

A PROPOSAL TO USE SEMANTIC WEB TECHNOLOGIES FOR IMPROVED ROAD NETWORK INFORMATION EXCHANGE

Michael G. Niestroj^{1,*}, David A. McMeekin², Petra Helmholtz², Michael Kuhn²

School of Earth and Planetary Sciences, Discipline of Spatial Sciences, Curtin University, Perth, Australia -

¹ Michael.Niestroj@postgrad.curtin.edu.au, ² (D.McMeekin, Petra.Helmholtz, M.Kuhn)@curtin.edu.au

Commission IV, WG IV/2

KEY WORDS: Asset Management, Data Conflation, Decision Tree, Metadata, Road Network, Semantic Rules, Semantic Web

ABSTRACT:

Data harmonisation improves the coherence between data sets within and across themes and is, therefore, a very helpful tool for governmental agencies, companies and other organisations that share their data. This research focuses on horizontal infrastructures, namely roads, and proposes a new strategy to apply Semantic Web Technologies. The aim is to understand if their application is efficient and effective in filling the gap of data harmonisation in Australia's and New Zealand's road asset management systems within the definition of location. The proposed strategy has three stages. First, available international data standards for road assets will be analysed to identify the gaps within these standards and create recommendations towards an improved standard. The second stage is for the location aspect within each stage of the life cycle of asset management with respect to existing road asset data standards. Finally, in a third stage Semantic Web Technologies, ontologies and semantic rules will be used to build a prototype solution for road asset data conflation by merging multiple data sources that share no common lineage. The application of these technologies will allow for easier search and discovery of this data as well as facilitate the automated processing and updating of this data over the Web.

1. INTRODUCTION

The challenge of not matching coordinates of objects within data sets from either different suppliers or data sources is well-known and has been researched by many groups, (e.g. Ruiz-Alarcon-Quintero, 2016; Zhang et al., 2005; Kim et al., 2016). The demand exists within transport agencies in Australia and New Zealand to share data and, therefore, road asset data conflation (in means of merging multiple road asset data sets) needs to be considered for a *Construction to Operations Network information exchange* (CONie) (Kenley and Harfield, 2015).

In a Geographic Information System (GIS), the combination of two data sources (with overlapping information) into a single source is called conflation (Longley et al., 2005; Seth and Samal, 2017). Although both data sets are individually accurate and perfectly adequate for one application, the history of GIS shows often that the task completely fails when an application is merged with new data sets that had no common lineage. For example, while merging Global Navigation Satellite System (GNSS) coordinates with data from the US Bureau of the Census TIGER files it occurs that points appear on the wrong side of the street because of the positional accuracy of 50 m in parts of the TIGER data sets (Longley et al., 2005).

Available road asset metadata definitions of international standards and standardisation organisations are used to maintain a common language. In the management of vertical infrastructure (buildings), solutions are available for the information exchange of asset data (e.g. *Construction to Operations Building information exchange* [COBie] (East, 2006) and Industry Foundation Classes [IFC]). A lack of open standards exists in the management of horizontal infrastructure, especially, concerning CONie related to location information (East et al., 2016b).

This research can benefit from other developments. For example, the buildingSMART (2017) IFCs are a data standard for open-BIM, the Austroads (2016b) road asset data standard is a framework for asset management of horizontal infrastructure, and the hierarchical definition of location of a construction project is called Location Breakdown Structure (LBS) (Austroads, 2016b; Kenley and Seppänen, 2009). These data standards are not optimal (Gelder, 2018; Amann et al., 2014) and, therefore, an improved data specification for road asset and road localisation with metadata representation needs to be identified.

Data harmonisation aims to improve the coherence (respecting geometry and semantics) between data sets within and across themes in such a way so that they fit together (Box et al., 2015; Tóth et al., 2012). By means of data harmonisation, two or more data sets are merged together to create a new data set. Another aspect is that of query federation whereby one data set is conflated to another data set (with possible different specifications and similar features) using queries to create a new data set for the lifetime of the query.

A number of studies exist that try to solve the problem of map conflation with a combination of road data analysis, image processing and rubber-sheeting methods (Song et al., 2006, 2009; Tong et al., 2009). However, image processing approaches are not the best tool for very large road networks (e.g. for Australia) because of many possible factors, such as intensive computing power, image data preparation and data merging (Abdollahi and Riyahi Bakhtiari, 2017; Cheng and Weng, 2017). Rubber-sheeting creates a new data set with adjusted polylines containing metadata information of conflated data sets (Haunert, 2005). Yu et al. (2018) uses Semantic Web Technologies to utilise a Points of Interest (POI) approach to conflate POI data sets from Landgate, Western Australian Police and the Department of Fire and Emergency services into one new POI data set.

*Corresponding author

This study proposes a new strategy to analyse, match and conflate road asset metadata information from multiple data sources such as Main Roads Western Australia (MRWA), Western Australian Land Information Authority (Landgate), as well as OpenStreetMap (OSM) and enable both harmonised new data and federated query data. The main focus is data federation with the use of Semantic Web Technologies. Data integration is the approach of including data into a data set without creating a new data set (Popovich et al., 2009).

The structure of this concept paper can be described as followed. The proposed data sources of this strategy will be introduced next. Then, the preliminary work, main work and evaluation of this proposed strategy will be explained with the help of a flow chart. After that, examples will be demonstrated related to the main work of this proposed strategy. At the end of this paper, this work will be concluded.

