INTRODUCING A FRAMEWORK FOR CONFLATING ROAD NETWORK DATA WITH SEMANTIC WEB TECHNOLOGIES

Michael G. Niestroj[1,*], David A. McMeekin[2], Petra Helmholz[2]

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.Helmholz)@curtin.edu.au

ICWG IWIDF

ABSTRACT:

Road network asset management is a challenging task as many data sources with different road asset location accuracies are available. In Australia and New Zealand transport agencies are investigating into harmonisation of road asset data, whereby two or more data sets are merged to create a new data set. Currently, identifying relations between road assets of the same meaning is not always possible, as road authorities of these countries use their own data structures and standards. This paper employs Semantic Web Technologies, such as RDF/Turtle ontologies and semantic rules to enable road network conflation (merge multiple data sets without creating a new data set) as a first step towards data harmonisation by means of information exchange, and shifts road network data from intersections and road nodes to data sets considering the accuracy of the data sets in the selected area. The data integration from *GeoJSON* into RDF/Turtle files is processed with *Python*. A geographic coordinates shifting algorithm reads unique data entries that have been extracted from RDF/Turtle into JSON-LD and saves the processed data in their origin file format, so that a closed data flow can be approached.

1. INTRODUCTION

Road network data harmonisation is an ongoing activity within transport agencies in Australia and New Zealand (Austroads, 2018). In future, transport agencies of these countries are supposed to migrate into an unified data standard for road management and investment (Martin et al., 2019), however, this data standard is currently not used. With the use of Semantic Web Technologies, data sets from different data sources can be merged even if they do not share the same metadata specifiers, as long as the data sets use data values of the same meaning (Niknam , Karshenas).

Within the Semantic Web, data is available in the Resource Description Framework (RDF) format, and is represented as triples in the form of a subject, predicate and object (Berners-Lee et al., 2001). For instance, a triple regarding the road network can have as structure 'National Route 1' (subject), 'has road name' (predicate) 'Ennis Avenue' (object). The Semantic conceptualisation (called ontology) is a set of classes, data types, object properties, data properties and individuals (Gruber, 2009). As the ontology classification is built on simple triples, an ontology can be enhanced with the use of the Semantic Web Rule Language (SWRL) (Horrocks et al., 2004). An ontology for the road network has been developed by Varadharajulu et al. (2016) with the application of an automated decision making processes for spatially related transactions, such as the verification of a valid road name (whereby a road name of a new road cannot be used again within 10 km in city areas and 50 km in rural areas).

The proccess of combining two data sources (with information of the same meaning) into a single source is called conflation (Longley et al., 2005; Seth , Samal). Ruiz et al. (2011)

provide an overview of theoretical and practicable data conflation techniques, while differentiating conflation applications into process categorisation, geospatial database models and entities used. For example, Zhang et al. (2016) developed an automated road network conflation approach to transfer pedestrian ways from ATKIS into NAVTEQ with a conflation correctness of 99.35% for whole of Germany (about 360,000 km^2).

This paper presents the use of Semantic Web Technologies regarding a selected road network selection of about 0.2 km^2 (containing intersections, road nodes, roundabouts and signs) within Western Australia. The road network data is taken from different data sources, whereby each data source is represented by an own ontology. The origin data format of the road network data is *GeoJSON*. The *GeoJSON* data sets will be processed into RDF/Turle ontology files with a developed *Python* script, and further merged into a single ontology. Semantic relations and semantic rules will be integrated to conflate the road network data (e.g. a road is available in different data sets, a road sign is part of an intersection and same geographic coordinates are used by other data sources).

A further application of this paper is to shift geographic coordinates (longitude and latitude) from intersections and road nodes using a spatial processing algorithm developed with *Python*, so that the different data sets will share the same road network shape. The shifted coordinates of intersections are then further saved with the added information about the shifted distances in meters in the origin data source file format (*GeoJSON*). This approach allows that the data sets with the shifted coordinates can be tranformed into ontologies with the above described *Python* script.

The structure of this paper can be described as followed. The proposed data sources of this approach will be introduced next. The Semantic Web data model with ontologies and semantic rules of this method will be then explained with the help of a

---

*Corresponding author

flow chart. After that, results will be demonstrated related to the coordinates shifting approach and the developed ontology approach of this paper. At the end of this paper, this work will be concluded.

2. DATA

The data sets presented in this paper are based on quality controlled data sources generated by public authorities, such as Main Roads Western Australia (MRWA) (intersections, road signs, and road network data) and the Western Australian Land Information Authority (Landgate) (simplified roads [publicly available data set] and Shared Location Information Platform [SLIP] data that include rich metadata information). The Open-StreetMap (OSMAP) ontology in this paper uses OSMAP data sets (map lines [roads], multistrings [e.g. state routes] and map points [e.g. traffic signals and turning circles]) that are edited by the OSMAP community on a trusted base.

The high-resolution aerial images used to visually compare the coordinates of intersections and road nodes to the road network shape are based on high-resolution airborne geo-referenced images by the company Spookfish/EagleView.

