

# AD HOC DATA ANALYTICS AND BUSINESS INTELLIGENCE SERVICE FRAMEWORK FOR CONSTRUCTION PROJECTS

Frank L. Wang<sup>1</sup>, Leonardo Rischmoller<sup>2</sup>, Dean Reed<sup>3</sup>, and Atul Khanzode<sup>4</sup>

## ABSTRACT

This paper presents a framework of an ad-hoc data analytic and Business Intelligence service tailored to a construction project. Mandates of delivering integrated information solutions and effective reporting are commonly required nowadays in large capital projects. Due to the nature of construction projects with schedule and budget constraints, poorly defined business problems prohibited the team to deploy full scale data analytic and Business Intelligence (BI) services on site. On the other hand, the increasingly complex data coming from multiple applications and organizations on projects requires more powerful data integration tools and techniques. The proposed framework outlines an agile and ad hoc best practice for job site data analytics and effective reporting based on a real use case from a large pharmaceutical project. Processes in the framework include data alignment, Level of Detail (LoD) data articulation and analytical model establishment. It also illustrates how to resolve complex data analytic challenges for unforeseen cost disputes and how to deliver solutions within a short period of time.

## KEYWORDS

Integration, waste, customization, complex, Integrated Information, Data Analytics

## INTRODUCTION

If design information is inadequate, materials are missing, or prerequisite work is incomplete, the assigned work will cost more, take more time, or be done incorrectly, if at all (Ballard & Howell 1998). Due to the nature of construction projects schedule and budget constraints have been identified by several authors (Alarcon & Ashley 1999;

---

<sup>1</sup> Business Intelligence Lead, DPR Construction, 1450 Veterans Boulevard, Redwood City, CA 94063, USA, Frankwan@dpr.com

<sup>2</sup> Business Analyst, DPR Construction, 1450 Veterans Boulevard, Redwood City, CA 94063, USA, LeonardoR@dpr.com

<sup>3</sup> Lean/Integration Advocate, DPR Construction, 1450 Veterans Boulevard, Redwood City, CA 94063, USA, DeanR@dpr.com

<sup>4</sup> Technology and Innovation Leader, DPR Construction, 1450 Veterans Boulevard, Redwood City, CA 94063, USA, AtulK@dpr.com

Gonzalez et al. 2006; Rischmoller 2006). These problems are aggravated by poorly defined business problems that include: the lack of customer requirements captured by the design team or available during the design, suitable capability of the production system, and construction requirements satisfaction; which are four of the five value generation principles proposed by Koskela (2000) and deepened by Rischmoller. The lack of these principles, as well as the loss of value expressed as ignorance of client requirements, bureaucracy and paper work, information not available, poor interdisciplinary coordination and rework (Rischmoller et al. 2006) are Lean Construction related problems identified in the case study presented in this paper.

A solution to the above-mentioned problems is presented, based on a solid conceptualization of the owner's requirements that led to outlining a data strategy supporting the development of a data model for analytics that helped to overcome miscommunication, misalignment, and loss of trust among the owner, general contractor and steel subcontractor, increasing value and reducing waste in the project.

Integrating information allows project teams to operate from single sources of truth, i.e., a single document containing current and accurate information on which to act (Fischer et al 2014). This eliminates the common industry practice where participating companies expect employees to create documents for their own use rather than share them across a project team. This wastes precious time and opens the door to misunderstanding because information is inaccurate and inconsistent (Flager et al. 2009).

## **PROBLEM AND CHALLENGE**

The owner team of a new life sciences laboratory building wanted a thorough review of the change order for significant modifications to the structural steel frame they had requested. Originally, the work was authorized without review in the interests of time so that steel members could be modified in the shop during scheduled fabrication. While a majority of the modifications were made in time, other steel had to be reworked during erection of the steel frame.

The steel fabricator was contracted lump sum and submitted a summary level description of changes along with the price. The entire change consisted of 69 individual change orders, called "Extra Work Estimation (EWE)." Causality of each EWE was documented in a "*Driver*." A driver could be either:

- 1) An updated drawing set or bulletin: 11 sets of drawings were listed.
- 2) Recorded RFI with attachments: 128 RFIs were associated with the changes.

The change orders were later rejected by the owner's project controls group that had been established later in the project. The general contractor asked the project manager for the steel fabricator to provide more documentation, which he did. The steel project manager provided fabrication control reports called Bills-of-Material (BoM) from the Manufacturing Information System (MIS), shop drawings showing changes and a color-coded erection plan showing the location of the changed steel for each EWE. The owner project controls team was still not satisfied and refused to approve the changes. They

wanted to see the material fabrication data together with shop and labor hours relative to change order requests and authorizations to proceed.

