RAILWAY MODELLING
The Case for Ontologies in the Rail Industry

J. M. Easton, J. R. Davies and C. Roberts
School of Electronic, Electrical and Computer Engineering, University of Birmingham, Birmingham, U.K.

Keywords: Railways, Data exchange, Domain ontology.

Abstract: As the demand for rail travel grows amongst the travelling public, capacity on the largely Victorian infrastructure is becoming an important issue. High construction costs and the need for environmental responsibility mean that simply building more lines is not necessarily a viable option, and therefore the rail industry is beginning to look for ways in which their current assets can be used more intelligently. In the first wave of this process, a number of new projects have been started that monitor the condition of vehicles and infrastructure, but the greater challenge of integrating that data and using it to support business operations is still largely unaddressed. This paper outlines a number of data handling and data sharing issues that face the rail industry, and presents the argument for the adoption of an ontology-based data standard across the sector.

1 INTRODUCTION

As rail transport becomes an increasingly attractive option to the travelling public, offering a fast, reliable means to beat traffic jams and travel straight into major urban centres, issues of capacity on the largely Victorian railway network in Great Britain are of ever greater importance. A recent report has shown a 7.3% increase in passenger kilometres between 2007-08 and 2008-09 (Office of Rail Regulation, 2009). The huge costs associated with alterations to infrastructure mean that increasing capacity through the large-scale construction of new lines is not always a viable option, and as a result the railway industry across Europe is beginning to look at ways in which the railways can be made smarter.

As part of the move towards a smarter railway, Network Rail (the owner of Britain's rail infrastructure) has started two complementary new initiatives - Intelligent Infrastructure (IIS) and LiveTrain. The Intelligent Infrastructure project, currently being trialed on a stretch of line between Edinburgh and Glasgow, monitors the condition of fixed assets such as track circuits, points, and signalling systems, processing the data and raising alarms for operators via a supervisory control and data acquisition (SCADA) system. The LiveTrain project by comparison, will see hundreds of older passenger and freight trains on the British railway network retrofitted for condition, power and location monitoring. In the first instance it is hoped that these projects will improve the reliability of existing assets, leading to more efficient maintenance practices and reduced delays across the network.

2 ISSUES AFFECTING DATA SHARING IN THE RAIL SECTOR

The data that will be collected by projects such as IIS and LiveTrain is an important step towards improving performance on the smarter railway; however, the greater tasks of storing, processing and presenting the information remain to be addressed. In the British railway industry these tasks are made more complex by a number of key factors, including a range of legacy systems, a large number of competing stakeholders, and historical working practices. When looking further afield to the European, US and Asian railway networks, the situation is further complicated by differences in nomenclature and units of measurement. In order to properly put forward a case for an ontology-based data standard in the sector, we must first elaborate on these issues.

M. Easton J., R. Davies J. and Roberts C..
RAILWAY MODELLING - The Case for Ontologies in the Rail Industry.
DOI: 10.5220/0003092102570262
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2010), pages 257-262
ISBN: 978-989-8425-29-4

2.1 Legacy Systems

Data logging in general, and the monitoring of the condition of assets in particular, is not a new idea on the British railway network (although IIS and Live-Train will perform these tasks on a much larger scale than was previously the case). Both the infrastructure operator, and various train operating companies have some ability to monitor the condition of their assets. Network Rail for example, has the WheelChex system, which records the impact force that a train's wheel makes on specially instrumented sections of track, and hot axle box detectors, which can monitor the temperature of the axles on passing vehicles in order to identify potential problems. These systems provide useful information and represent a significant amount of previous investment. Certainly, for management reasons if nothing else they must form an important part of the smarter railway. Unfortunately, many of the current monitoring systems were developed and installed in isolation, and may also be operated by third-parties. The raw data, if available at all, is in proprietary formats, greatly increasing the complexity of any attempt to combine it with data from other sources.

Equipment longevity on the rail network is expected to be high. In 2007 the average age of rolling stock on the British network was around 13 years, one of the lowest figures in Europe (Department for Transport, 2007). On vehicles that already have condition monitoring capabilities, different manufacturers may have chosen to instrument different systems, meaning that there is no standard set of measurements that can be used as a common basis for metrics across the network. Trackside equipment can be much older, with some side lines still using semaphore signalling instead of the newer electronic signals.

