



HAL
open science

Issues in Implementing a Data Integration Platform for Electric Vehicles Using the Internet of Things

Martin Smuts, Brenda Scholtz, Janet Wesson

► To cite this version:

Martin Smuts, Brenda Scholtz, Janet Wesson. Issues in Implementing a Data Integration Platform for Electric Vehicles Using the Internet of Things. 1st IFIP International Internet of Things Conference (IFIPIoT), Sep 2018, Poznan, Poland. pp.160-177, 10.1007/978-3-030-15651-0_14 . hal-03217376

HAL Id: hal-03217376

<https://inria.hal.science/hal-03217376v1>

Submitted on 4 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Issues in Implementing a Data Integration Platform for Electric Vehicles using the Internet of Things

Martin Smuts^{1✉}, Brenda Scholtz¹, and Janet Wesson¹

¹ Nelson Mandela University, Port Elizabeth, South Africa
{s210035447, Brenda.Scholtz, Janet.Wesson}@mandela.ac.za

Abstract. The emergence of the Internet of Things (IoT) has brought new improvement and development opportunities to the automotive industry, such as electric vehicles (EVs). EVs are well-known for their short ranges and many studies have reported on the challenges of trip planning and accurate remaining driving range (RDR) estimation. While the demand for connected vehicle applications and its enabling technology has progressed significantly in recent years, there are several constraints for connected and collaborative vehicle application deployments. Data integration issues are currently hindering the development of effective trip planning and RDR estimation solutions for drivers of EVs. Additional constraints have been identified in developing countries, including lack of charging station networks, EV data sources, and software applications. The purpose of this paper is to report on some of the main issues hindering EV data integration, as well as to report on an implementation in South Africa of a Data Integration Platform for EV data using the IoT. The findings show that data integration issues primarily relate to data availability, data quality, and interoperability between devices, IoT platforms, and EV service providers. The paper also identifies enabling technologies, drivers, and future directions for researchers in the IoT and EV domains.

Keywords: Internet of Things, Internet of Vehicles, Connected Vehicles, Electric Vehicles, Trip Planning, Telematics, Vehicular Networks.

1 Introduction

The Internet of Things (IoT) presents a range of potential opportunities for the improvement and development of Electric Vehicles (EVs). The emergence of the IoT has brought sensor-based “intelligent” technologies, high bandwidth networking systems, and cheaper memory to EVs, which has enabled more detailed travel/traffic data exchanges between transport infrastructure, mobile phones, and navigation systems [1]. Sensing devices transmit data on a massive scale and, in some cases, adapt and react to changes in the driving environment automatically as infrastructures, cloud computing models, and machine learning algorithms can manage real-time streams of EV and smart grid data [2]. Service providers and auto manufacturers, however, struggle to access all the required data as it differs in accuracy, resolution and structure [3], and they haven’t mastered the technologies needed to capture and analyze the valuable data

at their disposal [2]. A need exists for higher investments in vehicular technologies and infrastructure. The need for accurate trip planning applications is growing since their usage can reduce drivers' range anxiety [3]–[6].

One issue with EVs is the inaccuracy of the Remaining Driving Range (RDR) in on-board displays [3,7]. The RDR is the distance an EV can cover with the energy stored in the battery at any given time, whether fully charged or not. RDR displays are important as they can help drivers with “trip planning”, but are insufficient by themselves as the battery's residual energy can be utilized in many ways. Other issues relate to limited battery capacity, availability of public charging stations, and long charging cycles [8]. Existing on-board RDR calculations only consider a limited number of factors, and often ignore the effects that driving behaviour, weather, traffic, and terrain have on future trips [3,7]. EVs also suffer from a lack of suitable data integration technologies within the EV infrastructure [4,9]. A lack of standards and methodologies has led to a proliferation of proprietary systems and data formats, with limited interoperability [4,10]. These systems are tailored to work only within specific scenarios, e.g. in a given geographic area with specific target users or EV models [4]. Many connected platforms are not supported in all countries and/or lack consistency and reliable information [11]. For example, when charging station applications are not updated with the latest locations of chargers, drivers could potentially avoid taking a trip, as they fear they will not reach their destination. Drivers require a “trip plan” that assists them to drive in the most efficient way to obtain a desired range and trip time [4].

The lack of connectivity between EVs and software platforms cause these systems to remain unintegrated and force drivers to use several software products simultaneously. Drivers are forced to develop their own coping mechanisms such as finding charging possibilities during the day, plan detours, or use alternative transport modes [12]. These problems require an extension of traditional in-vehicle dashboard displays and navigation systems to consider EV limitations and drivers' preferences [1,11]. Integrating EVs and the IoT can benefit from the virtually unlimited capabilities and resources of cloud computing to compensate for EVs' technological constraints (e.g., storage, processing, communication) [13]. Cloud technologies provide the ideal solution to manage various data sources in a distributed and dynamic manner. The above mentioned issues are worsened in developing countries such as South Africa, as limited infrastructure and EV management software is available [14].

The purpose of this paper is to address the gap in the literature related to data integration issues for EVs using the IoT. The paper reports on a literature review of data integration challenges related to EVs, with a particular focus on ‘connected vehicles’ and the IoT. The main goal of the paper is to propose a Data Integration Platform to successfully integrate EV data that can be used as a foundation to develop trip planning and RDR estimation applications. The research reported on in this paper forms part of a larger study that aims to develop a solution to estimate the RDR based on various factors (weather, route typology, traffic, and driver behaviour) to display useful information to drivers about all aspects necessary for trip planning and EV management. A case study methodology was adopted in conjunction with the Design Science Research methodology. The case study was an e-mobility research organization in South Africa. For purposes of anonymity the organization is referred to as Emobi.

