View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Birmingham Research Portal

UNIVERSITY OF BIRMINGHAM

Research at Birmingham

From Data to Information: Provision of Railway Data to Passengers in the Information Age

Morris, Christopher; Easton, John; Roberts, Clive

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Morris, C, Easton, J & Roberts, C 2016, 'From Data to Information: Provision of Railway Data to Passengers in the Information Age' Paper presented at 11th World Congress on Railway Research, Milan, Italy, 29/05/16 - 2/07/16, .

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of ‘fair dealing’ under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 01. Feb. 2019

11th WCRR
World Congress on Railway Research
2016
29th May – 2nd June
Milan - Italy

FROM DATA TO INFORMATION: PROVISION OF RAILWAY DATA TO PASSENGERS IN THE INFORMATION AGE

Christopher Morris[1], John Easton[1], Clive Roberts[1]

[1] 1Birmingham Centre for Railway Research and Engineering, University Of Birmingham, Birmingham, United Kingdom

*Contact: chris@morrisdigital.co.uk

Abstract

A series of rapid developments in ICT over the last 30 years have fundamentally changed the ways in which we work & spend our leisure time. The advent of affordable personal computing (mid 80s), the growth of the World Wide Web (late 90s), & the ability to receive / digest information on mobile computing devices including tablets & smartphones (last 5 years), has allowed us to make decisions more effectively based on up-to-date real-time information on the state of the world we live in.

In the transport industry arguably the greatest impact of the new information-driven technologies has been in the area of passenger information, & in particular in journey planning. The railways, in common with other transport modes, have made good progress in the delivery of online timetables / journey planners, real-time services status reports & electronic ticketing to customers, but the real value of these tools will not be realised until a customers' entire journey can be captured across transport modes; information services of this type will enable the impact of disruption to service on one transport mode to be viewed in the context of knock-on effects to the rest of the journey, allowing the passenger to understand what the disruption means to them, & opening up new possibilities for use of the multimodal transport network as a single system.

The integration of data from across a number of transport modes is a challenge from the perspective of data modelling, as the meaning of terminology is often subtly different in each mode. As a result, large amounts of developer time can be spent ensuring that data is being used in the correct context, & reflecting that context in application code. The hard wiring of data context in this way means that applications frequently need to be updated in response to changes in one or more of the domain-specific data models.

This paper presents work on a new family of data models (ontology) designed for use with the next generation of the Web (the Semantic Web) that capture the context & meaning of data in a machine interpretable form, allowing information to be used without ambiguity across transport modes. It shows how, in combination with linked data architectures, ontology allows data resources to evolve over time without the need for substantial changes to the applications using them. Finally, it describes ways in which railway undertakings can publish their data to ensure maximum reusability in next-generation web applications.

Introduction

Linked data and ontology have been shown to be an effective means of storing knowledge and taking decisions across many domains. In this context ontology is defined by (Gruber 2009) as "a set of representational primitives with which to model a domain of knowledge or discourse". It may be thought of as a computational representation of knowledge. Linked data is less precisely defined; Sir Tim Berners-Lee (Berners-Lee 2006) suggests linked data that use three Uniform Resource

Identifiers (hereon URIs) to record the relationship between two items. Most ontologies, within the computer science definition of the term, rely upon linked data however not all linked data need be part of an ontology.

The benefits to the rail industry of using linked data with when confronted by ever increasing amounts of data from many disparate sources (Technical Strategy Leadership Group 2012) are shown by studies such as (Verstichel et al. 2011) or (Köpf 2010). These studies showed that there was a benefit to integrating that data, even when the benefits of integration didn't fall on the same party as the costs.

One of the key benefits of improved data integration is customer information, as shown by (Verstichel et al. 2014) ontology can enable advanced multi-modal passenger information systems which may be tailored to the users' needs.

