Research Proposal

Urban traffic data engineering using Semantic Web technologies to support Autonomous Vehicles

Eindhoven University of Technology (TU/e)

PhD Candidate:
Miloš Viktorović : m.viktorovic@tue.nl, +31 (0)40 247 6179

Promotors:
Professor Bauke de Vries : b.d.vries@tue.nl, +31 (0)40 247 2388
Professor Nico Baken : n.h.g.baken@tue.nl, +31 (0)65 142 8630
Dr. Dujuan Yang : d.yang@tue.nl, +31 40 247 3399

Eindhoven University of Technology,
Department of the Built Environment,
Information Systems in the Built Environment (ISBE).

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Content

1 Motivation

Nowadays we have various technologies for modelling the traffic infrastructure within urban areas. Representation of this infrastructure is done through different standards for spatial data sharing, like CityGML, Kadaster, and similar, that are being developed, not with the idea to be consumed by low performance computer systems like ones, for example, in cell phones or cars. In addition to relatively low computational power, autonomous vehicles, both ones present now and ones that will be coming in the future, mostly rely on mimicking behaviour of human controlled vehicles.

Other important thing to note is that mentioned technologies for modelling of traffic infrastructure do not include environmental variables that are being obtained by massive amounts of existing sensing devices. Therefore utilisation of IoT (internet of Things) infrastructure together with digital representation of physical road infrastructure would obviously be beneficial as source of information for all participants in traffic.

Thing to keep in mind also is that most of big car producers are anticipating that around 2021 there will be huge breakthroughs in autonomous driving, firstly on highways and then few years later in urban environments as well[1]. And as cars are already coming off from production lines equipped with DbW (Drive by Wire) systems, we could expect to see more projects like one developed by Comma.ai[1], which aim to transform all cars with mentioned DbW into autonomous vehicles.

This massive inflow of AVs (Autonomous Vehicles), will therefore create many challenges that we will have to mitigate in order to improve safety, efficiency and comfort of all traffic participants. By making these vehicles communicate and therefore collaborate with other participants in traffic, we can tackle challenges previously mentioned, but then we open completely new set of potential issues that come with it.

Main challenge, for this research would therefore be, how to integrate and codify information from different sources so that it can:

- Support processing of the shared information, by AI (Artificial Intelligence), in the most optimal way, using AV hardware.
- Optimally utilise the network bandwidth for sharing of the needed information.

As we will see in the next section, available research into fields of Autonomous vehicles is well developed, but still there are missing links in integration of vehicles, physical infrastructure and already present IoT traffic related devices, into digital representation of the whole traffic ecosystem on a data model level. This is reason why new approach has to be developed, how to integrate CAVs (Connected Autonomous Vehicles) with infrastructure and sensory data, but also how to re-model the existing infrastructure so that its digital representation is CAV friendly.

[1] https://comma.ai/

2 Review of literature and other related work

2.1 Connected Autonomous Vehicles (CAV)

Historically, idea of self-driving vehicles was present as long as the cars first started rolling down the roads, but first significant research and developments happened in 1940's and 1950's. With the development of robotics, in parallel came development of self-driving vehicles, in which first ground-breaking experiment came from Tsukuba Mechanical Engineering Lab, Japan in 1977[2]. This was the first computer-operated vehicle, using computer vision to identify road markings.

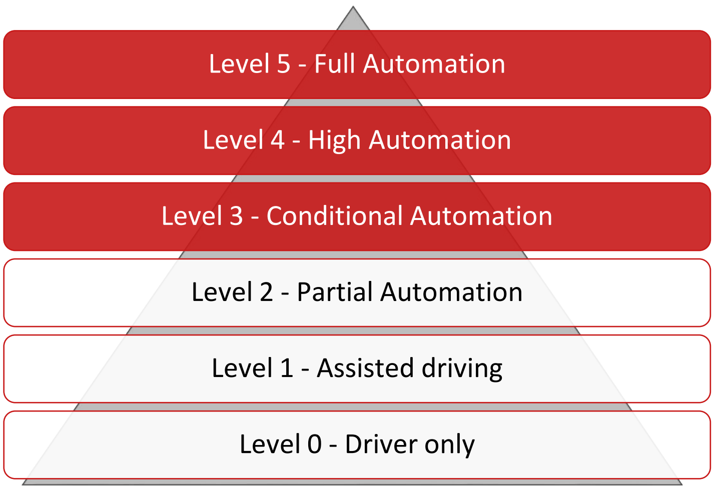
In recent years many companies developed prototypes with partial autonomy, most noticeable Tesla Motors and Google. As it can be noticed from those trials, we can distinguish two groups of approaches [3]:

i. Mediated perception approaches – multicomponent system, which takes all sensory inputs and with a help of AI, translates this complex information into driving outputs, based on set of pre-programmed rules.
ii. Behaviour reflex approaches – Using machine learning algorithms to map direct connections between sensory inputs and driving outputs, de-facto mimicking human driving, used to train the model.

Both approaches here use some sort of AI, and have to handle relatively big amount of information, coming from (currently) vehicle sensors only. If this input can be enriched with meaningful information from surrounding infrastructure, then we could optimise these algorithms even more, so that they provide more efficient vehicle control parameters.

2.1.1 Levels of autonomy

In current literature, 6 (0 to 5) levels of autonomy have been established as a standard[4], when it comes to road vehicles. Noticeably they can be subdivided into two groups, based on human involvement level. Second group, with autonomy levels from 3 to 5 we often refer to as Autonomous Systems or Autonomous Vehicles (AVs)[5].



For this research, interest would be in Levels 2 to 5, as they can benefit the most from uniform data model and information sharing. These levels can also be defined by their functional definition[4]:

- L2 – Traffic Jam Assist
- L3 – Highway Patrol
- L4 – Urban Driving
- L5 – E2E Fully Auto. driving

2.2 City Models

Vehicles today relay mostly on 2D GIS (Geo Information Systems), to navigate urban environment. In recent period, different models of cities have been provided to tackle different issues within cities. As they are also tackling modelling of traffic infrastructure, it is worth reviewing what has been developed, as it might give us insights on how to best represent infrastructure for CAV usage.

