Advanced Engineering Informatics 59 (2024) 102290

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

Semantic correction, enrichment and enhancement of social and transport infrastructure BIM models

Nicholas Nisbet *, Zijing Zhang, Ling Ma, Weiwei Chen, Mustafa Selçuk Çıdık

Bartlett School of Sustainable Construction, University College London, UK

A B S T R A C T

The use of Building Information Modelling (BIM) models in the design, construction and operation of buildings and infrastructure is leading to a stronger focus on the quality of the models. Models may need correction, enrichment or enhancement to meet the expectations for quality and completeness, especially if models are to be taken as legal documents, for example for regulatory approval. Past work on semantic development has looked at specific scenarios such as scanned geometry or missing classification. This paper describes an innovative unified approach to the documentation of semantic expectations by actors in the AECO (Architectural, Engineering, Construction and Operations) domain and the means to put them into effect. RASE (Requirements, Applications, Selections and Exceptions) semantic mark-up is used to make both the requirements and any supporting resources both human-readable and machine-operable. Two example models from industry, a motorway bridge and a healthcare space, are used to demonstrate applying geometric, schema and classification knowledge. This knowledge is represented in a number of different styles. This extends our understanding of the nature of the knowledge found in dictionaries, classifications and development specifications, demonstrating how this knowledge can be made operable. This bridges the gap between the application of static compliance knowledge and the accurate and efficient application of correction, enrichment and enhancement knowledge.

1. Introduction

The use of Building Information Modelling (BIM) models in the design, construction and operation of buildings and infrastructure is leading to a stronger focus on the quality of the models. For example, ISO19650 Part 4 [12] seeks to standardise on six priority quality aspects, emphasising the importance of information quality for the immediate, medium term and long-term use. For many reasons, including the focus of authoring tools on interacting with individual instances at the lowest level, models may be inconsistent or incomplete, needing semantic development for quality and completeness so that they can be taken as legal documents for contractual and regulatory purposes [2]. Semantic development includes semantic enhancement, semantic enrichment and semantic correction. Bloch [2] makes the distinction that “while [semantic enhancement] relies on … information stored in external data sources, [semantic enrichment] is concerned with retrieving new information from the original model …”. This paper adds ‘semantic correction’ for situations where errors are detected in the model. Semantic development can be used to create a revised model as a precursor to subsequent processes such as automated analysis or code compliance checking. Alternatively, semantic development could be integral with those processes, temporarily modifying or extending the model for the duration of the analysis or checking process.

Previous work on semantic development has used specialized databases [14] to represent expectations for semantic development. These have been mostly automated or semi-automated inference tools where rule-based checks (semi-)automatically generate new information in the target model (i.e., inferential tools). Sacks et al. [19] stated that one of the main challenges of this approach is the reliance on individual intuition and subjective experiences generally collected through interviews: as these rules need to cover the exact circumstances in which they are applicable, any alternative situations, and any exceptions along with the specific expectations for semantic development, collecting and collating them through interviews or similar methods proved to be unreliable for several use cases.

To bridge the gap between static knowledge and applications that allow AECO actors to apply the knowledge accurately and efficiently, this paper explores the efficacy of a novel approach to semantic development.

The paper extends RASE (Requirements, Applications, Selections and Exceptions) from representing normative content [15], into documenting definitive content about semantic development by using semantic mark-up. The proposed approach uses a unified method to the recording of expectations for semantic development as ‘Registrations’ of definitions and the means to put them into effect. The paper explores several different ways in which external knowledge about the expectations for

* Corresponding author.
E-mail address: n.nisbet@ucl.ac.uk (N. Nisbet).

https://doi.org/10.1016/j.aei.2023.102290
Received 3 June 2023; Received in revised form 23 November 2023; Accepted 24 November 2023
Available online 4 December 2023
1474-0346/

the content of descriptive and narrative models can be made accessible both for human development and for machine implementation. Overall, it extends our understanding of the nature of the knowledge found in dictionaries, classifications and development specifications, and how this knowledge can be made operable.

The paper applies the proposed approach to two example models from the UK construction industry (a motorway bridge and a healthcare space), where semantic development will be applied automatically and in bulk, so as to achieve a consistent standard of accuracy. This paper outlines the state of the art in semantic development in Section 2, the proposed method in Section 3 and the example expectations in Section 4. Section 5 describes the implementation and Sections 6 and 7 offer a discussion and conclusions respectively.

2. State of the art

Semantic development can be contrasted with automated validation and compliance checking. Several applications for the latter have sought to 'validate' information deliveries against requirements such as schema compliance [4], or explicit exchange information requirements [11]. Exchange requirements can be represented in technical specification formats such as mvdXML or idsXML [5] or interpreted as rules for commercial tools such as Solibri [20] or Verifi3d [23]. A UK digital compliance project [8] demonstrated automated code compliance checking on three major sections of the English Approved Documents. Various NLP approaches to automated compliance checking have been proposed to obtain the semantic meaning of individual sentences. Given sufficient training resources, some results can be obtained which can be compared against sentence templates to obtain the logical formulation needed, for example [25] but these are not applicable to whole documents with titles, paragraphs, tables and lists. Furthermore, there is no limit to the number of templates that may be required. In 2021 the same authors reported [17–26] that NLP had around twice the Type-1 error rate of rule-based approaches. buildingSMART [3] has published an initial investigation touching upon issues around accuracy and credibility of rule capture.

