Semantic Web Technologies Automate Geospatial Data Conflation: Conflating Points of Interest Data for Emergency Response Services

Feiyan Yu, David A. McMeekin, Lesley Arnold and Geoff West

Abstract Conflating multiple geospatial data sets into a single dataset is challenging. It requires resolving spatial and aspatial attribute conflicts between source data sets so the best value can be retained and duplicate features removed. Domain experts are able to conflate data using manual comparison techniques, but the task it is labour intensive when dealing with large data sets. This paper demonstrates how semantic technologies can be used to automate the geospatial data conflation process by showcasing how three Points of Interest (POI) data sets can be conflated into a single data set. First, an ontology is generated based on a multipurpose POI data model. Then the disparate source formats are transformed into the RDF format and linked to the designed POI Ontology during the conversion. When doing format transformations, SWRL rules take advantage of the relationships specified in the ontology to convert attribute data from different schemas to the same attribute granularity level. Finally, a chain of SWRL rules are used to replicate human logic and reasoning in the filtering process to find matched POIs and in the reasoning process to automatically make decisions where there is a conflict between attribute values. A conflated POI dataset reduces duplicates and improves the accuracy and confidence of POIs thus increasing the ability of emergency services agencies to respond quickly and correctly to emergency callouts where times are critical.

F. Yu (✉) · D. A. McMeekin · L. Arnold · G. West
Department of Spatial Sciences, Curtin University, Perth, WA 6845, Australia
e-mail: feiyan.yu@postgrad.curtin.edu.au

D. A. McMeekin
e-mail: D.Mcmeekin@curtin.edu.au

L. Arnold
e-mail: arnolds@iinet.net.au

G. West
e-mail: G.West@curtin.edu.au

D. A. McMeekin · L. Arnold
Cooperative Research Centre for Spatial Information, Carlton, Australia

P. Kiefer et al. (eds.), *Progress in Location Based Services 2018*, Lecture Notes
in Geoinformation and Cartography, https://doi.org/10.1007/978-3-319-71470-7_6

Keywords Geospatial • Points of interest • POI • Data conflation
Semantic web • Ontology • RDF • SWRL • Rules • Reasoning

1 Introduction

Open Linked Data and Semantic Web technologies have been accepted widely by the geospatial industry in the recent decade (Parekh et al. 2004; Patrick and Sven 2009; Janowicz et al. 2010; Zhang et al. 2013; Wiemann and Bernard 2016). The Australian government has been working closely with W3C and OGC[1] to standardize information and technologies and promote best practice in the management and use of spatial data on the web.[2] Australia has established its own government linked data working group (AGLDWG)[3] to develop government standards and set up Linked Data implementation techniques in response to its citizens and agencies' needs. More recently, the Australian and New Zealand Cooperative Research Centre for Spatial Information (CRCSI) published a white paper (Duckham et al. 2017) to propose moving traditional Spatial Data Infrastructures to a Next Generation Spatial Knowledge Infrastructure (SKI) which can automatically create, share, curate, deliver and use data or information, as well as knowledge creation to support decision making. Semantic Web technologies were identified as an essential element to support the SKI in connecting, integrating and analyzing data.

To be able to appreciate the benefit of data versatility as highlighted in the SKI and embrace the advantages of Linked Data for knowledge acquisition, data conflation is an essential process for creating a single point of truth data set from interrelated data sources, so that knowledge can be more easily derived.

Currently, duplicate geospatial data collection and maintenance exists across Australian government agencies, leading to data management and processing inefficiencies. Existing conflation processes are primarily manual and more automated conflation techniques are required (Yu et al. 2016).

The uniqueness of this research is the use of a SWRL Rule-based Data Conflation Framework to automatically match and link corresponding entities between similar data sets and conflate these entities into a single dataset by selecting the most accurate features while also removing duplicates without the need for human intervention. The framework consists of four stages. Stage 1 is the creation of an ontology based on a multipurpose data model. The multipurpose data model is one that can be used by government agencies for various business purposes. Stage 2, refers to the conversion of disparate source data sets into the RDF (Resource Description Framework) format so they can link to the ontology during the conversion; and the development of SWRL rules to align attributes from the various

[1] http://www.opengeospatial.org/.

[2] https://www.w3.org/2015/spatial/wiki/Main_Page.

[3] http://linked.data.gov.au/index.html.

sources so they can be more readily compared and assessed in the latter stages of the conflation process. Stage 3 uses location proxy and other similarity measurements based on semantic descriptions to find matching candidates across data sets. Stage 4 uses a reasoning process to model how domain experts make decisions on which feature attribute values are the best or most accurate when they are considering various data sources.

In addition to the data sets to be conflated, SWRL rules reference other information and knowledge, such as building footprints data. The process is ordered sequentially according to the decision logic used by domain experts. This is an important step in the conflation methodology. Domain experts often refer to other data set(s) to compare attributes in candidate data sets, or look for information in the associated metadata to understand the level of accuracy of each source data set. In many cases, decisions are based on personal knowledge of an area and experience accumulated over time.

This paper explains the Data Conflation Framework and processes, and is organized as follows: Sect. 2 introduces the research background and related works. Section 3 presents the motivating example of conflating three government agencies' Points of Interest[4] (POI) data into a single authoritative for use in the emergency services response domain. Sections 4 and 5 demonstrate the implementation and evaluation of this research, respectively. The paper concludes with a summary of the research and describes a plan for future work.

2 Related Work and Background

It is well recognized in the spatial data domain that Lynch and Saalfeld (1985) were the first to make 'map conflation' a reality in 1985. Their approach to map conflation was to build a prototype using mathematical algorithms to perform geometric alignment between two vector datasets (e.g., census block boundary and road centerline map) (Saalfeld 1988; Kang 2001). This method is typically used to overlay and integrate map layers. The key is to correctly identify matched feature pairs from both base maps. They use the Delaunay triangulation algorithm to partition spaces based on data matches and a rubber-sheeting method to align datasets in each triangle. The process is repeated until all possible corresponding pairs are identified (Saalfeld 1988). Subsequent researchers have improved the efficiency of this method (Chen et al. 2004, 2006, 2008; Dongcai 2013).

However, as technology advances, ways to capture, store and present geospatial data have become more diverse. Geospatial data is recorded in more formats than traditional maps and the data required to support decision-making is often now distributed across the web. Over the past decades, researchers have made significant

[4] A wide-ranging definition of a Point of Interest (POI) is any feature or service that people wish to visit or know the location of, and is of value to the community (WALIS).

attempts to bring multiple interrelated geospatial data sets into the same data set to simplify analysis and create a unified view for better data visualization (Uitermark et al. 1999; Fonseca et al. 2002; Lutz et al. 2009; Zhang et al. 2013). The process is normally referred to as spatial data integration (Flowerdew 1991).

One barrier that has impeded spatial data integration is the heterogeneous nature of data. Data heterogeneity is classified into three categories: (1) syntactic heterogeneity, (2) schematic heterogeneity and (3) semantic heterogeneity (Bishr 1998). Syntactic heterogeneity is due to the use of different database systems (relational, object oriented etc.) and geometric representations (e.g., raster or vector representations). Schematic heterogeneity occurs when different data models are used to represent the same real world objects. Semantic heterogeneity arises when different disciplines or user groups have different interpretations for the same real world object. Naming heterogeneity is another form of semantic heterogeneity, such as the same real world object having multiple different names or the same name but referring to different real world objects. The heterogeneous nature of geospatial data makes it difficult to share and leads to data duplication problems.

A study by Lutz et al. (2009) shows that semantic heterogeneity can occur at the metadata level, schema level and data content level; each level blocks the discovery, retrieval, interpretation and integration of geographic information, respectively. They suggest ontologies as an appropriate mechanism to overcome these problems. Parekh et al. (2004) added semantics into metadata based on ontologies to improve geospatial interoperability efficiency and data discovery according to data content. Uitermark et al. (1999) developed a conceptual framework for ontology-based geographic data integration. Their work included generating domain ontology for certain disciplines, and application ontology for each geographic dataset. They also created abstraction rules to define the relationship between the concepts of domain ontology and application ontologies.

Based on the idea that concepts from different application ontologies are semantically similar if they refer to the same concepts or related concepts in the domain ontology, then corresponding object instances can be defined as semantically matched. Fonseca et al. (2002) proposed an ontology-driven geographic information system (ODGIS) in which ontologies are presented hierarchically with the Top-level Ontology at the highest level, Domain Ontology and Task Ontology at the middle level and Application Ontology at the bottom level. Their basic principle was to integrate what was possible and accept that some kinds of information will never be completely integrated due to their fundamentally different nature. They proposed that integration should always be done as the first point of intersection at the lowest level and then propagated upwards in the ontology tree.

