Big Data Architecture for Construction Waste Analytics (CWA): A Conceptual Framework

ABSTRACT

In recent times, construction industry is enduring pressure to take drastic steps to minimise waste. Waste intelligence advocates retrospective measures to manage waste after it is produced. Existing waste intelligence based waste management software are fundamentally limited and cannot facilitate stakeholders in controlling wasteful activities. Paradoxically, despite a great amount of effort, the waste being produced by the construction industry is escalating. This undesirable situation motivates a radical change from waste intelligence to waste analytics (in which waste is propose to be tackle proactively right at design through sophisticated big data technologies). This paper highlight that waste minimisation at design (a.k.a. designing-out waste) is data-driven and computationally intensive challenge.

The aim of this paper is to propose a Big Data architecture for construction waste analytics. To this end, existing literature on big data technologies is reviewed to identify the critical components of the proposed Big Data based waste analytics architecture. At the crux, graph-based components are used: in particular, a graph database (Neo4J) is adopted to store highly voluminous and diverse datasets. To complement, Spark, a highly resilient graph processing system, is employed. Provision for extensions through Building Information Modelling (BIM) are also considered for synergy and greater adoption. This symbiotic integration of technologies enables a vibrant environment for design exploration and optimisation to tackle construction waste.

The main contribution of this paper is that it presents, to the best of our knowledge, the first Big Data based architecture for construction waste analytics. The architecture is validated for exploratory analytics of 200,000 waste disposal records from 900 completed projects. It is revealed that existing waste management software classify the bulk of construction waste as mixed waste, which exposes poor waste data management. The findings of this paper will be of interest, more generally to researchers, who are seeking to develop big data based simulation tools in similar non-trivial applications.

KEYWORDS: *Construction Waste; Big Data Analytics; Building Information Modelling (BIM); Design Optimisation; Construction Waste Analytics; Waste Prediction and Minimisation;*

1 INTRODUCTION

1.1 CONSTRUCTION WASTE—AN OVERVIEW

Rapid urbanisation and the appetite to build national infrastructure has escalated the construction activities globally. Notwithstanding the benefits, the adverse impact of construction activities on environment has serious implications worldwide [1]. Construction industry is noted for consuming bulk of rare natural resources and producing hefty amounts of construction and demolition (C&D) waste [2]. The construction industry is the UK's largest consumer of natural resources, using over 400 million tonnes of material per annum and is responsible for producing 120 million tonnes of construction, demolition and excavation waste yearly–around more than one third of all waste arising in the UK [3]. With the rising cost of construction projects and the growing environmental concerns, the construction industry is under immense pressure from government and environmental agencies to minimise construction waste and adopt more sustainable practices.

1.2 WASTE INTELLIGENCE—CURRENT STATE OF WASTE MANAGEMENT

Current waste management systems are based on what is called '*Waste Intelligence*' which is more about suggesting remedial measures to manage construction waste after it happens [4]. Waste intelligence based systems are mainly concerned about reports, dashboards, and queries on small amounts of current and past waste data [5]. These systems can efficiently answer close-ended questions such as project/site wise waste generated, progress towards defined waste targets, and understanding how a particular design strategy generates waste [6]. To answer such questions, these systems typically aggregate historical waste data or group it in some way (e.g. by RIBA stages, by material families, and so on). The end users are provided hindsight with limited insight on waste management activities.

1.3 WASTE ANALYTICS—NEXT GENERATION OF WASTE MANAGEMENT

In contrast to the static Waste Intelligence approaches, the methodology of `*Waste Analytics'* proposes to deploy data-driven decision making at the design stage to significantly cut down on construction waste [3,7]. Evidence from literature [8–12] has shown that utilising waste minimising at the design stage is most promising; this is leading to the development of a consensus that waste minimisation through design (a.k.a. designing out waste) is the future of mainstream research in construction waste management [13]. Waste Analytics is mainly concerned with holistically designing out construction waste.

Specifically, Waste Analytics is the process of proactively analysing disaggregated and huge construction datasets to uncover latent trends or non-obvious correlations pertaining to design, procurement, materials, and supply-chain within the construction delivery process, which lead to construction waste during the actual construction stage. Waste Analytics, by comparison, investigates waste-related data in a more forward-looking and exploratory way [12]. Through analysing historical data, it enables the development of robust predictive models for construction waste estimation. Waste estimation models proactively inform about the amounts of waste arising from building design. Thus, designers optimise design accordingly for waste minimisation from myriad perspectives by asking more open-ended questions [5]. Rather than just aggregating data, it employs advanced analytical approaches (such as time series analysis) to forecast waste and prescribe best course of actions to pre-emptively minimise construction waste. It provides insight on current waste trend of the design and foresight to optimise design for designing out construction waste.

1.4 BIG DATA FOR WASTE ANALYTICS

Big Data is the emerging capability to store and analyse large volumes of data scalably and reliably using a cluster of commodity servers [14,15]. There is tremendous interest in utilising the information in Big Data for analytics, not only to understand latent trends (exploratory analytics and descriptive analytics), but also for predictive & prescriptive analytics to forecast and shape future events [16]. Mostly, the advanced analytical techniques for Waste Analytics are supported by the Big Data technologies. Therefore, Big Data driven Waste Analytics is the next emerging trend that offers unprecedented opportunities to minimise construction waste through design. This synergistically integration of technologies (Big Data, Designing out Waste, and BIM) is a real game changer and promises the development of a resilient BIM based construction waste simulation tool to facilitate the designers in making right decisions to avoid construction waste in future construction projects.

Waste Analytics depends increasingly on high-performance computation and large-scale data storage. It requires large number of diverse datasets pertaining to building design, material properties, and construction domain knowledge for successfully carrying out the underpinning analytical tasks. Mostly, these datasets are highly complex, voluminous, heterogeneous, and incomplete [10,17,18]. Storing these datasets using traditional technologies and subjecting the data to real-time processing for sophisticated analytics is a very challenging proposition. This motivates the use of Big Data technologies to manage and analyse this data of unprecedented size.

1.5 Justification for Research and Contribution of This Paper

There exists an obvious technological gap in existing literature on designing out construction waste. In particular, there is very little work on using Big Data techniques for construction waste minimisation. Developing a robust construction waste simulation tool, in particular, is the ultimate objective of this ongoing R&D effort. The intended tool will equip designers with well-informed and data-driven insights to optimise design for designing out waste through their BIM authoring software (such as Revit, MicroStation, etc.). To this end, this study proposes a Big Data architecture for construction waste analytics—an essential first step towards the development of a non-trivial construction waste simulation tool. The components, and relevant technologies, of the proposed architecture are conceived to store and analyse the emerging construction datasets of unprecedented size for real-time design exploration and optimisation. Since the architecture is supposed to support lifecycle stages of Waste Analytics, the paper contributes by detailing the Waste Analytics lifecycle as well. The term 'Architecture' in this text, is not used as architectural profession used in the construction industry, rather it is used as computer architecture that refers to the high-level structures of a software system.

The remainder of this paper is organised as follows: In the next section, the research methodology, focus of the paper and research objectives are discussed. Section 3 expounds the literature review where the emerging concept of designing out construction waste and the complexities surrounding its true implementation are described: this paper also discuss the strengths and weaknesses of competing Big Data platforms. Section 4 deliberates the waste analytics lifecycle and its relevance to designing out construction waste. Section 5 explains the proposed Big Data architecture for construction waste analytics. In section 6, preliminary results are presented, and finally in section 7, conclusions are provided along with a discussion for future work.

2 Methodology and Focus of the Paper

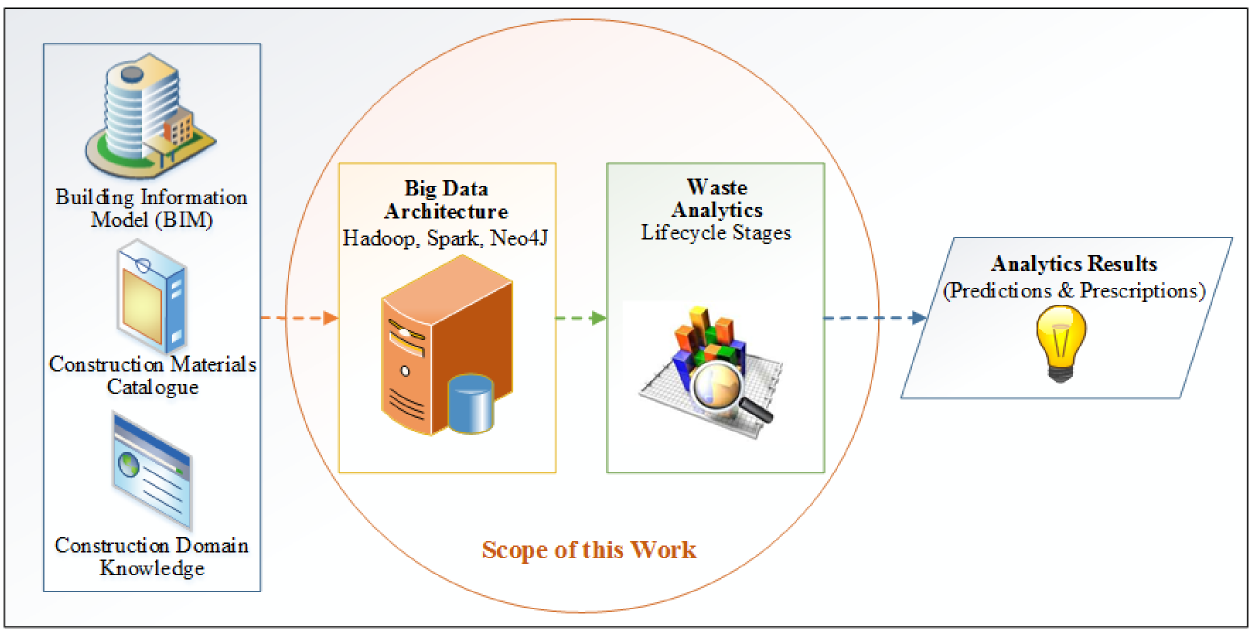

Figure 1: Focus of this paper

In this section, the twofold methodology adopted to carry out this research is discussed. An exhaustive literature review is initially conducted to propose the artefacts of intended waste analytics architecture and waste analytics lifecycle, which are later validated by employing them to perform the preliminary analysis over construction waste related data.