Similar problems can be seen with software systems, a good example being the Total Operations Processing System (TOPS) used by the Train Operating Companies to store information on locomotives. TOPS was created in the 1960s to run on mainframe computers, is written in its own programming language, and generates rather cryptic plain-text reports.

2.2 Stakeholders

The privatisation of the British railway network in the late 1990s has led to a confusing array of different stakeholders being responsible for different elements of the business. Amongst these entities are statutory authorities including the Department for Transport and Office of Rail Regulation, standards bodies such as the Rail Safety and Standards Board (RSSB), the train operating companies, rolling stock leasing companies, infrastructure operator, passenger groups, unions, and countless contracted maintenance and construction companies. Each stakeholder group has different interests in the network, operating practices, and levels of summary at which they need to display various items of data. Some of them, such as the train operating companies, may be in competition with each other and be unwilling to share information. Although the British system is more complex than most, railways throughout Europe (and the rest of the world) are tending to follow this trend.

2.3 Historical Practices

The safety critical approach which must be taken when dealing with any new addition to the railway means that there are numerous practices, procedures, data gathering methods, and data storage systems that must be maintained. Many of these were created before computing became main stream and present difficulties, both technical and financial, that must be overcome to create a unified data transfer mechanism. Particular challenges are posed by records that are only kept on paper, and information exchanges between stakeholders that are performed over the telephone. Currently, this information is essentially lost and can not be used in the future to inform business decisions.

2.4 Nomenclature

While the use of terms within the railway industry in a particular country are usually consistent, this is not necessarily the case when considering railways in different parts of the world. The terminology used in the British and US railways for example have a number of major differences. In Britain a shunt is an operation when coaches or trucks are moved from one track to another, usually to change the formation of a train; the equivalent US term for shunting is switching. While ambiguity of this type can usually be overcome by humans, it can be a very significant problem when automatically exchanging data between computer systems; an XML file with a tag “shunt” generated in Britain for example, may not have the same meaning as the tag in a file that originated in the US.

Although not strictly an issue on nomenclature, an important point when considering the difference between the British and European railway networks lies in units of measure. While European railways tend to be measured in metric units, the British system is measured according to imperial miles and chains for infrastructure purposes. Uncommunicated differences in units of measurement have been known to have catastrophic effects in the past, for example in the loss of the Mars Climate Orbiter (Mars Climate Orbiter Mishap Investigation Board, 1999).

3 ONTOLOGIES AS A DATA STANDARD FOR THE RAIL INDUSTRY

While it is clear that something must be done to improve data exchange and sharing within the industry in order to bring about the smarter railway, what that action should be may be less clear. It is unlikely that any one solution will be a “magic bullet”, resolving all the factors outlined above, but what are the options? Ruling out the replacement of current systems on a large scale, mainly on the basis of cost and the rigorous requirements for ensuring safety within the sector, we are left with creating a data transfer mechanism that is capable of meeting the needs of all the current solutions. From a technical perspective, ontologies have a number of advantages to offer: data transferred according to an ontology retains its context, making it possible for machines to reason on it and for intelligent software agents to perform a wide range of data processing tasks autonomously. Ontologies are also easily extended by individual companies to meet their own needs, an important consideration when they are in competition with each other and may need to protect elements of intellectual property. Finally, software that communicates using an ontology can interact with other packages as part of the semantic web.

3.1 Business Case

In the British railway industry, developing a solid business case for a major project is second only in importance to safety concerns. The following sections will discuss the potential advantages of an ontology-based data exchange standard to the industry.

3.1.1 Condition Monitoring and Predictive Maintenance

The benefits of the use of ontologies with regard to condition monitoring and predictive maintenance in railway vehicles has already been demonstrated in work done as part of the InteGRail project (Lewis and Roberts, 2010). Currently, maintenance in the British railway industry is performed according to a schedule; vehicles are taken to a depot and examined at specified intervals, irrespective of whether there is an apparent problem at the time. A lack of condition monitoring data means that failures, when they happen, are dealt with using a firefighting approach. As part of the LiveTrain project, much more condition monitoring data on older vehicles will become available. By transferring condition monitoring data according to an ontology model, ontological inference can be used to generate a consistent set of data items from vehicles with differing instrumentation sets. From this position it is much easier to generate metrics for the prediction of faults, since more data of a consistent form is available to train the classifier being produced. If it can be shown that the classification algorithm can predict faults a sufficiently long period before failure that maintenance work can be performed, then the scheduled maintenance processes, which are costly and potentially unnecessary, can be safely reduced in frequency.

