. A data model for distributed databases for data storage and extraction is also presented.

Another study combines cloud computing with big data processing techniques (Yang et al., 2020). The goal is to build a real-time energy monitoring system for a smart campus. The monitoring platform collects electricity consumption on campus buildings through intelligent meters and sensors and processes large amounts of data through big data processing techniques. A Hadoop ecosystem is built on a big data processing architecture to improve big data storage and processing capacity. In addition, the authors compare the efficiency of Hive and HBase in searching for energy data and the efficiency of relational databases and distributed databases for big data in searching for data. The proposed system has been implemented on the campus of Tunghai University. It allows administrators to monitor electricity consumption in real time and analyze historical data at any time and from anywhere.

The researched literature indicates that some construction companies have already implemented activities on big data analytics, but most of them are ad-hoc and situation specific. There is a need for the development of generalized approach which integrates big data analytics with the corporate strategy for digitalization of the construction companies.

6. An approach to performing big data analytics for the construction area

By using big data available in existing corporate systems, own databases and tools, construction companies can analyze their projects to make them better positioned, to seek and win new contracts, to control their current contracts, use their resources more efficiently and be better informed about their business.

Using as a basis the studied literature for big data analytics in construction industry, basic steps are proposed, which can be used as a starting point and resource by construction companies, seeking to build a comprehensive approach for big data analytics (Fig. 2).

The proposed approach for the analysis of big data includes the following main stages and tasks:

1. Goals and objectives
   a) Clearly defined problem or question
   b) Consent and support from managing stakeholders
2. Data evaluation
   a) Identification and mapping of data sources
   b) Defining a data set for analysis
   c) Selection of key attributes
3. Data extraction, transformation and loading
   a) Clear understanding of extraction requirements
   b) Data preparation for analysis
   c) Loading data into analytics software
4. Data analysis and validation
   a) Identification of tools and methods of analysis
   b) Creating a script or workflow for analysis
   c) Validation of results and study of deviations
5. Summary and presentation of results
   a) Selection of tools for presenting results
   b) Creating visualizations
   c) Reporting of feedback from stakeholders
6. Update/repeat analysis
   a) Development of an automated process
   b) Identification and standardization of KPIs
   c) Creation of executive dashboards