As mentioned, there were many proposed models for Smart Cities during last decade, which can be classified in three main categories[6]:

1. Abstract Smart City models
2. Specific-purpose application models
3. Multi-domain practical models

Interesting representative of Multi-Domain models is KM4City[2] (Knowledge Model for City) [7], a practical implementation of smart-city ontology, which is showing the usability and value of semantically structured open data within Smart Cities domain. More specifically it focuses on areas of: Administration, Local Public Transport, POI (Points of Interest), Sensors and Street guides.

Other interesting and relatively new model, which is not, in the literature, classified into one of previously mentioned categories, is CityGML[3]. This open data model, for 3D City and environment modelling, was developed by Open Geospatial Consortium (OGC) [8],as an application schema based on GML3 and ISO TC211 standards. CityGML uses GML3 to represent 3D geometries based on the ISO 19107 model, through multiple taxonomies:

- Digital Terrain Models as a combination of (including nested) triangulated irregular networks (TINs), regular raster, break and skeleton lines, mass points
- Sites (currently buildings, bridges, and tunnels)
- Vegetation (areas, volumes, and solitary objects with vegetation classification)
- Water bodies (volumes, surfaces)
- **Transportation facilities (both graph structures and 3D surface data)**
- Land use (representation of areas of the earth's surface dedicated to a specific land use)
- City furniture
- Generic city objects and attributes
- User-definable (recursive) grouping

All of these taxonomies are divide in 5 different LODs (Level of Detail). But most interesting part of this model, is that allows extension through built-in ADE (Application Domain Extension) Module. This gives the ability to formally specify any extension to provided schema. Some of interesting examples of extensions are:

- UtilityNetworkADE[4] – Developed at TU Berlin for the purpose of representing utility networks within 3D environments, using semantic ontologies[9].

[2] http://www.km4city.org/
[3] http://www.opengeospatial.org/standards/citygml
[4] http://www.citygmlwiki.org/index.php/CityGML_UtilityNetworkADE

- GeoBIM[5] - Integration of BIM with GIS systems, and combining IFC information with CityGML, created at Netherlands Organisation for Applied Scientific Research (TNO)[10].

New version of CityGML, named 3.0[6] is due to be fully published by the beginning of 2019. According to drafts, new standard will be fully compliant with ISO 191xx family. Also new modules have been added, most interesting of them being *Dynamizer*, which allows representation of time variable data for modelled object's properties[11].

Although developments as Dynamizer provide some integration of sensory data with physical infrastructure, these models are still lacking proper inter-tangling with IoT systems. Therefore they might be only useful to provide static GIS information in 3D format.

2.3 C-ITS

Currently there are a few pilot project within Netherlands and Europe [12], that are following European Commission legislatives[13], regarding standardization of ITS (Intelligent Transport Systems), under common umbrella of C-ITS (Cooperative ITS), which is to be developed by 2019.

C-ITS initiative has multiple use cases, under two categories "Day 1" and "Day 1.5". Within *Day 1*, use cases developed are all informative only communications, with a goal of providing drivers with useful traffic information. Here, main focus is on V2V and V2I communication, by providing framework for vehicle data message sharing.

Within this, CCAM (Cooperative, connected and Autonomous Mobility), in addition to legal frameworks, EU has directed efforts and funds towards creating platform for collaborative development of C-ITS environment, called *C-Road* [14].

Although this initiative, provides great framework for sharing information between vehicles and infrastructure, "Day 1 services" are intended mostly for informative purposes to drivers, and they are also not providing physical road information (surface, incline, etc.).

In addition to C-ITS, there are more projects aiming for standardisation in the field of TIS (Traffic Information Systems). One example is DATEX II[15], which provides general guidelines on how to interconnect data in order to make it exchangeable. This standard also defines element of vehicle, which contains basic vehicle information, and which, for example, researchers used for developing optimization for electric vehicles, based on their battery hardware info[16]. Unfortunately this digital representation of vehicle does not model vehicles' dynamic sensing capabilities [17]. For the purpose of research, teams from Belgium created OWL interpretation of DATEX II v2.0, which might be very useful for this research as well[7].

2.4 Linked Semantic Data

"Linked Data lies at the heart of what Semantic Web is all about: large scale integration of, and reasoning on, data on the Web."[18]. In essence Linked Data represents the collection of interrelated semantic datasets. This leads us to the question of what is semantic data.

Term *Semantic data* is used to describe concept of data model based on a relationship between stored symbols and real-world [19]. Key advantage of this kind of data organization is that it can be

[5] http://www.citygmlwiki.org/index.php/CityGML_GeoBIM_ADE
[6] https://www.citygml.org/ongoingdev/v3/
[7] http://vocab.datex.org/terms/#Ontology

interpreted easily and meaningfully by machines. Concept itself first time appeared in 70's, but in last decade it became extremely popular with the birth of concepts like Web of Data.

Other important thing to mention, when Semantic data is discussed, is the context. Context-awareness becomes important when building semantic models, and therefore defining it is of outmost importance, as it is the context what gives meaning to data.

"Context is any information that can be used to characterize the situation of an entity" [20].

This definition has been extended in 2007, with a set of elements' descriptions. Where these elements are:

- Individuality
- Activity
- Location
- Time
- Relations

"The activity predominantly determines the relevancy of context elements in specific situations, and the location and time primarily drive the creation of relations between entities and enable the exchange of context information among entities" [21].

Especially interesting for multi-domain use is the idea of *Shared Context* proposed by Zimmerman et al. "A shared context emerges, when the contexts of two entities overlap and parts of the context information become similar and shared." [21]

Semantic data is usually organised in triplets, which represent binary relationship between two objects of data. In addition these relationships are defined in other form of data structure, which represents ontology, which we use to interconnect our data. There are currently many domain specific ontologies especially in field of medicine and biomedicine, as well as some more general one, like DBpedia[8].

