Automated look-ahead schedule generation using linked-data based constraint checking for construction projects

Ranjith Kuttantharappel Soman (B. Tech, MSc)

(CID: 01322868)

November 2020

A thesis submitted towards the fulfilment of requirements for the PhD degree of Imperial College London and Diploma of the Imperial College (DIC)

Centre for Systems Engineering and Innovation

Department of Civil and Environmental Engineering

Imperial College London, United Kingdom

Declaration

I hereby declare that the entire work presented in this thesis has been personally carried out. Where sources of information or the work of others have been used, they are fully cited and referenced and/or appropriate acknowledgement is given.



…… ……

Ranjith Kuttantharappel Soman

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this license, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the license terms clear to others by naming the license and linking to the license text. Please seek permission from the copyright holder for uses of this work that are not included in this license or permitted under UK Copyright Law.

Dedicated to

Saeed Tahmasebi

A good friend, and a great engineer

Acknowledgement

This thesis marks the last leg of my journey to obtaining my PhD. degree. This PhD. study has been kept on track and been seen through to completion with the support and encouragement of many, including my supervisors, friends, colleagues, and various institutions. I want to thank all those people who made this thesis possible and making it an unforgettable experience for me.

'True mentors are the ones who inspire you to outdo yourself.' Nothing else describes my supervisors better. I want to express my special appreciation and thanks to Professor Jennifer Whyte, who has been a tremendous mentor for me. She gave me invaluable advice throughout my research, shared her knowledge, experience, and ideas. Her advice on both research as well as my career has been priceless. I am extremely grateful to my co-supervisor, Dr. Miguel Molina-Solana, for his valuable guidance, scholarly inputs, and consistent encouragement that I received throughout the research work. Besides, I would like to acknowledge the help and encouragement offered by Dr. David Birch, who co-supervised me during the first leg of my Ph.D. journey. Also, I would like to thank my industrial supervisor Dean Bowman, for sharing his endless enthusiasm and providing insight into the world of digital construction management. In addition to my supervisors, I acknowledge the guidance offered by Professor Jakob Beetz, Professor Rafael Sacks, Professor Koshy Varghese, Professor Benny Raphael, Dr. Pieter Pauwels , Dr. Ashwin Mahalingam, Dr. Vishal Singh, and Dr. Ioannis Brilakis. I would also like to thank my examiners Prof. Tony Thorpe and Dr. Panagiotis Angeloudis for their constructive comments.

Various institutions have supported me in the different phases of my PhD. Bentley Systems UK has been phenomenal in supporting this work throughout, I would like to take this opportunity to extend my gratitude to all my friends at Bentley. Special thanks to David

Robertson, director of digital advancement research and Dr. Nabil Abou-Rahme, global research officer at Bentley Systems for setting up the funding for the PhD and helping me get connected to key people in the industry. Most of the results described in this thesis would not have been obtained without close collaboration with leading construction projects. First, I would like to thank Peter Vale, who introduced me to different sites at the Tideway Project for conducting my case studies. Special thanks to Tom Inglis and Lampros Arvanitis who helped me get settled and connected within the Tideway offices, and thanks to engineers at FLO JV and Costain-Vinci JV for sparing their time for my research. Second, I would like to thank the engineers at Laing O' Rourke, who helped to understand the digital collaboration workflows at their sites. Third, I extend my gratitude to engineers and managers at the Nagpur metro project, who took the time to introduce me to their sites and their 5D BIM workflows.

I thank Department of Civil and Environmental Engineering for providing all necessary facilities and excellent support, including the Skempton scholarship. Many thanks to everyone in the Centre for Systems Engineering and Innovation; I gratefully appreciate all of your love, care and moral support. My warmest gratitude goes to Tim Gordon for all the ways he helped me throughout my time at Imperial. A journey is easier when you travel together.

Thank you does not seem sufficient to acknowledge the contribution of friends. They were always beside me during the happy and hard moments to push me and motivate me. Finally, the biggest thanks of all to my mom, my dad, and my brother; they gave me endless support and love in every phase of my personal and academic life and longed to see this achievement come true.

-Ranjith

Abstract

Poor planning on the construction site at the 'look-ahead' planning stage (where information from diverse sources is integrated as plans are being developed for the next six weeks) often result on cost overruns and schedule delays.

This thesis addresses the inefficiencies emerging from manual look-ahead planning by means of an information modelling approach to codify and validate detailed construction process information. The key contribution is a novel Linked-Data based Constraint Checking (LDCC) method to identify construction constraint violations from information distributed over multiple heterogeneous sources. LDCC can be integrated with machine learning methods to generate a constraint-free Look-Ahead Schedule (LAS) automatically. This approach can augment decision-making in look-ahead planning through data-driven constraint identification and automated LAS generation.

This research comprises of three main studies. First, digital information use in three construction projects was studied to inform the development of the information modelling approach. This study identified three codification challenges—software usage, information sharing, and missing construction process information. Second, building on the understanding of the codification challenges, the LDCC method was developed to codify and validate detailed construction process information, including complex construction constraints distributed over multiple databases. Third, the LDCC method was successfully integrated with two machine learning methods (Genetic Algorithm and Reinforcement Learning) to automate LAS generation.

When tested on a real construction project, the LDCC method identified all the constraint violations in the manually generated LAS. Also, LAS generation methods

automatically generated conflict-free LASs significantly faster than manual methods. Both results demonstrate the applicability of the developed methods on real construction projects.

In summary, the thesis extends existing knowledge in the construction informatics domain by demonstrating the benefits of using linked-data based methods to address the issue of data silos in construction information, and enabling the application of data-driven-decision support tools to aid look-ahead planning.

Table of contents

Declaration **2**

Acknowledgement **4**

Abstract **6**

Table of contents **8**

List of Figures **12**

List of Tables **14**

Glossary **15**

1. INTRODUCTION **17**

1.1. Background *17*

1.2. Aim and Objectives *20*

1.3. Justification for the study *20*

1.4. Delimitation of scope *23*

1.5. Thesis Structure *23*

2. REVIEW OF LITERATURE **27**

2.1. Introduction *27*

2.1.1. Constraint management 28

2.1.2. Look-ahead planning 30

2.1.3. Data-driven decision-support for look-ahead planning 32

2.2. Data quality issues in construction information *34*

2.2.1. Data quality dimensions 35

2.2.2. Machine Readability 36

2.2.3. Need for ensuring data quality 37

2.2.4. Prior research on data quality issues in construction 39

2.3. Modelling and validating construction process information *41*

2.3.1. Modelling detailed construction process information 41

2.3.2. Validating construction process information using rule-checking 45

2.4. Automating LAS generation *53*

2.4.1. Scheduling 53

2.4.2. Resource-Constrained Project Scheduling Problem 55
2.4.3. Limits of existing research 63
2.5. Summary: Research issues to be addressed 64
3. RESEARCH DESIGN 67
3.1. Introduction 67
3.2. Identifying codification challenges in construction (M1) 68
3.2.1. Case selection 69
3.2.2. Data collection 70
3.2.3. Data analysis 72
3.3. Codifying and validating construction process information using linked-data (M2) 75
3.4. Automating LAS generation (M3) 76
3.5. Implementation study 78
3.6. Summary 80
4. IDENTIFYING CODIFICATION CHALLENGES IN CONSTRUCTION 82
4.1. Introduction 82
4.2. Overview of the cases 84
4.3. Codification challenges in construction 86
4.3.1. Software usage 88
4.3.2. Information sharing 94
4.3.3. Missing information 110
4.4. Synthesis of the findings 116
4.5. Summary 117
5. CODIFYING AND VALIDATING CONSTRUCTION PROCESSES INFORMATION USING LINKED-DATA 120
5.1. Introduction 120
5.2. Linked-data-based constraint checking architecture 121
5.2.1. Ontology for linked-data-based constraint checking 124
5.2.2. Modelling scheduling constraints using Shapes Constraint Language (SHACL) 126
5.3. Linked-data-based constraint checking implementation 135
5.3.1. Prototype web application 135

5.3.2. Prototype evaluation 140

5.4. Implications of LDCC 144

5.5. Summary 148

6. AUTOMATING LOOK-AHEAD SCHEDULE GENERATION 150

6.1. Introduction 150

6.2. Formulating the LAS problem 151

6.3. Solving LAS problem using a GA 153

6.3.1. Population initialization 155

6.3.2. Fitness evaluation 157

6.3.3. Selection 159

6.3.4. Crossover 159

6.3.5. Mutation 161

6.4. Solving LAS problem using RL 161

6.4.1. Q-Learning 162

6.4.2. Modelling LAS problem 165

6.5. Computational Experiments 174

6.5.1. Comparison with traditional methods 175

6.5.2. Influence of the number of resource types 177

6.5.3. Influence on the number of processes and activities 179

6.5.4. Discussion 182

6.6. Summary 184

7. IMPLEMENTATION STUDY 185

7.1. Introduction 185

7.2. Case study description 186

7.3. Data import and enrichment 189

7.4. Results 194

7.4.1. Constraint checking using LDCC 194

7.4.2. Look-ahead-schedule (LAS) generation using GA and RL 197

7.5. Discussion 199

7.6. Summary 201

8. DISCUSSION AND CONCLUSIONS 202

8.1. Research summary 202

8.2. Combined discussion of the findings 205

8.2.1. Software usage 207

8.2.2. Information sharing 210

8.2.3. Missing process information 211

8.3. Thesis conclusions 213

8.4. Limitations of the research study 215

8.5. Implications to practice 216

8.6. Directions for future research 218

References 221

Appendix A: Data collection record -Method M1 263

Appendix B: Case Details-Method M1 265

Appendix C: Method M1-Emergent codes 269

Appendix D: Relationship between data quality dimensions and codification challenges 271

Appendix E: Relationship between levels of machine readability and codification challenges 277

Appendix F: LAS Generated during implementation study 281

Appendix G: Interview protocol for method M1 285

Appendix H: Protocol for the implementation study 293

Appendix I: Extended research summary 299

Appendix J: Examples of SHACL Snippets for scheduling constraints 302

Appendix K: Comparison of 4D tools 313

Appendix L: Solving RPSP using CPLEX 315

List of Figures

Figure 1.1: Thesis structure 26

Figure 2.1: RDF triple 44

Figure 3.1: Data analysis steps 73

Figure 4.1: Chapter overview 83

Figure 4.2: Federated BIM model used for discussion (Student Apartment- C1D2) 95

Figure 4.3: Snap shot from design BIM models (Student apartment -C1D2) a) Structural BIM model b) Architecture BIM model c) MEP BIM model 99

Figure 4.4: Screen shot of the hoist showing lack of constraint details (Student apartment, C2D2) 111

Figure 4.5: Visual representation showing the effect of the low level of detail in process information. The deck is completed before the completion of the supporting pier 112

Figure 4.6: Extract from construction programme [Student apartment-C1D1] 114

Figure 5.1: LDCC architecture 122

Figure 5.2: Schematic of the LinkOnt ontology, showing its classes and relationships between them 125

Figure 5.3: Functional details of the software implementing the LDCC method 136

Figure 5.4: User interface 138

Figure 5.5: Screenshot of the imported Schependomlaan dataset in an enriched site environment. 142

Figure 5.6: Task-edit interface for the imported dataset. The data shown refer to SubActivity S1 143

Figure 5.7: Successful validation of S1 activity. 143

Figure 5.8: Validation failure for S2 activity. 144

Figure 6.1: Genetic algorithm steps 153
Figure 6.2 LAS generation using genetic algorithm 154
Figure 6.3: Crossover operation illustration 160
Figure 6.4: Reinforcement learning model 161
Figure 6.5: Q-Learning steps 164
Figure 6.6 : Q-learning workflow for LAS generation 165
Figure 6.7: State in a LAS- Q-Learning problem 168
Figure 6.8: Action combinations 170
Figure 6.9: Gantt chart developed using different strategies 176
Figure 6.10: Influence of increasing number of resource type 178
Figure 6.11: Influence on number of processes and activities 181
Figure 7.1: STP foundation and Muck bin construction areas 187
Figure 7.2: Data imported to the LDCC prototype 190
Figure 7.3: Task edit UI showing sub activities details added to the processes Five activities can be seen in the left panel, of which BIN A is selected and their finer details and sub-activities shown in the right panel 191
Figure 7.4: Task edit UI showing resource details and constraint details added to the processes for BIN C 192
Figure 7.5: Constraint violation for Muck bin construction Bin A 195

List of Tables

Table 2.1: Levels of machine readability....37

Table 2.2: Methods to model construction process information....43

Table 2.3: Classification of algorithms for solving scheduling problems....54

Table 3.1: Research methods to address objectives....68

Table 4.1: Details of projects studied....84

Table 4.2: Code book for codification challenges used in the analysis....86

Table 4.3: Information sharing between teams in the water project....97

Table 4.4: Codification challenges across the cases....117

Table 5.1: Task details used as inputs for the evaluation....140

Table 5.2: Results after constraint checking....142

Table 6.1: Details of the computer in which the experiments were run....174

Table 6.2: Parameters for the computational experiment....175

Table 6.3: Input parameters....175

Table 6.4: Comparison of the results....176

Table 6.5: Input data....177

Table 6.6: Comparison of Genetic and Reinforcement learning algorithms....179

Table 6.7: Input data to the algorithm....179

Table 6.8 : Comparison of Genetic and Reinforcement learning algorithms....181

Table 7.1 Processes and sub-activities....188

Table 7.2: Equipment assignment....189

Table 7.3: Parameters for the LAS generation....194

Table 7.4: Violations detected in the LDCC prototype....195

Table 7.5: Schedule deviations....198

Table 8.1: Summary of the research....203

Glossary

ACO	Ant Colony Optimization
ADC	Automated Data Collection
AEC	Architecture, Engineering and Construction
API	Application Program Interface
BIM	Building Information Modelling
BOQ	Bill of Quantities
BRMS	Business Rule Management System
BSI	British Standards Institution
CDE	Common Data Environment
CPM	Critical Path Method
CPU	Central Processing Unit
DfMA	Design for Manufacture and Assembly
ERP	Enterprise Resource Planning
GA	Genetic Algorithm
GPS	Global Positioning System
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technologies
IFC	Industry Foundation Classes
ISO	International Standards Organization
IT	Information Technology
LAS	Look-Ahead Schedules
LDCC	Linked-Data based Constraint Checking
LPS	Last Planner System

MEP	Mechanical Engineering and Plumbing
O1	Objective 1
O2	Objective 2
O3	Objective 3
PDF	Portable Document Format
PSO	Particle Swarm Optimization
RL	Reinforcement Learning
RCMPSP	Resource Constrained Multi-Project Scheduling
RCPSP	Resource Constrained Project Scheduling
RDF	Resource Description Framework
RIF	Rule Interchange Format
SA	Simulated Annealing
SHACL	Shapes Constraint Language
SPIN	SPARQL Inferencing Notation
SPARQL	SPARQL Protocol and RDF Query Language
STP	Slurry Treatment Plant
SWRL	Semantic Web Rule Language
UI	User Interface
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
WIP	Work in Progress

1. INTRODUCTION

1.1. Background

Planning work on a construction site is a challenging task as there are several activities inextricably linked to each other. If the work is poorly planned, different stakeholders in the project would have a different understanding of the construction plans and constraints, resulting in workspace conflicts and rework, eventually leading to cost overruns and schedule delays. Can digital information help with this problem in the 'look-ahead' planning stage, where plans are being developed for the next six weeks? This research study describes the development of a data-driven decision-support system to aid look-ahead planning. It presents a method to identify the conflicts in construct plan by deriving insights from the construction information distributed over multiple projects servers and assist in creating a conflict-free plan of work.

Look-ahead planning is a medium-term planning method proposed by Ballard (1997) to shape workflow, and it aims to identify and remove constraint violations that may interfere with the continuous progression of work (Tommelein, 1998). The output of look-ahead planning is a Look Ahead Schedule (LAS), that stands between overall project coordination schedules and short-term weekly plans for crew-level commitments, filtering out scheduled activities that cannot be implemented yet (Ballard, 2000).

> "To achieve that, we review the 3 Weeks Lookahead in the beginning of the week with the use of Lean Construction Last Planner Boards and by having the 4D model open on screen for everyone to use […] in a 4D Collaborative Look Ahead Planning meeting where we bring different teams in the room to discuss safety, constructability, access, resources, any prerequisites for the works and eventually identify risks and opportunities […] we achieved a 20% programme reduction of the construction activities and we eliminated rework […] We have also estimated that if we hadn't resolved these issues beforehand with the use of look-ahead planning, we would have ended with more than £1,000,000 cost overruns" (Water Project, Digital engineer,C3R1)

When implemented properly, look-ahead planning has resulted in reduction in cost over runs, elimination of delays and rework and better coordination. However, there are inefficiencies associated with manual look-ahead planning such as difficulty in identifying and removing constraint violations (Hamzeh, 2009), resulting in scheduling tasks that are not ready for execution (Lindhard & Wandahl, 2014; Nieto-Morote & Ruz-Vila, 2012). Also, manual LAS generation is time-consuming (Dong *et al.*, 2013a).

There is great potential to develop data-driven approaches to support look-ahead planning in construction. In look-ahead planning, it is important to ensure that all prerequisites are complete and that constraints violations are removed for each work package before it is released to weekly work plans (Ballard, 1997). Prerequisites are the conditions necessary for a work package to be executed, such as availability of resource and completion of prior tasks. Constraint violations happen when the prerequisites are not complete. Research studies have focused on Automated Data Collection (ADC) technologies, such as Global Positioning Systems (GPS), Radio Frequency Identification (RFID), computer vision, and other sensor-based data collection methods, to track and monitor prerequisites, thereby informing the progress of activities at the construction sites (Alizadehsalehi & Yitmen, 2016; Omar & Nehdi, 2016; Kopsida, Brilakis & Vela, 2015). Although ADC technologies help in capturing information about the status of activities, research concerning the development of data-driven methods to identify constraint violations using this information have received insufficient attention.

The combination of the sensor data from ADC and the semantic relationships of the activities (including constraints relationships) embedded in the information models, such as Building Information Model (BIM), enable data-driven checks to identify constraint violations. However, it is often the case that the model breakdown structures in these information models do not match the work breakdown structures for construction site activities (Han & Golparvar-

Fard, 2017) as they only contain a subset of the relationships between activities and resources (Giretti *et al.*, 2012). Also, construction scheduling constraints such as disjunctive constraints and discrete resource capacity constraints (Morkos, 2014; Niederliński, 2011), are not codified in these information models. The current lack of detail and semantics in the information models limits the application of data-driven constraint checking systems.

Although embedding detailed construction information (including constraints) and associated semantics into a machine-readable form can enable data-driven constraint checking, it brings its own challenges. The first challenge is that the required data remains distributed in multiple heterogeneous datasets (Arayici *et al.*, 2018; Tchouanguem Djuedja *et al.*, 2019) over multiple servers, creating problems of interoperability. The second challenge is that the construction constraints are dynamic and span multiple domains (Wang *et al.*, 2016), requiring a constraint checking method that could traverse through heterogeneous data stored on different servers and domains to check for constraint violations. These challenges limit the codification of detailed construction process information and semantics into a machine-readable form.

This PhD study focuses on the development of an information modelling approach that would support the embedding of detailed construction information in a machine-readable format to enable data-driven constraint checks for look-ahead planning, thereby addressing the challenges stated above. To do that, first, the data quality issues in the late-design and construction stages are studied. Findings from this study are used to drive the development of the information modelling approach and constraint checking method. The constraint checking method and the information modelled using the developed method is then used to generate LAS automatically. This is an important contribution to practice as this approach will introduce data-driven decision-support for look-ahead planning, thus contributing to improving productivity in construction. The rest of this chapter describes the aims and objectives of the

study (Section 1.2), its justification (Section 1.3), and its scope (Section 1.4), and outlines the structure of the thesis (Section 1.5).

1.2. Aim and Objectives

This PhD study addresses the issues arising out of manual look-ahead planning. To address the issues, look-ahead planning is augmented using data-driven decision-support tools developed in this research.

Research aim: *to develop an information modelling approach to codify and validate detailed construction process information in BIM-enabled projects, in order to support data-driven decision-making for look-ahead planning in construction.*

Three objectives are framed around the research aim. The justification for articulating these objectives are explained in Section 1.3. The objectives are as follows.

Objective 1 (O1): To understand the codification challenges in the late-design and construction phase of BIM-enabled projects.

Objective 2 (O2): To develop an information modelling approach to codify and validate construction process information including dynamic constraint information; and

Objective 3 (O3): To develop a method to generate look-ahead-schedule for construction projects automatically using the codified process information.

1.3. Justification for the study

The research problem addressed in this PhD study has theoretical and practical implications. Section 1.1 presented the need to develop data-driven decision-support for look-ahead planning (research aim), to address the inefficiencies in manual look-ahead planning resulting in the generation of inaccurate LAS. The need for data-driven decision-support for

look-ahead planning is explained in detail in Subsection 2.1.3. This section presents the justification for choosing the three objectives of this research.

- To inform the development of data-driven decision-support tools for look-ahead planning, the issue of poor quality machine-readable information limiting the wider application of these tools (Bilal *et al.*, 2016; Sacks *et al.*, 2017b; Farias, Roxin & Nicolle, 2018) needs to be studied. Although there are advanced methods for gaining insights from poor quality data (Hossen & Sayeed, 2018; Carrillo, Harding & Choudhary, 2011; Wang, 2017), these methods are resource-intensive and may not yield good quality insights (Wang, 2017; Fan, 2015; Janssen, van der Voort & Wahyudi, 2017). Prior research has drawn attention to the factors which affect information use and BIM implementation (Dossick & Neff, 2010; Won *et al.*, 2013; Taylor, 2007). However, these studies do not focus on explaining how these practices cause data quality issues. By better characterizing the practice, such empirical work which understands technology use as a complex socialized activity (Cao, Li & Wang, 2014) inform future technical developments (such as data-driven decision-support for look-ahead planning) of the codification challenges and data quality issues that exist in the sector. This has led to the framing of objective O1. This justification is further explained in Section 2.2.
- As described earlier, there is a need to develop an approach to model detailed construction information (including construction constraints) and then validate them to identify constraint violations, taking into account the inherent challenges of construction process information as well as the challenges identified in objective O1. There is a trajectory of research looking at codifying and validating detailed construction process information, in order to use automated methods to generate construction programs (Moon *et al.*, 2015; Choi *et al.*, 2014; Liu, Al-Hussein & Lu,

2015; Wang, Lin & Zhang, 2020). Although these methods present a robust framework for the codification of construction process information, they do not address inherent challenges associated with construction datasets, as stated in Section 1.1. There is another strand of research which focus on validating information stored in BIM against a set of modeled rules (Schwabe, König & Teizer, 2016; Lee, Solihin & Eastman, 2019; Bloch & Sacks, 2018). The principles from this set of studies can be extended to develop constraint-checking method to support look-ahead planning. However, these methods require information to be in a single repository to perform rule-checking. Therefore, a new method to model information distribution of data across multiple databases and a constraint checking method which could validate the distributed information is required. This has led to the framing of objective O2. This justification is further explained in Section 2.3.

- Automatic generation of LAS can help look-ahead planning as previous research has reported that inaccurate LAS is being developed due to inefficiencies in manual look-ahead planning (described in Section 1.1). Prior research in this area has focused on a class of problems called Resource-constrained project scheduling problem (RCPSP) (Hartmann & Briskorn, 2010), a superset of LAS generation problem. These studies have looked into the development of heuristics (Kolisch, 1996b; Li & Willis, 1992; Valls, Ballestín & Quintanilla, 2005) and metaheuristics (Agarwal, Colak & Erenguc, 2015) to solve RCPSP. Although RCPSP is similar to LAS generation problems, LAS is characterized by a small number of interdependent activities happening over a short period (1 week to 6 weeks). In contrast, RCPSP problems are suited for a higher number of activities happening over longer periods (several years for master plan level) (Mossman, 2013). In contrast with RCPSP considering precedence and resource constraints, LAS

considers even more complex constraints such as workspace conflicts, logistic requirements, equipment operating criteria (Lindhard & Wandahl, 2014). Thus, the algorithms developed for RCPSP needs to be revisited to automate the generation of LAS. This has led to the framing of O3. This justification is further expanded in Section 2.4.

1.4. Delimitation of scope

The main effort in this study is directed at identifying the codification challenges which limit the automated generation of LAS (objective O1), modelling and validating detailed construction process information including constraints (objective O2), and finally using the detailed construction process information to generate LAS (objective O3). The current study does not concentrate on modelling geometric constraints in the construction site as the current geometric representation ontologies do not support direct inferencing. Significant research needs to be done to develop geometric representation which supports inferencing and is out of the scope of this study. Also, LAS generation in the current study focuses on integrating the constraint checking method into existing metaheuristics to solve RCPSP problems, so that these metaheuristics can be used to solve LAS generation problem. However, this study does intend on developing a novel method for solving RCPSP. Recommendations for further research that result from the current research are detailed in Section 8.5.

1.5. Thesis Structure

The thesis is structured into five parts consisting of introduction (Chapter 1), literature review (Chapter 2), research method (Chapter 3), findings (Chapter 4-7), and discussion and conclusions (Chapter 8). Chapters 4-6 present the main findings and Chapter 7 presents the validation of the findings. The relationships between the different parts of the thesis are presented in Figure 1.1, and the details of each chapter are given below.

- Chapter 2 **Review of literature** presents the areas relevant to the current research problem. These include data quality issues in construction, BIM based rule-checking for modelling and validating detailed construction process information and the existing research on generating LAS automatically.
- Chapter 3 **Research design** presents the choice of methods and the logic of choosing research methods for each of the objectives stated in Section 1.2.
- Chapter 4 **Identifying codification challenges in construction** presents findings, through analyzing a case study, to identify the factors which limit the data quality of the information in construction that limits the application of data-driven decision-support tools to aid look-ahead planning/
- Chapter 5 **Codifying and validating construction process information using linked-data** presents the findings of the development of the LDCC method to model detailed construction process information and validate them. This chapter details how detailed construction process information, including construction constraints, can be effectively modeled using linked-data.
- Chapter 6 **Automating Look-Ahead Schedule generation** presents the findings from the study where LDCC method is integrated with machine learning methods: 1) Genetic algorithm (GA); and 2) Reinforcement learning (RL) to generate LAS. This chapter presents the modelling of LAS problem using GA and RL and presents a comparison of the performance of these two algorithms to generate LAS.
- Chapter 7 **Implementation study** presents the findings of a study that tested the LDCC method and LAS generation method using real-world construction data to understand its efficacy.

- Chapter 8 **Discussion and conclusions** present the contributions, discussion of the combined findings and thesis conclusions. This chapter also describes the implication of the current research study and explains directions for future research.

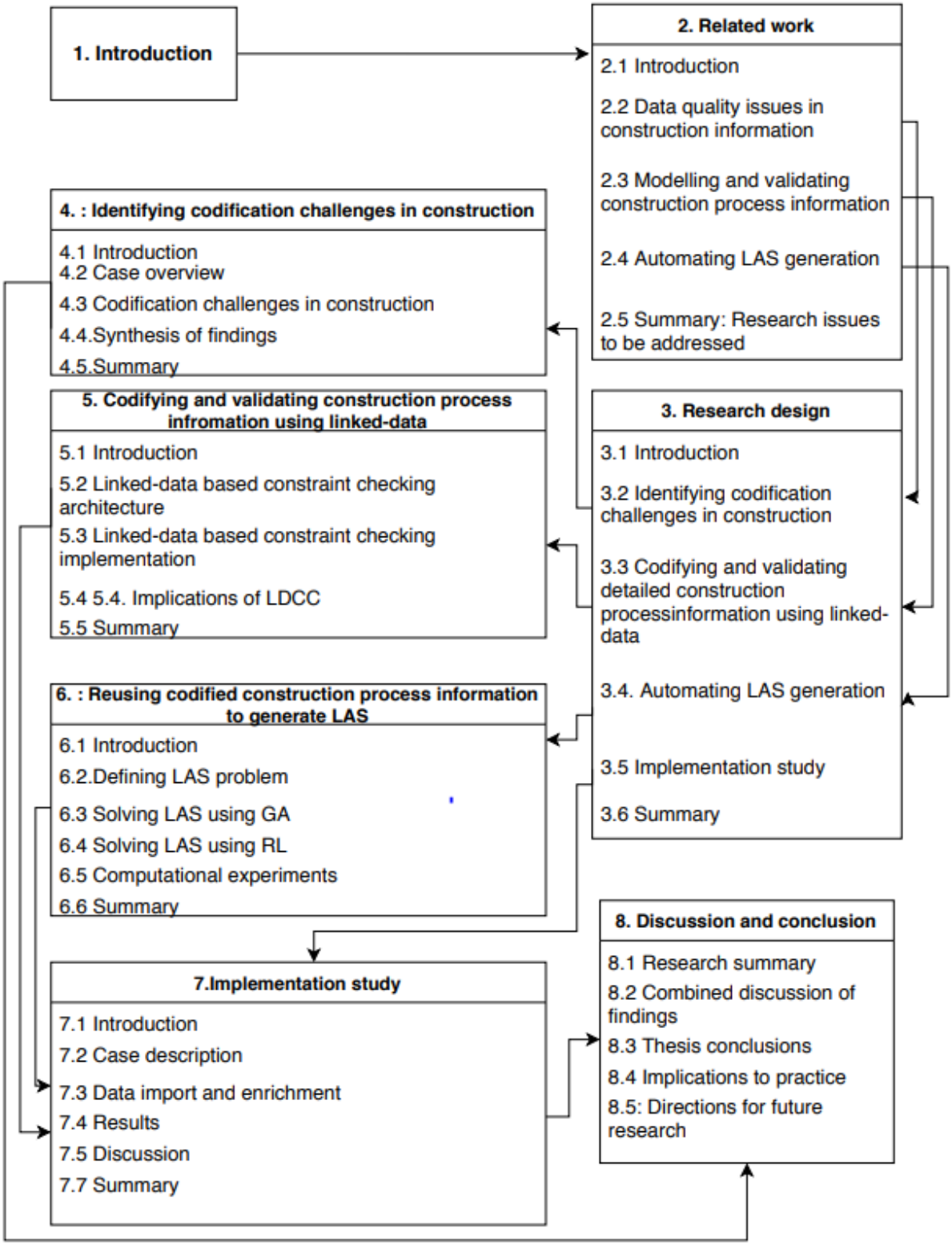


Figure 1.1: Thesis structure

2. REVIEW OF LITERATURE

2.1. Introduction

This chapter reviews relevant literature concerning data-driven decision-support for look-ahead planning. Within this section, Subsection 2.1.1 introduces constraint management (an important topic for the current study as constraint removal is vital to generating an error-free LAS), Subsection 2.1.2 introduces the look-ahead planning process (central to the current PhD), and Subsection 2.1.3 describes data-driven methods to support look-ahead planning. Constraint management, look-ahead planning and data-driven methods to support look-ahead planning provide a context for the study by introducing the relevant concepts and background necessary for the discussions in subsequent sections.

The other sections in this chapter discuss the prior research concerning the three objectives of the current study, and present a summary of the research issues identified in this chapter. Section 2.2 discusses the prior literature concerning objective O1, on data quality issues in construction information, which limits the application of data-driven decision-making. Section 2.3 provides background to objective O2 by describing the existing approaches to model and validate construction process information, including BIM based rule-checking and the application of linked-data in the construction sector. Section 2.4 provides the background to objective O3 by introducing key concepts related to automated generation of LAS, RCPSP, and heuristics and metaheuristics for solving RCPSP. Finally, Section 2.5 presents the research issues identified in the current chapter and provides a roadmap of how these issues are addressed in the following chapters.

2.1.1. **Constraint management**

Constraint management, first introduced by Goldratt and Cox (1984), is a management philosophy based on the theory of constraints. Prior research has shown that constraint management is a key issue in production planning and control (Goldratt & Cox, 1984; Koskela, 1992; Tommelein, 1998; Sacks *et al.*, 2017a). Theory of constraints is a scientific approach to improvement by identifying the most important limiting factors that stand in the way of achieving a goal and removing them (Goldratt & Cox, 1984). In the construction context, constraints are conditions which prevent successful executions of work plans (Ballard, 1997). They are base concepts of construction scheduling as they inform what can and cannot happen, or in other words, what can and cannot be scheduled at a point of time (Dechter, 2003). Constraints include conditions such as precedence conditions, resource availability conditions, safety conditions, logical conditions limiting the deployment of equipment, etc.(Blackmon, Saxena & Song, 2011). These constraints must be effectively managed to make sure work plans are effectively executed. Scholars have formalized these constraints into precedence, discrete resource capacity and disjunctive constraints (Morkos, 2014; Niederliński, 2011; Baykan & Fox, 1997).

- Precedence constraints make sure that the principles of physics are not violated, thus creating a sequence of work. This constraint is the basis of the Critical Path Method (Darwiche, Levitt & Hayes-Roth, 1988; Alzraiee, Zayed & Moselhi, 2015). These include constraints such as roof must be cast after all the columns are cured, or the third floor should be built after the second floor.
- Discrete resource capacity constraints are used to solve constraints related to the availability and assignment of resources. They are called discrete, because the resource may be available in numbers more than one, but in discrete integer values. This

constraint ensures that resources are allocated according to availability (Hu & Flood, 2012; El-Rayes & Jun, 2009).

- Disjunctive constraints are constraints which states that two activities cannot happen at the same time. These can be used to model spatial constraints (Baykan and Fox, 1997) and safety constraints. For example, if a construction activity A blocks the access to the workspace of another activity B, a disjunctive constraint must be assigned between these activities to ensure that they do not happen in parallel.

In addition, there are cardinality constraints to ensure that the relevant values are present in a model for computations to be performed on it. Also, Logical constraints that cater to construction site constraints and constraints related to resources and workspaces need to be managed.

Widely used methods in construction projects such as Critical Path Method (CPM) and Program Evaluation and Review Technique (PERT) are only able to consider precedence constraints. RCPSP do consider the discrete resource capacity condition in addition to the precedence constraints (Patterson, 1976). Ignoring constraints such as discrete resource capacity and disjunctive constraints results in problems related to identification and tracking of critical conditions for performing construction activity, thus limiting the execution of work plans (Pultar, 1990). This limitation of CPM and RCPSP methods emerges from the transformation view of production adopted in the CPM method, which follows push-based planning.

Transformation view considers the production process as a series of tasks, each with a start and end (Koskela, 2000). The tasks would happen one after the other, i.e., tasks are pushed one after the other. When the series of tasks are applied to the material, it is transformed from one form to another (raw material to end product). Transformation view assumes that the optimization of individual tasks would optimize overall production. However, the

transformation view holds true only if there are no variabilities in the task execution and when task continuity is maintained (Ballard & Howell, 1996). To accommodate the variability in the construction process and to reduce waste, Koskela (1999) has argued that construction project management should combine transformation view of production with flow management principles. Incorporating flow management principles would enable pull-based planning. Unlike a push-system which releases the work based on a pre-planned schedule, pull-system authorises the release of work based on the system status (Hopp & Spearman, 2001, p. 340). In construction, the system status includes the status of completion of preceding tasks, availability of resources and materials, availability of workspace, etc. (Hamzeh, Ballard & Tommelein, 2008).

Building on Koskela (1999), the Last Planner System (LPS) was developed to introduce pull-planning into construction management (Ballard, 2000). It addresses the limitation of the CPM method concerning the push-method, in particular lack of task continuity (Dave *et al.*, 2015). Last planner system has been widely used in construction projects to enable effective program management and production planning to ensure interdisciplinary constraint analysis and constraint removal, resulting in reliable work plans (Wang *et al.*, 2016; Mossman, 2013; AlSehaimi, Tzortzopoulos & Koskela, 2009; Lindhard & Wandahl, 2014; Sacks *et al.*, 2017a). Look-ahead planning is the phase in LPS where the construction programme for the next six weeks is efficiently developed. This phase concentrates in constraint identification and removal.

2.1.2. **Look-ahead planning**

Look-ahead planning is a crucial phase in construction planning and control where the construction programme for the next six weeks is efficiently developed (Ballard, 1997; Wang *et al.*, 2016; Sacks *et al.*, 2017a; Hamzeh, Zankoul & Rouhana, 2015). Look-ahead planning was first introduced as a part of the last planner system. In lean construction, look-ahead

planning enables pull-based scheduling. In contrast with the traditional planning, which follows a push-driven approach, look-ahead planning results in a reliable and accurate workable schedule (Mossman, 2013). It ensures that all prerequisites are complete, and constraints are removed for each work package before the work packages are released to weekly work plans, thus ensuring the flow of work. To do that, the activities that are planned for the next six weeks are identified from the master plan, followed by a constraint analysis on identified activities to make sure they are ready (Wang *et al.*, 2016; Hamzeh, Zankoul & Rouhana, 2015). The activities whose constraints are completely removed and are in a proper sequence are then scheduled for the next six weeks.

Ballard (1997, p. 8) prescribed a seven-step process to perform six-week look-ahead planning. These steps are:

1) The master plan is adjusted according to the latest status and forecast information. Then the activities, which are scheduled for the next six weeks after adjusting the master plan, are reviewed.
2) Week one should be assigned tasks for which all the prerequisites are completed. Scheduled activities having constraint violations are removed.
3) Examine the remaining weeks in the look-ahead, except for the last. Screen out the activity assignments that have constraint violations, considering that some activity assignments were screened out from the first week.
4) Identify the activities in the master schedule, which cannot be started and remove them from the look-ahead planning activities.
5) Translate the look-ahead week activities into assignments that can be performed, by grouping operations.
6) Match the resource availability with the workload. Remove assignments when the resource availability is less than the requirement for that activity.

7) Prepare a list of actions needed to make assignments ready when scheduled.

The LAS is the output of look-ahead planning, and it stands between overall project coordination schedules and short-term weekly plans for crew-level commitments. Activities that are scheduled in the overall project plan but cannot be implemented are filtered out from LAS. (Sacks *et al.*, 2017a). LAS improves the predictability of the cost, and schedule conformance of the project by ensuring the availability of detailed and realistic work sequences, such that workspace conflicts and rework are prevented (Lindhard & Wandahl, 2014).

2.1.3. **Data-driven decision-support for look-ahead planning**

Look-ahead planning is often done manually (Dong *et al.*, 2013a). Previous research has identified difficulty in tracking and monitoring prerequisites and constraints as a recurrent problem during manual look-ahead planning, resulting in LAS with errors (Hamzeh, 2009; Hamzeh, Zankoul & Rouhana, 2015; AlSehaimi, Tzortzopoulos & Koskela, 2009; Lindhard & Wandahl, 2014; Nieto-Morote & Ruz-Vila, 2012). These difficulties have led to investigate the use of data-driven methods to support look-ahead planning in two major areas: ADC technologies (Omar & Nehdi, 2016; Navon, 2007; Dave *et al.*, 2016a) and BIM (Sacks, Radosavljevic & Barak, 2010; Dave *et al.*, 2016b; Sacks *et al.*, 2013; Solihin *et al.*, 2017).

How can ADC technologies support look-ahead planning? ADC technologies have been used to collect the real-time status of a construction site. ADC technologies such as *GPS* (Razavi & Moselhi, 2012; Pradhananga & Teizer, 2013), *RFID* (Majrouhi Sardroud, 2012; Valero, Adán & Cerrada, 2015; Ren, Anumba & Tah, 2011), *computer vision* (Tuttas *et al.*, 2016; Kim, Kim & Kim, 2013; Zhang & Arditi, 2013), and other *sensor-based data collection methods* (Ibrahim & Moselhi, 2014; Soman, Raphael & Varghese, 2017) has been used to fetch as-built information from construction sites (Kopsida, Brilakis & Vela, 2015; Taneja *et al.*, 2011; Omar & Nehdi, 2016; Tang *et al.*, 2019). The data collected using these automated

methods are used to inform the status and progress of tasks happening at the construction site (Isaac & Navon, 2012) to assist decision-making in Step 1 (adjusting master plan according to the latest site update) of the look-ahead planning process Ballard (1997, p. 8). For realizing the benefits of using ADC technologies, the outputs should be linked to an information model such as BIM for ease of understanding and processing (Han, Cline & Golparvar-Fard, 2015).

How can BIM be used to support look-ahead planning? Sacks *et al.* (2010) have explored the synergy between BIM and lean construction. Their review suggests that the BIM functionalities (such as collaboration in construction using 4D BIM, rapid generation and evaluation of construction plan alternatives and object-based communication in BIM) supports look-ahead planning through digital management of the construction process. KanBIM, a prototype using BIM to support planning, negotiation, commitment, thereby supporting look-ahead planning steps 2 and 3, has shown promising results to visualise the status of the different work packages (Sacks *et al.*, 2013). Similarly, researchers have worked on cloud-based BIM to improve collaboration among project participants focusing on communicating constraints (Matthews *et al.*, 2015; Redmond *et al.*, 2012; Dave *et al.*, 2016a). Also, Song *et al.* (2017) have identified that a high level of detail BIM model can be used to generate daily work orders to support Steps 6 and 7 (Matching resource availability with the workload and prepare a list of actions to make the work-ready) in the look-ahead planning process. These work orders can help automatically generate activity assignments from a resource-loaded BIM model with embedded planning information.

Although there have been efforts to support look-ahead planning using ADC technologies and BIM, the success of these methods is highly dependent on the quality of the model and input from the relevant workforce. These methods relied on manual inputs to make sure that the work packages are ready to be released to weekly plans. For example, KanBIM (Sacks *et al.*, 2013) relies on manual input for tracking prerequisites and constraints. In order

to ensure all the constraints are removed, they need to be checked against the latest information When projects get more complex; it is difficult to do these constraint checks and create LAS manually (Hamzeh, Zankoul & Rouhana, 2015; AlSehaimi, Tzortzopoulos & Koskela, 2009), resulting in errors (Lindhard & Wandahl, 2014; Nieto-Morote & Ruz-Vila, 2012). Therefore, there is a need for data-driven decision-support systems to identify constraint violations to support look-ahead planning process. However, data-driven decision-support tools require high-quality data for performing their tasks. For example, Song *et al.* (2017) have shown that low-quality BIM models failed to generate work orders when compared to high-quality BIM models automatically. Thus, high-quality information needs to be modeled for data-driven decision making to be effectively used.

This section provided a brief overview of the use of ICT used to support LAS generation. Prior research concentrated on data acquisition using ADC, or visualization and enhanced collaboration using BIM. The current PhD focuses on developing an information modelling approach to codify and validate detailed construction process information in BIM-enabled projects to support data-driven decision making for look-ahead planning in construction. This requires understanding three main areas (associated with the objectives objective O1-3 respectively), namely: 1) data quality issues in construction information (Section 2.2), 2) modelling and validating detailed construction process information (Section 2.3), and 3) generating LAS automatically (Section 2.4).

2.2. Data quality issues in construction information

This section discusses the existing research on data quality issues in the construction sector. As explained in the previous section, data quality matters and the data need to be machine-readable for deploying data-driven decision-support systems. This means that the data should be accurate and with all the necessary attributes embedded in it. This section introduces

the metrics for evaluating data quality and reviews existing research on why data quality issues occur.

2.2.1. **Data quality dimensions**

What constitutes a good quality dataset? According to Wang and Strong (1996), a good quality dataset is the one that has enough information embedded in it for a particular use by the user. Researchers have set out multiple dimensions to assess the data quality for big data analytics (Batini & Scannapieco, 2016; Wang & Strong, 1996; Cai & Zhu, 2015; Delone & McLean, 2014; Naumann & Rolker, 2000). For this study, the focus is on the following data quality dimensions, adapted the book by Batini and Scannapieco (2016, 70).

- *Accuracy* is the closeness of the measured/represented data and reality. There are two kinds of accuracies: semantic and syntactic. Semantic accuracy relates to the closeness of the data value to reality, whereas syntactic accuracy refers to the closeness of the data representation to the expected data type/model.
- *Completeness* is the measure of information content present in the data compared to the extent of information content required to be present in the data to perform a particular task.
- Temporal dimensions refer to *currency, volatility, and timeliness*. *Currency* relates to the promptness of data updates. *Volatility* refers to the frequency at which the data variance occurs. *Timeliness* refers to the suitability of the current data to perform a task.
- *Consistency* refers to uniformity and constancy of data concerning the semantic rules defined over multiple data items.
- *Accessibility* refers to the ability of data to be accessed by a user (human user or computer program) and generate information from it; and

- *Data provenance* is the description of the origins of data and the process by which it is manipulated.

Jayawardene, Sadiq, and Indulska (2015) analyzed the dimensions stated above and consolidated the overlap of the dimensions, concluding that some of the data quality dimensions are interdependent. For example, the data quality dimension related to semantic accuracy might depend on the timeliness dimension as the data may be accurate with respect to time. However, the same value may be inaccurate at a different time. At the outset, a dataset is considered to be of good quality when the measure of the dimensions listed are high, leading to better inferences from the respective dataset.

2.2.2. **Machine Readability**

What constitutes machine-readable data? Data in a structured format that could quickly be processed by computers are considered as 'machine-readable'. Berners-Lee (2006) provides a set of' 'rules' for creating structured data so that it can be connected and interpreted easily by machines.

1. The first rule is to index the data to make it digitally accessible by storing it on online servers so that computers, as well as people, can easily access it. This rule relates to the *accessibility* dimension of data quality. Indexing the data and storing it online increases the accessibility as it is easy to find.
2. The second rule is to structure the data with relevant schemas for straightforward interpretation by machines. This rule makes the data structured such that semantic relations are embedded within it, resulting in better inferencing and, thus, improving the *syntactic accuracy* of the data.
3. The third rule is to make the schemas public and machine-readable by using open-source schemas to describe the data model so that computers can make interpretations without any proprietary data interfaces. Proprietary data formats

limit data inferencing as the schema against which data is modeled is only accessible to a few applications. Therefore, using an open format would increase the measure of *accessibility* dimension as more applications can use the open schema to derive the context of data for inferencing.

4. The last rule is to link the data with other datasets so that better inferences can be made by deriving the context information. This rule improves the *consistency* dimensions of data quality as the same data is linked to multiple datasets. Linking would ensure that there are no conflicts in the data about a concept stored in multiple databases.

Based on these rules, structured information can be classified into five types in the increasing order of machine readability, as shown in Table 2.1. Increasing the machine readability of the data, in turn, increases the data quality as the dimensions associated with *accuracy, consistency* and *accessibility* are improved in the process.

Table 2.1: Levels of machine readability

Levels of machine readability	Principles for publishing a machine-readable dataset
1-Star	Data is available on the web
2-Star	1-star data structured in a proprietary format
3-Star	1-star data structured in a non-proprietary format
4-Star	3-star data that is published using open standards
5-Star	4-star data with links to other 4-star datasets

2.2.3. **Need for ensuring data quality**

Despite having different methods to make inferences from poor quality data, it is resource-intensive (Hendler & Pardo, 2012; Wang, 2017; Quintero *et al.*, 2015; Loshin, 2011). The construction sector is becoming increasingly digital, such that most information is stored

digitally and is accessible through servers or a CDE held in firms or projects (Preidel *et al.*, 2017; BSI, 2015). However, being digitally accessible does not mean that the information is machine-readable as the semantics may not be embedded in the data (Hendler & Pardo, 2012). Semantics could be derived from the data using advanced machine learning techniques such as natural language processing and deep learning (Wang, 2017; Carrillo, Harding & Choudhary, 2011). Nonetheless, this is a resource-intensive process which requires training models to achieve satisfactory accuracy and has costs associated with it. The result of this process may also not be of high-quality (Wang, 2017). Despite the existence of data-cleaning algorithms across sectors, poor data quality is a big problem in construction, as in other sectors of the economy (costing $3.1 trillion in the United States (Quintero *et al.*, 2015)). In addition, the poor quality of data is increasing operational costs, decreasing revenue, and resulting in missed commercial opportunities (Loshin, 2011).

Scholars have conducted studies on how poor quality data limits the application of machine learning methods. Sacks, Ma *et al.* (2017) have described how the quality of input information influences semantic enrichment of BIM when using machine learning. Moreover, Farias, Roxin, and Nicolle (2018) had shown the effects of poor quality data resulting in wrong inferences when they tried to extract building views using a rule-based method. Whyte, Stasis, and Lindkvist (2016) articulated how a managing change in large datasets becomes a focus in an era of 'big data' in which project information is increasingly characterized by volume, velocity, and variety. Recent work further characterizes such data and also includes characteristics of veracity and value (e.g. Younas, 2019). In construction, data is heterogeneous (Tchouanguem Djuedja *et al.*, 2019), and its veracity is not always known. Data cleaning related to the variety (heterogeneity) and veracity characteristics of big data is especially tricky when compared to data cleaning related to other characteristics such as volume and velocity (Fan, 2015; Janssen *et al.*, 2017). Therefore, there is a need to keep the good quality and in a

machine-readable format (maintaining the data relationships) as far as possible to have the best use of this data, such as using the data to generate LAS automatically.

2.2.4. **Prior research on data quality issues in construction**

What is limiting the generation of good quality machine-readable information within the sector? It does not appear to be technical development as novel technical solutions are being developed by construction informatics researchers and software vendors with a focus on the integration of data in the sector, for example, through the use of *data standards* (Krijnen & Beetz, 2017; Pazlar & Turk, 2008), *cloud-based BIM* (Beetz *et al.*, 2010; Singh, Gu & Wang, 2011; Zhang *et al.*, 2017b), and *linked-data technologies* (Kim *et al.*, 2018; Pauwels *et al.*, 2015; Pedro *et al.*, 2017; Zhang & Beetz, 2015). It does not appear to be policy interventions either. Standards and public mandates have placed BIM at the heart of the information management required to coordinate processes in project delivery and operation of infrastructure, making BIM central to digital tools and workflows in projects (BSI, 2018; Sacks *et al.*, 2017a). Instead, the literature suggests that the issues may be both organizational and technical in nature.

Prior research on data quality issues within the construction sector has studied data quality issues in a model relating to accuracy, consistency and completeness. Solihin, Eastman and Lee (2015) identify requirements for good quality for BIM in an Industry Foundation Class (IFC) format. Building on this, Lee, Solihin and Eastman (2019) present a semantic rule-checking process to ensure data quality pertaining to semantic and syntactic accuracy is maintained while BIM in an IFC format is exchanged. In another study, the quality of the information in the design phase is assessed using a structured and quantifiable process based on a BIM quality assurance (Donato, Lo Turco and Bocconcino 2017). Mirarchi and Pavan (2019) analyze the data quality issues concerning accuracy, consistency, and completeness dimensions of the BIM models created during the design. For facility management, Zadeh *et*

al. (2017) propose a framework to assess the quality of BIM, focusing on the completeness dimension. These studies advance the knowledge of data quality issues associated with construction information stored in BIM models. However, construction information is not stored as BIM models alone. Instead, construction information remains distributed in paper documents and digital files too. A limitation of these studies is that they do not document the data quality issues emerging due to the wider practice of using document-based information sharing in the sector.

To identify the causes of data quality issues emerging due to such document-based and model-based information-sharing practices and their implications, there is a need to study information flow across project teams in detail. By better characterizing the practice, such empirical work can then inform future technical developments (e.g. Hartmann, 2008) and address challenges raised in prior work in areas such as automated scheduling (e.g. Han and Golparvar-Fard, 2017). Previous research on the use of BIM in organizations articulates antecedents to BIM uptake (Taylor, 2007) and identifies organizational issues which affect BIM implementation, such as organizational divisions (Dossick & Neff, 2010). It describes how practices are always 'hybrid', overlaying a range of old and new media and processes (Whyte, 2011; Harty & Whyte, 2010), with the roles of construction professionals also evolving (Akintola, Venkatachalam & Root, 2017; Jaradat, Whyte & Luck, 2013; Sebastian, 2011). Such work draws attention to the organizational factors associated with information use, whereby technological integration cannot be assumed to foster closer collaboration across companies (Dossick & Neff, 2010). This literature, which understands BIM use as a complex socialized activity (Cao, Li & Wang, 2014), provides an approach that can be used to study the codification challenges and data quality issues which emerge in leading practice due to document-based and model-based information-sharing practices.

2.3. Modelling and validating construction process information

The previous section discussed the characteristics of good quality dataset and reviewed prior research investigating data quality issues in construction information. This section discusses the state-of-the-art in modelling detailing construction information, including constraints and then validating it using BIM based constraint checking methods. Constraints are important in look-ahead planning as it is critical to remove constraints during the look-ahead planning phase to make sure the tasks assigned in the six-week window can be executed as discussed previously in Subsection 2.1.1. Below, Subsection 2.3.1 presents the prior research on modelling construction process information, and Subsection 2.3.2 presents the different methods to validate detailed construction process information

2.3.1. Modelling detailed construction process information

Modelling detailed construction information into a machine-readable format allows opportunities for the application of machine learning methods for data-driven decision-support (Han, Cline & Golparvar-Fard, 2015; Belsky, Sacks & Brilakis, 2016; Dong *et al.*, 2013a). However, the current codification of construction process information is limited to precedence relationships in the form of master plans and phase plans in the 4D BIM and other scheduling software systems (Han, Cline & Golparvar-Fard, 2015). The model breakdown structures in these knowledge bases do not match the work break down structures for construction site activities (Han & Golparvar-Fard, 2017) and contain limited relationships between activities and resources (Giretti *et al.*, 2012). To address the lack of detail and semantics in the information models, the construction process needs to be represented in detail with the semantics embedded in the information model.

2.3.1.1. *Modelling construction process information*

To provide decision-support for look-ahead planning, detailed construction process information can be modeled by different methods. Table 2.2 presents the prior research in the area. Darwiche *et al.* (1988) have proposed OARPLAN to embed construction planning knowledge into an information model by incorporating construction precedence logic and hierarchical planning. The planning knowledge in OARPLAN is represented as constraints based on activity constituents and their interrelationships. Activity constituents include objects, actions, and resources. This work is a great first step to converting planning information into a machine-readable representation using a generic framework. A limitation of this method is that it doesn't consider construction workspaces (the work areas required for construction activity to be executed). Thabet and Beliveau (1997) and Akinci *et al.* (2002) have addressed this limitation and proposed methods to model information to include workspaces and link them to construction activity model such as the one mentioned in Darwiche, Levitt and Hayes-Roth (1988). While Thabet and Beliveau (1997) present a specific model representing the information model for a multi-storey building, Akinci *et al.* (2002) can be applied to generic construction cases. Dong *et al.* (2013) has extended the construction methods model (Akinci *et al.*, 2002) and combined it with intelligent knowledge base system (Marshall, Barber & Boardman, 1987) to model processes in LAS so that attributes related to the organization (present in the intelligent knowledge base system) can be incorporated into the model. Dong *et al.* (2013) model also consider a sequencing model to label the activities as TO-DO, CAN-DO, DOING and ENDED which is favourable for decision-support systems for look-ahead planning, as the activity status can also be inferred from the construction process information. A limitation of this information model is that it is specific to finishing phase of construction (RIBA Stage 5).

Table 2.2: Methods to model construction process information

Work	Method	Advantage	Disadvantage
Darwiche, Levitt and Hayes-Roth, 1988	OARPLAN method	Hierarchical planning and precedence relation codification	Construction workspaces are not considered
Thabet and Beliveau, 1997	SCaRC method	Considers construction logic, resource constraints and workspace constraints	Construction methods are not considered
Akinci *et al.*, 2002	Extended construction method model	Considers the construction method for the activity	The process model of construction is not considered
Marshall, Barber and Boardman, 1987	Intelligent knowledge-based system	Divides model to product, process, and organization	Does not cater to construction domain
Dong *et al.*, 2013	ALASG method	Combines an extended construction model and Intelligent knowledge-based system	Specific to finishing phase of construction projects

Different methods of modelling the construction process information was described in the previous paragraph. These models provide a structure for representing construction knowledge. However, they assume that the data can be stored and accessed from a single database or a model. However, data remains distributed in multiple heterogeneous datasets (Arayici *et al.*, 2018; Tchouanguem Djuedja *et al.*, 2019) over multiple servers, creating problems of interoperability and data silos. Researchers have used linked-data methods to address such problems associated with data silos in construction datasets (Pauwels, Zhang & Lee, 2017).

2.3.1.2. *Addressing data silos using linked-data*

Researchers have attempted to use linked data to address the limitations of BIM, related to structuring and querying product information (Liu, Lu & Al-Hussein, 2016; Beetz, Van Leeuwen & De Vries, 2009; Zhang & Beetz, 2016a), facility management (Kim *et al.*, 2018; Terkaj, Schneider & Pauwels, 2017), and solving interoperability (Pauwels, De Meyer & Van Campenhout, 2011). Pauwels *et al.* (2011) combined IFC and semantic web technologies to develop a semantic rule-checking environment for BIM models. Pauwels *et al.* (2017) surveyed the application of linked data in the construction sector and found that researchers have used linked data to improve interoperability, to link across domains, and to generate logical proofs and inference, thus demonstrating the use of linked-data for connecting construction

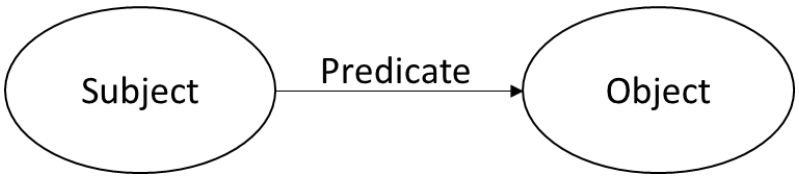


Figure 2.1: RDF triple

information.

Linked-data is a method for publishing data in a structured format so that the data can be connected and then interpreted by machines (Bizer, Heath & Berners-Lee, 2009). Following the linked-data method, information artefacts are modeled as *concepts* and are related to associated *properties* in the form of a *statement*. Each statement has three resources (*subject, predicate,* and *object*) modeled as a directed labelled graph (Figure 2.1). This graph structure is referred to as a Resource Description Framework (RDF) graph, which is formed by applying the logical AND operator to a series of information artefacts modeled as directed labelled graphs. Each statement in an RDF graph is referred to as an RDF triple (Pauwels, Zhang & Lee, 2017). The semantics of an RDF graph can be further improved using RDF vocabularies

and ontologies. The RDF schema vocabulary is one such vocabulary that enables the specification of classes, subclasses, data types, labels, etc. (RDF Schema 1.1, 2014.). OWL is a more expressive way to improve the semantic structure of an RDF graph (OWL Web Ontology Language Overview, 2012) and has RDF concepts as a subset. The graphs constructed using OWL are termed OWL ontologies, which can be used as vocabulary when constructing more complex RDF statements. Thus, RDF resources can be defined under a specific class in an OWL ontology, ensuring that new information can be inferred through a reasoning process. Publishing the construction process information following linked-data principles as an RDF graph ensures semantic interoperability. Semantic interoperability is achieved through the use of ontologies to define taxonomies and services to enable sharing and reusing domain knowledge (Davies, Fensel & Van Harmelen, 2003; Ei-Diraby, Lima & Feis, 2005).

2.3.2. **Validating construction process information using rule-checking**

The previous subsection described how detailed construction process information can be modeled using different methods and how previous research has used linked-data to address the issue of data silos which might occur while modelling construction information. This subsection discusses the different concepts for validating the construction information against constraints to support look ahead planning. Business Rule Management System (BRMS) are discussed first (Subtopic 2.3.2.1) as it provides a system to embed decision logic into an expert system to support decision making. Discussion on BRMS is followed by a discussion on BIM based rule-checking (Subtopic 2.3.2.2) reviewing how the rule-checking systems are combined with BIM. These discussions provide an understanding of how construction logic, such as construction constraints can be modeled and validated using an expert system.

2.3.2.1. *Business rule management systems*

Business rule management system is a system used to define, deploy, execute, and monitor the results of the rule-checking process (Fish, & Taylor, 2012). BRMS is a sub-domain of knowledge-based systems in the field of Artificial Intelligence. Knowledge in a BRMS consists of logic and data. Logic refers to the rules that need to be satisfied for the rule-checking process to happen. Data refer to the particular model that needs to be checked against the rules defined in the logic. Data are case-specific, whereas logic is generalizable and can be used across cases. Therefore, by separating logic from data, it is possible to manipulate both independently, as people who model data may or may not be the ones who define the rules. The advantages of using a BRMS instead of a custom hardcoded rule-checking software (Bali, 2009) are:

- Separating the data from logic enables their independent manipulation
- Rule languages are easier to understand than plain programming languages. Rule language enables domain users who are not programmers to declare rules.
- Inference machines are not locked to a specific data structure in a BRMS. Hence, efficient solving algorithms could be used.

To achieve the above advantages, BRMS should have essentially two components embedded in it: query languages and rule languages Query languages help the BRMS to retrieve data from the knowledge bases/ information models, whereas rule language is used to define custom rules against which the retrieved data is evaluated (Schwabe, Teizer & König, 2019).

a) Query languages

Query languages are an integral part of BRMS to query the knowledge base for facts that need to be checked against the rules. Query languages such as Structured Query Language (SQL) (ISO, 2016) have easy to understand syntax so that domain experts, who are not

necessarily programmers, can retrieve data from the databases. Query languages would enable the domain experts to formulate queries to create rules on the data. This relies on the principle that inherent knowledge can be exhaustively described in the information models/databases. In contrast, with other information models/databases, semantic knowledge about the data is not stored in the data structure of BIM models. Instead, it needs to be derived or calculated using semantic enrichment methods (Sacks, Ma *et al.*, 2017). Therefore, when rule-checking is done on information stored as BIM models, conventional BRMS cannot be used. Instead, there is a need for dedicated query languages which has extra functions to derive/calculate knowledge from the information model.

To address the challenge of the lack of semantic knowledge in BIM, there has been work on query languages for BIM as well as semantic enrichment of BIM models. Mazairac and Beetz (2013) develop a domain-specific query language BIMQL to select update and extract information from BIM models stored as IFC format. BIMQL enabled the extraction of partial models for rule checking. BIMQL still requires the knowledge of the underlying data structure to perform meaningful queries. Researchers have also used a different approach to enrich stored BIM data with additional information by performing operations on the existing data to derive more information (Belsky, Sacks and Brilakis, 2016; Sacks, Ma *et al.*, 2017). Although this method brings more knowledge into the information model, it still requires the domain expert to know the data structure to prescribe rules. Wülfing *et al.* (2014) introduced a visual BIM query language to ease data access through visual elements. However, this query language lacks semantic querying capability as it is designed to query for retrieving data and not essentially for rule checking. Zhang and Beetz (2016) address this issue through the use of linked-data to semantically query information from BIM models stored as RDF graphs. This method shows the potential for deriving semantic relationships within the data for rule checking. Zhang and Beetz (2015) use the concepts of linked-data to check and validate

information stored in IFC files. Their work presents the potential of using linked data to derive concepts which are not present in the existing model for rule checking. They do this by converting the native IFC to ifcOWL and then deriving the knowledge by traversing over the data represented in a graph structure (Gómez-Romero *et al.*, 2015). Therefore, from this review, the best approach for a query language for rule checking is to a have a visual interface to access the data and using linked-data to enrich and derive information out of the existing model.

b) Rule languages

Rule languages are the component in BRMS which checks the data retrieved through querying with a set of rules defined by the users. Rule languages can be structured into two parts: condition and consequence. In the condition part, the knowledge retrieved using query languages is evaluated against a set of rules. If the evaluation is a success, the consequence is executed (Schwabe, Teizer & König, 2019). There are rule languages developed to cater to BIM, such as BIM rule language (BIMRL) (Dimyadi *et al.*, 2016a). BIMRL was demonstrated in the context of compliance checking. However, this method has not been used outside of compliance checking, such as to check for constraint violations. Similarly, a domain-specific rule language, Building Environment Rule Analysis was developed to automate building design reviews (Lee, Eastman & Lee, 2015). This language has focused on the design verification and concentrates on the space-based information stored in BIM and less on the semantic construction information. Although these domain-specific rule languages assume the information to be in a specific model and stored in a central repository, in construction practice, it is distributed over multiple domains (Wang *et al.*, 2016). Hence the rule languages used in construction should have the flexibility to work together with query languages capable of retrieving data from multiple sources such as linked-data based query languages. Therefore,

for applications such as a rule checking for look-ahead planning, there needs to be an investigation of rule languages built on top of semantic web technologies.

