

1 **Big Data Architecture for Construction Waste Analytics (CWA): A**

2 **Conceptual Framework**

3 **ABSTRACT**

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

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

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

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

3 **1 INTRODUCTION**

4 1.1 CONSTRUCTION WASTE—AN OVERVIEW

5 Rapid urbanisation and the appetite to build national infrastructure has escalated the construction
6 activities globally. Notwithstanding the benefits, the adverse impact of construction activities on
7 environment has serious implications worldwide [1]. Construction industry is noted for
8 consuming bulk of rare natural resources and producing hefty amounts of construction and
9 demolition (C&D) waste [2]. The construction industry is the UK’s largest consumer of natural
10 resources, using over 400 million tonnes of material per annum and is responsible for producing
11 120 million tonnes of construction, demolition and excavation waste yearly—around more than
12 one third of all waste arising in the UK [3]. With the rising cost of construction projects and the
13 growing environmental concerns, the construction industry is under immense pressure from
14 government and environmental agencies to minimise construction waste and adopt more
15 sustainable practices.

16 1.2 WASTE INTELLIGENCE—CURRENT STATE OF WASTE MANAGEMENT

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

1 1.3 WASTE ANALYTICS—NEXT GENERATION OF WASTE MANAGEMENT

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

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

22 1.4 BIG DATA FOR WASTE ANALYTICS

23 Big Data is the emerging capability to store and analyse large volumes of data scalably and
24 reliably using a cluster of commodity servers [14,15]. There is tremendous interest in utilising
25 the information in Big Data for analytics, not only to understand latent trends (exploratory
26 analytics and descriptive analytics), but also for predictive & prescriptive analytics to forecast
27 and shape future events [16]. Mostly, the advanced analytical techniques for Waste Analytics
28 are supported by the Big Data technologies. Therefore, Big Data driven Waste Analytics is the

1 next emerging trend that offers unprecedented opportunities to minimise construction waste
2 through design. This synergistically integration of technologies (Big Data, Designing out Waste,
3 and BIM) is a real game changer and promises the development of a resilient BIM based
4 construction waste simulation tool to facilitate the designers in making right decisions to avoid
5 construction waste in future construction projects.

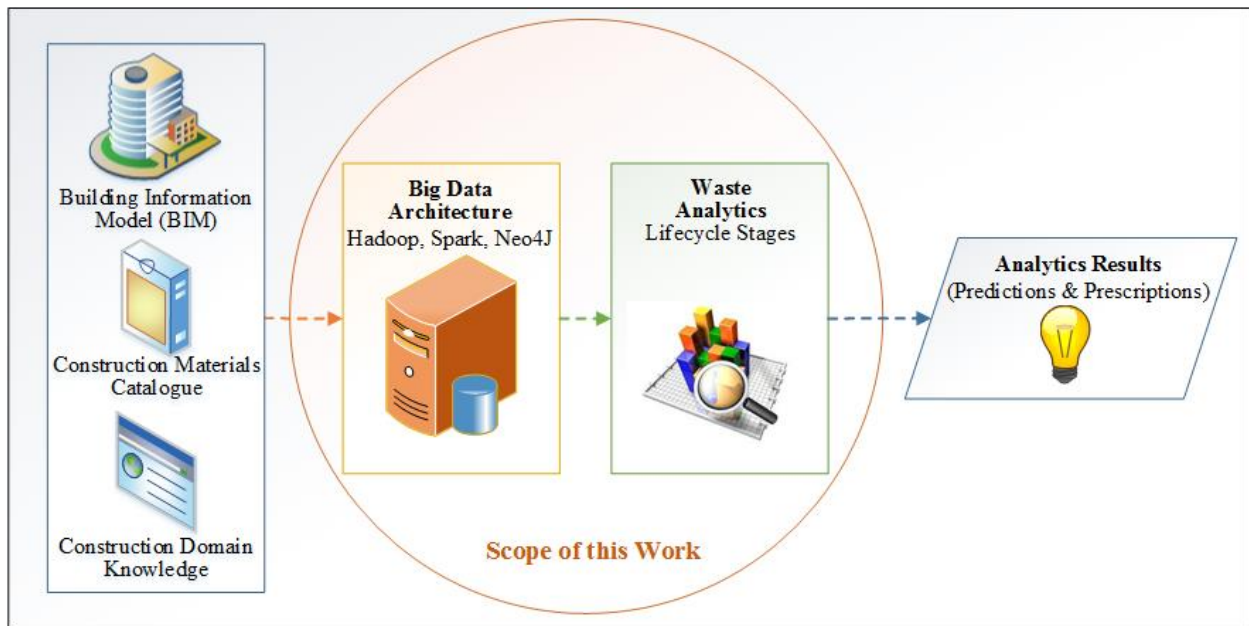
6 Waste Analytics depends increasingly on high-performance computation and large-scale data
7 storage. It requires large number of diverse datasets pertaining to building design, material
8 properties, and construction domain knowledge for successfully carrying out the underpinning
9 analytical tasks.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

14 1.5 JUSTIFICATION FOR RESEARCH AND CONTRIBUTION OF THIS PAPER

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

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

10 2 METHODOLOGY AND FOCUS OF THE PAPER



11

12

Figure 1: Focus of this paper

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

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

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

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

25 2.1 ANALYSIS & PRELIMINARY RESULTS

26 The proposed architecture is further ensured and validated using the objective data, taken from
27 the top waste management company in the UK, who offers broad range of recycling and waste
28 management services, including skip & container hire, onsite waste segregation, site waste

1 management services, including skip & container hire, onsite waste segregation, site waste
 2 management plans (SWMP), plasterboard recycling, etc. The company uses relational database
 3 to store the waste related data from a large number of other construction companies. The data is
 4 stored as individual waste movements from site, by project, with the major details of
 5 the waste transfer note being recorded. Every time it transports the waste, a digital record is
 6 created in their database. Full details about the name of fields for which the values are captured
 7 in these records are shown the Listing 1.