When it comes to technologies, main components used for semantic data are:

- RDF (Resource Description Framework)[9] – W3C's framework for representing Semantic data
- SPARQL[10] – A RDF query language
- Web Ontology Language (OWL)[11] – Knowledge (Ontology) representation language

Although a very promising technology, there are still some areas that are being developed and are not yet fully standardised. Primarily in domain of Dynamic (Streaming) data, which is very interesting from the point of IoT and therefore Smart Cities, we are lacking full consensus how it should be approached. For now there are two approaches defined:

- Time-stamp based
- Interval-based

[8] https://wiki.dbpedia.org/
[9] https://www.w3.org/RDF/
[10] https://www.w3.org/TR/rdf-sparql-query/
[11] https://www.w3.org/TR/owl2-overview/

Despite lacking proper standardization, there are developments to tackle the problems of data streams, developed as extensions to SPARQL language, C-SPARQL[12], SPARQLstream and few others. It is also important to mention that standardised frameworks like SSN[22] and SOSA [23], provide us a way of semantically linking sensory devices, and therefore might be also useful for connected vehicles, but this still remains to be proven.

2.5 PDOK & TOP10NL

"Publieke Dienstverlening Op de Kaart" (PDOK) is platform developed by Dutch authorities, with a goal of sharing geo-data services and sets. In addition to having datasets for whole Netherlands in GML formats, one of the groups has also focused efforts into transferring this data, using Linked data technologies[24]. This resulted in creation of ontology that can be queried[13] using SPARQL in order to extract basic information for infrastructure objects within Netherlands.

Unfortunately based on currently available information, Linked Data part of PDOK platform is in development phase. Thanks to documentation of other (GML) sets from Kadaster Netherland [25], basic data relationships and models can be deducted.

Based on practical research of available data online, we can see basic relationship in Figure 1 (*Wegdeel [NL] → Road section [EN]*; *Straat [NL] → Street [EN]*), which is provided, when exploring one road section. When we look deeper into TTL representation of SPARQL request for the same "Road Section 119076081"[14] (Figure 3) we can see that geo-ontology used is the one standardised under GeoSPARQL. We can also notice, that information provided is very limited, and would require work on extending it to the degree where it would bring true value, for consumption by CAVs.

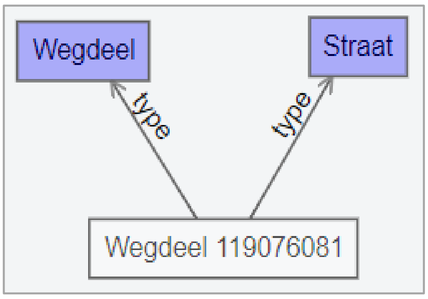


Figure 1 Road section linkage visualisation from TOP10NL

Road section
Smallest functional independent piece of road with constant, homogeneous characteristics and relationships for road traffic and air traffic on land.

HATE:	http://brt.basisregistraties.overheid.nl/def/top10nl#Wegdeel
Subclass of:	Top10nlObject
Has subclasses:	Motorway , main road , LokaleWeg , Overig wegdeel , parking , parking Carpool , ParkeerplaatsPR , RegionaleWeg , Rolbaan Platform , Runway Runway , Street ,Ferry line
Properties include:	aWegnummer , number of lanes , exit name , exit number , bridge name , eWegnummer , physical appearance , separated roadway , heart geometry , head geometry , head of movement , elevation , isBAGnaam , node name , nWegnummer , name , sWegnummer , status , tunnel name , type of infrastructure , paving width class ,hardening type
Inherited properties:	source current events , resource description , source accuracy , source type , the end of registration , identification , mutation type , object start time , object end time , tdnCode , time of recording , visualization code

Figure 2 "Road Section" (Wegdeel) class from TOP10NL

GeoSPARQL is standard for representing geospatial linked data [26], developed by OGC (Open Geospatial Consortium), the same consortium that standardised CityGML. On top of this ontology, TOP10NL vocabulary has been applied.

[12] http://streamreasoning.org/resources/c-sparql
[13] https://data.pdok.nl/sparql#
[14] https://brt.basisregistraties.overheid.nl/top10nl/doc/wegdeel/119076081

```
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>.
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>.
@prefix ns1: <http://brt.basisregistraties.overheid.nl/def/top10nl#>.
@prefix ns2: <http://rdfs.org/ns/void#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>.
@prefix n1: <http://brt.basisregistraties.overheid.nl/top10nl/id/wegdeel/>.
        n1:119076081
        rdf:type ns1:Wegdeel;
            rdf:type ns1:Straat;
            ns1:verhardingstype <http://brt.basisregistraties.overheid.nl/id/begrip/Verhard>;
            ns1:hartGeometrie
        <http://brt.basisregistraties.overheid.nl/top10nl/id/geometry/AE7866024ECE0C4E38D72BE6E20393F7>
        ;
            ns1:typeInfrastructuur
        <http://brt.basisregistraties.overheid.nl/id/begrip/Verbinding_typeInfrastructuur_wegdeel>;
            ns1:status <http://brt.basisregistraties.overheid.nl/id/begrip/InGebruik>;
            ns1:fysiekVoorkomen <http://brt.basisregistraties.overheid.nl/id/begrip/OpVastDeelVanBrug>;
            ns1:brontype <http://brt.basisregistraties.overheid.nl/id/begrip/Luchtfoto>;
            ns1:hoofdverkeersgebruik
        <http://brt.basisregistraties.overheid.nl/id/begrip/GemengdVerkeer>;
            ns1:verhardingsbreedteklasse <http://brt.basisregistraties.overheid.nl/id/begrip/_2-
        4Meter>;
            ns1:hoofdGeometrie
        <http://brt.basisregistraties.overheid.nl/top10nl/id/geometry/05A211AE3F3177EA816D98656F9CFB82>
        ;
            ns1:hoogteniveau 0;
            ns1:visualisatieCode 10601;
            ns1:bronnauwkeurigheid "0.1"^^xsd:float;
            ns1:gescheidenRijbaan false;
            ns1:tijdstipRegistratie "2016-11-01"^^xsd:date;
            ns1:bronactualiteit "2015-01-01"^^xsd:date;
            ns1:bronbeschrijving "Een orthogerectificeerde fotografische opname van een deel van het
        aardoppervlak. Gemaakt vanuit een vliegtuig.";
            ns1:objectBeginTijd "2010-07-02"^^xsd:date;
            ns1:tdnCode 353;
            rdfs:label "Wegdeel 119076081"@nl;
            ns2:inDataset <http://brt.basisregistraties.overheid.nl/id/dataset/top10nl>
```

Figure 3 SPARQL TTL response for resource "Wegdeel 119076081"

3 Problem description

In the age of Big Data, quite often we are confronted with the problem of extracting useful information from all that accumulated data. In other words, mining for knowledge can be quite complicated due to lack of structure and “richness” of, and between, different data sources.

