We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Big Data as a Project Risk Management Tool

Jarosław Górecki

Additional information is available at the end of the chapter

Abstract

Risk management plays a key role in project management methodology. Nowadays, it is quite common to manage projects with the use of digitally collected data. The Big Data (BD) analysis can enhance the quality of information taken from these records and be used for project management. Moreover, it can be a project risk management tool. Some examples of BD application can be found in the investment and construction projects which use both physical and virtual data derived from controlling, bids for tender, schedules from sub-contractors or some other specific information. An accurate analysis of all that data (BD) makes it possible to discover new phenomena characteristic for the project. Getting to know them can be helpful in reducing the risk of accomplishing the project objective. The chapter is intended to arouse the readers' interest in this new source of valuable information, required to improve a course of projects. An awareness of utility of Big Data was presented on the basis of the recent anonymous survey among Polish construction entrepreneurs at the turn of years 2016 and 2017. The research has been the first step towards the broad analysis of the propensity of construction companies for using BD in project risk management. Moreover, it points out some capabilities in this matter.

Keywords: project, risk, Big Data, management, construction industry

1. Introduction

Risk can be most generally defined as a diversion of the course of events from the expected ones. The risk stands for the probability of a lack of success in terms of the actions taken [1]. It should be associated mostly with the measure of deviation from the pre-planned values. It is usually defined as the probability of those deviations. Attempts at its parametrization refer to the estimation of the probability of achieving the objectives of the planned actions and the effects of the failure to accomplish them, expressed in physical or financial units [2].

In more detail the term of risk is also presented by the Project Management Institute (PMI) which defines the project risk as an event which is difficult to predict which, if exists, affects at least one project objective (e.g. quality, cost and time) [3]. Thus sensitivity to risk of a specific area of activity can be perceived as its vulnerability to the effect of disturbing factors. It can be objective or subjective. Objectively the susceptibility to risk can be determined as the exposition to risk. It mostly refers to the susceptibility to changes in objects or systems due to various factors. Subjective vulnerability to risk is related to the perception of risk factors and the scope of their effect by the work entities.

Risk management, on the other hand, is of fundamental importance for achieving the project objectives, not only minimizing bad results but also acting as a guide to maximize the positive results. Analysing the state of knowledge on risk management, one can state that the opinion about a necessity of considering risk in project management has been well established [4]. However, management tools have been still imperfect. In risk management heuristic models are still of dominant importance. Attempts at modelling mostly refer to the evaluation of the behaviour of some parameters exposed to random effects. An insight into the nature of phenomena affecting the risk level and diagnosing the extent of risk based on empirical research are, however, still an open research field.

The risk analysis can provide the grounds for building the security of accomplishing the objectives of a given project. One has to think not only about the weak links of primary nature, which occur in its respective processes, but also about a consequent risk which is the effect of mutual relations of those processes, developing dynamically. In project risk management one must also factor in the threats which can occur in the future. They are still constantly changing challenges faced by project managers, contributing to changes in project management. Such an approach facilitates orienting the actions in the project compliant with its objectives. For example, one can note that including the risk analysis to estimate the project cost is a chance for bringing those estimates closer to the reality [5].

With the development of IT the data in analogue format lose its importance since the computer is unable to process it. In response to that problem the digital technology emerged and the amount of data saved in that format is continuously growing. It has already been so high that it exceeded the technological possibilities of its collection. Besides, according to the International Data Corporation [6], each year it increases by 40% and from 2011 to 2016 it increased nine-fold. This phenomenon is referred to as Big Data (BD).

2. Problems of gathering the data for the purpose of Big Data analysis

2.1. General context

Database is an IT term. It is defined as the collection of digitally recorded data, which are appropriately specified and describe a chosen reality. Databases are able to store all the structures of data enumerated in the previous point, starting from numbers up to sound, including immeasurable features.

The key role in making available such big collections of data so that they could fulfil the BD criteria is played by the governments. Thanks to the efforts of the former President of the United States Barack Obama, in the USA, the database Data.gov, within which open access may be granted to the data from the government sources, was created [7]. In June 2013, at the meeting of G8 Group, its countries concluded agreements with reference to making public of their database in accordance with five rules: default open data, quality and quantity, usable for everybody, releasing the data for better management and releasing the data for innovations [8]. During the same year, the government of Japan made available its collections on the website: data.go.jp. The pioneers were followed by France (data.gouv.fr), China, who published commencing of the work on making available government data, and many others countries. More details can be found in the magazine "The Bridge" [7]. The countries mentioned above make available the data from the branches of economy selected thereby. However, there are also pure construction databases, for instance, the German database, created by German BKI [9]. This organization every year publishes well-organized and exhaustive data on the costs of construction projects; the structure of the report divides the buildings into different types, allocating a specific example to each of them. Then, it presents the standard deviation for price data, taking into account the localization (prices differ depending on the region) [10].

In spite of the fact that data collections from other countries or regions could be applied, it must be noticed that not all branches of the economy are highly globalized (as, for instance, construction). Thus, it is not always possible to use "foreign" collections without cleaning and adapting the data [10].

Taking into account the earlier, construction companies more and more often decide to create their own databases, especially big enterprises, implementing many projects yearly, are in a position to generate BD.

To introduce BD for the management of the risk of projects, the manner of data gathering must change. Based on visual inspections, measurements carried out with the tape, documentation relating to the transportation of material and talks to the workers or managers are not sufficient. The collections of data obtained in this manner are too small and therefore it is not possible to recognize general dependencies and connections with other cases [11]. Gathering of data is a huge challenge as technologically advanced sensors generating the data, their transfer and storage are necessary for that purpose.

2.2. Data collection

One of the possibilities of collection of data is the procedure suggested by Son and his team [12]. It consists of surveying the key units of the construction project (for instance: architect, investor, project manager and construction site manager) during face-to-face meetings. These persons have access to detailed information relating to the progress of works both at the stage of designing and at the stage of implementation. The questionnaire should be designed in such a manner as to picture the differences between adopted assumptions at the design stage and the results during the implementation of construction projects. The survey may be oriented on aspects of costs and time, and the conclusions relating to deviations should serve the future projects at every stage of implementation [12].

The suggestion of the system to assess the offer of the construction project given by Zhang with his team [13] may be also quoted. The data necessary to perform the analysis relate to costs. The main three sources of gathering thereof were presented. The first refers to the cumulated experience of the researchers relating to the implemented projects. The second source is the agreement concluded between the construction enterprises which obliged themselves to make available their data. The companies included customers, contractors, subcontractors and so on. The third source is the agreement on the cooperation that concluded with the government, giving the authorization to analyse the government collection of data. Combing the three, abovementioned, sources, one source was achieved which generates the sufficient amount of data to fulfil the conditions of BD. To ensure the timeliness and the reliability of sources, the data relating to the costs from all the units were connected to the system. It enables the quickest transfer thereof, which, in result, leads to their number growing constantly. The content of the abovementioned data is basic information on the project that includes the number and name of the project, date, localization of the contractor and total duration time of each stage of the project. Then, characteristic information depending on the type of project, for instance, in case of the underground station construction, is presented: type of station, depth of the excavation, surrounding environment, shape of station, geotechnical and geological conditions, construction methods and so on. Next, subsequent information of each cost that includes the total cost, cost of single units, costs and amounts of material and so on is presented. All these data, after being gathered, are subject to integration and consolidation [14].

Next possible way to collect data is using the global positioning system (GPS) receiver. The data gathered in this manner may deliver necessary information which may be used to minimize the risk relating to the safety of building sites, to improve the processes of new investment planning and to choose better their localization as well as to optimum architectural solutions fulfilling the needs of the users. The methods of collecting the data which are based on GPS receivers [14] include, among others:

- Data gathered from the vehicles of public transport, taxis in particulars: Public transport has a fleet of vehicles constantly crossing the town, which are in the position to deliver the most current data relating to, for instance, traffic intensity. Geographic coordinates, speed, time of drive and its directions allow one to learn the transport structures, templates of travels and size of traffic.