The paper is structured to present the main contributions in chronological order: A short introduction to the IoT is provided to help readers understand the complex data integration and data management issues with EVs and the IoT (Section 2). A real-world South African case study is presented of an electric mobility (e-mobility) organization (Section 3). A Data Integration Platform is proposed that uses cost effective and easily available resources such as smartphones, plug-in devices, cloud computing, the IoT and data services for EVs (Section 4). The data integration issues encountered and lessons learned from the case study are related to security, availability, quality and interoperability, and confirmed some of those from literature (Section 5). Finally, the paper concludes with remarks and future research directions (Section 6).

2 Background: Internet of Things and Connected Vehicles

2.1 Connected Vehicles: A subset of the Internet of Things

The development of enabling technologies in modern vehicles does not only provide computation capabilities via embedded chips and remote clouds, but also support by mobile and complex networks [15,16]. These networks are capable of sensing, wide-area connectivity, inference, and action consisting of up to 70 electronic control units (ECUs) capturing more than 2500 signals for the chassis, powertrain, user interfaces and safety networks [9]. The unity of the underlying technologies enable ‘Connected Vehicles’ or Vehicular Communication Networks (VCN), which fall into a category known as the Internet of Vehicles (IoV), a subset and indispensable member of the IoT paradigm [2,15,16]. The IoV allows for an environment where vehicles are equipped with dedicated on-board units (OBUs), capable of communicating with other vehicles, Vehicle-to-Vehicle (V2V) connections, and receiving data services from infrastructure (e.g. smart grids, road side units), cellular base stations, and Wi-Fi access points regarded as Vehicle-to-Infrastructure (V2I) communications [15,16]. The IoV relates to the scenario where drivers, passengers, and pedestrians can enjoy services provided through the Internet.

2.2 Enabling Technologies

For an EV to have an integrated data platform five types of technologies should form part of the connected vehicle environment [9]. These are: *Sensing, Intravehicular Connectivity, Intervehicle Connectivity, Inference, and Action and Feedback.*

Sensing. Electronic Control Units (ECUs) consist of various control modules, such as Engine Control Module or Brake Control Modules. The ECUs are typically embedded with software and sensors that monitor the internal systems of the vehicles. Vehicle sensors and systems fall into two categories: internal sensor systems and external sensor systems [17]. Internal sensors monitor the performance of vehicles such as wheel speed, yaw rate, steering inputs, driver inputs, powertrain outputs, or hydraulic braking [17]. Other internal sensors, specifically for EVs, can monitor battery state of charge (SoC),

battery temperature, and battery current and voltage from the Battery Management System (BMS) [7]. External sensor systems have grown exponentially in recent years and are focused on enhancing driver safety, perception and autonomous navigation. These systems include a combination of camera, GPS, RADAR, LIDAR, Ultrasonic, and Dedicated Shortrange Communications [17,18].

Intravehicle Connectivity. Intravehicle networks, or internal networks, are purposely built to share data among the different sub-systems, ECUs, sensors, and actuators to facilitate the operation of a single vehicle [9,19]. Standardized intravehicle networks, such as the Controller Area Network (CAN), Local Interconnect Network (LIN), FlexRay, and Media Oriented Systems Transport (MOST) are well documented and allow different technologies to communicate with each other [19,20]. The purpose of these networks is to ensure that On Board Diagnostics (OBD) services are readily available for drivers and technicians to monitor vehicle performance and health by offering fault tolerance, determinism, and flexibility [9]. Although the sensors and networks are proprietary to original vehicle manufacturers, these technologies typically follow standards that allow for vehicle diagnostics and future connected applications via interfaces, allowing external technologies from the vehicle network to communicate with the vehicle [19].

Intervehicle Connectivity. Intervehicle connectivity relates to networking approaches that allow data to move from within the vehicle to remote computing devices and other cloud computing infrastructures [9]. While local applications are contained within the vehicle, remote applications may make use of the vehicle data and combine it with data from external sources, such as traffic or weather data. Advancements in data collection and wireless transmission via telematics units or plug-in devices use cellular networks to share data among vehicles and infrastructures to facilitate data collection and optimization. Numerous competing IoT and intervehicle network standards and protocols exist, which can be a problem when considering interoperability and integration between vehicle and infrastructures [10]. Although a detailed discussion of different intervehicle networks are beyond the scope of this paper, popular networking communications found in connected vehicles are either categorised as mesh networks or cellular networks [8,9].

Cellular networks are ideal for machine-to-machine (M2M) connectivity as they are commoditized and ubiquitous [21]. They also offer benefits of being robust, long range, and are capable of sharing data as parallelized streams when traffic density is sparse [8,13]. As cellular networking technologies such as 4G/5G and LTE expand and costs fall, vehicles may rely on cellular technologies to facilitate connectivity for critical applications and often connect to users' personal devices [8,15,21].

Indirect cellular connectivity can refer to Bluetooth, Wi-Fi, and other vehicle interfacing hardware devices or visualization tools [15]. The signals from most sensors and systems flow through the CAN and can be captured through the OBD interface/port using either a wireless dongle [18] or wired equipment [22]. The dongles are usually equipped with Bluetooth or Wi-Fi functions that are able to interface with smartphones or computers, allowing data to be extracted over HTTP(S) [13,22].

Inference. The combination of sensors and connectivity in EVs is producing massive amounts of valuable data. The advancements in in-vehicle and cloud computing

power, as well as scalable data handling platforms, has made aggregation and synthesis easier in recent years. Cloud computing poses the perfect supplement for scalable server-side processing to transfer in-vehicle processing to remote locations [13]. Analyzing EV data with other vast datasets for insights can provide critical new services for EVs and their drivers. By applying big data and machine learning tools, applications can assess, learn, and adjust EV operations, and serve as a foundation for larger applications to make informed decisions. For example, to use live GPS and EV performance data and combine it with third-party devices like smartphone apps to predict or predict driving behaviour [18]. While in-vehicle analytics demonstrates the value of data in controlling EV functions in real-time, remote analytics demonstrates the potential to apply large-scale connectivity, computation and distributed information toward improving vehicle efficiency, reliability, and performance [9].

