Utilization of Data Analytics in the Construction Buyout Process

Andrew W Dowdle
California Polytechnic State University
San Luis Obispo, California

This is a qualitative research project designed to analyze the feasibility of using data analytics in the buyout process. The research methodology for the paper is presented. A series of interviews with industry professionals were conducted to gather data for this paper. These interviews targeted various fields relating to the topic, including; data analytics, information systems, and construction management. Themes that emerged in these interviews include industry application, data collection, the buyout process, similar applications, and subcontractor evaluation. Insights into how data analytics in the buyout process would affect the industry are also present in this paper. These preliminary results indicate that there are a few major obstacles that need to be overcome before data analytics in construction would become feasible to use.

Keywords: data analytics, subcontractor analysis, buyout process, construction data, construction buyout

Introduction

This paper is researching and discussing the potential use of data analytics in the construction buyout process. Research was conducted specifically in regards to three major fields; construction, data analytics, and information systems. The paper explores the relationship between the three and the strengths and weaknesses of using data analytics in the buyout process. Research was conducted through thorough examination of academic papers in the three subjects and through interviews with industry professionals in their respective fields. The buyout process of construction is an extremely important part of any construction job. In each and every job there is money to be made and money to be lost during this process. Many times human error, or lack of information, causes wrong decisions to be made in the buyout process that negatively affect the project. The use of data analytics in analyzing past subcontractor performance could help mitigate these issues in the buyout process. Many other industries utilize data analytics through information technology but construction has been unable to effectively implement something similar in its field. This may be due to the unique set of challenges each project faces or many other factors. This paper will discuss the possible implementation of data analytics in the construction buyout process and discuss whether it is feasible or not.

The Buyout Process

The buyout phase of construction is an integral step in every construction project. The buyout phase is the transitional period between the two major phases of a project, preconstruction and construction (Zwick 2004). During the buyout period the general contractor is tasked with creating and verifying the scopes of work for multiple different trades. The thoroughness of these scopes of work is necessary for a complete and successful project. Each scope is analyzed to ensure that each scope does not overlap and sums up to a whole project. Accurate scopes of work will ensure an accurate budget for the project and minimize change orders throughout the life cycle of the project (Zwick 2004). If the price agreed with the subcontractor is lower than the budgeted line item then money is already being made early on in the project. Because of this, it is incentivized to get the lowest possible contract sum during the buyout process. The final decision in choosing a subcontractor is often decided by two major factors, the price of the contract and the past work of the subcontractor. Although the low contract sum is of great value, the reliability of the subcontractor is incredibly important as well. Often times when there are two subcontractors with

competitive bids, the general contractor will reach out to other general contractors that the subcontractor as worked with to make sure the subcontractor does adequate work. If the subcontractor has performed poorly on other projects via schedule and budget then they will likely not receive the subcontract (Zwick 2004). After the subcontractor is selected the prequalification process starts. The prequalification process is intended to ensure that the subcontractor is capable of performing the work. The general contractor will make sure the subcontractor is financially stable, has evidence of workmen's compensation insurance, has proper licensing and has acceptable bonding requirements (Zwick 2004). With all of these moving parts project buyout is a highly complicated and critical part of the job. If the project buyout is completed diligently and accurately then the project will have fewer change orders. This will lead to a project that comes in under budget thus generating more profit for the general contractor.

Data Analytics

Data analytics is the science of analyzing raw data in order to make conclusions about information (Huang, McIntosh, Sobolevsky, Hung, 2017). Data analytics has been utilized in many fields in today's economy such as; geography, economics, finance, physics, biology, public health and many more (Huang et al, 2017). The concept of today's data analytics relies on the use of software. There are three main steps to data analytics; data collection, data normalization and storage, and analytics and reporting. Data collection is the process where information is collected at regular time intervals from many different sources (Ehrlich, 2014). In the normalization phase data is converted into a standard format and then stored into a communal database. Finally a series of programs or software inspect the data and provides an analysis of the data set (Ehrlich, 2014). In recent years there has been a new data analytics process that is emerging called "Big Data." Big data refers to the collection of large and complex datasets. The data sets are difficult to process using traditional data processing softwares. Big data has the capability to analyze a large amount of information that is unorganized and make an evaluation about the data set (Huang et al, 2017). Many companies have begun to utilize big data analytics to uncover unseen patterns or correlations to inform them in their business decisions.