The discovery of semantic or compliance lapses leads to the reporting of issues as a request for rework. Issues and potential remedial actions can be shared using BCF (BIM Collaboration Format) messages [6] so that they can be handled and managed systematically.

Semantic development applies similar rules as validation and compliance checking but directly effects change to the target model. Bloch [2] has provided a review of the motives and methods for the semantic development of BIM. Methods include machine learning, coding and templates. Bloch et al. [2] has emphasized that if the level of semantic and geometric content is low, as found in unlabelled apartment plans, then machine learning may be preferred over rule-based approaches, with the risk that the outcomes may be incomplete or inaccurate. Sacks et al. [19] have shown that semantic enrichment can be deployed if there is a sufficient set of geometric relationships between entities within a defined context such as simple concrete bridge forms. Results can include the completion of obscured forms, identification of the element type such as 'beam' and some specializations such as 'secondary beam'. The rules are embedded in procedural programme code. Ma et al. [14] developed this approach by encoding the semantic rules for execution in compressed strings. A commercial application 'SimpleBIM' [7] allows systematic enrichment to be carried out interactively with some ability to record the steps in 'macros' that can be re-enacted, as long as the rules fit to the available templates. Nisbet et al. [16] raised the need for more systematic approaches to the classification of entities within BIM models, for example to support the assignment of carbon density from rate libraries and the subsequent reporting of carbon results.

However, none of these approaches support the documentation of expectations by AECO actors in a manner that is generic, and both reviewable and operable. This reduces the transparency of the content and the usability of implementations.

3. Proposed method

The RASE (Requirements, Applications, Selections and Exceptions) semantic mark-up [15] is used to make both the requirements and any supporting resources both human-readable and machine-operable. A normative resource describes an idealized or aspirational set of expectations and is intended to be compared against a description of the real or an intended world. The result of such a comparison may be a 'pass' or 'fail'. In practice, a third outcome of 'unknown' is allowed for. Initially, RASE was targeted at normative resources such as regulations where individual clauses or 'checks' could be considered separately. Each clause typically contained phrases which identified the applicability of the clause, the requirements and sometimes exceptions. For example, a clause containing the phrase 'external doors' identifies two separate Application metrics, '(is) external' and '(is a) door'. Each applicability narrows down the aspects for the target domain, in this case buildings. The requirement 'fireproof and waterproof' identifies two distinct requirement metrics that can be tested in the building domain. A phrase 'except shopfronts and lorry docks' identifies two distinct exception metrics, excluding some aspects of the building domain. In some cases, the applicability included a range of situations and the concept of Selection was added. For example, 'Doors and windows' contains two parallel concepts that serve to broaden the aspects of the building domain that are relevant. Because each kind of metric has its own logical role, there is no limit to the number of RASE metrics that occur in a sentence or clause. As more complex clauses and entire documents were considered, it became clear that the concept of a 'check' was actually an example of a 'Requirement Section'. A section may correspond to a chapter or paragraph, but this is not always the case. For example, it is not unusual for exceptions to appear in a subsequent second paragraph. In some cases, complex subsidiary clauses were 'Application Sections', 'Selection Sections' or 'Exception Sections'.

This semantic mark-up can be used to execute automated code compliance of a description (or narration) about a building model, as long as the metrics identified refer to concepts that have an equivalence in the target building information. Descriptive/narrative text does not contain 'Requirements' but 'Reports'. Previous work has shown that a descriptive text describing the design of an existing hospital building can also be treated as a RASE source [17]. Treating this knowledge as 'Reports' allows for the possibility that the virtual account may differ from the real world, or that two virtual descriptions may be in contradiction. In the current examples, this target model is held in the IFC schema [10].

Summary of the RASE method:

a. Specific words and phrases indicate metrics which are testable. Any metric must be capable of giving a result that is 'true', 'false' or 'unknown', and so is typically a noun or adjective.
b. Sections indicate objectives which can be evaluated by considering the metrics and any other sections contained within them. A section should contain at least one metric or section. If a top-level section is the whole document, then its evaluation to 'true', 'false' or 'unknown' will indicate if the document has been satisfied.
c. Sections and metrics are of four kinds: Application, Selection, Exception and Requirement ('shall be').
d. Each Section can be evaluated by considering if it is 'As Required' or 'Not Applicable' or 'Not Selected' or 'Excepted' by considering the contained sections and metrics
e. 'As Required' is evaluated by using the logical connector 'and' applied to the set of all of the contained requirement metrics and sections.
f. 'Not Applicable' is evaluated by using the logical connector 'nand (notand)' applied to the set of all of the contained application metrics and sections, progressively narrowing the scope.

g. 'Not Selected' is evaluated by using the logical connector 'nor' applied to the set of all of the contained selection metrics and sections, each broadening the scope. There should be more than one selection.
h. 'Excepted' is evaluated by using the logical connector 'or' applied to the set of all of the contained exception metrics and sections, each eliminating scope.