Each BoM contained approximately 3,000 line-items including parts, cuts, welds and joints. The steel fabricator expressed concern when asked to provide detailed labor hours and BoM that were not an obligation under their original Lump Sum contract. The granularity of data was misaligned between the owner, GC and the subcontractor when the request for additional documentation was made.

On the owner side, without a structured layout and guided data navigation method, the owner's project controls staff was overwhelmed by piles of submitted documents. They also started analysing submitted data based on an incomplete understanding of the steel fabrication process, which created more ambiguity. On the other hand, the GC and steel fabricator considered that they had complied with owner's requirements and fulfilled the contractual obligation. This miscommunication and misalignment caused the parties to lose trust and threatened to jeopardize the long-term relationship between the client, GC and subcontractor (Phelps 2012). It also increased the possibility of unnecessary fee erosion to both GC and steel fabricator. Guaranteed Maximum Price (GMP) negotiations had broken down. At that point, the owner's Director of Project Engineering asked the project manager and controls team, the structural engineer, fabricator, and general contractor project managers to meet one last time to find a way through the impasse. The GC managers asked the owner project controls managers to describe the data they needed to determine that the owner was being charged a fair price. An urgent mandate was assigned to the GC and steel fabricator to provide an intuitive reporting mechanism so an outside auditor working with the owner's project control team could easily comprehend overall cost and impact. All parties agreed to give the GC and steel fabricator one week to see what they could produce.

The problem became a data analysis challenge given the nature, amount and complexity of data (Jin & Levitt 1996).

Table 1: Source documents and data to be analyzed

Document	QTY	Description
EWE summaries and invoices	69	<ul style="list-style-type: none"> <li>• Material Cost – Additive/Subtract steel tonnage and cost</li> <li>• Labor Hours Cost - Shop/Field man hours and cost, Field Cost - Additional field work cost</li> <li>• Sub-Tier Costs - Equipment rental, detailing service, freight delivery cost,</li> <li>• Other Costs - Fees, taxes and other costs</li> </ul>
Bill-of-Material per EWE	3000x69 lines (approx.)	<ul style="list-style-type: none"> <li>• All parts, joints, welds, etc.</li> </ul>
RFI	128	
Drawing Sets/ Bulletins	11	

## **SOLUTION**

The GC responded quickly with actions to turn around the rising distrust and resolve the dispute.

- 1) Invite industry lean experts and senior think-leaders to the job, interviewing all parties, analysing the current process, strategizing and facilitating team communication with the owner (Koskela et al. 2016).
- 2) Brought onboard a data analysis solution architect to meet with owner’s project controls team, understand the overall scope and intent, analyse the data, and deliver an acceptable solution (Dave et al. 2008).

## **SOLUTION DEVELOPMENT AND DATA ANALYSIS FOR JOB SITE**

Unlike normal software development, facility owners are often unable to provide detailed specifications or existing examples of an anticipated solution. Conceptualizing complex business problems into deliverables is a known challenge, analogous to a design process. In fact, the solution shown in this paper is no exception. It was a design problem, or arguably a design task with a data driven deliverable.

In conventional development and delivery processes, the conceptualization process is broken down into steps for a chain of specialists, e.g., business analyst, data analyst, developer, etc. However, this model does not suit a construction project very well due to a few impediments, which are:

- 1) Lack of agreement on scope and information. A typical lost-in-translation challenge often occurring during scope and requirement gathering where the project team cannot successfully communicate and provide the right amount of information

from the business to support solution development, causing scope creep or over/under development.

2) Time constraint. The given time for project team to deliver a solution is often very limited and does not allow for a conventional non-agile approach. For example, a week was given to deliver the final solution in this particular case.

3) High level of customization. Problems in hand often are specific to the project due to special reporting requirements from the owner. A certain customization is always needed, and a top-down corporate solution cannot fully resolve the problem.

A better approach learned from the past is to engage a solution architect who has both analytic and development skills to facilitate and advise on delivery of the solution. Such a solution architect must be able to: 1, capture the owner's intention; 2, digest the owner's intention into a solution design; 3, lead implementation of the solution inclusively on time; and 4, enable the scaling up of such a solution for other issues and to other projects later on. The delivered solution should not only fulfill the owner's expectation, but also prevent misuse that could lead to further disputes and undermine trust among parties.

## **REQUIREMENT AND SCOPE CONCEPTUALIZATION**

After a single scope meeting with the owner team, the solution architect was able to translate the owner's high-level guidance into the following "*Design*" criteria.

