Smart Compliance Checking Frameworks for BIM Standards

XIAOFENG ZHU

School of Engineering

January 29, 2024

A thesis submitted to Cardiff University for the degree of

Doctor of Philosophy

Abstract

As a process of generating and managing the digital representation of a built asset, the concept of Building information modelling (BIM) has been gradually accepted and standards for implementing BIM have also been released. Compliance checking for BIM standards plays a significant role in the AEC (Architectural Engineering and Construction) industry regarding ensuring quality, safety, and efficiency during the delivery of projects. As the concept and technology of BIM are accepted and popularised, BIM standards have undergone rapid development in recent years and gradually shaped comprehensive standard systems, which makes fast and accurate BIM compliance checking increasingly challenging. Although considerable effort has been made to address this issue over the past decades, existing checking methods are flawed in terms of granularity, comprehensiveness, and automation. Moreover, these methods suffer from systemic deficiencies that make them incapable of fulfilling the demand of checking compliance against complex BIM standard systems under a rapidly evolving environment.

To address the above gaps, a novel framework for BIM compliance checking is developed in this thesis, which is comprised of an ontology-driven subjective checking framework and an evidence-driven objective checking framework. The ontological checking framework adopts a domain ontology as the knowledge model and bridges users and domain knowledge through an interactive web-based service, which enables comprehensive and flexible BIM compliance checking. The evidence-driven checking framework leverages advanced NLP (Natural Language Processing) techniques, large language models, and graph learning to automictically extract information from regulations and project documents and convert it into knowledge graphs, then assess compliance via graph alignment.

The main outcomes of the research lie in the development of the ontological checking framework and the automatic checking framework. The ontological checking framework divides domain knowledge and compliance assessment into two separate components enabling flexible checking for multiple scenarios. The adoption of ontology allows the accumulation and integration of domain knowledge. The evidence-driven checking framework has successfully implemented fully automated BIM

compliance checking. This research significantly contributes to information extraction, knowledge engineering and BIM compliance checking in the AEC domain. The proposed ontological framework outperforms existing methods in terms of granularity and comprehensiveness and enables flexible compliance checking on various scenarios. As the first attempt, the evidence-driven checking framework fills the gap in fully automated BIM compliance checking. Due to the generic development principles adopted in this research, the proposed method and developed framework can be further extended for other research areas related to information extraction, knowledge extraction and compliance checking.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Professor Haijiang Li, for his valuable opinions and suggestions on my thesis. Without his meticulous supervision, I would not have had the opportunity to pursue this research. I feel fortunate and thankful to have such a patient and understanding supervisor. His continuous advice and guidance have been instrumental in steering this research in the right direction, and he has supported me in maintaining a focus on the broader goals for the future.

In addition, I must express my gratitude to the Engineering School of Cardiff University and the industrial partner (CCCC), who sponsored me to implement this research. Particular thanks must be given to our industrial partner (CCCC Second Harbor Consultant Co., Ltd.) who provided extensive research material and valuable feedback for the systems developed in this research.

Moreover, I want to extend my appreciation to all my colleagues at the Computational Mechanics and Engineering AI research group, including all academics and fellow PhD students, for your encouragement and inspiration from our discussions. Special thanks to Dr Yi An, Dr Honghong Song, and Lu Chen for their invaluable help on the technical route design, ontology development and key algorithm development.

I am deeply grateful to my family for their support and encouragement. They will always be my strong backing. This achievement would not have been possible without their constant assistance, valuable insights, and unwavering encouragement.

Eventually, I want to say thank you to all the incredible friends I met throughout my PhD journey, both those who remained in Cardiff and those who moved on. I feel privileged to have shared this experience with all of you. These four years have evolved into a remarkable life journey, encompassing happiness and sorrow, obstacles and success. These impressive experiences will be the motivation, ammunition, and ink to write the next chapter of my life.

Again, I appreciate all those who have supported my research.

Contents

List of Figures

List of Tables

List of Abbreviations

ACCC = Automated Code Compliance Checking

AEC = Architectural Engineering and Construction

AECOO = Architect, Engineer, Contractor, Owner, and Operator

AHP = Analytic Hierarchy Process

AI = Artificial Intelligence

AIM = Asset Information Model

ANN = Artificial Neural Network

BCF = BIM Collaboration Format

BEA = Building Energy Analysis

BEP = BIM Execution Plan

BERT = Bidirectional Encoder Representations from Transformers

BIAS = Building Industrialization Association of Shenzhen

BIM = Building Information Modelling

BIM-CAREM = BIM Capability Assessment REference Model

BMCHUD = Beijing Municipal Commission of Housing and Urban-Rural Development

BPie = Building Programming Information Exchange

BPMN = Business Process Modelling Notation

BRE = Building Research Establishment

BSI = British Standards Institution

CAAC = Civil Aviation Administration of China

CAD = Computer-aided Design

CAGR = Compound Average Growth Rate

CBIMS = Chinese BIM Standard System

CCCC = China Communications Construction Company Ltd.

CDE = Common Data Environment

CFG = Context-free Grammar

CI = Consistency Index

CMM = Capability Maturity Model

CNN = Convolutional Neural Network

COBie = Construction Operations Building Information Exchange

CR = Consistency Ratio

CRBA = China Railway BIM Alliance

CRF = Conditional Random Fields

CTESI = CCCC Second Harbor Consultant Co.,Ltd.

DG = Dependency Grammar

DL = Deep Learning

DP = Dependency Parsing

DSR = Design Science Research

DT = Decision Trees

EIR = Exchange Information Requirement

ER = Exchange Requirement

GAN = Generative Adversarial Network

GCN = Graph Convolutional Network

GHUCD = Guangdong Housing and Urban-Rural Construction Department

GNN = Graph Neural Network

GPBDA = Guizhou Province BIM Development Alliance

GSA = General Services Administration

HMM = Hidden Markov Models

HNUCD = Henan Urban-Rural Construction Department

HR = Human Resources

HVACie = HVAC (heating, ventilation, and air conditioning) Information Exchange

IAI = International Alliance of Interoperability

IBC = International Building Code

IDF = Inverse Document Frequency

IDM = Information Delivery Manual

IFC = Industry Foundation Class

IFD = International Framework of Dictionaries

ISO = International Organization for Standardization

KE = Knowledge Engineering

KPI = Key Performance Indicator

LLM = Large Language Model

LOD = Level of Detail

LOIN = Level of Information Need

LSTM = Long Short-Term Memory

MEP = Mechanical, Electrical, and Plumbing

ML = Machine Learning

MOHURD = Ministry of Housing and Urban-Rural Development (China)

MOT = Ministry of Transport (China)

MVD = Model View Definition

NBIMS = National BIM Standard

NIBS = National Institute of Building Sciences

NLP = Natural Language Processing

OIR = Organisational Information Requirement

PAS = Publicly Available Specification

PIM = Project Information Models

POS = Part of Speech

PSG = Phrase Structure Grammar

QTO = Quantity Take-off

RI = Random Consistency Index

RIBA = Royal Institute of British Architects

RM = Responsibility Matrix

RNN = Recurrent Neural Networks

ROC = Receiver Operating Characteristic

SHUCC = Shanghai Housing and Urban-Rural Construction Commission

SMCHURD = Shanghai Municipal Commission of Housing and Urban-Rural Development

SPARKie = Electrical System Information Exchange

SPV = Spatial Program Validation

SQWRL = Semantic Query-Enhanced Web Rule Language

SVM = Support Vector Machines

TF = Term Frequency

TIDP = Task Information Delivery Plan

WSie = Water Systems Information Exchange

XML = Extensible Markup Language

XSD = XML Schema Definition

Chapter 1. Introduction

1.1 Problem statements

As a data-based tool applied to engineering design, construction, and management, BIM (Building Information modelling) plays an important role in improving productivity, saving costs, and shortening schedules. After the concept of BIM was first introduced by Prof. Charles Eastman in the 1970s (Eastman et al. 1974), BIM-related research and technology have developed rapidly and been widely applied to various engineering projects in the AEC industry. According to China's National Bureau of Statistics, the market size of China's BIM industry increased from 4.05 billion RMB (2016) to 11.91 billion RMB (2020), with a CAGR (Compound Average Growth Rate) of 31.0%. Considering the market maintains a stable growth trend, the size of the BIM market is expected to reach 67.0 billion RMB by 2025. Accompanied by the popularity of BIM technology, the problem of uneven effectiveness of BIM projects has begun to appear. To ensure the project quality and standardise the workflow, plenty of international and national BIM standards have been released in recent years by different countries and organisations, such as the BSI series, ISO 19650 series, NBIMS series, etc. The surge of standards has spawned studies on compliance checking against BIM standards.

Compliance checking is the process of reviewing and analysis of actions to check whether the implemented actions and their output meet the requirements documented in the specific standards or regulations (Watson and Jones 2013). For BIM projects, compliance checking primarily focuses on the data schema, the information management process, the BIM model, deliverables, etc. With regards to the above checking objectives, several academics and industrial organisations have proposed some assessment methods, including the Interactive Capability Maturity Model (I-CMM), BIM Maturity Measure (Arup 2015b), BIM4VET (LIST 2016) and so on. However, all existing checking approaches are fully manual or semi-automated, which is labour-intensive, costly, error-prone and time-consuming (Hackitt 2017). Worse still, with the ever-increasing number of BIM standards, the scope of standards is expanding, and the interconnections between standards are becoming more complex. This makes it increasingly difficult to conduct comprehensive BIM compliance checking against BIM standards. Considering the urgent demands of enterprises in the AEC industry and the flawed spreadsheet-based compliance checking tools currently in use, a comprehensive BIM compliance checking framework is required to fill this gap.

1.2 Research motivations

The implementation of BIM has grown steadily over the past decade, from its initial release as a concept to projects that predominantly use BIM throughout the project lifecycle (Eadie et al. 2013). On the other hand, as a summary of human knowledge, standards or guidelines in the AEC industry are used for the standardization of workflow and technical references to guarantee not only the reliability and quality of the project but also the agility and efficiency in the delivery of the project (Preidel et al. 2015). Therefore, compliance checking against the applicable standards represents an essential process for different enterprises during the execution of an AEC project.

1.2.1 Lack of comprehensive comparative analyses between BIM standard systems

With the evolution of the BIM industry, a growing number of regulations, requirements and standards have been published. Currently, BIM standards can be broadly divided into four systems globally, namely:

- the British BIM standard system (BSI series, ISO 19650 series)
- the Chinese BIM standard system (GB/T 51212, GB/T 51235, etc.)
- the American BIM standard system (NBIMS series, IFC-related standards)
- the Open BIM framework

Due to differences in productivity and social conditions, there are variations between different BIM standard systems. Over the past decade, many efforts have been devoted to reviewing existing BIM standards. However, these studies usually only focus on one standard system (Barannik 2017) or one specific goal (e.g., sustainability (Chong et al. 2016), facilities management (Azzran et al. 2019), and collaboration (Lea et al. 2015)). Other studies have collated and analysed the content and structure of different standard systems but comparative analyses between the standard systems are absent (Edirisinghe and London 2015a) or only at a superficial level (document type and number) (Sarı and Pekeriçli 2020; Ibrahim and Al-Kazzaz 2021). There is a lack of comprehensive and in-depth comparative analyses of the contents of multiple BIM standard systems. In this research, a comprehensive comparative analysis is conducted to elaborate on the strengths and weaknesses of each standards system, thereby analysing the key requirements of BIM standards. The comprehensive comparative analysis is essential to help academics and industry better understand the current status of BIM standards.

1.2.2 Defects of existing BIM compliance checking approaches

In order to ensure the quality and efficiency of BIM projects, many enterprises in the AEC field that deliver BIM-related projects have raised the necessity for compliance checking (Beach et al. 2013). In this regard, academics have responded positively with a number of assessment models (e.g., i-CMM (National Institute of Building Sciences buildingSMART alliance 2007), BIM4VET (Guerriero et al. 2019), BeF(Chen 2015)). These assessment models usually use spreadsheets as the assessment tools to integrate the assessment indicators and corresponding weighting systems. During the assessment process, the users would first be evaluated against each indicator (or question) by domain experts or themselves, and then the spreadsheet calculates the assessment score based on the evaluation results, predefined formulas, and indicator weights. As spreadsheets can only do some simple calculations (e.g., Excel files) and cannot run logical relationship reasoning, this type of assessment model is merely capable of implementing simple and rough compliance checking with a fixed set of indicators. Considering the increasingly frequent collaborations and the growing complexity of BIM projects, the demand for more granular and multiple BIM standard compliance checking is emerging, which cannot be met by the existing assessment models. Therefore, a more advanced compliance checking framework is required to be proposed to resolve these challenges. It should support BIM compliance checking with more complex criteria and provide the function of switching criteria sets to conduct compliance checking under various scenarios. The advent of a comprehensive BIM compliance checking framework would have a profound impact on the development of the AEC industry. It could significantly reduce the time and costs of checking compliance against different standards. Additionally, a more comprehensive assessment of BIM compliance can help the stakeholders identify deficiencies in their existing project delivery methods and provide guidance for them to optimise and improve the workflow.

1.2.3 The challenge in automated BIM compliance checking and its potential solution

Although research on compliance checking of BIM standards has been conducted for many years, there is still a gap in fully automated assessment approaches. Currently, state-of-the-art methods only enable automatic calculation of the overall score for BIM compliance. The definition of indicators and the development of the assessment frameworks still rely on the manual work of domain experts. During the evaluation process, the users are also required to complete a self-evaluation by answering a series of questions. Therefore, checking compliance through these conventional methods is labour-intensive and time-consuming. In addition to this, the subjectivity of the domain expert in developing the assessment model and the uncertainty of the users in answering the questions can also affect the objectivity of the final assessment results. In view of the above shortcomings, academics have gradually begun to focus on how to realize automated consistency checking in recent years.

As the former Chair of the UK Health and Safety Executive, Dame Judith Hackitt, states in the Interim Report of Independent Review of Building Regulations and Fire Safety (Hackitt 2017), the contents of regulations and guidance are quite complex, leading to misinterpretation in their application in large and complex projects. This is the reason that most existing assessment methods still rely on domain experts to establish their knowledge models. Thus, it has been a challenge for automated compliance checking to accurately extract and interpret the requirements in the standard. Moreover, information exchanged among stakeholders is often complicated and duplicated between different stages of project delivery. Hence, it is also difficult to effectively extract information on the actual implementation of the project at each stage. Furthermore, the most critical challenge for automatic BIM compliance checking is how to automatically establish the mapping between requirements extracted from standards and the implementation of the actual project. The essence of compliance checking is to compare the gap between the requirements and the actual actions. The requirements in the standard and the actual actions of the stakeholders in the project are usually heterogeneous information. A solution that enables automatic mapping across heterogeneous information is one of the main tasks of this research.

Artificial intelligence (AI) technologies, which have been developed rapidly in recent years, show great potential in the automation of BIM compliance checking. With the help of deep learning (DL) techniques, rules and requirements in textual regulations can be efficiently identified and extracted. The entity and their relations in the documents can be automatically parsed and analysed through natural language processing (NLP) techniques. These advancements provide powerful tools for dealing with complex and heterogeneous information involved in BIM regulations and actual projects. Furthermore, research on automated code compliance checking (ACCC) has advanced greatly in recent years. Although the approaches proposed in these studies cannot handle the reasoning of complex logic and mapping of entities and relations for automated BIM compliance checking, they provide some fundamental inspirations for this research. Based on advanced NLP, DL tools, and related research in other domains, a fully automated BIM compliance checking framework is proposed via the incorporation of domain knowledge, integration and optimisation of existing methods to realise comprehensive, flexible and automatic BIM standard compliance checking.

A fully automated framework for BIM compliance checking is a landmark for the whole AEC industry. It can not only eliminate manual work the enterprises spend on BIM compliance checking but also enable fast compliance checking across different actual projects and target standards. The advent of automated checking methods can significantly save human resources and time for compliance checking and improve the efficiency of enterprises in managing and delivering BIM projects. This fully automated framework for BIM compliance checking can also help enterprises to rapidly and accurately assess the strengths and weaknesses of their existing BIM workflows, identifying gaps against target standards, thereby supporting them with precise quality control and the optimisation and improvement of existing workflows.

1.3 Research hypothesis

In view of the above-mentioned research problems and motivations, there are several gaps not only in the completeness but also in the automation of BIM compliance checking. To address the highlighted issues, this research aims to propose a comprehensive compliance checking framework for BIM standards through knowledge engineering (KE), deep learning (DL) and natural language processing (NLP). Based on the research objectives and the techniques utilised, the hypotheses adopted in this research are as follows:

> A smart BIM compliance checking framework, combining knowledge-driven subjective assessment and deep learning-based objective assessment can not only address the limitations of existing methods in comprehensiveness, granularity, and efficiency but also fulfil the industrial demands on BIM compliance checking under different application scenarios.

1.4 Research questions

Based on the motivations of this research, 7 research questions have been formulated to evaluate the above hypothesis and contribute to the body of knowledge.

Q1: What are the correlations and differences between different BIM standards (systems)? What are the KPIs for evaluating BIM compliance? – **Chapter 4**

Q2: How to assemble indicators into a comprehensive assessment framework that enables compliance checking against different BIM standards? – **Chapter 5**

Q3: How to automate the distinction and extraction of clauses and descriptions of clauses in the BIM standards? How to automatically parse statements, extract entities and relations within the statement, and assemble these constituents into a knowledge graph? – **Chapter 6**

Q4: How to automatically check the compliance between requirements specified in standards and the actual action recorded in the project documents? – **Chapter 6**

Q5: Can the proposed compliance checking framework provide an accurate assessment of BIM compliance in the practical scenario? What is the gap between the assessment results of the proposed methods versus the ground truth of experts' manual examination? – **Chapter 7**

1.5 Research innovations

This research contains work related to some theoretical analysis and practical developments regarding BIM standard ontology and compliance checking pipelines. The methodology incorporates a comprehensive analysis of BIM standards and several advanced techniques that have the potential to enable the automation of BIM compliance checking. The overall system has achieved promising performance on the validation of a practical industry project. The main innovations of this research are fourfold.

1.5.1 BIM standard analysis

Existing analyses of BIM standards are mostly confined to statistics or a superficial level of content analysis. In this research, a comprehensive BIM standard analysis is conducted for the first time by thoroughly comparing the analysis of various BIM standards from the UK, the USA, and China. It not only contains an in-depth analysis of the content in each of the BIM standards but also summarises and concludes the content from the perspective of the standard system. On this basis, different BIM standard systems are deconstructed, thereby enabling the mapping of KPIs in BIM standard systems.

1.5.2 Regulatory knowledge representation

Current BIM standards use textual documents as knowledge carriers, where the knowledge is poorly interoperable and prone to ambiguity. In this research, a more flexible and compatible semantic knowledge model is adopted as the digital carrier of BIM knowledge. Compared with textual documents, this semantic network-based knowledge representation (e.g., ontology and knowledge graph) provides a more powerful and flexible semantic representation, which makes it more desirable to represent knowledge with complex logic and relations. Semantic inference and retrieval can be easily implemented to facilitate querying the relevant knowledge. In addition, the semantic-based knowledge model can be easily reused and applied to various downstream applications. Moreover, this semantic representation shows great potential in knowledge fusion and knowledge completion,

1.5.3 comprehensive and flexible checking

As explained in Section 1.2.2, existing compliance checking methods are flawed in terms of flexibility and complexity. To address this, an ontology-driven approach is proposed in this research. The proposed method takes ontology as the core of knowledge and implements indicator filtering, weight calculation and result presentation through a series of algorithms. This comprehensive assessment method innovatively utilises an ontological knowledge model as the knowledge container to represent the requirements in the standards, enabling the linkage between different BIM standards. Through the ontological knowledge container, the flexibility and sustainability of the criteria representation in the BIM maturity assessment are further improved. Criteria can be easily modified, maintained and updated when dealing with different target standards, project documents and stakeholders. The indicator filtering and weight calculation algorithms allow a rapid and reliable assessment of overall BIM compliance as well as standard-specific or role-specific compliance checking.

1.5.4 Automation of compliance checking

The current academic research on automated compliance checking is all quantitive and design-orientated. Qualitative BIM compliance checking is challenging due to the complexity of entities and their relationships. In this research, a novel automated project-level BIM compliance checking pipeline is proposed through transfer learning, syntactic parsing, and graph learning. The proposed method innovatively converts the logical inspections into a quantifiable similarity calculation between knowledge graphs, thereby filling a vacuum in fully automated BIM compliance checking.

1.6 Structure of the thesis

This thesis is divided into several chapters corresponding to the above main research questions.

Chapter 1 – Introduction: This chapter serves as a comprehensive introduction to the whole research, covering various perspectives such as the objectives, research rationale, hypotheses, research contributions, and thesis structure.

Chapter 2 – Literature review: This chapter provides a comprehensive analysis of the relevant literature and offers valuable insights into the current situation of BIM standards and BIM compliance checking, and potential techniques for automatic compliance checking.

Chapter 3 – Research Methodology: This chapter presents the overarching methodology for conducting this research to articulate the principles and methods applied in this research.

Chapter 4 – BIM compliance KPIs and overarching compliance checking framework: In this chapter, a set of BIM compliance indicators is generated through a semi-automatic approach. On the basics of these KPIs, a smart BIM compliance checking framework is proposed, which is comprised of a subject-oriented ontology-driven checking framework and an evidence-driven automatic checking framework.

Chapter 5 – Ontology-driven compliance checking framework: This chapter demonstrates the manual development process of the ontological comprehensive compliance checking framework, which includes the development of an ontology, AHP-based weighting matrix and a Python-based platform.

Chapter 6 – Automated compliance checking framework: This chapter aims to introduce the development process of the automated compliance checking framework, which comprises a regulatory knowledge mining engine, a document information extraction engine, and a checking engine.

Chapter 7 – Validation and testing: This chapter addresses the testing and validation of the whole proposed framework. The validation constitutes three parts, including indicator validation, function validation and actual project validation.

Chapter 8 – Conclusion: This chapter serves as the conclusion of the research work presented in the previous chapters. As well as highlighting the research contributions, the research limitations and the future work are discussed respectively in the remainder of this chapter.

Chapter 2. Literature review

Due to its advantages in information exchange and collaborative work, BIM plays an important role in improving productivity, saving costs, and shortening the duration of the project (Bryde et al. 2013). Figure 1 is a visual explanation of the mission of BIM given by the BIM National Standards Committee of the buildingSMART Alliance in the United States (National Institute of Building Sciences buildingSMART alliance 2007).

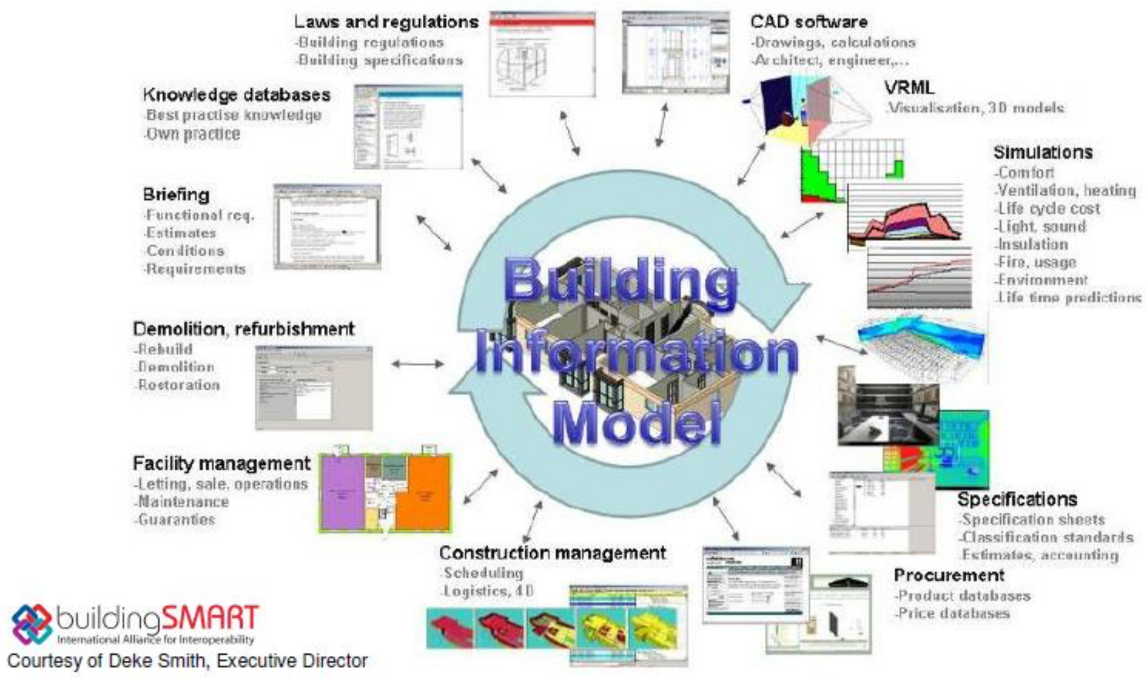


Figure 1 - BIM-enabled lifecycle process of design, construction, operation and maintenance (National Institute of Building Sciences buildingSMART alliance 2007)

2.1 BIM adoption

The adoption of BIM in the global AEC industry was intensely investigated at the beginning of the 21st century. With the popularity of BIM technology, this topic is now no longer of interest. As stated by Yan and Demian (Yan and Demian 2008), the adoption of BIM has brought significant improvement in quality, creativity, sustainability, and decrement in cost, time, and human resources (HR). After recognising the superiority of BIM, more and more countries have begun to adopt various means to promote the development of the BIM industry and gradually started to mandate the application of BIM-related concepts and technologies to deliver construction projects. The status of BIM adaption in different countries is shown in Figure 2 (Mcauley et al. 2017). According to Quantity Surveying Forecast, BIM's market share in the global AEC industry is still growing rapidly by 13% per annum (Edirisinghe and London 2015b), which means that the global BIM market share is expected to reach 14.7 billion US dollars by 2025. This can be considered as evidence that the implementation of BIM is accelerating globally in the construction industry.

Figure 2 - The status of global BIM adoption (Mcauley et al. 2017)

The figure above uses different colours to highlight the different statuses of BIM adoption. As illustrated in Figure 2, more than one-third of national and regional governments have required companies and organisations in the AEC industry to use BIM-related concepts and technologies to complete engineering projects in accordance with BIM standards. Among these countries, some have adopted open BIM standards published by buildingSMART Alliance, while others have developed BIM standards based on their industrial conditions. In addition, nearly one-fifth of countries have set a short-term goal of adopting BIM to be able to deliver construction projects using BIM-related technologies. Another one-fifth of countries are in the process of developing a BIM implementation plan. Some countries, including Brazil, Italy and New Zealand, are also planning to start promoting BIM adoption in the future.

Chronologically, the United States is considered the first country to adopt BIM across the world. The U.S. General Services Administration (GSA) launched the National 3D-4D-BIM Program in 2003 to improve productivity in construction and support the growth of information technology in the construction industry. The focus of this program is to develop policies to require model-based design, including native and IFC BIM and open-standard facility management data. Since 2007, GSA has been publishing a series of BIM guides to standardise and guide the use of BIM on real projects. On the other hand, BuildingSMART, a professional committee of the National Institute of Building Sciences in the field of information resources and technology, was established in 2007, and built on the foundation of the former International Alliance of Interoperability (IAI). With the help of BuildingSMART, the National Institute of Building Sciences (NIBS) published the first edition of the U.S. National BIM Standard (NBIMS-v1) in 2007, which was followed by the second, third and fourth editions in 2012, 2015 and 2023 respectively.

In the following 10 years, influenced by the United States, some European countries, such as Finland, Norway, Denmark, and the Netherlands, also began to introduce BIM mandatory for their construction projects. Another significant event in the BIM domain took place in 2011 when the UK government released their new Government Construction Strategy (Efficiency and Reform Group of the Cabinet Office 2011), which stated that the government required all constriction sectors to achieve full BIM adoption in asset information, documentation and data exchanging for government procured contracts by 2016. This policy drew a lot of global attention as this initiative helped elevate the BIM approach as key to the government's overall goal of increasing productivity. Moreover, unlike the requirements for BIM in most countries, the UK government made the adoption of BIM compulsory for all projects and a condition for securing government contracts, thereby significantly accelerating the adoption of BIM by contractors. Therefore, this policy is regarded as a starting point for BIM adoption and development in the UK (National Building Specification 2019). Guided by this policy, the UK government defined the definition, core work, and goals of BIM development and proposed a multi-level BIM roadmap (Figure 3).

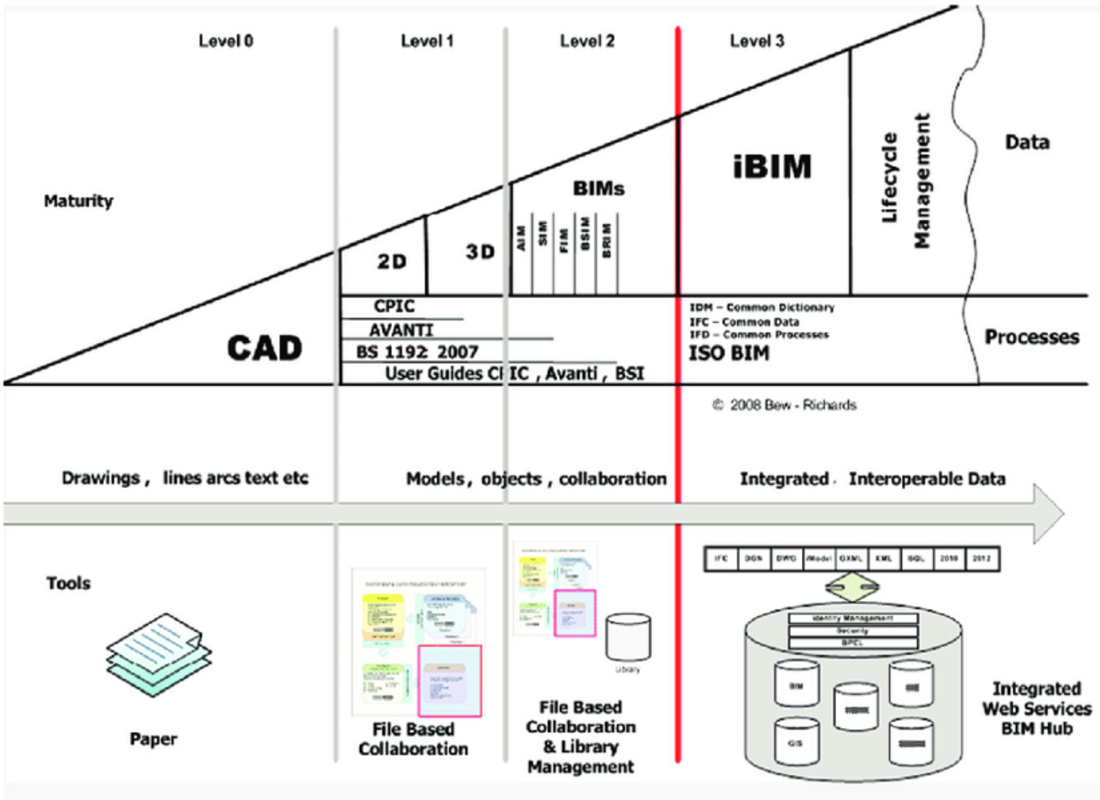


Figure 3 - The multi-level BIM roadmap proposed by the UK government (Bew and Richards 2008)

China, as the country with the largest engineering and construction scale in the world, its engineering and construction industry is greatly affected by BIM technology. In China, research on BIM technology began in the 1980s. Similar to the United States, the start-up BIM-related research in China was focused on the technical architecture and application methods of model-based architectural CAD software. In 1998, BIM technology research was first formally introduced into a national research project.

Since then, in the following 20 years of China's national economic and social development programme, the Chinese government has proposed a series of research topics to promote the development and application of BIM technology (Xinhua Press 2023), such as “Key Technology for Modern Building Design and Construction Integration”, “Research and Application of Key Technologies for Informatisation in the Construction Industry”, etc. Over the past two decades, China has made great progress in the research and application of BIM technology and has gradually formed a self-contained system of laws, specifications, and standards. These regulations have been used in a variety of projects, significantly improving ease of construction, budget control, management of the whole building lifecycle, and the productivity of the people involved.

2.2 BIM standards

As reviewed in the last section, BIM-related concepts and technologies have been widely accepted and applied in engineering projects around the world. With the development of BIM technology, many countries and organisations have published a series of BIM standards to standardise the way and the processes of using BIM to ensure the quality and efficiency of BIM projects. Due to the differences in production methods, there are variations in BIM standards among different countries and regions. Based on the review of BIM adoption, it can be observed that there are currently four representative and comprehensive BIM standard systems around the world, namely the US BIM standard system, the UK BIM standard system, the Chinese BIM standard system and the Open BIM standard system. In the remainder of this section, the above four typical BIM standard systems are selected to analyse the commonalities and differences between existing BIM standard systems through an in-depth content analysis.

2.2.1 American BIM standards

The BIM promotion system in the United States is relatively unique in the global context. The development of BIM technology in the United States is industry-led, where BIM technology and concepts are first applied in actual engineering cases, then through the accumulation of experience in specific engineering cases, enterprises, organisations, and the government are gradually required to formulate relevant policies and systems to accelerate the promotion of BIM, and enhance the productivity and value of the overall industry chain. The first version of the national BIM standard of American - National Building Information Model Standard (NBIMS) (National Institute of Building Sciences buildingSMART alliance 2007) was published by the BuildingSMART alliance in 2007, which was followed by the second and third versions in 2012 and 2015. The fourth version of NBIMS has just been released in August of this year. These four national BIM standards comprise the main body of the American BIM standards system.

1. NBIMS-v1

As the first national BIM standard of the United States, NBIMS-v1 (National Institute of Building Sciences buildingSMART alliance 2007) is considered an introduction of the whole standard system, which specifies principles, scope of investigation, organization, operations, development methodologies, and planned products. This first version consists of five main sections and three appendices. The first two sections illustrate the scope of the standard, interpretation of key terms, and related key concepts, and also provide a navigational guide for readers on how to use this BIM standard. The collaborative information exchange process proposed in NBIMS-v1 is shown in Figure 4.

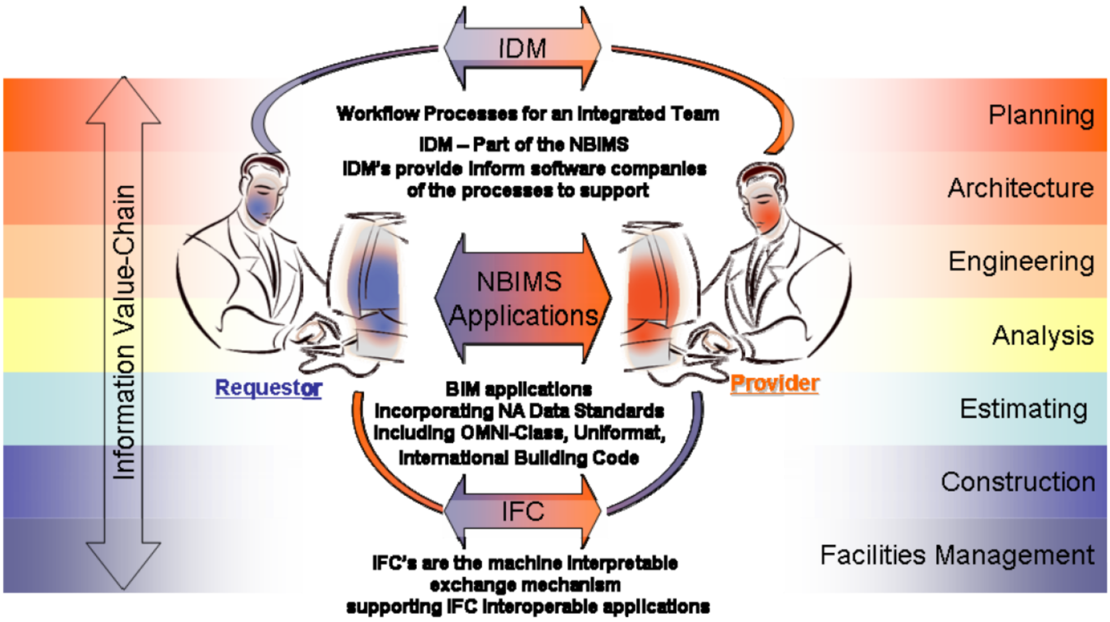


Figure 4 - Process of collaborative information exchange defined in NBIMS-v1 (National Institute of Building Sciences buildingSMART alliance 2007)

The third, fourth and fifth chapters of the NBIMS-v1 focus on the other three fundamental topics, including the information exchange concept, information exchange content, and development prose of the standard based on industry requirements, key factors that can enable effective BIM modelling, and software interoperability that can improve the efficiency of data exchange. Furthermore, the standard sets out a number of principles and requirements in dealing with information assurance, e.g., “Information assurance should start from creation and should include key information such as who created the data, the quality of the data, when and why the data was created”(National Institute of Building Sciences buildingSMART alliance 2007). In addition, the standard also proposes the concept of Minimum BIM and Capability Maturity Model (CMM). The Minimum BIM is intended to represent the minimum requirements on software tools, datasets and standards for project phases from design, and construction to handover. The Capability Maturity Model is developed for users to assess their business practices along a continuum or spectrum of desired technical level functionality. The final section of NBIMS-v1 highlights the process of developing and implementing this standard, which illustrates the requirements for Model View Definition (MVD) (buildingSMART 2022), discusses the process in the Information Delivery Manual (IDM) (International Organization for Standardization (ISO) 2016), and the tools used in the requirements definition activities. The coding system is introduced and explained in the appendices, including the Industrial Foundation Class (IFC) (International Organization for Standardization (ISO) 2018a), OmniClass (Construction Specifications Institute 2006), and the International Framework of Dictionaries (IFDLibrary) (International Organization for Standardization (ISO) 2022). This first version of the American BIM standard is relatively generic, which only concentrates on the fundamental concept, idea and tools of BIM and provides a basic explanation to help users understand what BIM is and how to use BIM.

2. **NBIMS-v2**

Five years after the release of NBIMS-v1, the second version of NBIMS was published. In comparison to the first edition, NBIMS-v2 (National Institute of Building Sciences buildingSMART alliance 2012) goes further beyond basic concepts and concentrates more on the project's information supply chain across its entire life cycle, from the stages of planning, design, construction to operation and maintenance. Architect, engineer, contractor, owner, and operator (AECOO) teams are encouraged to adopt more productive practices throughout the project lifecycle by using NBIMS-v2, a living and evolving set of guidelines for software developers, vendors, and implementers who work in the design, construction, and operation of the built environment. Additionally, NBIMS-v2 (National Institute of Building Sciences buildingSMART alliance 2012)provides standards for building information technologies and an integrated practice system and framework for all industry professionals working in an open, non-proprietary, standard-accessible, trustworthy, and collaborative digital environment.

The second version of NBIMS follows the data framework built in the first version, which is constituted by Industry Foundation Class (IFC), OmniClass and IFD Library. Based on this data framework, NBIMS-v2 (National Institute of Building Sciences buildingSMART alliance 2012) outlines the requirements for information sharing in modelling and information delivery. To facilitate users better following the requirements of information exchanging, some project temples are provided in the standard, for example, construction operations building information exchange (COBie) (National Institute of Building Sciences 2011), spatial program validation (SPV), building energy analysis (BEA) and quantity take-off for cost estimating (QTO). In addition, it also lists some actual project documents as sample cases to help users better understand the principles of developing the BIM project execution plan, the spatial coordination of the MEP and fire protection systems, and the handover of planning, executing, and managing information.

3. NBIMS-v3

The third version of the national BIM standard (NBIMS-v3) (National Institute of Building Sciences buildingSMART alliance 2015)was published in 2015 by the buildingSMART alliance and is still valid now. Compared with the second version, the NBIMS-v3 only saw a few minor changes. Along with the addition of some new standards and the update of outdated documents, some missing and incomplete actual sample cases were supplemented and enhanced respectively.

The third version of NBIMS consists of two parts. The first part is an introduction to the applicable scope of this BIM standard and explanations of both the additional concepts accepted in version 3 and the concepts established in previous versions (NBIMS-v1 (National Institute of Building Sciences buildingSMART alliance 2007), NBIMS-v2 (National Institute of Building Sciences buildingSMART alliance 2012)). The second part is the primary body of this standard, where the outdated information exchange standards and practical documents listed in previous editions are replaced by new versions, for example, the Minimum BIM, COBie, SPV, BEA, QTO, and guidelines for BIM project execution plans. In addition, a few more information exchange standards and practice documents have been supplemented into this third version of NBIMS (National Institute of Building Sciences buildingSMART alliance 2015), such as the requirements for a practical BIM contract, the building programming information exchange (BPie), electrical information exchange (SPARKie), HVAC (heating, ventilation, and air conditioning) information exchange (HVACie), and water systems information exchange (WSie).

4. NBIMS-v4

The NBIMS-v4 is the latest version of the American national BIM standard, whose draft version was released in August of 2023 and is still under public review. Influenced by the second part of ISO 19650, the international standard for BIM information management processes published in 2018, the fourth edition of the NBIMS supplements the requirements for information exchange processes on top of the previous version. In addition, concepts and technical terms in the previous standard are updated and modified to align with the ISO19650 series of standards.

This standard system (NBIMS) is normally considered to be the key component of the digital transformation of the American construction industry. It establishes standardised definitions and concepts for the exchange of building information to achieve the goal of using standard semantics and ontologies for the development of business-critical environments. This set of standards facilitates accurate and efficient communication and business dealings, which are desired by the construction industry and indispensable for the transformation of the building industry. Furthermore, it can help all parties involved in the supply chain get more reliable outcomes from commercial agreements and significantly reduce implementation costs simultaneously.

2.2.2 British BIM standards

Due to the government's mandatory promotion strategy in 2011, British companies that delivered government-procured AEC projects were striving to achieve BIM Level 2 when delivering construction projects. However, the lack of BIM standards obstructed the promotion of BIM to achieve the objective of BIM Level 2 (Efficiency and Reform Group of the Cabinet Office 2011). To fill this gap, the British Standards Institution (BSI) became the forefront of British BIM standardisation. With support from industry bodies, regulators, government, and construction professionals, the BSI has published a series of BIM-related standards (see Table 1) to fulfil the demands of both government and industry. These standards constitute the main part of the UK's BIM standards system. So far, the standards that comprise the UK BIM Standards System are organised into three main categories, the BS (British Standard) series, the PAS (Publicly Available Specifications) series, and the ISO (International Organization for Standardization) series. The PAS series is rapidly developed standards, which aim to meet the urgent demands of the market. The PAS 1192 series was built on BS 1192:2007 + A2:2016, which is designed to provide a 'best-practice' method for the development, organisation and management of production information for the construction industry. To cover all aspects in the process of project delivery, 6 BIM standards are planned to be published within this series, focusing on the project delivery phase (PAS 1192-2), operational phase (PAS 1192-3), COBie (PAS 1192-4), security-minded management (PAS 1192-5), and health and safety (PAS 1192-6) respectively. Following the publication of the PAS 1192 series, international asset owners and clients (particularly in the Middle East and Australia) recognized their benefit and began to require the adoption of the management processes defined within the UK 1192 series on their projects. Considering this demand, the international community proposed to elevate the PAS 1192 series into a series of international standards (ISO 19650 series). The ISO 19650 series was released in 2018. It uses the UK 1192 series as its basis and retains the same concepts and principles behind the collaborative production of information and the common data environment. More details of the ISO 19650 series BIM standard can be found in the 8th part of this section.

Table 1 – British BIM standards and their referencing relationships

No.	Document title	Issuer	Year	Reference
1	BS 1192 Collaborative production of architectural, engineering and construction information – Code of practice	BSI	2007	
2	BS 1192-4 Collaborative production of information: Part 4 – Code of practice	BSI	2014	1,3,4
3	PAS 1192-2 Specification for information management for the capital/delivery phase of construction projects using BIM	BSI	2013	1,4
4	PAS 1192-3 Specification for information management for the operational phase of assets using BIM	BSI	2014	1,3,12,13, 14,15,16
5	AEC (UK) BIM Protocol Implementing UK BIM Standards for the Architectural, Engineering and Construction industry	AEC (UK) Initiative	2012	1,3,7,8
6	BIM Overlay to the RIBA Outline Plan of Work	RIBA	2012	1
7	BS 8541-1 Library objects for AEC – Part 1: Identification and classification – Code of practice	BSI	2012	1,8,9,10
8	BS 8541-2 Library objects for AEC – Part 2: Recommended 2D symbols of building elements for use in BIM	BSI	2011	7,9,10
9	BS 8541-3 Library objects for AEC – Part 3: Shape and measurement – Code of practice	BSI	2012	7,8,10
10	BS 8541-4 Library objects for AEC – Part 4: Attributes for specification and assessment – Code of practice	BSI	2012	7,8,9
11	BS 7000-4 Design management systems: Guide to managing design in construction	BSI	2013	1,3
12	BS ISO 55000 Series: Asset management	BSI	2014	
13	BS 8210 Guide to facilities maintenance management	BSI	2012	12,14,15,16
14	BS 8587 Guide for facilities information management	BSI	2012	1,3,15,16
15	BS 8572 Procurement of facility–related services – Guide	BSI	2011	16
16	BS 8536 Facilities management briefing – Code of practice	BSI	2010	11

Figure 5 presents the publishing timeline of some key standards in the UK BIM standard system. The remainder of this section will introduce the content of these standards in chronological order of release.

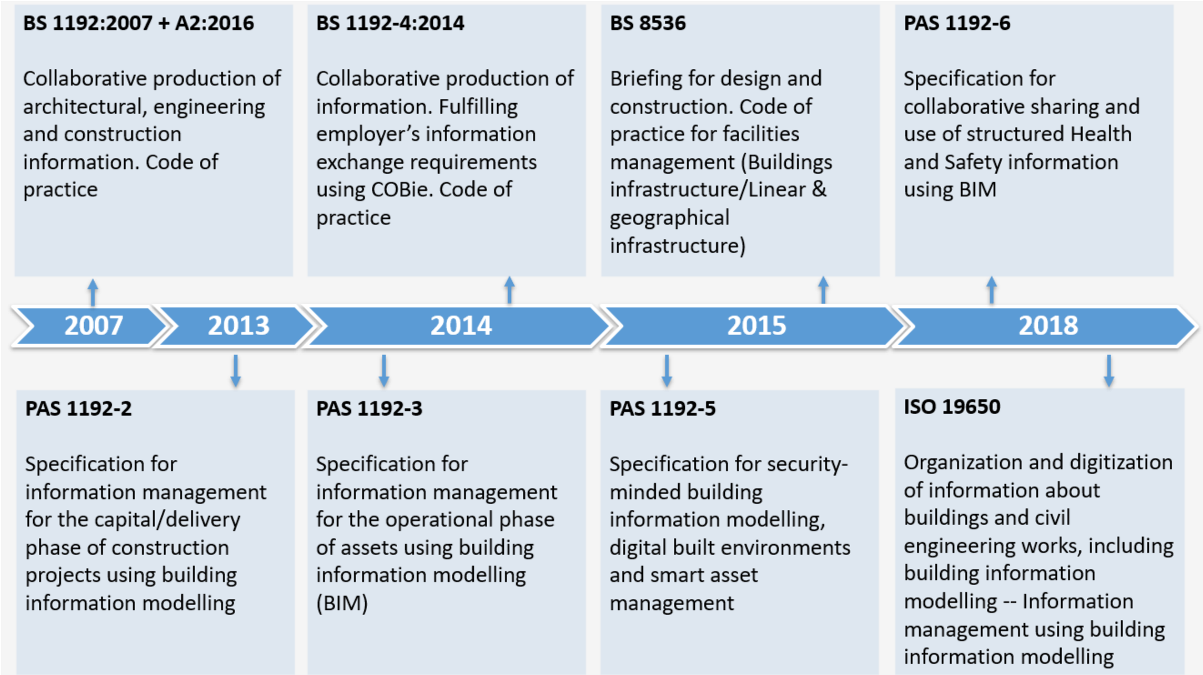


Figure 5 - Publishing timeline of key standards in the UK BIM standard system

1. BS 1192:2007 + A2:2016

BS 1192 (British Standards Institution (BSI) 2007), which was initially designed for computer-aid design (CAD) is considered as the first BIM standard published by the BSI. It provides a standardised approach to managing information related to production, distribution, and construction quality. This standard outlines naming conventions and information exchange principles that are applicable to all BIM projects over the entire supply chain and the whole lifecycle of the project. Figure 6 illustrates an example based on the standard naming conventions. In addition, this standard also provides instructions for application and software developers to help them develop appropriate profiles or application add-ons.

Fields	Directories (see 5.4.2)	Files (see 5.4.3)	Containers within files including layers (see 5.4.4)	Clause
Project	PR1	PR1		6
Originator		XYZ		7
A1) Volume or system (A1		A2) 01 (A2		8.1.2
Levels and locations		01		8.1.3
Type		M3		9
Role		A	A	10
Classification		A1) Uniclass (A1 (optional)	A1) Uniclass (A1	11
Presentation			M	12
Number		0001		13
Description (optional)			Doors	14
Suitability (optional)	S1	S1		15.2.2
Revision (optional)	A1) P02 (A1	A1) P02 (A1		15.2.3
Name	PR1-S1-P2	A2) PR1-XYZ-01-01-M3-A-0001 (A2	A-A1) Uniclass (A1-M_Doors	

Figure 6 - An example of the naming convention defined in BS 1192:2007+A2:2016 (British Standards Institution (BSI) 2007)

2. PAS 1192-2:2013

PAS 1192-2 (British Standards Institution (BSI) 2013) is considered one of the core standards in the UK BIM standard system, which has been further developed as an international BIM standard (ISO 19650-2 (International Organization for Standardization (ISO) 2018c)). This standard specifically concentrates on the process of project information delivery. It specifies the requirements for information management to guide managing information in a consistent and structured manner. Moreover, the standard also describes ways in which information can be accessed, exchanged and managed with a higher efficiency and accuracy. More details on this standard are illustrated in the 8th part of this section (ISO 19650).

3. PAS 1192-3:2014

The third part of PAS 1192 is usually considered a companion to the second part of PAS 1192 (British Standards Institution (BSI) 2013), which also focuses on the process of project information management but is in the operational phase. PAS 1192-3 (British Standards Institution (BSI) 2014b) explains the asset information management process in detail, regulates the responsibilities of each party in the management process, and sets out the requirements for the documents, information, and interactions of each participant at each stage of the process, which is shown in Figure 7. In addition, the standard sets out a framework for information management during the operational phase of an asset to provide guidance on the usage and maintenance of the Asset Information Model (AIM) and describes how it supports AIMs from the perspective of the Common Data Environment and data interaction.

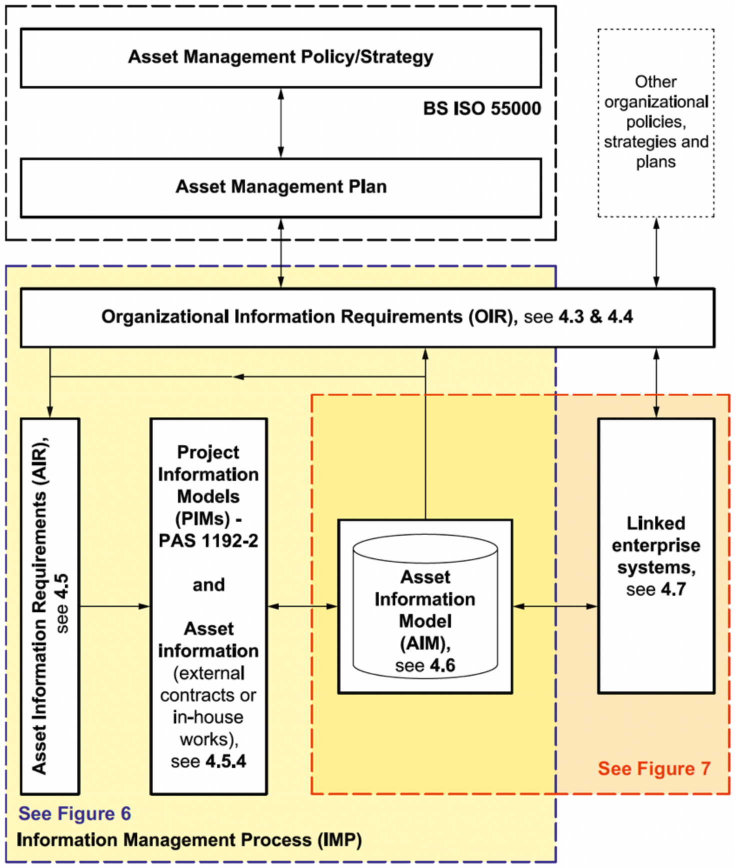


Figure 7 - High-level asset information process map defined in PAS 1192-3 (British Standards Institution (BSI) 2014b)

4. BS 1192-4:2014

Following BS 1192: 2007(British Standards Institution (BSI) 2007), PAS 1192-2 (British Standards Institution (BSI) 2013), and PAS 1192-3(British Standards Institution (BSI) 2014b), BS 1192-4 (British Standards Institution (BSI) 2014a) the fourth published standard in the 1192 series, which focuses on information collaboration and interaction. The core of BS 1192-4:2014 is the definition of the requirements for information interactions during the whole life of an asset. It mandates the exchange of information must be in accordance with COBie and specifies the use of COBie in the UK, as shown in Figure 8.

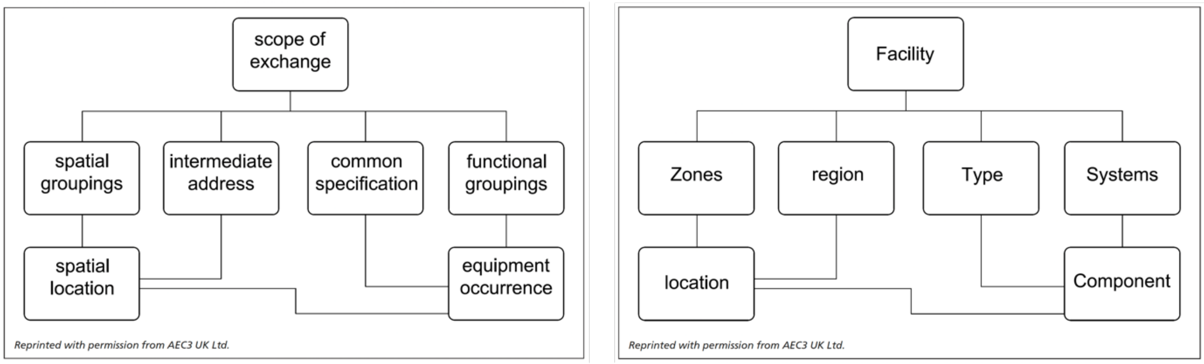


Figure 8 – Architecture of the exchange scope and object information of COBie in BS 1192-4 (British Standards Institution (BSI) 2014a)

COBie is a non-proprietary multi-page spreadsheet data format for the delivery of information during the operational phase of an asset. It is designed as a deliverable of BIM to provide asset data rather than geometric information. Its simple format ensures that BIM information is easy to create by suppliers and easy for employers to assess and use. During the planning and operations phases of a project, information can be provided by the owner and received by the supplier. During the design and construction phases, the information can be provided by the supplier and delivered to the owner. As the project evolves, the COBie deliverables will become more complete.