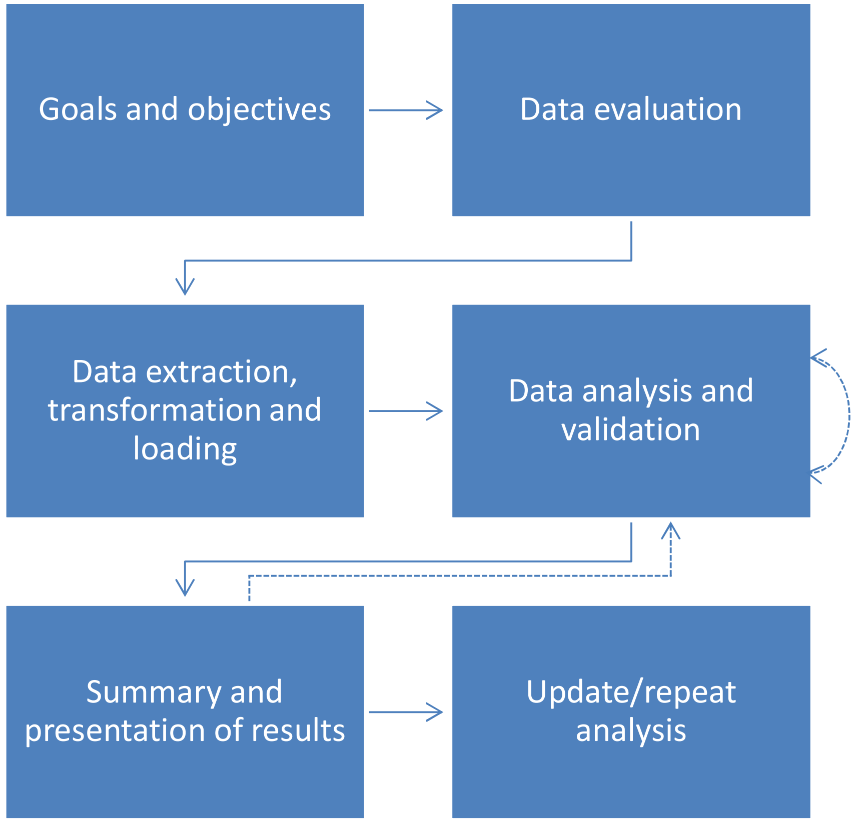

Figure 2. An approach for big data analytics in construction
Source: Own elaboration

The first step of the proposed approach is related to the clear definition of the problems or questions (goals of the construction company) that should be resolved or answered as a result of the analysis. Obtaining the consent and support of managing stakeholders and measuring the level of effort are also crucial factors for success.

As already mentioned, data in construction companies is usually extracted from many different systems. The initial **data evaluation** should include the following:

- Data sources – enterprises should create a map that identifies all data sources, shows how these different sources interact and describes in detail the format of the data. Once the available data is determined, the specific sources, necessary for the analysis, should be identified.
- Data set – in addition to identifying the sources, establishing a set of data to be used in analysis helps to further refine the purpose of the analysis and narrow the focus. Also, documentation and approval are needed on how many projects and for what period will be included in the analysis. Data that are incomplete or inaccurate should be considered when determining the set.
- Data attributes – after identifying and mapping the data sources, it is important to understand the specific data attributes that will be used in the analysis. It is appropriate to identify common fields (e.g., a unique identifier) that can be found in multiple data sources to facilitate the comparison of different data sets.

It is important to take the time to establish the quality and sources of data. If it is not possible to determine how the different data sets are interdependent, an enterprise may make decisions based on an erroneous analysis that contains incomplete or inaccurate data.

Once the sources, data set and data attributes are identified, qualified personnel are needed to extract the data from existing systems and tools to ensure its completeness. Data is extracted from one or more sources, which are mostly available in data warehouses or data lakes, maintained by the construction company. If necessary, the data obtained from multiple sources **should be transformed** (e.g., changes in the format) **and/or**

integrated to ensure consistency during the analysis. The data should then **be loaded** into an appropriate software tool for further analysis.

Conducting the data analysis and validating the results usually takes the most time and effort. Validation should be performed to ensure that the enterprise has reached complete and well-founded conclusions. Actual data analysis can be performed using a variety of tools and methods, from simple Microsoft Excel analysis, performed by average users, to more reliable analysis, performed by qualified data specialists or professionals with significant experience in big data analytics. The two key components needed at this stage include:

- Development and conduction of analysis - when developing the analysis, it is important to assess which tools and methods to use. They may vary depending on the technical capabilities of the people, performing the analysis, and the complexity of the initial goal. It is recommended that construction companies also consider the possibility of automating the analysis, which would allow easier updates in the future.
- Validation of the results – after the completion of the analysis it is necessary to make a review by experts to verify the initial results. The accuracy of the input and output data should be confirmed and any significant deviations or anomalies should be investigated. After validation of the results, it may be necessary to perform the analysis again with corrected data.

Intuitive and clear visualization allows easier **understanding, summarizing and presenting the results** of complex analyzes. For this purpose, it is appropriate to consider specialized software packages designed specifically for the visualization of large volumes of data. Moreover, visualization is an iterative process that allows data entry by the end user or experts in the field, who best understand the data. Using presentation as a brainstorming session to attract alternative ideas on how to display data can lead to different or new conclusions.

If necessary, construction companies may regularly **update or repeat the analysis**. Where possible, it is appropriate to automate the update process by linking existing data sources with analytical tools. By conducting regular analyzes, companies can develop internal key performance indicators (KPIs) or benchmarks.

