Work package-based information modeling for resource-constrained scheduling of construction projects

Heng-Wei Wang[a,b], Jia-Rui Lin[a,b,*], Jian-Ping Zhang[a,b]

[a]*Department of Civil Engineering, Tsinghua University, Beijing, China 100084*

[b]*Tsinghua University-Glodon Joint Research Centre for Building Information Model (RCBIM), Tsinghua University, Beijing, China 100084*

Abstract

As an essential problem in construction management, the resource-constrained project scheduling problem (RCPSP) has been studied for decades; however, an integrated information model that fully supports the RCPSP solving process is still lacking. Though building information modeling (BIM) was proposed to meet the data requirements in the building life cycle, some scheduling and resource information are not considered in information transfers between the information model and the RCPSP mathematical model. This paper presents an integrated approach that enables fluent data flow from the information model to the RCPSP model for construction scheduling. Within this approach, a work package-based information model is proposed to capture all the required data of the RCPSP. Then, a semiautomatic method that integrates multisource data is introduced to form the proposed information model, and an adaptive data transmission method is used to support a designed multimode resource-constrained project scheduling problem (MRCPSP) model. The models and approaches are validated using the data of an actual project, demonstrating the feasibility and efficiency of this approach. This study contributes a novel integrated approach that covers the information requirement and enables fluent data flow in the RCPSP solving process by formalizing a construction information model with a semiautomatic data integration approach. Meanwhile, the work package-based information model is a successful attempt to introduce previously-gained knowledge into automatic schedule generation processes. Future work like extending the information model, creating new methods for RCPSP model generation, and data analytics could bring a new chance to apply more complex and intelligent methods in project scheduling and construction management.

Key words: information modeling; data integration; resource-constrained scheduling; work package; constraint programming.

1. Introduction

Project scheduling is an essential problem in construction management[1]. The critical path method (CPM) is an early solution that can appropriately allocate resources at a minimum total cost when the objective is to minimize the total duration. However, the CPM may not obtain the optimal solution, and it is not able to consider resource constraints and deadlines[2]. The RCPSP, which is an NP-hard problem, is more comprehensive than the CPM. Heuristic algorithms are used to solve the RCPSP, including priority rule-based scheduling heuristics such as truncated branch-and-bound and local search techniques[3] and metaheuristics such as genetic algorithms (GA), particle swarm optimization (PSO), and the tabu search[2]. For construction projects, many studies have established more complex problem models to meet practical needs, such as considering multiple projects[4], changing resource constraints[5], and simultaneously considering resource-constrained and time-cost trade-off problems[2]. In general, all these problems can be classified as RCPSP. However, a significant flaw in these studies is that they do not take the difficulty of data acquisition into account. Although the scheduling problem model can meet the demands of actual projects and can be solved within a short time, it is difficult to apply to actual projects because the data input may require a significant amount of time and labor. Therefore, a trusted structured data source is needed to provide the schedule, resource, and cost information for project scheduling. Also, a method to transfer data from the data source to the problem's mathematical model is necessary.

BIM can meet this requirement. BIM provides a full life-cycle and extensible data integration approach for construction projects[6]. Researchers have associated building information models with known data structures to generate BIM-based integration models[7]. At the full life cycle scale, the model extended from BIM is defined as the nD model[8]. Then, during the process of converting from 3D to nD, the schedule information is mainly integrated with BIM through 4D modeling[9]. Using the geometric information from the BIM, the 4D model can be used to visualize the construction progress, display the construction process in advance, assist with scheduling, or help solve potential construction collisions[10,11]. Resource information can be directly integrated with BIM. For example, associating BIM with enterprise resource planning information can assist in schedule monitoring and material management[12]. Meanwhile, some studies associate BIM with a bill of quantities to be applied during the life-cycle cost analysis[13]. It is possible to generate a schedule using BIM. Kim et al.[14] provided a method to export schedule data using complete data from the Industrial Foundation Class (IFC). Lu and Thomas[15] proposed a method for BIM extension and discrete event simulation. Wang et al.[16] presented a method that used BIM and productivity to analyze process time and generate a schedule model. Liu et al.[17] proposed a method to generate a simulation model of RCPSP by correlating BIM with resource information and solved the simulation model using PSO. Sina et al.[18] proposed a method that used BIM and other information to generate a simulation model for the formwork scheduling problem by considering path and spatial constraints and then optimized the simulation model[18].

The above studies indicate that the information provided by BIM is incomplete for scheduling. Although it is possible to enrich BIM by extending IFC, few studies use the extended IFC as a data source. Instead, most use the information modeling method based on BIM to compensate for the lack of information. However, integration models developed in this way typically mainly consider usability, while completeness, which is essential for prospective research and applications, is rarely discussed. Moreover, the data source and preparation method for the information not provided by BIM is neglected. Besides, few studies have considered the efficiency of information integration, which will affect the practicability of the research results. Using the integration model for scheduling, no defined RCPSP mathematical model but customized simulation models with less general exist. The existing integration methods for scheduling not only make few contributions to RCPSP modeling, but also be hard to be promoted in most engineering projects.

It is necessary to design a new information model, find a data integration method for the model, and realize data transfer from the model to the RCPSP for solving the data acquisition problem when generating RCPSPs in construction projects. By analyzing the existing research on RCPSP, this study proposes a work package-based integrated information model that can be used as the data source for RCPSP models and has high completeness. Moreover, a semiautomatic data acquisition and integration approach is proposed to generate the information model for quick data preparation. A unique MRCPSP model is designed by constraint programming (CP) to fit the integrated model, providing scheduling results for analysis after obtaining data from the information model, and solving the problem.

The main contributions of this study to the body of knowledge include the proposed integrated information model that covers the information required by most RCPSPs, the semiautomatic data acquisition and integration approach that enables high efficiency, the knowledge-based work package data basis that reduces duplication of works, and bringing more possibilities to define more sophisticated and practical RCPSP models.

In this paper, Section 2 summarizes the research on RCPSP in single construction projects and related research on BIM-based information modeling with schedule and resource information. Section 3 describes the overall research methodology. Section 4 presents the information modeling results—the integrated information model. Section 5 introduces the data integration approach to generate the integration model. Section 6 explains the designed MRCPSP model and the method for obtaining data from the integration model. Section 7 provides the results and discussion of the model verification. Finally, Section 8 presents the conclusions.

2. Literature review

2.1. Resource-constrained scheduling problems of single construction projects

The basic RCPSP defines a series of activities with precedence relations. Each activity has a fixed duration and occupies several reusable resources during operation. These resources may have fixed capacities. On this basis, a variety of problems appear, among which the MRCPSP is mainly used in construction. The basic MRCPSP adds only some combinations of the duration, resource requests, and cost of each activity. Each combination is usually called a mode. MRCPSP can derive many variations by introducing different variables or different constraints[19]. For example, the discrete time-resource tradeoff problem (DTRTP) considers the relationship between the activity duration, workload, and resource requests. Because this relationship can also be considered as several combinations of duration and resource requests, DTRTP is a particular case of MRCPSP. Another particular type of MRCPSP is the discrete time-cost tradeoff problem (DTCTP), which introduces nonrenewable resources and uses resource-based or cost-based objectives to consider resource or activity costs.

Through a literature search, we obtained 24 RCPSP studies for single construction projects; Table 1 presents the information required by the RCPSP models. The activity durations, precedence relations, and resource capacities are the basic information requirements. Many studies also consider multiple modes and costs. Only three studies consider resources with different renewability; however, both renewable and nonrenewable resources exist and should be considered in construction projects. None of the studies above considers all seven information requirements. This study covers all of the requirements to improve the complicacy and completeness to serve to schedule in practical construction projects.

In these studies, the activity durations are set to fixed values or influenced by resource usage or activity cost through multiple modes. Different types of modes exist, for example, the combination of duration and resource usage, the combination of duration and cost, or the combination of all three factors above. When considering resource usage, only one renewable resource, labor, is generally considered. Cost is divided into direct costs and indirect costs. The direct costs include the cost of each activity, but resource prices are rarely considered, while indirect costs are generally obtained by estimation. For the precedence relations, finish-start relations are the primary consideration, some of which have lags. The resource capacity is generally defined as a constant upper limit, but some studies consider both upper and lower limits, and some studies consider the changes in resource capacity over time.

Table 1. Information requirement of RCPSP studies for single construction projects

Works	Activity Durations	Multiple Modes	Precedence relations	Release dates and deadlines	Resource capacities	Resource renewable type	Cost
Leu et al.[20]	√	√	√		√		
Zhao et al.[21]	√		√		√		
Chung-Wei et al.[22]	√		√		√		√
Christodoulou[23]	√		√		√		

Zhang et al.[24]	√	√	√		√		√
Menesi et al.[25]	√	√	√	√	√		√
Ghoddousi et al.[26]	√	√	√		√		√
Cheng et al.[27]	√		√		√		
Liu et al.[17]	√		√		√		
Bettemir et al.[28]	√		√		√		
Menesi et al.[29]	√	√	√		√		
Cheng et al.[30]	√		√		√		
Abuwarda et al.[5]	√	√	√		√	√	√
Abuwarda et al.[31]	√	√	√		√		√
Zhang et al.[32]	√		√		√		√
Sonmez et al.[2]	√	√	√		√		√
Kaveh et al.[33]	√		√		√		
Sonmez et al.[34]	√	√	√		√		
Kim et al.[35]	√	√	√				√
Elbeltagi et al.[36]	√	√	√				√
Siu et al.[37]	√		√		√		
Duc-Hoc et al.[38]	√		√		√		
Liu et al.[39]	√		√		√		
Giran et al.[40]	√		√		√		
This paper	√	√	√	√	√	√	√

The information requirement can indicate the complexity of the scheduling problem model. Six of the 24 studies considered five or more of the five information types mentioned above. Three of these studies used the CP approach, which is suitable for solving RCPSP[41], for problem modeling. For construction projects, Menesi proved that CP performs better for solving large-scale multiconstraint RCPSPs compared with the performance of metaheuristic methods[25] and demonstrated a method for solving multiobjective problems with CP[29]. Abuwarda used CP for scheduling problem modeling in construction. The model considers the path selection, external activities, multimode and multitype activity relations, time interval, nonrenewable resources and resource constraints changing over time and it uses total time or total cost as the optimization objective[5]. The complex activity relations caused by multiple modes were also considered[31]. However, none of the above studies using CP for scheduling considers the data source issue; therefore, it is difficult to apply them in practical projects.

2.2. BIM-based schedule and resource information modeling

The major studies regarding BIM-based schedule and resource information modeling in recent years are listed in Table 2. [10] and [42] performed typical works of nD modeling and application. The information integration of the nD model is based primarily on the work breakdown structure (WBS)[43]. It is common to generate schedules based on WBS. When associated with WBS, BIM can be associated with the schedule. The main advantage of this model is that the existing schedule can be used for modeling, which means that the existing scheduling method can be retained. However, the drawback of this approach is that it is necessary to make unique rules for the WBS to make the automatic association between BIM and WBS possible. Nevertheless, such WBSs are too rigid when used in actual projects, and they may require more labor. More importantly, changes to the WBS will affect the associations in the nD model and may even invalidate some data.

Some studies integrated BIM with a process simulation model, which can effectively optimize the construction process[44], for scheduling. Wang et al.[16] used BIM for quantity take-off and then associated BIM with the process simulation model by using construction work zones. In this study, each zone has a subsimulation model that uses the resource quantity from the quantity take-off and parameters such as productivity and resource limits to achieve automatic scheduling. The focus of this study was reinforced concrete engineering. The types of components involved included walls, beams, slabs, and columns, while the resources involved included formwork, concrete, reinforcement, workers, and cranes. Compared with that by Wang et al.[16], the method proposed by Liu et al.[17] has a broader scope of application. They used a work package to define the construction process and utilized the required labor resources and corresponding productivity of each category of building components. The work package can be attached to the building components in BIM by the component categories, and the simulation model is defined as the process pattern for different component categories. This integration model is component-centric, and some of the construction sequences can be generated automatically. WoonSeong et al.[45] integrated BIM with a simulation model for construction operations. For scheduling, they used the materials information defined in BIM and added labor requests in the simulation model of construction operations. However, the models related to the process simulation model mentioned above are limited to unique engineering domains and unique construction processes. Rarely is a general scheduling model addressed that can cover most engineering projects, such as the basic RCPSP model. Moreover, few of them consider optimizing the cost.

Schedule and resource information can also be integrated by defining new data formats. Koenig et al.[46] defined process sequences and the association between BIM and process patterns by partial model query language (PMQL). Xue et al.[47] extended the construction management content for IFC. Mehrdad and Saeed[48] proposed a knowledge base for construction scheduling. The efficiency of information integration is augmented by retrieving the knowledge bases for BIM and the schedule using SPARQL. These studies provided new data formats for integrating BIM and schedule and resource information, but some created only common data formats, without considering suitable integration approaches. Moreover, even when an integration approach was proposed, none of the studies considered whether the source data were reliable and whether the approach was appropriate for actual construction.

In addition to data integration, the importance of the scheduling data acquisition efficiency is also recognized by researchers. Knowledge-based methods show feasibility to improve the scheduling data acquisition efficiency. Sigalov and Konig[49] presented a method for automatically detecting the construction process patterns to generate process templates, which can reduce the planning time when scheduling based on BIM. Yeoh et al.[50] proposed a kind of construction method template to generate construction requirements from BIM and other data for further automatic scheduling generation. The existing knowledge bases for scheduling rarely support both schedule and resource information, and contain limited details, making them difficult to apply in practical projects.

