A Model for Integrating Construction Design and Schedule Data

Zhigang Shen  
*University of Nebraska - Lincoln*, shen@unl.edu

Raja R. A. Issa  
*University of Florida*, raymond-issa@ufl.edu

William O'Brien  
*University of Florida*, wjob@ufl.edu

Follow this and additional works at:  [http://digitalcommons.unl.edu/constructionmgmt](http://digitalcommons.unl.edu/constructionmgmt)

Part of the *Construction Engineering and Management Commons*

http://digitalcommons.unl.edu/constructionmgmt/7

This Article is brought to you for free and open access by the Construction Systems at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Construction Management by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
A Model for Integrating Construction Design and Schedule Data

Zhigang Shen, Raja R.A. Issa, and William O’Brien

Affiliation as of 2011: University of Nebraska-Lincoln; zshen3@unl.edu
Rinker School of Building Construction, University of Florida, Gainesville, Florida, USA; raymond-issa@ufl.edu and wjob@ufl.edu

Abstract

The difficulty of timely retrieval of useful information from heterogeneous data sources is a major cause of low productivity in the construction industry. The goal of this research is to provide a new methodology to handle the data-heterogeneity problems encountered in a construction project. This research proposes a tree-structured product model, which binds design knowledge, cost data and schedule data together, as a feasible solution for the data integration problem in construction projects. The methodology uses the knowledge representation of construction projects based on an ontology. Metadata are used to describe the conceptual structure of the project knowledge. The concept of views is used to answer particular queries from different domains.

A general methodology is under development to support queries about a particular construction project from different user perspectives based on heterogeneous construction data sources in a dynamic environment. The research is composed of two major parts: (1) organize the heterogeneous construction data into a tree structure; and (2) retrieve information and obtain domain views by specifying the ways of traversing the tree. The future goal is to develop a prototype that will support some of the major functions needed in today’s construction projects. Functional examples are used to demonstrate the validity of the proposed methodology.

Introduction

Significant improvements in productivity have been achieved in the manufacturing industry through the application of information integration tools, such as CAD/CAM, in their daily operations. However, in the construction industry, the improvement has been quite limited even though CAD, scheduling and estimating software have been widely used in construction for quite a long time. Various reasons have been identified for this discrepancy—an important one is the lack of information integration across the fragmented participants in construction projects. Indeed, around 60-70 percent of the working hours of a general contractor’s staff can be spent clarifying and confirming design or construction information (Nambayashi et al. 2000).

In the majority of current construction projects, CAD systems are still used only as the drawing tool for the architect, instead of being used as an integrated design information database. As a result, information retrieval has to be done manually in spite of its electronic format. In a typical scenario, when engineers, project managers, estimators or even other architects want to retrieve design information from the drawings, they might have to search dozens, if not hundreds of documents to find the scattered pieces and put them...
The information search and retrieval process becomes an extremely time-consuming job, especially for large and complex projects. Furthermore, this manual retrieval process is error prone. It has been estimated that up to 50% of all engineering changes in the manufacturing industry are corrections of errors rather than changed requirements or real improvements (Engelke 1987). Considering the fragmented nature of the construction industry, the error rate is most likely higher than that in the manufacturing industry.

Even though the design, cost estimating and project scheduling processes are computerized individually, the information retrieval situation does not improve much because each application is using its own data model. The high cost associated with information searching and retrieval is one of the causes associated with the low productivity of the construction industry.

With the increased use of construction software, the traditional cross-disciplinary communication problem is becoming increasingly an issue of data exchange and data sharing between different software applications. Unfortunately, most of the software packages used in the construction industry have their own proprietary data format, because they were developed by different vendors. Several international standardization efforts, including STEP (STandard for the Exchange of Product model data) and IFC (the Industry Foundation Classes), have been undertaken to address this communication problem, however, the results have been mixed so far. A common data standard does not, in itself, provide the necessary means to integrate information and provide views useful to humans. Hence, the goal of this study is propose and develop a new data integration model for construction that to some degree extends our current capabilities to provide relevant information to decision makers.