As Semantic Web and Linked Data concepts become increasingly popular, more techniques have been studied in the geospatial integration process. There now exist ontologies designed to add semantics into the metadata through the Web Ontology Language (OWL) so computers can understand the meaning of the information and automatically operate actions on it (Parekh et al. 2004). Using the data integration system KARMA (Szekely et al. 2011; Zhang et al. 2013), geospatial data sets can be linked with design ontologies to transform various source formats into the RDF format so data being integrated can be published and reused with rich semantic descriptions on the Web. Zhang et al. (2013) also model integration steps using an ontology, so these processes can read RDF triples as input and also return results as RDF triples. As a result, the system is able to offer some meaningful match and link suggestions across data sets. A tool named FAGI-gis further explores semantic web technologies in the geospatial data domain (Giannopoulos et al. 2015). The input to the tool is two separate geospatial data sets converted to the RDF format and stored in PostGIS databases. SPARQL endpoints are used to pull linkages between entities from both data sets and their associated attributes. The tool uses Virtuoso as its RDF triple repository to store output and it supports GeoSPARQL[5] vocabularies so geospatial features are presented as GeoSPARQL WKT serialization and Basic Geo.

However, literature about spatial data integration has either focused on part of the integration processes, such as data discovery (Parekh et al. 2004), data retrieval (Walter and Fritsch 1999), data matching and linking separately (Sehgal et al. 2006; Wiegand and García 2007). Even when the processes have been studied as a whole, results only link the matched entities together and display all attribute values from each source (Zhang et al. 2013). The value conflicts between different sources for a same attribute haven't been resolved so the duplicate datasets still exist in silos.

There is more geospatial data conflation research required to combine overlapping geospatial data sources into a single source with richer attributes by reconciling conflicts and minimizing redundancy amongst source data sets while still retaining the best attributes from each source. Unlike traditional map conflation, once base maps for conflation are identified, much of the essential information required during the process is also known, such as, coordinate system, map scale, date created etc. So the conventional map conflation processes usually set the base map with higher geometry accuracy as the target map, then align each other map with the target map and transform attributes to the target map.

Contemporary spatial data conflation processes not only need to deal with all the difficulties associated with data integration, but furthermore to merge or fuse multiple data sets into a single data set. This involves decision making, such as "which data is most accurate?" and "which data is more up-to-date?" etc. However, the relevant information to support these kinds of decisions is usually vague.

Fusion can be further categorized. For example Szekely et al. (2011) merged point data with the latitude/longitude representing buildings or structures with address information from Yellow or White Pages. The connection between these datasets is the vector data attributed with street information. It uses latitude/longitude information for each vertex so it can calculate distance to point data. Having street names means it can compare with addresses extracted from Yellow or White Pages. Because each data set contains only one aspect of the real world object, the main challenge is finding matches. Once the nearest distance is identified and the name strings matched, the data sets can be fused. This method showcases

[5] The OGC GeoSPARQL standard supports representing and querying geospatial data on the Semantic Web. http://www.opengeospatial.org/standards/geosparql.

the ‘attribute enrichment’ aspect of data conflation, which involves combining the complementary properties.

The other part of the data conflation mission which is to resolve conflicts and reduce duplicates has not been well addressed. The work of Zhang et al. (2013) reduced data redundancy wherever attribute values from both data sets were exactly matched such as, exact name for a country/state or coordinates for a building. However, when the attribute value is different, the conflicts are not resolved. Instead they ‘union’ the attributes into a single list. Hence, there are multiple values for the same attribute in the resulting integrated list, such as two coordinate pairs representing the same building. The problem here is that two locations create confusion for a user when navigating to the building.

While matching and linking processes have been done semi-automatically or automatically using computer algorithms, the fusion process is difficult to automate with algorithms because it requires decision making not only to look at the data themselves but also requires reference to other information or knowledge. It is hard for the computer to do this because it needs domain expert’s knowledge and intervention.

The fusion process requires holistic information, human logic and the sequencing of logic into a set of reasoning steps. Data sources that enable holistic reasoning include but not limited to, reference data, business rules, metadata, provenance, topological relationships or even domain expert’s experience and knowledge stemming from years of work. The motivating example used in this research endeavors to replicate and sequence human logic through a series of automated reasoning steps and reference data sets to achieve a more holistic approach.

3 Motivating Example

The problem of duplication in the collection and management of spatial datasets is twofold. Firstly, duplication is costly for governments as it creates an unnecessary overhead in human and computing resources. Secondly, there is inconsistency between datasets meaning that the source of truth is not clearly understood and end-users may make decisions using incorrect or outdated information.

This is particularly a problem for emergency services. Incidents are often attended by more than one emergency service organization—ambulance, State and Federal police, fire and rescue, defense organisations and emergency volunteer associations. If each agency is using their own datasets there is a risk that information may be different leading to poor communication and coordination between first responders. For example, each organisation typically collects location data (points of interest), such as education institutions, pubs and clubs, pharmacies and civic places, to enable dispatch operations and incident management. However, these location features are often collected using different means, from distinct sources and at different times. The characteristics of these features are also recorded differently. Sometimes this is for unique and specific business purposes e.g., police record locations where licensed firearms are held, where restraining orders exist, and where violent behavior has occurred previously; whereas the fire department records the age and maintenance cycle of fire hydrants, location of arson and building floor plans. However, the more common reason why information is recorded differently is simply because there was no agreed standard for capturing and modeling information when these systems were first built.

Agencies are now coming to realise that collaborative data collection and shared resources is a more attractive alternative and one that makes incident management more effective. However, bringing multiple agency datasets together is problematic.

The data conflation case study used in this research is based on a project named LOC8WA, which was managed by Landgate (Western Australian Land Information Authority) in collaboration with WAPOL (Western Australian Police) and DFES (Department of Fire and Emergency Services). LOC8WA sought to conflate the POI data sets managed by each department into a single authoritative data set. The objective of LOC8WA was to improve the accuracy and confidence of emergency location information to increase the ability of emergency services to respond quickly and correctly to emergency callouts.

Identifying matched POIs across three datasets and conflating them into a single POI is a complex process. A scenario where all three POIs datasets related to a same region are combined is shown in Fig. 1. A point representing a shopping centre is highlighted inside a red circle. This point is from the Landgate data set and is represented by a small dot inside a building footprint. Whereas, the shopping centre is recorded in the DFES dataset as two red diamond shape points (within blue circles) located in a road intersection.

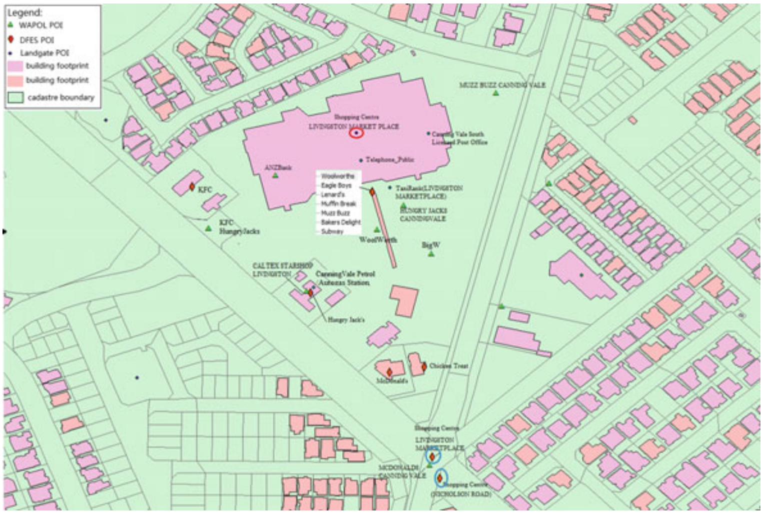


Fig. 1 POIs distributed around a shopping centre area

Noticeably, there are points inside the shopping centre with different categories such as supermarkets, bank branches and the post office. Around the shopping centre, there are other feature class points, bus stations, taxi ranks and fast-food outlets. The complexity or "confusion" in this situation is that some points are the same POI but their location is different. This is because they were sourced from different departments; or many POIs have the exact location but cannot be treated as the same POI as they have different names and attributes.

