Interoperable data models for infrastructural artefacts – a novel IFC extension method using RDF vocabularies exemplified with quay wall structures for harbors

J. Beetz & W. Coebergh, R. Botter
Eindhoven University of Technology, Eindhoven, Netherlands

S. Zlatanova
Delft University of Technology, Delft, Netherlands

R. de Laat
Logic Labs, The Hague, Netherlands

ABSTRACT: Currently, only a limited number of dedicated data models for infrastructural artefacts exist. To cover information exchange and interoperability requirements, a number of international initiatives have been started under the umbrella of the buildingSMART organization to extend the predominant IFC model. In this paper, we are introducing a light-weight approach that allows the flexible extension of the IFC model with RDF vocabularies and ontologies. Using real-world quay wall models from the port of Rotterdam we show how information from multiple networked data sources can be seamlessly integrated with IFC models as the main carrier of geometric representation. We demonstrate how these semantically enriched models can be used with unmodified legacy software systems to facilitate a number of interoperability scenarios throughout different lifecycle phases.

1 INTRODUCTION

1.1 Motivation

The Port of Rotterdam is one of the world's most busy transshipment centers and has a key logistic role in the trade between main land Europe with the rest of the world. A wide range of resources, goods and products ranging from gas to freight containers has to be loaded, unloaded, distributed and transferred in high frequencies. Similarly to the vibrant world economic cycles, costumers of the harbor have ever-changing needs and requirements: Less coal but more gas. A temporary dip in bulk goods and a rise in high tech products. For the Port of Rotterdam this means: A new container terminal today, massive tanks for natural gas tomorrow. To accommodate these evolving needs, the harbor infrastructure has to react dynamically too. Roads, train tracks and gas pipes have to be added or rerouted. Quay walls and berthing equipment to accommodate different and larger vessels with varying goods must be provided to customers in time. The harbor is in a permanent state of change. At times, there are more than ten building and construction projects on different parts of the vast areal going on at the same time. In order to plan, execute and manage these projects efficiently, a wide range of stakeholders have to collaborate within the boundaries of tight schedules. Effective and timely exchange of information between these various actors is a crucial factor. Information needs include large scale geographic conditions such as soil qualities, water depth and

cadastral information. They also cover high resolution details of engineering structures like buildings, pipeline networks and quay walls including their sub-components, material properties and spatial configuration. These have to be coordinated with logistic processes, as well as financial and management data. With the rise of computer-supported information technology, dealing with these various kinds of information has changed dramatically over the last few decades: Large paper plots of two-dimensional maps and construction drawings were the main media that dominated building and planning until a few years ago. Today, highly specialized technologies like Geographic Information Systems (GIS) and Building Information Modeling (BIM) support planners and engineers alike. These technologies allow to capture various aspects of the current and future built environment in data structures that can also include three dimensional geometry models.

In close collaboration with the stakeholders from the Port of Rotterdam and the City of Rotterdam, the two year research project "3D Spatial Data Infrastructures – 3DSDI", financed by the Next Generation Infrastructures (NGI) research program was executed to tackle two selected use cases. The main aim of the project were the integration of heterogeneous information from different actors into coupled information models that include three dimensional geometry. The combination of geospatial and building related information (GIS and BIM) into a homogenous data environment allows logistic experts,

planners and engineers to get better insight into on-going and future activities. The Information models can be integrated via networks which helps to save resources by avoiding communication delays and overhead. In a number of software prototypes the researchers demonstrated how the combination of different approaches from the areas of knowledge engineering, semantic networks and web-based 3D graphics and visualization technologies can be tailored to the complex information needs of large organizations in efficient ways. The suggested solution approaches have been published for international scientific communities and have been suggested to national and international standardization organizations such as OGC and buildingSMART. The coordination and collaboration with other initiatives like the national "Concept Library of the Netherlands" (CB-NL) or the international effort to enhance the interoperability for the infrastructural sector (Open-Infra) will ensure that the results of these developments will be sustainable. Possible future activities will refine the suggested approaches in concrete projects.

