

LINKED DATA FOR CROSS-DOMAIN CONSTRAINT INFORMATION SHARING IN LNG CONSTRUCTION

Jun Wang¹, Wenchi Shou², Peng Wu³, and Xiangyu Wang⁴

Abstract: Reliable construction plans are vital for effective collaboration across a project's design, procurement and construction. Numerous constraints arising from engineering, supply chains and construction site are the main factors affecting planning reliability. Currently, there is not an efficient way to access all these constraint information because they are stored in various systems and managed by multiple domain participants. This paper aims to utilise Linked Data Technology to enable links to be set between data (i.e. constraint information) in different systems and therefore connect these systems into a single global data space. A prototype of the proposed approach was developed and tested on a sample LNG project.

Keywords: Linked data; Ontology; Constraint information sharing; Liquefied Natural Gas; Lean construction.

1 INTRODUCTION

Owners and contractors face enormous challenges to complete billion dollar engineering and construction projects in Liquefied Natural Gas (LNG) industry (Chi et al. 2015). Virtually every project in Western Australia has been over budget and late, for instance, the latest Wheatstone LNG suffers a six-month delay due to the slow schedule of off-site construction of the LNG plant modules in Malaysia. Reliable construction plans and work flows are vital for effective collaboration across design, procurement and construction so as to reduce schedule delay and cost overrun (Wang et al 2016a).

The "Pull-driven" method has been proved as an efficient approach to produce finished products as optimally as possible in terms of quality, time, and cost (Tommelein 1998). Advanced Work Packaging (AWP) developed by a joint venture between the Construction Industry Institute and the Construction Owners Association of Alberta, is one of the "Pull-driven" methods which has been widely used in the LNG industry (Hamdi 2013). Within AWP, the pull concept is embodied in the process of work package releasing. A Work package cannot be released to a construction site until all of the related constraints are removed (CII 2013). Constraint concept in this paper is defined as anything which prevents work packages successfully executed in the field, such as engineering drawings, permits, materials, work crews, tool, equipment, etc. According to Wang, et al (2016b), these constraints are further classified into three main categories: engineering constraints (i.e. design calculation, general arrangement, isometric drawing, process & instrumentation diagram, specification, bill of material, and three-dimensional model),

¹ Ph.D. student, Australasian Joint Research Center for BIM, Curtin University, Perth, Australia, jun.wang15@postgrad.curtin.edu.au

² Ph.D. student, Australasian Joint Research Center for BIM, Curtin University., Perth, Australia, wenchi.shou@postgrad.curtin.edu.au

³ Senior Lecture, Australasian Joint Research Center for BIM, Curtin University., Perth, Australia, peng.wu@curtin.edu

⁴ Professor, Australasian Joint Research Center for BIM, Curtin University, Perth, Australia, xiangyu.wang@curtin.edu.au

supply chain constraints (i.e. bulk material, instrument, and off-site LNG modules), and site constraints (i.e. labour, permit, equipment, temporary structure, tool, weather, workspace, and preceding work).

This paper aims to develop a semantic approach for cross-domain constraint information sharing by using linked data technology. Compared with conventional approaches of data integration, linked data principles enable data to be delivered in both machine- and human-readable formats (Lee et al. 2016). Making constraint data on the Web enables greater transparency and accountability, and makes project participants get required information more efficiently

2 A LINKED DATA-ENABLED APPROACH FOR CROSS-DOMAIN CONSTRAINT INFORMATION SHARING

Figure 1 illustrates the architecture of the linked data-enabled approach which includes three layers: (1) Data wrapper and publication; (2) Linked constraint cloud data; and (3) Data access and inference.

In the first layer, the linked data wrappers perform the “RDFisation” process, which transforms existing legacy data into linked data according to the relevant vocabularies or ontologies (Heath and Bizer 2011). Four types of constraint data need to be converted to RDF format which include: Engineering constraint data, Supply chain constraint data, Site constraint data, and Project planning data. The process of data wrapper and publication consists of two main steps, namely, vocabulary & ontology selection and development, and RDF data transformation. The purpose of the first step is to provide domain-specific terms for describing resources in the world and how they relate to each other. The second step is to utilise the developed terms to convert existing multiple data formats into RDF. Detailed explanations of the vocabularies & ontologies selection and development, and RDF data transformation is presented in Sections 3 and 4 respectively.