Some of the rule languages that are already available are Semantic Web Rule Language (SWRL)(World Wide Web Consortium, 2004), Rule Interchange Format (RIF) (RIF, 2005) and the Notation3 Logic (N3Logic) language (Berners-Lee *et al.*, 2008). These languages facilitate rule-based reasoning. In the current PhD, there is a particular interest in constraint checking and less attention to other aspects of reasoning, such as deriving new knowledge. Specifically, the focus is on identifying constraint violations in construction schedules and thus, on validating the data about a construction process information against a set of constraints (Wang *et al.*, 2016). SPARQL inferencing notation (SPIN) (World Wide Web Consortium (W3C), 2011) is a SPARQL-based constraint language which offers a way to do constraint checking with closed world semantics leveraging the power of SPARQL ASK, and SPARQL CONSTRUCT queries. It links class definitions with SPARQL queries to capture constraints that formalize the behaviour of the classes. Shapes Constraint Language (SHACL), a data modelling language to model such constraints recommended by W3C (SHACL, 2017), is a successor to SPIN. Like SPIN, SHACL is a language for validating RDF graphs against a set of conditions (Knoblauch, Allemang and Steyskal, 2017). However, SHACL offers greater flexibility than SPIN in terms of defining the targets for constraints: While SPIN is limited to classes, SHACL can be applied to classes or set of nodes defined by the different target mechanisms, including custom targets using SHACL advanced features. Hence, SHACL can be used as an effective language for modelling construction information and constraints as it offers the power of linked-data and the flexibility in validating target constraints.

The use of SHACL to define construction scheduling constraints means that a dedicated constraint-checking language does not need to be developed from scratch. Validating an RDF model with a set of constraints using SHACL addresses the inherent issue of the open-world

assumption in RDF graphs, as SHACL offers the opportunity to define structural constraints (Ekaputra & Xiashuo Lin, 2016). Moreover, SHACL has two components:

- *Data graph*: contains the data to be validated.
- *Shape graph:* contains the constraints against which resources in the data graph are validated (called shapes).

This follows the BRMS principle of dividing knowledge into logic and data. Data graphs and shape graphs may be coded as a single RDF graph or multiple RDF graphs because the triples in the data graph are validated against triples in the shape graph. There are two types of shapes in SHACL: the node shape and property shape.

- *Node shapes* are constraints that act on subject resources (or concept instances) of a specific type (the type can be defined while modelling the constraints) in the data graph.
- *Property shapes* are constraints that act on the predicate resources (or attribute instances).

An advanced feature of SHACL, SHACL SPARQL contains all the features mentioned of SHACL and offers the expressive power of SPARQL to target specific instances of a *subject* resource satisfying certain conditions (flexible in defining the conditions) (Knublauch, Allemang & Steyskal, 2017). When a *data graph* is validated against a *shape graph*, the resources in the *data graph* are checked against the applicable shapes in the *shape graph*. Then, a validation report (in the form of another RDF graph) is produced, which details the resources that have passed or violated the constraints. As mentioned in Eastman *et al.* (2009), this is similar to how a rule-checking system produces a report as its output. The SHACL-based constraint checking is ideal for validating construction scheduling constraints (which are distributed across multiple servers), as SHACL uses the power of linked-data to traverse across constraints distributed across heterogeneous data sources and to validate them.

2.3.2.2. *BIM based rule-checking*

Researchers have worked on integrating the BRMS concepts with BIM to perform rule checking on BIM (Lee, Solihin & Eastman, 2019; Pauwels *et al.*, 2011; Dimyadi *et al.*, 2016b; Solihin *et al.*, 2017).This section discusses how BRMS is used in conjunction with BIM to perform rule checking. BIM based rule-checking systems ensures that the data (detailed construction process information) in BIM models conforms to the rules/constraints modeled in the BIM based rule-checking system.

a) BIM based rule-checking concepts

Prior research on BIM based rule-checking has presented four classes of functionality for rule-checking as well as introducing different complexities of rule-checking systems. Eastman *et al.* (2009, p. 1016) provide four classes of functionality to perform BIM based rule-checking. They are 1) *Rule interpretation:* Textual rules are converted into a machine-interpretable format; 2) *Building model preparation:* Information in the form of models is converted into a format required for the execution of the rules; 3) *Rule execution:* The prepared model is checked against the defined machine-interpretable rules; and 4) *Reporting:* The results of the rule-checking process are communicated to the user or raised to another system.

Each of these steps may be fully automated or semi-automated. Solihin *et al.* (2017) described methods to automate each of the four steps. To automate rule interpretation, this approach suggests distilling the rule content from written text using tools (e.g. LegalRuleML (OASIS LegalRuleML, 2018), RASE (requirement, applicability, selection, and exception) markup (Hjelseth & Nisbet, 2010)), transcribing the rule content into a rule language (e.g. Drools (Red Hat Inc, 2019)). To automate building model preparation, the existing data models are transformed into the required knowledge representation using converters (e.g. the conversion of IFC to ifcOWL to process rules in the SWRL (World Wide Web Consortium, 2004)). Rule execution is automated using BRMS. The results of rule-checking are taken from the BRMS

and reported to the user after analysis. These approaches to automated BIM based rule-checking is of relevance to the current research problem aiming to develop constraint checking systems to validate construction scheduling constraints.

Depending on how inferences are made from data, BIM based rule-checking systems can be classified into four classes according to its complexity (Solihin & Eastman, 2015). The complexity is regarded as Class 1 when the rule-checking system queries entities and attributes against a value. For Class 2 complexity, the rule-checking system needs to calculate/derive additional values for rule execution. In rule-checking systems classified as Class 3, the data structure of the information model needs to be extended to perform rule execution. Finally, new model objects are created from the existing data sets to perform rule execution in rule-checking systems with Class 4 complexity. Thus, to perform constraint checking on distributed construction data, a rule-checking system with a complexity of Class 3 is needed as the data structure needs to be extended to link data distributed in multiple servers.

b) Prior research on BIM based rule-checking

There is increasing interest in BIM based rule-checking to increase model quality and to provide decision-support in the construction sector. Existing proposals use different methods, such as using rule engines (Tan, Hammad & Fazio, 2010), query languages (Nawari, 2012), rule languages (Lee, Eastman & Lee, 2015) and machine learning methods (Bloch & Sacks, 2018) for BIM based rule-checking. Prior research in this area has extended knowledge on using BIM based rule-checking for ensuring code compliance (Martins & Monteiro, 2013) and construction safety (Zhang *et al.*, 2013), and for aiding design verification (Lee, Eastman & Lee, 2015) and construction-site planning (Schwabe *et al.*, 2016). These studies have assumed that construction information is available as a BIM model in a single repository. However, this assumption may not hold true in the case of construction information. Therefore, the BIM based rule-checking methods focusing on the construction stage should be able to

check for violation in data distributed over multiple databases. To address the issue of construction information siloes, attention is now turned to the prior research on the use of linked-data to model the knowledge required for rule-checking by exploiting the potential of linked-data to link across domains to create semantic information models (Zhang, Boukamp & Teizer, 2015; Zhong *et al.*, 2012).

2.4. Automating LAS generation

Previous sections in this chapter have introduced the data quality issues in construction as well as methods to model and validate construction process information. This section synthesises the prior work on scheduling and details the need of a new scheduling algorithm for LAS generation.

2.4.1. **Scheduling**

Scheduling is the decision making process of assigning resources to a set of tasks over a period to optimize one or more objectives (Pinedo, 2008). Early work on scheduling focused on machine scheduling problems (Chen, Potts & Woeginger, 1998). Machine scheduling problems are classified according to sources of demand, number of machines, complexity of the production system, performance index, characteristics of a production environment, processing characteristics of the operation and resource constraints (Lin *et al.*, 2012; Graves, 1981). Most common types of machine scheduling problems are flow-shop problems (Minella, Ruiz & Ciavotta, 2008) and job-shop problems (Zhang *et al.*, 2017a).

The algorithms for solving scheduling problems are divided into exact optimisation and approximate methods (Table 2.3). The exact optimisation algorithms include efficient rule approaches, mathematical programming approaches, branch definition methods, and etc. However, exact optimization methods can solve small scale problems only. Although, branch and bound methods can obtain the optimal solution in theory, it is difficult to have real practical

applications as it takes polynomial time to solve it (Chen, Potts & Woeginger, 1998).Therefore, exact optimization methods are not recommended for large scale problems as they cannot complete calculations in a good responding time. With the advancement of computing powers and intelligent algorithms, the research methods related to scheduling have gradually changed from exact optimization methods to the approximation methods (Rabadi, 2016).

Table 2.3: Classification of algorithms for solving scheduling problems

Classification			Example
Exact method	Efficient rule approach		(Johnson, 1954)
	Mathematical program approach		(Wagner, 1959)
	Branch and bound method		(Lomnicki, 1965)
Approximate method	Constructive method	Priority dispatch rule	(Barker & McMahon, 1985)
		Insert algorithm	(Nawaz, Enscore & Ham, 1983; Lian & Mesghouni, 2014)
		Bottleneck based heuristics	(Liu & Kozan, 2012; Balas & Vazacopoulos, 1998)
	Artificial intelligence methods	Constraint satisfaction techniques	(Jain & Meeran, 1998)
		Neural network	(Akyol & Bayhan, 2007)
		Expert system and knowledge based methods	(Ju, 2007)
		Fuzzy logic	(Canbolat & Gundogar, 2004)
	Meta-heuristic methods	Genetic algorithm	(Zhang *et al.*, 2017c)
		Simulated annealing	(Güçdemir & Selim, 2017)
		Tabu search	(Glover, 1986; Li & Gao, 2016)
		Ant colony algorithm	(Blum & Sampels, 2004; Huang & Yu, 2017)
		Particle swarm optimization	(Xia & Wu, 2005; Teekeng *et al.*, 2016)
		Differential evolution algorithm	(Storn & Price, 1997; Zhao *et al.*, 2016)
		Firefly algorithm	(Łukasik & Żak, 2009; Khadwilard *et al.*, 2011)

LAS generation problem falls under the class of project scheduling problems and it is concerned with small batch production where limited resources are allocated to interdependent

activities over time. Project scheduling problems varies from the classic machine scheduling problems in the following ways.

1. In machine scheduling problem, resource consumption of tasks and capacities of resources (machines) are unitary. Whereas in a project scheduling problem, resource consumption of tasks and capacities of resources may not be unitary. Thus, two tasks may share the resources.
2. In a machine scheduling problem, one task (operation) requires only one resource (machine) to be executed. Whereas in a project scheduling problem, one task may require several resources simultaneously to be executed.
3. Precedence constraints in machine scheduling are generally linear whereas the precedence constraints in project scheduling can form any acyclic graphs.

Over the last two decades, there has been numerous studies looking at solving project scheduling using Resource-Constrained Project Scheduling Problem(RCPSP) (Karam & Lazarova-Molnar, 2013). RCPSP and its generalization encompasses most of the complex machine scheduling problems as special cases (Brucker *et al.*, 1999). Look-ahead Schedule generation under resource constraints can be classified as a resource-constrained project scheduling problem (RCPSP)(Dong, 2012).

2.4.2. **Resource-Constrained Project Scheduling Problem**

RCPSP is one of the most studied optimization problems in project scheduling. It consists of a single project which consists of n activities that are scheduled subject to precedence and resource constraints, to reduce the makespan (i.e. minimize the time required to complete all the activities) (Schwindt & Zimmermann, 2015). An RCPSP problem consists of a series of activities $V= \{1, 2,..j\}$ with precedence relations (finish to start) among them with zero-time lag. For each activity $i \in V$, d_i represents the duration of the activity and F_i represents the finish time of the activity. Also, r_{ik} represents the number of the renewable resource of type $k \in K$ assigned to the activity $i \in V$ for a unit time. The number of resources of type k available at each unit time period is R_k. The parameters d_i, r_{ik} and R_k are all positive integers.

These activities and resources are related by precedence and discrete resource constraints (Christofides, Alvarez-Valdes & Tamarit, 1987).

Precedence constraints guarantee that each activity $i \in V$ starts only after all the predecessor activities ($h \in P_i$) have finished, where P_i is the set of predecessor activities for the activity i. This constraint is denoted by the following equation:

$$F_h \leq F_i - d_i, i = 2, \ldots j; h \in P_i$$

Here F_h is the finish time of the predecessor activity h, F_i denoted the finish time of the activity i, d_i represent the duration of activity I, and j represent the index of the last activity.

Discrete resource constraints ensure that the total amount of resource k required by all activities $A(t)$, $A(t)=i|\ i \in V$ at a time period t do not exceed R_k. This is represented by the equation:

$$\sum_{i \in A(t)} r_{ik} \leq R_k, k \in K, t \geq 0$$

Since, finishing a project in the shortest possible time and incurring the least possible cost represents the success of a project, the objective of RCPSPs is to minimise the makespan of the project Fj (finish time of the last activity j) and minimise the cost of the project.

In the case of look-ahead planning, the project can be composed by multiple sub-projects, as there are independent workspaces within a project that draws from the same pool of resources. Each workspace is relatively independent of other workspaces in terms of activities; however, they are dependent on the use of resources and interdependent workspaces. These problems come under a subset of RCPSP called as Resource-Constrained Multi-Project Scheduling Problem (RCMPSP) (Chiu & Tsai, 2002).

The RCPSP has been classified as an NP-hard problem (a problem is of complexity NP-hard if an algorithm designed for solving it can be translated into one for solving any

nondeterministic polynomial time problem)(Brcic, Kalpic & Fertalj, 2012). Due to its combinatorial nature, it is difficult to be solved using exact methods, such as constraint programming, linear programming and branch and bound methods, when the number of activities is more than 50 (Agarwal, Colak & Erenguc, 2015). Therefore, heuristics and metaheuristics algorithms are used to obtain near-optimal solutions. A heuristic algorithm is a problem-specific algorithm which aims to generate a valid solution (though the optimal solution is not guaranteed) in a reasonable amount of time. Metaheuristics are problem-independent algorithms that can be applied to generate a solution. In this section, various heuristics, and metaheuristics for the generation of RCPSP are reviewed.

2.4.2.1. *Heuristics*

There are different heuristics to solve RCPSP based on priority rules such as minimum latest finish time, longest finish time, etc. Priority rules are the rules which decide what activity happens next at each step resulting in a unique activity list. An activity list is a list of activities in a precedence feasible order. The order is determined by the chosen priority rule (Kolisch, 1996a). Given an activity list, a schedule is generated using a schedule generation scheme (Kolisch, 1996b). The normal heuristic methods have lesser complexity when compared to metaheuristics; however, they are not optimal (Kolisch & Hartmann, 1999). To reduce optimality gaps, there are several improvement techniques proposed in the literature such as schedule generation schemes , forward and backward scheduling and double justification.

a) Schedule generation scheme

A schedule generation scheme is used to develop an activity list into a schedule. There are two types of schedule generation schemes. They are serial schedule generation scheme and parallel schedule generation scheme. (Kolisch, 1996b).

- *Serial schedule generation scheme:* Serial schedule generation scheme is an activity-oriented schedule generation scheme. Activities in the activity list are scheduled at their earliest start time in this schedule generation scheme.
- *Parallel schedule generation scheme:* In a parallel schedule generation scheme, the clock is maintained at each unit of clock time and set of activities that are precedence and resource feasible are determined. If the set contains one activity, that activity is scheduled. Then the clock is incremented to the next time step. If there is more than one activity in the set, priority is given to the activity which occurs first in the activity list.

These two schedule generation schemes may produce different schedules. Dong (2012) has used the serial generation scheme to generate a LAS. However, there are no studies which used parallel schedule generation scheme to solve the LAS generation problem.

b) Forward and backward scheduling

Li and Willis (1992) introduced forward and backward scheduling in conjunction with serial and parallel schedule generation scheme to generate multiple schedules. An RCPSP Graph *G= (V, E)* where e *V* represents nodes and *E* represents arcs, or precedence constraints can be calculated as a forward or a backward scheduling problem depending on where the calculation is started. In the backward scheduling problem, the last activity node in the graph is taken as the first node for the calculation and vice versa. Using the same schedule generation scheme and same priority rules, forward and backward scheduling produces different schedules of different makespan (Özdamar & Ulusoy, 1996).

c) Double justification scheme

After a schedule is created using a schedule generation scheme, there may be empty spaces in the schedule. Valls *et al.* (2005) introduced a double justification technique to remove these gaps in the Gantt chart to improve the optimality of the solution. This technique is

performed by scanning the Gantt chart created from the schedule from right to left and shifting the activities to the right to fill empty spaces. This procedure is called backward justification. Once this is done, a similar process is performed by scanning the Gantt chart from left to right, and then shifting the activities from right to left to fill the empty spaces. This double procedure is called forward-backwards improvement or double justification (Valls, Ballestín & Quintanilla, 2006).

2.4.2.2. *Metaheuristics*

Metaheuristics are problem-independent methods to solve optimization problems. Several metaheuristic algorithms have been applied to solve RCPSP (Munlin, 2018; Servranckx & Vanhoucke, 2019; Das & Acharyya, 2011; Liu *et al.*, 2020; Myszkowski *et al.*, 2015). Metaheuristics can further improve the optimality of the solutions after performing heuristic methods such as forward and backward scheduling and double justification. Many of the metaheuristics have been inspired by processes observed in nature. These algorithms generate multiple solutions randomly within the search space aiming at finding a near-optimal solution (Agarwal, Colak & Erenguc, 2015). Different metaheuristics use different strategies to produce new solutions, in such a way that better solutions emerge as the search progresses. The search in every metaheuristic aims at doing two tasks. The first task is to converge the search for the best solution in a neighbourhood (local search, exploitation). The second task is to steer the search to the best neighbourhoods in the overall search space (global search, exploration). In addition to these, strategies should be developed so that the search is not stuck in a non-optimal neighbourhood (Ratajczak-Ropel & Skakovski, 2018). Therefore, every metaheuristic should have an element for local search, global search, and strategies to avoid getting stuck in local neighbourhoods. The following subsections describe different metaheuristics used for solving RCPSP.

a) Genetic algorithms

Genetic algorithms, proposed by Goldberg and Holland (1988), is one of the most-used metaheuristics to solve RCPSP. It is a population-based metaheuristic that uses the survival-of-the-fittest policy in evolution to converge towards an optimal solution. GA produced improved sets of the solution after every generation, by the process of selection, crossover and mutation. When GA is applied to RCPSP, activity list acts as an individual. From a given population of activity lists, a new population is produced through the reproduction process using crossover and mutation. Diversification of the population is achieved through the crossover mechanisms. Intensification is achieved through the use of selection strategy in GA. To promote escape from a non-optimal neighbourhood, a mutation operator is used.

Several variations of GA have been used to solve RCPSP by varying the 1) choice of varying chromosomes. 2) crossover mechanism and 3) mutation strategies (Hartmann, 2015; Gonçalves, De Mendes & Resende, 2015; Ortiz-Pimiento & Diaz-Serna, 2018). *Scatter search* is a variant of GA with a heavy emphasis on diversification. To ensure diversity, the population will be divided into a set of high-quality solutions and a set of diverse solutions. Scatter search approach has been applied to RCPSP and has produced the best results.in the RCPSP literature. (Mahdi Mobini *et al.*, 2009). *Electromagnetism like* search is another variation of GA inspired by Coulomb's law which states that force between particles is inversely proportional to the square of the distance between them (Debels *et al.*, 2006). In the RCPSP, each activity list is considered as a particle with a certain charge. The charge is the function of the objective function. Each particle exerts a force (attraction or repulsion) on other particles. The principle behind this algorithm is that inferior particles will move in a direction such that it gets attracted to the superior particles. Debels *et al.* (2006) have detailed how charges are defined, and forces are calculated.

b) Simulated annealing

Simulated annealing (SA) is a metaheuristic that has been inspired by the process of annealing used in metallurgy for hardening metals (Kirkpatrick, Gelatt & Vecchi, 1983). In SA, a current solution is maintained at all times. The current solution is evaluated with neighbourhood solutions. If a neighbourhood solution is better than the current solution, the current solution is replaced with the neighbourhood solution. To escape a local neighbourhood, the current solution may be replaced by a worse solution based on probability $\pi = e^{(-\Delta/Temp)}$. If π is less than a random number $u \in [0,1]$, a worse solution is accepted. Δ is the difference between the value of the objective function in the current solution and new solution. *Temp* is a parameter used to control the acceptance probability. The higher the value of *Temp*, the greater the probability of acceptance is. A higher value of temp increases diversification as worse solutions is accepted (van Laarhoven & Aarts, 1987). In a good SA, the algorithm should incorporate strategies to improve the local neighbourhood as well as to escape the neighbourhood to evaluate diverse neighbourhoods (Agarwal, Colak & Erenguc, 2015). Boctor (1996) and Cho and Kim (1997) have applied simulated annealing to RCPSP. Their observation states that SA is stronger for local search but not effective for global search. Hence, if the solution space is smaller in the LAS generation, SA can be used.

c) Tabu search

Tabu search (Glover, 1986) is a metaheuristic which employs the intelligent use of memory to exploit past experience. Memory is a list of previously visited solutions. There are three lists used in Tabu search. The first one is a short-term Tabu list. It stored the recently visited locations. This list avoids further search in a local neighbourhood and improves diversification. The second list maintains poor solutions. This list is used to direct the search away from neighbourhoods with poor solutions. The third list holds a good solution. This list directs the search for optimal neighbourhoods (Glover & Laguna, 1998). Neighbourhood

search is performed in a way similar to SA. Researchers have applied Tabu search to RCPSP problem (Nonobe & Ibaraki, 2002; Thomas & Salhi, 1998; Poppenborg & Knust, 2016; Skowroński *et al.*, 2013). Their solutions suggest that the RCPSP does not produce results as good as GAs. However, when implemented efficiently, Tabu search performs better than Simulated annealing (Reeves, 1993).

d) Ant colony optimization

Ant colony optimization (ACO) metaheuristic is inspired by the behaviour of ants discharging chemical substance pheromone along its path between its colony and the location of food to guide other ants (Dorigo & Colorni, 1996). The stronger the concentration of pheromone along a path means that more ants have passed through that path depicting a higher probability of getting food if this path is followed. In an RCPSP, the ant selects an activity order. The ant uses heuristic information and pheromone information to determine whether an activity can be placed at a particular position in the activity lists. Heuristics follows the idea of using a priority rule, whereas pheromone uses the information of the past experience of ants using a path to find a good solution (Dorigo & Caro, 1999). Escaping a bad neighbourhood is achieved by choosing a worse solution after accepting a predetermined number of good solutions. Merkle *et al.* (2002) proposed an ant-colony approach to the RCPSP. This method requires a good amount of computer memory as it tests out multiple paths before arriving at a solution. Also, this method is good for local search and less effective for global search (Agarwal, Colak & Erenguc, 2015).

e) Particle swarm optimization

Particle swarm optimization (PSO) is inspired by how bees find flower patches for honey, and this algorithm was proposed by Kennedy & Eberhart (1995). It is different from ACO, as the flying pattern of bees is dependent on the direct flight route rather than a dedicated route as in ACO. In a bee colony, scout bees are sent in a different direction to search for flower

patches. The scout bees returned to bee colony and communicated the information about the distance, direction, and quality of the flower patch back to the bee colony. Follower bees follow the scout bees, which has reported good quality flower patches (Zhang, Wang & Ji, 2015). In the PSO algorithm, a random initial set of solutions are created, and these serve as scout bees. Each solution is then evaluated for quality. Scout bees which reported good quality solutions are chosen for local neighbourhood search. Researchers have reported the use of PSO for RCPSP (Fahmy, Hassan & Bassioni, 2014; Nasiri, 2012; Chen, 2011; Jia & Seo, 2013). Their results suggest that PSO is more suitable for global search and less effective for local search.

f) Reinforcement learning

Finally, Reinforcement Learning is an area of machine learning, inspired by how learning takes place in nature, wherein the actions taken by an entity are dependent on the outcomes obtained from previous actions. Positive rewards would reinforce the behaviour or give extra weight to that particular action and the actions leading to it and vice versa. RL is built on top of Markov Decision Processes (Sutton and Barto 2018). RL differs from supervised learning since it does not require labelled input-output pairs for learning to happen (Kaelbling, Littman, and Moore 1996). RL has been used to solve RCPSP (Jędrzejowicz & Ratajczak-Ropel, 2014). In RCPSP, ach action in RL determines what activity needs to be started next. If the activities lead to a good solution, a higher reward is obtained. RL has shown the ability to incorporate complex constraints in addition to the precedence and resource constraints in RCPSP, making it an ideal candidate for LAS generation.

2.4.3. **Limits of existing research**

This section has detailed different heuristics and metaheuristics in the literature looking at solving RCPSP, a class of problems of which LAS is a subset of. Unfortunately, these methods cannot be directly applied to the LAS generation problem. In addition to the precedence and discrete resource capacity constraints that are handled within RCPSP, LAS

problems includes more construction site constraints such as disjunctive constraints (for example blocking constraints such as limited accessibility) (Dong *et al.*, 2013b) and logical constraints (For example Matching resource capability with requirements) . In addition, one of the assumptions in RCPSP is that multiple activities cannot share the resources at the same time (Ratajczak-Ropel & Skakovski, 2018). However, at times in construction, resources such as cranes and dump trucks may be shared between multiple activities. Hence the LAS generation algorithm should be capable of processing these complex constraints. Furthermore, unlike the common project networks in RCPSP, the LAS problem is characterized by a small number of highly interdependent activities (Dong *et al.*, 2013a). Thus, the metaheuristics need to be adapted for solving LAS problems.

Dong *et al.* (2013) have developed a method to generate LAS for the finishing phase of construction (RIBA Stage 5) automatically. They developed an information modelling approach and LAS generation approach and implemented it on a six-room case. In addition to the precedence and discrete constraints typically found in construction programmes, they introduced blocking and zone constraints to accommodate activities that need to be performed in parallel and ones which cannot be performed in parallel. Although their study shows positive results, this study is specific to finishing phases of building projects and cannot be generalized as the constraints are explicitly designed to cater to building construction.

2.5. Summary: Research issues to be addressed

This chapter reviewed the prior work on codification challenges limiting LAS generation, modelling and validating detail construction process information for automated LAS generation and finally methods to generate LAS automatically. Here these results are summarised and synthesized to identify the research issues to be addressed. Section 2.1

introduced the topics related to the main aim such as constraints management, look-ahead planning and data-driven methods to support look ahead planning. Section 2.2, 2.3 and 2.4 reviewed the relevant literature related to objectives O1, O2 and O3, respectively.

Section 2.2 discussed the existing research on data quality issues in the construction sector. This section detailed the characteristics of a good quality machine-readable dataset (Subsections 2.2.1 and 2.2.2). Further, the need for ensuring data quality was also established in this section (Subsection 2.2.3). Although researchers have studied the data quality issues within the sector, the focus of their study was limited to model-based information sharing. However, the literature has suggested that there is a combination of model-based and file-based information sharing (Subsection 2.2.4). The influence of these practices (a combination of the use of file-based and model-based information sharing) on data quality and how these issues emerge is an area that is less explored in the prior literature. Data quality issues emerging due to the broader practice of using document-based and model-based information sharing is critical for the current study because good quality data is necessary to perform data-driven decision-support for look-ahead planning. This limitation in the current literature has led to the first objective of the PhD research (objective O1), which is addressed in chapter 4.

Section 2.3 discussed state-of-the-art in modelling detailing construction information (Subsection 2.3.1) and then validating it using BIM based constraint checking methods (Subsection 2.3.2). The existing research on BIM based rule-checking assumed that the construction information remains stored as BIM models (Subtopic 2.3.2.2). The current state-of-the-art in BIM based rule-checking cannot handle this data silos in construction information. Although researchers have used the linked-data method to address the data silos in construction information (Subtopic2.3.1.2), this has not been applied in BIM based rule checking to validate construction process information distributed in multiple databases. This limitation has led to the second objective of this PhD (O2), which is addressed in Chapter 5.

Section 2.4 presents the different methods to solve the RCPSP problem, which is a superset of the LAS generation problem. Different heuristics (Subsection 2.4.2.1) and metaheuristics (Subsection 2.4.2.2) used for solving the RCPSP problem are presented. Although there are established methods to solve for RCPSP, these methods cannot be directly applied to the LAS generation problem. LAS problem deviates from the conventional RCPSP problem as the LAS has high interdependency between activities and complex constraint relationship between them (Subsection 2.4.3). The present methods have not discussed how to incorporate complex construction constraints into the RCPSP problem. This has led to the third objective of the PhD study (objective O3), which is addressed in Chapter 6.

3. RESEARCH DESIGN

3.1. Introduction

This chapter presents the research methods that were followed to achieve the objectives stated in Section 1.2. In the fields of construction engineering and management, there is a tradition of research to develop technologies that support decision-making (Morkos, 2014; Dong, 2012; Sacks *et al.*, 2013). Researchers have done so using a method comprising of six steps: 1) identification of the problem, 2) definition of the objectives of the solution, 3) development of the solution, 4) demonstration of the solution, 5) evaluation of a prototype, and 6) communication of the results (Hinkka & Tatila, 2013; Sacks *et al.*, 2017b; Ben-Alon & Sacks, 2017; Arayici *et al.*, 2018; Hevner *et al.*, 2004). This thesis follows these steps. First, problem identification is performed through a review of existing literature (Chapter 2) and through an interpretive case study (O1: Chapter 4). Then, the objectives for the solution development was defined on the basis of the identified problem and the knowledge of what is possible and feasible (Section 5.1 and Section 6.1). Following the objective definition, a prototype was developed to partially address the problems identified, and the developed prototype was demonstrated using a public dataset (O2: Chapter 5 and O3: Chapter 6). This prototype was then evaluated using real-world construction data (Chapter 7). Conclusions from this research, its implications and the directions for future research are presented (Chapter 8), constituting to the communication.

The previous paragraph described how the overall research is carried out. Table 3.1 presents the methods used to achieve each of the objectives of this research. The following sections explain the research method adopted for each objective in detail.

Table 3.1: Research methods to address objectives

Aim	Objectives	Method	Approach to ensuring research quality/validity
To develop an information modelling approach to codify and validate detailed construction process information in BIM-enabled projects, in order to support data-driven decision-making for look-ahead planning in construction	O1) To understand the codification challenges in the late-design and construction phase of BIM-enabled projects	M1) Interpretive case study research (Section 3.2)	Theoretical sampling (Subsection 3.2.1) Member checks for the interview summary and emergent findings (Subsection 3.2.2)
	O2) To develop an information modelling approach to codify and validate construction process information including dynamic constraint information	M2) Prototype development (Section 3.3)	Prototype testing using a public dataset (Subsection 5.3.2) Prototype testing using real-world data (Section 3.5 and Subsection 7.4.1)
	O3) To develop a method to generate look-ahead-schedule for construction projects automatically using the codified process information.	M3) Machine learning model development and testing (Section 3.4)	Accuracy check of model predictions for a test dataset using manual and automated methods. (Section 6.5) Accuracy check for model predictions using real-world data and comparison with real-world prediction (Section 3.5 and Subsection 7.4.2)

3.2. Identifying codification challenges in construction (M1)

This section presents the research design for achieving objective O1, which seeks to understand the codification challenges in the late-design and construction phases of BIM-enabled projects. As described in Section 2.4.3, this objective addresses the limitation of existing research by documenting the data quality issues in construction.

An interpretive case study approach (Eisenhardt & Graebner, 2007) was chosen for this study. Case studies offer a framework for in-depth investigation of the emergence of data quality issues within the construction projects through the analysis of multiple sources of data (documents, the output of software tools, interviews observations etc.). This study follows the tradition of research that uses a case study approach to study aspects of digital construction (Hartmann & Fischer, 2007; Dossick & Neff, 2010).In the current work, the focus of the case study is to identify the codification challenges causing data quality issues in the construction sector. The starting assumption of this study is that there are data quality issues in construction information due to the way digital tools and workflows are used in late design and construction stages. However, the questions about how and why these data qualities emerge are investigated through case studies.

3.2.1. **Case selection**

Approach to case selection follows theoretical sampling as it provides opportunities to illuminate and extend relationships and logic among constructs (Eisenhardt & Graebner, 2007). To understand the challenges of codification in construction, three construction projects are studied qualitatively. These projects were selected as cases as they follow a level of digital collaboration categorized as BIM Level 2. The study focuses on investigating the digital tools and workflows used in the projects, structured and unstructured information flows in these projects, and the problems associated with the information flow. These three construction projects are a multi-storey residential student apartment block in the United Kingdom (*Student apartment*), a metro rail infrastructure project in India (*Metro project*), and a major water infrastructure megaproject in the United Kingdom (*Water project*). The *Student apartment* is an exemplary project exhibiting the use of BIM in the United Kingdom, constructed by a leading contractor, and using state-of-the-art offsite manufacturing approaches in construction. The *Metro project* is pioneering BIM implementation amongst the metro projects in India,

incorporating learning on digital implementation from global megaprojects. The *Water project* is one of the biggest construction projects in the United Kingdom, using innovative technological solutions to futureproof construction and deliver a physical asset as well as a digital asset for operation.

BIM level 2 was mandated in *Water project* and seen as best practice in the *Metro project* and *Student apartment* (BSI, 2018). Although the *Metro project* is constructed in a country (India) which does not have a regulatory framework for BIM level 2, the owner required the adoption of BIM level 2 as international best practice, which justifies the choice for selecting the case. Information sharing across project teams in the late design and construction phases of the three projects is studied qualitatively by visiting the projects, analyzing internal and publicly available documents, observing meetings, and conducting informal and formal interviews on the use of product and process information during the construction stage.

3.2.2. **Data collection**

Within each project, there was an initial setup meeting to present the study and identify interviewees. The interview protocol covered questions of communication, software tools used, BIM, collaboration, and information flow. The data analysis phase overlapped with the data collection phase. The taped interviews were transcribed, and field notes were typed up. These were read and reread between the project meetings. Summaries of interviews were sent back to the interviewees for member checks. All the data was organized into cases and stored into the qualitative analysis software. These methods draw on a qualitative case study approach (Eisenhardt and Graebner, 2007), building insights across the three cases from multiple sources (site visits, documents, field notes, and interview transcripts) (Hartmann & Fischer, 2007; Dossick & Neff, 2010). The *Student apartment* was investigated first, followed by the *Metro project* and finally, the *Water project.* The findings from the *Student apartment* informed me

of the codification problems, and this information was used to direct the data collection and analysis of the *Metro project* and the *Water project*. The complete details of the data collection are provided in Appendix A, while data collection performed for each case is given below.

- *Student apartment*: To study this case, I visited the construction site and offices of the projects, had informal conversations with the digital and planning engineers, examined construction documents and models, and studied the software tools used to understand the information embedded in the BIMs, construction schedules, and other reports such as design calculation, and method statements. These documents were centrally hosted on a CDE, and I had access to it while being at the office.
- *Metro project*: To examine this case, documents such as the BIM execution plan, presentation documents for training, and press releases were studied to understand digital information management practices. The project manager, chief site engineer, casting yard engineer, and BIM consultants, who form a cohort of the key decision-makers during the construction stage, were interviewed informally to get an insight into the extent of codification in the information flow during their daily work practices. Field notes were taken during the interviews. In addition to these interviews, the casting yard, a viaduct construction site, and a station site were visited to understand the on-ground practice of various activities. Further insight into this case was obtained through a workshop with 40 participants, including client representatives of six major Indian metro-rail projects along with technology providers and delivery teams. The workshop provided a perspective on the digitization of this project in the broader landscape of Indian metro rail construction.
- *Water project*: To understand the codification challenges, in this case, eight semi-structured interviews were conducted. All eight interviewees had more than ten years of experience in the construction sector and had worked with different major

projects in the United Kingdom and abroad. The interviewees' areas of expertise covered design, planning, project engineering, digital engineering, prefabricated construction, and information management. All interviewees had teams working with them on their areas of specialization and also interacted with the other stakeholders in the project. These characteristics make them ideal for the study as the interviewees were highly knowledgeable, and they viewed the information sharing issues from diverse perspectives. Following a semi-structured approach ensured participants would talk broadly on their experiences with information flow using digital collaboration tools. This allowed the collection of data regarding constructs, while unexplored issues could be explored. Seven out of the eight interviews were recorded, and transcripts were made from the recordings. In addition to the interviews, I conducted multiple visits over two weeks to the project office, observing meetings and the work practice. I also had access to the CDE and several documents such as a construction programme, LAS, and method statements.

3.2.3. **Data analysis**

A starting assumption of this study is that there are data quality issues caused by the way digital tools and workflows are used in late design and construction stages. Building on this assumption, the data analyses were conducted to identify the codification challenges arising because of the different digital workflows and working practices across projects and understand the implication of these codification challenges on the data quality. Drawing on the six data quality dimensions-Accuracy, completeness, temporality, consistency, accessibility and data provenance (Batini & Scannapieco, 2016) and five levels of machine readability—one star to five stars (Bizer, Heath & Berners-Lee, 2009; Berners-Lee, 2006) explained in Section 2.2, coding was done on the collected data to identify different issues related to

codification. NVivo qualitative data analysis software was used to track the patterns emerging from these data. The analyses steps are as shown in Figure 3.1

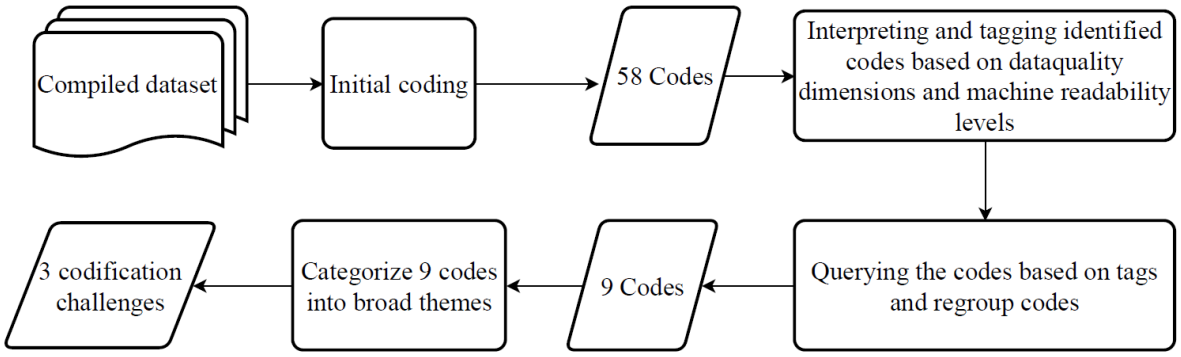


Figure 3.1: Data analysis steps

As mentioned earlier, the collected data was imported into a software package for analysis. The data was read, re-read. Initial coding of the dataset was done using open coding, an approach to identifying key ideas that reoccur across the dataset.

> "We're actually going to **export** that out of the CDE, we're going to **print** it out, we're going to staple it together, we're going to put our own cover sheet on the front of it, with the exact same details on the back and we're going to go off and go and get **three signatures**, **scan** it back in, put it back into [CDE1] and submit it." (Water project, Project engineer, C3I1)

For example, the quote shown above was coded[1] as 'export', 'printing and scanning' and 'signatures'. These codes were stored as nodes in the software. Similarly, the whole dataset was coded. The resultant codes were then reconsidered, and similarly, named codes were merged and renamed to a unique name. For example, 'wrong-modelling technique' and 'different-modelling techniques' were combined into 'multiple-modelling technique'. During

[1] In this chapter, coding is the process of analyzing qualitative text data by taking them apart to see what they yield before putting the data back together in a meaningful way. This is different from codification used in rest of the thesis. In the thesis, codification is defined as the process of conversion of information into a structured and machine-readable format to support the application of data science.

this initial coding, 58 codes were identified. The data set was queried for the identified codes using the software, which resulted in further coding of the dataset that was missed during the manual coding. The automatically coded portions of the dataset using the queries were then manually checked to ensure that the code matched the dataset. If not, the coding was removed from that portion of the data set.

Once the whole dataset was coded using the 58 codes, then each of these codes was interpreted based on the data quality dimensions and levels of machine readability. For example, the code 'printing and scanning' were tagged with level 1 machine-readability. It is categorized as level 1 as the document is available on the CDE and accessible as a scanned image in PDF document. However, scanned document as an image stored in PDF format does not have relevant information regarding the content in the image (e.g. an image of a table) in a structured format. Similarly, printing and scanning would result in the loss of metadata from the document and its origin information. Thus, the data provenance dimension of the piece of information is affected and hence tagged to printing and scanning. These tags were stored as another set of nodes in the software.

After the codes from the initial coding were interpreted based on the data quality dimensions and machine readability, the analysis software NVivo was used to query the tags (machine readability levels and data quality dimensions). This resulted in some of the codes (saved as nodes in the software) occurring together in the dataset. For example, the codes 'export' and 'loss of data' came together in multiple instances when queried for level zero machine readability. Thus, whenever these two instances came together, they were grouped into a new code 'loss of information during conversions. Similarly, querying for data provenance resulted in 'export' and 'multiple CDE' in different instances. These were grouped into the code 'lack of process change'. By following multiple combinations of queries, nine such codes were identified, which are interoperability, information loss during conversions,

multiple modelling techniques, unstructured information sharing, drawing and file-based sharing, document control bottlenecks, lack of process change, loss of constraints, and low level of detail.

These nine codes were then categorised into broad themes based on the functionality from which they emerged. For example, interoperability and loss of information during conversions were always related to the use of the software. Hence those were categorised under software usage. Similarly, file-based sharing and document control bottlenecks emerged because of the process used for sharing the information. Hence, they were categorized under information sharing. Such categorization resulted in three codification challenges: software usage, information sharing and missing construction process information. The nine codes were then compared across the cases (As mentioned in Subsection 3.2.2, the data collected from each project was stored as a case in the software). Reasons for the occurrence of certain codes in certain cases were also investigated. The results of the data analyses are presented in Section 4.2.

3.3. Codifying and validating construction process information using linked-data (M2)

This section presents the research design for achieving O2, to develop an information modelling approach to codify and validate construction process information, including dynamic constraint information. As described in Subsection 2.4.3, this objective is framed to develop an information modelling approach to model detailed construction information and validate it using BIM based constraint checking systems, taking it to account that the construction information remains fragmented across multiple sources.

Following the research approach followed by researchers working on ICT to support decision-making in construction (Schwabe, Teizer & König, 2019; Dave *et al.*, 2016a; Sacks

et al., 2013, 2017b), a prototype web application demonstrating the proposed solution was developed to achieve this objective. This is done through the development of a linked-data based constraint checking approach towards modelling and validating detailed construction process information and developing a prototype web application to implement the developed approach. Modelling of detailed construction process information, including construction scheduling constraints, is described in Section 5.2. The development of the prototype web application implementing the approach is described in Section 5.3.1.

After its development, the prototype needs to be evaluated to understand its capability to solve the problems (Dave *et al.*, 2016a; Sacks *et al.*, 2017b; Peffers *et al.*, 2007). Worked example method has been used by researchers to demonstrate research prototypes (Clayton, Kunz & Fischer, 1998; Schwabe *et al.*, 2016). This method involves describing a plausible scenario using the developed method in order to demonstrate that the ideas being promulgated are applicable. This method was followed to evaluate the prototype developed in this study. Following this method, the developed prototype was tested using an enriched public dataset (Van Berlo, 2018). The testing was designed to check whether the developed approach could identify constraints from construction information distributed over multiple sources. Accuracy of detection of constraint violation was used as a metric for demonstration. The results of the demonstration are presented in Section 5.3.2.

3.4. Automating LAS generation (M3)

This section presents the research design for achieving objective O3 to develop a method to generate LAS for construction projects automatically using the codified process information. As described in Section 2.4.3, this objective is framed to develop a method to extend the RCPSP problem to incorporate complex construction constraints using the method developed to address O2.

Following the research approaches towards solving of RCPSP problems (Lin, Zhu & Gao, 2020; Jędrzejowicz & Ratajczak-Ropel, 2014; Dong, 2012), a machine learning model was developed to generate LAS using the codified information. Different machine learning methods for solving RCPSP, such as GA, SA, Tabu search, ACO, PSO and RL were presented in Subsection 2.4.2. Of these metaheuristics, GA and RL were chosen for the current study. GA was chosen as it has been widely acclaimed for solving RCPSP problem (Elbeltagi, Hegazy & Grierson, 2005; Faghihi *et al.*, 2015). Also, GA has been used to solve the LAS generation problem for finishing phase of residential buildings (Dong, 2012). This makes GA an ideal candidate metaheuristic for the current study. Similarly, recent research has shown RL to be effective in solving highly constrained optimization problems (Koulinas *et al.*, 2018; Ratajczak-Ropel, 2018; Shi *et al.*, 2020). Reasons listed above make GA and RL the candidate algorithms for testing the use of codified construction process information to generate LAS. Modelling the LAS problems using GA and RL to incorporate the method developed in Chapter 5 is described in Section 6.3 and Section 6.4, respectively.

Two different strategies were used for testing the selected candidate algorithm for comparing their performance to generate LAS. These testing scenarios were so chosen to emulate the differences between classic RCPSP and LAS generation problem. Influence of the number of the resource type is selected as a scenario for testing because the LAS generation differs from the conventional RCPSP problems as there are multiple shared resources. Influence of the number of processes and activities are used as a scenario for testing the performance as LAS has multiple interdependent processes and activities under them when compared to conventional RCPSP problem (Dong *et al.*, 2013a). The interdependencies are codified as construction scheduling constraints, such as precedence constraints, discrete resource constraints, disjunctive constraints (Morkos, 2014) and logical constraints.

The ability of the developed algorithm to generate LAS was demonstrated using a computational experiment. The computational experiments analyzed the ability of the developed algorithms to converge to a solution. The metrics used to analyze the performance are the minimum predicted project duration and the time required to converge to that solution. The computational experiment compared how these algorithms behaved under varying conditions. Conditions included an increasing number of activities in the project and an increasing number of resource types in the project. The computational experiment evaluating the performance of these algorithms to solve the LAS problem and the results of the experiment are presented in Section 6.5.

3.5. Implementation study

This section presents the research design for the evaluation of the developed prototype to support look-ahead planning. The implementation study tests the applicability of the method developed for codifying and validating constraints (presented in Chapter 5) and the methods developed for using the codified information to generate LAS (presented in Chapter 6) on a real construction project information.

In this study, a worked example method (Clayton, Kunz & Fischer, 1998; Sacks *et al.*, 2017b; Schwabe, Teizer & König, 2019) is used for evaluating the prototype. A worked example describes a simulated scenario to test an innovative process to show the features of the innovative process. However, worked example method can also be used to test an innovative process replicating a real scenario. When testing the innovative process in a real-life situation, the worked example provides stronger evidence for the validity of the research than a simulated problem. It also offers the opportunity to compare the innovative process with the conventional process.

For the case selection, purposive sampling (Creswell, 2007) was used to identify a construction project, which uses 4D BIM intensively. Out of the three projects studied to identify codification challenge (M1), a construction site belonging to the water project was chosen for the current study This project was one among the finalists in the Bentley Year In Infrastructure 2019 and is leading in terms of the use of 4D BIM for look-ahead planning. These characteristics made it a suitable site for testing the implementation, as it provides an opportunity to compare the innovative process with state-of-the-art in practice. Although this site was part of the water project, it does not belong to the same section of water project studied in method M1. Snowball sampling is used to recruit participants for this study. Initially, contact was made with the digital engineer of the construction project. Based on the discussions with the digital engineer, remaining participants who would be included in the weekly meetings were added to the study.

In the 4D BIM review meeting of the project, I acted as a silent observer and captured the information regarding construction scheduling constraints and site constraints. 4D BIM models, method statements and construction programme used for these meetings were collected from the construction project for analysis. 4D BIM was then imported to the developed prototype (described in Section5.2). The information from the 4D BIM was enriched with the data captured during the review meeting using the prototype. After the data is enriched with all the information, constraint violations were identified using the prototype. The identified violations were discussed with the digital engineer to understand the problems with the developed prototype.

To test the LAS generation method, schedule data (start-time and end-time of activities) were removed from the prototype, and the prototype was used to generate a LAS for the week. The generated plan was then compared with the plan created by the engineers in the project. Variances between generated plan and the one created by engineers were studied based on

metrics such as time efficiency, and identification of constraint violation. The generated plan and the observed variances were presented to the engineers in projects and feedback on the limitations of the framework was collected from the digital engineer.

This study tests the efficacy of the developed prototype on a case and identifies the opportunities and limitations of using the developed prototype. However, this study does not claim that the developed prototype can replace the existing robust processes. The worked example has been argued to be subjective (Clayton, Kunz & Fischer, 1998), as the researcher who advocates the solution, collect the evidence for the study. To eliminate this threat to validity, the analysis involved a comparison of the outputs produced by the prototype with the outputs created by the practitioners. Thus, there is evidence from the systems that provides a basis for the comparison. Also, the practitioners were requested to comment on, and discuss the variations between the outputs produced by the prototype and the LAS developed by the practitioner. This comparison brings the practitioner's perspective of the prototype. This study was conducted to test the efficacy of the prototype to solve the research issues and does not focus particularly on the ease of use of the prototype.

3.6. Summary

The overall research aim is achieved using different research methods, as summarized in Section 3.1. The aim of the research is further subdivided into three objectives, and each of these objectives is achieved using a different method. The first objective (O1), to understand the codification challenges in construction, is achieved through an interpretive multiple case study approach. The second objective (O2), to develop an information modelling method to model detail construction process information and validate them), is achieved through prototype development and demonstration method. The third objective (O3), to reuse the construction process information to generate a LAS, is achieved through the development of a machine learning model and a computational experiment to demonstrate its ability to generate

LAS. The prototype developed for O2 and objective O3 was evaluated using an implementation study. The implementation study is performed using a worked example on real data. The trustworthiness and the authenticity of the research are ensured using different strategies. Using multiple methods in the research design ensures that the disadvantages of using one method are mitigated. This chapter has presented the different research methods to achieve each one of the objectives. The findings of the research for the objectives are presented in the subsequent chapters.

4. IDENTIFYING CODIFICATION CHALLENGES IN CONSTRUCTION[2]

4.1. Introduction

Machine-readable information enables new forms of data science, including machine learning and data analytics. These methods offer value to the construction sector through resource optimization, data-driven design, prescriptive analytics for rule-checking, visual analytics, performance predictions, operational analytics, and more (Bilal *et al.*, 2016). In the context of data-driven decision-support for look-ahead planning, these methods offer methods to identify constraints and generate LAS. However, the construction sector has not been able to take complete advantage of these developments, and data science is not widely deployed due to poor data quality (Bilal *et al.*, 2016).

This chapter contributes to objective O1 of the thesis, and it attempts to identify the codification challenges, and reasons for poor data quality, in the construction sector by qualitatively studying late design and construction stages of three construction projects. Section 2.2. provided a background for this chapter, including data quality dimensions (Batini & Scannapieco, 2016), machine readability (Berners-Lee, 2006; Bizer, Heath & Berners-Lee, 2009), and the prior research in this area. Further, the justification for identification of the codification challenges was explained in Section 1.3 and Section 2.5. Through the identification of codification challenges, this chapter informs the technology development in

[2] Parts of this chapter have been published as *Soman, R.K. & Whyte, J.K. (2020) Codification Challenges for Data Science in Construction. Journal of Construction Engineering and Management. [Online] 146 (7), 04020072.* and is available from*: doi:10.1061/(ASCE)CO.1943-7862.0001846.*

the later chapters of this thesis, about data quality issues in the sector and why these data qualities occur.

In the context of this thesis, codification is defined as the process of converting information into a structured and machine-readable format to enable data science (refer to Section 2.2.2), and codification challenges are the factors that limit the codification of information in the construction sector. The codification challenges are identified by examining the information sharing across project teams that use BIM in the late design and construction phase on three projects 1) Student apartment, 2) Metro project, and 3) Water project. The data collection and analysis(Eisenhardt & Graebner, 2007). The data collection and analysis steps used in this study are described in Section 3.2. The data analysis led to the identification of nine codification challenges (interoperability, information loss during conversions, multiple modelling techniques, unstructured information sharing, drawing and file-based sharing, document control bottlenecks, lack of process change, loss of constraints, and low level of detail). These codification challenges are further categorised into three themes. (software usage, information sharing and missing construction process information).

The rest of this chapter is divided into four sections (as illustrated in Figure 4.1) to present the codification challenges. An overview of the three cases studied (Student apartment,

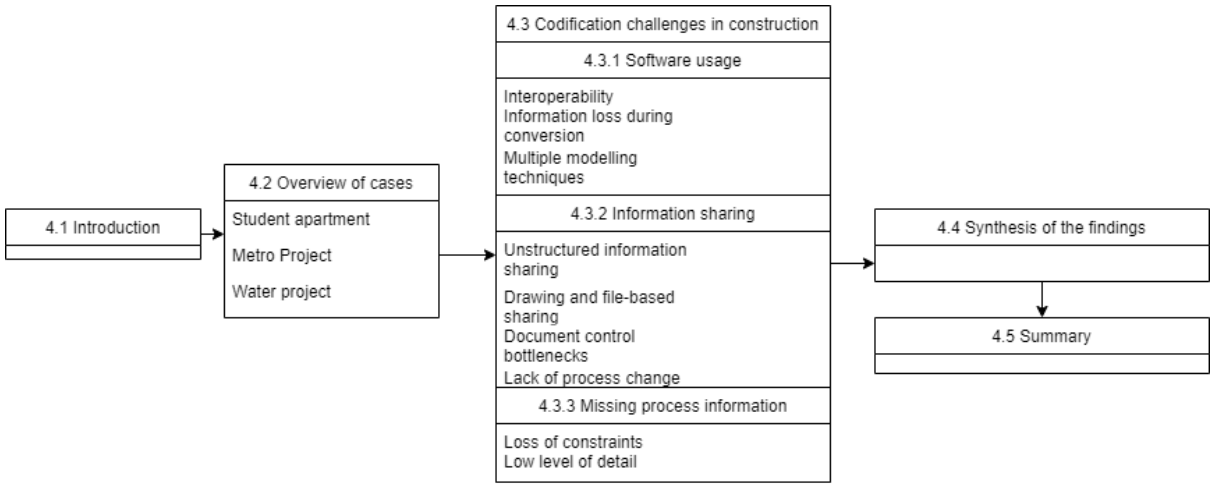


Figure 4.1: Chapter overview

Metro project and Water project) is presented in Section 4.2. Then, the identification of codification challenges in these cases is presented in Section 4.3. Following the identification of the codification challenges, a synthesis of the findings is presented in section 4.4. Finally, a summary of the chapter is presented in Section 4.5.

4.2. Overview of the cases

To identify challenges of codification in construction, three construction projects are studied qualitatively to investigate the digital tools and workflows used in the projects, structured and unstructured information flows in these projects, and the problems associated with the information flow. The three projects studied are 1) *Student apartment,* 2) *Metro project,* and 3) *Water project*. The project details and the digital strategy of these projects are provided in Table 4.1.

Table 4.1: Details of projects studied

Case	Project characteristics	Digital information use
Student apartment	*Cost:* >$ 75 Million *Location:* United Kingdom *Project start and end:* 2016-2018 *Project type:* Residential apartment • 34-storey residential tower to house 192 apartments. • The construction of this building used the Design for Manufacturing and Assembly (DfMA) philosophy and digital engineering to deliver and assemble structural steelwork, modules, risers, plant, and bathroom pods efficiently. • Different divisions did the design and the construction in the same firm.	• The engineers and managers in the project were exposed and trained in digital engineering through the prior projects. • CDE had inbuilt workflows for approval and publishing of drawings, models, and other pieces of construction information. • Different teams within the contractor's design division created the architectural, structural and Mechanical Engineering and Plumbing (MEP) models. • Models were aggregated and linked to the construction programme by the digital engineering team at the construction side to identify clashes. • Time sliced 4D models were used in the planning and progress meetings.

Metro project	*Cost:* >$ 1.2 Billion *Location:* India *Project start and end:* 2015-2020 *Project type:* Urban transit infrastructure • Urban rapid transit system consisting of 2 Corridors, with more than 40 km of elevated viaducts and 40 stations. • Megaproject with more than 190 sub-contractors. • Separate firms did design and construction. • Uses digital information management across the lifecycle of the project.	• Engineers and managers in the project were not exposed and trained in digital engineering. • A specialist team coordinated the digital project management and curated the models submitted by the supply chain. • Customized 5D BIM platform was developed for project management and visual project monitoring integrating the benefits of an Enterprise Resource Planning (ERP) software with construction design authoring and review software and project management software. • Design review and approval process are carried out in a CDE managed by document controllers.
Water project	*Cost:* >$ 6 Billion *Location:* United Kingdom *Project start and end:* 2016-2024 *Project type:* Wastewater infrastructure • Major sewage infrastructure megaproject in the United Kingdom using innovative technological solutions. • The project involves the construction of an underground sewer with a length of 25km, and associated infrastructure. The construction of the project was divided into three sections, each section handled by different joint ventures of leading sub-contractors in the country.	• The engineers and the managers in the project had prior experience in working with BIM-enabled project. • The digital strategy comprises of model-based design reviews. • Complete engineering data is stored and managed within CDEs, which utilizes BIM Level 2 workflows and has inbuilt approval and publishing processes. • CDE provided 1,200 users from more than 45 stakeholders to manage more than 190,000 documents, as well as 4,500 drawings and more than 1,500 models. • The design was handled through a design consultancy who created the BIM models and drawings which was later used by construction contractors.

I studied the information sharing across project teams in the late design and construction phases of the three projects by visiting the projects, analyzing internal and publicly available

documents, observing meetings, and conducting informal and formal interviews on the use of product and process information during the construction stage.

4.3. Codification challenges in construction

The data collected from the cases were compiled and analyzed to find the codification challenges in construction. (See also section 3.2). Table 4.2 gives a high-level overview of the dataset and the codes categories used for analysis. The occurrences of these issues are supported with evidence from the dataset in the remaining part of this section.

Table 4.2: Code book for codification challenges used in the analysis

Codification challenge	**Code**	**Example**
Software Usage	*Interoperability* Issues associated with the inability of computer programs/tools to read the output created by other computer programs/tools	"During conversion of files from the submitted file formats to the native file formats of Bentley, there were issues of distortion" (*Metro Project, Request for proposal, C2D5*) "I am using this MicroStation, but sometimes they were drafted from the client, let's say, in Autodesk. And transferring things from Autodesk to MicroStation, you lose data" (*Water project, Technical manager, C3I5*)
	Loss of information during conversions An issue associated with loss of data when the output of one computer program/tool is converted into a different format in order to process further with another computer program/tool	"The drawings are converted into PDF when uploaded to Asite" (*Student apartment, Digital engineer – Field notes, C1I1*) "There's a facility in [CDE1] that when you upload a PDF, you can also upload a secondary file" (*Water project, Information manager, C3I4*)
	Multiple modelling techniques Issues associated with different techniques to model the same information	"if they use the wrong tool to model something, you don't have the appropriate dataset to it " (*Water project, Digital engineer, C3I3*)
Information sharing issues	*Unstructured information sharing* Issues associated with sharing of information over channels that are not indexed and tracked such as telephone, meetings, email, face to face conversation	"Progress tracking is easily done over WhatsApp. We get updates and photos immediately" (*Metro project, Field notes – Project manager, C2I1*) "Then emails, meetings, usually types of communication." (*Water project, Project engineer, C3I1*)

	Drawing and file-based sharing Issues associated with the sharing of information as files instead of model-based collaboration	"Lead reviewer assigns internal reviewers depending on type of drawing" (*Metro project, Digital Project Management (PPT), C2D2*) "Because I have to open up every drawing individually and print them all, or even if you do it the other way around, it's very slow anyway.". (*Water project, Project engineer, C3I1*)
	Document control bottleneck Issues associated with the inaccessibility of information due to long processing time to publish the information	"there's certain people who are responsible for issuing information [...] high volume of these authorizations to do that it takes them a lot of time to get through" (*Water project, Project planner, C3I2*)
	Lack of process change Issues associated with using conventional workflows in place of digital workflows	"the new technology had to be implemented under circumstances where MS Excel, leave alone MS Project, was used predominantly for planning" (*Metro project, Report, C2D4*) "We're actually going to export that out of the CDE, we're going to print it out, we're going to staple it together, we're going to put our own cover sheet on the front of it, with the exact same details on the back" (*Water project, Project engineer, C3I1*)
Missing process information	*Low level of detail* Issues associated with lack of detail in the models for data-driven decision-support	"The schedules did not contain the required level of details and attributes and thereby could not be linked to the 3D model" (*Metro project, Report, C2D4)*
	Loss of Constraint Issues associated with inaccessibility to constraint information in schedules limiting data-driven decision-support	"we have to interpret engineering information, whether it be drawing or written constraints, written narratives and interpret those into a Gantt chart" *(Water project, Project planner, C3I2)*

Different codes observed in the cases have been categorized into three main codification challenges in this study.

1) *Software usage*: relating to challenges arising from the use of these software tools

2) *Information sharing*: relating to the practices of sharing of information

3) *Missing process information*: relating to issues of the codification in construction process information.

These findings are unpacked, and the rationale is explained in the following subsections. Each subsection provides evidence for the codification challenges, describes its occurrence across

the cases, its implication on data quality and machine readability, and situates the findings within the wider literature.

4.3.1. Software usage

This section discusses the codification challenges related to software usages such as interoperability, information loss during conversions, and multiple modelling techniques during the late design and construction phase. It also explains how these challenges are associated with data quality dimensions accuracy, completeness, accessibility, and data provenance.

4.3.1.1. *Interoperability*

Data interoperability is a significant problem when it comes to data quality and machine readability. If the data is locked to a proprietary format, it limits the application of data science. Software vendors provide a proprietary Application Program Interface (API) to access the data. However, access to the data through APIs is limited, and information access is limited to proprietary domain-specific programs. This issue was observed in the Metro project and Water project but at different scales. Uneven distribution of tools would result in issues of data interoperability.

In the Metro project, the request for proposal document which details how the BIM implementation is intended requires the submission to ensure openness and not have vendor locked submissions.

> "The system will need to interface with multiple disparate systems from external partners, agencies and contractors for bi-directional information flow. Openness comes from the use of open standards and using." (Metro Project, Request for proposal, C2D5)

Despite having taken measures to ensure submission in open formats, there are issues with interoperability which requires the data to be converted to be used by the client.

> "During conversion of files from the submitted file formats (For example: rvt files) to the native file formats of Bentley, there were issues of distortion" (Metro Project, Report, C2D4)

Although the request for proposal stated openness of construction information, the CDE was set up by a software vendor which required the conversion of files into their proprietary format to display the information in their view. This conversion resulted in the loss of information as well as distortion of information.

Interoperability was raised as a central problem by the interviewees in the Water project, especially when working with multiple CAD tools in projects, resulting in data loss during format conversions. Even within the same software environment, there are problems related to data compatibility while working between different software versions.

> "Sometimes the drawings are in Bentley format I am using this MicroStation, but sometimes they were drafted from the client, let's say, in Autodesk. And transferring things from Autodesk to MicroStation, you lose data […]2019 will open 2018, 2017, and 2016. But when 2020 comes into play, then you cannot open it anymore with the 2019 files that are generated with the 2020" (Water project, Technical manager, C3I5)

Here, the drawings are made using different CAD tools such as AutoCAD and MicroStation. However, the data may not be opened (or edited) using the same tools with which they were created. This creates problems of interoperability and loss of information when data in one format is converted to another. Similar problems also occur when using multiple versions of the same tool.

The Student apartment and the Metro project had predominantly used tools from a single software vendor (Autodesk for the Student apartment and Bentley for the Metro project) for executing most of their tasks, resulting in better interoperability when compared to the Water project, which uses tools from different software vendors. The scale of the project influences this diverse distribution of software use. The scale of the Student apartment was smaller than that of the Metro project and Water project, with a leading firm involved in both

design and construction, resulting in evenness of software usage (Contractor office visit, Student apartment-C1S1) resulting in having better interoperability. Even though the Metro project had different firms engaged in design and construction, the information management was handled by an owner support organization, resulting in evenness in the data resulting in lesser interoperability issues (BIM consultant 1, Metro Project-C2I4). The Water project, on the other hand, had multiple firms working on different phases of the projects with their own sets of tools, resulting in issues of data interoperability at a greater scale (Digital engineer, Water Project-C2I3).

Interoperability issue reduces the machine readability of the information to 2-star data as the data is not structured using an open format. In addition, this codification challenge limits application of machine learning algorithms as different systems cannot talk to each other and derive the context from the information, thereby reducing the accessibility dimension. Although there are open-data formats available for the exchange of information, these are not often used in the construction phase and are only submitted at specific data drops. Evidence also shows that there is a loss of information when converting between formats because of inefficient exporters and importers. This problem reflects data quality related to the accessibility dimension. If this data remains inaccessible, algorithms make inferences with limited data, resulting in incomplete inferences.