The LOC8WA project did not generate a conflated data set. Nonetheless, the importance of having an accurate POI data set for emergency services still remains and this has given rise to the importance of this research and the use of LOC8WA to case study automated conflation techniques using advanced semantic web technologies.

The amount of human effort required to complete the task was considered too great to correctly identify matches and make correct conflation decisions on a case-by-case basis. There are tens of thousands of POIs in total from these three agencies. Without the same ID to represent the same POI across agencies' data sets, the same POI's location varies from data set to data set, and there is no consistent naming convention. The research question is "How can it be known that the three points from the different data sets actually correspond to the same POI, which POI attributes (of each point) are the most correct and which points and attributes should be removed?".

4 Implementation

4.1 Stage 1: Ontology Development

Before ontology generation can be started, a fit for purpose output model should be defined which is able to satisfy multiple objectives and users. The data model represents the different models, each of which meets the business needs of each of the participating agencies. The choice of output model can affect the reasoning procedure design. For example, different models can define which data is ruled out and the final decision will consequently differ accordingly.

However, this research is not to define a completely new model from scratch; instead, the research will use existing models whenever possible (Yu et al. 2016). The LOC8WA project uses the Landgate's Points of Interest Data Model and participating agencies agreed that this model suited their business purposes. It was therefore adopted as the multipurpose mode for this study. The POI Ontology developed in this research is based on the Landgate data model and associated data dictionary. The POI ontology has potential to be adopted as a standard for all WA government agencies.

The essential knowledge in the data model was extracted and is shown in Fig. 2. It shows the classification system for the POIs which complies with a three-level hierarchy where red, blue and grey rectangles represent feature classes, feature

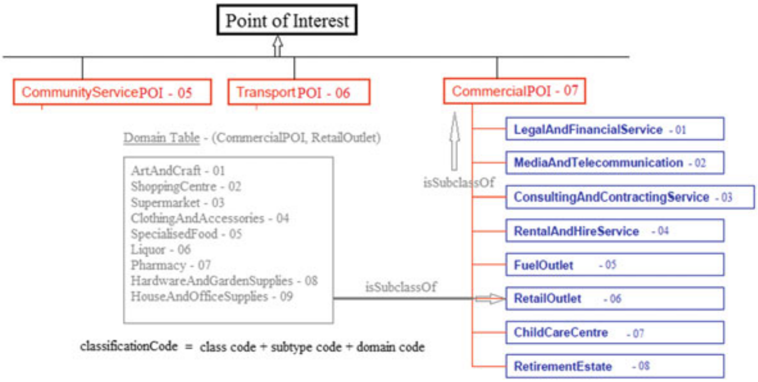
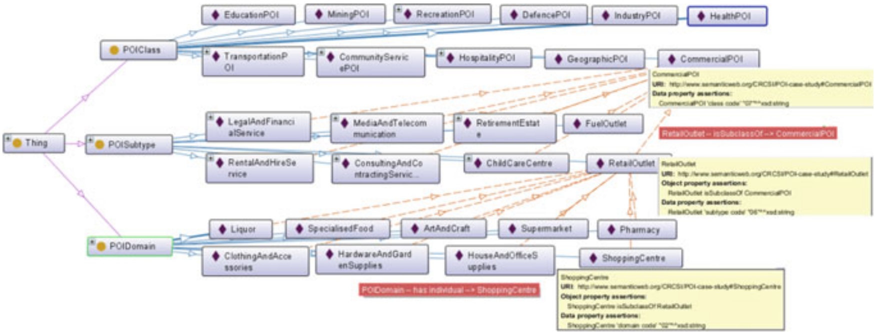

Point of Interest
CommunityServicePOI - 05
TransportPOI - 06
CommercialPOI - 07
Domain Table - (CommercialPOI, RetailOutlet)
ArtAndCraft - 01
ShoppingCentre - 02
Supermarket - 03
ClothingAndAccessories - 04
SpecialisedFood - 05
Liquor - 06
Pharmacy - 07
HardwareAndGardenSupplies - 08
HouseAndOfficeSupplies - 09
isSubclassOf
isSubclassOf
LegalAndFinancialService - 01
MediaAndTelecommunication - 02
ConsultingAndContractingService - 03
RentalAndHireService - 04
FuelOutlet - 05
RetailOutlet - 06
ChildCareCentre - 07
RetirementEstate - 08
classificationCode = class code + subtype code + domain code

Fig. 2 A portion of Landgate POI data model

subtype and feature domains, respectively. A two-digit number following each hierarchy level value is the class code, subtype code and domain code, which together form a six digit classification code number for each POI.

The POI Ontology, designed according to the above structure, formally captures the scope of knowledge for Points of Interest using the Web Ontology Language (OWL), so it is machine-readable and reasoning can be done on the ontology. A part of the ontology corresponds to the same part of the data model demonstrated in Fig. 2 is shown in Fig. 3. There are three classes *POIClass*, *POISubtype* and *POIDomain* in the ontology and each represents a concept in the classification system, i.e., feature class, feature subtype and feature domain. On the right hand side of each class are their individuals or instances, an example is highlighted in red color at the bottom of the figure. The individuals showcased in *POIDomain* correspond to the "Domain Table" values in Fig. 2. They are all feature domains relating to *RetailOutlet* feature subtype; hence all *POIDomain* individuals are pointing to the *RetailOutlet* individual which is a subclass of *CommercialPOI* as



CommercialPOI
URI: http://www.semanticweb.org/CRCSI/POI-case-study#CommercialPOI
Data property assertions:
CommercialPOI 'class code' "07"^^xsd:string

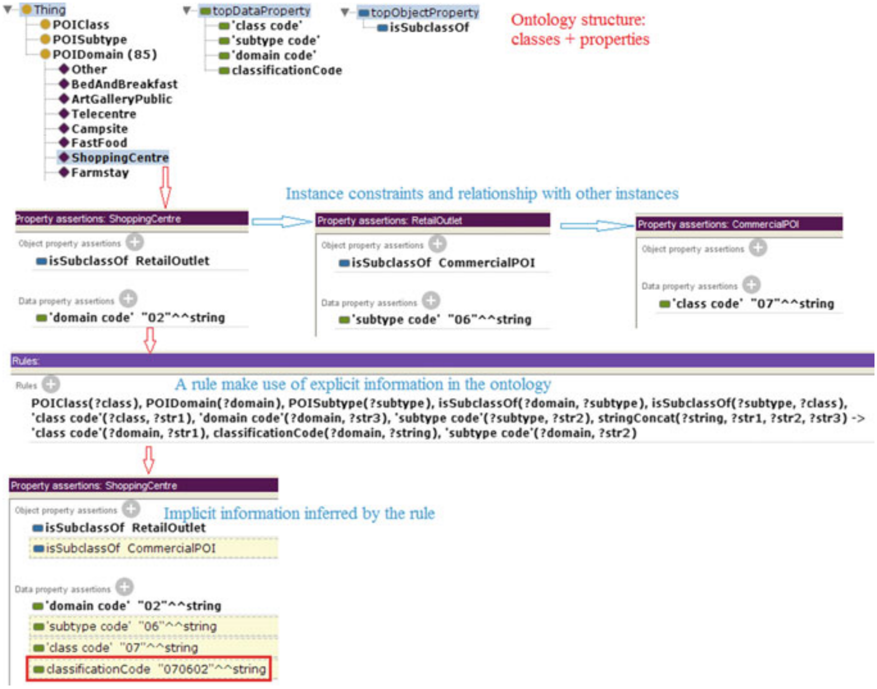
RetailOutlet
URI: http://www.semanticweb.org/CRCSI/POI-case-study#RetailOutlet
Object property assertions:
RetailOutlet isSubclassOf CommercialPOI
Data property assertions:
RetailOutlet 'subtype code' "06"^^xsd:string

ShoppingCentre
URI: http://www.semanticweb.org/CRCSI/POI-case-study#ShoppingCentre
Object property assertions:
ShoppingCentre isSubclassOf RetailOutlet
Data property assertions:
ShoppingCentre 'domain code' "02"^^xsd:string

Fig. 3 OntoGraf (https://protegewiki.stanford.edu/wiki/OntoGraf) representation for classes and instances based on POI data model

indicated by a yellow pointer. All other individuals enumerated in *POISubtype* class are subclasses of *CommercialPOI* as well. Individual features also have a data property to specify its two digit code (see yellow box Fig. 3) and information about whether it has a relationship with another feature using an object property (see yellow pointer Fig. 3). The ontology in Fig. 3 clearly demonstrates the information for individuals in each hierarchy level and their relationship with others; more importantly, these relationships are machine-readable so inferences can be drawn automatically.