In the second layer, the separate RDF graphs of data that result from the first layer are interconnected together to create a global data cloud. The interconnection network is created by adding external RDF links between data entities across isolated data graphs (Bizer et al. 2008). In Figure 1, the links within a domain data set are represented by the solid lines, and the external links among multiple data sets are represented by the dashed lines. With these links, a global, interconnected data space for cross-domain constraint information sharing and management can be generated. The details of the data interconnection process are discussed in Section 5.

In the third layer, the linked constraint data cloud can be searched and queried through linked data search engines. The basic approaches to access the data cloud is to dereference HTTP URIs into RDF descriptions and to discover additional data sources by traversing RDF links (Bizer et al. 2009). In addition, parts of the graph may also be accessed via SPARQL endpoints or downloaded in the form of RDF data set dumps. Because of the caching function of the linked data search engines, instead of directly accessing the original linked constraint data cloud, applications can also access the data via the APIs provided by these search engines. The inference rules in this layer concentrate on defining a general mechanism on discovering and generating new relationships based on existing data cloud. Inference based techniques can also detect potential inconsistencies within the linked constraint data cloud. Section 6 presents more details about these two modules.

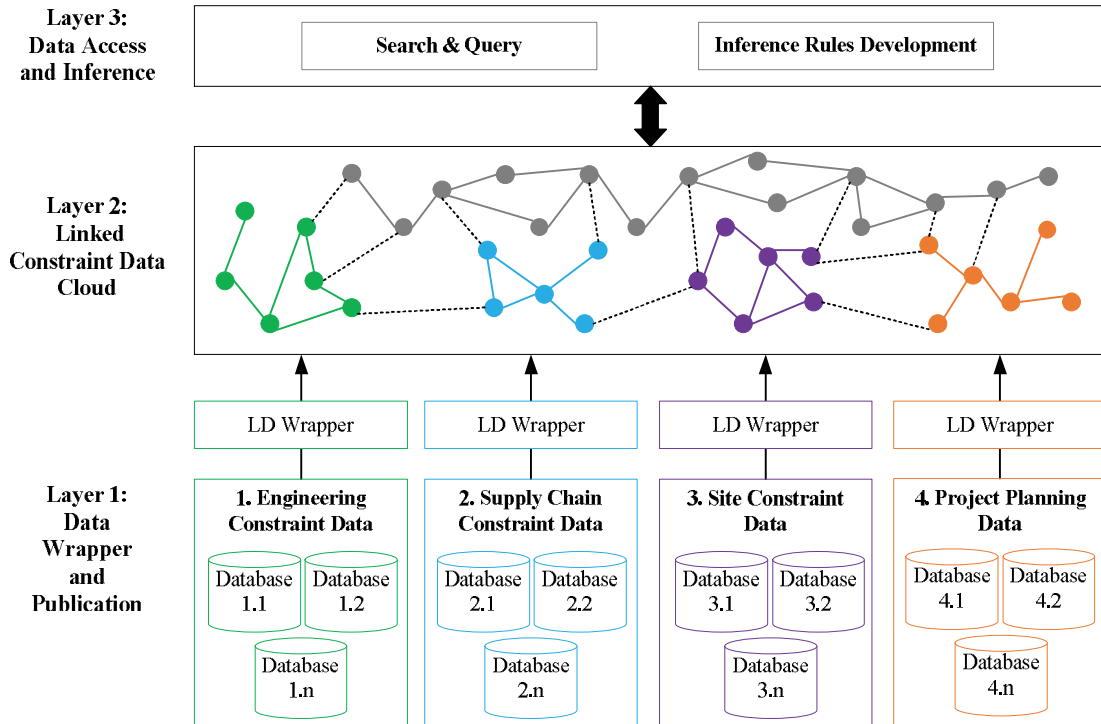


Figure 1: A Linked Data-enabled Approach for Cross-domain Constraint Information Sharing

3 VOCABULARY & ONTOLOGY SELECTION AND DEVELOPMENT

Reuse of existing terms is highly desirable as it maximises the probability that data can be consumed by applications that may be tuned to well-known vocabularies, without requiring further pre-processing of the data or modification of the application (Heath and Bizer 2011). Table 1 shows the reused vocabularies or ontologies and their descriptions.