4.3.1.2. *Information loss during the conversion*

Information loss during the conversion was raised as a key problem by the information manager in the Student apartment and the Water project. The CDE [Asite] used for collaboration within these projects need the primary information to be in PDF format, as the information in PDF format would ensure that information cannot be altered and remained locked. These projects required the information to be converted to PDF format to be visible in its viewer.

> "The drawings are converted into PDF when uploaded to Asite, Asite has a PDF viewer so that drawings can be viewed in the browser. It can be printed from there too" (Student apartment, Digital engineer – Field notes, C1I1)

The supporting documents in the native format may be uploaded as a supplementary file for that submission.

> "If we're conveying CAD information, it's being uploaded as a supporting file. There's a facility in [CDE1] that when you upload a PDF, you can also upload a secondary file." (Water project, Information manager, C3I4)

When structured information is converted to PDF format, they lose semantic information in the native files. For example, when a drawing is converted to a PDF file, the vector information embedded in the native file is lost.

This issue was observed in the Student apartment and the Water project. Both Student apartment and Water project used similar CDE which required the files to be converted to PDF to display the file in the CDE. The conversion from native formats to a PDF format results in the loss of semantic relationships embedded in the file. The loss of the semantic relationship between datasets means that the data structure associated with the data is lost, thereby reducing the machine readability to 1 Star data. This aspect of the loss of information results in incomplete information and lowers the data quality related to the completeness dimension. Furthermore, during the conversions, the metadata related to the original file is lost. This lowers the data quality dimension associated with data provenance. This also lowers the machine readability of the construction information as the data structure associated with the information is lost in the conversion.

4.3.1.3. *Multiple modelling techniques*

Software tools allow different methods for creation of information. However, not all methods lead to the information being reusable. The modeled information would have multiple uses, which may not be known to the creator of information, as stated by the digital engineer.

> "So, if they use the wrong tool to model something, you don't have the appropriate dataset to it [...] when you go into your authoring tool, do I use a slab tool or do I go and use a generic solid modelling tools then try and attach a dataset to it[…]if they're not, therefore we have to go in there and say, well I can't just say there's a slab now, that's just a piece of geometry" (Water project, Digital engineer, C3I3)

For example, a slab could be modeled as a generic solid model with a dataset attached to it or as a slab component. From the user's point of view, the information contained in both models is the same. However, during automated quantity take-off, the slab modeled as a generic solid would not be considered as the computer cannot classify it as a slab.

> "For tunneling purposes, when you try to extract 2D drawings from 3D BIM models, those drawings are not as correct and as detailed as they used to be. They have glitches, they have errors" (Water project, Technical manager, C3I5)

The technical manager in the Water project also stated that the 2D drawings from the BIM were less detailed than the required level for further processing. The modelling in the BIM tools resulted in glitches and errors, which affects the analysis further down the lane.

This problem was observed in the Water project with a large supply chain. Organizational divisions in large projects lead to lower machine readability and data quality because of differences in modelling approaches and software tools used. Software tools offer different approaches to model the same information at the same level of detail. Moreover, modelling approaches used by the firms are guided by the norms and practices followed in the firm. These norms may be different for the firms who use the data. Whilst examining the cases of the Student apartment and Metro project, where the information modelling is performed by

a single firm, the issues such as interoperability and improper modelling of information were limited. However, in the Water project, where the information modelling spanned over different firms, there was evidence of issues of interoperability of data and improper information modelling. Therefore, the information created could be used as a digital submission but limits further use. Even when the information exists in the model in the correct format, the level of detail of the modeled information is less than desired, making the information unfit for further use. This problem reduces the quality of the data concerning accuracy and completeness. The fact that the data exists but not in the way it was supposed to be, is a case of syntactic inaccuracy.

4.3.1.4. *Discussion*

The findings on software usage show that, despite significant digitization of work processes, data remains fragmented into different domains and formats because of the multiple software tools in use across the organizations involved in construction. Work in the construction information technology community is pioneering new data management solutions to improve interoperability (Hu *et al.*, 2016; Pauwels, De Meyer & Van Campenhout, 2011; Pazlar & Turk, 2008; Redmond *et al.*, 2012), and it is disappointing to find that construction projects still suffer from poor quality data as a result of problems of interoperability caused by the existence of multiple domain-specific tools and modelling practices. In their work, (Dossick & Neff, 2010) have previously shown how the organizational and cultural divisions between the designers and builders, contractors, and subcontractors stifle collaborative work. This study shows that these issues are not resolved. In the projects studied, organizational and cultural divisions between the firms involved in the late design and construction stages of projects cause software usage problems (interoperability, information loss during format conversion, multiple modelling techniques). Whilst there may be shared norms and tools within a firm for modelling information, these norms differ across the firms which modeled project information. Multiple

modelling techniques and data created using different software result in datasets which are not interoperable and require format conversions, resulting in loss of information and low machine readability. The Water project had multiple firms working on the data over different phases of the project, with interoperability problems more prevalent in this case in comparison with the student apartment, in which a single leading firm was involved with the creation and use of the model. Although the Metro project had different firms over the different phases of the project, digital data creation was handled through a single owner support organization, limiting the impact of this problem. Software usage problems are found to lead to challenges of codification for data science; hence, this work extends prior insights by Dossick and Neff (2010) to show how organizational and cultural divisions between designers and builders not only stifle collaborative work and joint problem-solving but also result in fragmented datasets in construction, leading to data silos and data loss and, thus, resulting in poor data quality which is more difficult to use in data science.

As it is relatively unusual and potentially undesirable to have one firm or owner with overall control of the model, to enable more distributed working, developers of new tools or digitally-enabled processes should consider the implication of organizational separation in the sector in addition to the technical requirements. To achieve better-quality data in projects, practitioners must focus beyond the individual scope of their firms towards the common goals of the project.

4.3.2. **Information sharing**

This section presents the codification challenges in construction, such as unstructured information sharing, drawing and file-based sharing, document control issues, and lack of

process change. The implications of these challenges on the data quality dimensions are presented in this section.

4.3.2.1. *Unstructured information sharing*

Even when there are structured workflows for information sharing, project participants find it easier to use the unstructured channels of communication. They often find structured information flow through the CDE slow and complicated. Despite being more traceable and accountable compared to unstructured information-exchange practices, the complexity of the new structured methods for information sharing and the poor understanding of workflows across the teams lead to the use of a combination of structured and unstructured channels for information sharing.



Figure 4.2: Federated BIM model used for discussion (Student Apartment- C1D2)

In the student apartment, although the formal submissions are communicated through a common data environment, the day to day decisions are taken through a variety of methods such as telephone calls, online meetings, interdisciplinary team meetings, emails etc. While being at the construction site office, I was a silent observer in a meeting attended by the digital engineering team, planning team and methods team. The meeting aimed to discuss a construction method and the sequence to remove a tower crane from the construction site and

do the finishing work of the part of the building where the hoist was anchored [Student apartment-C1S1]. The meeting took place in the construction site office. A federated BIM model was used to support the discussion (Figure 4.2).

> "…after every 5 floors, the spider crane used for holding the panels should be taken to the next location after it is used to support the cladding installation. Approximately, it would take half a day depending on the availability of the tower crane…" [Student apartment, Methods engineer-Field notes, C1S1]

The planner described the timeline of the events to be taken place, and the methods teams described the resources required and the constraints. After the meeting, the output of the meeting was coded into the construction programme as a sequence of activities to be constructed by the planner. The construction programme had the activity details and precedence relations between them. The methods team created a method statement. The method statement contained the resources required, constraints such as equipment constraints, safety constraints, operating procedures etc. This information was sent by email to the participants of the meeting for their feedback. Based on this information, which was sent over the email, the digital engineering team created a 4D model. The 4D model was used to create an animation which informed the site team of the construction process. Once all the participants approved the 4D model, the construction programme and the programme and method statements were published to the rest of the users to view.

In the Metro project, CDE was set up in the project for structured sharing of information to ensure that there is a single source of truth. However, the information was uploaded to the CDE through coercive pressures. This resulted in changes at the construction sites not updated in the CDE, thus resulting in incomplete information in the CDE.

> "Daily progress and labor reports were regularly uploaded only through coercive pressures. Whenever there were changes enacted in the site, the changes were not incorporated in the design management platform" (Metro project, Report, C2D4)

Although formal submissions are carried out in a CDE maintained by the client, the day to day decisions are made over unstructured channels of communication such as instant messaging, emails, telephone etc.

> "Progress tracking is easily done over WhatsApp. We get updates and photos immediately" (Field notes – Project manager, C2I1)

The project manager stated that they use instant messaging apps to get updates in the progress as they enable the sharing of media and is accessible to everyone.

In the Water project, information was shared over modes such as meetings, reports, e-mails, etc., contains relevant data for decision-making, as shown in Table 4.3. This information is embedded in documents, is shared in a human-readable format (reports, PowerPoint presentations, drawings, etc.), and is in formats which are both human and machine-readable (e.g. spreadsheet, BIM, etc.).

Table 4.3: Information sharing between teams in the water project

Media	Example evidence from the dataset
Common Data Environments (CDE1, CDE2 from different vendors)	*Formal submissions, drawing receipts, design and temporary works*: "if it's a formal document submission, we do it through [CDE1] [...] When I receive drawings from [design consultant], I get them through [CDE2]. And quite a lot of the designers use [CDE2] and[...] Well, we try get all the design functions, including temporary works to use [CDE2]." (Project engineer, C3I1)
Reports	*Spreadsheets and documents:* "There's lots of reporting on the project[...] And then that gets out into various outputs, so that could be just a schedule in Excel. Lots of Excel outputs as well, huge amount of Excel outputs. And, if it's a commercial discussion there may need to be some narrative around it, so using Microsoft Word to develop a narrative." (Project Planner lead, C3I2)
Meetings	*Design review* meetings: "You could do a design meeting, review something and then say, write comment on that[...]to understand what information, they're going to require at a particular stage. So that may consist of meetings; that might consist of face-to-face conversations, emails, etc." (Digital engineering lead, C3I3) *Buildability meetings* "attendance to buildability meetings and trying to get out of them what sort of temporary works may be needed to build something" (Project engineer, C3I1) *Client meetings*: "I will be going off-site to attend meetings with the client" (Information manager, C3I4) *Design for Manufacture and Assembly (DfMA) input*: "I'm trying to attend meetings and troubleshoot and try to help and provide technical input into the

	design and assisting the designer and support team" (Technical manager/DfMA coordinator, C3I5)
Email	*Highly used*: "So, obviously we do use emails a lot." (Project Planner lead, C3I2) *A normal type of communication*: "Then emails, meetings, usually types of communication." (Project engineer, C3I1) *Provides remote precision*: "If I'm communicating over longer distances or if I think to myself, I'd better make a precise request, then it'll be emails. We don't use a communicator-type facility in [CDE1]". (Information manager, C3I4) "Stakeholders, yes, it's certainly meetings and emails. Most of our stakeholders don't want to use [CDE1], because of the admin that comes with that" (Principal engineer, C3I8)
Remote conversations	*Online meetings*: "Generally, like a Skype, conference calls, linked meetings" (Senior digital engineer, C3I6) *Online communications and records of design logs, discussion points, online forums*: "Yes, so all the design data is held within a common data environment, which was [CDE2]. And so, I managed that area and access to that area. Then all communications were stored on SharePoint on Microsoft online, so everyone had access to registers or design logs or discussion points, almost used as an online forum where anyone could ask questions" (Senior digital engineer, C3I6) *Telephone calls with design consultants*: "So, I'll start with between us and design consultants: there's meetings, emails, and phone calls. I prefer meetings and phone calls" (Principal engineer, C3I8)

The main limitation of the information shared through these unstructured channels is its accessibility, which is limited to people involved in the meeting or e-mail conversation.

> "A guy made a design on a spreadsheet for quantities. Some people knew about it; he logged it as well. And I didn't know that at all. So, at the very end of the day, on the eleventh hour when I have finished everything, by the way, we have this spreadsheet, and it's exactly what I wanted to do" (Water project, Technical manager, C3I5)

In this case, the information required for a task was already available. However, it was not accessible for the person who needed it, who was not part of the group within which the information was shared. This led to the recreation of the information and loss of productive time.

Unstructured information sharing reduces the machine readability. Also, this challenge implies data quality, lowering it concerning the accessibility dimension since data is not available in a common repository. Instead, it is distributed in different silos and e-mail

databases, and the access is limited. In addition, it introduces inconsistency as the same information is distributed amongst different databases which are not connected or synchronized. Tracing the source of the data and its history is also difficult when using unstructured channels for information sharing, thereby reducing the quality of data associated with the provenance dimension. These issues limit the data science as the datasets for drawing inferences are siloed and disconnected. This reduces the data quality dimensions associated with consistency and data provenance

4.3.2.2. *Drawings and file-based sharing*

Even though the projects follow BIM level 2, the engineers are more comfortable performing submissions and approvals using drawings and files rather than model-based information sharing. This is mainly because they find it intuitive to use drawings

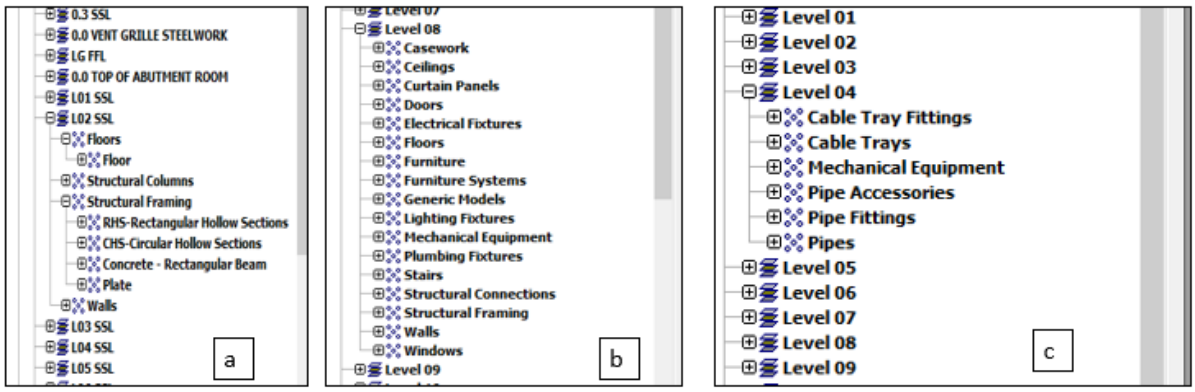


Figure 4.3: Snap shot from design BIM models (Student apartment -C1D2)
a) Structural BIM model b) Architecture BIM model c) MEP BIM model

While examining the files in the CDE in the Student apartment, there were predominantly four types of files. They are drawings, BIM files, reports (method statements, design logic, etc.) and construction programme. The drawings include design drawings, construction method drawings, site layouts etc. The drawings in the CDE are in both their native (.DWG) format and in PDFs (when published). BIM files in the CDE are of five types. 1) Architectural BIM files which contain information (geometry and coordinated engineering information) regarding the architectural components such as ceilings, curtain panels, doors,

electric fixtures, floor, walls, etc. 2) Structural BIM file which contains information regarding the structural components such as piles, floors, columns, beams, core etc. 3)Mechanical Engineering and Plumbing BIM file contain information on ducts, pipes, heating ventilation and air-conditioning components. 4) Federated BIM model contains an aggregation of the three of the above BIM files within a common geometric coordinate system. This model is used for checking clashes between disciplines and discussing the interfaces for construction [Student apartment-CID3]. 5) 4D BIM files contain all the information in a federated BIM model linked with schedule data imported from the Asta Powerproject. Most of the BIM files in the project were created using a set of authoring tools from Autodesk. The respective design teams created the first three types of BIM files. Figure 4.3 shows a snapshot from design BIM models selection tree. The digital engineering team uses the BIM files published by the design teams to create a federated BIM model and then further integrates with temporal data created by the planning team to develop the 4D BIM model [Student apartment-C1I1].

Although BIM is mandated for design submissions in the Metro project, drawings are still used for approvals and design checking.

> "Drawings will be attached against placeholders" (Metro project, Digital Project Management (PPT), C2D2)
>
> "Lead reviewer assigns internal reviewers depending on type of drawing" (Metro project, Digital Project Management (PPT), C2D2)

The digital project management PPT states have multiple references to the use of drawings in the design check workflows.

Despite having a BIM coordination tool in the water project, Bentley Navigator, the project engineers resort to using the CAD tool because they find it easier and intuitive to use drawings.

> "I use [Navigator], I'm not very good at, but I use all the navigator tools that we've got here. But I prefer to use AutoCAD because I find it a lot easier" (Water project, Project engineer, C3I1)

In the contractor's offices of the projects studied, the engineers had drawings on their desks, and the CAD software opened on the monitors. If they find errors in the drawings, they modify and edit them first on paper and then on the computer.

> "Because I have to open up every drawing individually and print them all, or even if you do it the other way around, it's very slow anyway. And then, once I've printed them, reviewed them all, you could do it on there [computer], but it's not the best way because we haven't got the technology. I haven't got a big screen" (Water project, Project engineer, C3I1)

The engineers use laptops with small screens. Some of the hot desks have an additional screen; however, these were also small (less than 23 inches). This is highly inconvenient when reviewing large drawings as they pan and zoom to detect mistakes.

The file-based sharing impacts the data quality dimensions associated with temporality and consistency. Most of the file-based manipulations happen within the computer and are uploaded only when complete. Hence, there is a mismatch between the rate of the volatility of data and currency of data. The volatility is high as the data is manipulated on the users' desks (for example, they are manipulating the information in a printed drawing). However, the currency of the data is low as it is uploaded as a batch. That means the data is updated at a lesser speed than it is varied. This has an implication on the data quality associated with the timeliness dimension as the data in the CDE is not the latest version. File-based sharing impacts consistency too. The files act as individual entities and have information from related files in them. Unlike a model, this information is not connected. Therefore, when the source is updated, the information in the file may not be updated, which introduces inconsistency problems.

4.3.2.3. *Document control bottlenecks*

Document control plays a vital role in the flow of structured information in the three studied projects, and document control professionals are tasked with managing the access, version control, and availability of documents. Before being uploaded to the CDE, those documents must be approved by the relevant authority (depending on the document). In the three projects, the number of documents uploaded to CDEs is enormous and thus processing information becomes a significant task, with bottlenecks in the process leading to workarounds and data quality issues. As the authorization of documents is limited to specific individuals, they tend to get asked for a huge volume of authorizations, and this slows down the information flow.

> "So, there's certain people who are responsible for issuing information or authorize certain communications, and if they're not available then things can stop. Or they may have a high volume of these authorizations to do that it takes them a lot of time to get through" (Water project, Project planner, C3I2)

This process of authorization means that multiple versions of designs can be circulating in different parts of the project. For example, whilst one design is in use by the construction team, in the meantime, the designers could have progressed the design, and the latest design is not uploaded as it is in the queue to get approved.

> "Design teams and checkers and approvers could be progressing designs but then it'd be held up when someone says high up needed to actually approve the whole design" (Water project, Digital engineer, C3I3)
>
> "It's just making sure that I'm getting the latest information and no-one's updating it in the background and then the right versions are going onto [CDE1] […] it's no longer the most current version anymore by the time I'm reviewing it" (Project engineer, C3I1)
>
> "[…]when this revision has been updated from the designer to revision 10 and I sit here on my desk checking the revision 1 and the designer has the revision ten, then that revision 10 is internal […]

> because he keeps on updating but he hasn't he hasn't put it on the [CDE1]." (Water project, Technical manager, C3I5)
>
> "the changes were not updated owing to the dual work repercussions that came along: the updated 2D and 3D drawings were to be first uploaded in Project-wise. After going through the chain of approval processes, the updated drawings were required to be then taken to Asset-wise." (Metro project, Report, C2D4)

Although it seems to be straightforward from the outset, many users find document control frustrating because submission ends up being a long process even when all the attributes are correct. For example, engineers submit the packages to the document controller in their firm, who sends it to the document controller in the other firm (receiving end), which is then sent to the team leader and, finally, to the user who would get the useful information out of that package. This is a long process with checks and iterative cycles involved in each stage. The document controller makes sure that the files in the CDE have the relevant attributes before they are published in the CDE. If there are missing attributes, the submission is rejected. In the case of a Request for Information (RFI), if the document controller does not understand the request, it gets rejected. At times, it takes more than two to three weeks for the document to reach the recipient while following the document control workflow. This is essentially slowing down the whole information flow by implementing a system which was supposed to speed up the process.

> "It took us three weeks to actually get the package with all the right documents and revisions in there, and that's a long time; again […] I gave it to my document control and my document control sent it to the other company's document control, the document control there sends it to whoever the lead is, and the lead then sends it on to whoever's doing the work – and that may take a week or two. And that's completely wasted time, and no one in the middle of that process has done any work[…] And actually, by the time it gets to the people who are reviewing the actual technical data, it may be three or four weeks later […] In fact, I did it yesterday, I sent a load of RFIs through to

> [designer] informally, five minutes after I'd sent it through my document control." (Water project, Principal engineer, C3I8)

To bypass this obstacle, workers send the information through an unstructured channel in addition to the structured workflows. This is because of poor understanding of document control workflows amongst the project participants regarding the requirement of these structured workflows and the CDE. The slow processes and the need for completing the task before deadlines force employees not to follow document control workflows.

> "There is generally a poor understanding of document control requirements, certification requirements […] we're finding that general good practice that people should have brought with them from other projects is being conveniently put to one side for the purposes of expediting the work that people are being asked to do" (Water project, Information manager, C3I4)

In addition, the workflows in the CDE are complex and not intuitive, making it difficult for users to follow the protocols for document control.

> "Just because of the way they need to store it in certain places and stuff like that, and it can't be done… The way that [CDE1] is set up here, I believe it is not easy to use" (Water project, Project engineer, C3I1)
>
> "It seems very complex, you open them up, there's lots of things going on […] I just want to know where I can get my latest drawing" (Water project, Digital engineer, C3I3)

This has resulted in employees bypassing the workflow, which leads to system conflicts and further delays in the processes and, at times, in the information being stored on the CDE.

> "I think someone within the doc management system had obviously circumnavigated it somehow, to get the drawings out. And then when we were trying to get the said revisions for our set out, the system wouldn't allow it because directory hadn't been properly created." (Water project, Technical manager, C3I5)

When there are conflicts in the system, this reduces the *accessibility* of that information to other stakeholders as well as limits the computer programs from accessing it. This challenge was observed in the Metro project and the Water project. The Student apartment, which is of

relatively lesser complexity when compared to other projects, did not show evidence of this issue.

The complexity of workflows and document control measures has implications for data quality. There are shared norms, values, and expectations for the users regarding the tools, such as speed, easy communication, and transparency, which were developed based on their previous experiences of collaboration. When the new tool does not meet the expected qualities, it reduces their productivity, and users move back to the older ways of information sharing to expedite the task. When users find it difficult to utilize the CDE for structured information flow, they bypass the workflows to get the work expedited. This leads to the loss of metadata, document trails, and information dependencies as these unstructured communication channels offer limited or no codification. In addition, the document control workflow itself makes the process slow. Not following the structured workflow, results in lowering the machine readability of the information to be rated no-star as it is not indexed. Document control bottlenecks have multiple implications for data quality. Firstly, the value for the timeliness dimension is lowered as the data which is published might not be the latest. Hence, the inferences are based on old data, which leads to false interpretations. Secondly, this lowers the semantic accuracy of the data as its attributes might no longer be true. This also introduces the problem of consistency. Depending on which database employees look at, they see different values. For instance, one CDE to which the information was packaged would have the latest value, whilst the one who must go through another document controller would have a different value. This tampers with the idea of a ‘single source of truth’. In addition, when users circumnavigate the workflow, there are more data quality issues related to unstructured information sharing such as accessibility and provenance.

4.3.2.4. *Lack of process change*

Lack of process change is when the engineers follow a conventional workflow even when they have a digital workflow for that task.

The CDE of the Student apartment was hosted on an online collaboration platform from Asite, ensuring that the documents and files uploaded to this platform are shared across all the stakeholders. Different stakeholders had different access permissions, granting them access to the data necessary to carry out their responsibilities. Files in the CDE have two status codes 1) Work in Progress (WIP) and 2) Published Documentation (CDE Access, C1D4). WIP documents can only be viewed by the team who is working on that document. Members from other teams can see the filenames in the CDE, but they cannot open or download these documents. Once a WIP document is completed, it goes through an authorisation process, wherein the team leader will check the completion of the work in the document and sent it to the document controller, who changes the status of the document to published.

> "Once the file is ready for sharing, it is reviewed by the designated team lead and then sent to the document controller. It is the document controller who makes sure that the file has relevant information so that I can share it in the CDE" (Student apartment, Digital engineer – Field notes, C1I1)

A published document can be accessed by users of other teams with relevant permissions [Student apartment -C1I1]. Most of the files in the CDE were PDF documents. These files included design drawings, temporary works drawings, method statements, construction programme, etc. The PDF drawings were linked to the source data in Autodesk DWG format through placeholders. This is true for the construction programme where the construction schedule created by the team using Asta power project was linked to from the PDF program [Student apartment C1D1, C1D4]. The files in the CDE are uploaded using a certain naming convention agreed between the stakeholders according to PAS1192-2:2013. The file names

inform the user of the project location, the author, spatial relation of that file to the project, type of the file, discipline which created the file and unique number [Student apartment C1D4].

In the Metro project, the client mandated the use of BIM and supporting tools such as Primavera, RIB and SAP to integrate the data and provide meaningful insights. However, the project manager has stated that the engineers could still use spreadsheets for design calculations and project management as they are used to those tools in their daily workflows.

> "the new technology had to be implemented under circumstances where MS Excel, leave alone MS Project, was used predominantly for planning" (Metro project, Report, C2D4)
>
> "The stakeholders found MS Excel to be very flexible to their needs and it was a one-stop solution for all their problems" (Metro project, Report, C2D4)
>
> "People still use spreadsheet – that's what they are used to" (Metro project, Field notes- Project manager, C2I1)

Although these tools are old, they are still robust for data handling in as they are embedded in the familiar workflows followed by the engineers. These practices create issues associated with data inconsistencies as the data is distributed in different files and not synchronized, as these tools are not designed for collaborative data management as required in a complex project.

In the *Water project*, three CDEs were used for the project, which created issues such as double handling and data inconsistencies. The presence of multiple CDEs in a project is another example of the lack of process change. Multiple CDEs resembles the paper-based workflows such as document flows between the designer and contractor, another set of document flows within the contractor's office, and another set of document flows to the clients for approvals.

> 'Our client prefers [CDE1], and we have the designer who stores things in [CDE2], so we have both of those tools, and we have to balance between those two. That can be very confusing when we have two platforms' (Water project, Technical manager, C3I5).

For the information submissions to the clients, CDE1 was used; for the information from the design consultants to the contractor, CDE2 was used; and for internal file handling and sharing with the contractor, CDE3 was used. CDE1 and CDE3 came from the same vendor. However, CDE2 was from a different vendor.

> "Everything had to be taken out of one data environment and pushed into another. One of the issues with that is the consistency or the compliance or knowing the latest versions of information" (Digital engineer, C3I6)
>
> "As the contractor, then we have to deliver it to a completely separate, disconnected CDE […] we're double-handling" (Water project , Digital engineer, C3I3)

When the CDEs are disconnected, the document trail is lost when a document is moved from one CDE to another, leading to the loss of traceability.

Even though structured information flow is digitized through the introduction of a CDE, the process to enable information flow remains unchanged. For example, for a piece of information to be approved, it must be printed, associated with a cover sheet, and signed.

> "We're actually going to export that out of the CDE, we're going to print it out, we're going to staple it together, we're going to put our own cover sheet on the front of it, with the exact same details on the back and we're going to go off and go and get three signatures, scan it back in, put it back into [CDE1] and submit it." (Project engineer, C3I1)

Printing and scanning the document results in loss of metadata. A scanned document in a PDF format has little machine-readable information in it.

This issue was observed in all cases. Lack of process change means that conventional workflows are followed even when a better digital workflow is available. Lack of process change has implications on machine readability as printing and scanning removes the metadata from the files resulting in 1-Star data as it removes metadata and data relationships from the files. A scanned version of the file would also have very little machine-readable information embedded in it and would require resource-intensive methods to extract insights from it. This

reduces the data quality dimensions such as accessibility (as the metadata and data relationships are removed), completeness (information is not complete), and provenance (document trail is lost in the process). In addition, having multiple CDEs introduces the data quality issue associated with provenance as files residing in multiple CDEs are disconnected, and the relationships of that particular file with another file are lost in the transition process. There are further issues with synchronization of the information when it is distributed in multiple CDEs. This introduces the data quality issues associated with consistency, which limits the quality of findings made by inference algorithms.

4.3.2.5. *Discussion*

The analysis suggests that the construction sector has not yet made the transition from document-based to model-based ways of organizing digital data. The use of drawings and file-based sharing, unstructured information sharing, printing, and scanning of documents, multiple CDEs, and so forth in information sharing has a significant impact on the machine readability of data. Paper-based practices are institutionalized in the sector, and while digital ways of working are replacing them, this change is slow, with users of construction information still conditioned to work with drawings and PDFs and unstructured information sharing. Even in projects which are championing newer BIM based workflows using CDEs, this work finds it is difficult to replace these practices, as evidenced by the problems associated with information sharing. The complexity and long processing times involved in these workflows force users to shift back to existing practices and workarounds to expedite their work. The findings from this study also support the previous characterization of users in construction combining different methods of information sharing, especially when they were hindered by bottlenecks in BIM based processes, such as document control. Thus, we can characterize that project participants use a range 'hybrid practices' (Harty and Whyte, 2010), for information sharing.

4.3.3. **Missing information**

This section presents the codification challenges related to missing values in construction process information. The data analysis shows that product information is relatively well structured as BIM models, analysis models, and CAD drawings. However, construction process information is relatively less structured and detailed when compared to product information. Process information is codified as Gantt chart models in scheduling software and then linked to the BIM model. The detailed process information is not structured into a model. Instead, it is shared as method statements in less structured PowerPoint presentations and PDF documents. Sharing information in these unstructured formats has implications for data quality, which are presented in this section.

4.3.3.1. *Loss of constraints*

In construction, the constraints for any activity execution are discussed during team meetings as the constraints span between different teams. For example, logistics constraints span amongst prefabrication, logistics, and site teams. These discussions lead to the removal of constraints by rearranging the start and end times for the activities. These are then translated to Gantt charts as an output. However, the constraints contributing to this translation are lost as the Gantt charts contain only precedence relationships. The knowledge that there was a constraint is not recorded, and thus the presence of that constraint is not codified. This means such constraints are not machine-readable as the access to this information was limited to the participants of a particular meeting. If an automatic scheduler is used to reschedule these activities, this rescheduling activity would not have access to this information, resulting in an unrealistic schedule.

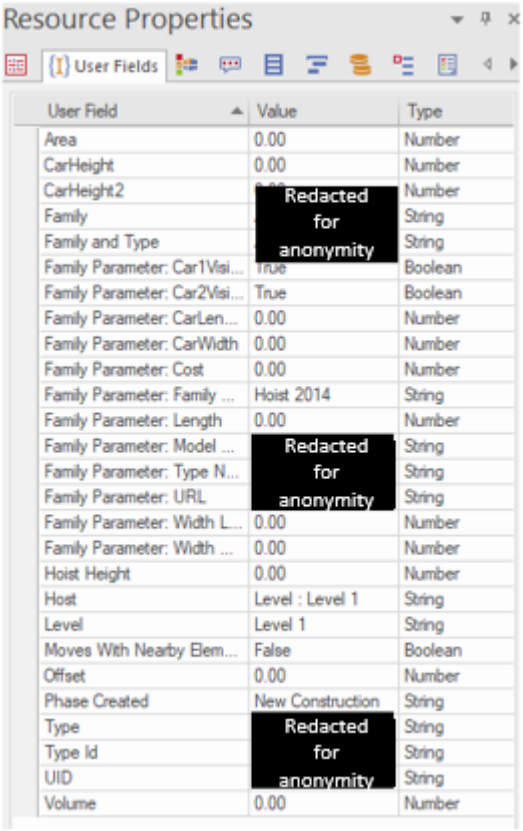


Figure 4.4: Screen shot of the hoist showing lack of constraint details (Student apartment, C2D2)

In the Student apartment, the constraints related to the resources that were discussed by the methods engineer was implicitly recorded into the method statements. The BIM model contained the crane as well as the hoist. The properties of the hoist were limited to the geometry in the 4D model. The semantic properties of the hoist, such as its car height, hoist height and lifting constraints, were not recorded in the model, as shown in the screenshot (Figure 4.4). The operating area of the tower crane was depicted as an extra geometric element in the 4D model.

The Metro project follows a 5D BIM workflow (digital project management PPT—C2D2), where the schedule is linked to a BIM model, Bill of Quantities (BOQ), and an Enterprise Resource Planning (ERP) system to compare the cost based on the quantities versus the cost stated in the work orders from the subcontractors. The progress information is also linked to this model to ensure that the work is done before sanctioning the bills for the work

orders. This enables the project to do automated progress and cost updates to the models based on the work order submissions from the contractor. However, the schedule linked to the BIM and BOQ did not have relevant constraints defined between the activities. The BIM consultant stated that the missing constraints resulted in conflicting reports.

> "Work package for three spans were linked to a work order. Model showed the deck for a span was completed before the pier supporting it was completed because the work package for the first span was reported as completed." (Metro project, Field notes- BIM Consultant 2, C2I5)

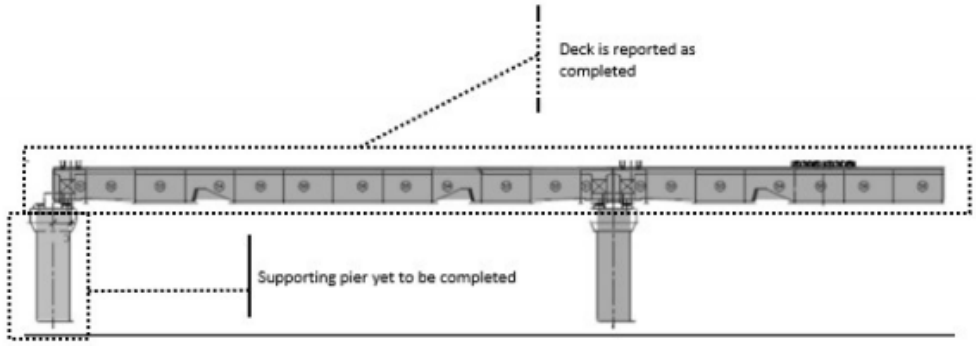


Figure 4.5: Visual representation showing the effect of the low level of detail in process information. The deck is completed before the completion of the supporting pier

The deck of the metro can be completed only when the pier supporting it is completed. However, the system recorded the deck assembly to be completed when the pier was not completed (Figure 4.5). This was because work packages for the deck and pier were different as different subcontractors did them, and the level of detail of the work package is low. A subcontractor who dealt with deck assembly had a part of the work package completed. However, the lack of detail in work packaging triggered the computer to record the whole work package as completed, resulting in the error. This is a clear case of lack of constraint codification in the model leading to wrong inferencing.

Project planners convert the constraints discussed in a meeting into a precedence relation in the Gantt chart for communication. However, during this process, many of the constraints themselves, and thus context information for rearranging the activities, are lost as the precedence relations only holds the sequence and duration of the activities.

> "So, from an engineering perspective, we have to interpret engineering information, whether it be drawing or written constraints, written narratives and interpret those into a Gantt chart. So, we physically need that information to know what we're building and what the constraints in building it are." (Water project, Project planner, C3I2).

For example, one of the meetings in the Water project had an issue with piling, where the pile-driving equipment did not have access to the site for a specific date as there was another activity going on which limited the width of the site access road. At the meeting, this was raised:

> "Access chamber works will conflict with access road for pile work, piling work package has to be moved back 2 weeks." (Water project, Progress review meeting notes, C3M3)

Here, there is a dependency between access chamber works, and the piling work package as the access chamber work would reduce the road width. Therefore, the piling activity was delayed to a later date. The constraint was removed. However, the knowledge that there was a constraint is not recorded, and thus the presence of that constraint is not codified. If there is a change in the plans in the case of a disruption, the dependency would be identified as a precedence relationship, but the real constraint would not be identified from the Gantt chart model.

Loss of constraints was observed in all cases. This issue introduces data quality problems associated with accessibility (information is limited to people who attended the meeting) and data completeness (the model does not include any constraint information; hence, it is incomplete). This also results in a Zero Star level of machine readability.

4.3.3.2. *Low level of detail*

The schedules need to be of a particular level of detail so that activities can be linked with the BIM object and then used for data analytics. However, in practice, the schedules are not codified in detail.

While examining the construction programme [Student apartment-C1D1] in the student apartment, I observed that the details of the hoist uninstall and finishing up is coded at a low level of detail (see Figure 4.6). The cladding installation was described as a lump of processes for levels 1-5 in the programme. Each activity had a duration of 4 days in the programme.

> "…after every 5 floors, the spider crane used for holding the panels should be taken to the next location after it is used to support the cladding installation. Approximately, it would take half a day depending on the availability of the tower crane…" [Student apartment-Methods engineer-Field notes, C1S1]

L01-05: Install Remaining Cladding	26/06/2018 08:00	4d	29/06/2018 17:00
L06-10: Install Remaining Cladding	02/07/2018 08:00	4d	05/07/2018 17:00
L11-15: Install Remaining Cladding	06/07/2018 08:00	4d	11/07/2018 17:00
L16-20: Install Remaining Cladding	12/07/2018 08:00	4d	17/07/2018 17:00
L21-25: Install Remaining Cladding	18/07/2018 08:00	4d	23/07/2018 17:00

Figure 4.6: Extract from construction programme [Student apartment-C1D1]

However, in the meeting, where the researcher was part of, the methods team had discussed processes related to the movement of spider crane from one floor to another, which takes at least half a day. However, these details were not recorded in the construction programme.

In the Metro project, the contractor who was new to the idea of project management had trouble submitting the schedules of required detail. This resulted in the schedules not linked to the BIM models.

> "Barring the schedules prepared by a few reputed contractors, the baseline schedules prepared by other contractors was poor. The schedules did not contain the required level of details and attributes and thereby could not be linked to the 3D model" (Metro project, Report, C2D4)

The attributes were missing in the schedule, and those parts of schedule not linked to BIM or BOQ remained inaccessible in the CDE and other model-based platforms. This leads to decision-making with incomplete information.

This issue was predominantly observed in the Student apartment and the Metro project. Schedules in the construction programme are made at a high level, and the detailed construction programmes are communicated as PDF documents, PowerPoint presentation or 2D Drawings. However, these are not always embedded in object-oriented machine-readable formats as 4D BIM or in planning software. When the required amount of information is not embedded in a machine-readable format, further tasks such as linking to a BIM model and ERP software cannot be done, as seen in the Metro project. Also, it may cause problems with understanding of the underlying activities (crane to lift to the next floor) in the student apartment. This information may be available in a human-readable format or may be accessible to the respected team members through unstructured media. However, the information remains inaccessible for data science.

These issues are caused by low data quality due to incomplete information relates to the completeness dimension and also results in Zero Star level of machine readability.

4.3.3.3. *Discussion*

Regarding missing process information, this study shows that the process codification is limited to the master planning or phase planning level and lacks the level of detail and linkage required for the application of data science tools. While product information is relatively well codified in BIM, process information is less detailed, and there is a lack of codification of constraint relationship across all cases. These challenges are identified by researchers implementing 4D BIM. For example, Han and Golparvar-Fard (2017) have stated that the process modelling methods fail to document field issues to be made available for further

analysis; for instance, the 4D BIM's *"Model Breakdown Structure typically does not match operational details or require creating complicated namespaces which, without visual representations, are difficult to communicate"* (p. 1733). Giretti et al. (2012) have further reported the lack of correlation between the resources employed hourly and work progress. This led to the decomposition of tasks into sub-tasks to determine causal relationships between the involved variables so the whole progress could be determined. This study suggests that to overcome such reported issues, the methods for codifying construction process information must be more detailed.

The practice of codifying planning information being limited to master planning and phase planning, without the focus on codifying the rationale for granular planning such as look-ahead planning and weekly planning, is causing this codification challenge. The lack of semantic relationships embedded in the model limits the application of automatic schedulers since the schedulers do not have access to the constraint relationship. These issues suggest both a change in the modelling of process information in construction, with the need to develop tools which support modelling of complex constraint information, and also a change in the practice to codify the process information in greater detail so that data science could be employed to augment decision making in construction

4.4. Synthesis of the findings

The previous section presented the identification of the codification challenges by analysing the data from the three construction projects based on the concepts drawn from the prior studies on data quality dimensions and machine readability. Table 4.4 summarizes the codification challenges observed from studying the projects. Low machine readability of data is a significant challenge for codification, which was observed across the projects. Product information is well codified through BIM, CAD drawings, analysis models, and so on in all the projects. However, the codified information is distributed amongst different formats and

databases, limiting the application of analytics. In addition, multiple modes of communication, multiple CDEs, and lack of process change also limit the codification of information in the projects studied Lack of codification of process information is identified as a critical issue amongst all the cases.

Table 4.4: Codification challenges across the cases

Codification challenges		Student apartment	Metro project	Water project
Software usage	Interoperability		X	X
	Information loss during conversions	X		X
	Multiple modelling techniques			X
Information sharing	Unstructured information sharing	X	X	X
	Drawing and file-based sharing	X	X	X
	Document control bottleneck		X	X
	Lack of process change	X	X	X
Missing process information	Loss of constraints	X	X	X
	Low level of detail	X	X	X

4.5. Summary

This chapter presented the identification of codification challenges in construction which limited the application of data science, by qualitatively studying three construction projects (Student apartment, Metro project and Water project. The identified codification challenges are categorised into three themes: 1) software usage—interoperability, information loss during conversion, and multiple modelling techniques, 2) information sharing—unstructured information sharing, drawing and file-based sharing document control, and lack of process change, and 3) missing process information—loss of constraints and low levels of detail. This chapter shows how codification challenges arise because of the different digital workflows and working practices across projects, and how these lead to fragmented data as a result of the use of multiple software packages, poor information sharing and only partially captured construction process information.

The codification challenges identified in this chapter has implications for the development of data-driven decision-support tools for look-ahead planning. For instance, codification challenges such as low-level of detail and loss of constraints resulting in missing construction process information limit the development of these decision-support tools as the required data are not present in a machine-readable format. In addition, information sharing issues such as file-based sharing, unstructured information sharing, document control bottlenecks reduce accuracy, completeness, consistency, and accessibility of data resulting in data-driven decision-support tools making incorrect inferences from the low-quality data. In addition, information sharing issues such as file-based sharing, unstructured information sharing, document control bottlenecks reduce accuracy, completeness, consistency, and accessibility of data resulting in data-driven decision-support tools making incorrect inferences from the low-quality data. Furthermore, software usage issues also limit the application of data-driven decision-support tools as the accessibility to the data is limited due to proprietary data formats which are not interoperable and loss of information during conversion to open formats. Although there are techniques to improve the quality of data using advanced machine learning techniques such as natural language processing and deep learning (Wang, 2017; Carrillo, Harding & Choudhary, 2011), this is a resource-intensive process, and the result of this process may also not be of high-quality (Wang, 2017).

This chapter has provided a foundation for technology development in the thesis by extending the knowledge on data quality issues in construction through the identification of codification challenges, considering the wider practice of model and document-based information sharing. Building on the findings from this chapter, the following chapter presents a linked-data based information modelling method based the principles stated by Berners-Lee (2006) to generate machine-readable construction process information by codifying by detailed construction information including construction scheduling constraints. This novel information

modelling approach addresses the issue of software usage such as interoperability and loss of information during conversions leveraging the power of linked-data to link heterogeneous datasets and making logical inferences from it. Linking of datasets would increase the data quality pertaining to accuracy, consistency, and timeliness dimensions. The following chapter describes the development and testing of this novel information modelling method.

5. CODIFYING AND VALIDATING CONSTRUCTION PROCESSES INFORMATION USING LINKED-DATA[3]

5.1. Introduction

This chapter presents a linked-data-based method for codifying and validating construction process information. The findings from the previous chapter inform that the knowledge about construction processes is predominantly codified as precedence relationships in the form of master plans and phase plans in 4D Building Information Model (BIM) and other scheduling software systems. However, it was observed that this information was of low-level of detail and lacking constraint codification (Subsection 4.3.3), thereby limiting the application of data-driven decision-support tools. Codification of detailed construction process information such as construction constraints is challenging, as they are highly dynamic and complex with numerous interdependencies, making it difficult to be modeled using relational databases or similar informational models (Wang *et al.*, 2016). Linked-data-based method for codifying and validating construction process information addresses these challenges to model detailed construction information so that data-driven decision-support tools can be applied.

Recent research has used linked-data based modelling approaches to address the issues associated with non-expressivity while modelling with relational databases and similar technologies. Linked-data enables linking of data across domains and heterogeneous sources, without being limited by the underlying schema. Therefore, linked-data can address the problem of complexities and interdependencies in construction process information, by

[3] Parts of this chapter have been accepted for publication as *Soman, R.K., Molina-Solana, M. & Whyte, J.K. (2020) Linked-Data based Constraint-Checking (LDCC) to support look-ahead planning in construction. Automation in Construction. (In press).*

enabling methods to encode the semantic description of the relationships of the constraints by linking across heterogeneous datasets in different domains. Building on the research on linked-data in the construction sector (Pauwels, Zhang & Lee, 2017), this chapter presents a method for codifying detailed information about construction processes. A further contribution is the ability to automatically check them for constraint violation using linked-data and Shapes Query Language (SHACL). The process information is modeled in Resource Description Framework (RDF) on top of the ifcOWL ontology, and constraints are modeled in SHACL. SHACL is a language for validating RDF graphs against a set of conditions. I developed a software prototype based on this framework to validate the process information in the model against the modeled constraints, checking whether a schedule is feasible or not.

The rest of this chapter is organized as follows. Section 5.2 outlines Linked-data based Constraint Checking (LDCC), the method that combines linked-data-based approaches with BIM based rule-checking to model and validate scheduling constraints. Section 5.3 presents a prototype web application demonstrating this LDCC method. Section 5.4 discusses the results of evaluating using an OpenBIM dataset. Section 5.5 presents the conclusions and implications

5.2. Linked-data-based constraint checking architecture

The architecture for checking the dynamic constraints is shown in Figure 5.1. The LDCC method is intended to be used at look-ahead planning meetings to aid decision-making. Detailed construction information shared during look-ahead-planning meetings (including subtasks and constraints) will be the input to LDCC. It will be linked to the 4D BIM, modeled as an RDF graph, and checked for constraint violations using the SHACL processing engine. This method for identifying constraint violation helps the engineers to make decisions during look-ahead planning meetings. Thus, the LDCC method act as a decision-support tool during look-ahead planning meetings. To provide decision-support functionalities, there are three layers in this architecture:

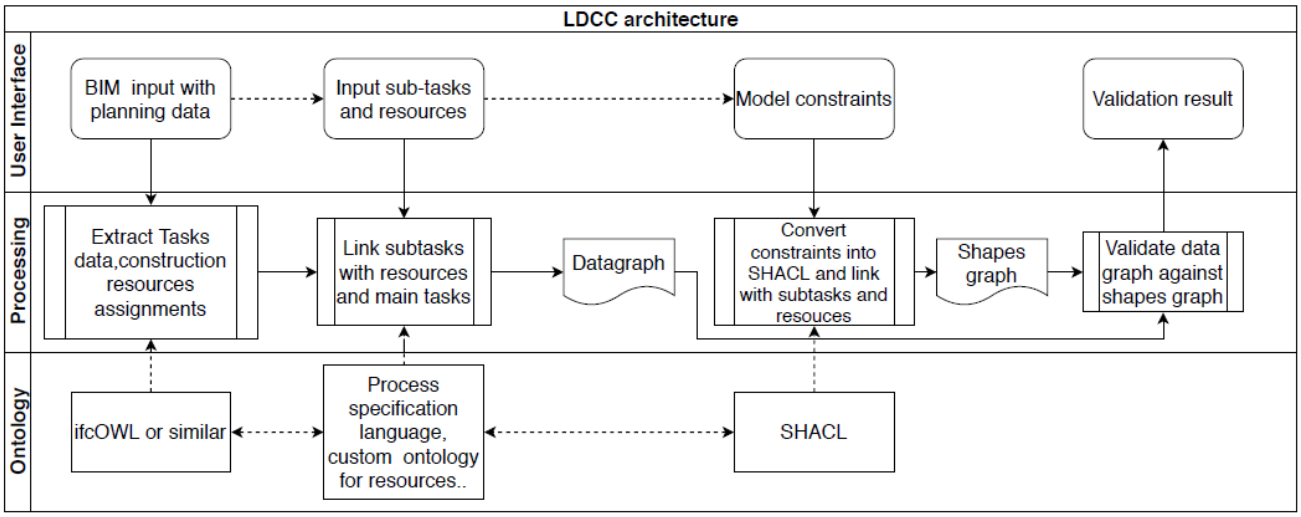


Figure 5.1: LDCC architecture

1 User interface (UI) layer: This layer is for interaction with the user. In an effort to make the tool work along with the existing work practices for successful adaptation (Hartmann, 2008), I designed the UI layer as an increment to a current 4D BIM UI with three major inputs from the user. The first input is a BIM model with embedded planning information (planning data at the master planning/phase planning level). The second input contains information required to enrich the planning data (including subtasks for main tasks/activities in master planning, renewable and non-renewable resources needed for executing the tasks, etc.) The third user input includes information on constraints and the relationship between activities, subtasks, resources, approvals, etc. The output from this layer is the feasibility report of activities scheduled using the information from the input layer.

2 Processing layer: This layer prepares the data for processing, processes the data, and generates the validation report following three steps. The first step involves extracting the planning data from the BIM model (using a parser and then converting it to an RDF graph following a suitable ontology) and enriching it with the detailed data input from the user (adding triples related to subtasks and

resources). This enrichment should include describing the data using a suitable ontology automatically, such as process specification language or LinkOnt (explained in Subsection 5.2.1) as in this chapter. This step results in the generation of a "data graph" in RDF for the look-ahead window. The second step involves creating a "shape graph" from the constraint information taken as input from the user. This step also imports the previous shapes generated from the prior meetings into the shapes graph. The shape graph is generated using SHACL vocabulary. The third step involves validating the data graph against the shape graph to identify constraint violations. The resources in the data graph containing planning information from the BIM and information added from the user during the look-ahead planning meetings are validated against corresponding shapes (constraints) in the shape graph. A validation report is generated as an output of this step.

3 Ontology layer: This layer holds the schema for different data (e.g., the BIM data, resource data, or SHACL). The schemas are stored in a repository on the web and published as per the principles stated by Berners-Lee (Berners-Lee, 2006). As this layer is accessed during the execution phase, the latest updates to the ontologies and constraint relationships introduce dynamicity in the LDCC architecture. Schemas are referenced and not hardcoded.

This architecture enables the separation between the data graph and shape graph, supporting re-evaluation with new information. In any case, constraints are not hardcoded into the system but are modeled and edited during look-ahead-planning meetings. Modeled constraints apply to the data in the data graph, enabling the capture of knowledge generated during meetings to be codified into a machine-readable format and reused. For example, if constraints (e.g., lifting capacity) related to a resource (e.g., crane) for an activity (e.g. lifting

of modules) are discussed in a look-ahead-planning meeting and codified as a constraint using the LDCC prototype, a SHACL shape is generated for the resource (e.g., a SHACL shape containing information for the lifting operation, such as lifting capacity < weight of the module). Because the shape is generated for this resource, the resource is validated against that shape each time the resource is used. Hence, if the same resource is assigned to a similar activity (lifting another module) in a subsequent meeting, the resource (crane) is validated against the previously created shape (lifting capacity<weight of the module) for the new activity, until the shape, is deleted from the shape graph. This means that constraints are always automatically reused in further meetings, even when they are not explicitly mentioned. However, if the nature of the activity changes completely, a new constraint needs to be modeled.

5.2.1. Ontology for linked-data-based constraint checking

This section details the data sources, naming strategy, and associated ontologies used for supporting LDCC. For LDCC to work, the data need to be published following the linked-data principles. To do so, several tasks are involved, linked-data including identifying and obtaining access to the data source, analyzing the data source, defining the resource-naming strategy, developing the ontology, transforming the data source, and linking with another dataset (Radulovic *et al.*, 2015). Within the scope of the objectives of this study, the generation of the linked-data is to support the look-ahead planning tasks and check the look-ahead schedule by traversing across multiple datasets. Therefore, the sources of data include the BIM model (including the construction site information), construction process data, resource information, the approval information, etc. The next step is to analyze the data sources to determine the relevant information necessary for performing look-ahead scheduling, including the geometry of the construction site and resources, operating requirements of the resources, and project management information (Dong *et al.*, 2013a). Regarding the naming strategy for

the entities, I chose the "hashtag" strategy to avoid redirection issues associated with the "slash" format (World Wide Web Consortium (W3C), 2006). The relevant data are hosted within the respective servers. Because the LDCC aims to support look-ahead planning, by linking across domains and making logical inferences to perform constraint checking on construction process information, existing ontologies are reused. Particularly, I use IfcOWL to capture the model information, LinkOnt (a custom ontology detailed below) to capture the resource information and links, and SHACL to model constraints.

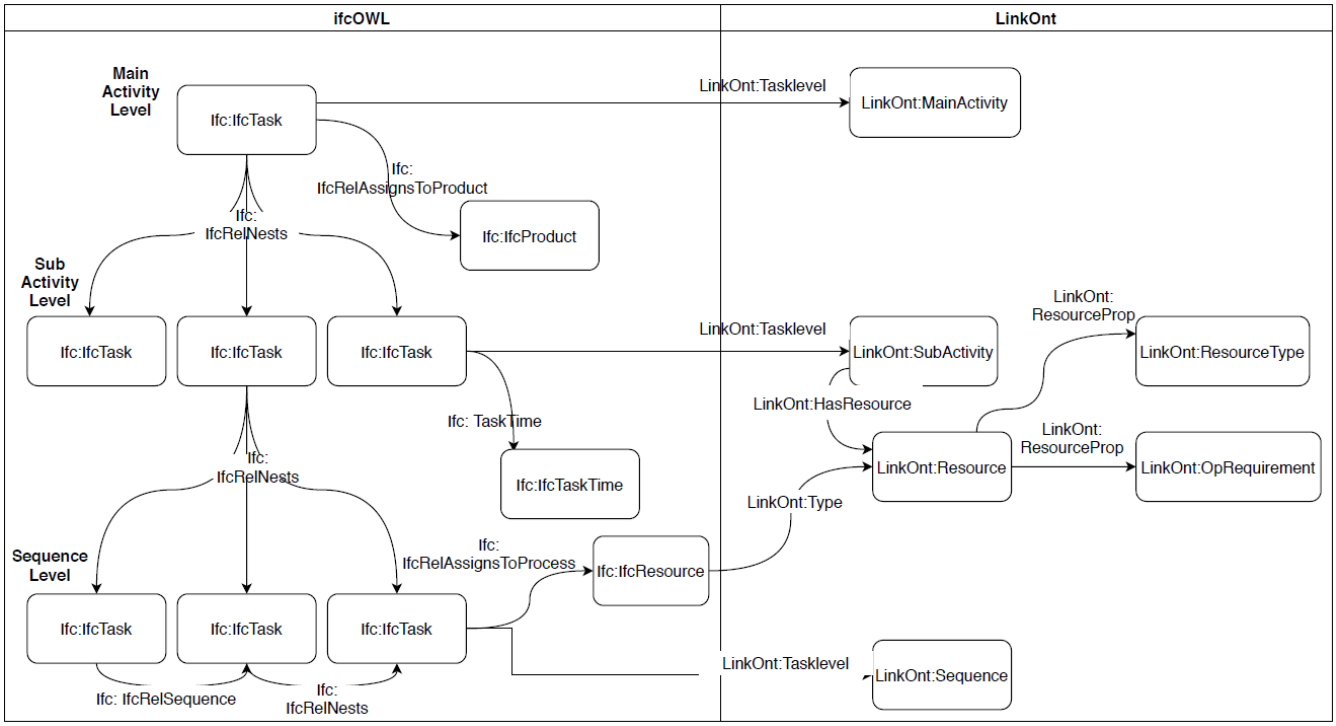


Figure 5.2: Schematic of the LinkOnt ontology, showing its classes and relationships between them

To perform dynamic constraint codification, we require several classes from *IfcOWL* (*ifcTask, ifcProduct, ifcResource* and *ifcTaskTime*) to capture the project management and model information. *ifcTask* describes a construction activity, and *ifcResource* describes the associated resources required to complete an activity. In addition, *ifcProduct* represents the

abstract concept of an entity related to the spatial or geometric context, whereas *ifcTaskTime* captures the time-related information for the activity to be performed.

LinkOnt is the ontology I propose to support dynamic constraint checking using linked-data. This ontology introduces classes that are missing from IfcOWL but are required to do constraint checking. Because the *ifcTask* does not define the levels of the task (activity, sub activity, and sequence), a new class *TaskLevel* class is used to define the task level. The three task levels are *MainActivity, SubActivity,* and *Sequence*. To capture the URI of the geometry and the operating requirements of resources, *ifcResource* is linked with the *Resource* class using its *Type* property. The *Resource* class has a *ResourceProp* property that links it to the *ResourceType* class (which holds the information about the resource) and *OpRequirement* (which holds the requirements about the operating conditions of that resource). To link a resource to a *SubActivity,* a property *HasResource* is used. In summary, the *LinkOnt* ontology includes the following classes (Figure 5.2): *TaskLevel, MainActivity, SubActivity, Sequence, Type, HasResource, Resource, ResourceProp, ResourceType,* and *OpRequirement*.

5.2.2. Modelling scheduling constraints using Shapes Constraint Language (SHACL)

Scheduling constraints ensure that the activities in the construction program are semantically linked to each other. This section describes how these scheduling constraints (described in Subsection 2.1.1) can be modeled as SHACL shapes to perform linked-data-based constraint checking. The constraints ensure that the relationship between activities, between activities and the workspace, and between activities and resources assigned to the activities are captured. The following subsections describe how these scheduling constraints are modeled as SHACL shapes.

5.2.2.1. *Cardinality constraints*

Cardinality constraints (which are part of SHACL core) represent restrictions on the number of values for a given property. This is important in scheduling, as different properties required different property values. For example, all scheduled activities must only have one start and one end date. Listing 5.1 contains the SHACL shape implementing this cardinality constraint. Similar cardinality constraints can be defined for other properties.

```
1. Ifc:hasTaskTime
2.  rdf:type sh:PropertyShape;
3.  sh:targetClass Ifc:IfcTask ;
4.  sh:path Ifc: taskTime_IfcTask ;
5.  sh:minCount 1;
6.  sh:maxCount 1.
7.
8. Ifc:hasDates
9.   rdf:type sh:PropertyShape;
10.  sh:targetClass Ifc:taskTime_IfcTask ;
11.  sh:path Ifc:ScheduleStart_IfcTaskTime,
   Ifc:ScheduleFinish_IfcTaskTime;
12.   sh:minCount 1;
13.   sh:maxCount 1
14.
```

Listing 5.1: Cardinality constraints SHACL shapes

In Listing 5.1, lines 1-6 ensure that each target class *Ifc:IfcTask* in the data graph has a property *Ifc:taskTime_IfcTask*. This is defined in such a way that the SHACL processing engine identifies all the classes defined in target class defined by *sh:targetClass* (in this case *Ifc:IfcTask*). Once the target class instances are identified in the data graph, the processing engine searches for the property shape (because the current property is defined as a property shape as shown in line 2 – *rdf:type sh:PropertyShape*) and the path for the property shape (as shown in line 4: *sh:path Ifc: taskTime_IfcTask*). In addition, *minCount* and *maxCount* are applied on that property path (in this case *Ifc:taskTime_IfcTask*). Similarly, lines 8-13 ensure

that each target class *taskTime_IfcTask* has at least and at most one start and end date defined by *Ifc:ScheduleStart_IfcTaskTime* and *Ifc:ScheduleFinish_IfcTaskTime* properties.

5.2.2.2. *Logical constraints*

Logical constraints are used to represent the site constraints that need to be satisfied for an activity to take place. These constraints might relate to space availability, resource parameters, and activity relations. For example, the lifting capacity of a crane should always be higher than the weight of the module to be lifted. Listing 5.2 defines the *less than* constraint as a SHACL shape. These shapes can be reused when the crane is used to support another activity.

```
1. LinkOnt:SubActivity
2.  rdf:type sh:NodeShape ;
3.  rdfs:subClassOf owl:Class ;
4.  sh:sparql [
5.  sh:message "Constraint LESS THAN is violated" ;
6.  sh:prefixes < http://purl.org/ConstructLinkOnt/LinkOnt > ;
7.  sh:select """SELECT $this
8. WHERE {
9.  $this rdf:type LinkOnt:SubActivity.
10.      $this LinkOnt:HasResource ?<Resource1>.
11.      ?<Resource1> LinkOnt:ResourceType <Resource1Type>.
12.      $this LinkOnt:HasResource ?<Resource2>.
13.      ?<Resource2> LinkOnt:ResourceType <Resource2Type>.
14.     …..?<Resource1> LinkOnt:OpRequirement ?<Resource1Prop>.
15.     …..?<Resource2> LinkOnt:OpRequirement ?<Resource2Prop>.
16.      FILTER (?<Resource1Prop> <= ?<Resource2Prop>).
17.     }""" ; ];
```

Listing 5.2: SHACL shape for Less-than constraint

To model this constraint, we need to use SHACL SPARQL, an advanced feature of SHACL that adds the expressive flexibility in targeting focus nodes. In Listing 5.2, the aim of the shape is to detect constraint violations where the activities have two resources (<*Resource1*> line 10, and <*Resource2*>line 12) of two types (*Resource1Type*> line 11, and <*Resource2Type*>, line 13) and operating requirements (<*Resource1Prop*> line 14, and <*Resource2Prop*>, line 15). The shape checks whether the value of <*Resource1Prop*> is less

than <*Resource2Prop*>. The constraint is modeled as a node shape and applied to the *LinkOnt:SubActivity* (Listing 5.2, lines 1-4). The *LinkOnt:SubActivity* is related to resources assigned to it (Listing 5.2, lines 9, 10, and 13). Therefore, all the assertions of *LinkOnt:SubActivity,* which has resources of type <*Resource1Type*> and <*Resource2Type*>, are checked for the condition <*Resource1Prop*> *is less than* <*Resource2Prop*>. The assertions of *LinkOnt:SubActivity* which does not have the particular resource type are ignored. The properties <*Resource1Prop*> and <*Resource2Prop*> reside in the respective resource servers. The SELECT queries filter all the *LinkOnt:SubActivity* assertions with the said resource associations and filter the *LinkOnt:SubActivity* assertions where the condition is violated. This raises the message *sh:message "Constraint LESS THAN is violated* " to the user.