Action and Feedback. The data insights and intelligence of ‘inference’ technologies will enable data-informed control over the connected vehicle to attain maximum impact on the IoV environment. The control can either be a direct or an indirect approach [9]. A direct approach relies on ECU controllers and networked data to manipulate the vehicle functions. For example, light sensors will switch on the vehicles headlights automatically if daylight diminishes. An indirect approach uses the data from the sensors and other sources to provide feedback to a human operator on an in-vehicle display, or a second-screen interface to provide occupant feedback [9]. Examples of in-vehicle displays are used to monitor and improve energy economy at a glance. Second-screen feedback displays are typically found on smartphones, tablets, and more recently, smart watches, to increase the level of interactivity drivers have with their vehicles and allow applications to run on upgradable hardware [4,8,23].

2.3 Drivers, Platforms, and Applications of the IoT for Electric Vehicles

The evolving market of IoT is expected to offer promising solutions to transform transportation systems and automobile services [13]. In the context of EVs, solutions are needed to better communicate with their users, charging stations, and utilities to effectively manage energy resources [8]. Applications and platforms that connect EVs, users and infrastructure have been created to address this need. Telematics applications have been developed to integrate EVs into the IoT environment to provide applications such as roadside assistance, remote door unlocking, charging activity feedback, navigation services and collision notifications [9]. Many automobile manufactures allow drivers to check the status of their EVs and remotely control their charging through mobile apps [8]. Some examples of smart charging software applications are ChargePoint, PlugShare, and BMW iDrive [8]. Other researchers have combined IoT and cloud computing with machine learning approaches to identify driving behaviour [18], estimate the RDR and battery SoC [3], V2I applications and charge recommendations [16], predict cost of charging [6], trip planning decision support and navigation suggestions [4], and calculation of CO² emissions [22].

3 Research Methods and Contributions

The research reported on in this paper adopted the Design Science Research (DSR) methodology of Johannesson and Perjons [24]. A case study approach was used in conjunction with DSR and the case study was an e-mobility organization in South Africa, which for anonymity purposes is referred to as Emobi. The case study illustrates the application of the IoT in the EV domain and provides a real-world context for implementing the proposed EV Data Integration Platform using the IoT to support a solution for integrating EV data. An interview was conducted with a senior engineer at Emobi to establish an overview of their EV environment and the data integration issues faced. Emobi owns a fleet of three Nissan Leaf model vehicles, which are used for general transport activities and experimentation activities. Emobi needs to collect data from the EVs to better manage their driving experience and improve their energy usage. Prior to this study, Emobi had no data integration platform. Staff at Emobi stated that they have no sophisticated navigation or trip planning system to help them deal with the limited range issues inherent to EVs. Staff keep track of drivers and trips by entering driver details and EV parameters manually in a logbook. These records are then given to a receptionist to type manually into a spreadsheet and to perform manual summary statistics, e.g. distance travelled.

To collect and store data, a GPS was installed into the EVs by a service provider (called LogCo). The available data was generated with a logger device recording the CAN bus signals of a 2015 Nissan Leaf (24 kWh Li-Ion battery, 80 kW electric motor, 1700 kg vehicle mass). The logger registered the battery's current and voltage, the SoC, the GPS coordinates, and the timestamp for a period of seven months. The data can be retrieved using a Simple Object Access Protocol (SOAP) API provided by LogCo in either a summarized format (e.g. duration, distance, energy used, path, date etc.) or in finer granularity that records data directly from the CAN bus (e.g. time stamp, latitude, longitude, speed).

Several novel contributions are made by this study when compared to existing literature. Prior studies do not consider all factors when estimating the energy consumption. This study considers five main factors as summarized in the authors' prior paper to estimate energy consumption [25]: weather (wind direction and temperature), route typology (highway vs urban routes, traffic), battery parameters (voltage, state of health, and SoC), as well as historic driving behaviour. Most studies focus on efficient routing algorithms based on graph-theory concepts to navigate drivers to destinations or chargers. This study proposes to integrate existing services and technologies (e.g. Google Roads API, ChargeNow API). To the best of the authors' knowledge, this is the first study that has proposed a Data Integration Platform based on a case study situated in a developing country, such as South Africa. Further, many studies are evaluated in simulated EV environments, whereas this study worked with a real-life e-mobility company to analyze challenges when implementing an EV data management platform.

4 Data Integration Platform for Electric Vehicles using the Internet of Things

The main, high-level objective of the EV Data Integration Platform is to provide EV drivers the tools to manage EV data and to receive services such as RDR estimations as part of a trip planning solution. The RDR estimations and trip planning services, such as charge recommendations, are enhanced by using machine learning algorithms to learn driver behaviour patterns. Deploying machine learning algorithms is important as energy consumption is heavily dependent on the driver behaviour. The platform will allow the driver to plan an EV itinerary considering spatial (route typology and length), temporal (duration of charging and route) and costing issues (cost of charging).

4.1 Overview and Requirements

The EV Data Integration Platform had three main requirements: (1) allow for data interoperability, (2) allow for service expandability, and (3) allow for device heterogeneity. In order to meet these three requirements, the platform incorporates the Microsoft Azure IoT and the HDInsight suite [26]. One reason for selecting the Azure IoT suite was that it provided various configuration and interconnected services to build large scale data analytics applications. Furthermore, Azure offered scalable and flexible services, such as facilitating a central IoT gateway hub, stream analytics engine, distributed databases, and configurable machine learning algorithms in the cloud. The Azure suite was critical for data integration and extensibility for developing trip planning services.