In conclusion, although the proposed approach for big data analytics in construction is illustrative, it includes some of the key elements of developing a reliable analytical process, as well as considerations for companies that are beginning to analyze their existing data on projects and programs. Each construction company should develop its own approach to conducting analyzes and improve it through evaluation.

7. Conclusion

The construction sector is a data-intensive area that is experiencing rapid growth in data generation and collection. In line with advances in technology, construction has entered the digital age, where data volumes are growing at an unprecedented rate. Different types of data such as numerical, graphical, textual, multimedia and other construction information are collected from different sources. The effective use and analysis of this heterogeneous data allows the discovery of knowledge in the construction sector, which facilitates the right decisions for better implementation of construction projects. Improving the efficiency of construction projects and/or enterprises requires to a large extent the analysis and transformation of the vast wealth of data into useful knowledge.

The proposed approach for big data analytics in construction industry provides opportunities for analysis and interpretation of big data in a fast, easy to use and accurate way. It can be used to automatically detect hidden knowledge from large and complex data stored in databases, data warehouses or data lakes. The main steps included in the approach ensure the development of a reliable analytical process that can be improved and adapted to the specifics of each construction company.

Literature

Aleksandrova, Y. (2021). Predictive analytics implementation in the logistic industry, *Electronic journal "Economics and computer science"*, Issue 2, Varna

Bello, S. A., Oyedele, L. O., Akinade, O. O., Bilal, M., Davila Delgado, J. M., Akanbi, L. A., Ajayi, A. O., & Owolabi, H. A. (2021). Cloud computing in construction industry: Use cases, benefits and challenges. In *Automation in Construction* (Vol. 122). https://doi.org/10.1016/j.autcon.2020.103441

Bello-Orgaz, G., Jung, J. J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*. https://doi.org/10.1016/j.inffus.2015.08.005

Bilal, M., Oyedele, L. O., Kusimo, H. O., Owolabi, H. A., Akanbi, L. A., Ajayi, A. O., Akinade, O. O., & Davila Delgado, J. M. (2019). Investigating profitability performance of construction projects using big data: A project analytics approach. *Journal of Building Engineering*, *26*. https://doi.org/10.1016/j.jobe.2019.100850

Bilal, M., Oyedele, L. O., Qadir, J., Munir, K., Ajayi, S. O., Akinade, O. O., Owolabi, H. A., Alaka, H. A., & Pasha, M. (2016). Big Data in the construction industry: A review of present status, opportunities, and future trends. In *Advanced Engineering Informatics* (Vol. 30, Issue 3). https://doi.org/10.1016/j.aei.2016.07.001

Chen, X., & Lu, W. S. (2018). Scenarios for applying big data in boosting construction: A review. *Proceedings of the 21st International Symposium on Advancement of Construction Management and Real Estate, 2016*, *209889*. https://doi.org/10.1007/978-981-10-6190-5_114

Daissaoui, A., Boulmakoul, A., Karim, L., & Lbath, A. (2020). IoT and Big Data Analytics for Smart Buildings: A Survey. *Procedia Computer Science*, *170*. https://doi.org/10.1016/j.procs.2020.03.021

Das, M., Cheng, J. C. P., & Law, K. H. (2015). An ontology-based web service framework for construction supply chain collaboration and management. *Engineering, Construction and Architectural Management*, *22*(5). https://doi.org/10.1108/ECAM-07-2014-0089

del Vecchio, P., di Minin, A., Petruzzelli, A. M., Panniello, U., & Pirri, S. (2018). Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges. *Creativity and Innovation Management*, *27*(1). https://doi.org/10.1111/caim.12224

Guo, S., Luo, H., & Yong, L. (2015). A Big Data-based Workers Behavior Observation in China Metro Construction. *Procedia Engineering*, *123*. https://doi.org/10.1016/j.proeng.2015.10.077

Harenberg, M. (2019, October 18). *PlanRadar*. 5 Ways to Boost Construction Productivity.