The classification code, which can be acquired by string concatenation of class code, subtype code and domain code, is an attribute of each feature domain. It has not been specified individually in the ontology as it is considered common knowledge for all the feature domains and can be inferred using a SWRL rule, as shown in Fig. 4. Consider the *ShoppingCentre* feature domain as an example. Its inferred classification code is inside the red rectangle. The rule together with all classes, instances for each class, object property and data properties presented are considered as the top-level ontology for Points of Interest (Fig. 4). The Top-level ontology includes the minimum information required to express the essential knowledge in this POI study area.



Property assertions: ShoppingCentre
Object property assertions
isSubclassOf RetailOutlet
Data property assertions
'domain code' "02"^^string

Property assertions: RetailOutlet
Object property assertions
isSubclassOf CommercialPOI
Data property assertions
'subtype code' "06"^^string

Property assertions: CommercialPOI
Object property assertions
Data property assertions
'class code' "07"^^string

```
POIClass(?class), POIDomain(?domain), POISubtype(?subtype), isSubclassOf(?domain, ?subtype), isSubclassOf(?subtype, ?class),
'class code'(?class, ?str1), 'domain code'(?domain, ?str3), 'subtype code'(?subtype, ?str2), stringConcat(?string, ?str1, ?str2, ?str3) ->
'class code'(?domain, ?str1), classificationCode(?domain, ?string), 'subtype code'(?domain, ?str2)
```

Property assertions: ShoppingCentre
Object property assertions
isSubclassOf RetailOutlet
isSubclassOf CommercialPOI
Data property assertions
'domain code' "02"^^string
'subtype code' "06"^^string
'class code' "07"^^string
classificationCode "070602"^^string

Fig. 4 POI top level ontology

4.2 Stage 2: Data Conversion and Alignment

When dealing with a specific project or application, the top-level ontology can be expanded to accommodate specific business needs. For example, the data property and object property lists are expanded so they can be used to transform the source data into RDF triples and used in reasoning processes (Fig. 5).

The three source datasets have quite different schemas including different levels of granularity. For example, even though the classification system for the POI was adopted by each source they represent it diversely. The WAPOL data set has three columns recording the POIs' feature class, feature subtype and feature domain values while DFES only contains the feature domain. The Landgate data set has six digital numbers to present the classification code. In order to automatically compare whether two POIs are in a same category, they need to all have a same attribute, either the feature domain value or classification code.

SWRL rules are used to read in the different kinds of classification attributes from each source and infer the missing information contained in the POI classification system so they can have the same attribute granularity. In the top-level ontology (Fig. 4), the 6-digit classification code has already been inferred for each feature domain. Hence, if a POI has a feature domain as "ShoppingCentre", its classification code can be retrieved from the ontology via a SWRL rule as well. This is because data is linked to the ontology during the RDF conversion process and therefore the data has the same semantic description as the ontology. Conversely, if a POI classification code is known, the relevant classification information can also be retrieved by a rule. The rules are shown in Fig. 6. Properties shown in yellow are inferred by the rules while the other data properties are drawn directly from RDF conversion. After alignment, the three example POIs shown below have the same attribute granularity.

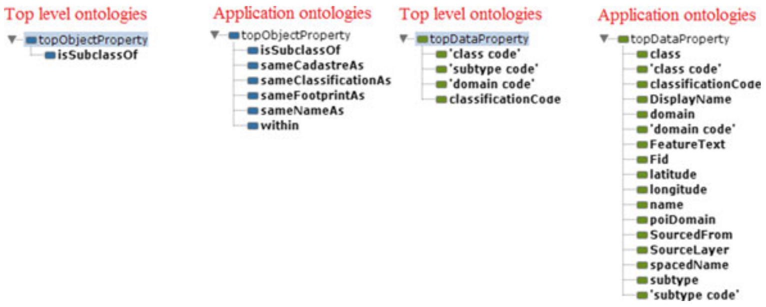
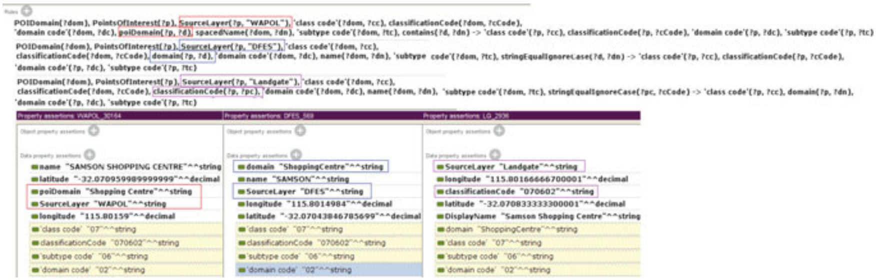


Fig. 5 Developed application ontology based on top level ontology



```
Rules
POIDomain(?dom), PointsOfInterest(?p), SourceLayer(?p, "WAPOL"), 'class code'(?dom, ?cc), classificationCode(?dom, ?cCode),
'domain code'(?dom, ?dc), poiDomain(?p, ?d), spacedName(?dom, ?dn), 'subtype code'(?dom, ?tc), contains(?d, ?dn) -> 'class code'(?p, ?cc), classificationCode(?p, ?cCode), 'domain code'(?p, ?dc), 'subtype code'(?p, ?tc)
POIDomain(?dom), PointsOfInterest(?p), SourceLayer(?p, "DFES"), 'class code'(?dom, ?cc),
classificationCode(?dom, ?cCode), domain(?p, ?d), 'domain code'(?dom, ?dc), name(?dom, ?dn), 'subtype code'(?dom, ?tc), stringEqualIgnoreCase(?d, ?dn) -> 'class code'(?p, ?cc), classificationCode(?p, ?cCode),
'domain code'(?p, ?dc), 'subtype code'(?p, ?tc)
POIDomain(?dom), PointsOfInterest(?p), SourceLayer(?p, "Landgate"), 'class code'(?dom, ?cc),
classificationCode(?dom, ?cCode), classificationCode(?p, ?pc), 'domain code'(?dom, ?dc), name(?dom, ?dn), 'subtype code'(?dom, ?tc), stringEqualIgnoreCase(?pc, ?cCode) -> 'class code'(?p, ?cc), domain(?p, ?dn),
'domain code'(?p, ?dc), 'subtype code'(?p, ?tc)
```

```
Property assertions: WAPOL_30164
Object property assertions
Data property assertions
name "SAMSON SHOPPING CENTRE"^^string
latitude "-32.070959989999999"^^decimal
poiDomain "Shopping Centre"^^string
SourceLayer "WAPOL"^^string
longitude "115.80159"^^decimal
'class code' "07"^^string
classificationCode "070602"^^string
'subtype code' "06"^^string
'domain code' "02"^^string
```

```
Property assertions: DFES_569
Object property assertions
Data property assertions
domain "ShoppingCentre"^^string
name "SAMSON"^^string
SourceLayer "DFES"^^string
longitude "115.8014984"^^decimal
latitude "-32.07043846785699"^^decimal
'class code' "07"^^string
classificationCode "070602"^^string
'subtype code' "06"^^string
'domain code' "02"^^string
```

```
Property assertions: LG_2936
Object property assertions
Data property assertions
SourceLayer "Landgate"^^string
longitude "115.80166666700001"^^decimal
classificationCode "070602"^^string
latitude "-32.07083333300001"^^decimal
DisplayName "Samson Shopping Centre"^^string
domain "ShoppingCentre"^^string
'class code' "07"^^string
'subtype code' "06"^^string
'domain code' "02"^^string
```

Fig. 6 Using SWRL rules to align disparate attributes

4.3 Stage 3 and Stage 4: Finding POI Matches and Attribute Conflation

The logic of finding matches and conflation is as follows:

1. Search points in buffer zone: The spatial (geographic location) characteristic is used as the first step in finding matches. For a selected POI, a buffer size is given by the user and used to calculate the distance between the POI and its surrounding POIs. Only points that fall inside the buffer zone of the selected point will be considered for conflation. This is because points that are close are more likely to be the same point than those further away. This is a mathematic calculation, so a rule is not used.
2. Compare classification code (Rule 1): the second step takes advantage of the POI classification system. As shown in Fig. 1, shopping centre, supermarket, fast food, bus station and taxi rank etc., they could all cluster within a buffer zone. However, each of them belongs to a different feature domain in the POI classification system so their classification code is different. Only points with the same classification code as the selected POI are considered as potential matches to be used in the next comparison step.
3. Compare by name string (Rule 2): For example, even though all POIs may belong to the *FastFood* feature domain, a POI named McDonalds® and another one named KFC® must not be conflated into a single POI because they represent different fast food stores. Following the classification code comparison, the matching list is further narrowed down by doing a name string measure. A POI named "KFC Cannington" and "Kentucky Fried Chicken Cannington" will be the matched points and a POI named "McDonald's Cannington" will not be in the matched list.