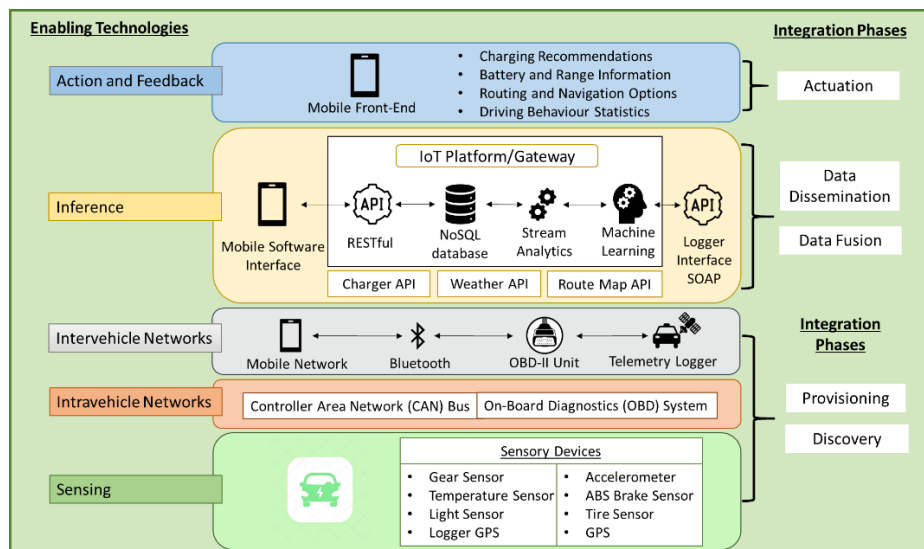


Fig. 1. Proposed Data Integration Platform for EVs using the IoT (constructed by author).

The proposed Data Integration Platform for EVs using the IoT (Fig. 1) is based on the five categories of enabling technologies (Section 2.2) and the five integration phases for integrating connected vehicles in the IoT ecosystem [10]. The platform describes the flow of data from the EV, the IoT, and cloud technologies through the five phases proposed in [10]: from the Resource Discovery phase, the Provisioning phase, to the Data Fusion, Data Dissemination and Actuation phases.

4.2 Resource Discovery Phase and Provisioning Phase

The Resource Discovery phase includes the configuration and uniform descriptions of available resources such as sensors, actuators and other associated OBUs that form part of the EV and its infrastructure. The platform mostly relies on the sensors installed in the Nissan Leaf EV, such as temperature, tyre pressure, light, rain, ABS braking, speedometer, and gear sensors. Third-party sensors are also considered such as the GPS installed via the logger unit and the smartphone accelerometer and GPS sensors. The measurements from these sensors are communicated via the CAN as raw data that need to be interpreted. This phase also includes the discovery of other sensors such as the GPS and accelerometers in the smartphone used, and the GPS from LogCo.

While the Resource Discovery phase retrieves a set of available sensors to provide raw data, the Provisioning phase monitors the rules, ontologies, protocols, and semantic reasoning for data to be transmitted via the networks. For example, the CAN coordinates the data that is transmitted from various sensors in the EV and enforces the permissions to prevent unauthorised devices to retrieve data from the network.

When adopting the platform in the case study of Emobi, two popular methods for extracting raw data were used as described in Wang et al. [18] and Tseng et al. [22]. The first method used a GPS logger installed by LogCo to extract data from the CAN. This approach is like the ‘sniffing tool’ approach where CAN message traffic is observed and requires knowledge of CAN protocols to program firmware to record and extract the CAN packets [22]. The second method, and an easier way to extract data, is to use an OBD dongle, which plugs into an OBD-II port generally situated in the cabin of the EV. The OBD dongle is programmed to interpret the CAN message traffic and parses them to return human readable parameters based on OBD conversion rules. The data were logged at a 5Hz frequency as in Wang et al. [18] and De Cauwer et al. [12].

4.3 Data Fusion Phase

The Data Fusion phase acts as a distributed knowledge base within the EV and IoT ecosystem. Data interoperability is important in this phase. It is worth noting that service providers (e.g. charging station providers) are responsible for their own data management platforms, which are characterised by their own data representation formats [4]. As mentioned before, the Data Integration Platform does not aim to replace existing systems but is rather aimed at serving as a data collector or integrated repository for EV data management. Raw data originating from the EV and other smart devices (smartphone GPS sensors) are integrated or ‘fused’ to attain a higher level of intelligence.

When adopting the platform at Emobi, the technical complexities of retrieving data from the CAN had to be considered. The choice was made to use a mobile application called LeafSpy, which has gained popularity in the EV community, to monitor Nissan Leaf diagnostics from a smartphone using an OBD dongle. LeafSpy has a large support community, allows a large variety of EV parameters to be traced from the CAN, and supports configuration settings to send data in JSON (JavaScript Object Notation) format to a server using a cellular network. This allowed for quick access to EV data. Although LeafSpy allows additional configuration to setup a server, a RESTful (representational state transfer) API had to be developed to divert the streamed data to the Azure IoT gateway server. This API was developed using the Node.js programming language, which is invoked when the data is sent from the smartphone.