2. DATA

The data sets presented in this paper are either based on data sources from Australia's National Map, Australian Capital Territory (ACT) (e.g. road signs), City of Greater Geelong (e.g. road signs), Landgate (e.g. freeways, main and minor roads), MRWA (e.g. roads and regulatory signs), OSM or synthetically generated. The corresponding data sources are pointed out throughout this proposal wherever necessary.

The ACT, City of Greater Geelong, Landgate and MRWA data sets are quality controlled, as they are generated by public authorities, and the measurements are based on ground surveys or photogrammetric mapping. The OSM data sets are crowd-sourced, edited by the OSM community on a trust basis, as everybody can contribute data using aerial imagery and GNSS devices.

In the future, additional road asset data sources can be employed, such as these from the Department of State Growth (Tasmania), Department of Transport and Main Roads, NZ Transport Agency, Roads and Maritime Services, VicRoads, as well as from other transport agencies.

3. PROPOSED STRATEGY

The workflow of the proposed strategy is visualised in the flow chart illustrated in Figure 1 and is subdivided into three stages. The first stage is about the identification of an improved standard specification for road asset management and the establishment of a Semantic Web supported environment. The second stage consists of collecting road map data and the available methods for the location transformation. The first two stages can be regarded as a preliminary requirement for the third stage. Therefore, the main focus here is related to the third stage, developing a prototype solution using Semantic Web Technologies for road asset data conflation. Finally, the results of the three stages will be evaluated.

3.1 Stage 1: Identify an improved data specification and establish the Semantic Web based environment

This stage is required to establish a base for further development processes within this strategy. Firstly, an improved data model (specification) can be discovered, and secondly, the programming environment can be established. Both approaches are required to be developed with the support of Semantic Web Technologies.

In the scope of this stage, an overview has been developed by Niestroj et al. (2018) to review standards and current developments for road asset information exchange. This research focuses on the most significant standards and standardisation organisations related to the proposed strategy (East et al., 2016b), such as:

- buildingSMART: IFC Alignment and IFC Road.
- OGC: International industry consortium which develops publicly available interface standards.
 - OGC LandInfra / InfraGML for land and civil engineering infrastructure facilities.
 - CityGML is an XML-based open data model for storage and data exchange of virtual 3D city models.
- ISO 19115-1: Geographic information - metadata.
- AS/NZS ISO/NP 19148: Geographic information - linear referencing.
- Austroads: Road data harmonisation project.
- ISO 55000-55002: Asset management standards suite.

With the influence of the above-mentioned standards, a new improved data specification can be explored supporting project relevant key elements, such as road asset support, geo-referenced coordinates, datums, common metadata identifiers and LBS.

An appropriate programming environment with Semantic Web Technology support will be established. The World Wide Web Consortium (W3C, 2017) provides information about available tools for the Semantic Web development, such as AllegroGraph Resource Description Framework (RDF) Store, Apache Jena, OpenLink Virtuoso, Oracle Spatial 11g or Sesame. The Apache Jena framework is an interesting programming environment which among others, contains features such as a triple store, a rule reasoner as well as a parser. Currently, the Apache Jena framework seems to be the best solution for use within this research project and will therefore be employed. In addition, the ontology and semantic rules editor Protégé will be used. Both tools are Open Source and used by other researchers, such as Yu et al. (2018).

3.2 Stage 2: Implement methods for location transformations

At the beginning of stage two, the data model will be established within a Semantic Web programming environment. This developed system enables, data sets being entered interactively from different data sources into the data model to provide an initial data set. The main task of this stage is to apply transitions between datums and coordinate systems so that a unified data set will be accessible.

Road asset data will be collected from different data sources such as MRWA, Landgate and OSM and further analysed. After analysing the data sets, an algorithm will be developed to transfer the data into a Semantic Web triple store using the *Python* programming language. Inside the RDF triple store, the data is now written as an ontology in the form subject – predicate – object. For instance, 'a road', 'has name' 'Plain Street' (Berners-Lee et al., 2001). Although Sequeda et al. (2011) presented seven different ways of how to automatically map to the Semantic Web, the strategy will first, analyse data sources interactive and find the corresponding matching parameters. This process will be then integrated, automated and tested.

