BIM-GIS INTEGRATED GEOSPATIAL INFORMATION MODEL USING SEMANTIC WEB AND RDF GRAPHS

A-H. Hor\(^*\), A. Jadidi\(^*\), , G. Sohn\(^*\)

\(^*\) GeoICT Lab, ESSE Dept., York University, 4700 Keele St, Toronto, ON M3J1P3, Canada

(ehladi, mjadidi, gsohn)@yorku.ca

Commission IV, WG IV/7

KEY WORDS: BIM, GIS, IFC, CityGML, Semantic Web, RDF Graph, SPARQL

ABSTRACT:

In recent years, 3D virtual indoor/outdoor urban modelling becomes a key spatial information framework for many civil and engineering applications such as evacuation planning, emergency and facility management. For accomplishing such sophisticated decision tasks, there is a large demands for building multi-scale and multi-sourced 3D urban models. Currently, Building Information Model (BIM) and Geographical Information Systems (GIS) are broadly used as the modelling sources. However, data sharing and exchanging information between two modelling domains is still a huge challenge; while the syntactic or semantic approaches do not fully provide exchanging of rich semantic and geometric information of BIM into GIS or vice-versa. This paper proposes a novel approach for integrating BIM and GIS using semantic web technologies and Resources Description Framework (RDF) graphs. The novelty of the proposed solution comes from the benefits of integrating BIM and GIS technologies into one unified model, so-called Integrated Geospatial Information Model (IGIM). The proposed approach consists of three main modules: BIM-RDF and GIS-RDF graphs construction, integrating of two RDF graphs, and query of information through IGIM-RDF graph using SPARQL. The IGIM generates queries from both the BIM and GIS RDF graphs resulting a semantically integrated model with entities representing both BIM classes and GIS feature objects with respect to the target-client application. The linkage between BIM-RDF and GIS-RDF is achieved through SPARQL endpoints and defined by a query using set of datasets and entity classes with complementary properties, relationships and geometries. To validate the proposed approach and its performance, a case study was also tested using IGIM system design.

1. INTRODUCTION

3D virtual indoor/outdoor urban models provides the key information for many purposes such as evacuation planning, emergency and facility management (Chan et al., 2013). Such applications require both 3D geometry and complex semantic information in different scales and domains (Karan and Irizarry, 2015). Building Information Model (BIM) and Geographical Information Systems (GIS) are broadly used as sources of such applications. However, integrating BIM and GIS poses great challenge since these two data formats are semantically inconsistent for any building systems and sub-systems analysis (Borrmann et al., 2014; Cheng et al., 2013; El-Mekawy et al., 2012; Irizarry et al., 2013; Isikdag et al., 2013; Li and He, 2008; Mignard and Nicolle, 2014). In fact, BIM is rich intelligent digital repository of building information and uses an Object Oriented (OO) approach to describe the characteristics (semantics and geometry) and behaviour of each building element as well as its relationships with others (Eastman et al., 2010). Industry Foundation Classes (IFC) is the open standard format for BIM by establishing interoperability in the construction industry (El-Mekawy et al., 2012). On the other hand, GIS is a platform for managing and presenting spatially referenced information. CityGML is established by Open Geospatial Consortium (OGC) standard data model for exchange of geospatial data and the interoperability between 3D GIS systems (Yao et al., 2015). Indeed, BIM and GIS share the foundation concepts that bring these two domains become closer together. However, the dissimilarities such as spatial scale, level of granularity, geometry representation methods, storage and access methods, and semantics mismatches make the integration of BIM and GIS complex and difficult to achieve (Borrmann et al., 2014; Cheng et al., 2013; El-Mekawy et al., 2012; Irizarry et al., 2013; Isikdag et al., 2013; Li and He, 2008; Mignard and Nicolle, 2014). Many researchers from both GIS and BIM communities have been working on developing conceptual geometric models, mainly designed for visualization purposes (Borrmann et al., 2014; Cheng et al., 2013; El-Mekawy et al., 2012; Irizarry et al., 2013; Isikdag et al., 2013; Li and He, 2008; Mignard and Nicolle, 2014). However semantic models are more needed for different engineering and planning applications to enable complex queries and analysis for facilitating sophisticated decision tasks (Stadler and Kolbe, 2007).