- Data from the personal devices of physical persons: The individual users deliver the data concerning the use of urban spaces and the traffic roads. Additional data such as age, gender, education or individual templates of behaviour may help to optimize the architectural solutions of the buildings and public spaces.

- Data from the receivers of the workers at the building site: The workers, wearing the receivers attached to their helmets, generate the data relating to their activities and localization at the moment of performing the activities of increased risk. The data are helpful to determine the safety zones and inflict real influence on the improvement of life and health of employees.

The disadvantage of the global positioning system (GPS) is that it loses the signal inside the buildings and therefore may only be applied in case of outside works—in the open area [14].

The solution to this problem may be a Polish technical novelty—Beacon, a small transmitter based on Bluetooth technology (radio waves), which may be attached to any place on the building site. It can monitor the activities of the employees in the range up to 70 m.

The real mine of information is also the effect of not removing the data from the construction projects implemented in the past. In spite of the fact that the projects differ in many macro-conditions, for instance, requirements of the customer, localization of the implementation or the content, there are similarities in the micro-scale, such as tools, technologies, workers' skills, structure of the team and so on. Therefore, the knowledge obtained during solving the problems should be preserved in the format allowing to manipulate thereof and to disseminate it in the forthcoming projects [15].

It is not possible to enumerate all sources of data. Apart from those mentioned above, monitoring may be distinguished (for instance, at the construction site), sensors (to control the operation of machines), pictures of the employees - to register dangerous behaviour, etc. All methods of generating and gathering data involve the challenge which is their format. Text data may be stored on a PC, tabular data, pictures, recordings and many other data recorded on paper, which require different processings. Therefore, the best manner is the transformation of the unstructured data into semi-structured or structured. They must be transformed manually, sorting the documents or automatically applying algorithms based on artificial intelligence (automatic scanning of the text and recognizing texts and pictures) [13].

2.3. Computer-aided data analysis

Presently, there are enough techniques and technologies to face BD. However, the choice of the proper ones is the most important. As a branch report indicates [16, 17], the application of the unsuitable programmes leads to the fact that the implementation of BD becomes unprofitable or significantly difficult for the organizations [18].

For BD to correspond to purposes and expectations the time between the gathering of data and the results obtained from their analysis is of unusual significance. The priority is to strive towards real time. The closest to that aim is cloud computing. Cloud computing is a big data-base which contains lots of supporting technologies and algorithms. It was introduced in 2006 by Google. The development of various internal applications relating to BD is supported by a series of tools, which integrate cloud computing with a platform called Hadoop [14]. Hadoop became a complete ecosystem, which contains a module such as database (H Base, Cassandra), file system (HDFS), the processing of data (MapReduce) and others. To some extent, it can be claimed that Hadoop became a standard application having the necessary tools to face BD [13].

Introducing the cloud as an analytical solution for an enterprise, the following models of implementation should be taken into account [19]: private, public and hybrid.

Private models are deployed in a private network, managed by enterprise or by external units. Private cloud is recommended for the companies which require the highest control over the data, their safety and privacy. Using this model, data and services provided by the cloud may be used more effectively by all departments of big enterprises.

Public models may be located in the Internet and be accessible for the public. Public clouds offer high efficiency and low costs. The supplier deals with analytical services and managing the data, and the degree of safety, privacy and accessibility is written in the contract. Organizations may use this model to carry out the analysis against lowered costs and to share the observations relating to the results of the analysis of public data.

Hybrid models join two models, where the additional source of data may be added to private collection. The users may develop and implement analytical applications with the use of private surroundings and, in the same way, take advantage of the flexibility and higher level of safety than in the case of public model.

Due to latest progress, the technologies like cloud, Hadoop and MapReduce enable the collection of large amounts of co-structured and non-structured data within reasonable, close to real, time [7]. More information relating to technologies, mechanisms and the construction of the systems to store and process the data can be found in the literature, see [20].

2.4. Quality of data

Poor data quality leads to poor results of analysis. The key challenge is how to improve the quality of data in relation to reliability, completeness, consistence and resolution. Two main reasons of the lack of high quality data are lack of generally accepted templates identifying parameters which should be monitored and lack of suitable sensors which would ensure the generation of reliable data of high resolution [21]. Generally speaking, when the size of data increases, their quality remains the same. However, proportionally to the scale of the size of data increase, the amount of problems related thereto increases as well. If the size of data was multiplied by 10, 10 times more hindrances would appear in the collections of data [22]. The quality of data is defined by detailed principles depending upon the branch they are used in. These principles determine specific requirements relating to collection of data in the dimensions of precision, preciseness, coherence, metrics, timeliness and meaning, the breach of one of the abovementioned principles constitutes the insufficient quality of data [22]. It must be taken into account that depending on the context and use, the quality of data should correspond to other requirements.

According to NASA [23], the errors generated by humans are the most difficult to detect, understand and correct. They are semantic in their nature; however, they are practically undetectable for the machines. The ratio of their occurrence is from 5 to 10%, significantly more often than any other type of errors. It is a red zone. Automation of the data sources significantly increases their quality, whereas, by replacing humans, it increases the effectiveness of work as well as eliminates the errors generated manually. The advantages of automation data generating include also their precision and the ease of detection of possible errors. It is characterized by high regularity, due to which it is possible to create appropriate algorithms which would easily detect and repair irregularities. It is a green zone. In spite of that, these are the situations in which the errors in data may be ignored. Everything depends on the required level of the quality of data. If the user searches exclusively for templates or trends they need not pay special attention to errors, which will be shown in the results of analysis as outliers. Thus, it is very important to precisely know the level of the data quality on which one works.

It may turn out that the same collection of data would fulfil the quality criteria for one user and would not for the other [22].

To improve the quality of data it is necessary to implement the management of meta-data (data about the data) and master-data (collection of trusted data disseminated in the organization between different branch systems, for instance, customers files) [24]. Both in the collection of data of traditional size and in BD the proper management of meta-data and master data are crucial. The errors contained therein have disproportionally big influence on the quality of data with which they are connected. One error in the fraction of meta-data would make a perfect collection of data completely useless [22].

Also the use of too many sources does not serve to improve the quality of data. The more the sources, the less the probability of the analysis success. This happens due to technological problems. The merger of several dozens of data sources nearly always creates the difficulties relating to the transformation of formats and data structures. Besides, the unification of the gathered data often takes too much time [22]. The challenge relating to the quality of data is also the lack of their consistency. Four types of consistencies may be distinguished [25]: time, space, text and the consistency of functional dependencies breach. Time inconsistency occurs when in databases containing the time attribute, the data components overlap in time or are contrary with the circumstances. Space inconsistency may result from the geometric representation of the object (object having many localizations), space relations between the objects (breaking of space relations) or their connections. Text inconsistency occurs in non-structured texts written by a human. Inconsistency resulting from the breach of functional dependencies causes the inconsistency within the data and information.

The manner to improve the data inconsistency consists of the identification of the cause of inconsistency and introducing the heuristic procedures, which aim at solving the problem at its source [25]. For instance, if the inconsistency arises out of improper features, the algorithm is introduced and it replaces the improper features with the proper ones [26]. Also the history of carried-out analysis may deliver lots of information. In the BD era the leaving of errors detected in the collections of data while performing the activities relating to the improvement of their quality is possible and significant. All corrections and other interferences performed should also be left. This serves to create the templates which may prove to be helpful in the detection and correcting of data [27].

