POSITION PAPER: Ontology in the Rail Domain
The Railway Core Ontologies

Mr Christopher Morris[1], Dr John Easton[1] and Professor Clive Roberts[1]
[1]*Birmingham Centre for Railway Research and Engineering,University Of Birmingham, Birmingham, United Kingdom*
{cxm210, J.M.Easton,C.Roberts.20}@bham.ac.uk

Keywords: Railway Domain Ontology, Data Integration, Railway Core Ontologies

Abstract: This paper presents the railway core ontologies, a group of related ontologies designed to model the rail domain in detail. The purpose of these ontologies is to enable improved data integration in the rail domain, which will deliver business benefits in the form of improved customer perceptions and more efficient use of the rail network. The modularity of the ontologies allows for both detailed modelling of the domain at a high level and the storing of instance data at lower levels. It concludes that the benefits of improved rail data integration are best realised through the use of the railway core ontologies.

1 INTRODUCTION

The railway core ontologies are a set of related models that capture the rail domain in depth, using a modular structure to ensure that the ontology is as lightweight as possible for any given application.

In the UK the technical strategy leadership group reported that “Excluding Network Rail’s own information systems, research discovered over 130 information systems maintained by approximately 20 suppliers” (Technical Strategy Leadership Group, 2012). Previous projects also note the fragmented and proprietary nature of the existing information systems. It has been found that “Where electronic data exchange standards for rail do exist, many are proprietary binary formats used to provide point-to-point interfaces between specific systems and not intended for use in a generalised context.” (Easton et al., 2013a). This situation caused the regulatory body responsible for the oversight of the UK rail domain, the Office of Rail Regulation, to express concerns as to the current state of asset condition monitoring (Office of Rail Regulation, 2012).

One area which will give rise to increased volumes of data is that of remote condition monitoring as described in (Garcia Márquez et al., 2003). Typical applications are remote sensors to report on the impending failure of railway systems such as points, in the case of infrastructure, or bearings in the case of vehicles. Similarly capable of generating large volumes of data are inspection vehicles, which regularly run over the rail network, to inspect the fixed infrastructure. These often record video, and many sensor readings per second along with the time and location of the reading. Data volumes are discussed in the context of the US rail network by Allan Zarembski (Zarembski, 2014) and in the context of European rail networks by Núñez et al (Núñez et al., 2014). Numerous studies, including (García Márquez et al., 2008), suggest benefits in the form of a reduction in unplanned, reactive, maintenance and a reduction in manual inspection, may be derived from asset condition monitoring.

Another area experiencing growth is passenger information. Studies, such as those reported by Dziekan and Kottenhoff in (Dziekan and Kottenhoff, 2007) show that customer perceptions and ridership can be improved by improving customer information.

There already exist a number of ontologies, such as the REWERSE project, reported in (Lorenz, 2005), that cover the broader transport domain, though these do not aim to create a domain ontology for the rail industry. This does include the locations of interchanges between modes (including rail) and the path (both geographic and network) taken by rail lines, along with journey planning information. The ArkTRANS project serves the multi-modal transport domain and has been employed by the Norwegian State Railway company (NSB) to produce a functioning journey planner along with being the basis for a number of other projects. ArkTRANS is introduced in (Natvig and Westerheim, 2003). Further work has been done implementing the concepts set out in the original project, including a domain ontology for freight transport, set out in (Gönczy et al., 2012) with some cross over to the rail network and a project relating to the maritime domain (Ørnulf, 2011).

Technology has matured since the use of ontology for the rail domain was first proposed by the InteGRail project, the final report for which was completed in 2010 (Köpf, 2010). Much of the work done for that project however pre-dated the availability of a number of the tools employed by RaCoOn.

The railway core ontologies aim to allow the rail domain to share information more readily, for example, when a new type of asset is added, its status can be inferred without further software development. TThis is discussed by Tutcher et al (Tutcher et al., 2013), where it is used by both industry and academia.

2 BENEFITS OF USING THE RAIL CORE ONTOLOGY

The benefits of using the rail core ontology stem from the ability to draw together disparate data sources. The following section outlines usage scenarios for the railway core ontologies.

As an example consider component life-cycle information stored in a maintenance database. There is little value in knowing how many times a given component has functioned without knowing how many times it can be expected to function. Taking this a step further, knowing how a given component performs when it is in good working order and how it degrades allows for better informed maintenance - it is possible to replace it before it fails. Another example in passenger information systems is knowing that a train passed a certain signal and was travelling at a certain speed is useful for basic platform boards, however, knowing that there are speed restrictions further along the line allows for more accurate information.