Especially now, with the expansion of AI, it is crucial to provide machines with well-structured data, which contains useful information, so that connected devices can fully benefit from shared-knowledge and IoT connectivity.

One example of these machines, that enhanced with AI will change the way we move through our cities, are connected autonomous vehicles (CAVs). Almost all big car manufacturers [26] are working on creating their own versions of CAV, and they all have predictions to surpass Level 3 autonomy, between years 2020 and 2025. Of course, first models will focus on autonomous highway driving, and later they will evolve to urban complex environments.

By analysing previous section (Section 2) and mentioned models, information sources and agents, we can conclude that there are a lot of initiatives to standardise different domains within Smart Cites.

We can also see that, there is a lot research into physical connectivity for V2V (vehicle-to-vehicle) and V2I (vehicle-to-infrastructure), extending into concept of V2X or V2E (vehicle-to-everything).

However, to the author’s best of knowledge, there is a gap when it comes to standardised data model of urban traffic infrastructure, which would include, not only model of physical infrastructure, but also sensory information from relevant IoT systems and connected vehicles.

This model would have to be organised in such a way, that CAVs, which we can refer to as *AI Agents*, can easily extract information, which they can, with a very little processing use in order to optimise their actions while minimally stressing communication networks’ resources.

Taking into account complexity of computations that AI algorithms for self-driving vehicles have, due to great amount of sensory inputs and limitations of hardware in CAVs in terms of processing power, external inputs, from V2V and V2I communication, should be presented in such a way, so that they require minimum processing, and therefore minimal computational resources. Only in this case, information provided by various sources will, indeed, be vehicle friendly.

This would be especially important within urban environments, where density of traffic is greatest. In addition to this, in cities vehicles must interact with much more different agents, (pedestrians, cyclists, etc.), which don’t have such predictable behaviour, therefore putting extra pressure on limited hardware. Also, shared data, should be structured and codified in such a way that it requires minimal bandwidth for transmission.

By solving the problem of data structuring, we would therefore enable efficient information sharing, which would lead to increase in safety, comfort, road and vehicle utilisation.

3.1 Use cases for research

During research, many potential use cases are expected to crystalize, therefore planning accounts for validation of three of them. Examples of use-cases to be tackled are presented below.

3.1.1 Optimal Speed Recommendation (Use Case 1)

A vehicle is approaching the traffic light that is currently displaying red. A local RSU (Road Side Units) or the central server connected with local vehicles, shares relevant information with this vehicle, traffic lights and other entities (data sources).

The local RSU or the central server then calculates a specific optimal speed for the vehicle based on the collected information to reach the traffic light at the beginning of the next green phase and if necessary to avoid potential dangers and avoid increased energy consumption and therefore contribute to lowering pollution. The resulting recommended speed is finally delivered to the driver who can avoid the unnecessary stop-start scenario at the traffic light, mitigating potential risks (e.g. accident happened) and unnecessary environmental impact.

3.1.2 Traffic Accident Consequence Mitigation (Use Case 2)

The Intelligent Transportation System (ITS) is mainly used for avoiding collision of vehicles. Although it is for sure that the ITS will save a lot of lives, still some traffic accidents will occur any way. So rescue teams are still required to go to the accident sites to help the victims and police to ease the traffic jam caused by the accident.

As it is known that time is curtail when it comes to medical interventions, ITS will have to not only provide the fastest route to Emergency Vehicles, but quite often, in cities, will have to create or unblock certain route. Therefore System will have to re-route all other vehicles, and provide Information to police about current state of traffic, so that bottlenecks could be resolved or avoided. This process of changing traffic flows would have to be initiated immediately after accident occurs, in order to avoid jams that would subsequently delay emergency response services.

4 Aim and objectives

The aim of this research would be to identify technologies and methodologies and develop semantic linked data model, together with required ontologies, of an urban traffic infrastructure, which would be CAV (Connected autonomous vehicle) and network friendly.

4.1 Research gap

As demonstrated in previous sections, there are multiple initiatives to create uniform C-ITS systems, which would in the future, provide information that AVs (autonomous vehicles), could use, in optimizing their handling characteristics. Unfortunately there is a gap, when it comes to integrating physical characteristic of road, with traffic information provided by various ITS databases and physical sensors and connected vehicles.

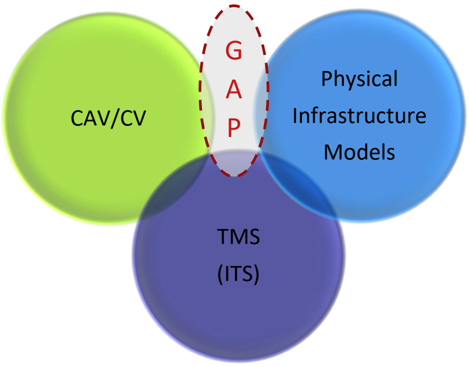


Figure 4 Illustration of missing link between CAVs and Physical Infrastructure in a digital domain

From the other hand we can see clear push for standardization of ITS solutions and traffic information sharing platforms. Clearly need for sharing information has been identified by all parties, including regulatory bodies of EU, US, and other states, which are funding these projects. Currently these are mostly focused on "Day 1 scenarios"[14], but if we look at the predictions[27], in few years, ITS, would have to involve much more knowledge sharing in order to accommodate AVs. Most promising development have, so far, been done within "Km4City" project[15], in which many ontologies, related to Smart Cities have been developed[28]. In addition, during this project, teams developed methodologies for comparison of different technologies, for smart city uses[29], [30]. Unfortunately, although this projects tackles sensory infrastructure, there are no indications, at this stage, that CAV's will be included. In addition, road infrastructure has been represented in a way author believes is not detailed enough or structured in a way that it would make CAVs fully benefit from it. Similar is with research done at TTI (Toyota Technological Institute), where 3 core ontologies have been developed. These ontologies represent vehicle control, map and vehicle itself, and as such cover only limited amount of knowledge.