Despite the cloud's scalability and relative low cost of operation, increasing computing power does not address data management challenges [9]. Traditional databases cannot handle real-time requests and require technologies that use distributed storage and parallel computing. For this reason, the data is streamed from the smartphone and is stored in a MongoDB instance (NoSQL database) hosted in Azure, which allows for quicker access. Due to popular issues with telematics devices being intermittent [13], the platform collects EV data from both the OBD dongle and the tracking unit (LogCo) as a backup for potential data loss. The benefits and challenges of having both devices are compared in **Table 1**. Nine criteria were used to perform this comparison: the graphical user interface, supported APIs for data access, data governance, flexibility, cost, mobile connectivity, data backup and driver identification. Effort refers to the amount of effort to configure the data collection approach [27]. Graphical User Interface refers to information displays about the EV or driver performance [28]. API support relates to additional web services that allow data to exchange between the EV and a server to either render data to an end-user device or another server for storage or processing purposes [4,13]. Data governance is the extent to which a service provider, or EV owner, is responsible for the installation, processing, and safeguarding of the data retrieved from the EV [9,27]. The cost relates to the installation cost and ongoing costs to retrieve and maintain the data [28]. Flexibility refers to the ability to manipulate the approach to collect, transform and use data for applications [19,22]. Data transport protocol refers to the protocol that is used to transport data from the EV to an end-user device or storage device. Driver identification relates to the ability to identify the driver [18].

The EV Data Integration Platform relies on web services from third parties to collect and integrate data for the five factors influencing RDR. Data sources were selected based on their ability to provide data on each of the following five factors: *Weather (W)*, *Route and Terrain (RT)*, *Vehicle Model (VM)*, *Battery Model (BM)*, and *Driving Behavior (DB)*. The available data sources were mapped to the data attributes that they contribute to as inputs to the machine learning and prediction algorithms (Table 2). The data sources for some of these factors, are webservice APIs available through RESTful formats (JSON). Other APIs that were used are the AccuWeather API and the Google API, which provided both XML and JSON data formats, as per Table 2. A smartphone application, called the Trip Planner App (TP-App) was developed as a user interface to

consume services from the Data Integration Platform and perform trip planning activities (e. g. input destination and driver ID).

Table 1. Comparison between LeafSpy OBD approach and LogCo tracking services.

Criteria	LogCo (Telematics CAN-Based)	Leaf Spy (OBD -Based)
Effort [27]	Once-off	On-going
Graphical user interface [28]	none	Live monitoring of EV data
APIs support [4,13]	SOAP/XML (Provided)	RESTful/JSON (self-managed)
Data governance [9,27]	Service provider (web service provided)	Self-managed (own server and database)
Cost [28]	Monthly subscription Initial installation	Mobile data costs Once-off
Flexibility [19,22]	Limited to technical knowledge and service level agreement	Limited to software
Cost [27]	Installation fee + monthly subscription	Initial purchase fee for equipment + cellular network
Data transport protocol [13,22]	Satellite-based (GPS/GSM)	Cellular network- based/HTTP
Driver identification [18]	none	none

AccuWeather. Weather data from hourly forecasts based on the location’s coordinates was obtained from the AccuWeather API [29]. The coordinates can be provided from the Google Maps API for the current location and segments of the planned route, if the destination is known. Data can be retrieved for all *Weather* data needs, such as temperature, wind speed and direction, visibility, cloud cover and precipitation. Of particular importance is the head on wind speeds that create additional forces on the EV, and possible rain that will cause the EV to use its headlight and windscreen wipers.

Google APIs provide a variety of services that satisfy the data needs for the *Route and Terrain*, and *Driving Behaviour* factors [30]. These APIs make use of a smartphone’s GPS and mobile network to communicate with cell towers and Wi-Fi nodes to provide real-time trip data. Support is provided for collecting and retrieving data on points of interests, travel times, live traffic alerts, directions, elevation data, and offers geocoding services to track full-length trips.

Charging APIs such as PlugShare [31], Google Places, and ChargeNow [32] were used to supplement missing data related to the other factors. PlugShare was selected as it supported the most charging stations in South Africa at the time of this study. Data can be retrieved in a JSON format, which allows developers to integrate details for listed charging stations across all three platforms based on location information. Data can be retrieved related to charging stations, such as connector type, charging station type, operating hours, cost, and location. Google maps can select the most suitable route based on charging station availability and route.

Table 2. Comparison of available data attributes for each factor from sources.

	Data Attributes	LeafSp y	LogCo	Accu-Weather	TP-App	Google Maps	Plug-Share
R T	Latitude and longitude	✓	✓	x	✓	x	x
	Elevation	✓	x	x	x	✓	x
	GPS heading	x	✓	x	x	x	x
	Time stamp	✓	✓	x	✓	x	x
	Distance	x	x	x	x	x	✓
	Gradient and terrain	x	x	x	x	✓	x
	Speed limit	x	x	x	x	✓	x
	Traffic	x	x	x	x	✓	x
	Charging location	x	x	x	x	✓	✓
	W	Wind speed/direction	x	x	✓	x	x
Ambient temperature		✓	x	✓	x	x	x
V M	Speed	✓	✓	x	x	x	x
	Braking	✓	x	x	x	x	x
	Tyre pressure	✓	x	x	x	x	x
M	Auxiliary items	✓	x	x	x	x	x
	Acceleration	✓	x	x	x	x	x
	Odometer	✓	✓	x	x	x	x
B M	SoC	✓	✓	x	x	x	x
	Battery health	✓	x	x	x	x	x
	Voltage and current	✓	✓	x	x	x	x
	Battery temperature	✓	x	x	x	x	x
D	Driver ID	x	x	x	✓	x	x
M	Driver score	x	x	x	✓	x	x

4.4 Data Dissemination Phase

Future applications of EVs will provide powerful analytical and machine learning services to drivers, thus requiring heterogeneous data collected in the Data Fusion phase. In the Data Dissemination phase, two analytical components are included: The Batch Processing Layer and the Streaming Layer.