3. METHOD

The data flow of the approach of this paper is shown in the flow chart in Figure 1. Data is collected from Landgate, MRWA and OSMAP. To enable the work within a specific area data sets will be initially loaded into the QGIS3 software. After selecting the road network, this selection will be extracted with QGIS3 into *GeoJSON* files. This has to be done interactivelly for each data set. The horizontal sections of the flow chart explained next.

The *Python* script (P1) processes road asset *GeoJSON* files and creates an RDF/Turtle ontology format for each source. Further details on the ontologies (O1–O4) are provided in Section 3.1.1. The 'MRWA.ttl' (O1), 'OSMAP.ttl' (O2) and 'Landgate.ttl' (O3) ontologies contain data individuals with separate entries for the location representation (points, lines and multilines). The created TTL files will then be integrated into their respective ontologies that are developed with Protégé and saved as OWL files. A further ontology 'Dataset.owl' (O4) is used as the main ontology with links to the ontologies (O1–O3) and semantic rules written in SWRL for the communication between these ontologies (see Section 3.1.2).

The complete ontology is then merged into one file and saved with Protégé as an JSON-LD (JSON for Linking Data) ontology file (see Section 3.2.2). The single JSON-LD file now contains the information of the previous described ontologies. The *Python* script (P2) reads the JSON-LD ontology and extracts all individuals. The benefit of extracting the individuals is that the resulting data set 'Dataset_Json-LD_individuals.json' will contain only the required information for the developed coordinates shifting algorithm of this paper.

The final work of this data flow is done with the *Python* script (P3) while shifting geographic coordinates from the MRWA ontology (intersections and road nodes) to the matching Landgate SLIP data road nodes. The outcome of this process is saved in the origin MRWA *GeoJSON* data structure with the additional metadata information about the shifted distances in meters. The

created *GeoJSON* files of the MRWA road network and intersections can be transfered with the *Python* script (P1) into RDF/Turtle ontologies, such as the initial MRWA data sets. This approach activates that the data flow is performed in a closed data circuit.

3.1 Semantic Web data model

The Semantic Web data model of this paper shows in Figure 2 that the ontology is subdivided into four ontologies OSMAP, MRWA, Landgate and conflated data. Each ontology shares the classes features and road network. The features' classes include information about the location of each individual and contain, depending on the data source, the information about point coordinates, linestrings and multilinestrings. Each data source has its own data structure depending on the provided information by the given data set and, therefore, the road network class has a different structure in each ontology.

3.1.1 Ontologies This section gives an overview of the developed ontology classes. The content of the classes is entered by semantic rules and ontology relations and are explained later. The ontology classes have been developed for the used Landgate, MRWA, OSMAP and conflated data ontologies.

MRWA (O1)
The road network class of the MRWA ontology class structure (see Figure 3) is categorised into: inventory (e.g. signs), intersections, roads, road nodes and signs regulatory (e.g all signs, stop and give way). The features class contains the geometry of each MRWA individual and is grouped into linestrings and point coordinates. Linestrings are used to represent the location of road nodes. Point coordinates are used for the location determination of intersections, signs and the point coordinates extracted from linestrings. The ontology further contains a location class to list regions and towns.

Validation rules have been implemented to evaluate a valid/invalid position of a give way and stop signs, such as a stop sign must be within 15.0 m to the corresponding line, as well as between 0.6 and 5.0 m to the carriageway. However, the MRWA provided information about the location of signs is not accuratate, as for MRWA it is sufficient to know at which intersection a sign is located but not its exact location. Furthermore, data sets of carriageways and stop sign line markings are usually not available within the data set, therefore, this paper will not include the road signs validation rules in greater detail.

OSMAP (O2)
The road network class of the OSMAP ontology class structure (see Figure 4) contains information about cycle ways, traffic signals, roundabouts, tertiary roads, secondary roads, service roads and others. The OSMAP data set has richer information compared to the MRWA and Landgate data sets, due to the fact that OSMAP features cannot be selected before downloading the data sets. In addition to road network data, the OSMAP data set provides information about captured buildings that are not availabe within the available MRWA and Landgate data sets, such as shops, kindergartens, police stations and fast food restaurants but not used in this research. The location of elements from the OSMAP data sets is described with the use of multilinestrings, linestrings and point coordinates.

Landgate (O3)
The Landgate ontology class structure (see Figure 5) has categorised the road network into roads, roundabouts, connectors

Action / Data sets	Dataset 1: Main Roads Western Australia	Dataset 2: OpenStreetMap	Dataset 3: Landgate
	- MRWA_Intersections.geojson - MRWA_Signs.geojson - MRWA_Road_Network.geojson	- MapLines.geojson - Multistring.geojson - MapPoints.geojson	- RoadSimplified.geojson - SlipDataset.geojson
Creates RDF/Turtle ontology of GeoJSON (P1) see Section 3.2.1	**Ontology (O1)** see Section 3.1.1 Create MRWA.ttl -> import into MRWA.owl	**Ontology (O2)** see Section 3.1.1 Create OSMAP.ttl -> import into OSMAP.owl	**Ontology (O3)** see Section 3.1.1 Create Landgate.ttl -> import into Landgate.owl
4th ontology	**Ontology (O4)** see Section 3.1.1 Dataset.owl - Including Semantic rules for the comparison of the ontologies (O1-O4). see Section 3.1.2		
Merge ontology	**Merge (O1) (O2) (O3) into (O4)** Dataset.owl (single file used for further processing)		
Save ontology **Output JSON-LD**	Use Protégé to save RDF/Turtle ontology as JSON-LD - > Dataset.json (renamed from OWL to JSON) The JSON-LD file can be read in a further process with a Python script for the processing of the coordinates shifting algorithm.		
Extract individuals (P2) see Section 3.2.2	Generate a JSON file of the JSON-LD ontology that contains only individuals and coordinates, as the JSON-LD document includes the whole ontology information. - File 'Dataset_JsonLD_individuals.json'		

