Zheng, Yuan; Törmä, Seppo; Seppänen, Olli

A shared ontology suite for digital construction workflow

Published in:
Automation in Construction

DOI:
10.1016/j.autcon.2021.103930

Published: 01/12/2021

Please cite the original version:
Zheng, Y., Törmä, S., & Seppänen, O. (2021). A shared ontology suite for digital construction workflow. Automation in Construction, 132, [103930]. https://doi.org/10.1016/j.autcon.2021.103930
A shared ontology suite for digital construction workflow

Yuan Zheng a,*, Seppo Törnä b, Olli Seppänen a

a Department of Civil Engineering, Aalto University, Finland
b Visualynk Oy, Finland

A R T I C L E   I N F O

Keywords:
Construction workflow
Ontology
Digital construction
Information management

A B S T R A C T

With ongoing advancements in information and communication technologies (ICTs) in all stages of the construction lifecycle, information from entities related to construction workflow (CW) can now be automatically collected. These implementations are point solutions, which require systematic integration to combine their information to enable a holistic picture of CW. The major barrier to such integration is information heterogeneity, where the information is collected from different systems under multiple contexts. Scholars in the construction domain have explored the use of ontology to solve the information-integration problem, although an ontology that both adequately represents the CW and integrates the digitalized information of CW via various systems and multiple contexts is currently missing from the existing literature. This research thus presents an ontology set for formalizing and integrating CW information within the digital construction context. The proposed digital construction ontologies (DiCon) are shared representations of construction domain knowledge that specify the terms and relations of CWs and their related information. We developed the DiCon based on a hybrid ontology development approach. The DiCon includes six modules: Entities, Processes, Information, Agents, Variables, and Contexts. The developed DiCon was further evaluated by approaches including automatic consistency checking, criteria-based evaluation, expert workshops, and task-based evaluation and involved two use cases by answering relevant competency questions via SPARQL queries. The results of the evaluation demonstrate that the DiCon ontologies are sufficient to represent domain knowledge and can formalize and integrate CW information within the digital construction context.

1. Introduction

The construction process is driven by information. Information is vital for site teams to be aware of the actual situations of construction workflows (CWs), and their related resources and constraints, in order to support timely decision-making and action-taking [1]. In the past, CW information was primarily acquired using manual methods, which tend to be expensive, inefficient, inaccurate, and subject to delay [2,3]. Various researchers have suggested that tracking project-related entities and phenomena at worksites can provide useful information for interpreting the situation of the CW [3–5]. The increasing implementation of information and communication technologies (ICTs) in the construction industry provides an opportunity to automatically capture and enrich information from CW-related entities. Previous works have shown that various CW-related entities can be tracked or monitored by using digitalized technologies including sensors and the Internet of Things (IoT) [6–8], indoor positioning systems (IPSs) to track labor flows [9], and computer vision/image processing to monitor equipment [10,11]. Simultaneously, necessary design information for workflow-related entities such as materials, objects, and workspaces can be retrieved from building information models (BIMs) [12]. Construction management (CM), enterprise resource planning (ERP) [13], and supply chain management (SCM) [14] systems can also provide large amounts of information on related entities including activity, labor, equipment, and material.

Although these ICT implementations enable automatic information collection and provide large amounts of digital information and data of CW-related entities, they are point solutions that cannot result in a comprehensive situational picture of the CW. Therefore, a systematic integration of all the information from these systems based on construction domain knowledge should be explored to build a holistic picture of CW. Such integration could also be considered as the foundation for achieving the “third era” of implementing information technology in construction, addressed by Froese [15]. The primary barrier to achieving integrated information, however, is information heterogeneity, which is a common problem in the construction domain [16,17].

* Corresponding author.
E-mail address: yuan.zheng@aalto.fi (Y. Zheng).

https://doi.org/10.1016/j.autcon.2021.103930
Received 7 January 2021; Received in revised form 23 August 2021; Accepted 27 August 2021
Available online 15 September 2021
0926-5805/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Information from related workflow entities is often acquired via various information sources and from different stakeholders working in different construction disciplines using a variety of tools, systems, and software. As a result, the distributed information from one system may not be available to other systems. This situation also leads to insufficient interaction between the different systems and ineffective data sharing and exchange [18]. As a consequence, integrating data from multiple sources within a construction project is usually a labor-intensive and costly procedure [19,20].

The heterogeneity of CW information is also generally characterized as a multi-context representation, because the construction lifecycle involves an evolution of information in many different areas, including requirements, designs, plans, and the progress of execution [21]. For example, in the design area, specified levels of detail/development (LODs) contain presuppecified levels with unambiguous content requirements [22,23]. In contrast, in the construction planning domain, the levels are usually more loosely defined and are based on prevailing practices; for instance, the Last Planner® system has a master schedule, a phase schedule, look-ahead planning, and weekly and daily plans. The different levels of the schedules and plans provide distinct planning contents and different information details [24]. In addition, the actual CWs almost always deviate from the as-planned context, which also leads to unexpected and unplanned changes that then cause modifications in existing designs or plans [25]. Additional contexts of information thus are dynamically created during the execution. To effectively and comprehensively represent the CW, besides accounting for the heterogeneity caused by multiple information sources, the distributed information in different contexts should also be accounted for during the integration.

In facing such challenges of information integration, Froese [26,27] emphasizes that standard construction information models are required as cornerstones to support information management and system interoperability. An information model should be developed based on domain knowledge that can represent the detailed mechanism of CW and then be used as a reference for integrating the heterogeneous information. Various scholars have developed several conceptual models that describe CW information over the past few decades [26,28–30], but they are conceptual models that lack computer interpretability. Known as an “explicit specification of a conceptualization” [31], an ontology can act as such an information model by effectively integrating heterogeneous information by providing common unambiguous terminologies of concepts and relations based on domain knowledge with a computer-interpretable format [32].

Ontologies, as the foundation of implementing Semantic Web and linked data technologies, could also enhance the sharing or reuse of information, data, and domain knowledge [33]. Various scholars in the construction domain have addressed ontologies’ efficacy in systematically formalizing domain knowledge and information [34,35]. But previous studies on ontology [36–40] have focused on formalizing CW knowledge without considering the information aspect or recent ICT implementations. An ontology to address the inherent challenges of integrating CW information from heterogeneous sources and representing multi-context information has not been developed to date.

To fill this research gap, in this paper we propose a suite of ontologies called the digital construction ontologies (DiCon), where we aim to offer a higher-level conceptualization and formalization of CW with shared and reusable domain knowledge representation. The ontology is novel in its provision of an unambiguous formalized information structure. This structure can serve as a reference to structure and integrate the data and information from multiple heterogeneous systems in order to build a holistic picture of the CW. Because the ontology suite has been implemented in a machine-readable format, it can work as a foundation for developing extensions or applications and systems based on Semantic Web and linked data to aid CW management. The ontologies developed in this paper will be valuable to both industrial users and academic researchers working in the lean construction, digital construction, and linked data domains, as it demonstrates how heterogeneous information can be integrated and further utilized under the linked data framework to support construction management.

The remainder of the paper is organized as follows. In Section 2, we review related works about modeling CWs as the background of this research. In Section 3, we outline the method for developing the proposed ontology suite. We then propose a detailed description of DiCon modules in Section 4, which covers the ontological model and the content of each module in the ontology suite. In Section 5, we provide detailed descriptions of our evaluation approaches and results, including task-based evaluation, with two specific use cases demonstrating the ontology application. This section is followed by a discussion in Section 6 of the study’s contributions and limitations, as well as future research avenues using the proposed ontology suite. Finally, we present conclusions from our research in Section 7.

2. Literature review

This section discusses prior research on modeling CW and related ontology works to provide a background and references for the development of the DiCon. Section 2.1 introduces the related domain knowledge of conceptualizing the CW and existing CW information models. In Section 2.2, we review existing ontologies that directly model the CW. At the end of this section, a review of other related ontologies is presented.

2.1. Modeling construction workflow

To explore CW modeling, our attention first turned to the lean construction domain. Those using the lean construction perspective consider construction activities to be assembly-type operations that are enabled by a variety of ingredients called “flows,” including labor, equipment, workspaces, components, information, external conditions, and prerequisite tasks. This flow view was first introduced by Koskela [41,42] and further refined as the activity flow model (AFM) by Garcia-Lopez [43]. The AFM conceptualization represents a detailed mechanism of construction activities with different types of flows. The contribution of AFM is that in order to model CWs, knowing just the flows associated with each construction activity is insufficient; the states of different flows should also be tracked since they are critical for construction stakeholders to gain a picture of the current situation.