The railway core ontologies serve this purpose well; not only do they serve as a standard for data interchange, as do XML standards used within the rail industry (railML), but they allow for any stakeholder to add their own custom extensions without a need to consult standards committees and thus avoid the delay that imposes. If, for example, a company is working on a new product then they can simply create any extensions to the ontology it may require. These extensions could then be released at the same time as the product. Furthermore as the ontologies naturally grow less functionality will need modelling - it will be possible to make available product documentation using existing ontological constructs. For this reason nothing is limited to pre-defined lists.

The use of ontological reasoning has benefits above and beyond those from the data integration. Reasoning can for example be employed to infer the condition of assets, the location of trains or availability of routes, from incomplete data.

There are other benefits of moving the logic out of the 'front end' and into the ontologies. When the front end is updated to reflect a changing IT environment the ontologies need not be altered, conversely editing the ontology need not imply development work. Since changes to the ontology need not be a task for a developer then as changes are made to the rail network any engineer can update the relevant ontology to reflect them. Providing tools to make this possible are part of the scope of this project.

2.1 Railway Core Ontology usage scenarios

Discussion with stakeholders, such as the UK's infrastructure manager, Network Rail, and a large engineering firm along with work stemming from the Factor-20 project, (Roberts et al., 2011), has identified the following scenarios where the railway core ontologies would add value.

2.1.1 Scenario 1 – Forward Planning

A scenario suggested by the UK infrastructure manager was the ability to answer forward looking "what if" questions relating to the costs and benefits of asset maintenance. The example given was of a bridge which is found to be in a condition such that it will soon need heavy maintenance. In this example the available options were to repair or replace it, with a further choice as to the loading gauge of the replacement. Whilst exact costs will vary on a case by case basis it should be possible to estimate costs from previous work. Questions such as how much revenue would be lost were the bridge to be replaced more cheaply with one not gauged for freight as well as the relative cost of repair versus replacement needed expert assessment, which was done on the basis of less than ideal data. Data currently resides in a number of silos: the finance system, the technical drawing document management system and train movement databases being three of the most prominent.

2.1.2 Scenario 2 – Maintenance Timing

A similar forward looking question answering scenario, put forward by the UK infrastructure manager, that brings together cost centre and operational information is determining the best time to carry out maintenance. When maintenance is best timed, from a revenue loss perspective is a non trivial question. To answer it information as to the number and ticket type of passengers must be brought together with information regarding outstanding maintenance items, whilst also considering the needs of freight operating companies.

The ontologies could also be of use more directly in the domain of asset maintenance. This would take available information from the physical network and make it available at both an operational level for day to day work and a corporate level for long term planning.

2.1.3 Scenario 3 – Train Identification

The linking of trains with alerts from track-side monitoring equipment could be more efficient using the ontologies. Wheel impact load detectors, a device for detecting trains with misshapen wheels, are a good example of this. These track mounted sensors measure how much force a wheel strikes the track with and, if the force is too great, the wheel is known to be defective and liable to damage the track. The only data available from the detector is the time and force of the impact. Where data from these sensors is found to be exceptional it is currently a manual process to identify the offending wheel and repair it. Data from one system is printed, assessed, phone calls made, the problematic train identified and the maintenance department of the train operating company (who are distinct from the infrastructure owners in most European countries) informed. Train movement data can tell you in what order trains passed down a given line. Making that information available, along with the time of impact and axle count from the wheel impact load detector would allow a system using ontologies to infer which wheel of a particular train had triggered the detector. This concept was also studied as part of the FuTRO project (Tutcher et al., 2013).

2.1.4 Scenario 4 – Predictive maintenance

Whilst planned maintenance has a lower total cost than corrective or condition-based maintenance, predictive maintenance is more cost effective still, as stated by Umiliacchi et al. (Umiliacchi et al., 2011). In the case of planned maintenance, assets are repaired on a fixed schedule, regardless of condition; in the case of predictive maintenance, work is done when a computer model, informed by sensor data, predicts it will be necessary. The intervening step of condition-based maintenance has been taken up for certain asset types, but often only with the use of crude thresholds and manual inspection. Generally, measurement equipment, be it integrated in a points motor or attached to a measurement vehicle, sends back continuous values for sensor readings, but these are only taken into account should they be outside of the specified envelope. When the reading exceeds the specified maximum or minimum, the asset is maintained, thus remaining in a safe condition.