5. PAS 1192-5:2015

PAS 1192-5 (British Standards Institution (BSI) 2015) is a companion document to PAS 1192-2 (British Standards Institution (BSI) 2013), PAS 1192-3 (British Standards Institution (BSI) 2014b) and BS 1192-4 (British Standards Institution (BSI) 2014a), which extensively references the definitions and concepts contained in these standards. However, the scope of PAS 1192-5 is wider than the concepts covered by the remaining standards in the 1192 series, extending to security-aware approaches in the digital environment and the management of new and existing building assets. PAS 1192-5 (British Standards Institution (BSI) 2015) highlights the inherent security issues of BIM technology in an environment where the construction process is increasingly using and relying on information and communications technology. In this regard, PAS 1192-5 (British Standards Institution (BSI) 2015)describes the cybersecurity vulnerabilities that may give rise to malicious attacks when using BIM technology and provides an assessment process that can be used to determine the level of cybersecurity of the BIM collaboration environment throughout the life of a project. In addition, it also proposes appropriate and proportionate solutions and sets out the requirements for the responsibilities, workflows, and information security management programmes of the various parties involved in the construction process from the perspective of the security of project management documentation.

6. BS 8536

BS 8536 (British Standards Institution (BSI) 2022) is a BIM operational standard published by BSI in 2015 and its main aim is to set out the working requirements for the design and construction phases to ensure that participants achieve information management with the mindset of building operations management from the design stage. The standard is expected to pull building operations teams and suppliers from the periphery of design and construction into the management process of construction. Simultaneously, the standard also expects to manage design and construction at an early stage based on environmental, social, safety, and economic efficiency objectives to ensure that assets and facilities are operated correctly, safely and efficiently after delivery.

7. PAS 1192-6:2018

The sixth part of PAS 1192 (British Standards Institution (BSI) 2018) aims to reduce hazards and risks throughout the project lifecycle, from deconstruction to design, including management of the construction process, and to make sure that health and safety information is managed by the right people at the right time. The standard emphasises the need to anticipate sources of risk at the design stage (anticipation of risk, the designer must identify "foreseeable risk"). Furthermore, PAS 1192-6 (British Standards Institution (BSI) 2018) also emphasises that safety management, like the BIM concept of collaboration, should involve everyone in sharing their knowledge and the sources of risk they witness in safety management. Finally, this standard requires participants to co-ordinately utilise structured health and safety information through the four basic components provided by the standard, enabling it to be iterated and shared throughout the project lifecycle.

8. ISO 19650

The ISO 19650 series of standards is an international standard for managing information over the whole life cycle of a built asset using building information modelling (BIM). It contains all the same principles and high-level requirements as the UK BIM Framework and is closely aligned with the current UK 1192 standards.

ISO19650 series standard consists of four parts:

(1) BS EN ISO 19650-1: Organization and digitization of information about buildings and civil engineering works, including building information modelling -- Information management using building information modelling: Concepts and principles (International Organization for Standardization (ISO) 2018b).
(2) BS EN ISO 19650-2: Organization and digitization of information about buildings and civil engineering works, including building information modelling -- Information management using building information modelling: Delivery phase of the assets (International Organization for Standardization (ISO) 2018c).
(3) BS EN ISO 19650-3: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM). Information management using building information modelling. Operational phase of the assets (International Organization for Standardization (ISO) 2020a).
(4) BS EN ISO 19650-5: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM). Information management using building information modelling. Security-minded approach to information management (International Organization for Standardization (ISO) 2020b).

These standards are founded on the UK's standards for information management using building information modelling, namely BS 1192:2007 + A2:2016 (British Standards Institution (BSI) 2007) and PAS 1192-2:2013 (British Standards Institution (BSI) 2013). The principles remain as per these standards with terminology changes being preserved via the UK National Forewords and National Annex (Figure 9).

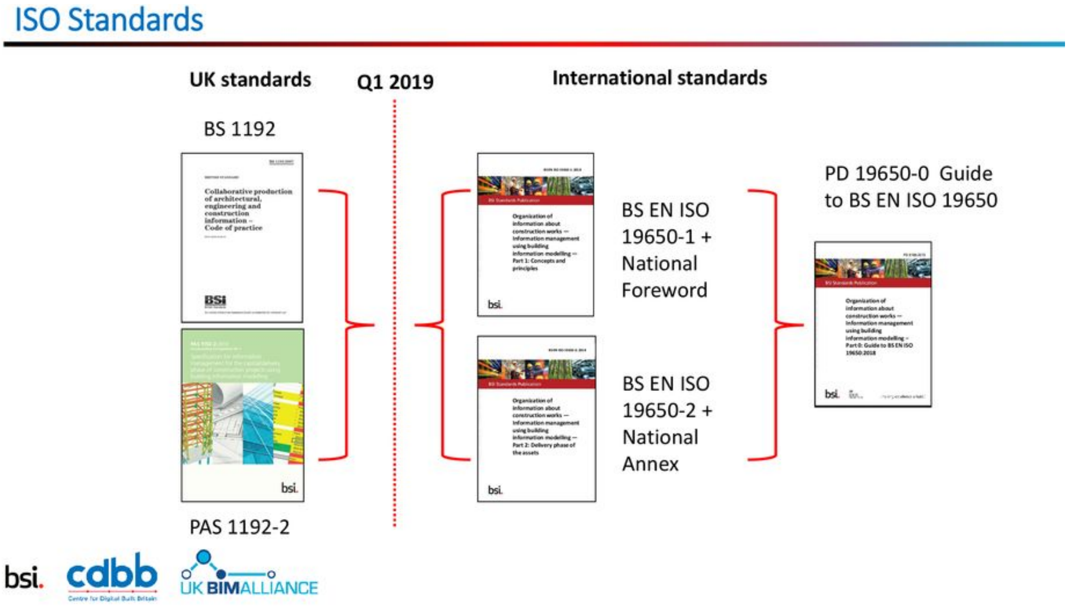


Figure 9 – Relationships between British BS and PAS standard and ISO 19650 series standard (Dadmehr and Coates 2019)

ISO 19650 series set out the recommended concepts and principles for business processes across the built environment sector in support of the management and production of information during the life cycle of built assets (referred to as “information management”) when using BIM. This series of standards is designed to apply to the whole life cycle of a built asset, including strategic planning, initial design, engineering, development, documentation and construction, day-to-day operation, maintenance, refurbishment, repair, and end-of-life. It defines several concepts and principles from information requirements, capability and capacity assessment to common data environment and provides a recommended framework to manage information including exchanging, recording, versioning and organizing for all actors.

In this series of standards, the concept and function of asset information models (AIM) and project information models (PIM) are defined and explained. In this regard, an information management framework has been designed to achieve organisational alignment. In addition, this standard proposes a list of information requirements and information models that the stakeholders should establish to state their purposes for requiring information deliverables, including asset register, support for compliance and regulatory responsibilities, risk management, and support for business questions. The relationships between these information requirements and information models are also thoroughly explained in the standard (**Error! Reference source not found.**). Moreover, ISO 19650 (International Organization for Standardization (ISO) 2018c)

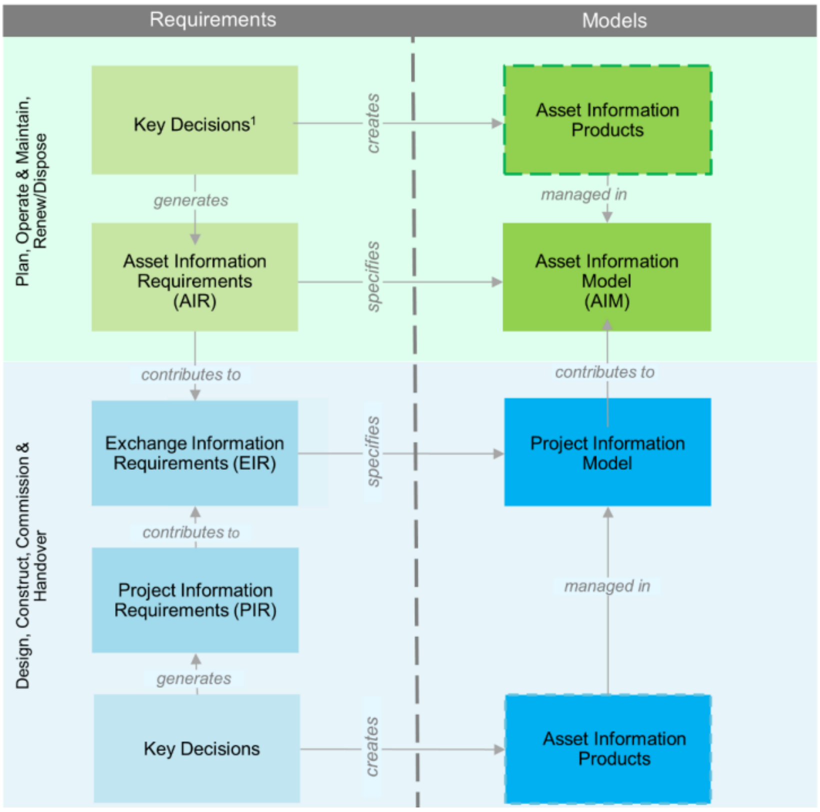


Figure 10 - Relationships between information requirements and information models defined in ISO 19650

sets specific requirements for asset information management functions, project information management functions and task information management functions in the aspect of complexity, assignment, and involvement in different project stages. It highlights the importance of capability and capacity reviews of the prospective delivery team by one of three groups: the appointing party, the delivery team or a nominated third party. The meaning and extent of capability and capacity are clarified as well. In terms of information delivery planning, ISO 19650 sets requirements for planning in terms of scope, timing, content, and relationships with information requirements, where two documents, Timing of information delivery and Responsibility matrix, are emphasised to be included in the information delivery planning process. To ensure the quality of information delivery, the purpose of the federation strategy is explained, and some development specifications are also provided for them. For collaborative information production, this standard defines the Level of Information Need and Information quality and further emphasises the conformity of the classification with ISO 12006. Finally, ISO 19650 illustrates the common data environment (CDE) defined in this standard from the perspective of its structure and workflow, which helps both the stakeholders and developers understand what a CDE is and how it works. On this basis, the state of revision in each information container and the transition process between different states within the CDE are specified (Figure 11).

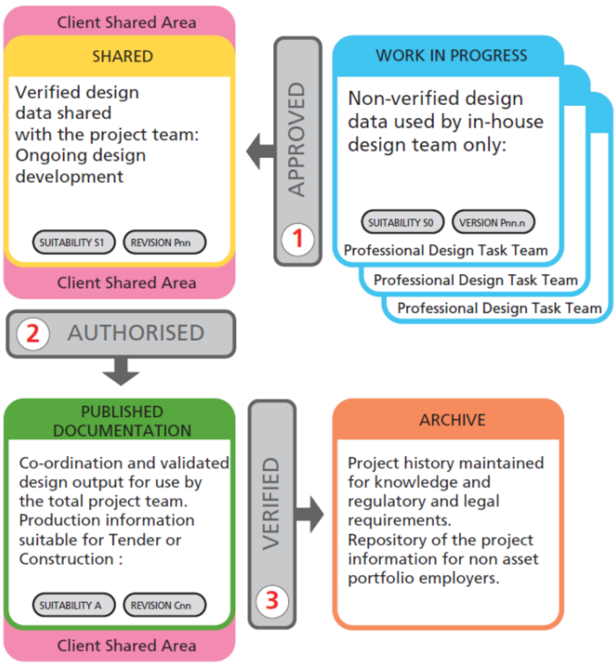


Figure 11 - Structure and workflow of a common data environment (International Organization for Standardization (ISO) 2018b)

Apart from these concepts and principles for the business processes, ISO 19650 (British Standards Institution (BSI) 2013) also sets out the specific requirements for information management during the delivery of built assets, as shown in Figure 12.

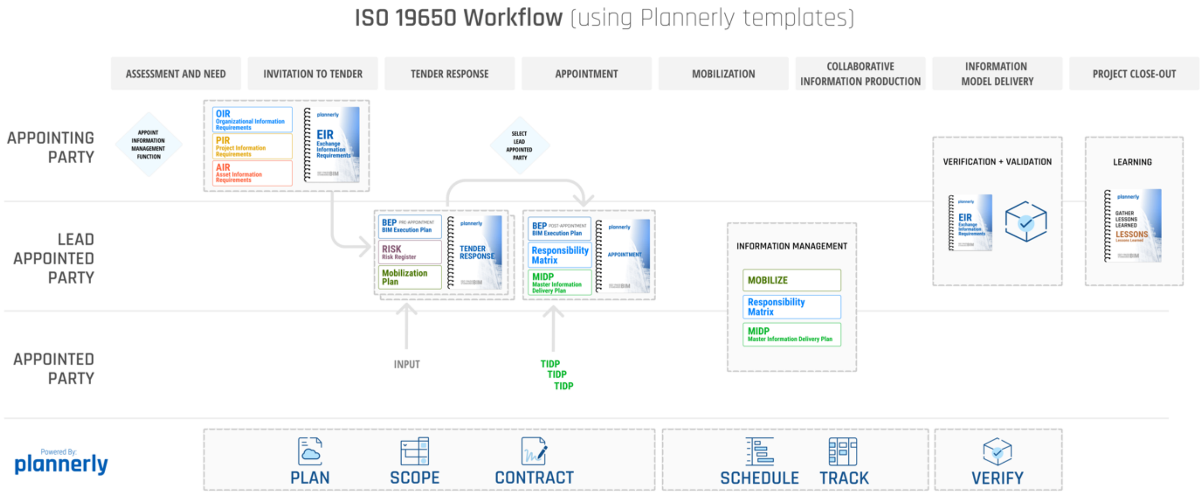


Figure 12 - The information management workflow defined in ISO 19650 (Plannerly 2024)

These processes can deliver beneficial business outcomes to asset owners/operators, clients, their supply chains, and those involved in project funding including increase of opportunity, reduction of risk and reduction of cost through the production and use of asset and project information models. The use of this standard will help remove barriers to collaborative working and competitive tendering across borders and increase opportunities. As well, the production and use of asset and project information models can also reduce risks and cost.

2.2.3 Chinese BIM standards

The development of China's BIM standard began in 2012, building upon earlier initiatives. The groundwork was laid with the release of JG/T 198-2007 "Digital Definition of Building Objects" (Ministry of Housing and Urban-Rural Development (China) 2007) by the China Academy of Building Standards and Design in 2007, and GB/T 25507-2010 "Industrial Foundation Classes" (Administration of Quality Supervision Inspection and Quarantine of People's Republic of China 2010) by the China Academy of Building Research in 2010. These standards, both rooted in the IFC (ISO 16739:2005 (International Organization for Standardization (ISO) 2018a)) standard, underwent some modifications to align with the format of the Chinese national standard. In terms of the research on the standard framework, the School of Software at Tsinghua University established the BIM group in 2009 and dedicated three years to formulating the concept and methodology of the Chinese BIM Standard System (CBIMS) (BIM Group of Tsinghua University 2010). Table 2 lists some representative national and local Chinese BIM standards.

Table 2 - Some representative Chinese BIM standards

Number	Name	Issuer	Year
GB/T 51212	Uniform Standard for Building Information Model Application	MOHURD	2016
GB/T 51269	Standard for classification and coding of building information model	MOHURD	2017
GB/T 51301	Standard for design delivery of building information modeling	MOHURD	2019
GB/T 51235	Standard for building information modelling in construction	MOHURD	2017
GB/T 51447	Building Information Model Storage Standard	MOHURD	2021
JTS/T 198-1	Unified Standard for Application of Building Information Modeling in Port and Waterway Engineering	MOT	2019
JTS/T 198-2	Standard for Application of Building Information Modeling in Port and Waterway Engineering Design	MOT	2019
JTG/T 2420	Unified Standard for Application of Building Information Modeling in Highway Engineering	MOT	2021
JTG/T 2421	Highway Engineering Design Information Model Application Standards	MOT	2021
JTG/T 2422	Standard for Application of Building Information Modeling in Highway Engineering Construction	MOT	2021
MH/T 5042	Uniform Standard for Building Information Model Application in Civil Transport Airports	CAAC	2020
DB11/T 1069	Design Standard for Information Model of Civil Building	BMCHUD	2014
DB11/T 1610	Standard for Level of Development of Detail Design Model of Civil Building Information Modeling	BMCHUD	2018
DBJ/T 15-142	Unified standard for building information modelling in Guangdong province	GHUCD	2018
DBJ/T 15-160	Standard for BIM Modeling and Delivery of Urban Rail Transit	GHUCD	2019
DG/TJ 08-2201	Building Information Model Application Standard	SMCHURD	2016
T/GZJXC 01	Standard for building information model (BIM) construction in Guizhou Province	GPBDA	2020

DBJ41/T235	Application standard for urban rail transit information model	HNUCD	2020
DG/TJ 08-2311	Application standard for building information modelling in municipal underground space	SHUCC	2020
T/BIAS 8	Standard for BIM technology of assembled buildings with concrete structure	BIAS	2020
Q/CCCC GL501	Unified Standard of Building Information Modeling in Highway Engineering	CCCC	2019
Q/CCCC GL502	Standard for Application of Building Information Modeling in Highway Engineering Design	CCCC	2019
CRBIM 1004	Standard for classification and coding of building information model in railway engineering	CRBA	2014

The compilation of China's national BIM standards commenced in 2012, with the Ministry of Housing and Construction initiating a compilation plan of national BIM standards in the Engineering Construction Standards and Specifications Development and Revision Plan of 2012 (Ministry of Housing and Urban-Rural Development (China) 2012). This plan included the formulation of four BIM standards: Uniform Standard for Building Information Model Application (Ministry of Housing and Urban-Rural Development (China) 2016), Building Information Model Storage Standard (Ministry of Housing and Urban-Rural Development (China) 2021a), Standard for Building Information Model Classification and Coding (Ministry of Housing and Urban-Rural Development (China) 2017d), and Standard for Design Delivery of Building Information Modeling (Ministry of Housing and Urban-Rural Development (China) 2018b). Additionally, the Engineering Construction Standard Specification Development and Revision Plan of 2013 introduced another standard: GB/T 51235-2017 - Standard for Building Information Modeling in Construction (Ministry of Housing and Urban-Rural Development (China) 2017c). Specifically, the standards GB/T51212-2016 (Ministry of Housing and Urban-Rural Development (China) 2016) and GB/T51235-2017 (Ministry of Housing and Urban-Rural Development (China) 2017a) focus on the application of building information models in a uniform and construction context. The Standard for Building Information Model Classification and Coding and the Standard for Design Delivery of Building Information Model (Ministry of Housing and Urban-Rural Development (China) 2017b) concentrate on the naming convention and delivery process of BIM projects. As of now, the Uniform Standard for Building Information Model Application and Standard for Building Information Model Construction Application, Standard for Building Information Model Classification and Coding, and Standard for Building Information Model Design Delivery have been released and are in implementation. The Standard for Building Information Model Storage (Ministry of Housing and Urban-Rural Development (China) 2021b) was published in September 2021 and is set to take effect from February 1, 2022.

The first Chinese engineering construction standard for building information models is the "Uniform Standard for Building Information Model Application" (Ministry of Housing and Urban-Rural Development (China) 2016), which was published by the Ministry of Housing and Construction on 2 December 2016. It outlines fundamental requirements for the application of building information models, serving as the foundational standard in this domain. This standard is also utilized as a basis for the research and formulation of building information model applications and related standards in China. After that, the Ministry of Housing and Construction released the “Standard for building information modelling in construction” (Ministry of Housing and Urban-Rural Development (China) 2017a) on May 4, 2017. This standard stands as the inaugural engineering construction standard for building information model applications in the realm of building construction in China. It introduces requirements for the creation, utilization, and management of building information models, specifically focusing on aspects such as detailed design, construction simulation, prefabrication and processing, schedule management, budget and cost management, quality and safety management, as well as construction supervision and completion acceptance. The "Classification and Coding Standard for Building Information Model" (Ministry of Housing and Urban-Rural Development (China) 2017b) is the third national BIM standard, which was issued by the Ministry of Housing and Construction on the 25th of October 2017. This standard establishes rules for the classification and coding of information related to building construction at all stages, encompassing both the construction phase and the subsequent use of building projects. On the 10th of April 2019, the Ministry of Housing and Urban-Rural Development released another national BIM standard - “Standard for design delivery of building information modelling” (Ministry of Housing and Urban-Rural Development (China) 2018a). This standard primarily focuses on the design phase and outlines the procedures, processes, and outcomes related to BIM delivery. It delineates the preparation and delivery processes, specifies model precision in terms of geometric accuracy and information depth, and presents fundamental requirements for collaboration and application for all stakeholders engaged in engineering and construction.

With the introduction of national BIM standards, local BIM standards have become a focus of study and development. Various provincial and municipal governments in China, such as Beijing, Shanghai, Shenzhen, Sichuan, Chengdu, Guangxi, and Jiangsu, have sequentially unveiled plans for the formulation of local BIM standards. Beijing took the lead by launching DB11T 1069-2014 " Design Standard for Information Model of Civil Building" (Beijing Municipal Commission of Housing and Urban-Rural Development 2014) in 2014, primarily offering principles regarding the depth and delivery of BIM models from a modelling perspective. Shanghai, being an early adopter with a more advanced BIM application, has published the most BIM standards and application guidelines. In 2015, the Shanghai Housing and Construction Commission organized the "Shanghai Building Information Model Application Technology Guide." It outlines 23 BIM application points covering the entire project life cycle to facilitate the practical implementation of BIM applications and assist enterprises in rapidly acquiring BIM technology expertise. Each application point is detailed in terms of its significance, data preparation, operational process, and application results. Subsequently, Shanghai has introduced several BIM standards, including DG/TJ 08-2201-2016 "Building Information Model Application Standard" (Shanghai Municipal Commission of Housing and Urban-Rural Development 2016) and some other local BIM standards specific to municipal roads, municipal water supply and drainage, rail transportation, underground space, and other industries.

With the development of national and local BIM standards, some large enterprises are independently developing their own BIM standards. These initiatives are undertaken by diverse organisations, including developers such as Wanda Group and Greenland Group, design institutes like China Academy of Architecture and Modern Group, and construction companies such as China Construction Group and Construction Engineering Group. In contrast to national and local BIM standards, enterprise-level and project-level BIM standards place a greater emphasis on the practical application of BIM. They are particularly concerned with application requirements related to organizational structure, division of responsibilities, hardware and software configuration, workflow, deliverables, and other on-the-ground considerations (Zhenqing et al. 2016).

Apart from housing construction projects, various types of construction projects, including municipal engineering, rail transportation, and railway engineering, are actively embracing BIM applications. These sectors, characterized by large scale, high investment, and diverse specialities, present a more complex construction management environment (Minsheng et al. 2016). The nature of these projects imposes elevated demands on BIM technology, underscoring the importance of developing BIM standards tailored to these specific domains.

According to the Standardization Law of the People's Republic of China (National People’s Congress of the People’s Republic of China 1988), Chinese standards are categorized into four levels: national standards, industry standards, local standards, and enterprise standards. The ongoing development of BIM standards in China reflects simultaneous progress across all these levels, with certain local standards being released ahead of national ones. Moreover, despite the strong promotion of BIM in Chinese policy, BIM is currently still treated as an innovative technology that is recommended to be adopted rather than mandatory.

2.2.4 OpenBIM standards

Open BIM is an approach proposed by several leading software vendors to enable collaborative design, construction and operation of buildings using open data models based on open standards and workflows (buildingSMART 2018c). Open BIM endeavours to build an interoperable working environment where all project stakeholders can participate in the exchange of building information and share information across the lifecycle of any built-environment asset, regardless of the software they use. This means that team members can use the most appropriate software to meet the requirements of the task and still retain control of their design data and collaborate with other team members. To achieve the goal of Open BIM, a neutral and non-profit international organisation (buildingSMART) has been working to create and adopt open international standards and solutions for infrastructure and buildings to drive the digital transformation of the built asset industry (buildingSMART [no date]). To date, buildingSMART has published five basic standards, twelve candidate standards and several related standards to support the application of Open BIM. Table 3 summarises all the published standards related to Open BIM (buildingSMART [no date]). The remainder of this section will be the content of the five basic standards to analyse the workflow of the Open BIM framework.

Table 3 - Summary of the five basic open BIM standards published by buildingSMART

Name	Standard	Purpose
Industry Foundation Class (IFC)	ISO 16739	Transports information
Information Delivery Manual (IDM)	ISO 29481	Describes processes
BIM Collaboration Format (BCF)	buildingSmart BCF	Change coordination
International Framework for Dictionaries (IFD)	ISO 12006	Mapping of terms
Model View Definitions (MVD)	buildingSmart MVD	Translates processes into technical requirements

1. **ISO 16739**

ISO 16739 was first published by the International Organisation for Standardisation in 2013 for the registration of Industry Foundation Class (IFC) specifications. IFC is an international, platform-neutral, standardized open data schema for describing building information model data, aiming to establish a standard data representation and storage methodology that enables a wide range of software to import and export building data in such a format, thereby facilitating data sharing across different disciplines and different software throughout the lifecycle (Venugopal et al. 2012).

ISO 16739 specifies the conceptual data schema and the structure of the exchange file format, defined in the EXPRESS Data Specification Language and the XML Schema Definition (XSD) Language. In addition, the standard also provides a standard format for data exchange and sharing. Figure 13 presents the architecture of the IFC data schema defined in ISO 16739. The Resource layer contains all resource definitions in individual schemas and the core layer focuses on the kernel and core extension schemas. The interoperability layer focuses on entity definition schemas specific to general products, processes and resources, while the Domain layer concentrates on intra-domain construction information sharing and exchange.

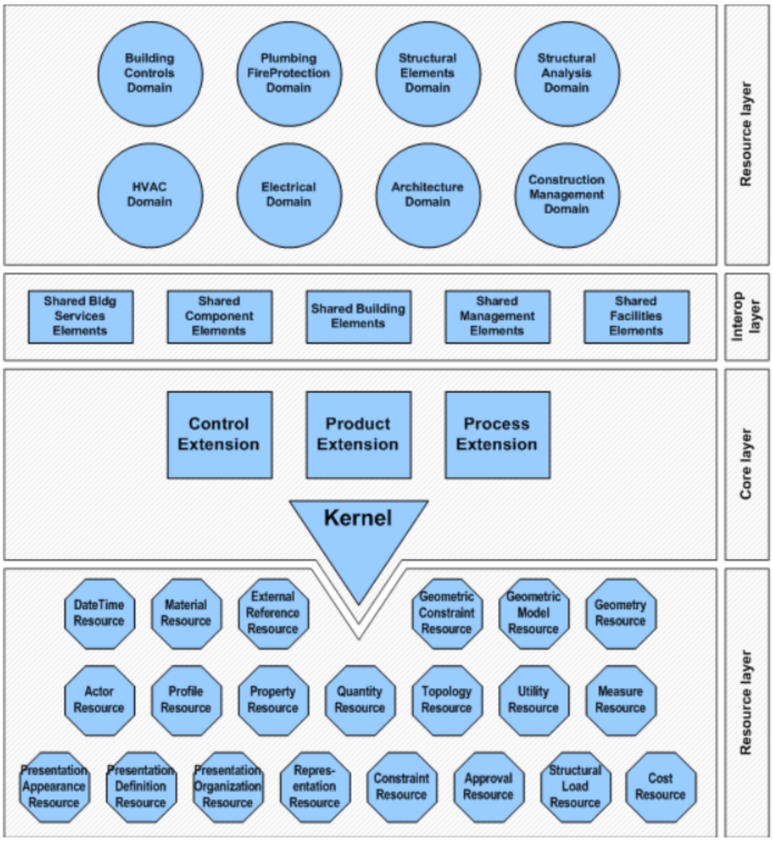


Figure 13 - Architecture of the IFC data schema (International Organization for Standardization (ISO) 2018a)

2. ISO 29481

Motivated by the requirement for dependability in information transmission, ISO 29481 (International Organization for Standardization (ISO) 2016) was released by the Technical Committee ISO/TC 59 in 2016, which is centred on information delivery manuals (IDM). IDM is a technique for recording and outlining every step of the information flow and process during the construction life cycle. With the help of IDM, pertinent data is transferred in a manner that the receiving software can interpret. As a wide range of stakeholders normally are involved in the construction and maintenance of a building, it is critical to understand what information needs to be shared and when it is communicated.

To address this, ISO 29481 (International Organization for Standardization (ISO) 2016) specifies a methodology to facilitate the information communication process undertaken during the facility's construction based on the information requirements and business process modelling notation (BPMN). It provides an overview in terms of manual information delivery development techniques and the related state of implementation in software-readable formats. Additionally, this standard also provides guidelines on requirement identification, information description, reuse of information and information configuration to create a reliable and high-quality information exchange environment.

3. **buildingSmart BCF**

The BIM collaboration format (BCF) is a simplified and open standard XML format designed by buildingSMART to encapsulate data and facilitate workflow communication across various BIM software tools (Jiang et al. 2019). To enable IFC model-based communications between different BIM software tools and applications, the BCF proposes an XML file-based data exchange format, namely bcfXML, and a web-based RESTful (bcfAPI) to enable project participants to coordinate the creation, modification, and management of BCF data in a centralized location (buildingSMART [no date]).

4. **ISO 12006**

The ISO 12006 (International Organization for Standardization (ISO) 2022) was published by the Technical Committee ISO/TC 59 in 2007, whose aim is to establish a language-independent information model to define and aggregate concepts, develop dictionaries for information storage, and delineate relationships between different concepts. This standard concentrates on three types of entities, which are objects, collections, and relationships. It offers a standard framework for categorizing objects, processes, and information systems and models. In addition to the attributes of units, values, actors, topics, activities, and measurements, objects defined in ISO 12006 can also be associated with formal categorization systems, which enables the definition of the context that includes objects. Collections use relationship collection to polymerise several types of things. Eight types of relationships are defined in ISO 12006 to bridge various objects, including property, involvement, association, specialization, composition, association, sequencing, and measure assignment.

5. **buildingSmart MVD**

The Model View Definition (MVD) is defined as a subset of IFC schema, whose purpose is to describe the process of data exchange for a specific project or workflow (buildingSMART [no date]). It is used to describe the specification of data exchange between different stakeholders that is required to fulfil the intended task. Generally, not all information needs to be delivered or received in the design, procurement, construction, and operational processes of a project during the information exchange process. Therefore, MVD can greatly improve the efficiency of information exchange by narrowing the scope of information exchanged according to the needs of a given workflow.

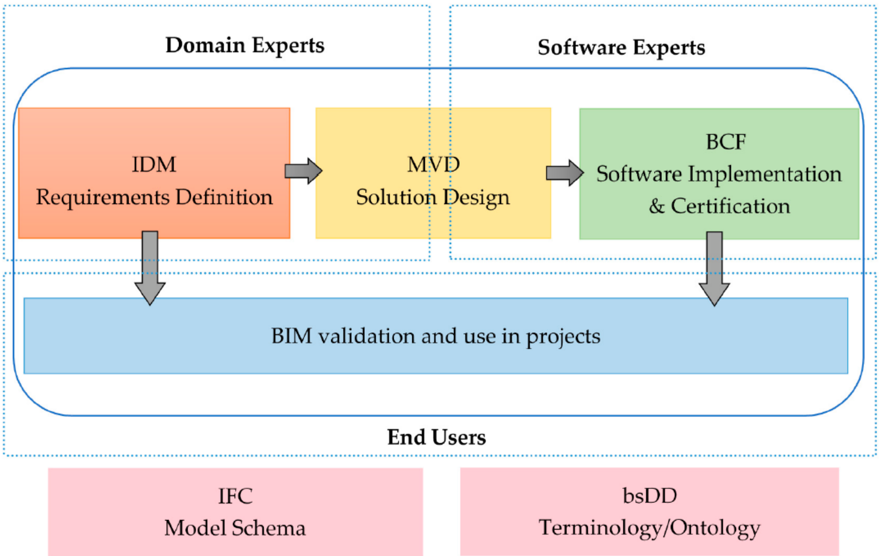


Figure 14 - Process of using the five basic standards for collaborative working (Jiang et al. 2019)

In summary, each of the above buildingSMART's basic standards plays a significant role in sharing and exchanging structured building information openly during the whole life cycle. Figure 14 shows the process of using these five basic standards for collaborative work. With the help of all these standards, the Open BIM framework can achieve not only open data compatibility but also open BIM workflows that are transparent and collaborative. The concept of Open BIM creates a common language for widely referenced processes to provide persistent project data throughout the asset lifecycle.

Through the survey of BIM applications, it is observed that the concept of BIM has been widely accepted and the BIM-driven applications have been popularised globally. Driven by this overall situation, BIM standards have flourished. With more than two decades of accumulation, several well-established BIM standard systems have been developed, the representative of which are the American BIM standard system, the British BIM standard system, the China BIM standard system, and the Open BIM standard system. After a comprehensive review of all the standards, some commonalities and differences among these BIM standard systems were noticed. The Open BIM standard system is committed to building an open and fluent data exchange environment. It specifies a unified data schema (IFC) to define the data involved in the AEC project, a domain dictionary to standardise the terms used in projects and several ancillary documents (IDM, MVD, etc.) to specify what and how information is changed during the delivery of a project. The American NBIMS standard system is a technical standard system developed based on open BIM standards. It uses the data exchange environment built by Open BIM and provides some practice documents and actual projects as templates to guide users in delivering BIM projects in the right way. Also based on the Open BIM standard, China has established its own set of standards. The Chinese BIM standard system draws on the experience of American BIM standards and incorporates the needs of engineering applications within the industry, making it highly practical. It divides the BIM standard into three aspects, which are information sharing, collaborative working, and disciplinary application. Different from the above three systems, the British BIM standard system focuses on the process of information management. It divides the whole process of a project from establishment to acceptance into 8 stages (Figure 12) and defines the tasks to be performed by each stakeholder (appointing party, lead appointed party, appointed party and task team) at each stage of the project and the corresponding requirements. Moreover, the British BIM standard system also establishes its own classification and coding system (Uniclass) to support their workflow of information management. Despite these systemic differences, there are commonalities among these BIM standard systems. They share many common concepts like level of information need (LOIN), Construction Operations Building information exchange (COBie), BIM execution plan (BEP), etc. In addition to the above, there are plenty of concepts that are named differently in different standard systems but perform the same or similar function when delivering BIM projects. For example, organisational information requirements (OIR) and owner's project requirements, exchange information requirements (EIR) and exchange requirements (ER), responsibility matrix (RM) and Roles & Stakeholders, etc. Based on the above comparisons, it is apparent that despite the differences in emphasis, there is uniformity in content and structure across different BIM standard systems. The draft of the fourth edition of the American National BIM Standard (National Institute of Building Sciences buildingSMART alliance 2023), which integrates relevant concepts and principles from the British BIM standards further proves this viewpoint.

2.3 BIM compliance checking

Compliance checking refers to short, focused reviews, typically providing a high-level investigation of the extent to which statistics meet the regulatory documents, such as standards, codes, guidelines, and specifications. In the most basic terms, BIM compliance checking is regarded as a measure of how well an organisation is delivering BIM projects against BIM standards. With the rapid development of BIM technology and standards in recent years, more and more organisations are concentrating on the compliance of their BIM workflows against BIM standards to guarantee the efficiency improvements brought by adopting BIM and further improve their productivity. On the other hand, many governments are now introducing mandatory requirements for construction projects to fully adhere to BIM standards. As a result, compliance checking of BIM is becoming a hotspot in the AEC field. So far, great progress has been made in BIM compliance checking and a number of approaches and frameworks have been proposed by various organisations and academics. The remainder of this chapter will review existing BIM compliance checking methods from the perspectives of industrial and academic respectively and analyse their strengths and weaknesses.

2.3.1 Industrial perspective

The research on BIM compliance in the AEC industry began shortly after the release of the BIM standards. In the early stage, as the BIM standard system was not well developed, the industry proposed the concept of BIM maturity to broadly measure the level of BIM adoption.

The BIM Maturity Levels is considered to be the first BIM maturity assessment tool. It was developed by Bew et al. (Bew and Richards 2008) and used as the main component of the UK BIM implementation strategy. This BIM maturity model was intended to provide an overall view of the technologies and processes that organisations can use for their projects. Figure 15 presents the overarching framework of the BIM Maturity Levels.

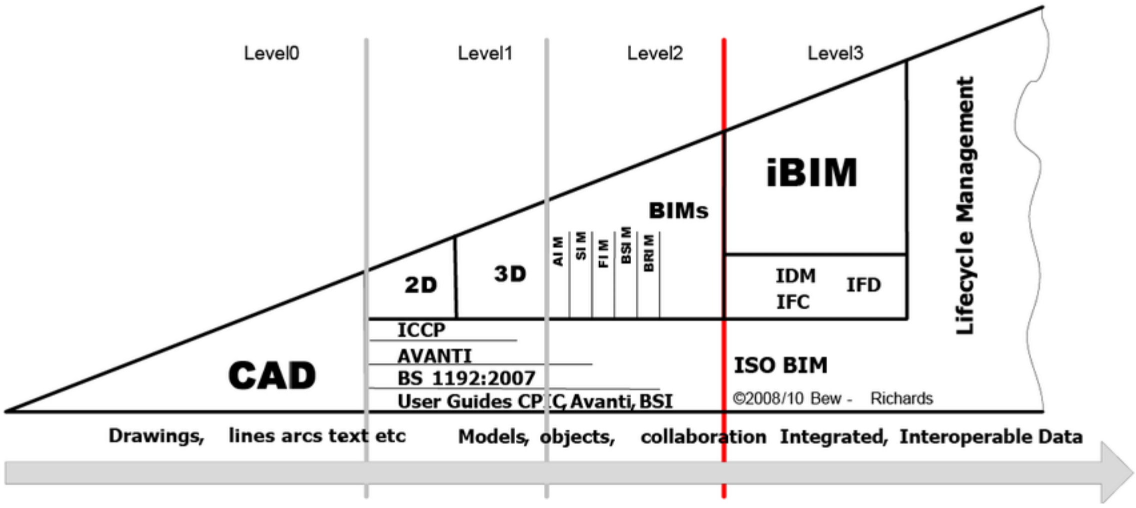


Figure 15 – The four levels of BIM Maturity (Bew and Richards 2008)

This model is straightforward and can be easily understood by most stakeholders. The organisation's compliance with the specifications listed in the model serves as the measurement system for maturity (Ammer et al. 2015). However, this maturity model stays at a low level of granularity and does not assess any specific details, which means it is only capable of a broad, strategic perspective for organisations and governments. Due to the above shortcomings, this model is not widely used in practice. However, as the first BIM maturity assessment model, it provided some basic ideas for subsequent models.

In the same year, another BIM maturity model was established and published as a part of the National BIM Standard of America, namely the Interactive capability maturity model (I-CMM) (National Institute of Building Sciences buildingSMART alliance 2015). It is an organisation-oriented self-assessment tool used for the evaluation of BIM maturity level (Figure 16).

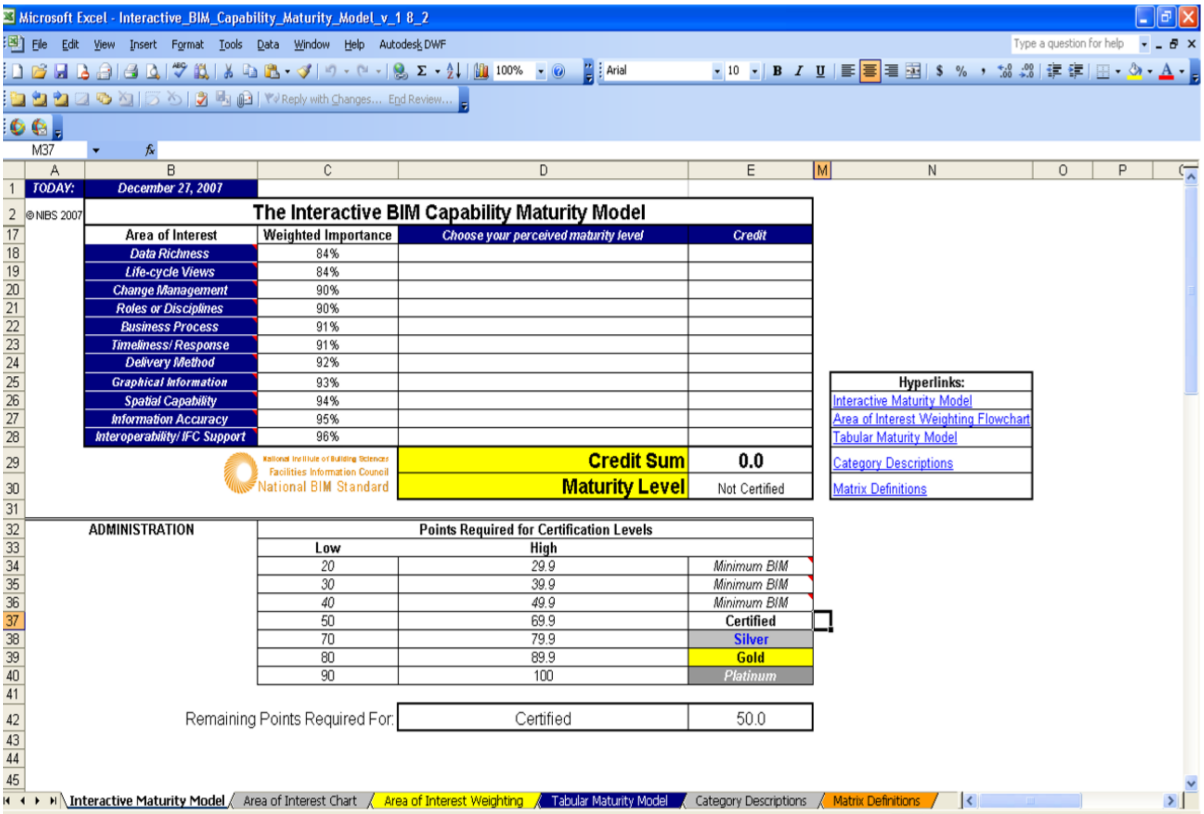
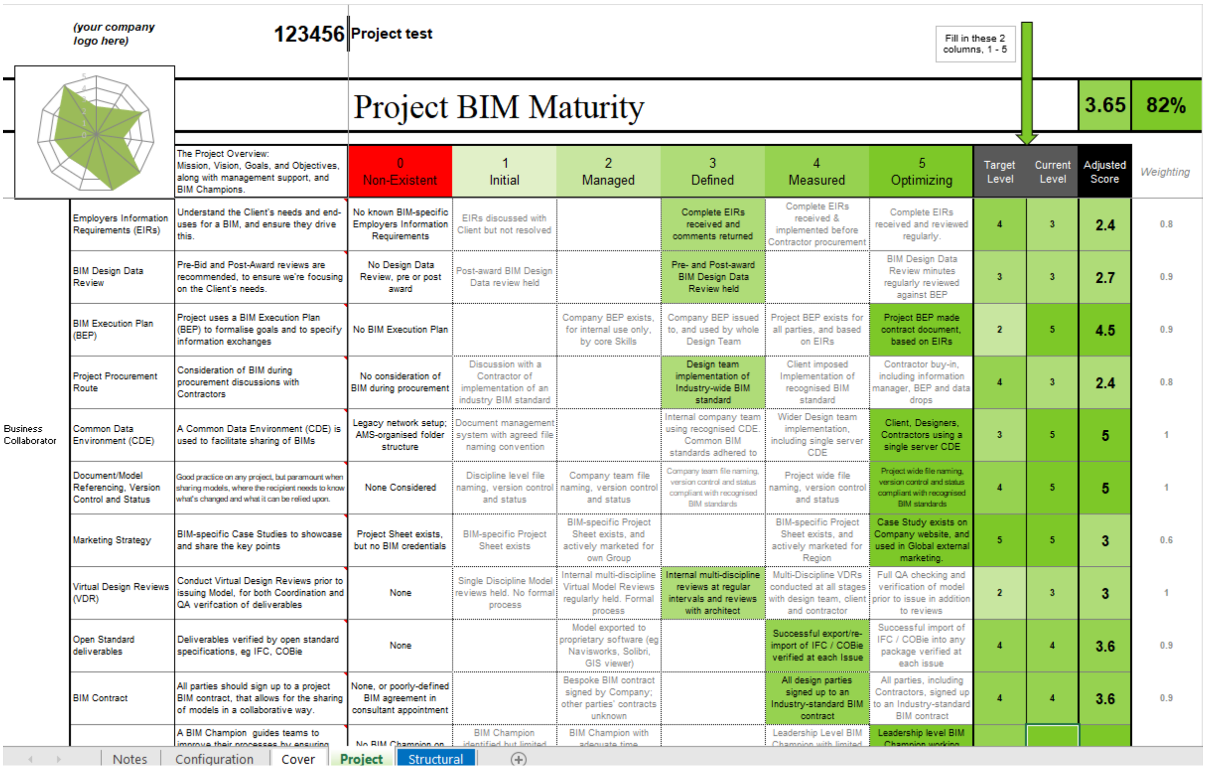


Figure 16 - The Interactive BIM Capacity Maturity Model (National Institute of Building Sciences buildingSMART alliance 2015)

Compared with BIM Maturity Levels, I-CMM is of a higher granularity level. It uses a spreadsheet-based method to help organisations "evaluate their business practices along a continuum of the desired technical level of functionality". The I-CMM used the areas of interest proposed in NBIMS as a basis for the framework establishing a BIM maturity assessment model. Each area of interest is assigned a weight, reflecting its relative importance. Figure 16 illustrates the areas included in the I-CMM model and their corresponding weightings. The user is asked to enter the perceived maturity level of their organisation or the project they delivered. This perception is based on guidance attached to the model. The perceived maturity level is a text-based entry and is chosen from a list. With each list entry is an associated score. Each score is multiplied by its associated weight and the total is added up to give a "Credit Sum". This is a score out of 100 which is compared against the table of certification levels. A maturity level of "Certified", "Silver", "Gold", or "Platinum" is then given based on score ranges of certification levels set in the model.

In 2013, a research group at Penn State produced a series of documents aimed to assist organisations with BIM implementations, and associated planning and strategy, which covers 6 planning elements, including process, information, strategy, infrastructure, uses, and personnel (Computer Integrated Construction Research Program 2013). Inspired by this guidance, Arup developed a spreadsheet-based tool (BIM Maturity Measure) to guide BIM users through the process of organisational assessment (Arup 2015). This tool (Figure 17) assesses BIM use in 25 areas to establish an overall view of an organisation's BIM implementation strategy.


(your company logo here)
123456 Project test
Fill in these 2 columns, 1 - 5
Project BIM Maturity
3.65 82%

		The Project Overview: Mission, Vision, Goals, and Objectives, along with management support, and BIM Champions.	0 Non-Existent	1 Initial	2 Managed	3 Defined	4 Measured	5 Optimizing	Target Level	Current Level	Adjusted Score	Weighting
	Employers Information Requirements (EIRs)	Understand the Client's needs and end-uses for a BIM, and ensure they drive this.	No known BIM-specific Employers Information Requirements	EIRs discussed with Client but not resolved		Complete EIRs received and comments returned	Complete EIRs received & implemented before Contractor procurement	Complete EIRs received and reviewed regularly.	4	3	2.4	0.8
	BIM Design Data Review	Pre-Bid and Post-Award reviews are recommended, to ensure we're focusing on the Client's needs.	No Design Data Review, pre or post award	Post-award BIM Design Data review held		Pre- and Post-award BIM Design Data Review held		BIM Design Data Review minutes regularly reviewed against BEP	3	3	2.7	0.9
	BIM Execution Plan (BEP)	Project uses a BIM Execution Plan (BEP) to formalise goals and to specify information exchanges	No BIM Execution Plan		Company BEP exists, for internal use only, by core Skills	Company BEP issued to, and used by whole Design Team	Project BEP exists for all parties, and based on EIRs	Project BEP made contract document, based on EIRs	2	5	4.5	0.9
	Project Procurement Route	Consideration of BIM during procurement discussions with Contractors	No consideration of BIM during procurement	Discussion with a Contractor of implementation of an industry BIM standard		Design team implementation of Industry-wide BIM standard	Client imposed Implementation of recognised BIM standard	Contractor buy-in, including information manager, BEP and data drops	4	3	2.4	0.8
Business Collaborator	Common Data Environment (CDE)	A Common Data Environment (CDE) is used to facilitate sharing of BIMs	Legacy network setup; AMS-organised folder structure	Document management system with agreed file naming convention		Internal company team using recognised CDE. Common BIM standards adhered to	Wider Design team implementation, including single server CDE	Client, Designers, Contractors using a single server CDE	3	5	5	1
	Document/Model Referencing, Version Control and Status	Good practice on any project, but paramount when sharing models, where the recipient needs to know what's changed and what it can be relied upon.	None Considered	Discipline level file naming, version control and status	Company team file naming, version control and status	Company team file naming, version control and status compliant with recognised BIM standards	Project wide file naming, version control and status	Project wide file naming, version control and status compliant with recognised BIM standards	4	5	5	1
	Marketing Strategy	BIM-specific Case Studies to showcase and share the key points	Project Sheet exists, but no BIM credentials	BIM-specific Project Sheet exists	BIM-specific Project Sheet exists, and actively marketed for own Group		BIM-specific Project Sheet exists, and actively marketed for Region	Case Study exists on Company website, and used in Global external marketing.	5	5	3	0.6
	Virtual Design Reviews (VDR)	Conduct Virtual Design Reviews prior to issuing Model, for both Coordination and QA verifcation of deliverables	None	Single Discipline Model reviews held. No formal process	Internal multi-discipline Virtual Model Reviews regularly held. Formal process	Internal multi-discipline reviews at regular intervals and reviews with architect	Multi-Discipline VDRs conducted at all stages with design team, client and contractor	Full QA checking and verification of model prior to issue in addition to reviews	2	3	3	1
	Open Standard deliverables	Deliverables verified by open standard specifications, eg IFC, COBie	None		Model exported to proprietary software (eg Navisworks, Solibri, GIS viewer)		Successful export/re-import of IFC / COBie verified at each Issue	Successful import of IFC / COBie into any package verified at each issue	4	4	3.6	0.9
	BIM Contract	All parties should sign up to a project BIM contract, that allows for the sharing of models in a collaborative way.	None, or poorly-defined BIM agreement in consultant appointment		Bespoke BIM contract signed by Company; other parties' contracts unknown		All design parties signed up to an Industry-standard BIM contract	All parties, including Contractors, signed up to an Industry-standard BIM contract	4	4	3.6	0.9
		A BIM Champion guides teams to improve their processes by ensuring	No BIM Champion on	BIM Champion identified but limited	BIM Champion with adequate time		Leadership Level BIM Champion with limited	Leadership level BIM Champion working				

Notes | Configuration | Cover | Project | Structural

Figure 17 - BIM Maturity Measure Tool developed by Arup (Arup 2015a)

As shown in Figure 17, the assessment of BIM maturity takes the form of a multiple-choice questionnaire. The questionnaire is generated from the Penn State BIM planning guide and tailored for different disciplines. Similar to the I-CMM, Arup's BIM Maturity Measure tool (Arup 2015a) adopts a weighted average as the calculation mechanism. The final assessment result is determined by summation of the indicator scores multiplied by their corresponding weights. Compared with other above-mentioned BIM maturity models, this tool is highly detailed and covers all levels of organisational structure, which allows organisations to clearly understand how well they are performing (in relation to BIM) and how they can improve.

BIM-profiler (BIMconnect 2019) is another BIM compliance assessment tool developed based on research of Penn State. It is a free web-based tool for self-assessing an organisation's BIM competency. The format of the tool is a series that covers the same six key areas of the study: strategy, BIM uses, process, data structure, infrastructure, and personnel. A total of 20 questions are asked over the six key areas. Each question has a choice of pre-written text-based responses and allows the person surveyed to set their organisation's current score, and their target score (Figure 18).

Figure 18 - Part of BIM-profiler questionnaire section (BIMconnect 2019)

Attaining an accurate readiness score is dependent on a deep knowledge of the organisation and subjective judgment when responding to the questions. The results from the questionnaire are shown in 2 different formats (Figure 19). The first is a percentage rating, giving a 1-dimensional analysis of the organisation's BIM readiness. The second is a radar chart which shows the degree of readiness in the 6 key areas.

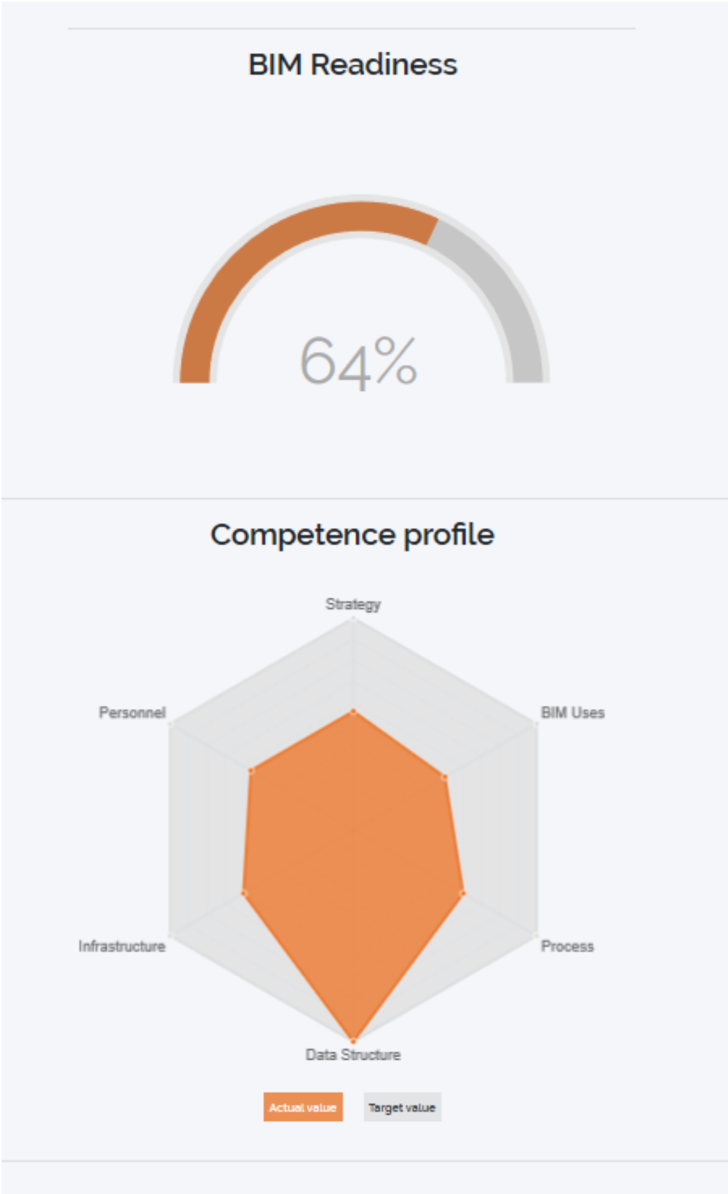


Figure 19 - BIM-profiler results output (BIMconnect 2019)

Compared with other models or tools, BIM-profiler (BIMconnect 2019) is the first web-based BIM compliance checking tool. However, similar to the I-CMM model, this tool is deficient in the level of granularity. It simply indicates the strengths and weaknesses of an organisation's BIM strategy and does not make any comment on whether an organisation is compliant with any particular BIM/collaboration standard.

In terms of compliance checking for BIM standards, there are some standard or research bodies providing standard-oriented BIM compliance checking services, for example, Building Research Establishment (BRE) and British Standards Institution (BSI). During the assessment, a dedicated BIM expert will be assigned to manually check the status of the BIM adoption on-site and assess the level of compliance with the standard by ticking a box on a requirement form.

In addition to the representative BIM maturity models and tools mentioned above, there are many other frameworks produced on BIM maturity assessment. Despite the variances in comprehensiveness and body of knowledge, these approaches share some commonalities. First, they all involve evidence-based approaches to demonstrate that compliance with standards is met. Second, most of them use spreadsheets as the knowledge carrier and questionnaires to capture the evidence of enterprises' BIM adoption. Finally, all these approaches are highly involved with manual labour, either in completing questionnaires or in expert auditing.