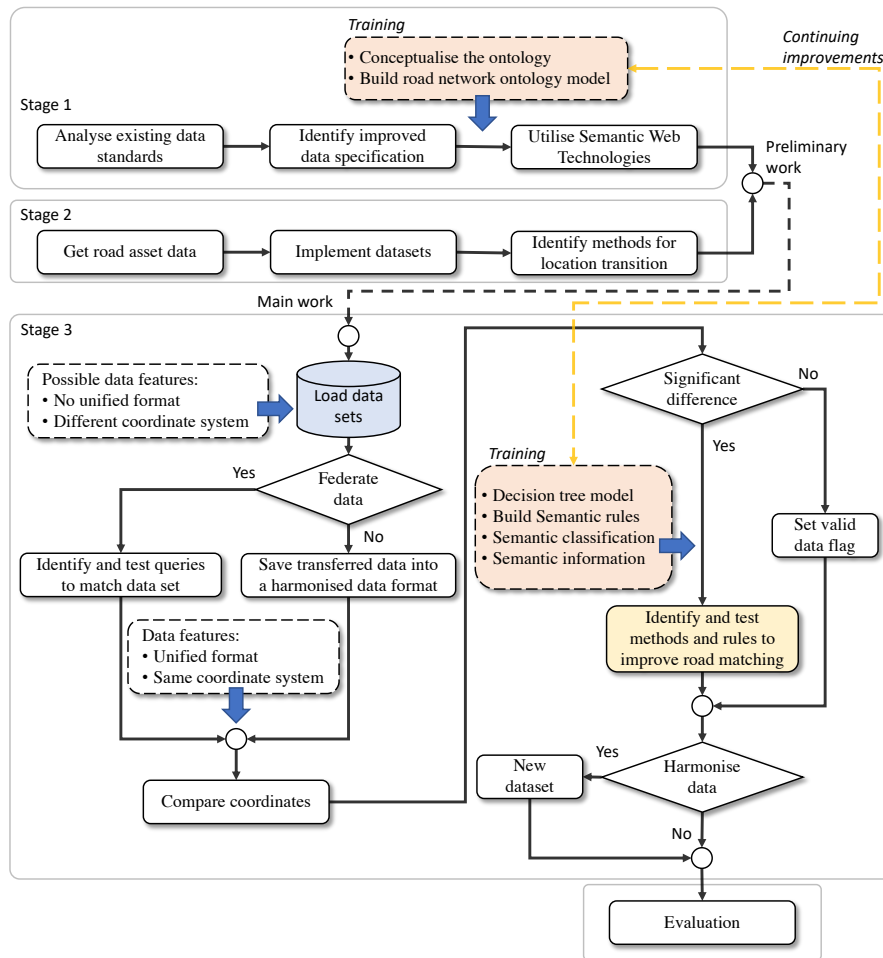


Figure 1. The workflow of this strategy is subdivided into three stages. The yellow box represents the primary task of this research and is connected to the training techniques in the orange boxes. The training symbolises testing of new semantic rules and decisions which flow as improvements back into the model.

The raw road asset data is not always available in the same datum. For example, road data from OSM is in the World Geodetic System (WGS) 84, while data from Landgate and MRWA are currently in Geocentric Datum of Australia (GDA) 94 but will change into GDA2020 (ICSM, 2018), and data from New Zealand is in New Zealand Geodetic Datum (NZGD) 2000. Therefore, this project will enable transformations between GDA (94–2020), ITRF (1992–2014), WGS84 and NZGD2000 by using available transformation parameters.

In the Australian context, transformation parameters will be taken from Altamimi et al. (2016, 2011, 2007) for the transitions between ITRF2008–2014, ITRF2005–2008, and ITRF2000–2005, respective. The transformation parameters between ITRF1996–2008 and GDA94 are defined in Dawson and Woods (2010). And the transformation parameters between ITRF1988–2008, WGS84 and NZGD2000, as well as the transformation parameters between ITRF1996–2008 and GDA94 are defined in Donnelly et al. (2014).

3.3 Stage 3: Develop a prototype for road data conflation using Semantic Web Technologies

The preliminary work of the preceding stages has to be completed before beginning this stage. This stage will focus on the strategy of implementing a prototype approach for road asset data con-

flation using Semantic Web Technologies. An example of an arbitrary road within multiple data sets is indicated in Figure 2. The left image shows an extracted data set from OSM plotted in a geo-referenced image. In contrast, the right image indicates Australia’s National Map with an activated MRWA data set of the same junction. One can see that these two data sets do not harmonise with each other because the road’s nodes are not matching (red ‘P’ in Figure 2). In fact, the road connecting points ‘P’ are about 12 m apart.

As mentioned, this proposal provides a strategy to utilise Semantic Web Technologies to conflate road asset data sets. Therefore, data sets will be loaded either from the RDF triple store or imported from an available source (such as OSM and MRWA). A first decision is whether the data set has to be federated or not. The benefit of federating data is that a new data set is not created, it exists only for the lifetime of the query. The disadvantage of federating data is that the query must be executed each time from scratch. The work from Buil-Aranda (2014) showed that queries can be optimised to affect the execution time. After treating the data sets in the selected way (federated or harmonised, see Figure 1), the data are in any way then available in a unified metadata format; using the same datum and coordinate system.

The geo-referenced coordinates of the loaded road assets will be then compared by utilising a commonly algorithm to calculate the



Figure 2. The picture on the left side shows a high-quality geo-referenced image with a data set of OSM, while the right image represents a National Map data set of the same junction with data from MRWA. The road connection points are indicated as 'P' (sources: <http://spookfish.com> (image left); <http://openstreetmap.org> (road data left); <http://nationalmap.gov.au> (image and road data right)).

distance in meters between geographic coordinates (MTL, 2018). If the calculated difference between two coordinates is within the tolerance range of transport agencies (meaning the distance is not significant), then a flag-variable will be set to validate the data sets as accurate. If the distance between the road coordinates is significant, test methods and approaches for an improved road matching will be addressed.