Table 2. Summary of recent BIM-based schedule and resource modeling works

Works	Integrated information (Beyond BIM)						Data integration method
	Schedule	Resource	Site	Cost	Quality	Structural	
Zhang and Hu[10]	√	√	√			√	Manual
Wu and Hsieh[42]	√			√	√		Manual
Wei-Chih et al.[16]	√	√					Semiautomatic
Liu et al.[17]	√	√					Automatic
WoonSeong et al.[45]	√	√					-
Koenig et al.[46]	√	√					Semiautomatic
Xue et al.[47]	√	√		√	√		-
Mehrdad and Saeed[48]	√	√					Manual

2.3. Summary of previous studies

An RCPSP can include seven types of information: activity durations, multiple modes, precedence relations, deadlines, resource capacities, resource renewability types, and costs. In the RCPSP models used by existing studies, the forms of these types of information are not unified. To simplify the problem model, these types of information were simplified or considered pertinently. At present, no research has considered all seven types of information, that is, a comprehensive RCPSP model is still lacking.

Meanwhile, less than a quarter of these studies considered more than five information categories, and half used CP for problem modeling. CP has solving efficiency and problem complexity advantages for solving RCPSP construction project problems. Therefore, CP is suitable for establishing a comprehensive RCPSP model.

Few studies have considered providing data for RCPSP through BIM-based information modeling. The integration models proposed by the studies that integrate schedule and resource information to BIM have a variety of limitations as follows: (1) few of them consider cost simultaneously for scheduling; (2) an automatic data integration method is missing;

and (3) most of them concentrate on unique engineering domains or unique construction processes for scheduling.

It is proven that knowledge-based methods can contribute to the efficiency of scheduling data acquisition, but the quality and scope of existing knowledge bases are hard to meet the requirement of scheduling in practical projects.

3. Methodology

The methodology of this study includes four main parts: information modeling, integration process designs, RCPSP modeling, and validation, as shown in Fig. 1. In information modeling, an integrated information model and the data source structure are designed based on the literature study of the RCPSP information requirement using an object-oriented modeling method. The following integration process design aims to produce a data integrator for the integrated information model by designing the data handling process with necessary algorithms using integration algorithm analysis methods. In the next part, the RCPSP modeling, an RCPSP mathematical model adapted to the integrated information model is designed using mathematical modeling and constraint programming. In the last part, the feasibility of the proposed data structures, processes, algorithms, and programs is validated by case study, while the efficiency when the approaches are applied in the practical construction projects is estimated using quantitative analysis, and the contributions are finally verified.

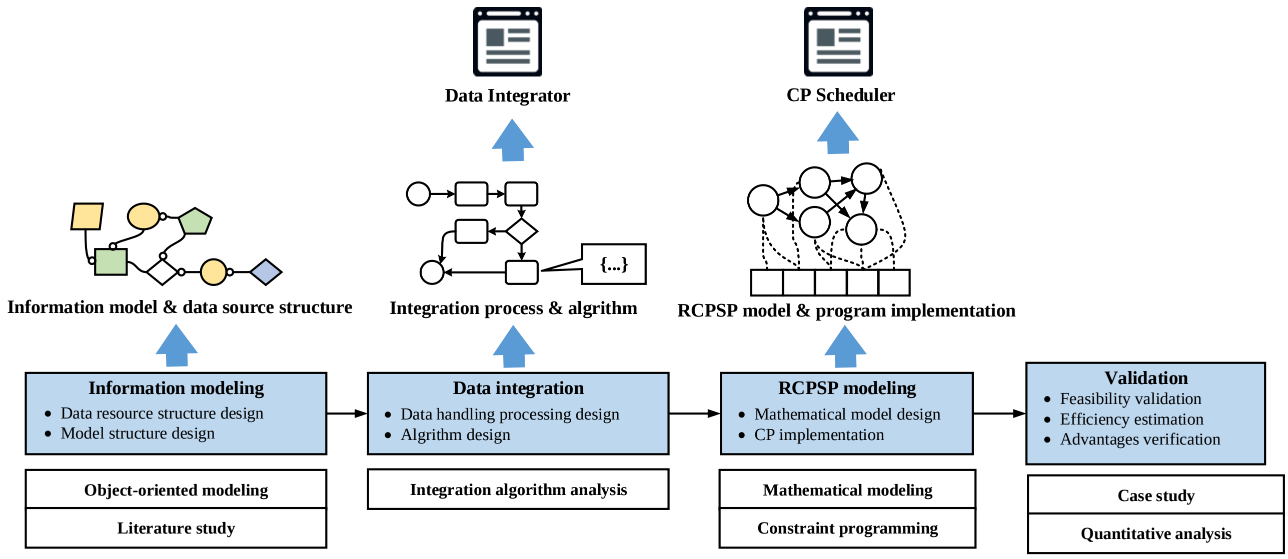


Fig. 1. Research methodology

The primary application process of this study consists of the three steps shown in Fig. 2. In the first step, data preparation, new work package templates are generated based on the designed structure. The required activity definition and resource empirical data form the body of the content, and the productivity relationships are converted into duration functions. The second step, data integration, uses both the work package templates from the first step and the existing work package template database. In this step, the work package templates, precedence rules, and BIM are integrated to generate immediate data for scheduling. The data are analyzed in the third step, scheduling, along with the scheduling settings required for producing scheduling results. The entire process is tested in feasibility validation, and the processing times are used in efficiency estimation.

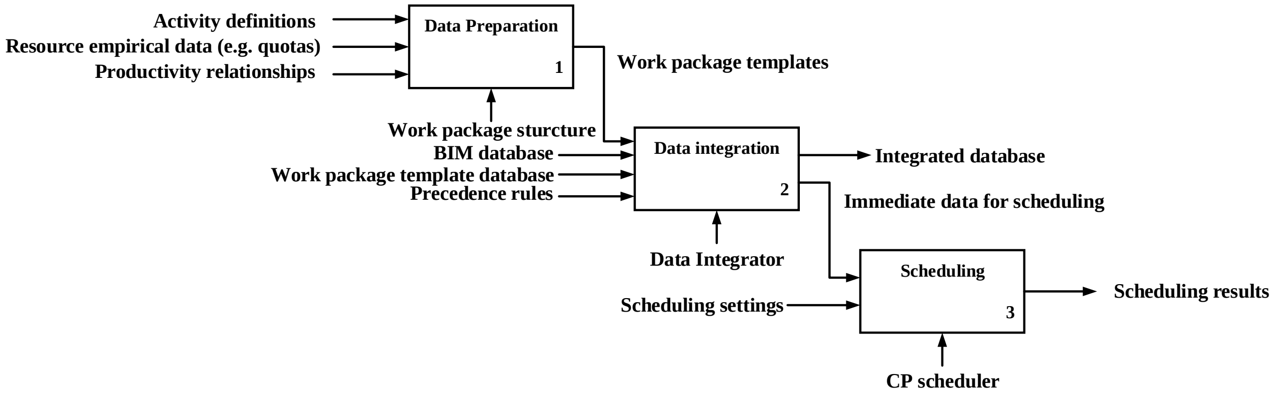


Fig. 2. IDEF0 diagram for the primary research application process

4. Work package-based information modeling for RCPSP

4.1. Work package definition

In this study, the work packages used in the schedules of practical projects are called work package instances. This term is distinct from the work package templates in the database. The main difference between the two concepts is that work package templates contain engineering experience, while work package instances contain construction project information. After being extracted from the database and associated with BIM, work package templates can be split and regrouped to generate multiple work package instances that can be regarded as activities in the scheduling problem model. The associated resource information serves as the data for resource-related constraints and objective functions.

In an RCPSP model, an activity contains schedule, quantity, resource, and cost information. Some include existing data sources, such as quantities of building elements associated with the activities, resource prices from price databases, and time-varying resource capacities from resource supply lists. Other information, including basic activity information and related resources, can be stored in work package templates. The template database enables information to be reused in similar projects. In practice, it is possible to use a work package template database during construction planning to quickly create work package instances that can store actual schedule and resource information and to update the database after the activities completed.

The content and structure of a work package are generally determined by actual needs because no standard reference exists. We can obtain the information required for a work package instance by analyzing the activity in the scheduling problem model. An activity may have multiple resource requests for selection; therefore, a work package instance should also contain multiple resource groups with resource requests concerning each activity. These are the basis for resource constraints and objective functions. The quota and quantity can be used to calculate the resource requests. Here, a quota refers to a list of resources required by a type of activities, and each item in the list provides the required amount of one resource per unit quantity according to the activity type. In China, quotas have national standards[51], and similar data are usually used for project budgets in construction enterprises.

Based on these considerations, we design the structure of the work package template, as shown in Fig. 3. Each work package template includes basic information, element classification, procedure information, and related quotas. Among these, the basic information and procedure information consist of text (e.g., notes, operational steps, and related documents). Element classifications are associated with building elements. A work package template contains one or more quotas. Each quota has a basic unit. For example, 10 m³ means that the rate in each quota item refers to the resource request for every 10 m³ of construction volume. Another key quota attribute is the usage condition which can support filtering the building elements associated with the work package template. For example, a work package template for concrete column construction needs to use different resource groups in the quotas according to the height of the column because the construction processes may differ for columns of different heights.

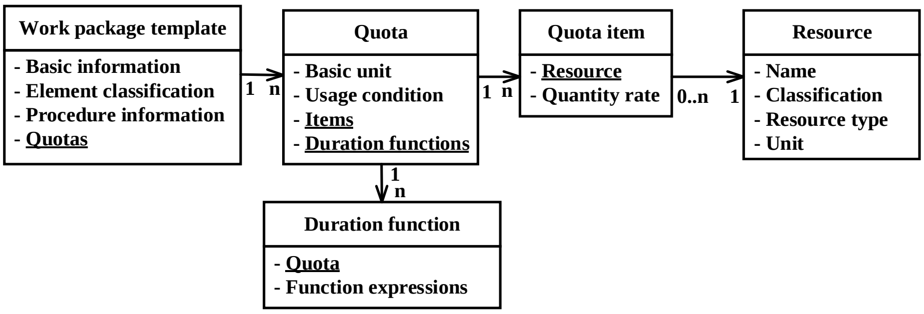


Fig. 3. Class diagram for the work package template database

The duration function is also a concept closely related to the scheduling model. The duration of an activity has a significant relationship with labor when other conditions are constant. We use the duration function for this relationship. This function may contain multiple equations that express the relationship between the all laborers in a quota and the activity duration. The simplest function is linear, such as a job that one person can complete in ten days, two people can complete in five days, or five people can complete in two days. However, the quota-duration relationship is actually nonlinear, and it is possible to have multiple human resources. Incorporating these relationships into the simulation model, assuming that the scheduler supports such relationships, should result in a more realistic scheduling result.

4.2. Work package based integrated information model

Based on work package templates and BIM, an integrated information model is designed, as shown in Fig. 4. The model consists of five basic entities, namely, the building element, the work package instance, the quota, the quota item, and the resources. Among these entities, the building element comes from BIM, while the other four entities come from the work package database. It is worth noting that in Fig. 4, we list only the entities and attributes that are used in this study. To ensure that the information is available, we need to identify the source and usage of all the attributes.

Each building element provides the basic quantities, element type, and main material attributes. The basic quantities include the volume, area, length, and weight of a building element. Among them, geometric quantities such as volume, area and length can be obtained directly from the geometric information, while others may not exist. For example, the weights of the steel bars in reinforced concrete may be added during modeling or estimated by geometric quantities. The element types and main materials are used to search for related work package templates. Element types can be represented using classification codes to solve problems in which custom types are not defined or are inconsistent with other element types in the work package templates. The main material is not usually defined in BIM and needs to be added manually. Similarly, to ensure consistency with work package templates, the main material is represented by unified resource classification and coding systems.

Each work package instance is associated with multiple building elements, and each building element may be associated with multiple work package instances, which usually use different quantities and units from those of the building element. A work package instance includes quotas, and each quota includes several quota items that each correspond to a resource. Work package instances are generated from templates in the work package template database, and resources are also extracted directly from the database. The related attributes of entities mostly come from the work package template database, while other properties with multiple candidates, such as the resource capacity and resource price, may come from other data sources.