These rules can be built into a simple procedure or rule engine and applied recursively to obtain a result thereby ensuring that knowledge documents can be understood and made operable. The result of the evaluation may be a true/false/unknown, indicating pass, fail or the need for more information. Work is in progress to discover the bounds and other limitations to this mathematically, linguistically and epistemologically.

Mark-up is the addition to a document of hidden or visible tags. RASE semantic mark-up is added to the HTML documents so as to make the sense of the rules explicit. RASE mark-up is typically added as HTML 'span' tag pair surrounding the section or phrase. Metrics can have additional attributes to capture any obvious interpretation of the phrase. The additional attributes include a 'data-raseProperty' attribute so that for example if the phrase contains a plural term such as 'doors', the singular term can be recorded 'door'. If the phrase represented an enumerable concept such as 'hospital', the 'data-raseProperty' attribute can be set to 'building type' and 'data-raseTarget' attribute can be set to 'hospital'. The 'data-raseComparator' and 'data-raseUnit' attributes can be used to clarify any numeric metrics.

The mark-up task can be carried out by AECO actors using either coloured pens during training or using an on-screen editor such as 'AEC3 Require1' [1]. If there is knowledge contained in a relational or graph database, its data dictionary can be extended to identify the RASE role of the fields as was done in the USACE Facility Capacity Analysis project [9]. Any mistakes or omissions can be identified and corrections and improvements made.

The proposed method begins with extending RASE to capture definitive content. RASE can be extended to support semantic development by allowing the definitive expectations to be first written as text or tables and then marked up. A definitive clause can contain 'Registrations' of the agreed equivalences, for example between '(is a) door' in English and 'element = IfcDoor' in IFC. Semantic development uses 'Registrations' to ensure that any clause that evaluates to true is added or corrected in the target domain model. In all other respects 'Registrations' behave as 'Requirements' and 'Reports', for example being joined using 'and' operators.

In order to explore this gap in the ability to document and execute semantic rules, a variety of presentations of the expectations are explored including free text, various tabular layouts and tree hierarchies. The examples shown are the result of iterating between various kinds of expectations, and candidate solutions in the form of increasingly accurate rules.

Two example BIM models from industry, a healthcare suite as an example of social infrastructure and a motorway bridge as an example of transport infrastructure (Fig. 2) are used to explore the need for semantic development of intrinsic properties and extrinsic relationships. A variety of such requirements for semantic development were discovered in the sample models.

The proposed implementation is summarized in Fig. 1. Each process (rectangle) is shown with inputs (left), outputs (right), constraints (above) and resources (below). The upper path was executed repeatedly to refine and extend the documentation of the expectations so as to converge on the expected outcomes. The lower path represents the use of the two example models. The transformation engine generates revised models and a detailed log which can be inspected and compared the original to check the result.

4. Experiments and development

Two models were selected for consideration, representing a hospital room as social infrastructure, and a motorway bridge as transport infrastructure. Because both models were available in the same open schema and format, the techniques needed to implement the expectation for semantic development could use the same transformation engine and dictionary resources.

One model represents a single-patient room with an ensuite shower room. The model is being prepared as a collection of recommended room layouts for use by a UK hospital building program. The scope of the models includes building fabric, spaces, furnishing, and mechanical and electrical services. One space represents a 'Single Bed (Outboard)' room with the ensuite bathroom placed on the external wall to maximise the opportunities for supervision of the patient. The second space 'Ensuite (Outboard)' contains a shower, WC basin and supporting equipment. An 'Inboard' variant is also available for patients with a lower dependency offering better views outside. The model contains a variety of IFC entities including Covering, Distribution Control Element, Distribution Port, Door, Flow Controller, Flow Fitting, Flow Segment, Flow Terminal, Furnishing Element, Slab, Wall Standard Case and Window. It also includes a number of Building Element Proxy elements which have not had an appropriate IFC element class assigned.

The second model represents a motorway bridge with its structural members and pavement deck. The model was being prepared as part of a major motorway and trunk-road improvement in North-West England. It includes Beam, Column, Element Assembly, Slab, Footing, Ramp, Railing, Ramp Flight, Reinforcing Bar and Wall Standard Case elements. It also includes a number of Building Element Proxy elements.