It may be noticed that data generating, gathering and storing are not easy tasks. These processes are very complicated and are accompanied by many challenges, on every stage of processing. One of the more important problems is the quality of data, which depends on the manner of their generation and gathering; however, this can be seen only at later stages. The solution that would significantly improve the quality of data and enable the detection of possible errors is replacing work of a human by sensors and automatic technologies of gathering. Thanks to it, the errors difficult to be detected which are generated by the semantic nature of people are avoided. However, these errors produced by the machines are easy to be detected and repaired. The data may be stored in private, public and hybrid databases. If an enterprise has at its disposal the collections of huge value, the application of private databases is recommended. It is the safest manner of storing, whereas proper protecting measures are to be preserved, that is, data

encryption and the policy of responsibility and the access authorization. The level of possibility of interference in the data gathered should be dependent on the position in the company. Based on visual inspections, measurements carried out with the tape, documentation relating to the transportation of material and talks to the workers or managers are not sufficient. Storing of any, even apparently not significant, data is recommended including, most of all, accomplished projects, recordings from the monitoring cameras, readings form GPS receivers or pictures. Since the governments have at their disposal the collections that are several teen times bigger, the collection of the whole construction industry making them public is also recommended. This would revive and increase the productivity of all segments of the economy.

3. Project management ontology

3.1. Theoretical framework

Each project is characterized by [28]:

- clearly specified objectives, the accomplishment of which closes the project,

- specific time frameworks (commencement and completion dates),

- use of resources (people, money, materials, machines, etc.),

- a number of interrelated processes (tasks) affecting the course and costs of the projects.

Thus the project, next to its innovativeness and untypical nature, has also a few other characteristic features: it is oriented on a specific objective (a specific result is expected), it has specific commencement and completion dates (it is limited in time), during its execution various resources (human, financial, etc.) are used, it is organizationally separated from other actions performed within a given organization, it has a specific organizational structure, it is a vast and complex project, it varies depending on the execution phases, it is related to investments, it is an interdisciplinary issue (it requires an involvement of specialists of many fields) and it requires an ongoing cost control. The risk of the actions taken is also the project's essential feature. However, there are no two such projects: one can always find at least one aspect which differentiates them: commercial, administrative or physical [29].

The term "project management" stands for all actions related to the preparation and execution of the decisions. However, it does not refer to the activities which directly refer to the project execution, especially with specialist aspects of solving it but to problem-solving process management [30]. Yet the term "problem" can refer to a threat of danger, an emerging chance as well as unsatisfactory situations or favourable situation in the course of project execution. The problems which need solving as part of project management are inseparably risk related.

Project maturity in project management is seen in the project team members' competencies, including the project manager. Project manager's project maturity can be defined as an ability of professional project management.

Searching for effective project management, the managers use risk management tools. It is necessary to assume the measures referring to the same risk of the project and accomplishing its objectives. One must also determine the size of any potential project losses in reference to the occurrence of independent risk areas. And independent risk is a risk which does not depend on the project manager's decision. The project risk factor analysis allows for defining and accounting for the events which can heal the project and it helps in defining possible strategies of counteracting such situations. The systematic approach to risk management stands for a global perception of the project via the role of each element of the whole, especially future effects of the decisions taken. Risk management mostly refers to a series of actions to reduce the risk effects. It thus seems that respective risk areas affect the accomplishment of project objectives to various extents, and some risk factors come from its environment. Risk management effectiveness is guaranteed by a systemic approach fully integrated with all the processes in the project. Practical risk management in the project appears mostly in a form of risk diversification.

It is also necessary to create a project risk management plan and risk capital facilitating the coverage of potential losses related to the project execution. Project maturity in project management is perceived mostly in potential possibilities of a competent project team selection, considering mostly planned risk management skills, including a search for methods on how to secure yourself from various risks, identified for the project or defining the active risk control methods. Risk management strategy in the project should foresee the necessity of creating reserves (risk capital) which can have a decisive effect on project profitability and which are important for decision-making about project execution. The key element of the strategy of securing yourself from risk is investing in the projects for which the expected rate of return is higher than the costs of the capital increased by the risk-related mark-up.

In project management there is a need for intergenerational management which requires understanding the differences in understanding values, the style of work and leadership as well as employee attitudes between the representatives of generations X and Y, as well as generation Z emerging in the market (persons born after the year 1995). The young generation articulates their willingness to participate in projects. Enterprises, willing to use their potential, formulate a number of their tasks as projects. Currently project managers are mostly representatives of generation Y (also referred to as the millennium generation), the persons born in the early 1980s. Their specific nature is related to a considerable acceleration of globalization processes accompanied by IT development and building the information society. Young people almost around the world feel like residents of the global village, from their childhood familiar with state-of-the art communication tools.

Interestingly, they usually have strong competencies of creative thinking, they adapt to changes easily, they are ready to face new challenges and ready to cooperate. Such features are especially needed for today's project managers, facing the need for implementing the system information modelling (SIM) ideas, see [31]. SIM, which refers to modelling information on the subject of the project, facilitates a permanent and immediate access to information on costs, schedules of actions and so on. One must note that the technical elements of that model have been, to much extent, operating already. Computers are powerful enough and the Internet is a huge communication tool. Specialist software does not only facilitate an

efficient performance of each element of multi-sector documentation but also supports the actions of the investor, contractors as well as the users of objects in which their life cycle is accompanied by SIM.

The competencies of project managers as the tools developed in terms of SIM make it possible to improve project management together with the process of operation and liquidation of the objects existing as part of the project.

3.2. Research results

Objects with the long cycle of operation are investment and construction project products. The opinions of managers executing projects in the construction sector are thus interesting. To acquire opinions on managing such projects, project managers in 29 Polish construction companies of the Kujawsko-Pomorskie region have been surveyed. The study was carried out in the first quarter of 2017. It showed that the risk of exceeding the essential project parameters, namely the project execution time, budget and the scope, shows varied intensity.

As for complying with the planned project execution time, exceeding the deadline was most often reported as ranging from 76 to 100%, which was the most frequent (48% respondents). Exceeding the planned budget was most often reported to fall within 0–25% (45% respondents) and 48% of the respondents pointed to the occurrence of exceeded project scope accounting for 26–50%. The studies also covered the evaluation of the level of project management complexity in the respective perspectives of essential project parameters, namely: project time, budget and scope as well as the construction workmanship quality.

As shown in **Figure 1**, the respondents see the highest level of complexity in project time and budget management.
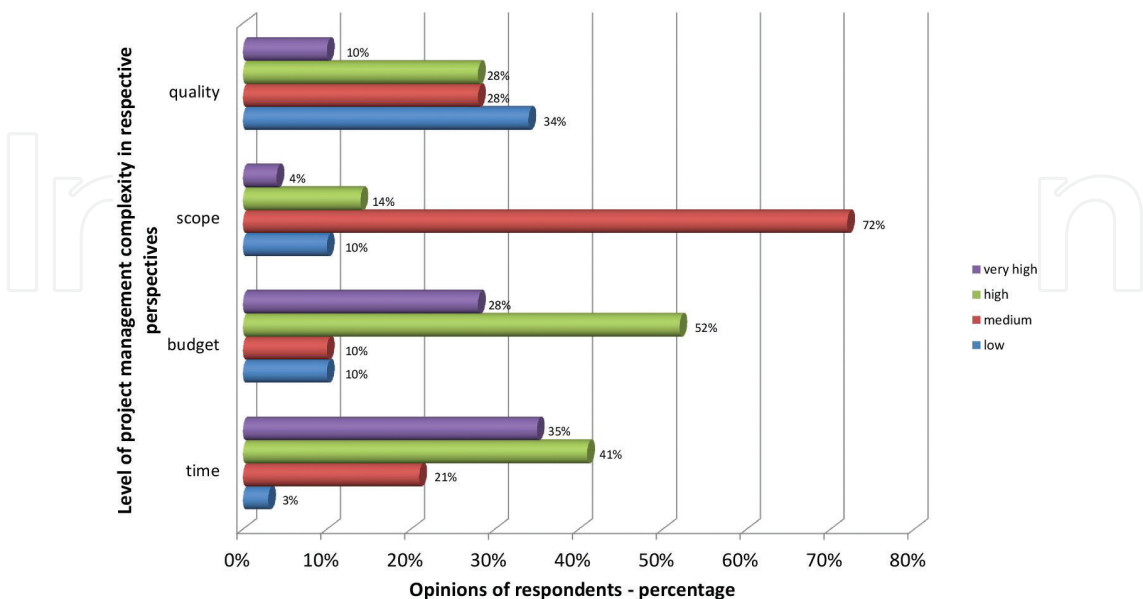
Figure 1. Opinions of respondents on the level of complexity in project management in terms of time, budget, scope and quality. Source: own study.