4.2 Closing research gap

As, to the author's best knowledge, there is no literature currently tackling integration of physical traffic infrastructure data model and IoT traffic relevant connected devices with connected vehicles, or with a goal to support AVs. Therefore research that is being proposed would aim to close the research gap, and prove that SWT can be used to structure data relevant to CAVs, in such a way that it will:

i. Support processing of the shared information, by AI (Artificial Intelligence), in the most optimal way, using AV hardware.
ii. Utilise the network bandwidth in the most efficient way, while satisfying first condition for sharing of the needed information.

Main purpose of using SWT and linked data, would be to provide machine-readable structured data, describing road's physical characteristics, together with current traffic conditions, so that AI in-

[15] http://www.km4city.org/?controlRoom

charge of handling autonomous vehicles, could consume this information in order to achieve more optimal calculation of control parameters of the vehicle while using minimal amount of network resources.

By conducting proposed research, previously mentioned gaps (Section 4.1) would be eliminated, and new step in development of ITS systems combined with the newly developed urban infrastructure linked data model, would be introduced, by using semantic web technologies, which hasn't been created, and therefore also not utilised for this purpose.

5 Research design and methods

Extremely promising approach to solving the problem of gap, in integration of physical characteristic of transport infrastructure (like Kadaster's Linked data) with C-ITS systems, is usage of Semantic Data structures, which can bring the necessary meaningfulness to existing data[31]. Through usage of suitable ontologies, possibilities of expanding the knowledge, by harnessing information from various Linked Open Data sources[32], can bring so far unlocked value to all stakeholders within the urban ecosystem.

5.1 Research questions

Main hypothesis (to be proven) would state that:

- Semantic web technologies can effectively structure heterogeneous data from multiple sources to allow for efficient data extraction and processing by CAVs.

With main research question being:

- How can SWT be used to structure data, so that information retrieval operations, performed by vehicle's AI, would be most optimal from the point of in-vehicle resources and network bandwidth usage?

In order to answer the main research question and achieve goals set (in chapter 4 above) number of sub-questions would have to be answered first:

1. Which information, regarding the traffic infrastructure, can contribute to more optimal vehicle control algorithms, from the perspective of performance? And what are the sources of this information?
2. How can this information be structured to enhance CAV's sensory information processing?
3. How can vehicle's sensory data be incorporated into previously mentioned data model of traffic infrastructure, to provide environmental info to other parties?
4. What are the time-sensitivity constrains for previously mentioned data?
5. Which time and spatial granularity of data leads to the most efficient in-vehicle processing?
6. What are possibilities for utilising SWT in order to incorporate previously mentioned models with systems, like DSH (Data Service Hub by KPN), that already in-place?

5.2 Research design and methods

Intended research would be divided into four phases. These phases would be organised in into three separate parts, and summary, which will culminate by PhD thesis.

First three phases would be connected also by iterative feedback loop, therefore they will not have direct linear relation with time scheduling.

First phase would focus on research of the matter in depth through review of literature, in order to answer 3 questions posed in Figure 5.

Second phase, is focusing on development of ontologies, in 3 sub-phases:

I. Static part
II. Dynamic (geostationary) data part
III. Vehicle sensory, non-geostationary ("Floating") data part

Third phase will focus on validation of development, which came as result of work in second phase, based on use-cases proposed, by using algorithms identified as interesting in the first phase of research.

Final phase will focus on summarising the results and is expected to deliver Doctoral Thesis as outcome.

More details on research design are shown in Figure 5.

5.3 Ontology generation, mapping & evolving methods

Within the Phase 1 of the research, main challenge would be to identify methodologies to support generation of new (parts of) ontologies, reuse of already existing development and their maintenance (evolving).

In the domains of building and extending ontologies, few projects deal with methods for generation and mapping [33], [34]. Most prominent ones are:

- InfoSleuth (MCC) [35]
- SKC (Stanford)[16]
- Ontology Learning (AIFB)[17]

Within generation methods, state-of-art are semi-automatic and automatic ontology creation methods, based on concept and relation extraction, while within ontology mapping area mentioned methods can't be automatized in most cases [34].

Development methodologies, described in the literature, in more details are rare. Based on literature review from 2013[36], almost all methods proposed are still in not fully-mature phase, and do not provide sufficient, formal, descriptions on how can they be used. Most promising one, according to authors of the review is METHONTOLOGY [37].

Methodologies given in literature also relate to ontology maintenance or evolving, where it is necessary to mention, that process, according to literature [34] still has to be done manually or in specific cases in a semi-automatized way. But main remark remains, toward all methods, that possible evolving of ontology has to be taken into consideration during generation phase.

Choice of one or several methodologies will crystallize during further research, in conjunction with available data and development already done by other researchers and organizations.

5.4 Perspective data sources

As a crucial requirement and enabler for development of data model, existing data sources have to be identified. Up to this moment, there are three sub-groups of data, that would be required and that correlates to three sub-phases of second (development) –phase of research. Public (governmental) sources for this data have been identified and classified in the next table:

[16] http://infolab.stanford.edu/SKC/
[17] http://www.aifb.kit.edu/web/Ontology_Learning/en

Table 1 Governmental data sources & sets

Sub-Phase	Source	Format	Standard	Description	Remark
I. Geospatial data	TOP10NL (Kadaster)	GML, Linked data	GML	most detailed product within the BRT	Incomplete – Still in development
II. Dynamic (geostationary) data	NDW.nu Traffic data	XML	DATEX II	Measuring locations (MST), Travel times, Speeds and intensities, Work and events, Bridge openings, Current flow information…	Samples only
III. **Vehicle “floating” data**	KITTI project (by KIT)	Txt, CSV	///	3D GPS/IMU data	Data from KITTI sensor suite only

As this research is part of SmartONE strategic partnership, all participant are willing to share their data and services. Collaboration with partners will open new data for research purposes, as the research progresses and need for such data arises.