Background

Researchers have looked at common product and/or process models as the solution to the problems caused by fragmented construction documents. Many models have been proposed in trying to address construction data integration issues by both industry and academia. Initially, these attempts could be divided into two (complementary) categories: (1) the standard (product and/or process) model; and (2) the Integrated Project Database approach. Both categories are discussed in detail below.

Standard (Product and/or Process) Model

Standardization requires all applications to use the same data model to solve the communication problem among the different domain applications, thus allowing all the applications to be able to exchange data with each other. Alternatively, it requires those applications to at least support a neutral data format as a medium for the data exchange. This neutral data format is an integrated data model, which captures the full semantics of a building system and its components. In this approach, each application only works with a subset of the model. This subset is often described as an aspect model. Data exchange can be achieved between all the aspect models by mapping through the integrated data model.

There are two international standardization efforts that address the representation of building designs: STEP being developed by the International Organization for Standardization; and IFC, which are specifications for a set of standardized object def-
A Model for Integrating Construction Design and Schedule Data

Inns, being developed by the International Alliance for Interoperability (IAI). Both standards define static building representations aimed at the transfer of information between applications.

In addition to the international standardization efforts, there exist several local standardization efforts, for example, the RATAS model (Enkovaara et al. 1988), the Integrated Data Model of the COMBINE project (Sun and Lockley 1997), and the Logical Product Model for CIMSteel (2000). All these models have adopted an object-oriented paradigm, which describes the building system using objects or entities and their inter-relationships and they are also compliant with the STEP or IFC data models.

Integrated Project Database Approach

This approach integrates AEC applications through an implemented project database. The aim of an Integrated Project Database is to provide consistent and reliable storage of the project information, and to serve as a data exchange hub for different tasks during the construction process. An example of this approach is Gallicon (Sun et al. 2000).

One problem with both approaches is that these representations lack extensibility. They are limited by a predefined semantic representation. Any changes to the predefined representation require the update of the whole system. In this sense, these systems are tightly coupled. Cases may occur over time in which changes are needed or some components may need to be added to the representation to adapt it to some particular context. Thus, the predefined schema approach may be useful in a general sense, but may be hard to handle in some particular situations (O’Brien et al. 2002). Another problem is that some of these representations (like RATAS) do not support multiple views from different domains because of their rigid representation requirements (Rivard and Fenves 2000).

Related Recent Models and Approaches

To overcome the drawbacks brought by the predefined schema, other integration approaches have been proposed in recent years. Some examples of these approaches are CIFE’s 4D CAD approach using CAD, schedule and cost integration (Staub-French and Fisher 2001), the Mediation/Wrapper approach (O’Brien et al. 2002), the Virtual Product Model (Clayton et al. 1996), and SEED (Software Environment to support the Early phases in building Design) (Rivard and Fenves 2000).

4D CAD Approach

CIFE’s case study of integrated scope-cost-time using off-the-shelf software tools showed that commercial tools can integrate design, cost and schedule to some extent using the 4D model. However, it also found some limitations to this approach:

- Most links between design objects and cost assemblies have to be done manually (CAD object-activity, object-cost assembly).
- Design Changes are hard to handle, so maintaining cost data over the project lifecycle is a problem.
- Schedule and cost are not linked, so the impact of schedule change cannot be readily evaluated.
The researchers concluded that “Commercial tools that can electronically integrate design, cost, and schedule information provide many benefits to project teams throughout the design and construction process. Specifically, they enable the early detection of design conflicts, shorten estimating time and improve estimating reliability, improve the communication of schedule intent (Staub-French and Fisher 2001).” The study also demonstrated that 3D CAD object data could be extracted from the application directly and that the extracted dimension data could be input into the estimating application by directly linking variables. Furthermore, the study also identified that there are research needs for developing a methodology to relate design and cost information at different levels of detail.