In order to support the data wrappers of the constraint and project planning data, two new domain-specific terms were developed, namely, Constraint ontology (as shown in Figure 2) and AWP ontology (as shown in Figure 3). Protégé (an open-source platform) (Musen 2015) was utilised to construct these two domain ontologies. The ontology development method deployed in this study is derived from the method developed by Noy and McGuinness (2001).

Table 1: The Reused Vocabularies or Ontologies and Their Descriptions

Reused Vocabularies or Ontologies	Descriptions	Example Usage
Simple Knowledge Organization System (SKOS)	Used to express conceptual hierarchies (i.e. taxonomies)	"Constraint" skos:definition "A constraint is ..."
RDFS and OWL	Used to describe conceptual models in terms of classes and their properties.	"EngineeringConstraint" rdfs:subClassOf "Constraint" "hasConstraint" owl:inverseOf "isConstraintOf"
Friend-of-a-Friend (FOAF)	Used to describe persons, their activities and their relationships	"Designer" foaf:mbox "example@gmail.com"
Dublin Core Metadata Initiative (DCMI)	Used to general metadata attributes such as title, creator, date and subject.	"EngineeringWorkPackage" dc:creator "Designer"
Basic Geo (WGS84)	Used to describe geographically-located things	"Perth" geo:lat "-31.95"
ISO 15926 (Kim et al. 2011)	Used to describe LNG plant product such as pipeline, instrumentation and equipment.	"PUMP" rdl:hasIdPCA "RDS327239"