The training part of this proposed strategy is connected to the ontology model which is represented by the data specification in stage one (see Figure 1). Added metadata properties and other improvements must be included into the ontology model. The adjustment of the ontology model has an influence on the training. For example, after adding an ontology into the model, the connected rules and the decision tree need to be updated and tested. The approach of an object-based semantic classification has been analysed by Gu et al. (2017) for images using ontologies. This strategy will also utilise ontology modelling, decision trees, and semantic classification based on semantic rules. However, the needed ontologies, decision trees, classification and semantic rules do not exist and are going to be identified, tested and analysed within stage three. The approach in this strategy recommends using metadata information of roads and road assets (such as lightning, traffic lights, crossing and turn signs) to provide effective road map conflation of multiple data sets and data sources. This information can then be used to formulate an ontology. To the best of the authors' knowledge, this approach has not yet been developed, tested, or implemented.

Ontologies can be realised with both, the Web Ontology Language (OWL) and the Semantic Web Rule Language (SWRL). Nodes and statements written in OWL can be further visualised as a LBS oriented decision tree model. The disadvantage of OWL is that the source code structure is rather long, and nodes are cumbersome to implement such as indicated in the OWL decision tree in the work of Gu et al. (2017). In comparison, semantic rules written in SWRL are simple to express. The disadvantage of SWRL is that a visualisation of semantic rules is not possible (Gu et al., 2017). To maintain a sophisticated network thousands of nodes and rules must be integrated. Tests will show which of the two methods is better suited for road network information exchange.

To be successful in stage three, it is necessary to apply the rules designed by East et al. (2016a):

1. *Be Specific, Not Abstract*: Allow any construction knowledge domain to be guided by their own component. Generic

approaches might be useful from a data modelling perspective, that does not mean automatic that they are efficient in the long term.

2. *Be Complete, Support Implementation*: The improved standard specification must be proofed for completeness, well-documented and tested before it is given to the public.
3. *Be Incremental, Not Aspirational*: Develop a project with the best use of current available techniques to provide an innovative solution, and the project's future is secured.

Currently, there is no existing data specification for the complete road network available in Australia and New Zealand. This required specification has to be identified in stage one and continuity improved after adding new data sets into the RDF triple store. For example, Laurini (2014) points out that roads can be categorised into motorways (centreline, lane number, emergency lane, verge/shoulder) or streets (sidewalk and lane). In fact, the representation of a road depends on the industrial usage: for traffic engineers, a road is represented by a graph; for street maintainers by a surface; for cadastre officers by two polylines; and for technical network engineers by volume (Laurini, 2014). However, a road asset network contains much more information (such as trees, street plates, traffic lights, turn signs and crossings) that can be used as POIs (Austroads, 2016b; Berdier, 2011).

3.4 Evaluation

This stage can evaluate the results of this proposed strategy for accuracy and industrial application. Further, it can be evaluated to what extent it is reliable to have both solutions OWL and SWRL running, and what are the advantages and disadvantages of each approach especially applied to the road network. To build ontologies, find available ontologies and proper ontology tools as well as to understand how to use them, is a difficult and time-consuming process (Gu et al., 2017; McMeekin, 2015).

Comparing OWL and SWRL techniques based on image processing decisions is conducted by Gu et al. (2017) showing that minor improvements (about 1%) in land cover classification can be achieved by comparing both techniques. The evaluation of a Semantic Web map conflation POI approach is guided by Yu et al. (2018) with the high accuracy of 98% in POI identification for shopping centres which are included in all data sets.

At this stage, it cannot be said whether data conflation relied on road asset metadata information is fast and effective because this is the first time that such an approach is proposed within Australia. In order to identify the most suitable method, the following evaluation can be performed.

Road asset data sets will be conflated with multiple data sets to validate road assets with same coordinates, and neighbouring road assets will be taken into account. If metadata information is not available in all data sources, then semantic rules and decisions will be processed to classify and to identify the required information from surrounding objects. For this missing information, a coherence probability will be calculated based on Bayesian networks employing a probabilistic framework, such as *BayesOWL* for uncertainty modelling from Ding et al. (2006). As a core part of this strategy, the accuracy of that process will be evaluated by analysing the number of correct adapted information. In addition, for road asset data sets (especially street sections) the correlation of these data points will be computed, in order to draw conclusions about how the data sources differ.

4. EXAMPLES

As mentioned in Section 1, an LBS defines the location hierarchical in a construction project, and can therefore be successful applied to the road network, such as the sample LBS of a road network with bridges, roadways, roads and road signs (see Figure 3).

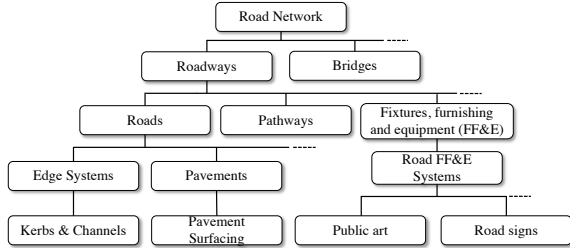


Figure 3. A sample LBS for the road network (Gelder, 2018).

Each of these assets within the LBS can contain a decision tree and a set of semantic rules. The following examples in this section are related to ‘road signs’.

