BENEFITS OF STANDARDISATION OF BUS
TRANSIT AUTOMATIC VEHICLE LOCATION DATA
OUTPUTS TO THE ACADEMIC COMMUNITY

Steve Robinson
London Buses
Transport for London
197 Blackfriars Road, Southwark
London, SE1 8NJ
United Kingdom
e-mail: stephen.robinson@tfl.gov.uk

NB. Since the original submission I have changed jobs. My new contact address is:
e-mail: stephen.robinson@smart.mit.edu

Text = 3200 words
Tables and Figures = 6 Tables and 2 figures – i.e. 2000 words
Total Words = 5200 words

Submission date = 15th November 2012

Submitted for presentation at the 92nd Transportation Research Board Annual Meeting, 13th to 17th January 2013, Washington, D.C. and publication in the Transportation Research Record
Automatic Vehicle Location (AVL) systems offer transit agencies and the academic community vast quantities of data on the operation of bus transit networks. However turning this raw data into useful metrics that are of interest can be quite cumbersome and require a significant amount of processing of AVL log files. There would be many benefits to both transit agencies and academia if such metrics were output in a standardised format by the AVL system.

This paper summarises existing data exchange standards used in the bus transit industry and explains the benefits of these. It then describes a recent project undertaken by London Buses to analyse dwell time data, and highlights the difficulties and costs involved in collecting the basic data required for this analysis. It proposes that more data standards should be defined for data output by AVL systems. If these standards were then implemented by AVL vendors this would mean that transit agencies and academics could get easy access to useful information which otherwise would require much time and cost in obtaining.

1. INTRODUCTION

Automatic Vehicle Location (AVL) systems offer transit agencies and the academic community vast quantities of data on the operation of bus transit networks. For example, many AVL systems provide second-by-second logging of the location of the bus, along with the recording of certain events such as the door opening and closing. However turning this raw data into useful metrics that are of interest can be quite cumbersome and require a significant amount of processing of AVL log files. There would be many benefits to both transit agencies and academia if such metrics were output in a standardised format by the AVL system.

This paper has been written to encourage transit agencies and the academic community to define certain metrics that AVL systems should output. It promotes the creation of clearly defined and published standards. Transit agencies and operators could then refer to these standards when specifying their requirements for AVL systems.

The paper is organised as follows. Section 2 provides a summary of existing standards in bus AVL systems. Section 3 summarises the benefits of standardisation. Section 4 provides a case study on an area where transit agencies and the academic community could benefit from having standards defined – the study of bus stop dwell time. Section 6 then provides conclusions.

The work was inspired by a bus speeds project that is currently being undertaken at London Buses. As part of this project it became evident that a large amount of work was being undertaken to obtain some quite basic information that was required by Transport for London. Since other transit agencies and the academic community may require the same information, it became obvious that all AVL systems should ideally provide this data as a standard output.

It should be noted that the opinions stated in this document reflect those of the author and not necessarily of London Buses.
2. STANDARDS USED IN THE BUS TRANSIT AVL INDUSTRY

In the past decade there has been a proliferation in the usage of standards in the bus transit AVL industry. A summary of several standards in common usage or proposed is given in Table 1. The area that the standard covers is one or many of the following types:

- Network: information about the bus route or bus stops.
- Schedule: information about the bus schedule
- Real Time: real time information about the bus service

<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>Type</th>
<th>Countries Used</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TransModel</td>
<td>Network &amp; Schedule</td>
<td>Europe</td>
<td>This standard provides an abstract model of concepts and data structures to allow the modelling of public transport in Europe. Version 5.1 is European standard, EN12896 [1]. This replaces an earlier version of the standard [2].</td>
</tr>
<tr>
<td>2</td>
<td>NaPTAN</td>
<td>Network</td>
<td>UK</td>
<td>The National Public Transport Access Node (NaPTAN) standard provides a way to model all locations that people can access public transport. This includes bus stops, train platforms, airport gates, etc. Each location has been given a unique identifier [3]. It is based on XML.</td>
</tr>
<tr>
<td>3</td>
<td>TransXChange</td>
<td>Network, Schedule</td>
<td>UK</td>
<td>This is a UK standard to exchange bus route and timetable information. Outside London all bus operators must register their bus services using this standard [4]. It is based on XML.</td>
</tr>
<tr>
<td>4</td>
<td>SIRI</td>
<td>Real Time</td>
<td>Europe</td>
<td>This is a European standard for storing and exchanging real time information about bus services [5]. It is based on XML. It should be noted that there is an additional implementation of SIRI in Germany which is based on the VDV453 and VDV454 standards (see below). This means that software built on the UK standard of SIRI may not necessarily be able to exchange data with software built on the German version of SIRI.</td>
</tr>
<tr>
<td>5</td>
<td>VDV452</td>
<td>Network &amp; Schedule</td>
<td>Germany</td>
<td>This is a German standard to define network and schedule data [6]. This is</td>
</tr>
</tbody>
</table>
Table 1 – Standards used by the bus transit AVL industry