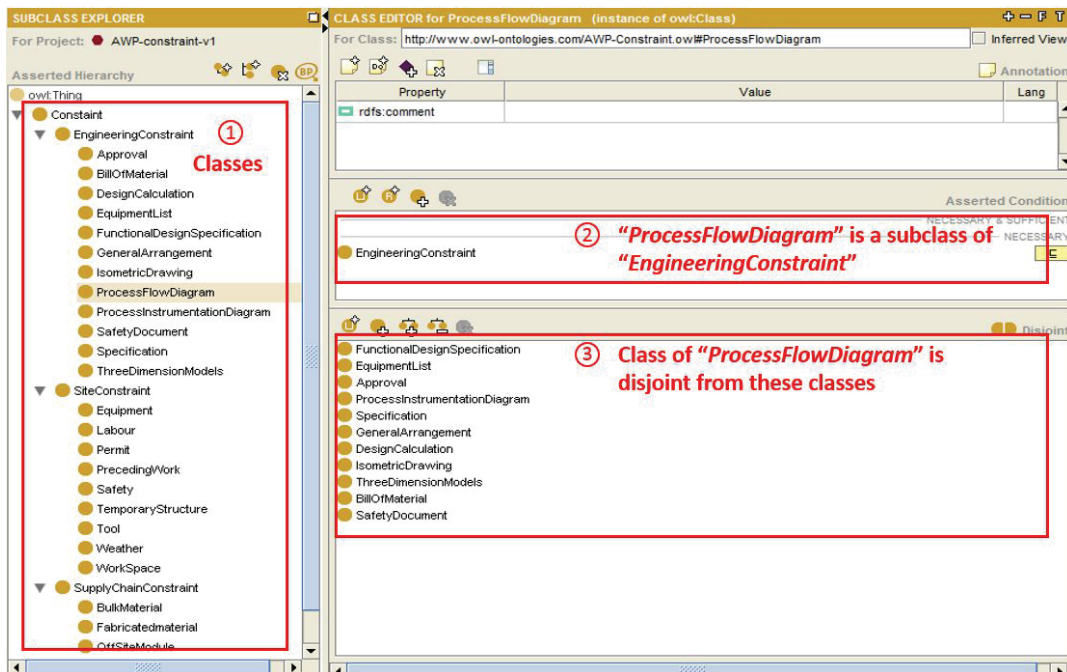


Figure 2: Constraint ontology

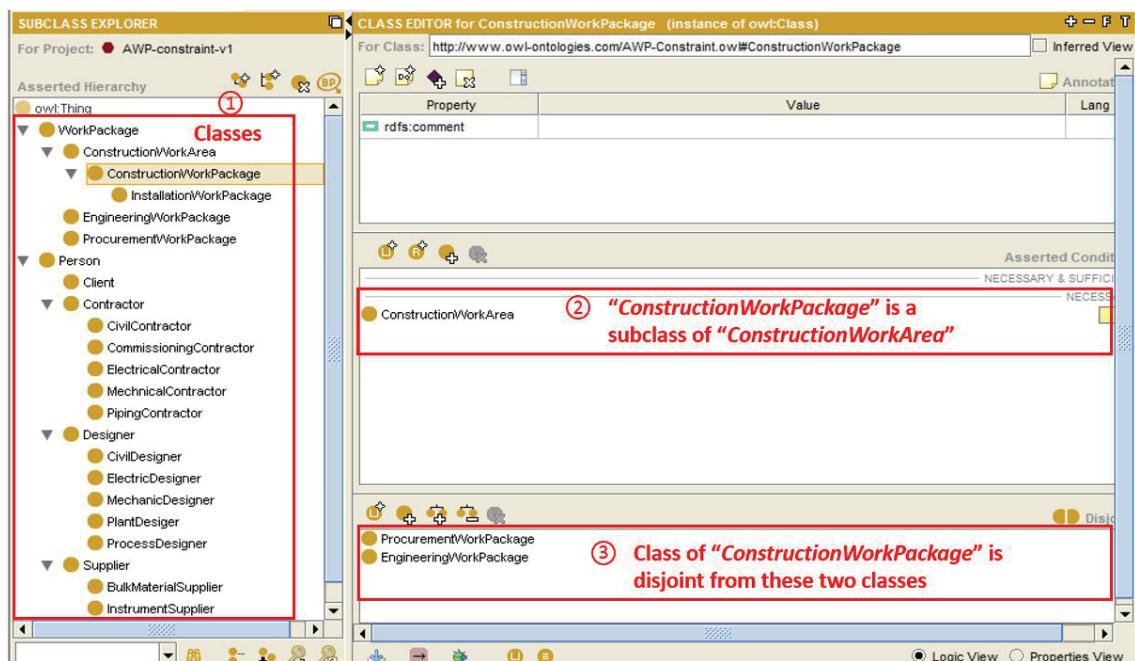


Figure 3: AWP ontology

Domain ontologies (i.e. ISO 15926 ontology, Constraint ontology, and AWP ontology) provide only data architectures and semantic formalisations for diverse data sources. The purpose of the merged ontology (as shown in Figure 5) is to interlink these isolated data islands into a unified data space so that the cross-domain constraint information can be shared. In this study, the three main developed ontologies were merged by matching synonymous concepts using the "equivalentClass" property. The mapping process could be semi-automated by performing ontology merging and alignment algorithms (Noy and Musen 2000). However, the efficiency and accuracy of the current mapping approaches are not high enough (Noy and Musen 2000). In order to assure the accuracy of the mappings, this study utilised a manual method to define semantic equivalence among entities from the proposed ontologies. The Constraint concepts in the constraint ontologies are linked to the Constraint in the AWP ontology through the "equivalentClass" property. In addition, sub-classes of "SupplyChainConstraint" (i.e. FabricatedMaterial) and "EngineeringConstraint" (i.e. IsometricDrawing) in the constraint ontology are also have corresponding equivalent classes in the AWP ontology. The "WorkPackage" concept is the bridge between ISO 15926 ontology and AWP ontology. The equivalent classes include "EngineeringWorkPackage", "ProcurementWorkPackage", "ConstructionWorkPackage", and "InstallationWorkPackage". Based on this merged ontology, any "WorkPackage" instances generated by engineers and Constraint instances generated by project planners can be legally linked to data graphs in the downstream phases.