8 **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]	√		

1

- 2 – Operation Code
- 3 – Reference No
- 4 – Business Stream
- 5 – Hub
- 6 – Project Name
- 7 – Site (including site number/ reference)
- 8 – Region
- 9 – Purchase Order No
- 10 – Ticket number (Unique identifier)
- 11 – Collection/uplift date
- 12 – Container size and type
- 13 – Movement type
- 14 – Classification
- 15 – Waste Type
- 16 – Waste Collected in Tonnes

- 1 – Division
- 2 – Hazardous/ Non-Hazardous
- 3 – Waste Transfer note number
- 4 – Hazardous/ Special waste consignment note number
- 5 – Hazardous waste premises code
- 6 – Waste Carrier
- 7 – Waste Carrier license
- 8 – Carrier expiry date
- 9 – Carrier Checked with EA/ SEPA and in compliance
- 10 – Total Cost
- 11 – Cost per Tonne
- 12 – Waste Management facility type
- 13 – Waste Management facility permit or exemption number
- 14 – Type and quantity in compliance with permit or exemption and checked with EA/ SEPA
- 15 – Description of Waste Management Innovation
- 16 – Tonnes to Landfill
- 17 – Tonnes to Other disposal location (not landfill)
- 18 – Tonnes Recycled/ recovered
- 19 – Tonnes to Other Recovery
- 20 – % to Landfill
- 21 – % to Landfill
- 22 – % Recycled

23 ***Listing 1: Structure of the waste disposal record***

24 As such, waste related data from construction projects for four consecutive years, starting from
25 2012 to 2015, are selected. Since the availability of this data has legal issues alongside its
26 significant commercial value, for preliminary evaluation, presented in this work, a small subset
27 of 900 randomly selected projects are made available. The criteria for this selection include
28 building types, such as residential, commercial, and educational, with projects mainly distributed
29 all across the whole UK. This location-wise distribution of data certainly helps in generating
30 advanced visualizations such as geographic heat map. Data from their relational database is
31 accessed via their front-end application, which is exported to comma-separated files (.csv).
32 Pointedly, by no means the data of just 900 projects can be labelled as Big Data and justified it to
33 use the data-intensive platforms for analysis. However, this approach can be used to analyse
34 larger sets of waste data.

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

10 **3 LITERATURE REVIEW**

11 3.1 DESIGNING OUT CONSTRUCTION WASTE

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

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

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

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

22 To sort out these vexing details of S&DC for unrelated sized building products and materials,
23 right at the design stage, is very challenging. Working out these details using automated
24 programs, as part of preparing detailed design document alone, depends increasingly on the high-
25 performance computation and large-scale data storage. This calls for a paradigm shift from waste
26 intelligence to waste analytics, which brings together theories and techniques together from
27 Designing out, Big Data, and BIM to enable the storage and analysis of massive datasets while
28 predicting and designing out construction waste. Big Data analytics has already been
29 successfully used in developing applications of similar complexity and nature [5,25]. Big Data

1 thus appears to be the right tool for not only storing the required datasets but also for executing
2 complex analytics underpinning the predicting and designing out construction waste efficiently.
3 It is on this basis, this paper focus on the development of the Big Data architecture for
4 construction waste analytics.

5 3.2 BIG DATA IN CONSTRUCTION INDUSTRY

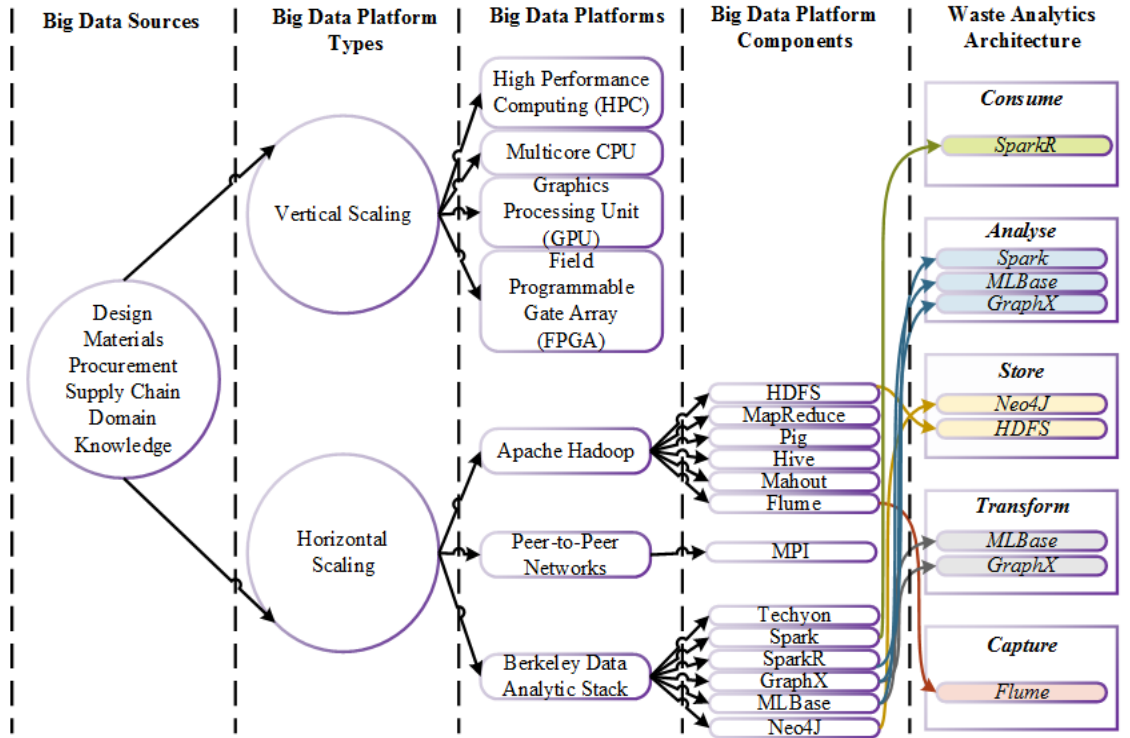
6 Construction industry generates huge data throughout the life cycle of a facility, which is
7 envisioned to be systematically captured through Building Information Modelling (BIM) [5].
8 BIM data are 3D geometric encoded, compute intensive (graphics and Boolean computing),
9 compressed, in diverse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