Information Systems

Business intelligence systems or information systems are where data analytics meets the industry. The Gartner Group defines business intelligence systems as "An umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance (Bogdan, 2014)." Business intelligence systems fall under a bigger umbrella of what is known as Information Systems or IS. IS assists with the professionalism of managers to ensure the quality of the decisions and to increase the performance of the organization. This may provide input for decision making, allow for the simulation of consequences and comparisons, and generate decision making solutions for the organization (Eugenia, 2018). There are four steps when using business intelligence solutions; analysis, finding causes, action and measurement of results. Analysis is filtering out the useful data information and identifying the performance indicators. Finding causes explains the causes of certain phenomenon that does not match the business model. Action is making a decision based on the analyzed information. Finally, measurement of results shows if the decisions that were taken are correct, indicating whether they need to be improved upon (Eugenia, 2018). Business intelligence systems have emerged as a necessity due to the competitive and complicated nature of current managerial activities. Business intelligence systems are capable of making informed decisions to help make up for human error and imperfect decision making.

Research Methodology

Semi-structured interviews were conducted with three different industry professionals in the construction management, data analytics, and information systems. These interviews were conducted to learn more about what each profession does and how they relate to implementing data analytics software to help with the construction buyout process. Questions were prepared beforehand to help guide the interview. These questions were created with the end goal of researching the utilization of data analytics in the buyout process of construction. Before each interview was conducted the interviewee was briefed on the topic of the research and how their field related to the topic.

Construction Management

An interview with Darren Thong, a Project Manager at Holland Partner Group, was conducted on May 22, 2019. Holland Partner Group is a vertically integrated development company that specializes in mixed use multifamily projects in urban areas. Holland is primarily focused on the West Coast but has properties in Colorado as well. During this interview we discussed the buyout process in construction, why it is important, how the subcontract is selected, how subcontract data is stored, and the challenges of implementing new software in construction. Darren Thong has 10 years of experience in construction management. During his time with Holland Construction he has delivered projects from cradle to grave with a large focus in the buyout phase of construction.

During the interview we discussed a few main ideas all relating to the buyout process. The first question asked about the buyout process and how Thong would describe it. Thong stated that they take care of the major trades such as concrete, excavation and shoring, at around 50% completed construction documents (CDs). Thong stated that there is not a lot of bidding going on with these trades because they have a good relationship with the subcontractors performing the work. In regards to the smaller trades such as framing, gypcrete, MEPs, or drywall, they go out to bid around 85% CDs. Bid invite will be given to 3 to 4 subcontractors. "In this invite we will include a draft scope rider, bid package, boilerplate contracts and links to the drawings and specs."

Next we spoke about the importance of the buyout process and the risks that are present in the buyout process. The biggest risk that Thong stated is exposure stemming from several different things. The first risk being escalation. If a scope or product is bought out to early, products could rise in price due to tariffs or other reasons and this would kill the budget. Another big risk that Thong takes on is scope gap. This causes issues between subcontractors and causes work to not be completed on time, often leading to delays in the project. The last major risk Thong spoke about is when buying out with incomplete CDs there is always a chance that there will be a large change in the project which would amount to a huge change order.

In the third part of the interview Thong spoke about the selection of a subcontract. Thong stated that the three most important factors are the scope, the cost, and the relationship. Specifically at Holland Partner Group, Thong believes that relationships are the most important factor. Thong states that this may differ for each company but this is the case at Holland. Although relationships are the most valued, for some scopes of work cost is the most important. "Like for example painting, I really don't care which painter I use as long as they've got a proven track record and they are the low guy." But of other trades Thong has a few proven guys that he talks to and then directly gives them a negotiated contract with no bidding involved. Thong states that this is because they have a proven track record of doing good work, meaning they will submit a competitive bid and he his confident they will do the job well. Thong states that most of the bidding occurs during budget sensitives scope items. Bid packages are sent out to multiple subcontractors and the end goal is to receive the best price and the best schedule because the project relies on it.

We then spoke about the historical data that is kept from project to project about the subcontracts, change orders, budget, and schedule. Thong stated that they are just now starting historical data collection on Procore. All of the information from each project is stored on a company wide database but it is just now being uploaded for use. Currently what Holland has been doing is uploading everything to Procore to show trends within regions and within project types. This helps them come up with a budget for a project in a location but is not focused on the work of past subcontractors. When asked if Procore has an analysis capability Thong stated that they cannot "spit information out of it just yet." Essentially what they do is they look at a similar project in the region and put a 3% escalation on the budget. This budget is changed throughout the project with a budget modification tool to get a more accurate number.

Finally Thong and I discussed the feasibility of a software being utilized in the buyout process and the difficulties of using new software within a company. Thong stated that its a battle implementing new software and it would be a

two or three year process. Thong states that it is possible but that it all comes top down, "if the VPs and the president tell you to use it then you are using it."

Data Analytics

An interview with Spencer Weisberg a full time data analyst with the Mariners was conducted on June 2nd, 2019. Weisberg graduated from Cal Poly with a degree in statistics in 2018. During the interview we spoke about what data analytics is and how he uses it as a tool in his job. Parallels could be drawn from the "player" being the subcontractor and the "team" being the general contractor.