1.2 Information requirements

In a first phase of the 3DSDI project a requirements analysis was executed resulting in an inventory of current processes, information exchange scenarios and the in IT infrastructure in use within the Port Authority of Rotterdam and other external stakeholders such as the Municipality of Rotterdam and engineering offices as contractors for infrastructural building projects. The main findings of this project consisted of the insight that for a number of common use cases data and information from up to 21 different departments has to be processed. In many of the planning scenarios, the integration of spatial data from different sources was identified as a major issue to improve the overall efficiency. For multiple concurrent construction projects within in the harbor area timeliness is crucial. Interoperability has been identified as a crucial bottleneck in the heterogeneous landscape of software systems ranging from GIS systems to Enterprise Resource Planning (ERP) that are in use in the organization.

To address some of these issues two use cases have been picked out to develop solution approaches for the integration and interoperability for spatial data. While the integration of spatial data concerning underground structures such as pipes and cables has been described elsewhere, the work reported in this paper addresses the frequent construction of quay walls.

The information requirements can be roughly categorized on four levels: a) Generic information about objects that are universally relevant (e.g. quay wall structures tend to be constructed in similar fashions in many cultural contexts); b) Regional information that is only relevant in a specific context (e.g. building regulations and best practices concerning quay walls in the Netherlands); c) organizational information (e.g. the way information is disambiguated, labeled and structured within the Port Authority of Rotterdam); d) project information (e.g. information that is specific to a certain construction process but might re-use templates from earlier works).
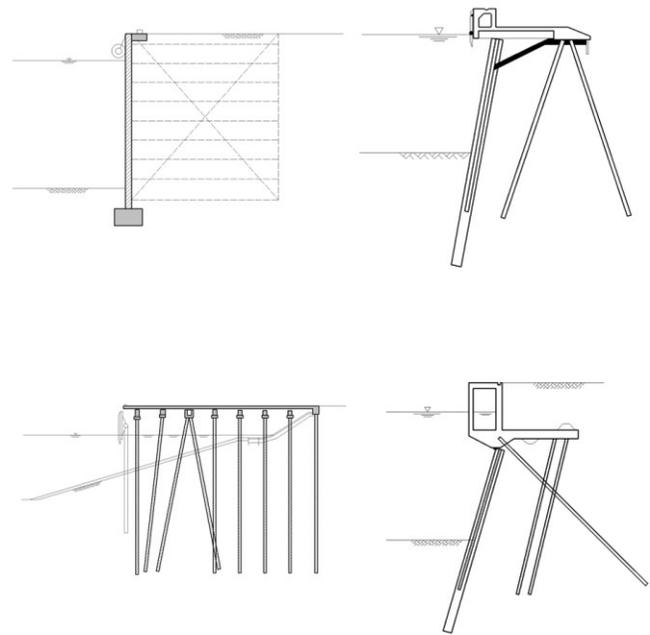

Figure 1 Examples of quay wall structure and super structure types covered by the IfcQuay domain model extension of the Industry Foundation Classes model

2 IFC MODELS FOR QUAY WALL STRUCTURES

In order to allow the efficient information exchange among the port authority management, the internal planning and engineering departments and external contracting engineering offices that during a the construction of new quay walls, interoperable data models are desirable that capture the specific semantics of these engineering structures. For the most widely accepted interoperability information exchange standard in building and construction, the Industry Foundation Classes (IFC) a number of extension initiatives have executed in the past that led to dedicated model extensions covering e.g. bridges and roads (Yabuki et al 2006, Lee & Kim 2011) However, no such extension is available to date that covers artefacts specific to quay wall structures. An inventory of such dedicated structures has been made together with the stakeholders, and using standard literature available in the field (de Gijt & Broeken 2013). Examples of a number of quay superstructures and quay wall types are illustrate in fig. 1. Specific objects include berthing equipment (bollards, fenders, rubbing posts, purlins etc.), jetties, retaining walls, quay wall superstructures etc.