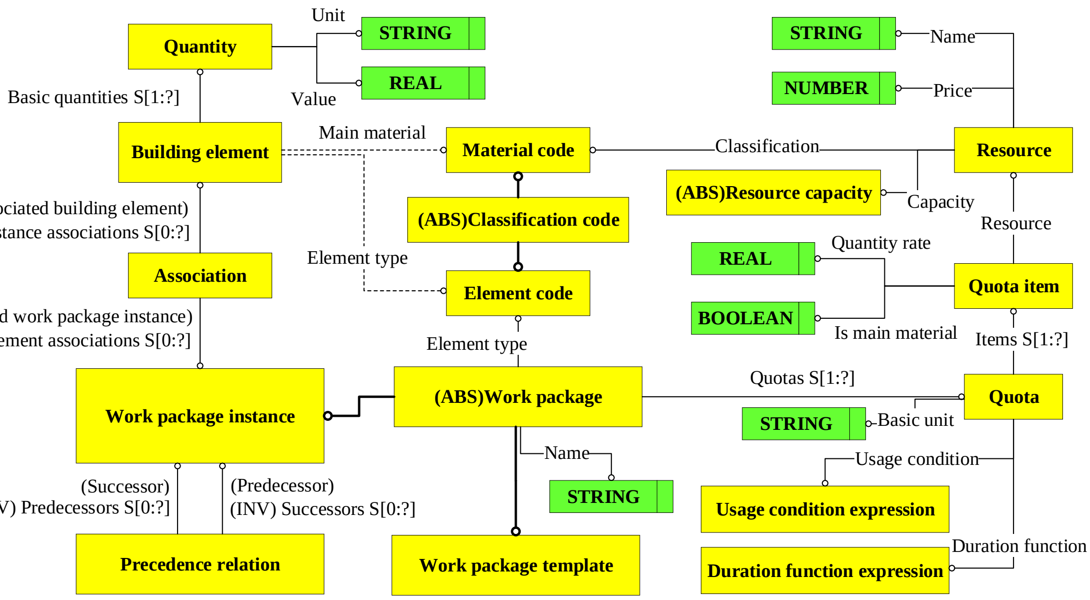


Fig. 4. EXPRESS-G diagram for the integrated information model

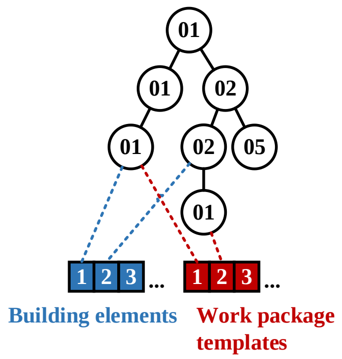
5. Data integration

5.1. Linkage between BIM and the work package

The classification and coding of information involve systematic regulation of the scope, structure, and external interface. Complete classification is conducive to associations among different data subjects and can ensure the completeness and unity of information. This study introduces two types of classification codes: element codes and material codes. Both codes can use existing classification and coding standards, such as OmniClass[52] and Uniclass[53]. When classification codes are subject to the same rules in both BIM, and the work package templates, the accuracy of the association between building elements and work package templates can be guaranteed. Table 21 of OmniClass or table Ef of Uniclass can be used as the element code. Because construction and operational information are not involved, this code can be added to the building elements in the design phase. One work package template corresponds to only one category of building elements; thus, it has only one element code. However, this correspondence is not a strict equation. The element code in the work package template represents all the categories of the classification node and its subnodes. Therefore, when the element code of a building element contains the code of a work package template, the two codes are associated. For process reinforcement works or formworks, which may include multiple element categories, different work package templates must be defined for every element category because the quotas always differ among different element categories. In this way, this type of quota filtering is complete after data association, which reduces the processing difficulty.

Table 23 of Omniclass and table Pr of Uniclass are candidates for the material code, which is not only used to associate BIM with the work package templates but is also used during work package instantiation. The quotas of a work package template include the material and its classification code. The materials that can influence the work result are the main materials. Their classification codes are used to generate associations with BIM by comparing them with the material codes of building elements.

A quick matching process, called the "first matching" of the work package templates and BIM, is conducted by building an element code tree. The first matching process consists of two steps: (1) building the element code tree and (2) traversing the building elements and completing the matching process using a tree search, as shown in Fig. 5. The process algorithm complexity depends on the depth of the code tree D, the number of work package templates Nw, the number of building elements Ne, and the number of possible associations Nc. The time complexity of the building step is $O(Nw \cdot D)$, and the time complexity of the matching process is $O(Ne \cdot D+Nc)$. Because Nw and D have much smaller orders of magnitude, the time complexity of the entire algorithm depends primarily on *Ne* and *Nc*. Because these values rarely reach into the trillions, the process can be completed within seconds.



a. An example of the code tree

```
Building process:
foreach t in work package templates {
  c = the code of t
  curnode = the root of the tree
  for i = 1 to top level of c {
    n = the number of level i in c
    if !curnode.children.contains(n)
      curnode.children.add(n)
    curnode = node n in curnode.children
  }
  attach t to curnode
}
```

```
Matching process:
foreach b in building elements {
  c = the code of b
  curnode = the root of the tree
  for i = 1 to top level of c {
    n = the number of level i in c
    if !curnode.children.contains(n) break
    curnode = node n in curnode.children
    create connection between b and every work
      package template attached to curnode
  }
}
```

b. Pseudo code for the first matching process

Fig. 5. Code tree and pseudo code of the first matching process

By traversing the first matching result, the second matching process can be completed by eliminating the associations that do not satisfy the matching principle of the material code. Multiple material codes may exist in both building elements and work package templates. One simple matching principle is to ensure that all the main material codes of the work package template include the material coding of the building element.

After these two matching processes, we generate the associations between the work package templates and building elements, allowing a building element to use the related work package templates for construction. Usually, the association is many-to-many. Various building elements may use the same work package templates, and a building element may have multiple work package templates.

5.2. Work package instantiation and regrouping

Work packages are instantiated by dividing the building elements into groups by construction areas and generating instances for each work package template of each group, as shown in Fig. 6.

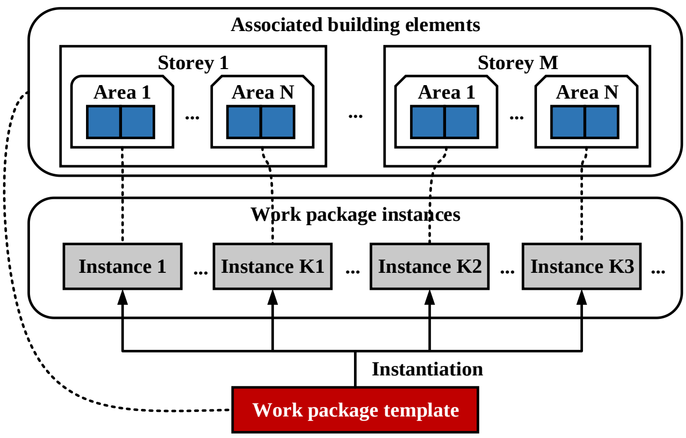


Fig. 6. Work package instantiation

Regrouping the work package instances ensures that the building elements associated with each instance use the same quotas. This work can be accomplished by traversing all the building elements and judging whether their properties meet the usage conditions of the quotas in their associated work package instances. A work package instance containing n quotas has 2^n-1 possible quota combinations. When building elements use a new quota combination, a new work package instance is generated. Fig. 7 shows an example when n = 4. After regrouping, the available quotas and corresponding quantities in all the work package instances have been determined and can be directly used for scheduling.

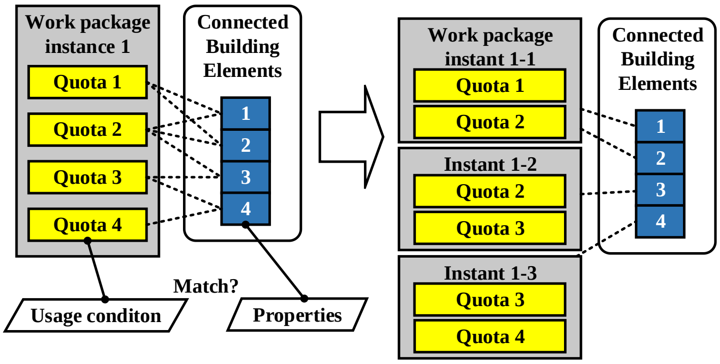
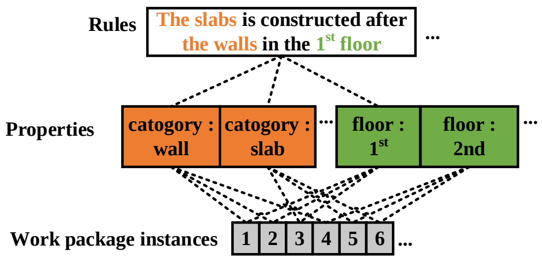


Fig. 7. An example of work package instances regrouping by quotas

5.3. Construction sequence generation

The construction sequence is generated by defining the precedence relations between work package instances. The information related to the work package instances makes automating this process possible. These relations can be divided into technical and nontechnical precedence relations[54]. This study focuses on the technical precedence relations, that is, the sequential logic that must be followed during construction. Such precedence relations are also a main consideration during the construction planning phase.

The generation of technical precedence relations is based primarily on rules. For example, one category of building elements must be constructed after the completion of another category of building elements, and upper-layer building elements must be constructed after the completion of the lower layer. The rule definitions are based on properties such as the spatial location, the element category, and the construction area. After defining the rules, by searching the properties of the related work package instances, the precedence relations among the work package instances can be generated according to the rules, as shown in Fig. 8. A rule represents two sets of work package instances and specifies their precedence relations or more specific settings, such as time lags and relationship types. Both sets are defined by one or more properties. A property represents a set of work package instances, and the intersection of these sets may be represented by multiple properties. Also, the union of these sets can be represented by defining multiple rules.



a. An example of relations among rules, properties and work package instances

```
generate a property set named φ
foreach wp in work package instances
  kv_w = all the key value pairs of wp
  foreach t in kv_w
    if t is not in φ, add t to φ
    generate the linkage between t and wp
foreach r in all the rules
  if not every key value pair in r is in φ, continue to next rule
  generate the first work package instance set s1 by key value pairs in r
  generate the second work package instance set s2 by key value pairs in r
  generate the precedence relations between s1 and s2 and set the
    relation type as r
```

b. Pseudo code for sequence generation

Fig. 8. Rule-based work package sequencing

6. Scheduling problem model and generation

6.1. Mathematical model

6.1.1. Constraints

The following five types of constraints are considered. Among them, constraints (1), (2), and (3) are commonly used in RCPSP for CP, while (4) and (5) define a new MRCPSP. In basic MRCPSP, one mode corresponds to the duration and cost or the duration and demands of one renewable resource. The duration and cost of each mode in the problem model in this study are calculated by r_{ik} and q_i, which means that the duration and cost are related to the quotas and the quantities, where, r_{ik} is the quantity ratio of resource k in activity i, which represents the requests of resource k per unit basic quantity, and q_i represents the basic quantity of the product of activity i, such as the volume, area, and weight.

(1) Precedence relations

For activity i, T_i to represent the start time SS_i or finish time SF_i, and the precedence relation between activities can be defined as:

$$minLag \le T_j - T_i \le maxLag \tag{1}$$

where T_i is the SS_i or SF_i of the predecessor activity i, T_j is the SS_j or SF_j of the successor activity j, $minLag$ is the minimum interval, and $maxLag$ is the maximum interval. In construction projects, $maxLag$ is seldom considered, and $minLag$ is zero in most cases.

(2) Milestones

In schedule management, setting a deadline is a general and effective method for controlling the schedule of critical activities. A milestone Mi is set as the deadline:

$$SF_i \le M_i . \tag{2}$$

(3) Resource capacities

At time t, the total demand RD_{kt} of resource k should be less than the total supply RS_{kt}, which is

$$RD_{kt} \le RS_{kt} . \tag{3}$$

The total demand calculation depends on whether the resources are renewable. For renewable resources, the total demand is

$$RD_{kt} = \max\left(\sum_{i \in DA_u} d_{ik}\right), \qquad u = 0,1,\ldots,t , \tag{4}$$

where d_{ik} is the amount of resource k required by activity i per day, which does not change when scheduling. $DA_t = \{i \mid SS_i \le t \le SF_i\}$, which is the set of ongoing activities at time t.

For nonrenewable resources,

$$RD_{kt} = \sum_{i \in ASA_t} d_{ik} , \tag{5}$$

where $ASA_t = \{i \mid t \ge SS_i\}$, which is the set of activities started before t.

(4) Activity internal relations

The amount of a renewable resource k required by activity i is qr_{ik}, which can be calculated by the basic quantity q_i and quantity ratio r_{ik}:

$$qr_{ik} = q_i r_{ik} , \tag{6}$$

where the unit of qr_{ik} is a man-day, that is, it requires one worker qr_{ik} days or qr_{ik} workers one day to complete activity i. The activity duration is related to d_{ik} and qr_{ik}. We can consider that the process duration SD_i is simply inversely proportional to d_{ik}:

$$qr_{ik} = d_{ik} SD_i . \tag{7}$$

Formula (7) is a typical duration function.

(5) Resource modes

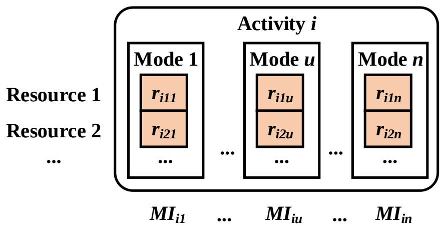


Fig. 9. Resource modes and selection indices in an activity

Different mode selections affect the cost and duration and cause different scheduling results. As shown in Fig. 9, we introduce MI_{iu} to indicate which mode u of activity i is selected. This variable can be equal to zero or one and satisfies the following formula:

$$\sum_{u} MI_{iu} = 1. \tag{8}$$

MI_{iu} is used to calculate r_{ik} as follows:

$$r_{ik} = \sum_{u} MI_{iu} r_{iku} . \tag{9}$$

6.1.2. Objectives

This RCPSP model considers the total duration and total cost as objectives.

(1) Total duration

The total duration TD is calculated as follows:

$$TD = \max\left(SF_i\right) - SS , \tag{10}$$

where SS is the project start time.

