AUSTROADS ROAD ASSET DATA STANDARD VERSION 4.0

Instructions for Machine Readable Files

1. Introduction

The *Austroads Road Asset Data Standard Version 4.0* (the standard) is published as a PDF document. To make it the standard useable in electronic systems such as databases, it needs to be available in a format that these systems can understand.

Three machine-readable versions of the standard have been created using the following widely understood, platform-independent formats:

- Comma Separated Value (CSV)
- JavaScript Object Notation (JSON)
- Extensible Markup Language (XML)

This document explains the background and purpose of these machine-readable versions of the standard.

1.1 Background

The *Austroads Road Asset Data Standard Version 4.0* is the most recent published version of the Austroads data standard (the standard). This 250-page PDF document is the primary source of supporting information for the road data harmonisation project (ASP6227). The project aims to introduce a level of standardisation to the collection and management of road asset data in Australia. The document explains the background and context behind the creation of the standard and details the various concepts and definitions that are used in the creation of the standard.

While the published standard document is comprehensive and an invaluable resource for organisations using the standard, it has limitations that make it both unwieldy to maintain and impractical for communicating the data definitions or measures that it contains to electronic record keeping systems.

In its current published document format, the data standard comprises several sections, each of which falls into one of the following two categories:

1. explanatory and background information
2. definitions.

Contents

1. Introduction 1
1.1 Background 1

2. Drivers for Change 3
2.1 Managing the standard as a data asset 3
2.2 Communicating the data standard to electronic systems 3

3. Data Formats 5
3.1 Comma Delimited (CSV) 5
3.2 Extensible Markup Language (XML (including XSD)) 6
3.3 JavaScript Object Notation (JSON) 6

4. Data Files 7

The explanatory and background information gives context and meaning to the standard. The definitional sections of the document (sections 7, 8 and 9) contain specific information about the naming, categorisation format and allowable values and ranges for each of the almost 1000 data measures that make up the standard.

Figure 1: Excerpt from the standard showing a data table

Ref	Name	Code	Definition	Example	Type	Precision	Scale	List	Purpose	Soph	Industry reference	PHS	Metrics
8.3.3.31	Number of components	br_comps	Number of same type of components with the same dimensions and material	4	I	2			D	1			
8.3.3.32	Width	br_wid_co	Width of the component in metres	2.45	DC	5	2		D	1			
8.3.3.33	Component type	br_co_type	Component type	TB – T Beam,	A	30		Code List 9.7	D	3			
8.3.3.34	Component material	br_co_mat	Component material	Wood	A	100		Code List 9.27	D	3			
8.3.3.35	Component code	br_co_code	Structure component code according to the Bridge Inspection Manual used in each jurisdiction	1S – Steel box girder	AN	6		Code List 9.6	I	3			
8.3.3.36 (NEW)	Structure Unique Identifier	stru_iden	Structure ID allocated to this asset. This structure ID is provided by the relevant authority. If no structure ID is provided, create an arbitrary ID so this table can be linked with Bridge/Major Culvert Component table							3		N	Structure unique ID (alphanumeric)

Specifically, the information in sections *7 Data Item Specifications (Common Classes)* and *8 Data Specifications (Data Classes)* define **the data standard.** These sections specify in detail the data measures that describe the road system and the associated works and costs required to construct and maintain the system. This is the data that local and state governments collect and maintain in various asset management and financial systems throughout the nation and the data that the project seeks to standardise and harmonise.

The information in section *9 Referenced Code Lists* is supplementary to the information in sections 7 and 8 and defines the allowable information for many of the data measures in the standard. For example, the data measure 8.3.3.33, br_co_type in Figure 1 is restricted to only the types contained in list 9.7 which is shown below.

Figure 2: Excerpt from the standard showing a Code List

Table 9.7: Bridge type

Code	Description
BG	Box girder
COL	Column
TB	T beam

Extracting this definitional information from the published document and creating new versions in format(s) that can be understood by electronic systems such as databases, enterprise resource planning applications and application programming interfaces is the next logical step in the operationalisation of the data standard.

2. Drivers for Change

As Austroads moves towards operationalising the standard, there is a requirement to have the standard available in machine-readable formats. There are two main drivers for this:

- so that the standard can be managed as a data asset
- so that both the definition and the content of the standard can be communicated to users of the standard in a format that is consumable by electronic systems such as databases.

These are explained further in the subsections below.

2.1 Managing the standard as a data asset

Over time there is an expectation that the data measures within the standard will change, primarily from the addition of new measures or deprecation of measures that are no longer required. Each time this happens, it will be necessary to update the standard.