In order to propose a holistic Big Data architecture and waste analytics lifecycle, a thorough review of the extant literature on designing out construction waste, Big Data, and BIM has been carried out. In this regard, online databases of journals such as Journal of Big Data, Big Data Research, VLDB Journal, Automation in Construction, American Society of Civil Engineering (ASCE), Waste Management, and Resources, Conversation and Recycling are searched for research articles between 2000 and 2015. Recent reviews of research and books on Big Data Analytics are also considered [5,19,20]. Some of the search words include: "designing out construction waste", "design strategies for construction waste minimisation", "BIM for waste minimisation", "Big Data in Construction", "Big Data based Application Architecture", and "Big Data Analytics". Overall, 83 publications were selected. While the literature search is not exhaustive (not all publications have been incorporated due to the great breadth of published literature), it is believed that the literature search has captured a representative balanced sample of the related research.

Studies where Big Data is used to develop enterprise applications are included. Studies that are focusing on construction-related waste (e.g., municipal or hazardous waste) are excluded. This reduced the number of selected articles to 64. Each of these articles is then further scrutinized for its relevance by reading their abstract, introduction, and conclusions. Eventually, 55 articles are selected for review in this study. Table 1 depicts how these selected articles are relevant and contributing to the development of proposed architecture, which is primarily based on three key constituents, namely Big Data, BIM, and construction waste. This paper proposes a Big Data architecture and waste analytics lifecycle stages of designing out construction waste. The focus of this study is shown in Fig. 1. The objectives of this study are:

- *Devising the lifecycle stages to carry out construction waste analytics*
- *Developing Big Data architecture for construction waste analytics*

2.1 Analysis & Preliminary Results

The proposed architecture is further ensured and validated using the objective data, taken from the top waste management company in the UK, who offers broad range of recycling and waste management services, including skip & container hire, onsite waste segregation, site waste management services, including skip & container hire, onsite waste segregation, site waste management plans (SWMP), plasterboard recycling, etc. The company uses relational database to store the waste related data from a large number of other construction companies. The data is stored as individual waste movements from site, by project, with the major details of the waste transfer note being recorded. Every time it transports the waste, a digital record is created in their database. Full details about the name of fields for which the values are captured in these records are shown the Listing 1.

Table 1: Summary of article w.r.t contribution for developing waste analytics architecture

Sr.#	*Article Referenced*	*Contribution to Waste Analytics Architecture*		
		Big Data	***BIM***	***Waste***
1	Oyedele et al. [1]		√	√
2	Osmani et al. [2]			√
3	Ekanayake & Ofori [3]		√	√
4	Wu et al. [4]		√	√
5	Camann et al. [5]	√		
6	Lu et al. [6]			√
7	Poon et al. [7]			√
8	Cheng & Ma [8]		√	√
9	Ajayi et al. [9]			√
10	Bilal et al. [10]	√	√	√
11	Akinade et al. [11]		√	√
12	Liu et al. [12]		√	√
13	Osmani [13]		√	√
14	Manyika et al. [14]	√		
15	Diebold [15]	√		
16	Siegel [16]	√		
17	Kim et al. [17]		√	
18	Radinger et al. [18]	√		
19	Basu [19]	√		
20	Ryza et al. [20]	√		
21	Panos et al. [21]			√
22	Keys et al. [22]		√	√
23	Langdon et al. [23]		√	√
24	Ajayi et al. [24]			√
25	Fan et al. [25]	√		
26	Jacobs et al. [26]	√		
27	Thomas et al. [27]	√		
28	Singh et al. [28]	√		
29	Stonebraker et al. [29]	√		

30	White et al. [30]	√		
31	Ghemawat et al. [31]	√		
32	Dean & Ghemawat [32]	√		
33	Berkeley [33]	√		
34	Apache Software Foundation [34]	√		
35	Beetz et al. [35]	√	√	
36	Robinson et al. [36]	√		
37	Halevy et al. [37]	√		
38	Garcia-Molina et al. [38]	√		
39	Chaudhuri & Dayal [39]	√		
40	Martínez et al. [40]		√	√
41	Shi & Xu [41]			√
42	Fatta et al. [42]		√	√
43	Solís-Guzmán et al. [43]			√
44	Shepperd & Kadoda [44]	√		
45	Mair et al. [45]	√		
46	Card et al. [46]	√		
47	Chen [47]	√		
48	Healey & Enns [48]	√		
49	Keim [49]	√		
50	Keim et al. [50]	√		
51	Lu et al. [51]			√
52	Wu et al. [52]			√
53	Laney et al. [53]	√		
54	Fayyad et al. [54]	√		
55	Wu et al. [55]	√		

- Operation Code
- Reference No
- Business Stream
- Hub
- Project Name
- Site (including site number/ reference)
- Region
- Purchase Order No
- Ticket number (Unique identifier)
- Collection/uplift date
- Container size and type
- Movement type
- Classification
- Waste Type
- Waste Collected in Tonnes

- Division
- Hazardous/ Non-Hazardous
- Waste Transfer note number
- Hazardous/ Special waste consignment note number
- Hazardous waste premises code
- Waste Carrier
- Waste Carrier license
- Carrier expiry date
- Carrier Checked with EA/ SEPA and in compliance
- Total Cost
- Cost per Tonne
- Waste Management facility type
- Waste Management facility permit or exemption number
- Type and quantity in compliance with permit or exemption and checked with EA/ SEPA
- Description of Waste Management Innovation
- Tonnes to Landfill
- Tonnes to Other disposal location (not landfill)
- Tonnes Recycled/ recovered
- Tonnes to Other Recovery
- % to Landfill
- % to Landfill
- % Recycled

Listing 1: Structure of the waste disposal record

As such, waste related data from construction projects for four consecutive years, starting from 2012 to 2015, are selected. Since the availability of this data has legal issues alongside its significant commercial value, for preliminary evaluation, presented in this work, a small subset of 900 randomly selected projects are made available. The criteria for this selection include building types, such as residential, commercial, and educational, with projects mainly distributed all across the whole UK. This location-wise distribution of data certainly helps in generating advanced visualizations such as geographic heat map. Data from their relational database is accessed via their front-end application, which is exported to comma-separated files (*.csv*). Pointedly, by no means the data of just 900 projects can be labelled as Big Data and justified it to use the data-intensive platforms for analysis. However, this approach can be used to analyse larger sets of waste data.

Exploratory data analytics is employed to understand the latent trends in the waste data using spatial and temporal dimensions. For this purpose, variety of visualizations such as bar plot, box plot, sankey diagram, geographic heat map, word cloud, etc. are used to investigate this data. It is revealed that large proportion of construction waste is segregated under light mixed and compatible waste, which is the key hindrance for understanding the potential sources of waste generation. It is also highlighted that despite substantial waste minimisation efforts, the amounts of construction waste keeps growing, calling for the advent of waste analytics to tackle this issue from every possible perspective. The findings in this research are in line, interestingly, with the findings of the literature.

3 LITERATURE REVIEW

3.1 DESIGNING OUT CONSTRUCTION WASTE

Designing out waste is highly desirable for managing waste effectively [7,21]. This emerging concept is offering numerous opportunities of preventing construction waste. However, designers are still long way off actually practicing it during their design activities [1,13]. Specifically, the lack of awareness of potential of waste management at design stage, extra time and effort needed to achieve it, and lack of design-based tools for designing out waste, are few of such barriers to achieve it. This reveals a clear opportunity to demonstrate its applicability by developing computer-assisted automated tools that involve designers to mitigate construction waste at early stages of the design.

Designing out waste, however, is non-trivial in the sense it presents myriad intricate challenges that must be resolved for it to deliver to its promise [10,11,22]. Even answering preliminary questions about the detailed design activities that cause construction waste is hard [11,12,52]. To this end, Waste and Resource Action Plan (WRAP) has provided a basic roadmap for researchers by identifying five design principles [23], namely, *(i)* design for re-use and recovery, *(ii)* design for resource optimisation, *(iii)* design for off-site construction, *(iv)* design for resource efficient procurement, and *(v)* design for the future. There exists numerous opportunities in each of the abovementioned design principles that guarantee to change this prevailing mantra.

This study mainly explores the opportunities laid out by resource optimisation and waste efficient procurement. Some of these potential opportunities are: *(i)* design layout optimisation, *(ii)* materials selection and optimisation, *(iii)* standardisation and dimensional coordination (masonry, rebar, tiling, carpets, timber, doors, and windows) *(iv)* building level & position optimisation, *(vi)* wall lining optimisation, *(vii)* materials packaging optimisation, *(viii)* procurement route optimisation, *(ix)* and supplier selection. Mostly, these opportunities require computationally intensive optimisation techniques, which are carried out in real time to facilitate designers on best-fit design decisions to minimise construction waste [10,52]. Nevertheless, if this is achieved (easier said than done), it would be a major breakthrough in the industry and would increase productivity of different stakeholders in unprecedented ways.

To elaborate how the abovementioned opportunities are computationally intensive, the issue of standardisation and dimensional coordination (S&DC) is discussed here. The abilities to lack the designing for S&DC has long been revealed a major culprit of producing a variety of unavoidable waste [3,9,13]. Masons usually require fine-grained details to coordinate dimensions of a vast assortment of building products and materials, specified in the detailed design. Although architects do to their best in specifying design elements under the given circumstances, however, such minor details of coordinating building products and materials are still missed out in the detailed design documents. Designers often argue that sorting out such details at early design are impracticable and completely irrelevant to their designing tasks [11,13]. Accordingly, field cutting and fitting happens which results in producing piles of broken backup tiles and masonry waste.