2.3.2 Academic perspective

Apart from industry, academia has also made great efforts in BIM compliance checking. For example, Berlo et al. (Van Berlo et al. 2012) developed a framework called BIM Quickscan to evaluate the current BIM performance of a company, which covers both quantitative and qualitative assessments of the 'hard' and 'soft' aspects of BIM. Yilmaz et al. (Yilmaz et al. 2023) proposed the BIM Capability Assessment REference Model (BIM-CAREM) through iterative expert reviews and exploratory case studies(Yilmaz et al. 2023). In witness to the extensive labour involved in industrial methods, academics began to explore automated compliance checking solutions. Typically, compliance checking can be roughly divided into three stages, namely rule interpretation, evidence acquisition and matching-based inspection (Sara Ismail et al. 2017). Based on this perspective, relevant research in academia mainly focuses on these three objectives.

For rule interpretation, existing methods can be broadly classified into three categories: rule-based method, machine learning (ML) based approach and ontology-based approach (Zhong et al. 2020). Rule-based methods were the first to be proposed and are now gradually being replaced by machine learning or combined with other methods. The rule-based approach refers to the capture of required factual information in text using predefined rules (Zhang and Liu 2014). Many efforts focused on the rule-based approach because of its fast speed and strong interpretability. Feijo and Krause (Feijó et al. 1994) combined hypertext with graphs and sentences in mathematical logic to navigate regulatory documents. Boukamp and Akinci (Boukamp and Akinci 2007) introduced an approach for automating the processing of construction specifications to support inspection and quality control tasks in construction projects. Hjelseth and Nisbet (Hjelseth and Nisbet 2011) proposed a semantic mark-up (RASE) methodology to capture normative constraints in target regulatory documents. Beach et al. (Beach et al. 2015) developed a rule-based semantic approach to extract regulations from textual documents by annotating regulatory documents. Li et al. (Li et al. 2016) proposed an information extraction method for utility regulations based on chunk-based rules. Lau and Law (Lau and Law 2004) developed a shallow parser that can consolidate different formats of regulations into extensible mark-up language (XML) to semi-automate the rule translational process. The ML-based approach uses ML algorithms such as decision trees (DT), Support Vector Machines (SVM), Hidden Markov Models (HMM), Conditional Random Fields (CRF), or artificial neural network (ANN) to learn the extraction pattern of rules from training texts (Tierney 2012). Unlike rule-based approaches, ML-based approaches perform better in acquiring semantic features and provide more flexibility in pattern recognition. Therefore, this method became another popular information extraction solution and has been widely utilised in related research. For example, Zhang and El-Gohary (Zhang and El-Gohary 2020) proposed an LSTM-based method to generate semantically enriched building-code sentences, which achieves an accuracy of 87% on their domain dataset. Phillip Schönfelder and Markus König (Schönfelder and König 2021) developed a Transformer-based deep learning model to extract entities in German public construction laws. A hybrid bi-directional long and short-term memory (BiLSTM) and convolutional neural network (CNN) model was proposed by Wang and El-Gohary (Wang and El-Gohary 2021), which can automatically identify entities in building safety regulations. Zhu and Li (Zhu et al. 2022) proposed an LSTM-based neural network model that can automatically recognise the qualitative rule sentences from engineering standards. An attention-based convolutional neural network model that can identify and classify relations mentioned in regulation documents was proposed by Wang and El-Gohary (Wang and El-Gohary 2022) in 2022. The ontology-based approach has recently emerged as a new solution for automated rule interpretation, which utilises the domain or application ontology to aid in extracting related semantic information from target documents (Moreno et al. 2013). Zhang and El-Gohary (Zhang et al. 2013) proposed a semantic approach, where extraction patterns are composed of a variety of syntactic and semantic features (captured via a domain ontology) to automatically extract information from building codes. Zhou and El-Gohary (Zhou and El-Gohary 2017) proposed a rule-based ontology-enhanced information extraction method for extracting building energy requirements from energy conservation codes and then formatted the extracted information into a B-Prolog representation. Additionally, Xu and Cai (Xu and Cai 2021) proposed an ontology and rule-based framework to automate the interpretation of utility regulations into deontic logic (DL) clauses, where pattern-matching rules are used for information extraction; pre-learned model and domain-specific hand-crafted mapping methods were also adopted for semantic alignment between rules and ontology.

In terms of automatic evidence acquisition, it normally focus on extracting entities and relationships from documents. For instance, Liu and El-Gohary (Liu and El-Gohary 2017) proposed an automated information extraction method for bridge inspection reports based on CRFs. Feng and Chen (Feng and Chen 2021) proposed a deep learning-based small sample train framework, which used the LSTM model to extract event-related information (e.g., date, location, and accident type) from accident news reports.

With regards to matching-based inspection, most of the research relies on SWRL rules, Semantic Query-Enhanced Web Rule Language (SQWRL) or RDF Query Language (SPARQL) to check the compliance between evidence and target standards. For instance, Yurchyshyna and Zarli (Yurchyshyna and Zarli 2009) proposed an ontology-based approach that can extract norms from the electronic regulations and formalise them as SPARQL queries in the IFC model. Moreover, Demir et al. (Demir et al. 2010) developed a semantic web-based approach to generating mappings between the vocabulary and the standardised building models in 2010. After that, Xu and Cai (Xu and Cai 2020) developed a semantic-based system, which can integrate heterogeneous data and achieve automatic compliance checking of underground utilities. Zheng et al. (Zheng et al. 2022) proposed an ontology-driven approach, which identified clauses with the help of predefined classes and properties in domain ontology, then these clauses were parsed and transformed into SPARQL queries by pattern-matching rules.

Despite significant efforts to automate compliance checking, existing methods still cannot eliminate manual work by domain experts. There is still a deficiency in the automation of regulatory knowledge interpretation. More importantly, existing studies of automated compliance checking mainly focus on quantitive requirements in technical standards, such as building design codes, energy conservation codes or fire and safety codes. Automated compliance checking for qualitative requirements (such as quality assurance, construction management and process management) currently remains a vacuum in academia due to complex logic and flexible relations in these standards.

2.4 Advanced techniques for BIM compliance checking

Compared with quantitive requirements in technical standards, the interpretation of qualitative or management requirements is much more challenging as these types of standards often contain complex logics and the entities and relations within them are flexible and varied. For example, (1) “the protection layer shall use non-combustible material and the thickness of the protection layer shall not be less than 10mm.” and (2) “the appointing party shall establish the requirements that tendering organizations shall meet within their tender response.” The first clause is taken from the International Building Code (IBC) (International Code Council 2018), which is a quantitive technical standard for building design. The second clause is extracted from ISO 19650-2 (International Organization for Standardization (ISO) 2018c), which is management standards focusing on the information delivery process. Compared to the first clause, the second clause is more complex in syntax and the relations between concepts are more flexible. In addition, the concepts in these clauses are dynamic and change with the project, such as the appointing party and tendering organizations. Therefore, correctly resolving complex logical relationships in qualitative clauses and establishing mappings between abstract concepts and dynamic entities are the challenges in the automation of BIM compliance checking. After reviewing a large number of related studies in other research areas, two advanced techniques are introduced into this study, which are promising to tackle the above difficulties.

2.4.1 Natural language processing

Natural language processing (NLP) is a discipline that combines computer science and linguistics. It provides computers with the ability to “understand” and process human language in the form of text via integrating computational linguistics with statistical, machine learning, and deep learning models (Nadkarni et al. 2011). The development of the NLP can be broadly divided into three phases, which are symbolic NLP (1950s -1990s), statistical NLP (1990s-2010s) and neural NLP (2010s-present) (Wikipedia 2024a).

Research on processing natural language can be traced back to the 1950s. At that time, the understanding of computer processing of natural language was limited to the way humans learn language. In the 1960s, Noam Chomsky, an American linguist, absorbed the idea of the finite state Markov process from Shannon's work and put forward the finite state model of natural language(Chomsky 1956a). He established the mathematical models of finite state grammar, context-free grammar, context-sensitive grammar, and type-0 grammar, trying to describe the infinite linguistic phenomena by using the finite rules, discovering the universal language mechanism of human beings, and establishing the "Universal Grammar" (Chomsky 1965). Influenced by this Transformational Generative (TG) grammar, it is widely recognised in academia that a machine must first be made to understand language so that it can perform NLP tasks (Khurana et al. 2023). Therefore, analysing utterances and obtaining semantics become the first task, which relies mainly on the manual summarisation of grammatical rules by linguists. However, human language is complex and flexible, which cannot be fully covered by hand-written grammar rules alone, and there may also be contradictions between the rules. Therefore, the performance of this symbolic NLP approach was not ideal. After more than two decades of stalemate, NLP research has seen a breakthrough after the introduction of the corpus approach in rule-based techniques, which is a sign of the birth of statistical NLP (Wikipedia 2024a). These Statistical-based NLP approaches are centred on the " Hidden Markov Model (HMM)", where the inputs and outputs are one-dimensional sequences of symbols and the original order is maintained (e.g., speech recognition, lexical analysis). Thereafter, research in natural language processing has changed considerably. Probabilistic and data-driven approaches have become almost standard in natural language processing. Probabilities are starting to be introduced into standard NLP tasks, such as syntactic parsing, speech tagging and co-reference resolution. Concerning NLP applications such as machine translation, text classification, information retrieval, question and answer systems, information extraction, linguistic knowledge mining, etc. (Khurana et al. 2023), statistical-based empiricist methods have gradually become mainstream. After the significant increase in the speed and storage capacity of computers in the 1990s, artificial neural network (ANN) based deep learning techniques were popularised. More and more NLP researchers are shifting their attention to deep learning methods, and various neural network models are being introduced to natural language processing tasks (e.g., recurrent neural networks - RNNs and convolutional neural networks – CNNs) (Sun et al. 2020). In 2017, Google proposed the Attention mechanism on the basis of previous work, providing a new idea for text encoding (Vaswani et al. 2017). The Transformer structure proposed in this study further guided the development of subsequent large NLP language models.

Currently, the research direction of natural language processing is focused on four main aspects: linguistic theory, text feature mining, language comprehension, and language cognition and generation (Sun et al. 2020). Common tasks for natural language processing include the following seven types (Wang and Yu 2021) :

- **Syntactic semantic analysis**: segmentation, lexical tagging, named entity recognition and linkage, syntactic analysis, semantic role recognition, and polysemous disambiguation for a given sentence.
- **Information extraction**: extracting important information from a given text, such as time, place, people, etc., which involves techniques such as entity recognition, time extraction, and causality extraction.
- **Text mining**: text clustering, classification, information extraction, sentiment analysis, summarization, and visual, interactive representation of extracted information and knowledge.

- **Machine translation**: transforming the input language text into another language text through automatic translation, which can be classified into text translation, speech translation, graphic translation, etc.
- **Information Retrieval**: Finding matching candidate documents in a large-scale document. After analysing the target search term or sentence, the algorithm sorts the candidate documents through a sorting mechanism and outputs the highest-scoring document in the index.
- **Question and Answer System**: forming logical expressions via semantic analysis of natural language query statements, and finding the best answer in the knowledge base through a sorting mechanism.
- **Dialogue system**: The system chats, answers, and completes a certain task with the user through a series of dialogues, involving technologies such as user intent understanding, generic chat engine, Q&A engine, and dialogue management.

So far, several studies in the field of AEC have attempted to use NLP techniques to address the relevant challenges and some of them have yielded superior results. For example, the studies of Zhang et al. (Zhang et al. 2013), Zhou et al. (Zhou et al. 2022b), and Xu and Cai (Xu and Cai 2021) have all involved NLP-based syntactic analysis methods. In addition, Zhang and El-Gohary (Zhang and El-Gohary 2015) proposed a semantic NLP-based approach named Regex-E, which annotates text in the building codes with the help of POS tags and domain ontologies and uses semantic mapping to transform single-requirement to logical clauses. Zheng et al. (Zheng et al. 2022) proposed a knowledge-informed framework, which identified clauses with the help of predefined classes and properties in domain ontology, and then these clauses were parsed and transformed into SPARQL queries by pattern-matching rules. Zhou et al. (Zhou et al. 2022b) developed an automated rule extraction method based on a deep learning model and a set of context-free grammars (CFGs), which can transfer textual regulatory rules into pseudocode formats. Chi et al. (Chi et al. 2017) propose a semi-automated approach to develop a gazetteer for construction safety management through an NLP-based Part of Speech Tagging and Chunking. Beyond the above studies, information extraction based on large language models (LLMs) is also becoming a growing trend. Currently, there are no cases in the AEC domain using large language models to extract information. Only Xue and Zhang (Xue and Zhang

2021) proposed a method to annotate building codes with part-of-speech (POS) tags generated by the BERT-base model. However, this LLM-driven approach has been proven effective in other domains. Taking BERT as an example, Shi and Lin (Shi et al. 2019) proposed a BERT-based deep learning model to extract relationships and annotate semantic roles from newswires and web text. Han and Wang (Han and Wang 2020) proposed a document-level entity mask method to recognise entities and their relationships from open-domain documents. Kim and Lee (Kim and Lee 2020) extracted clinical entities from diagnosis texts using a pre-trained BERT model. Chantrapornchai and Tunsakul (Chantrapornchai and Tunsakul 2021) developed a BERT-based model to classify and summarize tourists' reviews and extract the desired entities and relations. Qiao et al. (Qiao et al. 2022) proposed a BERT-based joint extraction model to extract entity relationships from the Agricultural Thesaurus. Wang et al. (Wang et al. 2022a) developed the Financial Regulation BERT (FR-BERT) model for relation extraction in the financial regulation field.

2.4.2 Graph learning

Graph (Figure 20) is a type of mathematical structure that is comprised of a set of nodes (or vertices) and edges (or links). A graph can be represented as a list of triples with nodes and their edge. The attributes of nodes, edges or the whole graph can be represented as vectors to carry more information for the graph (Jia et al. 2023). The graph can be divided according to the following principles:

(1) directed or undirected graphs, depending on whether the edges in the graph are directed or undirected.
(2) homogeneous or heterogeneous graphs, depending on whether the types of nodes or edges in the graph are various (Wang et al. 2019).
(3) static or dynamic graphs, depending on whether the attributes or the topologic structure of the graph change with time.

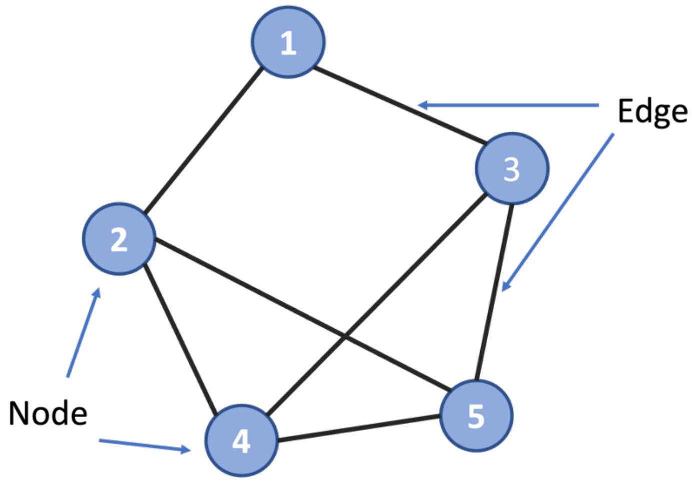


Figure 20 - An example of a graph with nodes and edges

In real life, graph data structures have been applied to describe information in numerous application areas, such as social networks, website links, road maps (Jiang and Luo 2022), protein molecules (Gligorijević et al. 2021), knowledge representation and reasoning (Wu et al. 2023a), etc.

Graph learning (also known as graph machine learning) refers to extracting the desired features of nodes, and edges in a graph or the features of the whole graph through analysing the structure of a graph. The majority of graph learning models or algorithms use deep learning techniques to encode graph data into low-dimensional embeddings in continuous space. These vectorized representations of a graph can be easily applied to downstream tasks, such as node classification, link prediction, graph completion, and entity alignment (Xia et al. 2021).

Although the concept of graph learning originated in computer science, graph learning techniques are now widely used in a variety of domains to solve customised tasks. In the construction domain, there have been some use cases for graph learning. Yan et al. (Yan et al. 2019) proposed a novel graph convolution method to learn the representation of building groups based on the topographic data of individual buildings. Nauata et al. (Nauata et al. 2020) developed a graph-based generative adversarial network (GAN) model, namely House-GAN, which can automatically generate floor plans based on a conceptual bubble diagram. Kim and Kim (Kim and Kim 2020) conducted a comparison analysis between the graph-based classification model and the image-based classification model for bridge component inspection. Hong et al. (Hong et al. 2021) proposed a graph-based method to find the efficient construction method patterns from historic projects through a graph convolutional network (GCN). Based on the same model, Feng et al. (Feng et al. 2022) introduced an approach for pavement crack detection based on high-density 3D point clouds collected by a mobile laser scanning (MLS) system. Wang et al. (Wang et al. 2022b) proposed a semantic enrichment approach for room type classification, which converts a BIM model into a layout graph and generates embeddings of each room through a graph neural network (GraphSAGE) for classification. In terms of knowledge engineering in the AEC domain, Wu et al. (Wu et al. 2023b) proposed a domain information-enhanced graph neural network (D-GNN) model to identify missing triples in the ontology of project constraints.

From the perspective of the AEC domain, a directed, heterogeneous graph consisting of entities (nodes) and relations (edges) is well suited to represent regulatory knowledge in qualitative standards and the information in project documents (Zhou et al. 2020). The above graph learning-related research has partially proven the feasibility of adopting graph-based techniques to address challenges in AEC-related research.

2.5 Summary of literature findings

After a detailed review from both industry and academic perspectives, the existing BIM compliance checking approach suffers from a variety of shortcomings.

- For industrial methods, most of them are inadequate in the level of granularity, which can only provide a high-level assessment for BIM compliance. Some of the methods are also deficient in the comprehensiveness of the assessment, only focusing on compliance with a certain BIM standard or area. Furthermore, the majority of industrial methods implement the evaluation through questionnaires filled out by users. The questionnaires do not require any supporting evidence and are completely disassociated with the actual project. Therefore, these methods require a great deal of manual labour in completing the questionnaires for different projects and the objectivity and reliability of the assessment results are inadequate due to the lack of supporting evidence.

- Additionally, huge efforts have been made by academics and several automated compliance checking methods have been proposed through rule-based algorithms, deep learning, and ontologies. These approaches reduce the amount of manual work spent on checking compliance but still rely on domain experts to define rules or knowledge models. In addition, current methods perform poorly in generalisation, and each method can only be applied to check compliance against a specific standard. Another significant gap in existing methods is that they only enable compliance checking for quantitative standards (codes or specifications). Regarding the complex logic and diverse entities in the qualitative standards, existing methods are not capable of parsing the relationships and establishing mappings among the entities. As a representation of qualitative standards, BIM standard-oriented automated compliance checking methods are currently still a vacuum.

In search of solutions for analysing complex logic and mapping dynamic entities, some related studies in other directions of the AEC field are reviewed. Through these related studies, some approaches were found to be theoretically capable of addressing these two challenges.

- For complex regulatory knowledge interpretation, several researches have adopted NLP techniques to analyse, and process information in textual documents. The result of these researches has proven that the effectiveness of applying NLP techniques to address the AEC-related tasks is desirable.

- In terms of automatic mapping of dynamic entities, graph learning is the most promising solution to address this challenge. Generally, the entities in the actual project are different from the concepts in the standard, and the entities in the project change with the project (e.g., companies and people). Despite the difference in naming, the tasks they perform and the roles they undertake in the project are similar or the same. Therefore, the mapping between the concepts in the standard and the entities in the project is established based on their relations to other concepts or entities, rather than relying on their semantic information. After transforming standards and projects into a graph, the relationships between concepts (entities) are represented as edges. In this case, the structural features of each entity can be captured through graph learning and represented as a vectorised embedding. The mapping of dynamic entities can be achieved by calculating the similarity of node embedding.

Chapter 3. Research Methodology

On the basis of the previous review findings, this chapter will initially introduce the overarching methodology through which this research was conducted. The methodology serves as a roadmap for the entire research process and an explanation of the principles and methods utilised. Following this, a more detailed approach for each research question will be discussed including the specific research theories and strategies.

3.1 Research philosophies

Research philosophy refers to a system of beliefs and assumptions about the development of knowledge (Saunders et al. 2019). In this research, epistemology is adopted as the core research philosophy, which is a part of the “research onion” (Dworkin et al. 2009). The vivid model (Figure 21) describes research methodologies from the outer layer to the inner layer according to the logic of the research and the way of thinking (Crotty 1998).

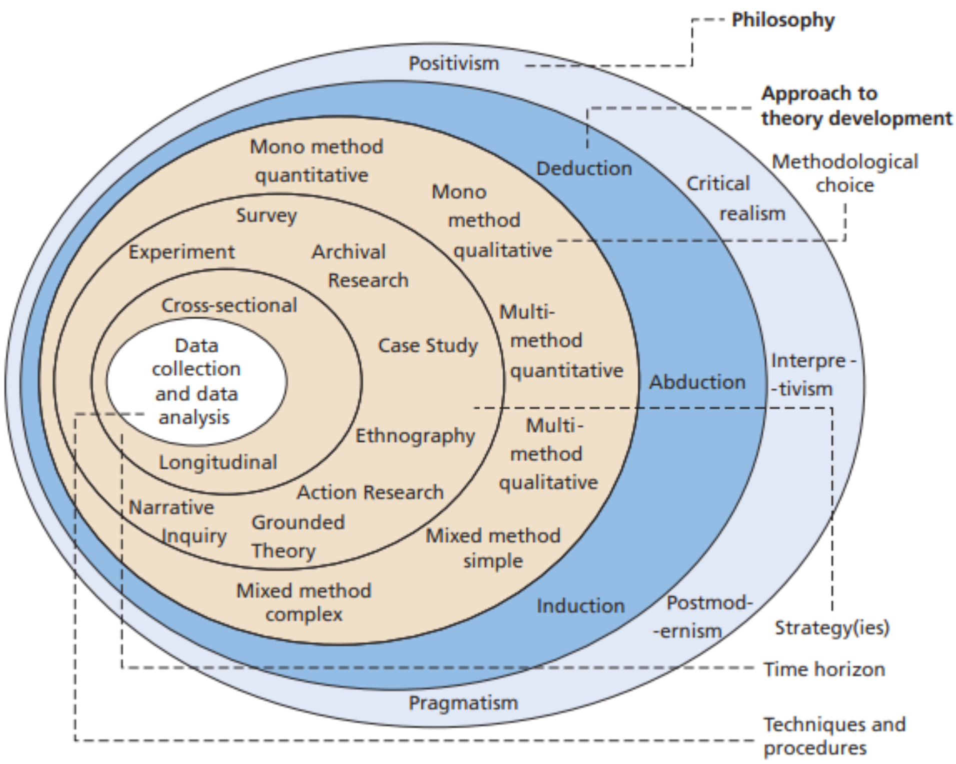


Figure 21 - Research “onion” (Melnikovas 2018)

In the outer layer of the “onion”, philosophical stances applied to different research are broadly classified into five main categories, namely positivism, realism, interpretivism, objectivism, and pragmatism. In particular, positivism is a philosophical thinking centred on "practical verification". It boils down the task of philosophy to the investigation of phenomena, arguing that scientific laws or knowledge can be obtained by induction on natural phenomena (Craib 1997). Comparable to the method and beliefs of positivism, realism emphasizes the independence of social reality and research. It does not consider the scientific method as a completely perfect solution and believes the theories proposed by humans can be revised in accordance with variations of reality. In realism, all beliefs and individuality in human values will be greatly valued. In opposition to positivism and realism, the interpretive method stresses the significance of human interaction in social science, which is significant to the research in the information system domain (Walsham 1995). Pragmatism overturns the previous hypothesis that thought is meant to describe and represent reality. It believes that human language and thought have limitations in reflecting the real world, which can only be used as tools for solving problems (William 1909). In addition, pragmatism argues that a singular method is not adequate to summarise the actual reality and mixed research should be incorporated to reach the predicted results. Objectivism and constructivism are two research methods derived from the perspective of pragmatism. Objectivism facilitates research to understand the various meanings that social phenomena may have on social actors, while constructivism takes the opposite stance, arguing that social phenomena are constructed by social actors (Avenier 2010). Given that the research carried out in this dissertation is comprised of both positivistic and interpretive elements, the pragmatist research philosophy has aligned with the research project via the analysis of the research questions.

3.2 Research design

Although the model of research onion has been widely accepted and used as guidance on constructing methodologies for various research projects (Melnikovas 2018), it has certain limitations when confronted with disciplines like engineering or natural sciences. As a model designed for business research, the “research onion” is more suitable for qualitative or mixed methods research. However, research in the engineering domain is usually dominated by quantitative methods and the design of experiments. Considering this research is carried out in the field of information processing and requires a mixed method to achieve the objectives, the Design Science Research (DSR) methodology was selected as the research strategy. The DSR methodology is developed based on the research onion but focuses more on the development and performance artefacts (Johannesson and Perjons 2014). Thus, this methodology is typically employed for categories of artefacts related to engineering and computer science disciplines to address a generic challenge encountered in practice (Hevner and Chatterjee 2010). Figure 22 indicates the six steps in the nominal process of DSR, which are: 1) problem identification; 2) objectives definition; 3) design and development; 4) demonstration; 5) evaluation; and 6) communication. To achieve comprehensive and smart BIM compliance checking, this research follows the principles of design science research (DSR) to construct the overarching research framework.

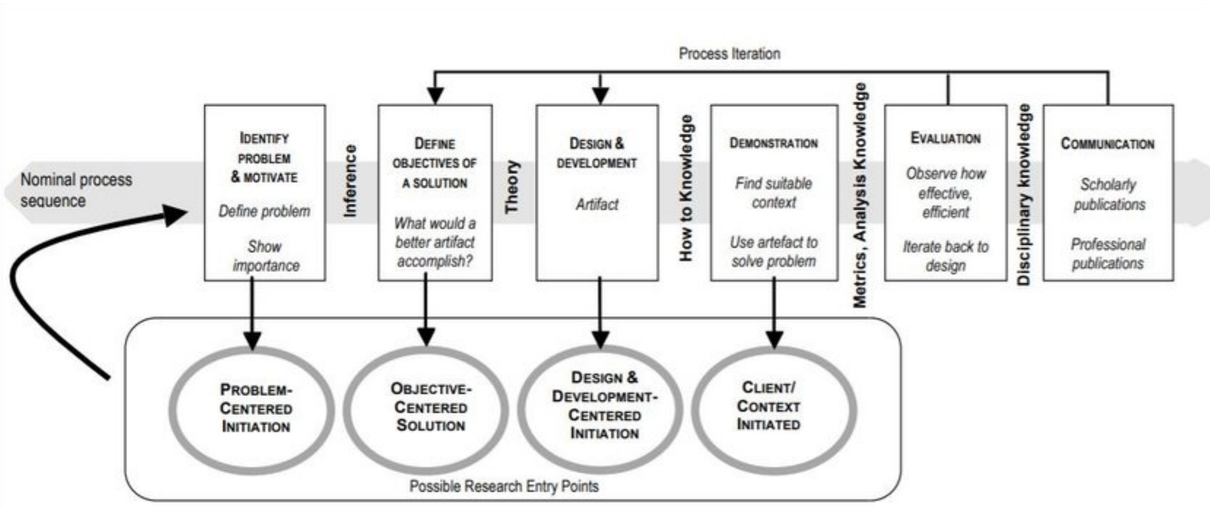


Figure 22 - Design Science Research Methodology (DSRM) process model (Lawrence et al. 2010)

In this research, a two-threaded and cross-validated research framework is proposed to address the deficiencies of the existing BIM compliance checking solutions in terms of comprehensiveness and automation. The overarching framework of this study is shown in Figure 23.

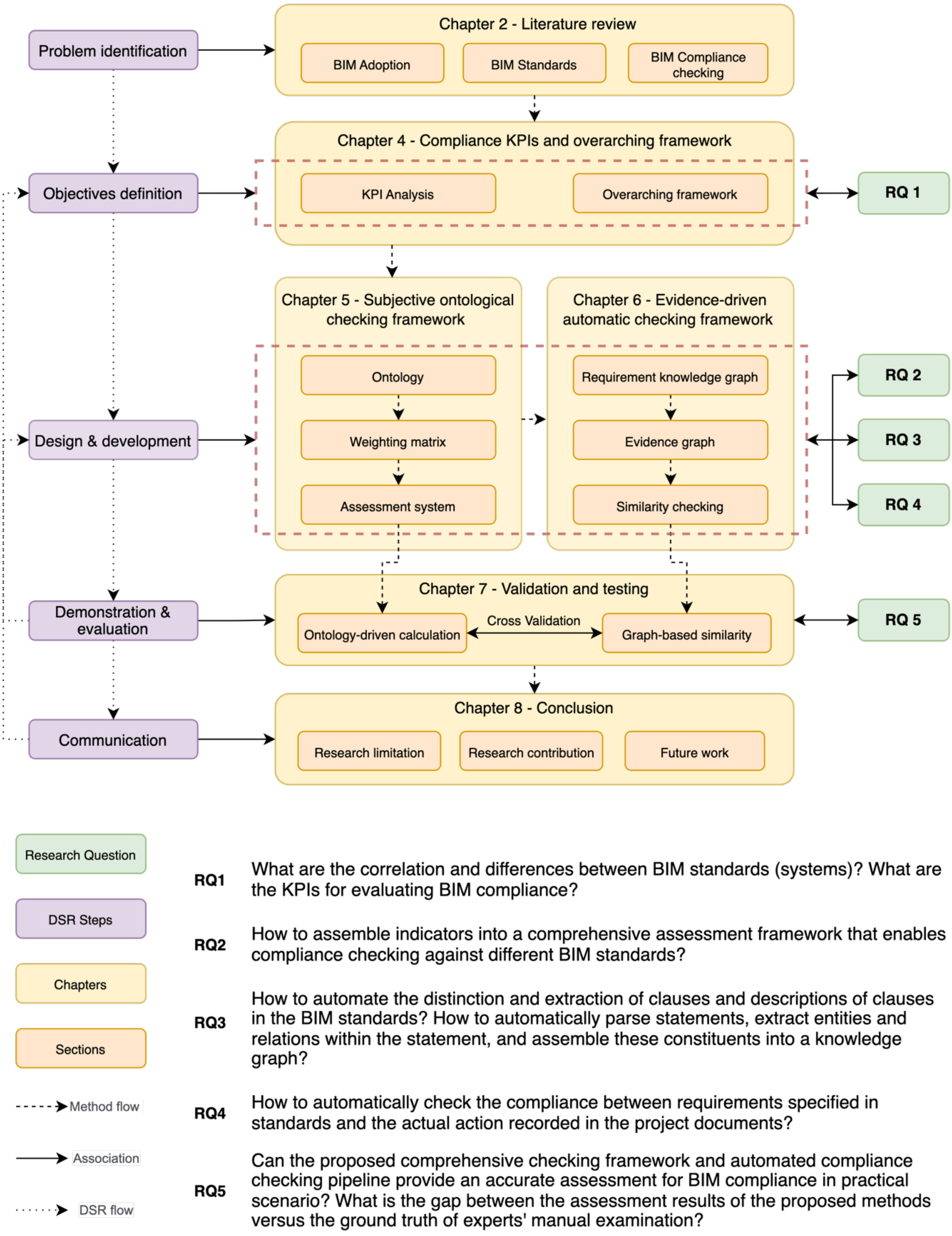


Figure 23 - The overarching framework for comprehensive and smart BIM compliance checking

In this study, the research methodology, the research questions, and the components in the proposed framework correspond to each other. Their associations are explained as follows:

1) **problem identification:** The related problem and its motivations have been investigated in the section of the introduction, indicating the significance of BIM compliance checking for the development of both enterprises and the AEC industry. Through the investigation of the international BIM market in Chapter 2, it was found that the BIM technology has been widely used in various AEC projects. Furthermore, great progress has been made in the development of BIM standards. BIM standards are not limited to a single standard, but a series of related standards referencing each other to form a BIM standard system. After reviewing the state-of-the-art compliance checking methods for BIM, existing approaches were found to suffer from various defects. Most industrial methods fall short in terms of granularity and automation. While the methods proposed by academics achieve a certain degree of automation, their performance in terms of comprehensiveness is still inadequate. In general, these methods can only be applied to check compliance against one specific standard. Considerable labour is still required to build the knowledge models if a comprehensive check for multiple standards is to be implemented.

2) **objectives definition:** This research aims to propose a comprehensive and smart BIM compliance checking solution, which can not only improve the existing BIM compliance checking method but also automate the checking process. Due to the limitations of existing studies on the scope of BIM standards, existing criteria flaws in terms of completeness and granularity. Therefore, a more comprehensive BIM compliance assessment indicator system needs to be established through comparison analysis of multiple BIM standard systems. In this part, the indicators are extracted from regulations by a text-mining algorithm and then assembled as a hierarchical indicator system under manual supervision. To ensure the reliability and effectiveness of the system, the Delphi method is adopted, where the indicators and hierarchical system are validated by a domain expert panel. This step is designed to build the foundation for comprehensive checking and provide validation for automated checking. Regarding to research questions, this step is presented to answer Q1.

   Q1: What are the correlations and differences between different BIM standards (systems)? What are the KPIs for evaluating BIM compliance?

3) **design and development:** This stage aims to develop a fully automated compliance checking solution for BIM projects. To eliminate the manual work involved in consistency checking, the requirements specified in regulations and the evidence recorded in the project documents need to be extracted and matched automatically. Through the investigation of technical difficulties and potential solutions in Chapter 2, an NLP and deep learning-based framework is proposed to automate BIM compliance checking. In this framework, deep learning and NLP techniques are introduced to address the challenge of requirement extraction and relation parsing. The Seven Clause theory in linguistics is innovatively integrated into existing syntactic parsing methods to enhance its parsing performance on complex logic and multiple relations. In addition, graph learning is adopted to capture the structural information of entities to achieve the mapping between dynamic entities. The proposed framework is divided into three parts, namely requirement (evidence) extraction, knowledge graph generation and compliance checking, aiming to answer the following three research questions respectively.

   Q2: How to assemble indicators into a comprehensive assessment framework that enables compliance checking against different BIM standards?
   Q3: How to automate the distinction and extraction of clauses and descriptions of clauses in the BIM standards? How to automatically parse statements, extract entities and relations within the statement, and assemble these constituents into a knowledge graph?
   Q4: How to automatically check the compliance between requirements specified in standards and the actual action recorded in the project documents?

4) **demonstration and evaluation:** After conducting the technical development in Chapter 5, a use case of an existing BIM project is set to demonstrate the workflow of the proposed framework. in this part, a cross-validation method is adopted to verify the comprehensive and automatic compliance checking system, where the feedback from domain experts, the indicators of the ontological model, and the result of the automated checking pipeline validate

each other to ensure the reliability of the validation results. This step aligns with Q5.

Q5: Can the proposed comprehensive checking framework and automated compliance checking pipeline provide an accurate assessment of BIM compliance in the practical scenario? What is the gap between the assessment results of proposed methods versus the ground truth of experts' manual examination?

Chapter 4. BIM compliance KPIs and overarching checking framework

Through the comprehensive review of the standard from four primary BIM standard systems, it can be found that the content of the BIM standard system is highly harmonised, regardless of differences in their emphasis. In the forthcoming fifth edition of the NBIMS standard, concepts and requirements from different BIM standard systems are referenced and integrated, which shows the trend towards standards fusion and proves the viability of the unified BIM standards system. In this chapter, a series of key performance indicators (KPIs) for BIM compliance are defined through a hybrid approach involving both automated extraction and manual calibration. Concepts and requirements from various standards are correlated through these KPIs. On the basis of these KPIs, a comprehensive BIM compliance checking framework is proposed to address the deficiencies of existing methods in terms of completeness, granularity, and automation.

4.1 BIM compliance KPI analysis

The requirements in BIM standards are basically management and operations-orientated, which are qualitative and challenging to quantify. In this research, a series of key performance indicators (KPI) are defined to measure conformity against specific requirements. To ensure comprehensiveness, a number of representative standards are selected from the four primary BIM standard systems (NBIMS, UK BIM, CBIMS and Open BIM) to conduct comprehensive analyses and extraction of key performance indicators in the knowledge model. Table 4 lists all the BIM standards involved in the development of an ontological BIM compliance checking framework.

Table 4 - BIM standards involved in generating indicators

Number	Name	Standard system	Issuer
GB/T 51212	Uniform Standard for Building Information Model Application	China	MOHURD
GB/T 51269	Standard for classification and coding of building information model	China	MOHURD
GB/T 51301	Standard for design delivery of building information modeling	China	MOHURD
GB/T 51235	Standard for building information modelling in construction	China	MOHURD
JTS/T 198-1	Unified Standard for Application of Building Information Modeling in Port and Waterway Engineering	China	MOT
JTS/T 198-2	Standard for Application of Building Information Modeling in Port and Waterway Engineering Design	China	MOT
Q/CCCC GL501	Unified Standard of Building Information Modeling in Highway Engineering	China	CCCC
Q/CCCC GL502	Standard for Application of Building Information Modeling in Highway Engineering Design	China	CCCC
ISO 19650-1	Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) - Information management using building information modelling - Part 1: Concepts and principles	UK	ISO
ISO 19650-2	Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) - Information management using building information modelling - Part 2: Delivery phase of the assets	UK	ISO
PAS 1192-3	Specification for information management for the operational phase of assets using building information modelling	UK	BSI
PAS 1192-4	Collaborative production of information. Fulfilling employer's information exchange requirements using COBie - Code of practice	UK	BSI
PAS 1192-6	Specification for collaborative sharing and use of structured Health and Safety information using BIM	UK	BSI
NBIMS-v1	United States National Building Information Modeling Standard - Version 1	USA	buildingSMART
NBIMS-v2	United States National Building Information Modeling Standard - Version 2	USA	buildingSMART
NBIMS-v3	United States National Building Information Modeling Standard - Version 3	USA	buildingSMART
ISO 16739	Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries	Open BIM	ISO
ISO 29481	Building information models - Information delivery manual	Open BIM	ISO
ISO 12006	Building construction - Organization of information about construction works	Open BIM	ISO

The scope of the comprehensive assessment framework includes 19 standards from four different BIM standard systems. To conserve the intensive labour in extracting KPIs from a large number of standard documents, a semi-automated method that combines manual modification with a text mining algorithm is proposed in this part to obtain the whole set of indicators. The specific process of indicator extraction is divided into two steps, namely automatic mining, and manual calibration.

4.1.1 Automatic KPI mining

The process of automatic KPI mining is comprised of the following three substeps:

1. Data conversion

All published BIM standard documents listed in Table 4 are formatted in PDF, where the content cannot be directly recognised and processed by the computer. Therefore, a data conversion algorithm is developed based on a Python library called PyPDF2 to transfer the format of standard documents from PDF files into computer-readable format (.txt). During the transformation, the figures and tables in the documents are filtered out as the indicator mining analyses are based on the textual content of the standard.

2. Content pre-processing

Although the expressions in the standards are normative, pre-processing is still required. In this stage, some redundant content is removed from the original content, including the header, footer, page numbers, and metadata of the documents. Moreover, the stop words (e.g., is, are, be, shall, should, etc.) that are not relevant to content are also dropped.

3. KPI mining

The statistics-based KPI extraction is the core of this PKI mining part. As stated by Xu and Zhang (Xu and Zhang 2021), the more times a concept appears in the standard, the more important that concept is. Therefore, the TF-IDF algorithm is adopted in this part to extract terms with higher frequency as KPIs.

The concept of TF-IDF (short for term frequency-inverse document frequency) was originally introduced by Karen Spärck Jones and Stephen Robertson (Jones 1972), which has been widely used in natural language processing for information retrieval. The TF-IDF method utilises the term frequency (TF) and inverse document frequency (IDF) to measure the significance of each word or term in the document, where the term frequency stands for how frequently the term/word appears in the document and the inverse document frequency represents how rare this term/word is across all documents in the corpus. The term frequency and inverse document frequency of a specific word/term can be calculated through the following formula respectively:

- Term frequency (TF) is calculated by dividing the number of times the term occurs in the document by the total number of terms in the document.

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \ . \tag{1}$$

  where $n_{i,j}$ represents the occurrence of term t_i in document d_j, $n_{k,j}$ represents the total number of occurrences of all terms in document d_j.

- Inverse document frequency (IDF) measures the amount of information a term provides. It is determined by dividing the total number of documents by the number of documents that contain the term, followed by computing the logarithm of this division result.

$$IDF_i = \log \frac{|D|}{|\{d: t_i \in d\}|} \ . \tag{2}$$

  where t_i represents the term in the document d and $|D|$ stands for the total number of documents.

- Finally, the value of TF-IDF can be calculated by multiplying the values of TF and IDF.

$$TF_{i,j} - IDF_i = TF_{i,j} \times IDF_i \ . \tag{3}$$

The final score indicates the significance of the term, where a higher score signifies greater importance and a lower score reflects lesser importance.

During the mining process, the pre-processed textual contents of the abovementioned 19 standards constitute the corpus. The content from each document is scanned individually by the TF-IDF algorithm to calculate the frequency of each term within it. Considering some terms are composed of multiple words, a dynamic window with sizes from 1 to 4 is used during the indicator extraction process. Table 5 and Table 6 list the 15 most frequent terms in ISO 19650-2 with a window size of 2 and 3 respectively.

Table 5 - Term frequency of ISO 19650-2 with a window size of 2

Rank	Terms	Frequency
1	appointing party	0.0416
2	appointed party	0.0390
3	delivery team	0.0341
4	information model	0.0226
5	information management	0.0208
6	task team	0.0186
7	information delivery	0.0155
8	information requirements	0.0142
9	information production	0.0124
10	shared resources	0.0097
11	exchange information	0.0093
12	information container	0.0089
13	reference information	0.0066
14	information standard	0.0062
15	delivery milestones	0.0053

Table 6 - Term frequency of ISO 19650-2 with a window size of 3

Rank	Terms	Frequency
1	lead appointed party	0.0239
2	information production methods	0.0084
3	information management process	0.0068
4	exchange information requirements	0.0058
5	common data environment	0.0055
6	project information standard	0.0045
7	bim execution plan	0.0045
8	information management function	0.0043
9	appointing party consider	0.0042
10	establish project information	0.0040
11	information delivery plan	0.0039
12	information delivery milestones	0.0032
13	project information delivery	0.0029
14	appointed party establish	0.0026
15	level information need	0.0024

Through a preliminary analysis of the statistical results, the 100 most frequent terms in the standard can roughly cover 80-90 per cent of the target indicators. Under the negotiation of accuracy and workload, the 150 highest frequency terms in each standard were collated as candidates for the indicators.

4.1.2 Manual KPI calibration

Although it is assumed that word frequency is positively correlated with importance (Xu and Zhang 2021) during the automatic extraction of KPIs, it is not strictly positive in the actual documents. Therefore, the frequency of several important indicators is not relatively high in the documents (e.g., the indicator of level information need in Table 6). In addition, not all terms with high frequency can be directly recognised as indicators, as some of them do not have practical meaning (e.g., exchange information in Table 5 and appointing party consider, establish project information, appointed party establish in Table 6). Based on the above considerations, manual calibration is necessary to supplement missing indicators and filter out meaningless indicators.

Before the manual calibration, the authors carefully reviewed all the BIM standards listed in Table 4 together with some related and referenced regulatory documents. After the review work, the author went through the list of indicator candidates generated by the algorithm and screened out the selected indicators for each standard based on the author's understanding of BIM standards. During the calibration process, three domain experts are invited to work as a calibration expert panel that provides feedback on calibration results to guarantee the reasonableness of indicators. All three experts have been engaged in BIM standards-related research for more than ten years. One of them was involved in the development of international standards, and the other two have extensive practical engineering experience and have published several papers related to BIM standards.

Before assembling these indicators into the knowledge model, there is one more step that needs to be done, which is the coreference resolution. As described at the beginning of this chapter, there are commonalities between different BIM standard systems. Different standards propose the same or similar requirements, for example, all the standards from NBIMS-v2, GB/T 51301, to ISO 19650-1 specify that the level of information need should be defined to clarify the information granularity during the information exchange process. Hence, different standards share some common indicators or some similar indicators that differ in naming but have the same function. To avoid ambiguity and unify the concepts, the coreference resolution of indicators should be conducted before the development of the comprehensive indicator system.

To solve this problem, a separate series of research meetings was organised with the expert panel. The work of coreference resolution was implemented based on the experts' domain knowledge and their consensus. Figure 24 and Figure 25 illustrate some of the mapping results of the same/similar indicators from different standards.

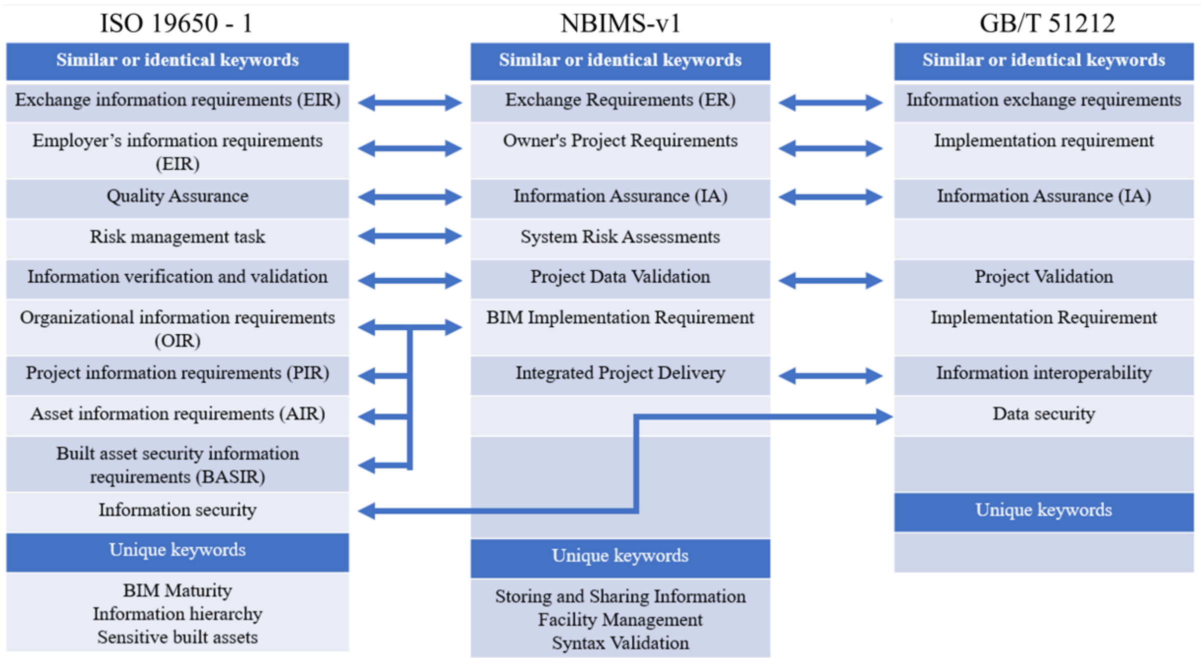


Figure 24 - Indicator mapping between ISO 19650-1, NBIMS-v1 and GB/T 51212

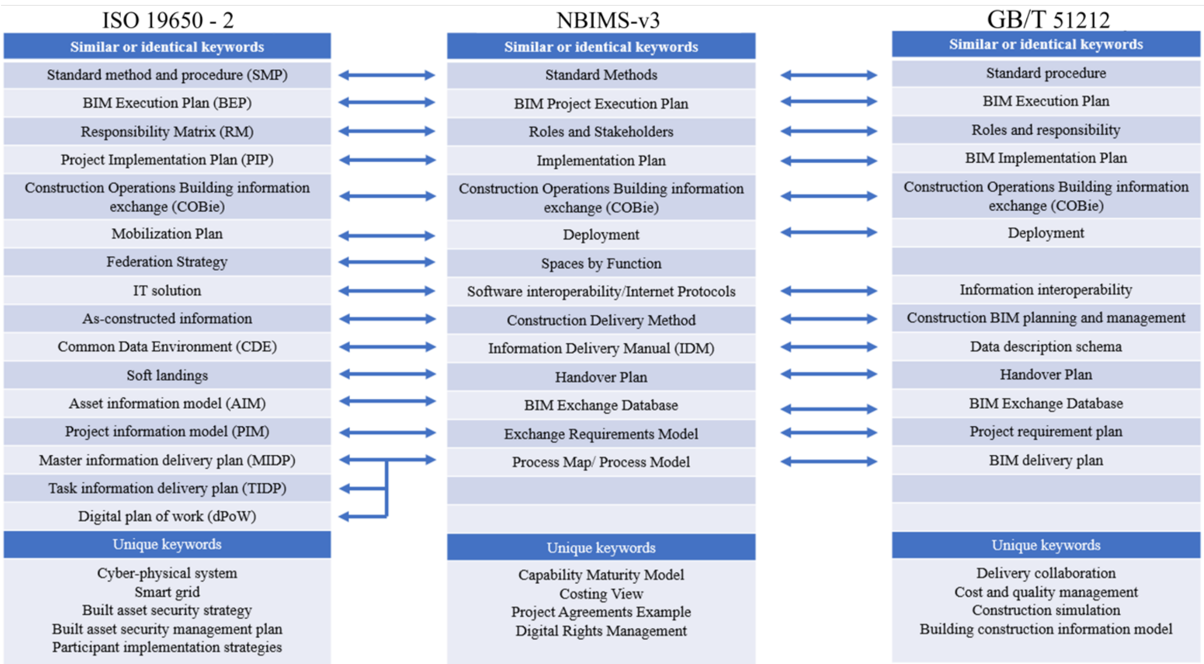


Figure 25 - Indicator mapping between ISO 19650-2, NBIMS-v3 and GB/T 51212

Through manual calibration, the 2,850 candidate indicators automatically extracted by the algorithm were refined into 510 final indicators, which are utilised to constitute the comprehensive BIM compliance checking framework.

4.2 The overarching framework for comprehensive BIM compliance checking

To bridge the gaps of existing methods in completeness, granularity, and automation, this research introduces a comprehensive BIM compliance checking framework based on some advanced techniques including semantic knowledge representation, deep learning, and natural language processing. The structure of the proposed comprehensive framework is illustrated in Figure 26.

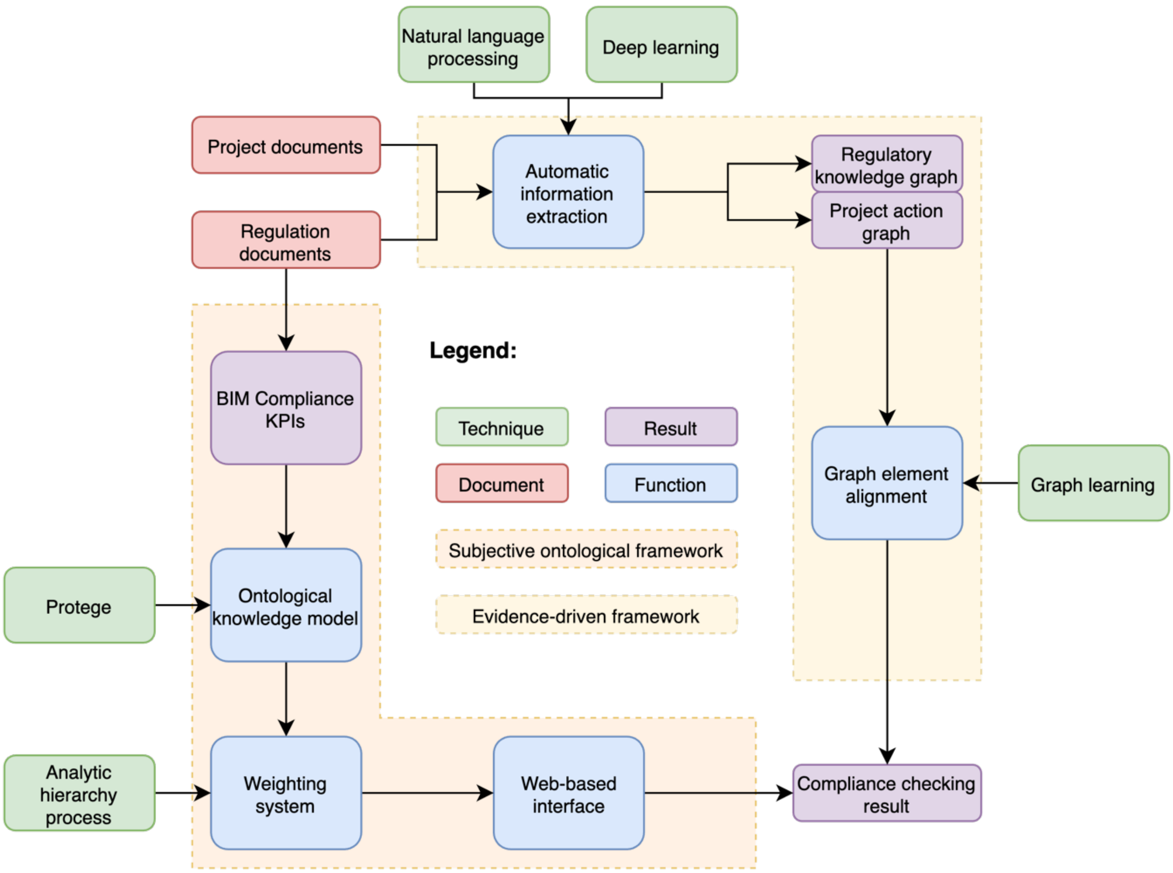


Figure 26 - Overarching framework for comprehensive BIM compliance checking

The framework can be divided into two main components: subject-oriented ontology-driven compliance checking framework and evidence-driven automatic compliance checking framework. The ontology-driven compliance checking framework consists of a knowledge model and an interactive web-based platform for user engagement. The knowledge model adopts an ontological representation, which is constructed by the abovementioned calibration expert panel based on extracted BIM indicators. In contrast to traditional spreadsheet-based knowledge representation, the ontology-based approach allows for a more flexible inspection mode through SPARQL queries by specifying and filtering the indicators to be examined. The web-based interactive platform is designed to assist users in customising their checking objectives for BIM compliance and provide a real-time representation of checking results, and recommendations for improvement. This framework has established a novel paradigm for BIM compliance checking, thereby addressing the deficiencies in existing approaches concerning completeness, flexibility, and granularity. However, akin to existing methods, the ontology-driven checking framework still requires users to respond to a series of pre-defined questions for compliance evaluation. This process is often intricate, and time-consuming, and user responses are often subjective and error-prone. Therefore, an evidence-driven automated inspection framework is proposed based on previous work to enhance the objectivity and reliability of checking results. The framework initially employs natural language processing and deep learning techniques to extract information from standard documents and project-related files and represents them as two separate knowledge graphs. A pre-trained language model is fine-tuned with a self-developed domain dataset, which can extract clauses from regulation documents. Subsequently, the information in the clauses and project documents are parsed into tuples by adopting syntactic parsing and further assembled as two knowledge graphs. Eventually, the concepts and relationships in these two graphs are aligned through graph learning techniques, thereby facilitating compliance checking. This framework is developed in the Python environment through some external libraries, such as TensorFlow, PyTorch Geometric, Pandas, etc., and composed of several deep learning models and specially designed algorithms, which renders the entire checking process fully automated. Thus, this evidence-driven framework not only enhances the reliability of BIM consistency checks but also addresses the research gap in automation.

The design and development details of the two aforementioned assessment frameworks will be comprehensively elucidated in Chapters 5 and 6.