BIM-GIS integration and data sharing can be composed into syntactic and semantic aspects of data integration. The syntactic aspect of the data integration is about access to the shared data in the other domains; while the semantic aspect involves incorporating the information from the domains into a new data structure of a consuming system. The syntactic approach focuses more on data conversion and developing translation algorithms from one format to the other resulting of many BIM-GIS packages and software extensions. This mainly allows access and use of geospatial data within BIM and BIM semantics data within GIS context (Alam et al., 2014; Cheng et al., 2013; El Meouche et al., 2013; El-Mekawy, 2010; El-Mekawy et al., 2012; Li and He, 2008). In the other hand, the
semantic integration is based on the ability to attach meaning to conventional concepts for structuring the domain knowledge about phenomenon or objects (El-Mekawy, 2010; Karan et al., 2015; Mignard et al., 2011; Stadler and Kolbe, 2007). The RDF graph as a data model used by semantic web technologies brings many in developing integrated data models that can have a huge impact in GIS, BIM integration mainly because:

- RDF and RDFs (RDF schema) are standards offering richer class and property modellling and inference compared to RDBMS, NoSQL counterparts
- Flexibility and ability to make ad-hoc RDF statements about any resource of BIM or GIS without the need to update global schemas. SPARQL queries can contain optional matching clauses that work well with sparse data representations.
- Shared ontologies from both BIM and GIS will facilitate merging data from different sources
- Graph theory is well understood and a lot of data related problems are better solved using graph data structures.

And, given that semantic web technologies are internet based and supports HTTP protocols and honour the Service Oriented Application (SOA) architectures the RDF graph-based integration is a web-wide scalable. Therefore such integration can be used as-is or merged with property data to increase the value of the in-house data stores.

This paper introduces a new semantic approach to coherently integrate BIM-GIS model, called Integrated Geospatial Information Model (IGIM). The proposed integration approach is based on the semantic web technology, where the integrated model is achieved by merging Resource Description Framework (RDF) graphs generated from respective BIM and GIS. In this paper, we employ RDF as platform to develop integrated information model, IGIM synergistically provides significant benefits providing holistic information of both indoor and outdoor spatial information. IGIM system architecture consists of three modules: 1) BIM-RDF and GIS-RDF graphs construction, 2) Integrated RDF graphs, and 3) Query of information through IGIM-RDF graphs using SPARQL endpoint. Proposed IGIM enables queries generation from both BIM-RDF and GIS-RDF graphs and resulting a semantically integrated model with entities representing BIM classes and GIS feature objects with target-client application design.

This paper is organized as follows: Section 2 presents the semantic web technologies and RDF graphs and their ability to facilitate geometrical and geospatial information selection and integration. Section 3 presents the processes of constructing the IGIM using ontologies and schema and mapping though RDF graphs. Section 4 presents the process of extracting and visualizing BIM elements and GIS classes as graphs to complete the integration for IGIM development. Section 5 discusses a case study for information retrieval from GIS and BIM for a potential evacuation planning application. Finally, Section 6 highlights the concluding remarks and ongoing research perspective.

2. SEMANTIC WEB, BIM AND GIS

The proposed approach in this paper is based on three crucial concepts: 1) Semantic web ontologies represented by RDF graphs, 2) BIM and 3) GIS. Different data formats, data structures and schemas respectively IFC and CityGML represent BIM and GIS. Semantic web ontologies is as nutshell between BIM and GIS. A brief description of each of these key concepts is provided hereafter.

2.1 Semantic Web Ontologies and RDF Graphs

The semantic web is a set of technologies used for the representation, publication and browsing of structural data on the Web. As shown in Table 1, the semantic web consists of four elements: 1) URIs (Uniform Resource Identifiers) for object identifiers, 2) RDF for the representation of data in a graph form, 3) Ontology Web Language(OWL) for representing conceptual schema, and 4) SPARQL, a SQL-type language for graph queries using RDF data. The core concept of semantic web is the ontology that defines the concepts and relationships used to describe an area of knowledge (a subject matter) and provide a vocabulary and shared language. Ontology is defined by a constructor for representing concepts or classes, the relationships and properties of object and data, their governing rules, axioms and constraints as well as the instances of individual data.

The main elements of RDF graph are the triple of subject, predicate and object (Figure 1). The RDF triple can be visualized as a directed labelled graph, consisting of subject, predicate, and Object (Figure 1). Also, predicate and object may be referred to as property and value, respectively (Figure 2). In addition, a triple can also be interpreted as subject and object which are nodes connecting with predicate edge of an RDF graph (Figure 3).