(2) Total cost

The total cost consists of the direct cost DC and the indirect cost IC:

$$DC = \sum_{i}\sum_{k} q r_{ik} p_k , \tag{11}$$

where p_k is the price of resource k.

The indirect costs include loan interest, site fees, design fees, and supervision fees. Because such costs are complicated and have little correlation with the activity sequence, only indirect costs related to the duration are generally considered when scheduling, and they are reasonably considered to be linearly related to the total duration:

$$IC = TD \cdot dc , \tag{12}$$

where dc is the direct cost per day.

6.2. Model implementation and data conversion

A scheduling problem model can be solved using a variety of algorithms. This study does not focus on how to solve the problem; therefore, it is appropriate to solve the problem using a CP solver. Here, *IBM ILOG CPLEX Optimization Studio* is used to solve construction scheduling problems. *Optimization Studio* provides the .NET API, and a CP model is designed for the mathematical model described above based on that API. The CP model is object-oriented to correspond to the integrated information model.

To store the key data for solving the problem, we designed three classes: *Activity*, *SelectionValue*, and *Resource*. They contain not only the data required for the solution but also the variables in the CP model, as shown in Fig. 10. Because including too many variables will increase the solution space and thus the complexity, we should consider the variables as expressions when converting them into the CP model from the mathematical model. *Activity* has a key variable of the type *IIntervalVar*, which is a class in the CP Optimizer .NET API. Using interval variables to represent the time required by activities is a typical approach in modeling CP scheduling problems[29]. An interval variable can represent the duration, start time, and finish time of an activity. The *IIntExpr*, an integer expression class in the CP Optimizer .NET API, can represent all three of these variables. Another key *Activity* variable is mode selection. *SelectionValue* is designed for this. Within a *SelectionValue*, a 2D integer array *source* is used to store activity quotas, while an integer array, *indices*, stores the resource IDs corresponding to the *source*. In addition, an *IIntVar* array *flag* whose length is equal to the number of quotas indicates the quota selection. By adding a constraint, the variables in the *flags* can satisfy Formula (8), and *flags* and *source* can be used to calculate r_{ik} in Formula (9). From r_{ik}, the amounts of all the resources required by an activity, the *Activity.ResourceDemands*, can be calculated. Furthermore, the *Activity.Cost* can be calculated according to the resource demands and resource prices. The key variable of *Resource* is the *Demands*, which belongs to the *ICumulFunctionExpr* class and indicates the demand distribution of a *Resource* over the entire time interval. The *Resource.Demands* are the cumulative demands of all activities for a *Resource*. When the resource is renewable, the *Pulse*() function is used for accumulation; otherwise, the *StepAtStart*() function is used. Subsequently, the *HeightAtStart*() function builds the relationships among *Activity.ResourceDemands*, *Resource. Demands* and *Activity.TimeInterval*. The code for this implementation is shown in Fig. 11.

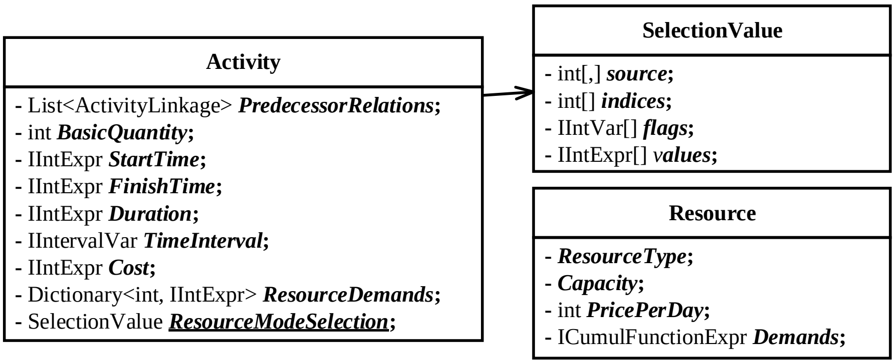


Fig. 10. Key classes of the CP model

```
foreach(var resource in D.Resources) {
    resource.Demands = cp.CumulFunctionExpr();
    foreach(var activity in D.Activities) {
        if(!activity.ResourceDemands.ContainsKey(resource.Id)) continue;
        ICumulFunctionExpr p = null;
        if(resource.Type == ResourceType.Labor || resource.Type == ResourceType.Machine) {
            p = cp.Pulse(activity.TimeInterval, 0, G.MaxResourceCurrentDemands[resource.Id]);
        } else {
            p = cp.StepAtStart(activity.TimeInterval, 0, G.MaxResourceCurrentDemands[resource.Id]);
        }
        resource.Demands.Add(p);
    }
    foreach(var activity in D.Activities) {
        if(!activity.ResourceDemands.ContainsKey(resource.Id)) continue;
        if(resource.Type == ResourceType.Labor || resource.Type == ResourceType.Machine)
            cp.AddGe(cp.HeightAtStart(activity.TimeInterval, resource.Demands, 0),
                cp.Max(activity.ResourceDemands[resource.Id], 1));
        else
            cp.AddGe(cp.HeightAtStart(acti
                activity.ResourceDemands[resource.Id]);
    }
}
```

Using *Pulse()* if resource is renewable, else using *StepAtStart()*

Renewable resources have a minimum demand of 1

Fig. 11. Cumulative function generation for renewable and nonrenewable resources

The formulas, including five constraints and two objective functions in the mathematical model, can be expressed directly by CP. Among the five constraints, the implementation of the precedence relations and milestones follows an existing study[29]. Regarding the resource capacities, the constant capacity, as assumed in traditional RCPSP, is defined by using *Le*() as the inequality constraint, and the time-varying capacity is defined by using *AlwaysIn*() as the cumulative constraint. The internal activity relation is considered when calculating the resource demands of an activity. When a resource is renewable, the resource demand should be divided by the duration according to Formula (7). Resource modes are generated when initializing the *SelectionValue*, as shown in Fig. 12.

```
IIntExpr[] fexprs = new IIntExpr[source.GetLength(0)];
for(int i = 0; i < source.GetLength(0); i++) {
    fexprs[i] = _flags[i];
}
cp.AddEq(cp.Sum(fexprs), 1);
```

Formula (8)

```
for(int i = 0; i < source.GetLength(1); i++) {
    IIntExpr[] vexprs = new IIntExpr[source.GetLength(0)];
    for(int j = 0; j < source.GetLength(0); j++) {
        vexprs[j] = cp.Prod(flags[j], source[j,i]);
    }
    values[i] = cp.Sum(vexprs);
}
```

Formula (9)

Fig. 12. Resource mode constraint implementation

The input of the CP model comes from the integrated information model. Because both models are built in an object-oriented way and there is a correspondence between the objects, some data, such as the price and type of resources, can be transferred directly. Nevertheless, data that need to be processed also exist. The data correspondence between the two models in this study is shown in Table 3. Two significant differences in content or form can be found. The first is that the quantity of a work package instance comes from the quantities of the associated building elements. In this process, the basic unit of the quota of the work package instance is filtered to obtain the basic quantity. The basic quantity of a work package instance is the sum of the basic quantities of all the associated building elements. The second is the conversion of quotas in the work package instance into the resource mode selection of the activity. The implementation method is described above.

Table 3. Key data correspondence between the two models

Source: integrated information model		Target: RCPSP model (CP)		Data processing methods
Entities	Properties (Entities)	Classes	Properties	
Work package instance	Predecessor relations	Activity	Predecessor relations	-
Building element	Basic quantities		Basic quantity	Filtered by the basic unit of the quota, and summed
Work package instance	Quotas		Resource mode selection	Fig. 12
Resource	Resource type/ supply type	Resource	Resource type/ supply type	-
	Capacity		Capacity	-
	Price		Price	-

7. Validation

7.1. Application development

The validation is mainly based on a client-server software called 4D-BIM that developed by our research group. The 4D-BIM software has been used as the basic developing platform for construction information integration and further management purpose[7]. It can export and import quite a few different file formats, including MS Project files, Excel files, IFC files, 3dxml, obj, dxf, and supports model management, 3D/4D visualization, schedule management, site management, cost management, and collision detection. Previously, it was also called 4D-GCPSU 2009[55].

Fig. 13 shows the structure of the application developed in this paper. A data integrator, a CP scheduler, and a work package template manager were developed as extension parts of the 4D-BIM software for this study. Moreover, the 4D-BIM database was also expanded to store the proposed integrated information model. The data integrator can integrate the work package templates into the 4D-BIM database, and can also produce the intermediate data, as shown in Fig. 18, for scheduling after data integration. Based on the data, the CP scheduler can export detailed scheduling results in Excel format for further analysis. Fig. 19, Fig. 20, Fig. 21, and Fig. 22 are all generated based on the raw scheduling results data.

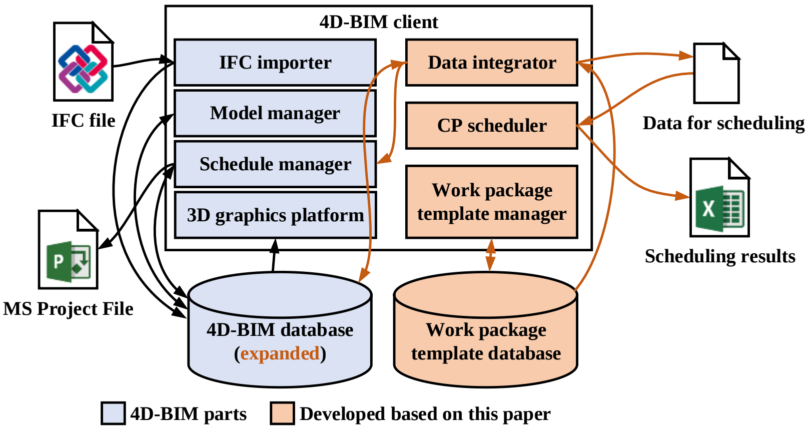


Fig. 13. Application structure for validation

7.2. Data source

The project is an apartment located in Hefei, China, and consists of 22 aboveground floors and one underground floor. The building has a reinforced-concrete structure. Along with all reinforced-concrete components, some of the external and internal walls are prefabricated. The construction processes and resource requests for these prefabricated components are different from those of cast-in-place components. This study uses the model built for the project using Revit. The model is exported as IFC format and imported into the 4D-BIM based application. The result is shown in Fig. 13. Before importing the model, we asked the modelers to add the necessary attributes to the building elements, such as the element codes, material codes, and wall thicknesses. This augmented information and the properties of the Revit model, such as the elevation will be used in the modeling and data integration. Among them, the element code and the material code use table Ef and table Pr of Uniclass, respectively.

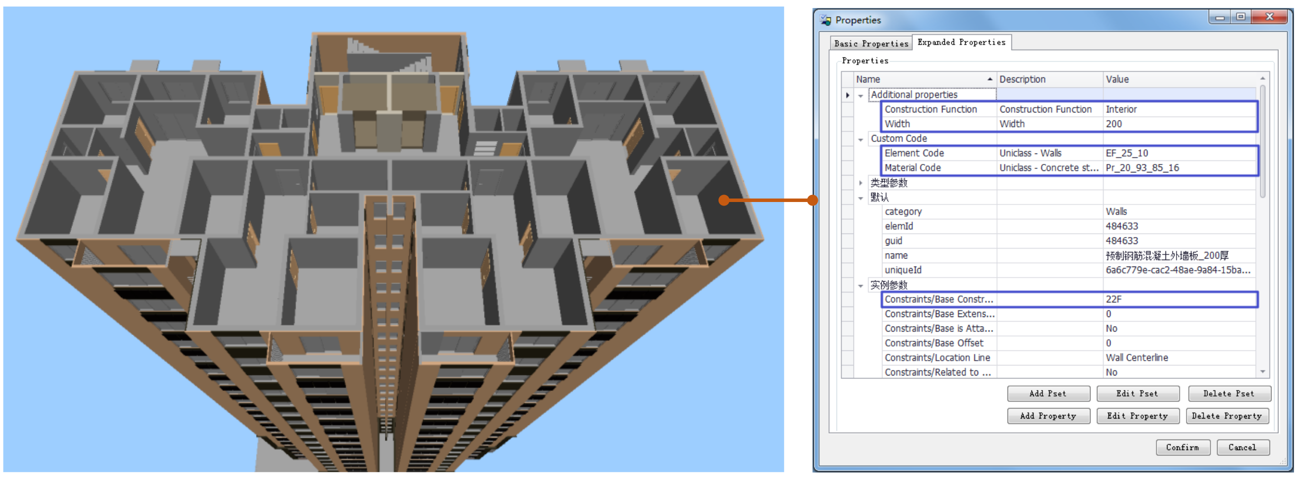


Fig. 14. 3D view of the BIM and the expanded properties

For clearly analyzing the results, this validation mainly focused on wall and slab construction because these are the main components in structural engineering. The related work package templates are created, as shown in Fig. 15. Because of the differences in construction materials, we divided the cast-in-place concrete construction into formworks, reinforcement works, and concrete works. We also created the work package templates for the prefabricated walls and slabs. There are eight work package templates for validation.