Batch Processing Layer. This layer is concerned with data from persistent or long-term storage, such as those large datasets found in the NoSQL database (MongoDB). Machine learning algorithms are executed here in regular time intervals to adapt to changes in patterns in data. The main machine learning algorithms applied here are for detecting patterns in driving behaviour and scoring drivers according to aggressive, normal or calm categories. Once drivers are classified, analytics are performed based on energy usage per route segment (e.g. highway, urban, mixed). Azure allows for configurable machine learning algorithms and scripts through a drag-and-drop interface. This enabled easy configuration for cleaning and aggregating datasets, feature selection and training, and evaluating and publishing of results in the Azure IoT platform. Two important processing steps were applied, namely pre-processing and model training.

Pre-processing required a number of Python scripts to process the raw data in order to obtain a list of features that is suitable for applying the machine learning model. For raw sensor data, this typically included further sub-stages such as a sampling stage, data cleaning, feature extraction and noise filtering [33]. The first Python script that was written handled the cleaning of the trip data to ensure that there are no missing or incomplete values. Once cleaning was complete, specific data attributes had to be derived that provide a suitable input to the machine learning model. For example, energy consumed, distanced travelled, trip number (ID), acceleration, elevation, and gradient. Other APIs that were used were Google's Roads and Elevation APIs, and the AccuWeather API to understand the route segment, traffic, and weather conditions at recorded GPS coordinates and timestamps. By doing this, driving behaviour and energy consumption patterns could be determined.

Streaming Layer. The purpose of the streaming layer is to analyze the incoming EV streams in real-time and combine them with the data from other sources required for the trip ahead. The data is also pre-processed in this step using the same pre-processing step as for the Batch Layer.

Azure's IoT Hub was used with its Stream Analytics job functionality to aggregate, filter, and compute data streamed from a smartphone. Once again, the data was combined with other APIs from other service providers, based on the user inputs. For example, the driver would input his/her identification and the planned destination in the TP-App. The input data was important when considering retrieving the driver category (aggressive, calm, normal), as the energy consumption estimation algorithm would adapt its calculations accordingly. Data was also retrieved from Google Roads and AccuWeather APIs to retrieve route and weather inputs for the specified route. This was mainly used to estimate energy consumed over the planned route and to provide charging recommendations. Azure's publish/subscribe approach was incorporated, where machine learning and stream analytics functionality can be published as a series of web services; thus, allowing for the results of the Stream Processing and Batch Processing learning layers to be published to a front-end application using HTTP RESTful interfaces.

4.5 Actuation Phase

During the Actuation Phase, the smart mobile devices of passengers and/or the on-board ECUs can take decisions and send commands to actuators allowing the EVs to react to the environment [10]. The developed mobile application (TP-App) subscribed to the RESTful services hosted on the Azure suite. An example of the smartphone's use is where the driver can view the estimated energy usage for a trip or view charging opportunities along the path to the inserted destination. If the driver chooses a new destination, the route and energy requirements are recalculated based on the machine learning model and a check is performed to know if enough energy is stored in the battery to reach the destination. The driver is informed of the amount of energy required to recharge the EV to reach the next destination. If no destination is entered, the model relies on historical driving cycle data, driver category, and real-time GPS data to predict driving routes.

5 Issues Identified and Lessons Learned

5.1 Data Availability and Governance

Despite the heterogeneity of data sources that are required for effective data integration in EVs, one of the major challenges is the lack of publicly available real-world data for EVs and their charging infrastructures [23]. The reasons are partially due to privacy issues of driver, non-disclosure agreements set with telematics service providers, or unsupported data sharing formats (e.g. poorly developed APIs) [4,13]. Some of these issues were experienced when attempting to connect to the LogCo APIs, which were often unavailable and inflexible in terms of upscaling the frequency of logging equipment and the number of parameters recorded. EV data and characteristics were often aggregated (e.g. trip summaries instead of logged data), which made simulation of real-life environments and the evaluation of predictive models difficult [23].

A strategy to protect EVs from security and privacy breaches is to shield network components using gateway devices and software to monitor network interceptions [20]. The retrieval of data from sensors is difficult to attain as the utility of sensors are governed to mask raw data in the EV [9,22]. A thorough understanding of CANs is required to extend the scope of data parameters.

5.2 Heterogeneity and Interoperability

IoT platforms need to support data interoperability, service-level expandability, and device heterogeneity in connected vehicles as a way to communicate effectively and efficiently [4]. The heterogeneity of existing sensing hardware, end-user devices (e.g. smartphones), and vehicle characteristics often differ in vehicular M2M communication [8,10]. Cloud-based IoT platforms and services depend heavily on RESTful web-services and IP technologies to provide interoperability and ease development [10]. For this reason, third-party apps that interpret OBD readings from CAN were purchased. Often these configurations are prone to cause higher latency and quality of service [10], whereas the ideal solution would be to send the data directly from the EV network to the cloud.

5.3 Security, Privacy, and Authentication

Security controls are required for EV data management and backup data storage facilities as sensitive information can be exposed to attacks [20]. Challenges were encountered when attempting to send data to Azure from the TP-App as well as to connect to the RESTful (MongoDB) and SOAP APIs (LogCo). This issue was due to computers that were not provisioned in the Azure suite, as well as computers trying to access the databases from unauthorized IP addresses. Additional programming was required to pass a Globally Unique Identifier (GUID) from the smartphone to Azure to establish a secure connection.