The AFM provides a theoretical foundation for representing the holistic picture of the CW and also indicates the interactions between various flows and information sources. But because it is a theoretical model that cannot be directly implemented as an information model to integrate the information from various information sources, an information model is demanded that can represent the concepts from the AFM. Several scholars in the construction domain have explored the research trajectory of modeling CW information since the 1990s [26,28–30]. These works have mainly focused on abstracting the construction process and conceptualizing related entities to provide references for managing construction information. Luiten et al. [28] developed a construction information model that identifies the central concepts in the construction domain, including products, activities, resources, and contracts, along with these concepts’ inter-concept relationships. This model was the initial contribution of IRMA, or the information reference model for the architectural, engineering, and construction domain [26]. Later, Luiten [29] developed the building project model (BPM), which integrates information from products, activities, and resources. Froese [26] reviewed a set of conceptual models of construction process information and proposed a core model of the construction process that includes the central elements and relationships. Meanwhile, Stumpf et al. [30] developed an object-oriented model that integrates information of the construction process and product; this integrated model defines the construction process components of construction processes and the relationships between them.
The Industry Foundation Classes (IFC) [44] are a common information standard for information exchange, especially for the BIM, which also has elements related to CW, including projects, actors, products, processes, resources, and relationships. The main elements in these models are shown in Table 1. These object-oriented conceptual models provide initial conceptualizations of CW with common terminologies. They cover the major entities of CW and have been adopted as references of later ontology development [36,37]. They are limited, however, when interpreting and representing detailed CW mechanisms with both computer-interpretable and human-readable forms.

### 2.2. Ontology works on modeling CW

The term “ontology” originated from philosophy and refers to the study of existence. In the computer science domain, Gruber [31,45] defines ontology as the “explicit formal specification of a conceptualization.” In other words, ontologies represent the knowledge in specific domains, with a formal description of the concepts and relationships. Currently, ontologies have been adopted in various domains as formal tools to share and reuse domain knowledge. In the Architecture Engineering and Construction (AEC) domain, ontology development started around 2001 and has become a crucial step to improve information and knowledge management [34]. Compared with the conventional conceptual information models mentioned previously, ontologies provide machine-readable knowledge formalization that can be used for further computer and artificial intelligence (AI) applications. Numerous ontologies have been developed to facilitate data integration and transfer, knowledge management, and information extraction under different themes in the construction domain [33,34].

Among the existing ontology works in the construction domain, a few efforts have formalized CW entities. Table 2 lists those ontologies whose vocabularies directly conceptualize CW entities, including their names and key concepts. The e-COGNOS project developed a process-based construction domain ontology in which an ontological model was applied such that “a group of Actors uses a set of Resources to produce a set of Products following certain Processes within a work environment and according to certain conditions” [36]. The e-COGNOS ontology provides portal documentation and supports consistent knowledge representation and organizational issues regarding access to and use of knowledge [37]. Under the e-COGNOS framework, El-Diraby et al. [38] developed a domain taxonomy that classifies construction concepts as processes, products, projects, actors, resources, technical topics, and systems. El-Gohary [39] presented a domain ontology for processes in infrastructure and construction (IC-PRO-Onto) that offers a conceptualization of process-centered infrastructure and construction knowledge. Besides e-COGNOS, El-Diraby [40] also introduced domain ontology for construction knowledge (DOCK 1.0) as a domain ontology for knowledge management in construction, which provides a skeleton to describe the key concepts pertinent to construction knowledge.

The key finding from reviewing these related existing ontology works is that although they provide general conceptualizations of the CW, they do not contain specific concepts about information such as information itself or the information systems that generate or obtain the information. This shortcoming leads to difficulties in directly utilizing these works as guides to integrating heterogeneous information sources to provide a comprehensive picture of the CW on the information level. Another limitation of these works is that few scholars have examined the problem of multiple contexts of information. Only in DOCK 1.0 [40] is the concept of context defined to refer to possible worlds, but no direct solution is provided for representing multi-context information based on the ontology.

These ontologies are also insufficient for modeling the detailed mechanism of activity- and flow-related entities addressed in the AFM [43]. For example, among these ontology works, the IC-PRO-Onto models activity-related constraints and provides a taxonomy of various types of constraints, but it fails to describe the exact content of constraints and relevant entities’ information. In the domain taxonomy developed by El-Diraby et al. [39], all entities addressed in the ontologies were designed with three dimensions (including state, stage, and situation), but the authors provided no detailed representation of these dimensions. We have thus not found an ontology that could adequately represent the CW and integrate the digitalized information from the CW via various information sources and systems and in different contexts based on domain knowledge.

### 2.3. Other related ontology works

In addition to ontologies that have conceptualized the CW, other noteworthy ontology works in both the construction and external domains exist that may be classified into two types: (1) ontologies with related concepts from CW-related entities and (2) ontologies that represent data and information from the ICT implementations in the construction industry. Table 3 summarizes the related ontologies.

The first type of ontologies includes the Basic Formal Ontology (BFO) [46], the PROV Ontology (PROV-O) [47], the Friend of a Friend (FOAF) ontology [48], the Organization Ontology (ORG) [49], ifcOWL [50,51], and the Building Topology Ontology (BOT) [52]. BFO is an upper-level ontology that defines the fundamental categories and their relations in order to support information integration and may be used to provide top-level terms to CW entities and relations. PROV-O represents the provenance of information generated in different systems and under different contexts. FOAF and ORG provide agent- and organization-related representation. The information from building elements and locations from BIM models that are essential to the CW can be represented with the ifcOWL or BOT ontologies.

The second type includes sensor data ontologies such as SSN/SOSA [53] and SAREF [54], the OWL-Time ontology [55] for temporal data, and the QUDT ontology [56] for units of measure. The SSN/SOSA and

| Table 1 | Models related to construction workflow (CW) and their primary elements. |
|---------|------------------------------------------------------------------------|
| **Model** | **Source** | **Primary elements in the model** |
| Information reference model for AEC (IRMA) [28] | Luiten et al., 1993 | Project objects, products, activities, resources, contracts, and agents |
| Building project model (BPM) [29] | Luiten, 1994 | Construction activities and resources, labor force, material, equipment, constructors, products, and construction states |
| Core model of the construction process [56] | Froese, 1996 | Methods, sequence constraints, utilization, input, physical entities, information, people, organizations, bulk materials, equipment, and product entities |
| Integrated construction information model (ICIM) [30] | Stumpf et al., 1996 | Projects, components, activities, schedules, construction zones, component groups, resources, and resource use |
| Industry Foundation Classes [44] (IFC) | buildingSmart (1996–2020) | Projects, elements, actors, products, tasks, resources, and relationships |

| Table 2 | CW-related ontologies. |
|---|---|
| **Ontology** | **Key concepts** |
| e-COGNOS and domain taxonomy [36–38] | Processes, products, projects, actors, resources, technical topics, and systems |
| IC-PRO-Onto [39] | Actors, resources, actions, products, projects, mechanisms, and constraints |
| DOCK 1.0 [40] | Entities (generic entities and secondary entities), abstract concepts, environmental elements, systems, attributes, and contexts |
Table 3
CW-related ontologies and ICT implementations.

| Ontology/ source | URL | Key concepts |
|------------------|-----|--------------|
| ifcOWL [50,51]   | https://standards.buildingSMART.org/IFC/DE ifcOWL/IFC4ADD2.TC1/OWL | IfcElement, IfcSpace, IfcActor |
| BOT [52]         | https://w3id.org/bot# | Buildings, elements, space, zones, stories, sites |
| PROV-O [47]      | R/prov-o/ | Entities, activities, agents |
| SAREF [54]       | https://saref.etsi.org/ | Observations, observation value, observed property, property, sensor, platform |
| SSSN and OWL-Time | https://www.w3.org/TR/vocab-ssn/ https://www.w3.org/TR/owl-time/ | Devices, tasks, services, measurements, properties |
| ORG [49]         | https://www.w3.org/TR/vocab-org/ | Organizations, sites, posts |
| FOAF [48]        | http://xmlns.com/foaf/spec/ | Agents, people, organizations, projects |
| QUDT [56]        | http://qudt.org/ | Quantities, kind, units, dimensions, data types |

SAREF ontologies provide semantical modeling of the sensor-observation process, observed properties and also the description of sensing devices. The OWL-Time ontology provides time-related concepts and properties, including time intervals and instants, durations, and these properties’ relations and value types. QUDT is a comprehensive ontology and vocabulary of quantity kinds, units of measure, and related data types. These ontologies are mature domain ontologies that formalize important aspects necessary in the representation of CW entities.

As a summary of the literature review, prior efforts and limitations are listed in Table 4. To address these limitations, the aim of the present study was to develop a set of higher-level ontologies to support the information integration of the CW within the digital construction context that could (1) represent the CW in detail with the connections of activities and flow entities, (2) integrate heterogeneous flow-related information sources from ICT-based systems, and (3) represent multi-context representational data.

3. Methodology

To achieve the aforementioned aims, we first reviewed several state-of-the-art ontology development approaches. A variety of methodologies have been established since the 1990s for building ontologies [34]. The currently most popular ontology development methodologies include the Grüniger and Fox approach [58], a system known as “METHONTOLOGY” [59], the “simple knowledge engineering methodology” (SKEM) [60], and the Uschold and Gruninger approach [61]. As shown in Table 5, these methodologies overlap but also have differences in their development processes. The Uschold and Gruninger approach provides a detailed guide to identifying the purpose and scope as well as ontology formalization, evaluation, and documentation, while the METHONTOLOGY approach emphasizes the processes of knowledge acquisition, conceptualization, implementation, and evaluation. The SKEM provides a detailed ontology building process of classes, properties, and axioms.