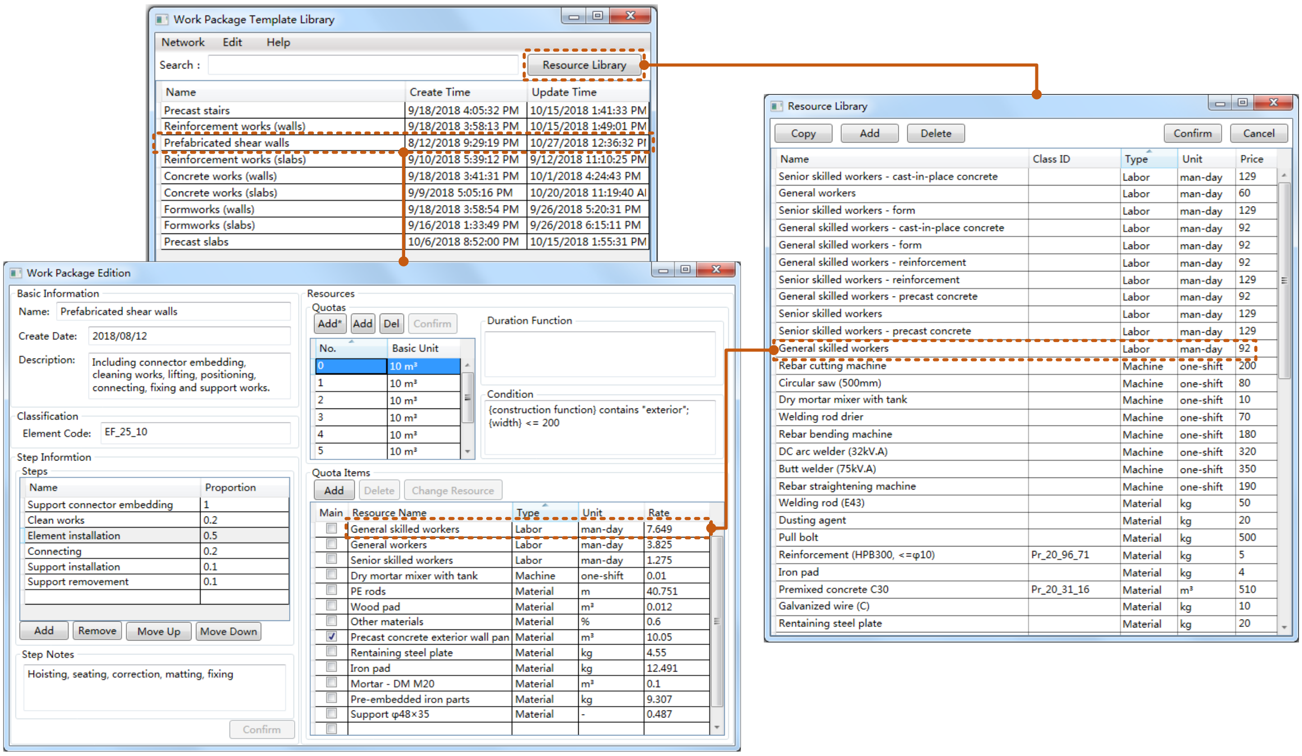


Fig. 15. Work package template database

Each work package template consists of four parts: basic information, classification, step information, and resources. The classification and resources parts contribute to the data integration and scheduling processes. The classification part includes only an element code for the first matching process, and the resources part contains several quotas, wherein each quota has several items that correspond to the resources in the resource database. The data for the quotas and resources were sourced from a Chinese national quota standard (TY 01-01(01)-2016).

7.3. Information integration

After the two-step matching process, 8,506 associations were generated between the work package templates and the building elements, as shown in Fig. 16. The prefabricated and cast-in-place structural walls were associated with different work package templates, and as expected, the three work package templates for cast-in-place concrete wall construction were associated with the same building elements. A similar situation applied to the three work package templates for cast-in-place concrete slabs. Finally, no building element was associated with the work package template for precast slabs because all the slabs are cast-in-place.

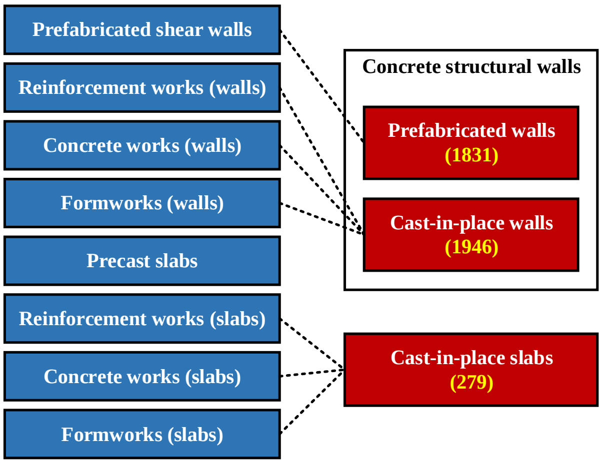


Fig. 16. Associations between the building elements and work package templates

After associating the building elements with the work package templates, the next step is to instantiate the work packages using construction areas. For this validation, we regarded each floor as a construction area; therefore, each work package template would generate an instance for each floor associated with the building elements. We obtained 167 work package instances after this step was completed. Afterward, during the regrouping process, the work package instances for the prefabricated shear walls of each floor were decomposed into two instances: one contained the quota, with ID 0, and the other contained the quotas with ID 2 and ID 4. Finally, 189 work package instances were generated.

Six types of relation rules were introduced to generate the sequence information in the integrated information model, as shown in Table 4. Among these, the first and second rules defined the spatial precedence relations, and the others defined the precedence relations in the process patterns. These rules were all limited by the construction area (in this validation, the story); thus, they were defined for every story. The first type defined 22 rules, and each of the remaining five types defined 23 rules. The total number of generated rules was 137. The precedence relations between the work package instances generated by the rules are shown in Fig. 17. We found that redundant relations existed, which may cause conflict. A manual check showed that no conflicts existed and that the integration model was available for scheduling.

Table 4. Global rules for sequence generation

ID	Relation rules		Explanations
	Predecessors	Successors	
1	Type: walls & Floor: (n)f	Type: walls & Floor: (n+1)f	The construction of each floor may start only after the construction of the walls on the previous floor.
2	Type: slabs & Floor: (n)f	Type: walls & Floor: (n)f	The construction of walls on each floor may start only after the construction of the floor.

3/4	Type: walls/slabs & Resource: formwork & Floor: (n)f	Type: walls/slabs & Resource: reinforcement & Floor: (n)f	Reinforcement works of the walls/floors on each floor must be started after the formworks are completed.
5/6	Type: walls/slabs & Resource: reinforcement & Floor: (n)f	Type: walls/slabs & Resource: concrete & Floor: (n)f	Concrete works of the walls/floors on each floor must be started after the reinforcement works are complete.

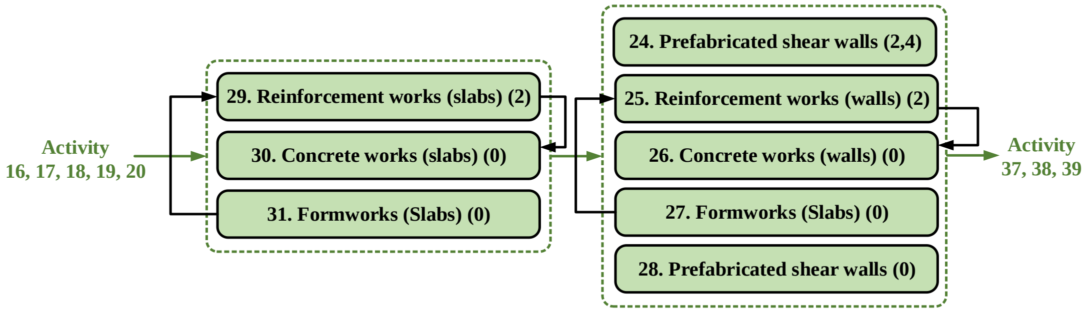
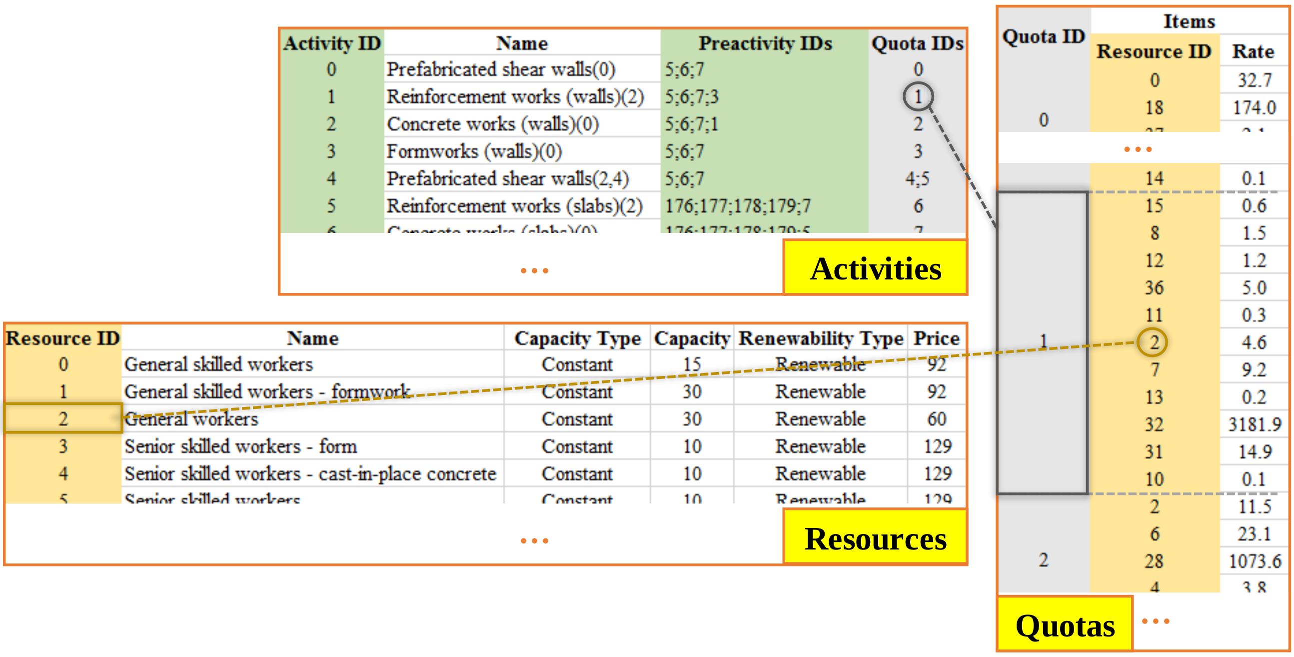


Fig. 17. Network graph of the activities of 7f (regenerated from Microsoft Project)

The data integration is complete when the precedence relations are generated. For transferring the data to the scheduling problem model, an intermediate data format was proposed. The format has three core concepts, namely, activities, resources, and quotas, and the format stores all the information extracted from the integrated information model that is required by the scheduling problem model. In the intermediate data format, the data included 189 activities, 40 resources, and quotas, some of which are shown in Fig. 18.



Activity ID	Name	Preactivity IDs	Quota IDs
0	Prefabricated shear walls(0)	5;6;7	0
1	Reinforcement works (walls)(2)	5;6;7;3	1
2	Concrete works (walls)(0)	5;6;7;1	2
3	Formworks (walls)(0)	5;6;7	3
4	Prefabricated shear walls(2,4)	5;6;7	4;5
5	Reinforcement works (slabs)(2)	176;177;178;179;7	6

Resource ID	Name	Capacity Type	Capacity	Renewability Type	Price
0	General skilled workers	Constant	15	Renewable	92
1	General skilled workers - formwork	Constant	30	Renewable	92
2	General workers	Constant	30	Renewable	60
3	Senior skilled workers - form	Constant	10	Renewable	129
4	Senior skilled workers - cast-in-place concrete	Constant	10	Renewable	129

Quota ID	Items: Resource ID	Items: Rate
0	0	32.7
	18	174.0
	…	
	14	0.1
1	15	0.6
	8	1.5
	12	1.2
	36	5.0
	11	0.3
	2	4.6
	7	9.2
	13	0.2
	32	3181.9
	31	14.9
	10	0.1
2	2	11.5
	6	23.1
	28	1073.6

Fig. 18. Part of the intermediate data

7.4. Scheduling results

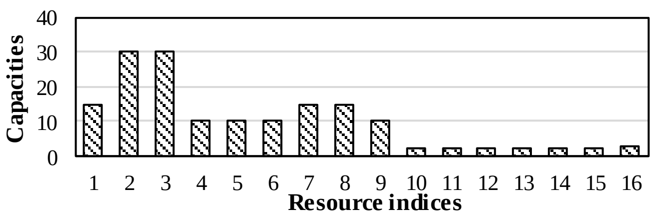
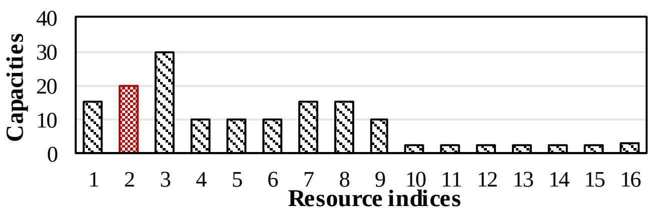
The four problem cases listed in Table 5 were solved for this validation. Study 1 served as a reference; Study 2 was used to verify the influence of the changing capacities of renewable resources; Study 3 was used to verify the influence of the changing supplies of nonrenewable resources, and Study 4 was used to verify the influence of changing the objective.