<table>
<thead>
<tr>
<th>No.</th>
<th>Standard</th>
<th>Category</th>
<th>Country</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>VDV453</td>
<td>Real Time</td>
<td>Germany</td>
<td>This standard provides real time information on aspects of the bus operations. In particular it provides data exchange functionality for on connection protection, dynamic passenger information, and the exchange of ad-hoc messages to bus stop signs [8].</td>
</tr>
<tr>
<td>7</td>
<td>VDV 454</td>
<td>Real Time</td>
<td>Germany</td>
<td>This standard provides real time information on aspects of bus operations. In particular it provides data exchange functionality of scheduled and real time bus arrivals [9].</td>
</tr>
<tr>
<td>8</td>
<td>TRIDENT</td>
<td>Network &amp; Schedule</td>
<td>France</td>
<td>This is an XML based standard to exchange inter-modal traffic and travel information. It includes both public transport and road traffic information [10].</td>
</tr>
<tr>
<td>9</td>
<td>GTFS</td>
<td>Network &amp; Schedule</td>
<td>USA</td>
<td>The Google Transit Feed Specification is a standard to exchange public transit network and schedule data. A total of 13 entities are defined in the standard. [11] The standard is based on CSV files.</td>
</tr>
<tr>
<td>10</td>
<td>GTFS – real time</td>
<td>Real Time</td>
<td>USA</td>
<td>GTFS-realtime is an extension to GTFS that provides real time updates to the schedule information provided by the GTFS feed [12].</td>
</tr>
<tr>
<td>11</td>
<td>RTIG: Digital Air</td>
<td>Real Time</td>
<td>UK</td>
<td>This standard defines a set of messages that can be communicated between bus-mounted devices and office-based systems, using a digital open protocol. These messages include bus location, driver logon, etc. [13].</td>
</tr>
<tr>
<td>12</td>
<td>JourneyWeb</td>
<td>Network &amp; Schedule</td>
<td>UK</td>
<td>This standard defines a way for Journey Planning servers to communicate with each other. This might be required if planning a journey that covers several transit regions [14].</td>
</tr>
<tr>
<td>13</td>
<td>IFOPT</td>
<td>Network</td>
<td>UK</td>
<td>IFOPT (Identification of Fixed Objects in Public Transport) is a proposed standard to provide a reference data model for describing the main fixed objects required for public access to public transport [15]. It is still in development.</td>
</tr>
</tbody>
</table>
The standards above are widely used in Europe. In particular TransXChange and VDV452 are used in many bus schedule and bus AVL systems. However no reference in the literature could be found of academics making use of these standards.

It is notable that all the above standards relate to data inputs to journey planner or AVL systems. No standard could found for outputting AVL derived data that can be used for operational analysis or academic research. This paper helps to address this point.

### 3. BENEFITS OF STANDARDISATION

From the literature review it is seen that the ITS industry is keen to promote standards. There are many advantages to standardisation. These are summarised in Table 2.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial cost</td>
<td>Building a system can be very expensive, involving many phases such as: requirements gathering, design, writing technical documentation, build, testing, and implementation. Buying a standard product means that many of these system building costs are shared amongst multiple parties or have already been paid for by an earlier customer.</td>
</tr>
<tr>
<td>Requirements</td>
<td>Standards are usually defined by a committee of people with differing initial requirements. This increases the likelihood that important requirements are not omitted before the product is built. It also increases the chances that the product is future proofed.</td>
</tr>
<tr>
<td>Testing</td>
<td>Keeping to widely used standards increases the likelihood that the end product has already been widely tested, and also proven to work in the field.</td>
</tr>
<tr>
<td>Time</td>
<td>In many cases ITS vendors will have already incorporated various standards into their core product. This helps to minimise the amount of bespoke customisation required for the individual implementation. This significantly reduces the time taken to install a new system.</td>
</tr>
</tbody>
</table>