   Up to this point the matching and linking process is finished and a list of candidate POIs is ready to be conflated. The list normally contains two or three points, so the next step is to decide which point to keep.
4. Interrelated Relationships (Rule 3 and Rule 4): During the conflation stage, human intervention is normally required as human logic is currently more

efficient than comparison algorithm logic. Domain experts usually use contextual validation to decide which point to keep for each POI. For example, points representing a building are typically overlaid on top of aerial imagery to manually inspect which point is closest to the actual location of the building. In order for the system to perform this task automatically, this contextual validation process is replaced by intersecting POIs with two polygon data sets, i.e., cadastral boundary data and building footprints. The reason is because of the topological relationship they have with POI data. A building footprint must fall into a cadastral boundary, and if a point represents that building, theoretically it must fall into the footprint too. The point is less accurate if it is outside of the footprint but inside the cadastral boundary. It is even less accurate if it is outside the cadastral boundary. Using this logic, if only one point is within the building footprint, then it is considered the most accurate point. This is the point kept and the other physical points will be removed and their attributes conflated into this point. The next choice is the single point within the cadastral boundary.

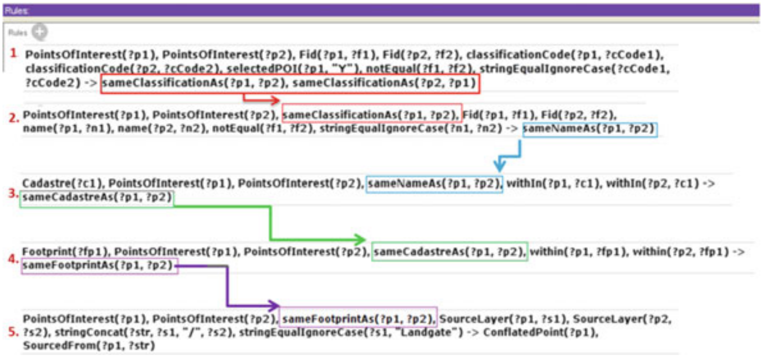
5. User purposes (Rule 5): In the situation where there are still multiple points within the building footprint or none inside the footprint but more than one inside the cadastral boundary, experts usually decide which point to keep based on different purposes and these purposes can be formulated into rules. There are three rules generated in this study:

   (1) *Provenance and Metadata Rule*: The order of reliability is determined by the combined information of metadata and interviews across agencies' experts. In the case study, the order is Landgate, WAPOL, and then DFES. The reason for selecting this option is the user wants to decide based on agencies authority.

   (2) *Statistical Rule*: The centroid (mean location) of all the points in the candidate list determines the conflated point. The reason for selecting this option is when all data from the various sources is to be treated equally.

   (3) *Random Rule*: Randomly select a point within the candidates list. The reason for selecting this option is when the location does not need a high level of accuracy, for example, for general navigation purposes.

According to the above logic, rules generated and are running in a sequential order, i.e., the result of previous rule will be used as a condition in the following rule, showcased in Fig. 7. It demonstrates a chain of rules to deal with the situation where multiple POIs are within a building footprint, the user makes a final decision based on *Provenance and Metadata rule* (Rule 5) and the result is output to a new class named *ConflatedPoint*.



```
1  PointsOfInterest(?p1), PointsOfInterest(?p2), Fid(?p1, ?f1), Fid(?p2, ?f2), classificationCode(?p1, ?cCode1),
   classificationCode(?p2, ?cCode2), selectedPOI(?p1, "Y"), notEqual(?f1, ?f2), stringEqualIgnoreCase(?cCode1,
   ?cCode2) -> sameClassificationAs(?p1, ?p2), sameClassificationAs(?p2, ?p1)

2. PointsOfInterest(?p1), PointsOfInterest(?p2), sameClassificationAs(?p1, ?p2), Fid(?p1, ?f1), Fid(?p2, ?f2),
   name(?p1, ?n1), name(?p2, ?n2), notEqual(?f1, ?f2), stringEqualIgnoreCase(?n1, ?n2) -> sameNameAs(?p1, ?p2)

3. Cadastre(?c1), PointsOfInterest(?p1), PointsOfInterest(?p2), sameNameAs(?p1, ?p2), withIn(?p1, ?c1), withIn(?p2, ?c1) ->
   sameCadastreAs(?p1, ?p2)

4. Footprint(?fp1), PointsOfInterest(?p1), PointsOfInterest(?p2), sameCadastreAs(?p1, ?p2), within(?p1, ?fp1), within(?p2, ?fp1) ->
   sameFootprintAs(?p1, ?p2)

5. PointsOfInterest(?p1), PointsOfInterest(?p2), sameFootprintAs(?p1, ?p2), SourceLayer(?p1, ?s1), SourceLayer(?p2,
   ?s2), stringConcat(?str, ?s1, "/", ?s2), stringEqualIgnoreCase(?s1, "Landgate") -> ConflatedPoint(?p1),
   SourcedFrom(?p1, ?str)
```

Fig. 7 Rule Chain for finding the best location based on provenance and metadata

5 Evaluation

5.1 Preliminary Testing

The methodology presented was tested with an example scenario shown in Fig. 8 and the process was run in Protégé.[6] A POI from the WAPOL dataset was selected (the blue point inside the basket icon) and a 250 m buffer around the point was calculated. Five points from Landgate, five points from WAPOL and one point from DFES (shown in yellow, blue and purple, respectively), all fall within the buffer zone.

The next stage compares the classification code of all points falling within the buffer zone. The selected WAPOL POI has the same code as one from DFES located in a roundabout and one from Landgate, which is located within the building footprint (represented by the green polygon). According to the conflation logic in Sect. 4.3, these three POIs will be conflated into a single point by taking the POI location from the Landgate dataset, shown using the star marker in Fig. 8.

All points in the example scenario and their relevant attributes were used in the reasoning processes listed in Fig. 9. These POIs were added to the same file as the designed POI ontology and SWRL rules so they could be run together with the Protégé reasoner. However, buffer distances are calculated using mathematical functions outside of Protégé. In addition, the comparison of POIs with the digital cadastre and building footprints is also pre-determined using methods, such as a layer intersection outside protégé. Here, the intersection results (listed in Fig. 9) show whether a POI is "within" a cadastral boundary or a building footprint (blue columns). The yellow columns represent data properties and the blue columns show the object properties.

[6] Protégé is a free, open-source platform that provides a suite of tools to construct domain models and knowledge-based applications with ontologies.