In this research, we have established a hybrid ontology development approach by taking the Uschold/Gruninger, METHONTOLOGY, and SKEM approaches into account as well as using the systematic framework for ontology building developed by Zhou et al. [34] as references. We made this choice to provide an explicit conceptualization of the CW and to align that conceptualization with the existing ontologies for workflow information. Fig. 1 demonstrates the ontology development approach in detail, including the four stages of specification, knowledge acquisition/conceptualization, implementation, and evaluation. The following subsections present details on the different stages in the research framework.

3.1. Ontology specification

The aim of ontology specification is to explicitly specify the scope and purpose of targeted ontologies and to determine the intended users and requirements of the ontology [59,61]. In the present research, the scope and purpose of the ontology were formalized by answering the following specification questions [59].

**What is the purpose of the ontologies?** The purpose of the digital construction ontologies (DiCon) is to offer a conceptualization of CWs, to allow the organized representation of multi-context data, to publish a shared reusable representation of the conceptualization, and to support the integration of CW information from various systems and sources.

**What is the scope of the ontologies?** The ontology will allow the representation of CWs and entities related to such workflows, the relations and attributes of entities relevant to construction management, and the representation of data in different contexts.

**Who are the ontologies’ end users?** The end users include (1) construction managers, (2) construction workers, (3) software developers in the construction domain, and (4) other stakeholders involved in the construction process, including relevant authorities and those involved in the design, logistics, and supply chain arenas. We should note that except for software developers, most users do not directly use the ontology, instead using software or applications that are further developed based on the ontology.

Based on the specified ontology scope and purpose, the functional
requirements of the ontology are typically identified by using competency questions (CQs). CQs are a set of requirements, formulated as questions in natural language, that the ontology should be able to answer [58]. These questions were used in formulating the ontology and defining its main concepts as well as the concepts’ modalities, attributes, relations, and axioms.

In this study, a series of monthly workshops were conducted with both ontology developers and domain experts to determine the CQs of the ontology during the entire period of the ontology development. The workshop participants represented a variety of construction industry sectors including an industrial software developing company, an engineering firm, a contractor, a prefabricator, a developer, and members of academia, all of whom had extensive experience and knowledge of the construction domain. Table 6 shows the details of each workshop, including their dates and themes.

Issues related to CW, construction management, ICT implementation, and information integration were discussed during the earlier workshops. The experts provided necessary knowledge of what information in the construction process is required by different stakeholders. By combining the intended purposes of the ontology defined previously and the domain knowledge provided by the experts, the ontology developers defined a set of core CQs. These CQs also contained the tentative terminology of ontology classes and relations and were also used for the later ontology evaluation process to check if the ontology had covered the desired information content and was able to represent the domain knowledge. The core CQs are presented in Table 7.

3.2. Knowledge acquisition and conceptualization

After the specification step, the next step is to determine what domain knowledge should be acquired for the ontology [59]. In this phase, relevant domain knowledge of the construction process was initially reviewed during the literature review phase. Relevant knowledge for the ontologies was also obtained from the same expert workshops that had determined the CQs.

The conceptualization phase was then conducted based on the knowledge obtained from the workshops and literature review. The major steps in this phase included listing the relevant terms in the ontology, defining a class hierarchy, defining class properties, and specifying the range and domain of the properties [60]. In the process of listing terms, most of the terminologies were directly extracted from related existing ontologies and models to ensure the clarity and unambiguity of the ontology. Then, based on the listing terms, a generic ontological model was first approached with definitions of the main classes and properties in order to facilitate the development of the ontology following a theoretical framework. In general, using an
ontological model helps to formalize the structure of the ontology and ensures that the vocabularies used in ontologies are coherent [62,63].

From an ontology engineering perspective, modularization should be considered as a way to structure ontologies. One possible definition for an ontology module is “a reusable component of a larger or more complex ontology, which is self-contained but bears a definite relationship to other ontology modules” [64,65]. Since our ontology is designed to cover the broad content of CWS entities, modularization can help to split the entities and relationships to smaller modules based on different themes to enable flexible usage and easy management of ontologies.

### 3.3. Ontology implementation

After the conceptualization, the ontologies were implemented by defining them in OWL to flesh out their details, to enable automatic ontology reasoning, and to make them machine-readable for new, ontology-based applications and tools. In OWL, concepts, relations, and attributes are modeled as classes, object properties, and data properties, respectively. The theoretical basis of OWL is in description logics [66]. This basis enables automatic reasoning to check the consistency of an ontology, to classify concepts that have been defined, and to aid in ontology integration and alignment. OWL is currently the standard language that is widely used in practice to define ontologies that can be linked to and from other ontologies in the broader ontology ecosystem. OWL allows interoperability with other ontologies, since most of the current ontologies are defined in OWL [67]. We defined the DiCon by using the Protégé ontology development environment, occasionally aided by a text editor. Protégé is a free, open-source OWL editor and framework for building intelligent systems [68] that has been broadly used by scholars and ontology engineers throughout the world.

Moreover, ontologies are generally developed to be reused [69]. According to Uschold and Gruninger’s approach [61], integration with existing ontologies should also be considered when developing ontologies by reusing or aligning relevant concepts and relations. This process can improve the reliability of the ontology and reduce the work of defining redundant concepts and relations in the ontology. A number of related ontologies have been investigated during the literature review. During the implementation phase, we thus imported and mapped these ontologies with the ontology modules to be developed.

### 3.4. Ontology evaluation

Ontology evaluation is an essential process in the development of ontologies. The aim is to check whether the new ontology satisfies specifications, fulfills its intended purpose, and meets all requirements [34]. El-Gohary and El-Diraby [39] define ontology evaluation as a “judgement of the ontology content with respect to a particular frame of reference.” Various ontology evaluation approaches have been developed in the ontology engineering domain, including gold standard evaluation [70], data driven evaluation [71], automated consistency checking, criteria-based evaluation [72], evaluation by humans, and task-based evaluation [72]. Appropriate formal evaluation approaches therefore must be taken based on the criteria, since some approaches may not fit well with the ontology and its application domain [34,72]. For example, the gold standard evaluation approach was not suitable for the present research because no existing published benchmark ontologies were available in the construction domain. These approaches also have their own target evaluation criteria that were not totally applicable to our present needs. Most previous ontology-related research in the construction domain has thus applied hybrid approaches that combine at least two approaches to test different aspects of an ontology [73].

Previous works [38,39] have used a set of criteria for evaluating the ontology. Based on the purpose of DiCon, we selected the following five evaluation criteria: clarity, coverage, consistency, extendibility, and usability. Accordingly, for this research we adopted a combination of automated consistency checking, expert workshops, criteria-based evaluation, the answering of CQs, and task-based evaluation. As shown in Table 8, each approach has various target criteria to be evaluated. Consistency checking aims to prevent contradictory facts in an ontology based on description logics, such as logical conflicts or inconsistent classes. Consistency checking is enabled by description logic reasoners, which perform various automated inferencing services [39]. Criteria-based evaluation mainly focuses on verifying the content of an ontology. El-Gohary and El-Diraby [39] have also emphasized the importance of domain expert participation in ontology evaluation, since the evaluation requires judgment with respect to abstraction, classification, and coverage based on domain knowledge. CQs serve as a frame of reference or requirement specification against which the ontology may be evaluated. The aim of task evaluation is to assess how the ontology can be used to accomplish certain tasks based on the designed purpose [63]. Further details of the description and results of the evaluation approaches are provided in Section 5.

### 4. Digital construction ontologies (DiCon)

This section presents the proposed digital construction ontologies (DiCon). We first present the ontological model in order to represent the knowledge and information of the CW before introducing the details of DiCon modules.

#### 4.1. Activities and flows

According to the theory of lean construction [41–43], each construction activity involves a set of flows: labor, equipment, workspace, components, information, external conditions, and prerequisite tasks. The DiCon need to represent these flows and their relations with activities in a formal manner. Classes should represent all the participating entities: activities, agents, equipment, locations, building objects, information entities, material batches, and so on. The representation required for the relations of activities and flow entities, however, is

| Criteria | Tools | Automated consistency checking | Expert workshops | Criteria-based evaluation | CQs | Task-based evaluation |
|----------|-------|-------------------------------|------------------|--------------------------|-----|----------------------|
| Clarity  | ✔️    | ✔️                            | ✔️               | ✔️                       | ✔️  | ✔️                   |
| Coverage | ✔️    |                               | ✔️               |                          | ✔️  |                      |
| Consistency | ✔️ |                               |                  |                          |     |                      |
| Extendibility | ✔️ |                               |                  |                          |     |                      |
| Usability | ✔️    |                               |                  |                          | ✔️  |                      |
somewhat more complicated than what may initially appear to be the case. Activities are actually related to the states of flow entities, not simply the flow entities themselves, as shown in Fig. 2, which captures the different flows of an activity. The activity can only be executed as planned if all its required flows are in proper states both before and during the execution. In a practical setting, the essential information of flow-related entities may be gathered from ICT systems such as BIM, CM, ERP, and SCM. The states of each flow can then be tracked by different systems (including sensor tracking, image processing, and CM systems) to check if they satisfy the prerequisites of an activity. Some conditions should hold continuously during the execution, such as the availability of resource-type flows. Once an activity is completed, it will have produced specific effects on some of its flows, which may then enable further activities.

