

ZENZIC²

SELF-DRIVING REVOLUTION

Geodata report - analysis and recommendations for self-driving vehicle testing

Zenzic consultation findings

June 2020



This document includes the initial Geodata report released in 2019 by Zenic and OS, as well as the new consultation findings.

Introduction

CAM Testbed UK is a collection of connected and self-driving technology test facilities which are collaboratively funded by government and industry. CAM Testbed UK provides a comprehensive set of testing capabilities within a defined geographical location. This collection of test facilities is independently operated but seek to complement each other and strive towards interoperability. This ensures that the UK has a globally-unique offering in this test and development space.

The members of CAM Testbed UK work together on a series of interoperability projects, exploring common areas of opportunity to align processes and provide interoperable systems. Zenic has recently published a common safety assessment framework and this work seeks to drive further interoperability in the area of geospatial data.

Many of the developers of Automated Driving Systems (ADS) regard the need for high quality, reliable and pervasive geospatial data as central to the development of a robust and safe ADS. However, there are limited commonly-used formats or standards in use by the majority of geospatial data owners. There are significant benefits to be had in aligning to an interoperable set of geospatial data infrastructure, firstly within CAM Testbed UK, but then with wider adoption. This will help accelerate the development and subsequent deployment of self-driving vehicles.

In July 2019 Zenic published Geodata Report – analysis and recommendations for self-driving vehicle testing authored by Ordnance Survey (Great Britain’s National Mapping Agency). The aims of the report were to examine interoperability requirements for geospatial data which are needed for testing and development, and then to set the foundations for operational deployments for connected and self-driving vehicles. The report was accompanied by a consultation form (See Annex 1) to gain further insight and input from wider industry. The consultation closed on the 30 October 2019 and the following sections explore the feedback and findings from the responses received.

Response to key recommendations

The original consultation document was framed around five key recommendations, for areas in which interoperability would be beneficial. These five areas were focussed on the most commonly discussed opportunity areas from initial research undertaken by Ordnance Survey.

The consultation opened these recommendations to a wider audience for comment. The consolidation of those comments is discussed in the below sections of this response document.

Data formats

The report recommended the adoption of the four key formats described in Table 1.

Table 1 – Baseline recommended data formats

Format	Purpose
LAS 1.2 or LAZ (compressed)	Point cloud data capture for identification, extraction and modelling of terrain and key features
OBJ	Good for representing the terrain and 3D objects such as buildings
OpenDrive	Good for describing track-based road networks
ESRI shapefile	A portable format good at representing a wide range of specific key features and their attributes

Consultation findings:

Overall, respondents tended to agree with the data formats as recommended in the report (and in Table 1). However, concern was raised about the suitability and use of ESRI shapefiles. ESRI shapefiles have several limitations that are likely to render the format less usable in the context of connected and self-driving vehicles, these include (ESRI, 2019):

- A file limit of 2GB – roughly 70 million-point features;
- Shapefiles do not contain x,y tolerance that are required to evaluate relationships between features;
- Compression limitations:
 - They do not support textures and lighting normals;
 - Are limited in their speed to deal with spatial queries;
 - Date fields only support date but not time;
- There are limits on field names (10-character limit), number of fields (225) as well as the maximum length for an attribute (4,000 bytes);
- Null values are recorded as 0

Concerns were also raised about using proprietary formats so alternative, more open, data formats were proposed. Respondents suggested the use of open formats such as GeoJSON and GeoPackage as alternatives. Other alternatives suggested by respondents is the use of LAZ (a LiDAR format) and/or E57 (similarly, a point cloud data format) for a similar purpose to shapefiles.

Other comments on data formats included some preference for the use of FBX over OBJ for a better store of material information. This file format was explored in the original report, and as there was limited support for FBX it is not included in Table 2. In addition, OpenSCENARIO has the additional ability to specify actors within a scene as such may be a stronger baseline format than OpenDRIVE. Updated baseline recommended data formats are shown in Table 2.

Future file formats and the evolution of technologies should also be considered, and baseline file formats should adjust accordingly over time

As trials and testing within CAM Testbed UK increases, more insights into the needs and requirements for geo-data will emerge. Over this time there will be a need to continually challenge the latest thinking and update best practice.

Table 2 – Updated baseline recommended data formats

Format	Purpose
LAS 1.2 or LAZ (compressed)	Point cloud data capture for identification, extraction and modelling of terrain and key features
OBJ	Good for representing the terrain and 3D objects such as buildings
OpenDrive/OpenSCENARIO	Good for describing track-based road networks
LAZ/E57/GeoPackage/GeoJSON	Good for representing a wide range of specific key features and their attributes

Data quality and resolution

Once requirements for data formats have been validated across industry, Zenic should facilitate discussions between emerging self-driving technology developers, CAM Testbed UK and the Geospatial Commission to ensure that requirements for self-driving vehicles have been adequately considered and represented.

Consultation findings:

It was clear from the responses that there are two main technical deployment styles for connected and self-driving vehicles that were not explicitly references in the original report:

- (a) Those which require HD maps for self-driving and in turn for the use in simulation testing and real-world; and
- (b) Those that do not require HD maps for self-driving. As such, any geospatial requirements would be for simulation and testing rather than real world use.

In both cases, the importance of data quality (as well as origin and limitations) was frequently highlighted in responses with greater emphasis placed upon accuracy and precision rather than resolution. Data quality from government bodies was highlighted as a key issue. The precision of sensors should also be considered in the validation of geospatial data.

In terms of resolution there were limited responses for the acceptable level – 10cm (with 5cm for lane boundary information) was posited as suitable resolution for AV systems that use HD maps and 2cm for those that do not. However, further dialogue and research is needed to understand resolution requirements for self-driving vehicles.

Terminology

Common standards for terminology should be developed across the connected and self-driving vehicles sector and fed into the BSI Standards Programme to ensure that common terminology is documented, maintained and adopted by the industry at large. A significant contributor to ensuring common terminology will be BSI, which has a pivotal role in the promotion of standards, both in the UK and internationally.

Consultation findings:

In addition to working closely with BSI, as well as the Open Geospatial Consortium, respondents suggested that wider cross-industry collaboration may be of benefit. In particular, the BIM, Gaming and Weather sectors were highlighted as ones that lessons and commonalities could be learnt (such as the top down approach taken by the BIM sector) as well as further engagement with the traditional transport sector. Where different sectors use differing terminologies for the same element it is important that these differences are captured so that subtle variances are known and can be accounted for. However, it would be preferable that, where common terminology already exists, these terms are then applied consistently. Alignment (and shaping) of international standards was also highlighted as a key factor for the UK to be globally competitive.

Minimum safe requirements and standards

Manufacturers may be reluctant to share data and methods, but ubiquitous, reliable and safe self-driving vehicle operation will nevertheless be dependent on conformity to a minimum set of requirements and associated standards. These requirements should be impartially captured, working alongside standards bodies such as BSI, security specialists and government, to ensure consistency, security and compliance.

Consultation findings:

The overall response to the question regarding minimum safe requirements and standards were rather limited however there were several key themes raised. Whilst the concept tended not to be questioned it was made clear that the purpose of any such standards and requirements should be clear and continued alignment with organisations such as BSI and the Law Commission is important. The need for a level of data redundancy of sources (which are safety critical) was highlighted.

There are currently two primary standards regarding the exchange of road related data the European TN-ITS (EC Inspire – and covers only 14 European countries) and the ISO 20542 standard. There is a gap between these two standards which is being examined at by the ISO TC211 and TC204 groups. Given the number of players in the market and dominance of certain manufacturers both of vehicles and equipment, the harmonisation of standards may be a difficult task. Therefore, facilities and testing regimes may need to accommodate both standards initially.

It is clear that further work is required in the development of minimum safe requirements and standards pertaining to geospatial data for self-driving vehicles.

Government data and Traffic Regulation Orders (TROs)

With multiple projects relating to self-driving vehicles being undertaken across industry and government, such as the exploration of the consistency of TROs and projects that examine neutral hosting of data, it is vital that strong governance is put in place to coordinate and align them. Zenzic is well positioned to provide this governance and ensure that activities are future-proofed. This is a non-trivial piece of work but should be developed with consideration of self-driving vehicles from the outset. CAM Testbed UK, funded by the Centre for Connected and Autonomous Vehicles and co-ordinated by Zenzic, provides an ideal environment in which to test and develop digital TRO's in tandem with self-driving technology.

Consultation findings:

It was clear from the responses that there are issues in the quality and reliability of government data. The problem of data reliability, responsibility and liability with the hosting of government data was raised in the context of creating a centralised source of data. The quality of public data can often be poor and have discrepancies within a dataset. There are over 200 Local Highway Authorities in Great Britain alone - which can present a problem with the number of different processes and data handling methods used. It was also suggested that a demonstration of two-way benefits (to both the industry and public data owner) would be a useful step in gain further 'buy-in' from public data owners. For example, to demonstrate the enhanced network efficiency benefits of connected and self-driving vehicles when able to access these types of data.

Access to data regarding Traffic Regulation Orders (TROs) and key information such as pick-up and drop-off points and parking spaces were frequently highlighted as an important public sector dataset that requires a greater strategic approach - it is vital that the future use of this type of data is considered to enable secure use of the data and in some cases this may require the digitisation of assets.

A recent TRO Discovery project between DfT, British Parking Association, Ordnance Survey and GeoPlace recommended a number of areas to be reviewed including data consistency, common standards, legislation and outputs included a draft data model for the spatial creation of TROs as well as a User Guide to assist Authorities to spatially create data to common standards and formats (GeoPlace, 2019).

Data hosting

Data is available from a variety of sources, spanning both public and private sectors. Datasets can be very challenging to access and integrate in a consistent and reliable manner. A neutrally-hosted geospatial environment that can make use of federated data from a variety of sources would provide a one-stop shop that will facilitate interoperability. Neutral hosting of data should be explored with relevant bodies, including the Geospatial Commission and the DfT, to consider available options. Zenzic must also ensure that the current work programmes creating data hosts are governed and aligned, and that lessons learned are shared frequently.

Consultation findings:

Overall, respondents were supportive of the concept of the large-scale neutral hosting of data. In response to the survey, **data quality** and **reliable data provenance** were ranked as the most important factors in delivering this, whilst **access to commercial data** was ranked least important of the seven factors. A number of key factors were also highlighted by respondents:

Access to data

It was posited that the access method to the data host should be considered as it would not be sustainable to have a model where users were all “takers” or a simple “pay to play” model. Alternatives such as a “share data to get data” could be considered. Similarly, it was suggested that competition between different business models could be considered during the early stages of development to allow the best solution to win out.

Data quality and service design

As reflected by the survey – data quality, accuracy and transparency regarding its origin is vital. This also includes consistent accuracy at a granular level to avoid data being aggregated/averaged at a regional level in a way that may distort the true reflection of the quality of data. Similarly, it is important that data consistency forms part of the neutral host i.e. regional variance in the quality of data should be known or, ideally, not exist. The data provenance issue should be captured in the service design and be assured through the process of delivering a neutral host.

The update or refresh rate of the data should be transparent. Furthermore, it was suggested that self-driving vehicles would, for those that use HD maps for real-world use, require short update cycles for this data to be useful – in the region of 24-48 hours.

Secure and scalable

As part of the delivery of a large-scale neutral host there would be an assumption that it would be secure through the services design and access agreements. Respondents highlighted the need for any such data hosting to be flexible enough to allow scalability of data and that this should be factored into the design at the very beginning.

Safety critical data

There is a need for redundancy of data sources for those which are safety critical (in the case of the self-driving vehicle requiring the use of HD map in the real-world). Trust and consistency are vital and there exist examples in the aviation industry that divide data in to that which is required for safety and that which is required for efficiency – this division may help clarify the exactly the type and form of data required to be hosted.

Annex

Annex 1 – consultation questionnaire [online]

- 1) Please enter your contact details
- 2) a) To what extent do you agree that the main data formats recommended, or those from Annex 4 (page 55) in the report, cover the needs of companies using simulation for development, testing and operation of automated control systems?
- 2) b) Do you have any specific comments on the data formats recommended or listed in Annex 4 of the report? Do you believe any data formats have been omitted that should be included?
- 3) Of the data formats identified in the report and any additional you may have added from question 2 b, how would you define required data resolution and quality?
- 4) How do you think the industry can create a common standard for terminology and ensure adoption across the industry?
- 5) What are your views given the issues raised in the report, regarding data sharing and cyber security risk, for setting minimum safety requirements and standards?
- 6) How do we ensure that data requirements for connected and self-driving vehicles are considered by public sector data owners such as local highway authorities? (For example, consideration of traffic regulation orders)
- 7) a) Rank, what you believe to be, the most important factors in delivering the large-scale neutral hosting of data. (Where 1 is the most important, and 7 the least important)
 - Frictionless access to data
 - A 'one-stop shop' service
 - Interoperable (with other data and services)
 - Accessible
 - Data quality
 - Reliable data provenance
 - Access to commercial data (via subscription or payment)
- 7) b) Are there any additional factors, not suggested in 7 a, that you believe are important in delivering the large-scale neutral hosting of data?
- 8) Do you have any other comments specific to either this report, or geospatial data requirements of connected self-driving vehicles?

References

ESRI (2019). Geoprocessing considerations for shapefile output. [Online] Available at <https://desktop.arcgis.com/en/arcmap/latest/manage-data/shapefiles/geoprocessing-considerations-for-shapefile-output.htm> [Accessed 20 November 2019]

GeoPlace (2019). Enabling transport systems of the future. [Online]. Available at <https://www.geoplace.co.uk/trodiscovery> [Accessed 03 December 2019].

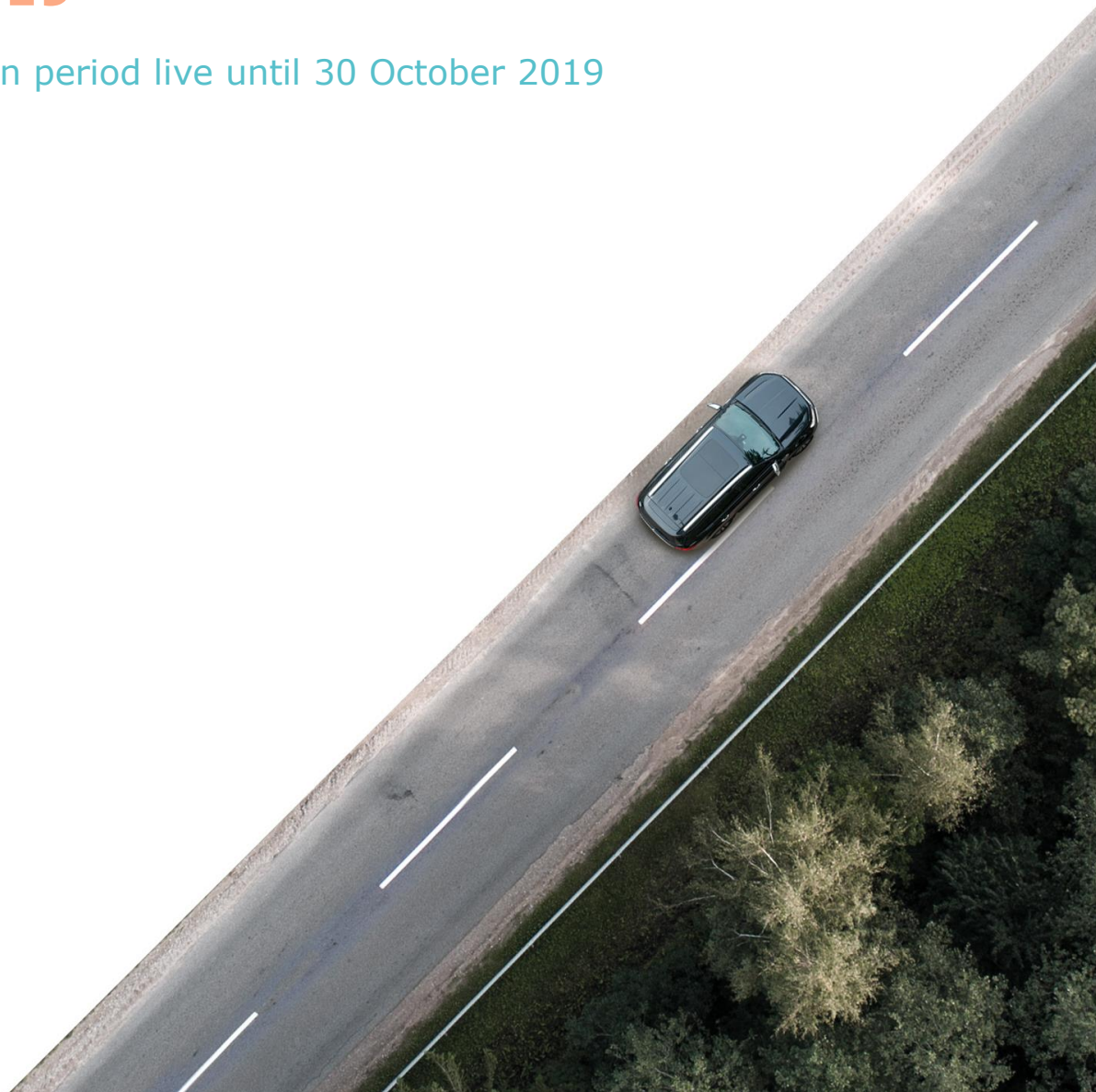
ZENZIC⁴

SELF-DRIVING REVOLUTION

Geodata Report – analysis and recommendations for self-driving vehicle testing

July 2019

Consultation period live until 30 October 2019



Copyright notice

This document has been produced by Ordnance Survey for Zenzic-UK Ltd to provide an insight into the use of geodata in support of self-driving vehicle testing and is pursuant to the consultancy agreement between Zenzic-UK Ltd and Ordnance Survey dated 1 March 2019.

