The "open web of transportation data" as a shared networked environment for the realization of the Single European Railway Area

Riccardo Santoro[1], Eric Lavillonnier[2], Clive Roberts[3], Paolo Umiliacchi[4]

[1] Ferrovie dello Stato Italiane spa, Rome - ITALY

[2] Mitsubishi Electric Research Centre Europe, Rennes – FRANCE

[3] University of Birmingham, Birmingham – UK

[4] CNC Centro Nuova Comunicazione, Bologna - ITALY

r.santoro@fsitaliane.it

Abstract

Railways connect people, cities and regions which are increasingly being instrumented with ICT capabilities, where everything and everybody is a sensor, a data producer and consumer, has local storage and processing capabilities, and is interconnected in an open-ended network of Railway Operators, Retailers, Travel related Service Companies, Public Agencies, Businesses and Individuals which expands driven by market dynamics.

In this paper we describe "passenger-centrality" as a statement of the fact that passengers are consumers of 'network products' that exhibit network externalities. We describe the specific dynamics of the markets for these products, showing that their self-sustainability is maintained only beyond a minimum threshold and below a maximum limit of their size, both of which depend on the cost of compatibility between complementary products for a given network size. We show that for information-intensive products this cost is one of interoperability of heterogeneous systems, and that their main driver is the problem of semantic heterogeneity. We describe a *semantic* interoperability approach to the challenge which leverages open data policies and semantic web technologies for the creation of an "open web" of transportation data, illustrating it with an example.

1. Problem description

1.1. The passenger at the centre of the transport system

The growing mobility demand from citizens in Europe and Worldwide is pushing the development and evolution of transportation systems, including railways, aviation, public transport, private services and vehicles. In parallel, travellers' needs and expectations are also growing, creating new challenges for transportation service providers, driving them to change from a traditional focus on transport assets (vehicles, infrastructure, staff, operations), putting the passenger at the centre of their systems. Transportation services need to go beyond the "travel" concept, moving towards a "full service model".

Ideally, passengers would like to organise, book and pay for their complete journey, from door to door, by means of a unique interface which allows them to express their needs from their own point of view (e.g. "I wish to attend WCRR") and which provides a complete journey solution, hiding all the multiplicity, heterogeneity and complexity of the interactions needed to achieve it. Moreover, passengers expect to be guided during each step of their journey, with prompt and automatic support in case of any disturbances, such as delays, service interruption or missed connections. Finally, passengers wish to receive seamless support in case of after-journey problems like lost luggage, refunds or changes of return plans.

European Regulations are driving such evolution, enforcing passenger rights before, during and after the journey [1], giving requirements for passenger oriented services and technical specifications for interoperability [2] and indicating specific objectives like integrated multimodal travel information, planning and ticketing services [3]. Such targets are far from being achieved today, due to the cost of integrating an open-ended network of Operators, Retailers, Travel related Service Companies, Public Agencies, Businesses and Individuals, that changes and expands driven by market dynamics.

ICT can offer a solution, but integration in a market context according to a self-sustaining business model is necessary to make it successful. Otherwise forcing a solution and its adoption can easily result only in partial and unsatisfactory results and huge waste of public money.

1.2. Passenger-centricity and Network Products markets

In this paper we restate the principle of "passenger at the centre of the transport system" in terms of the economics of network industries [9][10][11] in order to highlight the role and impact of open data and interoperability solutions in addressing certain specific issues in this context.

In this perspective we recognize that a transport system is 'passenger-centric' when it supplies mobility services as products that are complementary to other non-transport services that the passenger can source from specialized providers and assemble into personalized solutions that satisfy needs that originate in the activities passengers pursue in their daily life. On the other hand, mobility services are themselves combinations of complementary products supplied by multiple specialized mobility providers, the obvious example being 'multi-modal' services. We state, therefore, that in a passenger-centric view the passenger is a consumer of services which exhibit the features of *network products* [10], and that the ability of the transport industry 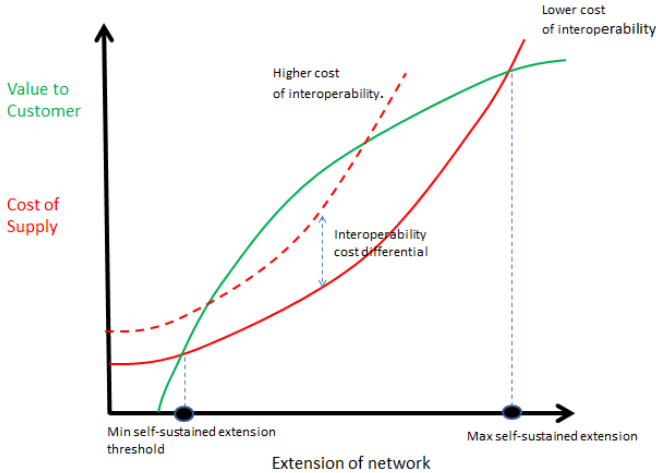to match their demand depends on an understanding of the particular characteristics of the markets that supply them. While a comprehensive discussion of network industries is beyond the scope of this paper, we do concentrate on the particular effect of the presence of externalities in markets for network products, i.e. the fact that the value of a supplied product which is complementary and compatible with others increases as a function of the availability of other such products that can be combined with it to form a solution, i.e. with the extension of a 'network' of such products. We illustrate the phenomenon and its consequences with the qualitative model in Figure 1.

