A novel environmental impact assessment ontology using a graph-based database

Abstract
Environmental impact assessment (EIA) is becoming a critical part of road asset management. EIA-related knowledge and information are often sporadic, and traditional management approaches are considered inefficient in collecting fragmented project information and making such information available to project managers. Some studies have tried to establish basic ontologies to solve such problems using the resource description framework (RDF). However, these studies have not sufficiently identified all special concerns for EIA ontology, and RDF-based ontologies have limitations such as large storage space when dealing with large volume of data and poor visualisation performance. To overcome these challenges, this study aims to investigate a framework for organising, transferring, and visualising EIA information flow using Labelled property graphs (LPGs) based on the Neo4j graph database. Two scenarios in a real-life project, including searching for featured and hidden information are investigated to validate the improved performance of LPG-based EIA ontology (EIAO). The results show that the framework can provide support to project managers through the development of a smart decision-making system to reduce the time required to collect and analyse EIA information. By applying a novel data model to store and present ontology, storage efficiency and visualisation performance for managing environmental information are improved. In addition, the EIAO provides quicker search and reasoning functions to replace part of the manual work in traditional EIA, which can reduce time and cost in performing such activities

Keywords: Environmental impact assessment, ontology, property graphs, decision-making

1 Introduction

Management of both the to-be-built and existing road assets is a growing environmental concern. The inclusion of environmental impact assessment (EIA) of a road project is mandatory in many countries, such as the United Kingdom, Australia, and China (Shah et al., 2010). EIA includes identifying, predicting, evaluating, and mitigating environmental and other relevant impacts (Macintosh, 2010a). Poor quality of EIA information may cause conflicts between organisations or an unbalanced relationship between humans and nature (Morgan, 2012).

Since EIA activities are often complex and require data from various databases, efficient relationship-based information system is needed to manage the information (Halfawy, 2008). As a data-driven technology, the ontology method has significant benefits in the information management field because understanding the properties and relationships of all assets can help derive useful information either directly or by simulations for decision-making (Noy & McGuinness, 2001). Ontologies are an excellent supporting tool to efficiently integrate and share information from different systems and aid in decision-making (Muñoz et al., 2013). However, most studies have focussed on developing frameworks for partial functions, such as EIA performance evaluation and information searching. For instance, Corry et al. (2015) established a performance assessment ontology for evaluating the environmental impact of buildings, but did not provide detailed algorithms and codes for searching information and using it for efficient project management.

Labelled property graphs (LPGs) can be adopted to overcome the above problems. Using graph-based structures, the subjects, objects, and relation types can be added to various properties in an integrated format (De Abreu et al., 2013). Compared to the previous approaches of RDF, LPGs show better query performance and use less storage space for high-volume data (Vicknair et al., 2010). Therefore, LPG-based ontology has been tested and used in the oilfield industry, where it significantly reduced the storage size and improved retrieval efficiency by more than 30 times (Gong et al., 2018).

Therefore, this study aims to propose a novel LPG-based ontology framework for EIA management. The specific objectives were to: 1) build the EIA ontology using the standard ontology establishment method and EIA knowledge pool and 2) validate the proposed framework through a real-world case study. The remainder of this paper is structured as follows: Section 2 provides a brief but concentrated literature review on the current EIA information management approaches and ontological applications in relevant fields (e.g., the architecture, engineering, and construction industry). In Section 3, the detailed process for developing the LPG-based environmental impact assessment ontology (EIAO) is explained. A case study for the evaluation and validation of the proposed ontology is discussed in Section 4. Section 5 presents the results and discussion. The final section presents the conclusions and future direction.