The ability to predict the failure of vehicles should also lead to a reduction in the number of in-service failures. On the British railway network train operating companies are charged by the minute for late-running services; as approximately 20% of delay minutes for the year 2006-07 were attributed to fleet causes (Department for Transport, 2007), a reduction in breakdowns has important financial implications for the train operators. These are aside from the customer service, timetabling and logistics difficulties that arise from a train being out of position.

3.1.2 Decision Support

While the large-scale gathering of information by projects such as IIS and LiveTrain is an important step towards the smarter railway, its value is never realised if it is not used by an organisation to effectively inform its business decisions. This is a complex issue with many different facets that must be considered; individuals with different roles within the industry may, for example, require radically different views of the information available to perform their assigned tasks. Data volume is also an important issue, and the information must be filtered and summarised so that humans, who are ultimately responsible for the decision-making process, are not overloaded with information that is not relevant. In a system where data is transferred according to an ontology model, the context of the data is retained. As a result of this, the combination of elements of the data for particular tasks could ultimately be performed by software agents, making the software applications themselves simpler and more easily extended and maintained.

3.2 Interoperability

While much of the current European rail network has evolved as a set of national systems, the realities of the European Union have meant that travelling across Europe is now much easier. With that in mind, the European Parliament has, for the last 10 years, been passing directives designed to encourage interoperability between the nation rail networks (European Parliament, 2001). The issues of nomenclature, units of measurement etc. outlined above make an ontology-based data transfer standard essential here, as the exact nature of the data is recorded in a machine-interpretable format, allowing appropriate steps to be taken in software to ensure that data is presented and processed using the correct scales.

3.2.1 Capacity Planning and Timetables

Railway scheduling on a national network is a complex task; among the factors that must be considered are demand for a service, peak travel times, track and vehicle maintenance, speed limits, load limits, connecting services, timings of other services on the same line, and overall journey times. The situation is even more involved when considering cross border routes, here the individual responsible for scheduling the service must have information on the railway network in each country the proposed service will pass through. Changes to local services in a country could easily have knock-on effects on services passing through it, although the way services and infrastructure are described will almost certainly change in the various national computer systems; this presents a clear need for a common mechanism for sharing route and scheduling information in a timely and unambiguous manner.

3.3 Customer Experience

The potential effects of an ontology-based data transfer standard for the rail industry on the customer experience can not be understated. While many of the business case benefits outlined above do, of course, impact on the customer experience (interoperability between networks, reduced delay minutes, and improved reliability), the most marked improvements will come through the rise of the semantic web. As part of the smarter railway, customers wishing to book train tickets to travel to a meeting for example, will only need to inform a software agent of their intended destination, and it will be able to find tickets for them. This might involve negotiating with the customer's diary and interpreting the timetable to find the best time for their journey, arranging for overnight accommodation in their preferred hotel, and booking a table for dinner. As the customer travels, agents will keep track of their progress and be able to adjust their itinerary if they are delayed or miss a connection, updating the time of the meeting if that fits with the diaries of the other parties attending. Obviously the smarter railway is only a small part of the semantic web vision, but it is an important element nonetheless.

4 THE CREATION OF AN ONTOLOGY TO MODEL THE RAIL INDUSTRY

When considering a modelling task of this scale, there are a number of important factors to be addressed. These include practical issues such as the sources of information available and compatibility with existing initiatives, along with the more strategic questions of how to get enough support from users to achieve a critical mass and force large-scale adoption of the model. The remainder of this paper discusses some initial thoughts on these matters, before going on to identify the key areas that should be captured in an initial model.