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

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

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

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



1

2

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

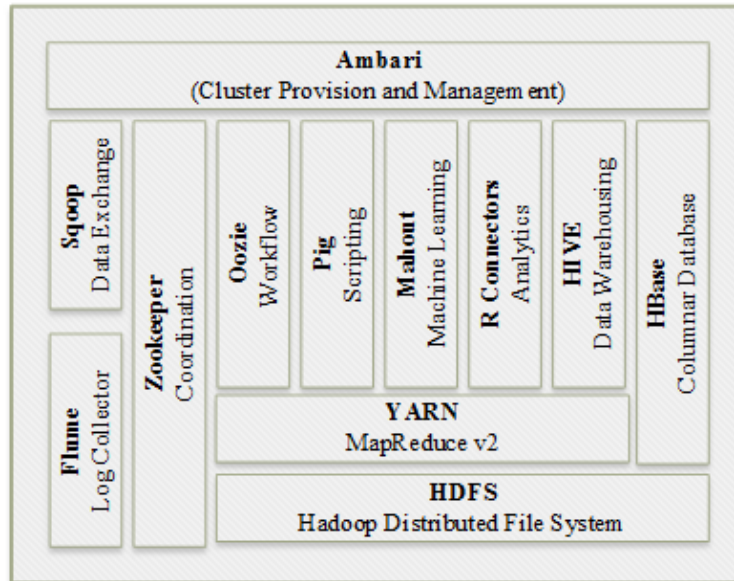
3

3.2.1 APACHE™ HADOOP PROJECT

4

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.

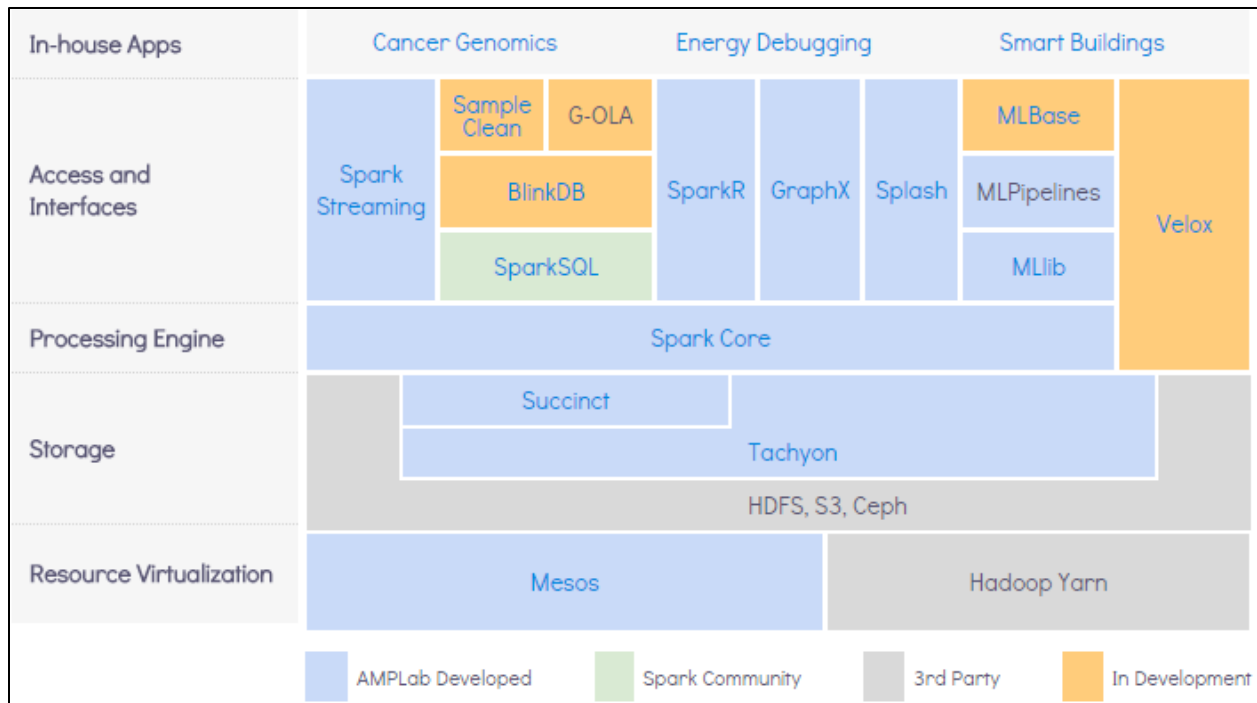
14



1

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

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



1

2

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

3

3.2.2 BERKELEY DATA ANALYTICS STACK (BDAS)

4

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

15

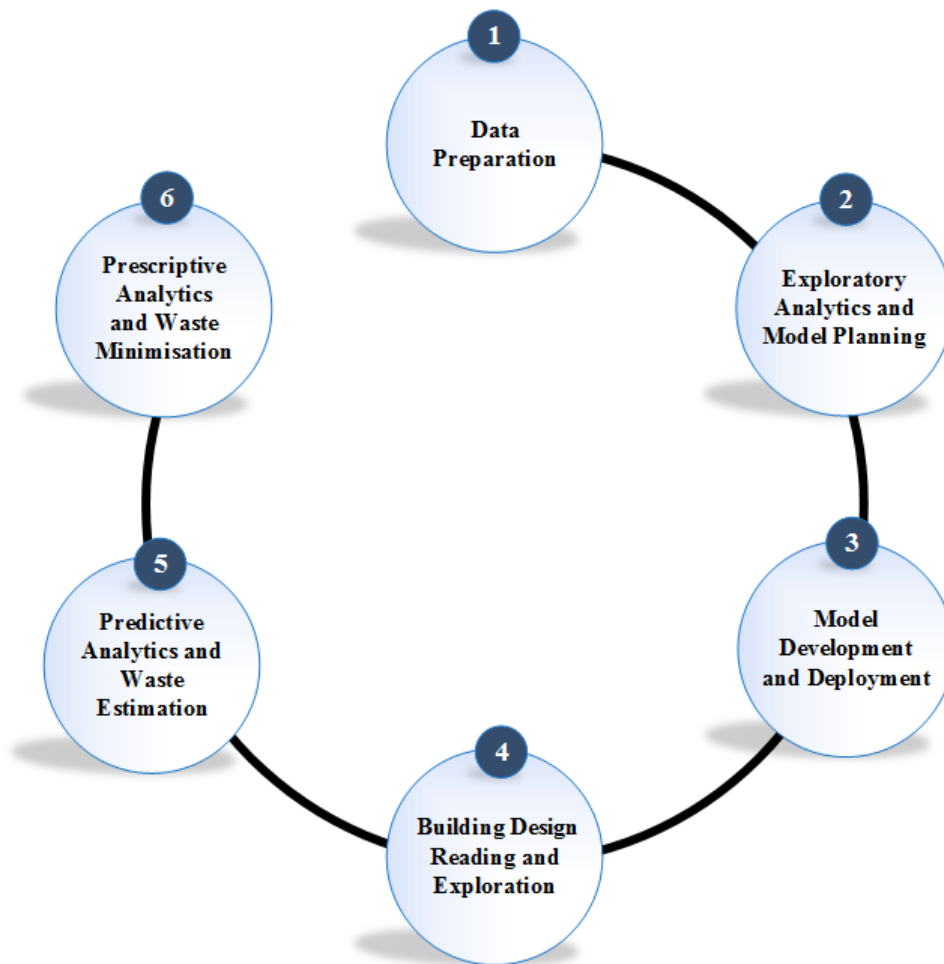
1 development. The tool can invoke Spark features through the code written in Scala or Java or
2 Python.