We first spoke about Weisberg's job and how he uses data analytics on a day to day basis. Weisberg takes data from various sources such as ball tracking data or player tracking data and utilizes the data to analyze player performance in comparison to the rest of the league. Weisberg does this by using data that measures different variables in relation to the player.

Next we spoke about the size of data sets that Weisberg uses in his models. Weisberg stated that the size of data sets can vary greatly depending on how the information is recorded. Some data sets are recorded every second of every game leading to large amounts of raw data. These data sets can have over 1 million rows in them. On the smaller side, Weisberg stated that they can be as low as 10,000 rows.

We then discussed how accurate the data models were at predicting the future performance of the player and how the size of the data set affects the certainty of the prediction. In addition to this the parameters around the data were discussed in regards to how similar it was to other sets of data. Weisberg said that the main factor on accuracy was how each team utilizes the data. If the data did not have the same parameters around it or if the data being compared was not "apples to apples" then the conclusions drawn from the data would have to be much broader. Another factor regarding certainty is the sample size. If the sample size is smaller, then Weisberg states that there will be less certainty regarding their conclusion or "their confidence level decreases."

Finally Weisberg discussed the use of software in his decision making and how it is utilized as a tool. Weisberg discussed how they will make models analyzing the performance of a player but will use other information and factor that into the decision. "There's a lot of things when evaluating a player that are hard to quantify that affect the quality of the team. Because of this it is important that we use the data as a tool to help make decisions instead of using it as the final decision maker."

Information Systems

An interview with Griffin Nordin a graduating senior at Cal Poly was conducted on June 4th, 2019 Nordin has experience with information systems, specifically enterprise resource planning. Nordin has secured a full time job from his previous work experience and will enter as a consultant into the field. Nordin will be recommending and implementing software that tracks cost of goods sold, accounting ledgers, and overall business performance.

In the first portion of the interview we spoke about what information systems is. Information systems focuses on how and when to use different information within a business. This entails the storage, collection, analysis, and utilization of data to come to decisions. Nordin specifically will enter into a business and gather business requirements, assess the situation, and then recommend solutions to their business practices.

We then spoke about the external analysis of vendors to a business. This relationship would be greatly similar to a subcontractor to contractor. Nordin stated that in the case of vendor analysis they will go into a company and start to collect data on the vendor that they are working with. They will look at a number of factors such as budget, schedule, quality of materials, and the relationship with the vendor. Once this historical and current data is collected an analysis will be performed to determine how the vendor is performing.

Next we talked about the struggles of collecting data from the companies that Nordin enters during consulting. With some of the larger organizations Nordin stated that they often have a data collection system already in place. This allows him to work with the data much more readily and implement it more effectively. As for smaller firms they often only have excel booklets with no consistent formatting across the board. This leads to more work for a consultant because they are then required to clean and format the data to be able to effectively utilize it. Nordin stressed the importance of having a system for formatting and storing the data because it allows for a consistent analysis.

Nordin and I spoke about the issue of having data that is not apples to apples. Something that Nordin suggested that would mitigate this issue is a "statement of work." Nordin stated "A statement of work helps to layout the parameters of a project. It makes it easier to go in and look at a project that had really clear defined rolls and see that a company excelled in that position. We can then look at a project where their performance was lacking and see that the statement of work for that specific project had very undefined rolls and the client didn't know what they wanted." This is not quantifiable data but it is useful to look at when making an analysis of a company.

Finally we spoke about the software that is implemented in information systems. Nordin stated there are quite a few but the major player when it comes to software are Oracle ERP Cloud System, SAP S4Hanna, MS Project, and Tablo. We spoke about any specific vendor analysis software but Nordin stated that he is not aware of any that are currently used in the industry.

Research Results

These interviews with industry professionals ended up being very informative. There were a few major takeaways from each interview and the respective research about each subject that will be discussed. The interview with Thong was very helpful for a few reasons. The interview helped establish how and why they choose subcontractors. They choose subcontractors for two main reasons; past relationships and the competitiveness of the bid. It also helped establish how construction companies store their past data. Thong confirmed that all of this data relating to subcontracts is stored in a historical database. Currently, their historical database is being uploaded to a cloud based service known as Procore. Procore is slowly becoming an industry standard for project management. It has vast project management capabilities but has yet to implement a tool to help analyze historical project data.

The data analytics proved useful in establishing what information is necessary to come up with useful historical analysis and projections. An important piece of information that Weisberg highlighted is how large the data sets need to be to come to accurate conclusions. The data sets that are used in his field are at the smallest 10,000 rows of information. This amount of data would be difficult to achieve in the construction industry because of the amount of work that is done each year. On an individual company basis there is not enough work performed with one subcontractor to generate enough historical data. A solution to this would potential be having an anonymous shared database between general contractors with subcontractor historical data. A cloud based system such as Procore could host this database since it is quickly becoming used industry wide. Since the data sets still would likely be smaller than what is utilized in Weisberg's industry, the conclusions drawn from data would have to be used as a tool to help make decisions. It is not a deciding factor and as Weisberg stated there is a lot of non quantifiable data that is taken into account when making a decision.