2.1 IfcQuay model schema extension

A gap analysis with the existing IFC 4 model, led to the creation of 43 specific entity classes in a top-down approach that capture information about these objects. These entities include parametric definitions of the most common types of retain walls that are
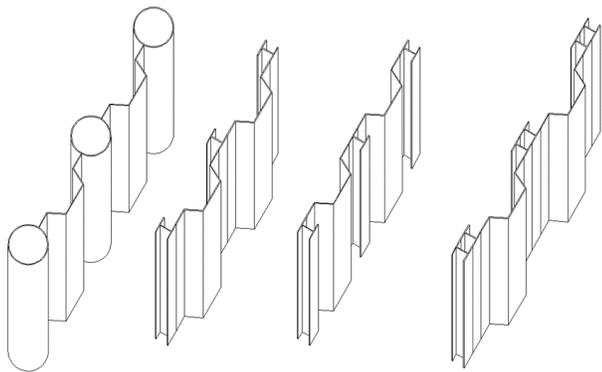

Figure 2 examples commonly employed retaining wall structures captured as parametric geometries in IfcQuay

used in quay walls. Similarly to work presented in (Ji et al 2013), such parametric objects can be used to capture such object in much more efficient ways than traditional static geometric representations. Examples of such re-occurring parametric objects are provided in figure 2. In order to further facilitate possible future implementations of this new schema into engineering software packages, a Model View Definition (MVD) has been created using the IfcDoc tool   (https://github.com/BuildingSMART/IfcDoc). The aim of the MVD was formally identify the minimal subset of the overall model that would be required to implement the model schema. Since at the time of the creation of the MVD no appropriate tooling was available, the IfcDoc tool was extended with functionality to import additional ISO 10303 part 11 schemas for such purposes.

This proposed IfcQuay model extension will be submitted to the buildingSMART organization as a potential candidate for standardization in the context of the OpenInfra initiative.

A considerable shortcoming of this traditional extension approach however is the high implementation effort of such model schemas. Similar domain-specific initiatives in the past have shown that this effort usually proves to be too high to be profitable for software vendors to be offered to a niche market. In a second phase of the 3DSI quay wall use case we thus devised and implemented an approach that could serve as a potential alternative. Furthermore, tailoring such models to regional or even organizational requirements is highly infeasible and require to resort to other solution approaches.

3 PARTIAL DOMAIN VOCABULARIES

Alternative to EXPRESS schema-based extension procedures described in section 2, a number of strategies can be identified to capture domain- and organization-specific information on the four indicated levels in interoperable ways. A common practice to date is to use properties that are associated to individual objects in the IFC instance model in an ad-hoc fashion. This practice however has the shortcoming that their semantics are limited to human-interpretable forms such as string identifiers and no standardized mechanisms exist that allow the sharing and reuse of such property sets among the different and changing stakeholders in projects.

3.1 The buildingSMART Data Dictionary

The disambiguation, internationalization and standardization of terms and vocabularies for the enrichment of IFC models can be organized in much more structured fashion using the ISO 12006 framework. While part 2 and 3 of this standard provide the conceptual approaches and data model respectively, its practical use is driven by a reference implementation provided by the buildingSMART standardization organization. The buildingSMART Data Dictionary (bSDD) (Bell & Bjørkhaug 2006) is a structured vocabulary that currently consists of some 80,000 concepts with translations in multiple languages that are accessible with a RESTful API. Using the GUIDs of the concepts and associated properties, references to common terms and definitions can be used to lend meaning objects in instance models that are not covered by the 'static' IFC model itself. Implementation agreements of these references are provided both in the IFC specification as well as in the Norwegian technical specification TS/NS 3489. The basic approach is to attach a chain of 'IfcPropertDepencyRelationship', 'IfcPropertyReferenceValue' and 'IfcClassificationReference' instances to an 'IfcPropertySingleValue' which in turn is associated to an 'IfcObject' with an objectified relationship. While this approach provides a uniform way to use the GUIDs of concepts itself, the implementation of retrieving the concepts from a repository like the bSDD is left to the respective vocabulary provider. In real-life situations this means that each repository defines its own service interfaces that might require different operations like authorization, retrieval and references depending on the API. In practice this means, that currently each additional concept repository resource means additional implementation effort that on the client software side that provides interfaces to engineering end-users. The notion to counter such fragmentation by gathering all possible definitions in a central globally valid repository as currently suggested by the bSDD approaches also has a number of severe