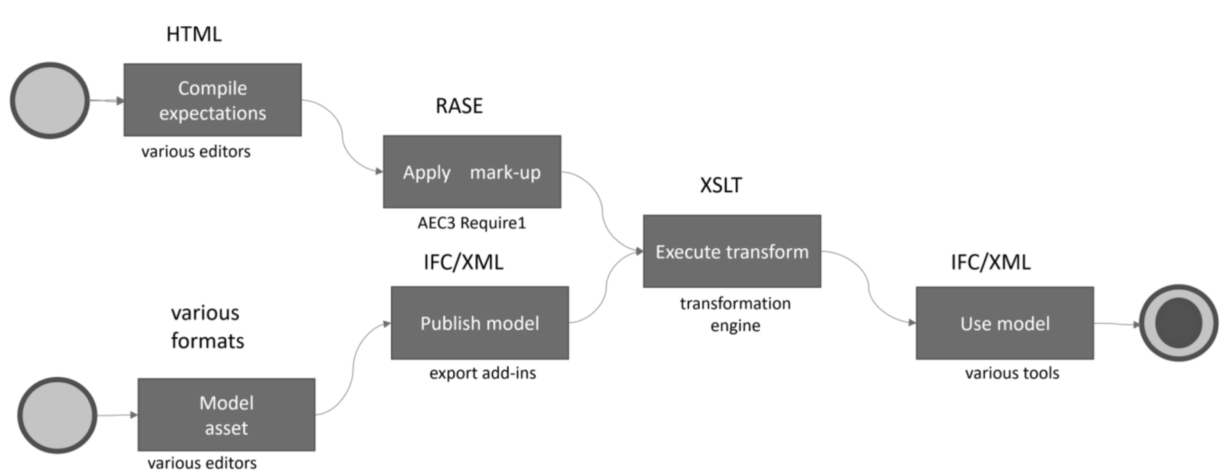

Fig. 1. Proposed Method.

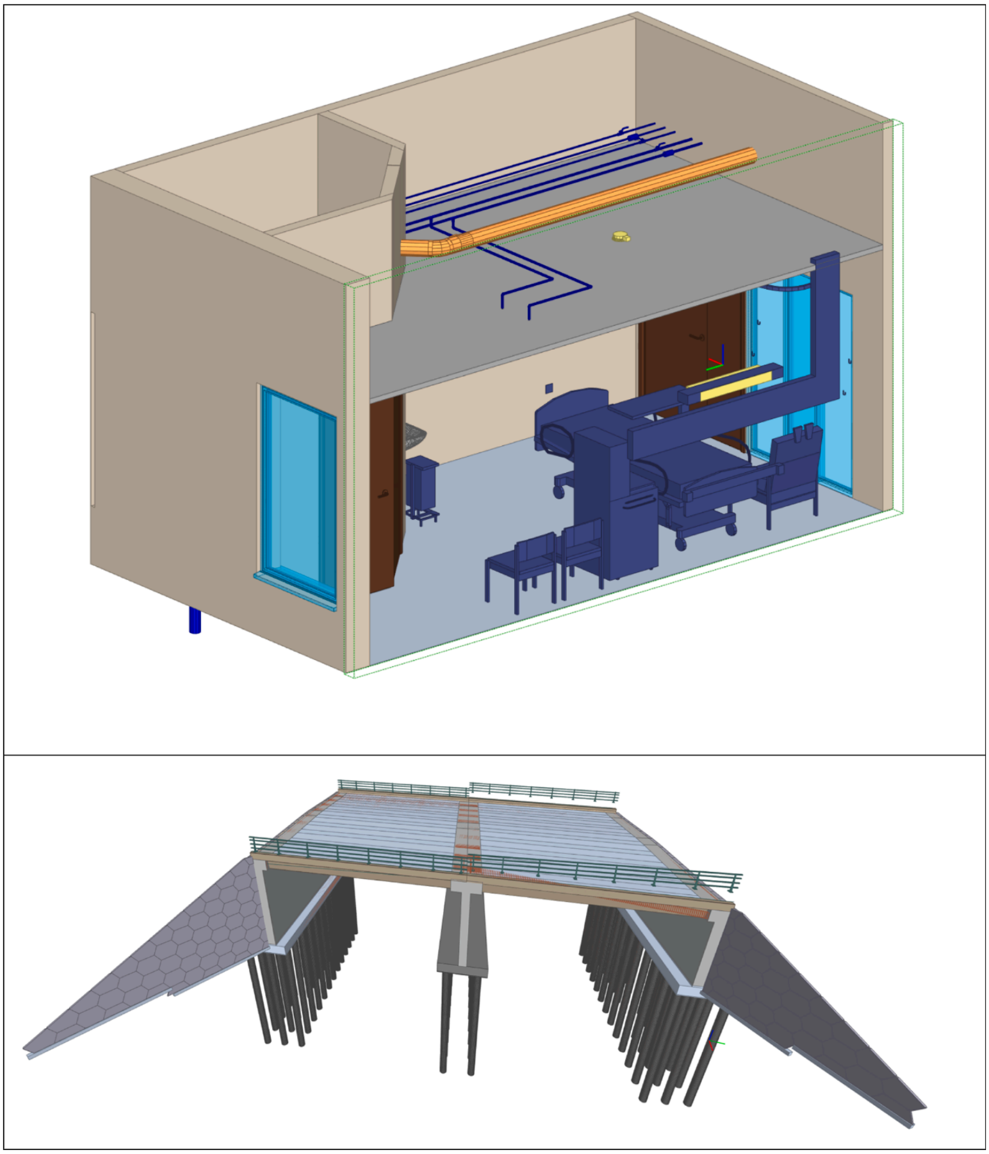

Fig. 2. The two target infrastructure models (see acknowledgements).

Both models contain extensive relationships between the entities and large numbers of properties (attributes) associated to the entities.