5.2.2.3. *Precedence constraints*

Precedence constraints validate the sequence of activities. They are the basis of network-based scheduling methods, such as the critical path method (Darwiche, Levitt & Hayes-Roth, 1988; Dong *et al.*, 2013a). An example of a precedence constraint is that the roof must be cast after all the columns are cured. Note that these constraints do not check the assignment and availability of resources. The constraint is mathematically defined as follows:

$$\textit{Start date of the successor activity} \leq \textit{End date of preceding activity.}$$

In IFC4, an *IfcTask* class is associated with the details of the activities and has the *IfcRelSequence* as an inverse attribute. *IfcRelSequence* has attributes that define precedence relations. *RelatingProcess* defines the predecessor, and *RelatedProcess* defined the successor activities. The successor and predecessor activities have the *taskTime_IfcTask* class as an attribute to hold the start and end times of the activities as described in Subsection 3.3.1.

```
1. Ifc:IfcTask
2.  rdf:type sh:NodeShape ;
3.  rdfs:subClassOf owl:Class ;
```

```
4. sh:sparql [
5.  sh:message "Constraint Precedence condition is violated" ;
6.  sh:prefixes            <            http://www.buildingsmart-
   tech.org/ifcOWL/IFC4#> ;
7.  sh:select """SELECT $this
8. WHERE {
9. ?Sequence a Ifc:IfcRelSequence.
10.     ?Sequence                                              Ifc:
   relatingProcess_IfcRelSequence ?predecessor.
11.     ?Sequence Ifc: relatedProcess_IfcRelSequence $this.
12.     ?predecessor Ifc: taskTime_IfcTask ?predecessortime
13.     $this Ifc: taskTime_IfcTask ?successortime
14.     ?successortime Ifc: ScheduleStart_IfcTaskTime ?start
15.     ?predecessortime Ifc: ScheduleFinish_IfcTaskTime ?end
16.     FILTER( ?start <?end)
17.     }""";];
```

Listing 5.3: SHACL shape for precedence constraint

Precedence constraint requires the expressive power of SPARQL to determine the activities that have precedence relations associated with them. Hence, SHACL SPARQL is used. Listing 5.3 presents the SHACL SPARQL definition of the precedence constraint Listing 5.3. It is assigned as a node shape on *IfcTask* (Listing 5.3, lines 1-3). During the constraint checking stage, all the assertions of *Ifc:IfcRelSequence* in the data graph are filtered (Listing 5.3, line 9). The corresponding activities *Ifc*:*relatingProcess_IfcRelSequence* (predecessor) and *Ifc*:*relatedProcess_IfcRelSequence* (successor) are then identified and stored to *?predecessor* and *$this,* respectively (Listing 3, lines 10-11). The *taskTime_IfcTask* assertion (which contains the time attributes) is stored to *?predecessortime* and *?successortime* (Listing 5.3, lines 12-13). The end date in *?predecessortime* and the start date of *?successortime* are then stored to *?start* and *?end,* respectively (Listing 5.3, lines 14-15). A filter function of the SPARQL is used to identify instances where *?start* is less than *?end* (Listing 5.3, line 16). If the start date of the related process happens before the end date of the relating process, a constraint violation is detected and a *sh:message "Constraint Precedence condition is*

violated" is raised. This constraint can be reused across all the problems where the IFC process data are defined using IfcOWL following IFC4.

5.2.2.4. *Disjunctive constraints*

Disjunctive constraints ensure that processes do not overlap. Thus, if a disjunctive constraint is assigned to two activities, they should not happen in parallel. These constraints support the modelling of spatial constraints (Baykan & Fox, 1997) and safety constraints (e.g. two activities corresponding to a unique space should not take place simultaneously) among others.

To mathematically define this constraint, consider two activities: *Activity 1* with start time *StartTime_1* and end time *EndTime_1* and *Activity 2* with start time *StartTime_2* and end time *EndTime_2*. The five conditions to check whether or not they overlap are as follows:

$$StartTime\_1<StartTime\_2<EndTime\_1<EndTime\_2;$$

$$StartTime\_1<StartTime\_2<EndTime\_2<EndTime\_1;$$

$$StartTime\_2<StartTime\_1<EndTime\_1<EndTime\_2;$$

$$StartTime\_2<StartTime\_1<EndTime\_2<EndTime\_1; \text{ and}$$

$$StartTime\_1=StartTime\_2 \;\&\&\; EndTime\_1=EndTime\_2.$$

These constraints can be represented as a single statement:

$$(StartTime\_2<EndTime\_1 \;\&\&\; StartTime\_1<EndTime\_2)||(\; StartTime\_1=StartTime\_2 \;\&\&\; EndTime\_1=EndTime\_2)$$

```
1. Ifc:IfcProduct
2.  rdf:type sh:NodeShape ;
3.  rdfs:subClassOf owl:Class ;
4. sh:sparql [
5.  sh:message "Constraint Disjunctive condition is violated" ;
```

```
6.  sh:prefixes            <            http://www.buildingsmart-
   tech.org/IfcOWL/IFC4#> ;
7.  sh:select """SELECT $this
8. WHERE {
9. ?activity1 a Ifc:IfcTask.
10.     ?activity2 a Ifc:IfcTask.
11.     ?activity1 Ifc: IfcRelAssignsToProduct $this
12.     ?activity2 Ifc: IfcRelAssignsToProduct $this
13.     ?activity1 Ifc: taskTime_IfcTask ?activity1time.
14.     ?activity2 Ifc: taskTime_IfcTask ?activity2time.
15.     ?activity1time Ifc: ScheduleStart_IfcTaskTime ?start1.
16.     ?activity1time Ifc: ScheduleFinish_IfcTaskTime ?end1.
17.     ?activity2time Ifc: ScheduleStart_IfcTaskTime ?start2.
18.     ?activity2time Ifc: ScheduleFinish_IfcTaskTime ?end2.
19.
20.     FILTER (!(?activity1=?activity2)).
21.     FILTER(                                    (?start2<?end1
   && ?start1<?end2)||( ?start1=?start2 && ?end1=?end2)).
22.     }""";];
```

Listing 5.4: SHACL shape for disjunctive constraint

Listing 5.4 demonstrates a disjunctive constraint modeled as a SHACL shape. For this example, a node shape is applied to the *IfcProduct* class, which replicates the space in which an activity is performed (Listing 5.4, lines 1-3). An example of a disjunctive constraint is that no more than one activity must happen on a particular object *$this* of class *IfcProduct*. During the constraint checking stage, all *IfcTask* elements related to *IfcProducts* are filtered using the expressive power of SHACL SPARQL and are stored into two activities: *?activity1* and *?activity2* (Listing 5.4, lines 9-12). To ensure that the activities are not the same, a filter operator is used (Listing 5.4, line 20). The start and end times of *?activity1* and *?activity2* are then stored to *?start1, ?end1,* and *?start2, ?end2,* respectively, through a set of code similar to the ones mentioned in Subsection 3.3.3 (Listing 5.4, lines 13-18). Then, *?activity1* and *?activity2* are checked for overlaps using a filter operator (Listing 5.4, line 21). If there is an overlap, the *sh:message* "*Constraint disjunctive condition is violated*" is shown (Listing 5.4, line 5).

5.2.2.5. *Discrete constraint*

Discrete constraints model the resource availability constraints in construction scheduling, which ensures that the allocation of resources is made according to the availability at any time point. This constraint is frequently used in research studies on resource-constrained project scheduling problems (El-Rayes & Jun, 2009; Hu & Flood, 2012). Because this constraint is predominantly applied to resources, the IfcOWL classes *Ifc:IfcResource, Ifc:IfcRelAssignsToProcess, Ifc:taskTime_IfcTask, ScheduleStart_IfcTaskTime,* and *Ifc: ScheduleFinish_IfcTaskTime* are used to define it.

```
1. Ifc:IfcResource
2.  rdf:type sh:NodeShape ;
3.  rdfs:subClassOf owl:Class ;
4.  sh:sparql [
5.  sh:message  "Discrete  constraint  for  the  resource  not
   satisfied" ;
6.  sh:prefixes           <           http://www.buildingsmart-
   tech.org/IfcOWL/IFC4#> ; sh:select """SELECT $this
7. WHERE {
8. SELECT ?Resource (count(distinct ?activity1) as ?count)
9. {
10.     ?Resource a Ifc:IfcResource.
11.     ?activity1 Ifc:IfcRelAssignsToProcess ?Resource.
12.     ?activity2 Ifc:IfcRelAssignsToProcess ?Resource.
13.
14.     ?activity1 Ifc: taskTime_IfcTask ?activity1time.
15.     ?activity2 Ifc: taskTime_IfcTask ?activity2time.
16.     ?activity1time Ifc: ScheduleStart_IfcTaskTime ?start1.
17.     ?activity1time Ifc: ScheduleFinish_IfcTaskTime ?end1.
18.     ?activity2time Ifc: ScheduleStart_IfcTaskTime ?start2.
19.     ?activity2time Ifc: ScheduleFinish_IfcTaskTime ?end2.
20.
21.     FILTER (!(?activity1=?activity2)).
22.     FILTER(                                    (?start2<?end1
   && ?start1<?end2)||( ?start1=?start2 && ?end1=?end2)).
23.     }
24.     FILTER (?count > <No of resource>)
25.     }""" ;
26.      ] ;
```

Listing 5.5: SHACL shape for discrete constraint

Again, SHACL SPARQL is used to model the shape graph for discrete constraints (see Listing 5.5), following a logic similar to that of the previous disjunctive modelling constraint, although now the constraint is defined as a node shape on the *IfcResource* class instead of *IfcProduct* class (Listing 5.5, lines 1-3). Unlike the disjunctive constraint, there is a possibility for several activities that use a particular resource type to be executed in parallel as long as it does not violate the resource constraint (i.e., when more than one resource is available for multiple resource types). Hence, a count is taken for the number of activities that use a type of resource, and the measured count should be less than the total number of resources of that type available at any point in time.

To achieve this, I use the expressive power of SHACL SPARQL defined as two nested SPARQL queries. The inner query (Listing 5.5, lines 9-24) identifies two activities that use the particular resource at a point of time. This loop is similar to the SPARQL query defined in Subsection 3.3.4. The *IfcTask* elements with a relation assigned to *IfcResource* are filtered using the expressive power of SHACL SPARQL. Then, these are stored into two activities: *?activity1* and *?activity2* (Listing 5.5, lines 12-13). To ensure that the activities are not the same, a filter operator is used (Listing 5.5, line 22). The start and end times of *?activity1* and *?activity2* are then stored into *?start1,end1* and *?start2, ?end2,* respectively, through a code similar to the one in Subsection 3.3.3 (Listing 5.5, lines 15-20). Then, *?activity1* and *?activity2* are checked for overlaps using a filter operator (Listing 5.5, line 22). The outer query (Listing 5.5, lines 7, 8, and 25) counts the number of instances of activities using resources in parallel at a point in time, and these are then compared with the total number of resources at that point. If the count is greater than the number of resources, a constraint violation is detected, and the user is presented with a message "*Discrete constraint for the resource not satisfied*" (Listing 5.5, line 5).

5.3. Linked-data-based constraint checking implementation

After the description of the LDCC method and how constraints are modeled as SHACL shapes in the previous section, this section presents its implementation (through a web application) to validate the construction scheduling constraints in a look-ahead planning environment. I tested the 4D model from the Schependomlaan OpenBIM dataset (Van Berlo, 2018). It contains construction-schedule information at the master planning level, and its activities were enriched by adding further subtasks and constraints. I manually checked the results from the LDCC application to confirm it was able to identify constraint violations.

5.3.1. Prototype web application

The LDCC architecture, discussed earlier in Section 5.2, was implemented as a prototype web application, with its functional details shown in Figure 5.3. The information required by the developed software is imported from an existing BIM model and enriched with user inputs. Users can interact with the web application through the front-end, which contains two UIs (see Figure 5.4):

4. Model interface – to interact with the geometric model; and
5. Task-edit interface – to interact with a Gantt chart to view task information, and to further enrich the imported data with user inputs.

The back-end of the web application consists of a processing web application and its associated data storage. Details of the different components are explained in this section.

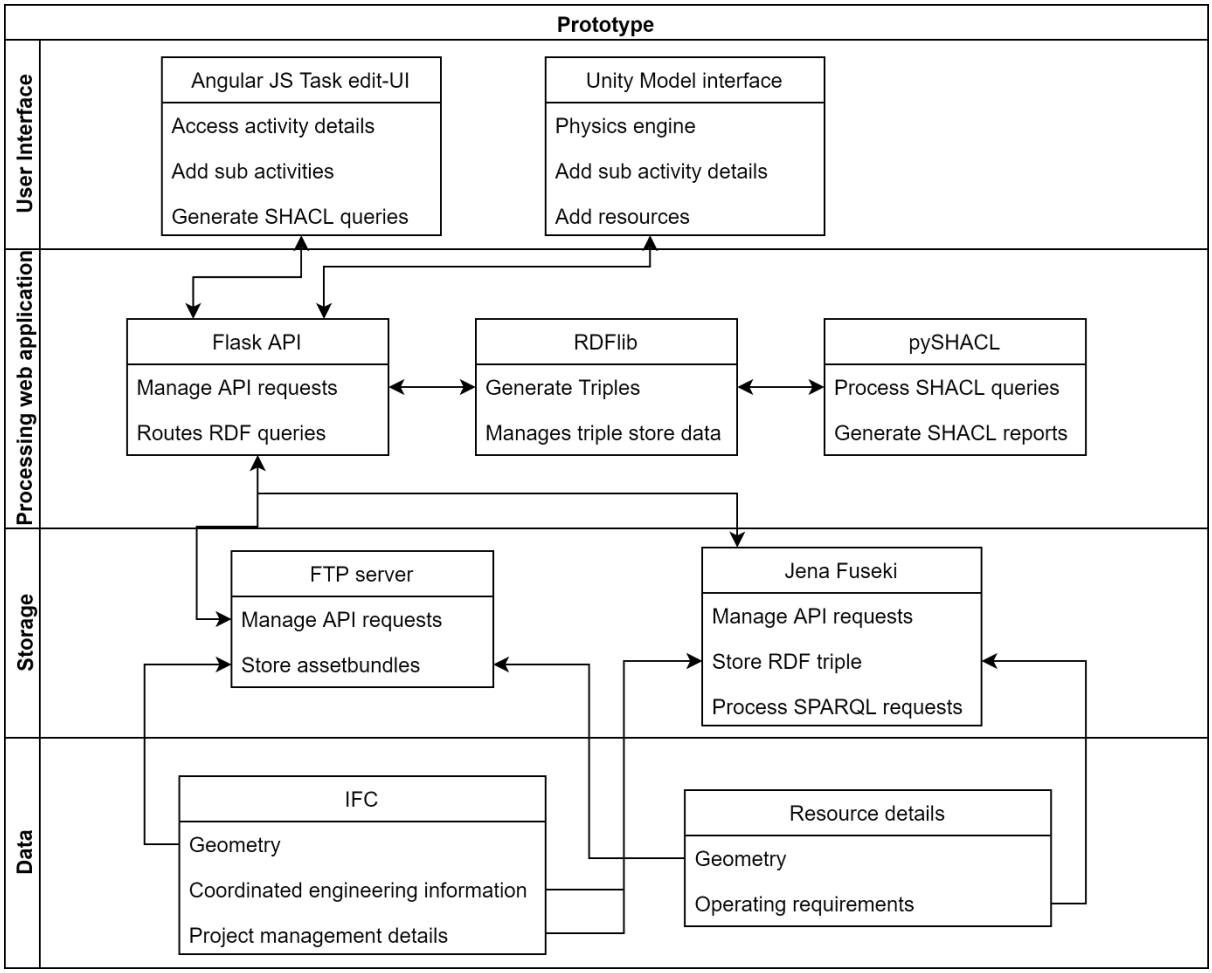


Figure 5.3: Functional details of the software implementing the LDCC method

5.3.1.1. *User interface*

The UI presents the imported data to the user, which can be further enriched with user inputs. The UI consists of the model interface and the task-edit interface, which were implemented separately. The model interface was developed using the Unity game engine (UnityTechnologies, 2019). The model interface enables 3D interaction with the construction site model (e.g., selecting the geometry of the objects, moving the resources, etc.) and can be deployed to any device, including augmented reality devices (e.g. Microsoft Hololens), virtual reality devices (e.g. HTC Vive), or a computer screen through a web browser. To interact with the model view, I used the 3Drepo.io unity plugin (Friston *et al.*, 2017). The Unity game engine

has an active physics engine which ensures that the model manipulations follow real-world physics.

The task-edit interface was developed using Angular JS (Google, 2019). It allows users to manage the project data (e.g., activities, sub-activities sequences, resources, and constraints) and facilitates the enrichment of tasks with further subtasks, resources, and constraints. The task-edit interface contains a Gantt chart, generated from the planning data that was embedded in the BIM model. When a task in the Gantt chart is clicked, the right side of the interface is activated. It shows existing task details, such as task name, resource details, and added constraints. There are provisions to enrich the existing information with additional resources and constraints. Imported data, as well as user inputs, are transformed into data graph and shape graph using the processing web application.

Both interfaces in the UI are connected. When a task is selected in the task-edit interface, that task could be manipulated using the model interface. All the manipulations in the model interface are recorded and transformed into sequences in the task-edit interface. Details on how this interface is used to perform constraint-checking are explained in Subsection 5.3.2 (Prototype evaluation).

1

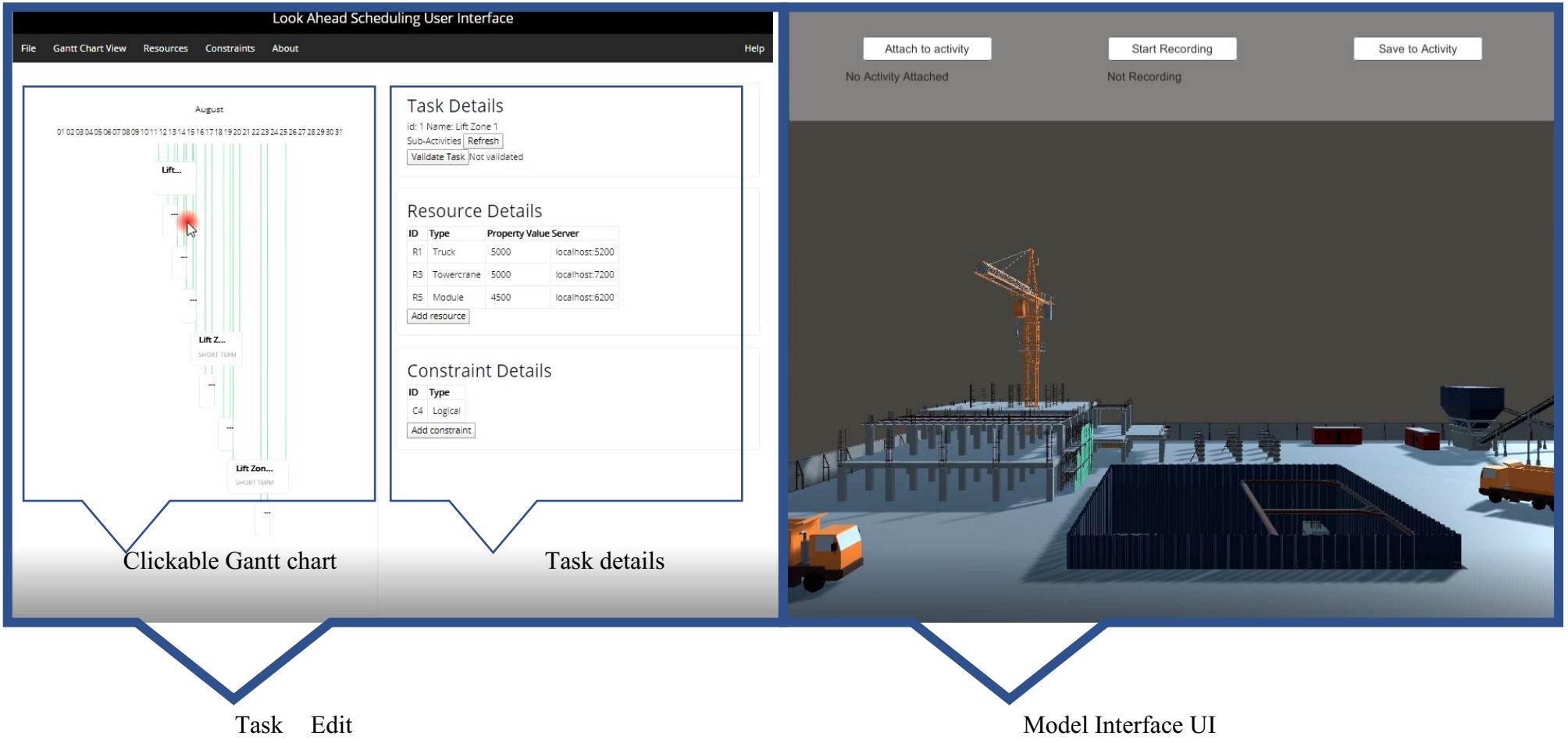


Figure 5.4: User interface

5.3.1.2. *Processing web application*

The processing web application is the main application that stands between UI and data storage. It processes the inputs from the users, converts them to RDF graphs, acts as a SHACL processing engine, and displays validation results. The backend of the processing web application is a Flask application (Flask, 2010) which uses the RDFLib library (RDFLib, n.d.) to process the data modeled in RDF, and validate the constraints expressed as SHACL SPARQL functions using the pySHACL library (pySHACL, n.d.). The construction information was stored as RDF triples following OWL, RDF, ifcOWL (Pauwels & Terkaj, 2016; Beetz, Van Leeuwen & De Vries, 2009; Pauwels *et al.*, 2017), LinkOnt and SHACL ontologies. The ontology layer was handled through RDFLib, which accesses the schema definitions over the internet to determine the relationships between data.

5.3.1.3. *Storage*

Two pieces of storage are required. The first one is an FTP server that holds the geometry of the resources and the construction site model. The model interface of the UI was developed using the Unity game engine and required an AssetBundles file (UnityTechnologies, n.d.) (AssetBundles are archive files that contain Unity-specific non-code assets, such as geometry and textures, that Unity can load at run time). Asset bundles are downloaded from the FTP server during the runtime using Unity API calls to the FTP server.

The second piece of storage is used for storing specific semantic information about the construction, such as operating conditions of the resources, engineering information embedded in the model, and the project management details. These are stored as RDF triples in a Jena Fuseki server (Apache Jena, n.d.). The data are accessed by the processing web application using HTTP requests. The Jena Fuseki server can also handle SPARQL requests from the processing web application to fetch data.

5.3.1.4. *Data*

For the application to work, two sets of data are required. The first set includes information required for managing the construction site, such as building information model, construction-site characteristics, and project management details. The construction information can be obtained from the BIM model exported in the IFC format. BIM model contains the geometry of the structure, the coordinated engineering information, and the project management details. The second dataset required by the model includes information for resource management. This dataset should include the details about the resource, including the geometry of the resource and its operating conditions. This needs to be obtained from subcontractor servers (e.g., using an API to the subcontractor common data environment or subcontractor enterprise resource planning systems).

5.3.2. **Prototype evaluation**

To evaluate the developed prototype, I used the Schependomlaan open BIM dataset (Van Berlo, 2018) as-planned model from Week 26. These constraints were then evaluated using the developed web application. This evaluation enabled us to demonstrate the potential of LDCC to model and check complex and dynamic scheduling constraints in construction. Evaluation is performed by testing LDCC to define and execute complex construction scheduling constraints, such as logical constraints, precedence constraints, disjunctive constraints, and discrete constraints.

Table 5.1: Task details used as inputs for the evaluation

Sub Activity	Main activity	Start date	End date	Resource assignment	Object assignment	Precedence
S1	T1	21/08/2015	23/08/2015	R1, R3	O1	
S2	T2	24/08/2015	27/08/2015	R2, R4	O2	S4
S3	T2	28/08/2015	30/08/2015	R1, R3	O2	
S4	T2	22/08/2015	25/08/2015	R2, |R4	O2	

Table 5.1 provides the data given as input to the application to enrich the information from the Week 26 planning file of the Schependomlaan dataset (Figure 5.5). Two instances of the *IfcTask* class were enriched (referenced as Tasks T1 and T2 in this study). Enrichment is done by clicking the task in the Gantt chart window shown in Figure 5.4. Once a task is clicked, the details of that task are displayed in the task details pane (Figure 5.6). Resources and constraints are added using that pane. During enrichment activities, S1, S2, S3, and S4 were added as *SubActivities* to tasks T1 and T2. S1, S2, S3, and S4 are lifting activities using cranes R3 and R4 (from trucks R1 and R2) to lift modules O1 and O2, respectively. Moreover, R1, R2, and R3 were assigned operating requirement values of 5 tons, while R4 was assigned an operating requirement of 3 tons. All resources had a discrete constraint applied to them to ensure the resources are allocated as per the availability. In addition, logical constraints were assigned to the resources. Two module objects O1 and O2 (with an operating requirement value of 4.5 tons) were associated with the activities S1, S2, S3, and S4, and disjunctive constraints were applied to ensure that only one subtask is performed on an object at one point in time. R1, R2, R3, R4, O1, and O2 were stored on external servers (to replicate the case of the construction, where this information is hosted in subcontractor servers). A precedence constraint was applied to activity S2 so that S2 can only start after S4. In addition, to demonstrate logical constraints, a simple *less than* was applied for all the resources. The less than logical constraint ensures that the value of the operating requirements of the resources is less than that of the operating requirement defined in the object associated with that activity.

When the *Validate Task* button for an activity is clicked (Figure 5.7), the data associated with the activity are checked against the shapes. Additionally, other activities happening at the same time and their associated resources are also checked for conflicts. Table 5.2: shows the results from the constraint checking process performed by the application.

Table 5.2: Results after constraint checking

Subtask	Result
S1	No constraint violation
S2	Logical constraint violation, disjunctive constraint violation, discrete constraint violation, precedence constraint violation
S3	No constraint violation
S4	Logical, disjunctive constraint violation, discrete constraint violation

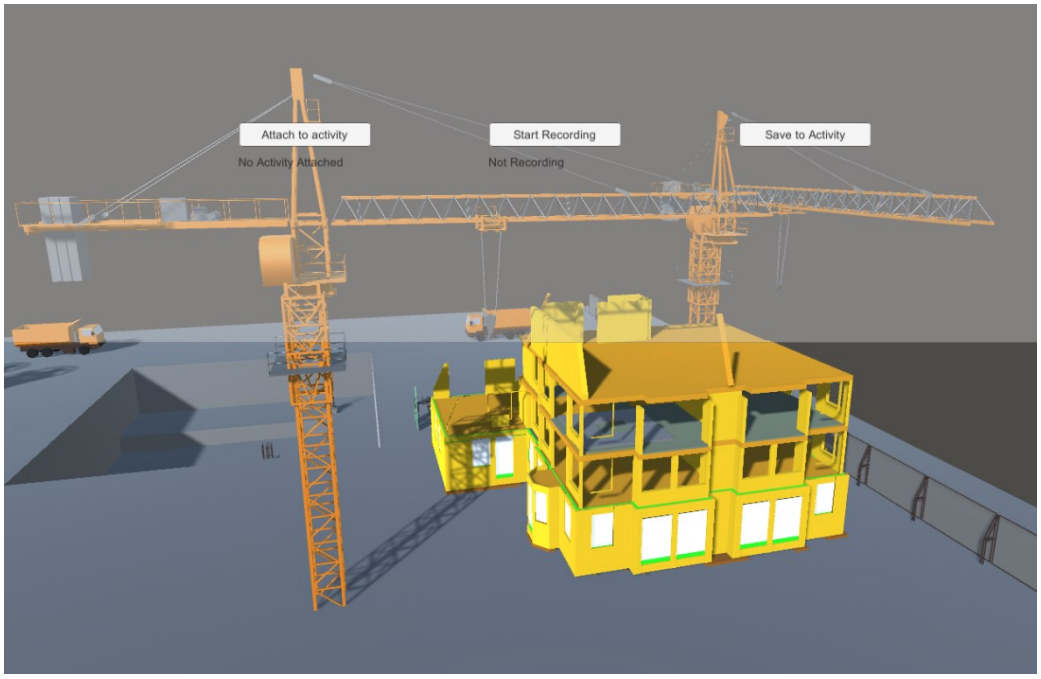


Figure 5.5: Screenshot of the imported Schependomlaan dataset in an enriched site environment.

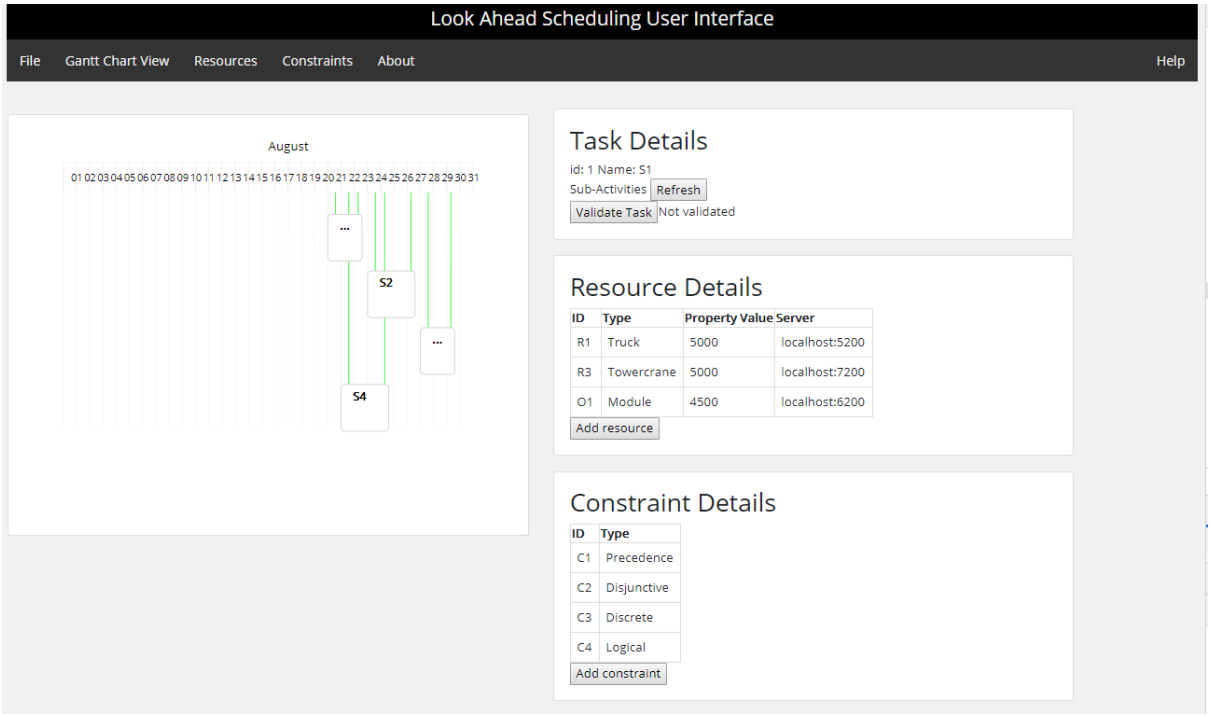


Figure 5.6: Task-edit interface for the imported dataset. The data shown refer to SubActivity S1

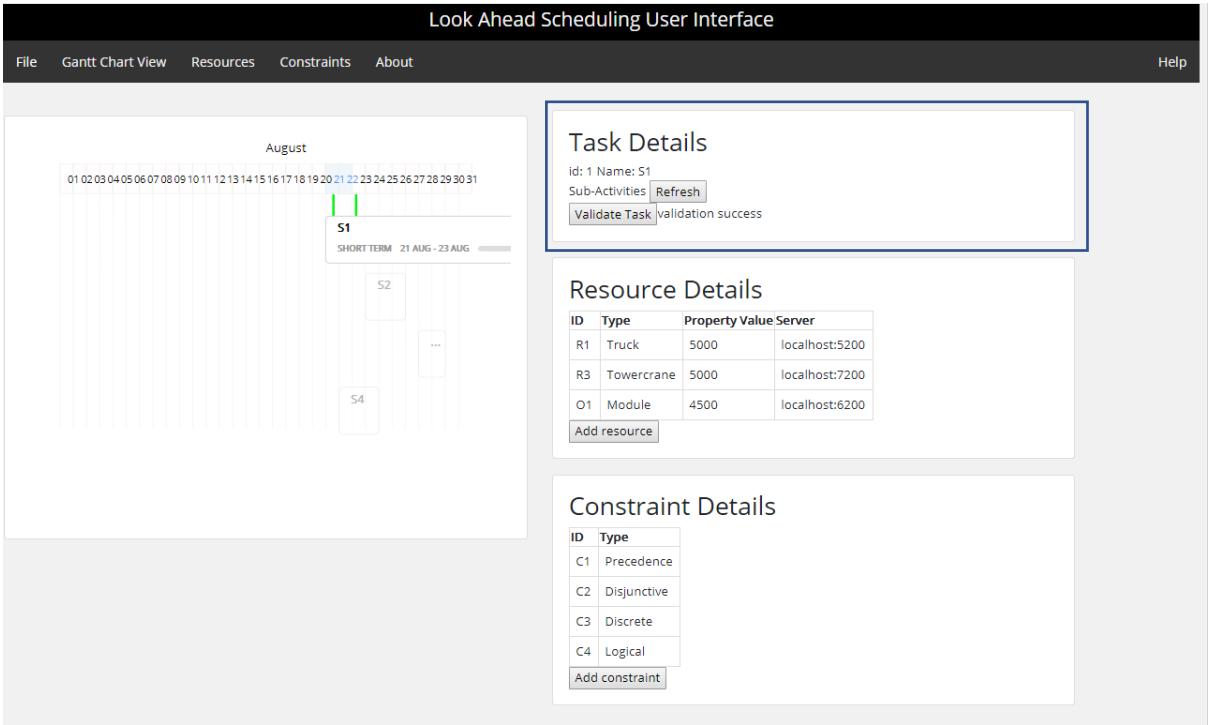


Figure 5.7: Successful validation of S1 activity.

S

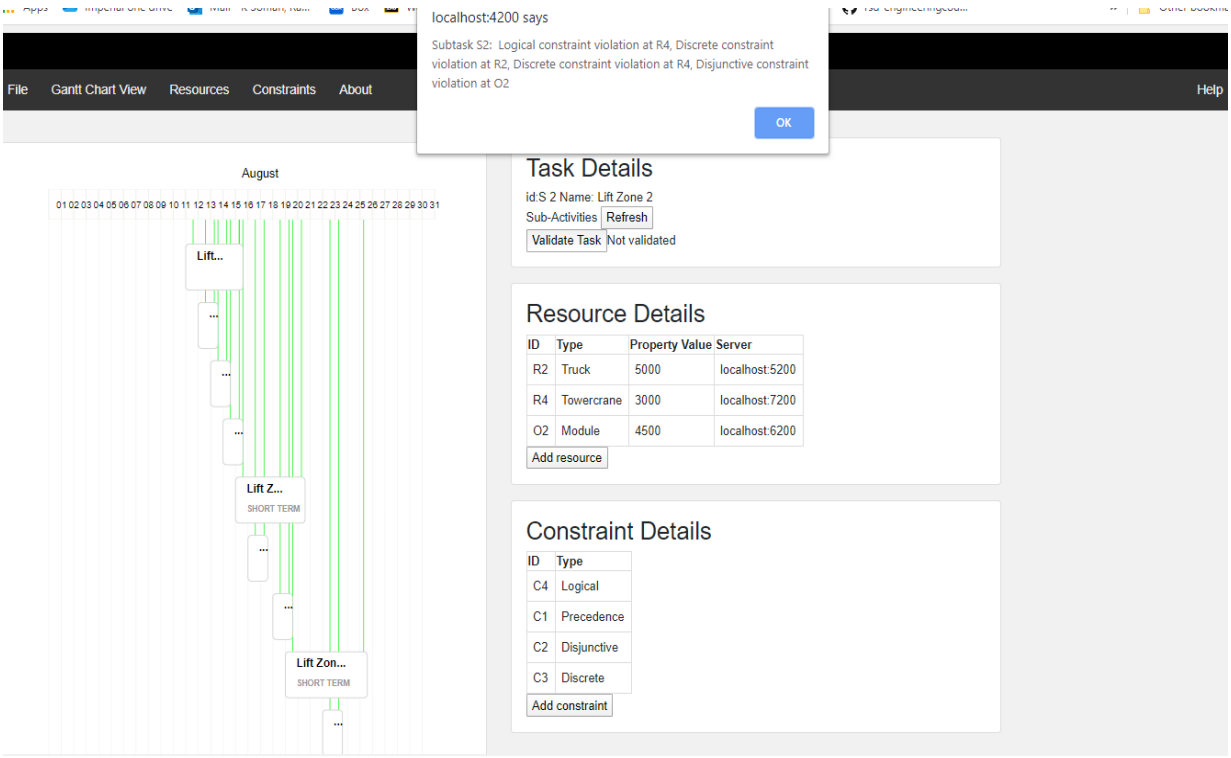


Figure 5.8: Validation failure for S2 activity.

The prototype application developed based on LDCC was able to identify several constraint violations using information distributed over multiple servers. There was no constraint violation for S1 (Figure 5.7) and S3, but S2 violates all the constraints (Figure 5.8), and S4 violates discrete and disjunctive constraints. In addition, S2 and S4 overlap each other on 24 August 2015 (which violates the precedence constraint applied on S2). Moreover, S2 and S4 share resources R2 and R4, thus violating the discrete resource constraints. Similarly, S2 and S4 operate on the same object O2, violating the disjunctive constraint, as two activities cannot happen in the same space simultaneously. The logical constraint is violated for these activities, as the operating requirement value of R4 is less than the operating requirement value of O2.

5.4. Implications of LDCC

The LDCC method follows the principles of BIM based rule checking. It follows the approach for rule checking suggested in Eastman *et al.* (Eastman *et al.*, 2009), discussed in Subsection 2.3.2, where rules are translated into SHACL automatically by engaging with a UI.

Construction information is converted to an RDF graph (machine-readable format) for easy execution of constraints. The SHACL constraints are then checked for constraint violation using a processing web application enabled by RDFLib and pySHACL. The results of constraint checking are reported back to the user through the UI. Instead of automated distilling of rules from the textual content, rules are defined as constraints by the user in LDCC. LDCC reaches Class 3 complexity, as described by Solihin and Eastman (Solihin & Eastman, 2015) because the data structure is extended and traversed to provide logical proofs for constraint checking.

The evaluation I performed demonstrates the potential of the proposed LDCC to test for constraint violations. In contrast with previous studies that used BIM based rule checking (e.g., (Schwabe, Teizer & König, 2019; Chua & Yeoh, 2009)), the implemented prototype was able to identify all the constraint violations and inform the user about them even when the data are distributed over multiple servers. When a discrete constraint was applied to the resources (R1, R2, R3, and R4), the resources do not have any schedule data associated with them in their servers. The schedule data were stored in the activities, as shown in Table 5.1, and resources and objects were in separate servers making the existing methods unsuitable for modelling this complex constraint. Despite this, the LDCC query could crawl through the data to identify relevant time data from the resource assignments and make logical inferences (also reported in prior research (Pauwels *et al.*, 2011; Pauwels, Zhang & Lee, 2017)), to identify whether there were multiple assignments during the same period. This is possible because the ontology layer is separate from the processing layer; therefore, queries need not have the know-how of the whole data structure schema but only the attributes necessary for making the traversal. The remaining semantics for the data traversal is handled through reasoning offered by the OWL and RDF ontologies.

Although example in the evaluation could have been done manually, the real power of the proposed LDCC method emerges as the scale of the problem becomes larger. As the scale of the problem increases (e.g., managing the resources and space allocation of multiple trades working in a large construction site with numerous shared resources), manual methods to identify constraint violations among multiple interdependent activities and resources becomes difficult (Hamzeh, 2009; AlSehaimi, Tzortzopoulos & Koskela, 2009) and results in errors (Lindhard & Wandahl, 2014). However, this would not be a problem in LDCC which uses SHACL to identify constraint violations. Prior studies which uses SHACL to perform constraint validation reported that SHACL can process the validation within reasonable time even for very large data graphs, providing evidence for scalability of this approach (Corman *et al.*, 2019). Thus, using computer-based methods such as LDCC, the scale of the problem is no longer a critical issue, as the computer can process these interdependencies as demonstrated as the processing mechanism remains the same, and the speed restricted by only the processing capability of the computer.

There are multiple benefits for using the proposed method in look-ahead planning. The first benefit of LDCC is that it not only provides automatic constraint checks during the look-ahead-planning meetings but also captures the construction knowledge generated and discussed in them and codifies it into a machine-readable format, which can be reused later (as described in section 3.1). This reduces the need for data input after a stage of saturation, as the shape graph is populated during subsequent sessions. The second benefit of this approach is that it can accommodate the dynamicity of constraints. Data and constraints are added to the incrementally, and constraint checks can be performed automatically whenever there is a change in data graph or shapes graph. For example, if the operating requirement of equipment changes and if the new operating requirement is modeled as a new shape (constraint), all the assignments of the equipment on activities, including the ones that had already been checked

for constraint violation, would be re-evaluated for the new constraint. This enables stakeholders to identify the effect of changes in the project variables as and when they arise. This feature also enables multiple parties to make changes to the schedule while maintaining the integrity of the schedules. When linked with other communication platforms, LDCC can inform the people responsible for the activities of the violations, if any. Therefore, LDCC acts as a "unit-test" environment for short-term schedule data.

This study is an important first step toward linked-data-based rule-checking for construction scheduling constraints. The LDCC method demonstrates the effectiveness of the approach to address the challenges existing in checking for construction-constraint violations. The limitations of the prototype include that the ifcOWL ontology that is used for demonstration is based on IFC, which is not designed to be used in linked-data applications (Pauwels & Roxin, 2016). Hence, defining SHACL constrained by following ifcOWL ontologies is complex and requires a good grasp of the structure of IFC. In addition, the SHACL shapes are defined as snippets (for common scheduling constraints), which are then used to create the exact constraints through the web application. This limits the definition of complex constraints with the help of the UI. However, if necessary, the SHACL shapes can be manually created and added to the shape graph with minimal effort by a trained engineer. In addition, this work did not explore the use of semantic relationships within the geometry (Zhang & Beetz, 2016b) to do more constraint checks, such as the ones demonstrated in (Schwabe, Teizer & König, 2019). These geometry-based checks are possible; however, they need further investigation before they can be defined as SHACL constraints. Evaluation of the developed prototype and framework was done using a public construction dataset as a proof of concept. There is a need to test this framework in actual look-ahead planning meetings to demonstrate the effectiveness and understanding of the limitations. As discussed in the next section, this opens up avenues for future research in the area of linked-data in architecture and construction,

focusing on geometry-based constraint checks and simple ontologies to describe construction process information.

5.5. Summary

This chapter extends existing research on BIM based rule-checking systems and linked-data technologies in the construction sector to develop a linked-data-based constraint checking method to model and validate construction scheduling constraints. This novel method addresses the inherent challenges in construction-constraint datasets, such as spanning across multiple domains and being distributed among heterogeneous siloed datasets. LDCC does this by leveraging the power of linked-data to traverse across the web of data created by linking across the domains and datasets. Construction-scheduling constraints, such as cardinality constraints, logical constraints, precedence constraints, disjunctive constraints, and discrete constraints, were modeled using SHACL. The study contributes by demonstrating the potential of this approach through the development and evaluation of a prototype web application based on LDCC. This application could identify construction scheduling constraint violations for a set of input data distributed over multiple servers. The LDCC demonstrated in this study is of value to researchers working on the lean construction, BIM based rule-checking and linked-data, as it demonstrates the power of the linked-data to check for constraint violations.

Further, the LDCC method can be combined with machine learning frameworks and discrete-event simulation tools to reschedule the processes and aid decision making during disruptions because the current state of work and the constraints are codified. In contrast with prior studies focused on generating optimal construction schedules addressing constraints, such as crew availability and material availability (Liu & Lu, 2018; Liu *et al.*, 2020), LDCC method does not require mathematical modelling of constraints to generate schedules. In addition, LDCC eases this task by taking in construction constraints spanning multiple domains and distributed over heterogeneous datasets from the user, converting them into a machine-readable

format. The LDCC environment thus acts as a processing engine for simulations and machine learning algorithms, where it is cumbersome to formulate numerous constraints mathematically. For example, in the case of genetic algorithms, the LDCC environment can be used to check the fitness value of the individuals when the constraints cannot be modeled a priori. Similarly, for a reinforcement learning algorithm, LDCC can reside within the environment to evaluate the reward for the actions taken by the agent. Thus, the proposed LDCC can be used to automatically reschedule tasks to create a feasible look-ahead schedule when there are disruptions, thereby supporting look-ahead planning. Next chapter describes the development of two algorithms to generate LAS automatically, that makes use of the LDCC method for constraint checking..

6. AUTOMATING LOOK-AHEAD SCHEDULE GENERATION

6.1. Introduction

The previous chapter focused on codifying the construction process information using linked-data principles and described LDCC to assist look-ahead planning. This chapter builds upon those two pieces to generate LAS automatically using machine learning algorithms in order to eliminate inefficiencies during the manual generation of LAS.

Look-Ahead-Scheduling is a type of RCPSP, that stand between overall project coordination scheduling and planning for short-term weekly plans for crew-level commitments, as explained in Subsection 2.1.2. Classic RCPSPs include schedules (at the masterplan or phase plan level) characterized by a large number of activities, happening over a long time (6 months to a few years) and constrained predominantly by precedence relationships. On the other hand, LAS is characterized by a small number of activities happening over a short time period (1 to 6 weeks) and constrained by a large number of factors including but not limited to precedence relationships, workspace conflicts, logistic requirements, and equipment operating criteria. In addition, the constraints in LAS are complex and spans multiple domains which increase the complexity of the problem. Although there are several algorithms to solve RCPSP (predominantly at masterplan and phase plan level), there is a need to investigate the applicability of these algorithms to generate LAS, to accommodate the differences between classic RCPSP and LAS generation problem.

In this chapter, I investigate two strategies to automate the generation of LAS. Exact optimisation methods take a lot of time to solve fairly complex RCPSP such as LAS problem

(Appendix L, Rodrigues & Yamashita, (2015)). Hence, they are not detailed in this chapter. The first strategy employs a GA—a type of algorithm which has been effectively used to solve RCPSP problems (Elbeltagi, Hegazy & Grierson, 2005; Faghihi *et al.*, 2015; Liu *et al.*, 2020)—and it uses an activity-oriented serial generation scheme to generate LAS. The second strategy makes use of reinforcement learning—recently used in solving highly constrained RCPSP problems (Koulinas *et al.*, 2018; Jedrzejowicz & Ratajczak-Ropel, 2014; Fonseca-Reyna *et al.*, 2019)—and takes a time-oriented parallel schedule generation scheme to generate LAS. Both these strategies use the LDCC method described in Chapter 5 to evaluate the constraints, thereby evaluating the applicability of using the dynamic constraint checking environment to automate LAS generation. I compare these strategies to check their effectiveness in solving LAS problems of increasing complexity through a series of experiments. By doing so, I can compare the suitability of serial and parallel schedule generation schemes for LAS generation.

The rest of the chapter is divided into five sections. Section 6.2 defines the LAS problem, Section 6.3 describes how to model it using a GA, and Section 6.4 describes how to model it using reinforcement learning. In Section 6.5, computational experiments are performed to test both approaches and compare them on the basis of the project duration and CPU time. Finally, Section 6.6 summarizes the chapter.

6.2. Formulating the LAS problem

The objective of the look -ahead planning is to the schedule the activities such that the project execution time and the cost of the projects are minimised. The LAS problem can be formulated as a maximisation problem as below:-

Maximise:

$$W_1 \text{ x } F_A + W_2 \text{ x } F_R$$

Where:

$$F_A = \frac{1000}{Project\ finish\ time}$$

$$F_R = -\frac{1}{100}\sum_{i=1}^{n} Cres_i$$

Here W_1 and W_2 are weight determined for the time and costs respectively. These may be altered as per the requirements of the project at the time of look-ahead planning. $Cres_i$ is the cost of deploying a resource i.

$$Cres_i = Dtotal * Cost_{per\ unit\ time} + (MobilisationCost + DemobilisationCost) * Nassigned$$

Dtotal is the Total duration assigned, and $N_{assigned}$= Number of activities the resources are assigned.

Subject to the constraints:

1. Precedence constraint: $F_h \leq F_i - d_i, i = 2, \ldots j; h \in P_i$;
2. Discrete resource constraint: $\sum_{i \in A(t)} r_{ik} \leq R_k, k \in K, t \geq 0$; and
3. Other site-based constraint violations: N_c =0.

Here F_h is the finish time of the predecessor activity *h,* where P_i is the set of predecessor activities for the activity *i*. F_i denoted the finish time of the activity $i \in V$ (*V* is the series of activities to be scheduled), d_i represent the duration of activity *i*, and *j* represent the index of the last activity. r_{ik} represents the number of the renewable resource of type $k \in K$ assigned to the activity *i* for a unit time. The number of resources of type k available at each unit time period is R_k. The parameters d_i, r_{ik} and R_k are all positive integers. *A(t)* is the set of all

activities which require the resource k at the time period t. N_c is the number of constraint violations detected by the LDCC method.

6.3. Solving LAS problem using a GA

Genetic algorithms are a type of metaheuristic algorithms which take a population-based evolution strategy to solve optimization problems in large search spaces. It mimics the biological evolution process in combinatorial optimization problems (Holland, 1992). The basic steps of a GA are represented in Figure 6.1, and Figure 6.2 shows how these steps can be used to generate LAS.

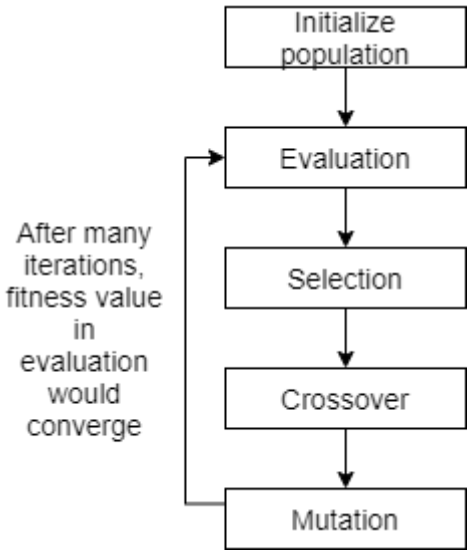


Figure 6.1: Genetic algorithm steps

The first step in a GA is to create a population of possible solutions. The solutions are called individuals or chromosomes. A fitness value is calculated for each individual in the population, by minimising or maximising an objective function. Once the fitness value is calculated, the best individuals (those with the best fitness) are selected, and a crossover operation is applied to them. This crossover operation combines the genetic information of two selected individuals, known as parents, to generate a new individual called offspring which will have some qualities of the parents. Multiple such crossovers are performed to create a new population. Once the new population is created (forming a new generation), some of the

individuals in the new population are altered in a step known as mutation. The goal of mutation is to maintain the genetic diversity of the population from previous generations (i.e. given some weight to exploration). The cycle then starts again with each individual in the new generation being evaluated for fitness. Multiple generations are executed to ensure an adequate balance between exploration and exploitation of most promising individuals is produced. The procedure is said to converge if, after a certain number of generations, it is unable to find solutions with a better fitness value. The rest of Section 6.3 will describe the different steps in

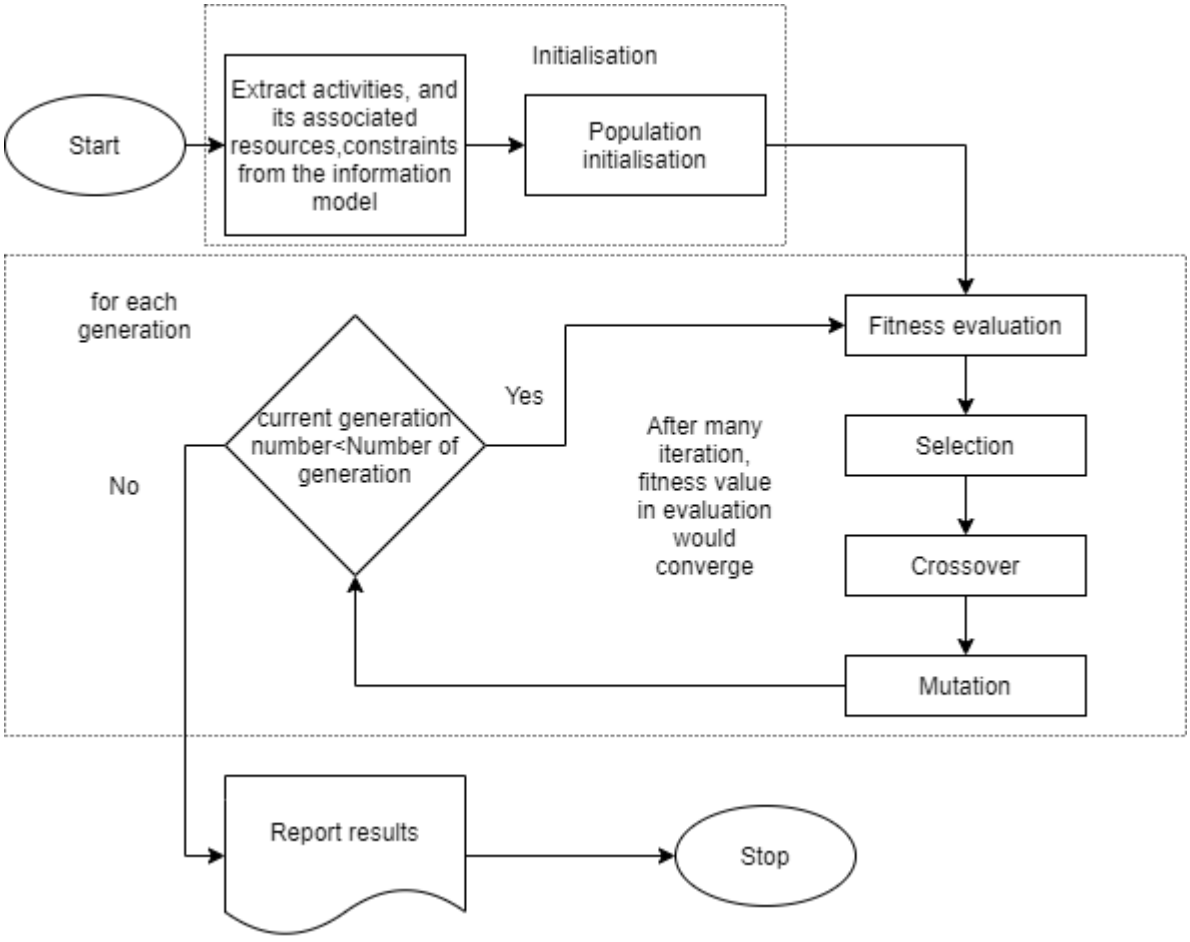


Figure 6.2 LAS generation using genetic algorithm

the GA (as shown in Figure 6.2) such as population initialization, fitness evaluation, selection, crossover and mutation for LAS generation.

6.3.1. **Population initialization**

Each individual in the population needs to be represented mathematically to prepare for the use of GA. This representation is called a chromosome, and the process of representing the information is called encoding (Goldberg & Holland, 1988). For encoding schedule information for GA, previous research (Kolisch & Hartmann, 2006) has shown activity list as one of the best methods followed by using a schedule generation scheme, and thus I will follow this approach to generate the LAS. In an activity list scheme, each activity is placed at a position before all its successors and after all its predecessors. In the LAS problem, there are multiple trades working in a space, and each space has a process to be completed comprising of multiple activities in a sequence. Therefore, the activity list matrix (A) is created with *m* rows as processes and *n* columns as activities associated to that process, where the right column comprises of an activity which is the successor of the activity on the immediate left in that column.

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix}$$

Each row in the matrix represents a process. Each column in the matrix represents an activity in that matrix. Some of the processes might not have the same number of activities. In such cases, the last empty columns of the process with a lesser number of activities are replaced with dummy activities. For example, if Process 1 has ten activities, and Process 2 contains seven activities, the last three columns of Process 2 would be replaced by dummy activities.

To create a schedule from the activity list, a schedule generation scheme is used. There are two types of schedule generation schemes: *Serial schedule generation* scheme (that is activity-oriented) and *parallel schedule generation scheme* (that is time-oriented). Usually, the output of the *parallel schedule generation scheme* is a subset of the *serial schedule generation*

scheme. An activity-oriented serial schedule generation scheme is used for the current implementation as it removes the issue of getting into a local optimum (Kolisch, 1996b). However, it is difficult to evaluate the time-oriented disjunctive and parallel constraints while using the serial schedule generation scheme. This problem can be mitigated by including the constraints in the fitness evaluation phase using the dynamic constraint checking method developed in the previous chapter. An activity list is converted to a schedule by adding the time increments to the activities in the list. Therefore, the activity list matrix A is converted to a Schedule matrix S by adding the start time of activities in place of the activity.

Before generating the initial population, the earliest start time and latest start time of each activity in the process is calculated using the critical path method. This would help the reduction of the search space. The first population of schedules is created by randomly choosing the start time of each activity which lies between the earliest start and latest finish time, satisfying the precedence and resource constraint. The earliest starting time is calculated by considering each process starts at the zero-time instance. However, since a schedule with earliest finish time may not satisfy the resource constraints, the latest finish time of the project is the earliest finished time multiplied by a factor. The factor is dynamically created by iterating to make sure that a certain number of schedules satisfies the precedence and resource constraints can be created. To do this, a factor of 2 is used initially, and a certain number of precedence feasible schedules are created. These schedules are checked for resource constraint feasibility, and the schedules violating this constraint are eliminated. The total number of feasible schedules (Both precedence and resource constraint) are counted. If the count is less than a certain population count, the factor is increased, and the process is repeated, until a certain number of feasible populations. Each chromosome in the population is represented by the following matrix where $S_{,m,n}$ is the start time of the activity where m is the index of the process and n is the index of the activity.

$$S = \begin{bmatrix} S_{1,1} & \cdots & S_{1,n} \\ \vdots & \ddots & \vdots \\ S_{m,1} & \cdots & S_{m,n} \end{bmatrix}$$

6.3.2. **Fitness evaluation**

The fitness of the chromosome indicates how good or bad a solution is in terms of an objective function, and as I have explained before, in the context of GA, this serves as the criteria to select a chromosome for the next generation, thus determining how the search of a suitable schedule is driven. For the current case, the domain of the function is (- ∞,∞). Therefore, the fitness function should consider all factors necessary for the generation of a good LAS in its formulation. The LAS problem focuses on scheduling the activities in such a way that there are no constraint violations. In addition, LAS should ensure that the resources employed should be utilized to the maximum (so that the cost is reduced), and the processes and activities to be scheduled should finish in the least amount of time (thereby reducing the total cost, as there are indirect costs associated with the additional time for process completion). Therefore, the fitness function will depend on:

1. Constraint satisfaction
2. Completion of the projects within minimal time
3. Lowest resource cost.

The constraint satisfaction is evaluated by LDCC, the SHACL-based dynamic constraint checking framework introduced in the previous chapter. Each chromosome in the population (i.e. a generated schedule) is converted to an RDF graph and is validated against the different constraints (precedence, discrete, disjunctive, and logical constraints). If no constraint is violated, a value zero is returned. Otherwise, the number of violations *Nc* is returned.

As chromosomes that do not satisfy all constraints are not really 'fit', we apply a penalty to their fitness score. This penalty is proportional to the degree of violation and is much higher than the normal fitness value (Smith & Raphael, 2013). Therefore, the first component (*Fc*) of the fitness function (accounting for constraint violations) is given by (note that with this formulation, solutions with no constraint violation get a value of zero):

$$Fc = -1000000 * Nc$$

Secondly, to promote chromosomes finishing the project at an earlier time, the fitness function should be inversely proportional to the end time. The second component (F_A) precisely captures that:

$$F_A = \frac{1000}{Project\ finish\ time}$$

Finally, to accommodate the cost of hiring resource and ensuring that the resources are used in a uniform manner, the third component of the fitness function (F_R) is associated with cost. Every resource is assumed to be hired for unit time costing (for example, cost is calculated per hour, per day or per week depending on the resource). In addition, there are costs for resources mobilization and demobilization of resources for each activity. The cost of deploying resources *i* is given as:

$$Cres_i = Dtotal * Cost_{per\ unit\ time} + (MobilisationCost + DemobilisationCost) * Nassigned$$

Dtotal= Total duration assigned, Nassigned= Number of activities the resources are assigned

F_R is calculated by reducing the cost of total resource usage.

$$F_R = -\frac{1}{100}\sum_{i=1}^{n} Cres_i$$

Therefore, the final fitness function is given by:

$$F=F_C+F_A+F_R$$

The relative weights on the equation can be changed to accommodate different scenarios in which criteria have different importance.

6.3.3. **Selection**

Once the fitness value of each individual in a population is calculated, the next step in the GA is to perform a selection. After going through multiple selection strategies such as roulette wheel, rank selection, and tournament selection, I chose the elitist strategy (Goldberg & Holland, 1988), a selection method which ensures that individuals with the best fitness are selected to remain in the next generation. Those individuals with maximum fitness are identified as elite individuals and directly transferred to the next generation. Remaining individuals are placed in a mating pool, where crossover and mutation operators will be applied to them in order to create offspring. Crossover and mutation add diversity to the population, resulting in wider exploration of the solution space and thus in higher chances of creating an offspring that will lead to higher fitness values in subsequent generations. The elitist strategy ensures that the best individuals survive into the next generation, thus accelerating the convergence rate.

6.3.4. **Crossover**

Crossover is a genetic operator to combine two parents to create an offspring. For the LAS problem, one-point crossover operator, well suited for RCPSP problem (Hartmann, 1998; Elbeltagi, Hegazy & Grierson, 2005) is used. Whether a crossover operator is used or not is determined by a set crossover probability *Cprob*. To apply the crossover operator, two-parent chromosomes are selected. Each chromosome is represented as a matrix of size m x n, where m is the number of processes, and n is the number of activities. The crossover point c is randomly generated where c satisfies $0<c<n$.

Two parents, $J^p = (J_1^p, \ldots\ldots J_n^p)$ and $J^q = (J_1^q, \ldots\ldots J_n^q)$ are used to illustrate the crossover operator. After a crossover point, c is generated, start times of the first c activities of the child J^d are selected from J^p for all the processes.

$$J_i^d = J_i^P, where\ 0 \leq i \leq c$$

The start times of the remaining n-c activities of J^d are selected from J^q.

$$J_i^d = J_i^q, where\ c < i \leq n$$

Resource distribution in the offspring would satisfy the resource constraint as the parent satisfies resource constraint. To make sure that the crossover operation results in a precedence feasible schedule, the offspring is then checked for precedence feasibility within the process. If offspring fails the precedence check, the offspring is not added to the next generation, else it added to the population of the next generation. This is illustrated in Figure 6.3.

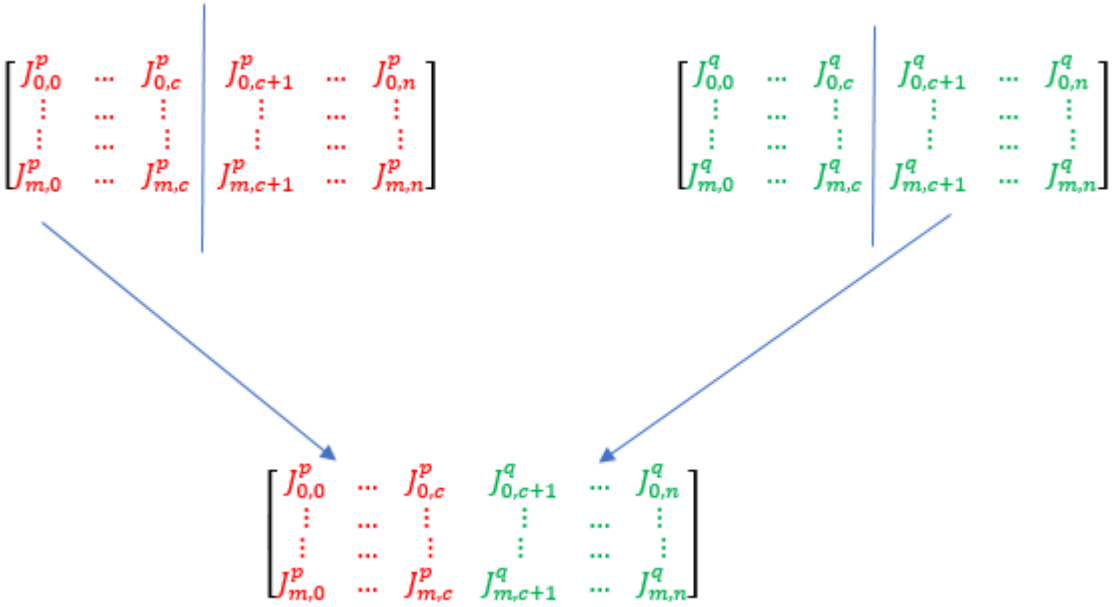


Figure 6.3: Crossover operation illustration

6.3.5. **Mutation**

The mutation operator changes certain genes in a population to maintain the diversity of the population. Instead of using a uniform mutation operator, I use a non-uniform resource operator so that initial generations have a higher mutation probability (and thus promoting exploration of the solution space at the beginning of the process). The mutation probability P_m^{gen} of a generation, *gen* is given by:

$$P_m^{gen} = Mprob.\frac{maxgen-gen}{maxgen-1}, gen = 1, \dots\dots, maxgen$$ (Lin, Zhu & Gao, 2020).

Where Mprob= mutation probability used for the initial population, gen=current generation number, maxgen= is the maximum number of generations

The mutation operator for a chromosome moves the genes a time step backwards. This increases the fitness of the individual because, all other things remaining equal, an earlier project finish leads to a better individual.

