*Williams Rail Review – Evidence Submission*

Paul Kelly, BR Fares Ltd.
22 January 2019

## Executive Summary

Rail fares are inherently very complex and have always been so, since the earliest days of the railways. To understand, manage and maintain the electronic databases supporting the sale of fares requires substantial and specialist expertise. Experience of recent IT projects relating to fares retail systems shows strong evidence of a lack of such expertise amongst the staff responsible for managing and maintaining these complex shared computer systems. This bodes poorly for the rail industry's capability to achieve any substantial innovation in, or simplification of fares.

We hypothesise that:

a) the loss of this expertise results from the dispersion of the staff responsible for these systems amongst multiple entities, due to the fragmented structure enforced by the privatised railway
b) the short-term nature of train operating company (TOC) franchises and the fixed term outsourced IT contracts awarded by Rail Settlement Plan (RSP) act as a disincentive to long-term investment in training staff and developing the necessary expertise

And make the following recommendations:

1. RSP should stop outsourcing IT contracts for the core systems which disseminate fares & timetable data and support retail of fares, and instead start to employ software developers in-house, as a first step to re-building the knowledge and "corporate memory" of how these systems work
2. An industry-wide technical forum for fares, timetable & routeing data should be initiated, where staff responsible for generating and consuming this data as well as those responsible for writing the computer software which does this, irrespective of their commercial affiliation, can come together to share in open and honest discussions about all aspects of the data and associated systems

## Introduction

BR Fares Ltd. develops software to calculate fares and ticket validity. Originally (beginning in 2012) this involved the operation of independent websites providing fares and timetable information by myself (Paul Kelly) purely as a hobby. Since 2015 this has been a full-time occupation and involves supply of software development services to the wider rail industry and used by 3rd party retailers in particular. Over years I have come to get to know the fares data and associated database structures very well.

The primary area I wish to offer evidence and comment on relates to the objective of offering good value fares without increasing subsidy, and touches on the broader topic of industry structures. As the subject matter is fairly complex, the majority of this document consists of fairly detailed descriptions, in an attempt to make it the evidence understandable to as wide an audience as possible. I will give an introduction to the complexity of rail fares, and follow that with two recent pieces of evidence which I believe show a lack of appreciation of such complexity on behalf of the staff who manage fares systems.

**Fares Complexity**

Railway fares are inherently very complex and have always been so, due to the need to accommodate flexible “walk-up” travel and the multiple journey options (connecting services, alternative routes etc.) available for travel between any two stations. When the Liverpool & Manchester Railway (L & M) opened in 1830 it ran just one route, all travel had to be booked at least 24 hours in advance and tickets were binding to a specific train. This allowed straightforward yield management by setting the price according to demand for the specific journey each passenger was booking.

The L & M had around 25 stations which, assuming tickets could be issued from any station to any other station means there were 600 different fares that could be issued, even without considering differential pricing based on peak demand. This simple example should make it obvious that individually setting fares between every possible combination of stations results in an impracticably large amount of data, a point to which I will return later. It appears that the L & M managed the situation in practice by calculating fares using a zonal distance-based system.

Such single train, advance purchase fares as were sold on the L & M are still sold on today's railways of course, and are probably some of the simplest to manage from a TOC's point of view since all that’s really required is to set up the available price tiers in advance, set quotas for the price tiers available on each train according to the TOC’s yield management strategy and the expected demand for that train, and then to “let the computer do the work” of calculating the appropriate price and letting the passenger book their journey.

However a requirement to book in advance and/or be tied to a specific train is too draconian for many passengers, and TOCs oblige by offering a wide range of more flexible tickets, which differ from each other in terms of validity, time & route restrictions, allowed discounts, etc. The data structures behind the fares are very complex, due to the need to accommodate all sorts of validity and restriction details in a computer-readable format which can be interpreted unambiguously by the various ticket issuing (TIS) and journey planner systems in use in the industry.

Considering this, combined with the large number of stations on the network today (over 2,500) and the consequently large number of fares available to be sold (e.g. if we assume there are on average 5 fares available between each pair of stations, we already have over 30 million individually saleable fares), it should be clear that there are real challenges in managing and maintaining such a vast and complex data set. To be clear, I am not disputing the capabilities of modern computer systems in handling such large amounts of data, but rather the ability of fallible humans (in the form of pricing managers, database administrators etc.) to keep track of and manage it all.

Traditionally the main tactic used in managing the data volumes is by grouping things together and standardising terms and conditions so that fares and restrictions do not need to be set individually for every pair of stations. A large part of the database structure still used today dates from the modernisation and standardisation carried out by British Rail as part of the introduction of the Accountancy and Passenger Ticket Issuing System (APTIS) in the 1980s.