4. Essence and development of Big Data (BD) analysis

According to the National Science Foundation [32], Big Data (BD) is "a big varied, complex and/or dispersed set of data generated from devices, sensors, Internet transactions, e-mails, films and/or other digital sources available at present and in the future" (own translation). Gartner, in the report of 2001, indicated that BD is characterized by a 3 Vs formula (volume, velocity, variety) [33].

"*Volume*" refers to generating, collecting and accumulating big, continuously increasing, amounts of data. It strongly depends on the existing equipment infrastructure which, if not continuously improved, can lead to data analysis becoming very quickly invalid [20].

"*Velocity*" is the speed, the data are produced and processed with.

Only when it is high, a maximum use of the value coming from BD is possible [19].

"*Variety*" points to various types of data: structured (data ready for analysis, e.g., spreadsheet data), semi-structured (not organized enough to be directly analysed) and non-structured (data which, before being processed to structured, cannot be analysed, e.g., sound, video) [20].

With time the 3Vs formula got transformed into 5Vs. It was considered that BD should also correspond to such features as: veracity and value.

"*Veracity*" refers to the source of data, whether it is trustworthy and whether the data quality is sufficient.

"*Value*" is most crucial of all "Vs". It is related to the difficulty in foreseeing whether the dataset is adequately adjusted to the question asked in the analysis [18].

All the "Vs" are both the features which describe BD and a challenge faced by the analysts. A missing success in any of the attributes results in a failure of the entire analysis [18].

The research reported by The Data Warehouse Institute shows that only 12% of the companies which had introduced BD were successful, 64% with a moderate success and 24% failed [17]. However, one must note that such BD analysis implementation experiences calls for further improvement.

BD uses a set of data of extreme sizes which exceed the possibilities of manual techniques and commonly applied computer programmes for capturing, management and processing in a tolerated time range [22]. Due to its size, BD requires brand new analytical tools. It is necessary to transform the semi-structured to structured data [7]. Structured data is adequate for a direct analysis and semi-structured data must be first adequately processed. Nevertheless the processing speed is essential. Datasets show some complexity and uncertainty. Complexity is due to a large amount and variety of data, whereas uncertainty comes from ongoing changes in nature. According to [34] the challenge is to determine how to acquire essential knowledge from complex and uncertain surface of data.

One should note the relations of non-homogenous data, non-homogenous knowledge and non-homogenous decisions. Decision-making depends on the knowledge acquired, while

knowledge depends on data. If the knowledge of the project manager who draws on his experience mixes with the knowledge acquired from automatically generated non-homogenous data, then we deal with non-homogenous knowledge. Due to non-homogenous knowledge the manager takes a non-homogenous decision. Manager's knowledge represents a group of non-structured data and thus it is difficult to use it in the analysis. There is also an inconsistency of data which is shown in various kinds of human behaviours and decision-making processes. We face it when one source describes a problem in another way than the other one [25].

Despite the above problems one faces while implementing BD, one must consider the advantages of the BD analysis. It allows for discovering some "hidden knowledge" [35] or "information providing grounds for actions" [36], which helps the decision-making process. Next to the technical and methodological difficulties, there is a premise that applying BD can eliminate the partiality of smaller datasets and ensure a more accurate image of the reality studied in the course of its objective assessment [37].

One must indicate the BD applicability in the course of project management and one can even assume a statement that being successful in investment and construction projects is conditioned by the access to data in a structured form, especially in a form of BD. At the same time one shall stress an innovative nature of BD as a project-management supporting tool [38].

The BD analysis can be included in the system information model (SIM), usually built with an object-oriented approach [39].

System information modelling (SIM) is a general term applied to describe the process of modelling complex information carrier systems. System information modules are digital representations of connected systems, such as, for example, tooling, power supply, control and communication systems. The objects modelled on the SIM card are closely related with the objects in the physical system. Components, connections and functions are defined and connected just like in the real world [40].

The BD application examples can be found in the investment and construction projects which use both physical and virtual data derived from controlling, building information modelling (BIM), bids for tender, schedules, from subcontractors and information from the construction site, to mention just a few. An accurate analysis of all that data (BD) makes it possible to discover new phenomena characteristic for the project. Getting to know them can be helpful in reducing the risk of accomplishing the project objective [41]. Currently on the construction market there appear centres dealing with data capturing, storage, protection and analysis [11].

5. Project risk management

Project risk must undergo a permanent control. The essence of risk is the occurrence of potential possibilities of unexpected events resulting in a change of the situation, people, objects, systems (e.g. economic systems, ecological systems, etc.), phenomenon and so on. Therefore foreseeing their future, made a priori, diverges from the reality. At the same time

the expectations of the planning entities do not come true, which is a definite planning failure. The factors of that failure, or rather the probability of their execution, are defined as risk.

The term risk itself thus has a definitely negative connotation. With the above in mind, it is hardly possible to agree with the views claiming a dualism of the term risk itself. The effects of the events, however, can already assume the form of both a loss or a profit as compared with the preliminary assumptions. Losses occur if the foreseen situations, referred to as the threats to planned objectives of the actions taken, come true. However, profits have a chance to appear if actions are taken contrary to the speculations about the threats which, however, will not come true or if there happens to be a confluence of circumstances making their occurrence not have a significant effect on accomplishing the objective of the challenges taken up.

Risk management, as an art of taking rational decisions, runs in a standard way in the following (basic) stages [4]:

- risk identification,

- risk assessment,

- risk control,

- risk financing,

- control of the actions taken.

Identifying risk involves defining what kinds of risk and to what extent the project is threatened. An analysis of respective processes from the point of view of their threat of risk and then their classification are necessary.

A detailed analysis covers external events which threaten the project from the outside and those which can emerge as part of the project and threaten others.

Risk estimation involves defining the probability of a possibility of the occurrence of damage and the loss size. An adequately performed risk evaluation allows for taking up the projects, reducing the level of project exposure to its objective accomplishment failure.

Risk control involves taking up actions limiting the risk to the assumed admissible size. Limiting the risk often becomes contrary to the other project objectives, including mostly meeting the desired parameters of project effectiveness.

The basic risk control objective is to determine the measures of prevention eliminating or limiting the risk evaluated. Each time the selection of adequate measures is a result of a detailed analysis of the effectiveness or costs of their introduction. The costs of limiting the risk cannot exceed the value of the damage which can happen (material and non-material damage). One must note that an increase in profit is often possible by increasing the size of risk related to project execution. Most often risky projects are related to high incomes. However, above a certain limit a risk of loss can threaten profit generation.

Similarly, an enhanced profitability is contrary to maintaining liquidity. Limiting the risk is also related to bearing additional costs. Determining the size of the admissible risk is necessary.

As part of risk control, two types of actions can be taken:

- actions affecting the causes of the occurrence of risk, the objective of which is to limit risk; they are defined as an active strategy of counteracting risk,

- actions influencing the effects, the objective of which is to reduce a negative effect of unexpected losses on the level of accomplishing the project objective; they are referred to as a passive strategy of counteracting risk.