In general, RDF triples represent a relationship (denoted by the predicate) between subject and object. Triples are sometimes referred to as statements or as assertions of relationships. A set of triples with common subjects or objects can be merged into one connected components. The integration can then be achieved between RDF graphs by using mapping entities based on semantic, geometries, and relationships as presented in Figure 4.

RDFs and OWL are the language to define semantics or mathematical basis for the meaning of each constructor. Semantics are defined based on inferences of each statement.
entails. Since the concepts in RDFs and OWL ontologies are expressed formally, computer programs can process them.

2.2 BIM and IFC

Industry foundation classes (IFC) is a data model for building and construction industry data (El-Mekawy et al., 2012). IFC platform neutral and it is an object-based file format with a data model developed by BuildingSMART (BuildingSMART, 2014) to facilitate interoperability in architecture, engineering and construction (AEC) industry. IFC is defined by EXPRESS-based entity-relationship model consisting of several hundred entities organized into an object-based inheritance hierarchy. Examples of entities include building elements such as IfcWall, geometry such as IfcExtrudedAreaSolid, and basic constructs such as IfcCartesianPoint. To represent and exchange IFC data, the following technologies are used: GUID (Global Unique Identifiers), STEP (ISO 10303 STEP (Standard for the Exchange of Product Data) for exchange format), EXPRESS (Object-Oriented data definition language) as presented in Table 1. The main challenge arise here is that the IFC EXPRESS schema cannot be directly translated to OWL, such global versus local names of relations, and mapping of datatypes. However, the integration is based on OWL due to its expressionless, flexibility, portability, and compatibility with semantic web and linked data platform.

2.3 GIS and CityGML

A detailed geographic information system datasets and their schema are represented in GML as well as in CityGML with given a level of details (LoD) in many other cases, a common information model for the representation, exchange and storage of 3D city models and urban objects, CityGML defines classes and relations for the most relevant topographic objects with respect to their geometrical, topological, semantical, and appearance properties (Kolbe and Nagel, 2012). CityGML provides the best framework for semantic-geometric relations of 3D objects above the earth surface and it is characterised by ISO 19100 using ObjectID, GML3 (Geography Mark-up Language) for exchange format, and UML (Unified Modelling Language) for a standard notation for the modelling of real-world objects in order to develop oriented object designs as presented in Table 1.

CityGML is structured as simple, single-scale models without topology and few semantics, or very complex multi-scale models with full topology and fine-grained semantic differences. CityGML enables lossless information exchange between different GIS applications and systems, and given that CityGML is implemented as a GML application schema as family of XML (eXtensible Markup Language). Thus, extracting and parsing information from CityGML into RDF-XML or OWL is likely straightforward process. Table 1 summarises the characteristics of IFC, CityGML and Semantic Web.

<table>
<thead>
<tr>
<th>IFC</th>
<th>CityGML</th>
<th>Semantic Web</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema</td>
<td>EXPRESS</td>
<td>OWL</td>
</tr>
<tr>
<td>Data</td>
<td>STEP</td>
<td>RDF</td>
</tr>
<tr>
<td>Identifier</td>
<td>GUID</td>
<td>ObjectID</td>
</tr>
<tr>
<td>Query</td>
<td>SQL</td>
<td>SPARQL</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of IFC, CityGML and Semantic Web

3. INTEGRATED GEOSPATIAL INFORMATION MODEL: ONTOLOGY AND SCHEMA DEFINITION

A novel integration method of BIM and GIS is proposed in this paper based on semantic web technologies and RDF graphs. Thus, a RDF-OWL data model is developed for IGIM to represent the relationships or predicate (i.e. relationship between object and subject). To do so, a series of process that transform building information IFC and CityGML objects from traditional format to RDF graphs are required. IFC is written in EXPRESS schema, and thus, a transformation from IFC ontology (O_BIM) into RDF-IFC instances is necessary. Also, CityGML data is stored in a database that needs to transform into GIS-RDF with GIS ontologies (O_GIS). As summary, the methodology consists of:

1. **Constructing O_BIM**: constructing IFC compliant ontology that describes the hierarchy structure of BIM objects, their relationships and properties. O_BIM is based on semantic indexing and retrieval of building information from IFC model.
2. **Constructing O_GIS**: constructing GIS ontology of geographic surrounding area of the building.
3. **Ontology Mapping**: linking similar concepts and relationships between O_BIM and O_GIS, a resulting ontology will have new RDF graph of concepts and relationships ontologically related and defined by (O_BIM, O_GIS).
4. **Querying O_GIS-BIM** based on application domain.
5. **Data loading onto O_GIS-BIM** and resulting the IGIM model.