4.1 Data identification

A unique identification of road network inventory is required to provide semantic heterogeneity. For instance, considering a STOP sign within the data sets MRWA-503 (regulatory signs), Road Signs from both the City of Greater Geelong and the ACT. Each data set has a different data structure, with data fields that need to be analysed and identified. The location of a road sign is given in all three data sets by the geographical latitude and longitude. The following longitude column name variations have been identified interactively: ‘Longitude’, ‘LONGITUDE’ and ‘long’. Further, the data value can be verified, which must be between 110 and 180 for Australia and New Zealand. Similar rules can be applied for the latitude. Each road sign has a unique identifier, such as “R1-1” is employed for the STOP sign. It cannot be said in which column the identifier is, as the data layout is not standardised. For example, the MRWA-503 data set allows up to four road signs in one data row if road signs are mounted to the same fixture. Whereby, the other two data sets allow only one sign per data row, and if more signs are attached to one mount, then a new data row is employed for each sign using the same coordinates. Therefore, each column can be checked to identify the content of a data set. However, the key information of identifying road signs are longitude, latitude and sign identifier, which is enough to determine a road sign and its location. The strategy focuses on extracting only the required information of road asset data sets, as this is fast and efficient for further processing.

4.2 OWL decision tree and SWRL semantic rules

An approach to conflate data sets could be oriented at the position of available road signs at an intersection. For instance, according to MRWA (2015) a rule for the placement of a STOP or GIVE WAY sign is that it must be placed as close as possible but not further than 15 m from the corresponding line. A further rule is that this sign must be placed between 0.6 m and 5 m from the edge of the carriageway (MRWA, 2015). With this information, a first ontology can be formalised to validate the position of these signs.

Figure 4 shows an image of an intersection with plotted road asset data. The indicated single placed STOP sign represents the data



Figure 4. The illustration shows multiple data sets, as well as synthetic data (the group of four STOP signs) plotted on a geo-referenced image (sources: <http://data.wa.gov.au> (data); <https://www.spookfish.com> (image)).

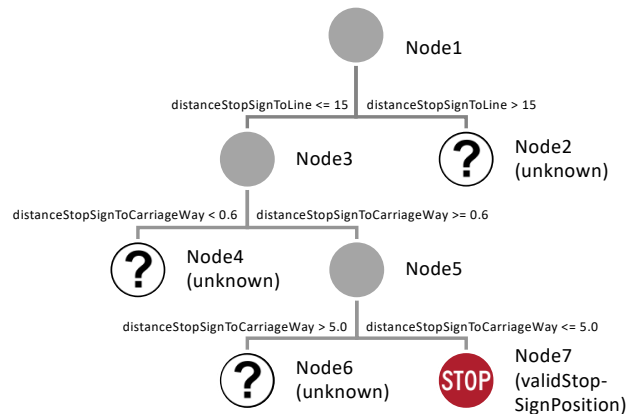


Figure 5. A visualised decision tree to validate the position of a STOP sign.

set MRWA-503 (regulatory signs), and the group of four STOP signs is a synthetic data set. It can be seen, that according to this example the MRWA data set will be set as a valid data set because the position of the STOP sign is within 4 m from the carriageway and about 10.5 m from the corresponding line away. In contrast, the synthetic data set indicates that the STOP signs are positioned more than 5 m from the carriageway. Therefore, the position of the synthetic data set can be set as invalid. In this specific situation, an exception handling will occur because of the fact, that multiple data sets of the same sign are not within the validation rules. A further investigation can be conducted in this case. For example, to either interactively validate the data or to renew the STOP sign location measurements.

The following two examples show both the realisation of the first ontology, respective, in the Web Ontology Language (OWL) and the Semantic Web Rule Language (SWRL).

Example 1 (decision tree)

Nodes and statements are written in OWL and can be further visualised as an LBS oriented decision tree (such as road network, roadways, roads and road barriers). Source Code 1 indicates the semantic decision rules of the above example written in OWL. To represent the first ontology in OWL it is required to use seven nodes because of the decision tree notation. In this example, there are unknown nodes and the validation of the STOP sign position is processed within node seven. The result of the OWL expression is visualised in the corresponding decision tree in Figure 5. The decision tree is simple to understand and a huge benefit of the rather long OWL syntax.

| | | |
|-------------|-------------------------------|-------|
| Individual: | Node1 | |
| Types: | Node | |
| Facts: | | |
| | GreaterThan | Node2 |
| | LessThanOrEqual | Node3 |
| | distanceStopSignToLine | 15.0 |
| Individual: | Node2 | |
| Types: | unknown | |
| Individual: | Node3 | |
| Types: | Node | |
| Facts: | | |
| | LessThan | Node4 |
| | GreaterThanOrEqual | Node5 |
| | distanceStopSignToCarriageWay | 0.6 |
| Individual: | Node4 | |
| Types: | unknown | |
| Individual: | Node5 | |
| Types: | Node | |
| Facts: | | |
| | GreaterThan | Node6 |
| | LessThanOrEqual | Node7 |
| | distanceStopSignToCarriageWay | 5.0 |
| Individual: | Node6 | |
| Types: | unknown | |
| Individual: | Node7 | |
| Types: | validStopSignPosition | |

Source Code 1. A decision tree expressed in OWL (adopted from Gu et al., 2017).

| |
|--|
| <code>distanceStopSignToLine(?x,?y), LessThanOrEqual(?y, 15.0) -> validStopSignPosition(?x); distanceStopSignToCarriageWay(?x,?y), GreaterThanOrEqual(?y, 0.6) -> validStopSignPosition(?x); distanceStopSignToCarriageWay(?x,?y), LessThanOrEqual(?y, 5.0) -> validStopSignPosition(?x);</code> |
|--|

Source Code 2. A set of semantic rules written in SWRL.