Wider benefits of data integration are discussed in (Easton et al. 2013) often enabling other work, notably predictive maintenance, as discussed in (Umiliacchi et al. 2011). Benefits of improved data integration extend into the multi modal transport domain, where there are many different data sources need to be integrated, often from competing commercial concerns. For this reason previous work exists using ontologies for data integration across transport modes. The ArkTRANS model is the first large scale attempt to integrate data across transport modes, developed by the Norwegian ministry of Transport and Communications, as discussed in (Natvig & Westerheim 2003). ArkTRANS has been taken as the basis for numerous future other projects, such as MultiRIT (Natvig & Westerheim 2008) which has applications in the multimodal transport domain. Other applications include the shipping domain. More recently the General Transit Feed Specification (hereon GTFS) has proven popular for multimodal transport data integration. The reasons for extensive uptake of the GTFS along with a though discussion of it implementation details may be found in (Santos & Moreira 2014). GTFS is a very simple, unlinked, data specification however a linked variation has been produced as discussed by (Colpaert et al. 2015). The road domain also has a requirement for data integration and has also looked to ontology for a solution, as discussed in (Seliverstov & Rossetti 2015).

Another key benefit of using ontology is decoupling of the decision making process from the implementation. A much touted benefit of good object orientated software design is separation of data and implementation; use of ontology takes that a step further and abstracts the decision. This is discussed eloquently and at a high level in: (Vanthienen & Caron 2013). Using ontological reasoning the decision making logic can be entirely separated from the software, allowing an entire new (more or less complex) decision making process to be added without development working being undertaken. Linked data and ontologies retain the advantage that relational databases have of separating the data from implementation, in many cases taking it further, as the data is self-describing. The study "Ontology-based data management for the GB rail industry Feasibility study" (referred to hereon as FuTRO) undertaken at this centre (Tutcher et al. 2013) goes some way to demonstrate this, by presenting cases studies in the context of the railway industry. Further work undertaken by industrial partners from the FuTRO project produced an extension to the system for monitoring the condition of railway points machines which, without changes to the front end, included circuit breaker condition monitoring.

Integrated multimodal transport

In multimodal transport data from different transport modes, including both planned and actual time tables along with interchange time and disruption information is relevant, as is cost to the customer. The TraPIST project outlined in (Verstichel et al. 2014) is a notable step forward in integration across transport modes, which produced, as a demonstrator, a journey planning tool for disabled users. The demonstrator takes account of users' needs, including disability, familiarity with transport interchanges and luggage then produces a route personalised to them. The framework upon which the demonstrator was built is intended for use by others as a stepping stone to further work. Amongst the goals of this project was dealing gracefully with the disappearance of individual data sources – the best available data should always be delivered.

Flexibility with respect to available data resources was also demonstrated by the FuTRO project[1] - in that instance using sources of differing accuracies for train location and seamlessly falling back to a less accurate technique when the more accurate data was removed. Other work in the area of multimodal transport focused on disruption, notably the transport disruption ontology (Corsar et al. 2015) which was employed as part of the social journeys project.

The FuTRO project (Tutcher et al. 2013) "shows how ontology will enable the industry to adopt a "build once" approach to application development, protecting software from changes to physical systems in the real world.". It does this by constructing an ontological model of the operating details of the railway, which enables it to be used not just for customer information, but also more general tasks including as an enabler to predictive maintenance or technical drawing retrieval.

The RaCoOn Ontologies – A case study from FuTRO

Aside from the train location demonstrator as discussed in the introduction the "Future Traffic Regulation Optimisation" project included collaboration with industrial partners on points condition monitoring. The use of ontology and reasoning made it possible to infer which track was inaccessible in the case of a failure. For an introduction ontological reasoning see (Horridge et al. 2011) which also discusses the use of a tool for ontology construction. The system was designed such that it could infer whether a points machine was faulty or not according to arbitrary criteria which could be altered at run time, without need to make changes to the software produced. As such if a new model of points machine came along with different criteria for failure (for example oil pressure rather than current draw) then altering the criteria would not require further development. This is one of the key advantages of using ontological reasoning.

The model used to store the data in the FuTRO project was the railway core ontologies, here on RaCoOn. These are a set of related ontologies that aim to model the rail domain in depth. They consist of a number of modules, not all of which need to be used for a given task. The modules break down as follows, as illustrated in Figure 1:

- An upper level ontology which holds the highest level concepts, which apply across domains. This can be mapped directly to other cross domain ontologies (in particular: DOLCE ultralite) for compatibility;
- A core ontology which holds the high level rail concepts which apply throughout the rail domain such as "Rolling Stock Concept" which can then be extended by lower levels
- Task Ontologies: rolling stock, time tabling and infrastructure have so far been developed, though all need more detail adding. These will model specific sub domains in depth;
- Application ontologies. These are the lowest level of detail and not needed in all instances, but can hold information relevant to specific applications.