5.4 Accuracy, Reliability and Completeness

Data accuracy and reliability is critical for connected applications in the context of EVs [9,13]. During test runs with the EV, the telematics solution and cellular network for the smartphone were often intermittent and suffered from imprecision and signal loss. Some of the logs were incomplete and missing data could be averaged by following a similar approach as in [12]. For example, speed values were missing from the GPS logs. If the speed was monotonically increasing (positive acceleration) or decreasing (negative acceleration) over multiple measurement points, the acceleration was averaged up to a maximum of four measurement points (4 seconds). Although averaging is a work-around method for incomplete data, this method negatively affects the accuracy of data. Another issue was that the telematics logger often became intermittent, impacting its reliability. The logger often indicated that the EV was moving, but did not show changes in SoC measurements. In this case, if the SoC readings could not be recovered from the alternative LeafSpy logs, the trips had to be disregarded as the energy usage could not be determined and would skew the training results of the machine learning model. Another issue was that LeafSpy user interface often stopped, but still logged the data to the server during a trip. The issue was overcome by purchasing a newer version of LeafSpy.

5.5 Training and Human Error

Using the OBD dongle and smartphone equipment for each trip relies on people, which means that improper configuration or forgetting to charge the gateway smartphone can cause a data loss [9]. Although the issue never occurred, it was a challenge to install preventative measures such as training drivers to use the OBD and smartphone app. Reminders were placed inside the EV to charge the smartphone, check that the phone was connected to the OBD dongle, and to power the phone off after each trip to save battery consumption.

5.6 Timeliness and Temporal issues

The timeliness of data is often impacted by unpredictable issues, which causes real-time applications to experience poor performance. While a significant challenge exists to attain stable, acceptable network performance between devices and cloud resources, it is often difficult to ensure timeliness of data. IoT applications that require quick reactivity for provisioning and authentication often suffer due to delays in network traffic, which negatively impacts performance, usability, and user experience. One issue experienced after analyzing the logged data was that often trips are recorded for short periods where the EV is either idle or travelling short distances (under 1km). This issue related to Emobi drivers switching on the EV to check the SoC or odometer readings, or moving the EV from one parking lot to another. These trips are insignificant for the broader RDR estimation problem and its consequences caused outliers in the analyzed data. It was important to automate the cleaning of such trips in the data set. A similar approach was taken to clean data as described in Section 5.4.

6 Conclusions and Future Directions

This paper forms part of a larger study and extends the work on enhanced RDR estimation models to support drivers in efficient trip planning for EVs. The main theoretical contribution of this paper is a Data Integration Platform for EVs using the IoT and several enabling technologies found in modern EVs. The platform was implemented at an e-mobility organization in South Africa to demonstrate the practicality of the proposed platform. A second contribution is the data integration issues identified in a real-world case study in South Africa. These issues verified some of those reported in other literature, but additional issues were also reported. Furthermore, this paper highlights some of the main drivers and applications for connecting EVs using the IoT. A comparison was derived between investing in data retrieval services (e.g. telematics provider) and self-managed data retrieval techniques (OBD and smartphone).

The Data Integration Platform can support drivers to accurately estimate the RDR and to plan their trips. To fully benefit from the platform and its overall design, associated enabling technologies and IoT platforms need to support a variety of data sources and analytical techniques. Data should be collected from heterogeneous charging management software, weather websites, telematics devices (OBD and tracker unit) for live EV modelling and battery modelling data, as well as route and terrain data from mapping packages such as Google Maps. The issues and lessons learned can provide guidance to other researchers to avoid similar data integration pitfalls. Practitioners can use the platform for designing EV systems, thereby reducing development costs and supporting the creation of innovative services in the IoT and EV domain. The limitations of this study were that it could not evaluate big data methods as the size of the data was still manageable with traditional data analytical techniques. Another limitation was that a full evaluation of the Data Integration Platform remains to be completed.

Future work will report on the results of automating driving behavior analysis to predict energy consumption based on driver style, as well as to make driver identification automatic. The evaluation of RDR estimation accuracy will also be investigated using the proposed platform. Lastly, user experience and user interface techniques will be evaluated based on the usefulness and helpfulness of information presented to drivers to plan trips more effectively.

Acknowledgements

The financial assistance of the National Research Foundation (NRF) and the Telkom Centre of Excellence (CoE) towards this research is hereby acknowledged. Opinions expressed, and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF or Telkom.