No part of this work may be reproduced, stored, or transmitted in any form, by electronic means or otherwise, without the written permission of Zenzic-UK Ltd.

For further information on this report please contact the Zenzic team

Email: info@zenzic.io
Web: zenzic.io

Table of Contents

Executive summary	4
Key conclusions	5
Key recommendations	7
Background	10
About Zenic	10
About OS	10
Study objectives	12
Scope	12
Approach	12
1. Key findings	13
1.1 Industry challenges	13
1.2 The ideal situation	13
1.3 Simulation, track and real-world testing	14
1.4 Terminology	15
1.5 Data reliability and authority	16
1.6 Data quality and resolution	16
1.7 Data provenance and currency	18
1.8 Standards	19
1.9 Governance	20
2. An overview of road/street environment: real-world elements	22
2.1 Static elements	22
2.1.1 Static element properties	23
2.2 Temporary and dynamic elements	23
2.3 Weather	24
3. Requirements for testbeds and simulation companies	25
3.1 Industry perspectives	25
3.2 Critical geospatial information for operation of self-driving vehicles	25
3.2.1 Street furniture	26
3.2.2 Road markings	27
3.2.3 Junctions	28
3.2.4 Road lanes	28
3.2.5 Pedestrian crossings.....	28
3.2.6 Traffic islands.....	29
3.2.7 Taxi stands and public transport stops.....	29
3.2.8 Parking locations	29

3.2.9	Vegetation	29
3.2.10	Road network.....	29
3.2.11	Buildings types and materials	29
4. Data sources for self-driving vehicles.....		31
4.1 The 'digital twin'.....		31
4.2 Ordnance Survey data		32
4.3 Local authority data.....		33
4.3.1	TRO Discovery Project	35
4.3.2	Transport Network Intelligent Transport Systems (TN-ITS)	35
4.4 Other sources		35
4.5 Data capture, processing and formats		36
4.5.1	Data capture	36
4.5.2	Data processing.....	37
4.5.3	Data formats.....	37
4.6 Data hosting		38
Recommendations		41
Annex 1 – Workshop output		46
Annex 2 – Feedback and comment		53
Annex 3 – Table of abbreviations and acronyms.....		54
Annex 4 – Data formats explored		55
References		56

Executive summary

The market for self-driving vehicles in the UK, specifically that for road vehicles using connected and self-driving technologies, is projected to be worth up to £52 billion in 2035, capturing 6% of the £907 billion global market. Ensuring the success of self-driving vehicles is a complex challenge, and one which will require collaboration between government and multiple industries to ensure its success. One of the major complexities relates to the use of geospatial data to aid the routing, navigation and safe operation of vehicles.

Zenitic commissioned Ordnance Survey (OS) to explore these geospatial data requirements when considering the life cycle of testing and the data interoperability requirements to enable UK plc to provide an exemplar test facility ecosystem and set the foundations for operational deployments.

Drivers have used paper maps, and more recently satellite navigation systems, for many years; maps for self-driving vehicles, however, need to be more detailed, and are often referred to as high-definition (or HD) maps. These HD maps are more complex than maps used for simple driver-based navigation systems. Mapping must be highly accurate (better than 5 cm in resolution) and needs to contain a minimum set of road information – for example, lane-level geometry, and information relating to street furniture.

This study has been conducted with the support of software simulation companies, testbed operators, the Met Office and the British Standards Institution (BSI), with the aim of understanding the data requirements, gaps and sources regarding geospatial data for self-driving vehicle testing, and in order to recommend follow-on work to address the issues raised.

Can self-driving vehicles really be safe?

Before self-driving vehicles are approved for commercial service, industry will have to clearly demonstrate to regulators that they are safe. Much of this work is likely to take place in international forums. A consultation by the Law Commission in early 2019 on the regulatory framework for self-driving vehicles included consideration of safety assurance procedures.

It is also essential that any road trials are conducted safely. In 2015, the UK published its world-leading Code of Practice for the safe trialling of self-driving vehicles on public roads. This guidance was updated in February 2019 to include the most up-to-date thinking on self-driving vehicle safety. OS/Zenitic are also investing, with industry, in test facilities to ensure that simulation systems are robust and road-ready before they are trialled on public roads. In addition, the Government has developed eight principles for good cyber security within the automotive sector. (Government, 2017)

It is widely accepted that self-driving vehicles would need to travel billions of miles in the physical world to demonstrate a significant improvement on safety for human drivers. To mitigate this blocker, simulation has already become ubiquitous across global development of self-driving vehicles.

We are now starting to see a shift in self-driving simulation technology from validation of safety cases towards its use for certification and regulation. Longer term, geospatial data for simulation will play a vital role in efficient and effective operation of self-driving technology. Particularly, if near real-time and highly accurate data is needed for safe operation and navigation.

Now is the time to examine not only what types of geospatial data are needed across the industry to deliver simulation services for testing and development of self-driving technology, but also to further explore standardisation and sharing of this data in preparation for operational deployments.

Ordnance Survey has vital role to play in self-driving technology development because of the complex and varied geospatial requirements which are emerging as the UK moves towards operational deployments. These include high definition data, localisation, navigation and data exchange. Our collaboration with OS on this report has already provided us some great insights and Zenic's initial responses to the recommendations are included in this paper.

Moving forward we hope that the results of the accompanying consultation will help us to further understand the needs of government and industry, and provide further detail on what action is needed in the coming years to unlock the potential power of geospatial data for development, delivery and operation of self-driving technology.

**Richard Porter
Technology and Innovation Director, Zenic**

Key conclusions

- Geospatial data is fundamental to life cycle support for self-driving vehicles. Maps can provide an important trusted baseline where the availability of sensor feeds cannot be guaranteed.
- As Great Britain's national mapping agency, OS holds valuable and nationally consistent data relevant to road geography within its OS MasterMap Topography Layer and Highways Network products. However, to provide the framework required for the testing and driving of self-driving vehicles, current mapping specifications will need to be enhanced to include relevant street-level features, with superior resolution (better than 5 cm), richer attribution and the ability to integrate with other sources of data.
- There is no authenticated single source of suitable geospatial data that users can locate today; that is, there is no access to a 'one-stop shop' service. Such a service, if authoritative and neutrally hosted, would aid interoperability and increase confidence in the data being sourced.
- Other data exists, in both the private and public sectors, that is needed to enhance the production of HD maps. Sources include local authorities, private geospatial companies and (potentially) crowdsourcing. Often this data is inaccessible or of variable quality and specification, making its use difficult and potentially unreliable, unless quality processes and approved standards are adopted.
- Standards within the UK (and globally) relating to geospatial data in the context of connected and self-driving vehicles are starting to emerge and will need to be developed rapidly and be flexible to meet the pace of the market.
- For effective use, real-time or near-real-time updates to mapping through sensor data feeds will play an important role in ensuring geospatial data remains up to date and could establish additional information relating to long-term but temporary features that do not appear on conventional maps.

Key recommendations

Data formats

During this study, more than 20 different formats which relate to the creation of real-world digital twins of road environments were investigated (see Annex 4 – Data formats explored).

Each format offers specific features and benefits, but to ensure interoperability and a good operational baseline, this report recommends the adoption of the four key formats described in Table 1. A full list of the data formats considered is included in Annex 4.

Table 1 – Baseline recommended data formats

Format	Purpose
LAS 1.2 or LAZ (compressed)	Point cloud data capture for identification, extraction and modelling of terrain and key features
OBJ	Good for representing the terrain and 3D objects such as buildings
OpenDrive	Good for describing track-based road networks
ESRI shapefile	A portable format good at representing a wide range of specific key features and their attributes

The OS MasterMap family of products and OpenStreetMap both offer a good maintenance base on which to build additional map layers that should ensure excellent interoperability across the UK and self-driving vehicle testbeds.

In addition to these defined formats, the Traffic Regulation Orders (TROs) should be made available in appropriate compatible formats (to be established) so they are easily accessible, interoperable and consistent across all local authorities and users of these critical rules defining driving parameters.

Zenic should collect feedback on this reduced set of file formats from organisations including testbeds, simulation companies and manufacturers, to establish any known limitations or challenges that they may present. Results of this consultation should be used to inform standards for self-driving vehicles.

Zenic agrees that early alignment on common data formats that can be used across industry to develop, validate and operate self-driving vehicles is a critical task. Zenic is already developing interventions to align data formats across Testbed UK and will use this report as a basis to further develop definition of these projects. Zenic will conduct a consultation on the data formats listed in this report to gather feedback from industry and other stakeholders.

Data quality and resolution

Many emerging technologies, including self-driving vehicles, demand rich high-resolution geospatial data. OS is currently in discussion with the UK Government's Geospatial Commission regarding its long-term service agreement for the public sector.

Once requirements for data formats have been validated across industry, Zenic should facilitate discussions between emerging self-driving technology developers, Testbed UK and the Geospatial Commission to ensure that requirements for self-driving vehicles have been adequately considered and represented.

Zenzic, with support of OS, will engage with the Geospatial commission to ensure requirements for self-driving vehicles are considered. This activity will be informed by collation of feedback from industry on the findings of this report.

Terminology

There is some evidence that terminology across the industry can be inconsistent and lacks definitive sources. **Common standards for terminology should be developed across the connected and self-driving vehicles sector and fed into the BSI Standards Programme to ensure that common terminology is documented, maintained and adopted by the industry at large.** A significant contributor to ensuring common terminology will be BSI, which has a pivotal role in the promotion of standards, both in the UK and internationally.

Zenzic is participating in the BSI CAV Standards programme board which will allow it to have visibility of and influence common terminology. We will seek to both collect input from our testbed partners and also to roll out standardised terminology across our ecosystem.

Minimum safe requirements and standards

Manufacturers may be reluctant to share data and methods, but ubiquitous, reliable and safe self-driving vehicle operation will nevertheless be dependent on conformity to a minimum set of requirements and associated standards. **These requirements should be impartially captured, working alongside standards bodies such as BSI, security specialists and government, to ensure consistency, security and compliance.**

Zenzic already sits on the steering group for the BSI CAV standards programme and is actively involved in defining cybersecurity investments and interventions with the Centre for Connected and Autonomous Vehicles (CCAV) and Innovate UK (IUK).

We believe the best way we can ensure delivery of safety critical data standards is by co-ordinating requirements from, and creating interventions across, our testbed ecosystem that will ensure approaches to standardisation can be empirically tested.

Government data and Traffic Regulation Orders

With multiple projects relating to self-driving vehicles being undertaken across industry and government, such as the exploration of the consistency of TROs and projects that examine neutral hosting of data, it is vital that strong governance is put in place to co-ordinate and align them. Zenzic is well positioned to provide this governance and ensure that activities are future-proofed. **This is a non-trivial piece of work but should be developed with consideration of self-driving vehicles from the outset. Testbed UK, funded by the Centre for Connected and Autonomous Vehicles and co-ordinated by Zenzic, provides an ideal environment in which to test and develop digital TRO's in tandem with self-driving technology.**

Zenzic is firmly of the belief that digitised highways codes and TROs will play an important part in the effective and safe operation of self-driving vehicles. Furthermore, there are significant near-term benefits to accurate digital TROs being available for 'connected' services, including dynamic advice on current speed limits and parking services.

Zenzic believes the best way to deliver these benefits is for Testbed UK to be at the centre of early roll out of digital TROs. This would ensure that Department for Transport (DfT) and industry organisations would have a focal point for testing implementation and additional services that might be beneficial for connected or self-driving vehicles.

Data hosting

Data is available from a variety of sources, spanning both public and private sectors. Datasets can be very challenging to access and integrate in a consistent and reliable manner. A neutrally-hosted geospatial environment that can make use of federated data from a variety of sources would provide a one-stop shop that will facilitate interoperability. ***Neutral hosting of data should be explored with relevant bodies, including the Geospatial Commission and the DfT, to consider available options. Zenzic must also ensure that the current work programmes creating data hosts are governed and aligned, and that lessons learned are shared frequently.***

Neutral hosting is a foundational element of delivering the data needed accelerate delivery of self-driving vehicles. However, Zenzic also believes there is a significant role for commercially focussed business models to play in unlocking the sharing of the highest quality data, which is often collected at significant cost to businesses. Zenzic is facilitating discussions between both for profit and not for profit organisations to encourage testing of data sharing modes where all parties understand the value gained up front.

Background

In support of the UK's Industrial Strategy, there are several UK Government-sponsored connected and self-driving vehicle research and development projects, supporting over 70 project themes with more than 200 partners, all having the aim of positioning the UK as the best location in the world for the testing of connected and self-driving vehicles.

The Government has committed to invest £100 million, matched by industry (£200m total), to create a world-leading ecosystem for the testing and development of connected and self-driving vehicles in the UK; this ecosystem is led by Zenzic and is called Testbed UK. Additionally, in 2018 the National Infrastructure Commission ran a 'Roads For The Future' competition (jointly won by City Science and Leeds City Council) (NIC, 2018) focused on future infrastructure requirements to support self-driving vehicles.

About Zenzic

Zenzic (formerly Meridian) was created by government and industry to champion the connected and self-driving ecosystem and accelerate the self-driving revolution in the UK. The company drives collaboration with partners across industry, government and academia to shape a world-class Testbed UK, and to deliver a comprehensive UK Connected and Automated Mobility Roadmap to 2030.

Zenzic was created by government and industry to focus on key areas of the UK's capability in the global connected and self-driving sector – a sector predicted to be worth £907 billion by 2035.

There are six testbeds currently co-ordinated through this programme of work:

- CAVWAY
- ConVEx (Connected Vehicle data Exchange)
- Midlands Future Mobility
- Millbrook–Culham Urban Testbed
- Trusted Intelligent Connected Autonomous Vehicles + Park-IT
- Smart Mobility Living Lab: London

About OS

Ordnance Survey (OS) is the national mapping agency for Great Britain, and a world-leading geospatial data and technology organisation. As a reliable partner to government, business and citizens across Britain and the world, OS helps its customers in virtually all sectors improve quality of life. OS expertise and data supports efficient public services and infrastructure, new technologies in transport and communications, national security and emergency services and exploring the great outdoors. By being at the forefront of geospatial capability since 1791, OS has built a reputation as the world's most inspiring and trusted geospatial partner.

OS has been appointed by Business Secretary Greg Clark to help shape a national infrastructure capable of supporting a nationwide network of connected and self-driving vehicles. E-CAVE (Ordnance Survey, 2018) is a four-year project that lies at the heart of the Government's Industrial Strategy, and further establishes the UK as one of the world's leading locations in this sector. E-CAVE focuses on the challenges of creating effective connected environments using OS digital data expertise. OS is also engaged in supporting and

collaborating with the testing of connected and self-driving vehicles across six testbed projects overseen by Zenzic.