Shift coordinates and create MRWA GeoJSON files (P3)
see Section 3.2.3
	Read the JSON file 'Dataset_JsonLD_individuals.json' that contains all individuals.
	
	- Shift MRWA intersections to Landgate Slip data road nodes
	- Shift MRWA road nodes to Landgate Slip data road nodes
	➢ Write results as GeoJSON files.
	  ▪ MRWA_intersections.geojson
	  ▪ MRWA_Road_Network.geojson.
	  The generated GeoJSON files are in the same format as the origin MRWA GeoJSON files with the difference that the coordinates are shifted. The metadata information about the shifted distances in meters will be added. The created MRWA GeoJSON files with the shifted coordinates will be be used to create the shifted RDF/Turtle 'MRWA.ttl' document.		

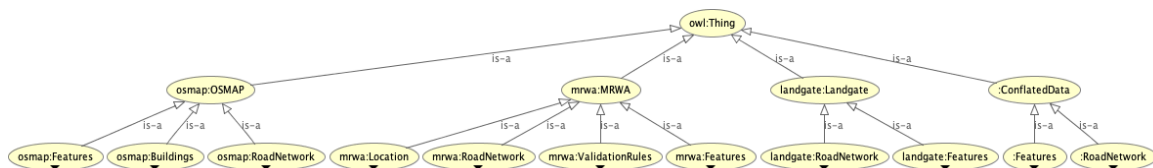Figure 1. Workflow overview of the approach of this paper.

Figure 2. The ontology model is subdivided into the ontologies OSMAP, MRWA, Landgate and conflated data. Each ontology has has an individual structure and the illustration shows the first subclass of each ontology.
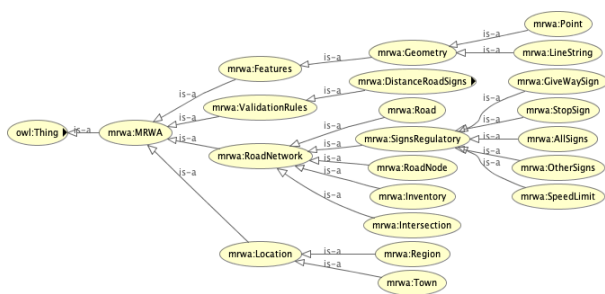

Figure 3. Representation of the MRWA ontology.

Conflated Data (O4)

The conflated data ontology class structure (see Figure 6) describe elements (e.g. intersections, road nodes and roundabouts) that have been determined in the MRWA, Landgate and OSMAP data sets. The intersection class lists elements that are connected to an intersection, and the road class lists road nodes. The geometry class contain multilinestrings, linestrings and point coordinates that are used by all of the data sets.

3.1.2 Semantic Rules   A collection of 88 semantic rules have been developed to integrate the data sets in the above described ontology classes and to let these different data sets from different data sources communicate with each other. In the scope of this paper not every single Semantic rule is explained, rather, this paper will focus on a set of seven selected rules indicated in Source Code 1. The explanation of these semantic rules is indicated in the following list:

1. Allocate an MRWA road node to the Landgate SLIP data set. The connection is done by comparing the MRWA 'ROAD' attribute with the Landgate 'mrwaroadnumber' attribute. If both values are equal, then the road data of both data sources refers to the same road.