4 RDF DATA TRANSFORMATION

In order to support the interlinkage of data from isolated sources, these data are required to be converted into the RDF format. There are two components in a data wrapper, namely, domain ontology and mapping rules. During the data transformation process, the domain ontology is used as the source of vocabularies, and mapping rules are utilised to assure the

converted RDF data model is expressed in a structure and target vocabulary. In this study, three types of data wrappers were developed to translate the existing legacy data (i.e. engineering constraint data, supply chain constraint data, site constraint data, and project planning data) to the RDF format: (1) Drawing data to RDF; (2) Table data to RDF; and (3) Document data to RDF.

5 LINKED CONSTRAINT DATA CLOUD

The result of the data transformation is a set of disparate RDF data sets. To fully support the decision making, these disparate resources are required to be interconnected to each other. The objective of linking constraint data in the cloud is to create a global, interconnected data space for cross-domain constraint information sharing and management. In this study, two types of data interconnections are developed, namely, data interconnection among RDF data sets and data interconnection between RDF triples and documents. The former one uses the merged ontology developed in Section 3 to set external RDF links among multiple RDF data sets. The latter one builds a virtual connection between RDF triple store and Enterprise Document Management System (EDMS) so that for each document in EDMS, there is a corresponding small RDF graph in Triple Store. Users can perform an SPARQL query based on the RDF Triple Store to get a document URI value and using this value to conduct a GET call from the EDMS resource to retrieve the raw document.

6 DATA ACCESS AND INFERENCE

Once these data are linked, query strategies and reasoning rules can be applied to extract specific information based on the objective of the cross-domain constraint information searching and sharing. The suite of the SPARQL1.1 specification was used in this study to query and update triples and graphs. There are four types of SPARQL queries: SELECT, CONSTRUCT, DESCRIBE, and ASK. An SPARQL SELECT query returns a solution, which is a set of bindings of variables and values. An SPARQL CONSTRUCT query returns triples as a sequence of triple values in an RDF graph. These triples are constructed by substituting variables in a set of triple templates to create new triples from existing triples. An SPARQL DESCRIBE query returns a sequence of triple values as an RDF graph that describes the resources found. An SPARQL ASK query returns a boolean (true or false) indicating whether a query pattern matches the dataset. Figure 4 illustrates a simple example when users try to find the description of the "IWP20001-Piping", and its related constraint information.

The inference is the process of discovering new facts from existing triples on a set of rules. In this study, a type of automatic inference method was applied which used rulesets and ontologies. New facts can be added to the RDF triple store through forward-chaining inference or be inferred at query time through backward chaining inference (Meditkos and Bassiliades. 2010). Rulesets utilised in this study include the standard rulesets, such as *rdfs*, *rdfs+* and *OWL-Horst*, and pre-defined rulesets. Users can also create their own rulesets by importing some of these rulesets and/or writing their own rules. Figure 5 shows an example of the rule *domain.rules* which states that if all the things in the second set of braces matches a triple ("p" has domain "o" - that is, for every triple that has the predicate "p", the object must be in the domain "o"), then construct the triple in the first set of braces (if you see "s p x", then "s" is a "o").

The screenshot shows a web interface titled "Cross-Domain Constraint Information Sharing". At the top left is a network graph. Below it is a search bar with the text "Input Constraint Name or Work Pack Id" and a "Search" button. The search results for "IWP20001-Piping" are displayed below. The URL is "http://data.example.com/awp/IWP20001-Piping". The results are organized into sections: "awp:hasCreator" (Jun), "awp:hasCreationDate" (13/09/2016), "awp:hasDescription" (A 'Installation Work Package' which aims to connect the Vessel-101 and Compressor 101), "awp:hasType" (Installation Work Package), and "awp:hasManhour" (1000). There are also sections for "Schedule", "EngineeringConstraint", "SupplyChainConstraint", and "SiteConstraint", each with associated constraint names and values.