**Table 2 – Benefits of standardisation**

However there are some negative points in having standards. These are summarised in Table 3.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does not meet requirements</td>
<td>A standard might not meet all the requirements that a user has. This may rule out the use of the standard or require the standard to be changed.</td>
</tr>
<tr>
<td></td>
<td>As an example the UK NaPTAN standard originally only allowed for $7^5 = 16,807$ bus stops in each area. Since London has over 19,000 bus stops this standard could not initially be used for London. This</td>
</tr>
</tbody>
</table>
required the NaPTAN standard to be updated in version 2.4 to allow an area to have up to 99,999 stops.

The use of standards may hinder the introduction of novel or better solutions as people will be encouraged to agree with the current standard. The academic community should be especially aware of this and ensure that if they choose to use standard data feeds they check whether there would be a more appropriate way of modelling the system.

Standards that have been ratified by an international standards agency often cost a significant amount of money to buy. For example, to buy the TransModel standard from the official British Standards agency currently costs approximately USD500. The SIRI and VDV453 and VDV454 are similarly expensive to procure. It is not known why government agencies charge so much to provide these standards to industry and academia.

<table>
<thead>
<tr>
<th>Stifles creativity</th>
<th>Cost of documentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The use of standards may hinder the introduction of novel or better solutions as people will be encouraged to agree with the current standard. The academic community should be especially aware of this and ensure that if they choose to use standard data feeds they check whether there would be a more appropriate way of modelling the system.</td>
<td>Standards that have been ratified by an international standards agency often cost a significant amount of money to buy. For example, to buy the TransModel standard from the official British Standards agency currently costs approximately USD500. The SIRI and VDV453 and VDV454 are similarly expensive to procure. It is not known why government agencies charge so much to provide these standards to industry and academia.</td>
</tr>
</tbody>
</table>

Table 3 – Negative aspects of standardisation

4. CASE STUDY: BUS STOP DWELL TIME ANALYSIS

As an example of the benefits of standardisation of AVL data outputs to transit agencies and the academic community, an example is provided on the study of bus stop dwell time. Such a project was undertaken by Robinson [16] and acted as the inspiration for this paper. This case study will highlight how much time and effort can be saved by transit agencies and the academic community if certain data were to be provided as a standard output by AVL systems. Section 4.1 introduces the problem on bus stop dwell time. Section 4.2 identifies the raw data from the AVL log files needed to analyse bus stop dwell time. Section 4.3 summarises how the raw AVL data was converted into the data of section 4.2 and highlights what a time consuming task this is. Section 4.4 states the advantage if the AVL system were to provide the bus stop dwell time of section 4.2 as a standard output.

4.1 Characteristics of bus stop dwell time

Work by Robinson (16) identified 8 key features of a bus stop visit event. These are shown in Figure 1 and listed below. Figure 1 shows the speed-distance plot of a bus serving Acton Street bus stop on the line 45 in central London. Data was obtained from the log file of the iBus AVL system that is installed on all buses in London. It should be noted that the stop zone of a bus stop is defined form the point 50 metres upstream of the bus stop to 30 metres downstream of the bus stop.
Standardisation of Bus Transit AVL data

Steve Robinson

Figure 1 – Features of a bus stop visit

The features are defined below in typical chronological order:

- **F1 – peak speed before the stop**: This is the point immediately before the bus stop where the bus was travelling fastest.
- **F2 – stop entry**: This is the point where the bus enters the stop zone. This should be 50 metres before the bus stop. NB. since the peak speed might occur less than 50 metres before the stop then F2 might occur before F1.
- **F3 – zero speed start**: This is the point where the bus becomes stationary immediately before the bus first opens its doors.
- **F4 – door open**: This is the point where the bus firsts opens its doors in the stop zone.
- **F5 – door close**: This is the point where the bus last closes its doors in the stop zone.
- **F6 – zero speed end**: This is the point where bus begins to accelerate away from the stop immediately after the bus last closes its doors.
- **F7 – stop exit**: This is the point where the bus exits the stop zone. This should be 30 metres after the bus stop.
- **F8 – peak speed after the stop**: This is the point immediately after the bus stop where the bus was travelling fastest. NB. since the peak speed might occur less than 50 metres after the stop then F8 might occur before F7.