OS is also involved in several collaborative 'smart' projects supporting Britain's connected and self-driving vehicle infrastructure, such as:

- Atlas, which studied and identified data critical to the efficient operation of self-driving vehicles
- the UK's Internet of Things demonstrator in Manchester (Cityverve, 2019)
- a connected and self-driving vehicle simulation testbed (Ordnance Survey, 2018)

OS is collaborating across sectors to support the planning and site selection of infrastructure for 5G mobile communications (Rogerson & Donegan, 2018), which will rely on a detailed model of the built and natural environment. 5G will provide a core element of future smart services by enabling services to join up in real-time over distance, and by underpinning a range of other applications, including connected and self-driving vehicles, advanced manufacturing and robotics, augmented reality, smart agriculture, and smart homes and cities. By doing this, OS is helping pave the way for smarter, more connected communities.

Study objectives

Scope

The full spectrum of potential self-driving vehicles is vast and could include air, land, sea and subsea vehicles. This project is focused on geospatial information within the overall context of land-based self-driving vehicles – typically cars, vans and lorries – and is limited to the information sourced through this study.

Figure 1 – Scope of study (source: OS)

The scope of this study (Figure 1) focuses on the geospatial component of the self-driving vehicle environment, and consists of three main elements:

1. the identification of the data assets that need to be accessible via a consistent and interoperable format through appropriate licensing terms
2. considerations for data hosting
3. the provision of expert and trusted advice in the use of data assets to support the accelerated testing of self-driving vehicles



This report covers:

- the key self-driving vehicle challenges that can be addressed using geospatial data
- the key datasets or data categories that are required for the full self-driving vehicle life cycle
- the critical features and key attributes within these data categories required for the full self-driving vehicle life cycle
- the current sources of data available today, their limitations, where they are sourced, how they are accessed and delivered, and their suitability for geospatial requirements
- a common approach to the standards and formats of geospatial data
- recommendations

Approach

A workshop and a subsequent questionnaire were used to explore in further detail what data simulation companies and testbed operators require in support of self-driving vehicle testing, what datasets are available, and how these are currently generated and captured.

The relevance of these datasets was examined to determine the level of fitness for purpose and level of discoverability, accessibility and quality as an enabler of interoperability. The aim was to:

- establish baseline content and requirements for the data and potential services
- establish the requirements for common standards and interoperability between existing and planned data assets
- undertake a gap analysis to determine which data, services, standards and processes exist and what is lacking and may need to be developed

1. Key findings

1.1 Industry challenges

The key challenges (see Annex 1) for those involved in the testing of self-driving vehicles were identified at a workshop by a core team of participants composed of owners of vehicle testbeds in the UK, simulation companies, the BSI and Zenzic. This was only a sample of the industry, but was deemed adequate to identify the core challenges.

Given that the participants came from differing backgrounds, there was a spectrum of challenges that included the commercial and promotional aspects of testing, the details of gathering specific requirements from users of test facilities, and the capture of critical infrastructure data. As this assignment was focused on the geospatial aspects of the test and trial cycle, commercial and promotional aspects will not be discussed further in this report.

Nineteen categories of challenges were identified from the workshop, and participants were asked to vote, based on their own knowledge, experience and perspective, on which must take priority. As there was a larger representation from the Zenzic testbeds during the workshop than from simulation companies, the voting results need to be viewed with caution. Some clear topics and areas of focus emerged:

- Clarifying the requirements of the users – what do manufactures need in order to achieve certification
- Updates of data, including real-time requirements
- Data interoperability across the full life cycle and across testbeds
- File formats and data sharing
- Security, accuracy and detail
- Road markings and road signs, and their relationship to traffic regulations

1.2 The ideal situation

A key objective of UK plc is to attract interest from global industries who are investing in the development of self-driving vehicles and associated technologies, and offer a world-class capability that provides full support for the testing of vehicles through simulation, and on physical testbeds that faithfully and reliably reflect real-world road conditions. This will provide assurance that they are safe to be used on public roads. To achieve this, development and test facilities will need to conform to a range of standards to ensure the test results are consistent – full interoperability between these environments should be embedded from the outset.

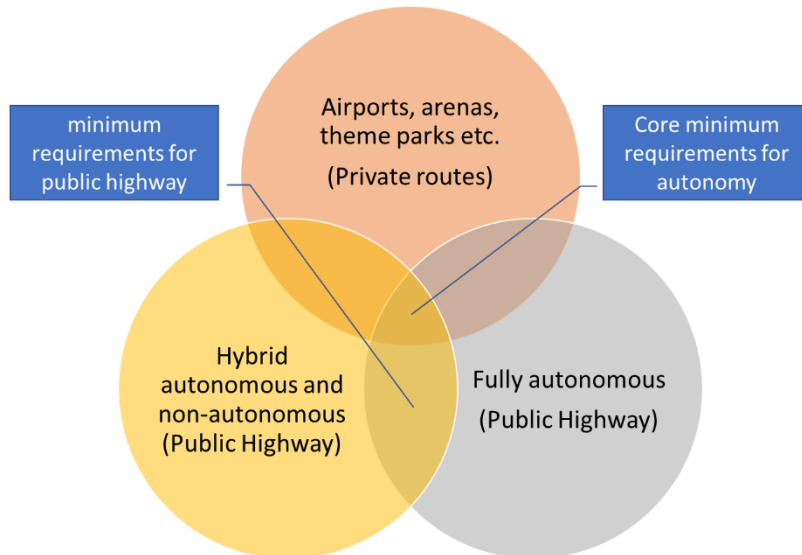
The development of self-driving vehicle standards is currently at an embryonic point, with several countries attempting to advance and lead the agenda. Within the UK, BSI is taking the lead in this role. However, to develop standards, a good understanding of requirements, definitions and terminology is required to ensure that best practices are adopted that will support safe deployment of self-driving vehicles.

The geospatial element plays a significant and potentially pivotal role in leading the development of such standards, as the road network defines the environment on which vehicles travel.

In the ideal future situation, the minimum geospatial requirements will be clearly defined, consistently referenced (using common terminology) and readily accessible to organisations involved in vehicle testing. These requirements (Figure 2) are likely to be different for varying self-driving vehicle driving scenarios, such as:

- vehicles operating in controlled environments: airports, arenas or theme parks
- vehicles operating alongside non-self-driving vehicles on public roads
- vehicles operating solely in a self-driving and connected environment

Figure 2 – Minimum requirement scenarios (source: OS)



It is expected there would be some commonality across datasets in each of these scenarios derived from a single authenticated and maintained source. An example of such commonality across all three categories might be lane markings, universally needed for ensuring that the vehicle operates within safe boundaries.

1.3 Simulation, track and real-world testing

Before a manufacturer trials a vehicle in a designated environment, typically the public highway, there is an expectation that its operation will not pose an unreasonable level of risk to any passengers or persons in proximity to that vehicle; in short, it is safe to operate. Defining what is 'unreasonable' will need to be determined, and there are several studies that discuss this, two of which are cited here:

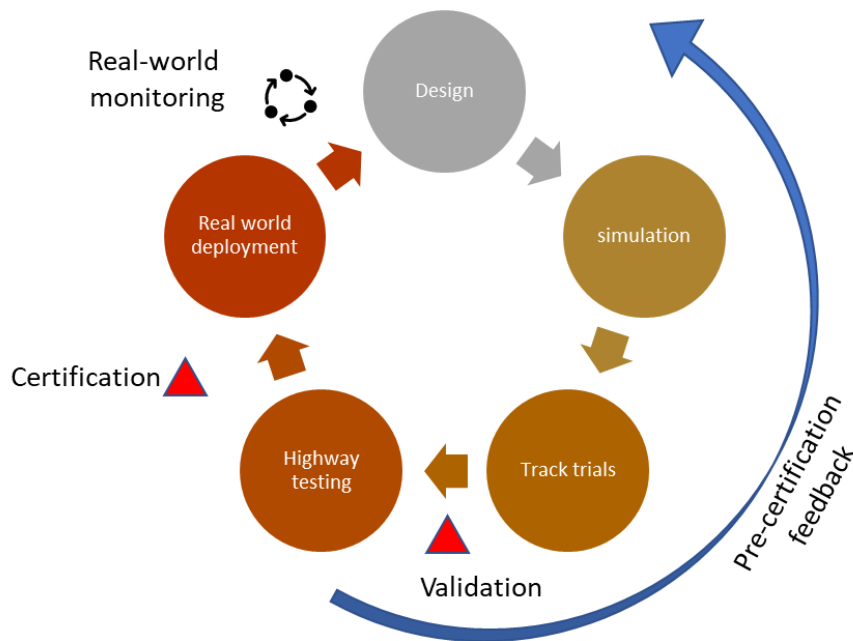
1. General Motors' safety report (General Motors, 2018) talks about:
 - zero crashes
 - zero emissions
 - zero congestion

as measures that will define a safe self-driving vehicle environment.

2. The Association of British Insurers issued a document (ABI, 2019) which discusses the role of the insurer and the expectations that self-driving vehicles will lead to safer driving.

Figure 3 provides a simplified view of the testing life cycle, from design through to real-world operation and deployment.

Figure 3 - Testing life cycle (source: OS)



Simulation of the real world enables a variety of scenarios to be tested, and typically includes how the vehicle will behave:

- in a simulated real-world environment of a chosen geographic area
- in a traffic scenario simulation, featuring vehicle types, operation volumes etc
- when confronted with events that require the vehicle to react, such as someone running into the carriageway

By means of these testing scenarios, manufacturers can determine risk levels and modify vehicle software algorithms so the vehicle can respond accordingly.

The next stage of testing is to replicate some of the environmental conditions created in the simulation process described above on testbed facilities, by introducing equivalent physical features and hazards. This will require test track operators to geolocate any specific features or hazards in accordance with agreed criteria, allowing the vehicle to be 'driven' along similar routes to those trialled through simulation. After such testing, risks will be better understood, and further refinements made if required.

Provision of these features needs to replicate the physical attributes of TROs to ensure that the vehicle can respond in accordance with either the source TRO data or the features indicating the presence of TROs.

This is an interactive cycle of development and test which will ultimately demonstrate that the vehicle meets the agreed legal criteria. Laws governing self-driving vehicles for use on the public highway are due to be reviewed during 2019.

1.4 Terminology

One of the challenges confronting the industry is the use of common terminology to describe the environment and the routing rules in which self-driving vehicles will operate.

Within the UK, OS has developed a comprehensive schema for defining both the highways network and the routing and asset management that will continue to be built upon as the needs of the self-driving vehicle environment develop. The development of this schema will require close collaboration with government stakeholders, as well as local authorities and the self-driving vehicle industry at large. Without this collaboration it is difficult to envisage a situation where full interoperability can be achieved across geographic regions, between simulation companies and across testbed facilities. This study has drawn upon several commercial and government documents referenced throughout this report which clearly show a lack of terminological consistency and of any definitive supporting sources.

1.5 Data reliability and authority

Any geospatial data used must be reliable and authoritative so that the decisions being made by a vehicle control system comply with both national legislation and local bylaws. Dependence on signage alone is not adequate, as it may be in a maintenance cycle, not in place, or damaged. However, even in this event, the current rules of the road (defined by TROs) will still apply; so, for example, if a sign is missing, a vehicle will still be under legal obligation to obey it, but unable to depend on that visual identification to make driving behaviour decisions.

It would therefore be reasonable to expect the rules of the road to have geospatial reference data associated with them to ensure that they can be obeyed by simulators, applied to testbeds, and adhered to for testing on highways. The challenge is that in the case of local authorities (which are legally bound to apply TROs), this data can vary significantly in terms of its currency, format, accuracy and completeness. This data has historically been fit for purpose, but in the context of self-driving vehicles, throws up new challenges – both as to its reliability and in terms of how it can be harmonised across geographic locations.

1.6 Data quality and resolution

The above sections have highlighted the challenges related to reliability and dependability of third-party datasets, but there is a need to consider the quality of these datasets. Quality means different things to different people, but the most useful definition for business is that it means 'fit for purpose' (Marr, 2013).

In the context of this report, 'fit for purpose' is concerned primarily with vehicle navigation that ensures safe operation: does the data used increase or decrease the risk to human life, damage to the vehicle or a third-party vehicle, or property damage? Determining whether data is fit for purpose needs careful consideration when planning safe, reliable operation and routing for self-driving vehicles. For example, when examining a typical road sign, several attributes can be associated with it. How these attributes impact the risk then needs to be considered as defined above – what is the information used for, and how well defined it is for the purpose? Table 2, which is non-exhaustive, considers some of those attributes as an example and suggests, for illustration only, potential measures and a risk factor assigned to each one.

Table 2 – Example feature and attributes

Road sign (generic)	What is it for?	Units /definitions	Accuracy /limits	Risk factor if not available or to specification
Positional in X,Y,Z	To aid identification of an instruction at a specific location	<ul style="list-style-type: none"> Centimetres 	<ul style="list-style-type: none"> Centimetres 	High
Type, shape, dimensions	For recognition of instruction type	<ul style="list-style-type: none"> Colour – RGB Shape style Centimetres 	<ul style="list-style-type: none"> Adequate to distinguish type Centimetres 	High
Material	Collision behaviour Sensor behaviour	<ul style="list-style-type: none"> Material type, ductility Reflectivity 	<ul style="list-style-type: none"> Within tolerances to ensure repeatable behaviour 	Low
Instruction	To inform vehicle behaviour	<ul style="list-style-type: none"> Legible text suitable for machine learning 	<ul style="list-style-type: none"> Contrast tolerances Standardised fonts 	High
Orientation	Direction Readability	<ul style="list-style-type: none"> Absolute to a georeferenced point Relative, to position of natural lighting (sun) 	<ul style="list-style-type: none"> $\pm n$ radians from vehicle direction of travel $\pm n$ radians azimuth from road surface $\pm 2n$ radians from sign surface 	Medium
Illumination	Direction Readability	<ul style="list-style-type: none"> Illumination status None, top lit, bottom lit etc. 	<ul style="list-style-type: none"> Light source type Light source, lux 	Medium

It is possible to represent the real world at very high (millimetric) resolution, depending on the technology applied. In the case of safe operation of self-driving vehicles, millimetric resolution is unlikely to be required, but a resolution to a few centimetres is more likely to be needed to ensure vehicle positioning is adequate and collisions can be avoided. Simulation and test track operation aims to align itself to the real world, but, equally, features in the real world must be georeferenced. In practice, the location accuracy of any feature, such as a signpost or a kerb, will be in the region of centimetres. The cost of raw data capture and post-processing to extract (identify) key features can be high, so it is important to strike the right balance between cost and resolution to ensure safe operation.

Consistent data capture and feature extraction measured against a defined specification will be required for the self-driving vehicle market, to ensure interoperability. The specification could be owned by an independent body such as Zenzic as part of its governance role, or defined through development of new standards from, for example, BSI. Under this arrangement, data capture and extraction of key features would take place against this defined specification and could be conducted either:

1. by an independent organisation such as OS as part of its public task; or
2. under direct contract to commercial suppliers.

Option 1 – OS is investing in data improvement to support the emerging markets, and this option has the key advantage of also supporting the upcoming Public Sector Geospatial Agreement (PSGA), the scope of which is currently under discussion with the Geospatial Commission. Any additional market insight from Zenzic on this topic would be useful for the Geospatial Commission to consider.

Option 2 – This approach will attract commercial rates, but given the growing number of organisations capable of performing this work, scope for negotiation exists. It is not usual for such organisations to store and maintain the acquired data, and if such services were required on a national scale, these would need to be separately discussed and negotiated. However, given that other organisations – for instance, partners in the ConVEx project – are developing a neutrally-hosted data repository (see section 0), such a service from the data suppliers may not be required.

1.7 Data provenance and currency

A critical aspect of data acquisition is understanding how it was obtained. Ideally, users should have assurance that data conforms to an agreed standard and update regime. Standards will be discussed in section 0.

Geospatial data can be complemented with key attributes from multiple sources – for example, local rules and restrictions applied to segments of highway or across geographic areas. These data sources can be used by a self-driving vehicle to make routing decisions; examples include a one-way street, parking (subject to time-based rules) and restriction zones. These rules are always applicable regardless of whether a vehicle can or cannot ‘see’ (detect) supporting signage or markings. The rules as presented to a vehicle therefore need to be trustworthy.

Local authority datasets form a primary source of road data, but this data is generally understood to be of variable quality (North Highland, 2018), and its availability is often in various formats, not all of which are machine readable. For its original purposes, such as asset registers to plan maintenance, this data or information is perfectly adequate; however, it is inadequate for the precision location of assets that a self-driving vehicle may encounter.