Example 2 (semantic rules)

The semantic rules of the first ontology expressed in SWRL are realised with just three semantic rules in Source Code 2; whereby, each rule is separated by a semicolon. That is enough to define that the STOP sign must be within 15 m to the corresponding line, as well as between 0.6 m and 5.0 m to the carriageway. A visualisation of the semantic rules is not possible.

By comparing these two approaches, one can see that the second example is shorter and simpler than the source code from the first example. This is related to the fact that the semantic rules are defined, independently. Whereby, within the definition of a decision tree always two cases must be specified and, therefore, the number of nodes is growing very fast. Overall, these examples validate the position of a STOP or GIVE WAY sign. For the validation of a whole road asset network, many more semantic nodes and rules are required to be identified, implemented and tested.

4.3 Road asset data conflation

An example of road asset data sets from different data sources (e.g. Landgate and MRWA) is illustrated in Figure 6. The yellow polylines are from the Landgate data set LGATE-195 (main roads, freeways and national highways). The STOP and GIVE WAY signs are from the data set MRWA-503 (regulatory signs), the road stopping places (highlighted as green circles) are from the data set MRWA-513, and the blue and cyan polylines are from the data set MRWA-515 (road hierarchy).

A road data set is usually oriented at the road centreline. Therefore, the plotted data sets should share a similar representation

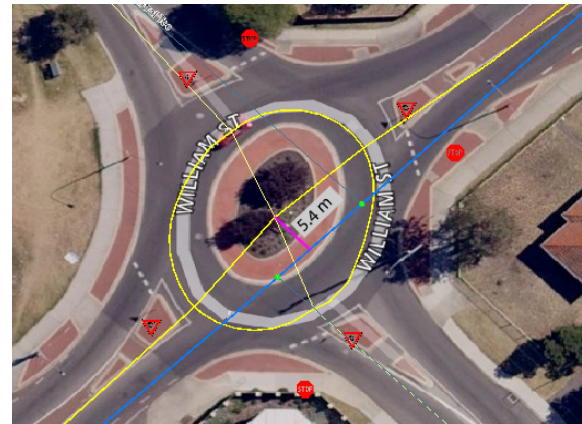


Figure 6. The illustration shows multiple data sets plotted on a geo-referenced image (sources: <http://data.wa.gov.au> (data); <https://www.spookfish.com> (image)).

of the centreline. However, only the east-south exit shows a harmonised data set with matching coordinates, whereby the other three exits do not share a path (yellow to blue polylines and yellow to cyan polylines). In this example, the data set LGATE-195 (yellow polylines) represents the centreline slightly with an offset to the red coloured pavement. The corresponding MRWA-515 data set (blue and cyan polylines) shows the road hierarchy connected to the road stopping places. The distance of about 5.4 m indicated in pink between the centerlines is calculated using the *haversine* formulae described in MTL (2018) to derive the great circle distance (d) between two points.

$$\Delta\varphi = \varphi_2 - \varphi_1; \quad \Delta\lambda = \lambda_2 - \lambda_1 \quad (1)$$

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1) \cdot \cos(\varphi_2) \cdot \sin^2\left(\frac{\Delta\lambda}{2}\right) \quad (2)$$

$$c = 2 \cdot \text{atan2}\left(\sqrt{a}, \sqrt{1-a}\right) \quad (3)$$

$$d = R_E \cdot c \quad (4)$$

where φ = latitude in radians
 λ = longitude in radians
 a = square of half of the chord length between the points
 c = angular distance in radians
 $R_E = 6,371,000$ m

First, the differences Δ in geographic latitude φ and longitude λ are calculated. The square of half of the chord length a is then calculated between the points. After that, the angular distance c in radians is calculated, and then finally multiplied with the Earth's mean radius R_E to convert the spherical distance d into meters.

However, there is more to be analysed in Figure 6. For example, there are four exits, four GIVE WAY signs and three STOP signs. It seems that there is a fourth STOP sign missing. In reality, it looks quite different because none of the STOP signs exist. The MRWA-503 (regulatory signs) data set is from the 6th September 2017 (MRWA, 2017). The high-resolution airborne images in Figure 7 are taken, before and after the MRWA data set release date. As a result, if the STOP signs did exist, then they must be on at least one of the images in Figure 7.



Figure 7. The left image shows a high-resolution satellite image from the 5th September 2017, and the right image shows the same road asset section with data from the 18th September 2017 (source: <https://www.spookfish.com>). The green rectangles indicate the location of placed GIVE WAY signs.

By comparing Figure 6 with Figure 7 one can see that the STOP signs cannot be found in Figure 7. The solution to this issue is related to a representation error within QGIS (<https://qgis.org>) and the road asset data set. Although metadata information (e.g. longitude, latitude, regulatory and road name) exists for each STOP sign within the MRWA-503 data set, there is no information available in the data fields 'date approved' and 'date installed', whereby the GIVE WAY signs have their entered dates '2005-11-24' and '2006-05-22', respectively. This means, that the placement of the STOP signs was not approved, and therefore the signs have not been installed. In a further development process of this strategy, data sets will be analysed for completeness to avoid misinterpretation.

5. CONCLUSIONS

The paper provided the concept of a new framework to conflate road asset data sets by employing Semantic Web Technologies. In the first stage of this framework, road asset data standards will be analysed with the focus on current international developments (such from Austroads, buildingSMART and Open Geospatial Consortium (OGC)) to provide a recommendation for an improved specification for road asset and road localisation. The second stage is to identify and to apply the best available transition between coordinate systems (such as Geocentric Datum of Australia (GDA) 94, GDA2020, New Zealand Geodetic Datum (NZGD) 2000) to make available unified coordinates.

