Connecting the Construction Site

EXPERIENCES REGARDING ESTABLISHMENT OF DIGITAL INFRASTRUCTURE ON CONSTRUCTION SITES, 2021-06-25

SCOTT FOWLER, LIU

LINKÖPINGS UNIVERSITET
LULEÅ TEKNISKA UNIVERSITET
SMART BUILT ENVIRONMENT
UPPKOPPLAD BYGGPLATS

Connecting the Construction Site

Principle form of Big IoT Data Analytics on construction sites

Scott Fowler, LiU

.

IoT data with Machine Learning (ML)-integrated smart analytic tools have been adopted in various sectors but are not fully implemented in the construction industry. Data analytics refers to techniques that uncover patterns from raw data and extract valuable insights for decision making, assessment, and forecasting, while Machine Learning (ML) provides optimum mathematical methods for such techniques.

In this project, we will explore how the construction industry will be benefitted from IoT-generated Big Data in combination with advanced analytic tools, taking the concrete industry as an example. Using digital IoT, we have obtained Big Data on multiple aspects of the concrete hardening process to optimize the production procedure in a friendly manner. The outcome will lead us to the next phase, where we gain new insights about the future usage of new concrete types that have been otherwise “hidden” or “undiscovered” by traditional analytic methods.

.

1 Introduction

The construction industry produces buildings and facilities on construction sites that are "temporary factories" and supply chains built up that are dismantled again and reassemble again in new places. The industry's productivity has trailed compared to other sectors for decades. At least 20% of the total construction volume is waste, 30% of construction is rework, 40% of job site work is unproductive, 40% of projects are over budget, and 90% of projects are late [1]. Typically, this is blamed on the complex and chaotic nature of construction production. This apparent complexity arises from the fact that currently, no one has an accurate real-time picture of what is taking place on construction sites. Increasing the usage of technology on the construction site will allow new applications and innovation, as well as improving the engagement of relevant external stakeholders. To date, digitalization of the worksite has primarily benefited the office functions but has not resulted in higher productivity in operation. One way to improve the situation is the implementation of Internet-of-Things (IoT) with analytic tools (e.g., Artificial Intelligence (AI), optimization) in the construction industry.

For the past year, Linköping University has been conducting various tests on the digitization of the construction industry in collaboration with leading construction companies, aiming for the industry's development in Sweden. A tool that has been applied is the so-called Industrial Internet of Things (IIoT) that allowed us to gather a vast amount of data.

Concrete is one of the most important materials in the construction industry as it is used in almost all parts of infrastructure projects. Applying IIoT (sensors) in concrete production is a good start in the digitalization of the industry, and so far, it has been successful in gathering a huge volume of data in the various aspects of the concrete production process. However, data collection by itself provides no benefit. Instead, the data generated by IIoT has to be translated into a somewhat understandable form. Martin Laninge from Peab expressed the need more clearly:

"We have used sensors in concrete before, but this is the first time that we put the measurement data into a larger context and connect it to digital plant models in 3D."

Using 3D visualisation is the important first step, and the natural progression from that is the utilization of the various data sets generated by IIoT. A large amount of Data (know as Big Data) itself does not result in improved productivity, safety, or environmental issues in the construction. The data needs to be properly inserted into the management of construction and logistics processes in the construction stage. To achieve this, data analytics are required. Big Data analytic techniques uncover patterns from raw data and extract valuable insights for making informed decisions.

Our long-term goal is to create a technological leap for the construction industry that results in digital working methods, more automated production, more efficient logistics, and more integrated planning. To study these objectives, better control over the construction process is required. Therefore, our team consists of leading companies in the industry, Cementa, and Peab who are responsible for one of the testbeds. The companies have been working with the experts in the IIoT (sensor) systems, EzeSystem and Celsicom, to investigate how IIoT wireless sensors in the concrete connected to data and forecasting tools can provide better and faster control over the temperature development and drying out of moisture when the concrete hardens, among other things.

“In a complex building like Olskroken, it is extremely important to keep track of the curing cycle.” says Robert Larsson of Cementa, testbed manager within the project Connected construction site.

“In a complex building like Olskroken, it is extremely important to keep track of the hardening cycle. Here, 660 trains pass per day and the Swedish Transport Administration has an advance of 18 months if one of the lanes has to be closed for a while”

Within the framework of the Connecting the Construction Site, digital infrastructure for the construction industry, we have been conducting research partly through tests of the technology in the sub-projects, and partly through studies of other industries such as the process industry. This white paper focuses on the base data monitoring and understanding of IIoT-generated data in concrete curing.