**Mediation/Wrapper Approach**

The Mediation/Wrapper approach proposed by O’Brien et al. (2002) was intended to handle the heterogeneity of construction legacy information and does not require that the involved application subscribe to data standards. It semiautomatically configures a wrapper for each particular application and answers the query from its mediator. The wrapper serves as an interface between the mediator and the domain specific database. A particular mediator is responsible for a particular domain view by extracting and summarizing the extracted data. This approach is intended for integrating distributed legacy databases to support supply chain management. The queries in this proposed approach require complicated query rewriting and transformation. Moreover, each query requires multi-level sub-queries and summation and cleansing of the sub-query results, which could lead to problems in distributed data sources. Data consistency among the distributed databases is another issue that is hard to address in this approach.

**Virtual Product Model**

The Virtual Product Model (Clayton et al. 1996) supports early stage preliminary conceptual design. It differs from the general object-oriented models in that it maps the explicitly represented semantic and graphic entities rather than unifying the two. The system is composed of four modules: form; function; behavior; and reasoning. It uses a separate symbolic model to represent the architectural design plan in 3D. The symbolic model is generated after the designer establishes the 3D drawing by interactively assigning 3D forms and function specifications to the symbolic nodes. The reasoning module subsequently uses the annotated symbolic module to predict the behavior of the design.

This approach avoids the rigidity of the predefined data model and achieves extensibility at the cost of manually mapping the design components into the symbolic nodes, which will be subsequently used in the reasoning module. Another limitation of this approach is that it only applies to the very early stages of the design, which consists of highly aggregated building components and thus will require only minor manual mapping between CAD and the symbolic model. This approach will be very inefficient if more detailed design information is required, for example, when the design is in the development stage and a design alternative needs to be evaluated. The third limitation is that this model ignores the none-graphic information, which is an integrated part of the design documents. The none-graphic information is equally important to the success of a project.
**SEED Model**

The building representation of SEED (Rivard and Fenves 2000) consists of two levels of abstraction models defined on top of the object-oriented data model. The first level consists of the information model, called the building entity and the technology model, which stores design data as they are generated during the conceptual design, supports case based reasoning, and shares data among all design participants. The technology model represents each building entity as a generic container that encompasses its properties, taxonomy, geometry, composition, relationships, and design knowledge applied in its generation. The second level is a conceptual model that defines the types of objects, relationships, and data needed to fully represent the information in a given design domain. The conceptual model specifies the semantics of the design information for the domain using the syntax defined in the information model. The representation has the following advantages: (1) it integrates multiple views; (2) supports design evolution; (3) supports design exploration; and (4) is extensible.

The SEED representation model shares some characteristics with the Virtual Product Model. One important overlap in the two models is that both of them use data in the representation model as the base for design reasoning. The major difference between the two is that the SEED model is a self-contained application, which is independent from CAD objects. The SEED model itself contains the geospatial relations among design objects. The contribution of this model is that it explicitly states that a Primitive-Composite hierarchical product model could be used to represent the building design to support multiple views, design evolution, and design exploration, and it is extensible. However, it does not solve the problem of how to dynamically generate the different domain views in a collaborative design environment.

**Information Model for Design Coordination**

Hegazy (2001) proposed an integration information model for improving design coordination, which, as shown in Figure 1, represents an information model for storing design information, recording design rationale, and managing design changes. The model is composed of three main parts: (1) a building project hierarchy (BPH); (2) a building components library (BCL); and (3) design-change management (DCM) procedures.

![Figure 1. Information Model for Design Coordination (Hegazy et. al )](image)
The first part, the BPH, stores all building data and represents this data as a hierarchy of active objects. Each object has information about its values, documents, design rationale, and communication paths. The second part, the BCL, is a central repository of common building components that allows a building project hierarchy to be described with little effort. The third part, the DCM procedures, manages the design changes made to any object in the BPH and keeps track of the history of changes made by all disciplines. This information model provides a good reference for the proposed model. But there are also problems with this model: (1) the BPH and BCL do not link to the 3D CAD component, instead they link to the CAD document, which makes it difficult in practice to check the consistency and redundancy between the design component and the BCL; and (2) because only a small portion of the information about each domain needs to be exchanged for design coordination, the four central databases, one for each professional (architect, structural engineer, mechanical engineer and electrical engineer), may store a lot of information that does not need to be exchanged. This model could be further improved by linking to the true objects in the CAD tool.