Each feature, $F_i$, has the following attributes:

- $D_{F_i}$: Distance of the bus from the bus stop – measured along the road.
- $T_{F_i}$: Time until bus is closest to the bus stop
- $V_{F_i}$: Speed of the bus in kph
Robinson (16) also defined three useful metrics that should be calculated for each bus stop visit:

- **Time that doors are open – i.e. Dwell Time:** This metric is useful for understanding the performance of London Buses in quickly getting passengers onto and off the bus. This metric is henceforth termed “Dwell Time”.

- **Time lost serving stop:** This metric is necessary to accurately measure drive time bus speeds. It could also be used to estimate the travel time costs and benefits of adding or removing bus stops to the network.

- **Time lost accelerating to cruise speed:** This metric can be used to identify bus stops that may need to be re-designed to allow the bus to return to the main traffic flow quicker.

These three metrics can allow transit agencies to better manage and understand the performance of the bus network.

### 4.2 Proposed data model

A data model has been proposed that can store the various features and metrics identified in the previous section. The data model defines the various entities that are required. Two entities have been defined which store bus dwell time specific data: “Stop_Visit_Event” and “Bus_Location”. The “Stop_Visit_Event” entity relates to other entities which are defined in the existing NaPTAN [3] and TransXChange [4] standards. These entities can be implemented in a relational database or as java objects. The data model below assumes they have been implemented as Java objects.

#### 4.2.1 Data schema

The data schema is comprised of several entities. These are described in detail below. The relationship between the entities is shown in Figure 2.

![Figure 2 – Data schema for storing bus stop dwell time data](image)
4.2.2 Bus_Location

This entity represents the location of the bus. It is typically produced once per second by the AVL system. Events recorded by the bus will also have a “Bus_Location” record.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>bl_index</td>
<td>Number</td>
<td>Unique identifier of the record</td>
</tr>
<tr>
<td>longitude</td>
<td>Number</td>
<td>Longitude of the bus – WGS84</td>
</tr>
<tr>
<td>latitude</td>
<td>Number</td>
<td>Latitude of the bus – WGS84</td>
</tr>
<tr>
<td>dist_from_stop</td>
<td>Number</td>
<td>This is the distance of the bus from the bus stop. This data is used to normalise the odo reading. Measured in metres</td>
</tr>
<tr>
<td>odo</td>
<td>Number</td>
<td>This is the reading of the odometer. It is normalised so that at the point the bus is closest to the bus stop the odo reading is 0. Measured in metres. Negative values imply the bus is upstream of the stop. Positive values imply the bus is downstream of the stop.</td>
</tr>
<tr>
<td>speed</td>
<td>Number</td>
<td>This is the speed of the bus in kph</td>
</tr>
<tr>
<td>heading</td>
<td>Number</td>
<td>This is the heading of the bus measured in degrees from 0 to 359.</td>
</tr>
<tr>
<td>time</td>
<td>Number</td>
<td>This is the time in seconds. It is normalised so that at the point the bus is closest to the bus stop the time is 0.</td>
</tr>
</tbody>
</table>