(nodes within roundabouts), as well as tracks. The Landgate data sets use multilinestrings to represent the location. However, a multilinestring consists of at least one linestring, and one linestring has at least two point coordinates. The Landgate simplified and SLIP data sets use the same data structure, so that the data set can be categorised into the described classes. The difference between these two data sets is that the simplified data set uses a simplified road network representation (e.g. a road is described by less road nodes) and a minimised metadata collection compared to the Landgate SLIP data set.

```
1. mrwa:RoadNode(?a)
     ^ mrwa:ROAD(?a, ?aValue)
     ^ swrlb:notEqual(?aValue, "")
     ^ landgate:Road(?b)
     ^ landgate:mrwaroadnumber(?b, ?bValue)
     ^ swrlb:notEqual(?bValue, "")
     ^ swrlb:contains(?aValue, ?bValue)
     -> dataset:isSameRoadAs(?a, ?b)

2. landgate:isPointCoordinatesOf(?a, ?aValue)
     -> dataset:PointCoordinates(?a)

3. landgate:Connector(?a)
     ^ landgate:MultilineString(?b)
     ^ landgate:hasMultilineCoordinates(?a, ?b)
     ^ landgate:hasLineCoordinates(?b, ?c)
     ^ landgate:LineString(?c)
     ^ landgate:Point(?d)
     ^ landgate:hasPointCoordinates(?c, ?d)
     ^ landgate:LATITUDE(?d, ?dLat)
     ^ landgate:LONGITUDE(?d, ?dLong)
     ^ mrwa:Intersection(?a1)
     ^ mrwa:Point(?b1)
     ^ mrwa:hasPointCoordinates(?a1, ?b1)
     ^ mrwa:LATITUDE(?b1, ?b1Lat)
     ^ mrwa:LONGITUDE(?b1, ?b1Long)
     ^ swrlb:subtract(?rangeLong, ?dLong, ?b1Long)
     ^ swrlb:abs(?absrangeLong, ?rangeLong)
     ^ swrlb:subtract(?rangeLat, ?dLat, ?b1Lat)
     ^ swrlb:abs(?absrangeLat, ?rangeLat)
     ^ swrlb:lessThanOrEqual(?absrangeLat,
          "0.00005"^^xsd:decimal)
     ^ swrlb:lessThanOrEqual(?absrangeLong,
          "0.00005"^^xsd:decimal)
     -> dataset:hasIntersectionPart(?a1, ?a)

4. mrwa:AllSigns>(?a)
     ^ mrwa:hasPointCoordinates(?a, ?b)
     ^ mrwa:Point(?b)
     ^ mrwa:LONGITUDE(?b, ?bLong)
     ^ mrwa:LATITUDE>(?b, ?bLat)
     ^ mrwa:Intersection(?a1)
     ^ mrwa:hasPointCoordinates(?a1, ?b1)
     ^ mrwa:Point(?b1)
     ^ mrwa:LONGITUDE(?b1, ?b1Long)
     ^ mrwa:LATITUDE(?b1, ?b1Lat)
     ^ swrlb:subtract(?rangeLong, ?bLong, ?b1Long)
     ^ swrlb:abs(?absrangeLong, ?rangeLong)
     ^ swrlb:subtract(?rangeLat, ?bLat, ?b1Lat)
     ^ swrlb:abs(?absrangeLat, ?rangeLat)
     ^ swrlb:lessThanOrEqual(?absrangeLong,
          "0.00025"^^xsd:decimal)
     ^ swrlb:lessThanOrEqual(?absrangeLat,
          "0.00025"^^xsd:decimal)
     -> dataset:hasIntersectionPart(?a1, ?a)

5. landgate:SPECIAL_PU(?a, ?b)
     ^ swrlb:stringEqualIgnoreCase(?b, "RO")
     -> landgate:Roundabout(?a)

6. osmap:Residential(?a)
     -> osmap:Roads(?a)

7. mrwa:RoadNode(?g)
     ^ mrwa:ROAD(?g, ?ss)
     ^ mrwa:AllSigns(?s)
     ^ mrwa:START_SLK(?g, ?kk)
     ^ mrwa:END_SLK(?g, ?ll)
     ^ mrwa:ROAD>(?s, ?ss)
     ^ mrwa:SLK(?s, ?ff)
     ^ swrlb:greaterThanOrEqual(?ff, ?kk)
     ^ swrlb:lessThanOrEqual(?ff, ?ll)
     -> mrwa:hasRoadNode(?s, ?g)
```

Source Code 1. Seven semantic rules written in SWRL.
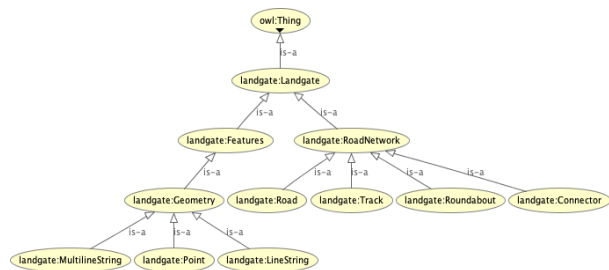

Figure 4. Representation of the OSMAP ontology.

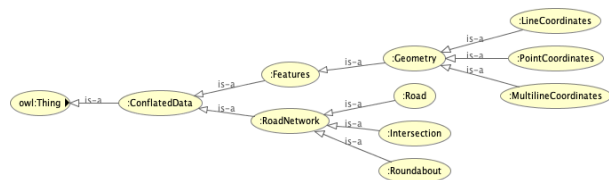Figure 5. Representation of the Landgate ontology.

Figure 6. Representation of the conflated data ontology.

2. When an individual of the Landgate data set has the attribute 'isPointCoordinatesOf', then it is a point coordinate of the ontology 'dataset'.

3. Connect a Landgate roundabout connector to an MRWA intersection. The semantic rules' condition is that a point coordinate of a Landgate connector must be within a range of 0.00005 longitude ($\sim$ 4.7 m in this area) and 0.00005 latitude ($\sim$ 5.5 m in this area) of an intersection to be part of it. The exact measurement in meters is not required as it will most likely not occur that a second intersection will apear in such a small range.