References

1. Gambuti, R., Canale, S., Facchinei, F., Lanna, A., Di Giorgio, A. Electric vehicle trip planning integrating range constraints and charging facilities. In: 23rd Proceedings on Mediterranean Conference on Control Automation, pp. 472–479 (2015).
2. Lee, C., Wu, C. Collecting and mining Big Data for electric vehicle systems using battery modeling data. In: 12th International Proceedings on Information Technology - New Generations, pp.626–631 (2015). IEEE.
3. Rahimi-Eichi, A., Chow, M. Y. Big-data framework for electric vehicle range estimation. In: 40th International Proceedings on Annual Conference of the IEEE Industrial Electronics Society, IECON, pp. 5628–5634 (2014). IEEE.
4. Bedogni, L., Bononi, L., Di Felice, M., D’Elia, A., T. S. Cinotti. A route planner service with recharging reservation: electric itinerary with a click. *IEEE Intelligent Transport Systems Magazine*, 8(3), 75–84 (2016). IEEE.
5. Sarrafan, K., Sutanto, D., Muttaqi, K. M., Town, G. Accurate range estimation for an electric vehicle including changing environmental conditions and traction system efficiency. *Electrical Systems in Transportation*, 7(2), 117–124 (2017). IEEE.
6. Ferreira, J., Monteiro, V., Afonso, J. dynamic range prediction for an electric vehicle. In: 27th Proceedings on World Electric Vehicle Symposium and Exhibition, (EVS27), pp. 1–11 (2014).
7. Sarrafan, K., Muttaqi, M., Town, G. An intelligent driver alerting system for real-time range indicator embedded in electric vehicles. *Proceedings - 2016 IEEE Industry Applications Society Annual Meeting*, 53(3), pp.1751–1760 (2016).
8. Mahmud, K., Town, G.E., Morsalin, S., Hossain, M. J. Integration of electric vehicles and management in the internet of energy. *Renewable and Sustainable Energy Reviews*, pp.1-25 (2017).
9. Siegel, J. E., Erb, D. C., Sarma, S.E. a survey of the connected vehicle landscape, architectures, enabling technologies, applications, and development areas *IEEE Transactions on Intelligent Transportation Systems*, pp.1–16 (2017).
10. Datta, S. K., Da Costa, R. P. F., Harri, J., Bonnet, C. integrating connected vehicles in internet of things ecosystems: challenges and solutions. *Proceedings - 17th International Symposium on a World of Wireless, Mobile and Multimedia Network (WoWMoM)*, pp.1–6 (2016).
11. Monigatti, P., Apperley, M., Rogers, B. Smart energy interfaces for electric vehicles. *Proceedings - International Working Conference on Advanced Visual Interfaces(AVI)*, pp.413–416 (2014).
12. De Cauwer, C., Van Mierlo, J., Coosemans, T. Energy consumption prediction for electric vehicles based on real-world data. *Energies*, 8(1), pp.8573–8593 (2015). doi:10.3390/en8088573
13. Botta, A., De Donato, W., Persico, V., Pescapé, A. Integration of cloud computing and internet of things: a survey. *Future Generation Computer Systems*, vol. 56, pp. 684–700 (2016).
14. EVIA. Unity in sustainable mobility: roadmap towards building a unified electro mobility industry in South Africa. (2016). <http://www.evia.org.za/EVIA2016Booklet.pdf>, last accessed 2018/11/01.
15. Chen, J., Zhou, H., Zhang, N., Yang, P., Gui, L., Shen, X. software defined internet of vehicles: architecture, challenges and solutions. *Communications and Information Networks*, 1(1), pp.14–26 (2016).

16. Cao, Y., Song, H., Kaiwartya, O., Lei, A., Wang, Y., Putrus, G. Electric vehicle charging recommendation and enabling ICT technologies: recent advances and future directions. *IEEE COMSOC MMTC Communications – Frontiers*, 11(2), 1-10 (2016).
17. Varghese J. Z., Boone, R. G. Overview of autonomous vehicle sensors and systems Proceedings - 2015 International Conference on Operations Excellence and Service Engineering, pp.178–191 (2015).
18. Wang, B., Panigrahi, S., Narsude, M., Mohanty, A. Driver identification using vehicle telematics data. Proceedings - WCX™ 17: SAE World Congress Experience, 1–8 (2017).
19. Hank, P. Muller, S. Vermesan, O., Van Den Keybus, J. Automotive ethernet: in-vehicle networking and smart mobility, Design, Automation Test in Europe Conference Exhibition (DATE), pp.1735–1739 (2013).
10. Parkinson, S., Ward, P., Wilson, K., Miller, J. Cyber threats facing autonomous and connected vehicles: future challenges. *IEEE Transactions on Intelligent Transportation Systems*, 18(11), pp.2898–2915 (2017).
21. Shladover, S. E. Connected and automated vehicle systems: introduction and overview. *Journal of Intelligent Transportation Systems*, 22(3), pp.190-200, DOI: 10.1080/15472450.2017.1336053.
22. Tseng, C. M., Zhou, W., Al Hashmi, M., Chau, C. K., Song, S. G., Wilhelm, E. Data extraction from electric vehicles through OBD and application of carbon footprint evaluation. In: Proceedings on Electric Vehicle Systems, Data, and Applications (EV-SYS), pp.1–6 (2016).
23. Li, B., Kisacikoglu, M. C., Liu, C., Singh, N., Erol-Kantarci, M. Big data analytics for electric vehicle integration in green smart cities. *IEEE Communications Magazine*, 55(11), 19–25 (2017). IEEE.
24. Johannesson, P., Perjons, E.: An introduction to design science. Springer International Publishing Switzerland, Kista (2014).
25. Smuts, M., Scholtz, B., Wesson, J. A critical review of factors influencing the remaining driving range of electric vehicles. Proceedings - 1st International Conference on Next Generation Computing Applications (NextComp), 1-6 (2017).
26. Microsoft Azure Homepage, <https://azure.microsoft.com/en-us/services/hdinsight/>, last accessed 2018/11/01.
27. Kolarova, V., Kuhnimhof, T., Trommer, S. Assessment of real-world vehicle data from electric vehicles – potentials and challenges. Proceedings - 11th International Conference on Transport Survey Methods (ISCTSC), 1-11(2017).
28. Castignani, G., Derrmann, T., Frank, R., and Engel, T. Driver behavior profiling using smartphones: a low-cost platform for driver monitoring. *Intelligent Transport Systems Magazine*, 7(1), 91–102 (2015).
29. AccuWeather Homepage, <https://developer.accuweather.com/apis>, last accessed 2018/11/01.
30. Google Maps Homepage, <https://cloud.google.com/maps-platform/>, last accessed 2018/11/01.
31. PlugShare Homepage, <http://developer.plugshare.com/>, last accessed 2018/11/01.
32. ChargeNow Global Homepage, <https://www.chargenow.com/web/chargenow-global>, last accessed 2018/11/01.
33. Strohbach M., Ziekow H., Gazis V., Akiva N. (2015) Towards a Big Data Analytics Framework for IoT and Smart City Applications. In: Xhafa F., Barolli L., Barolli A., Papajorgji P. (eds) Modeling and Processing for Next-Generation Big-Data Technologies. Modeling and Optimization in Science and Technologies, vol. 4, pp. 257-282. Springer, Cham.