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

15

1 **4 PROPOSED CONSTRUCTION WASTE ANALYTICS LIFECYCLE**

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



10

11

Figure 5: Construction waste analytics lifecycle stages

1 4.1 DATA PREPARATION

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

15 4.2 EXPLORATORY ANALYTICS AND MODEL PLANNING

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

1 4.3 MODEL DEVELOPMENT AND DEPLOYMENT

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

11 4.4 BUILDING MODEL READING AND EXPLORATION

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

22 Since BIM data is generally accessed through the Industry Foundation Classes (IFC), dealing
23 with IFC data is likely to pose some of the special data integration challenges relating to
24 semantic heterogeneities [37,38]. To this end, these files are accessed through either IFC
25 Toolboxes (such as ST-Developer) or IFC query languages (such as EQL, PMQL, BIMQL, etc.)
26 or open standards (like ifcOWL ontology). A series of transformations are applied to bring the
27 contents into application-friendly format, specifically, the contents are standardised using

1 ifcOWL ontology [39]. Finally, this data is stored as graph-annotated formats for supporting
2 wider computation underlying the CWA.

3 4.5 PREDICTIVE ANALYTICS AND WASTE ESTIMATION

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

17 4.6 PRESCRIPTIVE ANALYTICS AND WASTE MINIMISATION

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

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

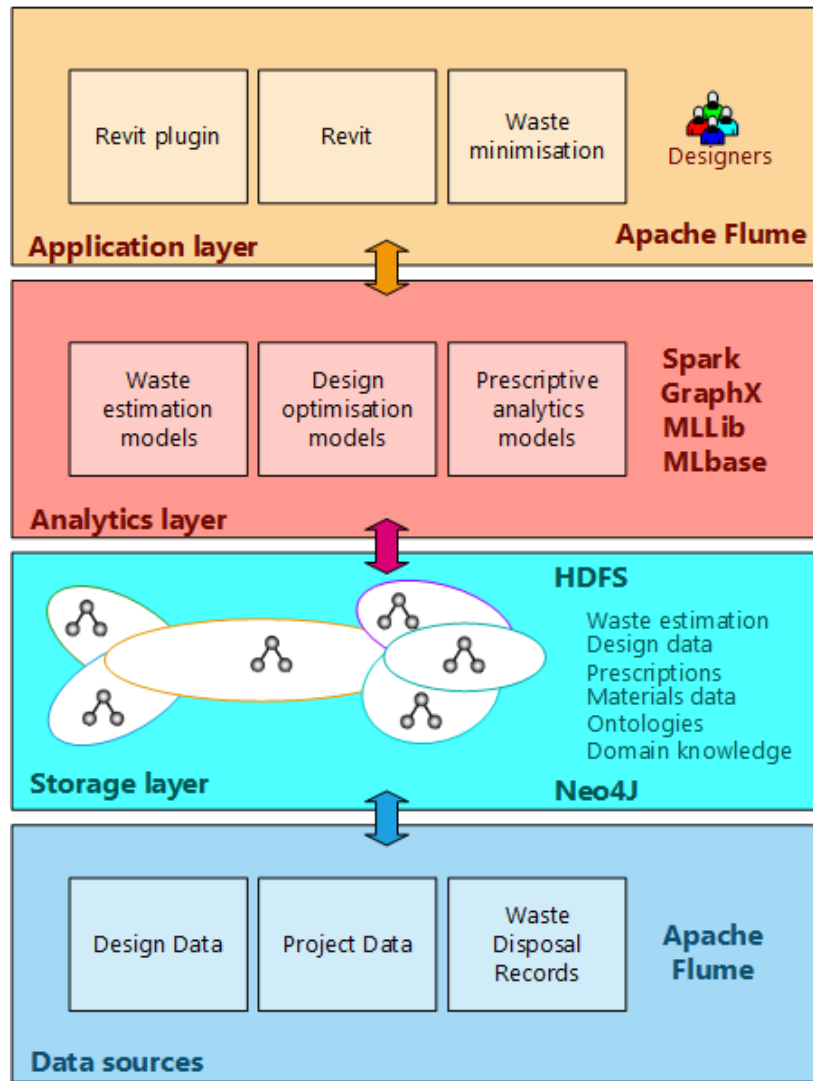
12 **5 PROPOSED BIG DATA ARCHITECTURE FOR CONSTRUCTION WASTE ANALYTICS**

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

17 **5.1 STORAGE LAYER**

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

1 respective ifcOWL constructs programmatically. ifcOWL has recently become a standard
 2 ontology to capture BIM knowledge using Web Ontology Language (OWL) constructs [35].