Figure 1 - Network product market dynamics

Network externalities are shown in the fact that the "value to Customer" (demand) increases with the extension of the network beyond a non-zero value, providing an incentive for suppliers to extend it wherever the cost of supply (dashed red line) is lower than customer value. In this diagram two equilibria are shown where the curves intersect:

- A minimum self-sustained extension threshold which is unstable: a smaller network than this becomes smaller whereas a larger one becomes larger under market dynamics (i.e. self-sustainability).

- A maximum self-sustained extension which is stable: a smaller network than this becomes larger and a larger one becomes smaller around this point[1].

We note that because the minimum threshold is unstable under market conditions, non-market forces are required to extend the network beyond the minimum, and because the maximum extension is stable under market conditions, non-market forces are needed to extend it beyond the maximum, both in the form of some Public Authority intervention.

If we consider a different cost of supply curve which is lower for every value of the network's extension (solid red line) we see that the equilibria are shifted: the minimum self-sustained extension threshold is lower and at a lower cost of supply, and the maximum self-sustained extension is higher and at a higher value to customer than in the previous case, allowing for a decrease of non-market forces needed to establish or extend the network. In order to benefit from these effects, we may therefore study conditions that can lower the *cost* of supply for a given *degree* of product complementarity and compatibility reflected in the given value to customer curve[2].

In the case of information-intensive products we may consider that the cost of supply is largely determined by the cost of interoperability between distributed and heterogeneous information systems.

1.3. Semantic Interoperability

Interoperability refers to the ability of devices or systems to participate in the coordinated performance of tasks and functions in the execution of some business process, exchanging data as means, but not as the purpose of interoperability itself. In fact, interoperability is predicated on the partners involved in the exchange of the data agreeing on the computational model that is applicable to such data and in processing them accordingly, i.e. according to some shared logical interpretation of what the data mean and what can be *meaningfully* be done with them [5]. Experience accumulated over years of attempts at making systems not originally designed for distributed computing interoperate through various forms of common 'data exchange' formats and protocols has provided ample evidence that by far the highest contributor to the costs of interoperability of such systems is the effort required to share and understand differing sets of assumptions about the *interpretation* of data, exchanged *in any format or with any protocol,* made and held by local applications – or, more correctly, by local programmers: this is the problem of *semantic heterogeneity* [12].

In this respect, while open data *policies* designed to make data accessible do constitute a significant reduction of some costs of interoperability, namely the cost of *access* to quality data, and while general standardization of data formats and protocols can also contribute to a reduction, albeit at the expense of introducing other costs, e.g. switching costs, lock-in at an inferior level of technology, lengthening of the innovation cycle, and the costs of the standardization process itself [11], addressing the problem of *semantic* heterogeneity *across* open data and 'local' standards is the fundamental engineering challenge that must be met in order to provide the necessary significant reduction of interoperability costs in ICT intensive product networks.

Another driver of the costs of interoperability that must be addressed is related to the adoption of the technology by independent operators with significant investments already encased in ICT infrastructure, tools and training, and/or with on-going such investment programs. Research in the market of network

---

[1] The presence of an unstable 'threshold' equilibrium and, beyond it, the self-sustained expansion of the network to its maximum are the defining characteristics of this market. They are determined by the 'shape' of the value to Customer curve, itself a reflection of the presence of network externalities.

[2] The same benefits can of course be obtained by increasing the value to customer for a given cost of supply. However the value to customer depends on product complementarity and compatibility, and these two parameters are affected by many non-ICT related factors such as legal frameworks or jurisdictions that may prevent compatibility. In this paper we assume, therefore, that they are *given* a certain level, de facto or desired, e.g. by Public Authorities.

products shows in fact that the adoption of interoperability technologies can be seriously limited by the need for centrally directed re-tooling or synchronization on a common roadmap, limiting therefore the extension of the network, locking it in this state and preventing innovation [10][11].