1) Multi-directional navigation. The owner demanded flexible navigation methods; for example: 1, navigating data from Change Orders (EWE) to correlated drivers and vice versa; and 2, see costs in different aggregated levels such as detailed unit price per part for both reused and scraped and labor cost breakdown at the shop and field level.

2) Correlation and finding insight. Enable correlation of all EWEs relative to a particular driver and visa versa, e.g., find all EWEs that are correlated to the "GMP drawing set."

3) Trending by time. The owner wanted to understand progression and magnitude of change orders throughout the project timeline; that is, make change orders visual as they evolved through time in both count and amount.

4) Data accuracy and transparency. The owner wanted to understand and visualize how cost and tonnage summed up from the BoM level to the final cost per EWE.

## **METHODS AND ACTIONS**

The solution architect outlined a data strategy and step-by-step processes after conceptualizing the owner's requirement into design criteria.

### **Articulate Missing and Derivable Fields**

1) Shop versus Field classification. Submitted BoMs did not obtain a distinct classifier for shop versus field per steel part. Such classification was needed and derivable from several existing fields within the BoM.

- 2) Scrap versus reuse classification. Submitted BoMs did not include a distinct classifier for reused or scrapped steel per part. Such classification was derivable from several existing fields with certain logic.
- 3) Unit Price for each part. Submitted BoMs did not include a unit price on a per-part level. Each part needed to be associated with a unit price in order to aggregate cost to an assembly, type, and entire change order level.

### **Data granularity and LoD (Level-of-Detail)**

Proper data granularity for reporting is normally driven by owner's requirements and contractual obligations. In this particular case, the granularity was readjusted from a summary level down to the BoM level. Because classifications of Shop/Field and additive/subtractive information resided in the detail BoM level, the team needed to readjust reporting data granularity to a proper level that accommodates these classifications. With the adjusted data LoD, the team not only could fulfill the owner's requirement on reporting, but also was able to provide additional insights. For example, the team could explore many intermediate levels of aggregations from the same data, and gain insight such as:

- 1) Summarize by type of part, e.g. channel shapes (C) or angles (L).
- 2) Summarize by assembly level.
- 3) Summarize by Extra Work Estimation (EWE).
- 4) Summarize by all 69 EWEs together.

There were also consequences when readjusting data granularity and collection processes in the middle of the project, especially to a more detailed level. These were:

- 1) Inability to reproduce data from the past – a system may not preserve a transaction log so retrieving historical data therefore is impossible.
- 2) Discontinuity on existing trending.

The requirement to breakdown Shop Labor Hours (SLH) into a per-part level was a good example to illustrate the problem of adjusting granularity to a more detailed level. The labor hours were collected and reported at the EWE summary level prior to the owner's new mandates. By adjusting the data granularity to the BoM level, the subcontractor could not easily provide shop hours associated with each line item from the BoM. The subcontractor may have had the standard SLH and material cost for cost-control internally but faced a tedious data process to ensure the total invoiced amount aligned with the breakdown at a per-part level.

As a summary, an appropriate data strategy, consisting of 1, the right data granularity and 2, establishing a collection process, should be planned ahead by future project team as part of an execution plan early as possible. A good data strategy (Pasquire 2012) could avoid possible discontinuity in trending and provide tolerance to granularity change.

### **Data Model for Analytics**

An analytic data model was created to transform raw data into consumable forms for connected reporting systems. The data model design complied with a typical data

warehouse approach of articulating data into dimensions and facts. Dimensions allowed the analyst to create filters that partitioned data across all datasets, i.e., facts. Dimensions were driven from owner’s requirements for example:

- 1) Extra Work Estimation (EWE) Number: a sequential number of individual change orders.
- 2) Shop/Field: Shop versus Field.
- 3) Drivers: Drawing Sets or RFI.

### Data preparation and cleansing

Like all typical projects of data scientists and analysts, the majority of work went into data preparation and a cleansing process. The analyst needed to process multiple versions of EWE summaries into a single dataset while also preserving all levels of cost hierarchy from the submitted summaries. An EWE number was added to the BoM to correlate two data sets as the joining key. The analyst was able to accomplish processing all EWE summaries and BoM files in two workdays by utilizing superior business analysis tools.

### Reporting and Dashboard Design

A robust dashboard was created by following the four “Design Criteria” listed above. A navigation panel was created to fulfil the multi-directional navigation request, shown in Figure 1 and Figure 2. The analyst was also able to create side filters allowing users to slice and dice the data for more specific focus and insight.