The first two stages are preliminary activities to enable the main and third stage namely, the implementation of a novel prototype solution with the task of road matching, map conflation, data harmonisation and data federation utilising Semantic Web Technologies. The evaluation compares the efficiency of OWL and SWRL techniques applied to a road network approach. The accuracy of the coherence probability using semantic rules and decisions to guess missing information can be evaluated with a Bayesian network approach.

This project is a significant component in the development of a digitised information exchange for horizontal infrastructure with the support of conflating road asset data sets. A business case developed by Austroads demonstrated that significant benefits and cost savings between \$65 and \$130 million per annum can be obtained by harmonising road asset data (Austroads, 2016a). This proposed strategy has the potential to contribute to these cost savings by developing a common accepted data standard specification within the use in transport agencies in Australia and New Zealand.

ACKNOWLEDGEMENTS

We thank Spookfish for using their data. This work is supported financially by Australian Research Council project LP160100524, Curtin University, Australia's Sustainable Built Environment National Research Centre (SBEnc) Project 2.33 and its partners. Core partners are Swinburne University of Technology, Queensland University of Technology, University of Melbourne, New Zealand Transport Agency, MRWA, Roads and Maritime Services and Aurecon Australasia PTY LTD.

REFERENCES

- Abdollahi, A. and Riyahi Bakhtiari, H. R., 2017. Roads Data Conflation using Update High Resolution Satellite Images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences IV-4/W4*, pp. 3–7.
- Altamimi, Z., Collilieux, X. and Métivier, L., 2011. ITRF2008: an improved solution of the international terrestrial reference frame. *Journal of Geodesy* 85(8), pp. 457–473.
- Altamimi, Z., Collilieux, X., LeGrand, J., Garayt, B. and Boucher, C., 2007. ITRF2005: A new release of the International Terrestrial Reference Frame based on time series of station positions and Earth Orientation Parameters. *Journal of Geophysical Research* 112(B9), pp. 175–19.
- Altamimi, Z., Rebischung, P., Métivier, L. and Collilieux, X., 2016. ITRF2014: A new release of the International Terrestrial Reference Frame modeling nonlinear station motions. *Journal of Geophysical Research: Solid Earth* 121(8), pp. 6109–6131.
- Amann, J., Flurl, M., Jubierre, J. R. and Borrmann, A., 2014. An Approach to Describe Arbitrary Transition Curves in an IFC-Based Alignment Product Data Model. In: *Computing in Civil and Building Engineering (2014)*, American Society of Civil Engineers.
- Austroads, 2016a. Austroads Annual Report 2015-16. Technical Report AP-C20-16, Austroads.
- Austroads, 2016b. Data Standard for Road Management and Investment in Australia and New Zealand: Draft 1 Version 2.
- Berdier, C., 2011. Road System Ontology: Organisation and Feedback. In: *Ontologies in Urban Development Projects*, Springer London, London, pp. 211–216.
- Berners-Lee, T., Hendler, J. and Lassila, O., 2001. The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *Scientific American*.
- Box, P., Simons, B., Cox, S. and Maguire, S., 2015. A Data Specification Framework for the Foundation Spatial Data Framework. Technical report, CSIRO.
- Buil-Aranda, C., 2014. *Federated Query Processing for the Semantic Web*. Vol. 15, IOS Press.
- buildingSMART, 2017. Industry Foundation Classes. Version 4.1 Final Release.
- Cheng, M. and Weng, Q., 2017. Urban Road Extraction from Combined Data Sets of High-Resolution Satellite Imagery and Lidar Data Using GEOBIA. In: D. A. Quattrochi, E. Wentz, N. S.-N. Lam and C. W. Emerson (eds), *Integrating Scale in Remote Sensing and GIS*, CRC Press, pp. 283–302.
- Dawson, J. and Woods, A., 2010. ITRF to GDA94 coordinate transformations. *Journal of Applied Geodesy* 4(4), pp. 189–199.