Integration of the various IoT data enables comprehensive monitoring and decision-making of the project. Real-time (or near to real-time) data collected from IIoT on a construction site allows technologies, such as BIM, to model and monitor the work process. Data fusion is a necessary step to achieve such sophisticated technologies, as they demand extremely large big and complex data set. Converting the raw data sets into more consistent, accurate, and useful information is critical in developing ubiquitous environments, which then can establish interoperability between, for example, IIoT data and digital models like BIM. Data fusion is particularly important for IIoT applications, where a timely fusion of data is needed to bring all pieces of data together for analysis, and consequently, provide reliable, accurate, and actionable insights. Data fusion, *per se*, is beyond the scope of the current white paper, but we will focus on the insight of the data that has been collected.

IoT communication protocols aim for ubiquitous connections among objects that allow seamless real-time data collection, processing, and analysis from a range of physical equipment, processes, and operations. We conclude and emphasize the necessity of a greater scale approach to meet the needs of the construction industry, i.e. implementation of Big Data Analytics to digitalize construction sites [2], which this report revolves around. The next section will discuss some of the Big Data Analytics.

2 Big Data Analytics

The data produced via IIoT accelerates the development of all areas of technologies and businesses. IoT sensor nodes gather data from the environment daily, generating a large volume of data. Such a high volume of data is typically referred to as “Big Data.” Big Data can be categorized according to three aspects: (a) volume, (b) variety, and (c) velocity [3]. Note that we only mention these three parameters because our focus is to find patterns for better decision-making in construction.

- Volume: the volume of generated data dictates whether the dataset is considered Big Data or traditional massive data.
- Velocity: the data generation rate that can be calculated in time or frequency and it relates to the need for data analysis.
- Variety: Big Data consists of the datasets in different forms and types, and this diversity and heterogeneity are a major challenge

Big Data in the construction industry signifies the data generated from the life cycle of the project, for instance, the phases of planning, design, tendering and bidding, construction, and operation management. Big data can be generated from several sources, such as IoT (i.e. sensors or RFID readers) [4], information systems (i.e., BIM, project management system (PMS), and historical project documents) [5]. Integration of such multiple datasets is a process called data fusion, where the datasets become more consistent, accurate, and useful information. This process is critical in developing ubiquitous environments based on IoT data for establishing interoperability, e.g. linking the IoT data to BIM.

In this project, we study Big Data in concrete production. The curing of concrete is dependent on various factors including climate and atmosphere, while on-site, concretes are often cured in extreme weather conditions. To improve this, IoT devices were applied to collect a vast amount of data to help to understand the hardening process of concrete types,

3 IIoT and Big Data Analytics

Big Data analytics in IoT process a massive amount of data, whether it is offline or real-time, and store it with various data storage technologies. The data format can be labeled (structure) or non-labeled (unstructured), and is gathered directly from web-enabled things or, so-called connected things. Big Data analytics' performance is super-high-speed for large queries so that companies/organizations can gain insights for quick decisions and interact with people and other devices immediately. Once IIoT is applied in the construction sites, the volume of data collected via IIoT will grow exponentially, and it will become humanly impossible to review and analyze the data quickly. To address this, Big Data analytics tools (e.g., Machine Learning, optimization) should be deployed in construction for data-driven decision-making. Over the past few decades, various Big Data analytics tools and algorithms have been proposed and it became crucial to select the right tools for the right settings.

4 Big Data Analytics Methods

Big data analytics aim to extract useful information that helps in making predictions, extract trends, identify hidden information, to help in making decisions [6]. Data mining methods are widely used for both problem-specific and generalized data analytics. Although the requirements for efficient mechanisms lie in all aspects of big data management [7], such as capturing, storage, preprocessing, and analysis; for our discussion, big data analytics requires the same or faster processing speed than traditional data analytics with minimum cost for high-volume, high-velocity, and high-variety data [8].

Knowledge of available big data analytics options is crucial when evaluating and choosing an appropriate approach for decision-making. Machine Learning (ML) plays an important role in analytics since it is algorithms that process and extract information from data; they facilitate the automation of tasks. These advances have, in turn, fueled interest and progress in the field of ML to extract information from these data. These learning algorithms may be categorized into supervised, semi-supervised, and unsupervised learning (see Figure 1), depending on the information available about the data to the learning machine.