To sort out these vexing details of S&DC for unrelated sized building products and materials, right at the design stage, is very challenging. Working out these details using automated programs, as part of preparing detailed design document alone, depends increasingly on the high-performance computation and large-scale data storage. This calls for a paradigm shift from waste intelligence to waste analytics, which brings together theories and techniques together from Designing out, Big Data, and BIM to enable the storage and analysis of massive datasets while predicting and designing out construction waste. Big Data analytics has already been successfully used in developing applications of similar complexity and nature [5,25]. Big Data thus appears to be the right tool for not only storing the required datasets but also for executing complex analytics underpinning the predicting and designing out construction waste efficiently. It is on this basis, this paper focus on the development of the Big Data architecture for construction waste analytics.

3.2 Big Data in Construction Industry

Construction industry generates huge data throughout the life cycle of a facility, which is envisioned to be systematically captured through Building Information Modelling (BIM) [5]. BIM data are 3D geometric encoded, compute intensive (graphics and Boolean computing), compressed, in diverse proprietary formats, and intertwined. These data are enriched gradually and persisted beyond the *end-of-life* of facilities. BIM files quickly get voluminous; with the design of a 3-story building model surpass 50GB in size [5]. This data in any form and shape has intrinsic value to the performance of industry. With the advent of embedded devices and sensors, industry has even started generating massive datasets during the operations and maintenance stage, eventually leading to more rich sources of Big BIM Data. This vast accumulation of data has signalled the advent of Big Data era in the construction industry [26].

The term `Big Data` was coined by Diebold [15] as the emerging availability of massive amounts of potentially relevant data. Laney [53] has identified three defining characteristics of big data—often called 3Vs of Big Data—including *(i) volume* (terabytes, petabytes of data and beyond) *(ii) variety* (heterogeneous formats like text, sensors, audio, video, graphs and more) *(iii) velocity* (contineous streams of the data). According to Jacobs [26], Big Data is the data whose size compells the community to look beyond the traditional data management and analysis approaches.

Systematically analysing Big Data to identify latent trends is at the top IT agenda of most companies today [27]. Business insights are generally buried inside these trends, which can ultimately shape the future of many companies through data-driven decision-making. Accordingly, this ability to identify, understand and react to latent trends in timely manner is a true competitive advantage of businesses in this hyper-competitive era [53]. Big Data Analytics, in this emerging ecosystem, is the real enabling toolbox for knowledge discovery.

Since this paper intends to develop Big Data architecture for construction waste analytics, various Big Data platforms, developed so far, with varied characteristics, are discussed here. Their right selection requires in-depth knowledge of critical features of these platforms. Particularly, in case of analytical applications like waste analytics, the capability of platform to adapt to increasing workload outweighs rest of the selection criteria. Big Data platforms are generally divided into vertical scaling and horizontal scaling platforms [28]. Horizontal scaling platforms distribute processing across multiple servers and scale out to increasing workload by adding multiple machines to the cluster. Whereas, vertical scaling platforms carry out computation on single server and their scaling up require processor or memory or hardware upgrade.

This paper limits discussions to horizontal scaling Big Data platforms, particularly, Apache™ Hadoop and Berkeley Big Data Analytics Stack (BDAS). This selection is mainly influenced by the data and computational requirements of construction waste analytics, which include iterative algorithms, compute-intensive tasks, and near real-time visualisation. Singh et al. [28] explained the selection of platforms in more details. Fig. 2 illustrates the most prominent platforms along with their components that are used to build the proposed waste analytics architecture. Section 5 expounds which particular technology is used to achieve the functionality precisely for developing a robust waste simulation tool.

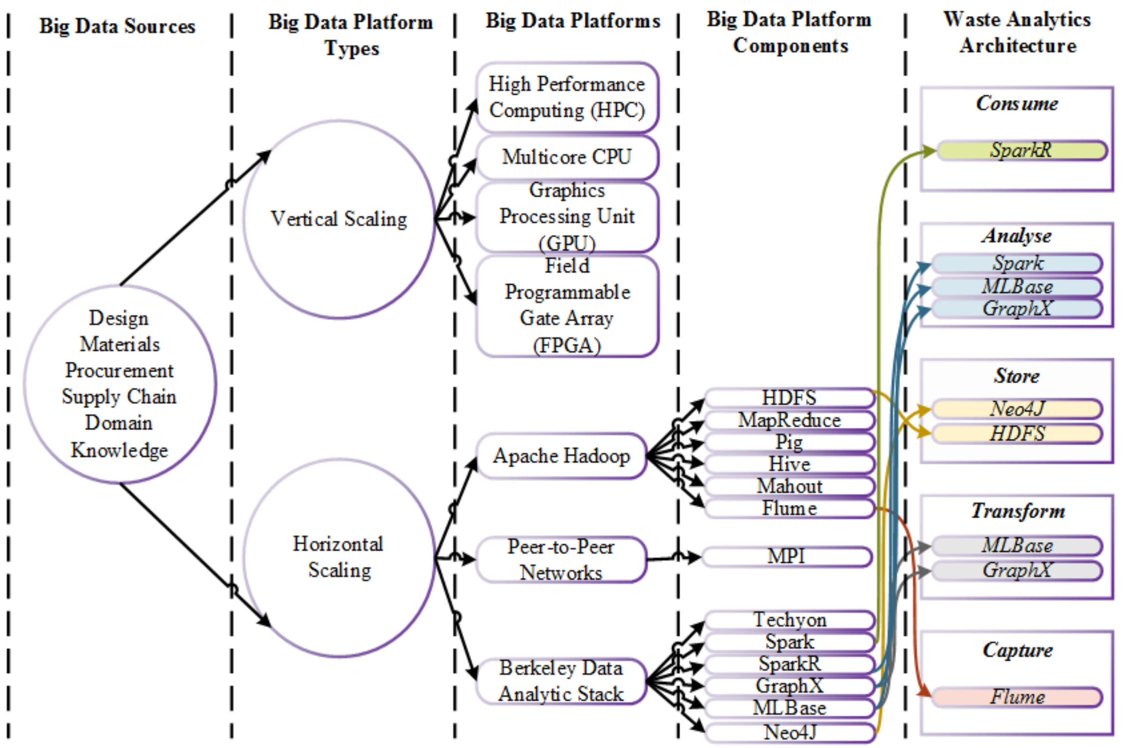

Figure 2: Big Data platform selection to develop waste analytics architecture

3.2.1 Apache™ Hadoop Project

Hadoop is a highly scalable and reliable open source framework for storing and processing data of unprecedented size. Hadoop ensures highly reliable computation at the top of commodity servers due to inherent capabilities of detecting and handling hardware failures at application layer [29,30]. Fig. 3 shows the core components of the Hadoop ecosystem. At the backstage of Hadoop are two integral components. Firstly, Hadoop distributed filesystem (HDFS) that is designed specifically to handle the storage of large datasets [31]. HDFS stores data as replicated blocks in files on the distributed file system for providing fault-tolerance and high availability. In case of block corruption or machine failures, HDFS reads replicated blocks from other machines to entertain the ongoing request seamlessly. In the case of construction waste analytics, storage of diverse datasets pertaining to design, procurement, construction, and waste disposal is the subject of distributed file storage.

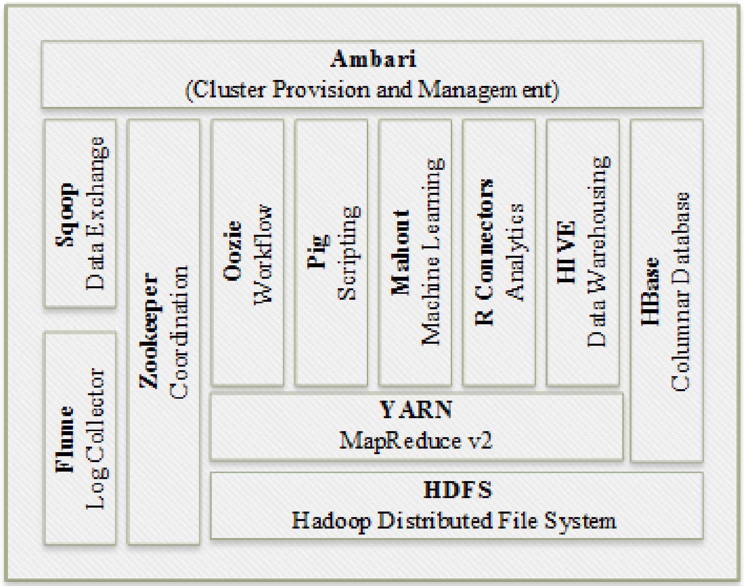

Figure 3: Critical components of the Hadoop software stack (adapted from [30])

Secondly, the MapReduce that is the programming model for large scale data processing [32]. In MapReduce, the entire computations are modelled as the map and reduce functions, which are undertaken by Mapper and Reducer processes. To contextualise, the waste estimation logic can be implemented as a series of several interlinked MapReduce jobs. During the job execution, mappers read data from the HDFS, processes it according to the user-specified *map* function, and generate the intermediate results. The framework sorts and aggregates intermediate results and assigns it to the reducers. The reducers finally process intermediate data based on *reduce* function and stores the results back to the HDFS. Several mappers and reducers run in Hadoop cluster simultaneously. The framework seamlessly takes care of intricacies related to job distribution and parallelism without any user intervention. Despite the fact that MapReduce has revolutionized Big Data applications with its simple programming model, it is not designed to undergo certain computational tasks efficiently such as iterative algorithms and reactive applications, where users are quickly responded by scanning datasets in-memory [20]. In the waste estimation, interactive and graph based processing is vital for developing real time waste simulation tool, so MapReduce is not the tool of choice for this study.