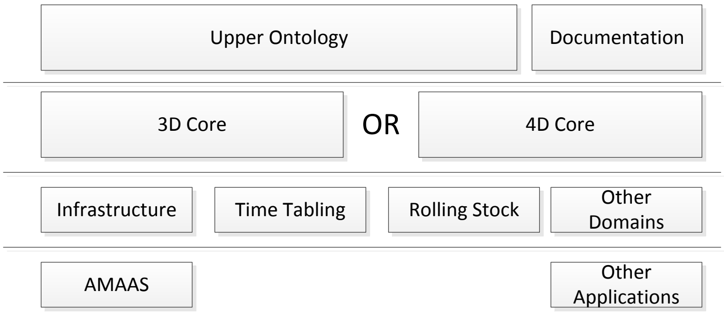

Figure 1 - Module Structure

---

[1] Available at http://purl.org/rail/trainlocator

The purpose of this modular structure is to allow developers to use only the required parts of the model. This has advantages in terms of computational workload: the larger the model the longer it takes to reason over it. All modules depend on those above them, but none depend on those below them or at the same level. As such the upper level ontology must always be used alongside one of the core ontologies of which two variations are presented for different scenarios, and then one or more of the task ontologies can be used if necessary. The core ontologies are presented as three or four dimensional – with different representations of time. The representation of time is directly relevant to the rail domain, for example: An individual loco of class 390 is built at a certain time, for some of that time it is part of a given train-consist and at this particular moment is hauling the 17:30 Birmingham New Street to London Euston. For the purposes of most tasks in the rail domain an endurant view is sufficient: the loco changes with time, rather than being a series of different objects at different moments in time. Whilst both these views have advantages where there are high volumes of data endurants are easier to reason over and thus they and the three dimensional version of the core ontology are generally preferred. ISO15926, which has been used in the oil and gas industry for representing plant specifications uses the 4D approach (Leal 2005), which is more suitable for very accurately modelling complex systems, but becomes challenging when large volumes of data are added.

The ontology aims to model the rail network in depth such that it can be presented to different users at different levels of abstraction: a maintenance engineer needs to know the longitude and latitude of a broken signal and the nearest access point such that he can drive to it by road, whilst a system for time table planning needs to know that between certain nodes of the rail network there are passing loops. Conversely a customer information system is interested chiefly in if and when a given train will reach a station.

In order as for the ontology to be relevant it needs to model the rail domain in depth. Currently the domain is only partially modelled, in particular detail is lacking at the lower levels. Work to remedy this must currently be done by people with knowledge of both the domain and of ontological modelling software; since there are few experts who meet these criteria this is not a sustainable way to model the entire domain in depth. To date this issue has been addressed using tools to automate the conversion of data from other formats. Whilst that approach has value – much knowledge about the rail domain exists, locked away in many and varied proprietary formats, it also has weaknesses: often the ontology produced is of a lower standard than one produced by hand containing redundancy and poor structure. An alternative approach explored here is that of creating tools to allow domain experts, who are not ontological modelling experts to edit the ontology.

Tool for extending RaCoOn

The objective of this tool is to allow any railway engineer with no more than basic IT literacy to add data to the ontology. In accordance with the wishes of UK's infrastructure operator the workflow of the data entry system would be thus:

User selects the subdomain of the asset to be modelled

The list of available subdomains is generated from the based on the task ontologies available. Meta data from within the ontologies, including RDFS:Label tags, will be used to present the user with the available sub-domains.

Using free text search the user selects the type of the asset to model

All available types containing the search term in the selected sub domain are listed. The previous search is stored since it is envisaged that multiple assets of the same class will be added at the same time.

The relationships suggested by the model are presented to the user

The user is prompted to enter all the details the model expects, including creating further related objects where they do not yet exist. If the object is inconsistent with the model this is highlighted and the user cannot store the object until it is rectified. The user is also free to enter any other relations that may be needed, so long as the provision for them exists in the model.

The changes automatically have provenance tags added to them

The name of the current user is recorded along with the time of the changes, as a minimum.

At a higher level the architecture may be broken into three parts:

The front end: This system presents a webpage that the user interacts with.

Middleware: This system presents web services to the front end and, crucially to other systems allowing integration.

Data store: A range of systems store data with the middleware selecting the appropriate one. Notably for storing triples i.e. linked data Stardog[2] is used.