practical shortcomings, which are further pointed out in (Beetz 2014). In particular, the gathering e.g. the different levels of information granularity that are partially organization-specific would prove problematic at best. For the example at hand, the unambiguous use of concepts in the context of the port of Rotterdam currently involves an organization-internal classification system 'objecten boom' (object tree) that is used to identify objects across the boundaries of the many software systems within the organization but should also be used by external parties such as the engineering offices designing new quay wall projects. The current conceptual approach and business model of the bSDD requires the organization to acquire an account on the repository in order to be able to store and of use this vocabulary in a controlled context. Each individual party outside the organization also need credentials to retrieve and use the company-specific information structures. To address this and other issues we employed the conceptual approaches, mechanisms and technologies from the Semantic Web and Linked Data initiatives and argues that these can be used as viable future alternatives.

3.2 Semantic Web and Linked Data

The 'Anyone can say Anything about Anything' (AAA) principle is of the fundamental notions of the Semantic Web and Linked Data efforts. The statement not only documents the idea of a democratic, non-centralized and independent approach to defining meaning, opinions and definitions of a particular Universe of Discourse (UoD). "Can say" also means that the requirements of the underlying technologies should allow anyone to share and expose their views with as a low a threshold as possible that can be used and interpreted with as little effort as possible. A minimal set of agreements that have been formulated as the 'five star' requirements for data (http://5stardata.info/) allow sharing and consuming such structured data in more interoperable ways than provided by the so-called 'information silos' – proprietary databases with thick layers of non-standard interfaces – that are currently predominant on the web and in service base architectures. At its struc-

tural layers of the Semantic Web technology stack, the Resource Description Framework (RDF) allows the creation, reference, and extension of data sets across the boundaries of network nodes. A further essential enabler is the formulation of a universal exploitation-mechanism and query language that is intended to work independent of underlying data base schemas and implementation specifics. The Simple Protocol and RDF Query Language (SPARQL) provides a mechanism to expose such data sets and vocabularies in straight-forward and affordable ways as so-called 'SPARQL endpoints'.

In the context of the 3DSDI project three such endpoints have been created where each represents a different layer of the information cascade described in section 1.2.

3.2.1 bSDD-RDF

As part of the work that was partially funded by the Dutch CB-NL initiative and the FP7 DURAARK project (www.duraark.eu), the bSDD vocabulary has been transformed into a configurable RDF dataset. On the meta-model level a number of different modeling approaches ranging from the Ontology Web Language (OWL) to RDF Schema and the Simple Knowledge Organization Structure (SKOS) have been implemented to evaluate the advantages and disadvantages of the respective modeling strategies that will be discussed in-depth outside the scope of this paper.

3.2.2 QuayWall vocabulary

From the initial IFC schema extension for quay wall structures light version represented in RDF was derived that partially draws on and extends concepts from the bSDD. While not harnessing the full potential e.g. of reusing e.g. existing RDF representations of the IFC model (Pauwels et al 2011), experiments with referencing different engineering vocabularies such as the QUDT ontology framework (http://qudt.org/) for the re-usable definition of quantities, units measures and values have been conducted during the transformation of the original EXPRESS data set.

