

What steps are needed to move a relational Roads Dataset to a Roads Linked Data Dataset?

Alexander Ramage; Head of Management Information Systems; Transport Scotland.

e-mail: alex.ramage@transportscotland.gsi.gov.uk

Background

Transport Scotland is responsible for the Trunk Road network in Scotland as well as setting the policy and running a number of other transport functions in Scotland. One of the datasets that we maintain in Transport Scotland is a detailed shape and referencing systems for the Roads that we are responsible for. We currently publish a Web Mapping service of the information¹ and will shortly (once infrastructure is put in place) publish a download service. The URL may be changing in the near future.

Relational Data set

Although the network data set shown in the URL below is quite a simple dataset, the Road network referencing information is the primary key that allows all data in our database to be identified against the physical location on the Road. The network referencing is held as a section of the Road (no more than 10km in length – usually significantly less) and a distance along the section. On the physical road there are node markings to allow users of the data to find points on the network; although most of the users now use GPS to identify a location and use “reverse segmentation” to link back to the network referencing.

Data that could be made into a controlled vocabulary

Although the data is fundamentally about the way in which the base network is described there are a number of fields in the data that need to be controlled, i.e. only those values from an approved data source are acceptable. To give a concrete example, a length of Road can be described by the number of lanes of traffic the Road is capable of holding. The valid values for this attribute can be found in the “Design Manual for Roads and Bridges”. Not all the information will come from defined data sources and how will Transport Scotland create ways of describing data that is interoperable with other sources.

Spatial Data

A substantial amount of the information in the data is spatial in character. The shape of the Road is defined as a Polyline, that is made up of a number of Points to describe the topological shape of the Road network. Other dataset may represent

¹<http://sedsh13.sedsh.gov.uk/ArcGIS/services/TransportScotland/MapServer/WMServer?request=GetCapabilities&service=WMS>

the same Road network, but the Polylines will not necessarily match the Polylines in our dataset. While the Roads represented in the two data sets are exactly the same, how in Linked Data can we make an assertion that these two Polylines represent the same Data. The issue of matching is made worse when the spatial element is a Polygon. Matching polygons which are made up of Polylines (which may themselves be different) is a big problem. An example of where this might be important relates to the purchase of Land, where a comparison of land holdings against the land occupied by the Road or a future Road is important.

What do we need to do – stepping stones to Linked Data

As identified above in the section on Controlled Vocabularies, we need to identify those elements in our data that could be validated against a Vocabulary. We then need to identify if that vocabulary exists (and use it if that is the case; however given the domain I suspect that there will not be many of these). If one does not exist, we need to create the vocabulary, and we need to make this vocabulary known to others in the same domain. This implies that some sort of domain registry will be required. Further a registry of registries will also need to be in place. To support this process Transport Scotland (or Scottish Government) on our behalf will need to “mint” a URI in order for Transport Scotland to “define” the vocabularies in a way that uses persistent identifiers. All this work is simply the preparation required to move to Linked Data.

The Challenges

Some of the steps that need to be taken are steps that are easily understood – those being the creation of vocabularies and registries of them. Although we have not actually taken these steps, we can see how these would benefit our business. Not only will the steps benefit the business directly, the same steps may well benefit the adoption of “Building Information Modelling (BIM)”, as objects will already have been defined in an unambiguous manner – a key requirement for BIM. The difficult steps are putting the spatial information together in a way that can be used to match across to other dataset – that do not necessarily have the same parentage. Finally the challenge that I think that scares people the most from two perspectives. First the tool sets that are available to work in this space are not necessarily mature both in the “Extract (from Relational) Transform (into RDF) and Load (into an appropriate database)” and tools to compare different RDF nets to identify similarity.

Conclusions

Transport Scotland (and the Scottish Government) have concluded that preparing to move to Linked Data is an approach that we want to pursue, although we do not expect to move in a big bang approach. We will slowly develop our thinking and implement appropriate registries and vocabularies before taking the step to move the data to a “tuple” format. Scottish Government have recently released a report on a

Trial Linked Data implementation to explore the benefits and difficulties of this approach.² The Pilot website can be found: www.opendatascotland.org.

² <http://www.scotland.gov.uk/Publications/2013/12/6550>