- Ding, Z., Peng, Y. and Pan, R., 2006. BayesOWL: Uncertainty Modeling in Semantic Web Ontologies. In: Z. Ma (ed.), *Soft Computing in Ontologies and Semantic Web*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–29.
- Donnelly, N., Crook, C., Haasdyk, J., Harrison, C., Rizos, C., Roberts, C. and Stanaway, R., 2014. Dynamic datum transformations in Australia and New Zealand. *Proceedings of Research at Locate'14* 1142, pp. 48–59.
- East, B., Kenley, R. and Harfield, T., 2016a. CONie: Lessons to be Learned from COBie Specifications and Standards. Technical report, Sustainable Built Environment National Research Centre (SBEncr).
- East, B., Kenley, R. and Harfield, T., 2016b. Network Asset Management Intervention: Designing CONie (Construction to Operations for Networks information exchange) for Australian and New Zealand Roads. Technical report, Sustainable Built Environment National Research Centre (SBEncr).
- East, E. W., 2006. Construction Operation Building Information Exchange. *USACE ERDC*.
- Gelder, J., 2018. Asset Data Harmonisation Stage III: BIM IFC Alignment Review. Technical report, Austroads.
- Gu, H., Li, H., Yan, L., Liu, Z., Blaschke, T. and Soergel, U., 2017. An Object-Based Semantic Classification Method for High Resolution Remote Sensing Imagery Using Ontology. *Remote Sensing* 9(4), pp. 329–21.
- Hauert, J., 2005. Link based exchange of information between geographic datasets. pp. 1–7.
- ICSM, 2018. Geocentric Datum of Australia 2020 Technical Manual. Intergovernmental Committee on Surveying and Mapping.
- Kenley, R. and Harfield, T., 2015. *New Project Management Models: Productivity Improvement for Infrastructure*. Sustainable Built Environment National Research Centre (SBEncr).
- Kenley, R. and Seppänen, O., 2009. *Location-Based Management for Construction: Planning, scheduling and control*. Planning, scheduling and control, Spon Press.
- Kim, I.-H., Feng, C.-C. and Wang, Y.-C., 2016. A simplified linear feature matching method using decision tree analysis, weighted linear directional mean, and topological relationships. *International Journal of Geographical Information Science* 31(5), pp. 1042–1060.
- Laurini, R., 2014. A conceptual framework for geographic knowledge engineering. *Special Issue on Multimedia Information Processing* 25(1), pp. 2–19.
- Longley, P. A., Goodchild, M. F., Maquire, D. J. and Rhine, D. W., 2005. *Geographical Information Systems and Science*. Wiley.
- McMeekin, D. A., 2015. Ontologies, Vocabularies and Various Tools Report. Technical report, CRC for Spatial Information.
- MRWA, 2015. Keep left, Keep right, Stop, Give Way and Divisional Marker Signs in WA. Main Roads Western Australia, http://www.mainroads.wa.gov.au/BUILDINGROADS/STANDARDSTECHNICAL/ROADANDTRAFFICENGINEERING/TRAFFICMANAGEMENT/Pages/Keep_left,_Stop,_Give_Way_and_Divisional_Marker_Signs_in_WA.aspx (4 May 2017).
- MRWA, 2017. Shapefile. Signs Regulatory (MRWA-503), data.gov.au, <http://catalogue.data.wa.gov.au/dataset/signs-regulatory-mrwa-503/resource/9b6ef363-8ffc0-3785-b3ef-82c52f97d11c> (6 September 2017).
- MTL, 2018. Movable Type Scripts. Movable Type Ltd. Calculate distance, bearing and more between Latitude/Longitude points, <https://www.movable-type.co.uk/scripts/latlong.html> (28 April 2018).
- Niestroj, M. G., McMeekin, D. A. and Helmholtz, P., 2018. Overview of standards towards road asset information exchange. *ISPRS Technical Commission IV Symposium (Volume XL-4) (accepted)*.
- Popovich, V. V., Claramunt, C., Schrenk, M. and Korolenko, K. V. (eds), 2009. Concept for Corporate Information Systems Development Based on GIS Technologies. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Ruiz-Alarcon-Quintero, C., 2016. Harmonization of Transport Data Sources According to INSPIRE Data Specification on Transport Networks. *Efficient, Safe and Intelligent Transport. Selected papers from the XII Conference on Transport Engineering, Valencia (Spain) 7-9 June*. 18, pp. 320–327.
- Sequeda, J. F., Tirmizi, S. H., Corcho, O. and Miranker, D. P., 2011. Survey of directly mapping SQL databases to the Semantic Web. *The Knowledge Engineering Review* 26(4), pp. 445–486.
- Seth, S. and Samal, A., 2017. Conflation of Features. In: S. Shekhar, H. Xiong and X. Zhou (eds), *Encyclopedia of GIS*, Springer International Publishing, Cham, pp. 291–297.
- Song, W., Haitcoat, T. L. and Keller, J. M., 2006. A Snake-based Approach for TIGER Road Data Conflation. *Cartography and Geographic Information Science* 33(4), pp. 287–298.
- Song, W., Keller, J. M., Haitcoat, T. L. and Davis, C. H., 2009. Automated Geospatial Conflation of Vector Road Maps to High Resolution Imagery. *IEEE Transactions on Image Processing* 18(2), pp. 388–400.
- Tong, X., Shi, W. and Deng, S., 2009. A probability-based multi-measure feature matching method in map conflation. *International Journal of Remote Sensing* 30(20), pp. 5453–5472.
- Tóth, K., Portele, C., Illert, A., Lutz, M. and de Lima, M. N., 2012. A Conceptual Model for Developing Interoperability Specifications in Spatial Data Infrastructures. Technical report, INSPIRE.
- W3C, 2017. Tools. Semantic Web Development Tools, World Wide Web Consortium (W3C), <http://www.w3.org/2001/sw/wiki/Tools> (16 November 2017).
- Yu, F., McMeekin, D. A., Arnold, L. and West, G., 2018. Semantic Web Technologies Automate Geospatial Data Conflation: Conflating Points of Interest Data for Emergency Response Services. In: P. Kiefer, H. Huang, N. Van de Weghe and M. Raubal (eds), *Progress in Location Based Services 2018*, Springer International Publishing, Cham, pp. 111–131.
- Zhang, M., Shi, W. and Meng, L., 2005. A generic matching algorithm for line networks of different resolutions. In: *Proceedings of the ICA Workshop on Generalisation and Multiple Representation*, Citeseer.