3.2.3 Organization vocabulary for administrative data

While bSDD-RDF and the domain QuayWall domain vocabulary could be reused in contexts outside the harbor, the data management vocabularies used in the HbR organization are of limited use outside the context. While there are need to share these e.g. with external contractors, it would be superfluous or even impermissible to share them in a global repository such as the bSDD database.

A project-specific and context vocabulary as introduced in section 2 has not yet been used in the scope
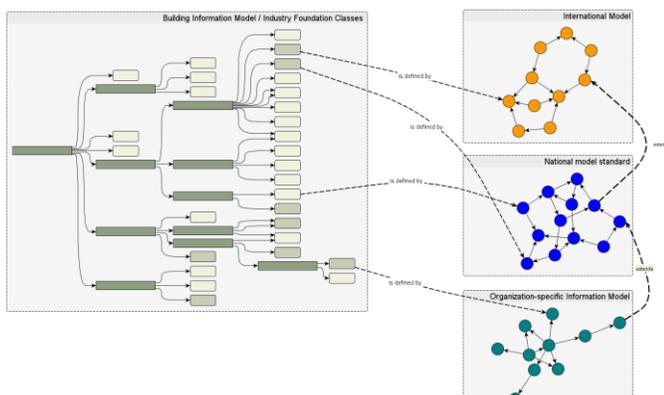

Figure 3 using external, linked vocabularies and data sets (right) to legacy IFC instance models (left)

of the 3DSDI project but the same mechanism could be applied.

4 ENRICHMENT OF PART 21 SPF INSTANCES WITH RDF

To allow the use of RDF vocabularies that might be exposed via SPARQL endpoints from within engineering models captured in IFC, a number of approaches are possible. While earlier work suggested to transform the core model or partial sub-model chunks itself into RDF, such approaches demand a considerable shift in technologies and would require an even higher implementation effort than the above mentioned schema extensions. While we think that such strategies are valid and useful approaches that the shift in technologies might take many more years and will come rather gradual than at once. Moreover, despite the promising developments with parametric geometries (Böhms et al 2008) for the IFC model, the largest part in IFC instance files are static geometric representation that cannot be stored as efficient in RDF as in the ISO 10303 part 21 Step Physical File (SPF) format.

In order harness the flexibility of the Semantic Web stack while at the same time preserving the efficiency of the SPF format and profiting from its many implementations in legacy applications, we are proposing a transitional approach to mix both technologies.
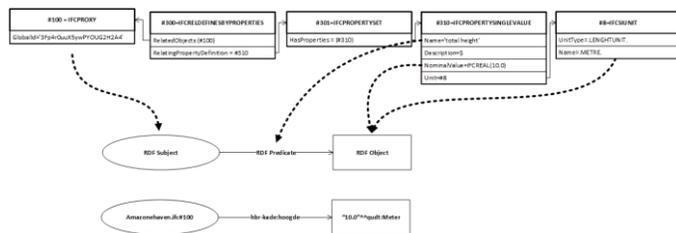

Figure 4 referencing arbitrary RDF triplets form legacy part 21 SPF IFC files using a three simple mappings

Based on the best practices proposed in TS/NS 3489 (see section 2.1) we are proposing three simple agreements to allow the use of RDF vocabularies in legacy part 21 files:
1 Every 'IfcObject' is regarded as a 'subject' of an RDF triplet <subject, predicate, object>
2 The 'Name' attribute of the 'IfcPropertySingleValue' entity is treated as an rdf:Property predicate in the triplet.
3 The 'NominalValue' attribute is treated as its object. This means that it can either be the reference to another resource with a URI or a literal value such as a xsd:string, xsd:float or complex entity e.g. stemming from the QUDT vocabulary