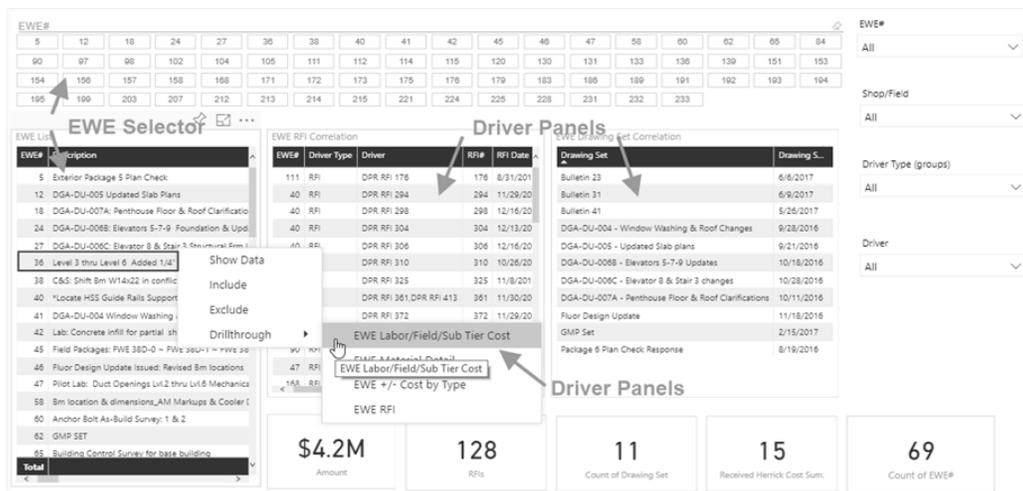


Figure 1: Navigation Panel

From the navigation panel, users were able to drill down to detail summary such as the BoM or Cost Hierarchy from selected EWE(s), and also collapse cost in hierarchy and materials by type, shown in Figure 3. Time trending charts, shown in Figure 4, were also created to illustrate EWEs in both the count and amount throughout the project history.

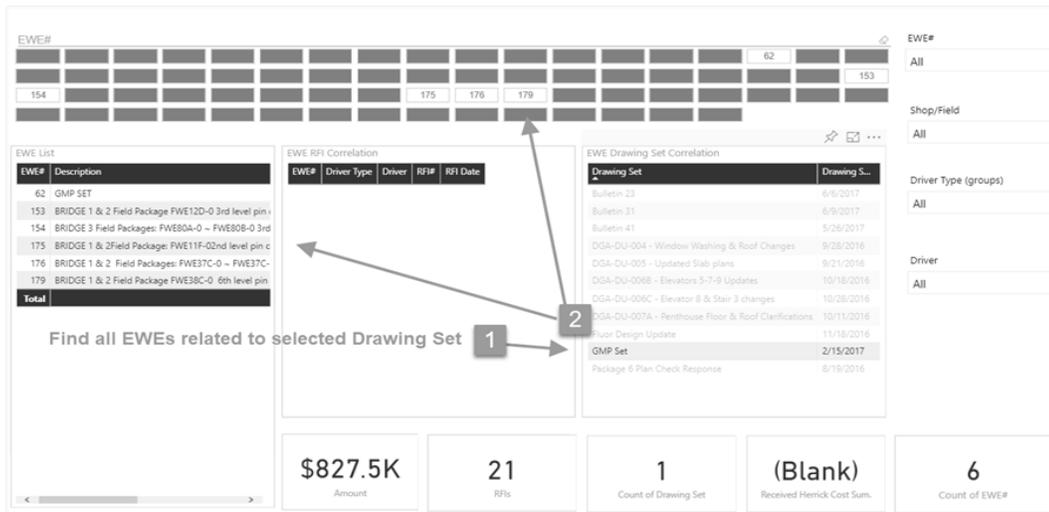


Figure 2: Finding correlation from selected Drawing Sets to related EWEs



Figure 3: Alternative Navigation – Finding correlation from selected Drawing Sets to related EWEs