Predictive maintenance goes a step beyond this using the trend exposed by the rate and direction of change of the underlying data to suggest when maintenance would best be scheduled. Whilst this is possible on an asset by asset, system by system basis, without the use of the railway core ontologies this process would be very slow and costly, to such an extent that it is unlikely to be undertaken on a large scale. The railway core ontologies would allow the implementation of a network wide system with only asset specific data being added, rather than extensive customisation.

2.1.5 Scenario 5 – Customer Information

Train location information is currently disseminated via a system known as DARWIN, for the purposes of passenger information. This takes the available train position data from track-side infrastructure, the accuracy of which varies depending on source, and combines it with timetable data. This is then displayed on the passenger information boards in stations as well as on a range of websites and mobile applications.

Efforts, described by Easton et al. (Easton et al., 2013b), are underway to integrate data from GPS units on trains with that from traditional track circuits (regularly over a km in length) to improve accuracy. Track circuits are electric circuits used to ascertain which section of track a train is on. This is an area in which the railway core ontologies can help; however there are far greater benefits to be derived when integration with other transport modes is considered. In order as to plan a multimodal journey it is necessary to have information from all modes available and integrated in one place. Here the rail core ontology is not the only one that needs consideration; work based on the Arktrans model (Natvig and Westerheim, 2003) also needs consideration, as does the General Transit Feed Specification, itself available as linked data – discussed in (Santos and Moreira, 2014).

3 IMPLEMENTATION

The strength of the railway core ontologies comes from their modularity. The rail domain is very large and a model which describes all of it in detail would, with the currently available technology, be unsuitable for high volume information retrieval tasks.

Several of the implementation decisions taken build upon the work of the first project focused on a domain ontology for the rail domain, the InteGRail project. This, as summarised in (Köpf, 2010), aimed to "allow IMs[Infrastructure Managers] and RUs[Railway Undertakings] across Europe to act as a single system." Amongst other deliverables a "Network Statement Checker" was produced as a demonstrator. This is a tool to ascertain whether a given train configuration can run over a (often transnational) route. This will be based on the availability of automated signalling systems, gauge, electrification and other factors which vary from area to area. This work was limited primarily by the available technology – whilst the report was published in 2010 it was written in 2009 and not all of the tools employed in the implementation of the railway core ontologies existed at that point.

The architecture of the broader system of which the ontologies will form a part also requires consideration. As originally suggested by InteGRail the system will use a Service Orientated Architecture (SOA). The system will by its nature be distributed across many stakeholders. Whilst questions of 'ownership' of information in this context are complex, the solution will involve physically disparate systems owned not merely by differing stakeholders, but at times, commercial competitors.

3.1 Provenance

The number of stakeholders gives rise naturally to issues of trust and provenance. These are being tackled, in part, by the use of the prov ontology[1]. The issue of access to information is further addressed by the use of triple stores that allow selective visibility - for example allowing users access to selected graphs.

3.2 Internal Organisation

The modular design of the ontologies is key to their practicality. The ontologies themselves are composed of a number of different namespaces only a subset of which need to be imported for any given task. The ontologies break into three broad categories:

1. An upper ontology, which may easily be mapped onto other upper level ontologies as needed. The upper level deals with concepts such as observation and place. Figure 1 shows a selection of classes from this layer.

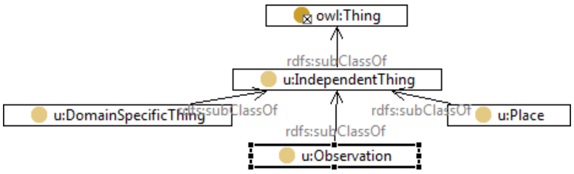

Figure 1: The upper ontology - http://purl.org/ub/upper

2. A core ontology, which can be either three or four dimensional, depending on which file is imported. When working with large volumes of data it is envisaged that the 3D version will be used, whereas the 4D version is better suited to model exchange. Reification as a means of representing changes over time was dismissed owing to the proliferation of triples it engenders and the reduction in OWL reasoning capability it causes.
3. Task ontologies: Task specific ontologies that represent a deep study into a particular area. Timetabling, Rolling Stock and Infrastructure have been implemented, with more expected to follow. Figure 2 shows a selection of classes from the Rolling stock module. Note that this includes such things as locomotive models.

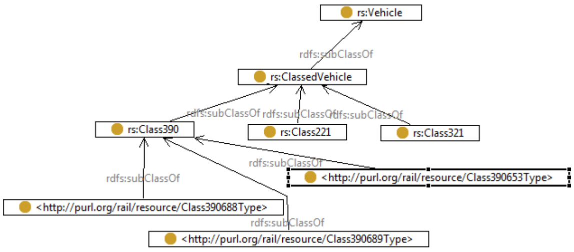

Figure 2: The Rolling Stock sub-ontology

Note that whilst all lower levels depend on the levels above, but none of the higher levels depend on any level below.