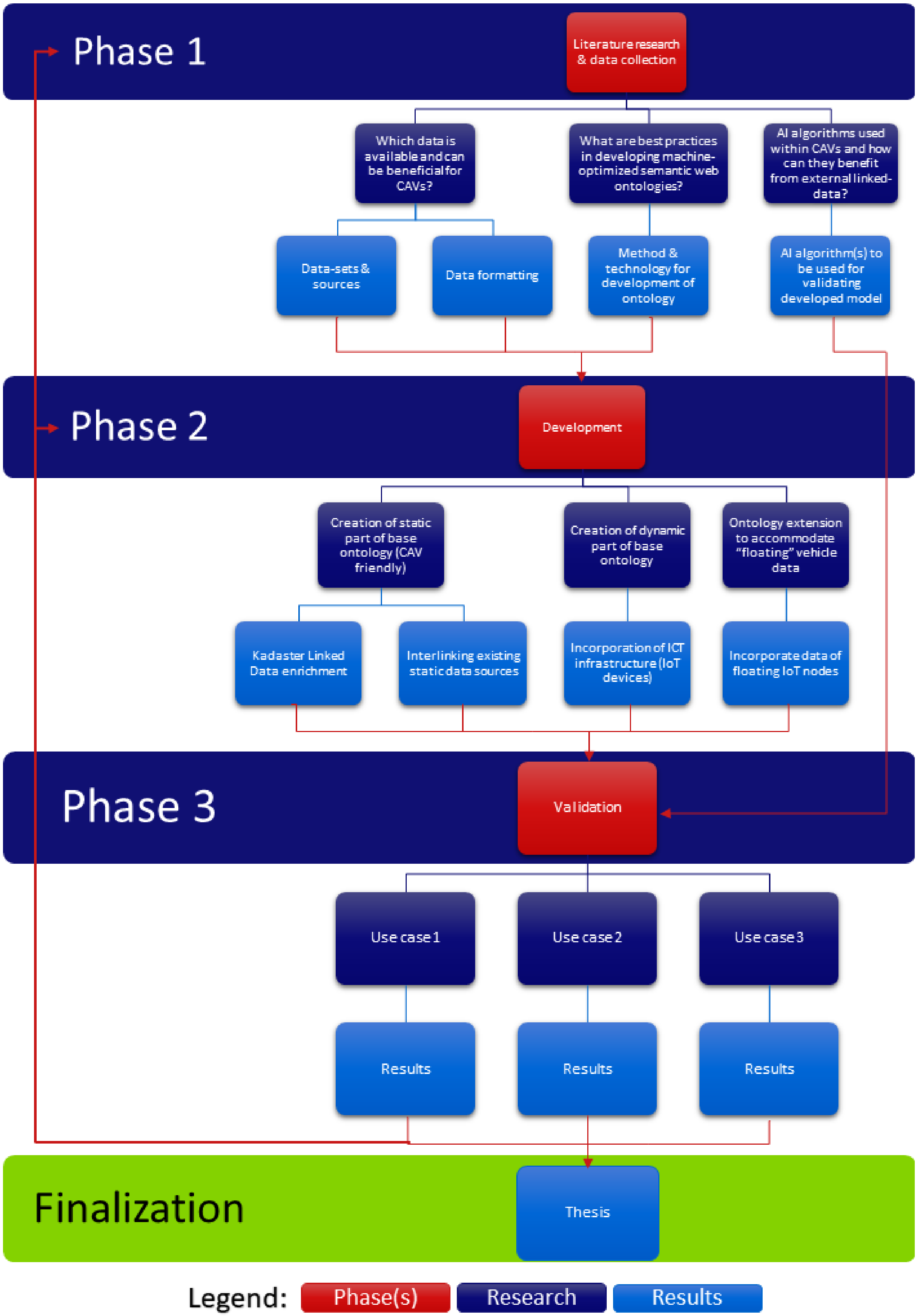


Figure 5 Research plan in Phases

6 Framework

As urban environments have different driving characteristics, approaches developed for "open road" conditions, will have to be re-designed in order to satisfy significantly more dynamic environments, in which, due to geography, information processing time will have to be significantly lower.

Challenges regarding such a model arise, due to a fact that data that is being streamed from the moving objects, has to be interlinked with geostationary data sets, in order to enable possibility of valuable information extraction.

Work on developing of proposed model(s) would have two conceptually separate parts:

1. Development of model based only on geostationary information (vehicles excluded)
    a. Creation of base ontology (self-driving car friendly)
    b. Extraction of information from City Models and Maps into semantic data networks based on previously designed ontology, using AI, and/or other suitable techniques.
    c. Incorporation of ICT infrastructure, through already existing standards for representing IoT devices in form of semantic data
2. Extension of the model and methodologies in order to incorporate data sources that are not stationary, and therefore, have to be dynamically linked to different nodes depending on their location info, as well as on type of data they are producing.