4. Connect an MRWA sign to an MRWA intersection. The semantic rules' condition is that a point coordinate of an MRWA sign must be within a range of 0.00025 longitude ($\sim$ 23.73 m in this area) and 0.00025 latitude ($\sim$ 27.68 m in this area) of an intersection to be part of it. The exact

```
### http://www.example.org/CONie/MRWA/PointCoordinates_115.716
51703_-31.69076300
mrwa:PointCoordinates_115.71651703_-31.69076300 rdf:type
    owl:NamedIndividual ;
mrwa:LATITUDE -31.69076300 ;
mrwa:LONGITUDE 115.71651703 ;
mrwa:CRSIDENTIFIER "urn:ogc:def:crs:OGC:1.3:CRS84" ;
mrwa:coordinates "[115.71651703, -31.69076300]" .


### http://www.example.org/CONie/MRWA/MRWA_Sign_Stop_15294395
mrwa:MRWA_Sign_Stop_15294395 rdf:type owl:NamedIndividual ,
owl:Thing ;
mrwa:hasInventory mrwa:Sign ;
mrwa:hasPointCoordinates
    mrwa:PointCoordinates_115.71651703_-31.69076300 ;
mrwa:hasRoad mrwa:MRWA_Road_PensacolaTce_1104583 ;
mrwa:ROAD_NAME_DECODE "Pensacola Terrace" ;
mrwa:OBJECTID 15294395 ;
mrwa:ROAD "1104583" ;
mrwa:ROAD_NAME "Pensacola Tce" ;
mrwa:COMMON_USAGE_NAME "Pensacola Tce" ;
mrwa:SLK 0.137 ;
mrwa:CWY "Single" ;
mrwa:TRUE_DIST 0.137 ;
mrwa:NETWORK_TYPE "Local Road" ;
mrwa:RA_NO "07" ;
mrwa:RA_NAME "Metropolitan" ;
mrwa:LG_NO "110" ;
mrwa:LG_NAME "Wanneroo (C)" ;
mrwa:XSP "Verge" ;
mrwa:PANEL_COUNT 1 ;
mrwa:PANEL_01_DESIGN "R1-1B" ;
mrwa:PANEL_01_DESIGN_MEANING "STOP (750 X 750)" ;
mrwa:LATITUDE -31.690763 ;
mrwa:LONGITUDE 115.716517 ;
mrwa:REGULATORY_SIGN_TYPE "Stop" ;
mrwa:ROUTE_NE_ID 16518969 .
```

Source Code 2. Example of an MRWA point coordinate and a STOP sign as RDF/Turtle format created by *Python* script (P1).

measurement in meters is not required as it will most likely not occur that a second intersection will apear in such a small range.

5. A Landgate roundabout will be classified, if the Landgate attribute 'SPECIAL_PU' has the value 'RO'.

6. An OSMAP residential road is an OSMAP road.

7. Allocate an MRWA road sign to an MRWA road node. An MRWA road node is identified by the attribute 'ROAD' and a start/end 'SLK' (straight line kilometres). If an MRWA regulatory sign shares the same road and its 'SLK' is between the start/end 'SLK' of the road node, then the MRWA sign is part of the road node.

The SWRL rules described in this section use prefixes instead of the URI. The prefixes are mrwa (MRWA), landgate (Landgate), osmap and dataset (conflated data) to refer to the ontology.

3.2 Data Processing

3.2.1 Create ontologies from GeoJSON (P1)
The *Python* scr-ipt (P1) is employed to create the Landgate, MRWA and OSMAP ontology individuals as RDF/Turtle files. This script includes features that are indicated in Table 1 to improve the result of the ontology, such as create additional individuals (e.g. roads, regions and towns), add data properties (e.g. decode road abbreviations) and allocate object properties (e.g. determine to which road node a road node is connected). An example of a processed MRWA point coordinate and STOP sign as RD-F/Turtle format is shown in Source Code 2. For the selected road network of this paper the generated RDF/Turtle files can be summarised by counting the created features as followed:

- MRWA.ttl: 23 intersections, 31 line coordinates, 109 point coordinates, 980 data properties, 1 region, 1 town, 12 roads, 31 road nodes, 7 signs, 96 is connected to, 38 has road, 57 has road node, 7 has inventory and 99 road names decoded.
- Landgate.ttl: 93 multiline coordinates, 93 line coordinates, 355 point coordinates, 2892 properties, 34 simplified roads, 74 SLIP roads, 116 is group of connectors, 51 is connected to, 506 is connected to node, 36 is part of same roundabout, 40 has road, 110 road names decoded and 76 road types decoded.
- OSMAP.ttl: 17 line coordinates, 129 point coordinates, 71 properties, 17 road nodes and 42 is connected to.

It has to be noted that the Landgate simplified and the Landgate SLIP data sets both describe the same road network (containing roads, connectors and roundabouts) in the 'Landgate.ttl' ontology and, therefore, the evaluation count is larger (e.g. for coordinates) compared to the 'MRWA.ttl' and 'OSMAP.ttl' files.

3.2.2 Extract individuals from ontology (P2)
The *Pyhton* script (P2) creates a data structure of the Landgate, MRWA and OSMAP ontologies that include only individuals (geographic coordinates are also represented as individuals), and saves the result as a JSON file.