Chapter 5. Subject-oriented ontology-driven compliance checking

This chapter elaborates on the process of developing a comprehensive and detailed framework for BIM compliance checking that covers the criteria extracted from multiple BIM standard systems, which enables compliance checking against one specific standard, multiple standard compliance checking, and overall compliance checking. This framework aims to fill the gap in comprehensive BIM compliance checking in the present AEC industry and also provide validation for automated compliance checking. This ontological compliance checking framework is divided into four parts, which are KPI analysis, ontology development, weighting matrix determination, and assessment system development. The remainder of this chapter will go through the above parts to illustrate the development process of this ontology-driven checking framework.

5.1 Knowledge model development

As introduced in Section 2.3, ontology has been widely used in the AEC domain as a carrier of knowledge to realize smart process management, and cost estimation (Pauwels et al. 2017). Except for ontology, there are also various other types of knowledge models, such as logical representation, semantic network representation, frames representation (Minsky 2019), and production rules. Compared with other forms of knowledge models, ontology provides a more powerful and flexible semantic representation, which makes it more desirable to represent knowledge with complex logic and relations. Semantic inference and retrieval can be easily implemented to facilitate querying the relevant regulation knowledge. Considering the above benefits, a domain ontology is developed in this study to build the correlations among indicators thereby forming a knowledge model for BIM compliance checking.

After sourcing all available BIM standard-related ontologies from the web, there are no reusable ontologies that can be applied to this part. Therefore, the whole ontology was manually constructed by the author under the supervision of the calibration expert panel. The details of the ontological knowledge can be found in Table 7 and Figure 27 presents a high-level overview of the BIM compliance checking ontology, which indicates the number of classes, properties and RDF triplets defined in the ontology.

Table 7 - Metrics of the comprehensive ontological knowledge model

Item	Number of instances
Axiom	3702
Logical axiom	1552
Declaration axiom	934
Class	510
Object property	15
Individual	409
Annotation assertion	1216

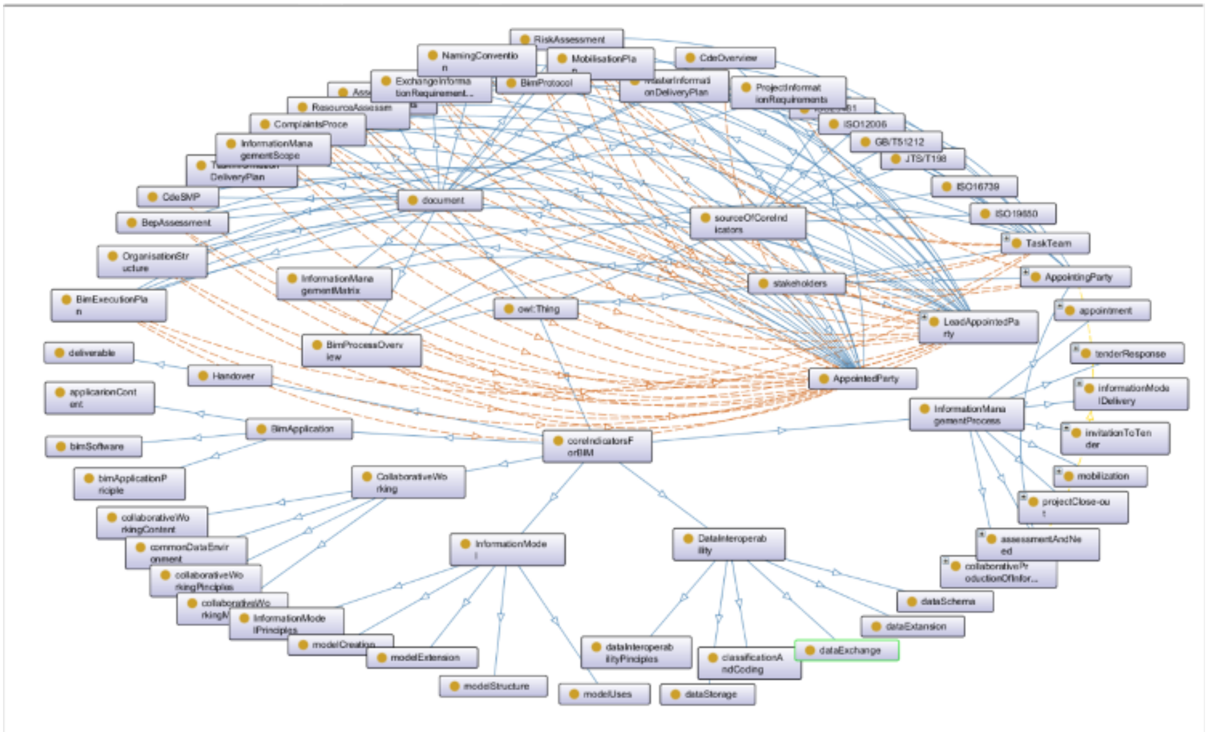


Figure 27 - Visualisation of the comprehensive ontology for BIM compliance

5.1.1 Ontology editor

In this research, an open-source ontology tool, Protégé (version 5.6.3), is used to build the knowledge model for BIM compliance checking. This software is developed by the Stanford Centre for Biomedical Informatics Research at the Stanford University School of Medicine and it enables both developers and users to develop and maintain an ontological knowledge model. Compared to other ontology editors, protege features greater flexibility and compatibility. It supports various reasoners and third-party plugins with diverse functions. Some important functions, such as SPARQL queries and visualisations, can be implemented directly in the software. Furthermore, Protégé is also compatible with various OWL-related syntaxes, such as OWL/DL, OWL/XML OWL functions and Turtle, and some other fundamental languages like LaTex and JSON-LD. During the development process, the built-in reasoner (Pellet) is used to check the logical relationships between elements in the ontology and a visualisation tool and a SPARQL query tool are used to check for missing concepts and relationships. The whole development process of the ontology follows the instructions listed in “Ontology Development 101” (Noy 2001) and the specific guidance on the software environment provided by Nagypál (Gábor 2007).

5.1.2 Classes

For the definition of classes, a top-down approach is adopted, where general domain concepts are first defined and then classified into specific hierarchies. In this research, the 681 previously summarised indicators are grouped into four main categories:

- **Documents**
  Document class lists all the documents that need to be established during the delivery of projects required by different BIM standards. These documents play significant roles in defining information exchange requirements and planning project information delivery. Moreover, some fundamental principles of the project are also clarified in these documents, such as a naming convention, assessment process, etc. Figure 28 shows the classes of required documents defined in the ontology.

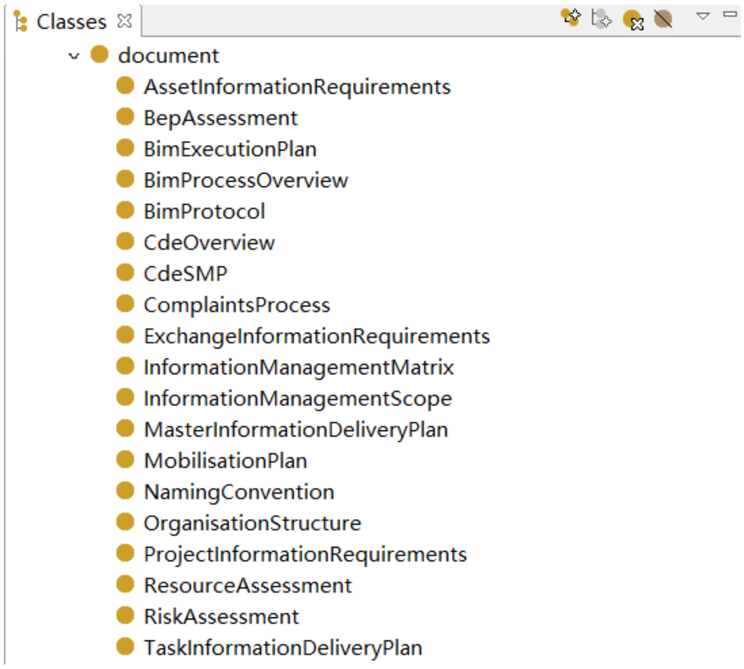


Figure 28 - Classes of documents defined in the knowledge model

- **Standards**

  The classes of standards include all the BIM standards involved in the knowledge model. These classes serve as markers for sources of indicators to enable compliance checking against a specific standard or a set of multiple standards. Figure 29 presents the classes of standards involved in the knowledge model.

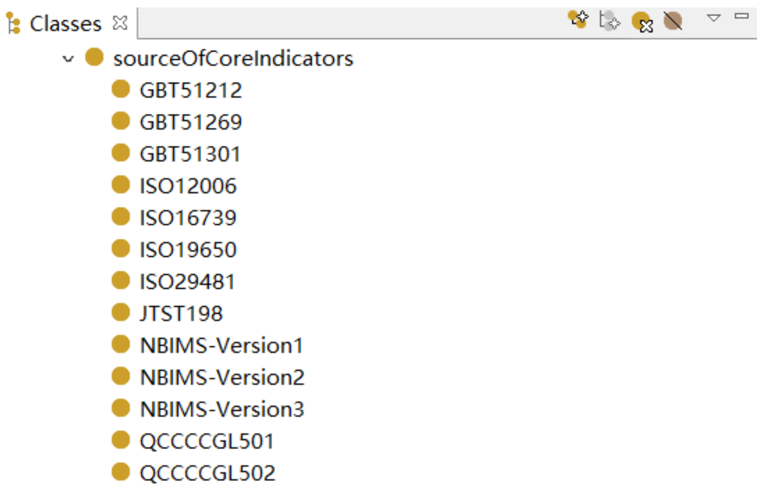


Figure 29 - Classes of standards involved in the knowledge model

- **Stakeholders**

The purpose of the stakeholders class is to outline all the responsible parties that are involved in the process of project delivery. Figure 30 lists all the subclasses of stakeholders. These four subclasses are standard terms extracted from the ISO 19650 series standard. Although similar concepts are found in other standards (e.g., first party and second party in GB/T 51301), these concepts can be unified into these four roles.

Figure 30 - Classes of stakeholders defined in the knowledge model

- **BIM compliance indicators**

  The classes of BIM compliance indicators are the core of the ontological knowledge model, which covers over 90% of the classes defined in this ontology. In order to explicitly represent the internal relationships among indicators, a hierarchical structure is adopted to organise the indicators. Figure 31 illustrates the classes of high-level indicators defined in the ontology. The whole set of indicators can be found in Appendix C.

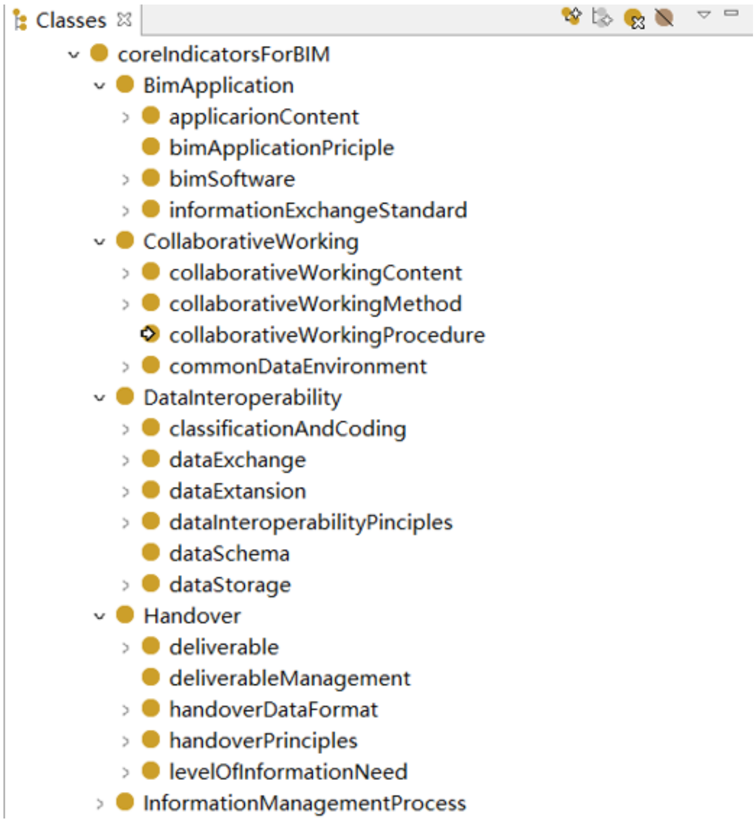


Figure 31 – Classes of BIM compliance indicators defined in the knowledge model

As shown in Figure 31, the indicators are classified into six high-level domains, which are information model, collaborative working, data interoperability, information management process, BIM application, and handover. In these domain indicators, the information model focuses on the requirements and principles of the modelling of information models, such as model structure, model creation method, model extensions, model usage, etc. For collaborative working, the content, method, and procedure are the main concentrations. Moreover, the requirements and structure of the common data environment (CDE) where the collaborative work is implemented are also included in this class. The domain of data interoperability concerns the classification and coding system, data exchange format, data extensions and data storage. In terms of the information management process, it specifies various required actions to be performed by the different stakeholders at different stages of the project. The last domain indicator is handover, which concentrates on the type and format of deliverables, delivery methods, and level of information (LOI) of the deliverables.

5.1.3 Property

To enable the function of reasoning, the properties are indispensable to the ontology. In general, the properties defined in OWL can be divided into three types: object property, datatype property and annotation property. Object properties are designed to describe the relationship or restriction between objects and instances in a class. The datatype properties aim to establish connections between objects and their quantitative or qualitative data values. The annotation properties are intended to provide additional information for the classes' properties and instances. In the process of constructing an ontology, the properties are generally built based on collective knowledge to connect different classes or instances.

In this research, the ontology is designed to be a comprehensive knowledge base that supports BIM standard compliance checking under four different scenarios. The first scenario is checking the BIM compliance against one or multiple target standards. The second scenario is checking the BIM compliance as a specific role (e.g., appointing party) of the project. The third scenario is checking the BIM compliance of one or several project documents. The last scenario is checking the BIM compliance against a specific indicator, several indicators or even the whole indicator system. To achieve the above-designed scenarios, 10 high-level object properties are defined in the knowledge model in Figure 32.

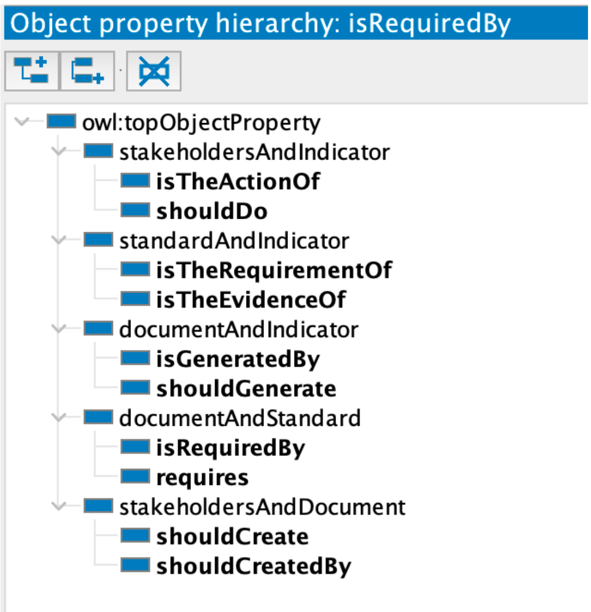


Figure 32 - Object properties defined in the knowledge model

As shown in Figure 33, these 10 high-level object properties can be categorised into five groups based on the classes connected to them. The two object properties in each group are inverse to each other. The arrows in the diagram represent the direction of properties. Group 1 is designed for the first scenario, which focuses on the relationships between BIM standards and compliance indicators. This group of object properties illustrates the source of each compliance indicator. For the second scenario, group 2 can address it through identifying the required actions for each stakeholder. The object property of group 3 concentrates on the relationships between project documents and compliance indicators. It indicates what indicator should be described in what document to enable the target of the third scenario. The fourth scenario can be directly achieved through the hierarchical properties defined among classes (e.g., subclass of). The rest of the object properties may not make a huge difference in the designed application scenarios but improve the completeness of the knowledge model. In addition to object properties, two annotation properties are applied in the knowledge model. They both use rdfs:comment as the property but are restricted in data type to strings and floats respectively.

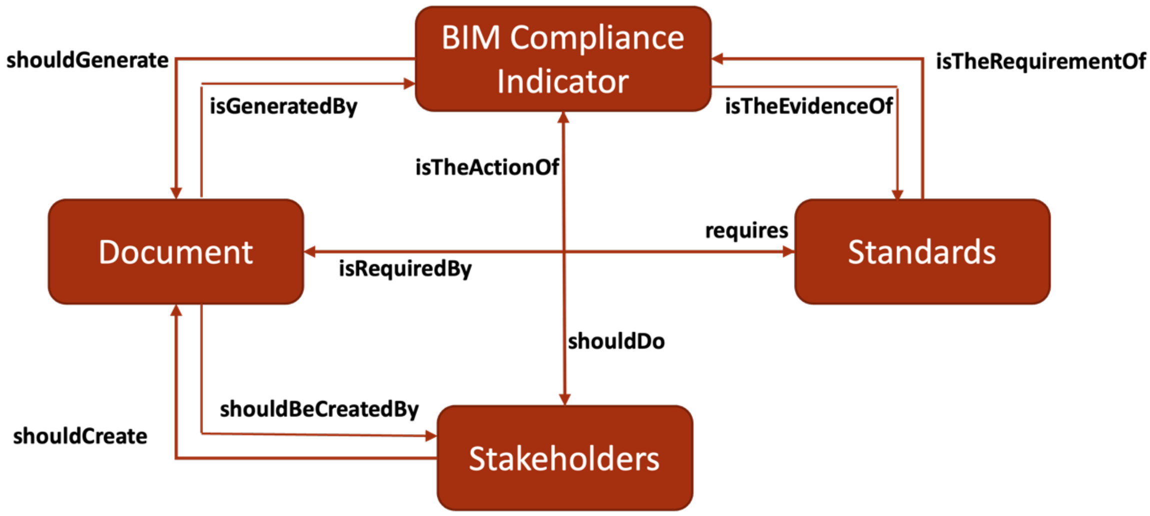


Figure 33 - Relationship between object properties and high-level classes defined in the proposed ontology

5.2 Weighting matrix

So far, the indicators for BIM compliance checking have been defined and organised in a sensible way. Considering the variations in the importance of different indicators for compliance checking, a weighting matrix is developed based on the analytic hierarchy process (AHP) method to aggregate the evaluation for each indicator and obtain the overall assessment results.

The analytic hierarchy process (AHP) is a structured technique introduced by Thomas L. Saaty in the 1970s (Figure 34). It was developed based on mathematics and psychology and was originally designed for complex decision-making (Saaty 1990). In the AHP method, different alternatives are rated against specific criteria. The ratio scales are derived from the principal eigenvectors and the consistency index is derived from the principal eigenvalue. Then these ratings are aggregated and sorted to help users find out the best alternative. In this research, the AHP method is not applied to make decisions but to calculate the weights for each indicator. This method provides an accurate way to quantify the weights of decision criteria. The relative magnitude of the indicators is estimated via pairwise comparisons conducted by experienced experts. A specially designed questionnaire is normally used to get feedback from each respondent on a comparison of the relative importance of each pair of items.

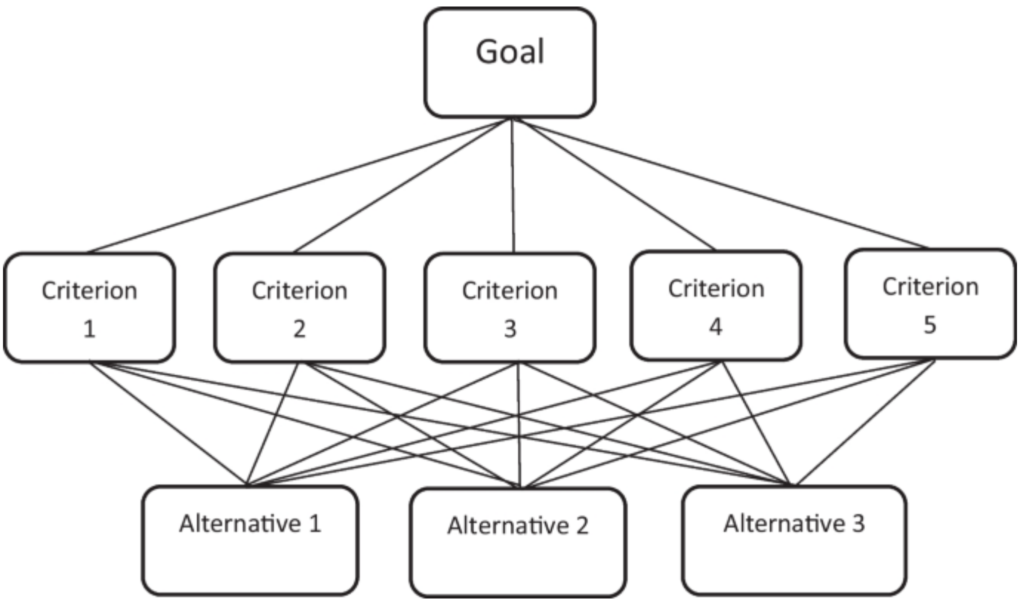


Figure 34 - The analytic hierarchy process (AHP) model (Vahidnia et al. 2022)

Based on the principles of the AHP method and the hierarchical structure pre-defined in the knowledge model, the indicators of BIM compliance checking are clustered into five levels. Figure 35 presents the hierarchical structure of the whole set of indicators. The sum of the weights of all subclasses that are within the sub-level of the same indicator equals 1, which can be expressed as formula 4.

$$\sum_{j=1}^{n} C_{ij} = 1 \tag{4}$$

where C_{ij} represents the j th subclass of the n subclasses of indicator i.

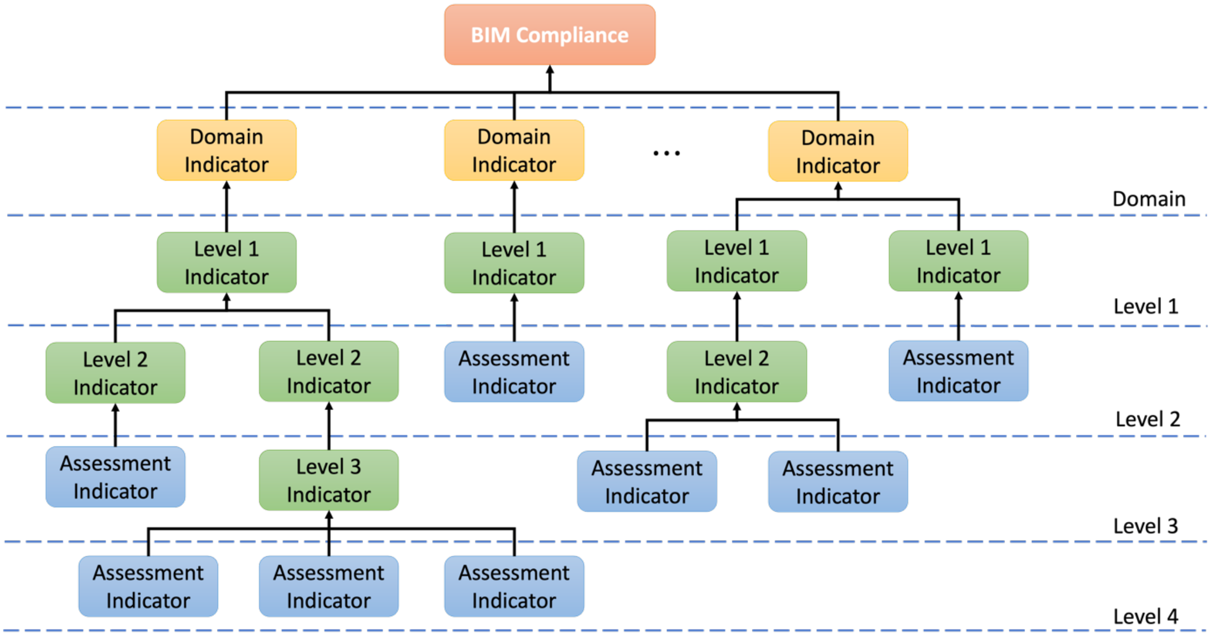


Figure 35 - Hierarchical structure of BIM compliance checking indicators

The specific value for each indicator is calculated through the AHP method, which can be broadly divided into three steps, namely pairwise comparison, weighting calculation, and consistency checking. In the following part, the six domain indicators will be used as an example to illustrate the process of calculating the weighting of each indicator through the AHP method.

- Pairwise comparison

  According to the research of Greenbaum (Greenbaum 1991), 5-7 experts are ideal for the AHP method. Hence, 5 domain experts from the validation panel were invited to participate in the survey of pairwise comparison. A specific questionnaire was developed and distributed to these experts to obtain the perceptions of experts regarding the significance of indicators. Table 8 presents the fundamental scale of the AHP method adopted in this research.

Table 8 - Scale of the importance adopted in this research

Intensity of importance	Definition	Explanation
1	Equal importance	Two factors contribute equally to the objective
3	Somewhat more important	Experience and judgement slightly favour one over the other.
5	Much more important	Experience and judgement strongly favour one over the other.
7	Very much more important	Experience and judgement very strongly favour one over the other. Its importance is demonstrated in practice.
9	Absolutely more important	The evidence favouring one over the other is of the highest possible validity.
2,4,6,8	Intermediate values	When compromise is needed.

- Weighting calculation

After acquiring feedback from the experts, a comparison matrix was generated through formula (5).

$$a_{ji} = \frac{1}{a_{ij}} \tag{5}$$

where a_{ij} represents the importance scale of indicator i versus indicator j.

Table 9 - Comparison matrix of the six domain indicators

Indicator	Information model	Collaborative working	BIM application	Data interoperability	Information management process	Handover
Information model	1.00	1.33	1.60	1.00	0.67	2.67
Collaborative working	0.75	1.00	1.20	1.00	0.50	2.00
BIM application	0.63	0.83	1.00	0.83	0.42	1.67
Data interoperability	1.00	1.00	1.20	1.00	0.50	2.00
Information management process	1.50	2.00	2.40	2.00	1.00	4.00
Handover	2.00	0.50	0.60	0.50	0.25	1.00

Based on the above comparison matrix (Table 9), the principal eigenvalue and eigenvector can be calculated, which are 6.639 and $[1.575\ \ 1.229\ \ 1.026\ \ 1.288\ \ 2.458\ \ 1]^T$ respectively. After the eigenvector was normalized, the weight of each indicator can be obtained (Table 10).

Table 10 - Weights for the domain indicators

Indicator	Information model	Collaborative working	BIM application	Data interoperability	Information management process	Handover
Weight	0.184	0.143	0.120	0.150	0.287	0.116

- Consistency checking

  Due to the transitive property, the consistency of the judgement needs to be further checked. The consistency ratio proposed by Saaty was adopted in this research.

  The consistency ratio (CR) can be calculated by dividing the consistency index (CI) by the random consistency index (RI), where the consistency index can be obtained using the formula (6) and the random consistency index can be directly picked from Table 11.

$$CI = \frac{\lambda_{MAX} - n}{n - 1} \tag{6}$$

$$CR = \frac{CI}{RI} \tag{7}$$

Table 11 - Random consistency index

Size of matrix	Random consistency index (RI)
1	0
2	0
3	0.58
4	0.90
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49

Through the above formulas, the consistency ratio is calculated as 0.055. It is smaller than the threshold of 10%, indicating the subjective judgement of experts is consistent and the calculated weights are reliable.

Following similar processes, the weight of each indicator was calculated and formed a weighting system which has identical hierarchical structures as the indicator classes within the knowledge model.

5.3 Assessment system

The existing assessment model runs basically on the format of a spreadsheet, where the knowledge is inflexible and difficult to manage and maintain. To address this problem, this part takes the first scenario as an example and a web-based platform is developed to achieve ontology-driven flexible compliance checking.

As illustrated in Sections 4.2 and 4.3, the indicators for BIM compliance checking are organised in a hierarchical format (Figure 35). Within it, the top of the pyramid is the final checking result obtained through integrating the evaluation result of indicators in the lower levels. Under the top level is the domain layer, which includes the 6 high-level domain indicators defined in the knowledge model. The subclasses of these domain indicators are lowered to the next level (Level 1). Following the same principle, subclasses of the indicators in Level 1 are represented in Level 2 and so on. As shown in Figure 35, there are three types of indicators in the framework. Apart from the domain indicators and normal indicators, there is another type of indicator called evaluation indicators. This type of indicator is the lowest level indicator in each evaluation branch, which is similar to the end nodes in a decision tree.

The assessment is a bottom-up process driven by evaluation indicators. It follows the mechanism of weighted average to calculate compliance scores from lower level to higher level. The initial value of the evaluation indicators is acquired according to manual quantification and the user's responses. Then the score of the normal indicators and domain indicators are calculated level by level from the evaluation indicators. Figure 36 provides a schematic diagram of calculating the compliance of project delivery with some dummy data. The red numbers are the coefficients in the calculation framework that are determined by the results calculated by the AHP method. This coefficient represents the weighting of each indicator and the sum of all the coefficients under the same higher-level indicator equals 1. For example, *Task Information Delivery Plan* and *Master Information Delivery Plan* are two sub-indicators of *Planning and Documenting* and the weightings of these two sub-indicators are 0.6 and 0.4.

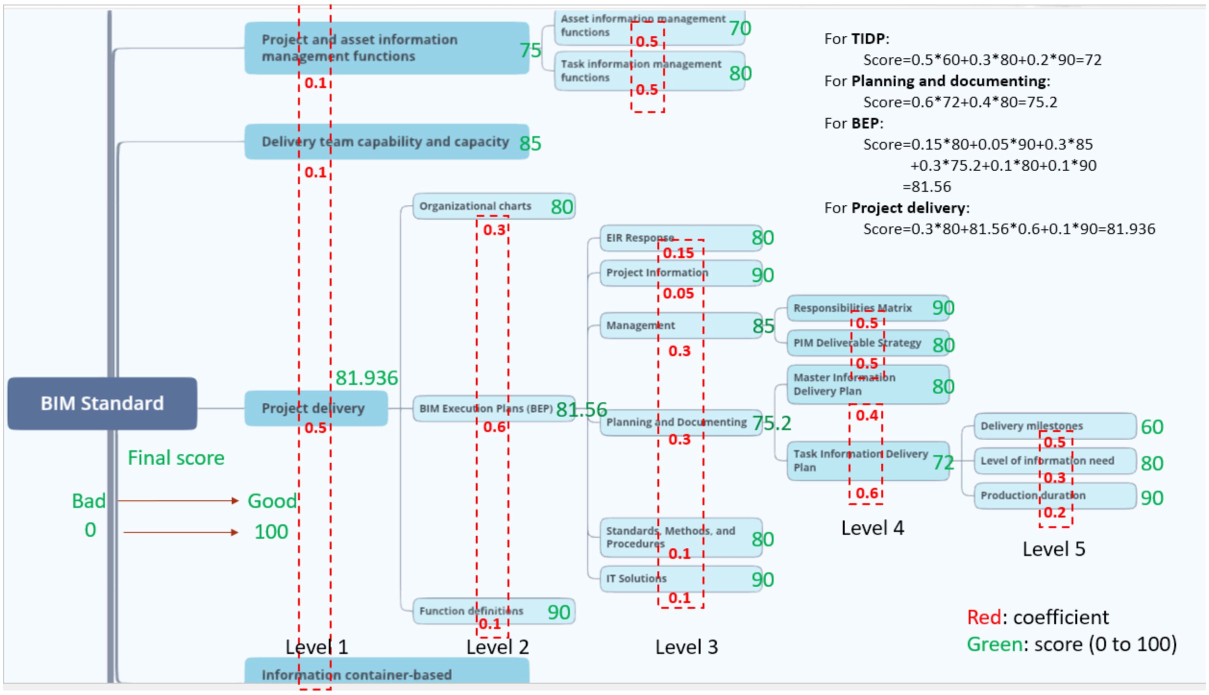


Figure 36 - Calculation mechanism of the assessment system

As shown in the graph, the score of each indicator is calculated from higher level to lower level. For instance, the score of the *Task Information Delivery Plan* is determined by its sub-indicators, which are Delivery milestones, Level of information need, and Production duration in Level 5. The score of *Planning and Documenting* is calculated by the score of the *Task Information Delivery Plan* and *Master Information Delivery Plan.* The detailed calculation process is presented on the top right of the diagram.

To enable switching between different application scenarios, a Python-based platform was developed, which enables flexible compliance checking against different numbers and types of BIM standards.

The platform is developed on a pure Python environment based on several third-party Python packages. The developed platform uses Streamlit to build a web interface and interact with users. This toolkit captures the information provided by the user and sends information to backend algorithms. After the information is processed, it presents the result and plot diagrams on the web page. Another Python library adopted in the platform is Owlready2, which is designed to operate ontology files with Python language. Through this toolkit, the information in the ontology can be easily queried, extracted and managed. The workflow of the platform can be divided into the following several steps:

1) Set target standards

   Setting a target standard is the first step, where the users are asked to select the BIM standards they want to check against on the interface of the platform. The selection can be a single standard or multiple standards. After setting standards, the user's selection will be sent to the backend platform, which is developed based on a Python library called RDFLib. With the help of this external library, the SPARQL queries will be automatically generated by the system to retrieve the required information from the knowledge model. This information includes the indicators to be checked, the weights of these indicators, the questions related to the indicators and a description of the requirements corresponding to each indicator in the standard.

   Here is an example. If the user selects GB/T 51212 as the target standard, the system will generate the following query.

   "PREFIX k: <http://www.semanticweb.org/zhu/ontologies/2021/6/untitled-ontology-96#> SELECT * WHERE { ?i k:isTheRequirementOf <http://www.semanticweb.org/zhu/ontologies/2021/6/untitled-ontology-96#" + *target_standard* + "> ; rdfs:comment ?b }"

   where the *target_standard* is a variable assigned by the user.

2) Complete questions

   After information retrieval, the questions related to the indicators are listed on the interface. The users need to answer all the questions according to their practical behaviour during the delivery of the project. These answers are then quantified based on predefined criteria and then used as the initial score of assessment indicators.

3) Present results

Considering some of the indicators are filtered out during the setting up of target standards, the weights of the selected indicators cannot be directly used for the final compliance score calculation. The updated weights of each indicator can be obtained through formula (8).

$$W_i^* = \frac{W_i}{\sum_i^j W_i} \tag{8}$$

where W_i represents the original weight of ith indicator in the knowledge model and W_i^* stands for the updated weight of ith indicator among all j selected indicators that belong to the same superclass and indicator level.

To realise the above function, a specific algorithm was developed in the backend of the platform, which can automatically update the weight of each indicator based on the standards users selected. The updated weightings and the quantified initial score are then sent to a calculation algorithm, where the final score for BIM compliance is calculated through formula (9).

$$S = \sum_j \sum_i W_i^{j+1} * S_i^{j+1} \tag{9}$$

where i, j represent the ith indicator in jth level (where the level counts down from 5 to 0). W_i^{j+1} and S_i^{j+1} stands for the weight and score of the ith indicator in the jth level.

In terms of result presentation, there are three types of results shown on the front end. The first one is the final score, which shows an overall evaluation of BIM compliance. The second one is a radar diagram, which provides more details on the user's strengths and weaknesses in BIM adoption. The last one is a recommendation list, which presents a series of suggestions to help users further improve their existing BIM workflow.

Chapter 6. Evidence-driven automated compliance checking

As summarised in Section 2.5, the existing BIM compliance checking approaches are flawed in comprehensiveness and automation. The ontological framework proposed in Chapter 5 resolves the issue of comprehensiveness. However, significant manual labour is still unavoidable to build knowledge models and complete questionnaires, resulting in subjectivity and uncertainty of compliance checking. To ensure the reliability of the assessment results, an evidence-driven automated compliance checking framework is proposed in this chapter. The framework takes the first scenario as the research case and uses project documents as input to conduct automatic BIM compliance checking. Figure 37 illustrates the overall structure of the proposed automatic compliance checking framework. The proposed framework comprises three main parts, namely regulatory knowledge extraction, project document information extraction, and autonomous compliance checking. The remainder of this chapter illustrates the development process following the three parts mentioned above and explains the design rationale and applied techniques.

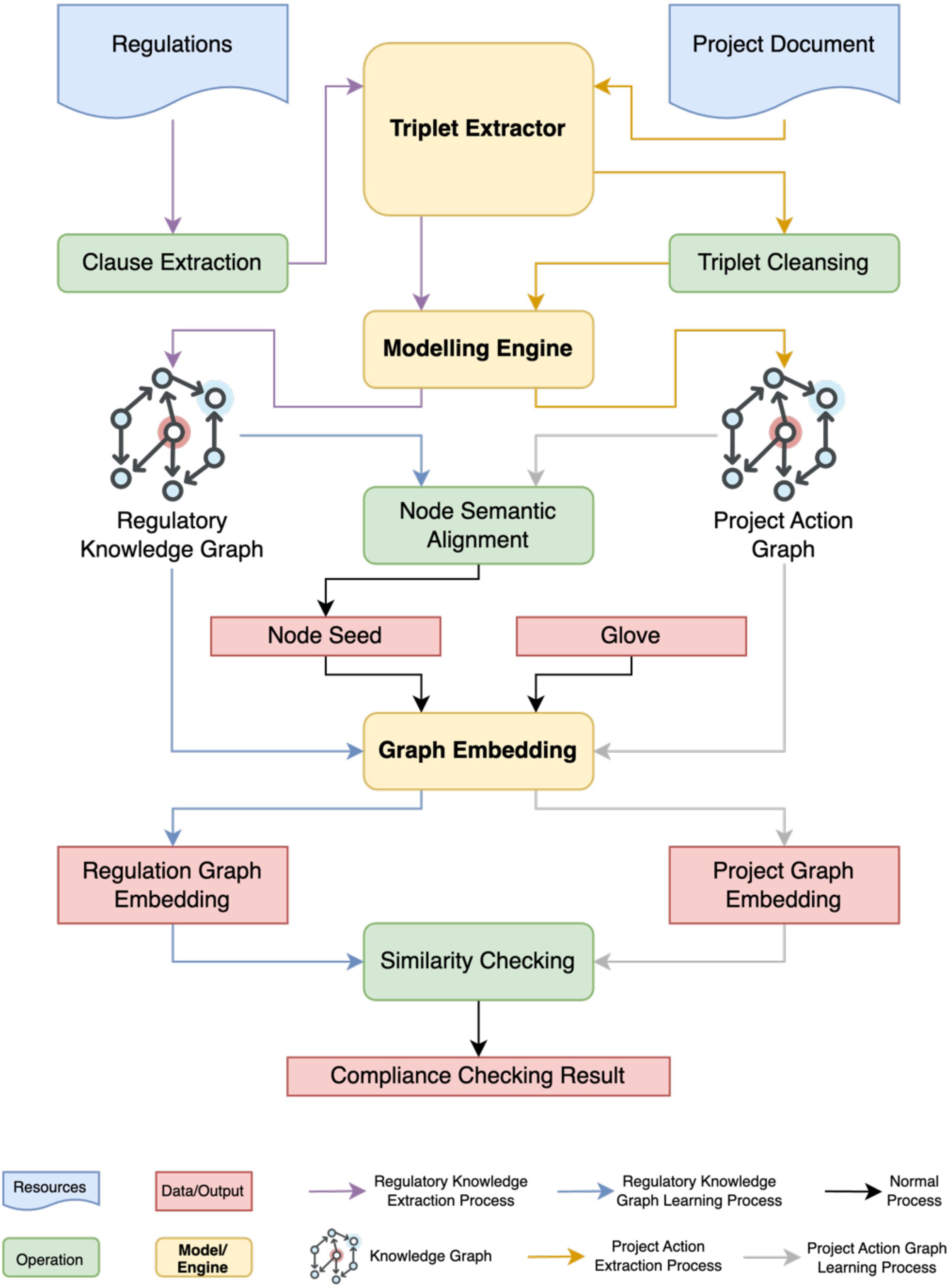


Figure 37 - Architecture of the evidence-driven compliance checking framework

6.1 Regulatory knowledge extraction

Regulatory knowledge extraction is the fundamental part of the proposed automatic compliance checking framework, which aims to extract the textual knowledge in the standard documents and convert it to a structured graph representation. The development work of this part can be divided into four steps, including the establishment of domain dataset, clause extraction, triplet extraction, and graph modelling. In the proposed framework, a domain dataset is first built through the Delphi method and data augmentation to fine-tune the pre-trained large language model (LLM) so that it can identify and extract clauses from standards automatically. Then, the extracted clauses are converted into knowledge triplets through natural language processing and joint label mapping. Finally, these knowledge triplets are assembled as a regulatory knowledge graph, which contains all the requirements specified in the regulation documents. The pipeline of automatic regulatory knowledge extraction developed in this research is shown in Figure 38.

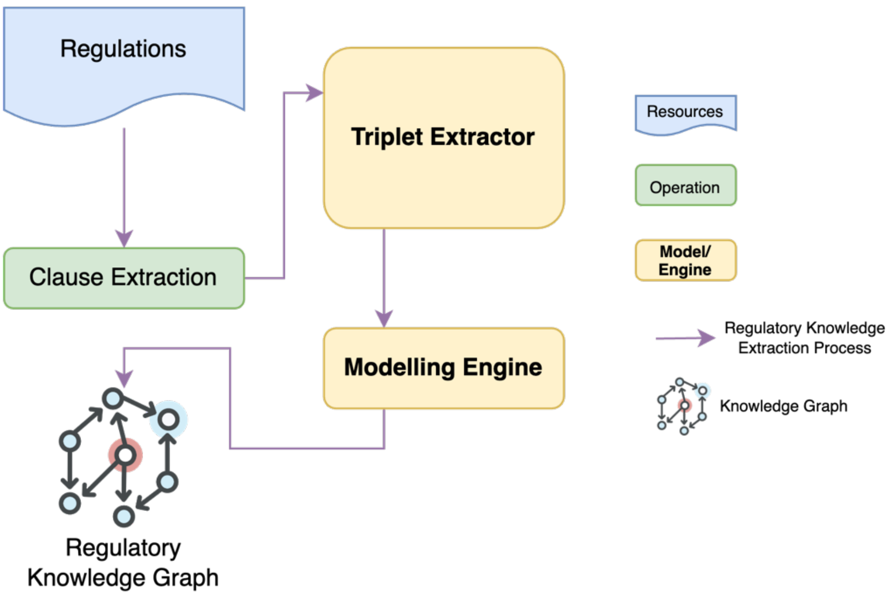


Figure 38 - The pipeline of automatic regulatory knowledge extraction approach

6.1.1 Domain dataset establishment

After scanning all available related datasets in the AEC domain, there are currently no suitable public datasets that can be directly used for the training of clause extraction. Consequently, the development of a domain dataset becomes imperative for training the deep learning model. This dataset is comprised of clause samples and description samples extracted from regulation documents.

For the requirement extraction task in this study, there is no available open dataset in the AEC domain that can be directly applied to this research. Although Zhou et al. published an open regulation dataset (Zhou et al. 2022b) with annotated sentences, the dataset does not fully satisfy the requirements of this research, as the labels in the dataset are self-defined and the samples are all from quantitive technical design code. In view of the circumstances, a specific domain dataset is required to be manually developed to fine-tune the deep learning model proposed in the framework of this study. To ensure the generalisation, several BIM standards, design codes and related qualitative standards are selected as sample sources. Table 12 illustrates the selected standards and rule samples extracted from each standard.

Table 12 - Engineering standards included in the domain dataset.

Standard Code	Description	Published by	No. of Rules extracted
ISO 9001	Quality management systems Requirements	ISO[a]	216
ISO 14001	Environmental management systems -Requirements with guidance for use	ISO	169
ISO 50001	Energy management systems - Requirements with guidance for use	ISO	153
ISO 19650-1	Organization and digitization of information about buildings and civil engineering works, including building information modeling (BIM)	ISO	199
2015 IBC	International Building Code	ICC[b]	48
GB/T 51212	Unified standard for building information modeling	MOHURD[c]	41
Total Number			826

[a]International Organization for Standardization, [b]ICC - International Code Council, [c]MOHURD – Ministry of Housing and Urban-Rural Development of China

To ensure diversity and representativeness, a variety of engineering standards released by different institutions were chosen as data sources. Additionally, samples from engineering-related standards, such as ISO 9001, were included to augment the generalisation capabilities of the neural network model. As detailed in **Error! Reference source not found.**, a total of 826 samples were manually retrieved from the selected standards. Due to the typical prevalence of a higher number of clauses than descriptions within regulatory documents, the extracted samples comprise 573 clause samples and 253 description samples.

The efficacy of the neural network model is widely acknowledged to be significantly dependent on both the quality and quantity of the data. Achieving satisfactory performance in training neural network models for classification tasks typically requires a minimum of 1000 samples for each category. Hence, the initial domain dataset suffers from data shifting and insufficient samples. Moreover, the manual labelling of clause and description samples in the original dataset introduces subjectivity and potential errors. To tackle these issues, data augmentation techniques are employed to expand and balance the samples in the original dataset. The Delphi method is adopted to mitigate subjectivity and uncertainty in the process of sample labelling.

6.1.1.1 Data Augmentation

Data Augmentation is a technique for artificially extending a training dataset by making a limited amount of data produce more equivalent data. It was first applied in the field of computer vision to overcome the problem of insufficient training data and is now widely used in various areas of deep learning (Maharana et al. 2022). There are four types of augmentation methods for textual data: character-level augmentation, word-level augmentation, phrase-level augmentation, and document-level augmentation (Bayer et al. 2021). Character-level data augmentation creates new training samples by changing individual characters in an existing training sample. Word-level data augmentation methods replace words in the text based on a lexicon or word vector without changing the main idea of the sentence. Phrase-level data augmentation generates new training samples by changing the structure of sentences. Document-level data augmentation creates new training samples by changing entire sentences in a document, including back translation (Coulombe 2018) and generative methods (Qiu et al. 2020). Given that the samples are sentences intended for training of classification, the preferred augmentation method is back translation, which generates more variants by running reverse translation in a different language to augment the unlabelled rule sentence samples.. Therefore, this research employs the back translation approach to augment the current samples and generate a larger, balanced dataset containing 1000 clause samples and 1000 description samples. Chinese and French were chosen as intermediary languages due to their widespread usage. The detailed procedure for back translation is depicted in Figure 39 and is elucidated as follows:

1) Determine the number of positive and negative samples slated for augmentation.
2) Randomly pick an equivalent number of positive and negative samples from the dataset based on the statistical outcomes.
3) Sequentially translate the chosen samples into French and Mandarin using a third-party translator (Google translation). Subsequently, translate them back into English using another translator (DeepL) to create the translated samples.
4) Compare the newly translated samples with the original ones, and if there are differences in expression, add the translated samples to the dataset as augmented samples.

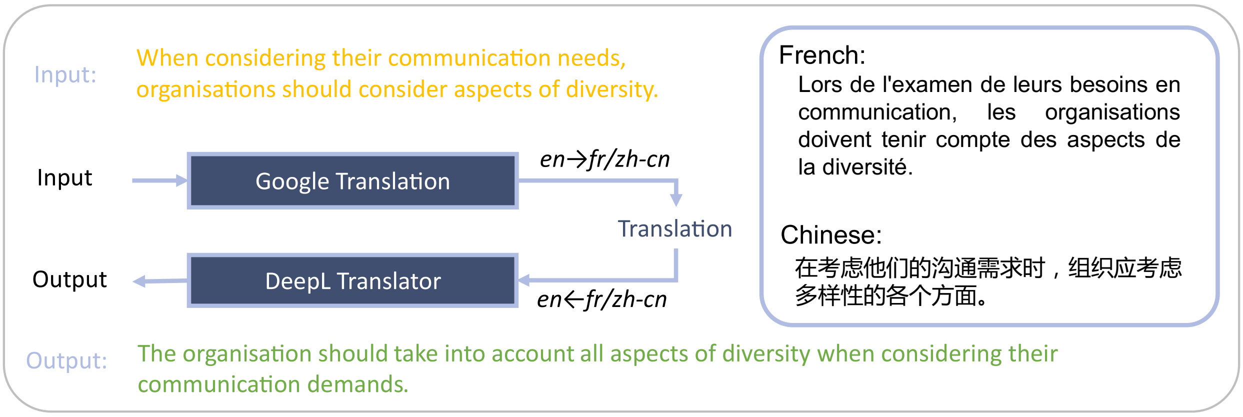


Figure 39 - The procedure of Back Translation with a specific example

6.1.1.2 Delphi validation

The Delphi Method (also called the Delphi technique) is essentially a feedback-anonymous correspondence method, which was pioneered by O. Helm and N. Dahlke in the 1940s (Dalkey and Helmer 1963) and further developed by T. J. Gordon and Rand Corporation. The general process is that after obtaining the opinions of experts on the issue to be predicted, they are collated, summarised, counted and then anonymously fed back to each expert, consulted again, pooled again, and fed back again until a consensus is obtained. The process can be simply expressed as Figure 40. The Delphi Method can obtain relatively objective information, opinions, and insights through the independent and repeated subjective judgment of several experts. There are currently no explicit regulations regarding the number of experts. The number of panellists mainly depends on the topic as well as the time and resources at the researchers' disposal. As stated by Beiderbeck et al. (Beiderbeck et al. 2021), 5 to 8 experts are sufficient to organise a Delphi survey. Even though the involvement of more experts can enhance the reliability and objectivity of the outcomes, too many experts can lead to an increase in the difficulty and complexity of the survey. Additionally, an excessive number of experts can make it extremely difficult to reach a consensus.

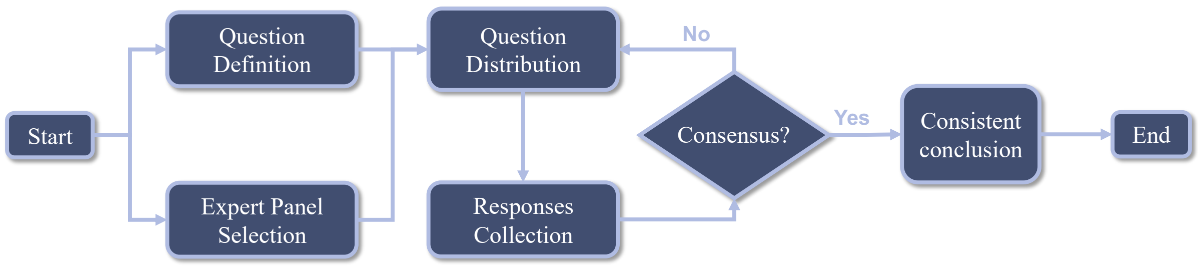


Figure 40 - General flow of the Delphi method

In this research, the Delphi method is applied to enhance the reliability of sample labelling. To conduct the Delphi method, the three experts from the calibration panel and two more PhD candidates who conduct research on building design are invited to validate the labels in the dataset. Hence, all these five experts have an excellent understanding of engineering standards. The detailed validation process is depicted in Figure 41 and is elucidated as follows:

1) Subject all sample and label data to the initial round of expert group validation.
2) The research group consolidates and summarizes the validation outcomes provided by the expert group. Samples garnering unanimous agreement from all experts pass validation directly. If more than half of the experts agree on a sample, the research group adjusts it according to expert opinions and proceeds to the next validation round. In cases where a sample receives validation support from only a few experts, a new sample of the same label type is generated through data augmentation, replacing the original, and undergoing validation in the subsequent round.

3) Repeat the validation process for all adjusted samples, iterating through the above steps until a unanimous agreement is reached among all experts.

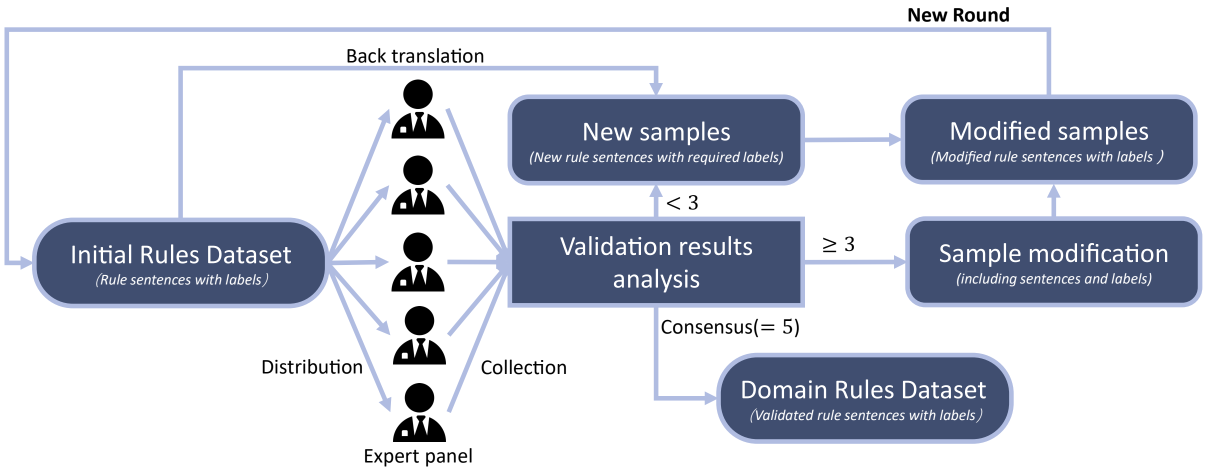


Figure 41 - The procedure of the Delphi method for sample labelling validation

Following four rounds of validation and modification, consensus among the five experts was ultimately achieved for all dataset sample labels. Through the implementation of data augmentation and the Delphi method, the domain dataset was effectively transformed into a well-balanced set of samples with reliable labels.

6.1.2 Transfer learning-based clause extraction

Recently, deep learning has gained more and more attention and development due to its excellence in the processing of complex forms of information (images, sound, text). As the complexity of the task increases, deep learning models also show a tendency to become more and more complex. In general, the complexity of a model is positively correlated with the amount of data required for training. As the labelling of training data still mainly relies on manual, the cost of acquiring large amounts of training data is quite high, which makes training or improving deep learning models a time-consuming and labour-intensive process (Pan and Yang 2010). In order to solve this problem, attempts have been made to extract the common knowledge generated during the deep learning process for relevant machine learning tasks, so that it is possible to reuse what has already been learned without having to start from scratch, saving a lot of resources and time. This is the basic idea of transfer learning. Based on this idea of reuse, transfer learning divides a complete training task into two phases: pre-training and fine-tuning (Yosinski et al. 2014). The pre-training phase aims to generate models that contain reusable knowledge, i.e., pre-trained models. The fine-tuning phase involves designing and adding fine-tuned layers to the pre-trained model based on specific task requirements. In the proposed framework, a large language model (BERT) is selected as the pre-trained model that has already learnt universal language representations. The pre-trained model captures the semantic features of clauses after being fine-tuned with domain samples and can be employed as an extractor to retrieve clauses from regulation documents.

6.1.2.1 Pre-trained language model

Extracting clauses from regulation documents is fundamentally a binary classification task for sentences (Zhu et al. 2023). In light of this, the *BertForSequenceClassification* model is chosen as the pre-training model. This model incorporates a sequence classification head atop the BERT base model. BERT (Devlin et al. 2018), namely Bidirectional Encoder Representations from Transformer, extracts feature information from the input sequence based on the bidirectional encoder provided by Transformer. With the help of its attention mechanism, the BERT model can capture long-distance dependencies and generate a feature vector for each sequence element (word) based on the contextual features of the input sequence. Therefore, it performs better than other deep neural network models (e.g., RNN, LSTM) in terms of efficiency and stability, with a wider range of applications (Xue and Zhang 2021). More details about the comparison between the BERT and other state-of-the-art models can be found in Chapter 7. The structure of the selected pre-trained BERT base model is illustrated in Figure 45.