With the standard in a printed format such as PDF, this involves the publishing of a new version of the printed standard. While this will still be a required step, it is impractical for users of the standard to manually compare the previous and new versions and then adapt their electronic systems based on the published changes. The manual nature of this process makes it susceptible to error and does not provide a simple way of checking the differences between versions of the standard.

In a machine-readable format these changes are simple to make and practical to communicate to users of the standard. Additional measures added to the data standard can be appended to the existing standard, with an incremented version number signifying that the measure belongs to a later version of the standard. In this way changes to the standard can be tracked.

2.2 Communicating the data standard to electronic systems

It is a long-term goal of the data harmonisation project that all the systems used in the creation, storage, application and communication of road asset data will eventually use or recognise the data standard. To do this, the developers of these systems and the road managers responsible for using the data need to be able to configure their systems, applications and processes to be compatible with the standard. Developing machine-readable versions of the standard facilitates this goal.

To communicate all the required information about the data standard, the machine-readable versions will contain, where appropriate, the following 17 attributes for each data measure, most of which are derived directly from the published version.

Table 1: Attribute definitions

Attribute	Description
PDS_Level	The level assigned to the data measure in the PDS levelling spreadsheet. This is similar to the PHS measure described below.
Version	The version of the standard at which the measure was added. In the initial version all measures will be version 3. Version is defined as a decimal so the contents of the standard can be incremented in minor (e.g. 3.1, 3.2) or major (e.g. 4) versions.
Code	Lower case descriptive name for the measure. Cannot contain spaces.
Name	A meaningful (lower case) name for the data item, e.g. ‘number of lanes’. Note that generally this does not include the name of the class (so not ‘carriageway number of lanes’). Abbreviations are avoided but may be included when very well known – these are explained in the definition part of the metadata.
Function	The data standard Functional group to which the data measure belongs
Asset_Type	The type of item described by the data measure. A subcategory of function.

Attribute	Description
Ref	Each Item is uniquely identified by a reference number, 9.9.9, or 9.9.9.9. As well as being unique, this identifies the document section where the item is located.
Definition	A generally relatively short definition of the item. E.g. 'A sequential number for every lane on a carriageway'.
Type	A classification identifying one of various types of data, such as alphanumeric, integer or Boolean. The list of Data Types is defined in Table 6.3 of the printed standard
Precision	Is the number of digits in a numerical value. For example, the number 123.45 has a precision of 5.
Scale	Scale is the number of digits to the right of the decimal point in a number. For example, the number 123.45 has a scale of 2.
List	A list of allowable values will be provided for data items where the item must be constrained to one of a set of values. E.g. the list of allowable materials that a deck can be constructed from
Purpose	A category of the main purpose the data is used for. This is either: L Location D Descriptive P Planning (forecasting the future asset state and financial liability) I Optional descriptive data – 'Information'.
Soph	The assessed level of sophistication as defined in Section 4 of the printed standard. This is a guide only and organisations will need to determine whether the data item is applicable for its asset management practice. Either 1, 2 or 3.
Industry_standard	The most relevant industry reference, which in most cases also formed the basis for the related data items.
PHS	Data item identified as a priority for implementation by Road Agencies for industry benefit and effective asset management practice. Codes represent: N Network Reporting (input to a reported network measure). M Management (asset and service).
Metrics	Further descriptive information

Note that:

- **PDS level** is an additional element to the published data standard
- **Version** is an additional element to the published data standard
- **Function** and **Asset_Type** are derived from chapter and section headings in the published data standard
- The **Example** field shown in Figure 1 has not been included

The following section details the different formats of the machine-readable files that are being supplied. Each of these formats have own pros and cons and each are appropriate to different use cases. The decision on which specific file type is most suitable for any given use case is one best taken by the data users at each individual organisation, however any organisations wishing to adopt or incorporate the standard will be able to find a version which is compatible with their particular use case, systems or skills.

For example, if a road condition survey organisation wanted to create a road asset condition database based entirely on the data standard, there is enough information contained in the accompanying files to create all the necessary tables and fields with data definitions such as data format, field size and constraints for all 900+ measures contained in the standard.

A second use case might involve a local government connecting to the road condition survey organisation via an API to import this condition data to their own asset management system. The provided files would allow the local government to create an import process that verified that the data being imported was compatible with their road asset management system, and if not perform the necessary transformations to make it compatible.

3. Data Formats

To provide maximum flexibility to users of the standard, three separate formats have been created:

- Comma Delimited (CSV)
- Extensible Markup Language (XML (including XSD))
- JavaScript Object Notation (JSON)

While each of the three formats has slightly different (but not exclusive) use cases, they all share the following characteristics:

- They are written in plain text
- They are platform independent (non-proprietary) and can be read/ingested/used by a variety of languages and applications
- They contain all of the necessary information from sections 7 and 8 of the standard to make the information of practical use to electronic systems (see section 2.2. above)

3.1 Comma Delimited (CSV)

This is a text-based format that is typically opened in a spreadsheet application such as Microsoft Excel or Google Sheets. Data is separated by commas in the application and a CSV file typically includes column headers in the first row to assist applications in the understanding of data.

