Data-Insight-Driven Project Delivery:

Approach to Accelerated Project Delivery Using Data Analytics, Data Mining and Data Visualization

ABSTRACT

Today, 98% of megaprojects face cost overruns or delays. The average cost increase is 80% and the average slippage is 20 months behind schedule (McKinsey 2015). It is becoming increasingly challenging to efficiently support the scale, complexity and ambition of these projects. Simultaneously, project data is being captured at growing rates. We continue to capture more data on a project than ever before. Total data captured back in 2009 in the construction industry reached over 51 petabytes, or 51 million gigabytes (Mckinsey 2016). It is becoming increasingly necessary to develop new ways to leverage our project data to better manage the complexity on our projects and allow the many stakeholders to make better more informed decisions.

This paper focuses on utilizing advances in data mining, data analytics and data visualization as means to extract project information from massive datasets in a timely fashion to assist in making key informed decisions for project delivery. As part of this paper, we present an innovative new use of these technologies as applied to a large-scale infrastructural megaproject, to deliver a set of over 4,000 construction documents in a six-month period that has the potential to dramatically transform our industry and the way we deliver projects in the future. This paper describes a framework used to measure production performance as part of any project’s set of project controls for accelerated project delivery.
INTRODUCTION

Today, 9 out of 10 infrastructural projects go over budget. According to KPMG’s “2016 Global Construction Survey,” project owners found that most projects failed to come within 10 percent of budget or deadline, with over half of respondents suffering one or more underperforming projects in the previous year (Armstrong and Gilge 2016). At the same time, the amount of information being collected on today’s job sites is rising exponentially. Advances in data analytics, data mining and data visualization are enabling the extraction of project information from massive datasets in a timely fashion and offer the opportunity to leverage these datasets in new ways. Yet the tools used to understand large volumes of data have yet to be comprehensively developed or integrated into the construction process. A survey of construction companies by software vendor Sage in 2014 found that 57% of construction companies want consistent, up-to-date project information, 48% want to be warned when specific situations occur, and 41% want forecasting, allowing them to prepare for best- and worst-case scenarios (Burger 2016).

In this paper we explore ways in which data analytics, data mining and data visualization can be used to provide insights into large datasets collected on projects or job sites to monitor key metrics for accelerated project delivery. The objective of this work is to define a framework for future project delivery that has the potential to dramatically transform the way we deliver future projects, and increase productivity and overall efficiency in our industry.

Data Analytics

Data analytics or analysis of data is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, suggesting conclusions, and supporting decision-making. There are a number of types of data analytics. Descriptive Analytics describes what is happening, Diagnostic Analytics describes why things are happening, Predictive Analytics describes what is going to happen and Prescriptive Analytics begin to offer recommendations. In the construction industry, data analytics applied to large datasets,
as in other sectors, offer the opportunity to better manage the complexity on our projects and allow the many stakeholders to make better, more informed decisions.

As an example, in motorsports, real-time data analysis is vital, and race teams have been at the cutting edge of data and analytics technology for decades. As a car and driver stream data while shooting past the pit wall at speeds in excess of 200 mph, real-time analysis can be the difference between winning and losing. The volume and velocity of data streaming from over 200 onboard sensors can generate around 400 GB of data throughout a race (Pritchard 2015).

Another example where descriptive analytics can be applied to gain insights into a project is in the analysis of production data on site or in the design office. Production data can take many forms, including production of drawings (Figure 2), production of assemblies or production of installations to name a few. For each of these activities, we can collect a number of metrics that help us to gain insights that allow us to achieve our production goals.

Data Mining
Data mining is a particular data analysis technique that focuses on computational process of discovering patterns in large datasets, involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. Data mining can be used as a key business tool to transform historical data into key decision support systems to increase reliability and accuracy within the construction industry to reduce costs.

There are few examples of applied data mining, artificial intelligence and machine learning techniques in the field of building design, construction and operations. One great example of applied machine learning to evaluate buildings can be found in a
paper titled "Evaluating Buildings with Computation and Machine Learning" (Davis 2016), which uses data collected from seventy buildings, using a room-booking application to evaluate meeting rooms through machine learning to identify maintenance patterns. This study demonstrates the advantages of machine learning techniques for the evaluation of existing buildings, and the impact the results may have on designing and operating future buildings.

Another example where useful data-mining techniques can be applied to support key decisionmaking processes in production can be found in production forecasting. By using historical production data, a project team can forecast (Figure 3) future production by technology, resource or team, and make the necessary adjustments on a weekly if not daily basis.

Data Visualization

Data visualization involves the creation and study of the visual representation of data, meaning information that has been abstracted in some schematic form. Data visualization can provide key insights into a dataset in a visual way that is not always evident in table format. Data visualizations can assist in communicating decisions across a project team.

An example where data visualizations can assist in the decision-making process is in project coordination. Data visualization (Figure 4) can be used to visualize the number of issues on a project, the number of open versus closed issues, the number of issues per zone, the number of issues per discipline or just keeping track of issues resolved over time. By visualizing coordination issues, project teams can use this information to begin to focus on high-priority issues and use this information to assist in the issue-resolution process.
METHOD
As part of a large-scale infrastructural project, a framework was developed to collect, process and analyze data for drawing production, which was then used to deliver a set of approximately 4,000 construction documents in a six-month period. A set of project controls were established to include a number of metrics that could be tracked overtime to achieve our project delivery goals. The framework, as illustrated in Figure 5, consisted of data collection, data analytics, data visualization, data synthesis and automated reporting to stakeholders.