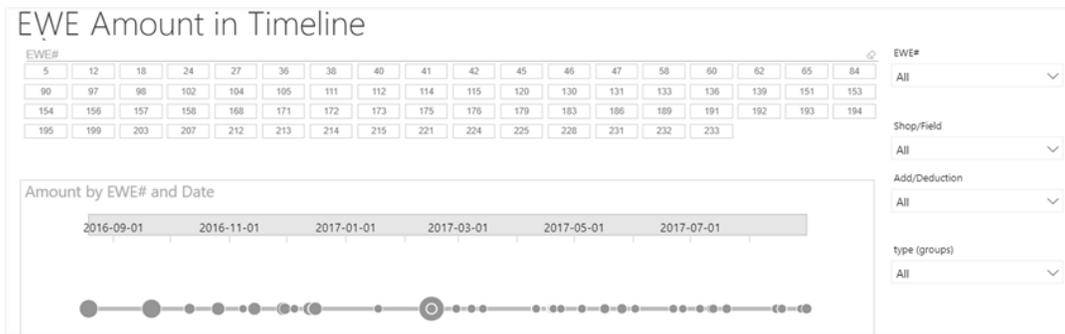


Figure 4: Browsing EWE amount over time showing the volume and amount

## SUMMARY

Working together, the team delivered a full solution within a week, including gathering business requirements, designing outcome visuals, aligning and collecting raw data, establishing the analytic model, and data transformation and cleaning. On the business side, the owner’s project controls team was able to verify that the fabricator and GC had properly accounted for all of the modified steel and priced the changes fairly, including credits for scrapped steel. Respect for good intentions and competence increased among the individual managers. The GC managers realized that things would have gone much faster if they had had asked the owner project control managers what they needed to see earlier rather than provide what they thought was sufficient.

The above framework presents an ad-hoc solution tailored for unforeseen and unexpected, one-time, data analytic challenge that frequently occurs during construction. It was a more agile and possibly cost-effective approach compared to full scale enterprise level data analytic solution. The framework was part of a data democratization effort from the GC IT corporate services, enabling data self-services for people to perform data analysis right on projects, that combines data science advances with value adding and waste reduction Lean principles. Future work based on this study includes continuously enhancing the framework into a formal curriculum along with proper tooling, and training for greater dissemination.

## **ACKNOWLEDGMENTS**

The authors are thankful for support from Rodney Spencley, Erik Kolstoe, Lauren King, Hrishi Maha, and the entire project team.

## **REFERENCES**

- Ballard, G., and Howell, G., 1998. "What kind of production is construction?" *Proc., 6th Annual Conf., International, Group for Lean Construction*, Guaruja, Brazil.
- Fischer, M., Reed, D., Khanzode, A. and Ashcraft, H. (2014). "A Simple Framework for Integrated Project Delivery" *22nd Annual Conference of the International Group for Lean Construction*. Oslo, Norway, 25-27 Jun 2014.
- Dave, B., Koskela, L., Kagioglou, M. & Bertelsen, S. (2008), "A Critical Look at Integrating People, Process and Information Systems Within the Construction Sector." *16th Annual Conference of the International Group for Lean Construction*. Manchester, UK, 16-18 Jul 2008.
- Flager, Forest; Adya, Akshay; and Haymaker, John (2009). "AEC Multidisciplinary Design Optimization: Impact of High Performance Computing." Technical Report 186, Center for Integrated Facility Engineering, Stanford University, CA.
- Gonzalez, V., Alarcon, L.F. and Gazmuri, P. "Design of Work in Process Buffers in Repetitive Building Projects: A Case Study." *Proceedings IGLC-14*. Santiago, Chile, July 2006
- Koskela, L., 2000. "An exploration towards a production theory and its application to construction." Ph.D. thesis, VTT Building Technology, Helsinki University of Technology, Helsinki, Finland.
- Koskela, L., Pikas, E., Gomes, D., Biotto, C., Talebi, S., Rahim, N. and Tzortzopoulos, P. (2016). "Towards Shared Understanding on Common Ground, Boundary Objects and Other Related Concepts." *24th Annual Conference of the International Group for Lean Construction*. Boston, USA, 20-22 Jul 2016.
- Pasquire, C. (2012). "The 8th Flow – Common Understanding." *20th Annual Conference of the International Group for Lean Construction*. San Diego, USA, 18-20 Jul 2012.
- Phelps, A.F. (2012). "Behavioral Factors Influencing Lean Information Flow in Complex Projects." *20th Annual Conference of the International Group for Lean Construction*. San Diego, USA, 18-20 Jul 2012.

- Rischmoller, L., Alarcon, L. F., and Koskela, L. (2006). "Improving Value Generation in the Design Process of Industrial Projects Using CAVT." *Journal of Management in Engineering*, 22(2), 52-60.
- Rischmoller, L. (2006). Chapter 5: Construction Scheduling: a Latin American Perspective, in *Constructing the Future-nD Modelling*. Edited by Ghassan Aouad, Angela Lee and Song Wu. Routledge; 1 edition (December 19, 2006).
- Jin, Y. and Levitt, R.E. (1996). "The virtual design team: A computational model of project organizations." *Computational & Mathematical Organization Theory*. 2(3), 171-195