As an example, Figure 4 depicts the position of lamp posts in a large town, and shows clearly that their actual positions and their positions as recorded by the local authority differ – and in some cases either the record does not exist or a record has no corresponding real lamp post. The vertical pink lines show the locations of lamp posts as defined by the local authority and the vertical white lines show their locations according to high-resolution survey data. The difference is not that important to the local authority when considering, for example, maintenance, as the location accuracy is good enough for an engineer to identify the correct lamp post to work on. However, for emerging markets, including 5G and self-driving vehicles,

positional accuracy is essential for location of millimetric wavelength transmitters and for navigation reference.

Figure 4 – Lamp post spatial error (source: OS)



Note: the pink lines show LA defined locations of lamp posts and the white show the real locations

Making such data suitable for use by self-driving vehicles requires both accurate location data and current attribution data, along with an appropriate maintenance regime. In some cases, mobile mapping can assist in identifying key attributes, such as by distinguishing between a 'stop' and 'give way' sign. Other key attributes may be known only by the local authority, who may then need to be consulted directly to obtain them.

1.8 Standards

In 2017 China released an initial national standard for the testing of self-driving vehicles (Intelligent Transport, 2018). However, it is fair to say that current self-driving vehicle standards are embryonic and have not been approved or adopted internationally. In the UK, BSI is taking the leading role in developing standards, and will be developing Publicly Available Specifications (PAS) documentation on an interactive basis to support this market. These PAS documents may cover a variety of areas, but, notably, geospatial and safety appear to be high on the focus of the initial agenda.

As self-driving vehicles develop, a good initial understanding of geospatial requirements will be needed early on to provide a baseline for initial testing and to enable BSI to progress its work on standards. Standards will become a focus in the next few years as a key enabler to getting self-driving vehicles from testbed to road. This importance can be evidenced by the attention that the 2019 'Autonomous Vehicle Safety Regulation World Congress' (UKi Media & Events, 2019) is putting into highlighting standards, safety, rules and regulations on its agenda.

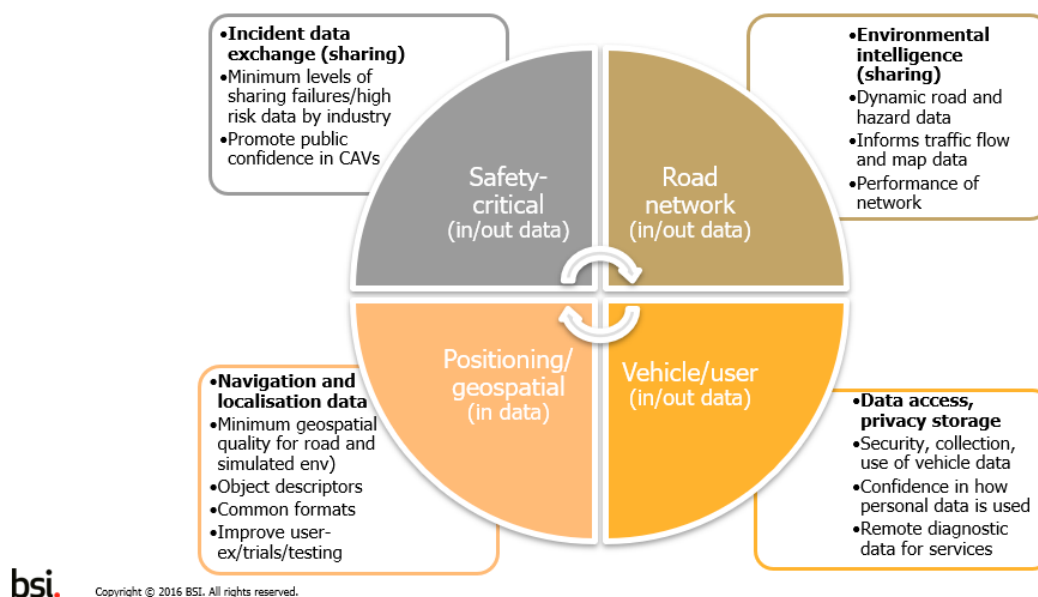
In 2016, BSI and the Transport Systems Catapult, supported by CCAV, began research to explore the priorities for standards to support the development and deployment of connected and self-driving vehicles in the UK (Fleming, 2017). Four key areas were identified relating to connected and self-driving vehicles, three of which have a strong geospatial dependence:

- incident data sharing;
- environmental intelligence sharing – especially traffic flow and dynamic hazards; and
- navigation and localisation data.

This is further elaborated in Figure 5.

Figure 5 – CAV data mapping: known challenges/gaps (source: BSI)

CAV Data – mapping known challenges/gaps



BSI is continuing to engage with industry and government to develop standards in the form of PAS documentation over the coming months. In March 2017 BSI published its connected and self-driving vehicle standards strategy summary report (BSI & Transport Catapult, 2017). Its primary objectives are to:

- map the current international standards landscape relevant to connected and self-driving vehicles
- provide better understanding of the key challenges and opportunities facing UK-based organisations working on connected and self-driving vehicle development and deployment
- identify areas where standardisation may be needed to help accelerate connected and self-driving vehicle deployment in the UK

The report contained five key recommendations that include harnessing experiences from connected and self-driving vehicle programmes and collaborating more closely with industry and government. It is likely that an 'agile' methodology (APM, 2019) will be adopted to meet the pace and needs of the market. OS intends to support this work by adding its experience from a geospatial perspective and ensuring that key geospatial issues for the emerging markets are considered.

1.9 Governance

There are many rich sources of data available that define the real world and can be utilised for the development of self-driving vehicles. It is unlikely all these sources will be required; for those that are, provenance, formats and overall quality must be considered.

OS has identified that there are a number of projects (see section 0) underway that all aim to provide a neutrally-hosted environment for their own investigative purposes, but this begs the question: why is there not one common environment today that these projects can all use for their own purposes? Such an environment does not currently exist, and Zenic has had the foresight to ensure that this issue will be addressed, and one will be established.

A government initiative is looking at aligning local authority TRO data (North Highland, 2018), but this explores primarily the current state of regulatory needs rather than future demands. Local authority data will be a key data source for self-driving vehicles, therefore these future demands must be considered. The wide range of initiatives in play need to be strongly governed to ensure that there are no gaps in the overall approach and that all activity maintains the focus, which is to deliver self-driving vehicles into the market by 2025.

2. An overview of road/street environment: real-world elements

To allow effective and reliable testing of self-driving vehicles, the environment in which a vehicle is going to be tested must be as representative of the real world as possible. This can be done either by building a realistic physical representation of road and street environment (such as testbeds) or by building their digital representations (for simulation).

A DfT guidance document *Manual for Streets* (Department for Transport, 2007) gives a clear explanation of the distinction between streets and roads.

"Roads are essentially highways whose main function is accommodating the movement of motor traffic. Streets are typically lined with buildings and public spaces, and while movement is still a key function, there are several others, of which the place function is the most important."

This chapter gives an overview of types of real-world elements that can be observed within both the road and the street environment while driving. These real-world elements include static, temporary and dynamic objects which should be taken into consideration while testing self-driving vehicles.

The static elements, outlined in the section below and examined in following sections, should be represented within the testing environment. However, the specifics and details of their representation should be decided, as they might have higher or lower relative importance to self-driving vehicles. Temporary and dynamic elements should also be represented within the testing environment; however, the analysis of these is not included in this report.

For the rest of this report, the terms 'road' and 'road environment' are used, but they refer to elements of both road and street environments.

2.1 Static elements

When driving one observes various elements which belong to a road. A driveable road surface is called a **carriageway** (Department for Transport, 2007). In the majority of urban areas, a road carriageway is bordered by a **kerb**, a border of stone, concrete or other materials formed at the edge of the carriageway (Ordnance Survey, 2001).

On the side of the road there might be a **pavement**, a paved surface adjacent to a road for use by pedestrians (Ordnance Survey, 2007) which is also referred to as a **footway** (Department for Transport, 2007) and **verge**, a natural area adjacent to a road between the carriageway/kerb and the pavement or any other adjacent delimiting features such as fences, hedges, trees or walls (Ordnance Survey, 2007)

The traffic flow and restrictions are being dictated by informative and regulatory **traffic signs, traffic lights (signals)** and various **road markings** painted onto the road metalling (Department for Transport, 2019). The traffic flow can be broken by different types of **pedestrian crossing**, a transverse strip of road where pedestrians should cross, indicated by road markings, dropped kerbs, and/or lights (Ordnance Survey, 2001) sometimes with a **traffic island**, a paved or planted island in the middle of a road sitting above the level of the carriageway, designed to guide or separate traffic.

Besides road signs and traffic lights, there are many other types of **street furniture**,¹ that is to say equipment installed along the street and roads for the benefit of the public. These include street lights, lamp posts, street signs, bus/tram stops and shelters, to name just a few.

The vehicles move on the road in **lanes**. These are delineated on a road carriageway to accommodate a single line of vehicles and limited by physical obstructions or road markings. Sometimes, especially in rural areas, the extent of the road surface can accommodate only one lane of vehicles but is used for driving in both directions. Some lanes have a specific user type, like bus, taxi or cycling. To indicate the specific user of the lane, a road sign and/or road marking is used.

An intersection where two or more roads meet or cross constitutes a **junction**. There exist many different types of junctions, which can be unmarked or marked, uncontrolled or controlled.

While driving, people pass on or through structures, such as **bridges**, which are built over rivers, railways, roads or ravines to permit the flow of the traffic (Ordnance Survey, 2001) or **tunnels**, long narrow passages under the ground or water (Ordnance Survey, 2001).

2.1.1 Static element properties

The road environment is very complex, comprising objects which need to be replicated on testbeds or represented in simulations to test and train self-driving vehicle behaviour. These include physical objects, such as kerbs, traffic signs and lamp posts, and also information painted on the surface of the road, which in turn may include different types of road marking. It is not only the representation of these objects which is important, but also information about their attributes – as these can have an impact on self-driving vehicle sensor system performance.

For example, through recent work with the University of Surrey's 5G Innovation Centre, 5GIC, (Ordnance Survey, 2018), it has become clearer that in the millimetric wavebands, radio communication can be compromised by different material types; for example, a large metal road sign could attenuate signals vital for vehicle-to-infrastructure (V2I) or vehicle-to-cloud (V2C) communications.

Each of these static elements of road environment have different properties/attributes which may be of importance to self-driving vehicles. This data is likely to include: name, unique number, shape, dimensions, type, class and material.

2.2 Temporary and dynamic elements

To create a reliable testing environment for self-driving vehicles it is necessary to consider, in addition to static elements, the representation of temporary and dynamic features that exist within the road environment.

Temporary objects, such as waste bins, cones, Portakabins or skips, exist in a location for a period of time. As with static objects, some of the attributes of temporary features (such as material type) can mislead self-driving vehicles' sensors or interfere in vehicle-to-vehicle (V2V) and V2I communication. An extended list of temporary elements (structures) identified during the workshop by participants can be found in Annex 1.

Dynamic objects, such as other moving vehicles, or pedestrians walking along the pavement or crossing the road, are of real significance to self-driving vehicles and need to be accounted for in testing environments to ensure that sensors accurately recognise moveable objects and can react in time to stop or to safely avoid them.

¹The term 'street furniture' has no formal industry definition and includes many objects within the road environment. This is a good example of where terminology would benefit from standardisation.

2.3 Weather

While physical objects such as kerbs, roundabouts and traffic islands are permanent in nature and durable under different weather conditions, weather of certain kinds can cause road markings to wear off and their visibility to be limited, potentially affecting self-driving vehicles' navigation.

Moreover, the direct impact of weather conditions on self-driving vehicles' system performance can be critical, both in terms of road holding and of sensor systems performance – for example on a wet road during a sunny day, where reflections may confuse visual sensors.

3. Requirements for testbeds and simulation companies

This section explores and examines the information gathered through the workshop with testbeds and companies that are creating simulation environments, and through a questionnaire where participants were asked what geospatial data is critical to support interoperability across self-driving vehicle life cycles. It also includes information gathered through OS's involvement and experience in self-driving vehicles projects, particularly E-CAVE and OmniCAV.

3.1 Industry perspectives

Communication between different stakeholders can be very challenging. Automotive industry, simulation and mapping players use different technical language to describe their respective working environments. Enabling interoperability across different stages of the self-driving vehicle's life cycle and effective communication between parties will require a common terminology.

A perfect example of where terminology differs amongst different stakeholders is **road geometry**, a critical element of the road environment. Without highly accurate information and representation of geometric parameters of the road, it is not possible to build an environment suitable for testing self-driving vehicles.

According to simulation companies, the road geometry includes information about its shape, curvature and elevation. Highways engineers focus on road alignment, profile and cross-section. Within the mapping domain, the extent of the driveable surface, and its dimensions and position above the sea level are captured and presented.

3.2 Critical geospatial information for operation of self-driving vehicles

This section explains which objects or features within the road environment have been recognised as critical in supporting interoperability and collaboration across a self-driving vehicle's life cycle, including effective and reliable testing.

The overall geometry associated with a road, including its inclination (slope) and general condition (including road defects such as potholes) (Makwana, 2018), together with the road surface, material classification, coefficient of friction, and surface reflectivity at different wavelengths, can all have an effect on a vehicle's actual or anticipated behaviour. DfT has conducted extensive research into road surfaces and road surface defects (DfT, 2019), which indicates some of the challenges self-driving that self-driving vehicles may have to encounter if they are to travel on existing infrastructure.

Potholes are a common occurrence on UK roads; 512,270 potholes were reported last year to 161 authorities (Cockburn, 2018), but they are not always known to local authorities. Sensor technology is likely to provide a valuable source of real-time information, but a road's surface finish may preclude a sensor from identifying potholes, resulting in the delivery of false information to the vehicle.

Similarly, whilst sensors have a key role to play, advance knowledge of slope, surface material and coefficient of friction will help inform a vehicle as to how it chooses its route and safely proceeds and stops.

The most significant information for navigation on the road relates to road markings and traffic signs. These usually complement each other (such as the 'give way' traffic sign and the equivalent road marking (a triangle symbol on the road surface and 'give way' line at

intersections), while TROs dictate the rules the driver must obey while moving vehicle on public roads.

Representation of these features within a simulation environment can be achieved by creating a highly accurate three-dimensional (3D) model of the road. Some of these features can be extracted and modelled as separate 3D objects (such as traffic signs or traffic lights); others can be extracted and visualised as vector data (such as road markings). There may be no need to model each 3D object separately; a library of objects (such as traffic lights) can be used to place each model in an accurate position for the feature it represents. Each of them should contain attributes which describe these features. The list of attributes related to each of the features described below is not exhaustive and can be extended by linking to information from different sources (such as local authorities).

The geospatial data requirements for future self-driving vehicles are unknown. Self-driving vehicles may conceivably not need any mapping at all to safely navigate on public roads. They might be self-sufficient when fully connected to each other (V2V), to infrastructure (V2I) and to the environment surrounding them (V2X). This is something for longer-term consideration. OS believes that to enable safe navigation of self-driving vehicles, high-resolution geospatial data is crucial, vitally providing a very accurate representation of the real world.

3.2.1 Street furniture

'Street furniture' is a collective term for objects and pieces of equipment installed along streets and roads for various purposes, but mainly for road safety and pedestrian mobility. Street furniture is recognised as a critical class of objects within the road landscape which should be represented in testing environments for the benefit of self-driving vehicles. Each of the groups below has a specific purpose and might have higher or lower relative importance to self-driving vehicles (reflected in their position within the list below). However, it is recommended that further investigation to recognise their importance be conducted. Some of them might be used only as an obstruction, others merely as a point of reference.

(1) Traffic signs

These have been erected at the side of the road or above the road to give instructions or provide information to road users. According to the *Highway Code* (Department for Transport, 2019), traffic signs include:

- signs giving orders;
- warning signs;
- direction signs;
- information signs; and
- roadwork signs.

Attributes such as accurate location, height, dimensions, shape, type, material, and orientation should be captured and stored. At the workshop, participants also mentioned other attributes of traffic signs that could be of importance to self-driving vehicles, such as age, maintenance, permanency, language, obscurity or linkage to other objects.

(2) Light signals for traffic flow control

These are signalling devices positioned at or above the road, at intersections, pedestrian crossings and other locations to control flows of traffic. According to the *Highway Code*, light signals controlling traffic include:

- traffic light signals;
- flashing red lights (at level crossings, lifting bridges, airfields, fire stations etc);

- motorway signals (informing of the maximum driving speed, closed lanes, reduced visibility, end of restriction); and
- lane control signals (indicating available or closed lane).

As with traffic signs, traffic lights should be attributed with their accurate location, their height, shape, type, material and orientation and so on.