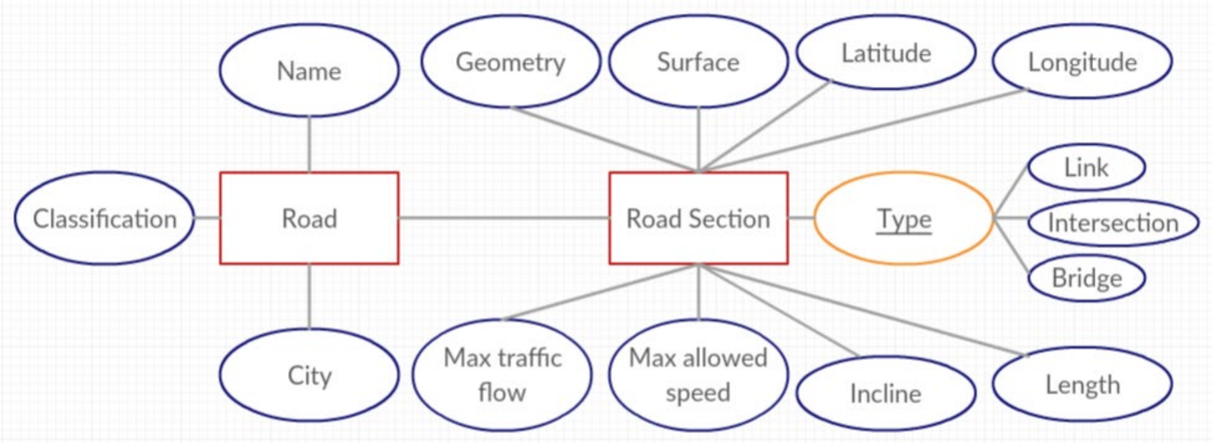


Figure 6 Example of geostationary semantic road model comprised of sections

In addition, when designing such models, thing to keep in mind is that they will be consumed by different AI algorithms, and therefore they have to be optimized for this.

6.1 Expected impact and contribution

Results of this research are expected to prove that SWT can provide a solution for efficient data integration and sharing, therefore tackling the mentioned research gap in academia. In addition to academic contribution, ontologies developed during this research will be result of multi-stakeholder collaboration, consisting of both academic, governmental and representatives of industrial sector.

As mentioned developments will be done in alignment with existing standards, interoperability between currently available information sharing standards would be easily achievable.

Such developed ontologies will in their nature be open for use, and therefore will provide developers and researchers means to adapt CAVs, so that they can consume available data. In this scenario vehicle manufacturers, would directly profit, as there would be a simple and standardised way to share vehicle important data.

In addition, achieving research goals would also mean that such information sharing systems would be efficient, when it comes to bandwidth utilization. Therefore providers of connectivity to such vehicles, will benefit from usage of developed linked data model.

Municipalities, and other governmental bodies, would be finally be able to have a proper ways of communicating with vehicles, which would make it easier to achieve greater comfort for everyone, especially within urban areas.

7 Deliverables and publication plan

In addition to the planned dissertation, the project will result in a number of papers to be submitted to international conferences and journals. Plan is to complete every mayor phase with at least one publication.

The work should be of interest for scholars working on information modelling and AI applications for Smart Cities, as well as professional practitioners in different fields. Papers related to AI, semantic data, Mobility and Smart Cities can be published in scientific journals like:

- “Sustainable Cities and Society“
- “IEEE Access”
- “IEEE Intelligent Transportation Systems Magazine“
- “IEEE Transactions on Knowledge and Data Engineering”
- “Ambient Intelligence and Humanized Computing”
- “Journal of Web Semantics”
- “Semantic Web journal (by IOS Press)”

And can be also presented at conferences such are: “Smart City Expo World Congress”, “International Conference on Ambient Systems, Networks and Technologies”, “Smart Cities International Symposium & Exhibition”, ”IOT World Congress”, “International Semantic Web Conference (ISWC)” “European Semantic Web Conference (ESWC)“ and many others.

7.1 Publication plan

Planned publications should be focused on next topics:

1. Linked data as enabler for Smart Cities – Case: Intelligent Mobility
2. CAV friendly ontology for representation of physical traffic infrastructure
3. Semantic data model of traffic infrastructure for AI-controlled vehicles within Smart Cities
4. Information extraction from Semantic data model of traffic infrastructure

7.2 Planning and tasks

Table 2 Research tasks planning

Phase	Task *Intended publication finish is marked with a symbol “↙”	Year 1				Year 2				Year 3				Year 4			
		2018		2019				2020				2021				2022	
1	Literature Review																
1	Data collection																
1	Methodologies & technologies review for Semantic Data				↙												
1	Research CAV’s AI algorithms and technologies																
2	Development of Ontology [Static data]																
2	Development of Ontology [Sensor geostationary data]																
2	Development of Ontology [Vehicle “Floating” data]																
3	Validation of results on Use Case 1							↙									
3	Validation of results on Use Case 2										↙						
3	Validation of results on Use Case 3														↙		
Final	Thesis																

8 Budget, team and collaboration

Proposed research has been funded by SmartONE strategic partnership, which is collaboration framework between TU Eindhoven and KPN. Scope of framework includes 4 PhD candidates in starting phase, with the plan to grow bigger in the upcoming phases. These four candidates (out of which, one is the author of this proposal) are:

- Yuhao Wang – Data Science Center, Data Mining Group
- Ngoc Quan Pham – Institute for Photonic Integration
- Robbert Schulpen – Center for Wireless Technology
- Miloš Viktorović – Department of Build Environment, ISBE research group

The research team for the proposed topic will consist of the PhD candidate Milos Viktorovic and Professors Bauke de Vries and Nico Baken and Dr. Dujuan Yang as his promotors. In addition Dr. Edgar van Boven will support the research team, and provide the link between KPN, research team and other related stakeholders.

For specific topics and issues, they will be supported by members of SmartONE flagship, together with colleagues from other departments and groups from the university.

9 Table of figures

10 List of abbreviations

Abbreviation	***Explanation***
SWT	Semantic Web Technologies
AV	Autonomous Vehicle
CAV	Connected Autonomous Vehicle
AI	Artificial Intelligence
IoT	Internet of Things
GML	Geography Markup Language
GIS	Geographic Information System
ITS	Intelligent Transport System

Abbreviation	***Explanation***
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
V2X/V2E	Vehicle-to-Everything
DB	Data Base
DbW	Drive by Wire
PDOK	Publieke Dienstverlening Op de Kaart
ICT	Information & communication tech.
C-ITS	Connected ITS

11 References

[1] "The Self-Driving Car Timeline - Predictions from the Top 11 Global Automakers." [Online]. Available: https://www.techemergence.com/self-driving-car-timeline-themselves-top-11-automakers/. [Accessed: 28-Sep-2018].

[2] Marc Weber, "Where to? A History of Autonomous Vehicles | Computer History Museum," 2014. [Online]. Available: http://www.computerhistory.org/atchm/where-to-a-history-of-autonomous-vehicles/. [Accessed: 30-Jul-2018].