Table 4 – Data dictionary for entity: Bus_Location

4.2.3 Stop_Visit_Event

A Stop_Visit_Event record is created for each visit by a bus to a stop. The attributes of the Stop_Visit_Event are described below. This entity joins onto NaPTAN and TransXChange entities.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>StopIndex_NaPTAN</td>
<td>String</td>
<td>This is the unique identifier of a bus stop. It relates to the NaPTAN attribute StopPoint.AtcoCode. It allows the Stop_Visit_Event record to be joined to a NaPTAN.StopPoint record. All other attributes of the bus stop such as the location can be obtained via the StopPoint entity.</td>
</tr>
<tr>
<td>VehicleJourneyCode</td>
<td>String</td>
<td>This is the unique identifier of the trip. It relates to the TransXChange attribute VehicleJourney.VehicleJourneyCode. All other attributes of the trip, such as the line and stop sequence can be obtained via the VehicleJourney entity.</td>
</tr>
<tr>
<td>Name</td>
<td>Type</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>blStopEntry</td>
<td>Bus_ Location</td>
<td>Bus Location record at the point where the stop zone entry event occurs. NB. Odometer are Time records are normalised by those values recorded in “blClosestToStop”.</td>
</tr>
<tr>
<td>blStopExit</td>
<td>Bus_ Location</td>
<td>Bus Location record at the point where the stop zone exit event occurs. NB. Odometer are Time records are normalised by those values recorded in “blClosestToStop”.</td>
</tr>
<tr>
<td>blClosestToStop</td>
<td>Bus_ Location</td>
<td>Bus Location record at the point where the bus is closest to the bus stop.</td>
</tr>
<tr>
<td>blPeakSpeedBeforeStop</td>
<td>Bus_ Location</td>
<td>Bus Location record at the point immediately before the bus stop where the speed was highest</td>
</tr>
<tr>
<td>blPeakSpeedAfterStop</td>
<td>Bus_ Location</td>
<td>Bus Location record at the point immediately after the bus stop where the speed was highest</td>
</tr>
<tr>
<td>vectorBlDoorOpen</td>
<td>Vector &lt; Bus_ Location &gt;</td>
<td>Vector of Bus_Location records. One record for each door opening event within the stop zone</td>
</tr>
<tr>
<td>vectorBlDoorClose</td>
<td>Vector &lt; Bus_ Location &gt;</td>
<td>Vector of Bus_Location records. One record for each door closing event within the stop zone</td>
</tr>
<tr>
<td>vectorBusLocation</td>
<td>Vector &lt; Bus_ Location &gt;</td>
<td>Vector of Bus_Location records. One record for each metre from 100m before the stop zone entry to 100m after the stop zone exit NB. Odometer and time records are normalised by those values recorded in “blClosestToStop”.</td>
</tr>
<tr>
<td>noDoorOpens</td>
<td>Int</td>
<td>This is the number of times that the door was opened in the stop zone</td>
</tr>
<tr>
<td>durationStopZone</td>
<td>Int</td>
<td>This is the time that the bus was in the stop zone: Time_Stop_Exit - Time_Stop_Entry</td>
</tr>
<tr>
<td>durationDoorOpen</td>
<td>Int</td>
<td>This is the time that the doors were open when the bus was in the stop zone</td>
</tr>
<tr>
<td>stationaryTimeBeforeDoorOpen</td>
<td>Int</td>
<td>This is the time that the bus was stopped immediately before the bus first opened its doors</td>
</tr>
<tr>
<td>stationaryTimeAfterDoorClose</td>
<td>Int</td>
<td>This is the time that the bus was stopped immediately after the bus last closed its doors</td>
</tr>
<tr>
<td>dataQuality</td>
<td>Int</td>
<td>This is a flag indicating if there are any data quality issues with the data</td>
</tr>
</tbody>
</table>

Table 5 – Data dictionary for entity: Stop_Visit_Event
4.2.4 Other entities


4.2.5 Points to note

The following points should be noted:

- The data model describes how data would be stored in Java objects. If it is desired to store data in a relational database then the following changes would have to be incorporated:
  
  o Define a unique key for each record in each table. This can be easily achieved by having an auto-incrementing numerical field
  
  o All “Vector” fields that describe a 1 to many relationship between the Stop_Visit_Event record and the Bus_Location record would have to be resolved. This could be done by defining a new table which referenced the index of both the Stop_Visit_Event and the Bus_Location table allowing for a many-to-many relationship between the two tables.

- It might be desirable to change the data type of some fields to allow data to be stored at decimal precision.

4.3 Process of converting log file data to required data

It is estimated that between 100 and 200 man hours were required to design, write, and then test the parser that converted data from the iBus log file into the desired data format given in section 4.2. The pseudo-code below outlines the steps that were required and highlights the difficulties in undertaking a task which at first appears quite trivial. The pseudo-code assumes that the user is coding in Java.

4.3.1 Stage 1: Preparation of required look-up tables

Many AVL systems will use unique indexes when referring to entities such as the bus stop, route, or trip. Such indexes allow the unambiguous and concise referencing of data – a bus stop can be referred to using only 2 bytes in London (i.e. $2^{16} = 65,536$ stops supported).

However, such indexes are meaningless to humans. For example the bus stop “Acton Street” shown in Figure 1 has an iBus stop index of 9503. It is therefore necessary to build look-up tables to convert the stop-index into useful bus stop information. Such look-up tables add complexity and extra work. Data sources have to be identified, transfer mechanisms introduced, and then a parser has to be designed, written and tested.