Data Collection
The first part of the framework consisted of collecting project data on a local server and pushing it up to a cloud repository—in our case Trimble Connect using Trimble Connect Sync—where data could be stored and further processed using the available API. Using the API, which allows one to read, write and update data in and out of the Connect platform, metadata from drawing files was extracted into a structured database for further sorting, filtering and analyzing through the available GET methods. Using the GET methods, our team extracted file names, author and timestamp metadata from drawing files onto a Google Spreadsheet.

Once this metadata had been processed, our team could parse out the relevant information based on a coding system in the file name, where the first character represented the discipline, followed by building, drawing package, level and drawing number, delimited by single dash characters. This allowed the team to sort and filter any relevant information by topic, such as drawing package, and correlate against other structured databases also hosted on Google Spreadsheets, which contained associated data such as source of authoring platform (e.g., Revit or CAD), thus allowing our team to extract further project insights.

Data Analysis / Data Visualization
Once collected and processed, our data could then be analyzed and visualized using a number of tools, including Google Sheets and Tableau business intelligence software through built-in statistical methods used in descriptive analytics, such as mean, median, mod and standard deviation, to begin to tell a visual story about the data collected, the project team and production status in order to make the necessary adjustments while using our project controls across the organization. By linking the data collected on Google Sheets into Tableau, we were able to develop a series of interactive dashboards to track weekly progress and published these dashboards onto Tableau Online for the project team to access. Visualizing this information allowed the project team to begin to find opportunities to optimize production and increase efficiency for that given week.

Data Synthesis
Over time, by collecting these datasets we began to synthesize production performance history and forecast production two
weeks in advance, using predictive analytic techniques (including regresional analysis) that are built in to Tableau. Given some historical data, our team could graph trend lines or plot a best-fit line to predict the following weeks’ estimated production if all other factors remained constant. By modeling production two weeks in advance, the project team was able to anticipate workload and increase resources where necessary.

Automated Reporting
Furthermore, at each stage of the framework, a series of automated scripts and procedures was developed to push, process and pull the large amounts of data throughout our data infrastructure, allowing our team to eventually automate our reporting and provide our project stakeholders with the most relevant and up-to-date information in a timely fashion. Because all of the project data was all hosted in the cloud, it was possible to refresh our production reports on a daily basis.

RESULTS
Our project team implemented the framework described above and established a set of metrics, including Number of Total Drawings, Drawings Produced, Missing Drawings, Drawings per Package, Drawings per Technology (CAD/BIM), Drawings drawn by Resource and Number of Revisions to track production performance.

Sample Metadata
As an example, Figure 6 below shows the type of metadata collected, which included: Drawing Name, Date, Type, and Authors. From this data and associated data stored on other databases, we could derive the metrics listed above using the drawing name as the unique identifier.

The included sample metadata represents only a small sample of the actual 200 GB of data that was collected, processed and analyzed on a daily basis. This data was collected, analyzed and fed into a series of interactive dashboards as described in the section above, which allowed the project team to gain key insights into the production process and make the necessary adjustments to achieve project goals.

Interactive Dashboards
Once collected, processed and analyzed, the collected metadata was used to produce our interactive dashboards. Figure 7 is an example of one of the many dashboards generated to measure drawing progress, including drawings produced to date, drawings missing and breakdown of drawings per package. Located at the top are the overall drawings produced, expressed both as a number and a percentage of the total drawings required. Below that, on the far left, are drawings produced and missing drawings per package. In the center, a visual representation of the same...
CONCLUSION

By implementing the described framework, the project team was able to accelerate their project delivery. The project example included in this paper demonstrates that the use of this framework and the application of data analytics, data mining and data visualization technologies has the potential to dramatically transform our industry and the way we deliver projects in the future.

As next steps we are looking to further develop the framework and techniques listed above across the design, construction and operations continuum. The same framework can be applied to other large datasets to measure the performance of equipment, people and buildings. One of the biggest challenges in achieving this goal will be in developing new methods to process unstructured data, data from disparate data systems and educating project teams in how to best utilize results to make better more informed decisions.

Lastly, as in motorsports, if we begin to adopt and leverage the advantages of using data analytics, data mining and data visualization techniques, we will begin to see a change in the way projects are delivered and begin to increase productivity and overall efficiency in our industry.

ACKNOWLEDGEMENTS

Special thanks to the Gehry Technologies Team who worked on the reference project above, helped shape the thinking behind this paper and served as a motivation to publish this work including: Pedro Alvito, Susana Martins, Ouafa Messaoudi, Rodrigo Moura, Allyn Polancic, Sean Ramsey, Lucas Reames Jim Rhone, Nassim Saoud, Lea Sattler and Chuck Xu.
REFERENCES


IMAGE CREDITS
Figures 1–7: Gehry Technologies / Trimble, © 2017

German W Aparicio Jr is Project Manager at Gehry Technologies Inc, a Trimble company, which provides consulting, strategic advisory and advanced project delivery services to leading owners, architects, engineers, builders, fabricators and industry professional worldwide. German Aparicio has both research and professional experience as a consultant for Ove Arup design engineering firm, consultant for AECOM, research associate at SENSEable CityLaboratory (MIT) and scholar in residence at the Autodesk IDEAS Studio. He received a Bachelor of Architecture from Cal Poly Pomona and a Masters of Science in Architectural Studies (SMArchS) Design and Computation from the Massachusetts Institute of Technology (MIT).