The Figure depicts and summarizes three learning categories 1) Supervised learning: learning from data labeled with expert knowledge, providing corrective information to the algorithm, 2) Semi-supervised learning: learning with partially labeled data (generative adversarial networks) or by interactions of the machine with its environment (reinforcement learning); 3) Unsupervised learning: learning without labeled training data. Each learning category is an ML function and involves many methods and algorithms to fulfill information extraction and analysis requirements.

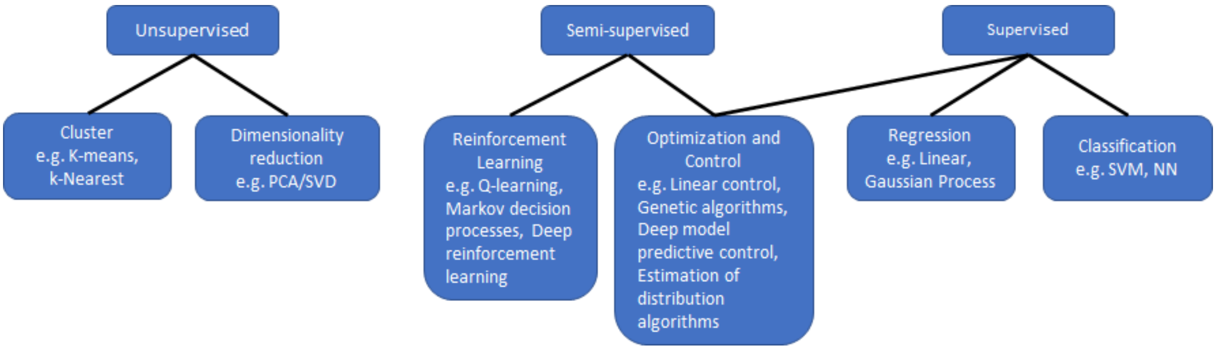

Figure 1 Machine learning algorithms may be categorized into supervised, unsupervised, and semi-supervised, depending on the extent and type of information available for the learning process. Abbreviations: PCA, principal component analysis; SVM, support vector machines; SVD, singular value decomposition; NN, Neural Network.

Various solutions are available for big data analytics, and the continuous advancements of these solutions make them suitable for new big data trends. Data mining plays an important role in analytics as most of the data analytics techniques are developed using data mining algorithms according to a particular scenario. There are many challenges to implementing analytics algorithms on the data, for example; whether to employ on the environment cloud or the premise; data warehouses for analytics and data storage; which approach should be performed when deploying data analytics models; whether to do real-time, semi or off-line data analysis; and so on. Knowledge of available big data analytics options is crucial when evaluating and choosing an appropriate approach. Next, we present some methods that can be implemented for several big data case studies.

5 The usage of Machine Learning (ML)

The study aims to apply advanced analytic tools for IoT/Big Data to mitigate the environmental risk associated with the construction industry, taking cement production as an example. A huge amount of concrete is used in construction, contributing to significant CO_2 emissions to the environment as well as heat generation during the cement hydration. To address such issues, alternatives or variations of cement mixtures have been explored to date. The current study will provide innovative insights about the performance of newly improved concrete types that have been otherwise undiscovered by traditional analytic methods, using advanced analytic tools for IoT/Big Data.

Here, we introduce the Singular Value Decomposition (SVD) and demonstrate to apply the SVD to our case. The SVD is a factorization of complex matrix and a foundation for nearly all of the data methods in ML as it provides a numerically stable matrix decomposition. The SVD is used as the underlying algorithm of Principal Component Analysis (PCA), and it decomposes data into the most statistically descriptive factors. Thus, engineering and technologies are widely benefitted from the SVD/PCA.

The first step is to see if it is possible to form a simple model ingredient to make concrete using the analytic tool SVD. We can build a simple SVD Machine Learning multilinear regression model that describes heat of cement generation data from the ingredient. The data is collected from the Portland cement [9] since Portland cement is the most common type of cement in general use around the world as a basic ingredient of concrete, mortar, stucco, and non-specialty grout. This simple dataset will serve as the baseline for optimum content of supplementary cementing materials to be blended. It is known that when the fillers, such as limestone, are blended with Portland cement properly there will be many environmental and technical advantages, such as enhancement of sustainability in the concrete industry, increase in physical properties, and reducing CO_2 emission. First, we need to represent the system.