2 Literature Review

2.1 Environmental impact assessment

Environment has become an important aspect in road construction and management in several countries (Chang et al., 2013; Li et al., 2010). EIA enables humans to recognise the significant and increasing impact of anthropogenic activities on the environment (Morgan, 2012). Since project data can be stored in databases owned by multiple stakeholders (e.g. clients, contractors, and suppliers), information management is an important part of EIA. Few studies have been conducted on the information management issue in EIA. For example, ISO 14001 (2015) has been established to help organisations enhance their environmental management performance. Many commercial asset management systems, such as Synergen, CityWorks, and Hansen, have been adopted by road agencies for EIA (Halfawy et al., 2006). An agent-based model was developed to reduce redundant and messy information using a generic behaviour model (Osman, 2012). However, several challenges, such as establishing alignment, formulating activity objectives, and managing multiple participants, still exist which can impede information flow and search efficiency (Schraven et al., 2011). A new methodology called multiple criteria data envelopment analysis was developed for EIA, which integrates qualitative data and quantitative factors to rank the alternatives (Zhao et al., 2006). Therefore, an innovative framework needs to be developed to establish an information management system that is more efficient in integrating data for sharing and decision-making.

2.2 Ontology in EIA

Ontology serves as explicitly defined metadata schemas, providing a controlled concept vocabulary and machine-processable semantics (Maedche & Staab, 2001). This machine-readable technology offers a wide range of opportunities in the asset management sector to increase the efficiency and improve data management. These studies prove that ontology can be effective in structuring and sharing information in relevant project management fields.

Ontology studies focussing on EIA areas can be divided into two main categories: those that establish a framework for information management and those that develop original data mapping. The first group introduces an ontology framework for knowledge management in a specific domain. For example, Garrido and Requena (2011) proposed an EIAO for knowledge mobilisation. The second category focuses on developing logical rules within a specific domain context. For instance, Niestroj et al. (2019) applied ontology to transfer road network data from geometry software into RDF format, and a rule-based technical documentation ontology was used to optimise the usage, maintenance, and configuration of the assets (Koukias & Kiritsis, 2015). In addition, in previous studies, RDF-based software (e.g. Protégé) were the most popular platform to establish ontology environments (Khondoker & Mueller, 2010). However, current RDFs use languages, such as SPARQL Protocol, RDF Query Language (SPARQL), and Structured Query Language (SQL), whose performance becomes weaker with increase in the number of relationships and data volume (De Abreu et al., 2013). LPG-based EIAO can address these limitations by providing a novel data structure with a special storage type and powerful query function. Furthermore, these advantages will be significant for the EIA that involves data and stakeholders from various systems (Morgan, 2012).

3 Development of the LPG-based EIAO

The development of the EIAO framework followed the most widely accepted practices outlined by Ontology Development 101 and the procedure of Noy and McGuinness (2001). The development procedure was adjusted according to the features of the EIA. As shown in Figure 1, it takes seven main steps to establish an ontology :1) defining the scope, 2) considering reusing ontologies, 3) knowledge acquisition, 4) defining the structure, 5) defining ontology establishing environment and data model, 6) ontology establishment, and 7) validation and improvement. The final product is a machine-readable ontology for EIA, and a few real-world scenarios will be tested to demonstrate the feasibility of the framework.

3.1 Scope of work

In this study, the scope of EIAO includes EIA planning and monitoring activities. The main components include identifying environmental hazards, environmental impacts, and EIA monitoring (audit). In these components, hazards and impact information can be treated as reasons for taking EIA audits, which are the most important and costly steps in the entire process (Morgan, 2012).

3.2 Reuse of existing ontologies.

When establishing a new ontology, existing ontologies are generally reused to avoid repetitive work and random errors. The work conducted by Garrido and Requena (2011), who proposed a knowledge mobilisation ontology for EIA, was the only relevant study in this field. However, it only covered activities to identify the environmental impact hazards. Some ontologies, such as common taxonomies of road management and construction projects (e.g. Das et al. (2015); Zhang et al. (2015)), fit the EIA development scope and hence were adopted as supporting materials in the following steps.