Before this step can be executed, the ontologies will be merged as one ontology file to allow simpler data handling. After that, the single file ontology will be converted from the RDF/Turtle format into the the JSON-LD file format. This allows that data will be readable with *Python* with a JSON parser. The JSON-LD file now contains the whole ontology, including semantic rules. However, the in the RDF/Turtle format used prefixes (e.g. landgate, mrwa, osmap, swrl, owl, rdfs and xsd) are no longer available after the conversion to JSON-LD. The prefixes will be interactively added so that the JSON-LD file will use the same prefixes as the initial RDF/Turtle document, such as 'owl:' for 'http://www.w3.org/2002/07/owl#' and 'landgate:' for 'http://ww-w.example.org/CONie/LANDGATE/').

After running (P2) and writing the results in the JSON document 'Dataset_JsonLD_individuals.json', the created data structure covers in the first data layer individuals, point coordinates, line coordinates and multiline coordinates as seen in Figure 7. The second layer includes specific information regarding the data sources (Landgate, MRWA and OSMAP). A third layer is used to group the individuals into road asset types, such as intersections, residential, road nodes, roundabouts, signs and other. With such a data structure, all individuals can be represented in an ordered way with respect to the different data sources.

3.2.3 Shift geographic coordinates and create GeoJSON (P3)
The *Python* script (P3) can be described as a 4 stage approach as visualised in the flow chart in Figure 8. In stage 1, the *Python* script (P2) created file 'Dataset_JsonLD_individuals.json' will be parsed and converted to a multi-level dictionary. A data structure will be created in stage 2 that contains all available road names in the first layer. The MRWA data sets (intersections and road nodes), as well as the Landgate SLIP data sets (connectors and road nodes) will be then added to the data structure and classified to the respective road names.

The developed geographic coordinates shifting algorithm is part of stage 3 of the *Python* script (P3) flow chart. MRWA intersections will be shifted to Landgate SLIP connectors at roundabouts, otherwise to Landgate road nodes, and the MRWA point

Table 1. Representation of the *Python* script (P1) functionality for the MRWA, Landgate and OSMAP data sets. This script creates individuals, writes data properties and allocates object properties. The results are written as RDF/Turtle files for each data source.

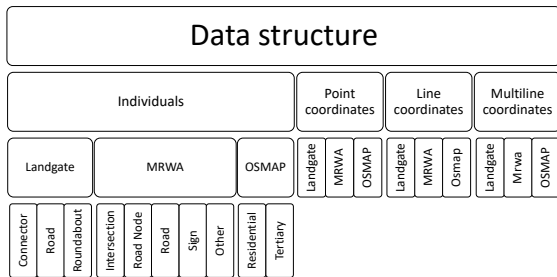| Data set / data processing	Create Individuals		Data Properties				Object Properties									
	Extract Road	Extract Town / Region	Road Type Decoded	Road Name Decoded	Road Name From / Middle / To + Decoded	Properties	has Point Coordinates	has Line Coordinates	has Multiline Coordinates	has Road	has Road Node	has Inventory	is Group OfConnectors	is Connected To	is Connected ToNode	isPart OfSame Roundabout
Signs_Regulatory_MRWA.geojson	X	X	X			X	X			X		X				
Road_Network_MRWA.geojson		X	X			X	X	X		X				X		
Intersections_MRWA.geojson					X	X	X				X					
Landgate_Roads_Simplified.geojson				X		X	X	X	X				X	X	X	
Landgate_Slip_Dataset.geojson			X	X		X	X	X	X	X			X	X	X	X
OSM_Map_Lines.geojson						X	X	X						X		
OSM_Map_Points.geojson						X	X							X		
OSM_Map_Multistring.geojson						X	X	X	X							

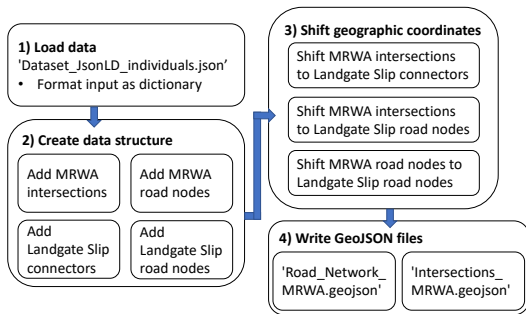Figure 7. Data structure generated with script (P2).

Figure 8. The approach of the *Python* script (P3) is indicated in four stages (load data, create data structure, shift geographic coordinates and write *GeoJSON* files).

coordinates of road nodes will be shifted to align with the pattern of the Landgate SLIP data set. Details of the developed geographic coordinates shifting algorithm is explained as followed:

1. If the coordinates of a Landgate connector are within a distance of 5.5 m to the intersection, shift an MRWA intersection to the nearest coordinates of a Landgate SLIP connector.
2. If the coordinates of a Landgate road nodes are within a distance of 5.5 m to the intersection, shift remaining shifted MRWA intersection to the nearest coordinates of a Landgate SLIP road node.
3. Shift an MRWA road node coordinates to an adjusted MRWA intersections.
4. Shift remaining MRWA road node coordinates to either a Landgate SLIP road node (if MRWA road node coordinates are within 5.5 m of Landgate SLIP road node coordinates) or shift it to the nearest calculated point of a

Landgate SLIP road node linestring. The key points 1–4 follow the shifting approach such as illustrated in Figure 9 a) and b).