Summary

Table 1 shows a comparison between the existing models and approaches aimed at solving the data integration issues faced by the construction industry and the proposed design component and schedule data integration model. Unlike the other models and approaches shown, the proposed model is extensible, handles design changes and allows for dynamic multiple views.

<table>
<thead>
<tr>
<th>Models/Approaches</th>
<th>Extensible</th>
<th>Handle Design Changes</th>
<th>Dynamic Multiple Views</th>
<th>Unnecessary Information Sharing</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEP (ISO) &amp; IFC (IAI) Data Models</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Integrated Project Database</td>
<td>No</td>
<td>Difficult</td>
<td>Difficult</td>
<td>Yes</td>
</tr>
<tr>
<td>SEEK Mediation/Wrapper Approach</td>
<td>Yes</td>
<td>Difficult</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>CIFE 4D CAD Approach</td>
<td>N/A</td>
<td>Difficult</td>
<td>Difficult</td>
<td>No</td>
</tr>
<tr>
<td>Virtual Product Model</td>
<td>Yes</td>
<td>Difficult</td>
<td>Difficult</td>
<td>No</td>
</tr>
<tr>
<td>SEED Model (Rivard and Fenves 2000)</td>
<td>Yes</td>
<td>Yes</td>
<td>Difficult</td>
<td>Yes</td>
</tr>
<tr>
<td>Info. Model For Design Coordination</td>
<td>Yes</td>
<td>Yes</td>
<td>Difficult</td>
<td>Yes</td>
</tr>
<tr>
<td>Proposed Model</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Case Study

Case Study Parameters

Cost-resource loading is currently used in CPM to load a construction project with the corresponding cost and resources it consumes. Because the CPM schedule is used for scheduling and controlling progress, the ‘dynamic accuracy’ of the cost and takeoff data for the corresponding activities is crucial for tight control of the project, especially for a project that needs to apply “Just In Time” (JIT) material management principles (Pheng and Meng 1997; Thomas et al. 1989). However, by far, most of the resource-cost loading in CPM applications is done manually using the takeoff data from the project estimators, which is usually in Construction Specifications Institute (CSI) format and is procurement-oriented. Most activities in CPM applications are building-system-component oriented. Furthermore, many useful estimating data for CPM are transitory during the estimating process and end up being summarized and lost in the final spreadsheet. This discrepancy makes it hard to build the explicit correspondence between the estimating data and the CPM activity. The discrepancy is more obvious when the projects are broken-down into a more detailed level in the Work Breakdown System (WBS).

Because of the dynamic nature of construction projects (frequent changes), the situation becomes worse. For example, cumulative updating of the cost data in the CPM may cause more inaccuracy. To handle unanticipated events on the jobsite, temporary activities may need to be added to the schedule. In such cases, the scheduler should be able to get the data input needed in a timely manner from the project information system, instead of having to regenerate it from the construction documents. Moreover, to better monitor and control the project, progress data from the field also needs to be integrated and reflected timely and accurately in the project documents. As a result, the manual application of resource-cost loading is hardly accurate and timely enough for the scheduling of the material delivery and control, and to apply the JIT principle.

Accordingly, what is required is a means to build the direct link between the estimate and field data about each building component and the corresponding activities, so that any data about the particular component can be input into the CPM activities in a timely and accurate manner. In a more general sense, there is a need for a data model, which is fine-grained enough to answer queries for any system components, while at the same time, this model needs to be flexible enough to answer queries from different perspectives at different abstraction levels. In the following case example a common scenario will be described to illustrate the need for accurate and timely data about building components in order to achieve better material management and to better control the progress of the project.

Case Example

A construction site located in a busy urban center with very tight storage spaces (an actual example, as shown in Figure 2, is the Tomorrow Square in Shanghai). Shipments are allowed only at night. The building includes a 6-story base and a 60-story tower, which has office space in the lower part and a hotel in the upper part. The hotel floors consisted of flat concrete slabs and the office floors were concrete beam and slab construction.