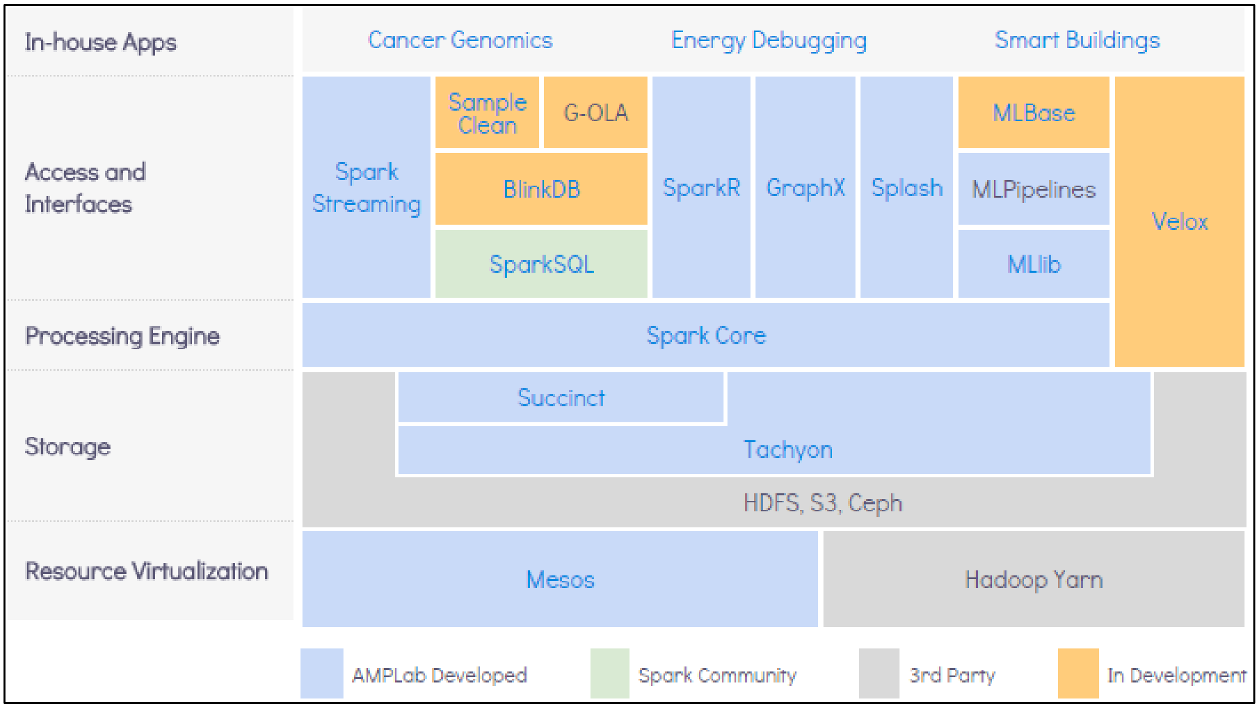

***Figure 4: Berkeley Data Analytics Stack (BDAS) and its Components** [33]*

3.2.2 Berkeley Data Analytics Stack (BDAS)

BDAS is another emerging Big Data platform (see Fig. 4). At the core of BDAS is the Spark, which is the next generation programming model for large scale data processing [20]. Spark improved MapReduce in three directions. Firstly, Spark relaxed the rigid map-then-reduce programming structure to a highly generic model. Secondly, Spark offers rich set of constructs to express computational logic more naturally and is developer-focused with APIs for writing complex tasks with just few lines of code. Lastly, it has in-memory processing capability through innovative idea of Resilient Distributed Datasets (RDD), in which data is materialized at any point in time in memory across the cluster for avoiding subsequent reloads. Consequently, Spark is reported to run 100x times faster than MapReduce where data can fit into memory, whereas, for disk-resident computations, it runs 10x times faster [28]. Additionally, Spark is ideal for iterative algorithms and reactive applications [20]. Since performance and interactive algorithms are at the core of waste simulation tool, Spark as the tool of choice is very applicable for this development. The tool can invoke Spark features through the code written in Scala or Java or Python.

Tachyon is the distributed file system of BDAS for providing the access to distributed data at memory speed. The salient features of Tachyon include: *(i) higher performance than HDFS*, since it caches frequently read data in-memory; and subsequent requests to cache data are made through memory, which reduces disk I/O and boosts performance dramatically. *(ii) It is backward compatible*; in addition to Spark, Tachyon can handle the MapReduce jobs without any modifications required to the programs. Due to the pertinent performance gains of BDAS over Hadoop, it is getting more attention recently. However, it is in its infancy with limited support and supporting tools. Whereas, Hadoop is still widely adopted and has become the de-facto framework for Big Data applications with strong support and variety of supporting data processing tools. For these reasons, this study chooses the best of the breed tools from these competing platforms and employs HDFS along with Spark to develop highly resilient waste simulation tool.

4 Proposed Construction Waste Analytics Lifecycle

Designing out waste is conceived as a huge and daunting task, which requires a well-articulated process to break it down into smaller actionable stages. Construction waste analytics (CWA) deviates from traditional waste intelligence here and asks for a rigorous process to ensure adequate provisions to support various analytical approaches thoroughly. To this end, this section discusses the CWA lifecycle stages. These stages offer a common framework for development of the proposed Big Data architecture. The lifecycle has mainly six stages, which are executed iteratively to closely suit the CWA requirements. Fig. 5 illustrates these lifecycle stages, which are discussed in the subsequent subsections.

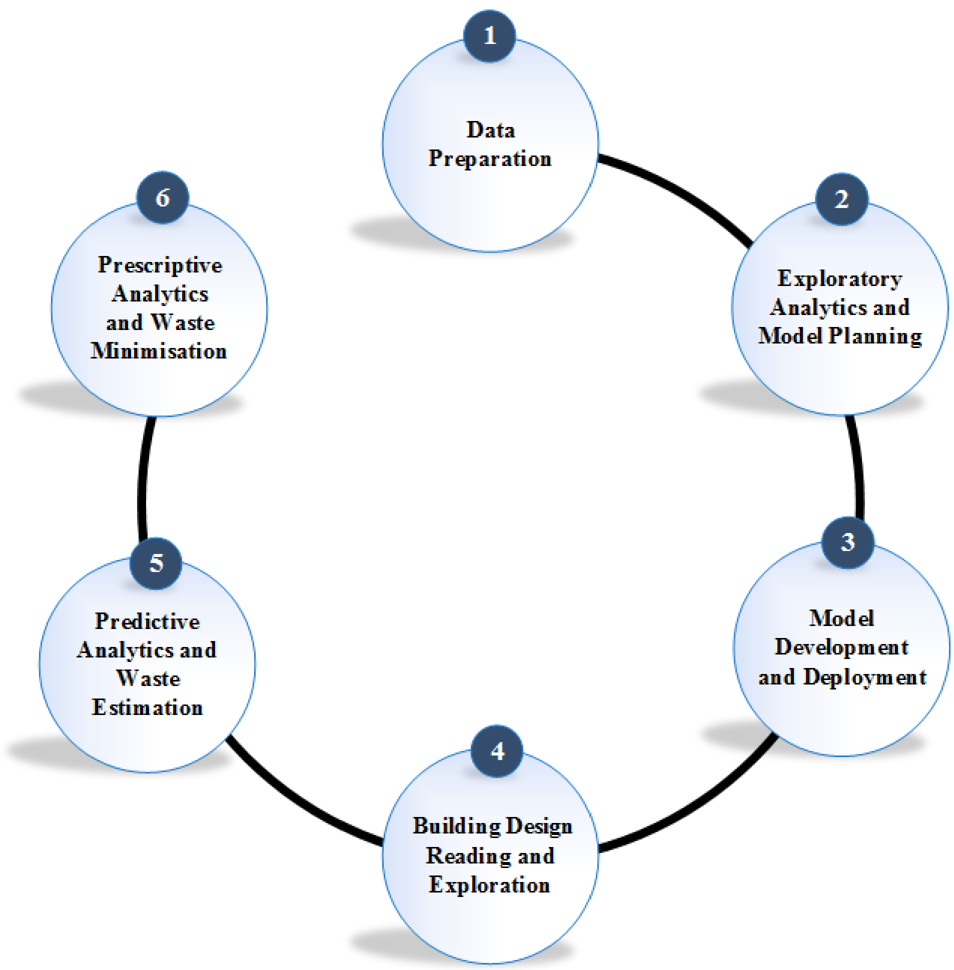

Figure 5: Construction waste analytics lifecycle stages

4.1 DATA PREPARATION

The CWA lifecycle begins with data preparation, which may include interacting with heterogeneous data sources. Data preparation is conceived as the most laborious activity [5,54]. In CWA, enough data of sufficient quality is central to performing high quality analytics. To this end, data from diverse sources are extracted, transformed, and loaded into the centralised data store. During the process, the outliers from the data are inadvertently filtered out. Since the data may be large, techniques of parallel data movement are also required, which may necessitates using the components from either of the Big Data platforms i.e., Hadoop or BDAS. Data is often analysed to get familiar with the construction waste domain. For the sake of preliminary analysis presented in this paper, the waste disposal records are supplied as the *.csv* files, which are loaded onto the Hadoop cluster. In addition, the BIM files of the respective construction projects are queried through IFC to retrieve the project-specific details such as gross floor area (GFA), material specification along with other non-trivial attributes. For this purpose, tools like Apache Flume are of immense relevance to capture the up to date versions of these datasets.