(3) Street signs

These provide the information signs used to identify named streets. There is no national standard for street name signs; each district council / local authority has its own standard and specification. Street name signs can be placed on posts or be attached to building facades. The priority is to accurately capture their location, dimensions, material and the text on them.

(4) Public transport stops

Public transport stops are designated places where public transport vehicles stop for passengers to board or alight. They are usually identified with a pole and flag to mark a location, therefore can be represented and attributed similarly to the traffic signs. In busy locations, public transport stops may also be identified with shelters. The impact of signage on communications at millimetric level for 5G can have an impact on signal propagation, meaning that key attributes of a sign, including material type, should ideally be captured.

(5) Street lights and lamp posts

Streetlights and lamp posts are raised sources of light on the edge of the road and path. It is not clear whether, in the context of self-driving vehicles, their precise location, height and orientation is of significance. In respect of 5G small cell radio deployment, lamp posts have a role to play as cost-effective candidates for antenna location, and it is anticipated that self-driving vehicles may have a dependency on 5G in the future. Therefore, it can be argued these features are relevant. If 5G is taken out of the equation, the value of the lamp post may not be so significant. Its location and attributes can help identify lighting conditions which may be beneficial for daylight cameras, but it can also be sensed by lidar and therefore may not be so significant as a static hazard.

(6) Other features limiting traffic movement

These include traffic barriers, bollards and other permanent obstructions intended to keep vehicles within their carriageway or prevent them from entering forbidden areas. It is important to capture their accurate location. Other attributes, for instance material type, may also be of importance to self-driving vehicles.

There are other types of street furniture – like benches, post boxes, phone boxes, public toilets, fire hydrants, litter bins, poster poles and advertising columns – and knowing their accurate location is critical. These do not have a direct impact on the driving behaviour of self-driving vehicles; however, their shape, width, height or material could obstruct the view and have an impact on self-driving vehicles' sensor system performance.

3.2.2 Road markings

Road markings on road surfaces exist to convey official information and provide guidance and information to drivers. Accurate location and representation of these markings is of very high importance to self-driving vehicles.

According to the Highway Code (Department for Transport, 2019), road markings include:

- across-the-carriageway road markings ('stop' lines and 'give way' lines);
- along-the-carriageway road marking (lane lines, lane edge lines, centre lines, hazard warning lines, double white lines, diagonal stripes or chevrons);
- markings along the edge of the carriageway – waiting restrictions indicated by yellow lines which could be accommodated with a traffic sign indicating waiting times restriction; this also includes parking bay markings with a traffic sign indicating the parking restriction time and red route stopping controls;
- markings on the kerb or at the edge of the carriageway (loading restriction on roads other than red routes);
- other road markings, such as school keep clear zigzag lines, 'give way' triangles, public transport stops and lanes, taxi stands, box junctions, and markings indicating the direction of traffic lanes.

3.2.3 Junctions

A road junction is where two or more roads meet. There are different types of junctions, including unmarked or marked, uncontrolled or controlled, box junctions, T-, Y-junctions and different types of roundabouts. Accurate location of the junction, its type and possible manoeuvres at the junction are of importance to self-driving vehicles, as this information dictates the rules for different types of driving behaviour.

3.2.4 Road lanes

In the context of traffic control, a lane is part of a carriageway that is designated to be used by a single line of vehicles, to control and guide drivers and reduce traffic conflicts; depiction and representation of lanes is therefore critical. Most highways have at least two lanes, one for traffic in each direction, separated by a central reservation or lane marking. For the benefits of self-driving vehicles, it is necessary to describe the direction and width of each lane, the lane type (such as flare lane, exit lane) and the lane priority or use (such as a bus, taxi or cycling lane). Incorporated into future attributes might be a 'self-driving vehicle priority lane'.

3.2.5 Pedestrian crossings

Pedestrian crossings are safe places designated for pedestrians to cross the road, where they are given priority. These crossings can be recognised within road environment by dropped kerbs and tactile paving, but can also be indicated with road markings, traffic signs, obstructions, lights and light signals. An accurate location for each of these feature types must be captured and represented for use by self-driving vehicles.

There are five different types of pedestrian crossings:

- Zebra crossings are identified by black and white stripes that form a path across a road and flashing yellow Belisha beacons at either side of the road;
- Pelican crossings use buttons, lights and sounds to allow pedestrians to cross the road safely;
- Puffin crossings operate in a very similar way to Pelican crossings, but are fitted with smart sensors; the other difference is that the signal for safe crossing is next to pedestrians, rather than on the opposite side of the road where it is in a Pelican crossing;
- Toucan crossings are generally wider to accommodate both pedestrians and cyclists safely crossing the road;
- Pegasus (also known as equestrian) crossings are similar to Toucan crossings but are designed for both pedestrians and horses to cross the road safely together.

There are also other places designed for pedestrians to cross the road, such:

- school crossings accommodated with specific road markings and street signs;
- 'refuge' crossings, usually not marked but characterised with tactile paving, dropped kerbs and a traffic island in the middle of the road with a white Belisha beacon.

3.2.6 Traffic islands

A traffic island is a solid or painted object in a road. It can also be a narrow strip of island between roads that intersect at an acute angle. If the island uses road markings only, without raised kerbs or other physical obstructions, it is called a painted island or ghost island. Its main purpose is to channel the traffic therefore should be accurately captured and represented within the data to allow reliable testing of self-driving vehicles.

3.2.7 Taxi stands and public transport stops

Taxi stands and public transport stops are usually identified by a single pole with relevant signage, shelters in busier locations, and associated road marking. It is important to capture and represent this information for the benefit of self-driving vehicles.

3.2.8 Parking locations

As with the features above, in most cases parking locations, such as bays and on-street parking, are also marked with both relevant signs and road marking, and are considered to be important for self-driving vehicles.

3.2.9 Vegetation

Vegetation in the roadside environment, especially hedges, rows of trees or single trees, are of great significance to self-driving vehicles. These objects are not permanent in the same sense as other physical objects within road environment – they change between the seasons and their shape changes over time through natural growth or management. For example, non-coniferous vegetation foliage changes during the year, and can obscure physical objects, such as traffic signs, which can affect a self-driving vehicle's behaviour. Vegetation not only obscures physical objects but also has a profound impact on millimetric wave radio communication (5G). Therefore, the vehicle needs to be able to ensure that it is not dependent solely on a single communication method for safe navigation.

3.2.10 Road network

A road network comprises interconnecting lines and points that represents the logic of vehicle movement on the road. It provides the foundation for network analysis and is used within driving simulation applications. A road network is based on available physical and painted information within the road environment, including the extent of driveable surface, street furniture, road markings and TROs. It can inform self-driving vehicles about traffic direction (one-way/two-way), turn restrictions and lane reservations (taxi, bus, cycle lanes).

3.2.11 Buildings types and materials

During the workshop it was identified that a knowledge of where buildings are relative to the highway is significant to self-driving vehicles in at least four ways:

- the building location, denoted by its address, which can be useful for navigation purposes;
- the building as a point of reference or a landmark to which a vehicle could refer if other source data or sensor data is uncertain;

- the building as a hazard, if it is very close to the highway – for example a street with terraced houses whose doors open directly onto the street or within, say, a metre of it; public buildings typically have increased footfall, and so awareness of pedestrian flows can be factored into predictions of heightened hazard or congestion;
- the building as a potentially radio-reflective object, if it is close to the highway and made of certain types of materials, such as material which is highly reflective both to optical spectrum and radio spectrum (as used by 5G sensors); this information can be used to predict whether this might disrupt vehicle operation.

Building information comes from a variety of sources, including OS for building type, public building identification, authoritative addresses and, in some cases, major landmarks.

Modern buildings may have plans that can be digitally accessed to provide construction material information, but this will not be centrally held. Information regarding older buildings may exist in records but is unlikely to be in digital format, or consistent. New data capture using aerial photography and lidar could be used to obtain such information. This information is not held centrally by any authoritative body.

4. Data sources for self-driving vehicles

Currently there is no central source of geospatial data that supports interoperability through the full life cycle of self-driving vehicles.

To allow effective and reliable testing of self-driving vehicles there is a requirement for highly accurate capture and representation of real-world objects. This can currently be achieved only for small areas, usually within individual projects.

There is therefore a need for creation of a standardised representation of the road environment which would enable interoperability between separate projects and geographies, and support scalability.

During the workshop, OS, local authorities and central government bodies were mentioned as key sources of information. These sources currently provide only limited support to self-driving vehicle requirements, but offer a valuable baseline for referencing or validating more accurate and detailed data which represents the real-world road environment in three dimensions.

This section describes in more detail the geospatial data currently provided by OS, local authorities and other sources.

4.1 The 'digital twin'

The term 'digital twin' (Various & Wikipedia, 2019)² is commonly used across industry, and in particular manufacturing, to define a digital facsimile of a physical object and the processes associated with it. For example, digital models of a machine can be built to predict its performance, and virtually modify components to optimise the performance, before embarking on costly investment in machinery and tooling to build the physical item.

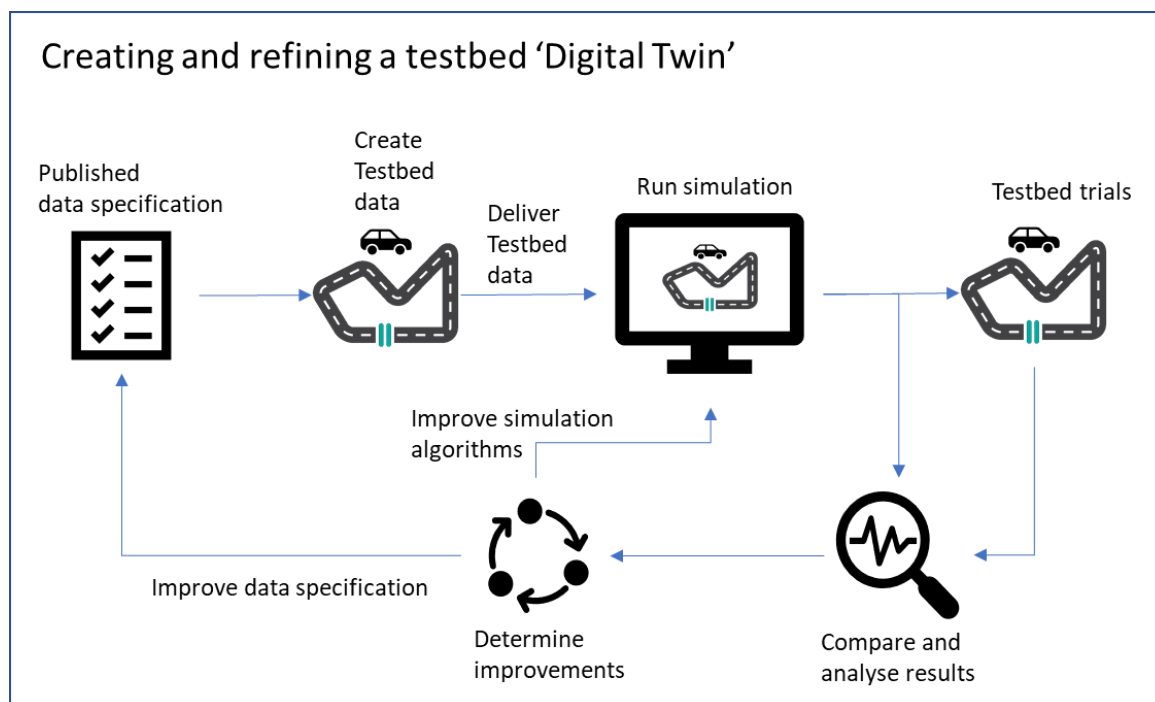
'Digital twin' is a term being rapidly adopted across many disciplines, and the relevance and appropriateness of its use should be carefully considered.

In the context of self-driving vehicles, it is equally important to be able to trial vehicle behaviour in a safe environment – and that vehicle behaviour can be simulated, provided the necessary data is made available. Determining the right level of data needed for reliable and trusted simulation will start with establishing a baseline, much of which is considered in this report; over time this baseline will be refined as simulation, trials and technology progress.

For testbed facilities, this will mean creating a digital twin of their physical test environments. This digital twin will need to meet a baseline specification for conformity and interoperability. It will also need to be delivered to an agreed level of spatial resolution and provide details of critical feature types, as defined in this report. This digital twin dataset can then be fed into simulation tools. Figure 6 demonstrates how such a digital twin may be used in the testing cycle.

² Wikipedia would not normally be cited as a reference, but given the diversity of the meaning of 'digital twin' across many industries, this reference is useful for highlighting the wide-reaching nature of the term.

Figure 6 – Digital twin for simulation testing (source: OS)



4.2 Ordnance Survey data

OS has been capturing and representing real-world objects on a national scale for over 220 years. It has skills, knowledge and experience of capturing geospatial data and creating a range of mapping products which are authoritative, trustworthy and standardised.

OS MasterMap (Ordnance Survey, 2019) is OS's most detailed product family. It is a consistent and maintained framework for the referencing of geographic information. It comprises separate but complementary layers that provide detailed topographic, photographic and network information positioned on the National Grid. Combining all these layers provides the definitive source of highly detailed geospatial data of Great Britain.

Data offered within OS MasterMap products can support markets as they are today, but an even more detailed representation of the real world is required to support the demands of new emerging markets, including self-driving vehicles. The paragraphs below focus on the geospatial data currently included within OS MasterMap that can be used as a reference for more detailed representation of the real-world road environment.

OS MasterMap Topography Layer (Ordnance Survey, 2019) contains features that represent objects in the physical environment. The data is delivered as a seamless, geographically contiguous entity. The basic unit of its data are points, polylines and polygons that represent real-world features. Each feature has its own unique topographic identifier (TOID) and contains attributes about accuracy and life cycle. Data is supplied in an open, tabular GML-based format, with a six-weekly product refresh.

The road extent is represented within Topography Layer as a set of topologically structured polygons representing road, track, pavement, verge, traffic calming or bridge. Obstructions along the road preventing access to forbidden areas, like barriers or bollards, are represented as polylines or points. Poles, posts, and single coniferous and non-coniferous trees are also represented as points, and buildings are represented as polygons. The third dimension can be added to this dataset by joining information included within Building Height Attribute (Ordnance Survey, 2019) – an enhancement to Topography Layer supplied in CSV (comma-

separated values) format. This enables the data to be visualised in 3D and to be used in a range of analytical applications to provide understanding of the built environment.

The AddressBase product range (Ordnance Survey, 2019) provides up-to-date, accurate information about addresses which can be easily linked to building features within Topography Layer by TOID. Address records include local authority and Royal Mail addresses, as well as pre-build, historic and alternative addresses, with coordinates for each address.

OS MasterMap Highways Network (Ordnance Survey, 2019) is the authoritative road network dataset for Great Britain, bringing together OS's large-scale road and path content, the National Street Gazetteer and the Trunk Road Street Gazetteer. It is made up of four product schemas: Linear Highway Network, Highways Dedication, Routing and Asset Management and Highways Water Transport Network, which together provide views of the physical network, navigation and road information. The data is supplied in GML format as topologically structured link and node networks, and includes information about speed limits, average speed, road junctions, traffic calming, access and turn restrictions, heights, weight, length and width restriction for vehicles, hazards such as fords and severe turns, and structures such as bridges, gates, level crossings, moveable barriers or rising bollards.

OS MasterMap Imagery Layer (Ordnance Survey, 2019) is a 25 cm-resolution orthoimagery product that uses the same source imagery that underpins OS large-scale data updates, and therefore provides a complementary visual background to all products from the OS MasterMap family.

OS Terrain 5 (Ordnance Survey, 2019) is a Digital Terrain Model of Great Britain, available as a grid of heightened points and contours at five-metre intervals. It adds the third dimension to OS MasterMap Topography Layer, models significant landscape features such as roads, railways, slopes, quarries and lakes, and offers a typical accuracy level greater than 2 m RMSE.

OS detailed data products support the current market, but it is recognised that richer, more detailed and more accurate geospatial data is required to support new uses such as self-driving vehicles. Through OS involvement in self-driving vehicle projects such as OmniCAV, E-CAVE and Atlas, and through the information obtained from this study, OS recognises that there are opportunities to significantly enhance its data to address some of the gaps anticipated by this report.

4.3 Local authority data

Local authority data provides a rich set of information that helps describe the highways environment, ranging from traffic restriction orders to signage and lighting, to name but a few. It is well understood that this data is hard to access in any meaningful way and can be inconsistent in accuracy, provenance and currency.