The new enriched ontology O_GIS-BIM contains all the classes, and properties from both BIM and GIS domains as shown in Figure 5. The BIM and GIS datasets are then translated into RDF standard format and loaded into the Ontology. The semantic web application can then have access to process these datasets.

![Figure 5. Merging O_BIM and O_GIS](image)

The mapping and ontology matching is a process that recognises the semantic correspondences between O_BIM and O_GIS triples as shown in Figure 5. The relationships between objects are semantically defined either by matching, equivalency or by establishing a relation with objects from other ontologies. Then, the new integrated semantic model O_GIS-BIM is achieved where triple from both ontologies are unified and merged onto one unique semantic model with unique format and schema. The information extracted from BIM and GIS sources then transformed into a semantic web format. In Figure 6, we illustrate the mapping and selection of triples process from GIS and BIM ontologies and it is defined either as:

- Equivalent: Object A (from GIS) is Equivalent to Object B (from BIM)
4. INTEGRATED GEOSPATIAL INFOMATION MODEL DEVELOPMENT

IGIM system architecture and data translation flow are presented in Figure 7. The IGIM system consists of three tiers: 1) Input data; 2) Processing and validation; 3) Configuration and application. The input tier contains BIM data represented as IFC EXPRESS and GIS data represented as CityGML formats. Processing and validation tier refers to the process for IFC and CityGML translation and validation into ontologies and RDF OWL format. Configuration and application tier includes mapping RDF-OWL both concepts and feeding at IGIM. Apache Jena API is used for IGIM system presentation. Following sections are a detailed description of each tier and procedure.

An application is defined to represent the semantic concepts from both GIS and BIM domains. This leads to a logical representation of an object-oriented model of concepts specifying relevant properties to each domain such as relationships, instances and values. In the core part, the model is based on two types of primitive and defined concepts (Karan et al., 2015). The primitive concept is used to represent the natural classes of the entity domain; where the necessary conditions are specified. The primitive concepts are represented by their definition and correspondences on the top of hierarchical structure of the ontology representation. The defined concepts represent subclasses of the primitive one; where this consider as ENTITY constructor. The ENTITY is organised into taxonomy via Super-type/Sub-type partial ordering relation. Let’s see how these concepts can be identified for IFC and CityGML.

4.1 Conceptualising IFC: As Example, IFCWindow

ENTITY is presented in original EXPRESS format (see Figure 8).

Figure 8. EXPRESS Class representation of IFCWindow.

ENTITY IfcWindow
SUBTYPE OF (IfcObject);
OverallHeight, OverallWidth, PredefinedType, PartitioningType, and UserDefinedPartitioningType; while there is rule specifying certain condition.

Figure 9. Subtype/Superotype relationship between IfcWindow and IfcWindowStandardCase

Let’s consider an individual \(W_n\), an instance of the primitive concept IfcWindow, then, \(W_n\) have automatically the properties of IfcWindow such as OverallHeight, OverallWidth, PredefinedType, PartitioningType. The profile of a standard window which is inserted into an opening, is represented by a rectangle through 2D plane of the opening, defined by IfcWindowStandardCase. This IFC Entity is considered as a defined concepts, therefore, the associated properties of the IfcWindowsStandardCase are necessary and sufficient as shown in Figure 10.

Conceptualizing CityGML: Given that GIS models are based on relational models using tables (relations), features classes, feature datasets, and raster imagery footprint tables. These tables are formed by a number of fields organised in tuple (rows) with same attributes. Each table is considered as an OWL Class as demonstrated in Figure 11.top. All associated attributes of GIS objects are used in establishing relationships between different instances and constraints (e.g. primary key). Code and domain values are used for transforming GIS objects to RDF graph such water pipes diameters range values. A constraint in relation GIS model are expressed by rules that is
called axiom in RDF graph is presented in Figure 7. bottom. GIS Extraction Script is used to extract GIS data from a CityGML file to a RDF graph. The process is followed by a deep look at IFC schema and classes based on EXPRESS (see Figure 4), attributes, data types, relationships, and their counterpart elements and objects from CityGML.