[3] C. Chen, A. Seff, A. Kornhauser, and J. Xiao, "DeepDriving: Learning Affordance for Direct Perception in Autonomous Driving."

[4] "AUTOMATED DRIVING LEVELS OF DRIVING AUTOMATION ARE DEFINED IN NEW SAE INTERNATIONAL STANDARD J3016," 2014.

[5] H. of Lords Science and T. Committee, "HOUSE OF LORDS Science and Technology Select Committee 2nd Report of Session 2016-17 Connected and Autonomous Vehicles: The future?"

[6] G. D'Aniello, M. Gaeta, and F. Orciuoli, "An approach based on semantic stream reasoning to support decision processes in smart cities," *Telemat. Informatics*, vol. 35, no. 1, pp. 68–81, 2018.

[7] P. Bellini, M. Benigni, R. Billero, P. Nesi, and N. Rauch, "Km4City ontology building vs data harvesting and cleaning for smart-city services," *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 827–839, Dec. 2014.

[8] G. Gröger, T. Kolbe, C. Nagel, and K.-H. Häfele, "OGC City Geography Markup Language (CityGML) Encoding Standard," *Ogc*, pp. 1–344, 2012.

[9] T. Becker, C. Nagel, and T. H. Kolbe, "Semantic 3D Modeling of Multi-Utility Networks in Cities for Analysis and 3D Visualization," Springer, Berlin, Heidelberg, 2013, pp. 41–62.

[10] R. de Laat and L. van Berlo, "Integration of BIM and GIS: The Development of the CityGML GeoBIM

Extension," Springer, Berlin, Heidelberg, 2011, pp. 211–225.

[11] K. Chaturvedi and T. H. Kolbe, "INTEGRATING DYNAMIC DATA AND SENSORS WITH SEMANTIC 3D CITY MODELS IN THE CONTEXT OF SMART CITIES."

[12] "C-ITS DEPLOYMENT IN THE NETHERLANDS."

[13] "COMMITTEE AND THE COMMITTEE OF THE REGIONS A European strategy on Cooperative Intelligent Transport Systems, a milestone towards cooperative, connected and automated mobility."

[14] "EUROPEAN PILOTS FOR INTEROPERABLE C-ITS SERVICES."

[15] "Road Infrastructure Publication Extension Proposal Introduction."

[16] M. J. Yoo, P. Kolyvakis, D. Kiritsis, D. Werthmann, and R. Hellbach, "Semantic model for IoT-Enabled electric vehicle services: Puzzling with ontologies," *Proc. - 2016 IEEE 4th Int. Conf. Futur. Internet Things Cloud, FiCloud 2016*, no. 688203, pp. 387–392, 2016.

[17] "DATEX II model." [Online]. Available: http://d2docs.ndwcloud.nu/_static/umlmodel/index.htm. [Accessed: 16-Jul-2018].

[18] "Linked Data - W3C." [Online]. Available: https://www.w3.org/standards/semanticweb/data. [Accessed: 30-May-2018].

[19] "What is Semantic Data." [Online]. Available: http://www.semagix.com/what-is-semantic-data.htm. [Accessed: 30-May-2018].

[20] A. K. Dey, "Understanding and Using Context."

[21] A. Zimmermann, A. Lorenz, and R. Oppermann, "An Operational Definition of Context," in *Modeling and Using Context*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 558–571.

[22] M. Compton *et al.*, "The SSN ontology of the W3C semantic sensor network incubator group," *J. Web Semant.*, vol. 17, pp. 25–32, Dec. 2012.

[23] K. Janowicz, A. Haller, S. J. D. Cox, D. Le Phuoc, and M. Lefrançois, "SOSA: A lightweight ontology for sensors, observations, samples, and actuators," *J. Web Semant.*, Jul. 2018.

[24] "Linked Data - PDOK." [Online]. Available: https://www.pdok.nl/linked-data?articleid=1976855. [Accessed: 23-Aug-2018].

[25] Kadaster Nederland, "Basisregistratie Topografie: Catalogus en Productspecificaties," vol. 2.2, p. 237, 2014.

[26] "The Self-Driving Car Timeline - Predictions from the Top 11 Global Automakers." [Online]. Available: https://www.techemergence.com/self-driving-car-timeline-themselves-top-11-automakers/. [Accessed: 30-Jul-2018].

[27] C. Morales Paulin *et al.*, "Connected Car Industry report," 2013.

[28] P. Bellini, P. Nesi, and M. Soderi, "Km4City - The Knowledge Model 4 the City Smart City Ontology," 2018.

[29] P. Bellini, I. Bruno, P. Nesi, and N. Rauch, "Graph databases methodology and tool supporting index/store versioning," *J. Vis. Lang. Comput.*, vol. 31, pp. 222–229, Dec. 2015.

[30] P. Bellini and P. Nesi, "Performance assessment of RDF graph databases for smart city services," *J. Vis. Lang. Comput.*, vol. 45, pp. 24–38, Apr. 2018.

[31] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Sci. Am.*, vol. 284, no. 5, pp. 34–43, 2001.

[32] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives, "DBpedia: A Nucleus for a Web of Open Data."

[33] Y. Ding and S. Foo, "Ontology research and development. Part 1 – a review of ontology generation,"

Journal of Information Science, pp. 123–136, 2001.

[34] Y. Ding and S. Foo, "Ontology research and development. Part 2 - a review of ontology mapping and evolving," *J. Inf. Sci.*, vol. 28, no. 5, pp. 375–388, Oct. 2002.

[35] R. J. Bayardo *et al.*, "InfoSleuth: Agent-Based Semantic Integration of Information in Open and Dynamic Environments."

[36] R. Iqbal, M. Azrifah Azmi Murad, A. Mustapha, and N. Mohd Sharef, "An Analysis of Ontology Engineering Methodologies: A Literature Review," *Res. J. Appl. Sci. Eng. Technol.*, vol. 6, no. 16, pp. 2993–3000, 2013.

[37] M. Ferndndez, A. Gomez-Perez, and N. Juristo, "METHONTOLOGY: From Ontological Art Towards Ontological Engineering," 1997.