The models were reviewed using a number of IFC viewers and reporting tools. A number of semantic issues were detected (Table 1) and documented using various presentations that may be familiar to different AECO actors (Tables 2–11).

A semantic issue was included if it had the potential to damage the usefulness of the model for reporting such as quantification and quantity take off (QTO), analysis or compliance checking. In practice each of these use cases might be subject to detailed documentation of expectations. Each issue was documented in HTML using a standard editor and then RASE mark-up added using a mark-up tool, 'AEC3 Require1′. The presence of RASE mark-up triggers Cascading Style Sheets (CSS) style rules, making the mark-up visible for review. The style rules used are introduced along with the examples.

4.1. *Wrong element types*

A number of entities were represented using an inappropriate IFC element [10–11]. This type of error may reflect difficulties with the configuration of the authoring application or user error. These issues

Table 1
Summary of semantic issues found.

Development	Semantic issue (discussion)	1) Hospital room	2) Motorway bridge
Correction	Wrong element types (4.1)	Damper Furniture Curtain track	Wall panels Mass concrete Baffle wall
	Confidential information (4.2)	Project, site and facility name and description	
Enrichment	Possible element misuse (4.3)	–	Slab, Beam, Footing, Column
	Groupings not used (4.2)	Systems Zones	–
	Missing connectivity (4.3)	–	Load transfer
	Missing classification (4.5)	Spaces and Products	
Enhancement	Missing common properties (4.5)	Function, Form, Usage, Material, Medium	
	Member weights missing (4.4)	–	Material and volume

reflect that there need be no coherence within the object representation in BIM between type (class), name, description, shape, properties and classification, if parametric capabilities have not been used to maintain that consistency. A simple sentence is composed for each issue, defining the expected correction. The RASE mark-up underlines various trigger selections in purple and the definitive semantic corrections are underlined in blue (Table 2). Similar rules are created to revise the type on mass concrete, baffle walls and hexagonal wall panels in the bridge model which had also been represented as a 'proxy' element, indicating that a more specific assignment had not been made. 'Proxy' entities lack semantic significance and so devalue the utility of the overall model.

A UK standard hospital equipment library had been imported to populate the hospital room layout. The import method had not assigned an appropriate type to the furniture, leaving it as a 'proxy' element. These issues are documented in a simple table. The three letter codes found in the name attribute such as 'BED' is used to identify the furniture entities. Description and classification enrichment rules are included.

The RASE approach to ordinary text has been summarised above. RASE is also applicable to tables. RASE takes a systematic approach to interpreting the content of tables as a number of rules. Data cells (HTML 'td' tag, shown here with italic text) each generate a distinct rule along with the content of any header/sider cells (HTML 'th' tag, shown here with normal text) aligned vertically or horizontally. The mark-up (Table 3) includes Applicability (shown underlined in green). For example, 'IfcSpaceHeater' typing and the other attributes will be set for any element named 'Radiant Panel'.

In the case of the privacy curtain railing, the user had selected a wall-sweep tool to achieve the continuous straight-curved-straight shape, without considering its semantic implications. The presentation advocated in ISO12911 for BIM implementation plans [13] has been used (Table 4) to document this semantic correction. For example, elements will be described as 'privacy curtain supports' on any 'IfcRailing', any element having a name including 'Curtain tracks or any element having object type including 'Tracks'.

4.2. Confidential information and groupings

Neither model had been anonymised prior to sharing. A rule was documented to correct the details of the project, site and facility. This can be considered as a correction, since a rule could have been developed to ensure that a model had the contractually agreed identifiers and descriptions for the project, site and facility. An enrichment is documented to ensure that the CAD-based layering of entities was re-interpreted as objectified element groupings (IfcSystem, IfcZone). This is achieved by mark-up including Exceptions (shown underlined in orange) in Table 5.

4.3. Verification of structural members

In contrast to the SEEBIM approach [19] which focussed on extrinsic relationships such as 'touching' and 'parallel', rules are developed based on the intrinsic properties of the individual loadbearing members. These focus on the relative size of the length, width and height of loadbearing members (Table 6). Three comparators are used: 'much smaller than', 'similar to' and 'much larger than'. A boundary ratio of '2′ was used when comparing the relative sizes of dimensions, but in practice local structural material codes could require different ratios. The 'data' cells identify the five main types of structural (load bearing) member. Diagonal members and struts are not considered. 'Header' and 'sider' cells than add Application and Selection rules for the plan proportions, elevation proportions and sectional proportions. For example a 'slab' is defined by being 'load-bearing' and by scanning sideways, having 'length similar to width' in plan, and scanning upwards, 'height much smaller than length' in elevation and 'width much larger than height' in section. Arranging the content as a grid makes it easier to validate that the rules span all possibilities without overlap or omissions. This is in contrast to the SEEBIM approach [14] where complex hashing algorithms were deployed to achieve the same assurance. Execution of this rule confirmed that all the structural objects already had the correct types assigned, so no enrichment was required.