In December 2018 DfT launched a £500,000 competition for opening local authority transport data. This competition followed a study and an independent report commissioned by DfT in August 2018, entitled Local Transport Data Discovery (North Highland, 2018). This focused not on the needs of self-driving vehicles, but on the general issues surrounding the challenges of providing and sharing open data and the benefits that this provides for all types of transport.

OmniCAV

OS is involved in the **OmniCAV** project, the objective of which is to develop a world-first artificial intelligence-based simulation model for testing self-driving car safety. Building on the unique strengths of OS capability and geospatial data, OmniCAV will help accelerate the safe deployment of self-driving vehicles on UK roads by creating testing experiences in digital environments that mirror the real world

OS's recent engagement with different organisations has demonstrated that geospatial aspects are complex and are key to the ambition for self-driving vehicles. These aspects include the creation of a realistic simulation environment to test in, and the eventual deployment of a new infrastructure to enable improved connectivity and mobility.

Many of the observations made in this study identified areas of opportunity that apply equally to self-driving vehicles. The key findings from this discovery work were as follows, with those particularly relevant to self-driving vehicles emphasised (North Highland, 2018: 6, emphasis added):

- Publishing open transport data offers potential commercial and societal benefits – as demonstrated by Transport for London – but there is currently limited commercial value associated with most local authority data outside London.
- There are pockets of excellence within local authorities, but much of the market is dominated by the private sector.
- **Early case studies are demonstrating the mutual benefits of collaboration across local authorities, the private sector, universities, and other government departments.**
- **Significant amounts of local authority data are currently closed – and there are barriers which need to be removed before the full benefits of open data can be realised.**
- **There is operational value in the data for managing road networks – and traffic data should be a priority dataset to open up.**
- **Investment is required to improve data quality and standardisation for operational and future commercial exploitation.**
- A significant volume and breadth of local transport data exists, which enforces the importance of a targeted approach to opening key datasets.
- Local authorities have not fully developed their approach for using transport data for land use planning, prioritising road maintenance investment and to support connected and self-driving (autonomous) vehicles.
- **There is significant enthusiasm within local authorities to progress the open data agenda, but guidance and support is needed to realise potential opportunities.**

The associated summary recommendations from the report echo the needs identified in the discovery workshop, in other words the key actions that DfT will need to consider in support of this market. The details can be found in the report (North Highland, 2018). The recommendations are broken down into five key themes:

1. Local authorities should be helped to focus on making more high-quality data open.
 - a. Establish sector-led programmes to identify data to be opened.
 - b. Work with local authorities to scale proven data initiatives.
 - c. Develop open data guidance for local authorities.
2. The DfT should sponsor identified data projects which encourage and foster better local authority transport services.
 - a. Create a framework and standards for local authorities to support current and future services.
 - b. Streamline and digitise TROs.
 - c. Develop a private/public national data catalogue.
3. More effective investment in infrastructure to harvest local authority data, and open data initiatives to improve data sharing.
 - a. Prioritise spend on infrastructure to capture data.
 - b. Increase DfT investment in open data initiatives.

4. Promote training and skills development within local authorities to develop internal capability.
 - a. Provide data procurement guidance.
 - b. Develop data skills and capability building.

5. Improve collaboration between local authorities, Highways England and the private sector.
 - a. Promoting cross-sector/boundary collaboration.
 - b. Improve data sharing across Highways England and local authorities.

If recommendations were addressed, many of the challenges relating to discovery, accessibility and interoperability could be overcome. Notably, themes 2 and 5 address several observations identified in this study.

4.3.1 TRO Discovery Project

A team comprising DfT, GeoPlace, The British Parking Association and OS worked on a project, that concluded in May 2019, that aims to address the current challenges relating to the management of TRO data within local authorities by compiling an initial spatial data model that, if adopted and developed, could provide a single framework that all local authorities can conform to when defining traffic regulations. In addition, it is proposed to accompany this with a user guide that local authorities across the UK will be able to use. It is important that the scope of this work considers not only the current state but also the future state of TROs.

4.3.2 Transport Network Intelligent Transport Systems (TN-ITS)

In addition to activity in the UK, DfT is a partner in the TN-ITS GO project to advance TN-ITS services in Europe (2018–21) (TN-ITS, 2019), which aims to develop a harmonised framework concerned with the exchange of information on changes in static road attributes. The continued engagement of DfT in this work is vital to ensure that the UK is established as being able to offer an authoritative infrastructure and environment for the testing of self-driving vehicles.

4.4 Other sources

Additional geospatial data can be sourced from a variety of sources, for example:

- The British Geological Survey (subsurface data, ground water, flooding, landslides, mines);
- the Met Office (weather, hydrometeors, probability of occurrence and simulation modelling);
- transport operators (timetables, routes);
- crowdsourced data (real-time traffic flow, road traffic collisions).

All these data types could be exploited to improve vehicle behaviour in the long term.

Other sources like OpenStreetMap (OSM), HERE maps, pre-existing transport models, and satellite and aerial mapping within Google or Bing platforms, were mentioned by workshop participants as sources of reference for geospatial information.

OSM (Various, 2019) is an interesting source of geospatial information, as it uses crowdsourcing to create a free editable map of the world. Volunteers gather location data using GPS, local knowledge, and other free sources of information, and this is then uploaded to OSM. The resulting free map can be viewed and downloaded from the OSM server. It is possible to view the map using four layers: standard, cycle map, transport map and humanitarian.

Standard and cycle maps contain the keys which refer to various types of features. Different types of roads, such as motorways, trunk, primary and secondary roads, tracks, cycleway and footway are represented as linear features. The maps provide information about street names and road numbers, location parking places, one-way streets, traffic lights, public transport stops, buildings, and points of interest such as shops, garages and pharmacies.

HERE maps (HERE technologies, 2019) is a web mapping and navigation service originally developed by Nokia and sold in December 2015 to a consortium of German automotive companies (comprising Audi, BMW and Daimler). The data includes map or satellite view, with layers displaying information about public transport (rail and bus connections, and rail, bus and coach stops) and traffic (traffic direction, intensity, lane restriction and roadworks). Maps are updated at two- to three-month intervals.

There is currently no one geospatial data source which would fulfil the requirements set out in this report for highly accurate data to underpin reliable testing of self-driving vehicles.

To capture high-resolution data, testbeds rely mainly on subcontracted private sector suppliers, such as KOREC, LandScope Engineering, MK Surveys and Getmapping. They capture data using mobile mapping, aerial survey, static, backpack and handheld scanning. Quality-checking of this data is then assessed internally.

Other players are entering the market too, such as Mobileye (producers of an advanced visually based collision avoidance AI system), which can tailor their solutions to be trained to identify certain feature types on the highway. The approach they offer can seek out changes in the environment and feed this back. Other entrants into the market who capture point cloud data may do this to varying degrees of accuracy and resolution, which may not permit interoperability unless exacting quality assurance is applied to the data or to extracted features before use.

Whilst there is potentially a wide variety of sources of data available, there is no single standards body which would specify the requirements for capturing this geospatial data and its representation. Data is being captured today, and digital representations are being created for specific and individual projects. This means that the data may be inconsistent and unstandardised, and therefore not scalable or interoperable.

Data capture processes need to be planned carefully, as they can have a big impact on the accuracy and quality of extracted data. There is therefore a real need for setting up specification requirements for 3D data capture, to support the interoperability through full life cycle of self-driving vehicles.

National or international interoperability will demand strong governance, common terminology, standards and ease of access to data that is known to be reliable and authenticated, and which can be readily modified to change with the environment.

4.5 Data capture, processing and formats

4.5.1 Data capture

To allow reliable testing of self-driving vehicles there is a need for a highly accurate representation of the real-world environment, something that is not currently fulfilled by any off-the-shelf product. The data available on the market, such as the OS MasterMap family of products or OSM, can be treated as base or reference for planning and capturing more detailed geospatial data.

Planning for data capture is key to ensuring accurate and reliable content. Additionally, ground control for any data capture mission must be carefully considered. Relative positional accuracy (which means the positional consistency of a feature in relation to other local features within

the same or another reference dataset (Ordnance Survey, 2019)) should be sufficient to enable the extraction and modelling of features such as dropped kerbs (25 mm above the carriageway for vehicular access, 12 mm for water checks (The Highways Agency, 1989)). Absolute positional accuracy (how closely the coordinates of a point in the dataset agree with the coordinates of the same point on the ground (Ordnance Survey, 2019)) will depend on ground control points, but should be better than 5 cm.

Highly accurate geospatial information is currently being captured for specific projects using a variety of methods. The most relevant for capturing the road environment is mobile mapping systems, where geospatial information is being collected using laser scanning (lidar) and optical sensors mounted on a moving vehicle. However, in using this method of capture, there are instances when the surrounding environment can be obstructed by other moving or parked vehicles, or large groups of pedestrians. To fill these gaps, other data capture techniques are being used, including aerial or satellite imagery or aerial lidar, static tripod-mounted laser scanners (total stations) or handheld scanners.

4.5.2 Data processing

Aerial or satellite imagery can be used to create high-resolution 'orthorectified' imagery. Orthorectification is a process of removing the effects of image perspective and relief (terrain) effects for creating a planimetrically correct image representation of the real world with all objects in their true position. Output imagery can be used as a 2D representation of the real-world environment or as a background for visualisation of 3D information. Common formats for this type of data are JPEG, GeoTIFF and ECW.

The output of scanning equipment is point cloud data, a set of 3D coordinates displayed as points that, when combined, define the physical shape of a surface. Point cloud data can also be acquired through dense image matching based on high-resolution satellite or aerial imagery. This approach has advantages, as it also enables the colourisation and classification of the point cloud data that can be achieved only by using information within imagery (RGB). Therefore, lidar and high-resolution camera are the most popular combination for capturing high-quality geospatial data. The most common point cloud data formats providing broad interoperability include LAS, LAZ, XYZ and E57.

To derive valuable 2D and 3D content from point cloud data, the data requires processing. Data processing involves cleansing the point cloud data by removing unnecessary noise (removing the points which are not consistent with the surface) and any temporary artefacts (such as moving vehicles, pedestrians, flying objects, parked cars and bins). In areas where objects have been removed, to achieve a contiguous representation, the point cloud can, if required and where possible, be filled in by using adjacent scans, or making use of other data (e.g. from oblique imagery) to provide an accurate base for the 3D modelling.

The processing also involves point cloud classification, which then – with varying degrees of automation – enables 3D modelling and extraction of essential features, such as road surfaces, kerbs, buildings, street furniture and vegetation.

4.5.3 Data formats

Point cloud data (in LAS, LAZ, XYZ or E57 format) can be used as a digital representation of the real world, and visualised within various free software packages such as MeshLab, CloudCompare, Open LAS Viewer and FugroViewer. It can also be further manipulated and transformed, to enable creation and extraction of:

- 3D elevation models, which can be provided as OBJ, GeoTIFF, ASCII grid, ESRI shapefile, GML, AutoCAD DXF or MicroStation DGN formats;
- 3D object models, for instance of street furniture, buildings or vegetation, provided in OBJ, FBX, DAE, STL, PLY or 3DS formats; or

- vector data depicting objects such as street furniture, kerbs and road markings, provided in ESRI shapefile format as point, line or polygon.

Extracted vector data can be enriched by cross-referencing it with additional datasets, for example information sourced from local authorities. This additional data, if stored in an interoperable format such as CSV, can be easily linked to relevant vector data using open GIS (geographical information systems) software, for instance QGIS. This could be information about road names or numbers, or specifics about street furniture such as unique ID, age or maintenance.

The most popular road network data formats within simulation companies are OpenDRIVE, OpenSCENARIO and OpenCRG; these are open formats describing road network for driving and traffic simulation. All three formats have been developed, owned and published by VIRES Simulationstechnologie GmbH and are currently being transferred to the Association for Standardization of Automation and Measuring Systems (ASAM) to become a public standard with an expectation of their first releases at ASAM during 2019.

OpenDRIVE (XODR) is an open file format for the precise analytical description of road network. All roads within OpenDRIVE are separated into road segments which consist of reference line (also called 'anchor line' or 'road centre line') to which various properties are attached, such as elevation, lateral profiles, lane records (width, markings, materials etc.), road signs, traffic lights, surface profiles, railway elements, tunnels, bridges and other road objects. Road networks are created by linking road segments either directly to predecessors and successors or via junctions. The data of OpenDRIVE is organised in a hierarchical structure and serialised in an Extensible Markup Language (XML) file format.

OpenDRIVE describes all static objects of a road network that allow realistic simulation of vehicles driving on roads. To render the complete environment, additional description formats for static 3D roadside objects, such as trees and buildings, are needed. Road surface profiles can be added by using OpenCRG file format (CRG). The dynamic content of driving simulations, such as vehicle manoeuvres, can be described with ASAM OpenSCENARIO (XOSC). These three standards complement each other, and together cover the static and dynamic content of in-the-loop vehicle simulation applications.

4.6 Data hosting

Key data for self-driving vehicles is likely to come from more than one source. This can give rise to several problems for those trying to use this data. Management of the data acquired, held and distributed will require a strong governance regime to be adopted.

Many of the challenges relating to standards, quality, format and provenance are discussed in this report; a further challenge is knowing where to source the data. An ideal situation would be to source data from one place, confident that it meets all the essential (minimum) requirements to be able to trust it for use in simulation or for designing testbed layout and feature positioning. This would ideally be built on a neutrally-hosted capability that would enable data for any given area to be obtained, in the assurance that it meets the minimum requirements.

With the adoption of cloud technologies, access to data is becoming easier; however, political and commercial pressures can still inhibit data from being shared widely and commercially. Cloud is likely to be the only technology able to cater for the vast volumes of real-time data that will be gathered from private and public sector bodies to meet the needs of simulation, the testbeds and the real world.

In the case of self-driving vehicles, data quality for safe operation and simulation needs to be governed. This does not mean that all the data needs to be held in one place and be centrally accessible. However, the federation of such data does demand an authoritative approach to

ensure compliance with the minimum requirements. Such an approach can ensure that the same data can be served to customers reliably, repeatably and to known specifications.

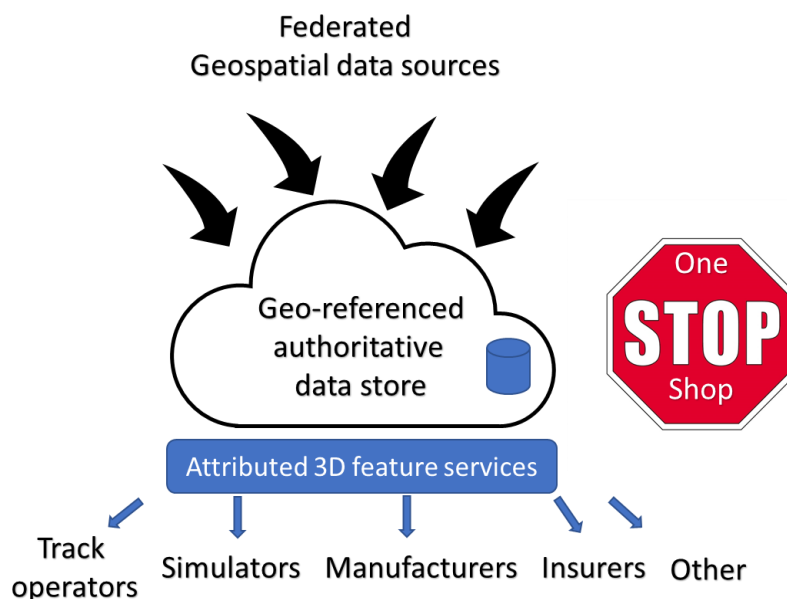
OS has sight of four publicly sponsored initiatives set up to explore the exchange of mobility data to support varying levels of technology readiness (TRLs), as outlined in Table 3. The very fact that there are four projects already leads to the conclusion that the concept of delivering a single data exchange to serve the market has some way to go, but through collaboration these organisations could – with the right governance – be brought together to meet the market demands from a single source.