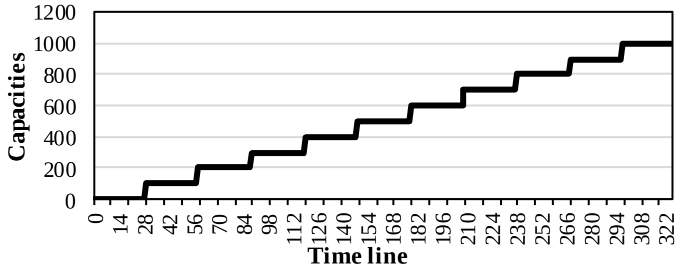
Fig. 19 shows the total costs and total durations of the scheduling results of the four studies, which are consistent with the theoretical expectations. Study 2 and Study 3 have more strict resource constraints than does Study 1. Moreover, the total cost in Study 4 is less than that in Study 1 because the total cost is the objective of Study 4. Similarly, the total duration of Study 1 was less than that of Study 4.

Considering that the activities and their precedence relations in each story were almost identical, we chose the scheduling results of the activities on the seventh floor for comparison purposes. Each activity in the scheduling results had a start time, a duration, and a selected quota. Because the start time was influenced by other activities, we compared only the durations and selected quotas.

The comparison results of the activity durations are shown in Fig. 20. Compared with Study 1, Study 2 changes the capacity of Resource 2, which is the general skilled formwork worker and is used by activities 27 and 31. The result shows that the durations of activity 27 and activity 31 increase after the capacity decreases from 30 to 20, while the durations of other activities are not changed. Also, compared with Study 1, Study 3 has a new time-varying capacity of Resource 38, which is the precast concrete exterior wall panel required by activity 24. Therefore, the duration of activity 24 changes. In Study 4, the durations of multiple activities change due to the scheduling objective change.

Table 5. Scheduling configurations of the studies

Study ID	Resource constraints	Objectives	
		Min (*TD*)	Min (*TC*)
1		√	
2		√	
3	Compared to study 1, add a time-varying constraint for Resource 38:	√	



4	Same to study 1	√

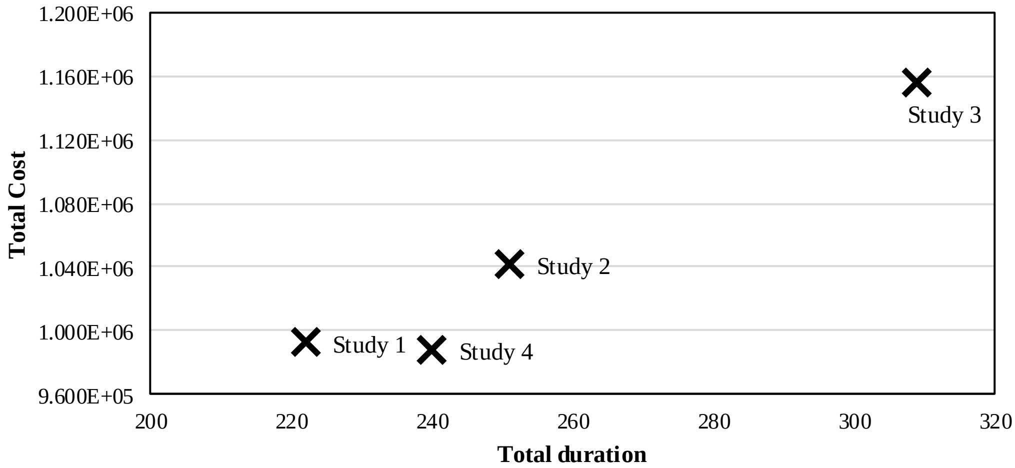


Fig. 19. Total durations and total costs of scheduling results

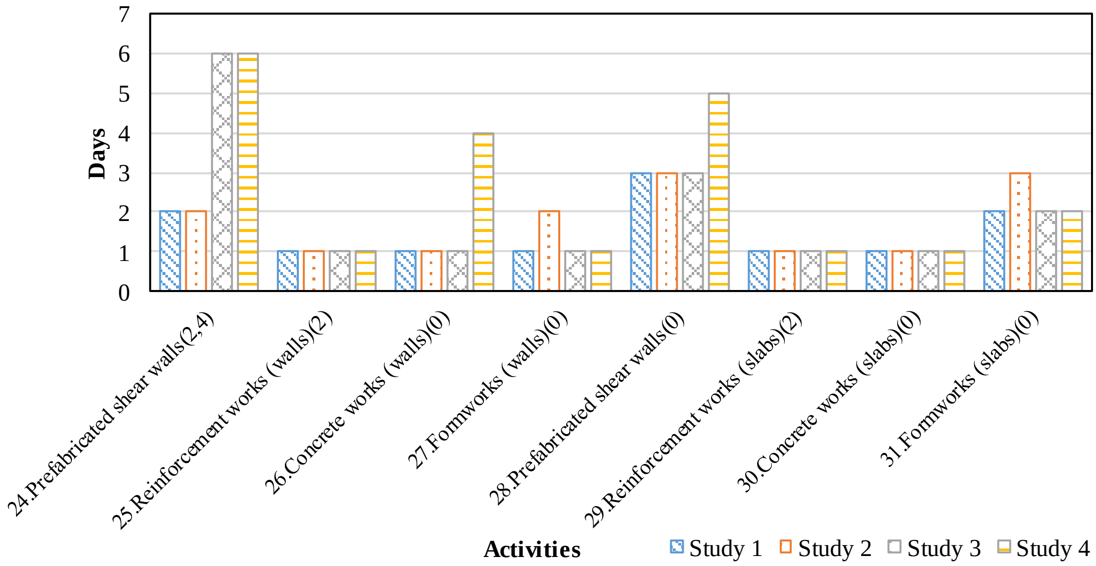


Fig. 20. Durations of activities on 7f in the scheduling results

Table 6. Quota selection of activities on 7F in the scheduling results

Activity ID	Activity names (available quota indices)	Quota IDs

		Study 1	Study 2	Study 3	Study 4
24	Prefabricated shear walls (2,4)	28	28	27	27

The comparison of the resource demands of Resource 2 in the scheduling results of Study 1 and Study 2 is shown in Fig. 21. A reduction in the capacity of a renewable resource not only reduces the daily requirement of the resource but also increases the total duration. In actual construction projects, similar results appear when the number of available laborers or machines changes.

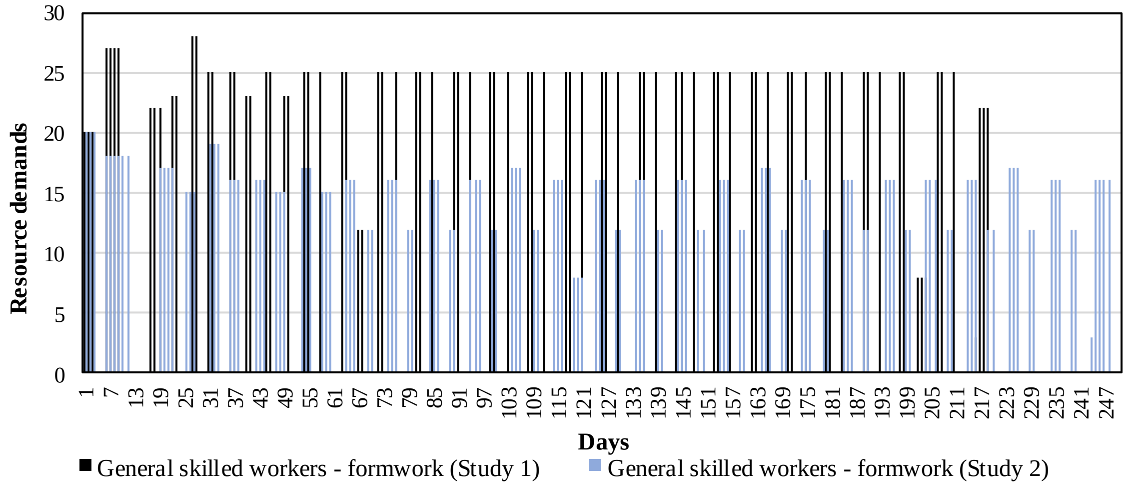


Fig. 21. Resource demands of Resource 1 in Study 1 and Study 2

A comparison of the demands of Resource 38 in the scheduling results of Study 1 and Study 3 is shown in Fig. 22. When time-varying capacity is added to Resource 38, the resource requirement is reduced, and the total duration is increased. In actual construction projects, when a specific material (such as a prefabricated component) has an arrival schedule, it has such a time-varying capacity.

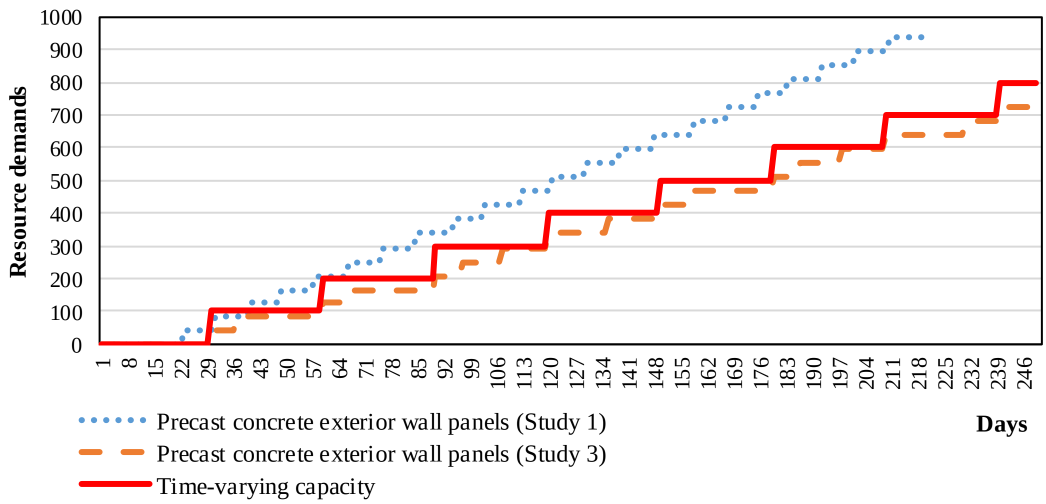


Fig. 22. Resource demands of Resource 38 in Study 1 and Study 3

7.5. Implementation efficiency estimation

Fig. 23a shows the scheduling workflow based on semiautomatic data integration in actual construction projects that includes nine tasks. Four of these tasks should be completed manually, but others can be completed automatically by a computer. The time estimations required by the workflow were made under the following assumptions:

(1) using the data in Section 7.2,

(2) conducting the workflow in ten similar projects and calculating the average time consumption in one project,

(3) estimating the time consumption when handling every data item and then summarizing,

(4) automated tasks did not consume any time, and

(5) performing five scheduling operations in one project.

Table 7 shows the time estimation results of the workflow shown in Fig. 23a. In the workflow, Task 1 creates eight work package templates. Creating each template requires up to five minutes to add the basic information; therefore, all eight templates consume 40 minutes. The next step is to add quotas. Assuming that it takes 15 seconds to add a quota item, it requires 51 minutes to add the 204 quota items. Simultaneously, assuming that it takes two minutes to add the basic unit and usage condition of each quota, this task requires 38 minutes. Consequently, Task 1 requires a total of 129 minutes.

Task 2 is a two-step loop. The first step is to filter the building elements, and the second step is to add the element and material codes. In this study, the building elements were filtered 26 times by the Revit family name. If each loop takes one minute, Task 2 requires a total of 26 minutes.

In Task 5, assuming that it takes five minutes to define each rule, defining all six rules requires a total of 30 minutes.

The time required by Task 7 is negligible and was conservatively set to 5 minutes.

In conclusion, considering that Task 1 and Task 5 need to be completed only once for ten similar projects, the average time required to complete one workflow is estimated to be (129+30)÷10+26+5*5≈67 minutes.

Table 7. Time estimation of each step of the proposed workflow

Manual task	Subtask	Time consumption (minutes)
1. Generating work package templates	1.1 Templates creation	5 * 8 templates = 40
	1.2 Quota items addition	0.25 * 204 quota items = 51
	1.3 Quota information addition	2 * 19 quotas = 38
2. Adding classification codes to building elements in BIM		1 * 26 times = 26
5. Defining sequence rules		5 * 6 rules = 30
7. Parameter input and configuring		5

Average time consumption considering assumption (2) and (5)	(40+51+38+30)/10+26+5*5≈67

The control group uses a general workflow of RCPSP modeling and scheduling, as shown in Fig. 23b. Conservatively, all the calculations are assumed to be performed automatically by the computer. The time estimation results of the workflow are shown in Table 8.

Task 1 generates a WBS with activities as leaf nodes and adds precedence relations. For this project, the WBS of one story can be created first, and then the WBS for the entire project can be generated by replication. Because this study has a simple WBS and precedence relations, it requires approximately 5 minutes.

Task 2 calculates the total quantities of building elements related to each activity. Assuming that looking for the relevant building elements of each activity and summing up their quantities takes nearly two minutes, the 189 activities require a total of 398 minutes.

Task 3 adds the modes for each activity. Each mode is a combination of the duration, cost, and resource requests. Among these, the cost and resource requests can be calculated according to the duration, the quota, and the quantity. Assuming that the data entry for one quota item takes five seconds, it requires a total of 7 minutes to complete the 84 quota items for the activities of one story. Assuming that each quota should produce four modes to ensure scheduling result accuracy, the nine quotas for each story produce 36 modes for the activities and require 6 minutes for data entry (assuming that data entry takes ten seconds for one mode).