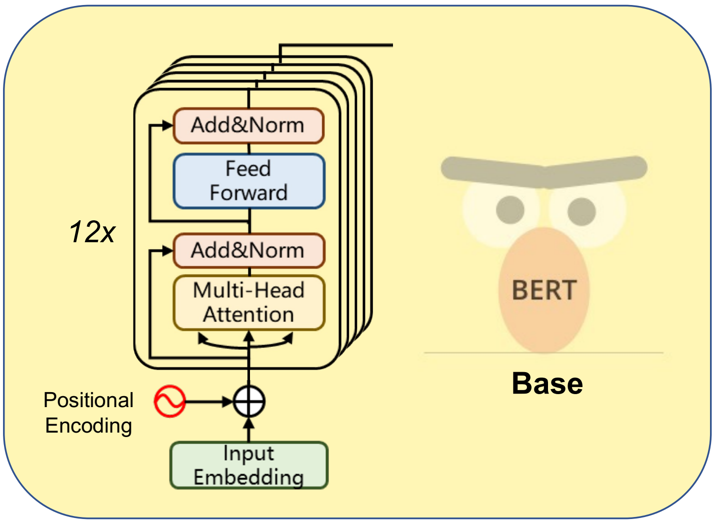


Figure 42 - The architecture of the pre-trained BERT base model

6.1.2.2 Fine-tuning

Fine-tuning the pre-trained BERT model for clause extraction involves a series of essential steps: pre-processing, tokenization, data packing, training, and testing. The detailed process of fine-tuning, using a clause example, is depicted in Figure 43.

1) Pre-processing and tokenisation

The samples in the domain dataset are all derived from regulation documents, which have a normative expression. Hence, common pre-processing methods like eliminating web links, stop words, and special characters become unnecessary. Nevertheless, some additional steps are required to guarantee the effectiveness of the resulting model, including:

(1) Substitute semicolons with dots. Numerous regular texts are composed with semicolons, posing challenges for proper splitting and tokenization.
(2) Eliminate all redundant whitespaces and consecutive multiple spaces. This is a standard pre-processing step for textual data.
(3) Exclude sentences that exceed 200 words. Sentences of such length often contain numerous words or numbers that do not form coherent sentences.
(4) Calculate dataset characteristics, including the number of samples, maximum sequence length, lexicon size, etc. These parameters are utilised in subsequent steps.

Prior to loading data into the BERT model, two additional steps are essential in the pre-processing stage. Firstly, two special tags (CLS and SEP) should be inserted to convert the original sample into the *[CLS]* + *sentence* + *[SEP]* format. Secondly, the words in the original sentence should be randomly replaced with *[MASK]* and random words (rnd) to enable attention mechanism. Following the completion of data pre-processing, a tokenizer (BertTokenizer) is employed to convert the textual samples into a sequence of IDs in the corpus and simultaneously generate a token type tensor and a tensor of attention mask.

2) Data packing

The generated dataset comprises a modest size of 2000 samples. The partitioning of training, validation, and test sets is set with a ratio of 80%, 10%, and 10%, respectively. In the context of fine-tuning tasks, a batch size of 16 or 32 is deemed suitable. After experimentation, the authors determined a batch size of 32 as optimal for fine-tuning.

3) Training and testing

Before initiating training, several hyperparameters need to be determined, including batch size, training epochs, and learning rate. Following the minimal hyperparameter tuning strategy recommended by Devlin et al. (Devlin et al. 2018), learning rates are advised to be within the range of 2e-5 to 5e-5, and the number of training epochs should typically be set to 3 or 4. To mitigate underfitting, the training epoch is set as 5 and the learning rate is chosen as the minimum value (2e-5). Additionally, a learning rate optimization strategy known as linear warmup is implemented to prevent overfitting and maintain model stability. The AdamW optimizer is employed for gradient computation and parameter updates. The testing process mirrors the training process, with the exception that gradients are set to zero during backpropagation. Further details on testing results can be found in Chapter 6.

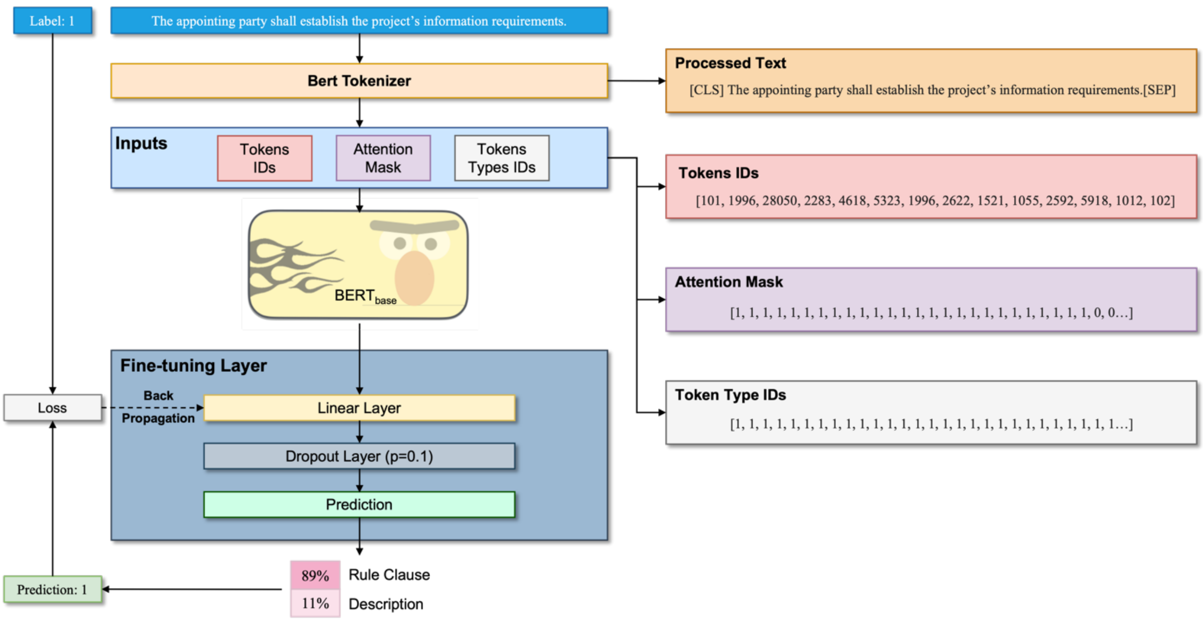


Figure 43 - The fine-tuning process in the proposed framework with a specific clause sample

6.1.3 Regulatory knowledge triplet extraction

Unlike other studies, the subject of this study is the regulation documents, where the expressions are flexible but normative. Through the literature review and analysis of existing methodologies, a hybrid approach that combines rule-based label mapping and syntactic parsing is optimal.

Syntactic parsing (also called parsing) (Wikipedia 2024c) is one of the key techniques in natural language processing. It is the process of analysing an input text sentence to obtain the syntactic structure of the sentence according to a given formal grammar. According to the representation of syntactic structure, the most common syntactic analysis techniques can be divided into two categories (Zhang 2020): 1) phrase structure parsing, which aims to identify the phrase structure in the sentence and the hierarchical syntactic relationships between the phrases. 2) dependency syntactic parsing, or dependency parsing, which focuses on the recognition of the interdependencies between words in a sentence.

In linguistics, sentences are constructed with progressively nested constituents (i.e. phrases, words), which are words organised by phrase structure. Theoretically, a fixed number of phrase structures can generate an infinite number of sentences. In other words, every sentence can be decomposed into several nested phrase structures. Phrase structure grammar (PSG) is a context-free grammar (CFG) based parsing method, which was first proposed by Noam Chomsky in 1956 (Chomsky 1956). It demonstrates the structure of constituents in sentences relying on constituency relations. According to Chomsky (1956), “a phrase-structure grammar is defined by a finite vocabulary (alphabet) V_p, a finite set Σ of initial strings in V_p, and a finite set F of rules of the form: X to Y, where X and Y are strings in V_p” (Zhang et al. 2013). Phrase structure parsing is the parsing method developed based on pre-defined PSG and POS tags. Figure 44 shows an example rule clause (i.e., the appointing party shall establish the requirements that tendering organizations shall meet within their tender response) derivation based on phrase structure parsing.

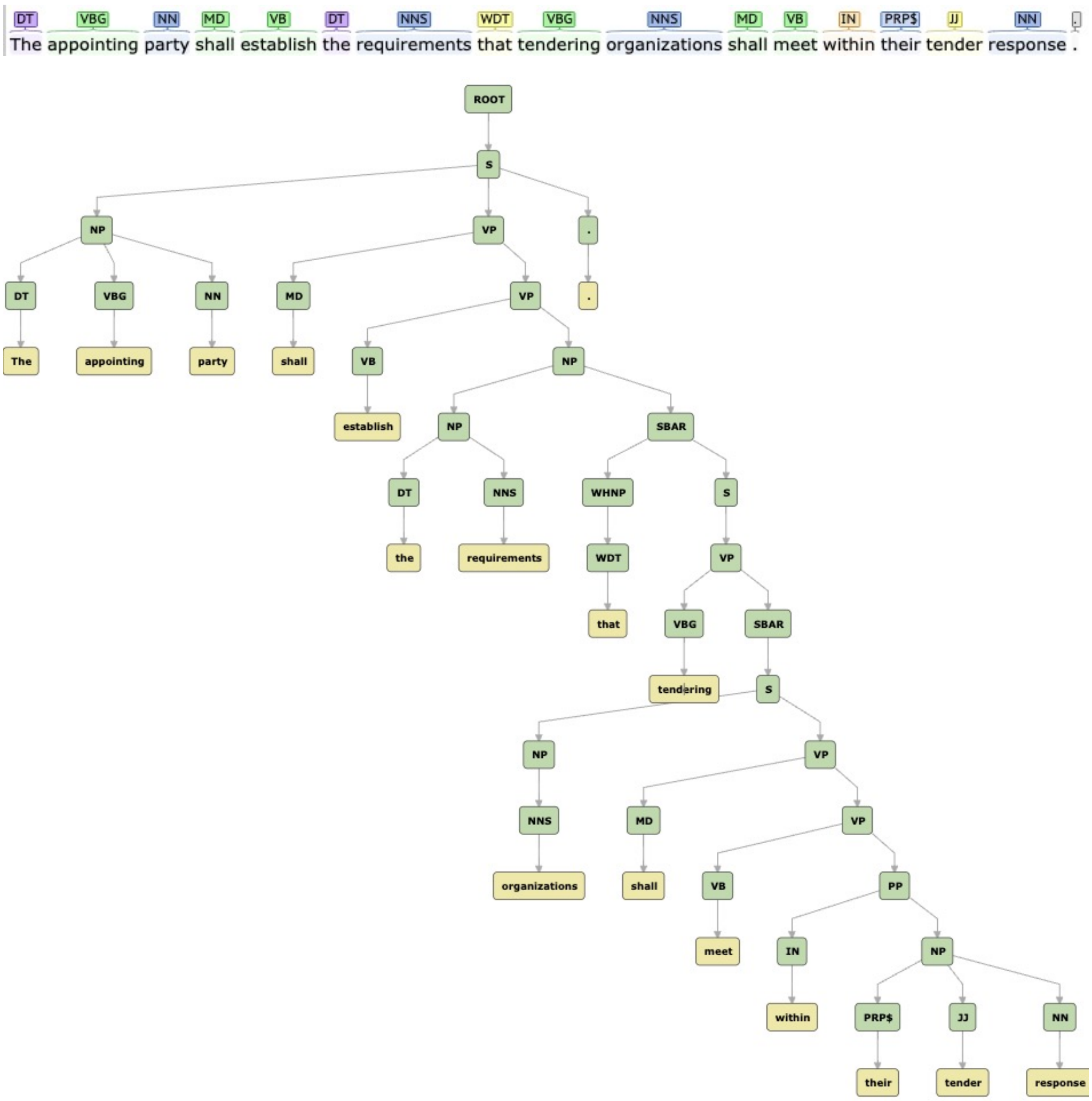


Figure 44 - An example of phrase structure parsing-based rule clause derivation

As can be seen from the example (Figure 44), phrase structure has various combinations and its number is also uncertain, making it quite challenging to automate syntactic parsing by mapping specific phrase structures. Although many sentences are manually decomposed and some corpus (e.g., Penn Treebank, Chinese Treebank) is established with the aid of the PSG developed by linguists, the accuracy of the phrase structure parser is not satisfying, especially for long sentences. For short sentences or phrases, the phrase structure parser is fast and accurate. Since normally the concept and its attributes are in the same phrase, the authors adopt phrase structure parsing to extract the attributes of the concepts and leave the rest of the constituents to neural network-based dependency parsing.

Dependency parsing refers to examining the dependencies between the words of a sentence to analyse its grammatical structure and define the relationships between "head" words and words, which modify those heads. Relations among the words are illustrated above the sentence with directed, labelled arcs from heads to dependents. An example of dependency parsing is shown in Figure 45.

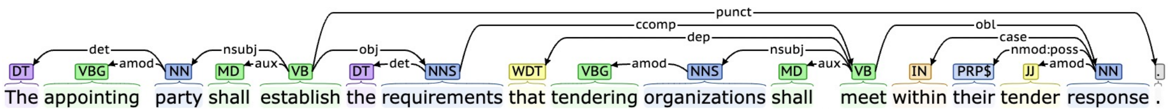


Figure 45 - An example of dependency parsing-based rule clause derivation

In this research, a constituent extraction engine is crafted by leveraging these two techniques to automate the information extraction of clauses. Its architecture comprises a clause classifier, two clause processors, and a constituent extractor. Two external toolkits (DependencyParser and Tagger) provided by SpaCy are integrated into the parsing engine to generate dependency parsing (DP) labels and POS tags, respectively. The process of constituent extraction involves the following three stages:

1) Clause classification

During this phase, the parser conducts an initial analysis of raw clauses that are extracted by the clause extractor and generates the parsing label for each word within the clause based on syntactic grammar. Subsequently, the raw clauses undergo a classification process based on the parsing labels, wherein they are categorized into coordinate clauses, compound clauses, and simple clauses. This classification is conducted through the clause classifier, employing a joint mapping approach that considers specific dependency labels and POS tags. For instance, a clause is identified as an adverbial clause if it incorporates a word with a dependency label of "*mark*" and a POS tag of "IN". The details of various complex clause examples, their corresponding markers, the association between clause types, and the parsing labels of these markers are presented in Table 13.

Table 13 - Correspondence between clause types and marker's parsing labels

Examples with Markers	Clause Type	DP Label	POS Tag
The appointing party should understand **what** information is required concerning their asset(s) or project(s). (Clause 5.1 of ISO19650-1)	Compound (Object)	ccomp, mark	IN, WDT
If the review is successful, the lead appointed party shall authorize the information model and instruct each task team to submit their information. (Clause 5.7.2 of ISO 19650-2)	Compound (Adverbial)	advcl, mark	IN, WRB
The requirements should be expressed in such a way **that** they can be incorporated into project-related appointments. (Clause 5.5 of ISO19650-1)	Compound (Relative)	relcl	WDT, WRB, WP$
Exterior load-bearing walls **and** nonload-bearing walls shall be mass timber construction. (Clause 602.4 of IBC 2015)	Coordinate	cc, conj	CC

2) Clause simplification

Upon classification by the clause classifier, simple clauses undergo direct transferal to the extraction phase. In contrast, coordinate clauses and compound clauses undergo simplification through the designated coordinate and compound clauses processors, respectively. This simplification process coverts complex clauses into several simple clauses.

Concerning coordinate clauses, the processor identifies juxtaposed elements based on dependency labels and POS tags. Given the common omission of repeated content in coordinate clauses, the sentence parts (S, P, O, A, C) of the juxtaposed elements are determined with reference to their dependency labels and POS tags. Subsequently, the processor dissects the coordinate clause into two distinct clauses, both adhering to the same clause pattern (refer to **Error!**

Reference source not found.). As an illustrative example drawn from ISO 19650-1, the original clause reads, "The complexity of project information management functions should reflect the extent and complexity of project information." The juxtaposed elements, namely "the extent" and "the complexity," serve as the objects in the clause. Following the aforementioned principle, the missing components (subject and predicate) are required to be supplemented to form a complete sentence when decomposing the sentence. Consequently, the two simplified sentences are: "The complexity of project information management functions should reflect the extent of project information" and "The complexity of project information management functions should reflect the complexity of project information."

In the context of compound clauses, with the consideration that subject clauses seldom occur in regulatory documents, the compound clause processor primarily concentrates on predicative clauses, object clauses, attributive clauses, and adverbial clauses. Several examples of compound clauses in regulatory documents are provided in Table 13. Similar to the approach employed by the coordinate clause processor, the compound clause processor utilises a joint mapping of dependency labels and POS tags to ascertain the sentence part of the subordinate clause. The identified subordinate clauses are preserved separately and are distinguished by a separate label of corresponding sentence parts. The detailed process of classification and simplification is elucidated in Figure 46, utilising a complex clause extracted from ISO 19650 as a representative example.

The process of clause classification and simplification operates cyclically. Each simplified clause is returned to the original set of clauses and subjected to reclassification iteratively. This cyclic procedure continues until all clauses have been subjected to the simplification process, culminating in their progression to the next stage as simple clauses.

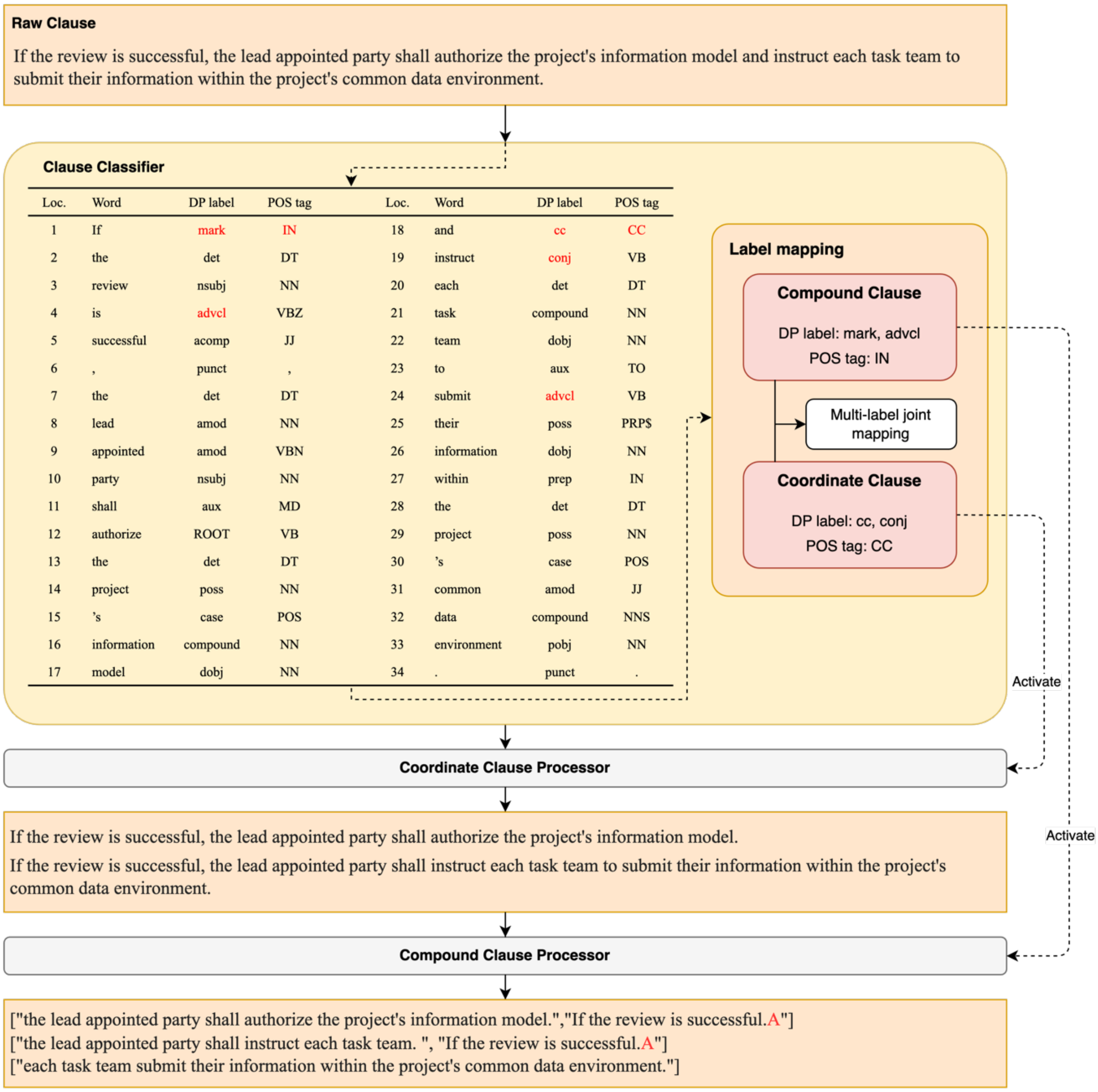


Figure 46 - The specific procedure of clause classification and simplification with a clause example from ISO 19650-1

3) Constituent extraction

The constituent extractor is devised by integrating syntactic parsing principles and the Seven Clause theory. The Seven Clause theory is a linguistic categorisation of English sentences. According to the Seven Clause theory, A clause is a part of a sentence that expresses some coherent piece of information, which comprises subject (S), verb (V), indirect object (O_i), direct object (O), complement (C), and adverbials (A) (Corro and Gemulla 2013). From the perspective of linguistics, not all combinations of these constituents appear in the English language. A sentence consisting of one subject (S), one verb (V), and optionally of an indirect object (O_i), a direct object (O), a complement (C), and one or more adverbials (A) is defined as a basic clause. All basic clauses can be classified into seven different types according to the grammatical function of their constituents (Weiss 1987). A complete list of all seven basic clause types is given in Table 14. Additionally, there may be occasionally more than one subject or verb in a sentence, which is called a coordinate clause. There is also another common clause type named compound clause, where some constituents (subject, object, adverbial, attribute) of the sentence are expressed as individual sentences rather than words or phrases. Nevertheless, the coordinate clause and the compound clause can be decomposed into two or more basic clauses.

Table 14 - Patterns and clause types defined by the Seven Clause theory

Pattern	Clause type	Example
		Basic patterns
SV_i	SV	A. Einstein died.
SV_cA	SVA	A. Einstein remained in Princeton.
SV_cC	SVC	A. Einstein is smart.
$SV_{mt}O$	SVO	A. Einstein has won the Nobel Prize.
$SV_{dt}O_iO$	SVOO	RSAS gave A. Einstein the Nobel Prize.
$SV_{ct}OA$	SVOA	The doorman showed A. Einstein to his office.
$SV_{ct}OC$	SVOC	A. Einstein declared the meeting open.
		Some extended patterns
SVAA	SV	A. Einstein died in Princeton in 1955.
SV_iAA	SVA	A. Einstein remained in Princeton until his death.
SV_eCA	SVC	A. Einstein is a scientist of the 20th century.
$SV_{mt}OA$	SVO	A. Einstein has won the Nobel Prize in 1921.
$ASV_{mt}O$	SVO	In 1921, A. Einstein has won the Nobel Prize.

S: Subject, V: Verb, C: Complement, O: Direct object, O_i: Indirect object, A: Adverbial, V_i: Intransitive verb, V_c: Copular verb, V_e: Extended-copular verb, V_{mt}: Monotransitive verb, V_{dt}: Ditransitive verb, V_{ct}: Complex-transitive verb

Based on this theory, a tuple extraction algorithm and an attribute extraction algorithm are developed in the constituent extractor to realise the function of sentence simplification and constituent extraction. The tuple extraction algorithm, rooted in dependency parsing, is crafted to identify constituents within clauses by mapping specific tags or sequences of tags. In alignment with the Seven Clauses theory, all simple clauses consist of five essential components: subject (S), predicate (P), object (O), complement (C), and adverbial (A), or a subset thereof. Consequently, a quintuple (S, P, O, V, C) is generated to encapsulate the corresponding constituents. Meanwhile, the attribute extraction algorithm, grounded in phrase structure parsing, is designed to extract attributes of entities mentioned in clauses. The extracted attributes and their associated entities are stored in an attribute matrix. Figure 47 visually depicts the constituent extraction process and the outcomes for the first simplified clause presented in Figure 46.

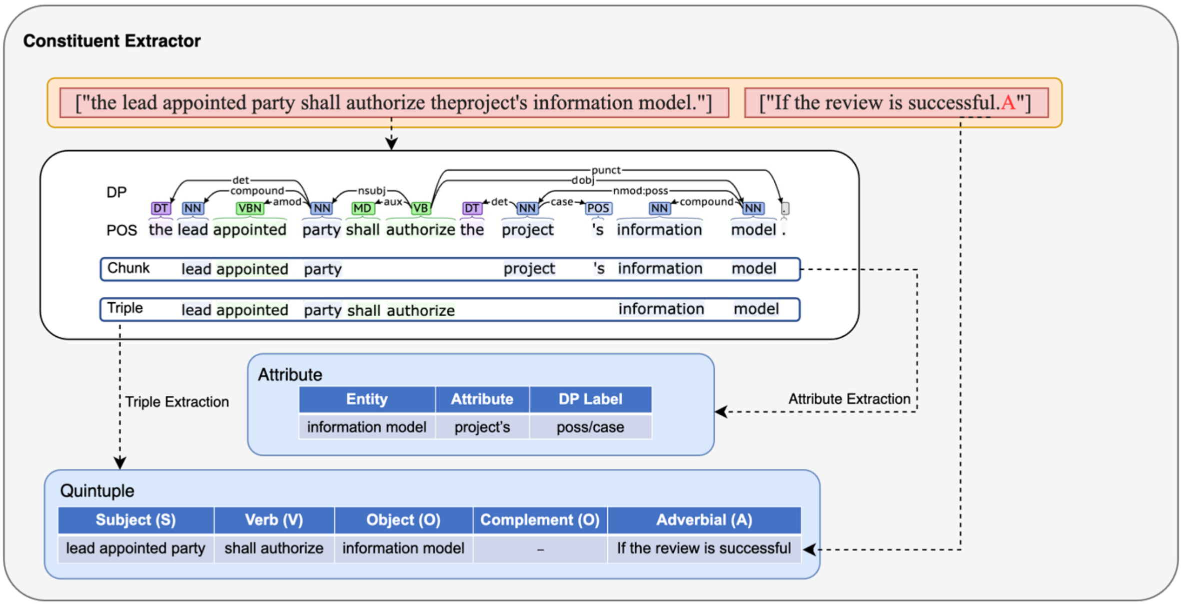


Figure 47 - The constituent extraction process on simplified clause

The specific extraction process of the constituent can be delineated into the following five primary steps:

(1) Predicate extraction. The initial step involves extracting words parsed with the ROOT (DP label) and those possessing verb-related POS tags (VB/VBD/VBG/VBN/VBP/VBZ) as predicates. Additionally, if the predicate is preceded by words labelled as neg or aux, these words are amalgamated to form an integral part of the predicate.

(2) Subject extraction. Subjects are extracted based on the fulfilment of two conditions: the head word of the subject has a DP label of ROOT, and the subject itself is tagged with a DP label of either nsubj or nsubjpass.

(3) Object extraction. Objects are extracted by identifying words with DP labels such as dobj or dative, where the head word's DP label is ROOT.

(4) Adverbial extraction. Adverbials are extracted through the application of specific label combinations within clauses, including patterns such as prep + (pcomp+) dobj, advmod, agent + pobj, etc. The triple extraction algorithm facilitates the extraction of adverbials by mapping these label sequences.

(5) Attributes extraction. Attributes are identified through the attribute extraction algorithm, operating concurrently with the triplet extraction algorithm. This algorithm initially identifies noun chunks within the clauses. Then, words with labels such as nummod, quantmod, poss, case, etc., are extracted as attributes of the central noun (entity). Additionally, complements extracted by the tuple extraction algorithm are also preserved as attributes.

6.1.4 Regulatory knowledge graph modelling

A knowledge model is a data structure that describes the entities, events, processes, and relations between them, which can be used by computers to solve complex tasks. Typical forms of knowledge representation are production systems, frame systems, and semantic webs/ontologies. Ontologies describe the relationships between events, concepts, situations, actions, and objects using a directed graph structure by nodes and marked edges. Due to the semantic network's intuitiveness, scope, and representational power, it is currently the most widely used knowledge representation method. In the proposed framework, a knowledge modelling engine is developed to automate the modelling of regulation knowledge.

Within the framework of the proposed knowledge modelling engine, two algorithms play pivotal roles: the tuple transformation algorithm and the knowledge integration algorithm. The tuple transformation algorithm is designed based on the Seven Clause Theory, with the specific objective of converting the quintuples and attribute matrices extracted by the constituent extractor into either RDF triples (node, edge, node) or RDF reifications (statement, subject, predicate, object). In parallel, the knowledge integration algorithm utilises an external Python library (RDFLib) to assemble the generated RDF triples and RDF reifications into a graph-based regulatory knowledge model, built upon the foundations of OWL and RDF schema.

In the implementation phase, all the extracted quintuples (S, P, O, C, A) underwent initial classification into rule triples and rule quaternions based on the Seven Clause theory by the tuple transformation algorithm. Quintuples with compositions of (S, P, O), (S, P, A), or (S, P, C) were categorized as rule triples, which can be directly reconstructed as RDF triples. Quintuples with compositions of (S, P, O, A) or (S, P, O, C) were designated as rule quaternions, necessitating a conversion into RDF reifications in the form (A/C, S, P, O) prior to integration. Attributes within the attribute matrices were represented as RDF triples following a fixed pattern (entity, should_be, attribute). Following the tuple transformation, the resulting RDF triples and RDF reifications were assembled into a regulatory graph by the knowledge integration algorithm. This algorithm automatically generated Internationalized Resource Identifiers (IRIs) for each element in the tuples based on a predefined namespace and associated other triples and reifications according to these unique IRIs.

To consolidate the extracted knowledge, four knowledge representation rules were established within the integration algorithm based on RDF syntax:

1) subject (S)/object (O) → rdf:type → OWL.NamedIndividual;
2) predicate (P) → subPropertyOf → topObjectProperty;
3) subject (S) → should_be → complement (C);
4) adverbials (A) → rdf:type → rdf:Statement.

Following processing by these two knowledge modelling algorithms, a graph-based knowledge model was established, encompassing all regulatory knowledge extracted from regulation documents. This model is amenable to querying and visualization by external services. Figure 48 provides a visual representation of the tuple transformation and knowledge integration process based on the preceding extraction results.

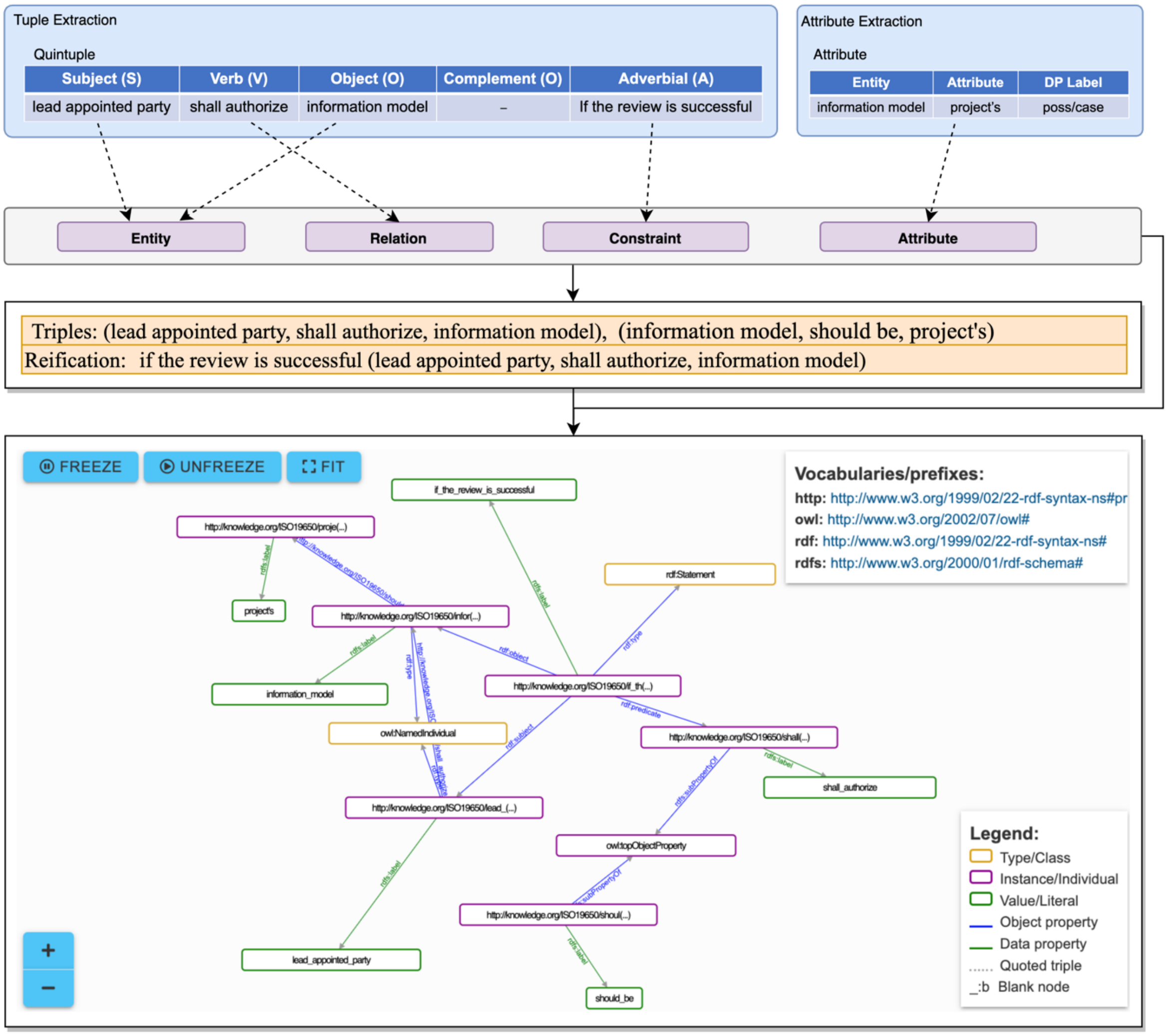


Figure 48 - The process of knowledge modelling based on the previous extraction results

6.2 Project document information extraction

Information extraction from project documents is another important module in the automatic compliance checking framework, which aims to extract the actual behaviour of different stakeholders during project delivery from the project record documents and transform them into graph representations. The extraction of actual behaviours is similar to regulatory knowledge extraction regarding the research objectives, which focuses on mining the factual actions rather than regulations from textual documents. Therefore, the mechanism of combining an extractor and a filter can also be implemented for fact extraction, where the extractor is adopted to extract information and the filter is utilised to remove irrelevant or redundant information. However, the action data relevant or irrelevant to the BIM compliance checking are identical in terms of textual features. Hence, the deep learning-based approach is incapable of data cleansing and a new cleansing algorithm is required to be developed. Compared with regulatory knowledge extraction (Figure 49), the extraction of project action (Figure 50) follows a similar process, where the triplet extractor and modelling engine developed in regulatory knowledge extraction are utilised to extract factual triplet and model action graphs.

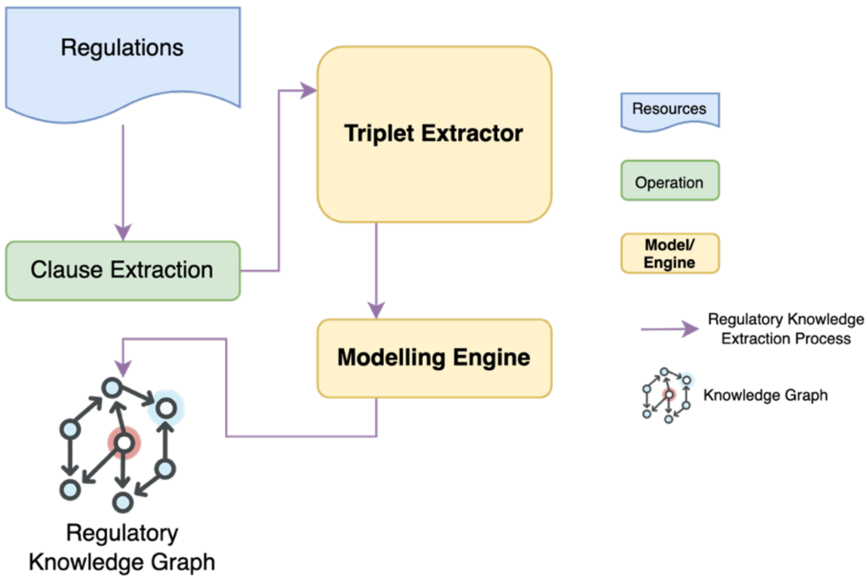


Figure 49 - Process of regulatory knowledge extraction

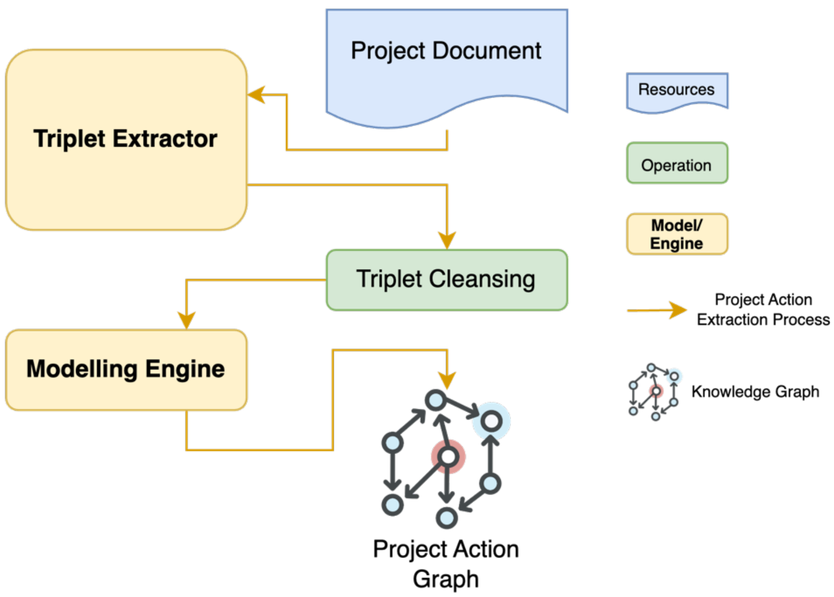


Figure 50 - Process of project action extraction

As shown in Figure 50, the whole process of project document information extraction can be divided into 3 main steps, including factual triplet extraction, triplet cleansing and graph modelling. Firstly, the triplet extractor split the record document into sentences. Then, all actual behaviour implemented by different stakeholders is extracted from these sentences and converted into action triplets. After triplet extraction, triplet cleansing is conducted to remove irrelevant or redundant action triplets, which can reduce computational costs and improve accuracy for subsequent compliance assessment. Eventually, the filtered triplets are transferred to graph representations and assembled as a project action graph by the modelling engine.

6.2.1 Factual triplet extraction

During the extraction process, the record documents of the project are inputted into the system as the data source. Then a pre-processing is conducted with the help of some Python libraries (e.g., PyPDF2 and TensorFlow), where the textual content in the documents is extracted and split as a bunch of separate sentences.

After pre-processing, the raw sentences are directly sent to the triplet extractor. Similar to regulatory knowledge extraction, these sentences with descriptions of actual behaviours are parsed based on dependency grammar and part of speech in the extractor and each word in the sentence is tagged with DP and POS labels. Following this, the sentence classification algorithm developed based on the multi-label joint mapping method classifies the raw sentences into coordinate sentences, compound sentences and simple sentences. The coordinate sentences and compound sentences are further processed by a sentence simplification algorithm that can convert one coordinate sentence or compound sentence into several separate simple sentences. Following the process of sentence simplification, all simple sentences are reparsed by dependency grammar and part of speech, and the DP and POS labels of each word in the new sentence are updated. On this point, these simple sentences with updated semantic labels are ready for factual constituent extraction.

The specific extraction process for factual constituents is identical to rule constituent extraction, which includes predicate extraction, subject extraction, object extraction, adverbial extraction, and attributes extraction. The extraction principles for the above constituent are also the same as the rules explained in Section 5.1.3.

6.2.2 Triplet cleansing

For project documents, it is difficult to distinguish between target sentences that describe the actual actions and redundant sentences not relevant to BIM compliance assessment through textual representation. Inspired by the statistical method used to extract indicators, a term frequency-based triplet cleansing method is proposed in the framework to automatically filter out irrelevant tuples in project documents for BIM compliance assessment. Since the regulatory knowledge graph represents all the requirements for BIM compliance as a series of triplets, the tuple cleansing method references the regulatory knowledge graph in filtering the target triplets.

The specific process of triplet cleansing is divided into two steps. The first step is constituent collection. In this step, all nodes and relations in the triplets that make up the regulatory knowledge graph are sorted into two separate lists, which record the nodes and relations of the regulatory graph respectively. The same process is applied to all the triplets extracted from project documents and two separate lists for nodes and relations in project documents are generated as well. After counting the nodes and relations in the two sets of triplets, the second step, triplet screening, is implemented. Considering the differences in the naming of entities and relations between the actual project document and the regulation document, filtering triplets by their subject, predicate, and object all being mentioned in the node and relation lists of the regulatory graph is not appropriate. To ensure sufficient relevance, too few shared nodes or relations are also undesirable. After several rounds of testing, the result of filtering by triplets containing at least one identical constituent (subject, relation, or object) in the diagram is the most promising.

Based on the above analysis and testing, a triplet cleansing algorithm is developed. The algorithm first sorts the nodes and relations of both regulatory triplets and factual triplets generated previously. Then the nodes and relations in factual triplet are checked against that of the regulatory graph. If the factual triplet shares at least one identical constituent with the regulatory graph, this triplet will be kept. Otherwise, this triplet is filtered out.

6.2.3 Action graph modelling

After triplet cleansing, factual triplets not relevant to the compliance assessment have been filtered out. The remaining triplets constitute a factual action graph for BIM compliance checking. To assemble these triplets into a graph representation, the graph modelling engine designed for the regulatory graph is reused in this part. The raw triplets are firstly converted into simple triplets and quaternions through the tuple transformation algorithm. Then, the constituents in the factual triplets are automatically assigned different URIs according to the predefined namespaces and comprise the project action graph by establishing connections with each other based on their unique IRI.

6.3 Autonomous compliance assessment

As introduced in the summary of the literature review, parsing complex logical relationships in regulation documents and mapping abstract concepts in regulations with dynamic entities in actual projects are the challenges in the automation of BIM compliance checking. The first challenge has been resolved by the proposed approach for regulatory knowledge extraction and project document information extraction. This section aims to introduce an innovative method to achieve dynamic entity mapping for automatic compliance assessment.

Through the previous work, a regulatory knowledge graph containing all the requirements for BIM compliance checking and a project action graph consisting of all the actual behaviours related to compliance checking have been generated. The next step is to develop a method, through which the corresponding entities and relations in these two graphs can be automatically mapped. For actual projects, some entities in the record documents are various across projects, for example appointing party, appointed party, task teams, and so on. Furthermore, there are also some differences in the naming of concepts or documents due to the different naming conventions applied, for example, exchange information requirement and exchange requirement. Although the naming of these entities differs from the concepts mentioned in the standard, they implemented the same functions required in the standard in the actual project. Therefore, this situation should also be treated as compliant in the assessment process of BIM compliance. According to the above analysis, it can be summarized that the dynamic mapping of entities in the regulatory knowledge graph and the project action graph cannot fully depend on the semantic information of the entities. The mapping should be established based on the function that the entities embody in the project. This feature of functionality can be captured through the connections between the node (representing the entity in the graph) to other nodes, which are also usually referred to as the structural features of nodes in a graph. Through the review of related research (Section 2.4.2), graph learning approaches are the most promising in capturing the structural features of nodes. Therefore, a graph learning-based approach has been proposed in this framework to capture the structural features of nodes in the graphs to achieve dynamic entity mapping.

The whole process of graph learning-based automatic compliance checking is shown in Figure 51, which can be broadly divided into three main steps, namely semantic alignment, graph embedding, and compliance assessment. The regulatory knowledge graph and the project action graph automatically generated by the previous steps are directly utilised as the input data for this part. External pre-trained word vectors (GloVe) (Pennington et al. 2014) are used to generate embedding for each node (details can be found in Section 6.3.1). Subsequently, the semantic node alignment is implemented by calculating the cosine similarity of these node embeddings. Based on the semantic alignment result and node embeddings, the initial embeddings for graph learning are generated, which are further updated during the learning process to capture the structural feature of each node. Finally, the structural alignment can align nodes with similar connectivity in the graph and the compliance score can be checked by calculating the portion of aligned triplets.

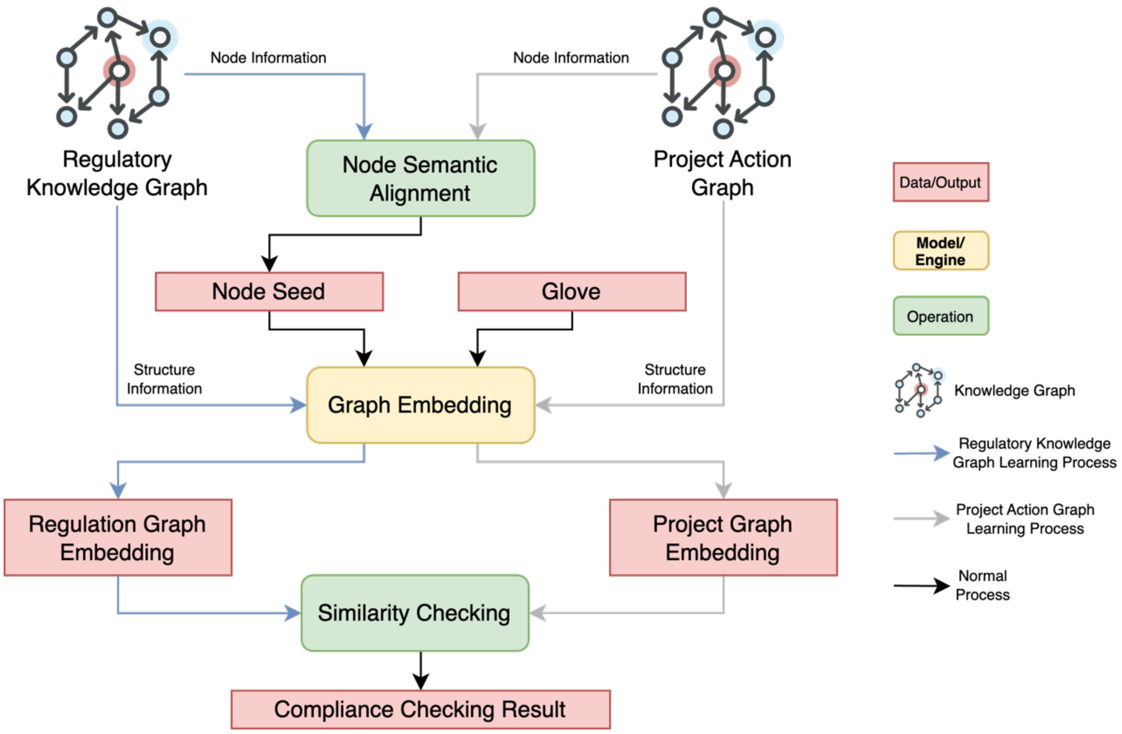


Figure 51 - The overall process of graph learning-based automatic compliance checking

6.3.1 Semantic alignment

Semantic mapping is the first step in automatic compliance assessment, which is also recognised as a preliminary alignment. The semantic alignment is designed to partially address the second case mentioned in Section 5.3, where some functionally identical nodes are described with slight differences in their names, such as information security and data security, or information requirement and information requirements.

The process of semantic mapping consists of the following three steps:

1. **Term pre-processing**

   In term pre-processing, the name of each node is extracted as an individual term. Then, these terms are processed by stemming and lemmatisation techniques to be converted to normalized term expressions.

2. **Semantic similarity calculation**

In this part, a word-level tokenizer from TensorFlow is adopted to split the normalized terms into separate tokens. Then an open-source word vector dictionary (GloVe) (Pennington et al. 2014) is utilised to embed each token with a 100-dimensional dense vector. The embedding of the term is determined by the mean value of its token embeddings. After acquiring all the term embeddings, the semantic similarity between each term is calculated based on the cosine similarity of their embeddings through the following formula:

$$S_c = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{10}$$

where A and B are two n-dimensional vectors, and A_i and B_i are the i th components of vectors A and B, respectively.

3. Node alignment

Based on the results of semantic similarity, the nodes can be aligned as an identical entity if their similarity reaches the threshold. After several rounds of experiments, the threshold of 0.85 was tested to be optimal. All the nodes in the two graphs whose semantic similarity is higher than 0.85 are automatically aligned and the nodes from the project action graph are replaced by their counterparts in the regulatory knowledge graph. The updated graphs are then sent to the next stage to generate graph embeddings.

In addition to the nodes in the graphs, the semantic alignment is also conducted to relations in the triplets.

Through the initial alignment of semantic mappings, nodes and relations with high semantic similarity are directly linked, which can remove some noise in the raw data. In addition, more accurate node correlations and more aligned nodes can improve the accuracy of node structure feature representation.

6.3.2 Graph embedding

Through the semantic alignment, the correspondences between the nodes in the generated regulatory knowledge graph and project action graph have been identified.

These corresponding nodes are used as seeds to learn the structural representation of each node in the graphs.

So far, there have been some graph learning methods proposed by different research to capture node features from static graphs, such as DeepWalk (Perozzi et al. 2014), node2vec (Grover and Leskovec 2016), GCN (Kipf and Welling 2016), GAT (Veličković et al. 2017), SDNE (Wang et al. 2016), DNGR (Cao et al. 2016), and so on. Some of these methods (such as DeepWalk and node2vec) are only applicable to large knowledge graphs as they can generate broader representations but with relatively low accuracy. Some other methods (including GAT, SDNE, and DNGR), have the ability to generate more accurate node feature representations but require more complex model design and more computational cost. Given the above considerations, the GCN model is the optimal choice at this stage. Hence, a GCN-based graph embedding generation method is developed in this study to obtain the structural features of nodes in a graph.

The GCN model is short for graph convolutional network model. As the name implies, the GCN models obtain a vector representation of each node by performing a convolution-like operation on the nodes in the graph. The GCN model is a deep neural network consisting of multiple graph convolutional layers. In GCN models, each convolutional layer only processes first-order neighbourhood information. The information transfer in multi-order neighbourhoods is achieved through stacking several convolutional layers. This mechanism makes the aggregated embedding of local neighbourhoods scalable and allows for multiple iterations of learning to capture the features of a node in the global neighbourhood.

Figure 52 illustrates the GCN-based graph embedding engine developed in the study, which consists of three types of layers, namely initial embedding layer, convolutional layer and output layer. As shown in Figure 52, the initial embedding layer takes in the node information and the word vectors to generate the initial embedding for each node in the graph. The convolutional layer focuses on the calculation of convolution and is made up of a GCN convolutional layer and an activation function layer. The output layer normalises all the embeddings computed through multiple rounds of convolution to obtain the final graph representation for each node.

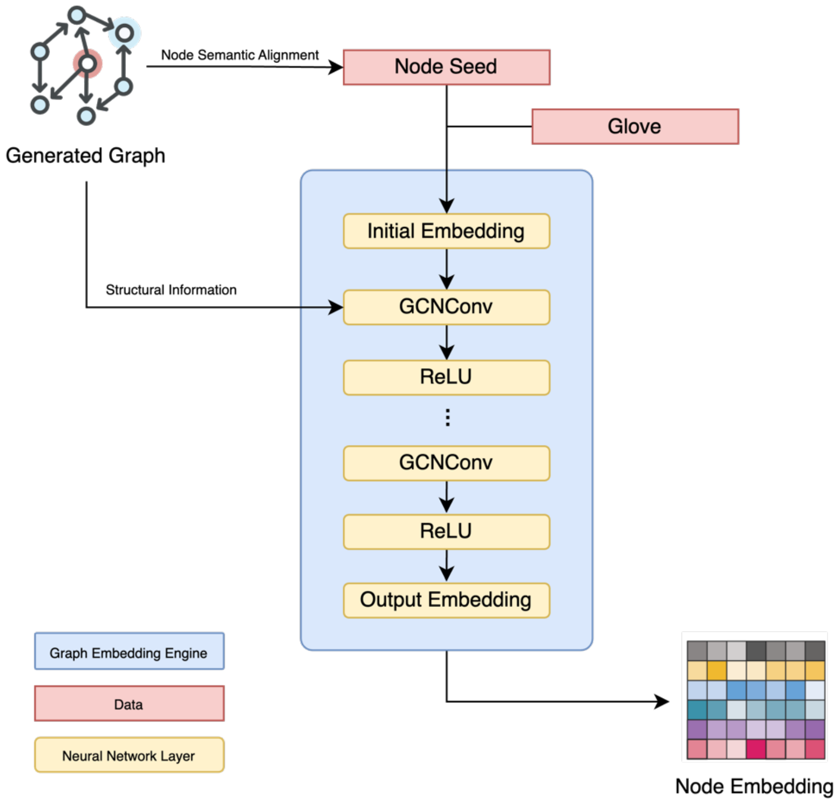


Figure 52 - Overall framework of graph embedding generation

During the process of graph embedding generation, the node information processed by semantic alignment and the GloVe word vector dictionary are fed into the initial embedding layer. For all the seed nodes, their initial value of graph embedding is the term embedding. To remove the influences of semantic information, the initial graph embeddings of all the unique nodes in each graph are set to zero vectors. After setting the initial value, the graph embeddings are sent to the convolutional layer to be updated based on the structural information of the nodes. The specific updating process of graph embedding in the convolutional layer can be divided into the following two steps:

1. **Sample neighbourhood**

During the calculation of convolution, the feature representation of a node is affected by its k-hop surrounding nodes. The terminology of “hop” is derived from the field of wired computer networks, the count of which aims to measure the distance between the source host and destination host (Kurose 1992). Specifically, the k-hop neighbourhood is the set of vertices that are reachable in k-hops or less from the source node (Xu 2020). Figure 53 shows an example of a 3-hop neighbourhood, where v_j (destination node) is one of the 3-hop neighbours of v_i (source node).

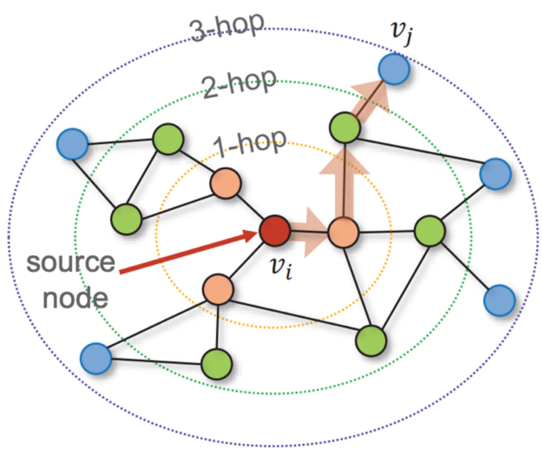


Figure 53 - Multi-hop neighbourhood sampling

In terms of the method proposed in this study, each of the GCN-based convolution layers only considers one-hop neighbourhood and an adjacency matrix that stores the one-hop connections of each node is generated for the convolution calculation in the next step. Due to the limitation of the GCN model, only the node information (calculated by word embedding) is involved during the learning process. The properties of edges are ignored. In addition, the GCN model is only capable of static graph learning, which means if the node or its connectivity changes, the graph embedding of each node needs to be relearned.

2. Information aggregation

In this step, the adjacency matrix A is first enforced on self-connections by adding the identity matrix I, which allows the previous information to be retained when updating the node's graph embedding. The enforcement of self-connection is implemented through the following formula:

$$\tilde{A} = A + I \tag{11}$$

where A is the adjacency matrix of the graph. I is the identity matrix. $\tilde{A}$ is the enforced adjacency matrix.