Fig. 3 illustrates the ontological model of the DiCon, which was built on the AFM and ICT information sources to describe the relations between activities and related entities. Activity is the central concept, and different aspects of the construction process are associated with activity. In the ontological model, an activity was thus modeled as a process that has various entities as participants, also known as flows, including agents, material batches, equipment, locations, information content entities, and precedence activities. The object of an activity can be any entities that are the main focus of an activity. For instance, a window is the object of a window-installation activity, and a material batch is the object of a material-shipping activity. The terminologies of the classes and relations were set up by using prior models and ontologies reviewed in the literature review phase. This step was taken to ensure that the terms employed in the model were unambiguous and easily understandable by users. In general, mapping the ontology to an upper-level abstraction model will also improve the interoperability of the ontology with other models [63]. The model thus describes the components of the CW (in the lower part) and their relations with the top-level concepts from the BFO ontology [46] (in the upper part).

In accordance with the BFO, Entities in DiCon are defined as the highest level of abstraction of all the things related to the CW. Entities refer to activities, agents, material batches, equipment, building objects, locations, or information content entities. In the BFO, Entities are classified as “Occurrent” or “Continuant.” The Occurrent class refers to an Entity that has temporal parts and happens, unfolds, or develops through time; the class is further divided into the Process class and the Temporal Region class. The activities in CW occupy certain time intervals for the execution and thus may be considered as a subclass of the Process class in the BFO. In contrast with Occurrent class, flow-related elements such as Locations, Material Batches, Agents, Equipment, and Information Content Entities are physical or virtual elements that have no relation with time and are therefore defined as subclasses of Continuant, since the Continuant class is defined in the BFO as “an entity exists in full at any time in which it exists at all, persists through time while maintaining its identity and has no temporal parts” [46].

As shown in Fig. 3, all the fundamental classes in the ontological model are combined as one Entities module. To describe more detailed models of the workflow entities, including the classes of Agent, Information Content Entity, and Activity, corresponding modules are further expanded. The Agents module illustrates the various concepts and relations to describe the capabilities, roles, and organization-related aspects of the CW. The Information module was developed to provide an unambiguous description of CW-related information content entities. The Process module was used for representing the detailed CW. Besides these four modules that provide basic representations of the relations of activities and which entities are the ingredients of flows, the Variables module was built to specify the state of an entity and also the condition of an activity. The Contexts module was also included to represent multi-context data, which is an essential feature of the DiCon. The Contexts module is the conceptualization of different contexts, and the information of the entities in the various contexts is represented in certain named graphs. These modules of the DiCon are presented in detail in the following section.

4.2. Entities module

The model of the Entities module is presented in Fig. 4, in which the classes and properties are annotated with the own prefix dica and are organized with respect to the fundamental categories of BFO (with the prefixes obo and iao). In this module, the ontological model was further expanded to provide detailed representations of entities. Each Entity can have any number of identifiers (both global and local) and can be associated with different categories—for example, based on the classification systems used in the construction domain. Additionally, in this module, a class called Group was added as a refinement to represent the member-and-group relation of entities, where a group can also be regarded as an entity. In the Entities module, the temporal and spatial entities are also defined to represent the detailed space-time data of entities, thus indicating their temporal and spatial attributes.

4.3. Information module

The various types of information contents that are produced or consumed in a construction process belong to the class Information Content Entity in the Entities module. Information content entities are always about some other entities, as represented with the relation isAbout. The Information module (prefix: dici) contains various subclasses, including Contract, Plan, Message, Design, Event, Point Cloud, Certificate, Scenario, Image, Labeling, and Issue, as shown in Fig. 5. These information content entities were identified within the expert workshops during the knowledge-acquisition phase; the participants considered these entities essential to the CW, as they are mandatory for information sharing and exchange during the construction process, regarding both conventional information objects and the wealth of current ICT-based implementations. Each Information Content Entity contains information about other entities in the CW. For example, Contract is an agreement that states the specified activities as the mutual obligation of agents who perform as clients and contractors. The Plan class refers to an entity that contains a set of schedule-related activities, with the associated constraints. Message is a common literal information content for different agents to communicate and essentially has a subject and a body. Notification is a message to inform recipients in a controlled manner about a situation they need to be aware of. The Design class describes the information contents that carry the design information, which has Drawing and Building Information Models as subclasses. The Event class, which represents an occurrence that happens with a time instant for obtaining information, can be classified into Observation and Status Update to indicate the state or status of a certain entity. Image and Point Cloud are information content entities that contain visual data. Certificate refers to a document in the construction domain that certifies that one has fulfilled the requirements of and may practice in a field. Finally, the Issue class contains information about an Entity problem that has been detected and requires certain agents to respond.

4.4. Agents module

Fig. 6 presents a detailed model of the Agents module (prefix: dica). The agent, which refers to an entity capable of taking responsibility for activities, has two subclasses: Person and Organization. Organization is divided into Companies and Crew, with a Crew having a certain size. Certain agents have their own capabilities and roles when participating in the construction process. Five types of agent capabilities are defined in the module, including Construction Capability, Data Gathering Capability,
Operation Capability, Management Capability, and Design Capability. Roles is divided into individual roles, consortium roles, and Stakeholder Roles. This module can also be used in organization-related information management use cases.

4.5. Processes module

To help describe the construction process and activities in detail, Fig. 7 provides a model of the Processes module (prefix: dicp), which expands the Activity class in the Entities module. Lima et al. [37] addressed an ontological model in which construction may be considered as “a group of Actors uses a set of Resources to produce a set of Products following certain Processes within a work environment and according to certain conditions.” According to this explanation, in the processes module, Activity is further modeled as a process in which certain Agents are responsible for dealing with certain Entities. This type of activity is defined as an Object Activity, which is directly connected to the focused Entity class with a hasObject relation. In the construction domain, Object Activity usually serves a variety of types; for example, a Project is defined as an Object Activity with specific goals to be achieved.
As a higher-level ontology, however, the Processes module is not intended to classify all the subclasses of the Object Activity. Mature classifications of the processes already exist in the construction domain that can be directly implemented, for example OmniClass [74], Uniformat [75], and Talo2000 [76]. To represent different types of activities, the processes module thus uses the ClassifiedBy relation in the Entities module to indicate the classification of the Object Activity based on a certain classification.

4.6. Variables

Activities have many properties from the perspective of workflow management, such as different kinds of resources and execution times, that are gradually specified in the planning and scheduling process or can be sensed during execution. Much of the essential knowledge found in workflow management can be represented in the form of Constraints, and over significant periods during project execution, the plans can have incomplete information about the exact values of some properties, such as the start and end times of tasks, resources assigned to tasks, the locations of entities, and so on. As shown in Fig. 8, to enable the interoperable sharing of such crucial and common information, the Variables module (prefix: dicv) includes the possibility to associate a Variable with any Property of an Entity and to represent Constraints between variables.

4.7. Contexts

The Contexts module (prefix: dicc) in the DiCon provides the basic representational capabilities for representing multi-context information at the metadata level. The module allows the definition of different context frameworks in order to create contexts within the frameworks, associate content to contexts, and to compare objects and values across contexts. As shown in Fig. 9, the Contexts module includes the class of Context, which is defined as an identified realm of data, representing the circumstances in which the data can be considered true. The Context framework refers to a collection of contexts that belong together based on a certain theme. The Context set describes a set of active contexts that are

Fig. 4. Entities module and the relation with BFO concepts.
The data in different contexts is managed in different named graphs of a Resource Description Framework (RDF) dataset. The use of named graphs is an orthogonal mechanism that allows any objects to be associated with different properties in different structures and contexts [77]. The context, provenance, and property metadata are recorded at the graph level, since a graph has an identity (a uniform resource identifier [URI]) that can be associated with information. The metadata is stored in the default graph of the RDF dataset. If the dataset contains data that has a different context, then the dataset must be split into different contexts, and each of the contexts has content of the named graph to store the information about them does. Meanwhile, various Semantic Web tools work well with the named graphs mechanism; RDF can be used to create and maintain databases containing multiple graphs, while SPARQL can be used to create queries for specific named graphs.