4.2 EXPLORATORY ANALYTICS AND MODEL PLANNING

In CWA, the analysis starts off with exploratory analytics and moves toward the predictive analytics. For every CWA activity, a well-defined objective is required, which dictates the selection of right analytical approaches, such as clustering, classification, regression, etc. The data exploration of waste disposal records is performed to understand the relationship among different attributes of the design and waste generation. This exploratory data analysis informs the selection of relevant variables to build a robust waste estimation model. In this paper, visualisation is used for exploratory data analysis. The overall aim of analysis at this phase is to capture the essential predictors and variables and abandon the ones, which are least relevant for model development. This selection is often iterative and requires a series of steps to identify the most useful variables for the given model. Tools like Python and SparkR could be exploited to build these analytical models.

4.3 Model Development And Deployment

In this stage, the CWA models are generated for waste estimation using robust Big Data analytics techniques and tools. Accordingly, the data are divided into training and testing sets. The CWA models are fitted to the training data and scored over the testing data. The models are evaluated for their accuracy and performance and the ones with greater predictive power are eventually selected. Oftentimes, this step involves certain optimisations; particularly the issues related to multi-collinearity are appropriately dealt. The most optimal model is eventually deployed and operationalized over the large volumes of actual waste data to predict the waste efficiency of the design. Many times the production environment may require the models to be adjusted and redeployed to cater for more realistic situations [5,55].

4.4 Building Model Reading And Exploration

In this CWA stage, the semantically rich data of building design is read for checking waste efficiency of the design. Particularly, the parameter values that are vital to execute the predictive models are extracted. In this regards, the relevant design elements (such as Walls, Doors, Windows, Roofs, etc.), material specifications, and construction strategies (like standard masonry wall build with stretcher bond type) are extracted from the design. This semantically rich data of building design enables the automated reasoning about not only the 3D geometric layout of the building design but also the relationships among various design attributes, construction materials, and construction design strategies. These relationships bring those latent trends into the focus that could be utilised to assess the waste efficiency of the given building design.

Since BIM data is generally accessed through the Industry Foundation Classes (IFC), dealing with IFC data is likely to pose some of the special data integration challenges relating to semantic heterogeneities [37,38]. To this end, these files are accessed through either IFC Toolboxes (such as ST-Developer) or IFC query languages (such as EQL, PMQL, BIMQL, etc.) or open standards (like ifcOWL ontology). A series of transformations are applied to bring the contents into application-friendly format, specifically, the contents are standardised using ifcOWL ontology [39]. Finally, this data is stored as graph-annotated formats for supporting wider computation underlying the CWA.

4.5 Predictive Analytics And Waste Estimation

Waste estimation provides the essential basis for understanding causes, types and quantities of construction waste arising from building designs [4]. Therefore, this phase employs the predictive models generated through contemporary big data analytics based approaches (applied in stage 3) to analyse building design to assess the amounts of construction waste it tends to generate. Indeed, real working horse behind this evaluation is the accuracy of waste estimation model being employed. Existing waste estimation models (developed so far) are unable to assess the true size of waste in a building design since they are only based on meta-heuristics of project characteristics such as material quantities [40,41] or gross floor area (GFA) [42,43]. However, more factors contributing to construction waste generation asides the material quantity and the GFA [4,11]. The development and deployment of a robust waste estimation model is the ultimate goal of the CWA lifecycle. Using the waste estimation model, a comprehensive waste forecast is generated which will provide the basis for design exploration and optimisation to design out construction waste.

4.6 Prescriptive Analytics And Waste Minimisation

Using the waste forecast, generated in the previous phase, this phase embarks on optimising the building design from various perspectives. Design artefacts, materials, and construction strategies are optimised based on myriad factors to reduce construction waste. The optimisation process is taken to the next level of prescriptive analytics where best course of actions (design artefacts, materials, and design strategies) are identified for designing out waste [19]. A large number of alternative optimisation plans are consequently generated and prescribed. These plans are then converted into user-friendly prescriptions, which are utilised by the designers while engaging their cognitive abilities to understand the building design for construction waste generation.

Since humans are best in identifying patterns in data [50], this phase mainly emphasizes to engage designers in understanding and proactively minimising waste. Surely, the visual data exploration techniques coupled with designer's experience enable broader understanding of building design from a large number of dimensions. As a result, a vibrant environment comes into being where design, materials, and construction strategies are investigated for their influence towards the waste generation. To this end, real-time predictions and prescriptions are generated to enable designers to get instant feedback on aspect of the design change, which certainly leads to data-driven decision-making for designing out waste. This CWA lifecycle is repeated unless a waste-efficient building design is produced. The final design is conceived to have optimised design attributes, material specifications, and construction strategies and is likely to produce less waste during the construction phase.

5 PROPOSED BIG DATA ARCHITECTURE FOR CONSTRUCTION WASTE ANALYTICS

This section discusses the proposed Big Data architecture for construction waste analytics (see Fig. 6). Various components comprising the architecture are categorised into three layers that include *(i)* Application Layer, *(ii)* Analytics Layer, and *(iii)* Storage Layer. These layers are discussed in the subsequent subsections in more details.

5.1 STORAGE LAYER

The storage layer is responsible for the data aspect of the waste simulation tool. There are two kinds of data. The historical data that contains design documents, project information, and waste disposal records. This data is uploaded once and used primarily to develop the robust waste estimation models. Another kind of data is the streaming data, which arrives whenever the designer modifies the design. Design changes are captured in the IFC files and are transferred to the storage layer of the proposed architecture by the Apache Flume, which is a reliable and distributed change listener and logging service [34]. All the unstructured data are initially stored over the Hadoop distributed file system (HDFS) and after the necessary transformations; it is loaded into the Triple store as the Resource Description Framework (RDF) graph triples. During the transformation, this data is standardised through different ontologies (like ifcOWL). The IFC toolbox (ST-Developer) are used for these transformations, where IFC objects are mapped onto respective ifcOWL constructs programmatically. ifcOWL has recently become a standard ontology to capture BIM knowledge using Web Ontology Language (OWL) constructs [35].

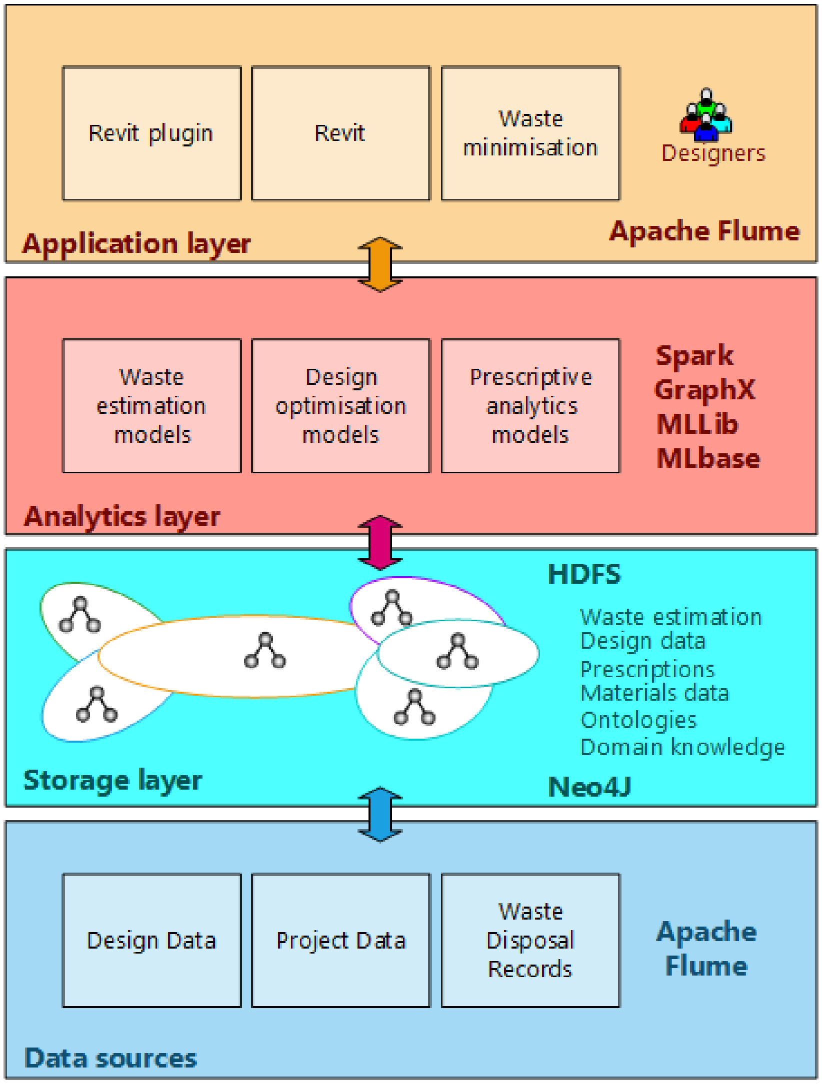

Figure 6: Big Data architecture for construction waste analytics

The Triple store is implemented using the Neo4J graph database, which can efficiently store and process massive graph data [36]. Besides these types of data, Triple store also stores the predictive models and prescriptions in Predictive Model Markup Language (PMML) format that are generated and utilised during the CWA lifecycle.

5.2 Analytics Layer

The true value waste simulation tool lies in the abilities to analyse and promptly act upon this data of unprecedented size. As discussed earlier, predicting and designing out waste is data-driven and compute-intensive task. Spark is used for these classy analytics, since it outperforms the predecessor MapReduce due to its inherent capabilities of in-memory storage and computation [20]. The analytical pipelines for waste estimation and minimisation are mainly implemented using SparkR, MLLib, and GraphX. With each iteration of analytical pipeline, the design is explored for myriad dimensions and optimised towards the waste efficiency. During the course of execution, waste efficiency of the design is instantaneously computed and disseminated to the designers, which can proactively react to mitigate the impact of design change towards the waste-efficiency. These intermediary predictions along with prescriptions are captured as RDF-annotated graphs in the Triple store.