Based on the enforced adjacency matrix $\tilde{A}$, the degree matrix of the graph $\widetilde{D}_{ij}$ can be calculated through formula 12:

$$\widetilde{D}_{ij} = \sum\nolimits_j \tilde{A}_{ij} \tag{12}$$

where i, j are the node's index.

To solve gradient exploding or vanishing, a symmetrically normalized Laplacian is utilised in the convolutional layer, which is expressed as formula 13:

$$L_{norm} = \widetilde{D}^{-\frac{1}{2}} \tilde{A} \widetilde{D}^{-\frac{1}{2}} \tag{13}$$

Based on the above, the updated graph embeddings in the GCN layer can be calculated according to the following formula:

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{-\frac{1}{2}} \tilde{A} \widetilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}\right) \tag{14}$$

where $H^{(l)}$ and $W^{(l)}$ represent the feature matrix and the trainable weight matrix in the lth convolutional layer respectively. From a node-wise perspective, the update embedding can be written as formula 15.

$$h_i^{(l)} = \sigma\left(\sum_{i \in N_j} c_{ij} W h_j\right) \tag{15}$$

where $c_{ij} = \frac{1}{\sqrt{|N_i||N_j|}}$, and N_i and N_j are the sizes of the node's neighbourhoods.

Figure 54 illustrates the process of information aggregation, where x_2 is the embedding of the target node and x_1, x_3, x_4, and x_5 are embeddings of its one-hop neighbour nodes.

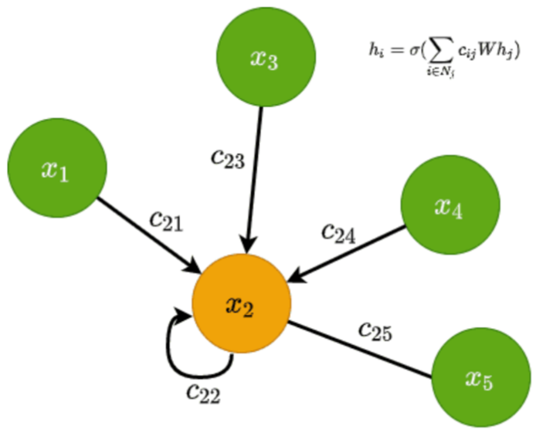


Figure 54 - Process of information aggregation from a node-wise perspective

In the proposed framework, the graph embedding engine is comprised of three convolutional layers, which means the updated node embedding is determined by its three-hop neighbourhoods. The embedding of each node in the graph is updated after one round of convolutional computation. After three rounds of convolution, the node embeddings capture adequate structural features for each node in the graph.

The last layer in the graph embedding engine is the output layer. In this layer, the updated node graph embeddings are passed to an activation function (Sigmoid), which aims to normalise the generated embeddings so that they can be directly applied to the compliance assessment. Figure 55 presents the normalisation process in the output layer.

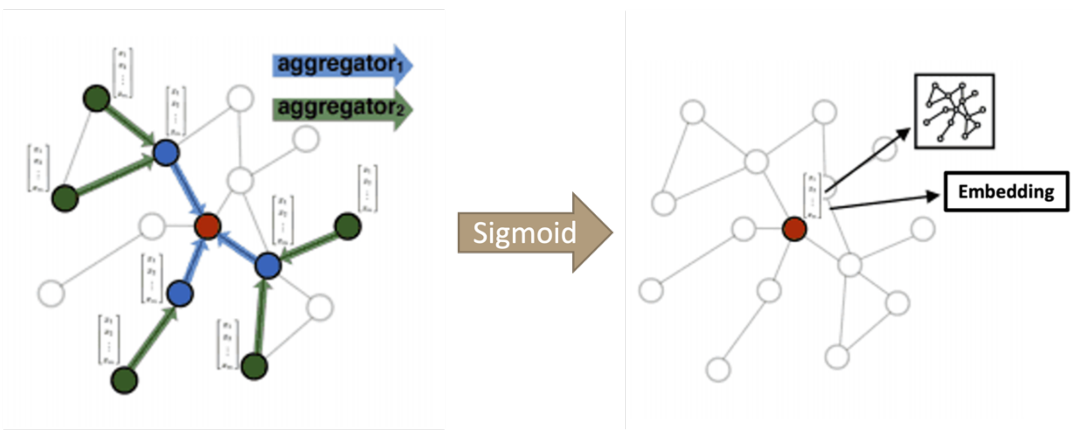


Figure 55 - Normalisation process in the output layer

6.3.3 Compliance assessment

Through the developed generator, two sets of embeddings that reflect the structural features of the nodes in the regulatory knowledge graph and project action graph are generated respectively. These graph embeddings are determined by the connectivity of nodes to other nodes in the graph, which is irrelevant to the name of the node. Thus, for nodes that perform the same function in the project, their graph representation is approximately invariant, even though their names change with the projects and some of the names also differ from the terminology defined in the standard. Based on the above analysis, the mapping of dynamic entities can be achieved based on the embedding of nodes. The final assessment of BIM compliance can be divided into two stages, which are structural alignment and triplet mapping.

6.3.3.1 Structural alignment

The structural alignment is designed to address the challenge of dynamic entity mapping. In this part, the similarity between the unseeded nodes in the project action graph and all the nodes in the regulatory knowledge graph is calculated one by one through the cosine similarity of their graph embedding. After calculation, the unseeded nodes in the project action graph are replaced by the nodes in the regulatory knowledge graph with the highest similarity to them.

6.3.3.2 Triplet mapping

Through semantic alignment and structural alignment, the potential connectivity of the nodes in the two graphs has been fully explored. The final step is to conduct a comparative analysis of the triplets in the two graphs to evaluate the consistency of the project with respect to the standard.

During the comparative analysis, each triplet in the regulatory knowledge graph is retrieved and mapped with triplets in the project action graph. If a triplet with identical subject, predicate, and object can be found in the project action graph, the triplet of the regulatory knowledge graph is considered to be matched. The final assessment result is calculated through the following formula:

$$Compliance\ Score = \frac{Number\ of\ matched\ triplets}{Total\ number\ of\ triplets\ in\ the\ graph} \tag{16}$$

More details on the compliance assessment can be found in Section 7.4, which explains the compliance assessment process by using an actual project and BIM standard as an example.

Chapter 7. Validation and testing

To improve the comprehensiveness and automation of the existing BIM compliance checking approach, this research proposed a comprehensive compliance checking framework with a subjective solution and an objective solution. The developed two solutions consist of several functionalities, including an ontological knowledge model, clause extractor, and triplet extractor. After completing the development work, the intended frameworks should be validated against specific theories and methodologies. In this research, validation work is divided into two parts, namely functionality validation and practical testing. Functionality validation focuses on the validation of functionalities developed in the frameworks, where the criteria are determined by their functions. In practical testing, the information of a real project is utilised to validate the two proposed frameworks.

Regarding the structure of this section, the first three sections constitute the functionality validation, where the ontological knowledge model, clause extractor and triplet extractor are verified separately. The fourth section concentrates on practical testing, which assesses the effectiveness of the whole framework in practical scenarios.

7.1 Compliance indicator system

The validation of the ontological knowledge model primarily lies in the elements (class, properties, and instances) within it. For this study, the high-level classes (Documents, Stakeholders, Standard and BIM compliance indicators) and object properties are designed for different checking scenarios. The assessment of BIM compliance mainly relies on the BIM compliance indicator system. Hence, this indicator system should be verified to ensure its reliability.

The compliance indicators in the system were initially derived from standard documents by statistical algorithm. Then these indicators are screened and collated into a hierarchical indicator system by the group work of domain experts. Since the indicator system was built based on the consensus of the calibration expert panel, a larger scope of the Delphi survey was implemented to validate both the indicators and the hierarchical structure of the indicator system. 15 domain experts from both academia and industry were involved in the validation work. This validation expert panel consists of seven academic experts and eight industry experts. Academic experts comprise three professors engaged in research related to BIM standards, two researchers involved in the development of BIM standards and two PhD students who have published articles related to BIM standards. The industry experts consist of four senior experts who have been involved in building design and construction for more than 20 years and four technical experts who have participated in the development of China's BIM standards.

The validation process followed the same procedure for validating sample labels of the domain dataset (Figure 41). Considering some experts may be difficult to physically encounter, an online survey was conducted using Google Forms. Figure 56 shows some screenshots of the questionnaire designed for indicator validation.

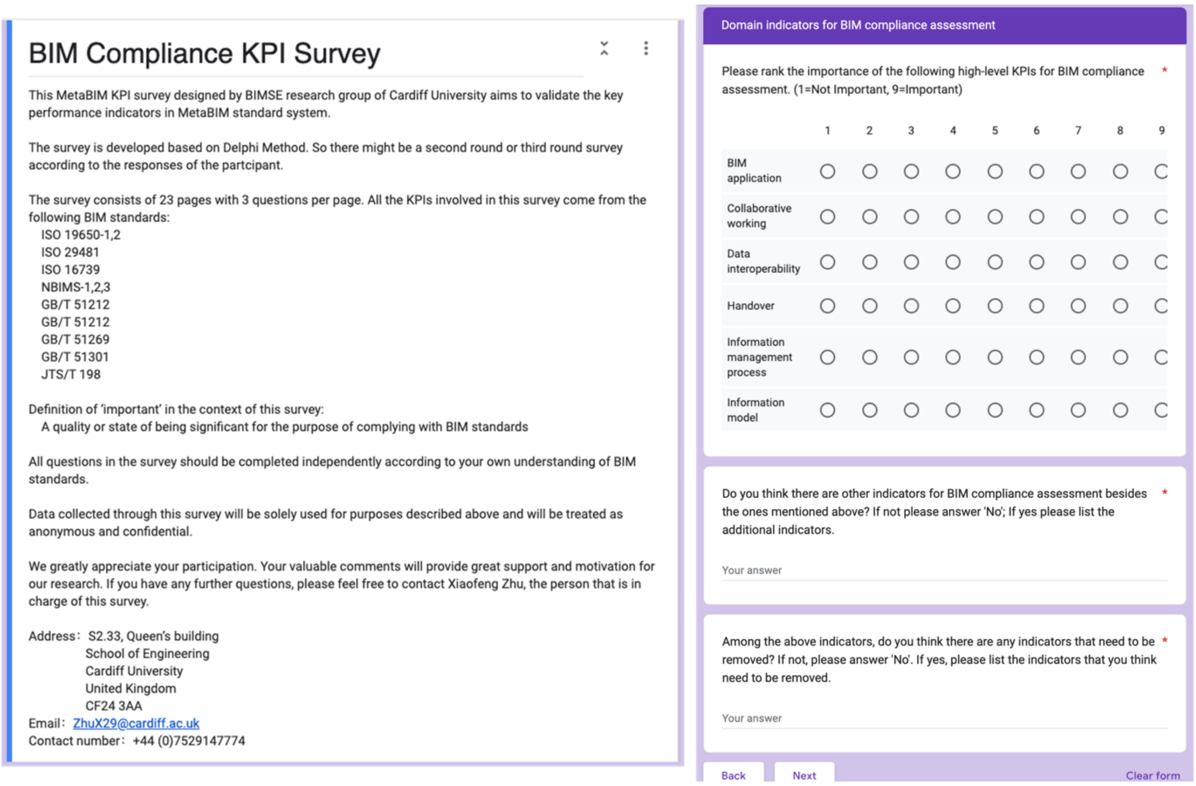


Figure 56 - Screenshot of part of the questionnaire designed for indicator validation

During the validation process, 20 questionnaires were initially distributed to different experts, of which 17 questionnaires were finally collected. Excluding the two incomplete questionnaires, the results of 15 questionnaires were utilised to validate the indicators. Among all the experts who submitted valid questionnaires, eight were from industry and seven from academia. This balanced composition makes the validation results more generalised.

As the proposed indicator system consists of approximately 500 indicators, it is unrealistic to present the validation process for all indicators in this thesis. Therefore, the six indicators in the domain layer were selected as an example to demonstrate the validation process. As shown in Figure 56, the questionnaire for domain indicator is comprised of three types of questions. The first one is used for the AHP method to calculate the weight of each indicator. The last two questions are used to verify the indicators.

Table 15 illustrates the answers from experts in the first round of the Delphi survey. Generally, there was a high degree of inconsistency in the experts' responses. Experts from the industry focus more on BIM applications, such as clash detection, cost estimation, planning, etc., while academic experts concern more on BIM principles like data format, data sharing and information operability. Based on this feedback, the domain indicators were modified, where data security and common data environment were removed and BIM application, data treatment and handover were supplemented. Apart from the domain indicators, some sub-indicators (e.g., clash detection, and project planning) have also been added to the system following the suggestions of experts. The questionnaire was revised accordingly for the second round of the survey.

Table 15 - Experts' answers to the designed questions in the first round of the Delphi survey

Domain indicators		
information model, collaborative working, information management process, data exchange, data security, common data environment.		
Experts' answers to questions		
No. of expert	Do you think there are other domain indicators for BIM compliance assessment besides the ones mentioned above?	Among the above KPIs, do you think there are any KPIs that need to be removed?
1	Information standard, BIM application requirements	data security
2	Yes, handover should be added.	No
3	No	common data environment
4	hand over	data security
5	Handover, data treatment	No
6	no	no
7	Operability of model information	No
8	Information standard	No
9	yes. clash detection, cost estimation, sustainability, etc	data security
10	clash detection, information standard	No
11	nD modelling	data security
12	Data format, classification and coding	No
13	Data sharing	data exchange and data security
14	BIM techniques/IoT (Any software requirements)	No
15	Project planning	common data environment

During the questionnaire distribution of the second-round survey, a brief explanation was included in the introduction of the questionnaire, which illustrated the modifications on indicators. In addition, it also clarified that the information standard mentioned by some experts in the first-round survey has already been defined as a high-level indicator in the knowledge model. The feedback of domain experts in the second round of the Delphi survey is shown in Table 16. Compared with the first round, the answers of experts were more consistent. Some experts were still insisting on their opinions in the first round (e.g., experts 5, 6, 14). Some experts objected to the changes made after the first-round survey (e.g., experts 3, 7, 10). In addition to this, it can be observed that some new ideas were triggered by the first round of modification, such as data schema, data storage, naming convention, etc. Based on the above observations, some modifications were implemented to satisfy the requirements of all experts as far as possible. A new indicator called data interoperability was introduced to integrate the two existing indicators (data exchange and data treatment) with the newly proposed ones (e.g., data storage, data schema, naming convention). These indicators were also added to the indicator system as the subclass of data interoperability. After the second modification, the questionnaire for the third round of survey was revised accordingly.

Table 16 - Experts' answers to the designed questions in the second round of the Delphi survey

Domain indicators		
information model, collaborative working, information management process, BIM application, data exchange, data treatment, handover.		
Experts' answers to questions		
No. of expert	Do you think there are other domain indicators for BIM compliance assessment besides the ones mentioned above?	Among the above KPIs, do you think there are any KPIs that need to be removed?
1	Data schema	No
2	Data storage	No
3	No	data treatment
4	Data exchange requirements	data exchange
5	Data treatment	No
6	no	no
7	Operability of model information	BIM application, data treatment
8	Naming convention	No
9	No	No
10	Common data environment	Data treatment
11	No	No
12	Data format, classification and coding	No
13	Data sharing, data storage	data exchange
14	BIM techniques/IoT (Any software requirements)	Handover
15	No	No

Table 17 presents the outcome of the third-round survey. The answers from different experts are highly consistent. It is acceptable that a small number of experts are not in full agreement, especially if some of the opinions are contradictory. At this point, a consensus of experts was considered to be reached on the definition of the domain indicators. The validation for domain indicators has been completed.

Table 17 - Experts' answers to the designed questions in the third round of the Delphi survey

Domain indicators		
information model, collaborative working, information management process, BIM application, data interoperability, handover.		
Experts' answers to questions		
No. of expert	Do you think there are other domain indicators for BIM compliance assessment besides the ones mentioned above?	Among the above KPIs, do you think there are any KPIs that need to be removed?
1	No	No
2	No	No
3	No	No
4	No	No
5	Data treatment	No
6	no	no
7	No	BIM application
8	No	No
9	No	No
10	Common data environment	No
11	No	No
12	No	No
13	No	No
14	No	Handover
15	No	No

Following a similar process, the entire indicator system was validated through 5 rounds of the Delphi survey. The whole ontological knowledge model and the weighting matrix were updated accordingly.

7.2 Clause extraction

In the developmental phase, the initial validation primarily concentrates on assessing the transfer learning training outcomes and evaluating the extraction performance of the constituent extractor.

Clause extraction aims to distinguish clauses from clause descriptions in regulatory documents. Therefore, it is essentially a deep learning-based binary classification. Given this consideration, the cross-entropy is selected as the loss function and the performance clause extraction is evaluated on the test set and measured by some commonly used indicators (accuracy, precision, recall, and F-measure), which are calculated using the following equations:

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \tag{17}$$

$$Precision\ (P) = \frac{TP}{FP + TP} \tag{18}$$

$$Recall\ (R) = \frac{TP}{FN + TP} \tag{19}$$

$$F1 = 2 * \frac{P * R}{P + R} \tag{20}$$

where TP, TN, FP and FN stand for the number of instances correctly identified as positive, correctly identified as negative, incorrectly identified as positive and incorrectly identified as negative, respectively.

Figure 57 illustrates the evolution of training loss across training epochs. The pre-trained BERT model attains its optimal performance after four epochs of fine-tuning, aligning with the parameter recommendations from Devlin's research (Devlin et al. 2018). The outcomes of the clause extraction are depicted in the confusion matrix (Figure 57). In adherence to the designated split ratio mentioned in Data packing of

Section 6.1.2.2, 400 samples are randomly selected from the domain dataset to establish the test set, encompassing 208 clauses and 192 description clauses. The results indicate a 94.4% precision and 98.1% recall in the clause extraction from the test set, yielding an F1 score of 0.96. These metrics collectively highlight the robust performance of the clause extraction process.

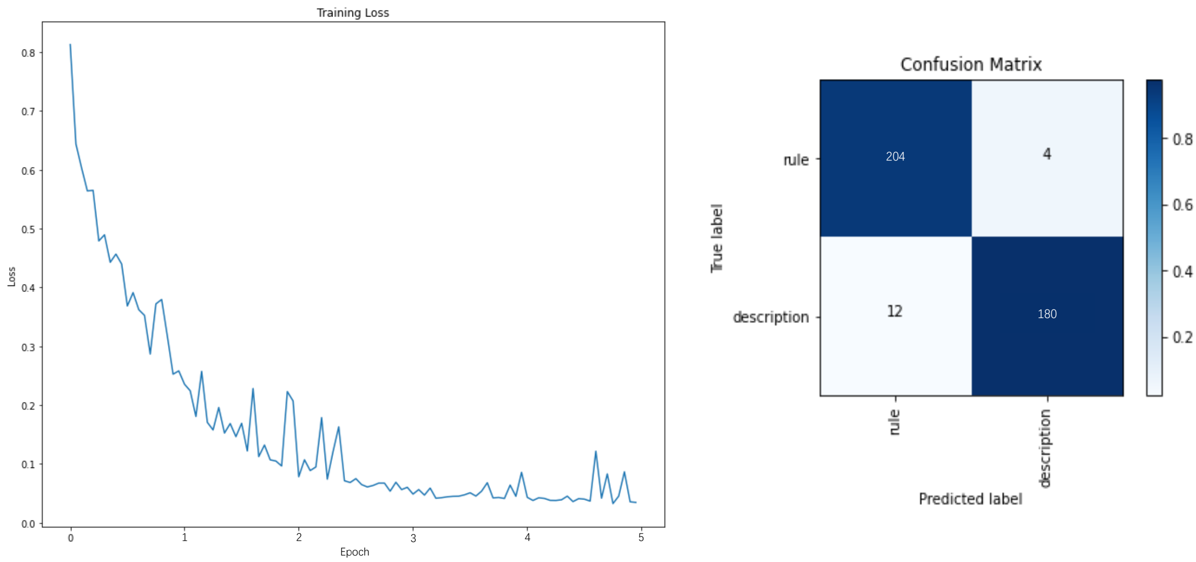


Figure 57 - The variation of training loss and confusion matrix of extraction result on the test set

To implement a thorough assessment of clause extraction performance, a comparative analysis is conducted between the proposed clause extractor and several state-of-the-art text classification models, including RNN, LSTM, and pre-trained BERT. The specifics of the extraction results are delineated in Table 20. Furthermore, Figure 58 provides insights into the distribution of classification results and depicts the Receiver Operating Characteristic (ROC) curves for each model.

Table 18 - Comparison of the clause extraction results between different deep learning models

Model	Accuracy	Precision	Recall	F1-value
RNN	0.798	0.828	0.752	0.789
LSTM	0.893	0.911	0.887	0.899
Bi-LSTM	0.932	0.947	0.919	0.933
BERT-pre	0.814	0.810	0.820	0.815
BERT-ft	0.960	0.944	0.981	0.962

RNN: recurrent neural network, LSTM: Long short-term memory, Bi-LSTM: bidirectional LSTM, BERT-pre: pre-trained BERT model, BERT-ft: fine-tuned BERT model

The outcomes presented in Table 18 underscore the superior performance of the proposed fine-tuned BERT extractor in clause extraction, as it achieves the highest accuracy. The corresponding Area Under the Curve (AUC) value, illustrated in Figure 58, further confirms the effectiveness of this extractor compared to state-of-the-art deep learning models. Furthermore, it is noteworthy that fine-tuning the pre-trained BERT model results in significant savings in training resources. In this experiment, the pre-trained BERT model achieves 96% accuracy on the test set after only 4 epochs of training, while the conventional models (RNN and LSTM) require approximately 30 epochs to reach their optimum performance.

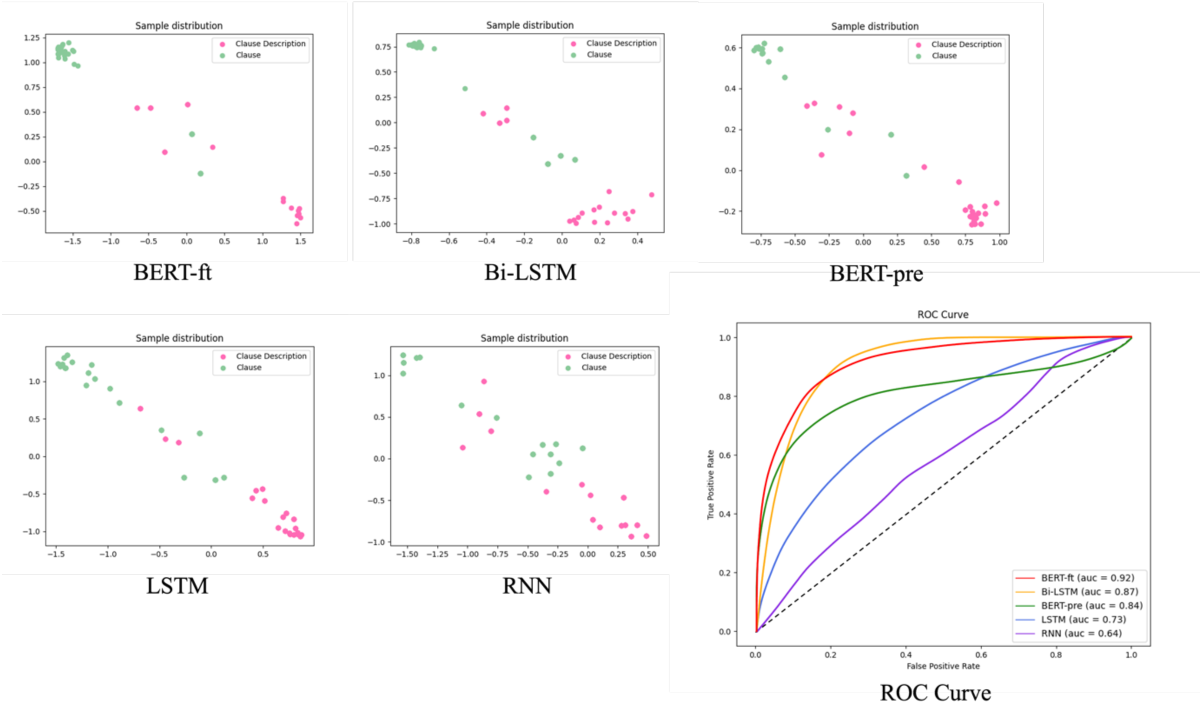


Figure 58 - Comparison of different models on the ROC curve and t-SNE visualisation of the classification result

7.3 Regulatory triplet extraction

Regarding constituent extraction, a comparative evaluation is conducted between the proposed extraction engine and two existing information extraction tools, namely OpenIE (Zhou et al. 2022a) and ClauseIE (Corro and Gemulla 2013). This assessment is performed on random samples of 50 clauses selected from the domain dataset, encompassing both complex and simple clauses. The total number of extractions for each method is summarized in Table 19, while Table 20 provides a detailed comparison of extractions by different information extraction tools on the same clauses.

Table 19 - Comparison of extraction numbers on selected samples

Clause samples	OpenIE	ClauseIE	Our extraction engine
50	72	123	176

In comparison with OpenIE and ClauseIE, the proposed extraction engine demonstrates similar information extraction capabilities but offers higher granularity. Existing information extraction approaches typically operate at the noun chunk level, capturing phrases such as "the lead appointed party" or "the information model." However, the proposed engine can delve deeper and extract attributes associated with the central noun, such as "project's "project's." This capability is particularly vital for quantitative clauses, where specific requirements are often embedded within noun chunks. For instance, in the clause "there shall be an approved alarm-initiating device at not more than 150-foot intervals" (Clause 415.5.2 of IBC 2015), "150-foot" represents the quantitative requirement for the intervals of the alarm-initiating device. Another notable advantage of the proposed engine is the construction of reifications (R1~R3), which enhances its ability to represent the conditions and constraints of required actions compared to individual triplets (C1). Based on the observed results, the proposed framework not only achieves high accuracy in clause extraction but also excels in the extraction and representation of constituents.

Table 20 - A comparison of the proposed framework and existing tools on information extraction for regulation clause

Clause: If the review is successful, the lead appointed party shall authorize the project's information model and instruct each task team to submit their information within the project's common data environment.
Triplets extracted by OpenIE:
O1: (the lead appointed party, shall authorize, information model) O2: (the lead appointed party, shall authorize, information model and instruct each task team to submit their information within project's common data environment)
Triplets extracted by ClausIE:
C1: (the review, is, successful) C2: (the lead appointed party, shall authorize, information model) C3: (the lead appointed party, instruct, each task team) C4: (the lead appointed party, submit, their information)
Triplets extracted by proposed framework:
T1: (lead appointed party, shall authorize, information model) T2: (information model, should be, project's) T3: (lead appointed party, shall instruct, task team) T4: (task team, submit, information) T5: (information, should be, their) T6: (common data environment, should be, project's)

R1: if the review is successful (lead appointed party, shall authorize, information model) R2: if the review is successful (lead appointed party, shall instruct, task team) R3: within the project's common data environment (task team, submit, information)

7.4 Actual project document testing

The previous validation work verifies the proposed ontological compliance checking framework and some functions of the automatic compliance checking framework from the perspective of framework development. To validate the proposed graph learning-based dynamic entity alignment mechanism and the whole automatic compliance checking framework, an actual construction project is taken as a use case to verify the performance of the proposed framework under practical application scenarios.

In this study, an actual construction project is taken as a use case to verify both the proposed ontological checking framework and the automated checking process, whose name is Hassyan Clean Coal Power Plant (*Hassyan Power Complex, Dubai, UAE* 2016). This finished project is located in Dubai and is a collaboration between several companies such as Hassyan Energy, CCCC, Acwa Power, etc. The project commenced in November 2016 and the first two units began operation in May May 2021. The whole set of the project documents is provided by one of the stakeholders, CCCC, for research on BIM compliance checking, which includes the project contract, project plan, project drawings and other related attachments. This project has also been audited by BSI for ISO 19560 kitemark certification. The feedback from BSI is utilised as the ground truth in the validation process.

The reason for choosing this project is that it has been submitted to the BSI for ISO 19650 certification. The professional team of the BSI has audited the project and listed all the gaps for ISO19650 certification, which can be used as the ground truth of BIM compliance checking results. Therefore, the documents of the Hassyan project and the ISO 19650 series of standards are used as validation materials for practical testing. A compliance check against ISO 19650 was implemented to verify the reliability of the proposed systems based on a comparison with the feedback given by the BSI.

Considering that the developed system can only recognise and process textual information, information such as pictures, tables and drawings in the project documents were manually removed during the document preparation phase. To prevent missing any important information, content related to BIM conformance checking in these removed diagrams and tables was manually converted to textual descriptions. After document preparation, the PDF files of ISO 19650 and the project document are sent into the system separately.

For ISO 19650, the second parts of the standard were fed into the pipeline of regulatory knowledge extraction (Figure 5959).

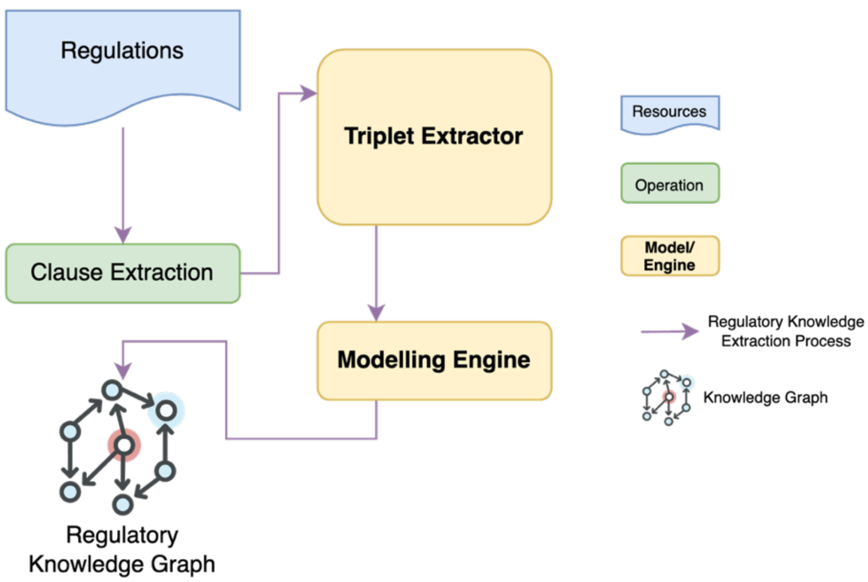


Figure 59 - The pipeline of regulatory knowledge extraction in the proposed framework

The selected standard documents were first converted into a pure textual format that can be processed by the system directly. The contents of them were split as individual sentences, which were then scanned by the fine-tuned BERT-based clause extraction model. Through clause extraction, 181 raw clauses are extracted from the standard documents. Then these clauses are processed by the algorithm of sentence classification and sentence simplification. A total number of 269 simple clauses are generated by the triplet extractor, which covers all the requirements defined in the standards. The generated simple clauses were then fed into the constituent extraction engine in the triplet extractor, which can transfer textual sentences into triplet and reifications. After triplet conversion, the constituent extraction engine converted 269 simple clauses into 578 triplets and reifications. Up to this point, the requirements in the standard document are all transformed into a structured knowledge representation with the format of triplets.

Based on the triplets generated from the triplet extractor, the modelling engine built a regulatory knowledge graph, which is shown in Figure 60.

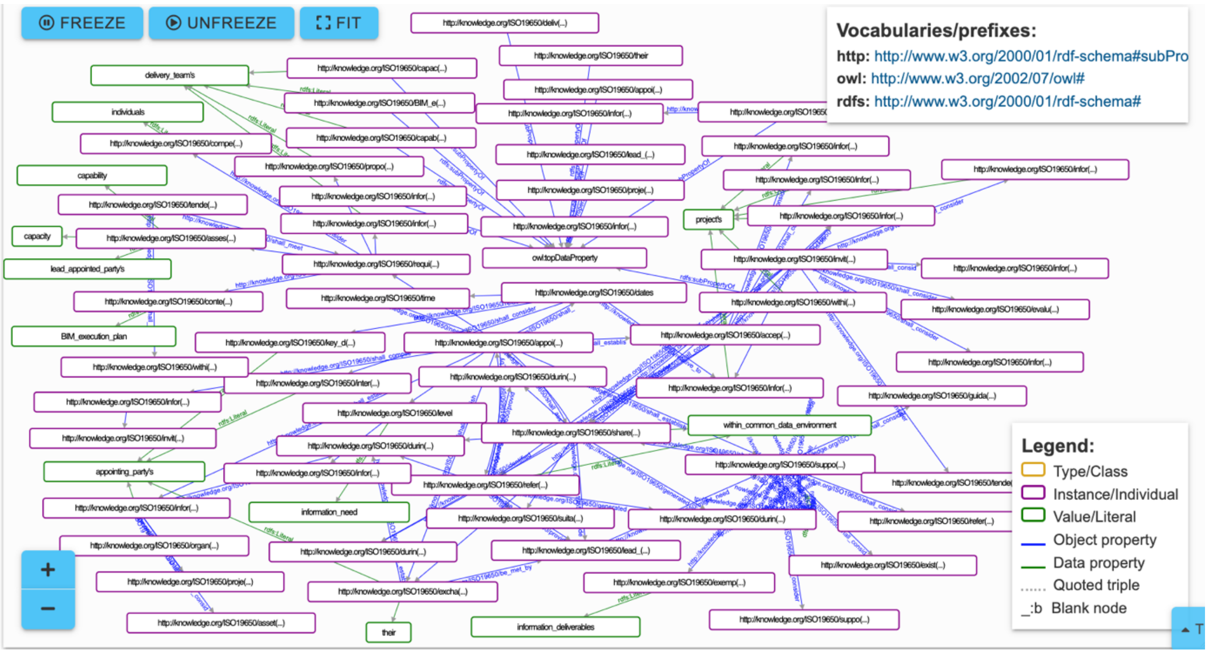


Figure 60 - Visualisation of part of the RDF graph generated by the proposed framework

To validate the generated regulatory knowledge graph, the nodes and edges in the graph are compared with the corresponding classes and properties in the ontological knowledge model. Since the indicators in the ontology have been validated through the Delphi survey, the results of alignment between the regulatory knowledge graph and expert ontology can evaluate the performance of regulatory knowledge extraction and modelling. The metrics of the ontological knowledge model constructed by experts from the calibration panel and the regulatory knowledge graph generated by the proposed framework are shown in Table 21.

Table 21 – Element metrics of the generated regulatory knowledge graph and expert ontology

	Regulatory graph	**Expert ontology**
Axiom	877	889
Logical axiom	644	449
Declaration axiom	233	354
Class	1	252
Object property	101	86
Individual	233	16
Annotation assertion	0	268

The validation process is segmented into two key components: 1) element checking - This phase involves verifying whether the required instances and relations are adequately defined; and 2) connectivity checking - In this step, the focus is on ensuring that the instances are correctly connected by the designated relations. The accuracy of the checking process is gauged using Intersection-over-Union (IoU), which is calculated through the following equation:

$$IoU = \frac{Number\ of\ Overlap}{Number\ of\ Union} \tag{21}$$

As delineated in Table 21, the proposed framework autonomously extracted 233 instances and 101 object properties, juxtaposed with the experts' stipulation of 268 instances and 86 object properties. Figure 61 delineates the correlation between these two sets of object properties. The RDF graph manifests a convergence of 78 same properties with the expert ontology. If the built-in object property "should_be" (used to connect instances and their attributes) is excluded, the accuracy of object property mining attains 72.2%. In the realm of entities, the RDF graph and expert ontology exhibit a shared presence of 189 instances, inclusive of both classes and individuals, thus culminating in an accuracy rate of 60.6%. It is important to note that some instances, though expressed in varying terms, convey similar meanings and correspond to identical entities within the regulatory documents (e.g., "existing asset information" and "asset information"). Upon considering these instances, the count of mutually shared instances increases to 211, indicating a refined accuracy of 72.8% in the context of instance extraction.

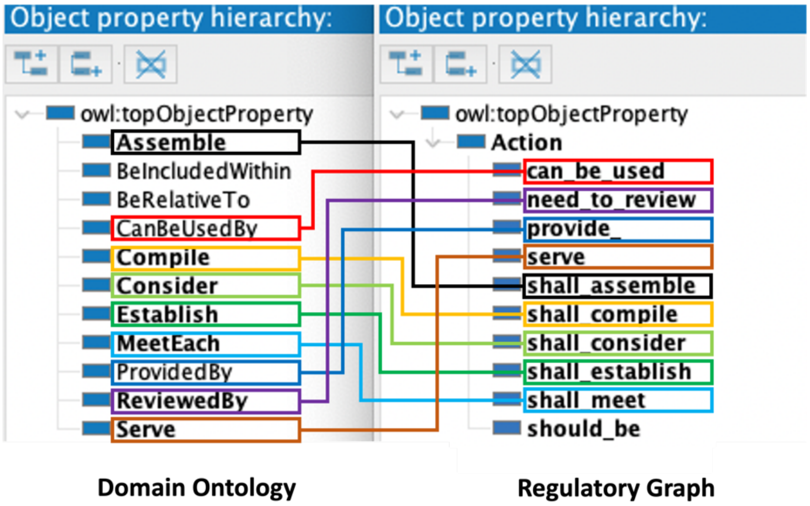
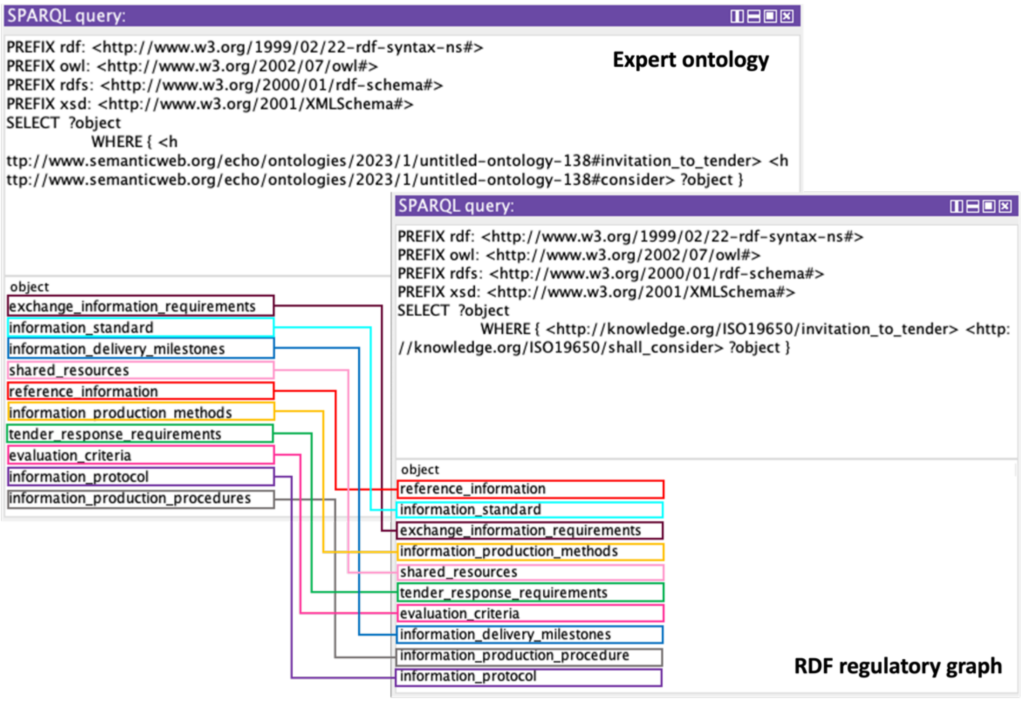
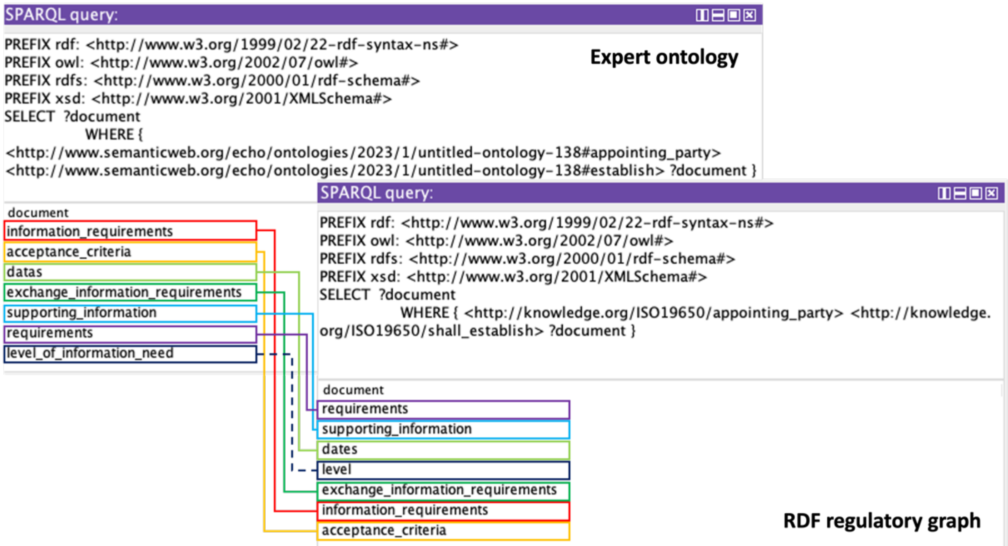


Figure 61 – Some examples of aligned object properties in the generated regulatory graph and the domain ontology

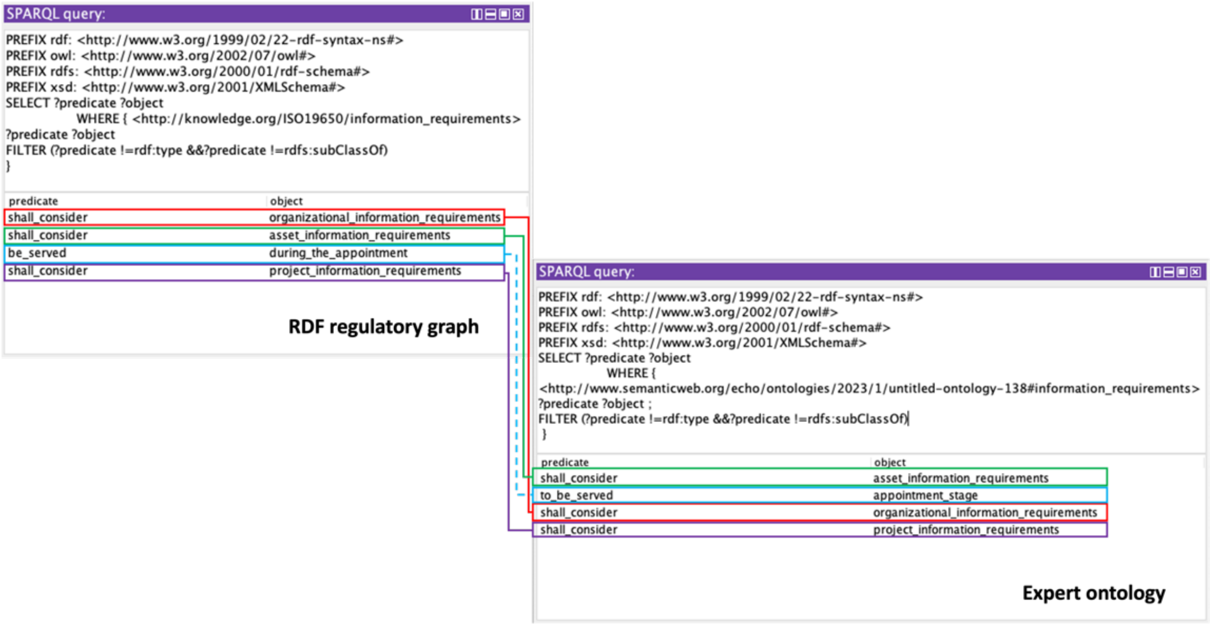
In terms of connectivity between instances and properties, it is challenging to be directly validated. Therefore, a sequence of SPARQL queries is employed to ascertain the connectivity of graph elements by juxtaposing the query outcomes of the knowledge graph and domain ontology. These queries are meticulously formulated by experts from calibration, drawing upon the content delineated in ISO 19650, and encompass the entirety of regulatory knowledge expounded upon in this standard. The assessment of graph element connectivity is conducted by contrasting the query results derived from the generated graph with those from the expert ontology. The mapping outcomes of three illustrative examples from the entirety of queries are depicted in Figure 62.



a) Mapping results for querying what document shall be considered in the invitation to tender stage.



b) Mapping results for querying what document the appointing party shall establish in the invitation to tender stage.



c) Mapping results for querying all regulatory knowledge related to information requirements.

Figure 62 - The mapping results of three query examples between the generated RDF graph and expert ontology

As demonstrated in Figure 62, queries involving aligned relationships and instances (e.g., “shall establish” and “establish”) yield essentially identical results. While there may be variations in the result list in terms of wording of some instances, they refer to the same instance (e.g., “level” and “level of information need”). However, queries incorporating unique instances or properties (e.g., might need, undertaking, etc.) in the RDF graph show no results in the expert ontology, and vice versa. This is understandable since the existence of an RDF triple is contingent upon the explicit definition of both the instances and the property forming that triple. Based on this analysis, the connectivity among the graph elements (instances and properties) in the generated RDF graph is complete and correct. Discrepancies in query results predominantly arise from differences in instances and properties. Considering that the proposed methodology attains an accuracy of 72.2% and 72.8% in the automated extraction of instances and properties, respectively, it is judicious to infer that the proposed framework achieves an overall accuracy of approximately 72.5% in the automated extraction and modelling of regulatory knowledge, in comparison to the manual approach undertaken by domain experts.

For project document information extraction, several project records documents, including the project contract, project plan, drawings, time schedule, etc., are pre-processed and merged into a master document, which provides a comprehensive description of the actions performed by the various stakeholders in the delivery of the project. After processing by the triplet extractor, 5803 action triplets were extracted from the master document. These raw action triplets were then cleansed by the proposed algorithm and 863 of them were eventually retained. The cleansed triplets were assembled into the project action graph by the modelling engine, which contains all the actual behaviours each stakeholder performed during the project delivery. Figure 63 visualises part of the project action graph.

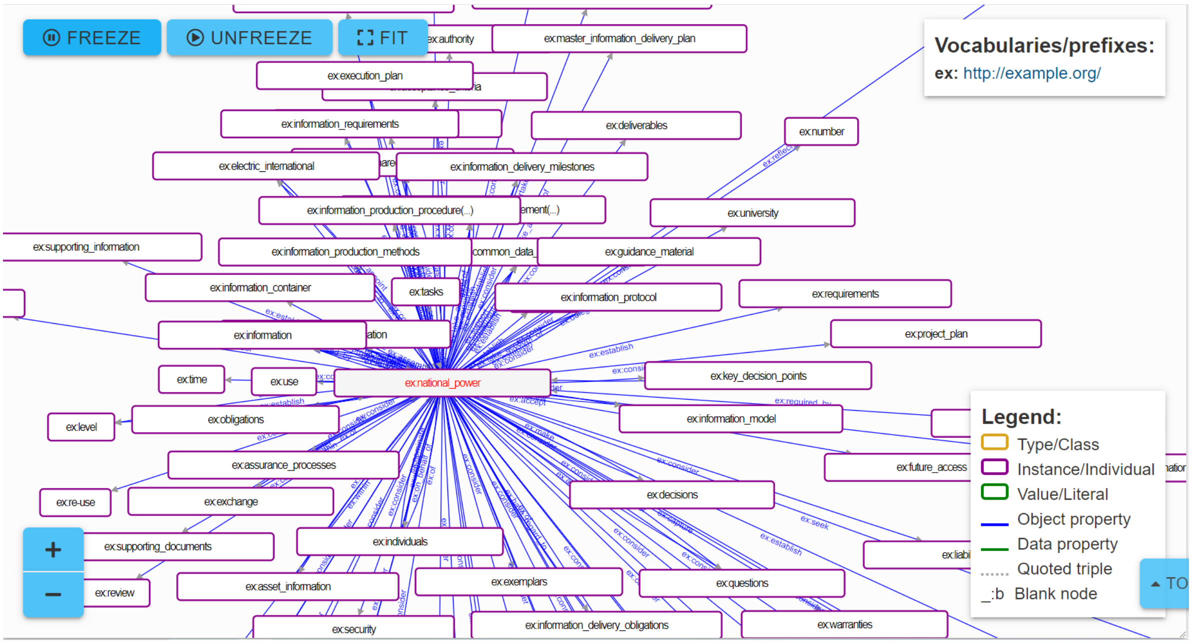


Figure 63 - A visualisation of part of the project action graph related to the National Power

After acquiring both the regulation graph and action graph, the proposed pipeline for autonomous compliance assessment was applied to these two graphs. Through the semantic alignment, the regulation graph and action graph share 182 nodes and 98 edges. The linkage and graphs are then populated into the graph embedding engine to obtain the embeddings of all the nodes, which represent the structural features of the nodes in the graph with a 100-dimensional dense vector. Based on these embeddings, nodes with similar structural features in the two graphs are aligned automatically. After checking the result of alignment, the proposed graph learning-based method has addressed the challenge of dynamic entity mapping. For example, the appointing party, lead appointed party, appointed party and task team are the abstract concepts defined in the standard. In a real project, these roles are taken by different businesses and these companies will change with the project. Hence, these roles of stakeholders are typical dynamic entities. Through the structural alignment, the node of the appointing party in the regulatory graph is aligned with the node of the ACWA Power in the action graph. Similarly, the lead appointed party is aligned with the Harbin Electric International Company and the appointed party is aligned with the CTESI. For the task team, several entities are taking this role in the different phases, such as the CCCC construction company and the NEPDI. Through comparing the embedding similarity, the node of the task team is aligned with the NEPDI, which has a higher similarity than the CCCC construction company. The above alignments fully comply with the roles assigned to each company in the contract.

After mapping the triplets of the two graphs, 151 triplets in the regulatory graph do not have a counterpart in the action graph, indicating the compliance score is about 0.74. To further investigate the actual performance of the proposed framework, a comparison analysis of the generated result and the feedback from the BSI is conducted, which is shown in Table 22.

Table 22 - Comparison of the generated missing triplet and gaps listed by the experts from the BSI

No.	Gaps outlined by experts from the BSI	Corresponding missing triplet
1	The lead appointed party has not developed an overall method of describing the entire process, meaning a repeatable system for use on all BIM projects was not demonstrated (All clause requirements).	Not found
2	As the lead appointed party delivers EPC projects, it intends to establish a standard PIR for the project. (5.1.2).	(lead appointed party, establish, project information requirement)
3	The appointing party intends to introduce a process to host, support and / or manage a CDE for the client if requested to do so (5.1.7).	Not found
4	The appointing party intends to develop its EIR document to ensure all requirements of ISO 19650- 2 are included. (5.2.1).	(appointing party, establish, exchange information requirement)
5	The lead appointed party intends to create an EIR checklist for use when receiving and confirming an appointing parties EIR. (5.2.1)	(lead appointed party, generate, exchange information requirement checklist)
6	The lead appointed party intends to merge its Pre and Post Appointment BEP templates to produce one Project BEP and intends to review this fully to remove any "PAS 1192-2" references. (5.3.2 and 5.4.1).	Not found
7	The lead appointed party intends to enhance its mobilisation plan and delivery activities to better describe how it will ensure these activities are completed. (5.3.5 and all of 5.5.X).	Not found
8	The lead appointed party does not create a "Lead Appointed Party EIR" to identify these requirements in its Post Contract BEP. (5.4.3)	(lead appointed party, establish, exchange information requirement)
9	The lead appointed party intends to establish a TIDP template. (5.4.4)	(lead appointed party, establish, task information delivery plan)
10	The appointing party intends to develop a CDE, and create a process guide to describe how the CDE functions in accordance with the UK national annex and how information is contained, approved and authorised between task and delivery teams. (5.6.X, 5.7.X and UK National Annex).	(appointing party, establish, common data environment)
11	The lead appointed party intends to create a process to ensure lessons learnt from the appointing party and task teams is effectively controlled. (5.8.2)	(lead appointed party, consider, lessons learnt process)
12	The appointing party intends to identify its file naming and protocols.	(appointing party, establish, information protocol)

As shown in the above table, 4 of the 12 feedbacks did not find a counterpart in the generated results of the missing triplets, indicating the whole automatic compliance checking framework achieves 53.3% accuracy in actual project testing. Compared with the gaps listed by domain experts, the proposed framework raised an additional 143 missing triplets. These additional triplets are composed of unessential tuples and tuples associated with listed gaps. During the generation of the regulatory graph, regulatory knowledge in the standard is extracted and processed in the form of sentences. Although these sentences contain target rules, they are sometimes sprinkled with unessential entities and relations. For example, “The appointing party shall establish the project’s information delivery milestones in accordance with the project’s plan of work.”, where the entity of “project” and its relation to “information delivery milestones” is not essential and will not be specifically checked during the assessment of compliance. These redundant entities and relations also result in a reduction in the accuracy of regulatory knowledge graph generation. Apart from the unessential tuples, the majority of the additional triplets are gap-associated triplets. The feedback provided by the experts is high-level, which does not cover all the details related to the gaps. For the task information delivery plan, the experts only stated the task information delivery plan (TIDP) should be established by the lead appointed party (Point 9 in Table 22) and did not list all the missing content within the TIDP. This situation results in the high-level triplet (lead appointed party, establish, task information delivery plan) being aligned while low-level tuples related to this gap, such as (task information delivery plan, consider, information delivery milestones), (task information delivery plan, identify, the level of information need), etc., cannot find its counterpart in the experts’ feedback. In addition to the aligned gaps, there are 4 gaps whose counterparts cannot be found in the missing triplets (Points 1,3,6,7 in Table 22). The reason for this lies in the format in which the proposed framework describes regulatory knowledge and project actions. As regulatory knowledge in standard and practical actions in the project documents are expressed as triplets in a graph, where the triplets have only two states (present or absent). Therefore, the proposed framework can only check whether the required action is implemented or not. It cannot check the implementation quality of the actions. For example, the experts suggested the lead appointed party intends to enhance its mobilisation plan (Point 7 in Table 22). Even though the quality of the mobilisation plan wasn't completely up to scratch, this mobilisation plan was technically established during the process of the project delivery. Therefore, the triplet of establishing mobilisation was detected in the project record document and the requirement of establishing mobilisation plan is fulfilled when checking the compliance. Furthermore, the proposed framework does not support consistency checking from a holistic perspective (e.g., Points 1 and 3) as these requirements are usually collated and summarised by examiners, which are not explicitly written in the standards.

Considering that feedback provided by experts is broad, and secondary indicators are commonly omitted when high-level indicators do not meet the requirements, an assessment result generated by the ontology-driven checking framework will be employed to further validate the aforementioned results. To ensure objectivity, three Ph.D. students involved in ISO 19650-related research were invited to thoroughly review project documents and respond to relevant inquiries. The final assessment results were determined by the average of the three evaluation outcomes. Table 23 presents the details of the compliance assessment results generated by the ontology-driven framework and some representative optimisation suggestions are listed in Table 24.

Table 23 - Compliance results generated by the ontology-driven framework

	Participant 1	Participant 2	Participant 3	Average
Compliance score	0.714	0.702	0.733	0.719

The compliance score assessed by the ontology-driven framework is about 0.72, while the result generated by the evidence-driven framework is around 0.74. The proximity of these two results can further substantiate the reliability of both methods to a considerable extent. Furthermore, comparing the optimisation recommendations provided by the ontological framework with the feedback of gaps from experts, there is a high degree of correlation between them (Table 24). Simultaneously, some other suggestions (Suggestions 9 to 11 in Table 24) corroborate the earlier analysis that the non-fulfilment of high-level indicator requirements contributed to the omission of secondary indicators gaps in expert feedback.