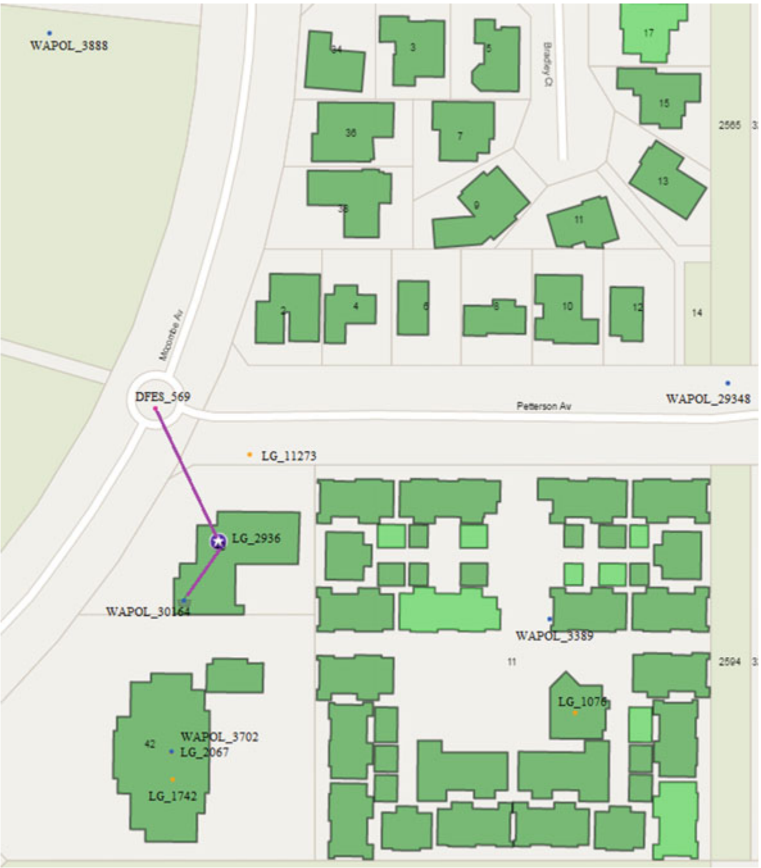
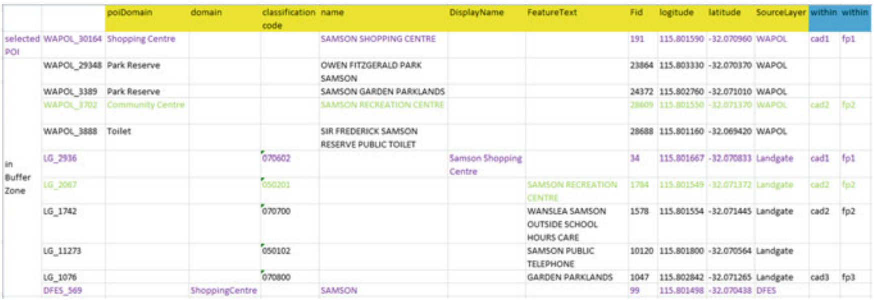
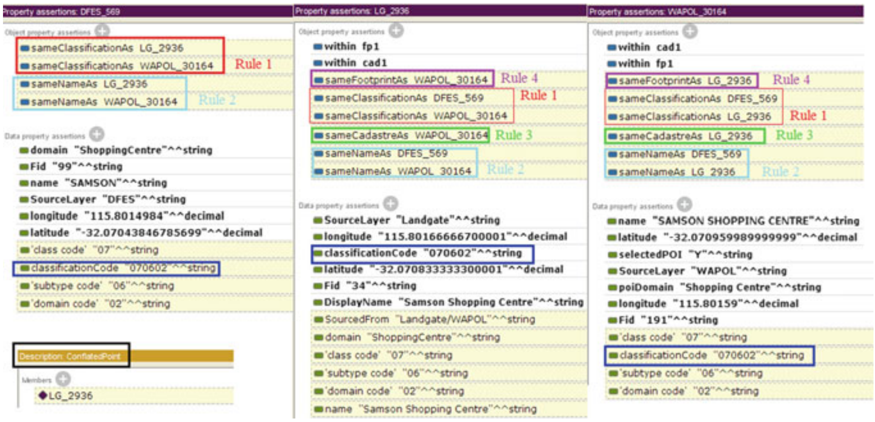


Fig. 8 Example scenario



		poiDomain	domain	classification code	name	DisplayName	FeatureText	Fid	logitude	latitude	SourceLayer	within	within
selected POI	WAPOL_30164	Shopping Centre			SAMSON SHOPPING CENTRE			191	115.801590	-32.070960	WAPOL	cad1	fp1
in Buffer Zone	WAPOL_29348	Park Reserve			OWEN FITZGERALD PARK SAMSON			23864	115.803330	-32.070370	WAPOL		
	WAPOL_3389	Park Reserve			SAMSON GARDEN PARKLANDS			24372	115.802760	-32.071010	WAPOL		
	WAPOL_3702	Community Centre			SAMSON RECREATION CENTRE			28609	115.801550	-32.071370	WAPOL	cad2	fp2
	WAPOL_3888	Toilet			SIR FREDERICK SAMSON RESERVE PUBLIC TOILET			28688	115.801160	-32.069420	WAPOL		
	LG_2936			070602		Samson Shopping Centre		34	115.801667	-32.070833	Landgate	cad1	fp1
	LG_2067			050201			SAMSON RECREATION CENTRE	1784	115.801549	-32.071372	Landgate	cad2	fp2
	LG_1742			070700			WANSLEA SAMSON OUTSIDE SCHOOL HOURS CARE	1578	115.801554	-32.071445	Landgate	cad2	fp2
	LG_11273			050102			SAMSON PUBLIC TELEPHONE	10120	115.801800	-32.070564	Landgate		
	LG_1076			070800			GARDEN PARKLANDS	1047	115.802842	-32.071265	Landgate	cad3	fp3
	DFES_569		ShoppingCentre		SAMSON			99	115.801498	-32.070438	DFES		

Fig. 9 Attribute list of example scenario POIs



```
Property assertions: DFES_569
Object property assertions
  sameClassificationAs  LG_2936
  sameClassificationAs  WAPOL_30164
  sameNameAs  LG_2936
  sameNameAs  WAPOL_30164
Data property assertions
  domain  "ShoppingCentre"^^string
  Fid  "99"^^string
  name  "SAMSON"^^string
  SourceLayer  "DFES"^^string
  longitude  "115.8014984"^^decimal
  latitude  "-32.07043846785699"^^decimal
  'class code'  "07"^^string
  classificationCode  "070602"^^string
  'subtype code'  "06"^^string
  'domain code'  "02"^^string

Description: ConflatedPoint
Members
  ◆LG_2936
```

```
Property assertions: LG_2936
Object property assertions
  within  fp1
  within  cad1
  sameFootprintAs  WAPOL_30164
  sameClassificationAs  DFES_569
  sameClassificationAs  WAPOL_30164
  sameCadastreAs  WAPOL_30164
  sameNameAs  DFES_569
  sameNameAs  WAPOL_30164
Data property assertions
  SourceLayer  "Landgate"^^string
  longitude  "115.80166666700001"^^decimal
  classificationCode  "070602"^^string
  latitude  "-32.070833333300001"^^decimal
  Fid  "34"^^string
  DisplayName  "Samson Shopping Centre"^^string
  SourcedFrom  "Landgate/WAPOL"^^string
  domain  "ShoppingCentre"^^string
  'class code'  "07"^^string
  'subtype code'  "06"^^string
  'domain code'  "02"^^string
  name  "Samson Shopping Centre"^^string
```

```
Property assertions: WAPOL_30164
Object property assertions
  within  cad1
  within  fp1
  sameFootprintAs  LG_2936
  sameClassificationAs  DFES_569
  sameClassificationAs  LG_2936
  sameCadastreAs  LG_2936
  sameNameAs  DFES_569
  sameNameAs  LG_2936
Data property assertions
  name  "SAMSON SHOPPING CENTRE"^^string
  latitude  "-32.070959989999999"^^decimal
  selectedPOI  "Y"^^string
  SourceLayer  "WAPOL"^^string
  poiDomain  "Shopping Centre"^^string
  longitude  "115.80159"^^decimal
  Fid  "191"^^string
  'class code'  "07"^^string
  classificationCode  "070602"^^string
  'subtype code'  "06"^^string
  'domain code'  "02"^^string
```

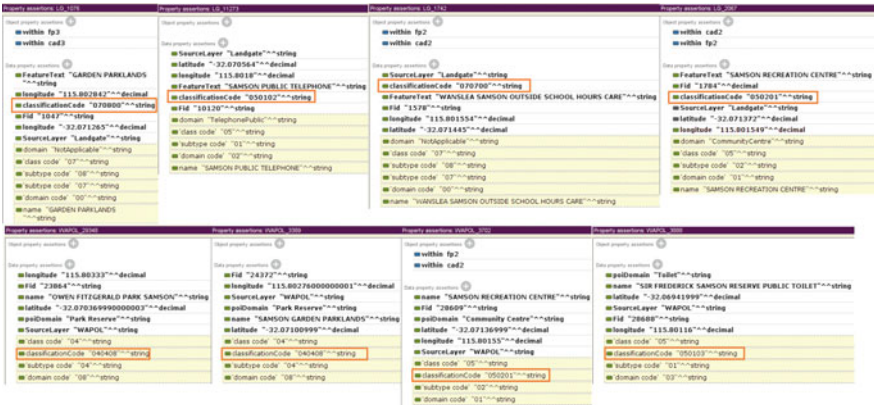
Fig. 10 Properties for POIs after running reasoner

The Protégé built-in reasoner Pellet[7] is run to check whether it can properly return inferred results for different POIs using each rule. As shown in Fig. 8, *DFES_569*, *LG_2936* and *WAPOL_30164* are supposed to be conflated into one, i.e., *LG_2936*. The inference results of the three POIs are showed in Fig. 10.