An example of grouping is the way stations are grouped into different-sized clusters for setting fares over longer distances. E.g. the fares from Reading to Birmingham New Street are the same as those from Reading to Dorridge, and from Reading to Sandwell & Dudley, and in fact to 57 other stations in the West Midlands. As fares only need to be set and maintained in these “flows” between clusters, rather than between individual stations, this saves a lot of complexity and keeps the data (which must be loaded and processed by all ticket issuing systems) to a manageable size.

It is fashionable to characterise the existing fares data structures as out-of-date and in need of replacement, but I would offer a contrary argument that the core data structures are actually very well-designed, have served their purpose well and continue to do so, and that the real problem is in fact a lack of understanding and appreciation of the existing data structures by those responsible for their ongoing management and use.

As evidence of this I would offer an example each from two IT projects managed by the Rail Delivery Group (RDG) in recent years: the Retail Control Service (RCS) and the Product Management System (PMS).

**Use of RCS for controlling availability of Ticket Fulfilment Methods**

The RCS data augments the main fares database with a number of additional restrictions on who may sell fares, and how. Of particular interest here are the way the restrictions on issuing fares as M-tickets, E-tickets, Self-Print tickets and on ITSO smartcards have been defined. The design of this sub-system ignores the lessons from the 1980s and earlier of grouping things together for efficiency and manageability, since the fulfilment method records are not aligned with the clustered flows from the main fares database, but instead defined individually between every possible pair of stations!

As an example of the wasteful inefficiency and complexity caused by this flawed design decision, consider the amount of database records that would be required in order to describe availability of M-tickets and E-tickets on CrossCountry advance fares between Wokingham in Berkshire and Uttoxeter in Staffordshire.

Wokingham to Uttoxeter is priced as a clustered flow between clusters containing 23 and 8 stations respectively. So, although this flow is represented in the main fares data as just 1 record, to describe it in the RCS data requires a separate record between every possible pair of stations in the two clusters: 23 x 8 = 184 records, i.e. 184 times as much data. Such inefficiency scales up very quickly and is further compounded by the inefficient XML transfer format used for the RCS data, which I would argue is another poor design choice for this sort of data. The most recent full refresh of the RCS fulfilment data, distributed by RDG on 12 December 2018, was 6 gigabytes in size. To put this in perspective, the entirety of the rest of the fares database, containing details of all stations, ticket types, prices, railcards, restriction and validity information is currently only 233 megabytes.

Or to put it another way, the RCS fulfilment data, which augments the fares database with a tiny bit of extra information, i.e. whether fares can be issued as M-tickets, E-tickets, Self-print or on ITSO cards, is *25 times the size of the entire fares database* – and still growing, as more flows are enabled for E-ticket and M-ticket fulfilment.

This means the data is very awkward to handle and has caused difficulties for rail industry suppliers in terms of requiring modifications to their systems to handle such large files. Possibly as an indirect consequence of the complexity, the data also contains many inconsistencies, e.g. within the above flow from which Wokingham to Uttoxeter is priced:

- Wokingham to Uttoxeter allows fulfilment on M-ticket and E-ticket
- Wokingham to Stoke-on-Trent allows fulfilment on M-ticket only
- Reading to Stoke-on-Trent allows fulfilment on neither format

N.B. all the above are exactly the same fare, at the same price, with a single record in the fares database, but many differing records in the RCS data. I find it hard to believe that such inconsistency is intentional and would suggest it is an error, an almost inevitable outcome of a system that appears to be too complex for its creators to manage.

**Maintenance of Time Restriction Data in PMS**

The Product Management System was introduced as a replacement for the old mainframe-based fares database in June 2017. My understanding is that the suppliers of the old fares database also maintained the data to a certain extent, but with the new PMS system TOC pricing managers maintain all the data themselves.

A specific example of the sort of maintenance needed is the adjustments to otherwise static data to keep abreast of public holiday dates. The fares database doesn't have a mechanism to describe public holidays; the dates of applicability of most restrictions and easements are described in terms of cascading date bands, defined by a starting and ending month and day, and in terms of which days of the week (within the date band) the restriction or easement applies on.

This system allows a lot of flexibility, but the data needs to be updated to reflect upcoming holiday dates in advance of fares going on sale. Towards the end of 2017 it became apparent in very many cases that the data hadn't been properly maintained since the switchover to PMS. I became aware of this due to complaints from clients of large numbers of apparently invalid fares being sold. In some of the worst cases, Super Off-Peak fares intended to be used only at weekends (e.g. from Cambridge to London) were being sold for travel in the morning peak on weekdays, because the restricted date bands had not been extended since the move to PMS and so by default the fares were valid all the time.

I was regularly told by my client that the problems with the PMS database had been raised at the highest levels with RDG, yet fixes and improvements, if they came, seemed to only ever happen at a snail's pace, which suggests to me that no one had any overall control and also perhaps that the impact of the problem on passengers and their confidence in the railway was not recognised, otherwise there surely would have been more urgency in fixing the problems.