6.4. Solving LAS problem using RL

Reinforcement Learning (RL) is an area of machine learning inspired by how learning takes place in nature, wherein the actions taken by an entity (called an agent) are dependent on the outcomes obtained from previous actions. Positive rewards would reinforce the behaviour or give extra weight to that particular action and the actions leading to it and vice versa (see Figure 6.4). RL is built on top of Markov Decision Processes (Sutton & Barto, 2018) and differs

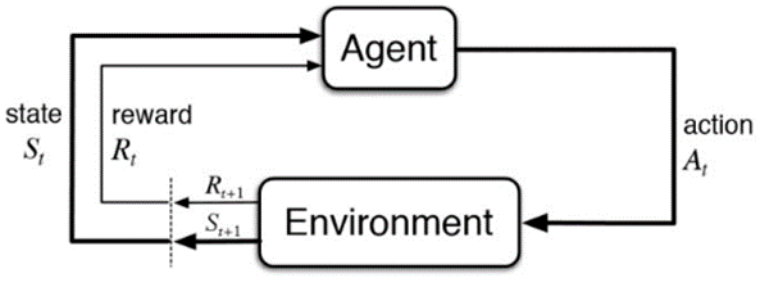


Figure 6.4: Reinforcement learning model

from supervised learning since it does not require labelled input-output pairs for learning to happen (Kaelbling, Littman & Moore, 1996). This is ideal for LAS generation as the states of a construction project are unique and a dataset of actions and consequences are non-existent, thus, making it challenging to apply supervised machine learning. Formally a reinforcement model consists of the following:

- A discrete set of states, *S*;
- A discrete set of agent actions, *A;* and
- A set of scalar rewards.

For every state, an agent takes action based on the reward obtained when the same state was visited previously. The role of an agent is to find a policy π, which maps states to actions such that reward is maximized.

In LAS generation process, the project manager chooses a project strategy similar to how an agent in RL algorithm takes action. At every time frame in a project, the state of the project can be represented as a combination of a set of activities that are complete, a set of activities in progress, a set of activities that can be done, and the number of resources available to execute the task. The project state resembles the state in an RL problem. Based on this information about the project state, the project manager iterates over multiple project execution strategy. Each strategy is evaluated based on the advantages and disadvantages and scored relatively, and the best strategy is taken. These scores are similar to the reward in an RL problem. Thus, when LAS is modeled as an RL problem, the role of the agent would do the tasks of the project manager.

6.4.1. **Q-Learning**

Q-learning is a subset of model-free reinforcement learning (i.e. it does not require a model of the environment to operate). It is not possible to use model-based algorithms for the

LAS problem, as the states and the transitions probabilities are not known in advance but are generated during algorithm execution. Q-learning stores only states and actions to determine the policy and discards the remaining information. Hence, Q-Learning is computationally cheaper when compared to other reinforcement learning algorithms as it required less memory.

As in Reinforcement Learning in general, the goal of Q-learning is to learn a policy, which tells an agent an optimal action to be taken for a state. The focus of the reinforcement learning algorithm is to find a balance between exploration and exploitation. (Kaelbling, Littman & Moore, 1996). There is a Q-value for every state-action pair denoted as $Q(s, a)$, and it serves to determine what action would be taken for that state.

$$Q(s, a) = r + \gamma \max_{a'} Q(s', a')$$

The Q-value is initialized randomly at first for all state-action pairs. Then it is updated after every timestep using the Bellman equation, which tells us that the maximum future reward is the sum of reward the agent received for entering the current state and the maximum future reward for the next state s′. The main concept behind Q-learning is a value iteration update. On a single state *s* and action *a*, we can express Q-value *Q(s, a)* of state action pair recursively, in terms of the Q-value of the next state s′: The value of the current Q-value is dependent on the reward of the future states that the action on current state would have. The discount factor γ determines the relevance of future rewards. A factor of 0 for the discount factor would mean that the agent is short-sighted and does not rely on the future rewards; whereas a factor of 1 means that the agent would strive for a long-term high reward. The Q-learning equation is given by:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_{a} Q_t(s_{t+1}, a) - Q_t(s_t, a_t))$$

Where α is the learning rate that controls how big is the difference between previous and new Q-value, the learning rate determines the extent to which the newly acquired information is used. A learning rate of 0 means that the agent relies on exploiting the existing information, whereas a value of 1 means that the agent relies on the latest information and completely ignores any prior knowledge. Q-values are initialized randomly. Then, they are updated as the

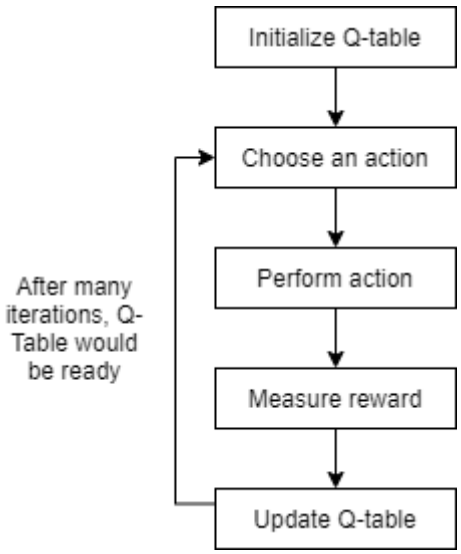


Figure 6.5: Q-Learning steps

agent takes action, and the reward is obtained from the environment. In most cases, Q-values start to converge after many iterations. Converged set of Q-values can be used for determining actions for each state. The Q learning steps are represented in Figure 6.5.

6.4.2. **Modelling LAS problem**

This section describes how the LAS problem is modeled using Q-Learning. To model the LAS problem, the time-oriented parallel schedule generation scheme is adopted (Kolisch, 1996b) over activity-oriented serial schedule generation scheme. The parallel schedule generation scheme is a time focused scheduling scheme consisting of *N* time periods. For each time period, a set of activities is scheduled. This scheme is applicable for LAS problem as it would cater to time-dependent constraints such as disjunctive constraint. In addition, it is

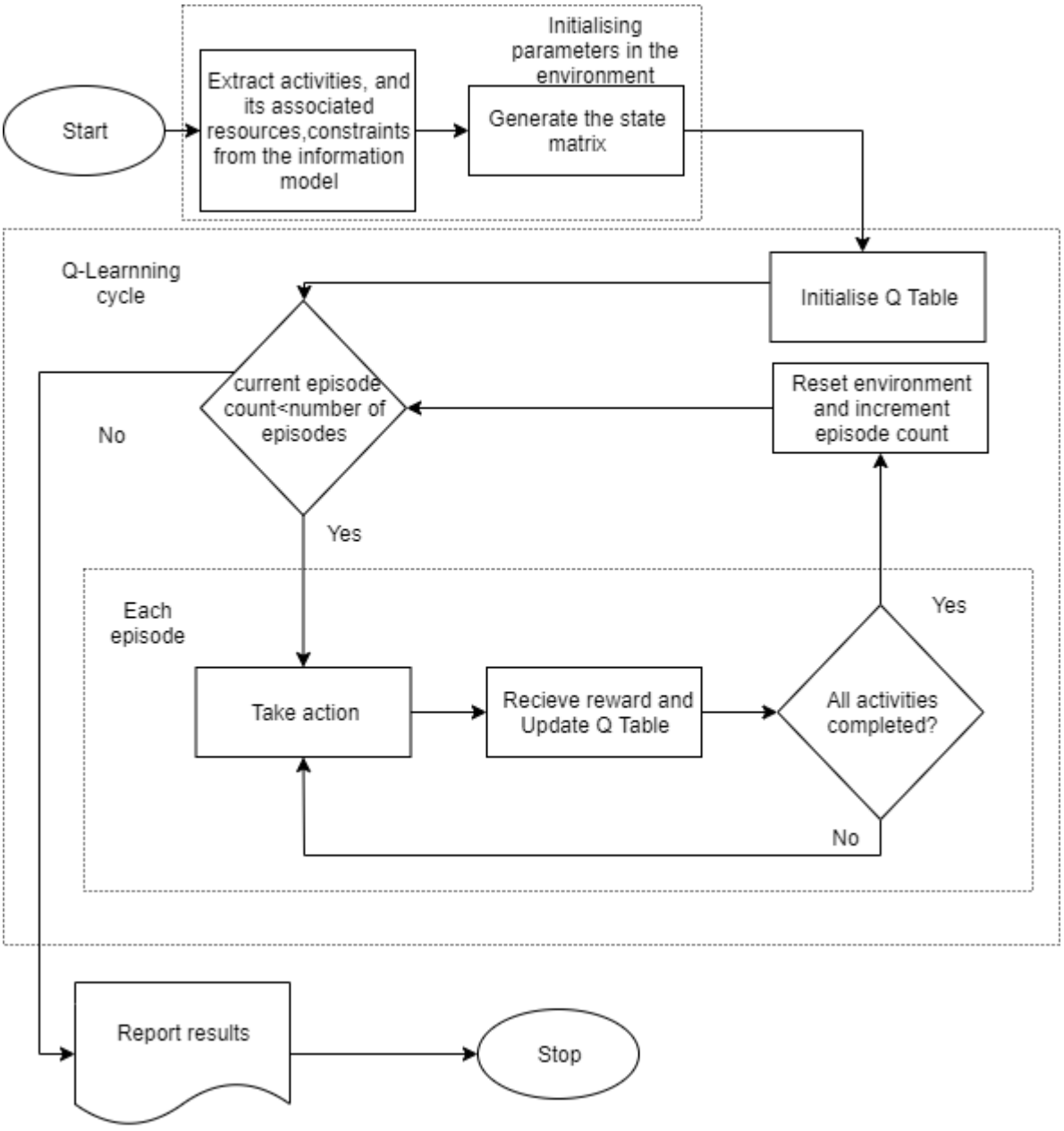


Figure 6.6 : Q-learning workflow for LAS generation

suitable to model Q-learning using a parallel generation scheme as each time step can be considered as a state. Figure 6.6 shows the workflow for modelling a LAS generation using Q-learning. State representation, Action representation, and Reward formulation are detailed in subsections 6.4.2.1, 6.4.2.2 and 6.4.2.3, respectively.

The first step in the Q-Learning workflow is the initialization of the environment, which is a simulation representing the construction site. Each state of the environment consists of activities completed, activities in progress, activities that need to be scheduled and the resources available at each time period. For the first initialization, the processes, their associated activities, and resources available, are imported from the information model and represented as a state matrix (to be detailed in 6.4.2.1). Processes are the major work packages, and the activities are the operations that need to be performed as a sequence for that process to be completed. The number of episodes Q-learning would run is also set during the initialization.

The second step is to generate the Q-table. Q-Table is a lookup table where the maximum expected future rewards for action at each state is stored. Every row in the Q-table represents a state. It is not possible to know the number of states that would appear during the learning. Hence, the Q table is first generated with a single row representing the initials state. A new row is added when there is a new state after an action is taken. The columns in the Q-table represent the combination of activities that could be scheduled. Hence, the first Q-table would represent the initial state and possible action combinations. There would be 2^n action combinations hence 2^n columns in the Q-table, where n is the number of processes. Every cell is initialized to zero.

Once a Q-table is generated, the environment gives the agent possible action combinations for the initial state. The agent takes action based on the Q-values of the state. In this context, an action is the activity combination that would be executed in this time period.

Based on the action, the environment returns the reward (detailed in 6.4.2.3), the next state and increments the time period. Once the reward is obtained, the Q-table is updated based on the returned reward and the Q-learning equation. The agent then takes action (schedules a set of possible activities) for the new state. This cycle is continued until all the activities are scheduled. Once all the activities are scheduled, an episode is complete. At this stage, the Q-table would be populated with a certain number of rows, with respect to the number of states visited in this episode. Once an episode is completed, every activity would be assigned a start time and the required resource constituting to the LAS. However, this LAS is not the optimal LAS.

After an episode is completed, the environment is reset, and the next episode is started reusing the Q-table from the previous episode. By doing so, the agent learns from actions in past episodes and keep updating it with new actions. New states might be added to the Q-table as the agent might have visited a new state in the second episode, as the Q-learning not only exploits the existing knowledge but also explores the state space for new knowledge. This means that the agent gains a little more knowledge with each episode. This cycle continues until a set number of episodes are completed. The LAS generated as the result of the last episode is reported back to the user.

6.4.2.1. *State*

Reinforcement learning algorithms use an incomplete Markov Decision Process framework to derive optimal strategies. Hence, every state in the reinforcement learning problem should be a Markov state (Sutton & Barto, 2018). This means that all the future states should depend on the current state and not on the events which occurred before. For the LAS problem, each state would correspond to the construction site characteristics at a time period. Therefore, each state (S) represents the outcome of previous actions, and comprises of the

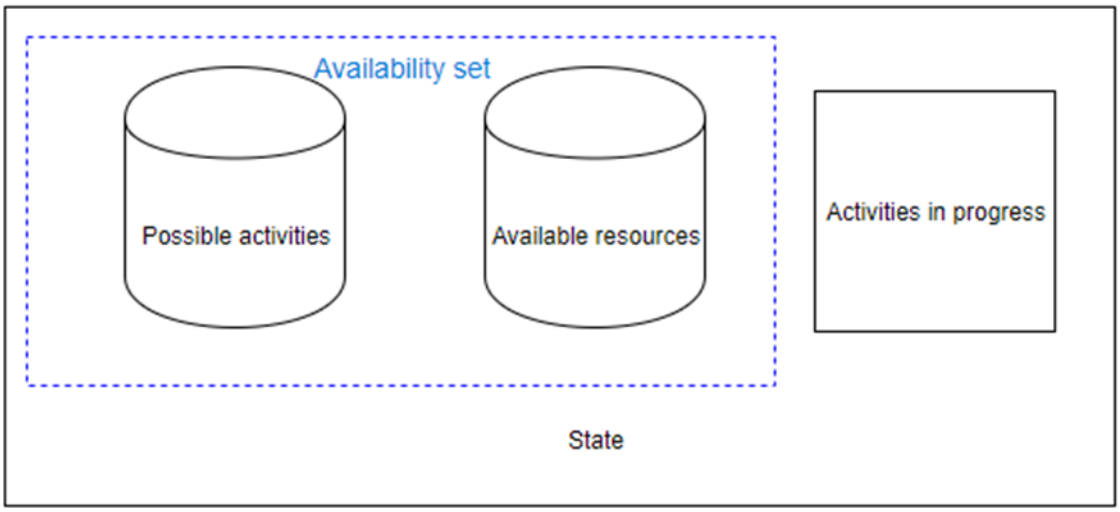


Figure 6.7: State in a LAS- Q-Learning problem

following piece of information (refer to Figure 6.7):

1. Possible activities: These includes activities that can be started. All the preceding activities in the process needs to be completed for an activity to be started.
2. Available resources: The current availability of the resources should be present in the state information as it would be required to assess which actions can be performed later.

3. Activities in progress: The state would give information on the activities in progress. Future states need this information to assess the activities that can be scheduled next.

The state (S) of the LAS problem can be represented by the following.

$$S = [X, R]$$

Where X is a *m* x *n* matrix representing the activity state with *m* processes and *n* activities, and R is *1* x *k* matrix representing resource availability of *k* resource types.

$$X = \begin{bmatrix} X_{1,1} & \cdots & X_{1,n} \\ \vdots & \ddots & \vdots \\ X_{m,1} & \cdots & X_{m,n} \end{bmatrix}$$

Rows in *X* denote processes. Columns in *X* denote the activities under each process. Value of each cell in *X* denotes the state of that activity.

$X_{m,n}$ = *-1(if the task has not started)*

$X_{m,n}j$ = *-2(if the task need not be visited)*

$X_{m,n}$= *duration* $_{m,n}$ *(if the task has ended)*

0<$X_{m,n}j$*< duration I,j (if the task is in progress)*

Matrix *X* gives information on activities in progress as well as activities that can be started. The activities that can be started are the ones whose preceding activities in the process are completed. As in the chromosome of a GA, the number of activities might not be the same for all the processes. Hence dummy activities are added to the last few columns to make the matrix rectangular. Dummy activities are marked a random negative value to inform that the algorithm that it may not be visited. In the current formulation, such dummy activities were marked as -2.

Array R represents the resource availability in each state and is represented as a *1* x *k* matrix, where each cell represents the number of items available of a resource type *j € k.*

$$R = [x \quad 0 \quad \dots \quad 0]$$

$$Ri=0, if\ not\ available$$

$$Ri=x,\ if\ x\ number\ of\ that\ resource\ type\ are\ available$$

6.4.2.2. *Action*

At each state, the agent takes action from the available ones, which depicts the combination of activities that can be executed at that time period. There might be multiple combinations that can be performed, as shown in Figure 6.8.

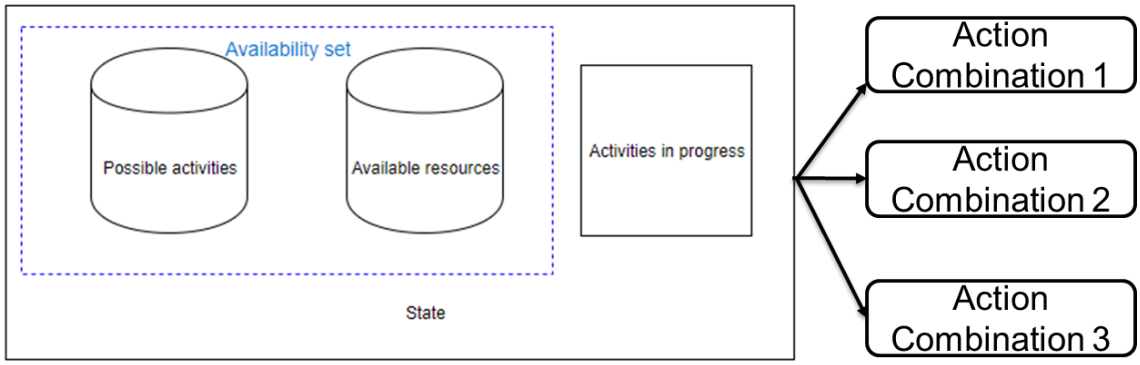


Figure 6.8: Action combinations

To model the action, there is no need to represent all the activities. As the activities happen only in a sequence, giving a binary value to the process would determine whether the next possible activity in that process is executed. Consider matrix *Ai* that represents an action combination

$$Ai = [0 \quad 1 \quad \dots \quad 0]$$

Each column in *Ai* represents the process at that index. A value of 1 means that the next activity corresponding to the process needs to be executed and zero means that the activity need not be executed.

$$Aj = [1 \quad 1 \quad \dots \quad 1]$$

Similarly, *Aj* means that activities under all the processes, which are to be executed next should be started.

$$Ak = [0 \quad 0 \quad \dots \quad 0]$$

Also, *Ak* means that activities no activities should be started. So, the action space is a combination of all these action combinations. Therefore, since each column is binary, every 2^n possible combinations that can be performed in every state, where n is the number of processes, hence, the action space can be presented as a matrix *A*.

$$A = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & o \end{bmatrix}$$

Rows denote a list of action combinations. There would be 2^n rows. Columns denote the process whose next possible activity must be started. The agent takes action by choosing one of the action combinations that can be executed (as reported by the environment).

Despite several action selection methods have been proposed in literature, I chose to follow an adaptive ε-greedy method (Richard S. Sutton, 2008) as this is a straight forward algorithm that aims at attaining global optimum by making locally optimal choices. In the ε-greedy method, the agent behaves greedily by selecting an action combination with the maximum Q-value for most of the time. However, to promote exploration, with a small probability ε, the agent selects a random action. To choose an action, the agent looks up the Q-table. The number of columns in Q-table equals number of rows in the *A*. Therefore, each column in Q-table is mapped to a row in the *A*. The agent chooses a combination by referencing the Q-table entry with the highest value for that state or by random. If the row corresponding to the Q-table entry is not found in the list of rows returned from the environment, a negative

reward is instantly given so that in the next iteration of the state, that particular value is not chosen.

6.4.2.3. *Reward*

The reward in Reinforcement Learning problems determines how good or bad the action was taken by the agent is, thus influencing its future behaviour. The LAS problem focuses on ensuring the works are scheduled in such a way that there is no constraint, the resources employed should be utilized to the maximum (so that the cost is reduced), and the processes and activities to be scheduled should finish in the least amount of time (thereby reducing the total cost, as there are indirect costs associated with the additional time for process completion), and the cost for hiring resources. As the reward determines how the agent chooses the next action, in the context of the LAS problem and similarly to the fitness value in the GA, the reward is a function of

1) Constraint satisfaction
2) Completion of activities and processes
3) Resource cost

Constraint satisfaction is determined by the linked-data based constraint checking algorithm determined in the previous chapter. Every time the agent takes action, the constraint checking module in the environment checks for constraint violations (precedence, discrete, disjunctive, logical constraints), returning 0 if no constraint is violated, and the number of violations *Nc* otherwise. Although there are provisions to determine the severity of constraint violations, I have opted for not implementing them and consider all violations equally severe. Again, to ensure that the agent does not choose solutions with constraint violation, a penalty function is applied. The penalty is proportional to the degree of violation and is much higher than the reward function (Smith & Raphael, 2013) so that the agent refrains from taking actions

resulting in a constraint violation. Therefore, the reward component for constraint violation *Rc* is given by:

$$Rc = -1000000 * Nc$$

More processes and activities completing in a time period means that the productivity in the time period would be higher. To ensure that productivity is higher, a positive reward is given to the activities that are progressing and completed. This reward is computed after every action. This is obtained by adding all element in matrix (*X'*), where *X'* is obtained by subtracting the Activity state matrix (refer to 6.4.2.1) after taking action (X_{t+1}) from the activity state matrix before taking action (X_t). Here we are assuming that the activity durations are the same. Therefore, the cost of activity would remain the same and the indirect cost would be the dependent factor in the cost computation. The reward component for activity completion *Ra* is given by:

$$Ra = 1000 * sum\ of\ all\ elements\ (X_{t+1} - X_t)$$

The last reward component to be modeled for LAS is the cost for resources. The logic for the resource cost mentioned in the fitness evaluation of GA (refer to 6.3.2) is used here. Every resource is assumed to be hired for unit time costing. In addition, there are costs for resources mobilization and demobilization of resources for each activity. The cost of deploying resources *i* is given as:

$$Cres_i = Dtotal * Cost_{per\ unit\ time} + (MobilisationCost + DemobilisationCost) * Nassigned$$

Dtotal= Total duration assigned, Nassigned= Number of activities the resources are assigned

The reward for *Rr* is calculated by reducing the cost of total resource usage.

$$Rr = -\frac{1}{100}\sum_{i=1}^{n} Cres_i$$

Therefore, the total reward for the agent is given by:

$$R = R_c + R_a + R_r$$

As in the case of GA, the relative weights on the equation can be changed to accommodate different scenarios in which criteria have different importance.

6.5. Computational Experiments

This section presents the results obtained from a series of experiments to measure the efficiency of both the proposed GA and reinforcement learning algorithm to solve the LAS problem. The algorithms were coded in python and executed on a workstation PC with the configuration shown in Table 6.1

Table 6.1: Details of the computer in which the experiments were run

Category	Details
System model	HP Z440 Workstation
Processor	Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz (16 CPUs)
Available OS memory	32GB RAM

Table 6.2 lists the values used in the different experiments for all parameters for the genetic and reinforcement learning algorithms. These parameters were obtained by testing the developed algorithms with various cases and scenarios and choosing an optimal value.

Table 6.2: Parameters for the computational experiment

Algorithm	Parameter	Value
GA	Crossover probability	0.8
	Mutation probability	0.2
	Population size	100
	Number of generations	100
Reinforcement Learning	Learning rate	0.4
	Discount factor	0.6
	Number of episodes	30

6.5.1. Comparison with traditional methods

To compare the performance of the developed algorithm with the traditional methods, a test case was prepared with three processes containing four activities each, as shown in Table 6.3. Precedence constraints, resource constraints and disjunctive constraints were present in the test case. Four strategies were used to generate the LAS for the data specified in Table 6.3. These strategies are:

1. Critical path method
2. Critical path method combined with the manual method
3. Automated method using developed GA
4. Automated method using developed RL

Table 6.3: Input parameters

Process	Activity	Duration	Precedence	Resource constraint	Disjunctive
A	1	3		R1	
A	2	4	1	R2	8
A	3	2	2	R3	
A	4	4	3,11	R3	
B	5	2		R3	
B	6	1	1,5	R1	
B	7	5	6	R3	
B	8	3	7	R2	2,12
C	9	1		R1	
C	10	4	9	R2	
C	11	3	6,10	R1	
C	12	2	11	R3	8

Gantt charts developed using each of these methods are presented in Figure 6.9. Table 6.4 compares the results of the LAS generation process using different strategies.

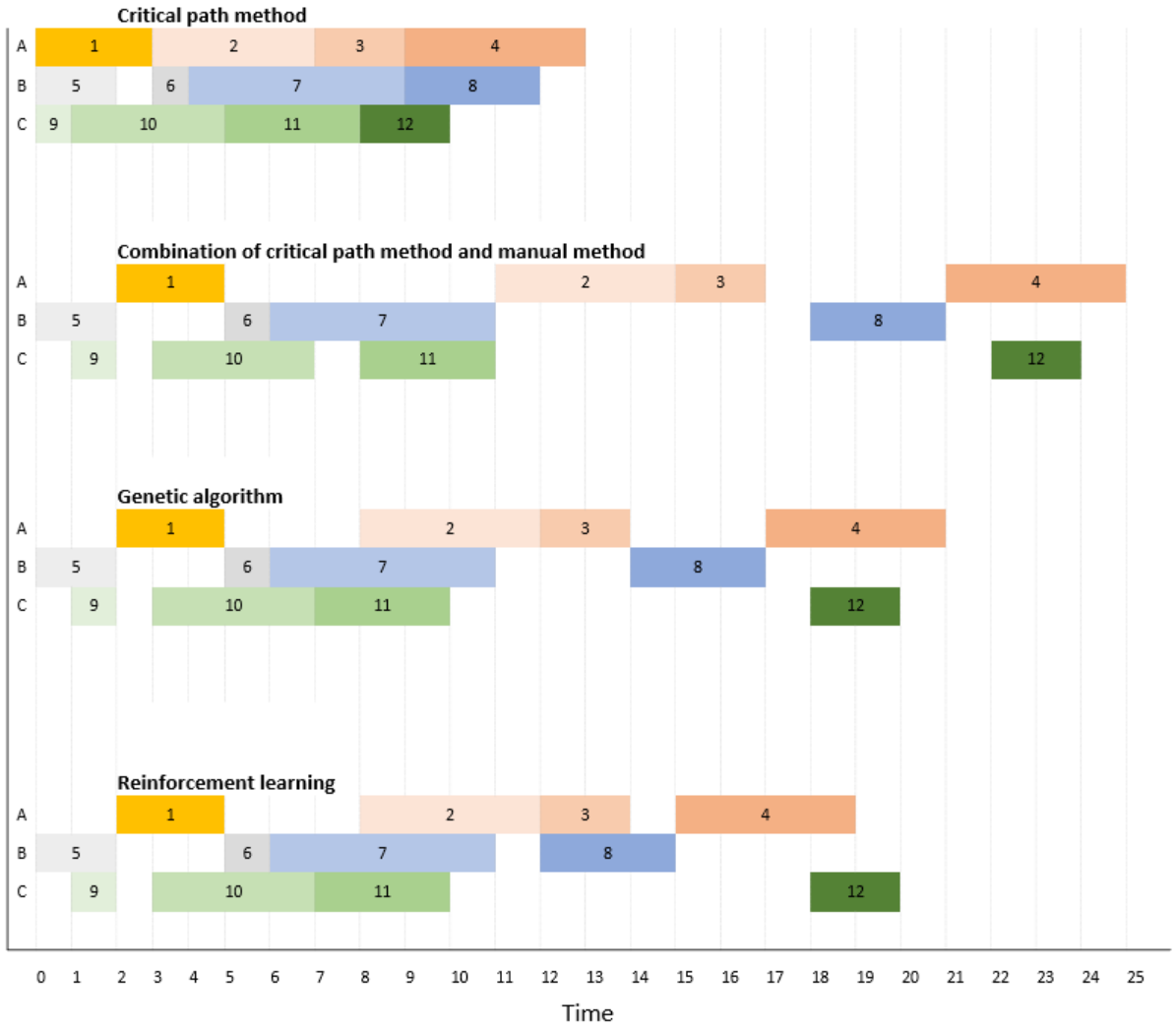


Figure 6.9: Gantt chart developed using different strategies

Table 6.4: Comparison of the results

	Project duration	Time taken	Number of constraint violations
Critical path method	13	2 seconds	6
Critical path method + Manual method	25	15 minutes	0
GA	22	15 seconds	0
RL	20	2 seconds	0

The best project duration was obtained using the critical path method. However, this LAS generated using critical path method had a number of constraint violations requiring manual intervention. The manual removal of constraints from the LAS generated during the critical path method took a long time even for the small problem. LAS generation using GA and RL resulted in automatically generating a LAS without any constraint violations. GA took 15 seconds to compute it and RL took just 2 seconds.

6.5.2. **Influence of the number of resource types**

Unlike a classic RCPSP, the LAS problem has a huge number of resource types (equipment's, crews, workspaces, objects, etc.), meaning that they have a considerably higher number of constraints that need to be solved. This section explores the performance of the proposed algorithms with respect to an increasing amount of resources as shown in Table 6.5 . The inputs specified in Table 6.5 were fed into the developed GA and RL algorithm and the results are presented in Figure 6.10.

Table 6.5: Input data

Number of Resource type	Number of Processes	Number of Activities	Number of Constraints			
			Precedence	Discrete	Disjunctive	Logical
5	10	10	113	10	2	2
10	10	10	105	20	2	5
15	10	10	102	30	5	7
20	10	10	103	40	12	10
50	10	10	115	100	19	25
100	10	10	101	200	90	50

Table 6.6 compares the project duration and processing time of the GA and the reinforcement learning algorithm for different resource types. It is clear from the results that the reinforcement learning algorithm fares better both in terms of the project duration (i.e. the best solution found) and the processing time (i.e. time taken to find it). Project duration of the GA for the number of resource type= 5 is similar to the reinforcement learning algorithm.

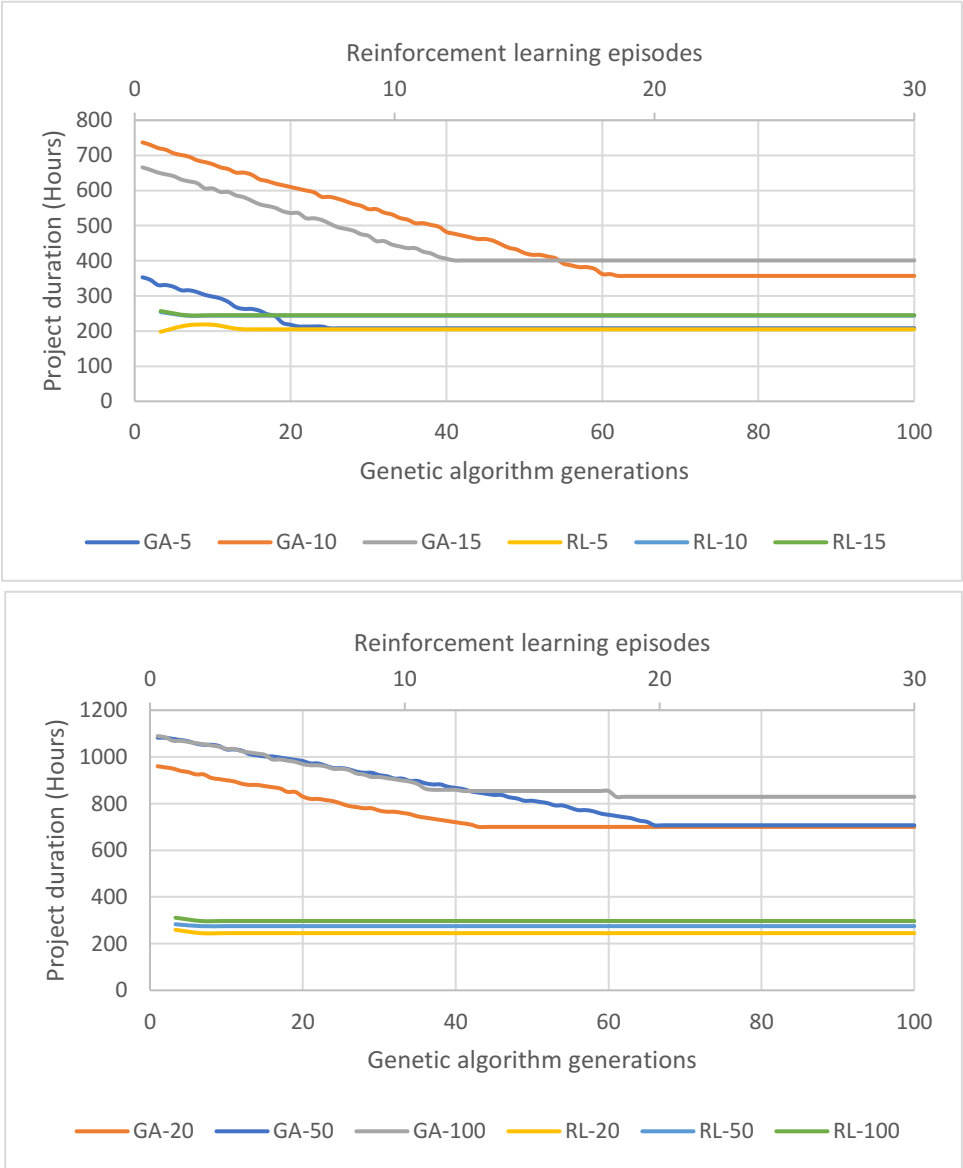


Figure 6.10: Influence of increasing number of resource type

However, as the complexity increases, the RL approach tends to find better solutions (in terms of shorter project durations) than the GA.

Table 6.6: Comparison of Genetic and Reinforcement learning algorithms

Number of resource types	Project duration (hours)		Processing time (s)	
	GA	Reinforcement Learning	GA	Reinforcement Learning
5	208	205	515	27
10	357	244	743	23
15	401	179	961	19
20	700	245	670	26
50	707	275	1773	41
100	829	297	1384	65

6.5.3. **Influence on the number of processes and activities**

LAS problem has a huge number of interdependent activities. To emulate the dependence on the number of activities, the GA and the reinforcement learning were tested for performance against the number of increasing processes and an increasing number of activities. Eight scenarios were created manually to test the dependency of the algorithms on an increasing number of processes and activities. Scenario S1 has the least number of process and activity, whereas scenario S8 has the highest number of activity and process. The scenarios and the inputs are given in Table 6.7. The inputs specified in Table 6.7 was fed into the GA and RL, and the results of convergence are shown in Figure 6.11.

Table 6.7: Input data to the algorithm

Scenario	Scale	Number of Processes	Number of Activities	Number of Resource type	Number of Constraints			
					Precedence	Discrete	Disjunctive	Logical
S1	25	5	5	10	26	20	2	5
S2	50	5	10	10	57	20	5	5
S3	75	5	15	10	81	20	8	5
S4	100	5	20	10	107	20	3	5
S5	50	10	5	10	58	20	1	5
S6	100	10	10	10	102	20	6	5
S7	150	10	15	10	152	20	2	5
S8	200	10	20	10	206	20	3	5

As the scale and complexity of the problem increases, the time for convergence is longer, as shown in Figure 6.11. For GA, the number of generations required for convergence increases as the scale and complexity of the problem increases. When the scale and complexity are lower (S1, S2, S5), the solutions converged in 20 generations, however for the scenarios with medium-scale and complexity (S4, S7, S6) the solutions converged around 60 generations. For scenario S8, GA did not converge to a solution even after 100 generations. For RL, the scenarios S1, S2, S3, solutions converged around the 3rd episode whereas S7 converged around the 6th episode. Scenario with the highest number of activity and process combination S8 converged around the 10th episode. In the case of S6, the solution converged to a solution that did not have the lowest duration. This is because the reward is not just comprised of the project duration but also resource availability. When comparing the processing times for generating solutions for different scenarios of increasing scale and complexity, the reinforcement learning algorithm fared significantly better than the GA in terms of CPU time, as shown in

Table 6.8. However, when comparing the project duration for activities with lower scale and complexity (S1, S2, S5), GA fares better than the reinforcement algorithm. However, as the scale and complexity increase, the GA loses its efficiency and ends up with worse solutions, whereas reinforcement learning algorithm shows a better capacity at finding good solutions.

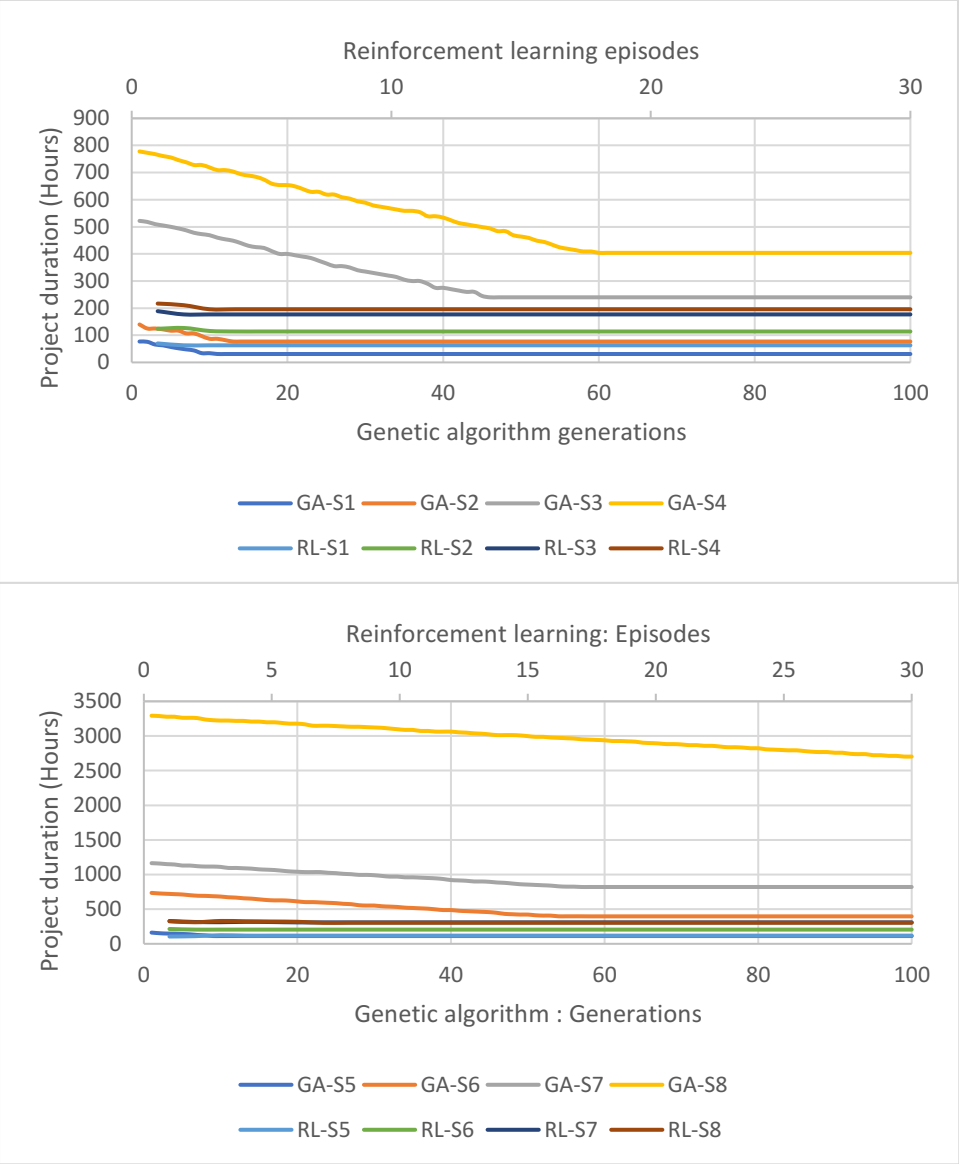


Figure 6.11: Influence on number of processes and activities

Table 6.8 : Comparison of Genetic and Reinforcement learning algorithms

Scenario	Project duration (hours)		Processing time (s)	
	GA	Reinforcement Learning	GA	Reinforcement Learning
S1	31	63	54	3
S2	77	114	138	6
S3	240	177	457	11
S4	404	196	753	15
S5	112	118	138	14
S6	396	205	848	22
S7	820	309	2060	52
S8	2703	304	6323	105

6.5.4. **Discussion**

Traditional methods for the generation of LAS was compared against the automated generation of LAS using GA and RL developed in this chapter. The results from the comparison reveal that the developed algorithms can generate a constraint-free LAS within a short duration in comparison with manual LAS generation. This demonstrates the capability of the developed algorithms to support look-ahead planning by generating LAS for different scenarios in near real-time, thus helping the engineers to assess the influence of different strategies and constraints on the project duration. When integrated with a costing software, these methods can be used to generate different cost profile to perform a cost-benefit analysis of different scenarios.

Also, experiments in this section have compared the results of the computational experiment conducted to solve the LAS problem using the LDCC method to evaluate the goodness of candidate solutions. From the results of these experiments, it can be concluded that the GA performs better than the reinforcement learning algorithm when the number of constraints and the scale and complexity of the problem is lower. However, when the number of constraints and the scale and complexity of the problem is higher, the reinforcement learning algorithm fares better than GA. As the number of resource types increases, the number of constraints increases, and the GA is slower as well as prone to produce a schedule with longer project duration.

The lower performance of the GA can be attributed to the convergence to a local optimum primarily due to the lack of diversity of the initial population. If the search space is relatively small (when the constraints are fewer) the initial population can cover most of it, and the GA converges to a good solution (often better than the one from the reinforcement learning algorithm). To define a good initial population, all the constraints need to be mathematically defined. However, this is not possible as the constraint definition takes place during the look-

ahead planning phase by trained engineers who may not be an expert in mathematical modelling. Instead, the constraint creation is supported by a prototype described in Chapter 5. The framework is designed to check for the violation of constraints by traversing over heterogeneous datasets. This aspect of constraint checking is embedded into the fitness function to guide to solutions with constraint satisfaction. However, this phase takes place after the initial population generation and does not help in the generation of a good quality initial population. This limits the efficiency of the GA.

Although GA is dependent on the initial population, reinforcement learning does not have this issue. Unlike GA, reinforcement learning is not a population-based method. In reinforcement learning, learning takes place as in nature, wherein the actions taken by an entity is dependent on the results of the rewards and punishments obtained from the previous actions. These types of algorithms are suited to work with the constraint checking mechanism mentioned in the previous chapter as the reward is obtained from the response of the environment, and actions are based on the environment. The constraint checking environment will act as the environment in the reinforcement learning setup, and the reward includes information about constraint violations. Therefore, each action will make use of the power of the constraint checking environment in RL, whereas GA uses the power of constraint checking environment only after obtaining a full solution. This results in faster convergence than the GA.

In contrast with GA using serial schedule generation scheme, using a parallel schedule generation scheme enables RL to construct the LAS incrementally. The constraints in LAS are predominantly time-oriented such as disjunctive constraints. Also, the activities in the LAS are highly interdependent. Therefore, an algorithm which uses parallel schedule generation scheme, which in this case is reinforcement learning would have better efficiency than the one which uses serial schedule generation scheme, as the parallel generation scheme allows

incremental construction of LAS leveraging the power of LDCC method using each increment. This reduces the number of inefficient full solutions being developed.

6.6. Summary

This chapter focused on generating LAS automatically with two different computational techniques. The LAS problem was modeled using both an activity-oriented GA and a time-oriented RL. Both these methods used LDCC method developed in Chapter 5 to evaluate complex construction constraints. The developed methods could generate constraint-free LAS automatically faster than the manual methods. Further, the influence of the number of resource types and the scale of the problem on these algorithms were evaluated through a series of experiments and the results were presented in this section. The experiments showed that the GA fared better than reinforcement learning algorithm for problems with lower scale and complexity. However, as the scale and the complexity increased, the reinforcement learning algorithm has higher performance over the GA.

7. IMPLEMENTATION STUDY

7.1. Introduction

This chapter presents the implementation of the data-driven decision-support tools developed in the previous chapters on a real construction project. The implementation study is conducted to investigate the applicability and limits of both the LDCC method and LAS generation methods when they are applied in a real project environment. This study focused on determining the accuracy of constraint violation identification for the LDCC method and the ability to generate a conflict-free LAS for the automated LAS generation methods.

A worked example of a real case (the *water project*, introduced in Chapter 4) followed by the feedback from practitioners, was used for demonstrating the capability of the developed methods. The LDCC prototype was used to import planning information from the 4D model and collate it with 1) process information generated during look-ahead planning meetings and 2) the process information recorded in method statements. The construction planning information imported from the 4D model was enriched with the information from the meetings and method statements, using the developed prototype web application. Construction process information, including constraints, was among the collected information.

After the import, a LAS generated by the practitioners was checked for constraint violation using the LDCC method and execution times for the tasks were removed from the codified information manually. The machine learning methods (GA and RL) discussed in Chapter 6 were used to automatically generate a LAS each, using the codified information. These two LAS were analyzed manually to check for constraint violations and then compared against the LAS generated by the practitioners. Deviations between both LAS were studied and then presented back to the practitioners to collect their feedback.

The rest of the chapter describes the testing in detail and is divided into five sections: Section 6.2 describes the case used for the study and informs the reader about the processes and sub-activities scheduled during the period, and the resources associated to these activities. Section 6.3 details how the data is imported into the LDCC to perform constraint checking. Section 6.4 presents the results of constraint checking using LDCC and describes the feedback from the practitioners on the results. Section 6.5 presents the results of the LAS generation, and Section 6.6 presents the summary.

7.2. Case study description

The *Water project* introduced in Chapter 4 was used as a case study for the evaluation. In Chapter 4, I studied several sites in the central section of the *Water project* to identify codification challenges. However, for the implementation study, a site from the east section of the *Water project* was selected because this section uses state-of-the-art 4D BIM based collaborative process to validate the construction programme. . The digital strategy followed in this section, one of the finalists in the model-based planning competition in Bentley's Year in Infrastructure 2019, has helped to manage logistics of the highly congested construction site situated within a dense residential area resulting in better productivity. Using this case for the implementation of the LDCC and the LAS generation method would enable a comparison with the state-of-the-art in practice, as this case pioneers the use of 4D BIM for look-ahead planning. For the implementation study, the focus was on a 30-day period for the treatment plant site in the *Water project*, which housed construction activities for a slurry treatment plant. During the 30-day period, two main construction processes were planned to be carried in the treatment plant site.

1) Muck bin construction: Muck bins are the components in the slurry treatment plant which collects the sludge from the slurry. The work package related to muck bin

construction includes work required for the construction of three muck bins (Bin A, Bin B and Bin C).

2) Slurry Treatment Plant (STP) foundation: The Slurry Treatment Plant (STP) foundation process is associated with building the reinforced concrete foundations for the plants.

Figure 7.1 shows the areas for construction for these activities in the construction site. Activities associated with Muck bin construction and STP foundation processes that are planned for the period of focus are given in Table 7.1.

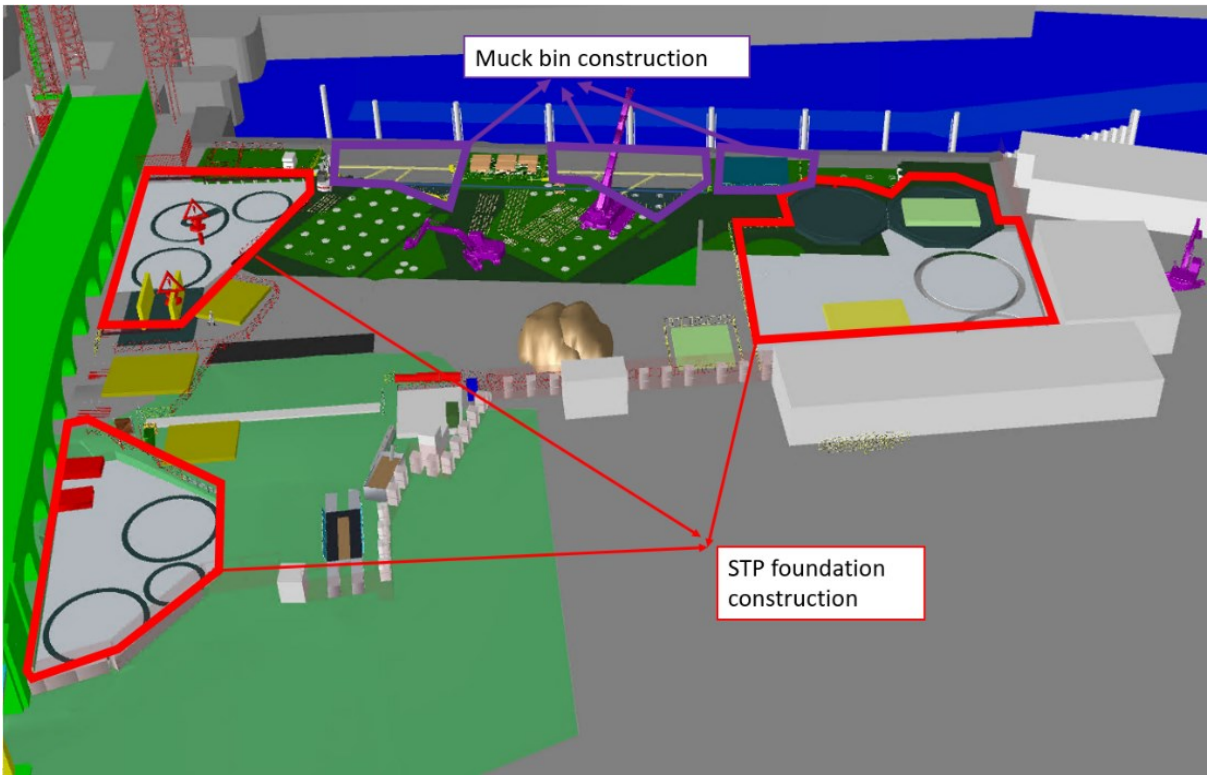


Figure 7.1: STP foundation and Muck bin construction areas

Resources (Plants and equipment) that are available for the construction processes and their assignment are shown in Table 7.2. Resources are shared between the processes *Muck bin construction Bin B and Bin C*. However, the processes *Muckbin construction A* and *STP Foundation* have resources available to them exclusively. The next section describes how this information is imported to the LDCC prototype.

Table 7.1 Processes and sub-activities

Process		Sub-activities
Muck bin construction	Bin A	Steel Delivery Bin A, Excavate to Prop Level & Create sump, Crop Piles and Cut Steel, Install River Wall Connection, Install Prop, Excavate to Base Level, Crop Piles - Ensure reo left for slab tie ins, Pour Blinding, Final Trim Piles - Integrity Test - Prep Base
	Bin B	Steel Base and Wall Starters, Inspection, Bin B Pour 1 - Base, Formwork for Kicker, Bin B Pour 2 - Kicker, Curing time (strike shutters), Remove Propping, Steel Walls
	Bin C	Final Trim Piles - Integrity Test - Prep Base, Steel Base and Wall Starters, Inspection, Bin C Pour 1 - Base, Formwork for Kicker, Curing time (strike shutters), Bin C Pour 2 - Kicker, Remove Propping
STP foundation		Placing Bottom Matt Steel (Outside Tanks), Concrete downstand Beam, Install Recess B6-3 + Levelled, Install Recess B6-2 + Levelled, Install Recess B6-1 + Levelled, Install Perimeter Shutters, Concrete Curing (To Strike Shutters), Placing Steel Top Matt B6-3, Placing Steel Top Matt B6-2, Placing Steel Top Matt B6-1, Strike Shutters (All bases and Perimeter), Steel delivery, Install Screed Rails + Levels, Prep area and blind southwest near siltbuster, Pour Watertight Concrete in Recess (Following Tank Install), Placing Top Matt Steel (Outside Tanks), Install Screed Rails + Levels, Concrete Curing, Placing Steel Bottom Matt B3-2, Placing Steel Bottom Matt B4, Placing Bottom Matt Steel (Outside Tanks), Excavate and surface prep for civil works (750mm Trench for recess), Strike Shutters (All bases and Perimeter), Prefabrication of Recess B5, Blinding, Placing Steel Top Matt B3-2, Placing Steel Top Matt B4, Placing Top Matt Steel (Outside Tanks), Install Perimeter Shutters, Excavate WTP Temporary Slab LA, Steel delivery, Pour, desanding building slab, place steel for desanding building slab, Concrete Curing, steel delivery, Blinding WTP Temporary Slab LA, Strike Shutters (All bases and Perimeter), placing steel for b5 Beam

Table 7.2: Equipment assignment

Equipment	Quantity	STP foundation	Muck Bin construction
22t excavator + breaker	1	X	
6t dumper	1	X	
40T crawler crane	1	X	
ABI TM 14/17 Leader Rig	1	X	
Forklift	2		X
8t excavator	2	X	
Roller	1	X	
25t excavator	2		X
40 T telescopic crane	2		X
Pile cropper	2		X
5t Excavator	2		X

7.3. Data import and enrichment

In the state-of-the-art practice in the construction site, schedule data for the 30 days was collated in the form of a 4D BIM file. The planners created the schedule in a planning software which was imported to the 4D BIM package by the digital engineer at the construction site. The BIM was linked to the schedule within the 4D BIM package to create a 4D model. This 4D model is used in the look-ahead meetings for decision-support. During the look-ahead planning meetings, planned activities for the look-ahead window are discussed with subcontractors and project engineers to check for feasibility. Construction sequence and requirements for sub-activities of a process are developed during this meeting. Based on the discussion, 3D elements, including resources and formwork, are added onto the 4D model to replicate the construction process. The construction sequence is then codified into a construction programme in the 4D BIM model. In addition, the constraints and requirements for the processes are recorded and developed in the form of method statements for planned processes. However, only the 4D BIM model is in a machine-readable format and used for applications such as BIM based rule-checking. The requirements and constraints are recorded

as method statements in PDF format. These PDF documents are intended to be used as a reference for the engineers on the site. The lack of semantics in the machine-readable construction process information greatly contributes to the codification challenge discussed in Chapter 4.

The proposed LDCC method has a significant advantage over the conventional method as it codifies the constraints and requirements in the construction process information into a machine-readable form so that BIM based rule-checking can be applied. To import the codified information into the prototype web application, the workflow discussed in Chapter 5 was followed. Geometry and project information in the 4D BIM model was converted to IFC using the inbuilt IFC exporter of the 4D BIM package. Then, the IFC file was converted to ifcOWL schema using the IFC2RDF converter (Vu Hoang, Nam Törmä, 2015) and stored as an RDF graph. Finally, construction process information in the form of RDF graph was imported to the prototype web application. The model UI after the import is shown in Figure 7.2. Sub-activities and constraints were added to the processes from the method statements manually. User interface outputs after adding sub-activities are shown in Figure 7.3 and Figure 7.4.

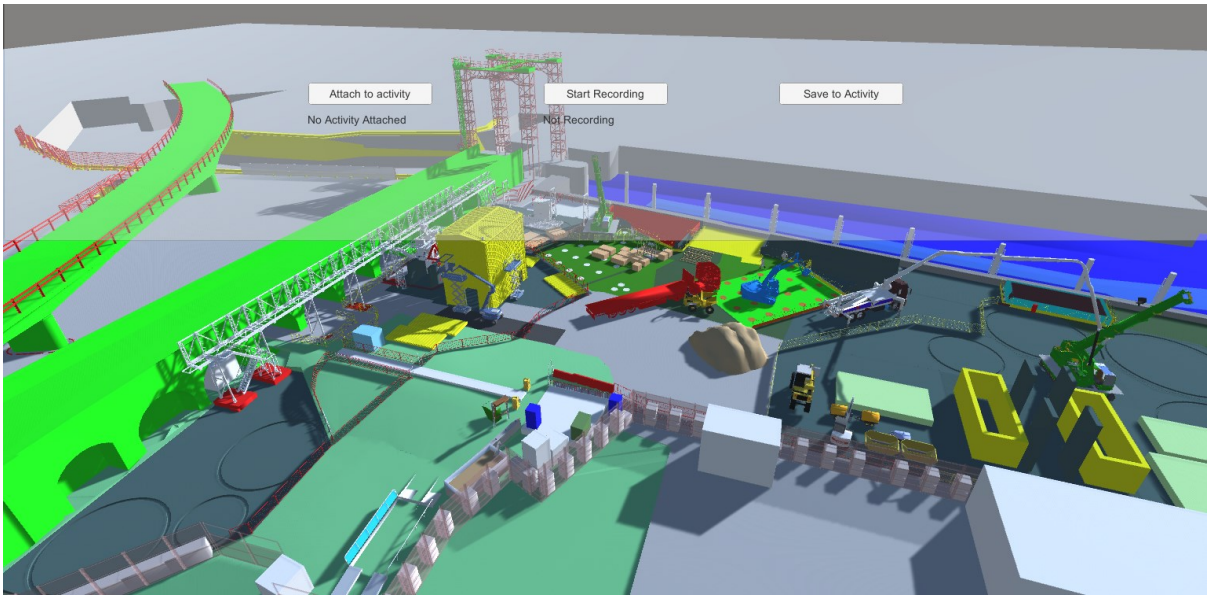


Figure 7.2: Data imported to the LDCC prototype

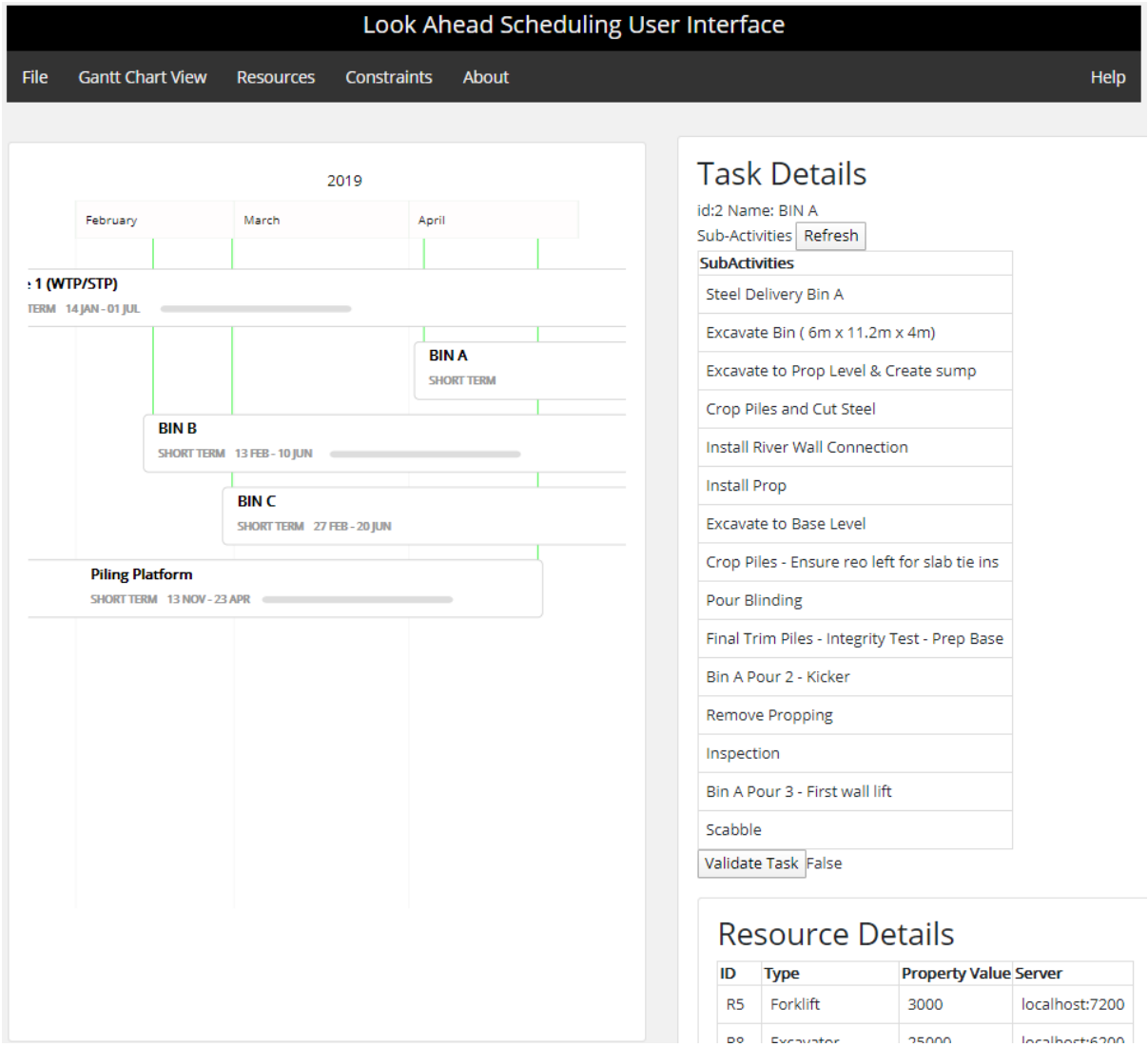


Figure 7.3: Task edit UI showing sub activities details added to the processes
Five activities can be seen in the left panel, of which BIN A is selected and their finer details and sub-activities shown in the right panel

The dates of the sub-activities were added as per the information in the method statements and construction plan. A total of 81 sub-activities for four processes (Slurry treatment plant foundation, Muck Bin A, Muck Bin B and Muck Bin C) were added. Information concerning the processes and sub-activities are stored as RDF graphs in the prototype web application. Triples in the RDF graph are stored in the triple store.

Constraints were added to the sub-activities and the resources. A discrete resource constraint (ensuring that used resources do not exceed the number of resources available at any point) was added to each type of resource. Finish-to-start precedence relationships were added

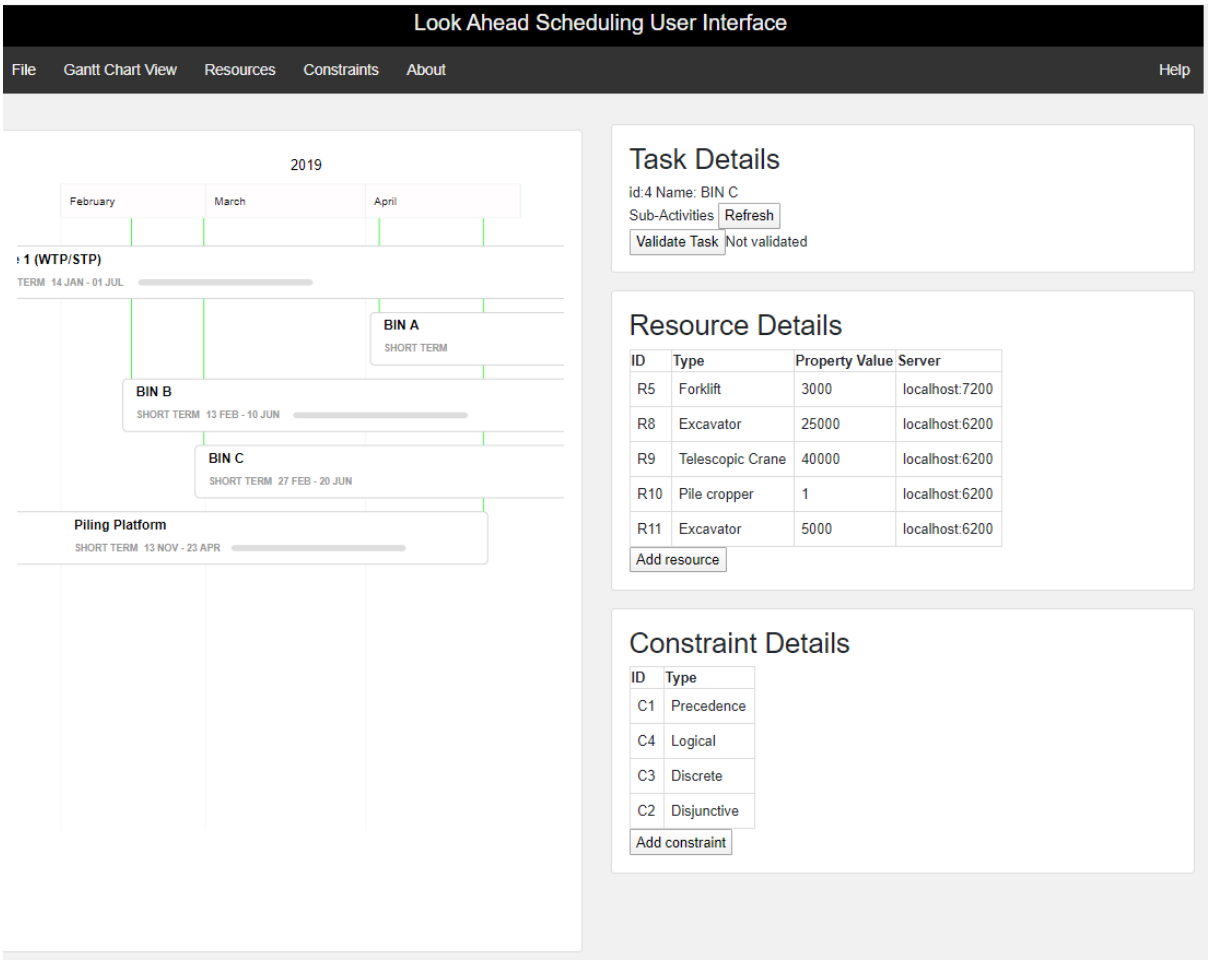


Figure 7.4: Task edit UI showing resource details and constraint details added to the processes for BIN C

to the sub-activities in the processes. Some activities, such as *STP foundation process* sub-activities *Concrete downstand Beam and Install Recess B6-3 + Levelled, Install Recess B6-2 + Levelled, Install Recess B6-1 + Levelled and Install Perimeter Shutters* had start-to-start relationships. These were modeled as SHACL shapes. Disjunctive constraints were added to *inspection* and *curing* sub-activities in *Muck bin construction* for *Bins B and C*, and also to the *curing* sub-activity in *STP foundation*. Disjunctive constraint ensures that no other activities take place during these sub-activities in that space. Logical constraints were applied to the 40T crawler crane (assigned for *STP foundation* process) and the 40T telescopic cranes (assigned for the *Muck bin construction* processes). The operating requirement of the cranes was set as 40 T. The material resources for each activity was taken from the RDF graph. The weights of steel, prop, formwork, perimeter shutters, steel top matt, screed rails, steel bottom matt and

others were added to the operating requirement value of the respective objects. Thus, the assigned logical constraints will check the operating requirement value of an object against the operating requirement value of the crane resource. In addition to the discrete, precedence, disjunctive and logical constraints, cardinality constraints were also applied to the schedule data to ensure that the schedule contains all the required parameters such as start-date, end-dates, and object assignment.