(1) Rule 1 returns results for the three POIs (see red rectangle). It correctly identifies one POI has the same classification code as the other two because they are all “070602” (see dark blue rectangle).
(2) Rule 2 also correctly returns inferred results for each POI. (See light blue rectangle). Each POI has the same name as the other two because the name values are “SAMSON”, “Samson Shopping Centre” and “SAMSON SHOPPING CENTRE”, so they are either an exact match when ignore case (e.g., “Samson Shopping Centre” and “SAMSON SHOPPING CENTRE”) or one is contained within the other (e.g., “SAMSON” and “SAMSON SHOPPING CENTRE”).
(3) Rule 3 and Rule 4 does not return any result for *DFES*_569 because it is not within any cadastral boundary or building footprint. Both rules return a result for the other two POIs because they all within “cad1” and “fp1”, so they have *sameCadastreAs* and *sameFootprintAs* with each other.
(4) Rule 5 returns the final result as *LG_2936*, which is an inferred member of *ConflatedPoint* class (see black rectangle in the lower left corner). This is the expected result for the test scenario based on the *Provenance and Metadata Rule*, i.e., Landgate data is more accurate than WAPOL data when two POIs from these two sources are both within a building footprint.

The inferred results for other points included in the test scenario are shown in Fig. 11. Because their classification codes are different than the selected POI, no

[7] Pellet is an open-source Java based OWL 2 reasoner https://www.w3.org/2001/sw/wiki/Pellet.



```
Property assertions: LG_1076
Object property assertions
within fp3
within cad3
Data property assertions
FeatureText "GARDEN PARKLANDS "^^string
longitude "115.802842"^^decimal
classificationCode "070800"^^string
Fid "1047"^^string
longitude "-32.071265"^^decimal
SourceLayer "Landgate"^^string
domain "NotApplicable"^^string
'class code' "07"^^string
'subtype code' "08"^^string
'subtype code' "07"^^string
'domain code' "00"^^string
name "GARDEN PARKLANDS "^^string
```

```
Property assertions: LG_11273
Object property assertions
Data property assertions
SourceLayer "Landgate"^^string
latitude "-32.070564"^^decimal
longitude "115.8018"^^decimal
FeatureText "SAMSON PUBLIC TELEPHONE"^^string
classificationCode "050102"^^string
Fid "10120"^^string
domain "TelephonePublic"^^string
'class code' "05"^^string
'subtype code' "01"^^string
'domain code' "02"^^string
name "SAMSON PUBLIC TELEPHONE"^^string
```

```
Property assertions: LG_1742
Object property assertions
within fp2
within cad2
Data property assertions
SourceLayer "Landgate"^^string
classificationCode "070700"^^string
FeatureText "WANSLEA SAMSON OUTSIDE SCHOOL HOURS CARE"^^string
Fid "1578"^^string
longitude "115.801554"^^decimal
latitude "-32.071445"^^decimal
domain "NotApplicable"^^string
'class code' "07"^^string
'subtype code' "08"^^string
'subtype code' "07"^^string
'domain code' "00"^^string
name "WANSLEA SAMSON OUTSIDE SCHOOL HOURS CARE"^^string
```

```
Property assertions: LG_2067
Object property assertions
within cad2
within fp2
Data property assertions
FeatureText "SAMSON RECREATION CENTRE"^^string
Fid "1784"^^decimal
classificationCode "050201"^^string
SourceLayer "Landgate"^^string
latitude "-32.071372"^^decimal
longitude "115.801549"^^decimal
domain "CommunityCentre"^^string
'class code' "05"^^string
'subtype code' "02"^^string
'domain code' "01"^^string
name "SAMSON RECREATION CENTRE"^^string
```

```
Property assertions: WAPOL_20348
Object property assertions
Data property assertions
longitude "115.80333"^^decimal
Fid "23864"^^string
name "OWEN FITZGERALD PARK SAMSON"^^string
latitude "-32.070369990000003"^^decimal
poiDomain "Park Reserve"^^string
SourceLayer "WAPOL"^^string
'class code' "04"^^string
classificationCode "040408"^^string
'subtype code' "04"^^string
'domain code' "08"^^string
```

```
Property assertions: WAPOL_3309
Object property assertions
Data property assertions
Fid "24372"^^string
longitude "115.80276000000001"^^decimal
SourceLayer "WAPOL"^^string
poiDomain "Park Reserve"^^string
name "SAMSON GARDEN PARKLANDS"^^string
latitude "-32.07100999"^^decimal
'class code' "04"^^string
classificationCode "040408"^^string
'subtype code' "04"^^string
'domain code' "08"^^string
```

```
Property assertions: WAPOL_3702
Object property assertions
within fp2
within cad2
Data property assertions
name "SAMSON RECREATION CENTRE"^^string
Fid "28609"^^string
poiDomain "Community Centre"^^string
latitude "-32.07136999"^^decimal
longitude "115.80155"^^decimal
SourceLayer "WAPOL"^^string
'class code' "05"^^string
classificationCode "050201"^^string
'subtype code' "02"^^string
'domain code' "01"^^string
```

```
Property assertions: WAPOL_3000
Object property assertions
Data property assertions
poiDomain "Toilet"^^string
name "SIR FREDERICK SAMSON RESERVE PUBLIC TOILET"^^string
latitude "-32.06941999"^^decimal
SourceLayer "WAPOL"^^string
Fid "28688"^^string
longitude "115.80116"^^decimal
'class code' "05"^^string
classificationCode "050103"^^string
'subtype code' "01"^^string
'domain code' "03"^^string
```

Fig. 11 Reasoning results for all other points

results are generated in Rule 1. Hence they are not carried any further in the reasoning process. This fulfills the expectation of the rules as only those candidates that meet the previous rules are carried into the next rule.

5.2 *Proof of Concept Web Portal and Further Evaluation Data*

The preliminary testing results demonstrate that the SWRL Rule-based Data Conflation methodology can model domain experts' decision making logic, thus enabling geospatial data to be conflated automatically. However, as Protégé is essentially an ontology and SWRL rule editor, there are many functions that cannot be performed, such as, calculate points within buffer zone, and intersect points with reference layers. Also, the example only demonstrates one scenario, which is two points within the same footprint and the final decision is based on *Provenance and Metadata Rule*. However, it is acknowledge that there could be other scenarios and different rules will come into play, such as a decision made by statistic rules or random rules, or if only one point is in a footprint, the point can be chosen automatically etc.

A Proof of Concept (PoC) web portal has been developed to integrate the aforementioned functions and automatically trigger different rules depending on the different situations.[8] The Data Conflation application server provides a visualisation layer so that the user can view the dataset points before and after conflation. The visualisation layer is developed using React JS. The user is able to access it through a common web browser such as Chrome and Firefox etc. The web application

[8] https://crcsi.amristar.com/automatedconflation; username: crcsi; password: 1@ndg@te.

server also hosts the Apache Jena Semantic Web business rules engine that the web application interfaces to execute the conflation processes.

As the PoC web portal is capable of dealing with larger datasets and more complicated scenarios, a further evaluation was able to be performed. The evaluation is based on conflating *ShoppingCentre* feature domain points from the three sources including 351 POIs from Landgate, 255 POIs from WAPOL and 381 POIs from DFES. These POIs are well distributed across Perth metropolitan area. The reason for using this particular feature domain is that these points exist in all three datasets in the study area. The WA Police dataset and Landgate dataset cover most of the feature domains, whereas the DFES dataset only records *FastFood*, *Supermarket* and *ShoppingCentre* feature domains. However, the Landgate dataset does not contain enough samples in the *FastFood* and *Supermarket* feature domains with only 8 and 28 points in each feature domain, respectively. Furthermore, the points in these two Landgate feature domains occur outside the Perth Metro area where no building footprint data is available to compare. Therefore, the *ShoppingCentre* feature domain data in this case is the best test data to evaluate whether conflation decisions can be correctly made between the three sources.

The buffer size is set as 250 m is based on trial and error. A manual check on a few of the larger shopping centres in the metropolitan region showed that 250 m is sufficient to return relevant points and it is not too larger an area to decrease system performance. Nonetheless, in the PoC web portal, a user is able to select an area of interest rather than the whole dataset search area.