Meeting these engineering challenges requires the adoption of two simultaneous design approaches:

- the creation of a shared domain ontology, i.e. of an explicit, formal, shareable, machine-readable and computable description of the computational model associated with data descriptions and exchanges in order to allow a higher degree of automation of distributed processes *across* multiple data formats and protocols, spanning unspecified actors.
- the provision of a set of semantic interoperability services that can be deployed in multiple architectures and configurations, and that do not mandate a specific set of communication protocols or frameworks, leaving the choice of deployment strategies to partners that may opt to re-use a shared enterprise service bus, perhaps on a virtual private network protected by specific security and authentication protocols, or decide to engage in pure peer-to-peer exchanges over the public world wide web, or a mixture of these or other options. This is also important to allow operators, including yet unknown companies who are not partners in an "integration project", to choose their own roadmap for adoption of the 'native' semantic language for their exchanges, using or discontinuing the semantic transformation services according to their own timeline.

We call this design approach "semantic interoperability".

2. An interoperability framework for an open Web of transportation data

2.1. An open framework for interoperability of distributed application services

The semantic interoperability approach is now integrated in the Rolling Plan for ICT Standardization as part of the European Digital Agenda [13], albeit with a focus on e-Government, and is generally pursued in many industries and by multiple "mega-vendors'. In the Railway Industry is has been introduced with the InteGRail project [4], which generated and demonstrated a first kernel of a Railway Domain Ontology [5] . Further investigation showed how such an approach can provide an enabler for the creation of new and advanced services, which can also facilitate opening the market to newcomers and promoting innovation [6].

The semantic interoperability approach is now an integral part of the ambitious Multi Annual Action Plan of the Shift2Rail Joint Undertaking, a Public Private Partnership between the European Commission and the Industry aimed at introducing breakthrough innovation to all parts of the railway sector [7], particularly of its IP4 "innovation program" targeted at generating innovative ICT Passenger solutions for an attractive Railway. An early "LightHouse" project of the IP4 program has been initiated in May 2015, the IT2Rail project [8], which includes an "Interoperability Framework" implementation of the semantic interoperability approach described above as an enabler for the interoperability of a number of rich passenger-centric applications operating on a distributed "web of transportation data" realized as a graph of semantically annotated and linked data and service resources from any number of providers in any formats. The project incorporates results and experience InteGRail, including lessons learned and the identification of the specific challenges associated with the adoption of the technology.

2.2. A simple example of a semantic language for railways

The classical way to achieve interoperability is to define interfaces. The exchanges are done following a strict syntax defining services with messages and parameters. When everything has been agreed and is stable this type of approach is working well. However for the transportation business case, there will be many actors involved. Furthermore both these actors and their offering will change with time and we will have to minimize both the costs of rebuilding again and again *and the costs of agreement on syntactically defined interfaces*. For this purpose it is necessary to exchange "meaning", i.e. to *automate*

the 'agreeing' process, much more than it is to exchange any messages. This has an impact on how the system is built.

As an example the user could perform a query such as "I want to make a journey from Rennes to Bologna Tuesday next week. I would like to start between 10h and 12h. Which are the possibilities?"

This question is clear for every human-being, is independent of the provider *or data formats*, was valid in the Middle Ages and will still be valid in 1000 years. The ideal interface would keep it like this only formalizing it. It loses its universality when translated into specific syntactical descriptions in strings and numbers.

The first step is to define the supporting domain logic, modelling the knowledge with concepts, relationships and identifiers. For this, the Ontology is used. It enables the construction of a system without assumptions about is its usage or implementation, but only on is logical assertions.

For example, the "StopPlace" concept (this example is extracted from a demonstration Transport Ontology being developed in the IT2Rail project) which defines a localized facility giving passengers access to transportation means, i.e. the place from which a user can initiate travel on the transportation system, demonstrates one of the powerful features of Ontology: machine inference.

Figure 2 shows how machine inferencing on Ontology can 'discover' many more information about the transportation system without programming but by means of simple logic, using the knowledge entered in the system. In the lower part of the figure, we see what was entered in the system, the fact that a hub is a stop place. The upper part shows what the system is able to "understand", in fact infer. For example the system has been instructed that Airports are departing/arriving points for Air Links, so it can conclude that Airports are "StopPlaces".