3.3 Knowledge acquisition

After defining the scope, the next step was to establish a knowledge pool for the EIAO. The main knowledge sources related to EIA of roads include standards (e.g. International Organization for Standardisation 55001: Asset management and 12006: Building construction), environmental authorities (e.g. Organisation for Economic Co-operation and Development), research institutions (e.g. Austroads and Environmental Protection Authority), books (e.g. *Methods of environmental impact assessment* by Morris and Therivel (2001); *Environmental impact assessment: theory and practice* by Wathern (2013)), and research (Jay et al., 2007; Macintosh, 2010a, 2010b; Morgan, 2012).

3.4 Defining ontology structure

A classic EIA system should be screened to identify any potential hazards. In case of hazard identification, its impact on the current environment and future projects must be determined (e.g. the impact level). A detailed EIA plan will be followed based on the documents and project conditions. If the plan is approved, relevant EIA action will be taken and recorded as an audit result.Based on this method, a linear relationship model was formed in this study.

Figure 1 presents the EIA audit actions and concepts in its high level as an example of the EIAO knowledge pool. Seven main auditing classes were identified: air quality monitoring, noise monitoring, landscape and visual impact assessment (potential impacts caused by construction activities), waste management, site inspection (on-site environmental issues), mitigation, and licence management (contractor qualification checking).

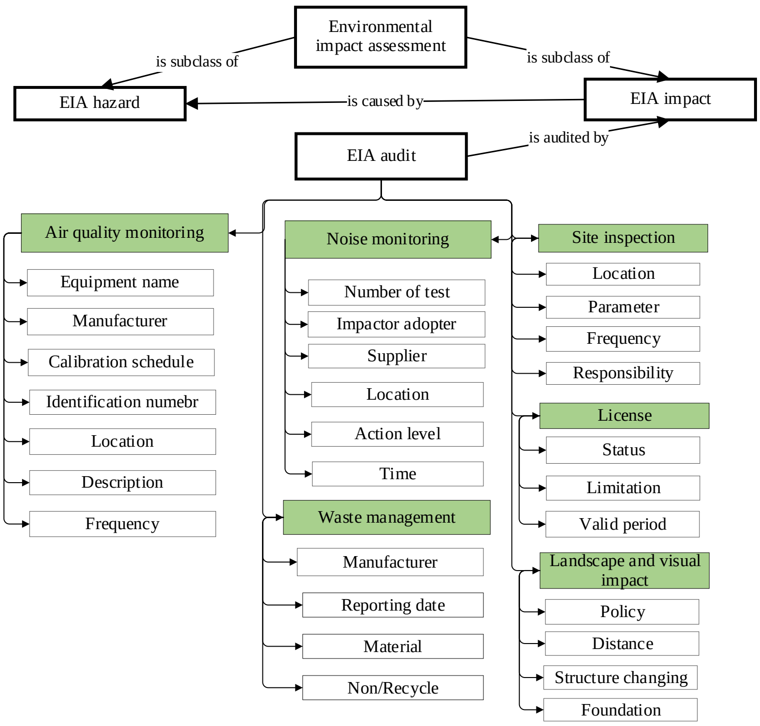


Figure 1 The EIAO structure.

3.5 Defining ontology establishing environment and data model

Several languages and data models use different structures, such as RDF, Web Ontology Language (OWL), and LPGs. RDF is one of the most popular standards for establishing an ontology, and is based on the expression *subject–predicate–object* (also known as *triples*) to represent relationships between instances. However, it also has limitations, such as requiring more storage space (Gong et al., 2018).

LPGs are a multiple labelled graph model, which was developed after RDF (Anikin et al., 2019). They present information by nodes, link those nodes by edges, and enrich them by embedded properties. Using graph-based structures, objects, and relation types in RDF can be effectively added to various properties (De Abreu et al., 2013).

3.6 Ontology establishment

This step converts the knowledge into LPG format. In EIA, there are four types of data: 1) drawings, e.g. inspection figures, 2) tabular data, e.g. spreadsheets, 3) raw digital documents, e.g. Adobe PDF and Microsoft Word, and 4) other paper-based materials, such as maps and paper records. A four-dimensional data model was utilised to convert these four types of data into the LPG format. The four steps of the model were related to defining the ontology instance (In), class (Cl), property (Pr) types, and relationships (R) (Zhou & Tao, 2011). The description of each dimension is as follows.