**4.2 Mapping of Concepts**

Once the RDF graphs of both BIM and GIS data sources are produced, next step is bridging gap between these two data sources, called mapping of concepts. To do so, the meaning, structural and syntactical differences of both data types are investigated, the two new BIM-RDF and GIS-RDF graphs are studied and the Graph Matching for Ontologies (GMO) bipartite algorithm is used to recognise the correspondences by measuring similarities of graphs. GMO compares the structure of entities of interest (which has a hierarchical tree structure), in order to quantify the degree of similarity of RDF triples between BIM and GIS triple stores. The Python RDFLib libraries was used to implement the GMO mapping algorithm for the IGIM is presented in Figure 12.

**4.3 Conversion and Integration into RDF Graphs**

This step consists of translating elements from GIS and BIM models into formal standard ontology language RDF/XML-OWL. To do so, a JAVA code has been developed using Apache Jena libraries for translating IFC to RDF (https://jena.apache.org/download/index.cgi). Also, Geotools API are used for CityGML to RDF conversion (http://docs.geotools.org/latest/javadocs/).

**4.4 Model Querying using SPARQL**

SPARQL is used for RDF graphs and semantic web query interogation. SPARQL is the standardized query language for semantic web to manipulate and retrieve data stored in the data models. The result of query can be represented as RDF, CSV, HTML or XML. The result will consists of a set of combining triples from both BIM and GIS ontologies provided in any of format that the semantic web supports Turtle, RDF/XML, N-TRIPLE and JSON-LD RDF formats.

The linkage between BIM-RDF and GIS-RDF is achieved through SPARQL endpoint defined by a set of data and entities with complementary properties, relationships and geometries unified in a fully integrated BIM-GIS RDF graph. In case of an unfamiliar vocabulary from either BIM or GIS sources, the application will resolve the Unified Resource Identification (URI) that identifies terms in order to the RDFs or the OWL definition. This provides a unified integrated view where applications have access to datasets as one integrated knowledgebase and one semantic schema. For further exploration of information, it is essential to translate the semantic web query results into XML format.

**5. CASE STUDY**

To validate the proposed solution in this paper, a case study is done using datasets of the IFC model (extracted from BIM Revit model) of Bergeron building Centre for engineering excellence (Figure 13), and the 3D GIS model (KML) of Petrie Science and Engineering Building (Figure 14) and also campus wall road at York University Campus, presenting a complete mixed version of different sources of datasets. The developed approached is applied on mentioned datasets as presented in Figure 15.

---

**Figure 10. Inter-relationships between IFC entities.**

**Figure 11. RDF graph of CityGML (top), rdf:Class and subclass OwI:Class, RDF schema rules (bottom left & right).**

**Figure 12. GMO Mapping algorithm for BIM-GIS Ontologies integration.**
An evacuation scenario is set up as a fire exposed through the ramp up to the room, floor and whole building. IGIM allows the retrieval of appropriate information about internal structure of building from BIM and link to the outdoor information from GIS such as roads, pedestrian path, etc. As study site, classrooms, corridors, exit doors, and stairwells are stored as IFC objects. The outdoor objects such as road pedestrian paths, vegetation area, etc. are extracted as CityGML objects. The components of IFCSpace and IFCStairs are retrieved from BIM data and the rest are extracted from GIS data, calculating the evacuee’s path to evacuee’s building (see Figure 16). This case study showed the proof of concepts, presented as solution in this paper. A wide-range of application can become possible using the proposed IGIM system to resolve the complex data sharing and integration situations while they rely on the integration of multiple sources of data and multiple scales in real-time computation.

6. CONCLUSION

This paper presented a highest level of interoperability between BIM and GIS data by introducing a new way of formatting and processing data using semantic web technologies and RDF graphs. The novelty of the proposed solution refers to the benefits of both BIM and GIS technologies together into one integrated model, called, IGIM. Using the GMO mapping algorithm with well-established classification rules, the IGIM offers the flexibility to access and process datasets from GIS and BIM through RDF directed graph, while there is no need to establish direct links between various terminologies that are represented in objects and classes only by cruising between them. IFC and CityGML were translated into IFC-RDF and GIS-RDF graphs, while the integration of both graphs was done at semantic level. The proposed IGIM system was tested for data retrieval for an evacuation planning case scenario. This brought the advantage of linked data from building elements to 3D/2D geospatial data into the IGIM unified domain ontology.

This is on-going research project. Our near future steps will be enhancing the GMO algorithm to take into account more semantics data from both GIS and BIM domains and make it able to provide as much rich data as possible for IGIM. Also we will look deeper on ways to maintain and optimise RDF graph generation and mapping procedure and creating an evaluation metrics for the RDF generated graphs.
REFERENCES