Table 24 - Correspondence between optimisation suggestions and compliance gaps

No.	Optimisation suggestions	Corresponding gaps
1	A complete set of project information requirements shall be established and take into consideration the information requirements which are needed at each key decision point.	Point 2
2	Appointing party shall establish their exchange information requirements.	Point 4
3	Appointing party shall establish the acceptance criteria for each information requirement: — the project's information standard, — the project's information production methods and procedures, and — the use of reference information or shared resources provided by the appointing party;	Point 5
4	The lead appointed party shall establish their exchange information requirements for each appointed party.	Point 8
5	A task information delivery plan (TIDP) shall be established by the lead appointed party and maintain throughout task team's appointment	Point 9
6	A common data environment shall be established for the projects.	Point 10
7	Appointing party shall capture lessons learned during the project and record them in a suitable knowledge store.	Point 11
8	A project information protocol shall be established.	Point 12
9	A Common data environment should be established before the invitation to tender stage.	–
10	Appointments should have the information protocol incorporated into them	–
11	Appointing party's information requirements shall consider: — organizational information requirements, — asset information requirements, and — project information requirements;	–

Based on the above analysis, it can be assumed the proposed automatic compliance checking framework demonstrates promising performance in practical project testing. Although there is still a gap in comparison with the manual work of domain experts, the proposed method realises a fully automated compliance checking and achieves an accuracy close to 60% against manual checking.

Chapter 8. Conclusions

Reflecting on the observations and results from previous sections, this chapter concludes this research by summarising the review and development work that has been done in this research (Section 7.1). Then, a summary of the research contributions of this thesis is presented in Section 7.2. Finally, the research limitations and the potential for improvement are discussed in Section 7.3.

8.1 Summary of research works

To explicate the research objective, the proposed research hypothesis was postulated as below:

> A smart BIM compliance checking framework, combining knowledge-driven subjective assessment and deep learning-based objective assessment can not only address the limitations of existing methods in comprehensiveness, granularity and efficiency but also fulfil the industrial demands on BIM compliance checking under different application scenarios.

Based on this hypothesis, a comprehensive review of BIM standards (systems) is conducted in this research and a smart compliance checking framework is developed based on the review findings and advanced techniques, such as semantic web, natural language processing, and deep learning.

Through the investigation of the current status of BIM adoption, it is apparent that the concept of BIM has been widely accepted worldwide. Many developed countries have already applied BIM technology to the actual construction process and some developing countries have also started to plan and formulate policies to promote the application and development of BIM technology. Driven by the expanding BIM market, BIM standards have also flourished. Currently, there are four relatively well-established BIM standard systems around the world, including the British BIM standard system, the American BIM standard system, the Chinese BIM standard system, and the Open BIM standard system. British BIM standard system proposed a comprehensive set of building information management processes, which aims to standardise the process of exchanging building information throughout its lifecycle.

Chinese BIM standard system focuses more on the application of BIM techniques and deliverables. American BIM standard system illustrates their principles of BIM application through practical use cases. The open BIM standard system strives to establish an open and standardised format and methodology for data exchange. Although each standard system has its characteristics, they are not completely irrelevant to each other. Many of the same or similar concepts are referenced in different standard systems. The IFC-based data exchange framework proposed by the Open BIM standard system is also adopted in the American Chinese BIM standard system and Chinese BIM standard system. According to the above findings, it is feasible to propose a checking framework that enables comprehensive and flexible BIM compliance checking.

After reviewing the studies related to BIM compliance checking over the last two decades, existing BIM compliance checking methods can be divided into two categories depending on the origin, namely the industrial approach and the academic approach. Most of the industrial assessment approaches utilise spreadsheets as assessment models, where only simple logic and calculation formulas are included. Therefore, these approaches are not capable of complex and fine-grained BIM compliance checking and fall short in comprehensiveness. Academic approaches are the other large collection of compliance checking methods. Many rule-based, deep learning-based, and ontological approaches have been developed to automatically interpret rules, acquire evidence from textual documents and match concepts and entities within them. However, the current methods are incapable of parsing complex logic and multiple requirements, which means they cannot handle checking complex qualitative rules. Since BIM standards are composed of a large number of complex qualitative rules, there is currently no available automatic approach to checking consistency against BIM standards. Additionally, existing methods are deficient in generalisability. Significant manual labour is still required to construct knowledge models when the target standard changes.

To address the shortcomings of existing approaches, a smart compliance checking framework for BIM standards is proposed in this research, which is composed of an ontology-driven subjective checking approach and an evidence-driven automatic checking approach.

The ontology-driven approach consists of three main components, namely ontological knowledge model, weighting matrix, and checking platform. The ontological knowledge model is composed of a hierarchical structure of key indicators and relationships between indicators, where the indicators are firstly extracted from BIM standards via text feature mining and then manually calibrated by the calibration expert panel. The weighting matrix follows the identical structure of the knowledge model, and the weighting of each indicator is determined based on the AHP method. The checking platform is developed based on Python, which aims to interact with users, backend ontological knowledge model, and developed weighting matrix to achieve compliance checking for different scenarios.

In terms of the evidence-driven approach, it evaluates compliance through comparing aligned triplets in graphs. Therefore, two separate pipelines are developed to transfer regulation documents and project records into knowledge graphs respectively. For regulatory knowledge graph generation, the pipeline is composed of a deep learning-based clause extractor, a triplet extractor, and a graph modelling engine. The clause extractor is a BERT-based fine-tuned DL model, which aims to extract clauses from regulatory documents. Then these clauses are parsed and processed by the triplet extractor, which is developed based on some advanced NLP technique (e.g., DP, PSG, and POS) and can convert textual clauses into regulatory knowledge tuples. Eventually, the regulatory tuples are assembled based on predefined rules in the modelling engine to form a regulatory knowledge graph that includes all the requirements mentioned in the standard. With regard to project action graph generation, the pipeline shares the same triplet extractor and graph modelling engine, while the clause extraction is replaced by tuple cleansing. In this pipeline, the documents related to the project are first processed by the triplet extractor. The actual behaviours of all the stakeholders are converted into action triplets. These raw triplets are screened by the tuple cleansing algorithm, where all the triplets irrelevant to the compliance against the standard are filtered out. The filtered triplets are finally integrated as a project action graph by the graph modelling engine. The graph alignment is the other core part of the proposed evidence-driven approach. To tackle the challenge of dynamic entity mapping, the proposed framework introduces a graph learning approach to capture the structural features of the entities in the graph and divides the whole alignment process into two stages. The first stage is semantic alignment, where the entities and relations in the two graphs are preliminarily aligned based on their semantic similarity. The aligned entities are then used as node seeds in the graph embedding engine to generate graph embeddings for each node. Then, the GCN-based graph embedding engine captures the structural information of the target node through graph convolution operations with neighbouring nodes and represents the structural information as a high-dimensional dense graph embedding. These embeddings indicate the structural features of the nodes, in other words, embeddings of nodes connected to similar nodes are similar. Based on this principle, the structural similarity of the entities in the two graphs is calculated in the second stage. The entities in the action graph are then replaced by their counterparts in the regulation graph according to their structural similarity. The overall compliance is evaluated by the proportion of the aligned triplets in the regulatory graph.

To verify the performance of the proposed framework under a practical scenario, the documents of a real construction project and a specific international BIM standard (ISO 19650) are applied to the framework. Through comparison of system-generated results and feedback from domain experts, it can be observed that the proposed method has successfully identified some inconsistencies that are also listed by experts but are flawed in overall assessment and quality checking. The overall accuracy of the generated results is about 60%, which needs to be further improved in the future.

8.2 Research contributions

This research encompasses efforts directed towards both theoretical and practical advancements, aiming to fill the gaps in comprehensive BIM compliance checking and attempting to propose a fully automated approach for BIM compliance checking. In light of the findings and development expounded upon in this dissertation, the principal contributions arising from this research are enumerated below:

1. The existing literature reviews of BIM standards mainly focus on one specific aspect (e.g., sustainability, facilities management, collaboration, etc.) or within a specific standard system. A few studies have touched on comparative analyses between BIM standard systems, but all are superficial. Therefore,

there is still a gap in comprehensive and in-depth comparative analyses of multiple BIM standard systems. To fill this gap, a comprehensive comparative analysis has been conducted in Chapter 2, which carefully reviews some representative standards in each BIM standard system and summarises the characteristics of each standard system. This detailed comprehensive comparative analysis can facilitate a more thorough understanding of the status of BIM standards and the correlations between BIM standard systems for both academics and industry professionals. Furthermore, the comparison also elaborates on the strengths and weaknesses of each standards system, which can help the government and organisations to optimise and improve the existing BIM standards.

2. In this research, a high-quality domain dataset is developed based on data augmentation and Delphi validation, which contains labelled samples of clauses and descriptions in the standards in the AEC domain. This domain dataset has been open-sourced and can be utilised by other researchers for similar research.

3. For BIM compliance checking, an ontological BIM compliance checking framework is developed in Chapter 4. This framework utilises an ontological knowledge model to represent the requirements stated in the standards, enabling the linkage between different BIM standards. With the help of the ontological knowledge model, this framework can deal with complex reasoning and flexible BIM compliance checking. For example, BIM compliance between specific stakeholders and specific BIM standards, BIM compliance between specific stakeholders and BIM documents, overall BIM compliance, etc. Additionally, the proposed framework outperforms other existing BIM compliance checking approaches in comprehensiveness and granularity level. The ontological representations make it easier to manage and update the knowledge model, which enables knowledge in the knowledge model to become dynamic and new knowledge can be continuously incorporated to form a larger knowledge model. Hence, the proposed ontological BIM compliance checking framework may profoundly impact the application of BIM in the AEC industry. This framework not only fills the gap in comprehensive BIM

compliance assessment but also provides a more nuanced compliance assessment compared to existing methods. The ontological knowledge model in the framework can constantly help the various enterprises identify deficiencies in their workflow and optimise them accordingly to improve productivity.

4. To achieve automatic BIM compliance checking, two innovative information extraction pipelines are developed in this research, which aim to convert textual documents into structured graph representations. These two pipelines incorporate linguistic knowledge with NLP and DL techniques to achieve sentence-to-tuple transformation, which further improves the capabilities of the pipelines in recognising complex relationships and parsing multiple requirements. These two automatic information extraction pipelines fill the gap in fully automated complex knowledge mining for regulation documents and significantly reduce the time and cost of project information extraction. The advent of the automated knowledge transformation method makes it possible to perform large-scale knowledge extraction from qualitative regulation documents, as well as facilitates the digitization of regulatory knowledge. It also brings some downstream applications, such as multi-regulation knowledge fusion, automated compliance checking, multi-objective optimisation, and holistic decision-making, one step closer to reality. In summary, these two pipelines may have a profound impact on knowledge transformation and information extraction in the AEC domain. Through these two pipelines, textual documents can be easily converted into structured tuple representations, which can be directly utilised by other functions or applications.

5. To realise the automatic alignment of dynamic entities, a graph learning-based compliance assessment method is developed in Chapter 5. Due to the difficulty in dynamic entity mapping, all existing automated compliance checking researches focus on quantitative technical standards (e.g., design code), where the entities are static (e.g., wall, beam, space, etc.) and most of the relation between them are simple (e.g., less than, more than). The approach for automated BIM compliance checking is a vacuum. The graph learning-based compliance checking approach is the first attempt to address this problem. This

method aligns dynamic entities by capturing the features that do not change with the entity, which are the structural embedding of the nodes in a graph. The results of the practical testing proved the effectiveness of this method, although its accuracy is not quite impressive. The contribution of this method is that it sets a precedent for automated BIM compliance checking and also provides a new solution idea for similar studies.

8.3 Research limitations and future works

This research suffers from deficiencies in the methodology, developed frameworks and use cases used in validation and testing. The specific limitations and future work are discussed below.

1. According to the result of practical testing, the accuracy of regulatory knowledge graph generation is approximately 72.5 and the overall accuracy of the proposed automatic compliance checking framework is about 60%. This level of accuracy is not sufficient for practical applications. The triplet extractor and the graph embedding engine need to be further improved. For triplet extractors, more specific rules can be predefined in the extractor or adopting deep learning methods to improve the performance of triplet assembly. For the graph embedding engine, more convolutional layers can be added to the engine to obtain more global structural information. A higher dimensional graph embeddings may also improve the performance of the proposed framework.

2. The proposed automatic framework is mainly developed based on NLP techniques. Hence, it can only process the textual content in the documents. The interoperation of information in tables and figures still relies on manual work. In future work, some computer vision techniques, such as optical character recognition (OCR), and Python libraries will be introduced into the system to improve the framework's capability of processing heterogeneous information.

3. In the process of regulatory knowledge graph generation, the content in the standards is firstly tokenised as separate sentences and then classified by the DL model. The extracted clauses are processed by the triplet extractor and all

the generated triplets are integrated into the regulatory graph. The entire processing above is performed at the sentence level, which means referential relationships between clauses may be lost during processing. For example, the “information requirements” established by the appointing party (Clause 5.2.1) are referenced as “requirements” in the following paragraph. Due to the missing of their correlation, the system will treat “information requirements” and “requirements” as two separate instances when assembling tuples and constructing graphs. To address this deficiency, some algorithm for coreference resolution will be applied to the proposed framework, which can identify the coreference of entities and relations between different clauses.

Furthermore, as a sentence-level classifier, the clause extraction model can only distinguish clauses from descriptions. It cannot filter out the unnecessary entities and relations in clauses, resulting in some redundant triplets. For example, Clause 5.2.1 of ISO 19650 states that “when establishing the exchange information requirements, the appointing party shall establish the supporting information that the prospective lead appointed party might need.", where the triplet (appointing party, shall establish, the supporting information) is necessary, while the triplet (lead appointed party, might need, the supporting information) is superfluous.

4. In terms of the graph learning method adopted in this research, it also suffers some limitations. For example, the method of generating graph embeddings through graph convolution is only applicable to static graphs, which means that the graph embeddings of the nodes are required to be re-computed after each change of the target standard or project document. Moreover, this graph convolution network only considers the connection between entities, and it cannot involve relations between entities in the generation of node embeddings. Finally, the proposed graph embedding-based alignment approach can only achieve the one-to-one entity mapping at the current stage. To solve the above deficiencies, some other graph learning models can be applied to the embedding engine, such as R-GCNs (Schlichtkrull et al. 2017), TransE (Bordes et al. 2013), RESCAL(Nickel et al. 2011), etc. In addition, some thresholds for

entity alignment can be set in the corresponding algorithm, which will enable one-to-many and many-to-one mapping of entities and relations.

5. To verify the practical performance of the proposed framework, a real construction project was selected as a use case in this research to test the framework. However, the selected cases only tentatively validated the performance of the proposed framework and more practical cases will be required to be applied to the framework to comprehensively test and evaluate its performance in practical application.

Bibliography

Administration of Quality Supervision Inspection and Quarantine of People's Republic of China. 2010. Industry foundation classes platform: GB/T 25507-2010. Available at: https://www.chinesestandard.net/PDF/English.aspx/GBT25507-2010 [Accessed: 25 April 2024].

Ammer, D., Mustafa Al, S. and Jason, U. 2015. BIM Client Maturity: Literature Review. In: *12 IPGRC Salford University*. Manchester, UK. Available at: https://www.researchgate.net/publication/279293516_BIM_Client_Maturity_Literature_Review [Accessed: 25 April 2024].

Arup. 2015a. *BIM Maturity Measure*. Available at: https://www.arup.com/perspectives/publications/research/section/bim-maturity-model [Accessed: 25 April 2024].

Arup. 2015b. *BIM Maturity Model - Arup*. Available at: https://www.arup.com/perspectives/publications/research/section/bim-maturity-model [Accessed: 25 April 2024].

Avenier, M.J. 2010. Shaping a Constructivist View of Organizational Design Science. *http://dx.doi.org/10.1177/0170840610374395* 31(9–10), pp. 1229–1255. Available at: https://journals.sagepub.com/doi/10.1177/0170840610374395 [Accessed: 25 April 2024].

Azzran, S.A., Ibrahim, K.F., Tah, J.H.M. and Abanda, F.H. 2019. Assessment of open BIM standards for facilities management. *Innovative Production And Construction: Transforming Construction Through Emerging Technologies*, pp. 247–259. doi: 10.1142/9789813272491_0015.

Barannik, S. V. 2017. Review of the USA National BIM standard NBIMS-US V3. *CAD&GIS for roads* (1(8)), pp. 4–8. doi: 10.17273/CADGIS.2017.1.1.

Bayer, M., Kaufhold, M.-A. and Reuter, C. 2021. A Survey on Data Augmentation for Text Classification. *ACM Computing Surveys*. Available at: https://dl.acm.org/doi/10.1145/3544558 [Accessed: 25 April 2024].

Beach, T.H., Kasim, T., Li, H., Nisbet, N. and Rezgui, Y. 2013. Towards automated compliance checking in the construction industry. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8055 LNCS(PART 1), pp. 366–380. Available at: https://link.springer.com/chapter/10.1007/978-3-642-40285-2_32 [Accessed: 25 April 2024].

Beach, T.H., Rezgui, Y., Li, H. and Kasim, T. 2015. A rule-based semantic approach for automated regulatory compliance in the construction sector. *Expert Systems with Applications* 42(12), pp. 5219–5231. doi: 10.1016/J.ESWA.2015.02.029.

Beiderbeck, D., Frevel, N., von der Gracht, H.A., Schmidt, S.L. and Schweitzer, V.M. 2021. Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX* 8, p. 101401. doi: 10.1016/J.MEX.2021.101401.

Beijing Municipal Commission of Housing and Urban-Rural Development. 2014. *Design Standard for Information Model of Civil Building*. Available at: https://www.codeofchina.com/standard/DB11T1069-2014.html [Accessed: 25 April 2024].

Van Berlo, L., Dijkmans, T., Hendriks, H., Consultant, Q. and Spekkink, D. 2012. BIM QUICKSCAN: BENCHMARK OF BIM PERFORMANCE IN THE NETHERLANDS. In: *Proceedings of the 29th International Conference CIB W78*. pp. 17–19.

Bew, M. and Richards, M. 2008. BIM Maturity Levels. In: *Construct IT Autumn 2008 Members*. Brighton, UK. Available at: https://www.researchgate.net/publication/279293516_BIM_Client_Maturity_Literature_Review [Accessed: 18 March 2024].

BIM Group of Tsinghua University. 2010. Research on Chinese Building Information Modeling Standard Framework. *Journal of Information Technologyin Civil Engineering and Architecture* 2(2), pp. 1–5. Available at: http://tmjzgcxxjs.manuscripts.cn/article/id/b50b14c7-23b2-40bd-8e6a-fa9ea8f572f0 [Accessed: 25 April 2024].

BIMconnect. 2019. *Are you BIM-Ready? The BIM-Profiler will help you assess your company's BIM maturity.* Available at: https://profiler.bimconnect.org/ [Accessed: 25 April 2024].

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J. and Yakhnenko, O. 2013. Translating Embeddings for Modeling Multi-relational Data. *Advances in Neural Information Processing Systems* 26.

Boukamp, F. and Akinci, B. 2007. Automated processing of construction specifications to support inspection and quality control. *Automation in Construction* 17(1), pp. 90–106. doi: 10.1016/J.AUTCON.2007.03.002.

British Standards Institution (BSI). 2007. *BS 1192:2007+A2:2016 - Collaborative production of architectural, engineering and construction information. Code of practice*. Available at: https://knowledge.bsigroup.com/products/collaborative-production-of-architectural-engineering-and-construction-information-code-of-practice?version=standard [Accessed: 19 March 2024].

British Standards Institution (BSI). 2013. *PAS 1192-2:2013 - Specification for information management for the capital/delivery phase of construction projects using building information modelling*. Available at: https://knowledge.bsigroup.com/products/specification-for-information-management-for-the-capital-delivery-phase-of-construction-projects-using-building-information-modelling?version=standard [Accessed: 19 March 2024].

British Standards Institution (BSI). 2014a. *BS 1192-4 : 2014 - Collaborative production of information Part 4 : Fulfilling employer' s information exchange requirements using COBie - Code of practice*. p. 58. Available at: http://shop.bsigroup.com/forms/BS-1192-4/ [Accessed: 19 March 2024].

British Standards Institution (BSI). 2014b. *PAS 1192-3:2014 - Specification for information management for the operational phase of assets using building information modelling Publishing and copyright information*.

British Standards Institution (BSI). 2015. *PAS 1192-5:2015 Specification for security-minded building information modelling, digital built environments and smart asset management*. British Standards Institution (BSI). pp. 1–66.

British Standards Institution (BSI). 2018. *PAS 1192-6:2018 Specification for collaborative sharing and use of structured Health and Safety information using BIM*. British Standards Institution (BSI). pp. 1–66.

British Standards Institution (BSI). 2022. *BS 8536:2022 Design, manufacture and construction for operability. Code of practice.* British Standards Institution (BSI). p. 110.

Bryde, D., Broquetas, M. and Volm, J.M. 2013. The project benefits of Building Information Modelling (BIM). *International Journal of Project Management* 31(7), pp. 971–980. doi: 10.1016/J.IJPROMAN.2012.12.001.

buildingSMART. 2018a. *BIM Collaboration Format (BCF) - buildingSMART International*. Available at: https://www.buildingsmart.org/standards/bsi-standards/bim-collaboration-format-bcf/ [Accessed: 25 April 2024].

buildingSMART. 2018b. *openBIM - buildingSMART International*. Available at: https://www.buildingsmart.org/about/openbim/ [Accessed: 25 April 2024].

buildingSMART. 2018c. *openBIM Definition - buildingSMART International*. Available at: https://www.buildingsmart.org/about/openbim/openbim-definition/ [Accessed: 25 April 2024].

buildingSMART. 2018d. *Standards Library - buildingSMART International*. Available at: https://www.buildingsmart.org/standards/bsi-standards/standards-library/ [Accessed: 25 April 2024].

buildingSMART. 2022. *Model View Definitions (MVD) - buildingSMART Technical*. Available at: https://technical.buildingsmart.org/standards/ifc/mvd/ [Accessed: 19 March 2024].

Cao, S., Lu, W. and Xu, Q. 2016. Deep Neural Networks for Learning Graph Representations. *Proceedings of the AAAI Conference on Artificial Intelligence* 30(1), pp. 1145–1152. Available at: https://ojs.aaai.org/index.php/AAAI/article/view/10179 [Accessed: 25 April 2024].

Chantrapornchai, C. and Tunsakul, A. 2021. Information Extraction Tasks based on BERT and SpaCy on Tourism Domain. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)* 15(1), pp. 108–122. Available at: https://ph01.tci-thaijo.org/index.php/ecticit/article/view/228621 [Accessed: 25 April 2024].

Chen, K. 2015. *A Strategic Decision Making framework for Organisational BIM implementation*.

Chi, N.-W., Lin, K.-Y., El-Gohary, N. and Hsieh, S.-H. 2017. Gazetteers for Information Extraction Applications in Construction Safety Management. pp. 401–408. Available at: https://ascelibrary.org/doi/10.1061/9780784480847.050 [Accessed: 25 April 2024].

Chiche, A. and Yitagesu, B. 2022. Part of speech tagging: a systematic review of deep learning and machine learning approaches. *Journal of Big Data* 9(1), pp. 1–25. Available at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00561-y [Accessed: 20 March 2024].

Chomsky, N. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2(3), pp. 113–124. doi: 10.1109/TIT.1956.1056813.

Chomsky, N. 1965. Aspects of the theory of syntax.

Chong, H.-Y., Lee, C.-Y. and Wang, X. 2016. A mixed review of the adoption of Building Information Modelling (BIM) for sustainability. Available at: http://dx.doi.org/10.1016/j.jclepro.2016.09.222 [Accessed: 25 April 2024].

Computer Integrated Construction Research Program. 2013. *BIM Planning Guide for Facility Owners | BIM Planning*. Available at: https://bim.psu.edu/owners_guide/ [Accessed: 25 April 2024].

Construction Specifications Institute. 2006. *OmniClass® - Construction Specifications Institute*. Available at: https://www.csiresources.org/standards/omniclass [Accessed: 19 March 2024].

Corro, L. Del and Gemulla, R. 2013. ClausIE: Clause-Based Open Information Extraction. *Proceedings of the 22nd international conference on World Wide Web*. Available at: http://www.mpi-inf.mpg.de/departments/d5/software/ [Accessed: 25 April 2024].

Coulombe, C. 2018. Text Data Augmentation Made Simple By Leveraging NLP Cloud APIs. Available at: https://arxiv.org/abs/1812.04718v1 [Accessed: 25 April 2024].

Craib, I. 1997. *Classical Social Theory*. Oxford University Press. Available at: https://philpapers.org/rec/CRACST [Accessed: 25 April 2024].

Crotty, M.J. 1998. *The Foundations of Social Research : Meaning and Perspective in the Research Process*. SAGE Publications Ltd.

Dadmehr, N. and Coates, S. 2019. AN APPROACH TO 'NATIONAL ANNEX TO ISO 19650-2'. In: *14th International Postgraduate Research Conference 2019 (IPGRC)*. Salford, UK. Available at: https://www.researchgate.net/publication/341915122_AN_APPROACH_TO_NATIONAL_ANNEX_TO_ISO_19650-2 [Accessed: 19 March 2024].

Dalkey, N. and Helmer, O. 1963. An Experimental Application of the DELPHI Method to the Use of Experts. *https://doi.org/10.1287/mnsc.9.3.458* 9(3), pp. 458–467. Available at: https://pubsonline.informs.org/doi/abs/10.1287/mnsc.9.3.458 [Accessed: 20 March 2024].

Demir, S., Garrett, J., Akinci, B., Akin, O. and Palmer, M. 2010. A Semantic Web-based Approach for Representing and Reasoning with Vocabulary for Computer-based Standards Processing.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference* 1, pp. 4171–4186. Available at: https://arxiv.org/abs/1810.04805v2 [Accessed: 25 April 2024].

Dworkin, M. et al. 2009. *Research Methods for Business Students*. Available at: https://books.google.com/books/about/Research_Methods_for_Business_Students.html?hl=zh-CN&id=u-txtfaCFiEC [Accessed: 25 April 2024].

Eadie, R., Browne, M., Odeyinka, H., McKeown, C. and McNiff, S. 2013. BIM implementation throughout the UK construction project lifecycle: An analysis. *Automation in Construction* 36, pp. 145–151. doi: 10.1016/J.AUTCON.2013.09.001.

Eastman, C., Fisher, D., Lafue, G., Lividini, J., Stoker, D. and Yessios, C. 1974. An Outline of the Building Description System. Research Report No. 50.

Edirisinghe, R. and London, K. 2015a. Comparative analysis of international and national level BIM standardization efforts and BIM adoption. *Proceedings of the 32nd International Conference of CIB W78*, pp. 149–158. Available at: https://researchrepository.rmit.edu.au/esploro/outputs/conferenceProceeding/Comparative-analysis-of-international-and-national-level-BIM-standardization-efforts-and-BIM-adoption/9921858856301341 [Accessed: 25 April 2024].

Edirisinghe, R. and London, K. 2015b. Comparative Analysis of International and National Level BIM Standardization Efforts and BIM adoption. In: *The 32nd International Conference of CIB W78*. pp. 149–158.

Efficiency and Reform Group of the Cabinet Office. 2011. *Government Construction Strategy*.

Feijó, B., Krause, W.G., Smith, D.L. and Dowling, P.J. 1994. A hypertext model for steel design codes. *Journal of Constructional Steel Research* 28(2), pp. 167–186. doi: 10.1016/0143-974X(94)90041-8.

Feng, D. and Chen, H. 2021. A small samples training framework for deep Learning-based automatic information extraction: Case study of construction accident news reports analysis. *Advanced Engineering Informatics* 47, p. 101256. doi: 10.1016/J.AEI.2021.101256.

Feng, H. et al. 2022. GCN-Based Pavement Crack Detection Using Mobile LiDAR Point Clouds. *IEEE Transactions on Intelligent Transportation Systems* 23(8), pp. 11052–11061. doi: 10.1109/TITS.2021.3099023.

Gábor, N. 2007. Ontology development: Methodologies for ontology engineering. In: *Semantic Web Services: Concepts, Technologies, and Applications*. Springer Berlin Heidelberg, pp. 107–134. Available at: https://link.springer.com/chapter/10.1007/3-540-70894-4_4 [Accessed: 25 April 2024].

Gligorijević, V. et al. 2021. Structure-based protein function prediction using graph convolutional networks. *Nature Communications 2021 12:1* 12(1), pp. 1–14. Available at: https://www.nature.com/articles/s41467-021-23303-9 [Accessed: 25 April 2024].

Greenbaum, J. 1991. A Design of One's Own: Towards Participatory Design in the US. *DAIMI Report Series* 20(375). doi: 10.7146/DPB.V20I375.6607.

Grover, A. and Leskovec, J. 2016. node2vec: Scalable Feature Learning for Networks. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-August-2016, pp. 855–864. Available at: https://arxiv.org/abs/1607.00653v1 [Accessed: 25 April 2024].

Guerriero, A. et al. 2019. BIM4VET, Towards BIM Training Recommendation for AEC Professionals. *Advances in Informatics and Computing in Civil and Construction Engineering*, pp. 833–840. Available at: https://link.springer.com/chapter/10.1007/978-3-030-00220-6_100 [Accessed: 18 March 2024].

Hackitt, J.D. 2017. *Building a Safer Future Independent Review of Building Regulations and Fire Safety: Interim Report*. Available at: www.gov.uk/government/publications [Accessed: 25 April 2024].

Han, X. and Wang, L. 2020. A Novel Document-Level Relation Extraction Method Based on BERT and Entity Information. *IEEE Access* 8, pp. 96912–96919. doi: 10.1109/ACCESS.2020.2996642.

Hassyan Power Complex, Dubai, UAE. 2016. Available at: https://www.power-technology.com/projects/hassyan-clean-coal-project-dubai/?cf-view [Accessed: 25 April 2024].

Hevner, A. and Chatterjee, S. 2010. Design Science Research in Information Systems. In: *Design Science Research in Information Systems. Integrated Series in Information*

Systems. Springer, Boston, MA, pp. 9–22. Available at: https://link.springer.com/chapter/10.1007/978-1-4419-5653-8_2 [Accessed: 25 April 2024].

Hjelseth, E. and Nisbet, N. 2011. Capturing Normative constraints by use of the semantic mark-up RASE methodology. In: *Proceedings of CIB W78-W102 Conference*. Sophia Antipolis, France, pp. 1–10. Available at: http://itc.scix.net/data/works/att/w78-2011-Paper-45.pdf [Accessed: 25 April 2024].

Hong, Y., Hovhannisyan, V., Xie, H. and Brilakis, I. 2021. Determining construction method patterns to automate and optimise scheduling – a graph-based approach. In: *Proceedings of the 2021 European Conference on Computing in Construction*. European Council on Computing in Construction, pp. 59–66. Available at: https://ec-3.org/publications/conference/paper/?id=EC32021_165 [Accessed: 25 April 2024].

Ibrahim, M.Y. and Al-Kazzaz, D.A. 2021. A comparative analysis of BIM standards and guidelines between UK and USA. *Journal of Physics: Conference Series* 1973(1), p. 012176. Available at: https://iopscience.iop.org/article/10.1088/1742-6596/1973/1/012176 [Accessed: 25 April 2024].

International Code Council. 2018. *International Building Code*. International Code Council. Available at: http://bcpad.net/Code%20Books/IBC%202006.pdf [Accessed: 20 March 2024].

International Organization for Standardization (ISO). 2016. *ISO 29481-1:2016 - Building information models — Information delivery manual — Part 1: Methodology and format*. Available at: https://www.iso.org/standard/60553.html [Accessed: 19 March 2024].

International Organization for Standardization (ISO). 2018a. *ISO 16739-1:2018 - Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema*. Available at: https://www.iso.org/standard/70303.html [Accessed: 19 March 2024].

International Organization for Standardization (ISO). 2018b. *ISO 19650-1:2018 - Organization and digitization of information about buildings and civil engineering works,* *including building information modelling (BIM) — Information management using building information modelling — Part 1: Concepts and principles*. International Organization for Standardization (ISO). Available at: https://www.iso.org/standard/68078.html [Accessed: 19 March 2024].

International Organization for Standardization (ISO). 2018c. *ISO 19650-2:2018 - Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 2: Delivery phase of the assets*. Available at: https://www.iso.org/standard/68080.html [Accessed: 19 March 2024].

International Organization for Standardization (ISO). 2020a. *ISO 19650-3:2020 - Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 3: Operational phase of the assets*. International Organization for Standardization (ISO). Available at: https://www.iso.org/standard/75109.html [Accessed: 19 March 2024].

International Organization for Standardization (ISO). 2020b. *ISO 19650-5:2020 - Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 5: Security-minded approach to information management*. International Organization for Standardization (ISO). Available at: https://www.iso.org/standard/74206.html [Accessed: 19 March 2024].

International Organization for Standardization (ISO). 2022. *ISO 12006-3:2022 - Building construction — Organization of information about construction works — Part 3: Framework for object-oriented information*. Available at: https://www.iso.org/standard/74932.html [Accessed: 19 March 2024].

Jia, Y., Wang, J., Shou, W., Hosseini, M.R. and Bai, Y. 2023. Graph neural networks for construction applications. Available at: https://doi.org/10.1016/j.autcon.2023.104984 [Accessed: 25 April 2024].

Jiang, S., Jiang, L., Han, Y., Wu, Z. and Wang, N. 2019. OpenBIM: An Enabling Solution for Information Interoperability. *Applied Sciences 2019, Vol. 9, Page 5358*

9(24), p. 5358. Available at: https://www.mdpi.com/2076-3417/9/24/5358/htm [Accessed: 25 April 2024].

Jiang, W. and Luo, J. 2022. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications* 207, p. 117921. doi: 10.1016/J.ESWA.2022.117921.

Johannesson, P. and Perjons, E. 2014. *An Introduction to Design Science*. 1st ed. Springer International Publishing. doi: 10.1007/978-3-319-10632-8/COVER.

Jones, K.S. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28(1), pp. 11–21. doi: 10.1108/EB026526/FULL/XML.

Khurana, D., Koli, A., Khatter, K. and Singh, S. 2023. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications* 82(3), pp. 3713–3744. Available at: https://link.springer.com/article/10.1007/s11042-022-13428-4 [Accessed: 20 March 2024].

Kim, H. and Kim, C. 2020. Deep-Learning-Based Classification of Point Clouds for Bridge Inspection. *Remote Sensing* 12. Available at: www.mdpi.com/journal/remotesensing [Accessed: 25 April 2024].

Kim, Y.M. and Lee, T.H. 2020. Korean clinical entity recognition from diagnosis text using BERT. *BMC Medical Informatics and Decision Making* 20(7), pp. 1–9. Available at: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01241-8 [Accessed: 25 April 2024].

Kipf, T.N. and Welling, M. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*. Available at: https://arxiv.org/abs/1609.02907v4 [Accessed: 25 April 2024].

Kurose, J. 1992. On computing per-session performance bounds in high-speed multi-hop computer networks. *Proceedings of the 1992 ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS/PERFORMANCE 1992*, pp. 128–139. Available at: https://dl.acm.org/doi/10.1145/133057.133097 [Accessed: 25 April 2024].

Lau, G.T. and Law, K. 2004. An Information Infrastructure for Comparing Accessibility Regulations and Related Information from Multiple Sources. In: Professur Informatik im Bauwesen ed. *Proceedings of the 10th International Conference on Computing in Civil and Building Engineering*. Weimar, Germany.

Lawrence, C., Tuunanen, T. and Myers, M.D. 2010. Extending design science research methodology for a multicultural world. *IFIP Advances in Information and Communication Technology* 318, pp. 108–121. Available at: https://link.springer.com/chapter/10.1007/978-3-642-12113-5_7 [Accessed: 20 March 2024].

Lea, G., Ganah, A., Goulding, J. and Ainsworth, N. 2015. Identification and analysis of UK and US BIM standards to aid collaboration. *Building Information Modelling (BIM) in Design, Construction and Operations* 1, pp. 505–516. doi: 10.2495/BIM150411.

Li, S., Cai, H., Asce, M. and Kamat, V.R. 2016. Integrating Natural Language Processing and Spatial Reasoning for Utility Compliance Checking. *Journal of Construction Engineering and Management* 142(12), p. 04016074. Available at: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CO.1943-7862.0001199 [Accessed: 25 April 2024].

LIST. 2016. *BIM4VET*. Available at: https://www.bim4vet.eu/ [Accessed: 25 April 2024].

Liu, K. and El-Gohary, N. 2017. Ontology-based semi-supervised conditional random fields for automated information extraction from bridge inspection reports. *Automation in Construction* 81, pp. 313–327. doi: 10.1016/J.AUTCON.2017.02.003.

Maharana, K., Mondal, S. and Nemade, B. 2022. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings* 3(1), pp. 91–99. doi: 10.1016/J.GLTP.2022.04.020.

Mcauley, B., Hore, A. and West, R. 2017. *BICP Global BIM Study-Lessons for Ireland's BIM Programme*. Dublin. Available at: https://arrow.tudublin.ie/beschrecrep [Accessed: 25 April 2024].

Melnikovas, A. 2018. Towards an explicit research methodology: Adapting research onion model for futures studies. *Journal of Futures Studies* 23(2), pp. 29–44. doi: 10.6531/JFS.201812_23(2).0003.

Ministry of Housing and Urban-Rural Development (China). 2007. *Building information model flatform: JG/T 198-2007*. Available at: https://www.chinesestandard.net/Default.aspx?StdID=JG/T%20198-2007 [Accessed: 25 April 2024].

Ministry of Housing and Urban-Rural Development (China). 2012. *Engineering Construction Standard Specification Formulation and Revision Plan of 2012*. Available at: https://www.mohurd.gov.cn/zcfg/jsbwj_0/jsbwjbzde/201202/W020120221105829.xls [Accessed: 25 April 2024].

Ministry of Housing and Urban-Rural Development (China). 2016. *Unified standard for building information modeling: GB/T 51212-2016*. Available at: https://www.chinesestandard.net/Default.aspx?StdID=GB/T%2051212-2016 [Accessed: 25 April 2024].

Ministry of Housing and Urban-Rural Development (China). 2017a. *GB/T 51235-2017 Standard for building information modeling in construction*. Ministry of Housing and Urban-Rural Development (China). Available at: https://www.webfree.net/download/4943 [Accessed: 19 March 2024].

Ministry of Housing and Urban-Rural Development (China). 2017b. *GB/T 51269-2017 Standard for classification and ending of building information model*. Ministry of Housing and Urban-Rural Development (China). Available at: https://www.chinesestandard.net/PDF/English.aspx/GBT51269-2017 [Accessed: 19 March 2024].

Ministry of Housing and Urban-Rural Development (China). 2017c. *Standard for building information modeling in construction*. Available at: https://www.chinesestandard.net/Default.aspx?StdID=GB/T51235 [Accessed: 25 April 2024].

Ministry of Housing and Urban-Rural Development (China). 2017d. *Standard for classification and coding of building information model*. Available at: https://www.chinesestandard.net/Default.aspx?StdID=GB/T%2051269-2017 [Accessed: 25 April 2024].

Ministry of Housing and Urban-Rural Development (China). 2018a. *GB/T 51301-2018 Standard for design delivery of building information modeling*. Available at: https://www.chinesestandard.net/PDF/English.aspx/GBT51301-2018 [Accessed: 20 March 2024].

Ministry of Housing and Urban-Rural Development (China). 2018b. *Standard for design delivery of building information modeling*. Available at: https://www.chinesestandard.net/PDF/English.aspx/GBT51301-2018 [Accessed: 25 April 2024].

Ministry of Housing and Urban-Rural Development (China). 2021a. *Building Information Model Storage Standard*. Available at: https://www.chinesestandard.net/PDF/English.aspx/GBT51447-2021 [Accessed: 25 April 2024].

Ministry of Housing and Urban-Rural Development (China). 2021b. *GB/T 51447-2021 Building Information Model Storage Standard*. Ministry of Housing and Urban-Rural Development (China). Available at: https://www.chinesestandard.net/PDF/English.aspx/GBT51447-2021 [Accessed: 20 March 2024].

Minsheng, X., Haitao, Y. and Weidong, L. 2016. *Municipal BIM Theory and Practice*. Shanghai: Tongji University Press.

Minsky, M. 2019. A Framework For Representing Knowledge. *Frame Conceptions and Text Understanding*, pp. 1–25. doi: 10.1515/9783110858778-003/PDF.

Moreno, A., Isern, D. and López Fuentes, A.C. 2013. Ontology-based information extraction of regulatory networks from scientific articles with case studies for Escherichia coli. *Expert Systems with Applications* 40(8), pp. 3266–3281. doi: 10.1016/J.ESWA.2012.12.090.

Nadkarni, P.M., Ohno-Machado, L. and Chapman, W.W. 2011. Natural language processing: An introduction. *Journal of the American Medical Informatics Association* 18(5), pp. 544–551. Available at: https://dx.doi.org/10.1136/amiajnl-2011-000464 [Accessed: 20 March 2024].

National Building Specification. 2019. *National BIM Report 2019*. Available at: https://www.thenbs.com/PublicationIndex/documents/details?Pub=NBS&DocID=326117 [Accessed: 25 April 2024].

National Institute of Building Sciences. 2011. *COBie Standard | National Institute of Building Sciences*. Available at: https://www.nibs.org/nbims/cobie/v3 [Accessed: 19 March 2024].

National Institute of Building Sciences buildingSMART alliance. 2007. *National BIM Standard – United States™ (NBIMS-US™) Version 1*.

National Institute of Building Sciences buildingSMART alliance. 2012. *National BIM Standard – United States™ (NBIMS-US™) Version 2*. Available at: https://buildinginformationmanagement.wordpress.com/2012/05/17/national-bim-standard-united-states-nbims-us-version-2-v2-released-may-2012/ [Accessed: 19 March 2024].

National Institute of Building Sciences buildingSMART alliance. 2015. *National BIM Standard – United States™ (NBIMS-US™) Version 3*. Available at: https://www.nationalbimstandard.org/nbims-us-v3 [Accessed: 19 March 2024].

National Institute of Building Sciences buildingSMART alliance. 2023. *NBIMS-US™ Public Review | National BIM Standard - United States*. Available at: https://www.nationalbimstandard.org/about/nbims-us-public-review [Accessed: 25 April 2024].

National People's Congress of the People's Republic of China. 1988. *Standardisation Law of the People's Republic of China*. Available at: https://www.jetro.go.jp/ext_images/world/asia/cn/ip/law/pdf/regulation/regulation20180101.pdf [Accessed: 25 April 2024].

Nauata, N., Chang, K.H., Cheng, C.Y., Mori, G. and Furukawa, Y. 2020. House-GAN: Relational Generative Adversarial Networks for Graph-constrained House Layout Generation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12346 LNCS, pp. 162–177. Available at: https://arxiv.org/abs/2003.06988v1 [Accessed: 25 April 2024].

Nickel, M., Tresp, V. and Kriegel, H.-P. 2011. A Three-Way Model for Collective Learning on Multi-Relational Data.

Noy, N. 2001. Ontology Development 101: A Guide to Creating Your First Ontology.

Pan, S.J. and Yang, Q. 2010. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10), pp. 1345–1359. doi: 10.1109/TKDE.2009.191.

Pauwels, P., Zhang, S. and Lee, Y.C. 2017. Semantic web technologies in AEC industry: A literature overview. *Automation in Construction* 73, pp. 145–165. doi: 10.1016/J.AUTCON.2016.10.003.

Pennington, J., Socher, R. and Manning, C.D. 2014. GloVe: Global Vectors for Word Representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 1532–1543. Available at: https://nlp.stanford.edu/projects/glove/ [Accessed: 25 April 2024].

Perozzi, B., Al-Rfou, R. and Skiena, S. 2014. DeepWalk: Online learning of social representations. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 701–710. Available at: https://dl.acm.org/doi/10.1145/2623330.2623732 [Accessed: 25 April 2024].

Plannerly. 2024. *ISO 19650 pdf explains the important steps in one diagram (includes ISO 19650 pdf download and high resolution image) - Plannerly*. Available at: https://plannerly.com/iso-19650-pdf/ [Accessed: 19 March 2024].

Preidel, C., International, A.B.-ISARC.P. of the and 2015, undefined. 2015. Automated code compliance checking based on a visual language and building information modeling. *researchgate.net*. Available at: https://www.researchgate.net/profile/Cornelius-

Preidel/publication/280316613_Automated_Code_Compliance_Checking_Based_on_a_Visual_Language_and_Building_Information_Modeling/links/55b240bd08ae9289a0852387/Automated-Code-Compliance-Checking-Based-on-a-Visual-Language-and-Building-Information-Modeling.pdf [Accessed: 25 April 2024].

Qiao, B., Zou, Z., Huang, Y., Fang, K., Zhu, X. and Chen, Y. 2022. A joint model for entity and relation extraction based on BERT. *Neural Computing and Applications* 34(5), pp. 3471–3481. Available at: https://link.springer.com/article/10.1007/s00521-021-05815-z [Accessed: 25 April 2024].

Qiu, S. et al. 2020. EasyAug: An Automatic Textual Data Augmentation Platform for Classification Tasks. *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020*, pp. 249–252. Available at: https://dl.acm.org/doi/10.1145/3366424.3383552 [Accessed: 25 April 2024].

Saaty, T.L. 1990. How to make a decision: The analytic hierarchy process. *European Journal of Operational Research* 48(1), pp. 9–26. doi: 10.1016/0377-2217(90)90057-I.

Sara Ismail, A., Nita Ali, K. and Iahad, N.A. 2017. A Review on BIM-Based Automated Code Compliance Checking System. In: *International Conference on Research and Innovation in Information Systems (ICRIIS)*. Langkawi, Malaysia: IEEE.

Sarı, R. and Pekeriçli, M.K. 2020. An investigation of comparison and evaluation of official BIM documents released in the USA, UK and Turkey. Available at: https://doi.org/10.31462/jcemi.2020.01067084www.goldenlightpublish.com [Accessed: 25 April 2024].

Saunders, M., Lewis, P. and Thornhill, A. 2019. Understanding Reseach Philisophy and Approaches to Theory Development. *Research Methods for Business Students*.

Schlichtkrull, M., Kipf, T.N., Amsterdam pbloem, V., Rianne van den Berg, vunl, Titov, I. and Welling, M. 2017. Modeling Relational Data with Graph Convolutional Networks Peter Bloem.

Schönfelder, P. and König, M. 2021. Deep Learning-Based Entity Recognition in Construction Regulatory Documents. In: *International Symposium on Automation and Robotics in Construction*. pp. 387–394.

Shanghai Municipal Commission of Housing and Urban-Rural Development. 2016. *Building Information Model Application Standard*. Available at: https://www.codeofchina.com/search/default.html?page=1&keyword=DG/TJ%2008-2201-2016 [Accessed: 25 April 2024].

Shi, P., Lin, J. and Cheriton, D.R. 2019. Simple BERT Models for Relation Extraction and Semantic Role Labeling. Available at: https://arxiv.org/abs/1904.05255v1 [Accessed: 25 April 2024].

Sun, M., Wang, J. and Zhang, D. 2020. The Progess That Natural Language Processing Has Made Towards Human-level AI. *Journal of Artificial Intelligence Practice* 3(1), pp. 38–47. Available at: http://www.clausiuspress.com/article/587.html [Accessed: 20 March 2024].

Tierney, P.J. 2012. A qualitative analysis framework using natural language processing and graph theory. *The International Review of Research in Open and Distributed Learning* 13(5), pp. 173–189. Available at: https://www.irrodl.org/index.php/irrodl/article/view/1240/2363 [Accessed: 25 April 2024].

Vahidnia, M.H., Vahidi, H., Hassanabad, M.G. and Shafiei, M. 2022. A Spatial Decision Support System Based on a Hybrid AHP and TOPSIS Method for Fire Station Site Selection. *Journal of Geovisualization and Spatial Analysis* 6(2), pp. 1–20. Available at: https://link.springer.com/article/10.1007/s41651-022-00125-x [Accessed: 25 March 2024].

Vaswani, A. et al. 2017. Attention is All you Need. *Advances in Neural Information Processing Systems* 30.

Veličković, P., Casanova, A., Liò, P., Cucurull, G., Romero, A. and Bengio, Y. 2017. Graph Attention Networks. *6th International Conference on Learning Representations,* *ICLR 2018 - Conference Track Proceedings*. Available at: https://arxiv.org/abs/1710.10903v3 [Accessed: 25 April 2024].

Venugopal, M., Eastman, C.M., Sacks, R. and Teizer, J. 2012. Semantics of model views for information exchanges using the industry foundation class schema. *Advanced Engineering Informatics* 26(2), pp. 411–428. doi: 10.1016/J.AEI.2012.01.005.

Walsham, G. 1995. The Emergence of Interpretivism in IS Research on JSTOR. *Information Systems Research* 6(4), pp. 376–394. Available at: https://www.jstor.org/stable/23010981 [Accessed: 25 April 2024].

Wang, D., Cui, P. and Zhu, W. 2016. Structural deep network embedding. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-August-2016, pp. 1225–1234. Available at: https://dl.acm.org/doi/10.1145/2939672.2939753 [Accessed: 25 April 2024].

Wang, H. and Yu, S. 2021. The Development of Natural Language Processing: Research on the Development of Electronic Information Engineering Technology in China. *The Development of Natural Language Processing: Research on the Development of Electronic Information Engineering Technology in China*, pp. 1–83. doi: 10.1007/978-981-16-1986-1/COVER.

Wang, X. and El-Gohary, N. 2021. Deep Learning-Based Named Entity Recognition from Construction Safety Regulations for Automated Field Compliance Checking. *Computing in Civil Engineering 2021 - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2021*, pp. 164–171. doi: 10.1061/9780784483893.021.

Wang, X. and El-Gohary, N. 2022. Deep Learning-Based Relation Extraction from Construction Safety Regulations for Automated Field Compliance Checking. *Construction Research Congress 2022: Computer Applications, Automation, and Data Analytics - Selected Papers from Construction Research Congress 2022* 2-B, pp. 290–297. doi: 10.1061/9780784483961.031.

Wang, X., Ji, H., Shi, C., Wang, B., Cui, P., Yu, P. and Ye, Y. 2019. Heterogeneous Graph Attention Network. Available at: https://www.imdb.com [Accessed: 25 April 2024].

Wang, X., Sun, Y., Chen, C. and Cui, J. 2022a. A Relation Extraction Model Based on BERT Model in the Financial Regulation Field. *2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology, CEI 2022*, pp. 496–501. doi: 10.1109/CEI57409.2022.9950087.

Wang, Z., Sacks, R. and Yeung, T. 2022b. Exploring graph neural networks for semantic enrichment: Room type classification. *Automation in Construction* 134, p. 104039. doi: 10.1016/J.AUTCON.2021.104039.

Watson, D. and Jones, A. 2013. Risk Management. *Digital Forensics Processing and Procedures*, pp. 109–176. Available at: https://linkinghub.elsevier.com/retrieve/pii/B9781597497428000054 [Accessed: 25 April 2024].

Weiss, H. 1987. Randolph Quirk/Sidney Greenbaum/Geoffrey Leech/Jan Svartvik, A Comprehensive Grammar of the English Language. *English World-Wide* 8(1), pp. 123–128. doi: 10.1075/EWW.8.1.10WEI.

Wikipedia. 2024a. *Natural language processing - Wikipedia*. Available at: https://en.wikipedia.org/wiki/Natural_language_processing [Accessed: 20 March 2024].

Wikipedia. 2024b. *Part of speech - Wikipedia*. Available at: https://en.wikipedia.org/wiki/Part_of_speech [Accessed: 20 March 2024].

Wikipedia. 2024c. *Syntactic parsing*. Available at: https://en.wikipedia.org/wiki/Syntactic_parsing_(computational_linguistics) [Accessed: 20 March 2024].

William, J. 1909. *The Meaning of Truth*. Harvard University Press. Available at: https://www.gutenberg.org/files/5117/5117-h/5117-h.htm [Accessed: 25 April 2024].

Wu, C., Li, X., Jiang, R., Guo, Y., Wang, J. and Yang, Z. 2023a. Graph-based deep learning model for knowledge base completion in constraint management of construction projects. *Computer-Aided Civil and Infrastructure Engineering* 38(6), pp. 702–719. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12904 [Accessed: 25 April 2024].

Wu, C., Li, X., Jiang, R., Guo, Y., Wang, J. and Yang, Z. 2023b. Graph-based deep learning model for knowledge base completion in constraint management of construction projects. *Computer-Aided Civil and Infrastructure Engineering* 38(6), pp. 702–719. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12904 [Accessed: 25 April 2024].

Xia, F. et al. 2021. Graph Learning: A Survey. *IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE* 00(0).

Xinhua Press. 2023. *Report on China's national economic, social development plan*. Available at: http://english.www.gov.cn/news/topnews/202303/16/content_WS64124df3c6d0f528699db539.html [Accessed: 18 March 2024].

Xu, M. 2020. Understanding Graph Embedding Methods and Their Applications. *Society for Industrial and Applied Mathematics* 63(4), pp. 825–853. Available at: https://doi.org/10.1137/20M1386062 [Accessed: 25 April 2024].

Xu, X. and Cai, H. 2020. Semantic approach to compliance checking of underground utilities. *Automation in Construction* 109, p. 103006. doi: 10.1016/J.AUTCON.2019.103006.

Xu, X. and Cai, H. 2021. Ontology and rule-based natural language processing approach for interpreting textual regulations on underground utility infrastructure. *Advanced Engineering Informatics* 48. doi: 10.1016/J.AEI.2021.101288.

Xu, Z. and Zhang, J. 2021. Extracting Keywords from Texts based on Word Frequency and Association Features. *Procedia Computer Science* 187, pp. 77–82. doi: 10.1016/J.PROCS.2021.04.035.

Xue, X. and Zhang, J. 2021. Part-of-speech tagging of building codes empowered by deep learning and transformational rules. *Advanced Engineering Informatics* 47. doi: 10.1016/J.AEI.2020.101235.

Yan, H. and Demian, P. 2008. Benefits and barriers of building information modelling. In: *Proc., 12th Int. Conf. on Computing in Civil and Building Engineering 2008*. Beijing: Loughborough University. Available at: /articles/conference_contribution/Benefits_and_barriers_of_building_information_modelling/9437141/1 [Accessed: 25 April 2024].

Yan, X., Ai, T., Yang, M. and Yin, H. 2019. A graph convolutional neural network for classification of building patterns using spatial vector data. *ISPRS Journal of Photogrammetry and Remote Sensing* 150, pp. 259–273. doi: 10.1016/J.ISPRSJPRS.2019.02.010.

Yilmaz, G., Akcamete, A. and Demirors, O. 2023. BIM-CAREM: Assessing the BIM capabilities of design, construction and facilities management processes in the construction industry. *Computers in Industry* 147, p. 103861. doi: 10.1016/J.COMPIND.2023.103861.

Yosinski, J., Clune, J., Bengio, Y. and Lipson, H. 2014. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems* 4(January), pp. 3320–3328. Available at: https://arxiv.org/abs/1411.1792v1 [Accessed: 25 April 2024].

Yurchyshyna, A. and Zarli, A. 2009. An ontology-based approach for formalisation and semantic organisation of conformance requirements in construction. *Automation in Construction* 18(8), pp. 1084–1098. doi: 10.1016/J.AUTCON.2009.07.008.

Zhang, J. and El-Gohary, N.M. 2015. Automated Information Transformation for Automated Regulatory Compliance Checking in Construction. *Journal of Computing in Civil Engineering* 29(4), p. B4015001. Available at: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CP.1943-5487.0000427 [Accessed: 25 April 2024].