Risk financing considered as a stage of risk management comes from the fact that all the risks which are not eliminated with preventive measures must be financed. The basic forms of risk financing are:

- self-insured retention:
  - without applying preventive measures,
  - applying preventive measures,
- transferring risk to other entities (e.g., suppliers, recipients, subcontractors, insurance companies):
  - complete,
  - partial (franchising, liability limit, exclusion of the subject and the scope of insurance, etc.).

Control of the actions taken is the last stage of risk management. Its objective should be to investigate the effectiveness of actions aiming at limiting the risk. A big role in risk control and limiting is played by internal control procedures.

The rationally performed risk management process allows finding optimal solutions, a compromise between insurance and self-insurance retention. That compromise constitutes an insurance programme. Risk management is also a professional approach to insurance.

The division of risk analysis into the qualitative and quantitative stimulates two different approaches to building risk measures. The project's qualitative risk analysis facilitates creating a risk matrix. Based on the forecasts, a project threats catalogue is made. Then the threats are attributed with the levels of probability of their occurrence and the effects for the project at the assumed scale, described verbally. Such an analysis is made originally for respective projects [42].

The quantitative project risk analysis aims at determining the probability of a failure to accomplish the project's objective or its phase and the consequences of such a situation. The quantitative sizes of the risk are shown by risk value *VaR(t)*, calculated as a product of risk probability *R(t)* and value exposed to risk *Va*. One must note that each action implies risks in many areas. The total risk value is a sum of products of risk probability in a given area "i" ($Ri(t)$) and value exposed to risk in that risk area (Va_i):

$$VaR(t) \; = \; \sum_{1}^{n} R_i(t) \cdot Va_i \tag{1}$$

where: VaR(t), risk value; Ri(t), risk probability in given area "i"; Va_i, value exposed to risk in given area "i".

More and more frequently one points to the necessity of a systemic approach to risk. Such an approach to risk management in the organization requires an interdisciplinary perception of the phenomena related to it. The necessity of a dynamic and interdisciplinary approach to risk problems is also indicated by [43].

The risk phenomenon has become the focal point of taking economic decisions and in project management the project scope risk, cost risk, time risk and the quality risk are of capital importance [44]. They must be treated as a probability of a failure to accomplish the planned level of respective project attributes and they are always a planning failure.

The project risk determinants are both in the specific nature of the trade the project refers to, most often sensitive to the variation in the economic environment, and they also have an inner nature as they are sensitive to the assumed technological and organizational solutions for the course of the processes in the project.

Project decision-making is often related to the settlement of problems in terms of costs [44, 45], quality and time. An efficient project management requires simultaneously obeying those three factors, covering the actions in the directions of:

- "cheaper and cheaper" strategy as a pursuit of making right decisions in terms of costs,

- "better and better" strategy as a pursuit of making right decisions in terms of quality,

- "faster and faster" strategy as a pursuit of making right decisions in terms of time.

Such perspectives, despite an apparent contradiction, determine contemporary tendencies in search of the optimal solutions for the course of projects. They are subject to compromise in project management and they constitute the basic determinants of their success.

The term "risk management" covers all the actions towards identifying, evaluating and approaching risk, namely its reduction, diversification or using the phenomenon of risk. Currently there is much interest observed in risk management which translates to applying good practices in project management and which helps in accomplishing the project objectives. The conditions of an effective risk management are created by a systemic approach which considers all areas of the organization holistically.

The project risk management (PRM) system should be based on the knowledge and skills of the employees willing to use them to achieve the project's objective. It should include tracking down all the sources and paths of the exposure of the processes which occur in the project, the circumstances generating risk and determining their effects. Respective risk types can be

grouped into single-name risk teams (e.g. costs, quality, environment, security), and detailed problems should be served by their adequate subsystems. The system information modelling (SIM) can be a subsystem supporting project cost risk management.

6. Big Data in construction projects

6.1. Background

Currently, the construction industry has been generating and storing more and more data. The reports with reference to the progress of the works, data from different types of sensors and equipment or pictures and recordings from the building site and so on may serve as examples. However, without an appropriate tool these data would be of no value. But, when the adequate tool to process the data is used, it is possible to reveal unique conclusions. The bigger the database, the more precise information extracted thereof.

The pioneers to use BD in the construction industry are the companies from the United States: Case Inc. and Terabuild USA. They process BD with the aim to, among others, monitor in real time the costs of project and to optimize the process of planning of investment and construction enterprises. However, the potential contained in BD is still not used properly. That is why, a research question should be asked on whether BD is a helpful tool to minimize the risk arising from the implementation of construction projects.

In the world literature from the end of twentieth century, a need to improve the management of risk in the construction projects is noticed as the research proved that less than a half of such projects were accomplished without exceeding the assumed budget [46].

6.2. Research results

In the search of the influence of BD onto the minimization of the risk of construction projects, the opinions of the Polish construction entrepreneurs were quoted. It was necessary to become familiar with the level of computerization of entrepreneurs, the type of their business and whether they use the systems supporting data mining. Furthermore they were asked if BD was used to manage the risk or if the enterprises intend to use it in future.

An anonymous Internet survey was used to carry out the poll. The questionnaires were sent via email to Polish construction entrepreneurs in the break of years 2016 and 2017. In total 739 questionnaires were sent.

A total of 32 replies were received and they represented different types of enterprises, including those acting in the domestic, international and local markets, dealing with construction, design works or trade, including micro-, small, medium and big enterprises. The ratio of respondents amounted to 4.3%, which is significantly below the norm of the reply to the questionnaires in the construction industry (20–30%) [47]. The low number of replies may testify to the confidentiality of the information but mostly to the low awareness of the need to use data on the level of BD.

Among the respondents there were 40.6% of the entrepreneurs performing the construction works at the building site, 25% dealing with design works, 15.6% dealing with trading in construction material, 9.4% dealing with production of construction material and 6.3% dealing with urban designing and 3.1% dealing with investor supervision. The respondents answered the question concerning the innovative management systems and IT systems currently implemented in the enterprises. They gave their assessment in the 5-grade scale, where "1" denominated a very low level and "5" very high. A total of 21.9% of the respondents graded their enterprise with very high level—level 5. Most of the answers were given to grade 4, that is, 34.4% of all respondents. Level 3 was indicated by 25% of the respondents, and level 2 was chosen by 15.6% of respondents, and 3.1% of the assessments had fallen to level 1.

As indicated in **Figure 2** only 3.1% of the respondents have installed the IT systems serving to mine data. In the examined group of enterprises nobody sees the possibility of implementing such IT systems in the nearest future.

The respondents assessed also the degree of utility of BD in the management of the risk of the construction investment project, which is illustrated in **Figure 3**.