Consideration was also given to identifying more specific roles of structural elements. For example, 'primary beam' is informally defined in various dictionaries as:

a. beam that supports secondary (smaller sized) beams.
b. beam (horizontal flexural structure member) that directly connects to supporting column (compressive structure member).
c. beam that connects to a column for transferring loads of a structure.
d. beam connecting to columns to transfer the load (from a secondary beam if present)

In summary, a primary beam can be defined as a beam that is not supported by other beams. This rule can be documented (Table 7) and can be executed if the 'supports' relationship is available in the model.

4.4. Addition of mass quantification

Semantic enhancement uses external resources. In a simple case of semantic enhancement, the rules set itself can be seen as an external resource. For the bridge model, rules (Table 8) are developed to deduce the mass (weight) of members where the material and volume is known. The document holds the density and the formula to map a volume to a mass. For example, each of the twenty primary beams in the motorway bridge is calculated to weigh almost 30 T, illustrating how semantic enhancement can contribute to other analysis, such as logistics and safety.

A second dictionary resource (Table 9) holds the documentation of 'mass' as a property. A dictionary may repeat terms both as a Selection and as a Registration. This ensures that a term used in one context can trigger the use of definitions in multiple other contexts.

4.5. Classification

Semantic enrichment rules are used to assign classification to entities

Table 2
Definitive clauses containing expectations.

Clauses
Any IFC Damper or entity with Name including Damper or having Object Type includes Damper should be an IFC Damper and be Described as Used to control or modulate the flow of air. and classified Uniclass: Pr_65_65_24 : Duct dampers .
Any entity with Name including Radiant Panel should be an IFC Energy Conversion Device and be Described as Used to convert between forms of energy. and classified Uniclass: Pr_70_60_36_71 : Radiant panels

Table 3
Table of definitions (truncated).

Applicabilities and Selections	Registrations
• Name includes Radiant Panel	• *IFC Space Heater* • *Space heaters utilize a combination of radiation and/or natural convection using a heating source such as electricity, steam or hot water to heat a limited space or area.* • *Uniclass: Pr_70_60_36_71 : Radiant panels*
• IFC Furnishing Element • Name includes ART • Name includes BBF • Name includes BED • Name includes Bedhead • Name includes BHD • Name includes BRA • Name includes CAL • …	• *IFC Furnishing Element* • *Described as A furniture related element, characterized as being pre-manufactured and assembled on-site, or manufactured on-site (built-in) and either movable or not.* • *Uniclass: Pr_40 : Signage, sanitary fittings and fittings, furnishing and equipment*

Table 4
Definitive table to ISO12911.

Objective:	Fix misuse of wall entities
Application:	(all)
Selections:	• IFC Railing • Name includes Curtain Tracks • Object Type includes Tracks
Exception:	(none)
Register:	• *IFC Railing* • *Described as Privacy curtain supports.*
	• *Uniclass: Pr_40_30_20_63 : Privacy curtain supports*
Note:	Mis-modelling can be the fault of authoring tools or user practice.

in the model. For example, the classification of Uniclass Spaces and Locations is strongly driven by the contents of spaces (Fig. 11). If a classification system has been marked up with RASE then this enhancement can be automated. Existing industry classifications such as Uniclass [22] have also had RASE mark-up systematically added to create hierarchical semantic resources that can be deployed to enhance the sample models.

Alternatively, the presence of a classification code can imply a number of properties which enhance the model for use in analysis and reporting. This requires systematically identify the driving properties implied by the naming of the classification entries (Table 11). The list of such properties includes 'function', 'form', 'medium' (for distribution services), 'user', 'material' and 'content/parts'.

5. Implementation

Having captured the expectations, the marked-up documents could have been transformed into a number of different rules and knowledge representations [18]. For this work, a rule-engine is used that can directly access both the semantic mark-up and the model, generating a revised model, without any need for an intermediate representation. A single 'eXtensible Stylesheet Language Transformation' (XSLT) [24] is used to parse the semantic content in HTML and the model in ifcXML to generate a revised model (Fig. 3). XSLT is a relatively high-level declarative language that supports the documentation of the trigger situations found when exploring the HTML structure and the actions to be taken in response. For example, if a rule has been evaluated that suggests that a description attached to an element needs to be updated, then prior to the description being transferred from the input to the output, the revised value can be substituted.

RASE expects that any rule in the document, when applied to any instance in the model, will resolve to being as Required, or not Applicable, or not Selected, or Excepted. This implies that the element in the model needs updating if there are any unsatisfied Requirements, no unsatisfied Applications, no satisfied Exceptions and at least one satisfied Selection if any are present. Considering each element in turn, an evaluation pass is made to consider any Applications, Selections and Exceptions, and determine if any Registration requirements are unsatisfied. This will evaluate to 'true' if semantic development is required: A second evaluation pass applies the required updates including associating a new 'Owner History' (IfcOwnerHistory) to the element, property or classification assignment. The 'Owner History' object can hold details

Table 5
Example table containing expectations arranged systematically in rows.