1. Class (Cl): All instances are sorted into a complete and logical hierarchy. It is defined from the knowledge pool and represented by a navigational relationship (direction).
2. Instance (In): This refers to the core of the ontology, also known as an entity in some studies.
3. Properties (Pr): This is an attribute that records the features and properties of an instance or relationship. Features and properties are embedded within nodes, which can also be treated as implicit nodes linked to nodes.
4. Relationship (R): This represents the relationship between instances and shows the direction of the information flow.

3.7 Validation and improvement

The validation main criterion chosen for EIAO was the time cost in querying, since information searching is the most concerned feature in a knowledge management system. The EIAO has been implemented in a standard Neo4j environment and uses Cypher as the query language (Zhang et al., 2015). A case study was conducted to test the actual usage and improved efficiency of the prototype. The measuring situations for querying were based two levels: searching for featured information. These are the most commonly used evaluation criterion when assessing an ontology (Scholer et al., 2002). Table 1 lists some of the query and reason functions that the EIAO can provide, where the accuracy is also manually tested.

Table 1 Description and used codes for evaluation.

Situation	Description	Project application	LPGs codes	Number
Featured	Find nodes with a certain feature	E.g. Find the works monitored by One Hour Total Suspended Particulate (1-hr TSP).	*MATCH (n:{parameter:*'1-hr TSP*'})* *RETURN n*	Query 1
	Find relationships with certain feature	E.g. Find the sub-class of 'Air Quality Monitoring'.	*MATCH p=()-[r: is subclass of]->(n: Air Quality Monitoring) RETURN p*	Query 2

Hidden	Reasoning requires nodes from an existing ontology	E.g. Find the unusual records in noise quality monitoring.	*Match (n: {result:}) AS one, (n: {limitation:}) AS two* *RETURN one > two AS result* *RETURN n, where result: 'true'*	Query 3
	Reasoning requires nodes from existing ontology	The manager tries to get the critical hazards, impact and auditing information of 'Air Monitoring Location 1'.	*MATCH (n:Air monitoring location {name:* Air Monitoring Location 1*}),* *p = shortestPath((n)-[*]-(Al))* *WHERE length(p) > 1* *RETURN p*	Query 4

4 Findings and Discussion

4.1 Case summary

The theoretical EIAO framework was tested and validated using a road project in Hong Kong, which required EIA for construction. The project aimed to improve the road network in the West Kowloon Reclamation Development area to meet future traffic needs. It started in 2015 and was completed in 2019. The EIA monitoring results during this period were reported based on the EIA audit manual.

The EIAO was implemented following steps 1–6 outlined in Section 3. Nodes, classes, and relationships were defined and linked by a knowledge pool and available project documents. This ontology scenario had six main EIA aspects: air quality monitoring, noise monitoring, waste management, site inspection, licence management, landscape, and visual impact assessment. A partial view of the EIAO is shown in Figure 2.