Humphreys, R. (2016). *Sourceable*. How Big Data Complements Information Management. https://sourceable.net/how-big-data-complements-information-management/

Ismail, S. A., Bandi, S., & Maaz, Z. N. (2018). An Appraisal into the Potential Application of Big Data in the Construction Industry. *International Journal of Built Environment and Sustainability*, *5*(2). https://doi.org/10.11113/ijbes.v5.n2.274

Koseleva, N., & Ropaite, G. (2017). Big Data in Building Energy Efficiency: Understanding of Big Data and Main Challenges. *Procedia Engineering*, *172*. https://doi.org/10.1016/j.proeng.2017.02.064

Le, P. L., Dao, T. M., & Chaabane, A. (2019). BIM-based framework for temporary facility layout planning in construction site: A hybrid approach. *Construction Innovation*, *19*(3). https://doi.org/10.1108/CI-06-2018-0052

Loyola, M. (2018). Big data in building design: A review. In *Journal of Information Technology in Construction* (Vol. 23).

Lu, W. (2019). Big data analytics to identify illegal construction waste dumping: A Hong Kong study. *Resources, Conservation and Recycling*, *141*. https://doi.org/10.1016/j.resconrec.2018.10.039

Lu, W., Webster, C., Peng, Y., Chen, X., & Chen, K. (2018). Big data in construction waste management: Prospects and challenges. *Detritus*, *4*(December). https://doi.org/10.31025/2611-4135/2018.13737

Meola, A. (2018). What is the Internet of Things? What IoT means and how it works. *Business Insider*.

Mileva, L., Petrov, P., Yankov, P., Vasilev, J., Petrova, S. (2021). Prototype Model for Big Data Predictive Analysis in Logistics Area with Apache Kudu, *Electronic journal "Economics and computer science"*, Issue 1, Varna

Nacheva, R. and Sulova, S. (2021). Research on the Overall Attitude Towards Mobile Learning in Social Media: Emotions Mining Approach, in *Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data*. Springer, Cham, pp. 429 440. doi: 10.1007/978-3-030-65722-2_27.

Özemre, M., & Kabadurmus, O. (2020). A big data analytics based methodology for strategic decision making. *Journal of Enterprise Information Management*, *33*(6). https://doi.org/10.1108/JEIM-08-2019-0222

Snyder, J., Menard, A., & Spare, N. (2018). Big data = big questions for the engineering and construction industry. *Fmi*.

Sulova, S. (2021). Big data processing in the logistics industry, *Electronic journal "Economics and computer science"*, Issue 1, Varna

Williams, S. P., Hardy, C. A., & Nitschke, P. (2019). Configuring the internet of things (IoT): A review and implications for big data analytics. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2019-January*. https://doi.org/10.24251/hicss.2019.706

Wood, C. (2016). *Betting on Big Data: How Construction Firms are Leveraging Digitized Job Sites*. ConstructionDive.

Wu, H., Gui, Z., & Yang, Z. (2020). Geospatial big data for urban planning and urban management. In *Geo-Spatial Information Science* (Vol. 23, Issue 4). https://doi.org/10.1080/10095020.2020.1854981

Yan, H., Yang, N., Peng, Y., & Ren, Y. (2020). Data mining in the construction industry: Present status, opportunities, and future trends. In *Automation in Construction* (Vol. 119). https://doi.org/10.1016/j.autcon.2020.103331

Yang, C. T., Chen, S. T., Liu, J. C., Liu, R. H., & Chang, C. L. (2020). On construction of an energy monitoring service using big data technology for the smart campus. *Cluster Computing*, *23*(1). https://doi.org/10.1007/s10586-019-02921-5

You, Z., & Wu, C. (2019). A framework for data-driven informatization of the construction company. *Advanced Engineering Informatics*, *39*. https://doi.org/10.1016/j.aei.2019.02.002