The way the namespaces and thus modules of the ontology inter-relate may be observed from figure 3 on the facing page, which shows the namespaces used by the rolling stock module. The namespace entitled "vocab" forms a part of the core ontology, "u" is the upper level. Note that, where possible, external namespaces are used.

Another implementation choice was that of complexity. Here there is a clear trade off: expressivity vs computational complexity. This trade-off has been the subject of much work, both at this centre and within the wider community. The result was that

---

[1] http://www.w3.org/TR/prov-o/

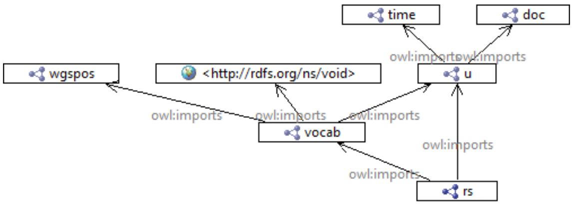

Figure 3: Rail Core Ontology Namespaces

each module is split into two parts: one containing the T-Box, which can itself be large, and another containing constraints which can be far more expressive. The constraints ontology must then be kept as small as possible. In this case the T-box is where possible OWL-RL compliant. A-Box raw data uses RDF schema reasoning whereas the constraints are as expressive as necessary. The triple store with which this was developed - stardog - allows for constraints to be reasoned over separately and at a different complexity to other parts of the ontology.

3.3 The FuTRO project

The FuTRO project "aimed to show how shared, open access ontologies and linked data could help the UK rail industry" (Tutcher et al., 2013). The Railway Core Ontologies provided the network model used by this project, in particular information as to which asset was monitored by which sensor and how these assets interacted was stored. Listing 1 is an exert of the pattern used to this end.

Listing 1: Observation Pattern. Note that comments and labels have been omitted to improve readability.

```
:associatedObservation
      a owl:ObjectProperty ;
      rdfs:domain u:IndependentThing ;
      rdfs:range u:Observation ;
      owl:propertyChainAxiom ( :monitoredBy :
          observedEvent ) ;
      owl:propertyChainAxiom ( :dependsOn :
          monitoredBy :observedEvent ) .
```

A demonstrator was developed, which used inference to show which lines were unavailable in the event of mechanical failure of railway switches. Similarly the ontology has provision for storing the physical location of the asset such that maintenance teams can get get to it.

The project has suggested that very high volume raw data be stored separately to the ontology and only a key to it be retained. In the case of the FuTRO project this was accomplished using the REDIS[2] key value store where appropriate. Further extensions are envisaged to allow the ontology to advertise the web services that could be called to retrieve that data.

Another factor aiding implementation at this time, which was not available at the time of the InteGRail project, is the existence of libraries for working with linked data. The FuTRO project made used of the dotNetRDF[3] to allow software written in the ".Net" family of languages to interface with triple stores, which removed one of the barriers to integration with the project industrial partners.

As is stated in the final report (Tutcher et al., 2013) two demonstrators were produced, along with that report. The train location demonstrator is publicly available and accessible at:

http://purl.org/rail/trainlocator

It demonstrates the use not only of linked data but also ontological reasoning, to infer a train's location even when it is no longer available in the original format.

4 FUTURE WORK

Although the core ontologies already capture enough of the rail domain to demonstrate value many areas require further modelling effort. Remedying this will require input from domain experts and as such tools need to be provided to allow railway engineers to edit the ontology. In order to align with existing industry best-practice for the procurement of new ICT systems, these tools should, where possible, be commercial off the shelf software (COTS).

There is also a need to enrich samples of existing data, documenting the process and making available sample code for use by stakeholders in the rail industry. Tools enabling conversion from the popular "railML" XML format would enable a quicker transition.

Tools to query the ontology and present the results to non-technical users are also required. Network Rail has expressed an interest in work around presenting Key Performance Indicators to senior managers for example, beyond the work outlined in section 2.1 on page 2.

The prioritisation and selection of data sources for conversion also requires consideration. The possibility of using a combination of contractual compulsion and reciprocal data exchange, to encourage those stakeholders that supply assets (railway infrastructure, vehicles etc.) to supply them along with data sheets in a format compatible with the railway core

---

[2] http://redis.io/

[3] http://dotnetrdf.org/ a comprehensive and well documented open source library for interfacing with triple stores

ontologies needs examination. Other existing data will need to be prioritised according to its value to stakeholders.