5.3 Application Layer

Autodesk Revit API is used for the plugin development. Since Revit is widely adopted worldwide and has the robust support for intended visual analytics, the application layer of the waste simulation tool is built by exploiting its powerful API programs. The design changes are captured through IFC files, which are loaded to the HDFS through the Apache Flume and eventually loaded to the Triple store. Spark Streaming triggers the analytics pipeline to estimate the waste and suggest actionable insights to optimise the design. These insights are presented as prescriptions, which are mapped onto visual attributes of Revit by exploiting the Model-View-Controller (MVC) design pattern. The predictions and prescriptions are communicated as the Predictive Model Markup Language (PMML). Designers are provided with instant feedback to optimise the building design. Overall, this is an iterative process; it begins with every design change and lasts after few iterations by producing the design, which is optimal in terms of materials, cost, and waste efficiency.

6 RESULTS & DISCUSSIONS

The prototype of the proposed architecture is implemented by interfacing diverse technology artefacts. In this section, the proposed architecture is evaluated using exploratory data analysis and some preliminary results are provided. The goal of this evaluation is to confirm the adequacy of the proposed architectural components. However, the complete validation of this framework worth another full paper. In this context, some initial results are provided that are obtained next. Interestingly, the findings of this research support findings in literature. The future goal is to conduct a more rigorous evaluation through predictive analytics, by exploiting preliminary analysis, presented in the next subsections.

Figure 7: Word cloud to display types of construction waste

6.1 WASTE DISTRIBUTION BY CONSTRUCTION MATERIALS

Since, construction waste disposal records includes the *waste type* field, which tells about the type of material producing construction waste. Understanding distribution of construction waste by material can highlight the *top-k* materials that are generally at the higher end of producing construction waste. Bar plot—a statistical graphic tool—is employed to compare different categories within the data. In bar charts, the grouped data is visualised as bars of the chart.

Additionally, the word cloud is used to check the most prominent types of materials that appear frequently within the waste data.

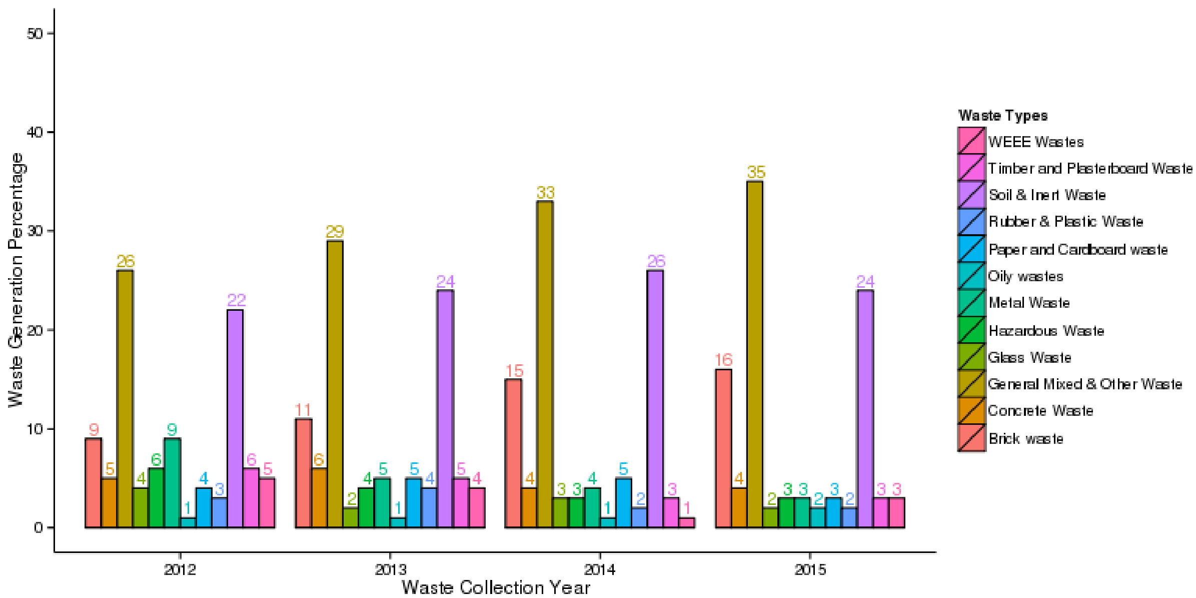

Figure 8: Bar plot to show yearly waste proportion by waste types

Through the evaluation (see Fig. 7 & Fig. 8), it is noticed that large proportion of construction waste is generally categorized as light mixed and compatible waste, mixed construction waste, and mixed inert waste. This means that either larger proportion of construction waste is not segregated as a general practice in the industry, which is due to no financial benefits associated with segregating it, or the constituents of waste are in a form of aggregates, which might have made this activity as non-trivial and unrealistic. More research needs to be done to understand the causes of this issue. In all the cases, this has revealed a major drawback of existing waste management approaches and has highlighted a serious shortcoming on their part. For effective waste minimisation, the waste categorised under these mixed and compatible waste types needs to be segregated. Incorporating these tiny details will improve the quality of waste disposal records data, which will certainly improve its usability and facilitate more sophisticated analytics. Eventually, the stakeholders (designers and waste managers) will be well informed and better guided for the sources of construction waste. This fine-grained knowledge is not only integral for developing a robust construction waste simulation tool but is also critical for the designers to use it to the best of their abilities in waste minimisation.

6.2 Waste Distribution by Waste Management Routes

Waste management hierarchy advocates the waste management routes such as reduce, reuse, recycle and dispose. The priority is given to reduce the waste in first place, however, if waste is generated, then the reuse, followed by recycle is preferred. As a last resort, the waste is disposed off to landfill which is undesirable due to its immense environmental impact. In the waste disposal data, the attributes such as *tonnes to landfill* and *tonnes recycled* capture the waste management routes with each record. The overall proportion of waste sent to various waste management routes worth further exploration. To this end, sankey graphs are employed for understanding the flow of waste based on their waste management routes. Sankey graph is the powerful tool to understand the flow of entities within a system, where weight of the flow is specified between the source and destination.

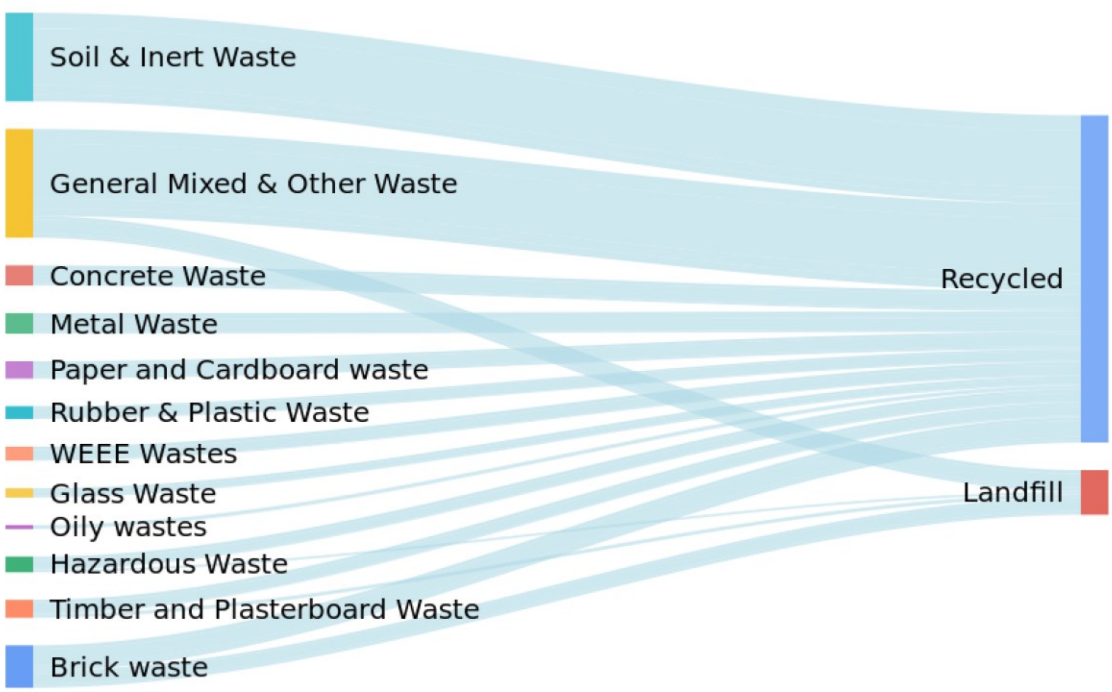

Figure 9: Sankey diagram to show waste flow by waste management routes

The *construction waste types* are mapped as the sources of graph while *waste management routes* as the destinations of graph. The amounts of waste generated are mapped onto the weights of the flows running between the sources and destinations. Fig. 9 illustrates this distribution. It is evident that the industry is quite successful in diverting the majority of construction waste going to landfill through recycling. Soil and inert waste comprise huge amounts, which is completely recycled. Whereas, some proportions of the mixed waste along with brick waste ends up in the landfill. This requires immediate attention of the research community to look into the matter. The literature revealed that field cutting due to lack of standardisation and dimensional coordination is one of the major causes of producing huge amounts of onsite offcuts [3,10]. More R&D is needed to develop such approaches that can successfully divert brick waste from going to the landfill.

Pertinently, the waste disposal data only captures the landfill and recycling routes whereas no consideration is given to the most desirable waste management routes of reducing and reusing the waste. This clearly depicts a deficiency on the part of the waste disposal records. Such details shall be captured to see how the industry is minimising construction waste in the first place (i.e., at the early design stages).