Applicability	Selection	Exception	Register
(Any)	• Project • Site • Building	(None)	*shall have* • *anonymized name* • *anonymized description* • *anonymized long name*
Any entity • having a layer assigned	(Any)	unless • space • already grouped	*shall be* • *grouped in an equivalent system*
(Any) • space • having a layer assigned	(Any)	unless • already grouped.	*shall be* • *grouped in an equivalent zone.*

Table 6
Example two-way grid containing registration of structural roles based on relative dimensional proportions.

Load Bearing beam	-	Length much smaller than Width	-	-	-	-
-	-	-	-	Length much larger than Width	*Load Bearing wall*	-
Height much smaller than Length	Height much smaller than Length	-	-	-	Width much smaller than Height	Width much smaller than Height
Width is similar to Height	-	-	*Load Bearing footing*	-	Height is similar to Length	-
-	Width much larger than Height	-	-	-	-	Height much larger than Length
-	*Load Bearing slab*	-	Length is similar to Width	-	-	*Load Bearing column*

Table 7
Example specialisation rules for bridge beams.

Type	Situation	Load Bearing
beam	not supported by beam	*Primary Girder Beam*
beam	supported by beam	*Secondary Transverse Beam*

of a user, an application and the date stamp for any change. This ensures that all changes are traceable to the user of the application, as expected by ISO 19650 on the implementation of collaborative BIM processes [11]. ISO 19650 addresses the management of distinct 'information containers' which are typically assumed to be files but in this case is taken as each element within the file. Detailed logs and visualisation tools are used to verify that all the expected semantic development has taken place, the model remained syntactically valid, and that no other changes were introduced.

6. Discussion

The RASE representation has previously been used for capturing normative knowledge. In this work the RASE methodology has been

Table 8
Example of semantic enhancement using material density.

Material	With known volume
Aluminium	*2700 kg/m3*
Concrete - Cast In Situ	*2300 kg/m3*
Concrete - Cast-in-Place Concrete - C35	*2300 kg/m3*
Concrete - Cast-in-Place Concrete - C25	*2300 kg/m3*
Concrete - Precast Concrete	*2100 kg/m3*
Concrete, Precast Panels	*2100 kg/m3*

extended to support semantic development. The semantic representation of dictionary and development resources can include alternative selections, exceptions and applicability's, reflecting the evolution and refinement of the expectations of AECO actors. RASE semantic mark-up links together different contexts (such as English and IFC) and different aspects (such as description, classification and shape). Since RASE is a recursive representation there is no limit to the complexity or subtly of the knowledge captured. The rules shown in this paper are kept relatively simple for clarity, with no subordinate objective sections or cross-references.

The paper demonstrates that external knowledge in dictionaries, classification tables and development specifications in a variety of presentations can remain human-readable, but can also be made machine

Table 9
Example of a supplementary definition concerning mass in various contexts.

Applicability and Selections	Register
• **other : Mass** • **IFC : Weight** • **en-GB : A mass measure is the value of the amount of matter that a body contains.** • **measure : mass**	• *other : Mass* • *IFC : Weight* • *en-GB : A mass measure is the value of the amount of matter that a body contains.* • *measure : mass*

Table 10
Example rules for deriving classification from content.

• *SL_45_10_09 : Bedrooms* • name includes 'bed' • content includes entities having Uniclass 'Pr_40_50_06 : Beds'
• *SL_35_80_80 : Showers* • name includes 'shower' • content includes entities having Uniclass 'Pr_40_20_06_79 : Shower enclosures'

Table 11
Example rule for deriving properties from classification.

• Uniclass: Pr_70_60_36_71 : Radiant panels • *Function: heating* • *Medium: electricity* • *Form: panel*

```
• for each element in the target model.
    • evaluate the semantic mark-up in the source document.
        • if an objective section: evaluate it by recursing down to evaluate
          the objective sections and metric phrases below.
        • if a metric phrase:
            • pass 1: evaluate against the current element
                • taking any Register metric as true
            • pass 2: evaluate against the current element
                • fix any failed Register metric phrase
                • consider type, name, description, property, classification
    • end evaluation
    • if evaluated to true on the first pass perform a second pass
• next element
```

Fig. 3. Outline of the XSLT transform.

operable and so used to effect the semantic development of descriptive and narrative models such as BIM.

Some specific limitations in this exploration are embedded in the simple declarative transformation XSLT file '_developed.ifcxml.xsl'. For example, the function to implement the multiplication of volume and density is hard-coded, pending the implementation of a more generic formula evaluation.

Secondly, the examples did not use any metrics relating to entity adjacency or connectivity. The RDF IfcEngine which is used to convert IFC to IFCXML and back also has a full suite of functions to detect and add adjacency relationships such as 'supports' or 'touches' using 'IfcRelConnects'. Adjacency relations may also be needed to identify new features such as hole openings being an emergent feature of slabs, or a vertical shaft space being an emergent feature of multiple storeys having aligned holes in slabs. Rules could ensure that the appropriate entities are added to the model.