Fig. 11 shows an example to illustrate more details of using the context module and the named graphs mechanism. Two context frameworks, including DesignContexts and ActivityContexts, refer to two themes of the data: design- and activity-related data. This metadata is stored in the default graph of the RDF dataset. If the dataset contains data that has a different context, then the dataset must be split into different contexts, and each of the contexts has content of the named graph to store the

---

**Fig. 5.** Information module of DiCon.

**Fig. 6.** Agents module of DiCon.
Fig. 7. Processes module of DiCon.

Fig. 8. Variables module of DiCon.

Fig. 9. Contexts module of DiCon.
information. In this case, the information of a particular building element loaded from different LODs (for example: LOD 200 and LOD 300) needs to be split into different LOD contexts under the DesignContexts to store the information. Two contexts for a particular activity, including planned and actual, are found under the ActivityContexts and store the as-planned and actual information.

4.8. Alignment with existing ontologies

As discussed previously, the DiCon was designed from the starting point to reuse or integrate existing ontologies as much as possible in order to enrich digital CW data content without redundant ontology modeling. Thus, besides applying the BFO as the upper-level foundation, relevant existing ontologies were also reviewed and then aligned with the DiCon. In Fig. 12, related ontologies for different information sources are mapped to corresponding classes in the DiCon with different types of relations. For example, the building element (ifc:IfcElement and bot:Element) and spatial element classes (ifc:IfcSpatialElement and bot:Zone) in both the IFC/ifcOWL [50, 51] and BOT [52] ontologies are respectively defined as subclasses of the BuildingObject or Location in the DiCon. Such linkages provide a portal to link CW with data from BIM or Linked Building Data (LBD) to use their product or spatial information, or for further related information.

Since the sensor and IoT are major data streams for monitoring the conditions of the CW, the SSN/SOSA [53] and SAREF [54] ontologies were reused to describe the sensor observation, in which the sosa:Observation and saref:Measurement were aligned to the Observation class. The FOAF [48] and ORG [49] ontologies were used as a reference to describe and formalize the information and knowledge of the agents who participated in the construction process. The related classes (foaf:Agent, foaf:Person, foaf:Organization, and org:Organization) were aligned to the corresponding classes in the DiCon. PROV-O represents provenance information generated in different systems and under different contexts. The reuse of PROV-O [47] was conducted by mapping the three basic classes prov:Entity, prov:Agent, and prov:Activity to the InformationContentEntity, Agent, and Activity classes in the DiCon. The OWL-Time ontology [55] was mapped to describe the time, while the QUDT ontology [56] was aligned to describe the unit of the property.

All the ontology modules of the DiCon were implemented in OWL during alignment to the previously mentioned existing ontologies. The current version of the DiCon has been published online as version 0.3 (BFO-ISO compliant). Only the core contents of the ontologies are shown.
5. Ontology evaluation and application

This section illustrates the evaluation process of the current version of the DiCon, in which we describe each ontology evaluation approach applied in this study and present the results of the evaluation.

5.1. Automated consistency checking

The aim of consistency checking is to ensure that no contradictory facts exist in an ontology based on description logic (DL) principles, such as logical conflicts or inconsistent classes. Consistency checking is enabled by DL reasoners, which perform various automated inferencing services [39]. In the present research, consistency checking of the proposed ontology was conducted using the Pellet reasoner, which is a built-in Protégé DL reasoner. Pellet is a complete open-source OWL-DL reasoner with reasoning support for individuals (instances), cardinality restrictions, user-defined datatypes, sub-property axioms, reflexivity restrictions, symmetric properties, and disjoint properties [78]. After using the debug function in Protégé with the Pellet reasoner, DiCon was confirmed to be consistent and coherent.

5.2. Criteria-based evaluation

The main focus of criteria-based evaluation is to verify the content of an ontology. We selected two criteria that matched the objectives of the DiCon from evaluation criteria proposed in existing studies [31,79], including clarity and extendibility.

5.2.1. Clarity

Clarity refers to whether an ontology effectively communicates the intended meaning of defined terms, which are clearly specified without ambiguity [79]. To ensure the clarity of the ontology, most of the concepts and their definitions in the DiCon were determined in compliance with existing models and domain knowledge. The concepts and relations in the DiCon were thus defined formally and unambiguously.

5.2.2. Extendibility

Extendibility refers to the ability of an ontology to be extended or expanded to describe specific application domains without changing the definitions within the current ontology [80]. Extendibility is an intrinsic feature of ontologies, which means that once the knowledge of a certain domain is captured, it can be reused and extended without changing the current definitions of concepts in the existing ontology [39]. The DiCon was designed to be a higher-level ontology, which makes the DiCon extendable by adding lower-level concepts.

5.3. Expert workshops

El-Gohary and El-Diraby [39] have stressed the importance of domain expert participation in ontology evaluation, since the evaluation requires judgment on abstraction, classification, and coverage based on given domain knowledge. When developing the DiCon, we conducted expert evaluation workshops to assess the ontology content from the user’s point of view. The participants of the evaluation included the intended end users of the ontologies and relevant researchers. A list of participants’ occupations is shown in Table 10. Although the participants were experienced in the construction domain, they were unfamiliar with the concept of ontology. The evaluation workshop was thus arranged to first provide an overview of the DiCon and then to provide a...
navigation game to ask the participants to search relevant classes and relationships in the ontology based on their domain knowledge and logics in order to answer a list of CQs. The results of the workshops indicated that workshop participants (1) found navigating the ontology to answer the CQs to be logical and easy; (2) found the terms and definitions of concepts and relations to be clear and familiar for the domain knowledge; and (3) agreed that the ontology suite covered the major concepts and inter-concept relations within the construction domain and could represent CW information from various digitalized sources.

5.4. Answering CQs

CQs are the requirement specification of the ontology, which the developed ontology should be able to answer [39]. In this research, the CQ answering was conducted in two ways. The first approach was manually conducted by the researchers and workshop participants by using logical navigation of the relevant concepts and relations to be clear and familiar for the domain knowledge; and (3) agreed that the ontology suite covered the major concepts and inter-concept relations within the construction domain and could represent CW information from various digitalized sources.

5.5. Task-based evaluation: case studies

The use of task evaluation was threefold: first, to assess how the ontology could be used to accomplish certain tasks based on the designed purpose [63]; second, to use the practical data as instance information of the ontology to answer specific CQs; third, to illustrate the practical application case of the ontology. Based on these principles, we selected two case studies for this work. The first use case involved subcontract monitoring, while the second involved resource-flow monitoring based on an indoor positioning system (IPS). Both cases were conducted using the practical project data. These two cases represent essential tasks during the construction process; they contain multiple information sources but require information integration based on domain knowledge and to achieve various objectives. The cases thus could be used to illustrate the capability of the ontologies in integrating information from multiple sources and in using the sources’ instance data to answer the CQs.

5.5.1. Case 1: subcontract monitoring

In general, the aim of subcontract monitoring is to support the general contractor in tracking and monitoring productivity and quality and to find any issues in the construction process related to a particular subcontracted scope. Subcontract monitoring data is currently entered from system to system. For example, digital tools are used to record the results of inspections and any other information about quality issues. These tools are typically not linked to other information sources, however, and therefore cannot yield a holistic picture of the CW. For this reason, in this practical case, the DiCon was used to facilitate the interchange of heterogeneous information related to subcontract activities. The key stakeholders identified in this case were the managers and engineers from the general contractor. The original data sources were obtained from various systems, including:

- the project’s construction schedule, to provide information about activities, responsible agents, and activity locations;
- a project’s architectural BIM model, to provide information on locations;
- indoor sensor data on relative humidity and temperature, to indicate the indoor environment data of a certain location at a certain time;

Table 10

| Workshop participants                      |
|--------------------------------------------|
| Construction contractor’s chief technical officer (CTO) |
| Industrial software company software developer |
| Manufacturing company engineer             |
| Construction company ICT specialist        |
| Research institute scientist               |
| University researcher                      |

Fig. 13. The process used to map the data sources to DiCon for the subcontract monitoring case.
quality-inspection information, to provide the progress and issue information of subcontract activities based on locations.

The process adopted to map the data sources to the DiCon and to generate the linked data set is shown in Fig. 13. The distributed data sources were first analyzed and manually mapped to the DiCon ontologies, as shown in Fig. 14. These sources were then converted into RDF [81] in order to instantiate the ontology. Data related to construction schedules, indoor sensors, and quality inspections was obtained in tabular format, which was then converted to RDF format by using Open Refine software. The BIM model, in IFC format, was directly converted to RDF by applying the IFC2LBD converter [82]. The converted RDF graphs of various sources were then aligned based on the common location element and stored in a triple store in the Graph DB tool together with the ontology.