4.3.2 Stage 2: Processing of each stop visit event

The second stage was to process all the log files written by the iBus AVL system on the bus. These log files contain second-by-second location information along with information about various events including door open times. A parser was written in Java to obtain the Stop
Visit Event data outlined in section 4.2 from these files. Some basic pseudo-code of how this was done is given below:

- **Step 1:** Download the log files from the bus and transfer to an appropriate directory
- **Step 2:** Read the zipped log file one by one into memory. For each file perform steps 3 to 14.
- **Step 3:** Identify all stop zone entries and exits. This had to take into account some quirks in the navigation logic which meant that a bus could enter or exit the same stop multiple times. Each matching stop zone entry and exit gives a stop visit event. For each stop visit event perform steps 4 to 13.
- **Step 4:** Identify all of the 8 features specified in section 4.1
- **Step 5:** Obtain additional information such as the stop location using the look-up tables generated in stage 1
- **Step 6:** Create a Bus_Location record for each metre between 100 metres before the stop zone entry to 100 metres after the stop zone entry.
- **Step 7:** For each Bus_Location calculate the direct distance between the bus and the stop. Identify the point that the bus is closest to the stop. This is BL\textsubscript{closest}.
- **Step 8:** Normalise the time and odometer of all Bus_Location records with the time and odometer reading of BL\textsubscript{closest}. Refer to Robinson [16] for more details.
- **Step 9:** Create a Bus_Location record for each door opening and door closing event in the stop zone
- **Step 10:** Calculate the dwell time of the bus at the stop
- **Step 11:** Calculate the time lost serving the stop using the technique described in [16].
- **Step 12:** Determine the quality of data in the Stop_Visit_Event and provide a quality indicator.
- **Step 13:** Store the data in a Stop_Visit_Event record
- **Step 14:** Write all Stop_Visit_Event records out in XML format

Once the parser was written this then had to be fully tested.

### 4.3.3 Stage 3: Analysis of data in Octave

A program was then written in Octave / Matlab to read the Stop_Visit_Event data output by stage 2. Functions to output the desired figures, such as that of Figure 1, were written. Other functions to determine the typical stop dwell time and time lost serving stop were written.

### 4.4 Benefits of standardisation to dwell time analysis

The three stage process described above was necessary to calculate the typical stop dwell time and typical time lost serving the stop. However the only useful information output was obtained in stage 3. Stages 1 and 2 were required to provide the raw Stop_Visit_Event records required for stage 3. Nevertheless it was stages 1 and 2 that required the bulk of the work. It is estimated that between 100 and 200 hours were required to design, build, and
test stages 1 and 2. This is time that could be more efficiently spent on the research of stage 3. It is therefore highly desirable if the AVL vendors were to introduce functionality in the AVL system to output data required for dwell time analysis in a standard like that proposed in section 4.1.

5. SUGGESTIONS TO ASSIST IN STANDARDISATION

The example provided in section 4 has demonstrated the benefits to academia in defining certain standard outputs from bus AVL systems. The next question to ask is how this can be achieved. This section suggests some initiatives that could be undertaken by academia in its relationship with both AVL suppliers and transit agencies. These are listed in Table 6.

<table>
<thead>
<tr>
<th>Suggestion</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publish proposed standards</td>
<td>There is nothing to stop academics identifying the data that they require from ITS systems and writing a small conference paper with this stated. A partial example of how this could be done is given in section 4 of this document. Publishing such a paper would provide a document which industry and transit agencies could then reference in a requirements specification.</td>
</tr>
</tbody>
</table>
| Engagement with industry community groups | In several countries there already exist industry community groups in the ITS industry. For example, in the UK there are two groups:  
- RTIG: The Real Time Information Group is an industry group for suppliers, local authorities and operators working in the bus Real Time Information industry [17].  
- PTIC: The Public Transport Coordination Group (PTIC) is an industry grouping to provide coordination on the storage, transmission and presentation of public transport information in the UK.  
These organisations have been active in defining the NaPTAN and TransXChange formats identified in Table 1. London Buses is a member of both organisations. However it is notable that there is little academic involvement in either organisation. Academics should aim to at the very least to review all standards proposed by these organisations. |
| Government drive for standardisation      | The UK has been at the forefront of defining standards in the bus transit industry due to support from the country’s Department for Transport, DfT. The DfT has sponsored much of the work on NaPTAN and TransXChange. Other countries could also sponsor such standardisation activities. The government agency could ask academia to review the proposed standard. |
| Partnerships between academia and industry| Academia should establish close connections with the bus AVL industry. There are many areas that industry may wish to do research but cannot justify the expense on. Academics, maybe able to win research grants from central government, to research these problems which industry is unable to fund. |
| Review of ITS tender                      | Local Authorities are obliged to tender ITS contracts in an open |