6.3 Waste Distribution by Time (Time Series Analysis

Tremendous R&D efforts have been carried out for minimising construction waste. An obvious assessment for checking the effectiveness of these R&D efforts is to observe the waste trend over time. Time series analysis, in this regard, advocates methods for analysing waste data with respect to time. It not only enables the exploratory waste analysis but also supports predictive waste analytics. In our waste data, *collection/uplift data* records the timestamp information that tells exactly the date and time, whenever waste skip is moved from client site to company premises. Successive waste measurements are typically recorded in this dataset over a time interval. Line plots are most commonly used graphics to illustrate the data for time series analysis.

As such, the waste data is summarized by the quarter (extracted from the data) to support the time series analysis. This derived data is plotted on line plot (see Fig. 9), where the *x-axis* of the chart is specified to the time dimension (i.e. quarter), and *y-axis* of the chart is specified to the amounts of waste produced for that particular quarter. The line plot clearly shows significantly sharp increase in the amounts of waste produced over the period. This confirms the fact that despite substantial R&D efforts for construction waste minimisation, the amounts of construction waste, produced by construction industry, keeps growing. This also informs about the failure of waste intelligence based approaches, and calls for the advent of waste analytics to tackle this issue from myriad possible perspectives proactively.

The existence of waste collection date in the waste disposal records could be leveraged for generating the site waste management plan, which is the subject of future research endeavour for predictive analytics.

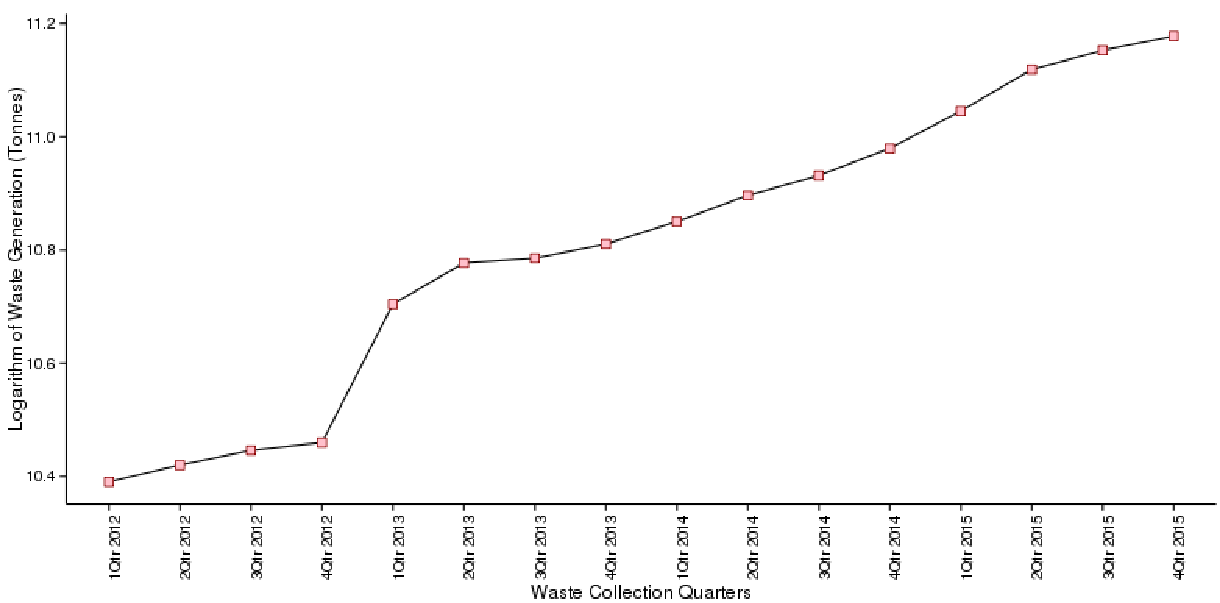

Figure 10: Line plot to display construction waste generation trend

6.4 WASTE DISTRIBUTION BY SPACE (SPATIAL ANALYSIS)

Oftentimes, the top management of a construction company is keen on seeing the locations, which are generating high volumes of construction waste. This can inform concerned managers to take proactive measures to reduce construction waste. Therefore, waste distribution by space is of immense importance in such situations, which is generally carried out using spatial analysis. Spatial analysis of waste data enables analysing construction waste over the topological, geometric, and geographic properties. In the waste data, the location information is captured in Site (including site number/ reference) field. Now for the purpose of spatial analysis, the latitude and longitude information of the given construction site location are needed. Waste data is pre-processed to get the UK postcodes for every waste record and joined with the respective latitude and longitude data obtained from the Doogal[1] . To visualise the resulting data, geographic heat map is employed for analysing the spatial distribution of construction waste. Fig. 10 shows the summary of this distribution, plotted over the geographic heat map, where size and colour of the dot informs on the volumes of waste produced at site.

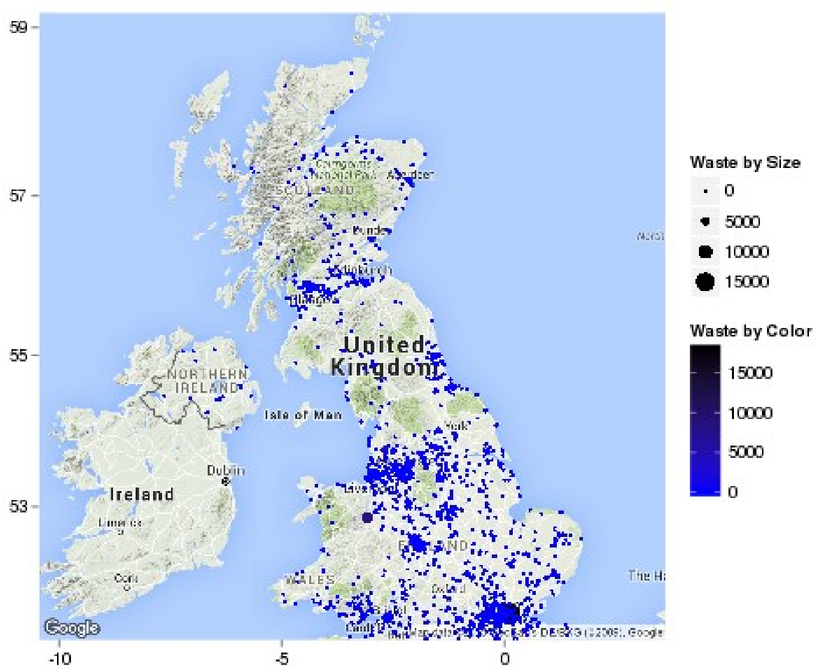

Figure 11: Geographic heat map to displaying spatial distribution of waste

The impact of location on construction waste generation also worth further exploration, which is again the focus of future research where the proposed architecture will be employed for predictive analytics.

[1] http://www.doogal.co.uk/UKPostcodes.php

7 CONCLUSIONS

In this paper, Big Data architecture for construction waste analytics is proposed. As discussed, designing out waste is the next generation of construction waste management. It is more aligned to waste analytics instead of waste intelligence. The size of the data and computational complexity underpinning the empirical tasks of designing out waste calls for the application of Big Data technologies, which are designed specifically to handle data storage and analysis at scale. Some of the chosen Big Data technologies include: *(i)* Spark, *(ii)* HDFS, and *(iii)* Neo4J. The proposed Big Data architecture is based on a well-articulated waste analytics lifecycle, which is also presented in this paper. The exploratory data analysis presented in this study revealed that despite immense R&D the construction waste generation is increasing. Similarly, majority of the construction waste is classified as mixed waste, which hinders the future innovation in construction waste minimisation.

This study is the part of an on-going research and development effort on developing a robust BIM-based construction waste simulation tool. The proposed Big Data architecture is the underlying backend of the intended construction waste simulation tool (which will be usable with existing BIM authoring tools such as Autodesk Revit and MicroStation for designing out construction waste). The BIM-based tool uses the support provided by BIM to provide waste analytics with interactive visualisation. This technological innovation is conceived to bring a vibrant environment of waste analytics, which will be certainly helpful to understand and mitigate this perennial issue of construction waste generation.

The findings in this study provide essential basis for the future work, particularly, the development of a rigorous waste estimation model using predictive analytics. Additionally, further studies are needed to validate the proposed Big Data architecture using the intended BIM based waste simulation tool. One of the critical matter to investigate is to check the accuracy of these software generated waste predictions and prescriptions, which could only be possible by deploying and using the waste simulation tool in real world setting (i.e., designer's office).

8 REFERENCES

[1] L.O. Oyedele, M. Regan, J. Von Meding, A. Ahmed, O. John, E. Amira, Reducing waste to landfill in the UK: identifying impediments and critical solutions, World J. Sci. Technol. Sustain. Dev. 10 (2013) 131–142.

[2] M. Osmani, a. Price, J. Glass, Architect and contractor attitudes to waste minimisation, Proc. ICE - Waste Resour. Manag. 159 (2006) 65–72. doi:10.1680/warm.2006.159.2.65.

[3] L.L. Ekanayake, G. Ofori, Building waste assessment score: Design-based tool, Build. Environ. 39 (2004) 851–861. doi:10.1016/j.buildenv.2004.01.007.

[4] Z. Wu, A.T.W. Yu, L. Shen, G. Liu, Quantifying construction and demolition waste: An analytical review, Waste Manag. 34 (2014) 1683–1692. doi:10.1016/j.wasman.2014.05.010.

[5] D.E. Camann, M.M. Zuniga, A.Y. Yau, L.P. Heilbrun, T.T. Walker, S. Miller, Data Science and Big Data Analytics, EMC, New York, 2011.

[6] W. Lu, H. Yuan, J. Li, J.J.L. Hao, X. Mi, Z. Ding, An empirical investigation of construction and demolition waste generation rates in Shenzhen city, South China, Waste Manag. 31 (2011) 680–687.

[7] C.S. Poon, A.T.W. Yu, L. Jaillon, Reducing building waste at construction sites in Hong Kong, Constr. Manag. Econ. 22 (2004) 461–470. doi:10.1080/0144619042000202816.