5. If a road consists of two lanes, it is often represented by two linestrings as seen in Figure 9 b). The related MRWA intersection (indicated as yellow dot) will be then shifted between these two road nodes such as shown in Figure 9 c) to represent the geographic centre of an intersection. To determine the right intersection location, road nodes in a radius of 12 m of the intersection will be considered by the shifting algorithm.
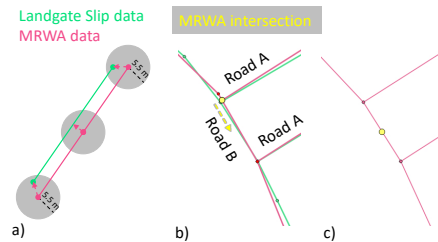

Figure 9. The shifting condition of MRWA geographic point coordinates is that a Landgate SLIP geographic point or line must be a) within 5 meters as indicated. If a road has two lanes as in b), then the related intersection will be shifted to the centre of the intersection as in c).

The results of the shifting algorithm will be written into *GeoJSON* files in stage 4. Therefore, the shifted MRWA intersections will be written in the file 'Intersections_MRWA.geojson' in the same format as the origin MRWA intersections data set, with the added information about the shifted distances in meters. The additional information of the shifted distances can be used to highlight the intersections respective to the length of the distance in different colours (e.g. green for a small shift, orange for an average shift and red for a large shift) and used for interactive quality control inspection.

The shifted MRWA road nodes will be written into the file 'Road_Network_MRWA.geojson' in the same format as the MRWA initial data set, with the added information about the shifted distances of each vertex in meters. This shifted distances are not further processed. The reason is that a linestring consists of multiple geographic points (at least two point coordinates) that are shifted individually and, therefore, currently a visualisation of the shifted geographic coordinates is not possible with the employed Open Source Geographic Information System (QGIS3).

4. RESULTS

4.1 Intersections and Road Nodes

The initial road network with the Landgate SLIP (green lines and vertex points) and MRWA (red lines and vertex points, and white intersections) data sets is displayed in Figure 10. It has been chosen to shift the MRWA road network to the Landgate road network as the Landgate SLIP data set is considered to be of higher quality, due to the fact that the Landgate data sets generally have a better representation of the road centreline.

Figure 10. Initial data set representation of the Landgate SLIP and MRWA road networks.

Figure 11. Result of the shifted MRWA road network.

The result of the shifting algorithm of MRWA intersections to Landgate SLIP data road nodes for the selected road network is indicated in Figure 11. The green, orange and red dots are used to visualise the shifted distances in meters. The intersections use the following colour filters: green (less or equal 2 m),

orange (greater 2 m and less or equal 5 m) and red (greater 5 m). One can see, that in most cases the intersection's shift takes place within 2 m. However, a shift between 2 and 5 m occurs three times in this area.

While exploring these intersections by visual inspection, it can be asserted that the new intersections better represent the centre of an intersection. In one instance an intersection has been shifted by more than 5 m due to the fact that an MRWA intersection is located between two road nodes, and the MRWA authority positioned the intersection near to one road node.

The final visualisation of the developed shifting algorithm of the MRWA road network (road nodes and intersections) is shown in Figure 11. The outcome enables a better displayed road network compared to the initial MRWA data set, as intersections are now centred, and road nodes are generally no longer outside of the pavement border. However, the representation of the road centreline is still not perfect while shifting to the Landgate SLIP data set. Therefore, this algorithm has the potential to benefit of an additional road map analysis, such as the identification of road centreline using aerial images.

4.2 Semantic Road Network

A reasoned ontology of an MRWA intersection of the selected road network is indicated in Figure 12. As one can see, there is sample information accessible that is mainly available due to evaluating geographic coordinates of surrounding individuals with SWRL rules. The Protégé property assertions view provides information about individuals and relations to other individuals. The first four rows of the object property assertions are in this example generated by the Python script (P1), and shows the connected point coordinates and MRWA road nodes.
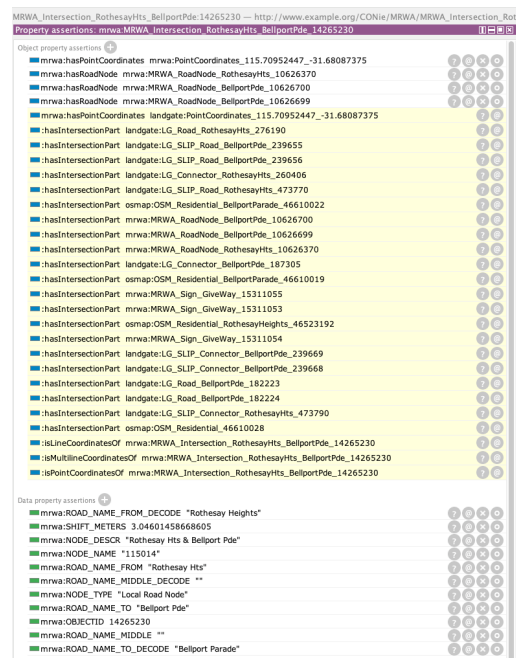

Figure 12. Property assertion view of an MRWA intersection after reasoning with *Pellet* in Protégé.