The research conducted on information systems was especially useful because it is where data analytics meets the industry. Information systems is an extremely useful tool that many businesses have been utilizing to see how their company is performing. There are a few major softwares that are used as business intelligence systems. They are geared towards an internal analysis of a company's performance. Their has yet to be a program made that is geared towards an external analysis of multiple vendors or subcontractors, but the framework to make this possible is laid out in other programs. There were a few major takeaways from Nordin's interview. The first takeaway is how he stressed the importance of the normalization of data. It is much easier to analyze the data and come to consistent conclusions if there is a system or framework for entering and storing the data. The next major takeaway is the use of a "statement of work." Due to the large changes in each project the subcontractor is working on a statement of work lays out a useful framework to see why the company succeeded or failed. This can be used as a tool to help

make a decision about the subcontractor as a non quantifiable data tool. Using non quantifiable data is a useful decision making tool that both Weisberg, Nordin, and Thong all commented on.

Barriers to Entry

There are a few major barriers to entry in implementing a software like this. The first of which being the collection of data. It would be extremely hard for any company to work with one subcontractor enough to generate an adequately large data set to analyze. This is due to the high turnover rate of subcontractors within the industry and the amount of work that each subcontractor is capable of taking on at one time. A potential solution to this would be to create a large shared database between general contractors that hosts their information about the subcontractor. This would allow for a comprehensive collection of data leading to a more complete analysis.

The next major barrier to entry is how unique each construction job is. It is very hard to compare data when it does not fall under the same parameters as another job. A possible solution to this is the "statement of work" that Nording spoke about. Although this is non-quantifiable data it could lay out parameters for a different analysis of the data and it would allow the general contractor buying out the subcontract to keep it in consideration.

Finally the major barrier to entry is the use of software in the construction industry. As Thong spoke about there are always issues implementing software within a company and especially in the construction industry. Thong mentioned that it has taken them at least two years to implement Procore within the company and it still is not fully integrated within. If a subcontractor analysis software solved the two previous barriers to entry, then it would take quite some time to be fully integrated within the industry.

Conclusions and Discussion

In conclusion there is the technology to make this idea become a reality but the construction industry is very unique causing some major difficulties to form. The biggest issue is that there is not a wide enough data collection resource within the industry. This could be solved with a large cloud database between General Contractors but this would likely be ill received by the industry. If this data collection issue is solved then there are programs that could be utilized to create an analysis of subcontractors. This analysis could be greatly beneficial to the buyout process saving companies money early on in the project. Overall it is unlikely that a system will be implemented in the construction industry that has this analysis capability. This is not due to the lack of technology in creating this system but due to the barriers to entry within the industry.

References

Bogdan Nedelcu. "Business Intelligence Systems." Database Systems Journal IV.4 (2014): 12-20. Web.

Ehrlich, Paul. "Data Analytics." Engineered Systems 31.6 (2014): 20. Web.

Eugenia Iancu. "THE BUSINESS INTELLIGENCE SYSTEMS." USV Annals of Economics and Public
    Administration 18.1(27) (2018): 144-48. Web.

Huang, Shih-Chia, Suzanne McIntosh, Stanislav Sobolevsky, and Patrick Hung. "Big Data Analytics and Business
    Intelligence in Industry." Information Systems Frontiers 19.6 (2017): 1229-232. Web.

Zwick, Darin C, and Kevin R Miller. "Project Buyout." *Journal of Construction Engineering and Management*130.2
    (2004): 245-48. Web.

Appendix A - Interview Questions

Information Systems - Griffin Nordin

Tell me a little bit about what you do?
How do the companies that you have worked with store their data?
How often do you analyze the performance of vendors?
How accurate do you believe these results are?
How do you compare two different companies performances against each other?
What programs do you use?

Construction Management - Darren Thong

How would you describe the buyout process?
Why is the buyout process important?
How do you select the subcontract?
How thorough are the records between the subcontractor and gc or owner. CO, Budget, Subcontract amount, Schedule?
Do you think it would be feasible to have a computer program assists in selecting a subcontractor?

Data Analytics - Spencer Weisberg

Tell me a little about your job?
How large of data sets do you use to analyze the performance of player?
How accurate are these data sets in predicting future performance of the player? How does the size of the data sets affect your certainty of the prediction?
How apples to apples is the data in comparison to other players? If it is the same data how much would it skew your results if it wasn't.
Do you implement the use of software in your decision making. Does it make your decision for you or do you implement it as a tool?