TRB 2013 Annual Meeting  Paper revised from original submittal.
Standardisation of Bus Transit AVL data

As part of this process requirements documents will be released to prospective suppliers. Academics could potentially review such requirement documents and suggest any changes to the requirements which would provide academia with valuable data sources. Any additional requirements suggested by academia could theoretically be stated as an optional requirement to be priced separately. However, it is likely that many suppliers would include these additional data outputs at no additional cost since the costs of supplying these are unlikely to be very high.

Academia should aim to be more professional in its relationship with transit agencies. Examples of this are given below:

- **Identify data requirements at the outset:** It is common for academics to approach transit agencies for collaboration in a research project. They will then ask the transit agency “what data do you have”? A far more professional approach is for the academics to initially identify exactly what data they would like to be provided to them, and a description of why this data is necessary, and what it will be used for.

- **Remember transit agencies’ operational requirements:** The data requirements of a transit agency are likely to be different to the data requirements of academia. For example, for most transit agencies an approximate indication of traffic flow every 15 minutes may be acceptable. However for many academic research projects far higher quality data at better temporal resolutions is required. However it typically costs transit agencies more money to gather such high quality data so there may not be justification to spend tax-payers money on providing such data.

- **Errors in data are inevitable:** The academic community should remember that errors in the output data of ITS systems are inevitable [18]. Academics should raise such data quality issues in a sensitive fashion to the transit agency and should actively design their models to be robust to such data errors.

### Table 6 – suggestions for enabling academia to define standard data outputs from bus AVL systems

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 1 | This paper has provided a summary of bus transit AVL standards that have been implemented and proposed by industry. Such standards tend to relate to the input data to AVL systems such as the network and schedule data required by AVL and journey planner systems. As yet no standards appear to have been defined for output data from AVL systems.
| 6 | This paper has demonstrated that a large amount of time can be required to simply convert raw AVL log data into a form that can be used for useful measurement on the performance of the transit system and for academic research. It was estimated that over 100 man hours was required.
| 11 | This paper has demonstrated that a large amount of time can be required to simply convert raw AVL log data into a form that can be used for useful measurement on the performance of the transit system and for academic research. It was estimated that over 100 man hours was required.
| 12 | This paper has demonstrated that a large amount of time can be required to simply convert raw AVL log data into a form that can be used for useful measurement on the performance of the transit system and for academic research. It was estimated that over 100 man hours was required.
| 13 | This paper has demonstrated that a large amount of time can be required to simply convert raw AVL log data into a form that can be used for useful measurement on the performance of the transit system and for academic research. It was estimated that over 100 man hours was required.

### 6. CONCLUSIONS

This paper has provided a summary of bus transit AVL standards that have been implemented and proposed by industry. Such standards tend to relate to the input data to AVL systems such as the network and schedule data required by AVL and journey planner systems. As yet no standards appear to have been defined for output data from AVL systems.

This paper has demonstrated that a large amount of time can be required to simply convert raw AVL log data into a form that can be used for useful measurement on the performance of the transit system and for academic research. It was estimated that over 100 man hours was required.
required to design, build, and test some parsers required to provide the basic data required
to analyse bus stop dwell time. If this work has to be repeated at each transit agency that
wishes to perform dwell time analysis, then a very large amount of time would be wasted on
writing basic parsing software. It would be far more efficient for transit agencies and
academia to identify the data structures that they require, codify this as a standard, and then
request AVL vendors to introduce this functionality into their core products. This would allow
transit agencies and academia to spend more time and resources on useful research to
identify issues with the transit network. There are many areas beyond dwell time analysis
that standardised AVL outputs could be defined. These include:

- Line performance indicators
- Bus bunching metrics
- Analysis of junctions
- Analysis of bus priority measures

Having standard data outputs from different AVL systems would also offer additional benefits
such as allowing for an easier comparison between transit agencies, or comparing Operators
running lines using different AVL systems within the same area.

7. REFERENCES


Robinson S, 2013, "Measuring bus stop dwell time and time lost serving stop using London Buses iBus AVL data", submitted to TRB annual meeting 2013