Figure 4: An SPARQL search example

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema/>

...

rule "domain_type" CONSTRUCT {
  ?s a ?o
} {
  ?s ?p ?x .
  ?p rdfs:domain ?o
}
    
```

Figure 5: Example of a “domain_type” rule

7 CONCLUSIONS

We have showcased a linked data-enabled approach in which, constraint data coming from several heterogeneous and large systems can be accessed and queried in a seamlessly integrated way, using a set of reference ontologies and inference rules. The finding of this research is expected to provide an effective and efficient means to facilitate seamless constraint information exchange throughout the life cycle of LNG projects. This proposed approach can also leverage the effective concurrent collaboration between multiple partners not only within the same LNG project but also outside the project. Once isolated constraint data sets can be instantly accessed by related contractors or subcontractors, better decision making in terms of schedule catch-up would be achieved.

This research is limited to the small size of the data sets. In the future, we are planning to apply the approach in a real LNG construction project.

8 REFERENCES

- Bizer, C., Heath, T., Idehen, K. and Berners-Lee, T. (2008). Linked data on the web (LDOW2008). In Proceedings of the 17th international conference on World Wide Web (pp. 1265-1266). ACM.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009). Linked data-the story so far. *Semantic services, interoperability and web applications: emerging concepts*, 205-227.
- Chi, H. L., Wang, J., Wang, X., Truijens, M. and Yung, P. (2015). A conceptual framework of quality-assured fabrication, delivery and installation processes for liquefied natural gas (LNG) plant construction. *Journal of Intelligent & Robotic Systems*, 79(3-4), 433.
- CII (2013). *Advanced Work Packaging: Design through Workface Execution. Construction Industry Institute, Implementation Resource 272-2*, Version 3.0, 2013.
- Hamdi, O. (2013). Advanced Work Packaging from project definition through site execution: driving successful implementation of WorkFace Planning.
- Heath, T. and Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), 1-136.
- Kim, B. C., Teijgeler, H., Mun, D. and Han, S. (2011). Integration of distributed plant lifecycle data using ISO 15926 and Web services. *Annals of nuclear energy*, 38(11), 2309-2318.
- Lee, D. Y., Chi, H. L., Wang, J., Wang, X. and Park, C. S. (2016). A linked data system framework for sharing construction defect information using ontologies and BIM environments. *Automation in Construction*, 68, 102-113.
- Meditskos, G. and Bassiliades, N. (2010). DLEJena: A practical forward-chaining OWL 2 RL reasoner combining Jena and Pellet. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(1), 89-94.
- Musen, M.A. (2015). The Protégé project: A look back and a look forward. *AI Matters*. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4).
- Noy, N. F. and Musen, M. A. (2000). Algorithm and tool for automated ontology merging and alignment. In Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00). Available as SMI technical report SMI-2000-0831.
- Noy, N. F. and McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Available at: http://liris.cnrs.fr/alain.mille/enseignements/Ecole_Centrale/What%20is%20an%20ontology%20and%20why%20we%20need%20it.htm
- Tommelein, I.D. (1998). Pull-driven Scheduling for Pipe-Spool Installation: Simulation of Lean Construction Technique. *ASCE J. Constr. Eng. Manage.*, 124(4), pp. 279-288, doi: 10.1061/(ASCE)0733-9364(1998)124%3A4(279).
- Wang, J., Wang, X., Shou, W., Chong, H. Y. and Guo, J. (2016a). Building information modeling-based integration of MEP layout designs and constructability. *Automation in Construction*, 61, 134-146.
- Wang, J., Shou, W., Wang, X. and Wu, P. (2016b). Developing and evaluating a framework of total constraint management for improving workflow in liquefied natural gas construction. *Construction Management and Economics*, 34(12), 859-874.