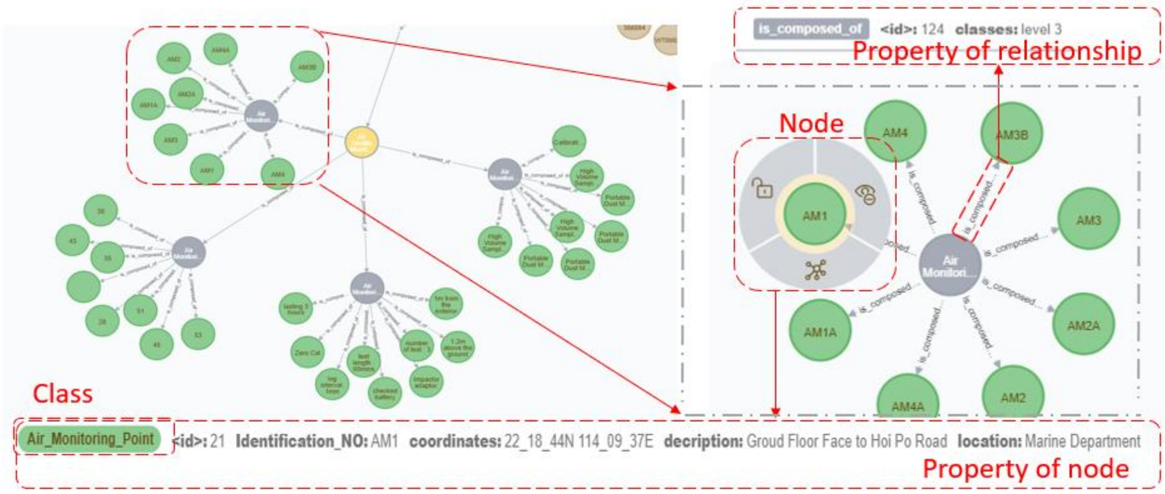


Figure 2 Partial view of the EIAO in Neo4j.

4.2 Scenarios

To validate the querying functions, the framework was investigated using three real-world scenarios with EIAO.

4.2.1 *Scenario 1 (Featured)*

The project manager wants to find a work procedure with a certain feature that is difficult to find manually among numerous unsorted documents. Thus, EIAO was implemented and used to search for the required information.

- Query 1 (Figure 3a) shows all the air monitoring work by 1-hr TSP during the projects. In total, four locations were found as 1-hr TSP measuring position: AM1-4.
- Query 2 (Figure 3b) shows all the locations for air monitoring in the 'Air Monitoring Location' class, and the director can view these locations to find out the person or body responsible for the site. There were eight monitoring locations in the entire project, and properties, such as road name, building name, floor number, and measurement parameters were also available.

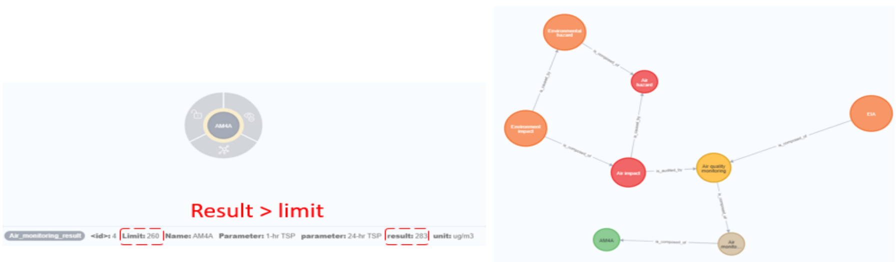
a b

Figure 3 Querying results for scenario 1.

4.2.2 *Scenario 2 (Hidden)*

Some unusual issues were recorded during EIA, and the manager wants to find the reason immediately and solve the problem. Traditional manual checking was inefficient and inaccurate. EIAO can provide a computer-based information-searching approach, where appropriate information can be digitally retrieved from existing databases.

- Query 3 (Figure 4a) shows all the observed records that exceeded the air quality limit. Using this code, Neo4j automatically compared the values under 'results and 'limitation', and showed the node whose 'results' value was higher than the 'limitation' value. 'AM4A' was found to fail to keep air quality within the standard level, and immediate responses, such as enhancing cleaning and monitoring frequency were taken on the construction site.
- Query 4 (Figure 4b) shows the critical steps from the beginning node to the node with the problem in the last query. The manager can use this code to find the shortest path from one node to node 'AM4A', which can help the team find the fastest way to solve the problem since the causes may exist in all relevant nodes.



a b

Figure 4 Querying results for scenario 2.

4.3 Comparison with popular ontology data model and manual checking

The improved querying time by EIAO in Neo4j was measured and compared with manual checking and RDF-based ontology. For each query, the duration from obtaining the order to manually finding the specific information in printed documents was recorded. For the RDF format, the Neosemantics (N10s) plug-in enables LPGs to transfer into RDF-based data, and SPARQL achieves most of the queries for RDF. The different results provided in each scenario and query are listed in Table 2.