3

4

Figure 6: Big Data architecture for construction waste analytics

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

1 5.2 ANALYTICS LAYER

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

13 5.3 APPLICATION LAYER

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

26

1 **6 RESULTS & DISCUSSIONS**

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



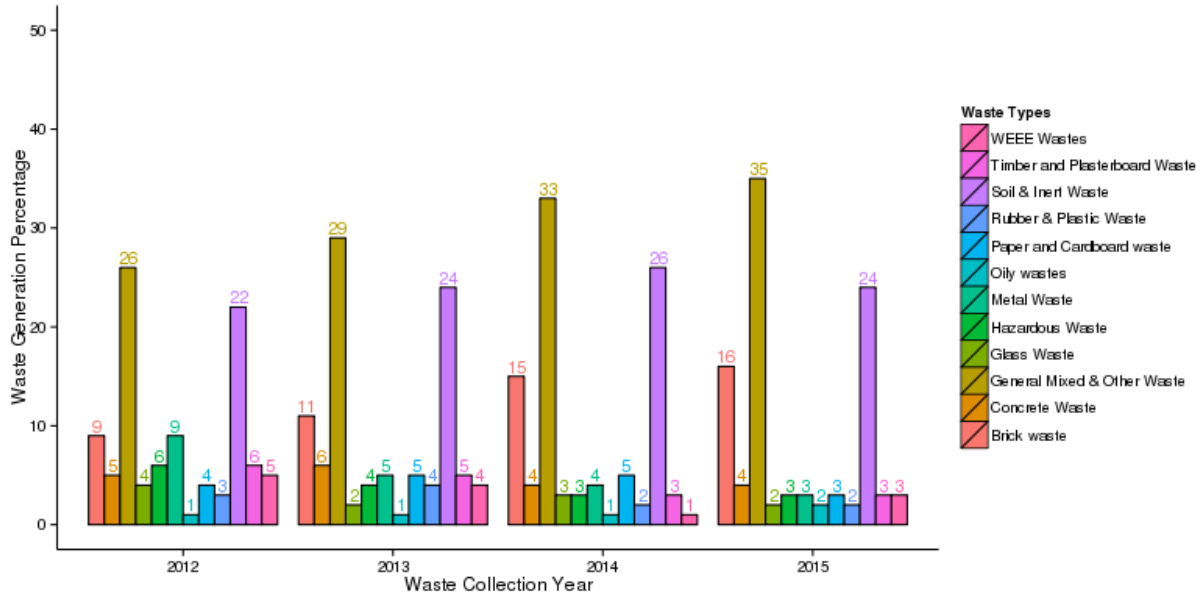
10

11 *Figure 7: Word cloud to display types of construction waste*

12 **6.1 WASTE DISTRIBUTION BY CONSTRUCTION MATERIALS**

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

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



3

4

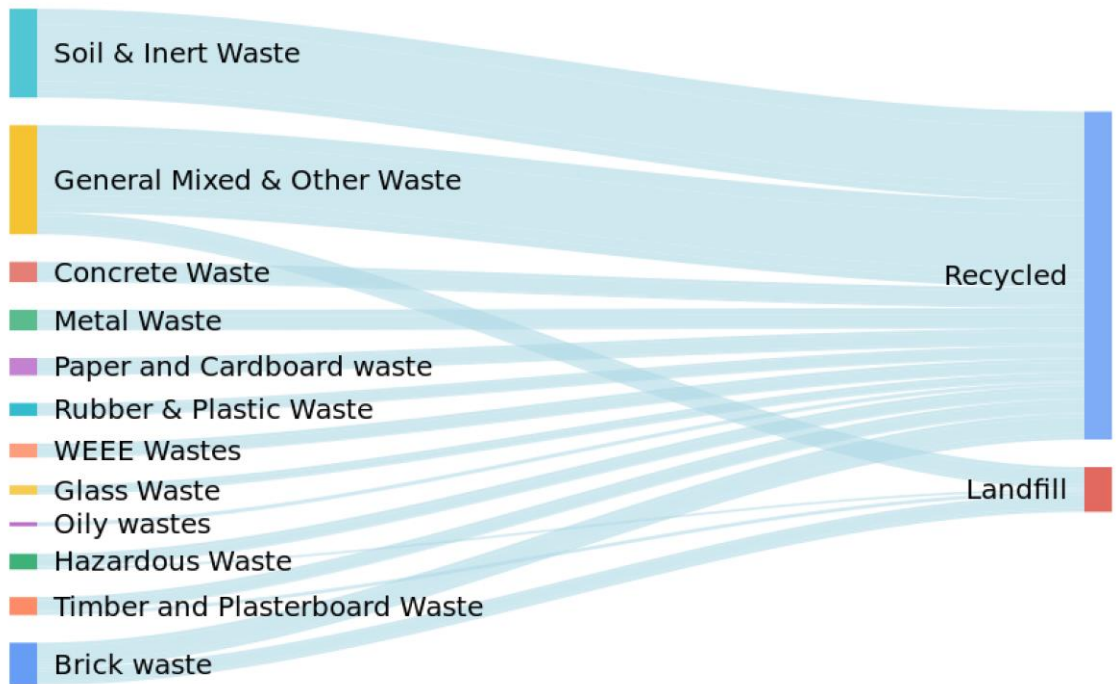
Figure 8: Bar plot to show yearly waste proportion by waste types

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

1 integral for developing a robust construction waste simulation tool but is also critical for the
2 designers to use it to the best of their abilities in waste minimisation.