Using these three mappings, illustrated in figure 4 arbitrary RDF triplets can be embedded and referenced from within part 21 IFC SPF filed without re-

quiring the adaption of existing legacy software implementations. To these existing tools the mapping are treated as text-encoded literals, without interfering with the code. To confirm this we have tested minimal examples using a number of commercial off the shelf tools such as Solibri, Archicad and Revit. Figure 5 shows the two screenshots of the free DDS CAD viewer interface inspecting the properties of an IfcProxy that has been semantically enriched using the above suggested approach to belong to the class 'QuayWall'. The lower part of the illustrations presents a HTML page that is served by the SPARQL endpoint when the user clicks the URI of e.g. the "Name" attribute of the instance.
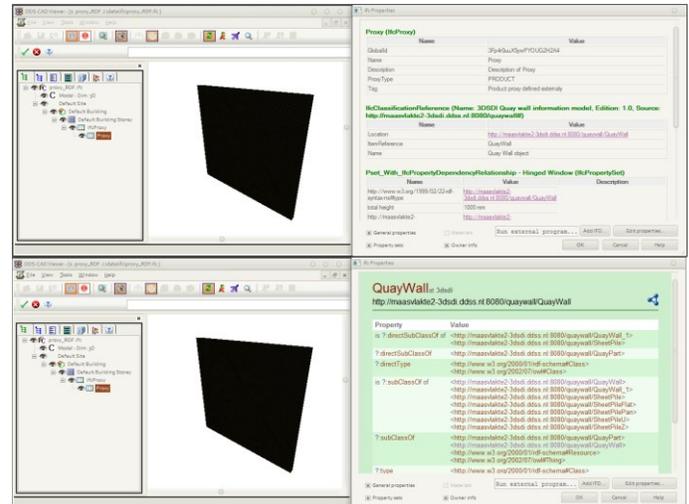

Figure 5 Unmodified, off the shelf legacy software application (DDS CAD) displaying properties of an IfcProxy object that has been enriched with RDF vocabularies

5 PROTOTYPE IMPLEMENTATION ON THE BIMSERVER.ORG PLATFORM

To evaluate and demonstrate the use of different vocabularies to enrich quay wall engineering models, we have implemented a prototype application that allows browsing and enriching IFC models using the mechanisms described in section 4:

The basis of the test scenario is an existing model of a quay wall for the new 'Amazonehaven' area of the port of Rotterdam. This is stored in a bimserver.org instance and its geometric representations are visualized as WebGL using the 'BIMSurfer' plugin (left hand part of the screenshot in figure 6). The meaning to all objects is assigned to the geometries using three different vocabularies (sections 3.2.1, 3.2.2 and 3.2.3 of this paper) that are exposed by three different SPARQL endpoint implementations (Jena Fuseki, OpenRDF Sesame and Virtuoso). When an object is selected in the browser, the properties attached to the objects via the 'IfcRelDefinesByProperties' relationship are listed on the left hand side of the web interface illustrated in figure 6. For each individual property that is defined using an RDF triplet from one of the vocabularies used, a live

SPARQL query is issued to display the associated rdfs:label either in Dutch or in English, depending on the preference toggle in the interface (figure 6, above).

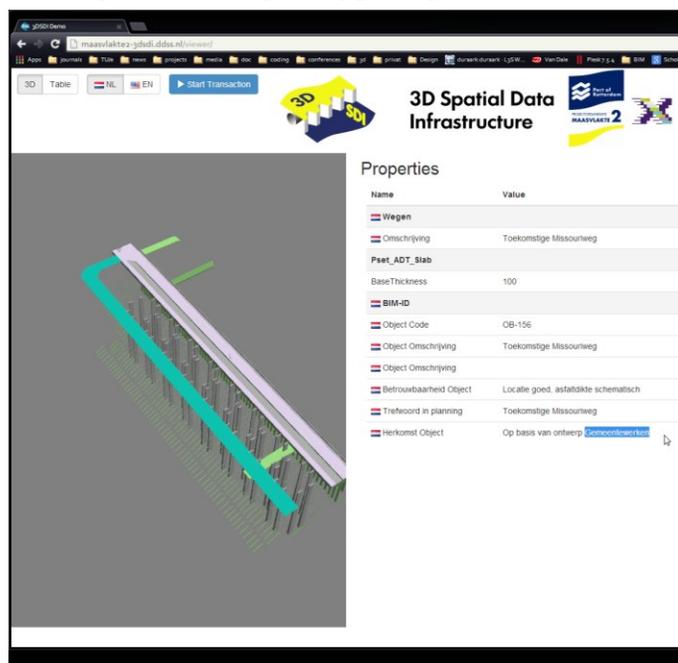In addition to browsing and looking up already existing values, the prototype implementation also