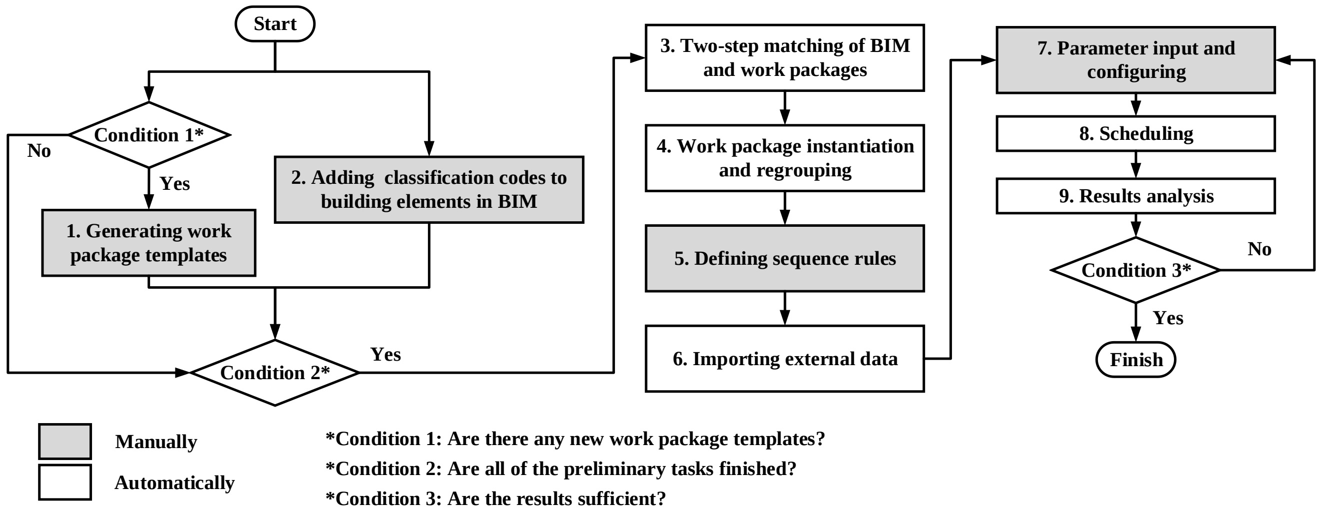
Task 4 is the same as Task 7 in Fig. 23a, and we assume that it takes five minutes.

In conclusion, the time required to complete one workflow for the control group is 5+398+7+6+5*5=441 minutes.

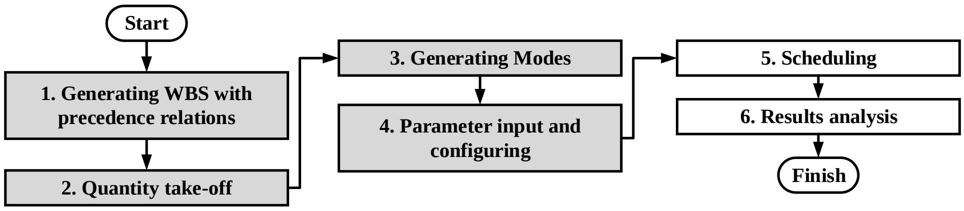
In comparison, the time required to produce the workflow proposed in this study is conservatively estimated to be less than 1/7 of the time required to produce a general RCPSP modeling and scheduling workflow.

Table 8. Time estimation of each step of the general scheduling workflow using RCPSP

Manual task	Subtask	Time consumption (minutes)
1. Generating WBS with precedence relations		5
2. Quantity take-off		2 * 189 activities = 398
3. Generating modes	3.1 Quota items input	1/12 * 84 quota items = 7
	3.2 Modes production	1/6 * 36 modes = 6
4. Parameter input and configuring		5
Average time consumption considering assumption (2) and (5)		5+398+7+6+5*5 = 441



a. Using RCPSP based on semiautomatic data integration



b. General using RCPSP

Fig. 23. Scheduling workflow in construction projects

7.6. Discussion

In the validation process, we generated an integrated information model based on BIM and work packages by adopting unified classification codes. The analysis of the results of each step shows that the process of data integration can be conducted in strict accordance with the theory; consequently, the process and the results of data integration are credible. Subsequently, the result of CP-based scheduling based on data from the integrated information model is also consistent with theoretical results; therefore, the scheduling implementation is also credible. Finally, the overall validation results are credible and are discussed below.

The results show that it is feasible to use BIM, the work package template database, and some external data as data sources for RCPSP. During the data integration process, BIM provides the basic quantities and properties, while the work package template database provides the activity, quota, and resource information. Note that manually defined precedence relations and information from external data sources, including resource prices and resource capacities, are also required. The fact that these data are fully used in scheduling indicates their necessity. Therefore, the information modeling conducted in this study is appropriate. Using the quota-centric work package is a possible choice to meet the data requirement of RCPSPs, and searching for (or designing) better data sources is a potential future research direction.

Moreover, the information model has the potential for use as the data basis to be further applied in scheduling management activities or to support related research works such as schedule monitoring[56].

Based on the results of the case study and efficiency estimation, the following four benefits can be concluded in this study.

1. **The information completeness is improved for RCPSPs in construction projects.** The completeness of the information model is validated according to the results. The proposed CP-based RCPSP model that contains all seven information types shown in Table 1 can generate scheduling results under the constraints defined based on the information. In this model, activity durations change when the mode selection or quantity is changed. Moreover, the multiple modes consist of several combinations of resources with coefficients to define the relations between the activity duration, resource requests, and element quantity. The correct RCPSP validation results under different constraints demonstrate that the scheduling problem modeling of the activity duration and multiple modes is successful. Besides, the precedence relations, deadlines, resource capacities, resource renewability type, and cost are defined in the CP model as existing studies[5,25] did; therefore, although the case study can only validate the limited conditions of these information definitions in the RCPSP model, it is reasonable to trust them.
2. **High efficiency is realized in data acquisition and integration for construction scheduling.** RCPSP data preparation becomes a semiautomatic process when using the approach in this study for construction projects, and it was shown to be considerably more efficient than the general RCPSP modeling and scheduling process. The significant advantage lies in the time consumption of the quantity take-off. Calculating the quantity of each activity is generally a time-intensive job. However, by generating the activities automatically based on BIM and the work package templates, the entire quantity take-off process only requires performing the addition of the work package templates and the classification codes manually. Because the resulting work package templates can be used repeatedly, the main factor that affects the efficiency is the data integration approach based on the classification codes. Also, defining and using unified classification codes is sometimes impossible in actual construction projects. Thus, finding a more efficient or practical data integration approach is one of the most significant factors in improving the results of this study.
3. **A large amount of duplication works in data preparation are avoided by introducing the knowledge-based data source for RCPSPs**. The work package-based information model provides a new approach to apply knowledge basis in automatic schedule generation. Moreover, the approach is also a first step of generating the RCPSP model using the knowledge-based data source. The case study is an epitome of practical construction project scheduling. A large number of activities in similar content exist in the schedule of one project or schedules of similar projects. Defining

work package templates can not only avoid duplication works but also make the data more credible.

4. **The information modeling and data integration methods bring more possibilities to RCPSP modeling**. The specific RCPSP model proposed in this study shows more possibilities to improve construction scheduling. Table 6 indicates that the selection of the activity mode, which represents a construction method, can be affected by the resource capacity and the scheduling objective. This condition has never been considered in the existing studies of RCPSP. Moreover, the activity duration function can express the complicated relationship between the duration and resource requests, making it possible to introduce research results about the construction productivity into scheduling. By further introducing the calendar, different types of work time, and different salaries, a more detailed RCPSP considering crash can be defined. Limited by the research time and the efficiency of the CP optimizer, the problem model in this study is still far from being suitable for practical construction projects, as it still takes several hours to solve the problem during validation. Therefore, it is worth considering developing new RCPSP models, along with new integrated information models, and improving the problem-solving efficiency.

8. Conclusion and future work

In this study, an integrated approach that enables fluent data flow from the information model to the RCPSP is proposed to solve the data acquisition difficulty of RCPSPs. First of all, an integrated information model that captures all the information required by most RCPSPs is established based on BIM and work package. Then, a semiautomatic data integration approach based on unified classification codes is proposed to form the integrated model from multiple data sources. A new MRCPSP model is designed with the proposed information model in mind and it is possible to consider more parameters and constraints than most of the previous MRCPSP studies in the construction field. Finally, a data transfer method is proposed to automatically generate the proposed MRCPSP model. Therefore, the data integration process and RCPSP model generation and solving process are automated.

With application in a construction project of a residential building, time used for manully RCPSP modeling and scheduling is measured and compared with the time used in the proposed approach, and it shows that only 1/7 of the time is needed with adopting the new approach proposed in this paper. The validation result also illustrates that the proposed information model is sufficient to capture the data required by most RCPSPs while providing the possibility to use a more comprehensive RCPSP model in an actual construction project. As demonstrated in this research, it is possible to consider different construction methods and productivity functions when modeling the RCPSP. Moreover, this kind of information model is also a rich data source for schedule tracking, construction management, cost estimation, etc., and would avoid potential rework as well as efforts and mistakes in data inputing and editing. Additionally, introduced database of work package templates enables reuse of previously-gained engineering knowledge and makes the whole process more efficient. Finally, as observed and summarized in the literature review section, previous work focuses on creating new models and methods for solving RCPSP, while few paied attention to the data integration and process automation part. That is, adopting complex RCPSP model will use more time in data collection and model solving, thus hindering the application of these complex models in real world projects. This research tries to shed light on how to automate the data integration, model generation and solving process and contributes to the state of practice in applying complex RCPSP models in real worl projects.

Nevertheless, improvements and potentional extensions of the proposed information model and approach are still needed. These include:

1. Extending the proposed information model to capture more data in the construction domain not only for schedule and resource optimization but also for cost estimation, safety management purpose.
2. Integrating different methods, such as GA and PSO, to solve the RCPSPs instead CP-based method, and exploring how to automate the model generation and solving process, and a unified framework for automatically generating and solving different models could be established.
3. Automatic and dynamic schedule optimization as the construction process goes can be realized by introducing automatic progress tracking techniques. Vision-based schedule tracking and information collection methods are suitable for updating the actual progress of the proposed information model.
4. As the information model becomes more complex and the data amount goes higher, user-friendly data retrieval methods[57] for request for information and sptio-temporal-semantic analysis for schedule valiation are also important.
5. The project data, especially the knowledge base, may encounter ownership issues in actual construction projects, cross-party collaboration environment with the consideration of data privacy[58] is important.
6. With proposed approach in this research, different RCPSP models can be generated and solved automatically, and it is possible to compare models with different parameters or factors in an efficient way. This would help the researchers and engineers determine key factors that influence the time, cost of a construction project.

Acknowledgments

This work was funded by the National Key R&D Program of China (No. 2018YFD1100900), the National High Technology Research and Development Program (863 Program) of China (No.2013AA041307). Dr. Lin is supported by the Beijing Natural Science Foundation (No. 8194067) and the Young Elite Scientists Sponsorship Program by China Association for Science and Technology (No. QNRC2016001).

References

[1] J.K. Larsen, G.Q. Shen, S.M. Lindhard, T.D. Brunoe, Factors Affecting Schedule Delay, Cost Overrun, and Quality Level in Public Construction Projects, Journal of Management in Engineering 32 (040150321) (2016), 10.1061/(ASCE)ME.1943-5479.0000391.

[2] R. Sonmez, M.A. Iranagh, F. Uysal, Critical Sequence Crashing Heuristic for Resource-Constrained Discrete Time-Cost Trade-Off Problem, Journal of Construction Engineering and Management 142 (040150903) (2016), 10.1061/(ASCE)CO.1943-7862.0001077.

[3] P. Brucker, A. Drexl, R. Mohring, K. Neumann, E. Pesch, Resource-constrained project scheduling: Notation, classification, models, and methods, European Journal of Operational Research 112 (1) (1999) 3-41, 10.1016/S0377-2217(98)00204-5.

[4] R. Sonmez, F. Uysal, Backward-Forward Hybrid Genetic Algorithm for Resource-Constrained Multiproject Scheduling Problem, Journal of Computing in Civil Engineering 29 (040140725) (2015), 10.1061/(ASCE)CP.1943-5487.0000382.

[5] Z. Abuwarda, T. Hegazy, Work-Package Planning and Schedule Optimization for Projects with Evolving Constraints, Journal of Computing in Civil Engineering 30 (040160226) (2016), 10.1061/(ASCE)CP.1943-5487.0000587.

[6] S. Pinheiro, R. Wimmer, J. O'Donnell, S. Muhic, V. Bazjanac, T. Maile, J. Frisch, C. van Treeck, MVD based information exchange between BIM and building energy performance simulation, Automation in Construction 90 (2018) 91-103, 10.1016/j.autcon.2018.02.009.

[7] J.R. Lin, J.P. Zhang, X.Y. Zhang, Z.Z. Hu, Automating closed-loop structural safety management for bridge construction through multisource data integration, Advances in Engineering Software 128 (2019) 152-168, 10.1016/j.advengsoft.2018.11.013.