The next step is to define how the different systems will interact. In a classical approach messages like this would be used:
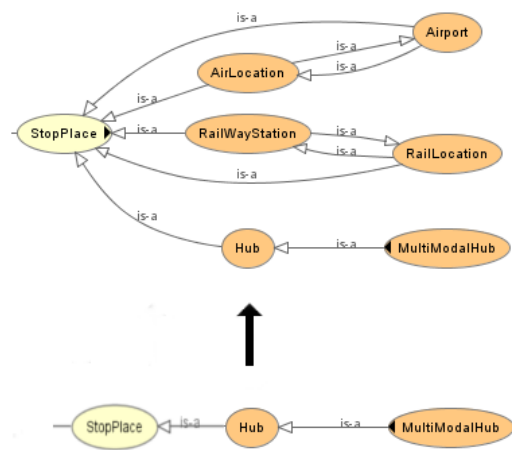

Figure 2 - Stop Place concept and inferred specializations

- (Hour hourStart = 10 , Minute minStart = 0, Hour hourEnd = 12, Minute minEnd = 0)

This interface is limited in two ways: the expression of the information is constrained and it refers to some pre-agreed knowledge necessary to understand its meaning. Another way is to exchange data using the Ontology. Simplifying the format for readability the exchange would look like:

- (date after thisWeek) (date dayOfWeek Tuesday) (date hasBeginning (startTime hour 10)) (date hasEnd (endTime hour 12))

The original request "I would like to start between 10h and 12h "is kept with less risk of error for users (who has never mistaken last year's calendar when preparing a journey?) and that expression is 'translated' into a more accurate, not necessarily numerical, but *machine* understandable formulation, in which the *meaning* of "after","dayOfWeek", "hasBeginning", "startTime" and "endTime" have been defined by W3C in the time Ontology [14] and is independent on how this meaning is encoded in a specific format.

Finally, without having to build a specific logic, only by knowing the different transport links, and the transitivity of transport function, the system will be able to indicate if there exist a transport link between Rennes and Bologna.

3. Conclusion: deployment of a railway data interoperability framework

This paper has highlighted the impact of ICT-interoperability costs on transportation providers' ability to supply passenger-centric products. It highlighted the importance of Open Data and Semantic Interoperability as essential building blocks for improving customer experience of railway networks. The critical areas for further focussed effort are identified, these are around the development of an ontology to define and model the railway data concepts and the need for specific service covering a broad range of applications. Initial work in this area has been previously undertaken as part of the InteGRail project, as well as in the current IT2Rail project; future work is planned within Shift2Rail.

At this stage further research is needed to provide demonstrators developed through the technology readiness levels to validate the effectiveness and benefits of the approach. This work does not require the development of new fundamental technologies, nonetheless it does require the extension of the focussed work on ontology, which has been already developed, and the agreement of a considered and extensible architecture that can be adapted over time and is sufficiently responsive for the service likely to be developed. These building blocks are an essential precursor to any wide scale realisation of the approaches discussed.

4. References

[1] Regulation (EC) No 1371/2007 of the European Parliament and of the Council of 23 October 2007 on rail passengers' rights and obligations.

[2] Commission Regulation (EU) No 454/2011 of 5 May 2011 on the technical specification for interoperability relating to the subsystem 'telematics applications for passenger services' of the trans-European rail system – OJEU Volume 54 – 12 May 2011

[3] P8_TA-PROV(2015)0246 - European Parliament resolution of 7 July 2015 on delivering multimodal integrated ticketing in Europe (2014/2244(INI))

[4] InteGRail – Publishable Final Activity Report – 13-04-2010

[5] J.M. Easton, J.R. Davies, C. Roberts – "The Development of a Domain Ontology for the Rail Industry" – WCRR 2011 – Lille - - May 2011

[6] Paolo Umiliacchi, Mauro Barella, Gianosvaldo Fadin, Eric Lavillonniere, Dobromil Nenutil – "Innovative platform for interoperability of multimedia applications on board trains paves the way for better services and new market perspectives" - WCRR 2013 – Sydney – November 2013

[7] Shift2Rail Joint Undertaking - Multi Annual Action Plan – November 2015

[8] Horizon 2020 - Grant Agreement No. 636078 – IT2Rail – 13-04-2015

[9] N. Economides -'The Economics of Networks', International Journal of Industrial Organization 14: 673–699, 1996

[10] O. Shy – Economics of Network Industries – Cambridge University Press, 2001

[11] H. W. Gottinger, - Economies of Network Industries (Routledge Studies in Business Organizations and Networks) (2003)

[12] V. Ventrone, S. Heiler - Semantic Heterogeneity as a Result of Domain Evolution. *SIGMOD RECORD,* 20, 4: pp. 16.20, 1991

[13] Rolling Plan for ICT Standardization - https://ec.europa.eu/digital-agenda/en/news/rolling-plan-ict-standardisation-0

[14] https://www.w3.org/TR/owl-time/