As a result of these problems I was commissioned to write a computer program that would manually modify the data and insert extra date band records to cover the gaps, in cases where a set of data records appeared to have been unmaintained since the PMS switchover. This modified data was then used by my client for its own retail channels as well as the many TOCs it provides "white label" websites to, and this limited the scale of the problem.

However it still took several months for the data to be corrected, and throughout 2018 we have noticed many further and ongoing problems with time bands in restriction data not being properly maintained. I can provide detail of specific incidents if required. To an outside observer, it seems that the skills in maintaining this data have had to be re-learned over an extended period, and that the complexities were not understood or appreciated by those responsible for introducing PMS, otherwise the relevant staff would have received training in advance of the new responsibilities they gained under PMS.

And this brings me to my core point, that it appears to me that there is a lack of appreciation and/or understanding of the immense complexity involved in the electronic databases behind rail fares by those personnel in RDG and TOCs responsible for their maintenance and renewal. There is perhaps also a lack of appreciation that this complexity is nothing new, is inherent to the railway network and cannot be removed or hidden by the introduction of new or improved IT systems.

There may also have been an over-optimism on the part of RDG project management staff in the ability of TOC staff to understand their data and keep it up-to-date, otherwise I can't see how they could have permitted the PMS system to go live with so few apparent safeguards against issues caused by poor data quality.

**Conclusions**

My hypothesis is as follows:

- Railway fares data is so complex that it requires many years of experience and/or training to achieve competency in its use and maintenance.
- The short-term nature of most TOC franchises together with the fixed-term nature of many contracts for out-sourced IT services discourages long-term investment in talent and training.
- The fragmented structure of the rail industry makes it difficult for the experienced staff that do exist to communicate and develop ideas jointly, since they all work in different organisations.

The difficulties in long-term development of human resources caused by short-term franchises and contracts should hopefully be obvious. As an example of the sort of difficulties caused by the fragmented structure of the industry, consider that the personnel responsible for:

1. generating and maintaining the fares data (over 20 TOCs)
2. consuming the fares data (TOCs, travel agents, 3rd party retailers)
3. writing the computer software that facilitates step 1 (e.g. Civica, iBlocks, CapGemini)
4. writing the computer software that facilitates step 2 (e.g. Worldline, Fujitsu, FastJP, Silverrail)
5. co-ordinating communication between the above (ATOC and RSP)

all sit in separate organisations with entirely separate management, purpose, commercial incentives etc.

There is no overall leadership of staff who work with the fares data, and no one has a whole-system view of all aspects. The problem is not the data complexity; that is inherent and can't be changed. The problem is both that the industry does not have enough trained and competent staff who understand the complexity and can work with it, and that the fragmented structure of the industry means those making decisions do not have access to the combined wisdom of those staff who *are* competent in working with the data.

Rail Settlement Plan (RSP) is the division of RDG responsible for provision of shared retail services used by the TOCs and retailers. It generally fulfils this role by procurement and management of a set of outsourced IT contracts for the provision of the various back-end computer systems necessary for the dissemination of timetable and fares information and for the sale of fares – including the PMS and RCS systems mentioned previously.

The original fragmentation of British Rail at the time of privatisation effectively made such outsourcing unavoidable, since the British Rail Business Systems division which provided those services at the time was sold to IT services companies which were then contracted to provide the same services to RSP.

However over time RSP has tended to migrate the outsourced contracts away from the incumbent IT services companies, for example with the contract for PMS, which was awarded to Civica. Before 2017 the fares database service was provided by Worldline, which had inherited it from British Rail Business Systems via various corporate mergers & acquisitions over the years.

It strikes me that it might go a long way towards improving the situation if, when these contracts expire, rather than outsourcing them to a new supplier with no experience of the fares data, it took the development in-house, employing software developers itself to develop and maintain the systems. I can see two main advantages from this:

1. It could justify long-term investment in training the relevant staff, as there would no longer

be a fixed term outsourcing contract involved.
2. Having such developers in-house would increase the pool of knowledge and experience of the fares data within RDG, which would be a great advantage when it came to decision-making regarding future enhancements to the fares system, re-structuring of fares etc.

However that is a medium-term option as RSP could only take contracts in-house when they expired. In the shorter term, it should still be possible to move things a long way towards better-informed decision making through a concerted effort to facilitate communication between the different staff responsible for generation, maintenance and consumption of the data throughout the industry, including suppliers of TISs, journey planners, RSP core services as well as TOC pricing managers and RSP/RDG project management and compliance staff.

I would suggest a bi-monthly technical forum as a place where the above staff could be encouraged to meet for open and honest discussions and sharing of plans, giving of feedback, answering of questions, etc.

Both these options are not really very radical and, as far as I can see, don't require any changes to legislation or even government policy; they could be taken forward almost immediately.