Table 3 – Data exchange programmes

Project	Lead	Timescales	Scope	TRL
E-CAVE	OS	2017-21 (4 years)	Open/neutral exchange, geospatial data, safety use case	5-8
ConVEx	Bosch	2019-20 (1 year build phase, 10 year project phase)	Neutral exchange for commercial value-add	7-9
Intelligent mobility data hub	Transport Systems Catapult	2019-25 (6 years)	Platform for mobility data analytics supporting small and medium-sized enterprises R&D	7-9
UK Mobility Data Institute	Warwick Manufacturing Group	2019 (1 year)	High-performance computing facility for research on large-scale mobility data	2-4

The data exchange programmes identified in the table have been set up to trial the specific needs of self-driving vehicles. They aim to explore a breadth of TRL from 2-9. These programmes will trial specific use cases broadly outlined in the scope column, but each will build its own data store to achieve this. For research purposes this may be acceptable, but it does not address the fundamental challenge of providing one central and neutrally-hosted data store that is trusted by all those who need access – in effect a one-stop shop for data exchange. The ConVEx project has a commercial focus and, along with E-CAVE, will focus on the core needs of neutrally-hosted data exchange. The high-level principles of such a data exchange are depicted in Figure 7.

Figure 7 – Neutral georeferenced data store (source: OS)



Several government organisations have the capability to host and manage volumes of data in geospatial formats – for example, the Geospatial Commission’s ‘Geo6’, comprising:

- The Coal Authority
- The British Geological Survey
- UK Hydrographic Office
- HM Land Registry
- Ordnance Survey
- The Valuation Office Agency

These organisations are driven primarily by public task activity, and commercial drivers are in general not a prime motivator for their existence. However, the ownership status of these organisations within government varies, which may make selecting any one of these to act as a neutral host problematic.

In addition, the private sector must be considered. As well as consortia like ConVEx, several organisations, such as Google, and a range of system integrators from across the globe may have the capacity, skills and appetite to offer themselves to this market as neutral hosts, but they will expect a return on their investment.

Real-time integration of data comes with a new set of challenges. Traditionally, data is incorporated into a data store through a defined process often referred to as Extract, Transform and Load. This simply will not work for real-time datasets. New techniques will need to be developed to ensure that data that is published in real-time datasets does not compromise safety. Selection of a neutral host (if adopted) needs careful consideration where provision of safety-critical and open data on demand is required.

Recommendations

The market for self-driving vehicles in the UK, specifically that for road vehicles using connected and self-driving vehicle technologies, is projected to be worth £52 billion in 2035, capturing 6% of the £907 billion global market (Transport Systems Catapult, 2017). Further, the industry is on an accelerated curve to have fully self-driving vehicles on UK roads by 2021, as part of the Government's modern Industrial Strategy (Norman MP & Harrington MP, 2019), with UK government and industry investing heavily in a variety of work programmes covering a wide range of issues from infrastructure and safety through to insurance and legal matters. This report has focused on the major geospatial aspects of self-driving vehicles, exploring what data is available, what formats it is available in, and how it might be hosted to ensure overall interoperability and conformity across the industry to support the full testing life cycle.

On publication of this report, Ordnance Survey (OS) recommends that it is shared with government, industry and academia to solicit feedback and comment.

Data formats

The UK has a world-class physical vehicle test environment ecosystem, but to allow further support in self-driving vehicle development, the creation of a virtual test environment is needed. To further the UK's prime position in the demonstration, testing, development and deployment of self-driving vehicles, it is important that customers be provided with the same experience no matter which UK testbed they use. To achieve this, there is a need to create a specification for standardised, digital representation – that is, digital twins – of these sites. These digital twins should be highly accurate, but at the same time offered in the formats that conform to well-known or open standards which support interoperability and scalability.

Data that can be used to assist in vehicle navigation can be provided in many formats, some of which are niche or preferred by certain industries and/or systems. However, to support full interoperability, the industry will benefit from adopting a selected set of formats that allow systems to interact throughout the full life cycle of vehicle testing and trials. In this report, OS has identified nearly 30 data formats referring to geospatial data, and at least 20 different formats which relate to the creation of real-world digital twins of the road environment.

The reason for such a variety in formats is the present rapid development in technology, in terms of both hardware and software, which enables capturing the data in higher resolution and with better accuracy, and which allows automation in data processing and creation of 3D content. Because of the diversity in formats available on the market, there is a need to agree on a standardised set of formats which would enable interoperability thorough the full life cycle of self-driving vehicles, and which would allow the same testing experience from one UK testbed to another.

It is worth noting that the recommended data formats described below are based in part on the requirements shared by the participants during the workshop and captured through the subsequent questionnaire responses, but are informed primarily by OS's involvement in collaborative projects, notably E-CAVE and OmniCAV.

It is recommended that the 2D and 2.5D (i.e. heighted 2D) datasets currently available on the market, such as the OS MasterMap family product and OpenStreetMap (OSM), should be seen as a reliable source of information for strategic planning in the testing of self-driving vehicles, and during the planning phase of the creation of a more detailed digital representation of road environments, or testbed digital twins. These data sources provide, in Extensible Markup Language (XML) syntax, essential information about static elements within the road environment, such as roads, houses and some kinds of street furniture.

However, to create a digital twin of the road environment, for high-definition visualisation, driving and traffic simulation tools, data needs to be provided in much greater detail than these sources can offer. This report recommends that significant consideration is given to planning how the digital representation of testbeds is undertaken. Planning for data capture is key to ensuring accurate and reliable content.

To generate the most accurate digital representation of the road environment in terms of surface details of the road, kerbs and street furniture, it is recommended that a point cloud is captured using terrestrial mobile mapping systems comprising lidar equipment and optical sensors.

The recommended point cloud data format to support interoperability is LAS 1.2 file (or LAZ, the compressed version), which, when tiled (broken into smaller geographic chunks), enables the data to be processed more efficiently. This format can also be ingested by most point cloud processing software available on the market. Point cloud data enables the creation and extraction of different types of geospatial datasets that can be stored as layers which, when combined, form a digital representation of a road environment.

Format recommendations for key derived layers are considered in turn below.

- **3D model of the terrain**, which forms a base for other layers of information stored as a textured mesh. Special care should be taken to produce an accurate model of the road surface and its immediate surroundings, such as kerbs. It is recommended that the 3D model data be generated in the interoperable and widely used OBJ format. The terrain should be separated into geometric objects, such as road, pavement/footway or verge, divided into smaller manageable sections if necessary. Special care should be taken to ensure that the road surface is as faithfully represented as possible and is free from any artefacts which do not provide a true representation of a road surface.
- **Object layers**, which provide information about all features existing within the road environment, such as buildings, kerbs, street furniture, trees and other vegetation, road markings. These should be represented in the formats outlined below:
 - As vector data: attributed and heighted points, lines or polygons where elevation information matches the terrain model. A minimum set of feature attributes should include height information, width, type and orientation, and can be expanded with additional information about the features they represent, such as material, age and maintenance. ESRI shapefile (SHP) is an interoperable and widely used vector data format for storing this type of information.
 - As 3D object models: for these, it is recommended that a library of objects be built, for example models of different types of traffic signs, lamp posts or trees, and that they be referenced within vector data. This allows 3D object models to be positioned accurately on the terrain layer for visualisation purposes. Please note that it is not possible to create a library for all features existing in a road environment. Some objects are unique, and there may be a need to model them separately. It is recommended to produce these models in OBJ format.
- The **road network** represents the logic of vehicle movement on the road and provides the foundation for network analysis, also forming a base for driving and traffic simulation. The road network can at a simple level be represented as vector data in ESRI shapefile format, with information about the direction of travel, number of lanes, lane width, lane use and connections through junctions.

Whilst ESRI shapefile format can be used, an increasingly adopted format within simulation environments is OpenDRIVE, an open XML format which provides a common

base for describing track-based road networks. The data stored in an OpenDRIVE file describes, in an analytical way, the geometry of roads as well as features along the roads that influence the logics, such as lanes, signs or signals. The format is currently being transferred to ASAM (the Association for Standardization of Automation and Measuring Systems) to become a public standard; it is therefore recommended that the development of this format is monitored.

In summary, there is a range of geospatial data formats available to the market, each offering specific features and benefits. The OS MasterMap family of products and OSM (XML) offer a good maintenance base on which to build additional map layers, with the objective of ensuring interoperability of high-definition content. The adoption of four further key formats is recommended, as summarised in Table 4.

Table 4 – Baseline recommended data formats

Format	Purpose
LAS 1.2 or LAZ (compressed form of LAS)	Point cloud data capture for identification, extraction and modelling of terrain and key features
OBJ	Good for representing the terrain and 3D objects such as buildings
OpenDRIVE	Good for describing track-based road networks
ESRI shapefile	A portable format good at representing a wide range of specific key features and their attributes

Zenic should collect feedback on this reduced set of file formats from organisations including testbeds, simulation companies and manufacturers, to establish any known limitations or challenges that they may present. Results of this consultation should be used to inform standards for self-driving vehicles.

Zenic agrees that early alignment on common data formats that can be used across industry to develop, validate and operate self-driving vehicles is a critical task. Zenic is already developing interventions to align data formats across Testbed UK and will use this report as a basis to further develop definition of these projects. Zenic will conduct a consultation on the data formats listed in this report to gather feedback from industry and other stakeholders.

Data quality and resolution

Accurate data is evidentially vital for safe operation of self-driving vehicles. Current OS data has been delivered to defined specifications to suit current government, industry and consumer needs for more than 220 years. However, emerging markets, including self-driving vehicles, are demanding richer high-resolution data. OS is currently in discussion with the UK government's Geospatial Commission about its long-term service agreement for the public sector. **Once requirements for data formats have been validated across industry, Zenic should facilitate discussions between emerging self-driving technology developers, Testbed UK and the Geospatial Commission to ensure that requirements for self-driving vehicles have been adequately considered and represented.**

Zenic, with support of OS, to engage with Geospatial commission on requirements for self-driving vehicles are considered. This activity will be informed by collation of feedback from industry on the findings of this report.

Terminology

Today this market is making use of differing definitions to describe geospatial features and their associated attributes, and as the market grows and additional players appear, it is important that the terminology used converges and becomes consistent, avoiding any ambiguity between manufacturers, simulation companies, test track operators, navigation and geospatial organisations, and government. **Common standards for terminology should be developed across the connected and self-driving vehicles sector and fed into the BSI Standards Programme to ensure that common terminology is documented, maintained and adopted by the industry at large.**

Zenic is participating in the BSI CAV Standards programme board which will allow it to have visibility of and influence in promoting common terminology. We will seek to both collect input from our testbed partners and also to roll out standardised terminology across our ecosystem.

Minimum safe requirements and standards

Simulation companies and manufacturers are currently developing control systems that will react to preprogrammed or sensed data, and safely manoeuvre a vehicle on the highway. As competition in this market is fierce, manufacturers may be reluctant to share data and methods, but ubiquitous, reliable and safe self-driving vehicle operation will nevertheless be dependent on conformity to a minimum set of requirements and associated standards. The risk of compromise of data and the communication of that data will be a major challenge, and the cybersecurity aspects will need particular focus – not only to ensure safety of operation but also to address the risk of compromising individuals and corporate information, intellectual property and national data. **These requirements should be impartially captured, working alongside standards bodies such as BSI, security specialists and government, to ensure consistency, security and compliance.**

Zenic already sits on the steering group for the BSI CAV standards programme and is actively involved in defining Cybersecurity investments and interventions with CCAV and IUK.

We believe that the best way we can ensure delivery of safety critical data standards is by co-ordinating requirements from, and creating interventions across our testbed ecosystem that will ensure approaches to standardisation can be empirically tested.

Government data and Traffic Regulation Orders

Zenic was instigated through CCAV, and it is important that it maintains its influence in, for example, the current DfT initiatives exploring the alignment of local authority data in a spatial context. Much of this is focused on the current state of play, and while self-driving vehicles are recognised as a key transport mode for the future of mobility, the specific future needs and standards to be adopted by these vehicles will need to be identified early and included in any spatial data requirements that local authorities will be requested to comply with.

Work by DfT is already underway to explore how local authority data can be improved across the country to ensure consistent methods for defining this data (in particular TROs, road marking and signage) and making sure that it is available in common formats. **This is a non-trivial piece of work but should be developed with consideration of self-driving**

vehicles from the outset. Testbed UK, funded by CCAV and co-ordinated by Zenzic, provides an ideal environment in which to test and develop digital traffic regulation orders in tandem with self-driving technology.

Zenzic is firmly of the belief that digitised highways codes and TROs will play an important part in the effective and safe operation of self-driving vehicles. Furthermore, there are significant near-term benefits to accurate digital TROs being available for 'connected' services including dynamic advice on current speed limits and parking services.

Zenzic believes that the best way to deliver these benefits is for Testbed UK to be at the centre of early roll out of digital TROs. This would ensure that DfT and industry organisations would have a focal point for testing implementation and additional services that might be beneficial for connected or self-driving vehicles.

Data hosting

Data is available from a variety of sources, notably OS and local authorities, who between them can provide a rich suite of data relating to public roads and highways. Other sources of data from private companies, transport operators, the Met Office and others can be hard to access and match reliably. A step to improving this will be to identify a neutrally-hosted geospatial environment (ideally not with a commercial focus) that can make use of federated data from a variety of sources, essentially providing the aforementioned trusted one-stop shop that will facilitate interoperability. **Neutral hosting of data should be explored with relevant bodies, including the Geospatial Commission and DfT, to consider available options. Zenzic must also ensure that the current work programmes creating data hosts are governed and aligned, and that lessons learned are shared frequently.**

Neutral hosting is a foundational element of delivering the data needed accelerate delivery of self-driving vehicles. However, Zenzic also believes there is a significant role for commercially focussed business models to play in unlocking the sharing of the highest quality data, which is often collected at significant cost to businesses. Zenzic is facilitating discussions between both for profit and not for profit organisations to encourage testing of data sharing modes where all parties understand the value gained up front.

Annex 1 – Workshop output

To support a fully self-driving vehicle life cycle, what geospatial data challenges must be overcome?

For each of these priority challenges, what key datasets, features, categories of data are required to support the full self-driving vehicle life cycle?

Voting key:

T = Test track
 S = Simulation
 B = BSI
 M = Met Office
 U = User

Groupings	Comment	Votes				
		T	S	B	M	U
Updates and ownership	<ul style="list-style-type: none"> Held by a trusted impartial body Ownership of centralised UK GSD Ownership of version control Ownership 	6	2	1	-	-
	<ul style="list-style-type: none"> Regularly updated Tracking changes in (near) real-time Version control Change in the environment Dates Accurate reflection of data throughout the life cycle Keeping location data up to date Regular updates Ensuring currency / real-time updates or regular How is it collected or kept up to date? 					
Financial & standards for resolution	<ul style="list-style-type: none"> Who hosts the data and commercial model? Data processing – labour intensive Cost of sets Commercial access and models Commercial models to allow data collection access Accessible software licences Common resolution to look; paid to see more Needs minimum resolution standard 	7	2	-	-	1
Data availability	<ul style="list-style-type: none"> Availability – surveys on demand currently 	1	-	-	-	-
Governance	<ul style="list-style-type: none"> Authority and liability 	2	1	-	-	1

Groupings	Comment	Votes				
		T	S	B	M	U
Access & standards	<ul style="list-style-type: none"> Ease of use access to data Bandwidth – transmission wirelessly Communications protocol? Access speed 	2	2	1	-	-
	<ul style="list-style-type: none"> Shared standards, formats and data types Standardisation of geospatial data for interoperability (between testbeds & life cycle) 					
	<ul style="list-style-type: none"> Before, during and after testing preparation and execution How are digital models used for simulation – standards / approach 					
Accuracy, security, detail	<ul style="list-style-type: none"> Critical National Infrastructure sensitivities 	5	3	1	-	1
	<ul style="list-style-type: none"> Resolution requirements for different Operational Design Domains (ODDs) Resolution and accuracy Data integrity accuracy Accuracy Accuracy of street furniture accurately positioned Is data accurate and reliable 					
	<ul style="list-style-type: none"> Develop a consistent shared understanding of the location Sophisticated geospatial operation versus secure data exchange Data is sufficiently detailed (where needed) such as surface of buildings, height of pavements 					
Requirements	<ul style="list-style-type: none"> Multiscale to range of simulation requirements What do test companies / OEMs want and when in the TRL life cycle? Real and virtual world specifications? Use case requirements on GSD Does full autonomy rely on 'maps'? 	8	1	-	-	-
Data interoperability	<ul style="list-style-type: none"> Data coherence, correlation Synthesis with other data Common understanding of geospatial data and assets. i.e. things mean the same thing Unique IDs for assets Common data approach to promote interoperability 	1	4	1	-	1