After the RDF graphs were stored in the Graph DB software, SPARQL [83] queries were set up to retrieve the instance data for answering the specified task-based CQs shown in Table 12. In this case, two queries were conducted. The aim of the first query was to answer CQ.1 to CQ.5, in which the location, agent, and status information of a certain activity is retrieved. The query follows the logic based on mapping to the DiCon shown in Fig. 14, in which an activity has certain agents and is located in certain locations defined in the schedule. Each location was modeled in the BIM model with its identifier. Locations are also the targets of inspection in order to report the issue in the location or status update, respectively. These logics were further developed as a SPARQL statement, shown in List 1.a. In this case we chose the activity K31 Korjaukset (English: “quality fixes”) as an example. As shown in Table 12, the result of the query answer indicates that activity K31 Korjaukset is scheduled to occur in apartments 1, 2, and 3 and is the responsibility of subcontractor B. The universally unique identifier (UUID) of each location in the BIM model is also given. The result also shows that the inspection of the activity in all three locations has been completed and that no issues have been reported; therefore, this activity has been completed.

The aim of the second query is to answer CQ.6 and CQ.7 by extracting the indoor condition information and checking if the condition satisfies the activity constraint. This query follows the logic shown in Fig. 14, in which a location hosts several sensors to observe different types of observed properties (and their values) of certain time instants. Activities have certain constraints of indoor conditions during their time interval in their location. By comparing the constraints with the observations during activity execution, we could check if any changes in indoor conditions could potentially affect activity execution. In this case, we specified activity KEINUL1_1_A as an example, and we assumed it had a constraint whereby the indoor temperature should be above 20 degrees Celsius, and relative humidity should fall within the range of 20% to 50%. The SPARQL statement of this query is shown in List 1.b. The answer, shown in Table 12, indicates that no abnormal conditions existed during the execution of KEINUL1_1_A.

List 1. SPARQL statements and results for the subcontract monitoring case: (a) query for retrieving the location, agent, and status information of activity; (b) query for extracting the indoor condition.
PRECEDES bct: <https://w3id.org/bct#>
PRECEDES dic: <https://w3id.org/digitalconstruction/Entities#>
PRECEDES dic: <https://w3id.org/digitalconstruction/Information#>
PRECEDES foaf: <http://xmlns.com/foaf/0.1/>
PRECEDES owl: <http://www.w3.org/2002/07/owl#>
PRECEDES inst: <https://www.example.com/Keinuuadankuja/>
PRECEDES props: <https://w3id.org/props#>
PRECEDES rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PRECEDES rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PRECEDES sosa: <http://www.w3.org/ns/sosa/>
PRECEDES time: <http://www.w3.org/2005/01/XMLSchema#>
PRECEDES xml: <http://www.w3.org/XML/1998/namespace>
PRECEDES xsd: <http://www.w3.org/2001/XMLSchema#

#(a)

SELECT ?ScheduleLocation ?Company ?UUID ?StatusUpdateValue
WHERE {
  ?Activity a dice:Activity.
  ?Activity dice:occursIn ?ScheduleLocation .
  ?Activity dice:hasAgent ?Company .
  ?Activity rdfs:label ?ActivityLabel .
  ?Company a dice:Agent .
  ?ScheduleLocation a dice:Location .
  ?BIMLocation owl:sameAs ?ScheduleLocation .
  ?BIMLocation prop:globalDiffRoot_attribute_simple ?UUID .
  ?StatusUpdate a dic:StatusUpdate.
  ?StatusUpdate dice:isAbout ?QualityInspectionLocation .
  ?QualityInspectionLocation owl:sameAs ?ScheduleLocation .
  ?QualityInspectionLocation a dice:Location .
  ?StatusUpdate dic:hasStatusValue ?StatusUpdateValue .

FILTER (?ActivityLabel="K31 Korjaukset")
}

#(b)

SELECT ?Observation ?ObservationValue ?Property
WHERE {
  ?Activity a dice:Activity.
  ?Activity dice:occursIn ?ScheduleLocation .
  ?Activity dice:hasAgent ?Company .
  ?Activity rdfs:label ?ActivityLabel .
  ?Activity time:hasBeginning ?ActivityStartTime .
  ?ActivityStartTime time:inXSDDateTimeStamp ?ActivityStartTimeValue .
  ?Activity time:hasEnd ?ActivityEndTime .
  ?ActivityEndTime time:inXSDDateTimeStamp ?ActivityEndTimeValue .
  ?Sensor sosa:madeObservation ?Observation .
  ?Observation sosa:resultTime ?ObservationTime .
  ?Observation sosa:hasSimpleResult ?ObservationValue .
  ?Observation sosa:observedProperty ?Property .
  ?ScheduleLocation a dice:Location .
  ?ObservationLocation owl:sameAs ?ScheduleLocation .
  ?ObservationLocation a dice:Location .
  ?ObservationLocation sosa:hosts ?Sensor.

FILTER ( ?ActivityLabel="KEINULI_1_A" \&\& ?ObservationTime > ?ActivityStartTimeValue \&\& ?ObservationTime < ?ActivityEndTimeValue )

5.5.2. Case 2: resource flow monitoring

The second use case involved resource flow monitoring. In general, tracking the indoor position of labor can be used to monitor worker behavior and to analyze the productivity index of a construction process [9]. This use case required the integration of data, including:

- an IPS to provide real-time worker location data;
- an architectural BIM model to provide location information;
- the project’s construction schedule information to provide information on activities, responsible agents, target objects, and activity locations.

The practical project involved a three-story residential building project during the interior operation phase. The IPS applied in this case was a Bluetooth beacon and a gateway system. The gateway, attached to a stationary location, detected the nearest signal from the beacons (attached to workers) to indicate where the workers were located during a given time interval. As shown in Fig. 15, the process of ontology mapping and conversion was similar to the first case, where the data from various streams was first mapped with the ontology (as shown in Fig. 16) and then converted to RDF format with Python codes and the RDF2LBD converter. The RDF graphs were then aligned and stored in Graph DB.

For the second case, SPARQL queries were developed to retrieve the resource information (person, in terms of practical data). For this case, three queries were conducted. The aim of the first query was to answer CQ.1 to CQ.3, which retrieve relevant information of activity, location, and building object for a worker. The logic of this query is that a certain agent is assigned to an activity that occurs in a location at a certain time for a given object. The SPARQL statement of this query is shown in List 2.

in this case, we chose carpenter 1 as an example. The result is shown in Table 13. According to the schedule, carpenter 1 should go to apartment 3 to execute door installation of door 26 on June 15. The second query (shown in List 2.b) was conducted to answer CQ.4 and CQ.5 by extracting the indoor positioning data to check a worker’s on-site presence. The query starts by checking the observation made by the gateway deployed in the location and the beacon attached to the worker during a certain time interval. The third query (shown in List 2.c) answers CQ.6 by extracting the indoor positioning data to trace a worker’s on-site patterns. This query follows the logic in which all the locations had the gateway observation for specific beacons attached to the workers. Table 13 shows the results, where both carpenter 1 and 2 appeared in the apartment on June 6 between 1 p.m. and 2 p.m. On June 5, between 7 a.m. and 8 a.m., carpenter 1 first visited the entrance of the building, then was observed in the staircase location, and then went to apartment 1.

List 2. SPARQL statements and results for the resource case: (a) query for retrieving the location, activity, and activity object information for an agent; (b) query for the presence of an agent; (c) query for the moving pattern of an agent.

![Fig. 15. The process used to map the data sources to the DiCon for the resource flow monitoring case.](image-url)
Y. Zheng et al.
The two cases illustrate that the DiCon is able to integrate the data from multiple digital systems and to answer the task-specified CQs, as it provides a representation of CW knowledge that could be used to integrate various pieces of CW information and data sources and to make the information retrievable by stakeholders. In the first case, the various data/information streams were combined to build up the integrated picture of subcontract workflows. The constraint of the activity was also used for comparison with the tracking data to identify if variabilities could occur. In the second case, the integrated information from the IPS, the BIM model, and the schedule could be used to support the workers and site managers in retrieving necessary information to support their jobs. These cases show that, based on the ontology, the integrated data could be used for further applications such as querying and information retrieval to support decision-making and action-taking in the construction process, which also demonstrates that the DiCon satisfied the criteria of coverage and usability.

5.6. Summary of ontology evaluation

In summary, we employed five approaches to evaluate the DiCon, and the evaluation results were positive. First, by using the Pellet reasoner in the Protégé environment, no inconsistencies have been detected, which confirmed that all the ontologies in the DiCon were consistent and coherent. Second, the results from the expert workshop suggested that the ontology could capture key concepts and relations within the domain with unambiguous terminologies and definitions and could be used for further applications. Third, the results of the criteria-based evaluation indicated that all the ontologies in the DiCon were clear and extendable. Fourth, we applied the answering CQs approach during the expert workshop and the task-based evaluation. The result of answering CQs shows the DiCon can answer the CQs, this is a proof of satisfactory of the coverage requirement. Finally, the task-based evaluation demonstrated that the DiCon was capable of achieving the competency tasks defined earlier via the competency questions. The DiCon captured all essential concepts to serve the design purpose. The coverage and usefulness of the DiCon were further proved by the task-based evaluation. Thus, the DiCon was assessed to be competent, consistent, concise, clear, and usable.