Many physical systems may be represented as a linear system of equations:

$$\boldsymbol{Ax} = \boldsymbol{b}, \tag{1}$$

where the constraint matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$ are known, and the vector $\boldsymbol{x}$ is unknown. In this problem, we are solving (1) where $\boldsymbol{A}$ is a 13×4 matrix with four ingredients as columns and heat measurements for 13 unique mixtures of these ingredients as rows. The vector $\boldsymbol{b}$ of 13 rows representing the output; how much heat is generated in the process of mixing the cement ingredients as it cures. The goal is to determine the weighting $\boldsymbol{x}$ that relates the proportions of the four ingredients to the heat generation. It is possible to find the estimator quality (also known as minimum error) solution using the SVD. Alternatives to SVD are using Multiple linear regression and Moore-Penrose pseudoinverse function (in MATLAB, the functions are labeled regress and pinv), also can be examined.

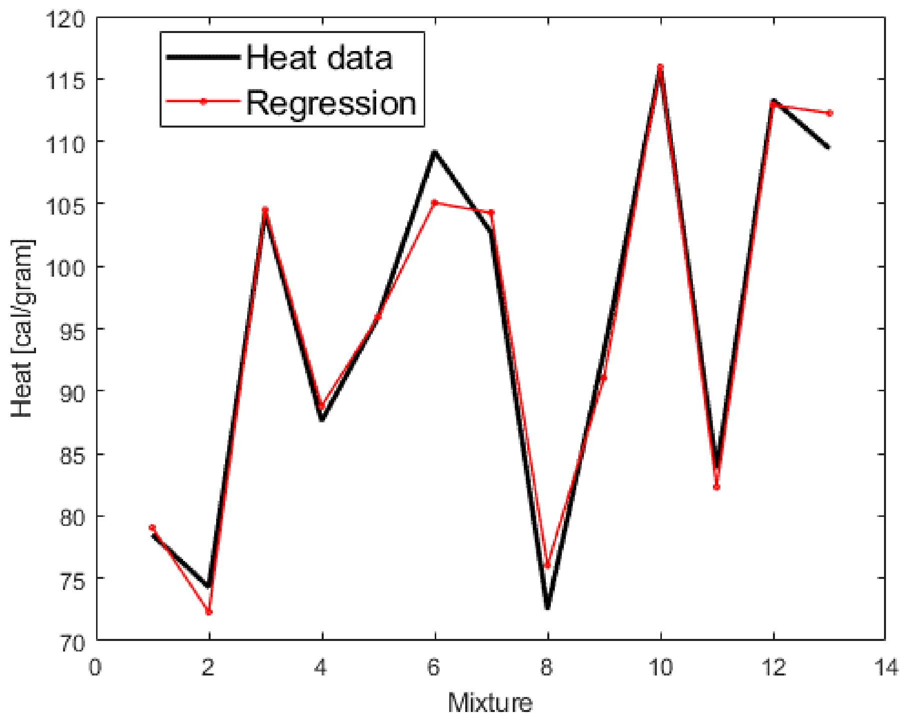

Figure 2 Heat data for cement mixtures containing four basic ingredients.

In Figure 2, the graph illustrates the heat generation across 13 experiments, the measured data in black and the regression model in red. The x-axis represents the 13 unique mixtures, and the y-axis is the heat generated from the 4 mixtures. It demonstrates that a linear regression model can faithfully

capture this data set. However, the heat measurements on the various ingredient composition are overly simplified plot and several other factors need to be taken into consideration. For example, the degree of maturity is a value representing the concrete hardening and taking into account both concrete temperature and hardening time. A calibration of concrete recipes and the drying process is also required. Thus, a vast amount of data needs to be collected and analyzed to be successful in a broad introduction of new concrete types, such as climate improved concrete, to the world market.
In the next section, we will discuss nonlinear curve fitting based on the drying time. The case deals with finding optimal balances between establishing a good environment for drying of the concrete and at the same time minimize energy costs and secure a good working environment.