Groupings	Comment	Votes				
		T	S	B	M	U
Real-time	<ul style="list-style-type: none"> Feedback mechanism as to how to use vehicle data to improved geospatial assets Data quality, reliability, access, real-time 	1				
Validation	<ul style="list-style-type: none"> Validation 	1	1			
Rules	<ul style="list-style-type: none"> Digitising the rules that apply to location to pass to vehicle 	1				
Platform	<ul style="list-style-type: none"> Run on a common platform but not exclusively 					

Groupings	Comment	Votes				
		T	S	B	M	U
Zenzic	<ul style="list-style-type: none"> Need to know the minimum shareable knowledge access testbeds Need to ensure useful across life cycle 	2	-	-	-	-
5G	<ul style="list-style-type: none"> 5G connection speed 	-	-	-	-	-
Buildings	<ul style="list-style-type: none"> Public buildings with high volumes of people Position of buildings Adjacent buildings materials, facades 	-	1	-	1	1
Road rules	<ul style="list-style-type: none"> LHD / RHD direction of travel Turn arrows on junctions, banned turns Cycle lanes One-way streets Road rules – one-way, accessibility, priority Toll charge zones Speed limits, bus lanes, parking, speed cameras Speed limits and where they apply 	2	2	1	-	-
Road networks	<ul style="list-style-type: none"> Detailed connected and self-driving vehicle-ready networks Road network vectors 	2	-	-	-	-
Road markings and associated features	<ul style="list-style-type: none"> Kerb width, road edge to building Road types by standard classifications Location of signal junction Road layout data Position and geometry of roads Road markings and signage Position / type of pedestrian crossings Traffic light position Route navigation, sensor for local navigation Road furniture position and type Roadside / street furniture – pipelines Public transport stops and network Public transport timetables Road crossings (school) low speed zones Road signals, pedestrian crossings, 'stop', 'give way' 	19	2	1	1	-

	<ul style="list-style-type: none"> • Roadside unit location and position – V2I communications 					
Road surfaces and materials	<ul style="list-style-type: none"> • Where required, road surfaces, building fascias • Material classification • Surface reflectivity at different wavelengths • Surface friction • Road surface / terrain model • Slope / contours • Road condition • Elevation / gradients 	3	3	1	1	1

Groupings	Comment	Votes				
		T	S	B	M	U
File formats and data sharing	<ul style="list-style-type: none"> • Vehicle dynamics Simulator 1 mm lidar • Driven in loop Simulator 10 mm lidar • Simulator – Shared standards across testbeds • Common file formats • Common software • Agreed levels of certification for test and development • Interoperable simulation models / APIs. • Highway pilot sensor only 	11	1	1	1	-
Restrictions	<ul style="list-style-type: none"> • Tunnels and bridges • Tunnels an urban canyons – GPS restrictions 	-	-	-	-	-
Dynamic features (on collection)	<ul style="list-style-type: none"> • Incidents • Weather • Roadworks • Traffic speed / flow 	-	1	-	-	-
Simulation	<ul style="list-style-type: none"> • Cost per mile in digital model 	-	-	-	-	-
Feature status	<ul style="list-style-type: none"> • Temporary, scheduled, permanent 	-	-	-	-	-
Real-world customer / user requirements	<ul style="list-style-type: none"> • What will 'Tesla' do but, what do they want? 	-	-	-	-	-
Natural environment	<ul style="list-style-type: none"> • Foliage • Seasonal time series data 	-	-	-	-	-
Asset locations	<ul style="list-style-type: none"> • Proposed infrastructure asset locations • Telecom infrastructure asset locations topology 					

Other comments noted during this exercise:

Baseline strategy going forward?
 Effort can be very high
 Consistency across test beds
 How does TRL impact?
 Different collection methods
 New customer education

What are the key attributes of the data that must be available to support the full self-driving vehicle life cycle?

The workshop was not a practical forum to identify the attributes of every identified feature discovered during the session. This would be a valuable activity but will take many hours to complete in a collaborative manner. The participants therefore focused on two areas of their choice; road signs and temporary structures. It should be noted that neither of the lists below are deemed to be exhaustive.

Feature attribution		
Road Signs	Temporary structures	
Position in X,Y,Z	Portakabins	Bollard lane markings
Type, shape, dimensions	Traffic lights	New road layouts
Height	Roadworks	Temporary textured markings
Material	Contraflows	Rising bollards
Instruction	Diversions	Merge in turn signs
Orientation	Sink holes	Rollout of various crossing types
Illumination	Speed bumps	Rollout of tunnels
Accuracy (relative to location)	Flooding	Pedestrian walkways
Age	Temporary signs	Rollout of bridges / gantries
Maintenance	Temporary road markings	Bicycle racks
Fixed or matrix	Raised/ lowered iron works	
Temporary or permanent	Ramps	
Obscured?	Bridging plates	
Language	Temporary access	
Double signs	Skips and bins	
Implicit signs	Dynamic hard shoulders / lanes	
Sensor enabled & communicable	Road blocks / collisions	
Brand / manufacturer	Temporary lighting	
Link to another object	Temporary road surfaces	
	Temporary ANPR (automatic number plate recognition)	
	Temporary roundabouts	

Notes:

1. Many temporary structures apply equally to the simulation and track environment, but this may not be true in all cases.
2. TSRGD (The Traffic Signs Regulations and General Directions 2016) and other specifications should be utilised.

Annex 2 – Feedback and comment

- 1.** Need to explore lessons from other domains such as defence.
- 2.** Need to understand specific requirement for different use cases.
- 3.** Input from regulators is required.
- 4.** Can we consider the biggest geospatial feature that need to be addressed first – what is it?
- 5.** Provision of example use cases to stimulate discussion.
- 6.** Data format options – what are they?
- 7.** To what extent can A.I. mitigate geospatial data gaps?
- 8.** End user requirements.
- 9.** Standards for digital models.
- 10.** Need to have more users in the room – unable to answer about what they need without them.
- 11.** Learn more on potential tool share: OSNET / GIS etc.
- 12.** Would be good to get into data formats.
- 13.** What minimum data requirements will Zenzi expect of the testbeds? This needs to be clarified.

Annex 3 – Abbreviations and acronyms

Abbreviation or acronym	Full description or meaning
ANPR	Automatic number plate recognition
ASAM	Association for Standardization of Automation and Measuring Systems
BSI	British Standards Institution
CCAV	Centre for Connected and Autonomous Vehicles
DfT	Department for Transport
IUK	Innovate UK
JPEG	JPG is a file extension for a lossy graphics file. The JPEG file extension is used interchangeably with JPG. JPEG stands for Joint Photographic Experts Group who created the standard. JPG files have two subformats, JPG/Exif (often used in digital cameras and photographic equipment), and JPG/JFIF (often used on the Web).
LAS 1.2	LAS files are binary files packed according to several specifications to represent lidar data
LAZ	LAZ is a compressed light detection and ranging (lidar) data format that is often used to transfer large amounts of lidar data.
Lidar	Lidar, which stands for light detection and ranging (originally a portmanteau of 'light' and 'radar'), is a remote sensing method that uses light in the form of a pulsed laser to measure ranges.
OBJ	An OBJ file is a standard 3D image format that can be exported and opened by various 3D image-editing programs. It contains a three-dimensional object including 3D coordinates, texture maps, polygonal faces and other object information.
OSM	OpenStreetMap
PAS	Publicly Available Specification produced by BSI
Point cloud	A point cloud is a collection of data points defined by a given coordinates system. In a 3D coordinates system, for example, a point cloud may define the shape of some real or created physical system.
PSGA	Public Sector Geospatial Agreement
SHP	Is a file extension for a Shapefile shape format used in geographical information systems (GIS) software. SHP is short for "shape".
TN-ITS	Transport Network Intelligent Transport Systems
TOID	topographic identifier
TRO	traffic regulation order
UK plc	The term used to describe the United Kingdom's commercial community considered as a single organisation, or the commercial interests of the United Kingdom considered as a whole, each of which form part of the wider economy of the United Kingdom
V2C	vehicle-to-cloud
V2I	vehicle-to-infrastructure; referring to the wireless connections between these two entities
V2V	vehicle-to-vehicle; referring to the wireless connections between these two entities
V2X	vehicle-to-everything
XML	Extensible Markup Language (XML) is used to describe data. The XML standard is a flexible way to create information formats and electronically share structured data via the public Internet, as well as via corporate networks.

Annex 4 – Data formats explored

The table below lists the different data formats identified through this study which relate to the creation of real-world digital twin of road environment.

Formats	Description
3DS	Autodesk 3D geometry file format.
ASCII grid	ESRI grid is a raster GIS file format.
AutoCAD DXF	Drawing Interchange Format or Drawing Exchange Format – Autodesk CAD format. Used for enabling data interoperability.
CSV	A comma-separated values.
DAE	COLLADA 3D graphic files.
E57	A format for storing point clouds, images, and metadata produced by 3D imaging systems.
ESRI shapefile	A vector data format for storing the location, shape, and geographic attributes of features
FBX	Autodesk owned format designed to provide interoperability between digital content creation applications.
GeoTIFF	Open file format and widely used standard based on the TIFF format. It is used as an interchange format for georeferenced raster imagery.
GML	An XML grammar for expressing geographical features.
LAS	The most common format for exchanging points cloud data
LAZ	Compressed lidar file format.
MicroStation DGN	DGN is a file extension for a Computer Aided Design (CAD) drawing.. DGN files are usually used for architectural and engineering designs.
OBJ	Open geometry definition file format
OpenCRG	Open file format for the detailed description, creation and evaluation of road surfaces.
OpenDRIVE	Open file format specification to describe a road network's logic
OpenSCENARIO	Open file format for the description of dynamic contents in driving simulation applications
PLY	Describes an object as a collection of vertices, faces and other elements, along with properties such as colour and normal direction.
STL	3D surface geometry file format
XML	Extensible Markup Language data file. A text-based database - it uses custom tags to define objects and the data within each object
JSON	Standard data interchange format that stores simple data structures and objects
XYZ	A chemical file format

References

ABI (Association of British Insurers), 2019. *Automated driving – Insurance industry’s role in automated driving*. [Online]

Available at: <https://www.abi.org.uk/products-and-issues/topics-and-issues/driverless-cars/>

Anon., n.d. [Online]

Available at: <https://www.here.com/>

APM, 2019. *Agile methods*. [Online]

Available at: <https://www.apm.org.uk/resources/find-a-resource/agile-project-management/agile-methods/>

BSI & Transport Catapult, 2017. *Connected and autonomous vehicles*. [Online]

Available at: <https://www.bsigroup.com/PageFiles/505209/CAV-standards-strategy-summary-report.pdf>

[Accessed 2019].

Cityverve, 2019. [Online]

Available at: <https://cityverve.org.uk/>

Cockburn, H., 2018. ‘*Shocking’ rise in number of potholes as Britain’s roads get worse, says RAC*. [Online]

Available at: <https://www.independent.co.uk/news/uk/home-news/potholes-number-increase-rac-foi-uk-east-sussex-london-yorkshire-a8701811.html>

Department for Transport, 2007. *Manual for Streets*. [Online]

Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/341513/pdfmanforstreets.pdf

[Accessed May 2019].

Department for Transport, 2019. *The Highway Code*. [Online]

Available at: <https://www.gov.uk/guidance/the-highway-code>

[Accessed May 2019].

DfT, 2019. *Technical Note: Road Condition and Maintenance data*. [Online]

Available at:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775107/road-conditions-technote.pdf

[Accessed 2019].

Fleming, N., 2017. *Connected and autonomous vehicles: Getting the standards right*. [Online]

Available at: <https://www.bsigroup.com/globalassets/localfiles/en-gb/events/standards-forum--november-2018/connected-and-autonomous-vehicles.pdf>

[Accessed May 2019].

General Motors, 2018. *2018 self driving safety report*. [Online]

Available at:

<https://www.gm.com/content/dam/company/docs/us/en/gmcom/gmsafetyreport.pdf>

HERE technologies, 2019. *HERE Technologies*. [Online]

Available at: <https://www.here.com/about-us>

Intelligent Transport, 2018. *China issues national standards for the testing of autonomous vehicles*. [Online]

Available at: <https://www.intelligenttransport.com/transport-news/70487/autonomous-technology-regulations-china/>

Makwana, 2018. *What is a pothole?*. [Online]
Available at: https://www.racfoundation.org/wp-content/uploads/What_is_a_pothole_final_Makwana_December_2018.pdf
[Accessed 2019].

Marr, A., 2013. *7 Definitions of Quality*. [Online]
Available at: <https://business.simplicable.com/business/new/7-definitions-of-quality>
[Accessed 2019].

NIC (National Infrastructure Commission), 2018. *Five shortlisted to develop 'Roads for the Future' and driverless cars*. [Online]
Available at: <https://www.nic.org.uk/our-work/roads-for-the-future/>

Norman MP & Harrington MP, 2019. *Government moves forward on advanced trials for self-driving vehicles*. [Online]
Available at: <https://www.gov.uk/government/news/government-moves-forward-on-advanced-trials-for-self-driving-vehicles>

North Highland, 2018. *Local transport data discovery*. [Online]
Available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730787/local-transport-data-summary.pdf
[Accessed 2019].

Ordnance Survey; 2018. *5G Planning – geospatial*. [Online]
Available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/684420/OS_Final_report_5g-planning-geospatial-considerations.pdf
[Accessed 2019].

Ordnance Survey;, 2019. *AddressBase products*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-products.html>

Ordnance Survey;, 2019. *OS MasterMap*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/business-and-government/products/mastermap-products.html>

Ordnance Survey;, 2019. *OS MasterMap Building Height Attribute*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/business-and-government/products/os-mastermap-building-height-attribute.html>

Ordnance Survey;, 2019. *OS MasterMap Highways Network products*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/business-and-government/products/os-mastermap-highways-network-products.html>

Ordnance Survey;, 2019. *OS MasterMap Imagery Layer*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/business-and-government/products/imagery-layer.html>

Ordnance Survey;, 2019. *OS MasterMap Topography Layer*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/business-and-government/products/topography-layer.html>

Ordnance Survey, 2019. *OS Terrain 5*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/business-and-government/products/os-terrain-5.html>

Ordnance Survey, 2001. *OS MasterMap real-world object catalogue*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/docs/legends/os-mastermap-real-world-object-catalogue.pdf>
[Accessed May 2019].

Ordnance Survey, 2007. *OS Mastermap Topography Layer Feature Catalogue*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/docs/legends/os-mastermap-topography-layer-feature-catalogue.zip>
[Accessed May 2019].

Ordnance Survey, 2018. [Online]
Available at: <https://www.gov.uk/government/publications/connected-and-autonomous-vehicle-research-and-development-projects>

Ordnance Survey, 2018. *Ordnance Survey underpins UK driverless car testing*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/about/news/2018/os-underpins-uk-driverless-car-testing.html>

Ordnance Survey, 2018. *Ordnance Survey underpins UK driverless car testing*. [Online]
Available at: <https://www.ordnancesurvey.co.uk/about/news/2018/os-underpins-uk-driverless-car-testing.html>

Rand Corporation, 2016. *How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* [Online]
Available at: https://www.rand.org/content/dam/rand/pubs/research_reports/RR1400/RR1478/RAND_RR1478.pdf

Rogerson, J. & Donegan, M., 2018. *What is 5G? | A comprehensive guide for all*. [Online]
Available at: <https://5g.co.uk/guides/what-is-5g/>

The Highways Agency, 1989. *DESIGN MANUAL FOR ROADS AND BRIDGES TA57/87*. [Online]
Available at:
<http://www.standardsforhighways.co.uk/ha/standards/dmrb/vol6/section3/ta5787.pdf>
[Accessed 8 May 2019].

TN-ITS, 2019. *A reference interface for European road data updates*. [Online]
Available at: http://go.tn-its.eu/wp-content/uploads/sites/4/2018/04/ERT-006-18-TN-ITS-GO-brochure_v6.pdf
[Accessed 2019].

Transport Systems Catapult, 2017. *Market Forecast for connected autonomous vehicles*. [Online]
Available at:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/642813/15780_TSC_Market_Forecast_for_CAV_Report_FINAL.pdf

UKi Media & Events, 2019. *Autonomous vehicle safety regulations world congress*. [Online]
Available at: <https://www.ukimediaevents.com/german/conferences.php>

Various, 2019. *OpenStreetMap*. [Online]
Available at: <https://www.openstreetmap.org/about>

Various & Wikipedia, 2019. *Digital twin – Wikipedia*. [Online]
Available at: https://en.wikipedia.org/wiki/Digital_twin
[Accessed 2019