Table 2 Information resources and comparison on search time.

Scenario	Query	Manual checking	EIAO	EIAO (RDF)
1	Q1	156s	0.06s	0.12s
	Q2	1278s	0.08s	0.15s
2	Q3	834s	24s	16s
	Q4	1592s	6s	11s

In the EIAO, querying general information is much faster than manually checking the results. For example, finding the monitoring parameter with 1-hr TSP (Q4) would take 156 s in daily reports, while in EIAO it requires only 0.06 seconds to run the prepared code and return the results. On an average, EIAO largely reduces the search time from 1827.3 s to 4.3 s, with 100 % accuracy compared to manual work.

4.4 Discussion

Traditionally, the professional knowledge pool for EIA is difficult to integrate because information is often scattered in various websites and documents. However, these attempts were not sufficiently comprehensive to be considered as smart decision-making processes. The LPG-based ontology has three significant contributions. First, environmental issues are a critical problem in society, especially in engineering projects (Morgan, 2012). Thus, an EIAO extends the available ontology research in EIA, especially the projects in the road asset management field. Second, this study chose LPGs as the data model and established a novel ontology in Neo4j instead of RDF, which enriches ontology development methods and demonstrating the application progress, advantages, and disadvantages by cases. The majority of ontologies in this field have been built by RDF/OWL (Das et al., 2014). This study proposes a novel ontology method by LPGs, which enables easier sorting and storing EIA knowledge. Third, the current EIA process involves manual information search and transfer, which delays the information flow and causes errors (Morgan, 2012). It also significantly improved the time to find featured information compared to manual work and RDF-based methods in two levels. It embedded relevant property information into the nodes, and can save a significant amount of time for the project team.

5 Conclusions and Further Research

Since EIA projects involve various knowledge resources, stages, and participants, integrating data for machine-reading purposes becomes a challenge. Ontology changes the management system using digital data storage, computer-based information processing, and data visualisation models. EIAO is designed to integrate comprehensive domain knowledge by reviewing relevant standards, guidance, project documents, and studies. By establishing an ontology with the LPG data model, information can be digitally stored as nodes and linked by edges to represent their relationships. Thus, the EIAO can support not only static information search but also continuous integration, reasoning, and information updates in ongoing projects. The proposed EIAO was validated in a real-world case, which proved its efficiency in information search (e.g. target step of work) along the project progress. Moreover, when new project information is automatically reasoned out by ontology, it can realise essential functions for project managers, such as finding the critical workflow and calculating basic delay, which provides the necessary resources for smart decision-making. The proposed EIAO extends the available EIAO research in the field of road asset management. Additionally, the proposed LPG-based ontology provides efficient storage, searching, and reasoning information. Consequently, the limitations of previous studies can be overcome, and project managers can save time and cost to achieve smart decision-making.

However, the proposed EIAO has several limitations. The knowledge pool was collected, sorted, and manually transferred. To make the ontology more practical and intelligent, it is necessary to automate its development using technologies, such as machine learning or big data applications. For example, Wu et al. (2021) applied natural language processing to a bridge maintenance ontology to automatically create new constraints. Although EIAO is capable of reasoning, it can be significantly improved in future by other supplementary programming languages through various application programming interfaces, such as UiO by Java and Session by C# (Hartig, 2019). In future studies, decision trees can be embedded into the EIAO itself to further automate the lifecycle process. Finally, this study did not design a unique and friendly interface for end-users.

6 Acknowledgments

This study was supported by the Australian Research Council Discovery Project (DP180104026).

7 References

Akiho. (2002). Overview of total productive maintenance, case studies of best practice of the Japanese manufacturing industry. *ChuSanRen (Central Japan Industries Association), Nagoya*.

Chang, Chou, Lin, & Hsu. (2013). *Road construction project environmental impact assessment scope definition using project definition rating index (PDRI).* Paper presented at the Advanced Materials Research.