The building footprints and cadastral boundaries reference datasets are provided by Landgate, which is the recognised authoritative source.

5.3 Evaluation Criteria and Results

The evaluation focuses on two aspects; (a) whether the system can effectively reduce duplicate data; and (b) the accuracy of conflated results.

In terms of duplication, the number of conflated POIs is 493, whereas the number of POIs from the combined datasets is 987 (Fig. 12). This means that over half of the points are duplicated, and hence have been removed. At the same time, each source dataset has an increased number of POIs and thus coverage is improved. This is shown in Fig. 12 where Landgate has increased the number of valid POIs by 40%, WAPOL by 93% and DFES by 29%.

In order to examine how accurate the results are, manual validation was performed. Among the 493 conflated POIs, 283 points were generated from multiple points, i.e., either from more than one source or more than one point from the same source. Each of these 283 points were loaded into ArcMap and overlaid with the three source datasets and the two reference datasets to check whether or not the SWRL rule system effectively selected the best location for each scenario. The statistical results are displayed in Table 1.

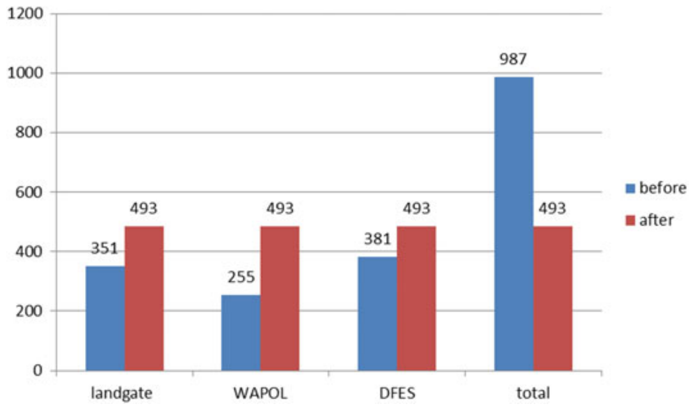


Fig. 12 Number of points before and after conflation for each source

Table 1 Evaluation result for conflate three datasets

Source	# Conflated POI				Total
	#Multi-sources			# Single source	
	Auto-select		Decided by rule		
	In footprint	In cadastre			
Landgate	58	2	156	60	276
WAPOL	15	4	24	63	106
DFES	15	0	9	87	111
Total	88	6	189	210	
					Total: 493

The test revealed that 88 points were conflated automatically because there was only one data source with the point inside the building footprint. There are 6 cases where no points were within a building footprint and only one point inside a cadastral boundary. The remaining 189 conflated points were decided by the *Provenance and Metadata Rule* as multiple source points existed in a same footprint or cadastral boundary. As the *Provenance and Metadata Rule* defines the Landgate dataset as the most accurate the result showing 156 points from Landgate source as the highest number of valid points was expected over the WAPOL (24 points) and DFES (9 points datasets. Changing the *Provenance and Metadata Rule* would achieve difference results.

Among the 283 conflated points, only 5 points were identified as incorrect and therefore, the conflation accuracy for *ShoppingCentre* POI is 98%.

There are 210 points in the conflated dataset, which were derived from a single source. However, 64 of these points should have match other points but were excluded due to the current name string method being too simple. The current string match method uses *SWRL Built-Ins for String*, which can only perform simple matches, such as match points with exactly the same name or where one name is contained within another *name* string. However, some name patterns such as full name (e.g., *Kentucky Fried Chicken*) and acronym name (e.g., *KFC*) will not return a result *as matched*. A better match method is required to deal with various name patterns across the datasets.

In future work, a more sophisticate string match algorithm will be used to generate custom Built-Ins for SWRL to improve the accuracy of the name string match in order to reduce the number of duplicate points further.

6 Conclusion and Future Work

Incidents are often attended by more than one emergency service organization. If each agency is using their own datasets there is a risk that information may be different leading to poor communication and coordination between first responders. A conflated single authoritative dataset is therefore desirable between agencies. This paper presents a new approach to data conflation where an ontology and RDF data conversion serve as the basis for the solution and SWRL rules are the core to automate the entire geospatial data conflation processes. By using a set of rules in a sequential order, human experts' logic can be used to find the most accurate or fit-for-purpose location and conflate the remaining attributes into the single location and removing duplicate features. In this way, the conflation processes can be run automatically without human intervention.

In the Proof of Concept web application, some other datasets are also used in the system, such as OpenStreetMap and BingImage. At this stage these are only used as based maps for visual reference and not included in the conflation process. Although the conflation with OpenStreetMap is not in the scope of this paper, including OpenStreetMap into the conflation reasoning process either as a reference dataset to facilitate decision making or used as a fourth source dataset to conflate into a single dataset is planned in the future work.

Acknowledgements The work has been supported by the Cooperative Research Centre for Spatial Information, whose activities are funded by the Australian Commonwealth's Cooperative Research Centre Program.

References

Bishr Y (1998) Overcoming the semantic and other barriers to GIS interoperability. Int J Geogr Inf Sci 12(4):299–314

Chen C-C et al (2004) Automatically and accurately conflating orthoimagery and street maps. In: Proceedings of the 12th annual ACM international workshop on Geographic information systems. ACM

Chen C-C et al (2006) Automatically conflating road vector data with orthoimagery. GeoInformatica 10(4):495–530

Chen C-C et al (2008) Automatically and accurately conflating raster maps with orthoimagery. GeoInformatica 12(3):377–410

Dongcai HE (2013) A study on theory and method of spatial vector data conflation. Res J Appl Sci Eng Technol 5(2):563–567

Duckham M et al (2017) Towards a spatial knowledge infrastructure Australia and New Zealand CRC for spatial information

Flowerdew R (1991) Spatial data integration

Fonseca F et al (2002) Using ontologies for integrated geographic information systems. Trans GIS 6(3):231–257

Giannopoulos G et al (2015) FAGI-gis: a tool for fusing geospatial RDF data. In: The semantic web: ESWC 2015 satellite events: ESWC 2015 satellite events, Portorož, Slovenia, May 31–June 4, 2015. Springer International Publishing, pp 51–57

Janowicz K et al (2010) Semantic enablement for spatial data infrastructures. Trans GIS 14 (2):111–129

Kang H (2001) Spatial data integration: a case study of map conflation with census bureau and local government data. http://www.cobblestoneconcepts.com/ucgis2summer/kang/kang_main.htm

Lutz M et al (2009) Overcoming semantic heterogeneity in spatial data infrastructures. Comput Geosci 35(4):739–752

Lynch MP, Saalfeld AJ (1985) Conflation: automated map compilation—a video game approach. Auto-Carto 7, Washington, DC, USA

Parekh V et al (2004) Ontology based semantic metadata for geoscience data. IKE

Patrick M, Sven S (2009) Data integration in the geospatial semantic web. J Cases Inf Technol (JCIT) 4(11):100–122

Saalfeld A (1988) Conflation automated map compilation. Int J Geogr Inf Syst 2(3):217–228

Sehgal V et al (2006) Entity resolution in geospatial data integration. In: Proceedings of the 14th annual ACM international symposium on advances in geographic information systems. Arlington, Virginia, USA, ACM, pp 83–90

Szekely P et al (2011) Exploiting semantics of web services for geospatial data fusion. In: Proceedings of the 1st ACM SIGSPATIAL international workshop on spatial semantics and ontologies. ACM, Chicago, Illinois, pp 32–39

Uitermark H et al (1999) Ontology-based geographic data set integration. In: Böhlen M, Jensen C, Scholl M (eds) Spatio-temporal database management, vol 1678. Springer, Berlin, Heidelberg, pp 60–78

Walter V, Fritsch D (1999) Matching spatial data sets: a statistical approach. Int J Geogr Inf Sci 13 (5):445–473

Wiegand N, García C (2007) A task-based ontology approach to automate geospatial data retrieval. Trans GIS 11(3):355–376

Wiemann S, Bernard L (2016) Spatial data fusion in spatial data infrastructures using linked data. Int J Geogr Inf Sci 30(4):613–636

Yu F et al (2016) Automatic geospatial data conflation using semantic web technologies. In: Proceedings of the Australasian computer science week multiconference, Canberra, Australia. ACM, pp 1–10

Zhang Y et al (2013) A semantic approach to retrieving, linking, and integrating heterogeneous geospatial data. In: Joint proceedings of the workshop on ai problems and approaches for intelligent environments and workshop on semantic cities, Beijing, China. ACM, pp 31–37