As an example, consider the 6th floor, the building is divided into three construction areas: A, B and core areas for fast track construction, which allows at any moment
for many parallel activities. Many of those activities have dependency relationships, which in the case of insufficient material purchase may cause chain delays. The drywall, plumbing and mechanical subcontractors are working on the ground floor; the masonry subcontractor is working on the 2nd floor. The concrete subcontractor is working on the 3rd floor: some workers are pouring the columns in the B area; other workers are pouring the concrete in the A and core areas. To prepare for the procurement of the resources needed for that day, the takeoff quantities have to be detailed enough to be able to answer queries regarding the specific building components in each of the building systems. An example situation occurs when the procurer needs to know the quantity takeoff of all the columns, beams and concrete walls in the core area and the quantity takeoffs of all interior drywalls plus the mechanical and plumbing system-component information in these areas.

This requires that that the procurer is able to quickly quantify all the material that is required for every building component. In addition, the summation needs to be at different levels of specificity. Ideally, the project manager and site manager will need to be able to track the progress of all major system components to accomplish this goal.

Problem Statement

Current CPM scheduling applications are not good enough for providing lower-level detail material management functions (such as JIT or pull-driven resource scheduling) and cost monitor-control functions (Chehayeb and AbouRizk 1998), because they are intended for high-level time-cost control. CPM scheduling methods use the concept of activity or task to describe the construction process. On the one hand, activities are generally defined according to the tangible deliverables or major building components. Activities are often grouped in the form of packages in a work-breakdown-system (WBS) for higher-level management. On the other hand, the resource-cost loading in CPM is based on the estimate sheet broken down by the CSI specification, which facilitates procurements and sub-contracts.

The lack of the necessary mechanism for mapping activities to the corresponding building components make it hard to implement accurate resource loading and to accurately record the real progress on site. The discrepancy of the views makes it difficult to get the cost estimate for a specific project deliverable, such as a specific building compo-
nent or a building space. Thus, although some scheduling application software packages allow for cost-resource loading, it is hard to accurately load the corresponding resources and costs into the relevant activities. Therefore, cost-resource control is less reliable than it should be.

The case example presented earlier puts a great challenge on CPM using the traditional resource-cost loading method. For example, while the project manager creates activities in a larger aggregation of components, the site manager generally arranges activities around smaller components to handle real site conditions such as daily material delivery to the site. However, without knowing the material quantities for any particular component, it is hard for the site manager to predict the material to be consumed in the subsequent day or even in subsequent weeks. So the traditional resource-cost loading method in CPM is incapable of handling the second scenario described in the case study. Thus, the project manager will not have access to real time information about material consumption and the percent-completion bar chart scheduling method is hard to interpret as actual progress (Chehayeb and AbouRizk 1998) – consequently, costly mistakes may result.

Clearly, in order to overcome the limitations of the models and approaches previously mentioned, a mapping mechanism needs to be developed that is able to explicitly map the schedule activities to the corresponding design components. As some of the approaches and models (Clayton et al. 1996; Staub-French and Fisher 2001) indicated manual mapping between the two applications significant limits the practical usage of the Integration models and approaches. In the case study presented, to achieve JIT or (partial JIT) material management, a model that is able to automatically map two applications; to dynamically (in order to handle frequent changes) update data; and to provide multiple views is needed. This requires that the model has a relatively fixed structure to achieve automatic mapping, and at the same time, it also requires the model to have the flexibility to reflect frequent changes and provide multiple views. The models and approaches reviewed earlier have met one requirement or another, but none of them have met all three requirements, as will be the case with the proposed design component and schedule data integration model.

**Proposed Model**

*Model Scope and Objectives*

The objective of this research is to develop an integration model to integrate design and schedule data and to support dynamic multiple views from various domains at various detail levels. One application of the model is to support accurate cost and resource loading to improve and extend CPM functionality (such as JIT material management). The expected research contribution of the model is that it provides a formal way to show how to use knowledge representation of construction and Object-Oriented programming to address the integration problem between different domain applications and to address the dynamic nature of construction information.