6 Curve Fitting based on drying time

Data measured from IIoT devices are noisy in nature, and often, only the trends in the data are sought. As presented in the previous section, a fundamental tool for data analysis and recognizing trends in physical systems is curve fitting in data measurements. The concept of curve fitting is fairly simple: use a simple function to describe a trend by minimizing the error between the selected function to fit a set of data. One such method, regression curve fitting, attempts to estimate the relationship among variables using a variety of statistical tools. One can consider the general relationship between independent variables $\boldsymbol{X}$, dependent variables $\boldsymbol{Y}$, and some unknown parameters $\boldsymbol{\beta}$:

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\beta}), \tag{2}$$

where the regression function $\boldsymbol{f}(\cdot)$ is typically prescribed and the parameters $\boldsymbol{\beta}$ are found by optimizing the goodness-of-fit of this function to data. Broadly speaking, ML is framed around regression techniques, which are themselves framed around optimization based on data. Thus, at its absolute mathematical core, ML and data science revolve around positing an optimization problem. The success of optimization depends on defining an objective function which too is optimized. Polynomial and exponential curve fitting are used commonly, however, such curve fits are highly specialized, and a more general mathematical framework is necessary for solving a broader set of problems. In the example data set, we wish to fit a nonlinear function of the form. In this case, general nonlinear curve-fitting leads to a system of nonlinear equations instead of a linear system. Thus, a nonlinear multivariable function that best fits the series of data points is to be described below.

The given dataset is nonlinear in the coefficients and requires formulation to fit a nonlinear model to data. Nonlinear models are more difficult to fit than linear models because the coefficients cannot be estimated using simple matrix techniques. To fit the nonlinear data set to a prediction function which define the data set to a correlation formula is:

$$y_p = p_1 + \frac{p_2}{x} + p_3 * ln(x), \tag{3}$$

where $\boldsymbol{x}$ is the length of the data set, p_1, p_2, and p_3, that are the three unknown values where we want to come up with the three best parameters to minimize the sum of error squares. Accordingly, the unknown coefficient parameters are an initial estimate for each coefficient. Next, we need to come up with the best parameters to be able to minimize the normalized sum of squared errors. Our objective is to minimize the y_m predicted minus y_m measured divided by the y_m. Then, we will square that and sum up the different measures from ***i*=*1*** to ***n*,** which then will become the objective. We will adjust the unknown parameters p_1, p_2, and p_3, and solve the problem using the following nonlinear least-squares (curve-fitting) approach. We define an objective function (scaled sum of squared errors) as:

$$\min_p \sum_{i=1}^{n} \left(\frac{y_p - y_m}{y_m}\right)^2, \tag{4}$$

where y_p is predicted and y_m is the measured value.

Figure 3 provides a visual understanding of the usage of (4). This example deals with finding the optimal environment for drying concrete. The x-axis is the length of time required for the heaters to dry the concrete, while the y-axis is the temperature of the IIoT heater. The blues points represent data measurements of the temperature, and the black line represents nonlinear least-squares fit which is labeled as optimal predicted. The graph demonstrates how the noise data from the heaters fit the optimal predicted line.