4.1 Industrial Support

As any data transfer standard will require buy-in from large proportions of the industry in order to be successful, it is vital that it reflects the needs of as large a group of the rail stakeholders as possible. To that end, it is suggested that a consortium be formed, consisting of members of rail governing bodies, operating companies, developers of existing standards and industrial partners to help guide the creation of the final model. In order to promote a sense of ownership within the industry, as much of the development work as possible should follow a collaborative approach, potentially even to the extent of using a community-based ontology engineering methodology like DILIGENT (Pinto et al., 2004b; Pinto et al., 2004a; Tempich et al., 2007). For collaborative modelling efforts to be effective however, it is desirable to have an initial "straw man" that can be used to guide the subsequent discussions. As such, in advance of the formation of a consortium, work has begun in Birmingham to investigate the available sources of domain knowledge.

4.2 Sources of Domain Knowledge

A number of sources of domain knowledge, aside from the expertise of the consortium members, are already in existence within the rail industry that could be drawn on to speed the creation of an initial model. These include, but are not limited to, existing ontology models, XML data standards, technical reports, and legislative guidelines.

Existing Ontologies

The InteGRail project (Fischer et al., 2009; Lewis and Roberts, 2010) was a European FP6 programme with a brief of "integrating current and future railway information systems across the European nations in an effort to achieve improved overall efficiency and performance, particularly in future European railway endeavours". As part of the project, a basic condition monitoring ontology for wheel impact faults was created, which could reason about the severity of reported faults and classify a vehicle status accordingly.

The ISO 15926 standard (Batres et al., 2007) was initially developed for the integration and exchange of information relating to process plants including oil and gas production facilities. ISO 15926 takes a very "ground-up" approach to the modelling of processes, allowing each piece of equipment to be described in terms of its component parts as well as its temporal existence. Recently, an Ontology Web Language (OWL) implementation of the ISO 15926 standard has been produced, which could serve as an upper-level ontology for the project.

Data Exchange Standards

The RailML (Nash et al., 2004) initiative provides an XML based approach to transmitting and receiving rail related data. The proposed standard covers a range of rail terminology, including vehicles, subsystems and infrastructure, and provides not only a good source of domain concepts and instances, but also basic hierarchies.

The Maintenance Information Open System Alliance (MIMOSA) are a trade association "dedicated to developing and encouraging the adoption of open information standards for operations and maintenance in manufacturing, fleet, and facility environments". Among the family of standards they produce is the Open System Architecture for Condition Based Maintenance (OSA-CBM) (Lebold and Byington, 2002), an ISO 13374 compliant data transfer architecture already being considered for use within the British railway industry. While the MIMOSA standards use their own data model, and may therefore not be suitable for direct use within the project, they do provide a good model for the implementation of an ISO 13374 complaint system.

Standards Bodies

In Britain there are several organisations that issue standards (both general and rail specific). The RSSB publish a set of safety related standards collectively referred to as the Railway Group Standards (RGS). These include approved codes of practice, guidance notes and railway industry standards. The British Standards institute (BSi) also issue a broad range of railway-applicable standards, some specific to the rail industry, others relating to broader topics. Many of these standards are closely related to standards issued by the International Organization for Standards (ISO).

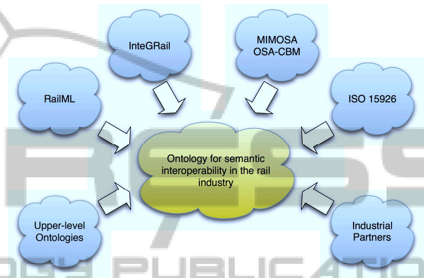

Figure 1: Sources of knowledge for a rail domain ontology.

4.3 Priorities for an Initial Model

The scale of the modelling task proposed by this paper should not be underestimated, and it is inevitable that certain areas of the domain would need to be prioritised during the development of any ontology. Fortunately, within the rail industry a clear group of topics exist that not only represent the minimum requirements for a useful data transfer standard, but would also facilitate a significant proportion of the business benefits discussed earlier; those being vehicles (both as individuals and as train formations), infrastructure (track types, operating speeds, clearances, power supply and signalling) and timetabling (including the scheduling of services).

Although these topics still represent a substantial implementation effort, they are all required in order to describe many events on the railway network. As an example consider the activation of trackside monitoring equipment such as a WheelChex device. The WheelChex system itself is part of the infrastructure, consisting of a section of instrumented track at a particular location. The defective wheel which activated WheelChex would be part of a particular train formation, but in order to uniquely identify it reference would have to be made to both the timetable for the line and to the details of the train formation so the individual vehicle (coach, wagon, locomotive etc.) could be singled out. Once the vehicle had been identified, the tracks which it had run over since it last successfully passed a WheelChex device (during which it may have been part of a number of different train formations) would need to be determined so that they could be checked for damage, again requiring access to the timetable, records of train formations, and the details of the tracks.