6. Discussion, limitations, and future research

As discussed previously, the major challenge to integrating the information and data acquired via various ICT implementations in order to interpret the comprehensive situation of the CW is information heterogeneity. Although ontologies are effective tools to manage heterogenous and unstructured information, previous efforts have mainly focused on representing the domain knowledge of CW but have been inadequate in representing information from current ICT systems. We thus developed the DiCon to address this challenge. The major contributions of the proposed DiCon are (1) to provide a comprehensive vocabulary to represent CW knowledge and information based on domain knowledge, including the activity flow model (AFM); (2) to fill the gaps of existing ontologies that lack information-related entities to support the integration of heterogenous and multi-context information; (3) to enable further applications based on linked data and Semantic Web technologies for information retrieval; and (4) to provide an unambiguous information model to promote construction management systems.
First, the DiCon was built by thoroughly acquiring domain knowledge of the CW, especially the theory of AFM [43], which makes the proposed ontology set a more specific and detailed formalization of the CW. The proposed ontologies are able to represent activities and conditions, related entities, and their states to build a picture of the CW. The ontology was also built on previous efforts regarding existing information models and ontologies and was refined through a set of expert workshops, where the experts provided suggestion-based practical experience. The DiCon thus was able to represent CW knowledge and information, which also form the foundation of information integration.

Second, existing data/information models or ontologies for CW focus on representing domain knowledge with common terminologies, but the information and multi-context aspects are neglected in such models and ontologies. The major purpose of the proposed ontology is to specifically integrate the information acquired from various systems and different contexts. The DiCon was built by collecting knowledge of CW, ICT implementation, existing ontologies, and data and thus can not only cover the domain knowledge of CW but can also be used to integrate heterogeneous information based on the current ICT tools in the industry. The combination of using the named-graph approach enables the DiCon to represent multi-context information. The reuse of the existing ontologies is another important feature of the DiCon that enhances the capability of integrating information acquired from ICT systems. Even though most previous ontologies with CW-related content were also coded with OWL/RDF, the creators of these ontologies merely discussed the reuse of or linkage with existing external ontologies that represent the contents of ICT implementations. The DiCon was specifically developed to reuse or merge existing ontologies. For instance, in the DiCon, the Building Object class can be directly linked to the IfcBuildingElement class in IFCOWL ontologies through the OWL property equivalent class. The DiCon thus is highly compatible with existing ontologies.

Third, the DiCon may be considered the foundation for implementing linked data and Semantic Web applications for CW information management. The cases in the task-driven evaluation not only demonstrated the feasibility of the DiCon in solving the integration problem of CW information but also illustrated the use of SPARQL to achieve information retrieval. Because the DiCon was coded based on description logic, it could be used to conduct reasoning based on defined logics and could also be combined with the Semantic Web Rule Language (SWRL) to enrich reasoning functionality. Once sufficient information is collected, the DiCon can be used for various essential management functions, for example detecting task variability, running automatic rescheduling procedures, identifying critical constraint conflicts, and calculating different performance indices, all of which are difficult to achieve using conventional tools alone.

Fourth, construction managers currently suffer from a lack of systemic information management and delayed information [84]. The increasing advancement of IoT and sensing technology has also enabled various comprehensive monitoring and controlling methods of the construction process, such as digital twins information systems for construction. The phrase “digital twins” refers to a data-centric management of the physical construction process by collecting the data streaming from a variety of site-monitoring technologies that collect accurate status information, thus supporting decision-making and action-taking [85]. The use of digital twins in construction, however, involves large amounts of data from different systems and requires formalized and systematic integration. The features of the proposed ontology could systematically represent the digital twin approach by providing an unambiguous data structure to handle and integrate the data. Based on the DiCon, the data could also be converted to RDF format and published online to support cloud data storage, access, and processes. The DiCon thus can support the establishment of novel construction management systems, especially those such as digital twins that involve enormous amounts of heterogeneous data.

The ontology suite developed in this research has the following limitations. First, the DiCon was designed to be a set of higher-level ontologies that are only intended to capture essential higher-level concepts and properties to represent CWs within the digitalized construction context. The ontologies thus do not build up a taxonomy that would cover the detailed classification hierarchies of the domain entities. For specific use cases, the ontology suite might need to be extended with detailed classifications of domain entities.

Second, although the DiCon was evaluated with five approaches, the expert-workshop and criteria-based evaluation approaches are subjective, and criteria such as clarity, extendibility, and ease of use are difficult to quantify and assess. Previous ontology-related works in the construction domain have reported on this issue [38,73]. In future studies, quantifiable ontology metrics should be considered for additional validation. For example, Tartir et al. [86] have proposed relationship richness, attributes richness, and readability as quantitative measures. These metrics can be used to evaluate the knowledge representation of the DiCon. Relationship richness, which refers to a schema metric that indicates the diversity of relations in the ontology, can be calculated by dividing the number of relationships defined in the DiCon by the sum of the number of subclasses and the number of relationships. Attributes richness, defined as the average number of attributes for all the classes, can be used to assess the richness of information pertaining to instance data. Attributes richness can be calculated by averaging the total number of attributes defined in the DiCon with the number of classes in the DiCon. In terms of readability, this metric indicates the level of human-readable descriptions for each class in the ontology, which can be represented by the total number of comments and labels of one class.

Third, during the task-based evaluation, although we did obtain various digital data sources to create the two use cases, the data we acquired could not cover all the content that we had designed in the ontology. This situation occurred because limited types of digital data could be acquired from the current practical project. The ontology we developed considers much more than the actual digital and ICT implementations in construction. With increasing ICT deployment in the construction domain, more digital data sources could be acquired to support more specific use cases and ontology evaluations. The emergence of the DiCon could also stimulate such ICT implementation in the construction domain. Since the DiCon can be used to improve the interoperability of heterogeneous data from different ICT systems, users would gain the benefits of the integrated database by implementing more digital applications.

We should also note that the process of building the RDF skeletons of the tabular data streams to be mapped with the DiCon for data conversion was a manual process. Some existing tools can be used to convert the tabular data to RDF: for example, OpenRefine, applied in the case study, and the Relational Databases to RDF Mapping Language (R2RML) [87], which is used to express the customized mappings from relational databases to RDF data sets. But the process still requires manual mapping. There is currently no automated solution to direct the bridge schema or the structure of various tabular data to the DiCon, and software providers do not yet support the DiCon. This manual mapping process is a bottleneck for potential ontology users to implement the DiCon or other ontology-based approaches, since users will encounter difficulties because they must be highly familiar with both the target data schema and the ontology [88]. The manual process can also be costly, especially because tabular data is the most common format for the structured data sources found in the digital construction arena.

The DiCon ways could be considered to facilitate automated ontology mapping for the DiCon. The first is to convince software providers that implementing data models based on ontologies such as the DiCon will be necessary in the future. Several discussions are already underway, and leading software providers have expressed interest in developing a standard. Because agreeing on a standard is a long process, however, an automatic solution for mapping tabular data to the DiCon ontology could be considered. The second approach is to achieve automatic mapping based on technical approaches. One major concern for
automatic mapping, however, is that data schemas are usually not unified but are instead customized by different users. Consequently, although the theme of the data may be same, the terminology or the content of the schema could be different for distinct users and applications. Therefore, for various tabular data, it is difficult to create a generalized converter, such as the IFC2LBD converter [82], based on the unified data standard. One possible solution to address this challenge would be first to use the Natural Language Processing (NLP) and Artificial Intelligence (AI) approaches to parse the terminology of tabular data headers [89], to map these headers to the corresponding classes or properties (object or datatype) in the ontology [90], and then to automatically convert the instance data [88,91].

There is also no “perfect” ontology, and ontologies require iterative maintenance and refinement [39], since the ontology representation should be in compliance with dynamic changes in domain knowledge and interests [81]. For example, the construction domain might see an increased implementation of robotics in the future. The executors of construction activities thus could be autonomous agents, which would also change the labor flow of the activity and lead to changes in related contents in the ontology.

The use of DiCon ontologies is just the first step in integrating the information of the CW. These ontologies are still under active maintenance, refinement, alignment, and refactoring, with the results being released in future versions of the DiCon. Future extensions and applications could be created in the following areas. First, the DiCon offers rich semantics for the CW in which ontology-based information and knowledge management systems could be developed to support the construction management and decision-making processes. The research group is currently developing a system for realizing the situational awareness of the CW based on the ontology suite to support operation management. The DiCon also has potential beyond the basic functions for knowledge and information management, including the ability to be combined with machine learning frameworks with a massive formalized semantic database, which could enable important practical applications.

Second, the further implementation of the ontology suite with specific use cases should also be explored, including supporting construction logistics management, quality management, and the implementation of AI based on workflow information. These further cases might require extensions to the basic DiCon ontologies. For example, we have developed a simple ontology extension for the construction logistics [92] and currently are working on building a DiCon extension for on-site image semantics to support AI-based computer vision applications. In addition, because implementing the DiCon requires proficient ontology and Semantic Web knowledge and skills, future studies should also consider developing an application programming interface (API) library that could help end users to more easily adopt the ontology, which also includes the aforementioned NLP-AI-based tabular conversion function.