The results of the poll carried out indicate a very low interest in the implementation of BD analysis in the Polish construction enterprises. The reasons of such conditions may be found in the structure of enterprises, among which small and medium enterprises dominate. The strategies of these enterprises refer to short-term periods and their role in the implementation of the construction project is usually seen as that at the level of subcontractors. Thus, increasing interest in the application of BD would be more visible on the side of big enterprises or organized groups of small and medium enterprises—which unite in clusters or around the regional centres of innovativeness.
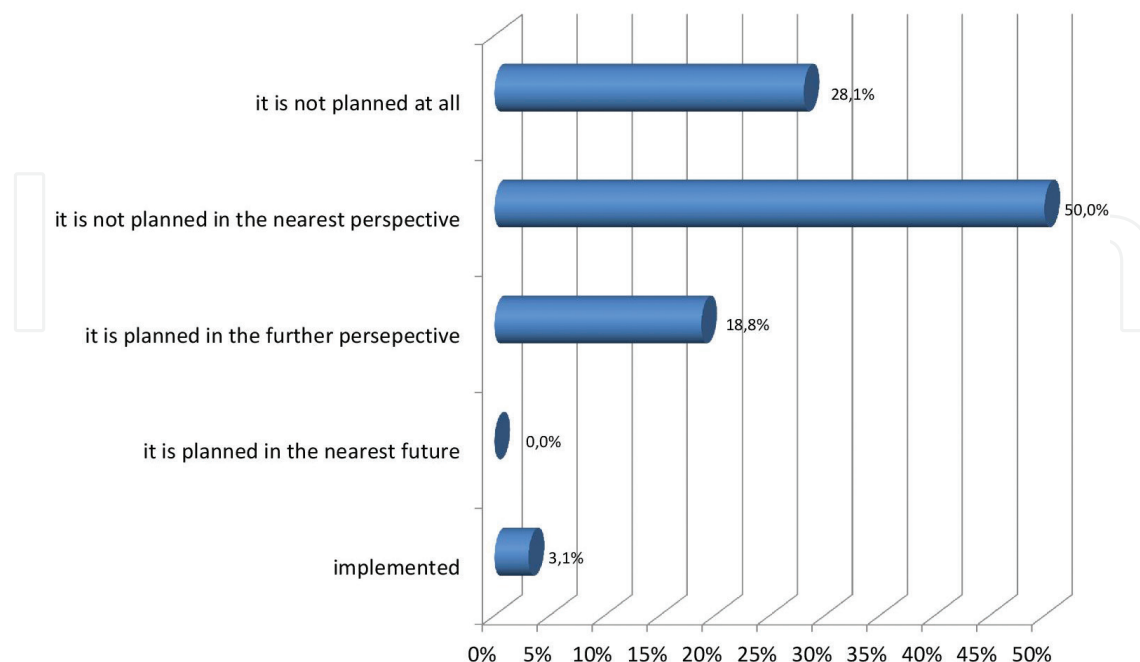
Figure 2. Implementation strategies of data mining software - according to Polish companies. Source: Górecki et al. [38].
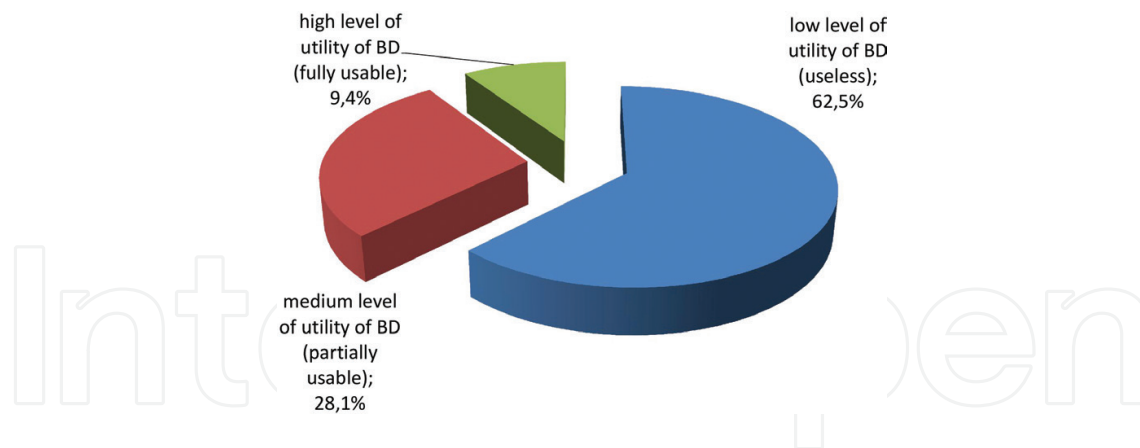
Figure 3. Level of utility of the Big Data – according to Polish respondents. Source: Górecki et al. [38].

7. BD as PRM tool: case studies

Examples of the real projects implemented in some Asian countries testify to the fact that BD has a significant influence on minimizing the risk during the implementation of construction projects. In the further part of this chapter such a relation was presented and it was described in the context of the choice of the most advantageous tender offer, the systems to ensure the safety at work at the building site and managing the construction waste during building works. The decisions which have the most influence on the success of the project (mainly in the aspect of costs) are adopted at its beginning stage. The control over the costs during the implementation of the construction projects may only be limited to preventing the exceeding of the established budget. It is not possible to correct the fundamental errors made at the beginning [10]. Thus, the effective assessment of the tender offer is the most significant in minimizing the risk relating to costs but also to quality and time.

7.1. Pre-contract stage

The pre-contract stage usually includes the following: inception, feasibility study, scheme and detail design, tendering arrangements and other pre-contract planning activities. In this part, one specific process of choosing the best offers with use of BD was taken into consideration.

The evaluation of the tender price of construction projects based on BD gathered at the building site was initially implemented in some of the projects of the underground in the Chinese town of Wuhan [14]. The system of evaluation of tender price decreases the risk resulting from poor quality, delays and exceeding budget. It consists of determination of the price frames for a given project. The enterprise taking part in the tender has the possibility of comparing their offer with the limits denominated by the system. If the offer proves to be lower than the bottom limit, the risk to exceed the budget is very high. Whereas, if the offer would be higher than the upper limit, the risk of failure in the tender rises. The system serves also for the units responsible for the selection of the tender offer. Thanks to it, they may reject the offers which,

by too low a price, expose the building works to poor quality or increase the probability of company bankruptcy. Possible insolvency would involve the delays in the implementation of construction projects. BD influences also positively the number of accidents at the building site and increases the self-control of the workers.

The effects of its operation were assessed as good but requires the improvement of the accuracy. It consists of the determination of the price frames for a given project with the use of the following information: type of station, the depth of excavation, shape, hydrological conditions, geotechnical conditions, the surrounding environment and the parameters of accomplished projects, together with the history of their costs. When the participant of the tender logs in, stating the tender data, the programme would automatically calculate the price frames and would set off the alarm should they be exceeded [14].

It is a promising system, which is helpful for the units responsible for the choice of the best tender offer, it does not need to take up decisions based on the criterion of the lowest price but also based on its justification. Furthermore, it may become also useful for the entrepreneurs taking part in the tender.

In the case of customers, it influences the minimization of the risk of the insufficient quality of performance, inadequate high costs and time. The choice is based on the marginalization of too high and too low offers. It limits the probability of bankruptcy of companies implementing construction projects; thus, there is no need to open new tenders, which would generate delays in the implementation of investment.

In turn, in case of the enterprises taking part in the tender, the system allows to minimize the risk of bankruptcy and to maximize the profits. It is worth noticing that it must be adjusted to differing construction standards in different countries or regions. For instance, because of the higher degree of seismic risk, Japan has adopted more stringent construction standards than in many other countries and, what follows, the costs of the construction projects significantly increased [46].

7.2. Operational stage

7.2.1. Occupational hazards

Significant number of the accidents at the building site is caused by inappropriate behaviour of the employees, which threatens their safety. Thus, focus would be put on human aspects. Therefore, the gathering of data from monitoring cameras (which have the ability to follow and detect the objects moving against the adopted rules) and setting off mobile application which allows the employees to register the dangerous behaviour are recommended. In consequence of the analysis of the data gathered, the safety zones, which change depending on the progress of work, are to be designated (in real time) and the employees must be made aware of specific examples of behaviour. In the same way, the level of self-control of the workers increases.

During the construction of the underground in China, the system of monitoring of the employees' behaviour was implemented for safety reasons. BD analysis was used for the implementation of the safety system at the building site of two underground lines (3 and 6) consisting of 15 stations and 8 tunnels, in the Chinese town of Wuhan. The data were generated from

the monitoring cameras, equipped with the abovementioned capacity to follow and detect the objects moving against the adopted rules, from mobile applications and from GPS receivers. It gave the possibility to take photos of dangerous behaviour by the employee. The mobile application itself generated 150,000 pictures registering the behaviour threatening safety. The analysis of data allowed one to create safety zones and to automatically capture the behaviour which may create threats and warn the employees. Additionally, they provided the building site manager with the access to information which was useful to optimize the schedule to include the safety component and gave the possibility of warning the employees in person in case of specific behaviour. All this was allowed to react in real time [48].