CSV files are the most easily readable version of the data standard for humans and are also readily ingested by most data-based applications. However, the simplicity of the files means that typical attributes of data such as structure, format, constraints or validation cannot be applied.

In text format a CSV file appears like this:

```
heading 1, heading 2, heading 3
value 1, value 2, value 3
value 1, value 3, value 3
```

where each row describes an object, in this case an address, e.g.

```
street number, street name, street type
17, McCubbin, Avenue
23, Roberts, Parade
```

Note that there are some fields in the original data standard that contain commas, mainly the “Definition” field. Before exporting to CSV these have been replaced with a double underscore, “__”, so as not to confuse the applications that are using the CSV based information.

3.2 Extensible Markup Language (XML (including XSD))

Extensible Markup Language is a more ordered and sophisticated way of displaying text-based, machine-readable data. XML uses tags in the same way that the HTML language used by web pages uses tags, to define different pieces of information. Information in an XML file consists of a piece of information between two tags that label the information. From the example above a street number would be stored as:

```
<street_number>17</street_number>
```

XML information can be nested to give information a hierarchical structure. Again, using the information above we can create an <address> tag that contains all of the address information, e.g.

```
<address>
	<street_number>17</street_number>
	<street_name>McCubbin</street_name>
	<street_type>Avenue</street_type>
</address>
```

And these objects can be nested in larger objects. In this example, we can create a <member_address> tag:

```
<member_address>
	<address>
		<street_number>17</street_number>
		<street_name>McCubbin</street_name>
		<street_type>Avenue</street_type>
	</address>
	<address>
		<street_number>23</street_number>
		<street_name>Roberts</street_name>
		<street_type>Parade</street_type>
	</address>
</member_address>
```

Unlike CSV files, information in XML files can be validated through the use of an XML schema or data definition file, also known as a XSD file. An XSD file can define the attributes that can be applied to an XML tag, for instance that a street number must only contain integers, or that a postcode can only be four characters long.

3.3 JavaScript Object Notation (JSON)

JSON is a plain text format most often used as for storing and transporting data. Information in a JSON file is stored in key:value pairs, where the key provides a label for the value, e.g.:

```
"street_number":17
```

JSON has similarities to both CSV and XML files. Like CSV, JSON data can be grouped into an object, but rather than using a new line the data is enclosed in curly braces, e.g.

```
{"street_number":17,"street_name":"McCubbin","street_type":"Avenue"}
```

And like XML these objects can be grouped in larger objects using arrays, defined by square brackets:

```
{"member_address":[
{"street_number":17,"street_name":"McCubbin","street_type":"Avenue"},
{"street_number":23,"street_name":"Roberts","street_type":"Parade"}
]}
```

JSON information can also be validated against a schema.

4. Data Files

The following files make up the machine readable versions of the standard.

Table 2: Data files

File ID	Asset Type	Asset file name
1	CSV file containing entire data standard	Austroads_Data_Standard_v4_0_Measures.csv
2	XML file containing entire data standard	Austroads_Data_Standard_v4_0_Measures.xml
3	JSON file containing entire data standard	Austroads_Data_Standard_v4_0_Measures.json
4	XSD schema file containing data standard definition	Austroads_Data_Standard_v4_0_Definition.xsd
5	JSON schema file containing data standard definition	Austroads_Data_Standard_v4_0_Definition.json
6	CSV file containing information from supporting tables	Austroads_Supporting_Table_Data.csv
7	XML file containing information from supporting tables	Austroads_Supporting_Table_Data.xml
8	JSON file containing information from supporting tables	Austroads_Supporting_Table_Data.json
9	Licencing information	Licence.txt

- File 1 is considered the master file for the data standard measures. It contains all of the reference. information for all 900+ elements of the data standard and is used as the basis file for the creation of the other machine-readable files.
- File 2 and file 3 are derived directly from File 1 using a standard data conversion tool.
- File 4 is the master reference file for data definitions. It defines the acceptable parameters of the attributes of the data measures, such as precision, data format and scale.
- File 5 is derived directly from File 4 using a standard data conversion tool.
- Files 6-8 are supporting files that contain the acceptable values for measures that have a restricted list of values noted in the “List” attribute as “Code List xx”. Note that some measures have restrictions on values that are directly defined in the “List” attribute e.g. measure 8.1.24, link_s_uni has acceptable values of U or V defined in the list attribute. For completeness these values could be converted to lists or other constrains (e.g. max / min).