Conclusions and Further work

Ontology and linked data can bring benefits to many stakeholders within the rail domain; passengers get better information whilst other stakeholders can expect reduced costs and improved decision making. Beyond the rail domain benefits accrue to uses of all modes of transport – data integration is necessary both within and between many different transport domains.

The tool for extending racoon will increase the coverage of the ontology thus making it possible to realise many of the benefits ontology has brought other domains within the rail domain. Adding detailed coverage of points machines for example, makes Failure Mode Effects Analysis possible in that sub-domain and this may be applied across the many subsystem that comprise the railway network.

Further modelling work needs to be done, both by converting existing data and making it possible for experts to model their knowledge directly.

References

Berners-Lee, T., 2006. Linked Data. Available at: https://www.w3.org/DesignIssues/LinkedData.html [Accessed January 14, 2016].

Colpaert, P. et al., 2015. Intermodal public transit routing using Linked Connections. In *International Semantic Web Conference*. pp. 5–8.

Corsar, D. et al., 2015. The Transport Disruption Ontology. In *The Semantic Web-ISWC 2015*. Springer, pp. 329–336.

Easton, J. et al., 2013. *RSSB Cross industry Railway Information Systems Workshop 7 th December 2012*,

Gruber, T., 2009. Ontology. *Encyclopedia of Database Systems*.

Horridge, M. et al., 2011. *A practical guide to building OWL Ontologies Using Protege 4 and CO-ODE tools*, The University Of Manchester. Available at: http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf.

---

[2] http://stardog.com/

Köpf, H., 2010. *InteGrail – Publishable Final Activity Report*,

Leal, D., 2005. ISO 15926 “Life cycle data for process plant”: An overview. *Oil and Gas Science and Technology*, 60(4), pp.629–637.

Natvig, M. & Westerheim, H., 2008. Refinement and evaluation of the ARKTRANS framework through use in travel information services. *15th World Congress on Intelligent Transport …*, pp.1–12. Available at: http://trid.trb.org/view.aspx?id=905537 [Accessed June 20, 2014].

Natvig, M.K. & Westerheim, H., 2003. JOINT EFFORT ON ESTABLISHMENT OF ARKTRANS – A SYSTEM FRAMEWORK ARCHITECTURE FOR MULTIMODAL TRANSPORT. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.2647&rep=rep1&type=pdf [Accessed March 5, 2014].

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

Seliverstov, A. & Rossetti, R.J.F., 2015. An Ontological Approach to Spatio-Temporal Information Modelling in Transportation. In *Smart Cities Conference (ISC2)*. IEEE, pp. 1–6.

Technical Strategy Leadership Group, 2012. *THE INDUSTRY ' S RAIL TECHNICAL STRATEGY 2012 THE RAIL TECHNICAL STRATEGY 2012*,

Tutcher, J. et al., 2013. *Ontology-based data management for the GB rail industry Feasibility study*, Available at: http://www.sparkrail.org/_layouts/Rssb.Spark/Attachments.ashx?Id=75NEMTS3ZVHP-8-7892.

Umiliacchi, P. et al., 2011. Predictive maintenance of railway subsystems using an Ontology based modelling approach. , pp.22–26. Available at: https://osl.eps.hw.ac.uk/files/uploads/publications/WCRR2011 Paper final v05.1.pdf\nhttp://osl.eps.hw.ac.uk/files/uploads/publications/WCRR2011 Paper final v05.1.pdf.

Vanthienen, J. & Caron, F., 2013. Modeling Business Decisions and Processes - Which Comes First? *{KDIR/KMIS} 2013 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval and the International Conference on Knowledge Management and Information Sharing, Vilamoura, Algarve, Portugal, 19 - 22 September, 2013*, pp.451–456. Available at: http://dx.doi.org/10.5220/0004623904510456.

Verstichel, S. et al., 2014. LimeDS and the TraPIST project : A Case Study An OSGi-based ontology-enabled framework targeted at developers in need of an agile solution for building REST / JSON-based server applications PoC demonstrator. , 2(Ic3k), pp.501–508.

Verstichel, S., Ongenae, F. & Loeve, L., 2011. Efficient data integration in the railway domain through an ontology-based methodology. *… Research Part C: …*, pp.1–64. Available at: http://www.sciencedirect.com/science/article/pii/S0968090X10001609 [Accessed February 18, 2014].