3 6.2 WASTE DISTRIBUTION BY WASTE MANAGEMENT ROUTES

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



14

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

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

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

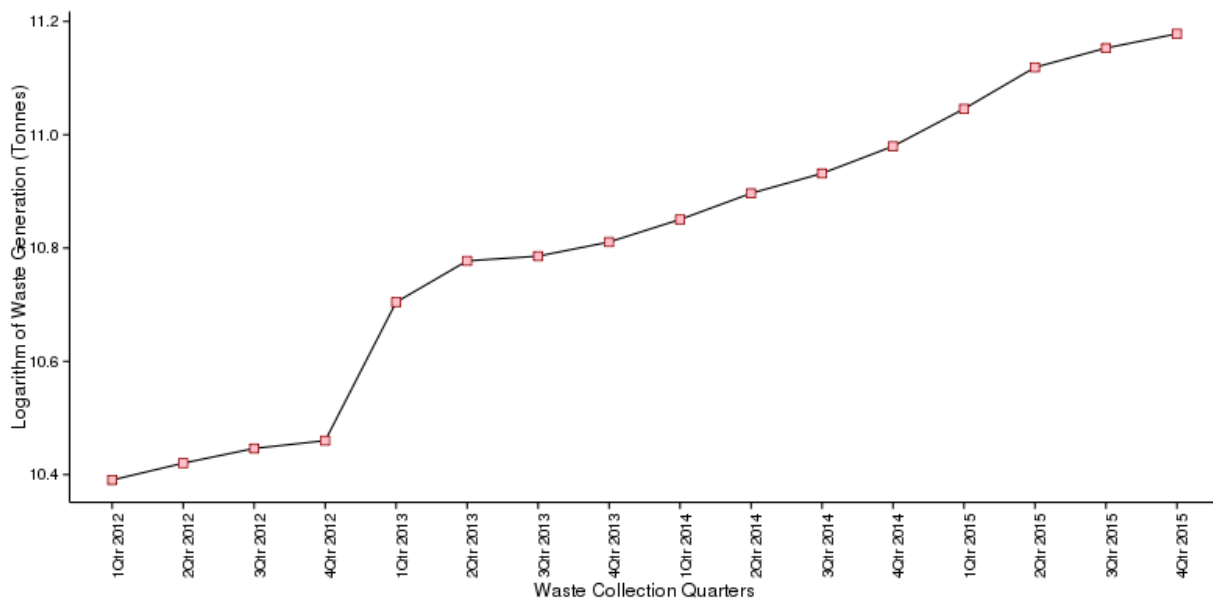
17 6.3 WASTE DISTRIBUTION BY TIME (TIME SERIES ANALYSIS)

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

27 As such, the waste data is summarized by the quarter (extracted from the data) to support the
28 time series analysis. This derived data is plotted on line plot (see Fig. 9), where the *x-axis* of the

1 chart is specified to the time dimension (i.e. quarter), and *y-axis* of the chart is specified to the
2 amounts of waste produced for that particular quarter. The line plot clearly shows significantly
3 sharp increase in the amounts of waste produced over the period. This confirms the fact that
4 despite substantial R&D efforts for construction waste minimisation, the amounts of construction
5 waste, produced by construction industry, keeps growing. This also informs about the failure of
6 waste intelligence based approaches, and calls for the advent of waste analytics to tackle this
7 issue from myriad possible perspectives proactively.

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



11

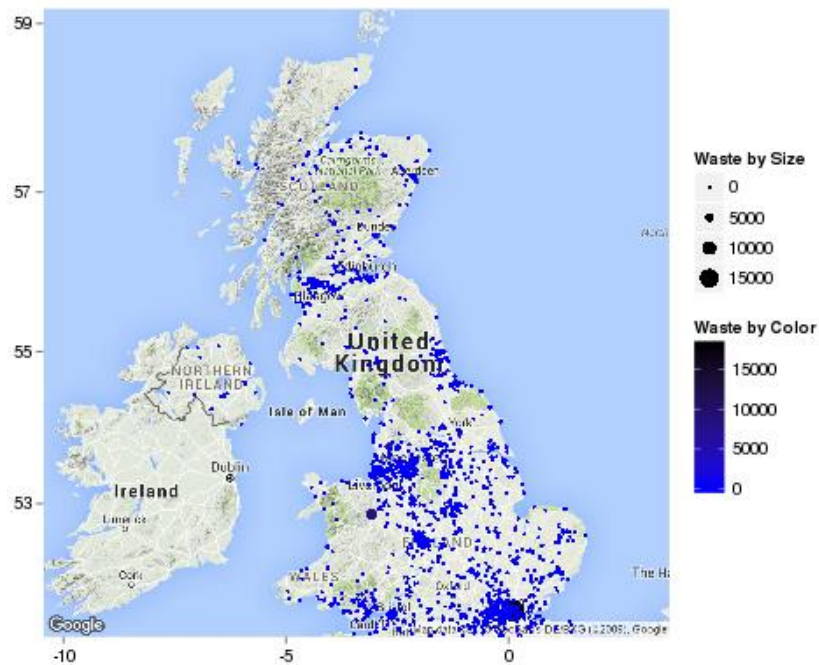
12

Figure 10: Line plot to display construction waste generation trend

13 6.4 WASTE DISTRIBUTION BY SPACE (SPATIAL ANALYSIS)

14 Oftentimes, the top management of a construction company is keen on seeing the locations,
15 which are generating high volumes of construction waste. This can inform concerned managers
16 to take proactive measures to reduce construction waste. Therefore, waste distribution by space

1 is of immense importance in such situations, which is generally carried out using spatial analysis.
2 Spatial analysis of waste data enables analysing construction waste over the topological,
3 geometric, and geographic properties. In the waste data, the location information is captured in
4 Site (including site number/ reference) field. Now for the purpose of spatial analysis, the latitude
5 and longitude information of the given construction site location are needed. Waste data is pre-
6 processed to get the UK postcodes for every waste record and joined with the respective latitude
7 and longitude data obtained from the Doogal¹. To visualise the resulting data, geographic heat
8 map is employed for analysing the spatial distribution of construction waste. Fig. 10 shows the
9 summary of this distribution, plotted over the geographic heat map, where size and colour of the
10 dot informs on the volumes of waste produced at site.