Finally, various domain ontologies and relevant works have been developed in the construction industry, including on the themes of construction safety [73,93], construction planning [94], and project management [95]. Any obvious overlaps among the classes and relations could be identified from these works. To improve the practical use of the DiCon, the ontology linking process should be automated with these ontologies. One possible option would be to use NLP-based processes to extract and link similar entities and relations, and then to map the ontologies together [90].

7. Conclusion

To support the information integration of the construction workflow (CW) found within the digital construction context, this paper has presented a suite of ontologies called the digital construction ontologies (DiCon). The DiCon was designed to be a unified representation of the detailed CW with related entities and relations. This ontology suite can also serve as a reference for integrating CW information from various information systems in the digital construction context and managing multi-context information by providing a standard semantic model. According to the definition provided by Gruber [31] and the result of ontology evaluation, the DiCon is a formal, explicit specification of a shared conceptualization of CW information that also takes account of the current digital construction context.

The proposed ontology suite offers a comprehensive ontology set for the construction domain, with a twofold core contribution. First, the DiCon offers a formal and enriched knowledge representation of the CW and can serve as a formal information reference to integrate heterogeneous CW information within the digital construction context. By implementing the ontology, construction stakeholders and software vendors can build up a customized construction information management system that may integrate distinct pieces of information from various sources within the digital construction context for their specific purposes. Second, the ontologies found in the DiCon were coded with the OWL language, which is machine readable and can potentially help further computing or Semantic Web applications in support of construction management. Further applications may thus be built on this ontology suite. Specifically, the DiCon may be utilized for applications to support the decision-making process of CW management, such as reasoning and information retrieval.

Declaration of Competing Interest

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

Acknowledgements

This research was partly supported by the DiCtion (Digitalizing Construction Workflows) project, which was funded by Business Finland and a consortium of six industrial partners (until 2020). It was also partly funded by the BIM4EEB project that has received funding from European Union’s H2020 research and innovation programme under grant agreement N. 820660 (from 2019–). Financial support was also provided by the Building 2030 consortium of 21 Finnish companies (from 2020–). The content of this publication reflects the authors’ view only, and Business Finland, the European Commission or the Building 2030 consortium are not responsible for any use that may be made of the information it contains. The authors are grateful for the study’s financial and technical support.

References

[1] B. Dave, S. Boddy, L. Koskela, Improving information flow within the production management system with web services, in: 18th Annual Conference of the International Group for Lean Construction (IGLC), Haifa, Israel, 2000, pp. 14–16. https://iglc.net/Papers/Details/738. Accessed date: 12 December, 2020.
[2] R. Sacks, R. Navon, I. Brodetskaia, A. Shakira, Feasibility of automated monitoring of lifting equipment in support of project control, J. Constr. Eng. Manag. 131 (2005) 604–614, https://doi.org/10.1061/(asce)0733-9364(2005)131:5(604).
[3] R. Kirkkäinen, R. Lavikka, O. Seppanen, A. Peltohoeri, Situation picture through construction information management, in: 10th Nordic Conference on Construction Economics and Organization, Tallinn, Estonia, 2019, pp. 155–161, https://doi.org/10.1108/S2150-2510-2020-00010.
[4] S. Taneja, B. Akinci, J.H. Garrett, L. Soibelman, E. Ergen, A. Pradhan, et al., Sensing and field data capture for construction and facility operations, J. Constr. Eng. Manag. 137 (2011) 870–881, https://doi.org/10.1061/(asce)co.1943-7862.0000332.
[5] M.W. Park, C. Koch, I. Brilakis, Three-dimensional tracking of construction resources using an on-site camera system, J. Comput. Civ. Eng. 26 (2012) 541–549, https://doi.org/10.1061/(asce)0887-3801(2012)26:6(541).
[6] M. Ibrahim, O. Moroselli, Experimental study of wireless sensor networks for indoor construction operations, in: Proceedings of 31st International Symposium on Automation and Robotics in Construction (ISARC), International Association for Automation and Robotics in Construction (ISARC), Sydney, Australia, 2014, pp. 805–812, https://doi.org/10.22260/narc2014/0109.
[7] B. Dave, S. Kuhler, K. Framling, L. Koskela, Opportunities for enhanced lean construction management using internet of things standards, Autom. Constr. 61 (2016) 86–97, https://doi.org/10.1016/j.autcon.2015.10.009.
M. Grüninger, M.S. Fox, Methodology for the design and evaluation of ontologies, in: Workshop on Basic Ontological Issues in Knowledge Sharing, Montreal, Canada, 1995. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.8723. Accessed date: 14 December, 2020.

M. Fernández-López, A. Gómez-Pérez, N. Juristo, METHODOLOGY: from ontological art towards ontological engineering, in: Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series, California, USA, 1997. http://oas.ua.ac.be/5484/. Accessed date: 14 December, 2020.

N.F. Noy, D.L. Mcguinness, Ontology Development 101: A Guide to Creating Your First ontology, Stanford Knowledge Systems Laboratory Technical Report KL-01-05 and Stanford Medical Informatics Technical Report SM-2001-0880. https://protege.stanford.edu/publications/ontology_development/ontology101.pdf. 2001. Accessed date: 14 December, 2020.

M. Uschold, M. Grüninger, Ontologies: Principles, Methods and Applications, The Knowledge Engineering Review vol. 11, 1996, pp. 93–136. https://doi.org/10.1017/S0269888900007797.

F. Baader, D. Calvanese, D. McGuinness, P. Patel-Schneider, D. Nardi, The Description Logic Handbook, second ed., Cambridge University Press, 2001. ISBN 0-521-78197-7. Accessed date: 14 December, 2020.

M. D'Aquin, A. Schlücht, H. Stuckenschmidt, M. Sabou, Ontology modularization for knowledge selection: Experiments and evaluations, in: R. Wagner, N. Revell, G. Pernul (Eds.), Database and Expert Systems Applications, Lecture Notes in Computer Science vol. 4653, Springer, Berlin, Heidelberg, 2007, pp. 874–883, https://doi.org/10.1007/978-3-540-74469-6_85.

J. Franc-Mensah, W. O'Brien, A shared ontology for integrated highway planning, Adv. Eng. Inform. 41 (2019) 100929. https://doi.org/10.1016/j.aei.2019.100929.

P. Doran, Ontology re-use via ontology modularisation, in: Knowledge Web PhD Symposium 2006, Budva, Montenegro. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.9581&rep=rep1&type=pdf. 2006. Accessed date: 14 December, 2020.

D. Kalibatiene, O. Vasilecas, Survey on ontology languages, in: J. Grabis, M. Kirikova (Eds.), Perspectives in Business Informatics Research, Business Informatics Research 2011, International Conference on Business Informatics Research, Lecture Notes in Business Information Processing vol. 90, Springer, Berlin, Heidelberg, 2011, pp. 124–141, https://doi.org/10.1007/978-3-642-24514-1_10.

M. Morridge, H. Knublauch, A. Rector, R. Stevens, C. Wroe, A Practical Guide to Building OWL Ontologies using the Protege-OWL Plugin and CO-ODE Tools Edition 1.0. http://mowel.cs.man.ac.uk/protegewotutorial/tools/ProtegeOWL TutorialP4 v1.3.pdf. 2004. Accessed date: 14 December, 2020.

W. Borst, Construction of Engineering Ontologies for Knowledge Sharing and Reuse, PhD Dissertation, University of Twente, Enschede, the Netherlands, 1997, ISBN 90-365-0988-2. https://research.ute.nl/en/publications/construction-of-ontology-ontologies-for-knowledge-sharing-and-reuse. Accessed date: 14 December, 2020.

K. Blößscheidt, S. Staab, On how to perform a gold standard based evaluation of ontology learning, in: L. Cruz (Ed.), The Semantic Web - ISWC 2006, International Semantic Web Conference (ISWC), Lecture Notes in Computer Science vol. 4273, Springer, Berlin, Heidelberg, 2006, pp. 228–241, https://doi.org/10.1007/11126078_17.

K. Brewster, H. Alani, S. Dasmahaptra, Y. Wilks, Data driven ontology evaluation, in: International Conference on Language Resources and Evaluation, 2004. https://www.ercim-quaero.org/collections/ercim-quaero-2004/010806/. Accessed date: 14 December, 2020.

P. Delièr Haghiö, F. Burstein, A. Zaslavsky, P. Arbon, Development and evaluation of ontology for intelligent decision support in medical emergency management for mass gatherings, Decis. Support. Syst. 53 (2012) 1192–1204, https://doi.org/10.1016/j.dss.2012.11.013.

B.H.W. Guo, Y.M. Goh, Ontology design for dynamic of active fall protection systems, Autom. Constr. 82 (2017) 138–153, https://doi.org/10.1016/j.autcon.2017.02.009.

Y. Zheng et al. Automation in Construction 132 (2021) 103930