[8] C.F. Fu, G. Aouad, A. Lee, A. Mashall-Ponting, S. Wu, IFC model viewer to support nD model application, Automation in Construction 15 (2) (2006) 178-185, 10.1016/j.autcon.2005.04.002.

[9] L. Ding, Y. Zhou, B. Akinci, Building Information Modeling (BIM) application framework: The process of expanding from 3D to computable nD, Automation in Construction 46 (2014) 82-93, 10.1016/j.autcon.2014.04.009.

[10] J.P. Zhang, Z.Z. Hu, BIM- and 4D-based integrated solution of analysis and management for conflicts and structural safety problems during construction: 1. Principles and methodologies, Automation in Construction 20 (2SI) (2011) 155-166, 10.1016/j.autcon2010.09.013.

[11] H. Moon, N. Dawood, L. Kang, Development of workspace conflict visualization system using 4D object of work schedule, Advanced Engineering Informatics 28 (1) (2014) 50-65, 10.1016/j.aei.2013.12.001.

[12] N.C. Babic, P. Podbreznik, D. Rebolj, Integrating resource production and construction using BIM, Automation in Construction 19 (5SI) (2010) 539-543, 10.1016/j.autcon.2009.11.005.

[13] D. Kehily, J. Underwood, EMBEDDING LIFE CYCLE COSTING IN 5D BIM, Journal of Information Technology in Construction 22 (2017) 145-167, http://www.itcon.org/2017/8 (May 7, 2019).

[14] H. Kim, K. Anderson, S. Lee, J. Hildreth, Generating construction schedules through automatic data extraction using open BIM (building information modeling) technology, Automation in Construction 35 (2013) 285-295, 10.1016/j.autcon2013.05.020.

[15] W. Lu, T. Olofsson, Building information modeling and discrete event simulation: Towards an integrated framework, Automation in Construction 44 (2014) 73-83, 10.1016/j.autcon.2014.04.001.

[16] W. Wang, S. Weng, S. Wang, C. Chen, Integrating building information models with construction process simulations for project scheduling support, Automation in Construction 37 (2014) 68-80, 10.1016/j.autcon.2013.10.009.

[17] H. Liu, M. Al-Hussein, M. Lu, BIM-based integrated approach for detailed construction scheduling under resource constraints, Automation in Construction 53 (2015) 29-43, 10.1016/j.autcon.2015.03.008.

[18] S. Mohammadi, M. Tavakolan, B. Zahraie, Automated Planning of Building Construction Considering the Amount of Available Floor Formwork, CONSTRUCTION RESEARCH CONGRESS 2016: OLD AND NEW CONSTRUCTION TECHNOLOGIES CONVERGE IN HISTORIC SAN JUAN (2016) 2197-2206, 10.1061/9780784479827.219.

[19] S. Hartmann, D. Briskorn, A survey of variants and extensions of the resource-constrained project scheduling problem, European Journal of Operational Research 207 (1) (2010) 1-14, 10.1016/j.ejor.2009.11.005.

[20] S.S. Leu, C.H. Yang, GA-based multicriteria optimal model for construction scheduling, JOURNAL OF CONSTRUCTION ENGINEERING AND MANAGEMENT-ASCE 125 (6) (1999) 420-427, 10.1061/(ASCE)0733-9364(1999)125:6(420).

[21] Z.Y. Zhao, W.Y. You, J. Zuo, Application of Innovative Critical Chain Method for Project Planning and Control under Resource Constraints and Uncertainty, Journal of Construction Engineering and Management 136 (9) (2010) 1056-1060, 10.1061/(ASCE)CO.1943-7862.0000209.

[22] C. Feng, Y. Chen, J. Huang, Using the MD CAD model to develop the time-cost integrated schedule for construction projects, Automation in Construction 19 (3SI) (2010) 347-356, 10.1016/j.autcon.2009.12.009.

[23] S.M. Christodoulou, Scheduling Resource-Constrained Projects with Ant Colony Optimization Artificial Agents, Journal of Computing in Civil Engineering 24 (1) (2010) 45-55, 10.1061/(ASCE)0887-3801(2010)24:1(45).

[24] Z. Zhang, J. Xu, A multi-mode resource-constrained project scheduling model with bi-random coefficients for drilling grouting construction project, INTERNATIONAL JOURNAL OF CIVIL ENGINEERING 11 (1) (2013) 1-13,

http://ijce.iust.ac.ir/article-1-699-en.html (May 7, 2019).

[25] W. Menesi, B. Golzarpoor, T. Hegazy, Fast and Near-Optimum Schedule Optimization for Large-Scale Projects, Journal of Construction Engineering and Management 139 (9) (2013) 1117-1124, 10.1061/(ASCE)CO.1943-7862.0000722.

[26] P. Ghoddousi, E. Eshtehardian, S. Jooybanpour, A. Javanmardi, Multi-mode resource-constrained discrete time-cost-resource optimization in project scheduling using non-dominated sorting genetic algorithm, Automation in Construction 30 (2013) 216-227, 10.1016/j.autcon.2012.11.014.

[27] M. Cheng, T. Duc-Hoc, Y. Wu, Using a fuzzy clustering chaotic-based differential evolution with serial method to solve resource-constrained project scheduling problems, Automation in Construction 37 (2014) 88-97, 10.1016/j.autcon.2013.10.002.

[28] O.H. Bettemir, R. Sonmez, Hybrid Genetic Algorithm with Simulated Annealing for Resource-Constrained Project Scheduling, Journal of Management in Engineering 31 (040140825) (2015), 10.1061/(ASCE)ME.1943-5479.0000323.

[29] W. Menesi, T. Hegazy, Multimode Resource-Constrained Scheduling and Leveling for Practical-Size Projects, Journal of Management in Engineering 31 (UNSP 040140926) (2015), 10.1061/(ASCE)ME.1943-5479.0000338.

[30] M. Cheng, D. Tran, An efficient hybrid differential evolution based serial method for multimode resource-constrained project scheduling, KSCE Journal of Civil Engineering 20 (1) (2016) 90-100, 10.1007/s12205-015-0414-0.

[31] Z. Abuwarda, T. Hegazy, Flexible Activity Relations to Support Optimum Schedule Acceleration, Journal of Construction Engineering and Management 142 (0601600411) (2016), 10.1061/(ASCE)CO.1943-7862.0001193.

[32] Z. Zhang, M. Liu, X. Zhou, L. Chen, A Multi-objective DCP Model for Bi-Level Resource-Constrained Project Scheduling Problems in Grounding Grid System Project under Hybrid Uncertainty, KSCE Journal of Civil Engineering 20 (5) (2016) 1631-1641, 10.1007/s12205-015-0615-6.

[33] A. Kaveh, M. Khanzadi, M. Alipour, Fuzzy Resource Constraint Project Scheduling Problem Using CBO and CSS Algorithms, INTERNATIONAL JOURNAL OF CIVIL ENGINEERING 14 (5A) (2016) 325-337, 10.1007/s40999-016-0031-4.

[34] R. Sonmez, M. Gurel, Hybrid Optimization Method for Large-Scale Multimode Resource-Constrained Project Scheduling Problem, Journal of Management in Engineering 32 (040160206) (2016), 10.1061/(ASCE)ME.1943-5479.0000468.

[35] K. Kim, J. Walewski, Y.K. Cho, Multiobjective Construction Schedule Optimization Using Modified Niched Pareto Genetic Algorithm, Journal of Management in Engineering 32 (040150382) (2016), 10.1061/(ASCE)ME.1943-5479.0000374.

[36] E. Elbeltagi, M. Ammar, H. Sanad, M. Kassab, Overall multiobjective optimization of construction projects

scheduling using particle swarm, Engineering Construction and Architectural Management 23 (3) (2016) 265-282, 10.1108/ECAM-11-2014-0135.

[37] M.F. Siu, M. Lu, S. AbouRizk, Resource Supply-Demand Matching Scheduling Approach for Construction Workface Planning, Journal of Construction Engineering and Management 142 (040150481) (2016), 10.1061/(ASCE)CO.1943-7862.0001027.

[38] T. Duc-Hoc, M. Cheng, M. Cao, Solving Resource-Constrained Project Scheduling Problems Using Hybrid Artificial Bee Colony with Differential Evolution, Journal of Computing in Civil Engineering 30 (040150654) (2016), 10.1061/(ASCE)CP.1943-5487.0000544.

[39] Z. Liu, L. Yang, R. Deng, J. Tian, An effective approach with feasible space decomposition to solve resource-constrained project scheduling problems, Automation in Construction 75 (2017) 1-9, 10.1016/j.autcon.2016.11.012.

[40] O. Giran, R. Temur, G. Bekdas, Resource Constrained Project Scheduling by Harmony Search Algorithm, KSCE Journal of Civil Engineering 21 (2SI) (2017) 479-487, 10.1007/s12205-017-1363-6.

[41] O. Liess, P. Michelon, A constraint programming approach for the resource-constrained project scheduling problem, Annals of Operations Research 157 (1) (2008) 25-36, 10.1007/s10479-007-0188-y.

[42] I. Wu, S. Hsieh, A framework for facilitating multi-dimensional information integration, management and visualization in engineering projects, Automation in Construction 23 (2012) 71-86, 10.1016/j.autcon.2011.12.010.

[43] L.S. Kang, H.S. Kim, H.S. Moon, S. Kim, Managing construction schedule by telepresence: Integration of site video feed with an active nD CAD simulation, Automation in Construction 68 (2016) 32-43, 10.1016/j.autcon.2016.04.003.

[44] H.X. Li, L. Zhang, D. Mah, H. Yu, An integrated simulation and optimization approach for reducing CO2 emissions from on-site construction process in cold regions, Energy and Buildings 138 (2017) 666-675, 10.1016/j.enbuild.2016.12.030.

[45] W. Jeong, S. Chang, J. Son, J. Yi, BIM-Integrated Construction Operation Simulation for Just-In-Time Production Management, Sustainability 8 (110611) (2016), 10.3390/su8111106.

[46] M. Konig, C. Koch, I. Habenicht, S. Spieckermann, INTELLIGENT BIM-BASED CONSTRUCTION SCHEDULING USING DISCRETE EVENT SIMULATION, in: C. Laroque, J. Himmelspach, R. Pasupathy (Eds.), Winter Simulation Conference Proceedings, IEEE, NEW YORK, 2012, 10.1109/WSC.2012.6465232.

[47] W. Xue, Y. Wang, Q. Man, Research on Information Models for the Construction Schedule Management Based on the IFC Standard, JOURNAL OF INDUSTRIAL ENGINEERING AND MANAGEMENT-JIEM 8 (3) (2015) 615-635, 10.3926/jiem.1283.

[48] M. Niknam, S. Karshenas, Integrating BIM and Project Schedule Information Using Semantic Web Technology,

CONSTRUCTION RESEARCH CONGRESS 2016: OLD AND NEW CONSTRUCTION TECHNOLOGIES CONVERGE IN HISTORIC SAN JUAN (2016) 689-697, 10.1061/9780784479827.070.

[49] K. Sigalov, M. Konig, Recognition of process patterns for BIM-based construction schedules, Advanced Engineering Informatics 33 (2017) 456-472, 10.1016/j.aei.2016.12.003.

[50] J. Yeoh, T.Q. Nguyen, E. Abbott, Construction Method Models Using Context Aware Construction Requirements for Automated Schedule Generation, in: K.Y. Lin, N. ElGohary, P. Tang (Eds.), AMER SOC CIVIL ENGINEERS, NEW YORK, 2017, pp. 60-67, 10.1061/9780784480847.008.

[51] Y. Fang, S.T. Ng, Z. Ma, H. Li, Quota-based carbon tracing model for construction processes in China, Journal of Cleaner Production 200 (2018) 657-666, 10.1016/j.jclepro.2018.08.028.

[52] OCCS Development Committee, OmniClass: a Strategy for Classifying the Built Environment, 2012.

[53] Construction Project Information Committee, Classification - Technical Support - NBS BIM Toolkit, 2015.

[54] D. Trang, H. Bargstaedt, 4D Relationships: The Missing Link in 4D Scheduling, Journal of Construction Engineering and Management 142 (040150722) (2016), 10.1061/(ASCE)CO.1943-7862.0001007.

[55] Z.Z. Hu, J.P. Zhang, BIM- and 4D-based integrated solution of analysis and management for conflicts and structural safety problems during construction: 2. Development and site trials, Automation in Construction 20 (2SI) (2011) 167-180, 10.1016/j.autcon.2010.09.014.

[56] X. Luo, H. Li, D. Cao, F. Dai, J. Seo, S. Lee, Recognizing Diverse Construction Activities in Site Images via Relevance Networks of Construction-Related Objects Detected by Convolutional Neural Networks, Journal of Computing in Civil Engineering 32 (UNSP 040180123) (2018), 10.1061/(ASCE)CP.1943-5487.0000756.

[57] J. Lin, Z. Hu, J. Zhang, F. Yu, A Natural-Language-Based Approach to Intelligent Data Retrieval and Representation for Cloud BIM, Computer-Aided Civil and Infrastructure Engineering 31 (1) (2016) 18-33, 10.1111/mice.12151.

[58] J. Zhang, Q. Liu, Z. Hu, J. Lin, F. Yu, A multi-server information-sharing environment for cross-party collaboration on a private cloud, Automation in Construction 81 (2017) 180-195, 10.1016/j.autcon.2017.06.021.