11

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

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

¹ <http://www.doogal.co.uk/UKPostcodes.php>

1 7 CONCLUSIONS

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

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

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

27

1 8 REFERENCES

- 2 [1] L.O. Oyedele, M. Regan, J. Von Meding, A. Ahmed, O. John, E. Amira, Reducing waste
3 to landfill in the UK: identifying impediments and critical solutions, *World J. Sci.*
4 *Technol. Sustain. Dev.* 10 (2013) 131–142.
- 5 [2] M. Osmani, a. Price, J. Glass, Architect and contractor attitudes to waste minimisation,
6 *Proc. ICE - Waste Resour. Manag.* 159 (2006) 65–72. doi:10.1680/warm.2006.159.2.65.
- 7 [3] L.L. Ekanayake, G. Ofori, Building waste assessment score: Design-based tool, *Build.*
8 *Environ.* 39 (2004) 851–861. doi:10.1016/j.buildenv.2004.01.007.
- 9 [4] Z. Wu, A.T.W. Yu, L. Shen, G. Liu, Quantifying construction and demolition waste: An
10 analytical review, *Waste Manag.* 34 (2014) 1683–1692.
11 doi:10.1016/j.wasman.2014.05.010.
- 12 [5] D.E. Camann, M.M. Zuniga, A.Y. Yau, L.P. Heilbrun, T.T. Walker, S. Miller, *Data*
13 *Science and Big Data Analytics*, EMC, New York, 2011.
- 14 [6] W. Lu, H. Yuan, J. Li, J.J.L. Hao, X. Mi, Z. Ding, An empirical investigation of
15 construction and demolition waste generation rates in Shenzhen city, South China, *Waste*
16 *Manag.* 31 (2011) 680–687.
- 17 [7] C.S. Poon, A.T.W. Yu, L. Jaillon, Reducing building waste at construction sites in Hong
18 Kong, *Constr. Manag. Econ.* 22 (2004) 461–470. doi:10.1080/0144619042000202816.
- 19 [8] J.C.P. Cheng, L.Y.H. Ma, A BIM-based system for demolition and renovation waste
20 estimation and planning, *Waste Manag.* 33 (2013) 1539–1551.
21 doi:10.1016/j.wasman.2013.01.001.
- 22 [9] S.O. Ajayi, L.O. Oyedele, M. Bilal, O.O. Akinade, H.A. Alaka, H.A. Owolabi, et al.,
23 *Resources , Conservation and Recycling Waste effectiveness of the construction industry :*
24 *Understanding the impediments and requisites for improvements*, "Resources, Conserv.
25 *Recycl.* 102 (2015) 101–112. doi:10.1016/j.resconrec.2015.06.001.
- 26 [10] M. Bilal, L.O. Oyedele, J. Qadir, K. Munir, O.O. Akinade, S.O. Ajayi, et al., Analysis of
27 critical features and evaluation of BIM software : towards a plug-in for construction waste
28 minimization using big data, *Int. J. Sustain. Build. Technol.* 7628 (2016).
29 doi:10.1080/2093761X.2015.1116415.
- 30 [11] O.O. Akinade, L.O. Oyedele, M. Bilal, S.O. Ajayi, H.A. Owolabi, H.A. Alaka, et al.,
31 *Waste minimisation through deconstruction: A BIM based Deconstructability Assessment*
32 *Score (BIM-DAS)*, *Resour. Conserv. Recycl.* 105 (2015) 167–176.

- 1 doi:10.1016/j.resconrec.2015.10.018.
- 2 [12] Z. Liu, D. Student, M. Osmani, S. Lecturer, P. Demian, A.N. Baldwin, The Potential Use
3 of Bim To Aid Construction Waste Minimisation, *Current. 2005* (2011) 26–28.
- 4 [13] M. Osmani, Design waste mapping : a project life cycle approach, *ICE-Waste Resour.*
5 *Manag.* 166 (2013) 114 – 127.
- 6 [14] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, et al., *Big data: The*
7 *next frontier for innovation, competition, and productivity*, McKinsey Global Institute,
8 New York, 2011.
- 9 [15] F.X. Diebold, “Big data” dynamic factor models for macroeconomic measuring and
10 forecasting, *Vasa.* 32 (2003) 54–55.
- 11 [16] E. Siegel, *Predictive Analytics: The Power to Predict who Will Click, Buy, Lie, Or Die*,
12 John Wiley & Sons, New York, 2013.
- 13 [17] K. Kim, G. Kim, D. Yoo, J. Yu, Semantic material name matching system for building
14 energy analysis, *Autom. Constr.* 30 (2013) 242–255. doi:10.1016/j.autcon.2012.11.011.
- 15 [18] A. Radinger, B. Rodriguez-Castro, A. Stolz, M. Hepp, *BauDataWeb*, in: *Proc. 9th Int.*
16 *Conf. Semant. Syst. - I-SEMANTICS '13*, 2013: p. 25. doi:10.1145/2506182.2506186.
- 17 [19] B.Y.A. Basu, Five pillars of prescriptive analytics success, *Exec. Edge.* (2013) 8–11.
- 18 [20] S. Ryza, U. Laserson, S. Owen, J. Wills, *Advanced Analytics with Spark*, O’Reilly,
19 Cambridge, 2015.
- 20 [21] K. Panos, G.I. Danai, Survey Regarding Control and Reduction of Construction Waste,
21 PLEA2012 - 28th Conf. Oppor. Limits Needs Towar. an Environ. Responsible Archit.
22 Lima, Perú 7-9 Novemb. 2012. (2012) 7–9.
- 23 [22] A. Keys, an Baldwin, Designing to encourage waste minimisation in the construction
24 industry, *Proc. CIBSE Natl. Conf. Dublin.* (2000).
- 25 [23] D. Langdon, *Designing out Waste : a design team guide for buildings*, 2011.
- 26 [24] S.O. Ajayi, L.O. Oyedele, B. Ceranic, M. Gallanagh, K.O. Kadiri, Life cycle
27 environmental performance of material specification: a BIM-enhanced comparative
28 assessment, *Int. J. Sustain. Build. Technol. Urban Dev.* 6 (2015) 14–24.
29 doi:10.1080/2093761X.2015.1006708.