In contrast to the practice, the LDCC prototype codifies the constraints and relationships into a machine-readable format. Codification of constraints would enable automated constraint checking to be applied to the codified information. To do the constraint checking on the codified process information, the validate button is clicked on the task edit UI. When the validate button for a process is clicked, each process is checked for constraint violation. The results of the constraint checking are presented in Subsection 7.4.1.

Diverging from the current best practice, which requires the manual generation of LAS, I proposed here the possibility of automating the generation of a LAS for the time-period by taking advantage of the codified information in the LDCC prototype. The start-dates and the end-dates of the activities of sub-activities were removed from codified information. Then the detailed construction process information (in the form of an RDF graph) and the modeled constraints (in the form of a shapes graph) were submitted as an input to the RL and GA algorithms. Two LAS was generated respectively using these algorithms with the parameters shown in Table 7.4 (also described in Chapter 6).

Table 7.3: Parameters for the LAS generation

Algorithm	Parameter	Value
GA	Crossover probability	0.8
	Mutation probability	0.2
	Population size	100
	Number of generations	100
Reinforcement Learning	Learning rate	0.4
	Discount factor	0.6
	Number of episodes	30

7.4. Results

This section presents the results of constraint checking using the LDCC prototype (Subsection 7.4.1) and LAS generation using GA and RL (Subsection 7.4.2). The constraint checking detected violations for five sub-activities. Also, the two LASs generated by the algorithms were broadly like the one generated by the practitioners. The results were presented to the digital engineer to get the practitioner feedback. The following subsections discuss the results in detail.

7.4.1. Constraint checking using LDCC

Table 7.4 presents the results of the constraint checking. Out of the 81 sub-activities belonging to four processes, LDCC detected constraint violations in five sub-activities. Of the five constraint violations, two sub-activities had cardinality constraint violations, and three sub-activities had precedence constraint violation. End-dates were not present in the sub-activities that had cardinality constraint violations. The constraint violation reported for Muck bin construction Bin A is shown in Figure 7.6. The violations detected by the software were presented back to the digital engineer responsible for maintaining the 4D models and the digital engineer was requested to comment on the constraint violations.

Table 7.4: Violations detected in the LDCC prototype

Process	Sub activity	Constraint violated
Muck bin construction Bin A	Crop Piles	Precedence constraint
Muck bin construction Bin A	Final Trim Piles-Integrity test – Prep base	Precedence constraint
STP foundation	Pour	Precedence constraint
Muck bin construction Bin C	Bin C Pour 2 - Kicker	Cardinality constraint
STP foundation	Concrete downstand Beam	Cardinality constraint

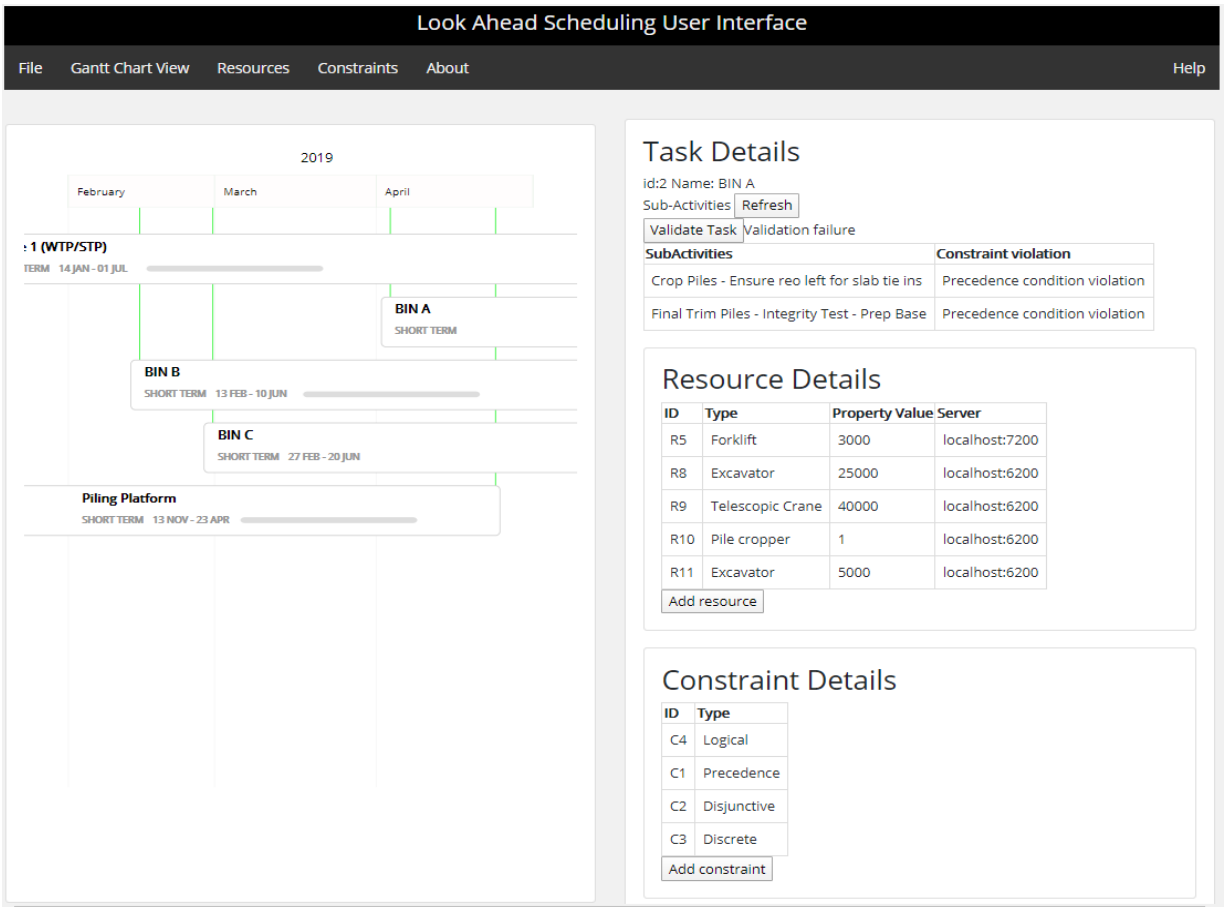


Figure 7.5: Constraint violation for Muck bin construction Bin A

The digital engineer compared the results of the constraint checking with the 4D BIM model and method statements, and the provided reasons for the constraint violation from a practitioner's perspective.

7.4.1.1. *Precedence constraint violations*

For the precedence constraint violation in sub-activity *MuckBin Construction Bin A crop piles*, the digital engineer was able to detect the error in the 4D BIM, where a precedence constraint was assigned. However, the manually created LAS did not conform to this condition.

> "So initially the excavation was excavation done first, and then when this was finishing, the next one was starting out. […] However, we were able to excavate, and while we were excavating, we were cropping some piles at the same time. So basically those [excavation] and cropping take took place at the same time." Digital engineer

Although there was a precedence constraint assigned to the Crop piles activity, the crop piles could be done in parallel with the excavation activity. Ideally, a start-to-start relationship could have been assigned. However, this was not done in the model. The site team realized the possibility for doing the parallel execution, and they adjusted the start-date on the look-ahead-schedule. However, the precedence relationship was still embedded in the model. Since the LDCC method extracted the constraint from the ifcOWL file, this was raised as a violation during the constraint checking. Similarly, there were precedence constraint violations for sub-activities *Muck Bin Construction A Final Trim Piles-Integrity test – Prep base* and *STP Foundation: Pour*.

7.4.1.2. *Cardinality constraint violations*

There was a cardinality constraint violation detected in the sub-activities Muck bin construction Bin C: *Bin C Pour 2 – Kicker and STP Foundation: Concrete downstand Beam*. The end-dates of these activities were not present in the schedule.

> "This is a start key date[…] it has a lot of predecessor and successor, but in the program, it has zero duration […] When we are updating project in [BIM package 1] and then importing to [BIM package 2], some activities get converted to start key date, and the duration is converted to zero, and the end-dates are removed. We have to manually add the end-dates to such activities after the import " Digital engineer

In this case, the construction programme was exchanged between two BIM packages and during the export and import process, some activities were converted to *start key date* activities. *Start key date* activities does not have a duration or end-date. However, the start key date definition does not apply to the IFC schema. Hence the prototype filtered out these activities for constraint violations. If it were not for the prototype, the digital engineer would have to filter out these inconsistencies manually.

7.4.2. Look-ahead-schedule (LAS) generation using GA and RL

This section presents the results of the LAS generation using GA and RL for the set of activities studied. GA came out with a LAS within 423 seconds, and RL obtained another LAS within 11 seconds. Both LAS were broadly similar to the LAS prepared by the practitioners. Out of the 81 sub-activities, LAS generated by both algorithms deviated from the LAS prepared by practitioners for 11 sub-activities. Remaining 70 sub-activities were scheduled identical to the manually developed LAS. The sub-activities which had deviations are presented in Table 7.5.

LAS generated by both algorithms conformed to all the modeled constraints leading to deviations from the LAS prepared by the practitioners for certain activities. In the LAS prepared by practitioners, cardinality constraint violations were detected for two sub-activities *Muck bin construction Bin C: Bin C Pour 2 – Kicker* and *STP Foundation: Concrete downstand Beam* as the end-date was not assigned for these sub-activities. However, in the LAS generated by the algorithms, end-date was assigned to these activities. Assigning an end-date to that activity resulted in the change in start-dates of the downstream sub-activities *Install Recess B6-3 + Levelled, Install Recess B6-2 + Levelled, Install Recess B6-1 + Levelled, Install Perimeter Shutters* and *Concrete Curing (To Strike Shutters)* for the *STP Foundation: Concrete downstand Beam*. Similarly, a precedence condition violation was identified for the *Muck Bin Construction Bin A:*

Table 7.5: Schedule deviations

Process	Subactivity	Planned by engineers		Predicted by GA		Predicted by RL	
		Start date	End date	Start date	End date	Start date	End date
Muck Bin Construction Bin A	Crop Piles	16/04/2019	24/04/2019	24/04/2019	2/5/2019	24/04/2019	2/5/2019
Muck Bin Construction Bin A	Pour Blinding	25/04/2019	25/04/2019	2/5/2019	2/5/2019	2/5/2019	2/5/2019
Muck Bin Construction Bin A	Final Trim Piles - Integrity Test - Prep Base	29/04/2019	14/01/2019	3/5/2019	18/5/2019	3/5/2019	18/5/2019
Muck bin construction Bin C	Bin C Pour 2 - Kicker	15/04/2019		15/04/2019	15/04/2019	15/04/2019	15/04/2019
STP Foundation	Concrete downstand Beam	3/4/2019		3/4/2019	4/4/2019	3/4/2019	4/4/2019
STP Foundation	Install Recess B6-3 + Levelled	3/4/2019	4/4/2019	4/4/2019	5/4/2019	4/4/2019	5/4/2019
STP Foundation	Install Recess B6-2 + Levelled	4/4/2019	5/4/2019	5/4/2019	6/4/2019	5/4/2019	6/4/2019
STP Foundation	Install Recess B6-1 + Levelled	4/4/2019	5/4/2019	5/4/2019	6/4/2019	5/4/2019	6/4/2019
STP Foundation	Install Perimeter Shutters	4/4/2019	5/4/2019	5/4/2019	6/4/2019	5/4/2019	6/4/2019
STP Foundation	Concrete Curing (To Strike Shutters)	5/4/2019	5/4/2019	6/4/2019	6/4/2019	6/4/2019	6/4/2019
STP Foundation	Pour	23/04/2019	23/04/2019	24/04/2019	24/04/2019	24/04/2019	24/04/2019

Crop Piles sub-activity in the LAS generated by practitioners, and this violation was not present in the LAS generated automatically. Rectifying it resulted in a deviation that was further propagated to start and end-dates for *Pour Blinding* and *Final Trim Piles - Integrity Test - Prep Base* sub-activities in the same process. Likewise, precedence violation detected for STP *Foundation: Pour* was not present in the LAS generated by the algorithms.

The LAS generated by the algorithms were presented to the digital engineer. The results from the automated LAS generation algorithms were broadly similar to the LAS prepared by the practitioners. Hence the results self-validate themselves. Because deviations in the LAS generated emerged as a result of correcting the constraint violations, the digital engineer did not comment on it.

7.5. Discussion

The implementation study presented in this chapter demonstrated the potential of the LDCC method to detect constraint violation in the LAS stored in 4D BIM models. In particular, the software was able to automatically detect cardinality constraint violations in the model. These violations were present in the LAS generated by the practitioners primarily due to the loss of information during conversion from one BIM package to another. In state-of-the-art practice, these violations should be manually searched by the practitioner. However, using LDCC, these violations can be automatically identified by the prototype, thereby saving the time and effort in manually identifying these constraints. Information loss during the conversion of formats reinforces the observation related to software usage made in Chapter 4, and the current LDCC method helps identify these inconsistencies in the dataset.

The software also detected precedence constraint violations in the model. However, the feedback from the digital engineer suggests that these were not constraint violations. Instead, these were last-minute changes that were not wholly incorporated back into the model, causing

LDCC to raise a constraint violation. This piece of information points to the fact that LDCC will be only as accurate as the data and constraints are given as an input. When the input is of low quality, LDCC may produce conflicting results. This also means that there would be social processes in the construction site that optimizes plan of work. Digital tools cannot replace these social processes.

The similarity of the LAS generated by the algorithms with the one generated by the practitioner demonstrates that it is indeed possibly to automate the generation of LAS, and that both GA and RL are suitable algorithms. When compared with the practice of creating LAS manually, the automated LAS generation creates the schedule accurately in a short time. The computation time required for generating LAS was minimal (423 seconds for GA and 11 seconds for RL) when compared with manual methods (a few hours of the practitioner). However, the LAS generated by the automated methods strictly adhere to the inputs of the model. This is evident from the deviations caused by correcting the precedence constraint violations, even when they are not constraint violations. Thus, the LAS generated is only as accurate as the inputs given into the model.

The implementation study presented in this chapter focused on testing for accuracy and speed. The study did not test the ease of use of the developed method in real practice as the worked example was done by me, the same person who developed the tool. When deployed in practice, the practitioners should be trained to use the tool, and they should have basic knowledge of modelling constraints in SHACL. However, the focus of this study was to demonstrate the potential of the method and not the software prototype. Future work needs to be done to improve the usability of the software to develop it into a production package. Despite these limitations, the effectiveness of the LDCC method and LAS generation method in terms of accuracy and speed was validated in this research.

7.6. Summary

This chapter presented how the LDCC method developed in Chapter 5 and the automated look-ahead-schedule generation algorithms presented in Chapter 6 can be applied to a real project to assist look-ahead planning. Implementation was done at a site in the Water project. When the schedule prepared by the engineers was enriched with construction constraints and checked for constraint violations, violations were detected for five sub-activities. These constraint violations were discussed with the digital engineer to make sure that the detected violations were indeed true positives. Further, the LAS generation methods(GA and RL) predicted an LAS each for the 30-day period consisting of 81 activities. The predicted schedules were identical to the ones planned by the engineers for 70 out of 81 activities. There were deviations for 11 activities. Deviations are justified as they emerged from correcting the five constraint violations. Thus, this chapter demonstrated the applicability of the LDCC method and LAS generation method using a worked example on real project data.

8. DISCUSSION AND CONCLUSIONS

8.1. Research summary

This thesis has achieved its aim of supporting data-driven decision-making for look-ahead planning in construction by developing an information modelling approach to codify and validate detailed construction process information in BIM-enabled projects. Prior research that studied the implementation of look-ahead planning in construction had identified inefficiencies such as late identification of constraints violations (Hamzeh, 2009) and errors in constraint analysis (Nieto-Morote & Ruz-Vila, 2012; Lindhard & Wandahl, 2014) resulting in error-prone LAS that are non-reliable. Although, data-driven approaches to look-ahead planning can address these inefficiencies, these require detailed construction process information to be codified into machine-readable formats. A key contribution of this thesis is the development of the LDCC method that can codify and validate detailed construction process information distributed across heterogeneous sources.

To understand issues of codification of construction information in late-design and construction stages, three construction projects were studied qualitatively (Chapter 4). Software usage, information sharing practices and missing construction process information were found to cause data-quality issues limiting the application of data-driven decision-support systems. Building on the understanding of these data quality issues in construction, a novel LDCC method was developed to codify detailed construction process information including complex construction constraints (Chapter 5). To support data-driven decision-making, the developed method checked the codified construction process information against the set of constraints to identify constraint violations. A key advantage of this method is that it could check for constraint violation even when the information is distributed over multiple servers in

heterogeneous formats. Further, to reduce the errors that might occur during the manual

Table 8.1: Summary of the research

Objective	Research method	Key findings	Contribution
O1) To understand the codification challenges in the late-design and construction phase of BIM-enabled projects	M1) Interpretive case study research (Section 3.2)	Codification challenges are identified to be 1) *software usage*—interoperability, information loss during conversion, and multiple modelling techniques, 2) *information sharing*—unstructured information sharing, drawing and file-based sharing document control, and lack of process change, and 3) *missing process information*—loss of constraints and low levels of detail (Chapter 4).	Extended the understanding of the data quality in construction information by identifying the codification challenges emerging from the broader practice of using model-based and document-based information sharing
O2) To develop an information modelling approach to codify and validate construction process information including dynamic constraint information	M2) Prototype development and testing (Section 3.3)	LDCC method using SHACL can be used to model complex constraints spanning multiple domains and identify constraint violations to perform data-driven constraint checking (Chapter 5).	Extended the knowledge on linked-data and BIM based rule checking to support data-driven decision-support for look-ahead planning
O3) To develop a method to generate look-ahead-schedule for construction projects automatically using the codified process information.	M3) Machine learning model development and testing (Section 3.4)	LDCC method can be integrated with GA and RL to generate LAS. RL that used parallel schedule generation scheme generated LAS faster than GA that used serial schedule generation scheme, and RL produced LAS with shorter project durations for complex scenarios (Chapter 6).	Extends the methods to solve RCPSP to include complex construction constraints in addition to the precedence and discrete resource constraints

development of LAS, automated methods to generate LAS were also developed (Chapter 6). These automated methods reduced the time and effort required to create the LAS. The developed methods were tested using data from a live construction site (Chapter 7). The methods were able to identify constraint violations and generate a constraint free LAS for a look-ahead window, thus demonstrating the efficacy of the method. Table 8.1 presents the summary of the research.

The main contribution of this thesis is that it addresses the implications of data silos in construction process information by developing a novel information modelling method to codify detailed construction process information distributed over multiple sources and using the codified information to provide data driven insights for look ahead planning. Further, the practical implication of this research is that it reduces the errors in manual look-ahead planning through the development of the novel LDCC method by performing data-driven constraint checking to identify constraint violations and using it to generate conflict free LAS automatically. In addition, the research presented in this thesis extends the body of knowledge concerning automated look-ahead planning as described below.

- This PhD study adds to the trajectory of research focusing on the development of decision-support systems to aid look-ahead planning (Dave *et al.*, 2010; Dong *et al.*, 2013a; Sacks *et al.*, 2013; Laine, Alhava & Kiviniemi, 2014) by presenting the LDCC method to model and validate complex construction constraints spanning multiple domains.
- The LDCC method extends the knowledge on BIM based rule-checking (J. Choi *et al.*, 2014; Dimyadi *et al.*, 2016; Schwabe *et al.*, 2019; Zhong *et al.*, 2012) as it demonstrates

the use of linked-data based methods to perform rule-checking without the need for the construction information to be stored in one model or database.

- This research adds to the understanding on the applications of linked-data to construction applications (Pedro *et al.*, 2017; Pauwels, 2014; Pauwels & Terkaj, 2016; Lee *et al.*, 2016) by documenting the use of SHACL to perform constraint checking to identify construction schedule violations. Evidence from this shows that this method can be used to validate semantic and syntactic accuracy of digital information for model checking.
- This study extended the knowledge on automated projected scheduling methods (Bruni, Beraldi & Guerriero, 2015; Munlin, 2018; Pellerin, Perrier & Berthaut, 2020) to include complex construction constraints such as disjunctive constraints and logical constraints (in addition to precedence and discrete resource capacity in traditional RCPSP). Also, this study demonstrated how RL could be used to solve heavily constrained LAS problems.
- This work extends the knowledge on data quality issues in construction information (Mirarchi & Pavan, 2019; Zadeh *et al.*, 2017) by documenting the codification challenges emerging from broader practice of information sharing using both file-based and model-based methods.

Th rest of this chapter presents the discussion of the combined findings (Section 8.2), main conclusions (Section 8.3), implications to the practice (Section 8.4) and directions for future research (Section 8.5).

8.2. Combined discussion of the findings

This section describes how the LDCC method and LAS generation method aids the data-driven decision-making in look-ahead planning and thus contribute to achieving the overall aim of the research. In conventional look-ahead planning, the activities that are

supposed to take place in the look-ahead window are taken from the master schedule and then analyzed to understand whether the activity can take place in the look-ahead window (typically six weeks). (Ballard, 2000; Sacks *et al.*, 2017a). Constraint identification and removal is a vital step to determine whether the activity can take place or not and it is manually done in conventional look-ahead planning. However, the constraints are complex and interdisciplinary (Wang *et al.*, 2016). Identifying constraints require inputs from different project participants and participants are often less transparent about the constraints (AlSehaimi, Tzortzopoulos & Koskela, 2009), resulting in sluggish removal of constraints (Hamzeh, Zankoul & Rouhana, 2015) and often releasing the tasks that are not ready to be executed into the LAS (Lindhard & Wandahl, 2014; Nieto-Morote & Ruz-Vila, 2012).

In contrast with manual identification of constraints, the LDCC method offers a data-driven method to automatically identify the constraints from the codified information. It uses linked-data principles to link multiple datasets across disciplines and model the constraints spanning these domains. The constraint identification happens automatically using the SHACL process engine. Thus, identifying the constraint violation becomes an automated data-driven exercise and results in lesser errors. Also, by integrating the LDCC method with machine learning methods, the practitioners are equipped with error-free LAS based on the codified constraints. Thus, through the development of LDCC and LAS generation method, this study contributes to the prior studies focusing on using computer-based tools to support look-ahead planning (Sacks *et al.*, 2013; Dave *et al.*, 2010). These methods can be integrated with Internet of Things (Dave *et al.*, 2016a) and ADC technologies (Cheng *et al.*, 2010) to further aid the practitioners in aiding look-ahead planning and generating error-free LAS.

This PhD study also extends prior information modelling approaches to codify construction process information (Darwiche, Levitt & Hayes-Roth, 1988; Akinci *et al.*, 2002b; Dong *et al.*, 2013a) and then use it to generate LAS. It addresses the limitation of the existing

studies which assumes that all the construction information is available as a single model. However, findings from this thesis (see Chapter 4) contests this assumption and gives evidence that the construction information is siloed and distributed in multiple heterogeneous datasets. This PhD study addresses these limitations and develops a method to codify and validate construction process information, including construction constraints such as disjunctive constraints and logical constraints. By doing so, the developed method also addresses the three main codification challenges (software usage, information sharing and missing process information) identified in Chapter 4. How these challenges are addressed through the development of the LDCC method and machine learning methods such as GA and RL to generate LAS automatically are discussed in the following subsections.

8.2.1. **Software usage**

The findings from the case studies reinforce the observations reported in the literature about construction information remaining siloed and distributed in multiple heterogeneous datasets due to software usage issues (Dossick & Neff, 2010; Abanda *et al.*, 2015; Tchouanguem Djuedja *et al.*, 2019). In their study, Dossick and Neff (2010) showed that the organizational and cultural divisions between designers, builders, and subcontractors stifle collaborative work. My own findings show that these organizational and cultural divisions also lead to the use of multiple tools and data formats leading to software usage issues, as discussed in Subsection 4.2.1. Although each stakeholder in the project may have a shared norm for modelling information, and they use a set of tools to do that, these norms differ for other stakeholders. Differences in shared norms for modelling information results in datasets which are not interoperable and require format conversions, resulting in loss of information and low machine-readability, thus limiting the application of machine learning algorithms. Even when using CDE in the projects—this siloed information now resides in an accessible cloud-based repository—the information is still stored as individual files which are not interoperable, thus

reducing the machine-readability of the information. In the case of construction information during the late design and construction stages, this is evident from the use of planning tools such as Microsoft Project, Asta Powerproject, and Oracle Primavera P6, CAD tools such as Autodesk AutoCAD and Bentley Microstation, BIM-review tools such as Autodesk Navisworks, Bentley Navigator, Bentley Synchro pro in combination with Microsoft Office tools such as Excel, PowerPoint and Word for communication. The siloed information limits the application of conventional BIM based rule-checking systems (Lee, Solihin & Eastman, 2019; Nawari, 2012; Schwabe, Teizer & König, 2019) to support look-ahead planning and also limits the application of machine learning algorithms to generate LAS.

To address this issue, I adopted a linked-data based approach (Berners-Lee, 2006) in my study. By doing so, this study extends the knowledge on the use of linked-data (presented in Subsection 2.3.3.1) in the construction sector by modelling construction processes as RDF graphs, with OWL being used to provide a semantic structure to these graphs. OWL enables the creation and sharing of domain-specific ontologies (Terkaj and Pauwels, 2017). They provide opportunities to store the context of the data specific to a particular stakeholder while maintaining links to the information required by other stakeholders in a machine-readable manner (Pauwels, Zhang & Lee, 2017; Ei-Diraby, Lima & Feis, 2005). Instead of proposing a completely new ontology for construction process information, the current study uses the IFC description of construction information, as it promotes a vendor-neutral description of construction information supported by most of the software vendors. The ifcOWL(Beetz, Van Leeuwen & De Vries, 2009; Pauwels & Terkaj, 2016) representation of IFC was used to leverage the semantic expressivity of OWL in the current study. As the ifcOWL ontology could not describe the data to the extent required for look-ahead planning, I developed the LinkOnt ontology (described in Section 5.2) to add the context information.

This study demonstrated the power of linked-data using a public BIM dataset and enriching it with further data required for look-ahead planning. The enriched data was used for constraint checking. The data, together with its context, were in the form of RDF graphs and resided in multiple servers; despite this, the proposed method could make the inferences from the siloed data (Subsection 5.3.2). Thus, this study provides practical evidence for the use of linked-data to solve the interoperability problems in the construction sector as the data published as RDF graph contains the context of the data embedded in itself while following linked-data principles.

By following the linked-data approach, different companies involved in a project are not required to follow a single modelling methodology or a single tool. Instead, they can use their preferred tool and modelling methodology. However, they need to use a domain-specific ontology to describe their data in a machine-readable format. Their domain-specific ontology should also be linked to the other standard ontologies so that semantic query languages can derive the context from the modeled information. In the current study, LinkOnt was designed as a domain-specific ontology for supporting look-ahead planning, while ifcOWL was the standard construction ontology. In this way, specialized tasks were performed without affecting the rest of the data required for the tasks by other teams. This method also avoids the need for converting between formats as the data is stored as triples. The format lies in the description of the ontologies and not in the data structure itself. As the same data can be described using multiple ontologies, there is no need for format conversions.

The current study provides a technical solution to the issue of construction information siloes and offers a solution to do that without affecting the individual activities of different project participants. In contrast with the conventional method of using multiple tools and converting between formats, the current method offers the ability to separate the data and context. For effective use of this method, some of the roles in look-ahead planning phase—

such as the role of the digital engineer—require adaptation. They should ensure that the detailed construction information, previously documented in the form of method statements should also be codified. In agreement with Akintola, Venkatachalam & Root, (2017) observations, there may be new roles created for this purpose in the beginning owing to the expertise required to model construction information using RDF and SHACL. However, these roles would fall within the scope of traditional functions of creating method statements, however being carried out using the LDCC method.

8.2.2. **Information sharing**

Participants in the investigated cases use a combination of different practices of information sharing, as discussed in Subsection 4.5.2 resulting in reduced data quality. These information sharing practices include drawings and file-based sharing, unstructured information sharing, printing, and scanning of documents, multiple CDEs. Even though the projects studied championed the BIM use and followed BIM based workflows using CDE, the findings from case studies give evidence that the existing robust practices of file-based information sharing exist alongside the BIM based workflows. This observation from the case studies is in agreement with the observations from Harty and Whyte (2010), where they observed the coexistence of paper-based drafting and computer-aided drafting practices in design and drafting offices. The project participants are accustomed to traditional practices of information sharing, and these practices are intuitive to them. However, the BIM-based information sharing workflows are not comfortable to the project participants, and they are not as robust as the prior information practices; this leads to the circumnavigation of workflows and results in unstructured information sharing. Unfortunately, the combination of different practices for information sharing reduces the data quality and limit the application of machine learning algorithms to support look-ahead planning.

Although it is impossible to replace the existing robust workflows to BIM-based workflows, Hartmann (2008) have recommended that the technologies that suite the existing workflows are more likely to be adopted by construction teams. Building on Hartmann (2008) argument, the LDCC method I developed for this study attempts to situate itself within the look-ahead planning workflow. In contrast with the conventional method of producing method statements in PDF format, LDCC method captures the decisions taken during the look-ahead planning meetings and codifies it into a model-based machine-readable format as described in Section 5.2. Thus, by developing the LDCC method, this study addresses the issue of document-based sharing of detailed construction process information, especially construction constraints. However, the LDCC method does not replace conventional methods for look-ahead planning. Instead, it supports look-ahead planning by checking for constraint violations, which if done manually, is a tedious process. The LDCC method is anticipated to be used in coexistence with other tools such as 4D planning tools and BIM review tools in look-ahead planning meetings to address the existing limitation of 4D planning tools to model complex construction constraints. However, it is necessary to acknowledge that there would always be coexistence of LDCC methods with conventional methods for documenting detailed construction processes.

8.2.3. **Missing process information**

The findings from the interpretive case study indicated that the product information is well codified in the form of BIM models and drawings, whereas construction process information is relatively unstructured often limited to the precedence relationship between activities at master plans stored in the planning tool. This observation is aligned with prior research (Han & Golparvar-Fard, 2017; Giretti *et al.*, 2012), that reported that the codified process information lacked necessary details to compute progress or productivity at an activity level. Also, constraint relationships between activities, resources and workspaces are not

codified in a machine-readable format (refer to Subsection 4.4.3). In the context of look-ahead planning, the lack of detailed information and semantic constraint relationships embedded in a model limits the application of BIM-rule-checking systems and automatic schedulers.

The LDCC method addresses the issue of lack of codifying detailed construction process information, including complex construction constraints. LDCC uses linked-data to model construction information that can be stored in several places. This method, when deployed during the look-ahead planning meeting, can encode the construction knowledge generated and discussed in the meeting. The developed prototype can codify the details about the relationship between the subtasks and resources, between the resources and constraints, between the constraints and subtasks, and between the subtasks and subtasks into a machine-readable format. The codified knowledge on the relationships and constraints can be reused to aid look-ahead planning by checking for constraint violations, as discussed in Section 5.4. The constraint checking follows the principles of BIM based rule-checking. It followed the approach for rule-checking suggested in Eastman *et al.* (2009) and reached the complexity of class 3, as stated in Solihin and Eastman (2015).

In Chapter 6, the LDCC method was combined with two machine learning methods (GA and RL) to automatically generate LAS, giving us evidence for the reuse of codified construction process information for automated look-ahead scheduling. What is more, the LAS generated in this way was broadly similar to the ones created by the engineers (Chapter 7). Previous studies to solve RCPSPs (Hartmann & Briskorn, 2010; Brcic, Kalpic & Fertalj, 2012) considered only discrete resource capacity and precedence constraints. Unlike them, the proposed method considers construction scheduling constraints such as disjunctive constraints (Morkos, 2014) and logical constraints necessary for LAS.

8.3. Thesis conclusions

Four conclusions drawn from this PhD study are presented in this section.

Conclusion 1) The partial codification of information in machine-readable formats is based on a range of challenges in late design and construction phases of BIM-enabled projects, related to software and information sharing, with a particular lack of process information.

While all studied projects were BIM-enabled, information sharing issues and software usage issues emerge within these projects because of their document-based—and sometimes work-based—practices. Project-wide standards and policies are created to streamline information sharing through structured workflows, and these workflows are aimed at improving collaboration, but the process of document control raises issues. Also, this study identified inefficiencies (such as long processing times, complexity, etc.) in these workflows, which pushes the users to bypass the structured methods in the existing workflows. These inefficiencies force the users to revert to older methods or use a combination of old and new methods, resulting in the generation of unstructured information. In addition, the data from the case studies suggest that the level of detail of the scheduling is still at a macro-planning level. Besides, the constraint relationships between resources and processes, resources and the site conditions, and site conditions and the processes discussed in weekly meetings are not codified into a model. These cause data quality issues relating to inaccessibility of data, completeness, timeliness of data, and data consistency.

Conclusion 2) The issue of data siloes in construction information can be addressed by leveraging the power of linked-data to traverse across the web of data created by linking across the heterogeneous construction information datasets.

The prototype evaluation of the LDCC method developed in this study has demonstrated the power of linked-data to traverse across the web of data and make meaningful inferences.

During the prototype evaluation (Subsection 5.3.2), the information stored in one server was automatically accessed through the discrete resource constraint applied on a resource stored on a separate server automatically. Thus, the current study has demonstrated the advantages of publishing construction process information to address inherent problems of interoperability associated with construction information. By storing the domain knowledge in the form of ontologies in a machine-readable format using OWL, the implicit knowledge and context concerning data can be inferred as the description logic is embedded in the ontology. The ontologies can be linked to each other to share the information necessary for common tasks. By doing so, each project participant can share the modeled information without causing software usage issues such as interoperability and loss of information during the conversion. By publishing the construction information as RDF graphs, the data quality related to accessibility, accuracy and consistency are improved as the data, and the context of data is available online and linked to other datasets.

Conclusion 3) LDCC method using SHACL can be effectively used to model and validate construction scheduling constraints distributed over multiple datasets, thus supporting look-ahead planning.

This study presented a novel method using linked-data technologies to model and validate complex scheduling constraints using LDCC. Construction-scheduling constraints, such as cardinality constraints, logical constraints, precedence constraints, disjunctive constraints, and discrete constraints, were modeled and validated using SHACL. SHACL leverages the power of linked-data to query information by utilizing the semantic information stored in the respective ontologies. This work contributes by demonstrating the potential of this approach through the development and evaluation of a prototype web application based on LDCC. The developed application could identify construction scheduling constraint violations for a set of input data distributed over multiple servers. The identification of constraint violations using

LDCC for a public dataset as well as real construction data demonstrates the capability of LDCC to model and validate construction scheduling constraints using SHACL.

Conclusion 4) Construction information codified using LDCC can be reused to generate LAS by incorporating LDCC within the machine learning framework.

The thesis also presented how the developed LDCC method can be integrated with machine learning frameworks such as GA and RL to generate LAS automatically. Integration of LDCC with machine learning frameworks removes the need for mathematical modelling of complex construction scheduling constraints. Instead, the constraints can be modeled in SHACL, and they will be automatically used while generating LAS. This separation between the modelling of constraints from the machine learning framework brings modularity to the LAS generation as all the constraints do not need to be previously known. LAS generation methods were tested on simulated data as well as a real construction case, and the results from the testing show that the method can produce LAS similar to the ones prepared by the engineers.

8.4. Limitations of the research study

The current study has achieved the objectives set out using different research methods. Although the research methods followed are rigorous, there are limitations to the current research which need to be addressed through future studies.

- To identify the codification challenges in construction, the current study has focused on late design and construction phases. However, the codification challenges may arise from the practices set out in the earlier stages of the project. The current study does not explain the influence of the information creation and uses on the data quality issues.
- The LDCC method has demonstrated how construction scheduling constraints can be modelled and validated using SHACL. However, the current study does not demonstrate how geometrical constraints can be modelled using SHACL. ifcOWL, the ontology used

for representing geometry in the current study are not expressive enough to describe and make inferences from geometric data, thus limiting the modelling of geometric constraints using SHACL.

- The prototype developed in the current study uses ifcOWL as the base ontology for representing construction information. However, the ifcOWL is based on the IFC schema, and IFC schema was not created for semantic web applications. This makes modelling SHACL constraint using the ifcOWL ontology difficult since the user needs a good grasp of IFC schema to model the SHACL constraints.
- The LAS generation method developed in the study is a proof of concept to demonstrate how LDCC can be integrated with GA and RL. However, this study was not aimed at arriving at an optimal machine learning method for the generation of LAS. Therefore, there are opportunities for improvements in the GA and RL methods through the fine-tuning of parameters and Objective function.
- LDCC method and LAS generation methods were validated using a worked example method. The focus of the validation study was to test whether the developed methods work well with real-world data. However, the testing was done on the look-ahead planning data for only one construction site in the project. To ensure generalizability, the same has to be tested in multiple projects. In addition, the method was tested within the project wherein the PhD researcher modelled the SHACL constraints. Thus, the validation study does not test the ease of modelling the constraints when a practitioner is performing the modelling.

8.5. Implications to practice

The PhD presented LDCC method to codify construction process information, including dynamic constraints in a machine-readable format to address the issue of lack of detailed construction information. When combined with ADC technologies, LDCC mitigates the difficulty to track and monitor the prerequisites and validate constraints in complex

construction projects through data-driven constraint checking. In contrast with manual methods of constraint identification, LDCC method provides a faster and accurate identification of constraint violations. Thus, the LDCC method provides decision-support during look-ahead planning meetings by identifying the constraint violations automatically. Also, the LDCC method offers the opportunity to capture the construction knowledge created during look-ahead meetings. Captured construction reduces the need for redundant input from the stakeholders in subsequent meetings, as the stakeholder requirements and constraints are codified into a machine-readable format. Furthermore, this method acts as a "unit-testing" environment for detailed construction process information. Unit testing fosters model-based collaboration among stakeholders, because every time a change occurs, one can evaluate the consequences of change to the program and check for constraint violations using the codified construction information. Unit testing is possible because the collective construction scheduling input of all the stakeholders is codified as detailed construction process information and constraints. Besides, when the LDCC method is combined with machine learning frameworks, it offers the ability to generate LAS automatically using the codified information. Automatic generation of LAS is of relevance to the practitioners as the manual generation of LAS is time-consuming and prone to errors. Thus, integrating the LDCC method and LAS generation within the look-ahead planning workflows would augment the decision-making in look-ahead planning meetings resulting in generating an error-free LAS.

How might codification challenges be overcome to enable higher uptake of data science in construction? The evidence from this work points towards a need for change in the policy and practice to ensure machine-readability and better-quality construction information. The current policies on information modelling are aimed at improving collaboration amongst the project participants. Building on the existing policies, newer policies should ensure that the machine-readability and quality of the data are maintained during information sharing. Care

should be taken to ensure that newer policies suit the existing work practices in construction so that the workflows recommended by such policies are not bypassed by the users. Newer policies also ensure the transition from file-based information sharing to model-based information sharing, where the model data is shared based on the principles stated by Berners-Lee (2006) to ensure machine-readability. Standards under development, such as ISO/DIS 21597—information container for data drop, exchange specification parts 1 and 2—are moving in this direction, where linked-data technologies are used to define the relationships between documents and datasets, thereby automating the document control processes (International Organization for Standardization, 2019). These standards would have a positive impact on data quality pertaining to dimensions, timeliness, accessibility, and consistency.

Thus, in addition to the policies, it is, therefore, necessary to adopt newer methods in practice to make construction information machine-readable. Despite having advanced workflows and innovative approaches to ensure structured information in the studied projects, lack of process change after digitization of workflows was evident in the datasets, which included multiple CDEs, printing and scanning of documents, need for wet signatures, and such. These practices continue to occur due to the lack of trust in digital workflows; thus, there must be a change in this mindset, and trust in digital data must be developed in the construction practice. There is also a need for a shift to model-based information sharing from the file-based information sharing between teams. To address this issue, related to a lack of detail in process information, there is a need for change in the look-ahead scheduling practice to incorporate the codification of different constraint relationships into information.

8.6. Directions for future research

The current study is the first step toward a linked-data-based constraint checking environment. To exploit the full potential of this method, further research should be done to extend the LDCC method.

- Researchers working on linked-data in the architecture and construction domain can extend LDCC to investigate the applicability of this method to detect violations related to workspace constraints, construction safety constraints, and construction-site layout constraints by exploiting the geometric data in the domain models.
- Researchers working in the area of lean construction can extend the LDCC approach to cater to the constraints in real construction sites and test its efficacy in improving the flow of work and in developing further methods for enhancing model-based collaboration in look-ahead planning.
- Researchers working on ADC can extend the LDCC method to predict the knock-on-effects of the deviations detected by sensors to predict future disruptions. In addition, this study has demonstrated how to reuse construction information codified using the LDCC method to generate LAS.
- Data science researchers can investigate the potential of the use of LDCC as a simulation or processing engine and develop further optimized machine learning methods to generate LAS. These methods can be connected to the ADC methods to verify and validate the constraints at different periods and report the results back to the decision-makers to support data-driven decision-making.

Further, future research can build on this study to develop recommendations for ensuring the machine-readability of construction information generated throughout different stages of the lifecycle of the project. To achieve this, future research should focus on different aspects. Researchers should elaborate on the organizational issues identified in the current study across different phases of the project and amongst different stakeholders involved in the project.

Accounting for the complexity of information-use trends in construction, there is a need for to develop new information modelling approaches which adapt to the current work processes as well as support codification of information, such as algorithms for crawling

through the disconnected information to draw insights and learning algorithms to detect discrepancies in data (such as with accuracy, completeness, timeliness, consistency, and provenance) and predict their consequences. The current research has demonstrated the power of linked-data based methods to achieve this; however, more in-depth research needs to be performed to develop it into a user-friendly method.

Finally, researchers and managers of construction projects should work together towards developing workflows and information sharing practices which ensure the machine-readability of construction information. This work has provided a foundation for such future research by extending the knowledge on data quality issues in construction through the identification of codification challenges, considering the wider practice of model and document-based information sharing.

References

Abanda, F.H., Vidalakis, C., Oti, A.H. & Tah, J.H.M. (2015) A critical analysis of Building Information Modelling systems used in construction projects. *Advances in Engineering Software*. [Online] 90, 183–201. Available from: doi:10.1016/j.advengsoft.2015.08.009.

Agarwal, A., Colak, S. & Erenguc, S. (2015) Metaheuristic methods. In: *Handbook on Project Management and Scheduling Vol. 1*. [Online]. Springer International Publishing. pp. 57–74. Available from: doi:10.1007/978-3-319-05443-8_4.

Akinci, B., Fischer, M. & Kunz, J. (2002) Automated Generation of Work Spaces Required by Construction Activities. *Journal of Construction Engineering and Management*. [Online] 128 (4), 306–315. Available from: doi:10.1061/(ASCE)0733-9364(2002)128:4(306).

Akinci, B., Fischer, M., Kunz, J. & Levitt, R. (2002a) Representing work spaces generically in construction method models. *Journal of Construction Engineering and Management*. [Online] 128 (4), 296–305. Available from: doi:10.1061/(ASCE)0733-9364(2002)128:4(296).

Akinci, B., Fischer, M., Levitt, R. & Carlson, R. (2002b) Formalization and Automation of Time-Space Conflict Analysis. *Journal of Computing in Civil Engineering*. [Online] 16 (2), 124–134. Available from: doi:10.1061/(ASCE)0887-3801(2002)16:2(124).

Akintola, A., Venkatachalam, S. & Root, D. (2017) New BIM Roles' Legitimacy and Changing Power Dynamics on BIM-Enabled Projects. *Journal of Construction Engineering and Management*. [Online] 143 (9), 04017066. Available from: doi:10.1061/(ASCE)CO.1943-7862.0001366.

Akyol, D.E. & Bayhan, G.M. (2007) A review on evolution of production scheduling with neural networks. *Computers & Industrial Engineering*. [Online] 53 (1), 95–122. Available

from: doi:10.1016/j.cie.2007.04.006.

Alizadehsalehi, S. & Yitmen, I. (2016) The Impact of Field Data Capturing Technologies on Automated Construction Project Progress Monitoring. *Procedia Engineering*. [Online] 161, 97–103. Available from: doi:10.1016/j.proeng.2016.08.504.

AlSehaimi, A.O., Tzortzopoulos, P. & Koskela, L. (2009) Last Planner System: Experiences From Pilot Implementation in the Middle East. In: *17th Annual Conference of the International Group for Lean Construction IGLC 17*. [Online]. 2009 Taipei. p. Available from: http://usir.salford.ac.uk/id/eprint/9545/1/2009_Last_planner_sys_exps_from_pilot_implementation_in_middle_east.pdf:public [Accessed: 22 May 2020].

Alzraiee, H., Zayed, T. & Moselhi, O. (2015) Dynamic planning of construction activities using hybrid simulation. *Automation in Construction*. [Online] 49, 176–192. Available from: doi:10.1016/j.autcon.2014.08.011.

Apache Jena (n.d.) *Apache Jena - Apache Jena Fuseki*. [Online]. Available from: https://jena.apache.org/documentation/fuseki2/ [Accessed: 16 June 2020].

Arayici, Y., Fernando, T., Munoz, V. & Bassanino, M. (2018) Interoperability specification development for integrated BIM use in performance based design. *Automation in Construction*. [Online] 85, 167–181. Available from: doi:10.1016/j.autcon.2017.10.018.

Balas, E. & Vazacopoulos, A. (1998) Guided Local Search with Shifting Bottleneck for Job Shop Scheduling. *Management Science*. [Online] 44 (2), 262–275. Available from: doi:10.1287/mnsc.44.2.262.

Bali, M. (2009) *Drools JBoss Rules 5.0 Developer's Guide Develop rules-based business logic using the Drools platform*. [Online]. Birmingham, Packt Publishing. Available from:

www.packtpub.com [Accessed: 4 June 2020].

Ballard, G. (1997) Lookahead planning: the missing link in production control. In: *Proceedings of the 5th Annual Meeting of the International Group for Lean Construction, Gold Coast, Australia*. [Online]. 1997 pp. 1–14. Available from: http://iglc.net/Papers/Details/17.

Ballard, H.G. (2000) *The Last Planner System of Production Control*. University of Birmingham.

Ballard, H.G. & Howell, G.A. (1996) Can Project Controls Do Its Job? In: *Proceedings of the 4th annual conference of the International Group for Lean Construction.* 1996 pp. 1–16.

Barker, J.R. & McMahon, G.B. (1985) Scheduling the General Job-Shop. *Management Science*. [Online] 31 (5), 594–598. Available from: doi:10.1287/mnsc.31.5.594.

Batini, C. & Scannapieco, M. (2016) *Data and information quality*. [Online]. Cham, Switzerland, Springer International Publishing. Available from: doi:10.4017/gt.2012.11.02.348.704.

Baykan, C.A. & Fox, M.S. (1997) Spatial synthesis by disjunctive constraint satisfaction. *Artificial intelligence for engineering design analysis and manufacturing*. [Online] 11 (04), 245–262. Available from: doi:10.1017/S0890060400003206.

Beetz, J., de Laat, R., van Berlo, L. & van den Helm, P. (2010) Towards an Open Building Information Model Server. In: *10th International Conference on Design & Decision Support Systems in Architecture and Urban Planning*. [Online]. 2010 pp. 1–8. Available from: http://bimserver.org/wp-content/uploads/2010/11/Beetz-Berlo_ddss2010.pdf.

Beetz, J., Van Leeuwen, J. & De Vries, B. (2009) IfcOWL: A case of transforming EXPRESS schemas into ontologies. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*. [Online] 23 (1), 89–101. Available from:

doi:10.1017/S0890060409000122.

Belsky, M., Sacks, R. & Brilakis, I. (2016) Semantic Enrichment for Building Information Modelling. *Computer-Aided Civil and Infrastructure Engineering*. [Online] 31 (4), 261–274. Available from: doi:10.1111/mice.12128.

Ben-Alon, L. & Sacks, R. (2017) Simulating the behavior of trade crews in construction using agents and building information modelling. *Automation in Construction*. [Online] 74, 12–27. Available from: doi:10.1016/j.autcon.2016.11.002.

Van Berlo, L. (2018) *Schependomlaan*. [Online]. 2018. Available from: doi:10.17605/OSF.IO/NE2YU [Accessed: 25 February 2020].

Berners-Lee, T. (2006) *Linked Data - Design Issues*. [Online]. 2006. Available from: https://www.w3.org/DesignIssues/LinkedData.html [Accessed: 14 January 2019].

Berners-Lee, T., Connolly, D., Kagal, L., Scharf, Y., et al. (2008) N3Logic: A logical framework for the World Wide Web. *Theory and Practice of Logic Programming*. [Online] 8 (3), 249–269. Available from: doi:10.1017/S1471068407003213.

Bilal, M., Oyedele, L.O., Qadir, J., Munir, K., et al. (2016) Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*. [Online] 30 (3), 500–521. Available from: doi:10.1016/j.aei.2016.07.001.

Bizer, C., Heath, T. & Berners-Lee, T. (2009) Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*. [Online] 5 (3), 1–22. Available from: doi:10.4018/jswis.2009081901.

Blackmon, T., Saxena, R. & Song, L. (2011) A conceptual framework for total constraint management in construction. *Proceedings of the 28th International Symposium on*

Automation and Robotics in Construction, ISARC 2011. [Online] 419–424. Available from: doi:10.22260/isarc2011/0077.

Bloch, T. & Sacks, R. (2018) Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models. *Automation in Construction*. [Online] 91, 256–272. Available from: doi:10.1016/j.autcon.2018.03.018.

Blum, C. & Sampels, M. (2004) An Ant Colony Optimization Algorithm for Shop Scheduling Problems. *Journal of Mathematical Modelling and Algorithms*. [Online] 3 (3), 285–308. Available from: doi:10.1023/B:JMMA.0000038614.39977.6f.

Boctor, F.F. (1996) Resource-constrained project scheduling by simulated annealing. *International Journal of Production Research*. [Online] 34 (8), 2335–2351. Available from: doi:10.1080/00207549608905028.

Brcic, M., Kalpic, D. & Fertalj, K. (2012) Resource Constrained Project Scheduling under Uncertainty: A Survey. *23rd Central European Conference on Information and Intelligent Systems*. [Online] (September), 401–409. Available from: http://bib.irb.hr/datoteka/590574.2012_Resource_Constrained_Project_Scheduling_under_Uncertainty-_A_Survey.pdf.

Brucker, P., Drexl, A., Möhring, R., Neumann, K., et al. (1999) Resource-constrained project scheduling: Notation, classification, models, and methods. *European Journal of Operational Research*. [Online] 112 (1), 3–41. Available from: doi:10.1016/S0377-2217(98)00204-5.

Bruni, M.E., Beraldi, P. & Guerriero, F. (2015) The stochastic resource-constrained project scheduling problem. In: *Handbook on Project Management and Scheduling Vol. 2*. [Online]. Springer International Publishing. pp. 811–835. Available from: doi:10.1007/978-3-319-05915-0_7.

BSI (2018) ISO19650 - 1 & 2 : Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM). Information management using building information modelling. *British Standards Institution (BSI)*.

BSI (2015) PAS 1192-5-2015: Specification for security-minded building information modelling, digital built environments and smart asset management. *British Standards Institution (BSI)*.

Cai, L. & Zhu, Y. (2015) The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*. [Online] 14 (2), 1–10. Available from: doi:10.5334/dsj-2015-002.

Canbolat, Y.B. & Gundogar, E. (2004) Fuzzy priority rule for job shop scheduling. *Journal of Intelligent Manufacturing*. [Online] 15 (4), 527–533. Available from: doi:10.1023/B:JIMS.0000034116.50789.df.

Cao, D., Li, H. & Wang, G. (2014) Impacts of Isomorphic Pressures on BIM Adoption in Construction Projects. *Journal of Construction Engineering and Management*. [Online] 140 (12), 04014056. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000903.

Carrillo, P., Harding, J. & Choudhary, A. (2011) Knowledge discovery from post-project reviews. *Construction Management and Economics*. [Online] 29 (7), 713–723. Available from: doi:10.1080/01446193.2011.588953.

Chen, B., Potts, C.N. & Woeginger, G.J. (1998) A Review of Machine Scheduling: Complexity, Algorithms and Approximability. *Handbook of Combinatorial Optimization*. [Online] 3, 1493–1641. Available from: doi:10.1007/978-1-4613-0303-9_25.

Chen, R.M. (2011) Particle swarm optimization with justification and designed mechanisms

for resource-constrained project scheduling problem. *Expert Systems with Applications*. [Online] 38 (6), 7102–7111. Available from: doi:10.1016/j.eswa.2010.12.059.

Cheng, T., Yang, J., Teizer, J. & Vela, P.A. (2010) Automated construction resource location tracking to support the analysis of lean principles. In: *Proceedings IGLC-18, Technion, Haifa, Israel*. [Online]. 2010 pp. 643–653. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84866071498&partnerID=tZOtx3y1.

Chiu, H.N. & Tsai, D.M. (2002) An efficient search procedure for the resource-constrained multi-project scheduling problem with discounted cash flows. *Construction Management and Economics*. [Online] 20 (1), 55–66. Available from: doi:10.1080/01446190110089718.

Cho, J.H. & Kim, Y.D. (1997) A simulated annealing algorithm for resource constrained project scheduling problems. *Journal of the Operational Research Society*. [Online] 48 (7), 736–744. Available from: doi:10.1057/palgrave.jors.2600416.

Choi, B., Lee, H.-S., Park, M., Cho, Y.K., et al. (2014) Framework for Work-Space Planning Using Four-Dimensional BIM in Construction Projects. *Journal of Construction Engineering and Management*. [Online] 140 (9), 04014041. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000885.

Choi, J., Choi, J. & Kim, I. (2014) Development of BIM-based evacuation regulation checking system for high-rise and complex buildings. *Automation in Construction*. [Online] 46, 38–49. Available from: doi:10.1016/j.autcon.2013.12.005.

Christofides, N., Alvarez-Valdes, R. & Tamarit, J.M. (1987) Project scheduling with resource constraints: A branch and bound approach. *European Journal of Operational Research*. [Online] 29 (3), 262–273. Available from: doi:10.1016/0377-2217(87)90240-2.

Chua, D.K.H.H. & Yeoh, K.W.W. (2009) A framework for construction requirements based planning utilizing constraints logic programming. *Proceedings of IGLC17: 17th Annual Conference of the International Group for Lean Construction*. (65), 327–336.

Clayton, M., Kunz, J. & Fischer, M. (1998) *The Charrette Test Method*. (July).

Corman, J., Florenzano, F., Reutter, J.L. & Savković, O. (2019) *Validating Shacl Constraints over a Sparql Endpoint*. In: [Online]. pp. 145–163. Available from: doi:10.1007/978-3-030-30793-6_9.

Darwiche, A., Levitt, R.E. & Hayes-Roth, B. (1988) OARPLAN: Generating project plans by reasoning about objects, actions and resources. *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing*. [Online] 2 (3), 169–181. Available from: doi:10.1017/S0890060400000639.

Das, P.P. & Acharyya, S. (2011) Simulated annealing variants for solving resource constrained project scheduling problem: A comparative study. In: *14th International Conference on Computer and Information Technology, ICCIT 2011*. [Online]. 2011 pp. 469–474. Available from: doi:10.1109/ICCITechn.2011.6164835.

Dave, B., Boddy, S., Koskela, L., Dave, et al. (2010) Improving information flow within the production management system with web services. *18th Annual Conference of the International Group for Lean Construction, IGLC 18*. (January), 445–455.

Dave, B., Hämäläinen, J.P., Kemmer, S., Koskela, L., et al. (2015) Suggestions to improve lean construction planning. In: *23rd Ann. Conf. of the Int'l. Group for Lean Construction*. [Online]. 2015 Perth, Australia. pp. 193–202. Available from: www.iglc.net [Accessed: 26 May 2020].

Dave, B., Kubler, S., Främling, K. & Koskela, L. (2016a) Opportunities for enhanced lean

construction management using Internet of Things standards. *Automation in Construction*. [Online] 61, 86–97. Available from: doi:10.1016/j.autcon.2015.10.009.

Dave, Seppänen, B.;, Modrich, O. & Ralf-Uwe (2016b) Modelling Information Flows Between Last Planner and Location Based Management System. In: *Proceedings of the 24th Annual Confenrence of the International Group for Lean Construction*. [Online]. 2016 p. Available from: https://research.aalto.fi/files/7207101/Dave_et_al._2016_Modelling_Information_Flows_Between_Last_Planner_and_Location_Based_Management_System.pdf [Accessed: 26 March 2018].

Davies, J., Fensel, D. & Van Harmelen, F. (2003) *Towards the semantic web : ontology-driven knowledge management*. J. Wiley.

Debels, D., De Reyck, B., Leus, R. & Vanhoucke, M. (2006) A hybrid scatter search/electromagnetism meta-heuristic for project scheduling. *European Journal of Operational Research*. [Online] 169 (2), 638–653. Available from: doi:10.1016/j.ejor.2004.08.020.

Dechter, R. (2003) *Constraint Processing*. [Online]. Sanfrancisco, US, Elsevier. Available from: doi:10.1016/B978-1-55860-890-0.X5000-2.

Delone, W. & McLean, E. (2014) The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*. [Online] 19 (4), 9–30. Available from: doi:10.1080/07421222.2003.11045748.

Dimyadi, J., Clifton, C., Spearpoint, M. & Amor, R. (2016a) Computerizing Regulatory Knowledge for Building Engineering Design. *Journal of Computing in Civil Engineering*. [Online] 30 (5). Available from: doi:10.1061/(ASCE)CP.1943-5487.0000572 [Accessed: 9 April 2020].

Dimyadi, J., Solihin, W., Eastman, C. & Amor, R. (2016b) *Integrating the BIM Rule Language into Compliant Design Audit Processes*.

Donato, V., Lo Turco, M. & Bocconcino, M.M. (2017) BIM-QA/QC in the architectural design process. *Architectural Engineering and Design Management*. [Online] 14 (3), 239–254. Available from: doi:10.1080/17452007.2017.1370995.

Dong, N. (2012) *Automated Look-Ahead Schedule Generation and Optimization for the Finishing Phase of Complex Construction Projects*. Stanford University.

Dong, N., Fischer, M., Haddad, Z. & Levitt, R. (2013a) A method to automate look-ahead schedule (LAS) generation for the finishing phase of construction projects. *Automation in Construction*. [Online] 35, 157–173. Available from: doi:10.1016/j.autcon.2013.05.023.

Dong, S., Behzadan, A.H., Chen, F. & Kamat, V.R. (2013b) Collaborative visualization of engineering processes using tabletop augmented reality. *Advances in Engineering Software*. [Online] 55, 45–55. Available from: doi:10.1016/j.advengsoft.2012.09.001 [Accessed: 16 January 2017].

Dorigo, M. & Caro, G. Di (1999) *The ant colony optimization meta-heuristic*.pp.11–32.

Dorigo, M. & Colorni, A. (1996) The Ant System: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*. 26 (1), 1–13.

Dossick, C.S. & Neff, G. (2010) Organizational Divisions in BIM-Enabled Commercial Construction. *Journal of Construction Engineering and Management*. [Online] 136 (4), 459–467. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000109.

Eastman, C., Lee, J., Jeong, Y. & Lee, J. (2009) Automatic rule-based checking of building designs. *Automation in Construction*. [Online] 18 (8), 1011–1033. Available from: doi:10.1016/j.autcon.2009.07.002.

Ei-Diraby, T.A., Lima, C. & Feis, B. (2005) Domain taxonomy for construction concepts: Toward a formal ontology for construction knowledge. *Journal of Computing in Civil Engineering*. [Online] 19 (4), 394–406. Available from: doi:10.1061/(ASCE)0887-3801(2005)19:4(394).

Eisenhardt, K.M. & Graebner, M.E. (2007) Theory building from cases: Opportunities and challenges. *Academy of Management Journal*. [Online] 50 (1), 25–32. Available from: doi:10.5465/AMJ.2007.24160888 [Accessed: 21 March 2018].

Ekaputra, F.J. & Xiashuo Lin (2016) SHACL4P: SHACL constraints validation within Protégé ontology editor. In: *2016 International Conference on Data and Software Engineering (ICoDSE)*. [Online]. October 2016 IEEE. pp. 1–6. Available from: doi:10.1109/ICODSE.2016.7936162.

El-Rayes, K. & Jun, D.H. (2009) Optimizing resource leveling in construction rojects. *Journal of Construction Engineering and Management*. [Online] 135 (11), 1172–1180. Available from: doi:(ASCE)CO.1943-7862.0000097.

Elbeltagi, E., Hegazy, T. & Grierson, D. (2005) Comparison among five evolutionary-based optimization algorithms. *Advanced Engineering Informatics*. [Online] 19 (1), 43–53. Available from: doi:10.1016/j.aei.2005.01.004.

Faghihi, V., Nejat, A., Reinschmidt, K.F. & Kang, J.H. (2015) Automation in construction scheduling: a review of the literature. *International Journal of Advanced Manufacturing Technology*. [Online] 81 (9–12), 1845–1856. Available from: doi:10.1007/s00170-015-7339-0.

Fahmy, A., Hassan, T.M. & Bassioni, H. (2014) Improving RCPSP solutions quality with Stacking Justification - Application with particle swarm optimization. *Expert Systems with Applications*. [Online] 41 (13), 5870–5881. Available from:

doi:10.1016/j.eswa.2014.03.027.

Fan, W. (2015) Data Quality: From theory to practice. *ACM SIGMOD Record*. [Online] 44 (3), 7–18. Available from: doi:10.1145/2854006.2854008.

Farias, T.M. de, Roxin, A. & Nicolle, C. (2018) A rule-based methodology to extract building model views. *Automation in Construction*. [Online] 92, 214–229. Available from: doi:10.1016/j.autcon.2018.03.035.

Fish, A.N. & Taylor, J. (2012) *Knowledge automation : how to implement decision management in business processes*. [Online]. New Jersey, Wiley. Available from: http://repositorio.unan.edu.ni/2986/1/5624.pdf.

Flask (2010) *Welcome to Flask — Flask Documentation (1.1.x)*. [Online]. 2010. Available from: https://flask.palletsprojects.com/en/1.1.x/ [Accessed: 16 June 2020].

Fonseca-Reyna, Y.C., Martínez-Jiménez, Y., Cabrera, A.V. & Rodríguez Sánchez, E.A. (2019) Optimization of heavily constrained hybrid-flexible flowshop problems using a multi-agent reinforcement learning approach. *Investigacion Operacional*. 40 (1), 100–111.