7. Conclusions

This paper demonstrates the feasibility of the innovative approach to capturing expectations for semantic development using plain text with RASE mark-up in documents which are both human readable and machine operable. This completely avoids any need for any computer coding in contrast to previous work in this area.. This empowers AECO actors to take direct control of their 'registering' their expectations. This allows review and endorsement prior to their application, and their refinement from experience. The variety of presentations of semantic development rules has shown that the most humanly accessible presentation can be chosen, with the mark-up ensuring that machine operability is also achieved. This confirms that semantic development is an example of knowledge capture, so that bodies of knowledge can be accumulated and used, without depending on tacit knowledge held by individuals and without depending on maintaining a separate body of computer code.

This ability to review and maintain sufficiently expressive semantic resources using RASE is intended to reduce the burden of manual BIM authoring tasks, and so act as a gateway to the wider deployment of automated analysis and code compliance checking. In addition to addressing the limitations of the implementation, future work could conduct controlled trials to establish metrics around accuracy and efficiency of mark-up compared to specialised interfaces or computer coding.

CRediT authorship contribution statement

Nicholas Nisbet: Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zijing Zhang:** Writing – review & editing. **Ling Ma:** Supervision. **Weiwei Chen:** Supervision. **Mustafa Selçuk Çıdık:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

Appreciation is expressed to Costain plc and to the UK NHS ProCure22 Consortia for permission to use their work-in-progress IFC models.

References

[1] AEC3 Require1, 2023. http://www.aec3.eu/require1/AEC3_Require1.html (accessed May 13, 2023).
[2] T. Bloch, Connecting research on semantic enrichment of BIM-review of approaches, methods and possible applications, J. Inf. Technol. Constr. 27 (2022) 416–440.
[3] buildingSMART, Report on Open Standards for Regulations, Requirements and Recommendations Content, 2016 https://www.buildingsmart.org/wp-content/uploads/2017/11/17-11-08-Open-Standards-for-Regulation.pdf (accessed May 13, 2023).
[4] buildingSMART, IFC4.3, 2023a https://ifc43-docs.standards.buildingsmart.org/ (accessed May 13, 2023).
[5] buildingSMART, IDM, 2023b https://technical.buildingsmart.org/projects/information-delivery-specification-ids/ (May 13, 2023).
[6] buildingSMART, BCF, 2023c https://technical.buildingsmart.org/standards/bcf/ (accessed May 13, 2023).
[7] DataCubist https://simplebim.com/ (accessed May 13, 2023).
[8] DCOM, 2023 https://www.dcom.org.uk/ (accessed May 13, 2023).
[9] E.W. East, N. Nisbet, Facility capacity analysis, in: Proceedings of the CIB W078 Conference, Sophia Antipolis, 2011.
[10] ISO (2018a) ISO 16739-1:2018 Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema https://www.iso.org/standard/70303.html (accessed May 13, 2023).
[11] 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. https://www.iso.org/standard/68078.html (accessed May 13, 2023).
[12] ISO (2022) ISO 19650-4:2022 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 4: Information exchange https://www.iso.org/standard/78246.html (accessed May 13, 2023).
[13] ISO (2023) ISO 12911:2023 Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Framework for specification of BIM implementation https://www.iso.org/standard/79692.html (accessed May 13, 2023).
[14] L. Ma, R. Sacks, U. Kattel, T. Bloch, 3D object classification using geometric features and pairwise relationships, Computer-Aided Civil and Infrastructure Engineering 33 (2) (2018) 152–164.
[15] N. Nisbet J. Wix D. Conover The future of virtual construction and regulation checking P. Brandon T. Kocaturk Virtual Futures for Design, Construction and Procurement 2008 Blackwell, Oxfordshire 10.1002/9781444302349.ch17.
[16] N. Nisbet, S. Lockley, et al., Automated enhancement of BIM models, eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2012.
[17] N. Nisbet, M. Aizlewood, J. Cartwright, B. Wakelam, Gaming Energy Refurbishment. In Proceedings (Vol. 1, No. 7, p. 688). MDPI, 2017, October.
[18] N. Nisbet, L. Ma, G. Aksenova, Presentations of RASE knowledge mark-up *EC3 Conference 2022*. Vol. 3. University of Turin, 2022.
[19] R. Sacks, L. Ma, R. Yosef, A. Borrmann, S. Daum, U. Kattel, Semantic Enrichment for Building Information Modeling: Procedure for Compiling Inference Rules and Operators for Complex Geometry, J. Comput. Civ. Eng. 31 (6) (2017) 04017062.
[20] Solibri. (2023). Solibri Model Checker (SMC)". https://www.solibri.com/ (May 13, 2023).
[21] T. Bloch, R. Sacks, Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models, Automation Constr. 91 (2018).
[22] Uniclass. 2023 https://www.thenbs.com/our-tools/uniclass (accessed May 13, 2023).
[23] Verifi3d https://verifi3d.xinaps.com/ (accessed May 13, 2023).
[24] W3C XSLT 1999 https://www.w3.org/Style/XSL/ (accessed May 13, 2023).
[25] J. Zhang, N.M. El-Gohary, Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking, Automation Constr. 73 (2017) 45–57.
[26] R. Zhang, N. El-Gohary, A deep neural network-based method for deep information extraction using transfer learning strategies to support automated compliance checking, Automation Constr. 132 (2021), 103834.