- 1 [25] W. Fan, A. Bifet, Mining Big Data : Current Status , and Forecast to the Future, ACM
2 SIGKDD Explor. Newsl. 14 (2013) 1–5. doi:10.1145/2481244.2481246.
- 3 [26] A. Jacobs, The Pathologies of Big Data, Commun. ACM. 7 (2009) 10.
4 doi:10.1145/1563821.1563874.
- 5 [27] R. Thomas, P. McSharry, Big Data Revolution, Wiley, West Sussex, UK, 2015.
- 6 [28] D. Singh, C.K. Reddy, A survey on platforms for big data analytics, J. Big Data. 2 (2015)
7 8. doi:10.1186/s40537-014-0008-6.
- 8 [29] M. Stonebraker, The Case for Shared Nothing 2 . A SIMPLE ANALYSIS, Contract. 9
9 (1986) 1–5.
- 10 [30] T. White, Hadoop. The definitive guide., 3rd ed., O’Reilly, Sebastopol, Canada, 2012.
- 11 [31] S. Ghemawat, H. Gobioff, S.-T. Leung, The Google file system, ACM SIGOPS Oper.
12 Syst. Rev. 37 (2003) 29.
- 13 [32] J. Dean, S. Ghemawat, MapReduce : Simplified Data Processing on Large Clusters,
14 Commun. ACM. 51 (2008) 1–13. doi:10.1145/1327452.1327492.
- 15 [33] A. Berkeley, Berkeley Data Analytics Stack (BDAS), (2015).
16 <https://amplab.cs.berkeley.edu/software/> (accessed April 1, 2015).
- 17 [34] Apache Software Foundation, Apache Flume, (2015). <https://flume.apache.org/> (accessed
18 July 1, 2015).
- 19 [35] J. Beetz, J. van Leeuwen, B. de Vries, IfcOWL: A case of transforming EXPRESS
20 schemas into ontologies, Artif. Intell. Eng. Des. Anal. Manuf. 23 (2009) 89.
21 doi:10.1017/S0890060409000122.
- 22 [36] I. Robinson, J. Webber, E. Eifrem, Graph Databases, 2nd ed., O’Reilly, Sebastopol,
23 Canada, 2015.
- 24 [37] A.Y. Halevy, N. Ashish, D. Bitton, M. Carey, D. Draper, J. Pollock, et al., Enterprise
25 information integration: successes, challenges and controversies, Proc. 2005 ACM
26 SIGMOD Int. Conf. Manag. Data SIGMOD 05. (2005) 778–787.
27 doi:10.1145/1066157.1066246.
- 28 [38] H. Garcia-Molina, Y. Papakonstantinou, D. Quass, A. Rajaraman, Y. Sagiv, J. Ullman, et
29 al., The TSIMMIS approach to mediation: Data models and languages, J. Intell. Inf. Syst.
30 8 (1997) 117–132. doi:10.1023/A:1008683107812.

- 1 [39] S. Chaudhuri, U. Dayal, An overview of data warehousing and OLAP technology, ACM
2 SIGMOD Rec. 26 (1997) 65–74. doi:10.1145/248603.248616.
- 3 [40] I. Martínez Lage, F. Martínez Abella, C.V. Herrero, J.L.P. Ordóñez, Estimation of the
4 annual production and composition of C&D Debris in Galicia (Spain), Waste Manag. 30
5 (2010) 636–645. doi:10.1016/j.wasman.2009.11.016.
- 6 [41] J. Shi, Y. Xu, Estimation and forecasting of concrete debris amount in China, Resour.
7 Conserv. Recycl. 49 (2006) 147–158. doi:10.1016/j.resconrec.2006.03.011.
- 8 [42] D. Fatta, a. Papadopoulos, E. Avramikos, E. Sgourou, K. Moustakas, F. Kourmoussis, et
9 al., Generation and management of construction and demolition waste in Greece - An
10 existing challenge, Resour. Conserv. Recycl. 40 (2003) 81–91. doi:10.1016/S0921-
11 3449(03)00035-1.
- 12 [43] J. Solís-Guzmán, M. Marrero, M.V. Montes-Delgado, A. Ramírez-de-Arellano, A Spanish
13 model for quantification and management of construction waste, Waste Manag. 29 (2009)
14 2542–2548. doi:10.1016/j.wasman.2009.05.009.
- 15 [44] M.J. Shepperd, G. Kadoda, Comparing software prediction techniques using simulation,
16 IEEE Trans. Softw. Eng. 27 (2001) 1014–1022.
- 17 [45] C. Mair, G. Kadoda, M. Le, K. Phalp, C. Scho, M. Shepperd, et al., An investigation of
18 machine learning based prediction systems, J. Syst. Softw. 53 (2000) 23–29.
19 doi:10.1016/S0164-1212(00)00005-4.
- 20 [46] S.K. Card, J.D. Mackinlay, B. Shneiderman, Readings in information visualization: using
21 vision to think, Morgan Kaufmann, San Francisco, CA, USA, 1999.
- 22 [47] C. Chen, Information visualization: Beyond the horizon, Springer Science & Business,
23 New York, 2006.
- 24 [48] C.G. Healey, J.T. Enns, Perception and painting: A search for effective, engaging
25 visualizations, IEEE Comput. Graph. Appl. 22 (2002) 10–15. doi:10.1109/38.988741.
- 26 [49] D. a Keim, Visual Exploration of Large Data Sets, Commun. ACM. 44 (2009) 39–44.
- 27 [50] D. a Keim, F. Mansmann, J. Thomas, D. Keim, Visual Analytics: How Much
28 Visualization and How Much Analytics?, ACM SIGKDD Explor. Newsl. 11 (2010) 5–8.
29 doi:10.1145/1809400.1809403.
- 30 [51] W. Lu, X. Chen, D.C.W. Ho, H. Wang, Analysis of the construction waste management
31 performance in Hong Kong: the public and private sectors compared using big data, J.

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

2 [52] Z. Wu, H. Fan, G. Liu, Forecasting Construction and Demolition Waste Using Gene
3 Expression Programming, 29 (2008) 1–8. doi:10.1061/(ASCE)CP.1943-5487.0000362.

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

6 [54] U. Fayyad, The KDD Process for Extracting Useful Knowledge from Volumes of Data, 39
7 (1996) 27–34. doi:10.1145/240455.240464.

8 [55] X. Wu, X. Zhu, G.-Q. Wu, W. Ding, Data mining with big data, IEEE Trans. Knowl. Data
9 Eng. 26 (2014) 97–107. doi:10.1109/TKDE.2013.109.

10