No matter how automated construction process or how complex the managing system is, people cannot be separated therefrom totally. The employees must constantly control the production processes and intervene in the case of unplanned events. It must be underlined that the accidents are caused in 50–90% by human error [49]. To decrease the number of incidents and to improve the efficiency of safety management at the building site it is necessary to pay particular attention to the dangerous behaviour of the employees [50]. The analysis of this behaviour is significant as it turns out that dangerous behaviour may be controlled by appropriate resources or even by the employees themselves [51]. Thus, it is worth knowing what influence the adopted measures inflicted on the employees.

Apart from the behaviour of employees the risk of preserving of their safety is also influenced by the schedule of construction works. The experts noticed certain phenomena occurring at the building site: the more tense the schedule is the more works overlap due to insignificant delays. This causes shrinking of the working space for the employees which, in turn, generates the danger. The safety of the workers at the building site decreases with their distance from construction material, equipment or other danger [14, 52, 53]. Besides, building site is a very dynamic environment where the working space relating to the performance of different tasks is subject to constant changes. The implementation of advanced automation would limit overcrowding while performing different construction activities, which often leads to threats to safety [54].

7.2.2. Waste management

Construction industry is the branch, which is not environment friendly. The waste that is generated thereby often makes up a huge portion of all communal wastes, which contribute to the degradation of the environment [55, 56]. Together with the constant increase of the pressure on sustainable development, taking up steps aiming at limiting the amount of generated waste by the construction industry is exceptionally important [11]. The comparative analysis described by Lu and his team in their papers may be of help [11, 37]. It presents the waste generation rate (WGR), which is used as the ratio of efficiency of construction waste management (CWM). BD is created by three joined databases. The first of them contains the registers of waste utilization, which include the information, among others, related to the dumping vehicle bringing the waste (for instance, number of the vehicle) or the amount of waste and name of the landfill. The second base contains, among others, the name of the project, category, location, technology and so on. The third base includes the information on waste.

The contractor may compare its achievements in CWM to its equivalents or based on its earlier achievements and denominate its practices as "good", "medium" or "not too good".

This allows one to optimize practices related to CWM. Furthermore, the government or the office responsible for the management of the waste may encourage the enterprises which were denominated as" not too good" in the analysis to increase the efficiency of CWM by, for instance, imposing fines thereto. What is more, this allows us to encourage the companies denominated as "good" to improve their achievements by rewarding them. In the analysis carried out, all outliers also deliver the information, on, for instance, the possibility of illegal disposing of waste in the case of unusually low results.

8. Conclusion

Currently the world faces the challenge of a growing inflow of digitally generated data, which is due to the technological advancement, becoming a new tool in human hands to optimize human actions and aims. The economy segments, such as banking, healthcare or insurance, already for some time have been benefiting from BD. Although the awareness of the need of project risk management is common and the tools to identify, evaluate and to manipulate risk are not a novelty, a great majority of project managers draw only on a deterministic approach to risk management.

Interestingly, the data derived from various sources, collected digitally in datasets with a high numerical amount, allows for building project risk management tools in terms of quantity. It is highly probable that already in the near future the state of the project managers' awareness will be drifting towards a greater use of digital data, including BD systems. It is also purposeful to publish the experience of the execution of projects applying BD as a recommendation for building and using BD-like data in project management. In search of effective project management methods, it is essential to inspire new actions and an ongoing search for innovative solutions. BD, next to BIM, SIM or Internet of Things, is one of those areas the development of which is worth monitoring, especially the applications supporting project risk management.

Author details

Jarosław Górecki

Address all correspondence to: gorecki@utp.edu.pl

UTP University of Science and Technology, Bydgoszcz, Poland

References

[1] Bizon-Górecka J, Górecki J. Risk of construction investment project in perspective of execution model. Studies & Proceedings of Polish Associations for Knowledge Management. 2015;**74**:4-15

[2] Bizon-Górecka J, Górecki J. Risk of construction investment project in perspective of construction costs. Organization Review. 2013;**6**:36-44

[3] Project Management Institute. A Guide to the Project Management Body of Knowledge (PMBOK® Guide). 4th ed. Newtown Square: Project Management Institute; 2008. 459 p

[4] Bizon-Górecka J. Modelling the Structure of the Risk Management System in the Enterprise – Holistic Perspective. Bydgoszcz: BGJ Consulting; 2009. 280 p

[5] Fenton RE, Cox RA, Carlock P. Incorporating contingency risk into project cost and benefit baselines: A way to enhance realism. In: Proceedings of the Ninth Annual International Symposium; International Council on Systems Engineering (INCOSE, Brighton, England). Wiley; 1999

[6] International Data Corporation. The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things [Internet]. April 2014. Available from: https://www.emc.com/collateral/analyst-reports/idc-digital-universe-2014.pdf [Accessed: 28 October 2017]

[7] Latanision RM, editor. The Bridge. Linking Engineering and Society: A Global View of Big Data. Vol. 44, No. 4. Washington: National Academy of Science; 2014. 84 p

[8] Cabinet Office. Policy paper. G8 Open Data Charter and Technical. Annex [Internet]. 18 June 2013. Available from: https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex [Accessed: 28 October 2017]

[9] Baukosteninformationszentrum Deutscher Architektenkammern. Baukosten Teil 1: Statistische Kostenkennwerte für Gebäude. In: BKI-Kostenplanung. Stuttgart: BKI; 2005

[10] Domer B, Raphael B, Saitta S. KnowPrice2: Intelligent cost estimation for construction projects. In: Smith IFC, editor. Intelligent Computing in Engineering and Architecture. Lecture Notes in Computer Science. 4200th ed. Berlin, Heidelberg: Springer; 2006. pp. 147-152. DOI: 10.1007/11888598_16

[11] Lu W, Chen X, Peng Y, Shenc L. Benchmarking construction waste management performance using big data. Resources, Conservation and Recycling. 2015;**105**(A):49-58. DOI: 10.1016/j.resconrec.2015.10.013

[12] Son H, Kim C. Early prediction of the performance of green building projects using pre-project planning variables: Data mining approaches. Journal of Cleaner Production. 2015;**109**:144-151. DOI: 10.1016/j.jclepro.2014.08.071

[13] Zhang Y, Luo H, He Y. A system for tender price evaluation of construction project based on big data. Procedia Engineering. 2015;**123**:606-614. DOI: 10.1016/j.proeng.2015.10.114

[14] Zhang S, Teizer J, Pradhananga N, Eastman CM. Workforce location tracking to model, visualize and analyze work space requirements in building information models for construction safety planning. Automation in Construction. 2015;**60**:74-86. DOI: 10.1016/j.autcon.2015.09.009

[15] Kamara JM, Anumba CJ, Carillo PM, Bouchlaghem N. Conceptual framework for live capture and reuse of project knowledge. In: Proceedings of the CIB W78's International Conference on Information Technology for Construction : Construction IT Bridging the Distance; Waiheke Island, New Zealand. Auckland: International Council for Building Research Studies and Documentation; 2003. pp. 178-185

[16] Kart L, Heudecker N, Buytendijk F. Survey Analysis: Big Data Adoption in 2013 Shows SubStance Behind the Hype [Internet]. 12 September 2013. Available from: https://www.gartner.com/doc/2589121/survey-analysis-big-data-adoption [Accessed: 28 October 2017]