Figure 6 Screenshot of a web-browser based prototype inteface demonstrating the enrichment and use of RDF vocabularies with IFC legacy part 21 SPF format instances. The geometric model (left) is enriche with RDF tiriples (right)

allows to add new properties to a selected object. If a an object has been tagged with type information, by e.g. stating <entity x, rdf:type, 3dsdi:QuayWall> than all rdf:Properties that include the 'QuayWall' concept as part of their 'rdfs:range' properties are offered for selection to the interface. A simple string-based search furthermore allows the associations of new types to the objects (e.g. the Query "Anchor*" yields some 40 different concept types from the vocabularies used to further specify the ground anchors of the retaining walls)

A plugin for a major CAD vendor that allows similar browsing and enrichment from within the application is currently in its final stages of development and will be released as Open Source.

6 DISCUSSION AND OUTLOOK

In this paper, we have suggested a transitional approach to enrich IFC Step Physical Files with arbitrary RDF triplets such as domain-specific vocabularies, model extensions or organizational data. The proposed approach is based on a three minimal implementation agreements and does not break the backwards compatibility with existing implementation of the IFCs. At the same time, it offers to use the rich landscape of existing vocabularies and datasets available e.g. as part of the Linked Open Data

cloud without demanding overly complicated implementation efforts. Using a browser based client prototype we have shown how using different open and readily available, accepted and established standards, technologies and software implementations can be used to semantically enrich engineering models. We proposes this approach as one of the possibilities for a flexible yet interoperable assembly of semantically rich information for organization and domain-specific purposes.

While our approach has a number of shortcomings such as a lack of rigidness and needs further work to fully understand, develop and evaluate its potential (RDF blank nodes, nested objects, potential incoherencies with the data model etc.) we recommend the consideration of this idea especially to address some of the urgent needs in the context infrastructural modeling and integration with e.g. the GIS domain.

7 ACKNOWLEDGEMENTS

8 REFERENCES

Beetz, J., 2014. A scalable network of concept libraries using distributed graph databases. To appear in *Proceedings of the joint ICCCBE and CIB W78 conferences 2014, Florida, USA*

Bell, H. & Bjørkhaug, L., 2006. A buildingSMART ontology. In *Ework and Ebusiness in Architecture, Engineering and Construction: Proceedings of the 6th European Conference on Product and Process Modelling, 13-15 September 2006, Valencia, Spain*.

Böhms, H.M. et al., 2008. Semantic Product Modelling with SWOP's PMO. *EWork and EBusiness in Architecture, Engineering and Construction: ECPPM 2008*, p.95.

De Gijt, J.G. Broeken, M.L, *Quay Walls, Second Edition*, CRC Press, ISBN-13: 978-1138000230

Ji, Y. et al., 2013. Exchange of Parametric Bridge Models Using a Neutral Data Format. *Journal of Computing in Civil Engineering*, 27(6), pp.593–606.

Lee, S.-H. & Kim, B.-G., 2011. IFC Extension for Road Structures and Digital Modeling. *Procedia Engineering*, 14(0), pp.1037–1042.

Pauwels, P. et al., 2011. Three-dimensional information exchange over the semantic web for the domain of architecture, engineering, and construction. *AI EDAM*, 25(Special Issue 04), pp.317–332.

Yabuki, N. et al., 2006. International Collaboration for Developing the Bridge Product Model "IFC-Bridge." In *Joint International Conference on Computing and Decision Making in Civil and Building Engineering*. Montreal, Canada.