The scope of this research is limited to providing the fundamental methodology to handle heterogeneous and dynamic data in the construction industry. The validity of the methodology will be demonstrated by showing that it is able to satisfy the research objective. A prototype application will be implemented to prove the concept.
Model Description

To address the problems noted with existing models and approaches, this study proposes a methodology that centers on an integrated product model. This model will be based on three observations:

1. It is not necessary to share in detail all domain knowledge among all participants. The information that needs to be shared among partners is typically only a small portion of that available in most of the domain applications. For example, estimators do not need to know the structural load on a floor or beam in order to develop the estimate nor do they need to know how the architect organized the functional spaces.

2. Despite the frequently unique and changing nature of construction projects, the structure of their functional components remains stable. This allows us to partition them into fixed hierarchical categories.

3. Despite the dynamic nature of construction projects, the changes or reconfigurations to a project are generally within the known, countable, usually limited combination of some basic building block. As such, the majority of the construction knowledge of one project could be reused in a subsequent project. For example, using a design component library we could represent the unique configuration of any given project and different aggregations and combinations of components can be done dynamically using Object Oriented programming techniques.

The proposed model will be composed of three major components: (1) the abstract components tree, representing the building; (2) the concrete components libraries storing the reusable design components; and (3) the cost database. The model interacts with the CAD tool and the project scheduling application through two separate interfaces.

A CAD drawing is an abstract representation of the building, which is aimed at representing abstract concepts such as the form, space transformation, traffic line, the spatial relation of different systems, component dimensions and positions etc. Its ability to represent these abstract concepts is up to several levels of granularity of the building components. The CAD drawing is not good at representing the detailed composition information of components, such as layers of paints on the wall and the composition of a reinforced concrete beam. A line or simple shape in a CAD drawing may need to be linked to the project specifications or an enlarged section drawing to explain its detail composition and/or connection. That is why the specifications and detailed enlarged section drawings are still needed to represent the information in addition to the CAD drawings. Therefore, in this proposed model, the CAD application is used only to provide geometric data of the components. The detailed information about the composition of the components will be extracted from the specifications, which are usually available in text format only. A typical simplified working procedure for this model will be as follows:

1. A completed design document is read as input data into the system. (The CAD output documents contain the components’ id, type, and geometric data and the specifications contain the detailed specification about the composition or combination information about the component.)

2. A component parser will parse the input document and build a component tree based on the internal abstract syntax logic. This tree is unique for each project. If any design changes occurred resulting in the addition or deletion of components, new design documents need to be read again by the parser; and the tree will be rebuilt.
3. Different visitor-applications are used to input or update or extract data from the tree node by traversing the tree. A variety of views could be produced based on the integrated information model.

4. Schedule information could also be entered into the corresponding tree nodes. The quantity and cost data residing in the tree nodes could be introduced into the corresponding activities at any time. Thus, the schedule application can always obtain the latest and most accurate quantity data.

System Architecture

In order to integrate the fragmented data in several domain applications, the model dynamically binds the most recent project data to the component tree node. Subsequently different visitors will trace the tree to conduct their assigned tasks. Figure 3 describes the major components of the proposed model. The model is composed of several loosely coupled modules, which allows for the flexibility of incrementally developing the system. The modules can be implemented one at a time and then assembled and new system functions can be added through new application modules. By using a visitor pattern, each visitor could represent a specific functional application. Visitors can be added at will to add more customized functions to the model.

How does the tree representation describe a building? The proposed model will use a neutral description to avoid domain bias, e.g. using 3D-Axis information to locate design objects in order to describe the building. How could multiple views be generated from a single tree-model? Various domain views will be generated via different traversal paths and there will be many virtual domain trees in existence from which user views can be generated. However, there will be only one tree structure that will be traversed.

Figure 3. Design and Schedule Data Integration System Architecture
Conclusions

The difficulty of timely retrieving of useful information from heterogeneous data sources is one of the major causes that lead to low productivity in the construction industry. This goal of this research is to provide a methodology to handle the data-heterogeneity problems encountered on a construction project. The proposed research will look at a tree-structured product model, which binds design knowledge, cost data and schedule data together, as a feasible solution for the data integration problem encountered on construction projects. The proposed methodology uses the knowledge representation of construction projects based on ontology. A prototype of the proposed model is under development and will be available in the near future.

References