The industry also requires design patterns for the larger system and guidance for using these patterns. Many means exist for webservices to advertise there availability and this needs to be standardised such that all stakeholders can access information. Distribution and ownership of information issues also need addressing - the current solution, discussed in section 3.1 on page 4 is only partial.

5 CONCLUSIONS

It is clear that much can be achieved by bringing together data from different sources; data currently resides in isolated silos and value for all stakeholders exists where it is brought together. This value exists as money saved by more efficient maintenance, lower costs incurred due to equipment failure, less staff time wasted manually compiling reports and less long term IT expenditure. Beyond the savings there is also value from improved passenger perceptions: improved customer information has been shown to improve customer perceptions of the service. Twinned with improved reliability this has potential to increase rider-ship and freight traffic.

The use of an ontology as opposed to lighter weight, data only, standards has multiple advantages: It is easier and quicker to modify, without affecting front-end applications. The ontologies can be modified more quickly than could a traditional standard. Information from one sub-domain can be reused in another, and can be combined to easily obtain further information. The ontology can be self documenting. The disadvantages are skills, which are less common than traditional IT skills and computational complexity. These can both be overcome; the lack of IT skills by provision of well designed software and complexity by carefully addressing the trade off between expressivity and complexity discussed in section 3 on page 3.

All these advantages suggest the Rail Domain Ontologies are the best way to bring together data in the UK Rail Domain.

REFERENCES

Dziekan, K. and Kottenhoff, K. (2007). Dynamic at-stop real-time information displays for public transport: effects on customers. *Transportation Research Part A: Policy and Practice*, 41(6):489–501.

Easton, J., Chen, L., Davies, R., Tutcher, J., and Roberts, C. (2013a). Ontime - Part D7.1. Technical report.

Easton, J., Chen, L., Davies, R., Tutcher, J., and Roberts, C. (2013b). RSSB Cross industry Railway Information Systems Workshop 7 th December 2012. Technical Report January.

García Márquez, F. P., Lewis, R. W., Tobias, A. M., and Roberts, C. (2008). Life cycle costs for railway condition monitoring. *Transportation Research Part E: Logistics and Transportation Review*, 44(6):1175–1187.

Garcia Márquez, F. P., Schmid, F., and Collado, J. C. (2003). Wear assessment employing remote condition monitoring: A case study. *Wear*, 255(7-12):1209–1220.

Gönczy, L., Kövi, A., Andreas, P., Thalhammer, A., and Stavrakantonakis, I. (2012). e-Freight ontology. Technical Report 233758.

Köpf, H. (2010). InteGrail – Publishable Final Activity Report. Technical report.

Lorenz, B. (2005). A1-d4 ontology of transportation networks.

Natvig, M. K. and Westerheim, H. (2003). Joint effort on establishment of ARKTRANS–a system framework architecture for multimodal transport. Technical report.

Núñez, A., Hendriks, J., Li, Z., Schutter, B. D., and Dollevoet, R. (2014). Facilitating Maintenance Decisions on the Dutch Railways Using Big Data : The ABA Case Study. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 48–53, Washington, DC.

Office of Rail Regulation (2012). Network Rail Monitor - Quarter 1 of Year 4 of CP4 |1 April 2012 – 21 July 2012. Technical Report July.

Roberts, C., Easton, J., Sharples, S., Golightly, D., and Davies, R. (2011). Rail Research UK Feasibility Account (FA): Factor 20 – reducing CO 2 emissions from inland transport by a major modal shift to rail (EP / H024743 / 1). Technical report, Birmingham.

Santos, M. Y. and Moreira, A. (2014). Integrating Public Transportation Data : Creation and Editing of GTFS Data. 2:53–62.

Technical Strategy Leadership Group (2012). THE INDUSTRY ’ S RAIL TECHNICAL STRATEGY 2012 THE RAIL TECHNICAL STRATEGY 2012. Technical report.

Tutcher, J., Easton, J., Roberts, C., Myall, R., Hargreaves, M., and Tiller, C. (2013). Ontology-based data management for the GB rail industry Feasibility study. Technical report, RSSB.

Umiliacchi, P., Lane, D., Romano, F., and SpA, A. (2011). Predictive maintenance of railway subsystems using an Ontology based modelling approach. pages 22–26.

Zarembski, A. M. (2014). Some Examples of Big Data in Railroad Engineering. In *IEEE BigData 2014*, Washington, DC.

Ørnulf, R. J. (2011). A Maritime ITS Architecture for e-Navigation and e-Maritime : Supporting Environment Friendly Ship Transport. pages 1156–1161.