Friston, S., Fan, C., Doboš, J., Scully, T., et al. (2017) 3DRepo4Unity: Dynamic Loading of Version Controlled 3D Assets into the Unity Game Engine. In: *Proceedings of the 22nd International Conference on 3D Web Technology - Web3D '17*. [Online]. 2017 New York, New York, USA, ACM Press. pp. 1–9. Available from: doi:10.1145/3055624.3075941.

Giretti, A., Carbonari, A., Novembri, G. & Robuffo, F. (2012) Estimation of job-site work progress through on-site monitoring. In: *2012 Proceedings of the 29th International Symposium of Automation and Robotics in Construction, ISARC 2012*. [Online]. 14 June 2012 p. Available from: doi:10.4017/gt.2012.11.02.348.704.

Glover, F. (1986) Future paths for integer programming and links to artificial intelligence.

Computers and Operations Research. [Online] 13 (5), 533–549. Available from: doi:10.1016/0305-0548(86)90048-1.

Glover, F. & Laguna, M. (1998) Tabu Search. In: *Handbook of Combinatorial Optimization*. [Online]. Springer US. pp. 2093–2229. Available from: doi:10.1007/978-1-4613-0303-9_33.

Goldberg, D.E. & Holland, J.H. (1988) Genetic Algorithms and Machine Learning. *Machine Learning*. [Online]. 3 (2) pp.95–99. Available from: doi:10.1023/A:1022602019183.

Goldratt, E.M. & Cox, J. (1984) *The Goal: A Process of Ongoing Improvement*. Third Edit. Great Barrington MA, North River Press.

Gómez-Romero, J., Bobillo, F., Ros, M., Molina-Solana, M., et al. (2015) A fuzzy extension of the semantic Building Information Model. *Automation in Construction*. [Online] 57, 202–212. Available from: doi:10.1016/j.autcon.2015.04.007 [Accessed: 26 May 2020].

Gonçalves, J.F., De Mendes, J.J.M. & Resende, M.G.C. (2015) The basic multi-project scheduling problem. In: *Handbook on Project Management and Scheduling Vol. 2*. [Online]. Springer International Publishing. pp. 667–683. Available from: doi:10.1007/978-3-319-05915-0_1.

Google (2019) *AngularJS — Superheroic JavaScript MVW Framework*. [Online]. 2019. Available from: https://angularjs.org/ [Accessed: 16 June 2020].

Graves, S.C. (1981) A Review of Production Scheduling. *Operations Research*. [Online] 29 (4), 646–675. Available from: doi:10.1287/opre.29.4.646.

Güçdemir, H. & Selim, H. (2017) Customer centric production planning and control in job shops: A simulation optimization approach. *Journal of Manufacturing Systems*. [Online] 43, 100–116. Available from: doi:10.1016/j.jmsy.2017.02.004.

Hamzeh, F.R. (2009) *Improving construction workflow- The role of production planning and control*. [Online]. University of California Berkley. Available from: doi:10.13140/RG.2.1.1182.9286 [Accessed: 27 May 2020].

Hamzeh, F.R., Ballard, G. & Tommelein, I.D. (2008) Improving construction work flow - The connective role of lookahead planning. *Proceedings of IGLC16: 16th Annual Conference of the International Group for Lean Construction*. [Online] (January), 635–646. Available from: doi:10.13140/RG.2.1.3804.3685.

Hamzeh, F.R., Zankoul, E. & Rouhana, C. (2015) How can 'tasks made ready' during lookahead planning impact reliable workflow and project duration? *Construction Management and Economics*. [Online] 33 (4), 243–258. Available from: doi:10.1080/01446193.2015.1047878 [Accessed: 22 May 2020].

Han, K.K., Cline, D. & Golparvar-Fard, M. (2015) Formalized knowledge of construction sequencing for visual monitoring of work-in-progress via incomplete point clouds and low-LoD 4D BIMs. *Advanced Engineering Informatics*. [Online] 29 (4), 889–901. Available from: doi:10.1016/j.aei.2015.10.006.

Han, K.K. & Golparvar-Fard, M. (2017) Potential of big visual data and building information modelling for construction performance analytics: An exploratory study. *Automation in Construction*. [Online] 73, 184–198. Available from: doi:10.1016/j.autcon.2016.11.004.

Hartmann, S. (1998) A competitive genetic algorithm for resource-constrained project scheduling. *Naval Research Logistics*. [Online] 45 (7), 733–750. Available from: doi:10.1002/(SICI)1520-6750(199810)45:7<733::AID-NAV5>3.0.CO;2-C [Accessed: 18 February 2020].

Hartmann, S. (2015) Time-varying resource requirements and capacities. In: *Handbook on Project Management and Scheduling Vol. 1*. [Online]. Springer International Publishing.

pp. 163–176. Available from: doi:10.1007/978-3-319-05443-8_8.

Hartmann, S. & Briskorn, D. (2010) A survey of variants and extensions of the resource-constrained project scheduling problem. *European Journal of Operational Research*. [Online] 207 (1), 1–14. Available from: doi:10.1016/j.ejor.2009.11.005.

Hartmann, T. (2008) *A grassroots model of decision support system implications by construction project teams*. [Online]. California, USA, Stanford University. Available from: https://stacks.stanford.edu/file/druid:gv950qy3147/WP108.pdf.

Hartmann, T. & Fischer, M. (2007) Supporting the constructability review with 3D/4D models. *Building Research & Information*. [Online] 35 (1), 70–80. Available from: doi:10.1080/09613210600942218.

Harty, C. & Whyte, J. (2010) Emerging Hybrid Practices in Construction Design Work: Role of Mixed Media. *Journal of Construction Engineering and Management*. [Online] 136 (4), 468–476. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000146.

Hendler, J. & Pardo, T.A. (2012) *A Primer on Machine Readability for Online Documents and Data*. [Online]. 2012. Available from: https://www.data.gov/developers/blog/primer-machine-readability-online-documents-and-data [Accessed: 10 October 2018].

Hevner, A.R., March, S.T., Park, J. & Ram, S. (2004) Design Science In IS Research. *Management Information Systems*. [Online] 28 (1), 75–105. Available from: doi:10.2307/25148625.

Hinkka, V. & Tatila, J. (2013) RFID tracking implementation model for the technical trade and construction supply chains. *Automation in Construction*. [Online] 35, 405–414. Available from: doi:10.1016/j.autcon.2013.05.024.

Hjelseth, E. & Nisbet, N. (2010) Exploring semantic based model checking. In: *Proceedings*

of CIB W78 Conference. [Online]. 2010 pp. 16–18. Available from: https://itc.scix.net/paper/w78-2010-54.

Holland, J.H. (John H. (1992) *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. MIT Press.

Hopp, W.J. & Spearman, M.L. (2001) *Factory physics : foundations of manufacturing management*. Irwin/McGraw-Hill.

Hossen, J. & Sayeed, S. (2018) *A Survey on Cleaning Dirty Data Using Machine Learning Paradigm for Big Data Analytics*. [Online] (April 2019), 1234–1243. Available from: doi:10.11591/ijeecs.v10.i3.pp1234-1243.

Hu, J. & Flood, I. (2012) A Multi-Objective Scheduling Model for Solving the Resource-Constrained Project Scheduling and Resource Leveling Problems. In: *Computing in Civil Engineering (2012)*. [Online]. 11 June 2012 Reston, VA, American Society of Civil Engineers. pp. 49–56. Available from: doi:10.1061/9780784412343.0007.

Hu, Z.-Z., Zhang, X.-Y., Wang, H.-W. & Kassem, M. (2016) Improving interoperability between architectural and structural design models: An industry foundation classes-based approach with web-based tools. *Automation in Construction*. [Online] 66, 29–42. Available from: doi:10.1016/j.autcon.2016.02.001.

Huang, R.-H. & Yu, T.-H. (2017) An effective ant colony optimization algorithm for multi-objective job-shop scheduling with equal-size lot-splitting. *Applied Soft Computing*. [Online] 57, 642–656. Available from: doi:10.1016/j.asoc.2017.04.062.

Ibrahim, M. & Moselhi, O. (2014) Experimental study of wireless sensor networks for indoor construction operations. In: *31st International Symposium on Automation and Robotics in Construction and Mining, ISARC*. [Online]. 2014 pp. 805–812. Available from:

https://www.iaarc.org/publications/fulltext/isarc2014_submission_12.pdf.

Isaac, S. & Navon, R. (2012) Combining automatically and manually collected data for project monitoring and control. In: *2012 Proceedings of the 29th International Symposium of Automation and Robotics in Construction, ISARC 2012*. [Online]. 2012 p. 325. Available from: doi:10.4017/gt.2012.11.02.202.676.

ISO (2016) *ISO - ISO/IEC 9075-1:2016 - Information technology — Database languages — SQL — Part 1: Framework (SQL/Framework)*. [Online]. 2016. Available from: https://www.iso.org/standard/63555.html [Accessed: 9 April 2020].

Jain, A.S. & Meeran, S. (1998) *A State-of-the-Art Review of Job-Shop Scheduling Techniques*. [Online]. (1). Available from: ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia707_1s04/textos/jain98stateart.pdf.

Janssen, M., van der Voort, H. & Wahyudi, A. (2017) Factors influencing big data decision-making quality. *Journal of Business Research*. [Online] 70 (1), 338–345. Available from: doi:10.1016/j.jbusres.2016.08.007.

Jaradat, S., Whyte, J. & Luck, R. (2013) Professionalism in digitally mediated project work. *Building Research & Information*. [Online] 41 (1), 51–59. Available from: doi:10.1080/09613218.2013.743398.

Jayawardene, V., Sadiq, S. & Indulska, M. (2015) An Analysis of Data Quality Dimensions . *ITEE Technical Report*.

Jedrzejowicz, P. & Ratajczak-Ropel, E. (2014) Reinforcement Learning strategies for A-Team solving the Resource-Constrained Project Scheduling Problem. *Neurocomputing*. [Online] 146, 301–307. Available from: doi:10.1016/j.neucom.2014.05.070.

Jędrzejowicz, P. & Ratajczak-Ropel, E. (2014) Reinforcement learning strategy for solving the

resource-constrained project scheduling problem by a team of A-teams. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online] 8398 LNAI (PART 2), 197–206. Available from: doi:10.1007/978-3-319-05458-2_21.

Jia, Q. & Seo, Y. (2013) An improved particle swarm optimization for the resource-constrained project scheduling problem. *International Journal of Advanced Manufacturing Technology*. [Online] 67 (9–12), 2627–2638. Available from: doi:10.1007/s00170-012-4679-x.

Johnson, S.M. (1954) Optimal two- and three-stage production schedules with setup times included. *Naval Research Logistics Quarterly*. [Online] 1 (1), 61–68. Available from: doi:10.1002/nav.3800010110.

Ju, Q.Y. (2007) *Planning and scheduling optimization of job-shop in intelligent manufacturing system*. Nanjing University of Aeronautics and Astronautics, Nanjing, China.

Kaelbling, L., Littman, M. & Moore, A. (1996) Reinforcemen t Learning? A Surv ey. *Journal of Artificial Intelligence Research*. 4 (1), 237–285.

Karam, A. & Lazarova-Molnar, S. (2013) Recent trends in solving the deterministic resource constrained Project Scheduling Problem. *2013 9th International Conference on Innovations in Information Technology, IIT 2013*. [Online] 124–129. Available from: doi:10.1109/Innovations.2013.6544405.

Kennedy, J. & Eberhart, R. (1995) Particle swarm optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. [Online]. 1995 IEEE. pp. 1942–1948. Available from: doi:10.1109/ICNN.1995.488968 [Accessed: 4 June 2020].

Khadwilard, A., Chansombat, S., Thepphakorn, T., Thapatsuwan, P., et al. (2011) Application

of firefly algorithm and its parameter setting for job shop scheduling. In: *First Symposium on Hands-On Research and Development*. 2011 p.

Kim, C., Kim, B. & Kim, H. (2013) 4D CAD model updating using image processing-based construction progress monitoring. *Automation in Construction*. [Online] 35, 44–52. Available from: doi:10.1016/j.autcon.2013.03.005.

Kim, K., Kim, H., Kim, W., Kim, C., et al. (2018) Integration of ifc objects and facility management work information using Semantic Web. *Automation in Construction*. [Online] 87 (November), 173–187. Available from: doi:10.1016/j.autcon.2017.12.019 [Accessed: 25 February 2018].

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983) Optimization by simulated annealing. *Science*. [Online] 220 (4598), 671–680. Available from: doi:10.1126/science.220.4598.671.

Knublauch, H., Allemang, D. & Steyskal, S. (2017) SHACL Advanced Features. *W3C Working Group* . [Online]. Available from: https://www.w3.org/TR/shacl-af/ [Accessed: 25 April 2019].

Kolisch, R. (1996a) Efficient priority rules for the resource-constrained project scheduling problem. *Journal of Operations Management*. [Online] 14 (3), 179–192. Available from: doi:10.1016/0272-6963(95)00032-1.

Kolisch, R. (1996b) Serial and parallel resource-constrained project scheduling methods revisited: Theory and computation. *European Journal of Operational Research*. [Online] 90 (2), 320–333. Available from: doi:10.1016/0377-2217(95)00357-6 [Accessed: 17 February 2020].

Kolisch, R. & Hartmann, S. (2006) Experimental investigation of heuristics for resource-

constrained project scheduling: An update. *European Journal of Operational Research*. [Online] 174 (1), 23–37. Available from: doi:10.1016/j.ejor.2005.01.065.

Kolisch, R. & Hartmann, S. (1999) *Heuristic Algorithms for the Resource-Constrained Project Scheduling Problem: Classification and Computational Analysis*. In: [Online]. pp. 147–178. Available from: doi:10.1007/978-1-4615-5533-9_7.

Kopsida, M., Brilakis, I. & Vela, P. (2015) A Review of Automated Construction Progress and Inspection Methods. In: *Proceedings of the 32nd CIB W78 Conference on Construction IT*. 2015 pp. 421–431.

Koskela, L. (2000) *An Exploration towards a Production Theory and its Application to Construction.* [Online]. Available from: doi:10.1073/pnas.1107281109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1107281109.

Koskela, L. (1992) Application of the new production philosophy to construciton. *Stanford University*.

Koskela, L. (1999) Management of production in construction: A theoretical view Title Management of production in construction: A theoretical view MANAGEMENT OF PRODUCTION IN CONSTRUCTION: A THEORETICAL VIEW. In: *Proceedings of Seventh Annual Conference of The International Group for Lean Construction*. [Online]. 1999 p. Available from: http://usir.salford.ac.uk/9429/ [Accessed: 22 April 2018].

Koulinas, G., Xanthopoulos, A., Kiatipis, A. & Koulouriotis, D. (2018) A summary of using reinforcement learning strategies for treating project and production management problems. *2018 13th International Conference on Digital Information Management, ICDIM 2018*. [Online] 33–38. Available from: doi:10.1109/ICDIM.2018.8847099.

Krijnen, T. & Beetz, J. (2017) An IFC schema extension and binary serialization format to

efficiently integrate point cloud data into building models. *Advanced Engineering Informatics*. [Online] 33, 473–490. Available from: doi:10.1016/j.aei.2017.03.008.

van Laarhoven, P.J.M. & Aarts, E.H.L. (1987) *Simulated Annealing: Theory and Applications*. [Online]. Springer Netherlands. Available from: doi:10.1007/978-94-015-7744-1.

Laine, E., Alhava, O. & Kiviniemi, A. (2014) Improving Built-In Quality by BIM Based Visual Management. In: *22nd Annual Conference of the International Group for Lean Construction: Understanding and Improving Project Based Production*. [Online]. 2014 p. Available from: https://iglcstorage.blob.core.windows.net/papers/attachment-6d0c60ea-3fc0-4967-a77b-6eaf1fc9ee2c.pdf [Accessed: 25 April 2018].

Lee, D.-Y.Y., Chi, H. lin, Wang, J., Wang, X., et al. (2016) A linked data system framework for sharing construction defect information using ontologies and BIM environments. *Automation in Construction*. [Online] 68, 102–113. Available from: doi:10.1016/j.autcon.2016.05.003 [Accessed: 26 February 2018].

Lee, J.K., Eastman, C. & Lee, Y.C. (2015) Implementation of a BIM Domain-specific Language for the Building Environment Rule and Analysis. *Journal of Intelligent and Robotic Systems: Theory and Applications*. [Online] 79 (3–4), 507–522. Available from: doi:10.1007/s10846-014-0117-7.

Lee, Y.-C., Solihin, W. & Eastman, C.M. (2019) The Mechanism and Challenges of Validating a Building Information Model regarding data exchange standards. *Automation in Construction*. [Online] 100 (December 2018), 118–128. Available from: doi:https://doi.org/10.1016/j.autcon.2018.12.025.

Li, K.Y. & Willis, R.J. (1992) An iterative scheduling technique for resource-constrained project scheduling. *European Journal of Operational Research*. [Online] 56 (3), 370–379. Available from: doi:10.1016/0377-2217(92)90320-9.

Li, X. & Gao, L. (2016) An effective hybrid genetic algorithm and tabu search for flexible job shop scheduling problem. *International Journal of Production Economics*. [Online] 174, 93–110. Available from: doi:10.1016/j.ijpe.2016.01.016.

Lian, L. & Mesghouni, K. (2014) Comparative study of heuristics algorithms in solving flexible job shop scheduling problem with condition based maintenance. *Journal of Industrial Engineering and Management*. [Online] 7 (2). Available from: doi:10.3926/jiem.1038.

Lin, J., Zhu, L. & Gao, K. (2020) A genetic programming hyper-heuristic approach for the multi-skill resource constrained project scheduling problem. *Expert Systems with Applications*. [Online] 140, 112915. Available from: doi:10.1016/j.eswa.2019.112915 [Accessed: 17 February 2020].

Lin, L., Hao, X.-C., Gen, M. & Jo, J.-B. (2012) Network modelling and evolutionary optimization for scheduling in manufacturing. *Journal of Intelligent Manufacturing*. [Online] 23 (6), 2237–2253. Available from: doi:10.1007/s10845-011-0569-6.

Lindhard, S. & Wandahl, S. (2014) Scheduling of large, complex, and constrained construction projects - an exploration of LPS application. *International Journal of Project Organisation and Management*. [Online] 6 (3), 237. Available from: doi:10.1504/IJPOM.2014.065258 [Accessed: 22 May 2020].

Liu, H., Al-Hussein, M. & Lu, M. (2015) BIM-based integrated approach for detailed construction scheduling under resource constraints. *Automation in Construction*. [Online] 53, 29–43. Available from: doi:10.1016/j.autcon.2015.03.008.

Liu, H., Lu, M. & Al-Hussein, M. (2016) Ontology-based semantic approach for construction-oriented quantity take-off from BIM models in the light-frame building industry. *Advanced Engineering Informatics*. [Online] 30 (2), 190–207. Available from: doi:10.1016/j.aei.2016.03.001.

Liu, J., Liu, Y., Shi, Y. & Li, J. (2020) Solving Resource-Constrained Project Scheduling Problem via Genetic Algorithm. *Journal of Computing in Civil Engineering*. [Online] 34 (2), 04019055. Available from: doi:10.1061/(ASCE)CP.1943-5487.0000874.

Liu, J. & Lu, M. (2018) Constraint Programming Approach to Optimizing Project Schedules under Material Logistics and Crew Availability Constraints. *Journal of Construction Engineering and Management*. [Online] 144 (7). Available from: doi:10.1061/(ASCE)CO.1943-7862.0001507.

Liu, S.Q. & Kozan, E. (2012) A hybrid shifting bottleneck procedure algorithm for the parallel-machine job-shop scheduling problem. *Journal of the Operational Research Society*. [Online] 63 (2), 168–182. Available from: doi:10.1057/jors.2011.4.

Lomnicki, Z.A. (1965) A "Branch-and-Bound" Algorithm for the Exact Solution of the Three-Machine Scheduling Problem. *Journal of the Operational Research Society*. [Online] 16 (1), 89–100. Available from: doi:10.1057/jors.1965.7.

Loshin (2011) *The Practitioner's Guide to Data Quality Improvement*. [Online]. Available from: doi:10.1016/b978-0-12-373717-5.00001-4.

Łukasik, S. & Żak, S. (2009) *Firefly Algorithm for Continuous Constrained Optimization Tasks*. In: [Online]. pp. 97–106. Available from: doi:10.1007/978-3-642-04441-0_8.

Mahdi Mobini, M.D., Rabbani, M., Amalnik, M.S., Razmi, J., et al. (2009) Using an enhanced scatter search algorithm for a resource-constrained project scheduling problem. *Soft Computing*. [Online] 13 (6), 597–610. Available from: doi:10.1007/s00500-008-0337-5.

Majrouhi Sardroud, J. (2012) Influence of RFID technology on automated management of construction materials and components. *Scientia Iranica*. [Online] 19 (3), 381–392. Available from: doi:10.1016/j.scient.2012.02.023.

Marshall, G., Barber, T.J. & Boardman, J.T. (1987) METHODOLOGY FOR MODELLING A PROJECT MANAGEMENT CONTROL ENVIRONMENT. *IEE Proceedings D: Control Theory and Applications*. [Online] 134 (4), 287–300. Available from: doi:10.1049/ip-d.1987.0047.

Martins, J.P. & Monteiro, A. (2013) LicA: A BIM based automated code-checking application for water distribution systems. *Automation in Construction*. [Online] 29 (23), 12–23. Available from: doi:10.1016/j.autcon.2012.08.008.

Matthews, J., Love, P.E.D.D., Heinemann, S., Chandler, R., et al. (2015) Real time progress management: Re-engineering processes for cloud-based BIM in construction. *Automation in Construction*. [Online] 58, 38–47. Available from: doi:10.1016/j.autcon.2015.07.004.

Mazairac, W. & Beetz, J. (2013) BIMQL - An open query language for building information models. *Advanced Engineering Informatics*. [Online] 27 (4), 444–456. Available from: doi:10.1016/j.aei.2013.06.001.

Merkle, D., Middendorf, M. & Schmeck, H. (2002) Ant colony optimization for resource-constrained project scheduling. *IEEE Transactions on Evolutionary Computation*. [Online] 6 (4), 333–346. Available from: doi:10.1109/TEVC.2002.802450.

Minella, G., Ruiz, R. & Ciavotta, M. (2008) A Review and Evaluation of Multiobjective Algorithms for the Flowshop Scheduling Problem. *INFORMS Journal on Computing*. [Online] 20 (3), 451–471. Available from: doi:10.1287/ijoc.1070.0258.

Mirarchi, C. & Pavan, A. (2019) Building information models are dirty. *2019 European Conference of Computing in Construction (2019 EC^3)*.

Moon, H., Kim, H., Kamat, V.R. & Kang, L. (2015) BIM-based construction scheduling method using optimization theory for reducing activity overlaps. *Journal of Computing in*

Civil Engineering. [Online] 29 (3). Available from: doi:10.1061/(ASCE)CP.1943-5487.0000342.

Morkos, R. (2014) *Operational efficiency frontier: Visualizing, manipulating, and navigating the construction scheduling state space with precedence, discrete, and disjunctive constraints*. [Online]. Stanford University. Available from: https://searchworks.stanford.edu/view/10701353?fbclid=IwAR3kNQkRysjSySxKfnGEDajJjxl9a2k2T9h5OO5Lkv-P1ktn3HbZgyDLh9s.

Mossman, A. (2013) Last Planner®: 5 + 1 crucial & collaborative conversations for predictable design & construction delivery. *Lean Construciton Institute*. [Online]. Available from: http://bit.ly/LPS-5cc.

Munlin, M. (2018) Solving resource-constrained project scheduling problem using metaheuristic algorithm. In: *2018 5th International Conference on Electrical and Electronics Engineering, ICEEE 2018*. [Online]. 20 June 2018 Institute of Electrical and Electronics Engineers Inc. pp. 344–349. Available from: doi:10.1109/ICEEE2.2018.8391359.

Musen, M.A. (2015) The protégé project: A Look Back and a Look Forward. *AI Matters*. [Online] 1 (4), 4–12. Available from: doi:10.1145/2757001.2757003.

Myszkowski, P.B., Skowroński, M.E., Olech, Ł.P. & Oślizło, K. (2015) Hybrid ant colony optimization in solving multi-skill resource-constrained project scheduling problem. *Soft Computing*. [Online] 19 (12), 3599–3619. Available from: doi:10.1007/s00500-014-1455-x.

Nasiri, M.M. (2012) A pseudo particle swarm optimization for the RCPSP. *International Journal of Advanced Manufacturing Technology*. [Online] 65 (5–8), 909–918. Available from: doi:10.1007/s00170-012-4227-8.

Naumann, F. & Rolker, C. (2000) Assessment Methods for Information Quality Criteria. *Fifth Conference on Information Quality (IQ 2000)*. [Online]. Available from: doi:http://dx.doi.org/10.18452/2441.

Navon, R. (2007) Research in automated measurement of project performance indicators. *Automation in Construction*. [Online] 16 (2), 176–188. Available from: doi:10.1016/j.autcon.2006.03.003.

Nawari, N.O. (2012) Automating codes conformance. *Journal of Architectural Engineering*. [Online] 18 (4), 315–323. Available from: doi:10.1061/(ASCE)AE.1943-5568.0000049.

Nawaz, M., Enscore, E.E. & Ham, I. (1983) A heuristic algorithm for the m-machine, n-job flow-shop sequencing problem. *Omega*. [Online] 11 (1), 91–95. Available from: doi:10.1016/0305-0483(83)90088-9.

Niederliński, A. (2011) *A Quick and Gentle Guide to Constraint Logic Programming Via ECLiPSe*. Gliwice, Poland.

Nieto-Morote, A. & Ruz-Vila, F. (2012) Last Planner Control System Applied to a Chemical Plant Construction. *Journal of Construction Engineering and Management*. [Online] 138 (2), 287–293. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000415 [Accessed: 22 May 2020].

Nonobe, K. & Ibaraki, T. (2002) *Formulation and Tabu Search Algorithm for the Resource Constrained Project Scheduling Problem*. In: [Online]. pp. 557–588. Available from: doi:10.1007/978-1-4615-1507-4_25.

OASIS LegalRuleML, T. (2018) *LegalRuleML Core Specification Version 1.0*. [Online]. 2018. Available from: http://docs.oasis-open.org/legalruleml/legalruleml-core-spec/v1.0/legalruleml-core-spec-v1.0.html [Accessed: 17 November 2019].

Omar, T. & Nehdi, M.L. (2016) Data acquisition technologies for construction progress tracking. *Automation in Construction*. [Online] 70, 143–155. Available from: doi:10.1016/j.autcon.2016.06.016.

Ortiz-Pimiento, N.R. & Diaz-Serna, F.J. (2018) The project scheduling problem with non-deterministic activities duration: A literature review. *Journal of Industrial Engineering and Management*. [Online] 11 (1), 116–134. Available from: doi:10.3926/jiem.2492.

Özdamar, L. & Ulusoy, G. (1996) A note on an iterative forward/backward scheduling technique with reference to a procedure by Li and Willis. *European Journal of Operational Research*. [Online] 89 (2), 400–407. Available from: doi:10.1016/0377-2217(94)00272-X.

Patterson, J.H. (1976) Project scheduling: The effects of problem structure on heuristic performance. *Naval Research Logistics Quarterly*. [Online] 23 (1), 95–123. Available from: doi:10.1002/nav.3800230110.

Pauwels, P. (2014) *Supporting Decision-Making in the Building Life-Cycle Using Linked Building Data*. [Online]. Available from: doi:10.3390/buildings4030549.

Pauwels, P., Van Deursen, D., Verstraeten, R., De Roo, J., et al. (2011) A semantic rule checking environment for building performance checking. *Automation in Construction*. [Online] 20 (5), 506–518. Available from: doi:10.1016/j.autcon.2010.11.017.

Pauwels, P., Krijnen, T., Terkaj, W. & Beetz, J. (2017) Enhancing the ifcOWL ontology with an alternative representation for geometric data. *Automation in Construction*. [Online] 80, 77–94. Available from: doi:10.1016/j.autcon.2017.03.001 [Accessed: 25 February 2018].

Pauwels, P., De Meyer, R. & Van Campenhout, J. (2011) Interoperability for the Design and Construction Industry through Semantic Web Technology. In: *Samt*. [Online]. pp. 143–

158. Available from: doi:10.1007/978-3-642-23017-2_10.

Pauwels, P. & Roxin, A. (2016) SimpleBIM : From full ifcOWL graphs to simplified building graphs. *11th European Conference on Product and Process Modelling*. [Online] (October), 11–18. Available from: https://www.semanticscholar.org/paper/SimpleBIM-%3A-from-full-ifcOWL-graphs-to-simplified-Pieter-Roxin/7ae58717e7b5786de66437823b6140fe2e1928f4.

Pauwels, P. & Terkaj, W. (2016) EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction*. [Online] 63, 100–133. Available from: doi:10.1016/j.autcon.2015.12.003 [Accessed: 25 January 2020].

Pauwels, P., Törmä, S., Beetz, J., Weise, M., et al. (2015) Linked Data in Architecture and Construction. *Automation in Construction*. [Online] 57, 175–177. Available from: doi:10.1016/j.autcon.2015.06.007.

Pauwels, P., Zhang, S. & Lee, Y.-C. (2017) Semantic web technologies in AEC industry: A literature overview. *Automation in Construction*. [Online] 73, 145–165. Available from: doi:10.1016/J.AUTCON.2016.10.003 [Accessed: 26 February 2018].

Pazlar, T. & Turk, Ž. (2008) Interoperability in practice: Geometric data exchange using the IFC standard. *Journal of Information Technology in Construction*. [Online] 13 (June), 362–380. Available from: http://www.itcon.org/2008/24.

Pedro, A., Lee, D.Y., Hussain, R. & Park, C.S. (2017) Linked data system for sharing construction safety information. *ISARC 2017 - Proceedings of the 34th International Symposium on Automation and Robotics in Construction*. [Online] (Isarc), 121–127. Available from: doi:10.22260/isarc2017/0016.

Peffers, K., Tuunanen, T., Rothenberger, M.A. & Chatterjee, S. (2007) A design science research methodology for information systems research. *Journal of Management Information Systems*. [Online] 24 (3), 45–77. Available from: doi:10.2753/MIS0742-1222240302.

Pellerin, R., Perrier, N. & Berthaut, F. (2020) A survey of hybrid metaheuristics for the resource-constrained project scheduling problem. *European Journal of Operational Research*. [Online] 280 (2), 395–416. Available from: doi:10.1016/j.ejor.2019.01.063 [Accessed: 17 February 2020].

Pinedo, M.L. (2008) *Scheduling: Theory, algorithms, and systems*. [Online]. Available from: doi:10.1007/978-0-387-78935-4.

Poppenborg, J. & Knust, S. (2016) A flow-based tabu search algorithm for the RCPSP with transfer times. *OR Spectrum*. [Online] 38 (2), 305–334. Available from: doi:10.1007/s00291-015-0402-2.

Pradhananga, N. & Teizer, J. (2013) Automatic spatio-temporal analysis of construction site equipment operations using GPS data. *Automation in Construction*. [Online] 29, 107–122. Available from: doi:10.1016/j.autcon.2012.09.004.

Preidel, C., Borrmann, A., Oberender, C. & Tretheway, M. (2017) Seamless Integration of Common Data Environment Access into BIM Authoring Applications: the BIM Integration Framework. In: *eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2016: Proceedings of the 11th European Conference on Product and Process Modelling (ECPPM 2016)*. [Online]. 2017 Limassol, Cyprus. p. 119. Available from: doi:10.13140/RG.2.1.4487.4488.

Pultar, M. (1990) Progress-Based Construction Scheduling. *Journal of Construction Engineering and Management*. [Online] 116 (4), 670–688. Available from:

doi:10.1061/(ASCE)0733-9364(1990)116:4(670).

pySHACL (n.d.) *RDFLib/pySHACL: A Python validator for SHACL*. [Online]. Available from: https://github.com/RDFLib/pySHACL [Accessed: 16 June 2020].

Quintero, D., Genovese, W.M., Kim, K., Li, M.J.M., et al. (2015) *IBM Software Defined Environment*. SanFrancisco, USA, IBM Red books.

Rabadi, G. (2016) *Heuristics, Metaheuristics and Approximate Methods in Planning and Scheduling*. International Series in Operations Research & Management Science. Ghaith Rabadi (ed.). [Online]. Cham, Springer International Publishing. Available from: doi:10.1007/978-3-319-26024-2.

Radulovic, F., Poveda-Villalón, M., Vila-Suero, D., Rodríguez-Doncel, V., et al. (2015) Guidelines for Linked Data generation and publication: An example in building energy consumption. *Automation in Construction*. [Online] 57, 178–187. Available from: doi:10.1016/j.autcon.2015.04.002.

Ratajczak-Ropel, E. (2018) Experimental Evaluation of Agent-based Approaches to Solving Multi-mode Resource-Constrained Project Scheduling Problem. *Cybernetics and Systems*. [Online] 49 (5–6), 296–316. Available from: doi:10.1080/01969722.2017.1418269.

Ratajczak-Ropel, E. & Skakovski, A. (2018) *Resource-constrained project scheduling*. [Online]. Available from: doi:10.1007/978-3-319-62893-6_4.

Razavi, S.N. & Moselhi, O. (2012) GPS-less indoor construction location sensing. *Automation in Construction*. [Online] 28, 128–136. Available from: doi:10.1016/j.autcon.2012.05.015.

RDFLib (n.d.) *rdflib 5.0.0 — rdflib 5.0.0 documentation*. [Online]. Available from:

https://rdflib.readthedocs.io/en/stable/ [Accessed: 16 June 2020].

Red Hat Inc (2019) *Drools*. [Online]. 2019. Available from: https://www.drools.org/ [Accessed: 17 November 2019].

Redmond, A., Hore, A., Alshawi, M. & West, R. (2012) Exploring how information exchanges can be enhanced through Cloud BIM. *Automation in Construction*. [Online] 24, 175–183. Available from: doi:10.1016/j.autcon.2012.02.003.

Reeves, C.R. (1993) Improving the Efficiency of Tabu Search for Machine Sequencing Problems. *The Journal of the Operational Research Society*. [Online] 44 (4), 375. Available from: doi:10.2307/2584415.

Ren, Z., Anumba, C.J. & Tah, J. (2011) RFID-facilitated construction materials management (RFID-CMM) - A case study of water-supply project. *Advanced Engineering Informatics*. [Online] 25 (2), 198–207. Available from: doi:10.1016/j.aei.2010.02.002.

Richard S. Sutton, A.G.B. (2008) *Reinforced Learning: An Introduction*. [Online]. Available from: doi:10.1007/978-3-540-29678-2_6199.

Rodrigues, S.B. & Yamashita, D.S. (2015) Exact methods for the resource availability cost problem. In: *Handbook on Project Management and Scheduling Vol. 1*. [Online]. Springer International Publishing. pp. 319–338. Available from: doi:10.1007/978-3-319-05443-8_15.

Sacks, R., Barak, R., Belaciano, B., Gurevich, U., et al. (2013) KanBIM workflow management system: Prototype implementation and field testing. *Lean Construction Journal*. [Online] 2013 (2012), 19–35. Available from: https://www.leanconstruction.org/media/docs/lcj/2013/LCJ_12_004.pdf.

Sacks, R., Korb, S., Barak, R. & Geva, G. (2017a) *Building lean, building BIM : improving*

construction the Tidhar way. Routledge.

Sacks, R., Koskela, L., Dave, B.A. & Owen, R. (2010) Interaction of Lean and Building Information Modelling in Construction. *Journal of Construction Engineering and Management*. [Online] 136 (9), 968–980. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000203.

Sacks, R., Ma, L., Yosef, R., Borrmann, A., et al. (2017b) Semantic Enrichment for Building Information Modelling: Procedure for Compiling Inference Rules and Operators for Complex Geometry. *Journal of Computing in Civil Engineering*. [Online] 31 (6), 04017062. Available from: doi:10.1061/(ASCE)CP.1943-5487.0000705.

Sacks, R., Radosavljevic, M. & Barak, R. (2010) Requirements for building information modelling based lean production management systems for construction. *Automation in Construction*. [Online] 19 (5), 641–655. Available from: doi:10.1016/j.autcon.2010.02.010 [Accessed: 23 April 2018].

Schwabe, K., König, M. & Teizer, J. (2016) BIM applications of rule-based checking in construction site layout planning tasks. *ISARC 2016 - 33rd International Symposium on Automation and Robotics in Construction*. [Online] (Isarc), 209–217. Available from: doi:10.22260/isarc2016/0026.

Schwabe, K., Liedtke, S., König, M. & Teizer, J. (2016) BIM-based Construction Site Layout Planning and Scheduling. In: *16th International Conference on Computing in Civil and Building Engineering*. [Online]. 2016 pp. 679–686. Available from: http://www.see.eng.osaka-u.ac.jp/seeit/icccbe2016/Proceedings/Full_Papers/086-352.pdf.

Schwabe, K., Teizer, J. & König, M. (2019) Applying rule-based model-checking to construction site layout planning tasks. *Automation in Construction*. [Online] 97

(November 2018), 205–219. Available from: doi:10.1016/j.autcon.2018.10.012.

Schwindt, C. & Zimmermann, J. (2015) *Handbook on project management and scheduling vol. 1*. [Online]. Springer International Publishing. Available from: doi:10.1007/978-3-319-05443-8.

Sebastian, R. (2011) Changing roles of the clients, architects and contractors through BIM. *Engineering, Construction and Architectural Management*. [Online] 18 (2), 176–187. Available from: doi:10.1108/09699981111111148.

Servranckx, T. & Vanhoucke, M. (2019) A tabu search procedure for the resource-constrained project scheduling problem with alternative subgraphs. *European Journal of Operational Research*. [Online] 273 (3), 841–860. Available from: doi:10.1016/j.ejor.2018.09.005.

Shi, F., Soman, R.K., Han, J. & Whyte, J.K. (2020) Automation in Construction Addressing adjacency constraints in rectangular floor plans using Monte- Carlo Tree Search. *Automation in Construction*. [Online] 115 (August 2019), 103187. Available from: doi:10.1016/j.autcon.2020.103187.

Singh, V., Gu, N. & Wang, X. (2011) A theoretical framework of a BIM-based multi-disciplinary collaboration platform. *Automation in Construction*. [Online] 20 (2), 134–144. Available from: doi:10.1016/j.autcon.2010.09.011.

Skowroński, M.E., Myszkowski, P.B., Adamski, M. & Kwiatek, P. (2013) Tabu search approach for Multi-Skill Resource-Constrained Project Scheduling Problem - IEEE Conference Publication. In: *2013 Federated Conference on Computer Science and Information Systems, Krako¿w*. [Online]. 2013 pp. 153–158. Available from: https://ieeexplore.ieee.org/abstract/document/6643991 [Accessed: 4 June 2020].

Smith, I.F.C. & Raphael, B. (2013) *Engineering Informatics:Fundementals of Computer Aided*

Engineering. 2nd edition. West Sussex, UK, John Wiley & Sons.Ltd.

Solihin, W., Dimyadi, J., Lee, Y.-C., Eastman, C., et al. (2017) The Critical Role of Accessible Data for BIM-Based Automated Rule Checking Systems. In: *Lean and Computing in Construction Congress - Volume 1: Proceedings of the Joint Conference on Computing in Construction*. [Online]. 4 July 2017 Edinburgh, Heriot-Watt University. pp. 53–60. Available from: doi:10.24928/JC3-2017/0161.

Solihin, W. & Eastman, C. (2015) Classification of rules for automated BIM rule checking development. *Automation in Construction*. [Online] 53, 69–82. Available from: doi:10.1016/j.autcon.2015.03.003.

Solihin, W., Eastman, C. & Lee, Y.-C. (2015) Toward robust and quantifiable automated IFC quality validation. *Advanced Engineering Informatics*. [Online] 29 (3), 739–756. Available from: doi:10.1016/j.aei.2015.07.006.

Soman, R.K. (2019a) Linked-data based dynamic constraint solving framework to support look-ahead-planning in construction. In: *36th CIB W78 conference*. 2019 New Castle, United Kingdom. p.

Soman, R.K. (2019b) Modelling construction scheduling constraints using Shapes Constraint Language (SHACL). In: *2019 European Conference on Computing in Construction*. [Online]. 2019 Chania, Greece. pp. 351–358. Available from: doi:10.35490/EC3.2019.170.

Soman, R.K., Raphael, B. & Varghese, K. (2017) A System Identification Methodology to monitor construction activities using structural responses. *Automation in Construction*. [Online] 75, 79–90. Available from: doi:10.1016/j.autcon.2016.12.006.

Soman, R.K. & Whyte, J.K. (2020) Codification Challenges for Data Science in Construction.

Journal of Construction Engineering and Management. [Online] 146 (7), 04020072. Available from: doi:10.1061/(ASCE)CO.1943-7862.0001846.

Song, M.H., Fischer, M. & Theis, P. (2017) Field Study on the Connection between BIM and Daily Work Orders. *Journal of Construction Engineering and Management*. [Online] 143 (5), 06016007. Available from: doi:10.1061/(ASCE)CO.1943-7862.0001267.

Storn, R. & Price, K. (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*. 11 (4), 341–359.

Sutton, R.S. & Barto, A.G. (2018) *Reinforcement Learning: An Introduction*. 2nd editio. Cambridge, Massachusetts., MIT Press.

Tan, X., Hammad, A. & Fazio, P. (2010) Automated code compliance checking for building envelope design. *Journal of Computing in Civil Engineering*. [Online] 24 (2), 203–211. Available from: doi:10.1061/(ASCE)0887-3801(2010)24:2(203).

Taneja, S., Akinci, B., Garrett, J.H., Soibelman, L., et al. (2011) Sensing and Field Data Capture for Construction and Facility Operations. *Journal of Construction Engineering and Management*. [Online] 137 (10), 870–881. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000332.

Tang, S., Shelden, D.R., Eastman, C.M., Pishdad-Bozorgi, P., et al. (2019) A review of building information modelling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. *Automation in Construction*. [Online] 101, 127–139. Available from: doi:10.1016/j.autcon.2019.01.020.

Taylor, J.E. (2007) Antecedents of Successful Three-Dimensional Computer-Aided Design Implementation in Design and Construction Networks. *Journal of Construction Engineering and Management*. [Online] 133 (12), 993–1002. Available from:

doi:10.1061/(ASCE)0733-9364(2007)133:12(993).

Tchouanguem Djuedja, J.F., Karray, M.H., Foguem, B.K., Magniont, C., et al. (2019) Interoperability Challenges in Building Information Modelling (BIM). In: *Proceedings of the I-ESA Conferences*. [Online]. Springer International Publishing. pp. 275–282. Available from: doi:10.1007/978-3-030-13693-2_23 [Accessed: 25 January 2020].

Teekeng, W., Thammano, A., Unkaw, P. & Kiatwuthiamorn, J. (2016) A new algorithm for flexible job-shop scheduling problem based on particle swarm optimization. *Artificial Life and Robotics*. [Online] 21 (1), 18–23. Available from: doi:10.1007/s10015-015-0259-0.

Terkaj, W. & Pauwels, P. (2017) A method to generate a modular ifcOWL ontology. *CEUR Workshop Proceedings*. 2050.

Terkaj, W., Schneider, G.F. & Pauwels, P. (2017) Reusing Domain Ontologies in Linked Building Data : the Case of Building Automation and Control. In: *Proceedings of the 8th Workshop Formal Ontologies Meet Industry, Joint Ontology Workshops 2017, CEUR Workshop Proceedings*. [Online]. 2017 p. Available from: https://www.semanticscholar.org/paper/Reusing-Domain-Ontologies-in-Linked-Building-Data%3A-Terkaj-Schneider/5533517181b3e78a52bb207d5ece00ee4cb0c562.

Thabet, W.Y. & Beliveau, Y.J. (1997) SCaRC: Space-Constrained Resource-Constrained Scheduling System. *Journal of Computing in Civil Engineering*. [Online] 11 (1), 48–59. Available from: doi:10.1061/(ASCE)0887-3801(1997)11:1(48).

Thomas, P.R. & Salhi, S. (1998) A Tabu Search Approach for the Resource Constrained Project Scheduling Problem. *Journal of Heuristics*. [Online] 4 (2), 123–139. Available from: doi:10.1023/A:1009673512884.

Tommelein, I.D. (1998) Pull-Driven Scheduling for Pipe-Spool Installation: Simulation of

Lean Construction Technique. *Journal of Construction Engineering and Management*. [Online] 124 (4), 279–288. Available from: doi:10.1061/(ASCE)0733-9364(1998)124:4(279).

Tuttas, S., Braun, A., Borrmann, A. & Stilla, U. (2016) Evaluation of acquisition strategies for image-based construction site monitoring. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*. [Online]. 2016 pp. 733–740. Available from: doi:10.5194/isprsarchives-XLI-B5-733-2016.

UnityTechnologies (2019) *3D Software for Architecture, Engineering & Construction | Unity*. [Online]. 2019. Available from: https://unity.com/solutions/architecture-engineering-construction [Accessed: 16 June 2020].

UnityTechnologies (n.d.) *Assets, Resources and AssetBundles - Unity Learn*. [Online]. Available from: https://learn.unity.com/tutorial/assets-resources-and-assetbundles?_ga=2.195376108.1174687737.1592338982-1034668825.1592338982#5c7f8528edbc2a002053b5a6 [Accessed: 16 June 2020].

Valero, E., Adán, A. & Cerrada, C. (2015) Evolution of RFID Applications in Construction: A Literature Review. *Sensors*. [Online] 15 (7), 15988–16008. Available from: doi:10.3390/s150715988.

Valls, V., Ballestín, F. & Quintanilla, S. (2005) Justification and RCPSP: A technique that pays. In: *European Journal of Operational Research*. [Online]. 1 September 2005 North-Holland. pp. 375–386. Available from: doi:10.1016/j.ejor.2004.04.008.

Valls, V., Ballestín, F. & Quintanilla, S. (2006) Justification technique generalizations. In: *International Series in Operations Research and Management Science*. [Online]. Springer New York LLC. pp. 205–223. Available from: doi:10.1007/978-0-387-33768-5_8.

Vu Hoang, Nam Törmä, S. (2015) Implementation and experiments with an IFC-to-Linked Data converter. In: *Proceedings of the 32nd international CIB W78 conference*. 2015 Eindhoven, Netherlands. p.

Wagner, H.M. (1959) An integer linear-programming model for machine scheduling. *Naval Research Logistics Quarterly*. [Online] 6 (2), 131–140. Available from: doi:10.1002/nav.3800060205.

Wang, H.W., Lin, J.R. & Zhang, J.P. (2020) Work package-based information modelling for resource-constrained scheduling of construction projects. *Automation in Construction*. [Online] 109 (September 2019), 102958. Available from: doi:10.1016/j.autcon.2019.102958.

Wang, J., Shou, W., Wang, X. & Wu, P. (2016) Developing and evaluating a framework of total constraint management for improving workflow in liquefied natural gas construction. *Construction Management and Economics*. [Online] 34 (12), 859–874. Available from: doi:10.1080/01446193.2016.1227460.

Wang, L. (2017) Heterogeneous Data and Big Data Analytics. *Automatic Control and Information Sciences*. [Online] 3 (1), 8–15. Available from: doi:10.12691/acis-3-1-3.

Wang, R.Y. & Strong, D.M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems,*. 12 (4), 5–33.

Whyte, J. (2011) Managing digital coordination of design: emerging hybrid practices in an institutionalized project setting. *Engineering Project Organization Journal*. [Online] 1 (3), 159–168. Available from: doi:10.1080/21573727.2011.597743.

Whyte, J., Stasis, A. & Lindkvist, C. (2016) Managing change in the delivery of complex projects: Configuration management, asset information and ‘big data’. *International*

Journal of Project Management. [Online] 34 (2), 339–351. Available from: doi:10.1016/j.ijproman.2015.02.006.

Won, J., Lee, G., Dossick, C. & Messner, J. (2013) Where to Focus for Successful Adoption of Building Information Modelling within Organization. *Journal of Construction Engineering and Management*. [Online] 139 (11), 04013014. Available from: doi:10.1061/(ASCE)CO.1943-7862.0000731.

World Wide Web Consortium, . (2004) *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. [Online]. 2004. Available from: https://www.w3.org/Submission/SWRL/ [Accessed: 17 November 2019].

World Wide Web Consortium (W3C) (2006) *Best Practice Recipes for Publishing RDF Vocabularies*. [Online]. 2006. Available from: https://www.w3.org/2001/sw/BestPractices/VM/http-examples/2006-01-18/ [Accessed: 19 June 2020].

World Wide Web Consortium (W3C) (2012) *OWL 2 Web Ontology Language Overview*. [Online]. 2012. Available from: https://www.w3.org/TR/owl2-overview/ [Accessed: 25 January 2020].

World Wide Web Consortium (W3C) (n.d.) *RDF Schema 1.1*. [Online]. 2014. Available from: https://www.w3.org/TR/rdf-schema/ [Accessed: 25 January 2020].

World Wide Web Consortium (W3C) (2013) *Rule Interchange Format (RIF)*. [Online]. 2013. Available from: https://www.w3.org/TR/rif-overview/ [Accessed: 25 January 2020].

World Wide Web Consortium (W3C) (2017) *Shapes Constraint Language (SHACL)*. [Online]. 2017. Available from: https://www.w3.org/TR/shacl/ [Accessed: 25 January 2020].

World Wide Web Consortium (W3C) (2011) *SPIN - Overview and Motivation*. [Online]. 2011.

Available from: https://www.w3.org/Submission/2011/SUBM-spin-overview-20110222/ [Accessed: 6 March 2020].

Wülfing, A., Windisch, R. & Scherer, R.J. (2014) A visual BIM query language A visual BIM query language. *10th European Conference on Product & Process Modelling*. [Online] (SEPTEMBER 2014), 157–164. Available from: doi:10.13140/2.1.2158.9769.

Xia, W. & Wu, Z. (2005) An effective hybrid optimization approach for multi-objective flexible job-shop scheduling problems. *Computers & Industrial Engineering*. [Online] 48 (2), 409–425. Available from: doi:10.1016/j.cie.2005.01.018.

Younas, M. (2019) Research challenges of big data. *Service Oriented Computing and Applications*. [Online] 13 (2), 105–107. Available from: doi:10.1007/s11761-019-00265-x.

Zadeh, P.A., Wang, G., Cavka, H.B., Staub-French, S., et al. (2017) Information Quality Assessment for Facility Management. *Advanced Engineering Informatics*. [Online] 33, 181–205. Available from: doi:10.1016/j.aei.2017.06.003.

Zhang, C. & Arditi, D. (2013) Automated progress control using laser scanning technology. *Automation in Construction*. [Online] 36, 108–116. Available from: doi:10.1016/j.autcon.2013.08.012.

Zhang, C. & Beetz, J. (2015) Model Checking on the Semantic Web: IFC Validation Using Modularized and Distributed Constraints. In: *Proc. of the 32nd CIB W78 Conference 2015, 27th-29th October 2015, Eindhoven, The Netherlands*. 2015 pp. 819–827.

Zhang, C. & Beetz, J. (2016a) Querying Linked Building Data Using SPARQL with Functional Extensions. In: *11th European Conference on Product & Process Modelling 2016*. [Online]. 2016 p. Available from: https://research.tue.nl/en/publications/querying-linked-

building-data-using-sparql-with-functional-extens.

Zhang, C. & Beetz, J. (2016b) Querying Linked Building Data Using SPARQL with Functional Extensions. *eWork and eBusiness in Architecture, Engineering and Construction (ECPPM 2016) - Proceedings of 11th European Conference on Product and Process Modelling. - S.l. : s.n., 2016*. (October).

Zhang, J., Ding, G., Zou, Y., Qin, S., et al. (2017a) Review of job shop scheduling research and its new perspectives under Industry 4.0. *Journal of Intelligent Manufacturing*. [Online] 30 (4), 1809–1830. Available from: doi:10.1007/s10845-017-1350-2.

Zhang, S., Boukamp, F. & Teizer, J. (2015) Ontology-based semantic modelling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Automation in Construction*. [Online] 52, 29–41. Available from: doi:10.1016/j.autcon.2015.02.005.

Zhang, S., Pan, F., Wang, C., Sun, Y., et al. (2017b) BIM-Based Collaboration Platform for the Management of EPC Projects in Hydropower Engineering. *Journal of Construction Engineering and Management*. [Online] 143 (12), 04017087. Available from: doi:10.1061/(ASCE)CO.1943-7862.0001403.

Zhang, S., Teizer, J., Lee, J.K., Eastman, C.M., et al. (2013) Building Information Modelling (BIM) and Safety: Automatic Safety Checking of Construction Models and Schedules. *Automation in Construction*. [Online] 29, 183–195. Available from: doi:10.1016/j.autcon.2012.05.006.

Zhang, W., Wen, J.B., Zhu, Y.C. & Hu, Y. (2017c) Multi-Objective Scheduling Simulation of Flexible Job-Shop Based on Multi-Population Genetic Algorithm. *International Journal of Simulation Modelling*. [Online] 16 (2), 313–321. Available from: doi:10.2507/IJSIMM16(2)CO6.

Zhang, Y., Wang, S. & Ji, G. (2015) *A Comprehensive Survey on Particle Swarm Optimization Algorithm and Its Applications*. [Online] Available from: doi:10.1155/2015/931256 [Accessed: 4 June 2020].

Zhao, F., Shao, Z., Wang, J. & Zhang, C. (2016) A hybrid differential evolution and estimation of distribution algorithm based on neighbourhood search for job shop scheduling problems. *International Journal of Production Research*. [Online] 54 (4), 1039–1060. Available from: doi:10.1080/00207543.2015.1041575.

Zhong, B.T., Ding, L.Y., Luo, H.B., Zhou, Y., et al. (2012) Ontology-based semantic modelling of regulation constraint for automated construction quality compliance checking. *Automation in Construction*. [Online] 28, 58–70. Available from: doi:10.1016/j.autcon.2012.06.006.

Appendix A: Data collection record -Method M1

Table A.1: Data collection record

Sources	Source ID	Details	Time in the field; documentation/ research records; and/or dates	
Case 1: Student appartment				
Document	C1D1	Construction Programme	3 years of planning data	
	C1D2	BIM Model	1.5 gigabytes	
	C1D3	BIM training document	5 pages	
	C1D4	Access to Common data environment	2 weeks of access	
Interviews	C1I1	Digital engineer (15/08/2017)	2 hours, Field note (3 pages)	
	C1I2	Planner (15/08/2017	1 hour, Field note (2 pages)	
Office visit	C1S1	Contractor's office	Multiple visits over 2 weeks, Fieldnotes (7 pages) 14/08/2017-18/08/2017	
Case 2: Metro project				
Documents	C2D1	BIM Execution Plan	17 Pages	
	C2D2	Digital Project Management (PPT)	26 slides	
	C2D3	Employee Information Requirements	32 pages	
	C2D4	Report	13 pages	
	C2D5	Request for proposal	484 pages	
Informal interviews	C2I1	Project manager	10/01/2018, 30 Minutes, Fieldnotes (4 pages)	
	C2I2	Chief site engineer	10/01/2018, 30 Minutes, Fieldnotes (3 pages)	
	C2I3	Casting yard engineer	10/01/2018, 20 Minutes, Fieldnotes (3 pages)	
	C2I4	BIM consultant 1	11/01/2018, 20 Minutes, Fieldnotes (2 pages)	
	C2I5	BIM consultant 2	11/01/2018, 20 Minutes, Fieldnotes (2 pages)	
	C2I6	BIM consultant 3	11/01/2018, 20 Minutes, Fieldnotes (2 pages)	
Site visit	C2S1	Casting yard	10/01/2018, 1 hour	Fieldnotes (5 pages); 22 Photographs
	C2S2	Station site	10/01/2018, 2 hours	
	C2S3	Viaduct site	10/01/2018, 1.5 hours	

Workshop	C2W1	Workshop with 6 metros, Nagpur, Ahmedabad, Kochi, Chennai, Bombay, Hyderabad, Delhi.	08/03/2018, 8 hours, 22 Photographs; 6 flipcharts; written summary (30 pages)
Case 3: Water project			
Set-up meeting	C3M1	Introducing the study	31/05/2018, 1 hour, 20 slides
Documents	C3D1	Design workflow	2 Pages
	C3D2	Access to CDE 1	2 months
	C3D3	Summary construction programme	4 year data
	C3D4	Construction programme update	1 Month data
	C3D5	Work package plan	45 pages
Semi-structured interviews	C3I1	Project Engineer	09/07/2018 1 hour, Taped & Transcribed (T&T) (10 pages)
	C3I2	Project Planner Lead	10/07/2018, 1 hour, T&T (10 pages)
	C3I3	Digital Engineer Lead	11/07/2018, 1 hour, T&T (11 pages)
	C3I4	Information manager	11/07/2018, 1 hour, T&T (15 pages)
	C3I5	Technical manager/ DfMA design lead	12/07/2018, 1 hour, T&T (15 pages)
	C3I6	Senior digital engineer	12/07/2018, 1 hour, T&T (8 pages)
	C3I7	Design manager	13/07/2018, 1 hour, Fieldnotes (4 pages)
	C3I8	Principal engineer	17/07/2018, 1 hour, T&T (12 pages)
Office visit	C3S1	Contractor's office	05/07/2018-19/07/2018, Multiple visits over 2 weeks, Fieldnotes (5 pages)
Meetings	C3M2	Temporary works design meeting	13/07/2018, 2 hours, Fieldnotes (3 pages)
	C3M3	Progress review meeting	16/07/2018, 4 hours, Fieldnotes (4 pages)
4D Review	C3R1	Digital Engineer	12/10/2019, 3 hours ,T & T (20 pages)

Appendix B: Case Details-Method M1

Table B.0.1: Details of the cases

Cases	Location	Type	Project cost
Student Apartment	United Kingdom	Residential	> $ 75 million
Metro Project	India	Urban transit infrastructure	> $1.2 Billion
Water Project	United Kingdom	Waster infrastructure	>$ 6 Billion

B.1 Student apartment

The *Student apartment* is a 34-storey residential tower costing more than $75 million and would house 192 apartments. The construction of the project started in 2016 and ended in 2018. The construction of this building used Design for Manufacturing and Assembly (DfMA) philosophy and digital engineering to deliver and assemble structural steelwork, modules, risers, plant, and bathroom pods efficiently. Multi-storey residential student apartment block in the United Kingdom is an exemplary project exhibiting the use of BIM in the United Kingdom, constructed by a leading contractor, and using state-of-the-art offsite manufacturing approaches in construction. The project was completed ahead of the completion time and within the stipulated cost.

The construction of the *Student apartment* was executed with the support of the contractor's state of the art digital engineering capabilities. The engineers and managers in the project were exposed and trained in digital engineering, through the prior projects in the firm. The design and the construction were done by different divisions in the same firm. To manage and control information flow in the project, a CDE from the vendor Asite were used. The CDE had inbuilt workflows for approval and publishing. Drawings, models, and other construction

information was published to the CDE so that there exists a single source of truth for decision making. The model-based collaboration was promoted in the project using BIM. The project used design authoring tools from Autodesk. Different teams within the contractor's design division created the architectural, structural and Mechanical Engineering and Plumbing (MEP) models. These models were aggregated by the digital engineering team at the construction side to identify clashes. The aggregated model was linked to the construction programme to inform the construction team of the sequence and discuss the construction methods. Time sliced 4D models were used in the planning and progress meetings. Workspace conflicts were identified using the time-sliced models. The use of digital engineering helped in the completion of project within time and cost.

B.2 Metro project

The metro project is an innovative rail infrastructure in India, costing more than $1.2 Billion. It is an urban rapid transit system consisting of 2 Corridors, with more than 40 KM of elevated viaducts and 40 stations. Construction of the project began on 31 May 2015. The Metro project is pioneering BIM implementation amongst the other metro projects in India, incorporating learning on digital implementation from global megaprojects. It is a mega project with more than 190 sub-contractors and to manage and coordinate the construction information across the large supply chain, and the Metro project uses digital information management across the lifecycle of the project. Unlike the Student apartment, the engineers and project managers in the Metro project had not previously worked in BIM-enabled projects. Hence, the client was actively involved in training the supply chain on digital project management strategies by collaborating with the software vendors. Also, client onboarded a specialist team to coordinate the digital project management and to curate the models submitted by the supply chain as well as supporting the client with digital information management.

The metro project uses a customized 5D BIM platform for project management and visual project monitoring. The platform integrates the benefits of an Enterprise Resource Planning (ERP) software from SAP with construction design authoring and review software and project management software from RIB and Primavera, thus allowing a 5-Dimensional view of the project in near-time. Design review and approval process are carried out in ProjectWise, a CDE from Bentley, to improve transparency and to enable tracking. Drawings will be attached against placeholders. The document will be transmitted to the document controller. Document controller acknowledges transmittal & forwards to Lead Reviewer to initiate the approval process. Reports are generated to track the status of review and review process requiring attention are flagged. Building information models are developed by building on the approved 2D drawings. Dimensions, volume and surface area, material definitions, quantity take-off, asset tags and link to related documents are mandated to be included in every component in the building information model so that it can be linked to the schedule and the ERP software. The combined ability of these multiple technology products integrated into a single 5-Dimensional Digital Project Management Platform has the potential to play a decisive role in ensuring project delivery timelines and cutting upfront expenditures.

B.3 Water project

The water project is a major sewage infrastructure megaproject in the United Kingdom. It is a major project costing more than $ 6 Billion and uses innovative technological solutions to futureproof construction and deliver a physical asset as well as a digital asset for operation. The project involves the construction of an underground sewer with a length of 25km, and associated infrastructure. The construction of the project started in 2016, construction of the *Water project* is aimed to take seven to eight years with all work completed by 2024. The construction of the project was divided into three sections, each section handled by different joint ventures of leading sub-contractors in the country.

To manage the complex construction and large supply chain, the *Water project* uses state of the art digital engineering and project management practices. The engineers and the managers in the project had prior experience in working with BIM-enabled project. This comprises of model-based design reviews to assist decision-making; and delivery teams using building information models for designing, sequencing; planning, live progress reporting and for communicating their works. Information flow is handled through CDEs. Complete engineering data is stored and managed within CDEs, which utilizes BIM Level 2 workflows. CDE provided 1,200 users from more than 45 stakeholders to manage more than 190,000 documents, as well as 4,500 drawings and more than 1,500 models. The CDE has inbuilt approval and publishing processes. These digital engineering and document management strategies assist in managing the complexity and scale of the project.