[17] Russom P. Managing Big Data, TDWI Best Practices Report [Internet]. 2013. Available from: https://www.pentaho.com/sites/default/files/uploads/resources/tdwi_best_practices_report-_managing_big_data.pdf [Accessed: 28 October 2017]

[18] Tardio R, Mate A, Trujillo J. An iterative methodology for Big Data management, analysis and visualization. In: Big Data '15 Proceedings of the 2015 IEEE International Conference on Big Data (Big Data); Alicante. Washington: IEEE Computer Society; 2015. pp. 545-550

[19] Assunção MD, Calheiros RN, Bianchi S, Netto MAS, Buyya R. Big Data computing and clouds: Trends and future directions. Journal of Parallel Distributed Computing. 2015;**79-80**:3-15. DOI: 10.1016/j.jpdc.2014.08.003

[20] Chen M, Mao S, Liu Y. Big Data: A survey. Mobile Networks and Applications. 2014;**19**(2):171-209. DOI: 10.1007/s11036-013-0489-0

[21] Yu Z, Haghighat F, Fung BCM. Advances and challenges in building engineering and data mining applications for energy-efficient communities. Sustainable Cities and Society. 2016;**25**:33-38. DOI: 10.1016/j.scs.2015.12.001

[22] Becker D, King TD, McMullen B. Big Data, Big Data quality problem. In: 2015 IEEE International Conference on Big Data. Santa Clara: IEEE; 2015. pp. 2644-2653. DOI: 10.13140/RG.2.1.2947.4164

[23] Chandler F, Heard IA, Presley M, Burg A, Midden E, Mongan P. NASA Human Error Analysis. National Aeronautics and Space Administration: Washington; 2010. Final Report

[24] Masewicz M. Oracle master data management – let's bring oneness in… our data. In: XVI PLOUG Conference. Kościelisko: PLOUG; 2010

[25] Zhang D. Inconsistencies in Big Data. In: 12th IEEE Internationa Conference on Cognitive Informatics & Cognitive Computing. New York: IEEE; 2013

[26] Zhang D. i²Learning: perpetual learning through bias shifting. In: Proceedings of the 24th International Conference on Software Engineering and Knowledge Engineering. San Francisco: SEKE; 2012. pp. 249-255

[27] Jonas J. Macro trends in data & sense making, keynote speech. In: Proceedings of 14th International Conference on Information Quality. Potsdam: ICIQ; 2009

[28] Marmel E. Microsoft Project 2000. Bible. Gliwice: Helion; 2001. 592 p

[29] Lock D. The Essentials of Project Management. Warsaw: PWE (Polish Economic Publishing House); 2009. 228 p

[30] Pawlak M. Project Management. Warsaw: PWN (Polish Scientific Publishers); 2007. 276 p

[31] Love PED, Zhou J. Documentation Errors in Instrumentation and Electrical Systems: Toward Systems Information Modelling. Perth: Curtin University; 2012. 31 p

[32] National Science Foundation. Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) [Internet]. Available from: https://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm [Accessed: 28 October 2017]

[33] Laney D. 3-D data Management: Controlling Data Volume, Velocity and Variety, META Group Research Note [Internet]. 6 February 2001. Available from: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [Accessed: 28 October 2017]

[34] Ouyang ZB, Shi Y. A fuzzy clustering algorithm for petroleum data. In: Proceedings of the 2011 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Vol. 03. 22-27 August 2011. Washington: IEEE Computer Society; 2011. pp. 233-236

[35] Agrawal A. Engaging the inventor: Exploring licensing strategies for university inventions and the role of latent knowledge source. Strategic Management Journal. 2006;**27**(1):63-79. DOI: 10.1002/smj.508

[36] World Economic Forum. Big Data, Big Impact: New Possibilities for International Development [Internet]. 2012. Available from: http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf [Accessed: 28 October 2017]

[37] Lu W, Chen X, Ho DCW, Wang H. Analysis of the construction waste management performance in Hong Kong: The public and private sectors compared using big data. Journal of Cleaner Production. 2016;**112**:521-531. DOI: 10.1016/j.jclepro.2015.06.106

[38] Górecki J, Bizon-Górecka J, Michałkiewicz K. Execution of construction projects in the use of Big Data. Scientific Review. Engineering and Environmental Science. 2017;**26**(2):241-249. DOI: 10.22630/PNIKS.2017.26.2.23

[39] Love PED, Zhou J, Matthews J, Sing MCP, Edwards DJ. System information modelling in practice: Analysis of tender documentation quality in a mining mega-project. Automation in Construction. 2017;**84**:176-183. DOI: 10.1016/j.autcon.2017.08.034

[40] Zhou J, Love PED, Matthews J, Carey B, Sing CP. Object-oriented model for life cycle management of electrical instrumentation control projects. Automation in Construction. 2015;**49**(A):142-151. DOI: 10.1016/j.autcon.2014.10.008

[41] Labrinidis A, Jagadish HV. Challenges and opportunities with big data. In: Proceedings of the VLDB Endowment; Vol. 5, No. 12. Springer-Verlag; 2012. pp. 2032-2033. DOI: 10.14778/2367502.2367572

[42] Jamroż J. Risk Management in the Project. Gdańsk: Training materials of the Center for Consultancy and Training; 2006

[43] Kaczmarek TT. Risk and Risk Management. Interdisciplinary Approach. Warsaw: Difin; 2005. 257 p

[44] Kasprowicz T. Contingency of Construction Projects. Warsaw: Military University of Technology; 2004

[45] Patrascu A. Construction Cost Engineering Handbook. 1st ed. New York: CRC Press; 1988. 512 p

[46] Wang J. Estimating system of project cost based on neural network. In: 2015 International Conference on Intelligent Transportation, Big Data & Smart City; 19-20 December 2015. Halong Bay: IEEE; 2015. DOI: 10.1109/ICITBS.2015.103

[47] Akintoye A. Analysis of factors influencing project cost estimating practice. Construction Management and Economics. 2000;**18**(1):77-89. DOI: 10.1080/014461900370979

[48] Guo S, Luo H, Yong L. A big data-based workers behavior observation in China metro. Procedia Engineering. 2015;**123**:190-197. DOI: 10.1016/j.proeng.2015.10.077

[49] Zou PXW, Sunindijo RY. Skills for managing safety risk, implementing safety task, and developing positive safety climate in construction project. Automation in Construction. 2013;**34**:92-100. DOI: 10.1016/j.autcon.2012.10.018

[50] Choudhry RM. Implementation of BBS and the impact of site-level commitment. Journal of Professional Issues in Engineering Education and Practice. 2012;**138**(4):296-304. DOI: 10.1061/(ASCE)EI.1943-5541.0000111

[51] Lingard H, Rowlinson S. Occupational Health and Safety in Construction Project Management. Abingdon-on-Thames: Routledge; 2005. 464 p

[52] Hinze JW, Teizer J. Visibility-related fatalities related to construction equipment. Safety Science. 2011;**49**(5):709-718. DOI: 10.1016/j.ssci.2011.01.007

[53] Teizer J, Allread BS, Fullerton CE, Hinze J. Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system. Automation in Construction. 2010;**19**(5):630-640. DOI: 10.1016/j.autcon.2010.02.009

[54] Melzner J, Zhang S, Teizer J, Bargstädt H-J. A case study on automated safety compliance checking to assist fall protection design and planning in building information models. Construction Management and Economics. 2013;**31**(6):661-674. DOI: 10.1080/01446193.2013.780662

[55] Boiral O, Henri JF. Modelling the impact of ISO 14001 on environmental performance: A comparative approach. Journal of Environmental Management. 2012;**99**:84-97. DOI: 10.1016/j.jenvman.2012.01.007

[56] Coelho A, de Brito J. Influence of construction and demolition waste management on the environmental impact of buildings. Waste Management. 2012;**32**(3):532-541. DOI: 10.1016/j.wasman.2011.11.011