[8] J.C.P. Cheng, L.Y.H. Ma, A BIM-based system for demolition and renovation waste estimation and planning, Waste Manag. 33 (2013) 1539–1551. doi:10.1016/j.wasman.2013.01.001.

[9] S.O. Ajayi, L.O. Oyedele, M. Bilal, O.O. Akinade, H.A. Alaka, H.A. Owolabi, et al., Resources , Conservation and Recycling Waste effectiveness of the construction industry : Understanding the impediments and requisites for improvements, "Resources, Conserv. Recycl. 102 (2015) 101–112. doi:10.1016/j.resconrec.2015.06.001.

[10] M. Bilal, L.O. Oyedele, J. Qadir, K. Munir, O.O. Akinade, S.O. Ajayi, et al., Analysis of critical features and evaluation of BIM software : towards a plug-in for construction waste minimization using big data, Int. J. Sustain. Build. Technol. 7628 (2016). doi:10.1080/2093761X.2015.1116415.

[11] O.O. Akinade, L.O. Oyedele, M. Bilal, S.O. Ajayi, H.A. Owolabi, H.A. Alaka, et al., Waste minimisation through deconstruction: A BIM based Deconstructability Assessment Score (BIM-DAS), Resour. Conserv. Recycl. 105 (2015) 167–176.

doi:10.1016/j.resconrec.2015.10.018.

[12] Z. Liu, D. Student, M. Osmani, S. Lecturer, P. Demian, A.N. Baldwin, The Potential Use of Bim To Aid Construction Waste Minimisation, Current. 2005 (2011) 26–28.

[13] M. Osmani, Design waste mapping : a project life cycle approach, ICE-Waste Resour. Manag. 166 (2013) 114 – 127.

[14] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, et al., Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, New York, 2011.

[15] F.X. Diebold, “Big data” dynamic factor models for macroeconomic measuring and forecasting, Vasa. 32 (2003) 54–55.

[16] E. Siegel, Predictive Analytics: The Power to Predict who Will Click, Buy, Lie, Or Die, John Wiley & Sons, New York, 2013.

[17] K. Kim, G. Kim, D. Yoo, J. Yu, Semantic material name matching system for building energy analysis, Autom. Constr. 30 (2013) 242–255. doi:10.1016/j.autcon.2012.11.011.

[18] A. Radinger, B. Rodriguez-Castro, A. Stolz, M. Hepp, BauDataWeb, in: Proc. 9th Int. Conf. Semant. Syst. - I-SEMANTICS ’13, 2013: p. 25. doi:10.1145/2506182.2506186.

[19] B.Y.A. Basu, Five pillars of prescriptive analytics success, Exec. Edge. (2013) 8–11.

[20] S. Ryza, U. Laserson, S. Owen, J. Wills, Advanced Analytics with Spark, O’Reilly, Cambridge, 2015.

[21] K. Panos, G.I. Danai, Survey Regarding Control and Reduction of Construction Waste, PLEA2012 - 28th Conf. Oppor. Limits Needs Towar. an Environ. Responsible Archit. Lima, Perú 7-9 Novemb. 2012. (2012) 7–9.

[22] A. Keys, an Baldwin, Designing to encourage waste minimisation in the construction industry, Proc. CIBSE Natl. Conf. Dublin. (2000).

[23] D. Langdon, Designing out Waste : a design team guide for buildings, 2011.

[24] S.O. Ajayi, L.O. Oyedele, B. Ceranic, M. Gallanagh, K.O. Kadiri, Life cycle environmental performance of material specification: a BIM-enhanced comparative assessment, Int. J. Sustain. Build. Technol. Urban Dev. 6 (2015) 14–24. doi:10.1080/2093761X.2015.1006708.

[25] W. Fan, A. Bifet, Mining Big Data : Current Status , and Forecast to the Future, ACM SIGKDD Explor. Newsl. 14 (2013) 1–5. doi:10.1145/2481244.2481246.

[26] A. Jacobs, The Pathologies of Big Data, Commun. ACM. 7 (2009) 10. doi:10.1145/1563821.1563874.

[27] R. Thomas, P. McSharry, Big Data Revolution, Wiley, West Sussex, UK, 2015.

[28] D. Singh, C.K. Reddy, A survey on platforms for big data analytics, J. Big Data. 2 (2015) 8. doi:10.1186/s40537-014-0008-6.

[29] M. Stonebraker, The Case for Shared Nothing 2 . A SIMPLE ANALYSIS, Contract. 9 (1986) 1–5.

[30] T. White, Hadoop. The definitive guide., 3rd ed., O'Reilly, Sebastopol, Canada, 2012.

[31] S. Ghemawat, H. Gobioff, S.-T. Leung, The Google file system, ACM SIGOPS Oper. Syst. Rev. 37 (2003) 29.

[32] J. Dean, S. Ghemawat, MapReduce : Simplified Data Processing on Large Clusters, Commun. ACM. 51 (2008) 1–13. doi:10.1145/1327452.1327492.

[33] A. Berkeley, Berkeley Data Analytics Stack (BDAS), (2015). https://amplab.cs.berkeley.edu/software/ (accessed April 1, 2015).

[34] Apache Software Foundation, Apache Flume, (2015). https://flume.apache.org/ (accessed July 1, 2015).

[35] J. Beetz, J. van Leeuwen, B. de Vries, IfcOWL: A case of transforming EXPRESS schemas into ontologies, Artif. Intell. Eng. Des. Anal. Manuf. 23 (2009) 89. doi:10.1017/S0890060409000122.

[36] I. Robinson, J. Webber, E. Eifrem, Graph Databases, 2nd ed., O'Reilly, Sebastopol, Canada, 2015.

[37] A.Y. Halevy, N. Ashish, D. Bitton, M. Carey, D. Draper, J. Pollock, et al., Enterprise information integration: successes, challenges and controversies, Proc. 2005 ACM SIGMOD Int. Conf. Manag. Data SIGMOD 05. (2005) 778–787. doi:10.1145/1066157.1066246.

[38] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. Ullman, et al., The TSIMMIS approach to mediation: Data models and languages, J. Intell. Inf. Syst. 8 (1997) 117–132. doi:10.1023/A:1008683107812.

[39] S. Chaudhuri, U. Dayal, An overview of data warehousing and OLAP technology, ACM SIGMOD Rec. 26 (1997) 65–74. doi:10.1145/248603.248616.

[40] I. Martínez Lage, F. Martínez Abella, C.V. Herrero, J.L.P. Ordóñez, Estimation of the annual production and composition of C&D Debris in Galicia (Spain), Waste Manag. 30 (2010) 636–645. doi:10.1016/j.wasman.2009.11.016.

[41] J. Shi, Y. Xu, Estimation and forecasting of concrete debris amount in China, Resour. Conserv. Recycl. 49 (2006) 147–158. doi:10.1016/j.resconrec.2006.03.011.

[42] D. Fatta, a. Papadopoulos, E. Avramikos, E. Sgourou, K. Moustakas, F. Kourmoussis, et al., Generation and management of construction and demolition waste in Greece - An existing challenge, Resour. Conserv. Recycl. 40 (2003) 81–91. doi:10.1016/S0921-3449(03)00035-1.

[43] J. Solís-Guzmán, M. Marrero, M.V. Montes-Delgado, A. Ramírez-de-Arellano, A Spanish model for quantification and management of construction waste, Waste Manag. 29 (2009) 2542–2548. doi:10.1016/j.wasman.2009.05.009.

[44] M.J. Shepperd, G. Kadoda, Comparing software prediction techniques using simulation, IEEE Trans. Softw. Eng. 27 (2001) 1014–1022.

[45] C. Mair, G. Kadoda, M. Le, K. Phalp, C. Scho, M. Shepperd, et al., An investigation of machine learning based prediction systems, J. Syst. Softw. 53 (2000) 23–29. doi:10.1016/S0164-1212(00)00005-4.

[46] S.K. Card, J.D. Mackinlay, B. Shneiderman, Readings in information visualization: using vision to think, Morgan Kaufmann, San Francisco, CA, USA, 1999.

[47] C. Chen, Information visualization: Beyond the horizon, Springer Science & Business, New York, 2006.

[48] C.G. Healey, J.T. Enns, Perception and painting: A search for effective, engaging visualizations, IEEE Comput. Graph. Appl. 22 (2002) 10–15. doi:10.1109/38.988741.

[49] D. a Keim, Visual Exploration of Large Data Sets, Commun. ACM. 44 (2009) 39–44.

[50] D. a Keim, F. Mansmann, J. Thomas, D. Keim, Visual Analytics : How Much Visualization and How Much Analytics ?, ACM SIGKDD Explor. Newsl. 11 (2010) 5–8. doi:10.1145/1809400.1809403.

[51] W. Lu, X. Chen, D.C.W. Ho, H. Wang, Analysis of the construction waste management performance in Hong Kong: the public and private sectors compared using big data, J.

Clean. Prod. (2015) 1–11. doi:10.1016/j.jclepro.2015.06.106.

[52] Z. Wu, H. Fan, G. Liu, Forecasting Construction and Demolition Waste Using Gene Expression Programming, 29 (2008) 1–8. doi:10.1061/(ASCE)CP.1943-5487.0000362.

[53] D. Laney, 3D Data Management: Controlling Data Volume, Velocity, and Variety, Appl. Deliv. Strateg. 949 (2001) 4.

[54] U. Fayyad, The KDD Process for Extracting Useful Knowledge from Volumes of Data, 39 (1996) 27–34. doi:10.1145/240455.240464.

[55] X. Wu, X. Zhu, G.-Q. Wu, W. Ding, Data mining with big data, IEEE Trans. Knowl. Data Eng. 26 (2014) 97–107. doi:10.1109/TKDE.2013.109.