Curry, Hasan, & O'Riain. (2012). *Enterprise energy management using a linked dataspace for energy intelligence.* Paper presented at the 2012 Sustainable Internet and ICT for Sustainability (SustainIT).

Das, Cheng, & Law. (2015). An ontology-based web service framework for construction supply chain collaboration and management. *Engineering Construction and Architectural Management, 22*(5), 551-572. doi:10.1108/Ecam-07-2014-0089

De Abreu, Flores, Palma, Pestana, Pinero, Queipo, Sánchez, & Vidal. (2013). *Choosing Between Graph Databases and RDF Engines for Consuming and Mining Linked Data.* Paper presented at the Cold.

Fernandes. (2000). Landscape ecology and conservation management—Evaluation of alternatives in a highway EIA process. *Environmental impact assessment review, 20*(6), 665-680.

Garrido, & Requena. (2011). Proposal of ontology for environmental impact assessment: An application with knowledge mobilization. *Expert Systems with Applications, 38*(3), 2462-2472.

Gong, Ma, Gong, Li, Li, & Yuan. (2018). Neo4j graph database realizes efficient storage performance of oilfield ontology. *PLoS One, 13*(11).

Halfawy. (2008). Integration of municipal infrastructure asset management processes: challenges and solutions. *Journal of Computing in Civil Engineering, 22*(3), 216-229.

Holzschuher, & Peinl. (2013). *Performance of graph query languages*. Paper presented at the Proceedings of the Joint EDBT/ICDT 2013 Workshops on - EDBT '13.

Jay, Jones, Slinn, & Wood. (2007). Environmental impact assessment: Retrospect and prospect. *Environmental impact assessment review, 27*(4), 287-300.

Koukias, & Kiritsis. (2015). Rule-based mechanism to optimize asset management using a technical documentation ontology. *IFAC-PapersOnLine, 48*(3), 1001-1006.

Li, Zhu, & Zhang. (2010). An LCA-based environmental impact assessment model for construction processes. *Building and Environment, 45*(3), 766-775.

Lohmann, Negru, Haag, & Ertl. (2016). Visualizing ontologies with VOWL. *Semantic Web, 7*(4), 399-419.

Macintosh. (2010a). The Australian Government's environmental impact assessment (EIA) regime: using surveys to identify proponent views on cost-effectiveness. *Impact Assessment and Project Appraisal, 28*(3), 175-188.

Macintosh. (2010b). Best practice environmental impact assessment: a model framework for Australia. *Australian Journal of Public Administration, 69*(4), 401-417.

Merdan, Koppensteiner, Hegny, & Favre-Bulle. (2008). *Application of an Ontology in a Transport Domain.* Paper presented at the 2008 IEEE International Conference on Industrial Technology.

Morgan. (2012). Environmental impact assessment: the state of the art. *Impact Assessment and Project Appraisal, 30*(1), 5-14.

Noy, & McGuinness. (2001). Ontology development 101: A guide to creating your first ontology. In: Stanford knowledge systems laboratory technical report KSL-01-05

Tang, Fan, Ni, & Shen. (2016). Environmental impact assessment in Hong Kong: a comparison study and lessons learnt. *Impact Assessment and Project Appraisal, 34*(3), 254-260. doi:10.1080/14615517.2016.1177934

Vicknair, Macias, Zhao, Nan, Chen, & Wilkins. (2010). *A comparison of a graph database and a relational database: a data provenance perspective.* Paper presented at the Proceedings of the 48th annual Southeast regional conference.

Wu, Wang, Wu, Wang, Jiang, Chen, & Swapan. (2021). Hybrid deep learning model for automating constraint modelling in advanced working packaging. *Automation in Construction, 127*, 103733.

Zhang, Boukamp, & Teizer. (2015). Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Automation in Construction, 52*, 29-41. doi:10.1016/j.autcon.2015.02.005

Zhou, & Tao. (2011). *A framework for ontology-based knowledge management.* Paper presented at the 2011 International Conference on Business Management and Electronic Information.