Zhang, J., El-Gohary, N.M. and Asce, A.M. 2013. Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. *Journal of Computing in Civil Engineering* 30(2), p. 04015014. Available at: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CP.1943-5487.0000346 [Accessed: 25 April 2024].

Zhang, L. and Liu, B. 2014. Aspect and Entity Extraction for Opinion Mining. *Studies in Big Data* 1, pp. 1–40. Available at: https://link.springer.com/chapter/10.1007/978-3-642-40837-3_1 [Accessed: 25 April 2024].

Zhang, M.S. 2020. A Survey of Syntactic-Semantic Parsing Based on Constituent and Dependency Structures. *Science China Technological Sciences* 63(10), pp. 1898–1920. Available at: https://arxiv.org/abs/2006.11056v1 [Accessed: 20 March 2024].

Zhang, R. and El-Gohary, N. 2020. A Machine-Learning Approach for Semantically-Enriched Building-Code Sentence Generation for Automatic Semantic Analysis. *Construction Research Congress 2020: Computer Applications - Selected Papers from the Construction Research Congress 2020*, pp. 1261–1270. doi: 10.1061/9780784482865.133.

Zheng, Z., Zhou, Y.C., Lu, X.Z. and Lin, J.R. 2022. Knowledge-informed semantic alignment and rule interpretation for automated compliance checking. *Automation in Construction* 142, p. 104524. doi: 10.1016/J.AUTCON.2022.104524.

Zhenqing, Y., Pingping, S., Juanli, N., Xiaoling, Z. and Dandan, Z. 2016. The application and significance of BIM standards in the organisation. *Building technology* 47(8), pp. 691–693.

Zhong, B., Xing, X., Luo, H., Zhou, Q., Li, H., Rose, T. and Fang, W. 2020. Deep learning-based extraction of construction procedural constraints from construction regulations. *Advanced Engineering Informatics* 43, p. 101003. doi: 10.1016/J.AEI.2019.101003.

Zhou, J. et al. 2020. Graph neural networks: A review of methods and applications. *AI Open* 1, pp. 57–81. doi: 10.1016/J.AIOPEN.2021.01.001.

Zhou, P. and El-Gohary, N. 2017. Ontology-based automated information extraction from building energy conservation codes. *Automation in Construction* 74, pp. 103–117. doi: 10.1016/J.AUTCON.2016.09.004.

Zhou, S., Yu, B., Sun, A., Long, C., Li, J. and Sun, J. 2022a. A Survey on Neural Open Information Extraction: Current Status and Future Directions. *IJCAI International Joint Conference on Artificial Intelligence*, pp. 5694–5701. Available at: https://arxiv.org/abs/2205.11725v2 [Accessed: 25 April 2024].

Zhou, Y.C., Zheng, Z., Lin, J.R. and Lu, X.Z. 2022b. Integrating NLP and context-free grammar for complex rule interpretation towards automated compliance checking. *Computers in Industry* 142, p. 103746. doi: 10.1016/J.COMPIND.2022.103746.

Zhu, X., Li, H. and Su, T. 2023. Autonomous complex knowledge mining and graph representation through natural language processing and transfer learning. Available at: https://doi.org/10.1016/j.autcon.2023.105074 [Accessed: 25 April 2024].

Zhu, X., Li, H., Xiong, G. and Song, H. 2022. Automated Qualitative Rule Extraction Based on Bidirectional Long Short- Term Memory Model. pp. 227–237. doi: 10.7146/AUL.455.C213.

Appendix A - Part-of-Speech

Part of speech (POS) (Wikipedia 2024b) is a category of words that are classified according to their functions in sentences. The English parts of speech can be classified into nine main categories, which are nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions, interjections, articles, and determiners. A detailed list of POS tags can be found below (Chiche and Yitagesu 2022). Every word in English sentences falls into some of the nine parts of speech. The parts of speech belonging to the open class (nouns, verbs, adjectives, and adverbs) can be altered and added to as language develops, while the parts of speech of the closed class (pronouns, prepositions, conjunctions, articles/determiners, and interjections) are set in stone. Due to the above characteristics, the part of speech can not only indicate how the word functions grammatically within the sentence but also be used as labels for parsing. As a result, part of speech tags is widely involved in syntactic parsing methods.

The following list presents the all the part-of-speech tags and their corresponding meanings.

Number	Part-of-speech tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Predeterminer
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun

20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*to*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

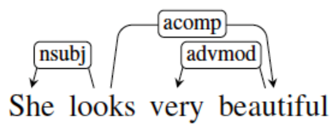
Appendix B - Dependency Parsing

Dependency grammar (DG) is a class of modern grammatical theories that are all based on the dependency relation, asserting that the root verb is the structural centre of clause structure, and all other syntactic units are either directly or indirectly connected to the verb in terms of the directed links, which are called dependencies. To provide a clear and simple description of these grammatical relationships, some dependency labels are created, which represents all sentence relationships uniformly as typed dependency relations. This Appendix aims to explain the definitions of dependency labels.

The current dependency labels cover approximately 50 grammatical relations, which are all represented as binary relations: a grammatical relation holds between a *governor* (also known as a regent or a head) and a *dependent*. The grammatical relations are defined below with the help of part-of-speech tags and phrasal labels, in alphabetical order according to the dependency's abbreviated name.

***acomp*: adjectival complement**

An adjectival complement of a verb is an adjectival phrase which functions as the complement (like an object of the verb).



***advcl*: adverbial clause modifier**

An adverbial clause modifier of a VP or S is a clause modifying the verb (temporal clause, consequence, conditional clause, purpose clause, etc.).

"The accident happened as the night was falling"	*advcl*(happened, falling)
"If you know who did it, you should tell the teacher"	*advcl*(tell, know)
"He talked to him in order to secure the account"	*advcl*(talked, secure)

advmod: adverb modifier

An adverb modifier of a word is a (non-clausal) adverb or adverb-headed phrase that serves to modify the meaning of the word.

"Genetically modified food"	*advmod*(modified, genetically)
"less often"	*advmod*(often, less)

agent: agent

An agent is the complement of a passive verb which is introduced by the preposition "by" and does the action. This relation only appears in the collapsed dependencies, where it can replace prep by, where appropriate. It does not appear in basic dependencies output.

"The man has been killed by the police"	*agent*(killed, police)
"Effects caused by the protein are important"	*agent*(caused, protein)

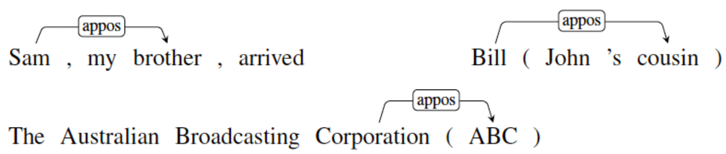
amod: adjectival modifier

An adjectival modifier of an NP is any adjectival phrase that serves to modify the meaning of the NP.

"Sam eats red meat"	*amod*(meat, red)
"Sam took out a 3 million dollar loan"	*amod*(loan, dollar)
"Sam took out a $ 3 million loan"	*amod*(loan, $)

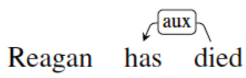
appos: appositional modifier

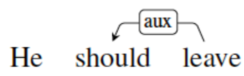
An appositional modifier of an NP is an NP immediately to the right of the first NP that serves to define or modify that NP. It includes parenthesized examples, as well as defining abbreviations in one of these structures.



***aux*: auxiliary**

An auxiliary of a clause is a non-main verb of the clause, e.g., a modal auxiliary, or a form of “be”, “do” or “have” in a periphrastic tense.





***auxpass*: passive auxiliary**

A passive auxiliary of a clause is a non-main verb of the clause which contains the passive information.

“Kennedy has been killed”	*auxpass*(killed, been) *aux*(killed,has)
“Kennedy was/got killed”	*auxpass*(killed, was/got)

***cc*: coordination**

A coordination is the relation between an element of a conjunct and the coordinating conjunction word of the conjunct. (Note: different dependency grammars have different treatments of coordination. We take one conjunct of a conjunction (normally the first) as the head of the conjunction.) A conjunction may also appear at the beginning of a sentence. This is also called a cc, and dependent on the root predicate of the sentence.

“Bill is big and honest”	*cc*(big, and)
“They either ski or snowboard”	*cc*(ski, or)
“And then we left.”	*cc*(left, And)

***ccomp*: clausal complement**

A clausal complement of a verb or adjective is a dependent clause with an internal subject which functions like an object of the verb, or adjective. Clausal complements for nouns are limited to complement clauses with a subset of nouns like “fact” or "report". Such clausal complements are usually finite (though there are occasional remnant English subjunctives).

"He says that you like to swim"	*ccomp*(says, like)
"I am certain that he did it"	*ccomp*(certain, did)
"I admire the fact that you are honest"	*ccomp*(fact, honest)

***conj*: conjunct**

A conjunct is the relation between two elements connected by a coordinating conjunction, such as "and", "or", etc. We treat conjunctions asymmetrically: The head of the relation is the first conjunct and other conjunctions depend on it via the conj relation.

"Bill is big and honest"	*conj*(big, honest)
"They either ski or snowboard"	*conj*(ski, snowboard)

***cop*: copula**

A copula is the relation between the complement of a copular verb and the copular verb.

"Bill is big"	*cop*(big, is)
"Bill is an honest man"	*cop*(man, is)

***csubj*: clausal subject**

A clausal subject is a clausal syntactic subject of a clause, i.e., the subject is itself a clause. The governor of this relation might not always be a verb: when the verb is a copular verb, the root of the clause is the complement of the copular verb. In the two following examples, "what she said" is the subject.

"What she said makes sense"	*csubj*(makes, said)
"What she said is not true"	*csubj*(true, said)

***csubjpass*: clausal passive subject**

A clausal passive subject is a clausal syntactic subject of a passive clause. In the example below, “that she lied” is the subject.

“That she lied was suspected by everyone”	*csubjpass*(suspected, lied)

***dep*: dependent**

A dependency is labelled as dep when the system is unable to determine a more precise dependency relation between two words.

“Then, as if to show that he could, . . . ”	*dep*(show, if)

***det*: determiner**

A determiner is the relation between the head of an NP and its determiner.

“The man is here”	*det*(man, the)
“Which book do you prefer?”	*det*(book, which)

***dobj*: direct object**

The direct object of a VP is the noun phrase which is the (accusative) object of the verb.

“She gave me a raise”	*dobj*(gave, raise)
“They win the lottery”	*dobj*(win, lottery)

***expl*: expletive**

This relation captures an existential “there”. The main verb of the clause is the governor.

“There is a ghost in the room”	*expl*(is, There)

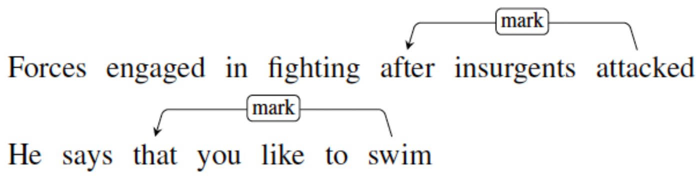
***iobj*: indirect object**

The indirect object of a VP is the noun phrase which is the (dative) object of the verb.

“She gave me a raise” *iobj*(gave, me)

mark: marker

A marker is the word introducing a finite clause subordinate to another clause. For a complement clause, this will typically be “that” or “whether”. For an adverbial clause, the marker is typically a preposition like “while” or “although”. The mark is a dependent of the subordinate clause head.



neg: negation modifier

The negation modifier is the relation between a negation word and the word it modifies.

“Bill is not a scientist” *neg*(scientist, not)
“Bill doesn’t drive” *neg*(drive, n’t)

nn: noun compound modifier

A noun compound modifier of an NP is any noun that serves to modify the head noun.

“Oil price futures” *nn*(futures, oil)
nn(futures, price)

npadvmod: noun phrase as adverbial modifier

This relation captures various places where something syntactically a noun phrase (NP) is used as an adverbial modifier in a sentence. These usages include: (i) a measure phrase, which is the relation between the head of an ADJP/ADVP/PP and the head of a measure phrase modifying the ADJP/ADVP; (ii) noun phrases giving an extent inside a VP which are not objects; (iii) financial constructions involving an adverbial or PP-like NP, notably the following construction $5 a share, where the second NP means "per share"; (iv) floating reflexives; and (v) certain other absolutive NP constructions. A temporal modifier (tmod) is a subclass of npadvmod which is distinguished as a separate relation.

"The director is 65 years old"	*npadvmod*(old, years)
"6 feet long"	*npadvmod*(long, feet)
"Shares eased a fraction"	*npadvmod*(eased, fraction)
"IBM earned $ 5 a share"	*npadvmod*($, share)
"The silence is itself significant"	*npadvmod*(significant, itself)
"90% of Australians like him, the most of any country"	*npadvmod*(like, most)

***nsubj*: nominal subject**

A nominal subject is a noun phrase which is the syntactic subject of a clause. The governor of this relation might not always be a verb: when the verb is a copular verb, the root of the clause is the complement of the copular verb, which can be an adjective or noun.

"Clinton defeated Dole"	*nsubj*(defeated, Clinton)
"The baby is cute"	*nsubj*(cute, baby)

***nsubjpass*: passive nominal subject**

A passive nominal subject is a noun phrase which is the syntactic subject of a passive clause.

"Dole was defeated by Clinton"	*nsubjpass*(defeated, Dole)

***num*: numeric modifier**

A numeric modifier of a noun is any number phrase that serves to modify the meaning of the noun with a quantity.

"Sam ate 3 sheep"	*num*(sheep, 3)
"Sam spent forty dollars"	*num*(dollars, 40)
"Sam spent $ 40"	*num*($, 40)

***number*: element of compound number**

An element of compound number is a part of a number phrase or currency amount. We regard a number as a specialized kind of multi-word expression.

"I have four thousand sheep"	*number*(thousand, four)
"I lost $ 3.2 billion"	*number*(billion, 3.2)

***parataxis*: parataxis**

The parataxis relation is a relation between the main verb of a clause and other sentential elements, such as a sentential parenthetical, a clause after a ":" or a ";", or two sentences placed side by side without any explicit coordination or subordination.

"The guy, John said, left early in the morning"	*parataxis*(left, said)
"Let's face it we're annoyed"	*parataxis*(Let, annoyed)

***pcomp*: prepositional complement**

This is used when the complement of a preposition is a clause or prepositional phrase (or occasionally, an adverbial phrase). The prepositional complement of a preposition is the head of a clause following the preposition, or the preposition head of the following PP.

"We have no information on whether users are at risk"	*pcomp*(on, are)
"They heard about you missing classes"	*pcomp*(about, missing)

***pobj*: object of a preposition**

The object of a preposition is the head of a noun phrase following the preposition, or the adverbs "here" and "there". (The preposition in turn may be modifying a noun, verb, etc.)

"I sat on the chair"	*pobj*(on, chair)

***poss*: possession modifier**

The possession modifier relation holds between the head of an NP and its possessive determiner, or a genitive 's complement.

"their offices"	*poss*(offices, their)
"Bill's clothes"	*poss*(clothes, Bill)

***possessive*: possessive modifier**

The possessive modifier relation appears between the head of an NP and the genitive's.

"Bill's clothes"	*possessive*(John, 's)

***preconj*: preconjunct**

A preconjunct is the relation between the head of an NP and a word that appears at the beginning bracketing a conjunction (and puts emphasis on it), such as "either", "both", "neither").

"Both the boys and the girls are here"	*preconj*(boys, both)

***predet*: predeterminer**

A predeterminer is the relation between the head of an NP and a word that precedes and modifies the meaning of the NP determiner.

"All the boys are here"	*predet*(boys, all)

***prep*: prepositional modifier**

A prepositional modifier of a verb, adjective, or noun is any prepositional phrase that serves to modify the meaning of the verb, adjective, noun, or even another preposition. In the collapsed representation, this is used only for prepositions with NP complements.

"I saw a cat in a hat"	*prep*(cat, in)
"I saw a cat with a telescope"	*prep*(saw, with)
"He is responsible for meals"	*prep*(responsible, for)

***prt*: phrasal verb particle**

The phrasal verb particle relation identifies a phrasal verb, and holds between the verb and its particle.

"They shut down the station"	*prt*(shut, down)

***punct*: punctuation**

This is used for any piece of punctuation in a clause, if punctuation is being retained in the typed dependencies. By default, punctuation is not retained in the output.

"Go home!"	*punct*(Go, !)

***quantmod*: quantifier phrase modifier**

A quantifier modifier is an element modifying the head of a QP constituent. (These are modifiers in complex numeric quantifiers, not other types of "quantification". Quantifiers like "all" become det.)

"About 200 people came to the party"	*quantmod*(200, About)

***rcmod*: relative clause modifier**

A relative clause modifier of an NP is a relative clause modifying the NP. The relation points from the head noun of the NP to the head of the relative clause, normally a verb.

"I saw the man you love"	*rcmod*(man, love)
"I saw the book which you bought"	*rcmod*(book,bought)

***ref*: referent**

A referent of the head of an NP is the relative word introducing the relative clause modifying the NP.

"I saw the book which you bought"	*ref*(book, which)

***root*: root**

The root grammatical relation points to the root of the sentence. A fake node "ROOT" is used as the governor.

"I love French fries."	*root*(ROOT, love)
"Bill is an honest man"	*root*(ROOT, man)

***tmod*: temporal modifier**

A temporal modifier (of a VP, NP, or an ADJP is a bare noun phrase constituent that serves to modify the meaning of the constituent by specifying a time.

"Last night, I swam in the pool"	*tmod*(swam, night)

***vmod*: reduced non-finite verbal modifier**

A reduced non-finite verbal modifier is a participial or infinitive form of a verb heading a phrase (which may have some arguments, roughly like a VP). These are used to modify the meaning of an NP or another verb. They are not core arguments of a verb or full finite relative clauses.

"Points to establish are . . . "	*vmod*(points, establish)
"I don't have anything to say to you"	*vmod*(anything, say)
"Truffles picked during the spring are tasty"	*vmod*(truffles, picked)
"Bill tried to shoot, demonstrating his incompetence"	*vmod*(shoot, demonstrating)

***xcomp*: open clausal complement**

An open clausal complement (xcomp) of a verb or an adjective is a predicative or clausal complement without its own subject. The reference of the subject is necessarily determined by an argument external to the xcomp (normally by the object of the next higher clause, if there is one, or else by the subject of the next higher clause. These complements are always non-finite, and they are complements (arguments of the higher verb or adjective) rather than adjuncts/modifiers, such as a purpose clause.

"He says that you like to swim"	*xcomp*(like, swim)
"I am ready to leave"	*xcomp*(ready, leave)
"Sue asked George to respond to her offer"	*xcomp*(ask, respond)
"I consider him a fool"	*xcomp*(consider, fool)
"I consider him honest"	*xcomp*(consider, honest)

***xsubj*: controlling subject**

A controlling subject is the relation between the head of a open clausal complement (xcomp) and the external subject of that clause. This is an additional dependency, not a basic dependency.

"Tom likes to eat fish"	*xsubj*(eat, Tom)

Appendix C - BIM Compliance Indicators

The following tables list the specific indicators and corresponding descriptions for BIM compliance checking. These indicators are presented in a hierarchical format according to their high-level domains.

Indicators for Information Model

Level 1	Level 2	Level 3	Description
Information model principles	capacity of information model		The information model should be able to carry the geometric information and non-geometric information
	consistency of mode information		Model data expressed in different ways should be consistent.
	interoperability of model information		The shared resource in the model should be able to be used by all relevant parties at all stages and tasks
	security of information model		The creation, application and management of information models should ensure information security.
	transmissibility of model information		The shared resources in the model should be able to be exchanged and shared between different stages, tasks
	uniqueness of model information		The information model shall have the uniqueness of the data source during the entire life of the project and shall be uniquely identified.
Model creation	level of definition		The LOD of the information model should be divided according to the work phase, and consistent with LOIN
	model consistency		There should be coordination and consistency between models created in different methods.
	model content	geometric attributes	Geometric attributes should include position information and size information
		non-geometric attributes	Non-geometric attributes should include identification information and design information
	model creation method		The information model can be created in an integrated way, or it can be created in a decentralized way by profession or task.
	model creation planning		Before the model is created, the types and quantities of models and sub-models should be planned
	model creation record		Information about the creation, modification and approval and the responsible person should be included in the information model.

	model creation scope		The scope of the information model creation should meet the project requirements of the corresponding phase.
	model data format	same or compatible data format	Information model creation should adopt the same or compatible data format
		data exchange standards or tools	When the data format is not compatible, it should be possible to realize data exchange through data exchange methods or tools.
	standard method and procedure		The relevant parties shall formulate a unified creation and management rules
model extension	extension requirement		
	extension consistency		The model extension should be coordinated with the original model structure.
	model extension method		The entity expansion method should be adopted to increase the types of model elements.
model structure	expandability		The model structure should be extensible.
	model level		The model system level should be organized in five levels: project, unit, discipline, component and equipment, rebar, and part.
	model structure composition		The model structure is composed of resource data, shared elements, and discipline elements
	mode structure requirement		Data of the information model must be able to be fully extracted and used
	openness		The model structure should be open.
model uses	access control		The management authority should be determined, and version control should be carried out for updates.
	data exchange and update		Methods can be used for model data exchange and update

Indicators for Collaborative Working

Level 1	Level 2	Level 3	Level 4	Description
collaborative working content	design stage collaboration	achievement quality control		Collaboration in the design phase should include the quality control of information models and related results.
		creation and coordination of information models		Collaboration in the design phase should include the creation and coordination of information models.

		docking with technologies of other disciplines		Collaboration in the design phase should include the docking with technologies of other disciplines.
		information model collaborative design	unit-level information model collaborative design	The collaborative design of the unit-level information model should include the coordination of the facility space between disciplines and the quality control process.
			component and equipment-level information model collaborative design	The collaborative design of component and equipment-level information models should include the following: (1) Coordinating the space usage requirements and positional relationship between the components and the steel bars, embedded parts, parts, etc. in the equipment; (2) Coordinating the connection relationship between the component and the equipment-level information model and the unit-level information model; (3) Coordinate the connection relationship between components and equipment-level information models and embedded parts.
			discipline-level information models	The collaborative design of discipline-level information models should include the following: (1) Coordinate the accuracy of design intent expression in the profession, the space usage requirements of components and equipment, and avoid the occurrence of conflicts in space usage; (2) Check the space use requirements of non-components and equipment in the discipline and complete the design quality control process of discipline-level information models.

			project-level information models	The collaborative design of project-level information models should include the following: (1) Associate the unit-level information model to the project-level information model, and coordinate its plane, elevation position and mutual relationship, etc.; (2) Check the correctness of the cross-monomer system connection relationship; (3) Complete the project-level information model design quality control process.
	construction stage collaboration			The coordination in the construction phase should be completed by all related parties based on the information model according to the construction process, and the content should include management of schedule, quality, safety, cost, procurement, personnel, materials, site, construction equipment, etc.
	operation stage collaboration			Collaboration in the operation and maintenance phase should be completed by all relevant parties in accordance with the management process based on the information model, and the content should include asset, equipment, facility, space, personnel, safety, quality, cost, risk management.
collaborative working method				Information model collaboration should use model collaboration, and when model collaboration cannot be achieved, data collaboration and file collaboration can be used.
collaborative working procedure	storage, update and backup requirements of files and data			Unified working rules should be established, which include storage, update and backup requirements of files and data

	communication and coordination rule			Unified working rules should be established, which include the method and content of labor division, data interaction and delivery of all related parties
	method and content of labour division, data interaction and delivery			Unified working rules should be established, which include the communication and coordination rules based on the information model of all relevant parties
common data environment	common data environment requirement	support the collaborative work method		The collaborative environment should be able to support the collaborative work method based on the information model, and realize the data collaboration and file collaboration of related information;
		realize the management and rights		The collaborative environment should be able to realize the management of information models and related information usage rights;
		support the exchange requirements		The collaborative environment should be able to support the exchange requirements of information model data;
		support software technology integration requirements		The collaborative environment should be able to support software technology integration requirements based on information models and with production management functions.
		support file version management		The collaborative environment should be able to support file version management and information sharing.
		ensure information security		The collaborative environment should be able to ensure information security.

Indicators for BIM Application

Level 1	Level 2	Level 3	Description
application content	application in design	clash detection	The information model application at the design stage should include collision inspection and pipeline synthesis
		construction plan simulation	The information model application at the design stage should include construction plan simulation
		drawing generation	The information model application at the design stage should include drawing generation
		plan comparison	The information model application at the design stage should include plan comparison
		quantity calculation	The information model application at the design stage should include quantity calculation
		site design	The information model application at the design stage should include site design
		virtual simulation	The information model application at the design stage should include virtual simulation
	application in construction	construction plan optimization	The information model application at the construction stage should include construction plan optimization
		construction resource management	The information model application at the construction stage should include construction resource management
		digital and assembly construction	The information model application at the construction stage should include digital and assembly construction
		dual product delivery	The information model application at the construction stage should include dual product delivery
		on-site quality monitoring	The information model application at the construction stage should include on-site quality monitoring
		on-site safety monitoring	The information model application at the construction stage should include on-site safety monitoring
		visual management of construction process	The information model application at the construction stage should include visual management of construction process
	application in operation	maintenance plan	The information model application at the operation stage should include maintenance plan
		project monitoring system	The information model application at the operation stage should include project monitoring system

		visual interactive inspection	The information model application at the operation stage should include visual interactive inspection
		visual operation management	The information model application at the operation stage should include visual operation management
BIM application principle	characteristics and needs		The application of the information model should be organized and carried out according to the characteristics and needs of each work stage.
	stage		The application of information models should be organized and carried out in stages.
	targeted manner		The application of information models should be organized and carried out in a targeted manner.
BIM software	BIM software testing		It is advisable to test and evaluate the technical level, data management and data interoperability of BIM software.
	software functions		BIM software should have corresponding professional functions and data interoperability functions

Indicators for Data Interoperability

Level 1	Level 2	Level 3	Description
classification and coding	IFD library		Specific "names" (types, attributes, etc.) in the project should follow the term dictionary
	classification		Projects should follow a unified classification system
	coding		Projects should follow unified coding rules
data exchange	data checking		The data check should include the following: 1 The data has been reviewed and cleaned up; 2 The data is a confirmed version; 3 Data content and format comply with data interoperability standards or data interoperability agreements.
	data exchange content		The application standards of a construction project should clarify the content and format of model data exchange.
	data exchange format		Data exchange should adopt an open data format
			When the open data format can't be used, the same or compatible data format agreed by all parties can be used.
	data exchange structure	core layer	The core framework and core extension objects should be defined and described at the core layer.

		domain layer	Data objects specific to the engineering field should be defined and described at the domain level
		interoperation layer	Objects shared by the engineering field should be defined and described in the interoperation layer
		resource layer	The most basic objects that are repeatedly referenced should be defined and described at the resource layer
data extension	core layer extension		Core layer data should expand according to data exchange requirements
	domain layer extension		Proprietary objects in the engineering domain should be expanded at the domain layer.
	interoperation layer extension		When the object that needs to be referenced is not defined, the interaction layer should be extended.
	resource layer extension		Resource layer data should expand according to data exchange requirements
data interoperability principles	data correctness		The format conversion of interoperable data should ensure the correctness.
	data format		Interoperable data should adopt the same format or compatible format;
	data integrity		The format conversion of interoperable data should ensure the integrity of the data.
data schema			It is recommended to use an open IFC data structure
data storage	data storage content	component and equipment-level	The storage of component and equipment-level information models should include the information model, data and documents of the component and equipment itself, as well as the association relationship with other-level information models.
			The storage of component and equipment-level information models should meet the relevant requirements of design work, and should include embedded parts information, hole information, and design quality control requirements within and between disciplines.
		discipline-level information model storage	the storage of discipline-level information models should include the discipline information models, data and documents, as well as the relationship with other-level information models.
			The storage of discipline-level information models should meet the requirements of discipline applications, and should include discipline quality control based on information model design work, construction plan simulation, etc.

		project-level information model storage	The project-level information model storage should include the associated information of the relevant single-level information model, quality management information, and related data and documents.
		rebar and part-level information model storage	The storage of rebar and part-level information models should include design information about component reinforcement and equipment parts processing, as well as the relationship with other levels of information models.
			The storage of reinforcement and part-level information models should meet the relevant requirements of design work, and should include data collaboration with calculation analysis models, construction drawing generation, engineering quantity statistics, design quality control, etc
		unit-level information model storage	The storage of a unit-level information model should include its own information model, data and documents, as well as the relationship with other-level information models.
			The storage of single-level information models should meet the requirements of discipline applications, and should include professional collaborative design based on information models, calculation and analysis of various disciplines, extraction of engineering quantities, professional quality control of design work, etc.
	data storage format		The storage of the information model should choose a data format that conforms to the data exchange structure, or an open data format according to application requirements.
	data storage requirement	data security	The storage of model data should meet the requirements of data security.
		mode information integrity	The storage method of the information model should not only preserve the complete information of the model itself, but also ensure the integrity of the associated information with other information models, related data and documents.

Indicators for Information management process

Level 1	Level 2	Level 3	Level 4	Description
assessment & need	5.1.1	AP information manager		An information manager shall be approved

		capability of information manager		Capability of information manager shall be assessed
		capacity of information manager		Capacity of information manager shall be assessed
	5.1.2	project information requirement		A complete set of project information requirements shall be established and take into consideration the information requirements which are needed at each key decision point
		project scope		The project scope shall be considered
		information purpose		The PIR shall detail the purpose for the information requirement
		plan of work		The information requirements shall list the project plan of work
		procurement route		Intended procurement route shall be detailed
		key decision points		Number of key decision points shall be included
		decisions		Required decisions at each key decision point shall be listed
		questions		Questions which require answers shall be considered
	5.1.3	milestones		The appointing party shall be aligned milestones with respect to the key decision points
		information delivery obligations		The appointing party shall be aligned its own information delivery obligations with the key decision points
		nature and substance of information delivery		The nature and substance of information delivery shall be aligned with the delivery milestone
		dates for milestone		The dates for each milestone shall be listed
	5.1.4	project specific standard		A project specific standard shall be produced
		information standard		Evidence that the information standard shall be accounted for the exchange of information: a.Within the appointing party's organisation b.Between the appointing party and external stakeholders c.Between the appointing party and external operators or maintainers d.Between the prospective lead appointed party and the appointing party e.Between prospective appointed

				parties on the same project f.Between independent projects.
		structuring and classifying method		A suitable method for structuring and classifying information shall be included
		level of information need		A suitable method for assignment of level of information need shall be included
		operational phase use method		A suitable method for use of information during the operational phase of the asset shall be included
	5.1.5	information production and procedures		Project specific information production and procedures which are required by the appointing party's organisation shall be produced.
		existing asset information capture		Project specific information production and procedures shall consider capture of existing asset information
		generation, review, or approval		Project specific information production and procedures shall consider generation, review, or approval of new information
		security or distribution		Project specific information production and procedures shall consider security or distribution of information
		delivery		Project specific information production and procedures shall consider delivery of information to the appointing party
	5.1.6	setup reference information and shared resources		Reference information and shared resources need to be setup
		open data standards		Open data standards shall be used
		existing asset information	within the appointing party's organisatio n	Existing asset information within the appointing party's organisation shall be considered
			from adjacent asset owners	Existing asset information from adjacent asset owners shall be considered
			under licence from external providers	Existing asset information under licence from external providers shall be considered

			within public libraries and other sources of historical records	Existing asset information from within public libraries and other sources of historical records shall be considered
		shared resources	process output templates	Process output templates shall be considered
			informatio n container templates	Information container templates shall be considered
			style libraries	Style libraries shall be considered
		library objects		Consideration shall be given to library objects within national and regional standards.
	5.1.7	common data environment		A common data environment shall be established for the projects.
		CDE establishment		A common data environment should be established before the invitation to tender stage
		unique ID		Each information container shall have a unique ID based upon an agree and documented convention comprised of field separated by a delimiter.
		field		Each field shall be assigned a value from an agreed and documented codification standard
		status attribute		Each information container shall have a status attribute
		revision attribute		Each information container shall have a revision attribute
		classification attribute		Each information container shall have a classification attribute
		transition		Each information container shall have the ability to transition between states
		record		Each information container shall record the name or the user and date when information container revision transition between each state
		controlled access		Each information container shall have controlled access
	5.1.8	information protocol		A project information protocol shall be established
		appointments		Appointments shall have the information protocol incorporated into them

		license agreements		Appointments shall have associated license agreements incorporated into them
		obligations		Obligations of appointing parties, prospective lead appointed parties and prospective appointed parties Warranties or liabilities associated to the project information model shall be included
		warranties or liabilities		Warranties or liabilities associated to the project information model shall be included
		intellectual property rights		Background and foreground intellectual property rights of information shall be included
		Use of existing asset information		Use of existing asset information shall be included
		Use of shared resources		Use of shared resources shall be included
		use of information during the project		Use of information during the project, including any licensing terms shall be included
		Re-use of information		Re-use of information following the appointment or in the event of termination shall be included.
Invitation to Tender	5.2.1	exchange information requirements		Appointing party's exchange information requirements should be established
		EIR consideration		Appointing party's information requirements shall consider: — organizational information requirements, — asset information requirements, and — project information requirements;
		level of information need		The level of information need shall be established
		acceptance criteria		Establishing the acceptance criteria for each information requirement shall consider: — the project's information standard, — the project's information production methods and procedures, and — the use of reference information or shared resources provided by the appointing party;

		supporting information		Establishing the supporting information that the prospective lead appointed party might need shall consider: — existing asset information, — shared resources, — supporting documents or guidance material, — references to relevant international, national or industry standards, and — exemplars of similar information deliverables;
		dates		Establishing the dates, relative to the project's information delivery milestones and appointing party's key decision points shall consider: — the time needed by the appointing party to review and accept information, and — the appointing party's internal assurance processes.
	5.2.2	assemble the reference information		The appointing party shall assemble the reference information or shared resources that they intend to provide to the prospective lead appointed party.
		reference information or shared resources		Reference information or shared resources identified during project initiation shall be considered.
		previous information		Information generated during previous stages of the project shall be considered.
		suitability		The suitability for which the information can be used by the prospective lead appointed party shall be considered.
	5.2.3	tender response requirements		The appointing party shall establish tender response requirements and evaluation criteria
		BIM execution plan		The contents of the delivery team's (pre-appointment) BIM execution plan shall be considered
		competency of individuals		The competency of the prospective individuals undertaking the information management function on behalf of the delivery team shall be considered.
		capability and capacity assessment		The prospective lead appointed party's assessment of the delivery team's capability and capacity shall be considered.

		mobilization plan		The delivery team's proposed mobilization plan shall be considered.
		risk assessment		The delivery team's information delivery risk assessment shall be considered.
	5.2.4	compile the information		The appointing party shall compile the information to be included within the invitation to tender package.
		consider EIR		The appointing party's exchange information requirements shall be considered.
		consider reference information and shared resources		The relevant reference information and shared resources shall be considered.
		tender response requirements and evaluation criteria		The tender response requirements and evaluation criteria (if applicable) shall be considered.
		consider milestones		The project information delivery milestones shall be considered.
		consider information standard		The project's information standard shall be considered.
		consider production methods and procedure		The project's information production methods and procedure shall be considered.
		information protocol		The project's information protocol shall be considered.
Tender Repsonse	5.3.1	LaP information manager		An information manager shall be successfully appointed
		AP EIR consideration		The appointing party's exchange information requirements shall be considered
		responsible task		The responsible tasks shall be accounted for
		authority		Authority to the appointed party shall be given
		competency consideration		Individual's competency at undertaking tasks set shall be considered
		probity arrangements		Probity arrangements for potential conflicts of interest shall be considered
	5.3.2	LaP BIM execution plan		Lead appointed party shall establish the delivery team's (pre-appointment) BIM execution plan

		LaP names and resumes		Information management functions shall require names and resumes of all actors
		LaP delivery strategy		Production of delivery strategy for information shall consider: — the delivery team's approach to meeting the exchange information requirements, — a set of objectives/goals for the collaborative production of information, — an overview of the delivery team's organizational structure and commercial relationships — an overview of the delivery team's composition
		LaP federation strategy		Production of a federation strategy to be adopted
		LaP responsibility matrix		Production of delivery team high level responsibility matrix — the allocated responsibility for each element of the information model — the key deliverables associated to each element
		LaP additions or amendments to SMP		Production of additions or amendments to the project's information production methods and procedures
		additions or amendments to information standard		Production of additions or amendments to the project's information standard
		LaP schedule of software		Production of a schedule of software, hardware, and IT infrastructure
	5.3.3	TT assessment		Task team shall undertake an assessment of their capability and capacity in accordance with: — appointing party's exchange information requirements — delivery team's proposed (pre-appointment) BIM execution plan
		TT capability to manage information		Capability and capacity to manage information shall be considered: — the relevant experience and number of task team members — the relevant education and training available to task team members
		TT capability to producee information		Capability and capacity to produce information — the relevant experience and number of task team members

				— the relevant education and training available to task team members
		TT availability of information technology		Availability of information technology (IT) within the task team shall consider: — the proposed IT schedule — the specification and quantity of the task team's hardware; — the architecture, maximum capacity and current utilization of the task team's IT infrastructure; — the associated support and service level agreements available to the task team.
	5.3.4	managing and producing information summary		Each prospective lead appointing party shall provide a summary of task team capabilities in managing and producing information
		timely delivery of information summary		Each prospective lead appointing party shall provide a summary of task team capacity for timely delivery of information
	5.3.5	mobilisation plan		A mobilisation plan shall be established to be initiated and implemented during the mobilisation.
		information production methods and procedures test		Lead appointed party shall consider their approach, timescales and responsibilities for testing and documenting the proposed information production methods and procedures
		information exchanges test		Lead appointed party shall consider their approach, timescales, and responsibilities for testing the information exchanges between task teams
		information delivery test		Lead appointed party shall consider their approach, timescales, and responsibilities for testing the information delivery to the appointing party
		CDE test		Lead appointed party shall consider their approach, timescales, and responsibilities for configuring and testing the project's CDE
		connectivity test		Lead appointed party shall consider their approach, timescales, and responsibilities for configuring and testing the delivery team's (distributed) CDE and its connectivity to the project CDE (if applicable)

		additional IT infrastructure test		Lead appointed party shall consider their approach, timescales, and responsibilities for procuring, implementing, configuring and testing additional software, hardware and IT infrastructure
		shared resources development		Lead appointed party shall consider their approach, timescales, and responsibilities for developing additional shared resources to be used by the delivery team
		education delivery		Lead appointed party shall consider their approach, timescales, and responsibilities for developing and delivering education (knowledge required) to delivery team members
		training delivery		Lead appointed party shall consider their approach, timescales, and responsibilities for developing and delivering training (skills required) to the delivery team members
		recruiting additional members		Lead appointed party shall consider their approach, timescales, and responsibilities for recruiting additional members of the delivery team to achieve the required capacity
		supporting		Lead appointed party shall consider their approach, timescales, and responsibilities for supporting individuals and organizations that join the delivery team during the appointment
	5.3.6	risk register		A risk register shall be produced indicating risks associated with timely delivery of information
		assumptions risk		Lead appointed party shall consider risks associated with assumptions the delivery team has made in relation to the appointing party's exchange information requirements
		milestones risk		Lead appointed party shall consider risks associated with meeting the appointing party's project information delivery milestones
		information protocol risk		Lead appointed party shall consider risks associated with the contents of the project's information protocol
		delivery strategy risk		Lead appointed party shall consider risks associated with achieving the proposed information delivery strategy

		production methods and procedures risk		Lead appointed party shall consider risks associated with adopting the project's information standard and information production methods and procedures
		information standard risk		Lead appointed party shall consider risks associated with inclusion (or non-inclusion) of proposed amendments to the project's information standard
		mobilization risk		Lead appointed party shall consider risks associated with the mobilization of the delivery team to achieve the required capability and capacity
	5.3.7	compilation		A compilation of the delivery team's tender response shall be completed
		BEP compilation		Lead appointed party shall compile (pre-appointment) BIM execution plan
		capability and capacity compilation		Lead appointed party shall compile capability and capacity assessment summary
		mobilization plan compilation		Lead appointed party shall compile mobilization plan
		risk assessment compilation		Lead appointed party shall compile information delivery risk assessment
Appointment	5.4.1	LaP BEP confirmation		Delivery team's BIM execution plan shall be confirmed by lead appointed party
		LaP information management individual confirmation		Lead appointed party shall confirm the names of the individual(s) who will undertake the information management function
		delivery strategy update		Lead appointed party shall update the delivery team's information delivery strategy
		responsibility matrix update		Lead appointed party shall update the delivery team's high-level responsibility matrix
		production methods and procedures confirmation		Lead appointed party shall confirm and document the delivery team's proposed information production methods and procedures
		additions or amendments agreement		Lead appointed party shall agree with the appointing party any additions or amendments to the project's information standard
		schedule of software confirmation		Lead appointed party shall confirm the schedule of software, hardware, and IT infrastructure the delivery team will use

	5.4.2	responsibility matrix refinement		The lead appointed party shall further refine the high-level responsibility matrix to establish the detailed responsibility matrix, which identifies: — what information is to be produced; — when the information is to be exchanged and with whom; and — which task team is responsible for its production.
		LaP milestones consideration		The information delivery milestones shall be considered by the lead appointed party
		LaP responsibility matrix consideration		The high-level responsibility matrix shall be considered by the lead appointed party
		LaP production methods and procedures consideration		The project's information production methods and procedures shall be considered by the lead appointed party
		container structure consideration		The elements of information container breakdown structure allocated to each task team shall be considered by the lead appointed party
		production process consideration		The dependencies on the information production process shall be considered by the lead appointed party
	5.4.3	LaP EIR		The lead appointed party shall establish their exchange information requirements for each appointed party.
		information requirement definition		The lead appointed party shall define each information requirement, and in doing so shall consider: — the appointing party's information requirements, which the lead appointed party requires the appointed party to meet, and — any additional information requirements that the lead appointed party requires the appointed party to meet
		LaP level of information need		The lead appointed party shall establish the level of information need required to meet each information requirement

		LaP acceptance criteria		The lead appointed party shall establish the acceptance criteria for each information requirement, and in doing so shall consider: — the project's information standard, — the project's information production methods and procedures, and — the use of reference information or shared resources provided by the appointing party or lead appointed party
		LaP dates		The lead appointed party shall establish the dates that need to be met for each requirement, relative to the project's information delivery milestones, and in doing so shall consider: — the time needed by the lead appointed party to review and authorize information, and — the lead appointed party's internal assurance processes
		LaP supporting information		The lead appointed party shall establish the supporting information that the appointed party might need, to fully understand or evaluate each information requirement or its acceptance criteria, and in doing so shall consider: — existing asset information, — shared resources, — supporting documents or guidance material, — references to relevant international, national or industry standards, and — exemplars of similar information deliverables.
	5.4.4	TIDP		A task information delivery plan (TIDP) shall be established by each task team and maintain throughout task team's appointment
		TT milestones consideration		The project's information delivery milestones shall be considered by each task team
		responsibility matrix consideration		The task team's responsibilities within the detailed responsibility matrix shall be considered by each task team
		information requirements consideration		The lead appointed party's information requirements shall be considered by each task team

		shared resources consideration		The availability of shared resources within the delivery team shall be considered by each task team
		time		The time the task team will need to produce (generate, coordinate, review and approve) information shall be considered by each task team
		content of TIDP		For each information container, the TIDP shall list and identify: the name and title; the predecessors or dependencies; the level of information need; the (estimated) production duration; the information author responsible for its production; and the delivery milestones.
	5.4.5	LaP MIDP		The lead appointed party shall aggregate the TIDP from each task team to establish the delivery team's master information delivery plan (MIDP).
		LaP responsibility consideration		The assigned responsibilities within the detailed responsibility matrix shall be considered by the lead appointed party
		predecessors or dependencies consideration		The information predecessors or dependencies on information between task teams shall be considered by the lead appointed party
		LaP time		The time the lead appointed party will need to review and authorize the information model shall be considered by the lead appointed party
		AP time		The time the appointing party will need to review and accept the information model shall be considered by the lead appointed party
		MIDP content		The lead appointed party shall baseline the deliverables and dates within the MIDP
		TIDP changes		The lead appointed party shall inform each task team and notify if any changes are required to the TIDP
		risks or issues		The lead appointed party shall inform the appointing party of any risks or issues which could impact on the project's information delivery milestones.
	5.4.6	AP appointment documents		The appointing party shall take account of the following, in that they are included within the completed appointment documents for the lead appointed party and managed via

				change control throughout the duration of the appointment: — the appointing party's exchange information requirements; — the project's information standard (including any agreed additions or amendments); — the project's information protocol (including any agreed additions or amendments); — the delivery team's BIM Execution plan; and — the delivery team's MIDP.
	5.4.7	LaP appointment documents		The lead appointed party shall take account of the following, in that they are included within the appointment documents for each appointed party and managed via change control throughout the duration of the appointment: — the lead appointed party's exchange information requirements; — the project's information standard (including any agreed additions or amendments) **(see 5.1.4);** — the project's information protocol (including any agreed additions or amendments); — the delivery team's BIM Execution plan; and — the agreed TIDP.
Mobilization	5.5.1	mobilize resources		The lead appointed party shall mobilize the resources, as defined within the delivery team's mobilization plan **(5.3.5).**
		resource availability confirmation		The lead appointed party shall confirm the resource availability of each task team
		develop and deliver education		The lead appointed party shall develop and deliver education on topics such as the project's scope, exchange information requirements and delivery milestones (knowledge required) to delivery team members
		develop and deliver training		The lead appointed party shall develop and deliver training (skills required) to the delivery team members
	5.5.2	mobilize the information technology		The lead appointed party shall mobilize the information technology, as defined within the delivery team's mobilization plan **(5.3.5).**

		IT infrastructure		The lead appointed party shall procure, implement, configure, and test software, hardware and IT infrastructure (as required)
		CDE		The lead appointed party shall configure and test the project's CDE **in accordance with 5.1.7**
		connectivity		The lead appointed party shall configure and test the delivery team's (distributed) CDE and its connectivity to the project CDE (if applicable) in accordance with **5.1.7**
		information exchanges		The lead appointed party shall test the information exchanges between task teams
		information delivery		The lead appointed party shall test the information delivery to the appointing party
	5.5.3	information production methods and procedures test		The lead appointed party shall test the project's information production methods and procedures, as defined within the delivery team's mobilization plan **(5.3.5).**
		information production methods and procedures document		The lead appointed party shall test and document the project's information production methods and procedures
		refine information container structure		The lead appointed party shall refine and verify the proposed information container breakdown structure is workable
		develop shared resources		The lead appointed party shall develop shared resources to be used by the delivery team
		information production methods and procedures communication		The lead appointed party shall communicate the project's information production methods and procedures to all task teams
Collaborative production of information	5.6.1	access check		Prior to generating information, each task team shall check that they have access to the relevant reference information and shared resources within the project's common data environment
		inform LaP		If a task team doesn't have access to the relevant reference information and shared resources within the project's common data environment, the task team shall inform the lead appointed

				party and assess the potential impact that this could have on the TIDP.
	5.6.2	TIDP accordance		Each task team shall generate information in accordance with their respective TIDP
		information generation		The task team shall generate information: — in compliance with the project's information standard, and — in accordance with the project's information production methods and procedures
		unnecessary information		The task team shall not generate information that: — exceeds the required level of information need, — extends beyond the allocated element of the information container breakdown structure, — duplicates information generated by other task teams, or — contains superfluous detail;
		cross-reference		The task team shall coordinate and cross-reference all information with information shared within the project's common data environment, in accordance with the project's information production methods and procedures
		spatially coordinate		The task team shall spatially coordinate geometrical models with other geometrical models shared with the appropriate suitability, residing within the project's common data environment
		coordination resolution		In the event of a coordination issue, the relevant task teams shall collaborate to identify a possible resolution. If a resolution cannot be found the task teams shall notify the lead appointed party
	5.6.3	quality assurance check		Each task team shall undertake a quality assurance check of each information container, in accordance with the project's information production methods and procedures, prior to undertaking a review of the information within it (5.6.4).
		information container check		The task team shall check the information container in accordance

				with the project's information standard.
		successful container check		If the check is successful, the task team shall: — mark the information container as checked, and — record the outcome of the check
		unsuccessful container check		If the check is unsuccessful, the task team shall: — reject the information container, and — inform the information author of the outcome and corrective action required.
	5.6.4	TT information review		Each task team shall undertake a review of the information within the information container prior to sharing within the project's common data environment.
		TT information requirement consideration		The lead appointed party's information requirements shall be considered by the task team
		TT level of information need consideration		The level of information need shall be considered by the task team
		information needed for coordination		Information needed for coordination by other task teams shall be considered by the task team
		successful information check		If the review is successful, the task team shall: — assign the suitability for which the information contained within the information container can be used, and — approve the information container for sharing
		unsuccessful information check		If the review is unsuccessful, the task team shall: — record why the review was unsuccessful, — record any amendments for the task team to complete, and — reject the information container.
	5.6.5	TT model review		The delivery team shall undertake a review of the information model, in accordance with the project's information production methods and procedures, to facilitate the continuous coordination of information across each element of the information model

		acceptance criteria consideration		The appointing party's information requirements and acceptance criteria shall be considered the delivery team
		information containers		The information containers listed within the master information delivery plan shall be considered the delivery team
Information model delivery	5.7.1	authorization		Prior to the delivery of the information model to the appointing party, each task team shall submit their information to the lead appointed party for authorization within the project's common data environment.
	5.7.2	LaP review		The lead appointed party shall undertake a review of the information model in accordance with the project's information production methods and procedures.
		LaP MIDP consideration		The deliverables listed in the master information delivery plan shall be considered by the lead appointed party
		AP EIR consideration		The appointing party's exchange information requirements shall be considered by the lead appointed party
		LaP EIR consideration		The lead appointed party's exchange information requirements shall be considered by the lead appointed party
		LaP acceptance criteria		The acceptance criteria for each information requirement shall be considered by the lead appointed party
		LaP level of information need consideration		The level of information need for each information requirement shall be considered by the lead appointed party
		LaP acceptation		If the review is successful, the lead appointed party shall authorize the information model and instruct each task team to submit their information for appointing party acceptance within the project's common data environment.
		LaP rejection		If the review is unsuccessful, the lead appointed party shall reject the information model and instruct the task teams to amend the information and re-submit for lead appointed party authorization.
	5.7.3	Task team submission		Each task team shall submit their information for appointing party review and acceptance within the project's common data environment

	5.7.4	AP review		The appointing party shall undertake a review of the information model in accordance with the project's information production methods and procedures.
		AP MIDP consideration		The deliverables listed in the master information delivery plan shall be considered by the appointing party
		AP EIR consideration		The appointing party's exchange information requirements shall be considered by the appointing party
		AP acceptance criteria		The acceptance criteria for each information requirement shall be considered by the appointing party
		AP level of information need consideration		The level of information need for each information requirement shall be considered by the appointing party
		AP acceptation		If the review is successful, the appointing party shall accept the information model as a deliverable within the project's common data environment.
		AP rejection		If the review is unsuccessful, the appointing party shall reject the information model and instruct the lead appointed party to amend the information and re-submit for appointing party's acceptance.
Project close-out	5.8.1	archive		The appointing party shall archive the information containers within the project's common data environment in accordance with the project's information production methods and procedures.
		information container		Information containers will be needed as part of the asset information model
		future access and reuse		Future access requirements and future re-use
		retention policies		Relevant retention policies to be applied
	5.8.2	lessons learned		The appointing party shall capture lessons learned during the project and record them in a suitable knowledge store

Indicators for Handover

Level 1	Level 2	Level 3	Level 4	Description
deliverable	bill of quantities	document purpose		Bill of quantities should include document purpose
		quantity and coding		Bill of quantities should include quantity and coding
	building index table	index name and coding		Building index table should include index name and coding
		index value		Building index table should include index value
	deliverable list			The deliverable list is included in deliverables
	delivered information model	model creation time		Model creation time should be included in the information model
		model creator		Model creator should be included in the information model
		model expression		Model expression should be included in the information model
		model reviewer		Model reviewer should be included in the information model
		model review time		Model review time should be included in the information model
		model updater		Model updater should be included in the information model
		model update time		Model update time should be included in the information model
		software version		Software version should be included in the information model
		state of ownership		State of ownership should be included in the information model
	execution plan	categories of deliverable		Categories of deliverable should be included in the information model
		geometric expression accuracy		Geometric expression accuracy should be included in the information model
		information depth		Information depth should be included in the information model
		model fineness description		Model fineness description should be included in the information model
		naming classification and coding rule		Naming classification and coding rule should be included in the information model
		project description		Project description should be included in the information model
		resource allocation		resource allocation should be included in the information model

		software and hardware working environment		Software and hardware working environment should be included in the information model
	information model description			Information model description should be included in the information model
	model property data	component and equipment property data		The component and equipment-level attribute data should include the main design indicators of the component and equipment.
		discipline-level property data		The discipline-level attribute data should include the main design indicators of the professional system.
		project-level property data	coordinate system	Project-level attribute data should include coordinate system
			elevation system	Project-level attribute data should include elevation system.
			main economic indicators	Project-level attribute data should include main economic indicators
			main technical indicators	Project-level attribute data should include main technical indicators
			project scope	Project-level attribute data should include construction scale
			project site	Project-level attribute data should include construction site
			unit system	Project-level attribute data should include unit system
		rebar and part-level property data		Rebar and part-level attribute data should include the main design indicators of rebar and parts in components and equipment.
		unit-level property data		The unit-level attribute data should include the main technical indicators of the unit.
	project requirements	application demand		The project requirements should include the application demand of the building information model
		collaboration methods		The project requirements document should include the collaboration method of the project participants
		data access method		The project requirements should include the data access methods of the project participants

		data access permissions		The project requirements should include the data access permissions of the project participants
		data storage method		The project requirements should include the data storage methods of the project participants
		deliverables type		The project requirements should include the category of deliverables
		delivery method		The project requirements should include the delivery method of the project deliverables
		model ownership		The project requirements should include the ownership of the project building information model
		project plan	project coordinate	The project plan should include project coordinates
			project elevation	Project plan should include project elevation
			project location	The project plan should include the project location
			project scale	The project plan should include the size of the project
			project type	The project plan should include the project type
	supplementary documents			Model-related supplementary documents should include technical documents supplementing model information, expanded model accuracy grade tables, attribute information documents
deliverable management				Deliverables should be managed centrally and set data access permissions.
handover data format	file association			The attribute data in the information model should be associated with the information model.
	Independent file			The attribute data in the information model should be organized in the form of independent files.
	Principle of least			The attribute data in the information model should be based on the principle of least necessary
handover principles	data relevance			The deliverable should maintain effective data association during the delivery process
	file naming			The naming of the file should conform to the naming rules

	handover completeness			Information model delivery should ensure the integrity of the delivered data
	handover consistency			The content and format of the data information should be consistent before and after delivery.
	handover correctness			Information model delivery should ensure the correctness of the delivered data
	Intellectual property right			The deliverables of the information model should protect the intellectual property rights of all relevant parties
level of information need	concept design stage			The main information model major in the feasibility study stage should include four levels: project-level information model, unit-level information model, discipline-level information model, and component and equipment-level information model.
	detail design stage			The information model of the construction drawing design stage should include five levels: project-level information model, unit-level information model, discipline-level information model, component and equipment-level information model, and rebar and part-level information model.
	preliminary design stage			The preliminary design stage information model should include four levels: project-level information model, unit-level information model, discipline-level information model, and component and equipment-level information model.
	level of information need requirement	project requirement		In each stage, the corresponding model depth level should be specified in detail according to the needs of the construction project.
		relative standards		In each stage, the corresponding model depth level shall be specified in detail according to the requirements of relevant standards.