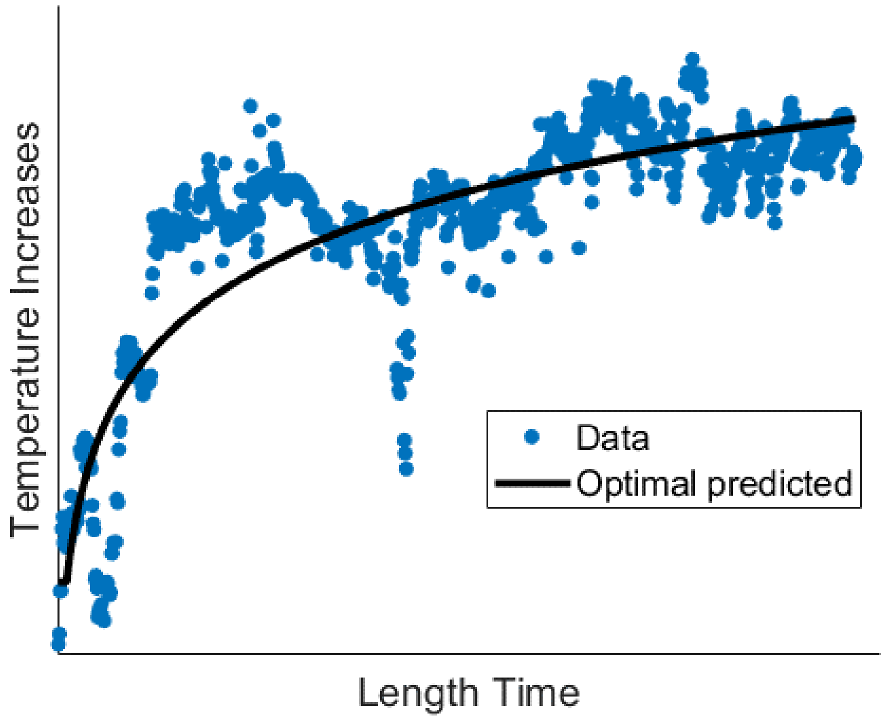

Figure 3 Nonlinear Regression Temperature

With the help of IIoT heater systems, one can measure as many variables as possible, and once an optimal predicted line is applied for such data, it becomes possible to make predictions of future states based on actual measurements[1]. In our case, the application of the advanced analytic tools provides a better understanding of optimal drying of concrete and their performance, which in turn results in insightful knowledge for decision making on, for example, means for lower life-cycle costs.

7 Conclusion

In this white paper, we explored an application of IoT with ML big data analytics in the construction sector, taking concrete production as an example. In the process of concrete production, there are many variables affecting

[1] *Note that it is important to consider calculating the Relative Humidity with Dew Point and Temperature to have even curing of the concrete.*

temperature, including the position of the sensor in the element, the weather conditions, the type of form and insulation used, or the use of heating cables, cement mixture, etc. Big Data regarding such parameters can be generated by IoT, and then translated into meaningful results that can be used for decision making by advanced analytic tools. The outcome of studies will lead us to the next phase, where we obtain new insights about the performance of new climate-improved concrete types that have been otherwise hidden or undiscovered by traditional analytic methods.

Today, the interaction between IoT and Big Data analytics is still at a stage where processing, transforming and analyzing large amounts of data at a high frequency is needed, while Big Data analytics methods were not fully studied. This paper presented examples of data analytics applied in cement production. We concluded that existing big IoT data analytics solutions remained in their early stages, and real-time analytics solutions that can provide quick insights need to be developed, especially for the construction field.

Reference

[1] Forbes, L., Ahmed, S. (2011). Modern Construction. Boca Raton: CRC Press.

[2] Dick, J., and Payne, D. (2005), "Regional Sectoral Support: A Review of the Construction Industry, SMEs and Regional Innovation Strategies across Europe ", International Journal of Strategic Property Management 9: 55-63

[3] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT Big Data and streaming analytics: A survey", IEEE Communications Surveys Tutorials, 20(4):2923–2960, 2018.

[4] F. Alam, R. Mehmood, I. Katib, N. N. Albogami, and A. Albeshri, "Data fusion and IoI for smart ubiquitous environments: A survey", IEEE Access, 5:9533–9554, 2017.

[5] Zhijia You and Chen Wu, "A framework for data-driven informatization of the construction company", Advanced Engineering Informatics, 39:269–277, 2019.

[6] C.-W. Tsai, "Big data analytics: A survey",' J. Big Data, vol. 2, no. 1, pp. 1_32, 2015.

[7] A. Siddiqa, “A survey of big data management: Taxonomy and state-of-the-art”,' J. Netw. Comput. Appl., vol. 71, pp. 151_166, Aug. 2016.

[8] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A. C. Coello, “A survey of multiobjective evolutionary algorithms for data mining: Part I”, IEEE Trans. Evol. Comput., vol. 18, no. 1, pp. 4_19, Feb. 2014.

[9] MATLAB Sample Data Sets, “hald.mat”, https://www.mathworks.com/help/stats/sample-data-sets.html

Contact

For more information on the content of this report or the project as a whole, please contact:

Co-Investigator – Connecting the Construction Site
Dr. Scott Fowler, scott.fowler@liu.se

Project Manager – Connecting the Construction Site
Professor Martin Rudberg, 0734-14 10 22, martin.rudberg@liu.se
Professor Lars Stehn, 0920-49 19 76, lars.stehn@ltu.se

SMART BUILT ENVIRONMENT
UPPKOPPLAD BYGGPLATS

LINKÖPINGS UNIVERSITET

LULEÅ TEKNISKA UNIVERSITET

Med stöd från

VINNOVA Sveriges innovationsmyndighet

Energimyndigheten

FORMAS

Strategiska innovations-program