EBSR Help Sheet

Introduction to XML and TransXChange

EBSR Help Sheet - Introduction to XML and TransXChange

What is XML?

XML stands for eXtensible Markup Language. It is a system for storing and transporting data. XML uses 'tags' to define the properties of different types of data.

You can think of tags as a bit like the headings in an address book. In plain text, a name would be tagged as <name>Chris<\name> to show that it was a name. A telephone number would be tagged as <telephone>01234 56789<\telephone> to show that it was a telephone number. As long as the piece of data keeps its tag, a computer can identify what kind of data it is.

Why are tags so useful? Some kinds of markup language, such as HTML (Hyper Text Markup Language), use a set of pre-defined tags. This means that every computer will interpret a tagged piece of data in the same way. HTML tells a computer how to display a piece of data to the user using a web browser. For instance, if a web browser sees a word tagged as <bold>bold<\bold>, it will show the word in **bold** on the screen. It won't show the word underlined or in *italics*. The advantage of pre-defined tags is that the fixed set of rules makes it possible for web-pages to display text and graphics in the same way in any browser on any computer in the world.

What makes XML so powerful is that you can define the properties of each tag and the order in which tags must appear in relation to one another. Software engineers can write software which reads the XML and applies rules to ensure that the data makes sense. For instance, a rule could be applied to all <telephone> tags, rejecting any pieces of data that did not contain only numbers. The ability to implement rules in this way makes it easier to capture, send and read accurate data between computer systems.

The set of rules that should be applied to a specific set of tags is called a 'schema'. As long as both you and I have the same schema version, we can ensure that the data that we exchange is accurate and intelligible.

What is TransXChange?

TransXChange is a special set of XML tags that can be used for describing public transport data. The latest published schema, version 2.1, defines a set of tags for labelling all of the different elements of a public transport journey – the service number, the unique bus stop reference numbers, the times at each stop and the operator name. TransXChange is designed to replace older methods for describing public transport journeys such as ATCO-CIF.

The TransXChange schema defines the order in which the elements of the journey should be presented, and contains rules about the data enclosed by specific tags. For instance, a NaPTAN bus stop code must begin with a valid 3-letter ATCO prefix. If it doesn't, the file can be rejected by the software reading the TransXChange file, as no-one wants to use a bus stop reference that's not in the NaPTAN database!

As already mentioned, TransXChange is a system for storing and transporting data about public transport journeys. It can be useful to display this information on-screen, to check that it makes sense to a human eye. A piece of software called TransXChange Publisher can be used to present TransXChange files in a normal timetable format as a PDF file on a computer.

TransXChange 2.1 is used as the standard format for exchanging files in Electronic Bus Service Registration.