Appendix C: Method M1-Emergent codes

1. Shared understanding
2. Document control
3. Speed of change
4. Volume of data
5. Poor meeting management
6. Loss of learning from previous project
7. Long review time
8. No traceability in emails
9. Approvals
10. Common platform
11. Slow data transfer
12. Wasted time
13. Data incompatibility
14. Change management -data inconsistency
15. Wrong modelling techniques
16. PDF viewer
17. Data Loss
18. Software versions
19. Different CDE-Data inconsistency
20. Manual process
21. Taxonomy
22. No upload accesses
23. 2 CDE
24. Complexity
25. Access control
26. Slow process
27. codifying constraints
28. Cost of software
29. Hard copy signatures
30. Improper information storage
31. Inefficient search
32. Lack of process change
33. Low level of detail
34. Multiple CDE
35. Need for signature
36. No obsolete status
37. Shared understanding of RFI
38. Unclear process
39. Variation in process/software
40. Approval process
41. Bypassing the workflow
42. Correct storage
43. Document control bottlenecks
44. Lack of document controller's domain expertise
45. lack of obsolete status
46. Lack of real time updating

47. Lack of understanding of common data environment
48. Lack of understanding of document control workflow
49. Version control
50. Improper modelling techniques
51. Lack of standardized information tagging
52. Access to latest decision
53. Asynchronous storage
54. Data consistency
55. Double handling
56. Ease of use
57. Longer times for organizing the data
58. Process slowdown

Appendix D: Relationship between data quality dimensions and codification challenges

Table D.1: The relationship between data quality dimensions (Batini and Scannapieco, 2016) and the identified codification challenges

Data quality dimension	Code	Codification challenge
Accuracy	*Multiple modelling techniques*	Software usage
	Document control bottlenecks	Information sharing
Completeness	*Interoperability* *Information loss during conversions* *Multiple modelling techniques*	Software usage
	Lack of process change	Information sharing
	Loss of constraint information *Low level of detail*	Missing process information
Temporality	*Unstructured information sharing* *Document control bottlenecks* *Lack of process change*	Information sharing
Consistency	*Document control bottlenecks*	Information sharing
Accessibility	*Interoperability*	Software usage
	Unstructured information sharing *Drawings and file-based sharing* *Lack of process change*	Information sharing
	Loss of constraint information *Low level of detail*	Missing process information
Data Provenance	*Information loss during conversion*	Software usage
	Lack of process change	Information sharing

Table D.2 : Data quality dimensions related to codes

Data quality dimension	Codes and data
Accuracy	*Multiple modelling techniques:* "use the wrong tool to model something […] I can't just say there's a slab now, that's just a piece of geometry" (Digital engineer, C3I3); "when you try to extract 2D drawings from 3D BIM models, those drawings are not as correct and as detailed as they used to be" (Technical manager, C3I5)

	Document control bottlenecks: “it’s no longer the most current version anymore by the time I’m reviewing it” (Project engineer, C3I1) “he keeps on updating but he hasn’t he hasn’t put it on the [CDE1].” (Technical manager, C3I5) “I just want to know where I can get my latest drawing” (Digital engineer, C3I3) “I think someone within the doc management system had obviously circumnavigated it somehow, to get the drawings out. And then when we were trying to get the said revisions for our set out, the system wouldn't allow it because directory hadn't been properly created.” (Technical manager, C3I5)
Completeness	*Interoperability:* “transferring things […], you lose data” (Technical manager, C3I5) *Information loss during conversion:* “when you upload a PDF.” (Information manager, C3I4) *Multiple modelling techniques:* “use the wrong tool to model something […] I can’t just say there’s a slab now, that’s just a piece of geometry” (Digital engineer, C3I3); “when you try to extract 2D drawings from 3D BIM models, those drawings are not as correct and as detailed as they used to be” (Technical manager, C3I5)
	Lack of process change: “We’re going to print it out, we’re going to staple it together […] get three signatures, scan it back in, put it back into [CDE1] and submit it.”(Project engineer, C3I1) “it’s not actually speeding everything up, it’s sort of making everything a lot slower; which I find very frustrating” (Principal engineer, C3I8) “can be very confusing when we have two platforms” (Technical manager, C3I5) “Everything had to be taken out of one data environment and pushed into another. One of the issues with that is the consistency or the compliance or knowing the latest versions of information” (Digital engineer, C3I6) “As the contractor, then we have to deliver it to a completely separate, disconnected CDE […] we’re double-handling” (Digital engineer, C3I3)
	Loss of constraint information: “we physically need that information to know what we’re building and what the constraints in building it are.” (Project planner, C3I2). “Access chamber works will conflict with access road for pile work, piling work package has to be moved back 2 weeks.” (Progress review meeting, C3M3); *Low level of detail:* “Work package for three spans were linked to a work order. Model showed the deck for a span was completed before the pier supporting it was completed because the work package for the first span was reported as completed.” (Field notes- BIM Consultant 2, C2I5)
Timeliness	*Unstructured information sharing:* “when I have finished everything - by the way, we have this spread sheet” (Technical manager, C3I5) *Document control bottlenecks:* “it’s no longer the most current version anymore by the time I’m reviewing it” (Project engineer, C3I1) “he keeps on updating but he hasn’t he hasn’t put it on the [CDE1].” (Technical manager, C3I5) “I just want to know where I can get my latest drawing” (Digital engineer, C3I3) “I think someone within the doc management system had obviously circumnavigated it somehow, to get the drawings out. And then when we were trying to get the said revisions for our set out, the system wouldn't allow it because directory hadn't been properly created.” (Technical manager, C3I5)
	Lack of process change: “We’re going to print it out, we’re going to staple it together […] get three signatures, scan it back in, put it back into [CDE1] and submit it.”(Project engineer, C3I1) “it’s not actually speeding everything up, it’s sort of making everything a lot slower; which I find very frustrating” (Principal engineer, C3I8) “can be very confusing when we have two platforms” (Technical manager, C3I5) “Everything had to be taken out of one data environment and pushed into another. One of the issues with that is the consistency or the compliance or knowing the latest versions of information” (Digital engineer, C3I6) “As the contractor, then we have to deliver it to a completely separate, disconnected CDE […] we’re double-handling” (Digital engineer, C3I3)
Consistency	*Document control bottlenecks:* “it’s no longer the most current version anymore by the time I’m reviewing it” (Project engineer, C3I1) “he keeps on updating but he hasn’t he hasn’t put it on the [CDE1].” (Technical manager, C3I5) “I just want to know where I can get my latest drawing” (Digital engineer, C3I3) “I think someone within the doc management system had obviously circumnavigated it somehow, to get the drawings out. And then when we were trying to get the

	said revisions for our set out, the system wouldn't allow it because directory hadn't been properly created." (Technical manager, C3I5)
Accessibility	*Interoperability:* "transferring things […], you lose data" (Technical manager, C3I5); *Unstructured information sharing:* "when I have finished everything - by the way, we have this spread sheet" (Technical manager, C3I5);
	Drawings and file-based sharing: "I use all the navigator tools that we've got here. But I prefer to use AutoCAD because I find it a lot easier" (Project engineer, C3I1) "It's not the best way because we haven't got the technology. I haven't got a big screen" (Project engineer, C3I1)
	Lack of process change: "We're going to print it out, we're going to staple it together […] get three signatures, scan it back in, put it back into [CDE1] and submit it."(Project engineer, C3I1) "it's not actually speeding everything up, it's sort of making everything a lot slower; which I find very frustrating" (Principal engineer, C3I8) "can be very confusing when we have two platforms" (Technical manager, C3I5) "Everything had to be taken out of one data environment and pushed into another. One of the issues with that is the consistency or the compliance or knowing the latest versions of information" (Digital engineer, C3I6) "As the contractor, then we have to deliver it to a completely separate, disconnected CDE […] we're double-handling" (Digital engineer, C3I3)
	Loss of constraint information: "we physically need that information to know what we're building and what the constraints in building it are." (Project planner, C3I2). "Access chamber works will conflict with access road for pile work, piling work package has to be moved back 2 weeks." (Progress review meeting, C3M3)
	Low level of detail: "Work package for three spans were linked to a work order. Model showed the deck for a span was completed before the pier supporting it was completed because the work package for the first span was reported as completed." (Field notes- BIM Consultant 2, C2I5)
Data Provenance	*Information loss during conversion:* "when you upload a PDF." (Information manager, C3I4); *Drawings and file-based sharing:* "I use all the navigator tools that we've got here. But I prefer to use AutoCAD because I find it a lot easier" (Project engineer, C3I1) "It's not the best way because we haven't got the technology. I haven't got a big screen" (Project engineer, C3I1)
	Lack of process change: "We're going to print it out, we're going to staple it together […] get three signatures, scan it back in, put it back into [CDE1] and submit it."(Project engineer, C3I1) "it's not actually speeding everything up, it's sort of making everything a lot slower; which I find very frustrating" (Principal engineer, C3I8) "can be very confusing when we have two platforms" (Technical manager, C3I5) "Everything had to be taken out of one data environment and pushed into another. One of the issues with that is the consistency or the compliance or knowing the latest versions of information" (Digital engineer, C3I6) "As the contractor, then we have to deliver it to a completely separate, disconnected CDE […] we're double-handling" (Digital engineer, C3I3)

Accuracy: The organizational and cultural divisions between different teams results in problems associated with multiple modelling techniques, leading to data quality issues concerning the syntactic accuracy of the data. This was evident from the dataset when different people had different perceptions of the model, as in the case of the slab example. Similarly, the hybrid practices associated with information sharing lead to lowering the semantic accuracy of the data as the data with which inferences are made are not accurate due to inefficiencies in

information sharing. When there are syntactic errors in the data, this leads to incorrect insights (for example, if a slab is modelled as a geometric object with attributes attached to it and when a software tool is used to compute quantity take-off from the model for all the slabs). The output would be zero as quantity required as the program fails to identify the geometric object as a slab. If this software is integrated with a costing tool used for cash flow analytics, this error gets propagated into that tool. These errors can be removed to an extent using semantic enrichment programs. However, even the accuracy of inferences of semantic enrichment programs is dependent on the quality of input datasets (Sacks et al., 2017).

Completeness: Concerning the completeness of the data, organizational and cultural divisions between the teams resulting in problems such as interoperability, loss of information during conversions, multiple modelling techniques and the implication of hybrid practices such as printing and scanning the documents play a role in reducing the data quality. For example, the software usage issues caused by the organizational divisions leads to format conversion resulting data loss leading incomplete dataset. Lack of process change also leads to similar problems such as loss of metadata when documents are printed and scanned. The institutionalized practices of process modelling with low levels of detail and the practice of not codifying constraints in the model result in incomplete data. Inferring insights from incomplete datasets reduces the quality of the output. For example, if the constraints are not codified in a schedule, an automatic scheduler would create an unrealistic schedule. This leads to further problems down the line. In the case of a product model, an incomplete dataset used for a structural capacity prediction would give incorrect results.

Temporality: This dimension of data quality is mostly affected by the information-sharing practices in construction using hybrid practices by using old and new practices simultaneously. Document control bottlenecks delaying the submission of data into the system, and the use of multiple CDEs requiring data transfer from one to another (resulting in a delay

in fetching the data) constitute to outdated data,-skewing analytics. When outdated data is used for analytics, the resulting inference is not time appropriate, and decisions taken with those inferences lead to problems. For example, if the construction plan is made using outdated resource availability data, the assigned resource might not be available on the actual construction date. When this is managed manually, the planner makes sure this does not happen. However, when performed automatically, it is necessary that the datasets are updated close to real time.

Consistency: Hybrid practices in information sharing, resulting in document control bottlenecks, multiple CDEs, and circumnavigating workflows, induce inconsistency in the data. The data in one CDE might be different from that in another CDE. Similarly, the document control bottlenecks in publishing the data result in different teams and their databases having different versions of the data. This creates a problem during the data analysis phase. For example, for the same content, different values might exist, of which only one is true. This induces problems of *semantic accuracy* if the computer takes the incorrect value for analysis.

Accessibility: This dimension of data quality is affected by organizational and cultural divisions between teams, hybrid practices, and institutionalized practices in process modelling. Organizational and cultural divisions between teams, resulting in the use of multiple software, multiple modelling techniques, and so forth, affect the accessibility of data as employees must convert the data to access it, and there are instances which show conversions result in loss of data. Hence, the data is only accessible in full content to the team, which created it. Similarly, the hybrid practices, particularly printing and subsequent scanning of the documents, remove the metadata as well as semantic information within those data, resulting in a less-efficient analysis. In addition, the lack of constraint codification restricts the information regarding constraints, making it accessible to the few people who attended the meeting. It is not

embedded in the model and, hence, not accessible to any analytic algorithms. This reduces the capability of such algorithms to infer accurate information and determine causalities.

Provenance: This dimension of the data quality is affected by both organizational and cultural divisions in the industry as well as hybrid practices. Fragmentation in the industry leads to many conversions and much manipulation of data to suit purposes, in the process removing the traceability and origins of the data. Similarly, multiple CDEs, printing and scanning of documents and unstructured information sharing also lead to loss of metadata and ends up removing information regarding origins of the data. This limits algorithms from making inferences based on data origins and their incremental manipulation.

Appendix E: Relationship between levels of machine readability and codification challenges

Table E.1: Relationship between levels of machine readability (Berners-Lee, 2006) and identified codification challenges

Level of machine readability	Codification challenge	Code	Interpretation of the code with respect to machine readability
No Star	Information sharing	Unstructured information sharing	Information communicated over these channels are not indexed nor available over the web.
		Document control bottle neck	
	Missing process information	Loss of constraints	
1-star data	Information sharing	Lack of process change	CDE makes sure that the information is indexed and available on the web. This makes it easier for the user and computer to search for this information.
2-star data	Software usage	Interoperability	The files are stored in a CDE; hence it satisfies the criteria for a 1-star data. In addition, the data is structured in a native format which can be read by a computer using the application program interfaces of the authoring application. Thus, the data is available on the web and structured in non-proprietary format.
	Information sharing	Drawing and file-based sharing	

Table E.2 :Machine readability levels related to codes

Level of machine readability	Code
No Star	*Unstructured information sharing :* "when I have finished everything - by the way, we have this spread sheet" (Technical manager, C3I5) *Document control bottlenecks:* "it's no longer the most current version anymore by the time I'm reviewing it" (Project engineer, C3I1) "he keeps on updating but he hasn't he hasn't put it on the [CDE1]." (Technical manager, C3I5) "I just want to know where I can get my latest drawing" (Digital engineer, C3I3) "I think someone within the doc management system had obviously circumnavigated it somehow, to get the drawings out. And then when we were trying to get the said revisions for our set out, the system wouldn't allow it because directory hadn't been properly created." (Technical manager, C3I5) *Loss of constraint information*: "we physically need that information to know what we're building and what the constraints in building it are." (Project planner, C3I2). "Access chamber works will conflict with access road for pile work, piling work package has to be moved back 2 weeks." (Progress review meeting, C3M3)
1-star data	*Lack of process change:* "We're going to print it out, we're going to staple it together […] get three signatures, scan it back in, put it back into [CDE1] and submit it."(Project engineer, C3I1) "it's not actually speeding everything up, it's sort of making everything a lot slower; which I find very frustrating" (Principal engineer, C3I8) "can be very confusing when we have two platforms" (Technical manager, C3I5) "Everything had to be taken out of one data environment and pushed into another. One of the issues with that is the consistency or the compliance or knowing the latest versions of information" (Digital engineer, C3I6) "As the contractor, then we have to deliver it to a completely separate, disconnected CDE […] we're double-handling" (Digital engineer, C3I3)
2-star data	*Interoperability "*Sometimes the drawings are in Bentley format I am using this MicroStation, but sometimes they were drafted from the client, let's say, in Autodesk. And transferring things from Autodesk to MicroStation, you lose data […] the 2019 will open the 2018, 2017, and 2016. But when the 2020 comes into play, then you cannot open it anymore with the 2019 files that are generated with the 2020" (Water project, Technical manager, C3I5) *Drawings and file-based sharing:* "I use all the navigator tools that we've got here. But I prefer to use AutoCAD because I find it a lot easier" (Project engineer, C3I1) "It's not the best way because we haven't got the technology. I haven't got a big screen" (Project engineer, C3I1)

Most of the construction information observed from the cases satisfies the requirement for a one-star category. The observed projects use a CDE for storing and managing project data, resulting in indexing the data and storing it on online servers, resulting in one-star data.

CDE makes the data easier to retrieve for the computers to make inferences on them. However, the complexity of the new structured methods for information sharing using CDE, and the poor understanding of workflows across the teams, leads to the use of a combination of structured and unstructured channels for information sharing. This aspect reduces the machine readability of the information distributed over unstructured channels as the information is not indexed nor available on a common server. The same issue occurs when the users circumnavigate the workflows to get the work expedited. Similarly, codification challenges associated with missing information also lower the machine readability as the information is not recorded (lack of detail and loss of construction information) and, hence, not indexed or stored in online servers. These issues make the information inaccessible for inferencing.

With regard to the structure of the construction information, construction datasets in the form of drawings and BIM, project management information in project management software (Primavera P6, Asta powerproject, etc.), outputs from Microsoft tools such as Excel, and so forth are structured, satisfying the requirements for two-star data. However, construction data are also unstructured in the forms of PDF documents and scanned documents as there is a lack of process change in information sharing. The lack of structure in the datasets makes inferencing from them difficult, leading to the need for complex algorithms. This results in the categorisation of machine readability to be 1-star.

Where the construction data is structured, the data structure is in proprietary formats which require the APIs to access the semantic relationships within the data. The observed projects do not use open formats or open standards for publishing the data. Proprietary tools for the authorship of construction data are far more advanced and easier to use than the open-source tools. Hence, construction projects resort to using the tested and robust proprietary tools, resulting in issues associated with interoperability and loss of information presented. Thus, the

construction information rarely achieves three-star classification, as mentioned by Berners-Lee (2006).

Low machine readability of the data has implications on data quality. When the construction information is not stored on servers due to information sharing issues or lack of detail in the models, the accessibility of that data is affected, thus reducing the accessibility dimension of the data quality. When information is stored in CDE (satisfying conditions for one-star data) as PDF documents, the structure of the data is not maintained, resulting in data quality issues associated with *syntactic accuracy* and *consistency*. Lack of a data structure removes the context from the information, thus resulting in the need for complex algorithms to infer the contexts and infer from data. This issue also introduces problems associated with *consistency* as the data value for a field might be different in different files, and the lack of context limits the computer programs to detect it. This problem is further worsened as the datasets are not linked since the links are lost when a file is moved from one CDE to another. Storing the information in proprietary formats also reduces the *accessibility* dimension for data quality.

Appendix F: LAS Generated during implementation study

Table F.1: LAS generated for Muck bin construction Bin A

Sub activity	Planned by engineers		Predicted by GA		Predicted by RL	
	Start date	End date	Start date	End date	Start date	End date
Steel Delivery Bin A	2/4/2019	2/4/2019	2/4/2019	2/4/2019	2/4/2019	2/4/2019
Excavate Bin (6m x 11.2m x 4m)	3/4/2019	24/04/2019	3/4/2019	24/04/2019	3/4/2019	24/04/2019
Excavate to Prop Level & Create sump	3/4/2019	5/4/2019	3/4/2019	5/4/2019	3/4/2019	5/4/2019
Crop Piles and Cut Steel	5/4/2019	9/4/2019	5/4/2019	9/4/2019	5/4/2019	9/4/2019
Install River Wall Connection	9/4/2019	10/4/2019	9/4/2019	10/4/2019	9/4/2019	10/4/2019
Install Prop	10/4/2019	12/4/2019	10/4/2019	12/4/2019	10/4/2019	12/4/2019
Excavate to Base Level	12/4/2019	24/04/2019	12/4/2019	24/04/2019	12/4/2019	24/04/2019
Crop Piles	16/04/2019	24/04/2019	24/04/2019	2/5/2019	24/04/2019	2/5/2019
Pour Blinding	25/04/2019	25/04/2019	2/5/2019	2/5/2019	2/5/2019	2/5/2019
Final Trim Piles - Integrity Test - Prep Base	29/04/2019	14/01/2019	3/5/2019	18/5/2019	3/5/2019	18/5/2019

Table F.2: LAS generated for Muck bin construction Bin B

Sub activity	Planned by engineers		Predicted by GA		Predicted by RL	
	Start date	End date	Start date	End date	Start date	End date
Steel Base and Wall Starters	26/03/2019	2/4/2019	26/03/2019	2/4/2019	26/03/2019	2/4/2019
Inspection	2/4/2019	2/4/2019	2/4/2019	2/4/2019	2/4/2019	2/4/2019
Bin B Pour 1 - Base	3/4/2019	3/4/2019	3/4/2019	3/4/2019	3/4/2019	3/4/2019
Formwork for Kicker	4/4/2019	8/4/2019	4/4/2019	8/4/2019	4/4/2019	8/4/2019
Bin B Pour 2 - Kicker	9/4/2019	9/4/2019	9/4/2019	9/4/2019	9/4/2019	9/4/2019
Curing time (strike shutters)	10/4/2019	15/04/2019	10/4/2019	15/04/2019	10/4/2019	15/04/2019
Remove Propping	16/04/2019	16/04/2019	16/04/2019	16/04/2019	16/04/2019	16/04/2019
Steel Walls	17/04/2019	1/5/2019	17/04/2019	1/5/2019	17/04/2019	1/5/2019

Table F.3: LAS generated for Muck bin construction Bin C

Sub activity	Planned by engineers		Predicted by GA		Predicted by RL	
	Start date	End date	Start date	End date	Start date	End date
Final Trim Piles - Integrity Test - Prep Base	1/4/2019	2/4/2019	1/4/2019	2/4/2019	1/4/2019	2/4/2019
Steel Base and Wall Starters	3/4/2019	10/4/2019	3/4/2019	10/4/2019	3/4/2019	10/4/2019
Inspection	10/4/2019	10/4/2019	10/4/2019	10/4/2019	10/4/2019	10/4/2019
Bin C Pour 1 - Base	11/4/2019	11/4/2019	11/4/2019	11/4/2019	11/4/2019	11/4/2019
Formwork for Kicker	12/4/2019	15/04/2019	12/4/2019	15/04/2019	12/4/2019	15/04/2019
Curing time (strike shutters)	15/04/2019	22/04/2019	15/04/2019	22/04/2019	15/04/2019	22/04/2019
Bin C Pour 2 - Kicker	15/04/2019		15/04/2019	15/04/2019	15/04/2019	15/04/2019
Remove Propping	24/04/2019	24/04/2019	24/04/2019	24/04/2019	24/04/2019	24/04/2019

Table F.4: LAS generated for STP Foundation

Sub activity	Planned by engineers		Predicted by GA		Predicted by RL	
	Start date	End date	Start date	End date	Start date	End date
B6-1	23/04/2019	29/04/2019	23/04/2019	29/04/2019	23/04/2019	29/04/2019
Settlement tank	12/4/2019	15/04/2019	12/4/2019	15/04/2019	12/4/2019	15/04/2019
Thickener tank	16/04/2019	16/04/2019	16/04/2019	16/04/2019	16/04/2019	16/04/2019
Thickener equipment	17/04/2019	17/04/2019	17/04/2019	17/04/2019	17/04/2019	17/04/2019
units and modules (Inside building)	2/4/2019	29/04/2019	2/4/2019	29/04/2019	2/4/2019	29/04/2019
Placing Steel Bottom Matt B6-3	29/03/2019	3/4/2019	29/03/2019	3/4/2019	29/03/2019	3/4/2019
Placing Steel Bottom Matt B6-2	29/03/2019	3/4/2019	29/03/2019	3/4/2019	29/03/2019	3/4/2019
Placing Steel Bottom Matt B6-1	29/03/2019	3/4/2019	29/03/2019	3/4/2019	29/03/2019	3/4/2019
Placing Bottom Matt Steel (Outside Tanks)	1/4/2019	3/4/2019	1/4/2019	3/4/2019	1/4/2019	3/4/2019
Concrete downstand Beam	3/4/2019		3/4/2019	4/4/2019	3/4/2019	4/4/2019

Install Recess B6-3 + Levelled	3/4/2019	4/4/2019	4/4/2019	5/4/2019	4/4/2019	5/4/2019
Install Recess B6-2 + Levelled	4/4/2019	5/4/2019	5/4/2019	6/4/2019	5/4/2019	6/4/2019
Install Recess B6-1 + Levelled	4/4/2019	5/4/2019	5/4/2019	6/4/2019	5/4/2019	6/4/2019
Install Perimeter Shutters	4/4/2019	5/4/2019	5/4/2019	6/4/2019	5/4/2019	6/4/2019
Concrete Curing (To Strike Shutters)	5/4/2019	5/4/2019	6/4/2019	6/4/2019	6/4/2019	6/4/2019
Placing Steel Top Matt B6-3	8/4/2019	10/4/2019	8/4/2019	10/4/2019	8/4/2019	10/4/2019
Placing Steel Top Matt B6-2	8/4/2019	10/4/2019	8/4/2019	10/4/2019	8/4/2019	10/4/2019
Placing Steel Top Matt B6-1	8/4/2019	10/4/2019	8/4/2019	10/4/2019	8/4/2019	10/4/2019
Strike Shutters (All bases and Perimeter)	8/4/2019	9/4/2019	8/4/2019	9/4/2019	8/4/2019	9/4/2019
Steel delivery	8/4/2019	8/4/2019	8/4/2019	8/4/2019	8/4/2019	8/4/2019
Install Screed Rails + Levels	9/4/2019	9/4/2019	9/4/2019	9/4/2019	9/4/2019	9/4/2019
Prep area and blind southwest near siltbuster	10/4/2019	10/4/2019	10/4/2019	10/4/2019	10/4/2019	10/4/2019
Pour Watertight Concrete in Recess (Following Tank Install)	10/4/2019	10/4/2019	10/4/2019	10/4/2019	10/4/2019	10/4/2019
Placing Top Matt Steel (Outside Tanks)	11/4/2019	12/4/2019	11/4/2019	12/4/2019	11/4/2019	12/4/2019
Install Screed Rails + Levels	15/04/2019	15/04/2019	15/04/2019	15/04/2019	15/04/2019	15/04/2019
Concrete Curing	15/04/2019	15/04/2019	15/04/2019	15/04/2019	15/04/2019	15/04/2019
Placing Steel Bottom Matt B3-2	15/04/2019	17/04/2019	15/04/2019	17/04/2019	15/04/2019	17/04/2019
Placing Steel Bottom Matt B4	15/04/2019	17/04/2019	15/04/2019	17/04/2019	15/04/2019	17/04/2019
Placing Bottom Matt Steel (Outside Tanks)	15/04/2019	17/04/2019	15/04/2019	17/04/2019	15/04/2019	17/04/2019
Excavate and surface prep for civil works (750mm Trench for recess)	15/04/2019	15/04/2019	15/04/2019	15/04/2019	15/04/2019	15/04/2019

Strike Shutters (All bases and Perimeter)	16/04/2019	26/04/2019	16/04/2019	26/04/2019	16/04/2019	26/04/2019
Prefabrication of Recess B5	16/04/2019	18/04/2019	16/04/2019	18/04/2019	16/04/2019	18/04/2019
Blinding	17/04/2019	17/04/2019	17/04/2019	17/04/2019	17/04/2019	17/04/2019
Placing Steel Top Matt B3-2	18/04/2019	24/04/2019	18/04/2019	24/04/2019	18/04/2019	24/04/2019
Placing Steel Top Matt B4	18/04/2019	24/04/2019	18/04/2019	24/04/2019	18/04/2019	24/04/2019
Placing Top Matt Steel (Outside Tanks)	18/04/2019	24/04/2019	18/04/2019	24/04/2019	18/04/2019	24/04/2019
Install Perimeter Shutters	18/04/2019	24/04/2019	18/04/2019	24/04/2019	18/04/2019	24/04/2019
Excavate WTP Temporary Slab LA	23/04/2019	26/04/2019	23/04/2019	26/04/2019	23/04/2019	26/04/2019
Steel delivery	23/04/2019	23/04/2019	23/04/2019	23/04/2019	23/04/2019	23/04/2019
Pour	23/04/2019	23/04/2019	24/04/2019	24/04/2019	24/04/2019	24/04/2019
place steel for desanding building slab	25/04/2019	3/5/2019	25/04/2019	3/5/2019	25/04/2019	3/5/2019
Concrete Curing	26/04/2019	26/04/2019	26/04/2019	26/04/2019	26/04/2019	26/04/2019
steel delivery	26/04/2019	26/04/2019	26/04/2019	26/04/2019	26/04/2019	26/04/2019
Blinding WTP Temporary Slab LA	29/04/2019	29/04/2019	29/04/2019	29/04/2019	29/04/2019	29/04/2019
Strike Shutters (All bases and Perimeter)	29/04/2019	30/04/2019	29/04/2019	30/04/2019	29/04/2019	30/04/2019
placing steel for b5 Beam	29/04/2019	29/04/2019	29/04/2019	29/04/2019	29/04/2019	29/04/2019

Appendix G: Interview protocol for method M1

Participant Information Sheet

Research Project Title:	Codifying information flow in 'construction as a manufacturing process' in BIM-enabled projects
Principal Investigator :	Ranjith K. Soman
Research Overview :	This doctoral research focuses on codifying information flow in 'construction as a manufacturing process' in BIM-enabled projects to support future automation. The focus is on determining the potential areas of improvement in information flow. For more information, please refer PhD project description document.
Funding :	This research is co-sponsored by a departmental Skempton scholarship and a grant from Bentley Systems UK.

Thank you for agreeing to be interviewed as part of the above research project. The interview is formatted as a semi-structured interview and will take approximately one hour. We don't anticipate that there are any risks associated with your participation, but you have the right to stop the interview or withdraw from the research at any time. To gain your informed consent to take part I'd like to make you aware of the following:

- Your contribution to this study is voluntary. You are under no obligation and are free to withdraw at any time without providing an explanation and without prejudice;
- With your permission, the interview will be recorded, and a transcript will be produced. You will be sent the transcript and given the opportunity to correct any factual errors;
- Access to the interview transcript will be limited to Ranjith K. Soman and academic colleagues and researchers with whom he might collaborate as part of the research process;
- Your name and any personal details will be kept confidential. Any summary interview content, or direct quotations from the interview, that are made available through academic publication or other academic outlets will be anonymized so that you cannot be identified, and care will be taken to ensure that other information in the interview that could identify yourself/your organisation/project is not revealed;
- Any variation of the conditions above will only occur with your further explicit approval;
- The anonymised content of your interview may be used in academic papers/spoken presentations/archive of the project/feedback events; and
- This study will be carried out in accordance with the Ethics Code of Imperial College London. This includes the assurance that audio recording and other data about you will be held securely, handled in accordance with the Data Protection Act 1998, and disposed of in line with Imperial's retention policy. The Ethics Code is available at https://www.imperial.ac.uk/admin-services/secretariat/secretariat/what-we-do/ethics/

and the Data Retention Policy at http://www.imperial.ac.uk/admin-services/legal-services-office/data-protection/.

Researcher's contact details

Ranjith K. Soman,
Research Postgraduate,
437, Skempton Building, Imperial College London,
South Kensington Campus,
London SW7 2AZ
Email: ranjithks17@imperial.ac.uk

The supervisor of the project is :
Prof Jennifer Whyte,
Dept. of Civil and Environmental Engineering,
436, Skempton Building, Imperial College London,
South Kensington Campus,
London SW7 2AZ
Email: j.whyte@imperial.ac.uk

INTERVIEW FOR CODIFYING INFORMATION FLOW IN 'CONSTRUCTION AS A MANUFACTURING PROCESS' IN BIM-ENABLED PROJECTS

INTERVIEW PROTOCOL

Aim: To develop and implement novel solutions to codify information flow in 'construction as a manufacturing process' in BIM-enabled projects to support future automation.
Method Check: Respondents will have an opportunity to read notes from the interview to make factual corrections, and will receive an executive summary of the final report. Individual names will not be reported in research outputs.
Interview code:
Tape recorder: [ON/ OFF]

PRELIMINARIES

Job title:
Length of time in current post:
Role in current post:
Career history of those projects that were important in developing the knowledge and skills in BIM for your current post:

Project name	Dates	Your role	Company/type of industry	What you learnt

GENERAL

1. Could you describe a typical day in your office?
   Prompt:

COMMUNICATION

2. Which 4-5 teams do you interact with on a daily basis?
   Prompt: Why/Need for this interaction?
   Prompt: Is there any other team that you interact with weekly/monthly?

	Need for Interaction	
i. ii. iii. iv. v. vi.		

3. What forms of communication do you use to communicate with these teams?

i. ii. iii. iv. v. vi. vii.		

4. What challenges do you face in communicating with these teams?
   Prompt: How frequently do you face these challenges? Did you face these in the last month?

i. ii. iii.		

BIM & COLLABORATION

5. What software tools do you use most frequently?
   Prompt: What do you use _____ for?
   Prompt: How frequently do you use it?

	Usage Frequency	
i. ii. iii. iv. v. vi.		

6. What data do you interact with on a daily basis?
   Prompt: Are these digitized?

	Usage	
i. ii. iii. iv. v. vi.		

7. While working with multiple software tools and data sources, did issues such as interoperability and incompatibilities influence your workflow?
   Prompt: if yes, can you give example?

8. What is your experience of using a common data environment (For example Asite, Projectwise, BIM360 etc)?
   Prompt: Did it influence/change your workflows?
   Prompt: Did it change how information access and transfer is done, especially when working with different teams?

9. Have you felt that the introduction of these tools and environment increase/reduce the time to query for information (clear queries/design changes/information clarification requests etc.)?
   Prompt: if yes, can you give examples

10. How much time did it take for the team to learn new tools?
    Prompt: How did it affect your productivity?
    Prompt: Was there training?

CHALLENGES

11. What are the main challenges that you faced in terms of information access/clarification of information/ communication in the past month?
    Prompt: Could you take the most recent issue you've been spending time on?
    Prompt: Is there any other issue that you would like to share?

12. If there is a software solution that would simplify your everyday challenges in terms of information flow, can you describe what it does and how it helps you?

13. Is there anything else that you would like to specify in terms of information flow?

Appendix H: Protocol for the implementation study

Testing of linked data based dynamic construction constraints checking framework to aid look ahead planning in construction

Study Management Group

Principal Investigator: Professor Jennifer Whyte

PhD Researcher: Ranjith Kuttantharappel Soman

Funder

This research is co-sponsored by a Skempton scholarship from the Department of Civil and Environmental Engineering, Imperial College London and a grant from Bentley Systems UK.

This protocol describes the study entitled 'testing of linked data based dynamic construction constraints checking framework to aid look ahead planning in construction' and provides information about procedures for entering participants. Every care was taken in its drafting, but corrections or amendments may be necessary. These will be circulated to investigators in the study. Problems relating to this study should be referred, in the first instance, to the Principal Investigator.

This study will adhere to the principles outlined in the UK Policy Frame Work for Health and Social Care Research It will be conducted in compliance with the protocol, Data Protection Act 2018 and General Data Protection Regulations (Europe) and other regulatory requirements as appropriate.

1.INTRODUCTION

An earlier research by the project investigators by examining three construction projects revealed the problem of lack of detail and semantics in a construction schedule. To address the issues of lack of detail and semantics, there exists a need to codify the relationships between tasks, resources and constraints in detail. However, the constraint relationship has inherent dynamicity and uncertainty, limiting the users from coding them using conventional BIM tools. BIM tools lack the ability to code these relationships in a dynamic nature as the relationships are complex and spanning across different domains. To embed the dynamic constraints into BIM and use it to aid look ahead planning, Soman (2019a) presented a dynamic constraint checking framework based on linked-data. A prototype was developed based on the framework to test the efficacy of the framework. The prototype was developed based on 4D planning tool so that it gets integrated into the workflow. The developed prototype was able to identify the constraint violations within the detailed schedule data on an open construction dataset. This demonstrated the potential of the frameworks to be used to look ahead planning meetings to assist in the decision-making process. However, there is a need to test the framework in a real project to identify and resolve the emerging issues when deployed the framework in a live project.

The current study is to further test this framework in a real project. By deploying the framework in a real-world construction project, the efficiency and limitations of the framework in codifying construction scheduling constraints can be determined through a structured approach.

1.1 Background

The constraints in construction processes are highly dynamic and complex with numerous interdependencies, making it difficult to be modelled using relational databases or similar informational models. Recent research has used linked-data based modelling approaches to address this issue. Linked-data enables linking of data across domains and heterogeneous sources, without being limited by the underlying schema. Linked-data addresses the problem of complexities and interdependencies in construction constraints. Although there is extensive research done on linked-data applications to address the limitations of current information modelling approaches in the constructions sector, the focus has been on the modelling of product information. Applications of linked-data to address the limitations in information models for representing process information such as constraints relationships and resource dependencies are relatively less explored. Building on the research on linked-data in the construction sector, Soman, (2019a)presented a framework for dynamic constraint checking using linked-data and Shapes Query Language (SHACL) to address this limitation. The framework is based on the modelling of construction scheduling constraints using SHACL as presented in Soman, (2019b).

In this study, Worked example method as mentioned in Clayton, et al. (1998) is used in for evaluating the framework. A worked example describes a simulated scenario to test an innovative process, to show the features of the innovative process. However, worked example method can also be used to describe real world situation and use it to demonstrate the innovative process. When testing the innovative process on a real life situation, the worked example provides stronger evidence for the validity of the research than a simulated problem. It also offers the opportunity to compare the innovative process with conventional process. In the current study, innovative process is look ahead planning using linked-data based constraint checking framework and 4D BIM tools while the conventional process uses 4D BIM tools only.

1.2 Study Rationale

The aim of the study is to test the efficacy of a framework developed to aid look ahead planning in construction. The framework uses a linked-data based constraint checking language to validate construction constraints against a set of requirements. By deploying the framework in a real-world construction project, the efficiency and limitations of the framework in codifying construction scheduling constraints can be determined through a structured approach.

2.STUDY OBJECTIVES

Objectives of this study are: -

1. To test the applicability of linked-data based constraint checking framework on a live construction project;
2. To determine the emerging issues while deploying the framework in a live construction project; and
3. Determine the limitations of linked-data based constraint checking framework on a live construction project.

3.STUDY DESIGN

Purposive sampling was used to identify a construction project, which uses 4D BIM intensively. The project was one among the finalists in the Bentley YII 2019. Snowball sampling is used to recruit participants to this study. Initially, contact was made with the digital engineer of the construction project. Based on the discussions with the digital engineer, remaining participants who would be included in the weekly meetings were added to the study.

PI would be a silent observer in the weekly 4D BIM review meetings (at least 4 meetings) of the project. PI would capture the decisions regarding construction scheduling constraints and site constraints as meeting notes. 4D BIM models, method statements and construction programme would be requested for analysis. PI would then import the dataset collected to the developed framework. Framework will be used to generate look ahead plan for the week. The generated plan will be compared with the plan created by the engineers in the project. Variances between generated plan and the one created by engineers will be studied based on metric such as time efficiency, cost, resource distribution, and constraint violation. The generated plan and the observed variances will be presented to the engineers in projects and feedback on the limitations of the framework will be collected from the engineers.

The study would be conducted from 2/12/2019 to 15/01/2019. Eight participants would be involved in this study.

4.PARTICIPANT RECRUITMENT

4.1 Pre-recruitment evaluations

There are no pre-recruitment evaluations for this study.

4.2 Inclusion Criteria

All the participants of the study should be involved in the construction stage of the selected project. These include project engineers, construction site supervisors, digital engineers responsible for maintaining digital models, and construction planners. They should attend at least one of the 4D project plan/review meeting on which the PI is a silent observer.

4.3 Exclusion Criteria

People who are attending 4D project meeting however not associated with the construction stage of the project are excluded.

4.4 Withdrawal Criteria

To withdraw from the research, participant needs to inform the PI of their decision to withdraw. PI would then remove the data that relates to the corresponding participant from the dataset.

5.ADVERSE EVENTS

All adverse events should be reported. Any questions concerning adverse event reporting should be directed to the Principal Investigator in the first instance.

5.1.1 Non serious AEs

All such events, whether expected or not, should be recorded.

5.1.2 Serious AEs

An SAE form should be completed and emailed to the Principal Investigator within 24 hours.

All SAEs should be reported to the Ethics and Research Governance Coordinator where in the opinion of the Principal Investigator, the event was:

- 'related', i.e. resulted from the administration of any of the research procedures; and
- 'unexpected', i.e. an event that is not listed in the protocol as an expected occurrence

Local investigators should report any SAEs as required by their Local Research Ethics Committee, Sponsor and/or Research & Development Office.

6.ASSESMENT AND FOLLOW UP

The findings will be communicated with all the participants of the study. Their feedback would be used for further improving the framework.

7. REGULATORY ISSUES

7.1 Ethics approval

The Principal Investigator has obtained approval from the Head of Department and Joint Research Compliance Office (JRCO). The study will be conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

7.2 Consent

Consent to enter the study must be sought from each participant only after a full explanation has been given, an information leaflet offered and time allowed for consideration. Signed participant consent should be obtained. The right of the participant to refuse to participate without giving reasons must be respected. All participants are free to withdraw at any time.

7.3 Confidentiality

The Principal Investigator will preserve the confidentiality of participants taking part in the study and fulfil transparency requirements under the General Data Protection Regulation for health and care research. Data and all appropriate documentation will be stored for a minimum of 10 years after the completion of the study, including the follow-up period.

7.4 Indemnity

Imperial College London holds negligent harm insurance policies which apply to this study.

7.5 Sponsor

Imperial College London will act as the main Sponsor for this study

7.6 Funding

This research is co-sponsored by a Skempton scholarship from the Department of Civil and Environmental Engineering, Imperial College London and a grant from Bentley Systems UK.

7.7 Audits

The study may be subject to inspection and audit by Imperial College London under their remit as sponsor.

8.PUBLICATION POLICY

Any content captured during this research study, that are made available through academic publication or other academic outlets will be anonymized so that participants cannot be identified, and care will be taken to ensure that other information in the data that could identify participant or the project participant is associated with is not revealed.

Appendix I: Extended research summary

The PhD research aims to develop an information modelling approach to codify and validate detailed construction process information in BIM-enabled projects to support decision-making for look-ahead planning in construction. This aim was achieved in the PhD study by individually attaining the three objectives (set out in Chapter 1) following the research methods stated in Chapter 3.

Objective 1 (to understand the codification challenges in construction) was achieved through an interpretive case study (Eisenhardt & Graebner, 2007), by studying three BIM-enabled construction projects. Findings from the study highlighted the codification challenges as 1) *software usage*—interoperability, information loss during conversion, and multiple modelling techniques, 2) *information sharing*—unstructured information sharing, drawing and file-based sharing document control, and lack of process change, and 3) *missing process information*—loss of constraints and low levels of detail. These challenges were mapped to data quality dimensions such as *accuracy, completeness, temporality, consistency, accessibility, and data provenance* (as described in Batini and Scannapieco (2016)) to understand their implications. By identifying the codification challenges in the late design and construction phase of the projects and their mapping to the data quality dimensions, our study extends the knowledge on data quality issues in construction. Although there are previous studies that looked at the reasons for successful implementation of BIM (leading to structured and machine-readable data) (Kassem & Succar, 2017; Chegu Badrinath & Hsieh, 2019; Won *et al.*, 2013), they did not investigate the reasons for data quality issues. Also, existing research on data quality issues in BIM investigated the issue from a model-based perspective (Lee, Solihin & Eastman, 2019; Zadeh *et al.*, 2017; Donato, Lo Turco & Bocconcino, 2017). However, existing research ignored the issues emerging from the broader practice of using

mode-based and document-based information sharing practices. The current study addressed this limitation and extended the understanding of the data quality issues arising from construction information use. Findings from this objective reinforce the need for a novel method for modelling detailed construction process information such that the data quality issues are minimized, and machine-readability is increased.

Objective 2 to develop the information modelling approach (to address the technical problems in the construction information identified in O1) was achieved by developing a novel information modelling approach using linked-data principles. The LDCC method, to define and check complex and dynamic scheduling constraints in construction, was developed and tested in this study. In the LDCC method, construction process information is published as an RDF graph to achieve semantic interoperability even if construction information is distributed among multiple domains. Semantic interoperability is achieved through the use of ontologies to define taxonomies and services to enable sharing and reusing domain knowledge (Davies, Fensel & Van Harmelen, 2003; El-Gohary & El-Diraby, 2010). To validate and detect violations in the construction process information modelled as RDF graphs, SHACL was used. The LDCC method ensures that that semantic interoperability is retained, enabling constraint-checking systems to infer from multiple databases without human intervention. Section 5.2 describes the LDCC method, its architecture, ontology, and how scheduling constraints are modelled using SHACL. To test the efficacy of the developed method, it was tested using a public dataset to detect constraint violations (described in Section 5.3) and on real-world construction information (with the results of the study presented in Section 7.4). The results from the testing give evidence for the applicability of the LDCC method to detect constraint violation in construction information. This objective addresses the limitation of existing information modelling approaches (Darwiche, Levitt & Hayes-Roth, 1988; Thabet & Beliveau, 1997; Akinci *et al.*, 2002b; Dong *et al.*, 2013; Marshall, Barber & Boardman, 1987) which

have not described methods to model construction constraints. Complex construction constraints are not modelled as they are highly dynamic and span multiple domains. In addition, this method extends the knowledge on BIM-based rule-checking systems to consider the siloed nature of construction information, which was not handled by existing BIM-based rule-checking systems (Schwabe, Teizer & König, 2019; Lee, Solihin & Eastman, 2019; Solihin *et al.*, 2017).

Objective 3 was to reuse the construction process information codified using the LDCC method to generate a LAS. This objective was achieved by developing a method to solve the RCPSP problem using a genetic algorithm and reinforcement learning. Both of them used the LDCC method developed in O2 to detect constraint violation to generate LAS. The performance of the algorithms was compared using a computational experiment in Section 6.5. Unlike previous studies to solve for RCPSP (Bruni, Beraldi & Guerriero, 2015; Ziarati, Akbari & Zeighami, 2011; Pellerin, Perrier & Berthaut, 2020), our study incorporates the complex construction constraints distributed over multiple domains to be checked for while generating LAS. This study extends the knowledge on solving RCPSP to include complex construction constraints in addition to the precedence and discrete resource constraints. The developed algorithms were tested on real construction data, and the results were presented in Section 7.4. The results from the study demonstrated the potential for the LDCC method to be incorporated with metaheuristics to generate LAS.

Appendix J: Examples of SHACL Snippets for scheduling constraints

The modelling of constraints in this section is based on a simplified lifting process as shown in Figure J.1. There are three classes modelled for representing this process as shown in

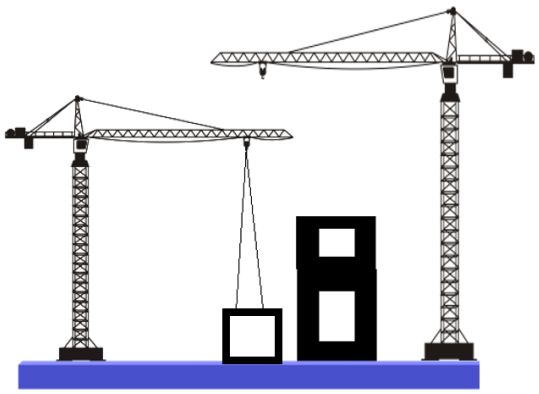


Figure J.35: Lifting process

Figure J.2. The classes have attributes which are data properties and object properties. Data properties are the attributes giving the information such as start date, the weight of module etc. Object properties are the attributes which define the relationship of the object with another object.

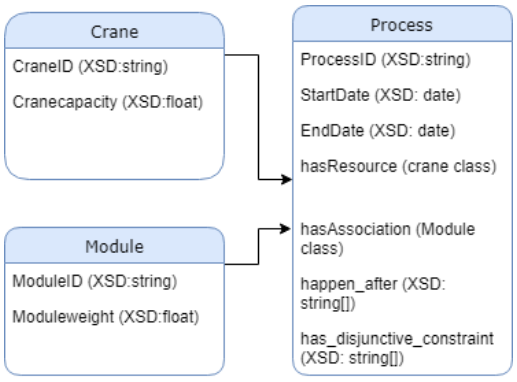


Figure J.34: Class diagram for lifting process

1. Process: It is the parent class and instances of this class stores information related to the task including the start and end dates as data properties, resources, module

association, precedence and disjunctive constraints as object properties.

2. Crane: Instances of this class stores the information related to the crane. For the current paper, the scope of this class is limited to oneattribute crane capacity modelled as a data property, in order to demonstrate the logical constraint.
3. Module: Instances of this class stores information related to the module. Module weight is the only attribute data property used.

Even though object relationships are not explicitly defined in module and crane classes, the inherent linking in the OWL applies the relationship defined in the process class to the crane and module classes.

J.1 Logical constraints

In the current problem setting, there is a simple logical constraint which is to be satisfied while scheduling. The Lifting capacity of the crane should always be higher than the weight of the module. The constraint is modelled as shown below.

```
OntProcess:Process
  rdf:type rdfs:Class ;
  rdf:type sh:NodeShape ;
  rdfs:subClassOf owl:Class ;
  sh:sparql [
     sh:message "Crane capacity should be greater than module weight" ;
     sh:prefixes <http://semanticprocess.x10host.com/Ontology/OntProcess> ;
     sh:select """SELECT $this
WHERE {
     $this  rdf:type OntProcess:Process.
     $this OntProcess:hasResource ?crane.
     $this OntProcess:hasAssociation ?module.
```

```
    ?crane OntProcess:Cranecapacity ?cc.
    ?module OntProcess:Moduleweight ?mw.
      FILTER (?cc <= ?mw).
}""" ;
  ]
```

The constraint is applied to the class Process as a node shape as it related to both module and the crane. Therefore, during validation, all the instances of 'OntProcess:Process' (OntProcess is the prefix defined in the ontology for the IRI), which has crane as a resource and module as an association are checked for this logical constraint. Instances of'OntProcess:Process' which do not have crane and module assigned to it will be ignored from this check. Targeting the instances which have assignment is achieved by using SHACL-SPARQL constraint. The SELECT query is used to access the attributes of instances of Crane and Module class. From all the process instances with the module and crane assignmentThe *FILTER (?cc <= ?mw).* statement filters the instances where crane capacity is less than module weight. In the case of violation of the constraint, the compiler would provide a message to the user as defined in the sh:message which in the current case is Crane capacity should be higher than module weight".

J.2 Precedence constraints

Precedence constraints may be used in the current problem to define which processes have to be completed so that a new process could take place. Therefore, precedence is an object property defined on the processes. The constraint is mathematically defined as:

$$Start\ date\ of\ current\ activity \leq End\ date\ of\ preceding\ activity$$

```
OntProcess:Process
  rdf:type rdfs:Class ;
```

```
  rdf:type sh:NodeShape ;
  rdfs:subClassOf owl:Class ;
sh:sparql [
    sh:message "Precedence condition is violated" ;
    sh:prefixes <http://semanticprocess.x10host.com/Ontology/OntProcess> ;
    sh:select """SELECT $this
WHERE {
$this  rdf:type OntProcess:Process.
$this owl:happen_after ?process2.                    $this owl:hasStartDate ?sd1.
        ?process2 owl:hasEndDate ?ed2.
FILTER( ?sd1 <?ed2)
}
```

The above code defines the precedence constraint in SHACL-SPARQL. During validation, the compiler searches for all the instances of class Process with a property happen after as represented in the statement *$this owl:happen_after ?process2* . The start dates of the current process (*$this*) are compared with *?process2* to check whether the precedence constraint is satisfied or violated. If it is violated, the user receives a message as defined in *sh:message.* All the instances of the class Process without an *owl:happen_after* property is ignored. If there are multiple precedence assignments, all the processes are checked for constraint violations.

J.3 Disjunctive constraints

Disjunctive constraints may be defined in this problem setting to make sure that one process shouldn't take place while another process is happening. There are five cases in which this might happen. Consider 2 processes, *Process1* and *Process 2* . The five cases where these two processes overlap are:-

i. *SD1<SD2<ED1<ED2;*

ii. *SD1<SD2<ED2<ED1;*

iii. *SD2<SD1<ED1<ED2; and*

iv. *SD2<SD1<ED2<ED1.*

v. *SD1=SD2 && ED1=ED2*

where *SD1*- start date of *Process 1, SD2* - start date of *Process 2, ED1* – end date of *Process 1* and *ED2* – end date of *Process 2.* This constraint can be represented as the following constraint.

(SD2<ED1 && SD1<ED2)||(SD1=SD2 && ED1=ED2)

This is modelled in the SHACL as follows.

```
OntProcess:Process
  rdf:type rdfs:Class ;
  rdf:type sh:NodeShape ;
  rdfs:subClassOf owl:Class ;
sh:sparql [
     sh:message "Disjunctive constraint is violated " ;
     sh:prefixes <http://semanticprocess.x10host.com/Ontology/OntProcess> ;
     sh:select """SELECT $this
  WHERE {
$this  rdf:type OntProcess:Process.
$this owl:has_disjunctive_constraint ?process2.
$this owl:hasEndDate ?ed1.
$this owl:hasStartDate ?sd1.
?process2 owl:hasEndDate ?ed2.
?process2 owl:hasStartDate ?sd2.

  FILTER((?sd2<?ed1&&?sd1<?ed2)||(?sd1=?sd2 && ?ed1=?ed2))
  }""" ;
    ]
```

As given in the previous definitions, the constraint is modelled as a node shape on the Process class. The processes which are disjunctive to the current is assigned to the current process using the *has_disjunctive_constraint* property. All the instances of this process with this assignment are checked for the violation using the above constraint representation inside the SPARQL query as shown in (*FILTER((?sd2<?ed1&&?sd1 <?ed2)||(?sd1=?sd2 && ?ed1=?ed2)))*. As before, the instances without this assignment are ignored.

J.4 Discrete constraint

Discrete constraints are required in this problems setting to ensure resources (cranes) are not assigned to two processes at the same time. Since this is a constraint related to the resource, the constraint is modelled as a node shape on the class *Crane* as given below.

Pairwise comparison feature of SPARQL is used to model this constraint. As a first step, all the processes which are assigned the particular resource is shortlisted into pairs in the statements:

?process1 OntProcess:hasResource $this.

?process2 OntProcess:hasResource $this.

From this list of pairs, identical processes are removed using *FILTER (!(?process1=?process2)).*

Then the data properties start date and end date for each process in the pair is retrieved and checked for the disjunctive constraint in the statement *FILTER((?sd2<?ed1&&?sd1<?ed2)||(?sd1=?sd2 && ?ed1=?ed2))*

```
OntProcess:Crane
  rdf:type rdfs:Class ;
  rdf:type sh:NodeShape ;
  rdfs:subClassOf owl:Class ;
  sh:sparql [
     sh:message "Discrete constrsint for the resource not satisfied" ;
     sh:prefixes <http://semanticprocess.x10host.com/Ontology/OntProcess> ;
     sh:select """SELECT $this
WHERE {
 $this rdf:type OntProcess:Crane .
?process1 rdf:type OntProcess:Process.
?process2 rdf:type OntProcess:Process.
?process1 OntProcess:hasResource  $this.
?process2 OntProcess:hasResource  $this.
```

```
?process1 owl:hasEndDate ?ed1.
?process1 owl:hasStartDate ?sd1.
?process2 owl:hasEndDate ?ed2.
?process2 owl:hasStartDate ?sd2.

FILTER (!(?process1=?process2))
FILTER((?sd2<?ed1&&?sd1<?ed2)||(?sd1=?sd2 && ?ed1=?ed2) )
}""" ;
   ] ;
.
```

The discrete constraint is necessary for all the resources. In case, the resources can have multiple assignments; this can also be defined in this constraint using SPARQL COUNT query.

Implementation

The above process was implemented using open source Python Libraries RDFLib and pySHACL in a Jupyter Notebook environment. The ontology for the process was developed using Protégé(Musen, 2015). The instances used for testing are shown in Table J.1. The results of the same are presented in the next section.

Table J.0.2: Instances of the process class

Process	Crane	Crane capacity (metric ton)	Module	Module Weight (metric ton)	Start date	End date	Precedence Constraint	Disjunctive constraint
Process_1	Crane_1	12	Module_1	15	01/01/2019	02/01/2019		
Process_2	Crane_2	18	Module_2	15	02/01/2019	03/01/2019	Process 1	
Process_3	Crane_2	18	Module_3	15	01/01/2019	05/01/2019	Process 1	Process 1,Process 2

Results and discussion

The results after the validation of the data graph against shapes graph are given in Table J.2.

From the results in Table J.2, we can identify that there are constraint violations in Process 1, Process 3 and Crane 2.

In *Process_1*, the logical constraint was violated, and the message *"Crane capacity should be greater than module weight"* was received as an output. This is an actual violation as *Crane_1* with a capacity of 12 metric ton was assigned to *Process_1* to lift *Module_1* which has a mass of 15 metric ton.

In *Process_3,* there are two violations which were reported in the messages "*Disjunctive constraint is violated"* and *"Precedence condition is violated"*. *Process_3* was assigned a disjunctive constraint with *Process_1* and *Process_2.* However, *Process_3* starts with *Process_1*

Table J.2: Output after constraint validation

Process	Output
Process 1	Constraint Violation in SPARQLConstraintComponent (http://www.w3.org/ns/shacl#SPARQLConstraintComponent): Severity: sh:Violation Source Shape: OntProcess:Process Focus Node: OntProcess:Process_1 Value Node: OntProcess:Process_1 Source Constraint: [sh:message Literal("Crane capacity should be greater than module weight") ; sh:prefixes <http://semanticprocess.x10host.com/Ontology/OntProcess> ; sh:select Literal("SELECT $this WHERE { $this rdf:type OntProcess:Process. $this OntProcess:hasResource ?crane. $this OntProcess:hasAssociation ?module. ?crane OntProcess:Cranecapacity ?cc. ?module OntProcess:Moduleweight ?mw. FILTER (?cc <= ?mw).}")] Message: **Crane capacity should be greater than module weight**
Process 2	No Violation
Process 3	Constraint Violation in SPARQLConstraintComponent (http://www.w3.org/ns/shacl#SPARQLConstraintComponent): Severity: sh:Violation Source Shape: OntProcess:Process Focus Node: OntProcess:Process_3 Value Node: OntProcess:Process_3 Source Constraint: [sh:message Literal("Precedence condition is violated") ; sh:prefixes <http://semanticprocess.x10host.com/Ontology/OntProcess> ; sh:select Literal("SELECT $this WHERE { $this rdf:type OntProcess:Process. $this owl:happen_after ?process2. $this owl:hasEndDate ?ed1. $this owl:hasStartDate ?sd1. ?process2 owl:hasEndDate ?ed2. ?process2 owl:hasStartDate ?sd2. FILTER(?sd1 <?ed2) }")] **Message: Precedence condition is violated** Constraint Violation in SPARQLConstraintComponent (http://www.w3.org/ns/shacl#SPARQLConstraintComponent): Severity: sh:Violation Source Shape: OntProcess:Process Focus Node: OntProcess:Process_3 Value Node: OntProcess:Process_3 Source Constraint: [sh:message Literal("Disjunctive constraint is violated ") ; sh:prefixes

	<http://semanticprocess.x10host.com/Ontology/OntProcess> ; sh:select Literal("SELECT $this WHERE { $this rdf:type OntProcess:Process. $this owl:has_disjunctive_constraint ?process2. $this owl:hasEndDate ?ed1. $this owl:hasStartDate ?sd1. ?process2 owl:hasEndDate ?ed2. ?process2 owl:hasStartDate ?sd2. FILTER((?sd2<?ed1&&?sd1<?ed2)||(?sd1=?sd2 && ?ed1=?ed2)) }")] **Message: Disjunctive constraint is violated**
Crane 1	No violation
Crane 2	Constraint Violation in SPARQLConstraintComponent (http://www.w3.org/ns/shacl#SPARQLConstraintComponent): Severity: sh:Violation Source Shape: OntProcess:Crane Focus Node: OntProcess:Crane_2 Value Node: OntProcess:Crane_2 Source Constraint: [sh:message Literal("Discrete constrsint for the resource not satisfied") ; sh:prefixes <http://semanticprocess.x10host.com/Ontology/OntProcess> ; sh:select Literal("SELECT $this WHERE { $this rdf:type OntProcess:Crane . ?process1 rdf:type OntProcess:Process. ?process2 rdf:type OntProcess:Process. ?process1 OntProcess:hasResource $this. ?process2 OntProcess:hasResource $this. ?process1 owl:hasEndDate ?ed1. ?process1 owl:hasStartDate ?sd1. ?process2 owl:hasEndDate ?ed2. ?process2 owl:hasStartDate ?sd2. FILTER (!(?process1=?process2)) FILTER((?sd2<?ed1&&?sd1<?ed2)||(?sd1=?sd2 && ?ed1=?ed2)) }")] **Message: Discrete constraint for the resource not satisfied**

and ends after *Process_2* which is a clear violation of the disjunctive constraint. In addition, *Process_3* was assigned a precedence constraint with *Process_1*. This is again violated as the start times of both processes are the same.

For *Crane_2, "Discrete constraint for the resource not satisfied"* was received as an output. This is true as the *Crane_1* was assigned to two processes *Process_2* and *Process_3* at the same time.

Hence from the above, we can conclude that the modelled constraints were able to detect the violations effectively. Although the model represented here is very simplistic, and the violations could be easily detected without any modelling, the same method can be applied to more massive process datasets where it would be tedious to identify the violations.

It is to be noted that the data (RDF containing instances) and the shapes (RDF containing constraints) are different components linked through the ontology of the lifting process. More constraints can be added dynamically to the shapes file, and the validation is done for the whole data. This is an excellent advantage in construction to incorporate new constraints as they arise.

Also, as the constraints are coded in a machine-readable format (RDF), it is possible to do analytics and machine learning on this model too and derive insights for future automation of constraint modelling.

Appendix K: Comparison of 4D tools

Table K.1: Comparison of existing 4D tools

Tool	Create 4D model easily	Insert Temporary Equipment	Ease of learning	Automating workflows	Model rendering quality	Interoperability	Dealing with schedule updates	Planning / scheduling Capability	Editing 3D geometry	Handling large models	Sharing models	Advanced scenario analysis	Embedding constraints
Alice Technologies	Good capability	Some capability	Good capability	Very good capability	Some capability	Limited capability	Good capability	Good capability	Limited capability	Good capability	Very good capability	Very good capability	Some capability
Asta PowerProject BIM	Good capability	Some capability	Good capability	Good capability	Limited capability	Limited capability	Good capability	Very good capability	Some capability	Good capability	Good capability	Good capability	Limited capability
Bexel Manager	Good capability	Some capability	Good capability	Good capability	Some capability	Some capability	Good capability	Good capability	Some capability	Good capability	Some capability	Good capability	Limited capability
Desite BIM	Some capability	Some capability	Some capability	Good capability	Some capability	Some capability	Some capability	Some capability	Some capability	Good capability	Some capability	Some capability	Limited capability
Deswik	Some capability	Some capability	Some capability	Good capability	Some capability	Limited capability	Good capability	Good capability	Some capability	Good capability	Some capability	Good capability	Limited capability
Fordify	Some capability	Limited capability	Some capability	Good capability	Limited capability	No capability	Some capability	Limited capability	Limited capability	Some capability	Limited capability	Limited capability	Limited capability
Fuzor	Good capability	Very good capability	Good capability	Good capability	Very good capability	Limited capability	Good capability	Some capability	Good capability	Good capability	Good capability	Some capability	Limited capability

Navisworks	Good capability	Some capability	Good capability	Some capability	Some capability	Some capability	Good capability	Limited capability	Limited capability	Good capability	Good capability	Limited capability	Limited capability
Synchro	Good capability	Good capability	Good capability	Good capability	Good capability	Good capability	Very good capability	Very good capability	Good capability	Good capability	Good capability	Good capability	Limited capability
RIB iTWO	Good capability	Limited capability	Good capability	Some capability	Some capability	Some capability	Very good capability	Good capability	Limited capability	Good capability	Good capability	Some capability	Limited capability

Appendix L: Solving LAS generation problem using CPLEX

L.1 Algorithm

LAS generation problem was attempted to be solved using IBM ILOG CPLEX Optimization Studio. Since, LAS generation problem required embedding of constraints in which cannot be modelled directly inside CPLEX studio (such as disjunctive and Logical constraint), constraint programming modelling for python feature of IBM Decision Optimization CPLEX Modelling for Python. The algorithm used for solving LAS using CPLEX is as follows.

1. Parse the inputs from the detailed process graph into CPLEX format.
2. Extract durations for each task.
3. Extract resource demand for each task.
4. Extract successors for each task.
5. Create CPLEX model.
6. Create task interval variables.
7. Create precedence constraints.
8. Create constraint capacity constraints.
9. Create disjunctive capacity constraints.
10. Minimize end time for all tasks.
11. Solve the model.
12. Show results.

L.2 Experiment

The algorithms were coded in python and executed on a workstation PC with the configuration shown in Table 6.1

Table L.3: Details of the computer in which the experiments were run

Category	Details
System model	HP Z440 Workstation
Processor	Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz (16 CPUs)
Available OS memory	32GB RAM

Scenarios tested are and the processing times for the same are given in Table L.2.

Table L.2: Scenarios and processing times

Scenario	Number of activities	Number of resource type	Number of precedence constraints	Number of disjunctive constraints	Processing times
S1	35	4	10	10	1
S2	35	8	50	50	3
S3	35	12	100	100	5
S4	35	16	200	200	NA
S5	70	4	20	10	2
S6	70	8	100	50	6
S7	70	12	300	100	8
S8	70	16	700	200	NA
S9	105	4	100	10	4
S10	105	8	300	50	8
S11	105	12	800	100	NA
S12	105	16	1000	200	NA
S13	140	4	300	10	5
S14	140	8	500	50	10
S15	140	12	900	100	NA
S16	140	16	1000	200	NA

L.3 Observations

- There are no provisions for modelling logical constraints as the optimization model deals with schedule data only.
- Algorithm works well for simpler cases with lesser number of constraints and produces results within reasonable time.

- Algorithm fails to produce results for complex problems where there are more activities and higher number of constraints.

L.3 Observations

This method cannot be used for LAS problems which is characterised by higher number of constraints. Further, logical constraint, which is a critical constraint cannot be embedded while using this optimisation method.