The yellow block indicates object property assertions that has been reasoned by *Pellet* (Java based OWL 2 reasoner) with respect to the developed SWRL rules. For instance, it has been

determined that a Landgate data set exists with same point coordinates, three Give Way signs are at the intersection, this intersection is a roundabout (connectors are only available at roundabouts) and OSMAP roads are connected.

The data property assertions (e.g. object id, road name, road type and road description) are provided by the MRWA data source. The Python script (P1) has delivered the information about roads that are connected to the intersection by separating the description at the '&' sign (see road name from/middle/to) and has further decoded road name abbreviations (e.g. 'Wy' into 'Way' and 'Pde' into 'Parade'). The attribute 'SHIFT_-METERS' indicates that this data set has been shifted by Python Script (P3).

5. CONCLUSIONS

The paper demonstrated that it is possible to conflate different road network data sets and sources using Semantic Web Technologies. The advantage of using these technologies is to implement data sets directly as such called Linked Data without creating a new data set (provided that the data source enables a Linked Data capable format, such as RDF/Turtle or JSON-LD). Data sets are converted from different data sources (e.g. Landgate, MRWA and OSMAP) using *Python* into RDF/Turtle files. An ontology for each data source has been developed. And a further ontology is used for the data conflation between the different data sources.

It has been detected that the Landgate data sets have a higher accurancy regarding the centre position of intersections and the representation of the road centrelines. Therefore, investigation was carried out into a geographic coordinates shifting algorithm to shift MRWA intersections and road nodes to Landgate road nodes. The shifted coordinates are then saved in the origin MRWA *GeoJSON* format with all metadata information retained. A metadata information is added to the MRWA intersections *GeoJSON* file to inform about the shifted distances in meters. The created data sets are evaluated, as well as displayed with QGIS3, and transformed into RDF/Turtle ontologies. In the future the aim is to nominate one data set and to shift all others to this data set, as well as to use available metadata such as 'last update' or 'method of capturing' to introduce more compare rules.

The evaluation of the Semantic road network verified that surrounding objects can be determined with the use of SWRL. For instance, while considering an MRWA intersection the connected Landgate, MRWA and OSMAP road nodes are determined with the help of geographic calculations regarding distances in longitude and latitude. Currently, the framework was tested on a small road network.

In future developments, the framework will be tested on a larger area. The implementation of the Shapes Constraint Language (SHACL) to evaluate the road network is also a possible next contribution, provided that a working SHACL plugin will be available for Protégé.

Furthermore, the outcome of this paper can be used as a solid base for further contributions. Therefore, investigation can be done in multiple areas, such as ontology based support of IFC (Industry Foundation Classes) and OGC (Open Geospation Consortium) standards to unify metadata definitions, as well as the improvement of the road centreline representation with involving geographic referenced aerial images.

References

Austroads, 2018. Austroads 2017-18 Annual Report. 1–80.

Berners-Lee, Tim, Hendler, James, Lassila, Ora, 2001. The semantic web. *Scientific american*, 284, 34–43.

Gruber, Tom, 2009. *Ontology*. Springer US, Boston, MA, 1963–1965.

Horrocks, Ian, Patel-Schneider, Peter F, Boley, Harold, Tabet, Said, Grosof, Benjamin, Dean, Mike et al., 2004. SWRL: A semantic web rule language combining OWL and RuleML. *W3C Member submission*, 21, 79.

Longley, Paul A., Goodchild, Michael F., Maquire, David J., Rhine, David W., 2005. *Geographical Information Systems and Science*. Wiley.

Martin, T, Davey, S, Clarke, M, 2019. Data standard for road management and investment in australia and new zealand. Technical report.

Niknam, Mehrdad, Karshenas, Saeed, 2017. A shared ontology approach to semantic representation of BIM data. *Automation in Construction*, 80, 22–36.

Ruiz, Juan J, Ariza, F Javier, Ureña, Manuel A, Blázquez, Elidia B, 2011. Digital map conflation: a review of the process and a proposal for classification. *International Journal of Geographical Information Science*, 25, 1439–1466.

Seth, Sharad, Samal, Ashok, 2017. Conflation of Features. S. Shekhar, H. Xiong, X. Zhou (eds), *Encyclopedia of GIS*, Springer International Publishing, Cham, 291–297.

Varadharajulu, Premalatha, Arnold, Lesley, McMeekin, David A, West, Geoff, Moncrieff, Simon, 2016. Swrl rule development to automate spatial transactions in government. *International Conference on Geographical Information Systems Theory, Applications and Management*, Springer, 122–142.

Zhang, Meng, Yao, Wei, Meng, Liqiu, 2016. Automatic and Accurate Conflation of Different Road-Network Vector Data towards Multi-Modal Navigation. *ISPRS International Journal of Geo-Information*, 5, 68–17.