As the example shows, information on vehicles, infrastructure and timetabling are sufficient to report the activation of lineside equipment. Information from the same topics would also allow the scheduling of services to be performed, where the vehicles proposed for use must be checked for compatibility with the infrastructure over which they are to run (clearances to lineside equipment and other vehicles, track gauge, availability of power), and an appropriate free timeslot must be identified in the timetable. It would also allow basic semantic web services to be created, improving the customer experience.

On completion of the initial model the first phase of expansion should include information on maintenance practices, allowing the correct responses to the activation of WheelChex and similar equipment to be determined, and details of the various roles that employees have within the industry, enabling appropriate information to be displayed for each worker. This could be followed later by business logic, which although providing the key financial benefits to the industry, would serve little purpose without the data modelled by the earlier phases.

5 CONCLUSIONS

An ontology-based standard for data transfer would have much to offer the rail industry, allowing existing, largely incompatible, legacy systems to exchange information in a meaningful way, without the need for costly changes that would potentially pose a risk to safety. More importantly, it would provide a framework on which modern, semi-autonomous processing agents could be built, improving the efficiency of the railway network, reducing the risk of human errors in mundane tasks, and enhancing the experience of the travelling public. This paper has attempted to present the arguments for, and the challenges to be addressed during, the creation of such a standard; along with some initial thoughts on the methods that could be adopted for its development.

REFERENCES

Batres, R., West, M., Leal, D., Price, D., Masaki, K., Shimada, Y., Fuchino, T., and Naka, Y. (2007). An Upper Ontology Based on ISO 15926. *Computers & Chemical Engineering*, 31:519–534.

Department for Transport (2007). Delivering a sustainable railway. Technical report, Department for Transport, http://www.dft.gov.uk/about/strategy/whitepapers/whitepapercm7176/.

European Parliament (2001). Directive 2001/16/EC of the European Parliament and of the Council of 19 March 2001 on the interoperability of the conventional rail system. Technical report, European Parliament, http://europa.eu/legislation_summaries/internal_market/single_market_for_goods/technical_harmonisation/l24229_en.htm.

Fischer, J., Roshchin, M., Langer, G., and Pirker, M. (2009). Semantic data integration and monitoring in the railway domain. In *Proceedings of the 10th IEEE international conference on Information Reuse & Integration*, pages 11–16.

Lebold, M. and Byington, C. (2002). OSA-CBM architecture development with emphasis on XML implementations. In *Proceedings of the 2002 Maintenance and Reliability Conference (MARCON)*.

Lewis, R. and Roberts, C. (2010). Using non-monotonic reasoning to manage uncertainty in railway asset diagnostics. *Expert Systems with Applications*, 37(5):3616–3623.

Mars Climate Orbiter Mishap Investigation Board (1999). Mars Climate Orbiter Mishap Investigation Board Phase 1 Report. Technical report, NASA, ftp://ftp.hq.nasa.gov/pub/pao/reports/1999/MCO_report.pdf.

Nash, A., Huerlimann, D., Schuette, J., and Krauss, V. (2004). *Computers in Railways IX*, chapter RailML: a standard data interface for railroad applications, pages 233–242. WIT Press.

Office of Rail Regulation (2009). National Rail Trends: 2008-2009 Quarter One. Technical report, Office of Rail Regulation, http://www.rail-reg.gov.uk/upload/pdf/382.pdf.

Pinto, H., Staab, S., and Tempich, C. (2004a). DILIGENT: Towards a fine-grained methodology for DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)*, pages 393–397.

Pinto, S., Staab, S., Sure, Y., and Tempich, C. (2004b). OnToEdit empowering SWAP: a case study in supporting DIstributed, Loosely-controlled and evolvInG Engineering of oNTologies (DILIGENT). In *Proceedings of the 1st European Semantic Web Symposium (ESWS 2004)*, pages 16–30.

Tempich, C., Studer, R., Simperl, E., Luczak, M., and Pinto, H. (2007). Argumentation-based ontology engineering. 22(6):52–59.