

## Real Time Train Information Consultation – response by Tom Cairns (swlines Ltd)

### Background

[REDACTED]

Realtime Trains (RTT) is my newest consumer product. RTT is a website available free of charge to the general public at <http://www.realtimetrains.co.uk>, and was developed by myself. Its intended purpose is to provide a realistic and viable alternative to Darwin, providing predictions on future arrival, departure and passing times for all trains on the network.

RTT makes use of the Network Rail open data feed platform in order to provide real time train running information for passenger and freight trains across the rail network. RTT also uses data from Transport for London in order to facilitate additional reporting in areas from which Network Rail does not have any data sources.

Realtime Trains site was launched in October 2012 superseding previous two previous versions of the site, which were made available at [traintimes.im](http://traintimes.im) and [rail.staging.swlines.co.uk](http://rail.staging.swlines.co.uk)<sup>1</sup> under the name "Timetables". RTT's October 2012 launch represented the first website making use of the TRUST data feed, not making use of any element of the Darwin service.

There are two functionality modes: 'basic' and 'advanced'. The 'basic' functionality, aimed at the majority of rail passengers, shows public arrival and departure times of in-service passenger trains at stations. Users can drill down to a specific train to access certain information, such as the train's departure times en-route to their station, expected times of arrival, and also some platform detail. The 'advanced' functionality, aimed at those with a need for greater detail (enthusiasts, for instance), provides information on all trains, including empty passenger, special and freight services, on the rail network. It also provides detail of the Working Timetable (as opposed to the public timetable aimed at passengers) and passing times of trains (both at stations and at passing points and junctions).

Since the initial launch of RTT, I have steadily expanded the site with added data from the Train Describer (TD) feed which makes use of signalling movements in order to expand upon where TRUST is not able to provide data. The TD input currently covers the entire South West Trains operating area. I intend to extend the use of this data in the future to cover the entire railway network in due course, where possible.

In late January 2013, the site was further expanded following the release of anonymised freight schedule and associated real-time running data from Network Rail. [REDACTED] The usage spread is now broadly even between those looking for passenger train running information and those looking for detail on freight and empty passenger trains. The former category is continuing to grow, with the growth for information outpacing that of freight and non-passenger services.

[REDACTED]

The site is funded through the use of advertisements on the bottom of pages featuring real-time running data. [REDACTED]

Prior to this, I wrote a website called RailMiles which is a commercial product allowing rail enthusiasts to log their journeys by rail. I have always aimed to streamline the process of inserting journeys, by making use of real-time running information in order to reduce the amount of information that needs to be entered. For the short time the Darwin feed was open, I was able to use this but it was a halfway step and not in any way capable of meeting all the requirements of my users. Nonetheless, RailMiles is slowly gaining traction within the, albeit niche, market. I expect a small growth in users when real-time data is placed into the user flow.

---

<sup>1</sup> This was a test site, which I made publicly available for a period of time.

As part of the RailMiles service, I created a system called the RailMiles Mileage Engine (RMME). RMME contains a mathematical graph (or geographic model) of the national railway network and is intended to provide the shortest distance between two given points. This is freely available at <http://mileage.railmiles.org>. The commercial RailMiles service makes use of RMME by calculating distances between the calling points of a service. This is primarily due to being created and developed prior to the open release of timetabling data from either ATOC or Network Rail.

RTT makes use of RMME by augmenting the planned schedules with additional passing points<sup>2</sup>. The primary intention is to facilitate an enhanced prediction model, through the use of additional reporting the TD feed, but it has also enabled a complete list of trains being made available to users in the advanced mode.

I have, personally, worked at ITSO for a period of a few months reviewing the travel smartcard market in the UK. My time involved travelling the UK discussing smartcard implementations with operators as well as experiencing, first hand, the differences between various operators. The outcome was a report containing recommendations for ITSO, to take forward, in order to attempt to create greater conformity within schemes while ensuring that each scheme is able to be distinct. I maintain a continuing interest in this – and it is specifically relevant in terms of the Darwin ‘code’.

---

<sup>2</sup> For example, a train between London Waterloo and Clapham Junction on the *fast* lines will pass through several junctions and Vauxhall station en route. Through the use of RMME, Realtime Trains will add these en route.

**Question 1: We are looking for stakeholder comments on NRE's proposed changes to its Code and where changes have not been made, comments on NRE's reasoning.**

*This answer will discuss both my viewpoint of Darwin and how I see the proposed code as operating.*

The changes that NRE are proposing to make are to be welcomed. However, this does not mean that they are moving in the right direction.

Darwin is an excellent system with a stated goal of providing information to the travelling public. It cannot be disputed that the system, by and large, meets the targets that it sets out. NRE, however, frequently appear to act in ways that raise questions as to whether they are the most appropriate custodians of real-time information aimed at passenger consumption.

Unfortunately, NRE are moving into a position where a developer/company will be guaranteed to be reliant upon them should they wish to provide real-time information about rail services. At present, there is a potential that it can be done, for some TOCs, solely through the real-time data feeds provided by Network Rail.

This move, surrounded by the phrase 'one source of the truth', should be of extreme concern to both passengers and the industry. During normal rail operations, if a service runs late it can be expected that most passengers are only likely to see **minor** discrepancies in predicted arrival and departure times. Commendably, Darwin aims to resolve this problem by providing a single source of predictions across the country to resolve these discrepancies.

'Darwin CIS' is a rollout programme that aims to drive the output of Darwin predictions across the country throughout all Customer Information Systems (CIS) systems during 2014. In this day and age of data being increasingly open, this programme should be alarming to developers and the industry on the whole. TOCs and NRE are rolling out a system aimed at removing a problem and, simultaneously, recreating the exact problem again, within the online ecosystem, by not liberating their data.

It could be asked whether the one source of predictions that NRE aim to provide are correct and I have touched upon this later in my response. [REDACTED] Personally, the idea of a centralised data source is appealing and could be considered a good move for the industry by reducing the potential for failure in the data flow.

For every potential non-NRE source that is lost, it adds another hurdle that must be crossed in order to provide alternatives without going straight to the central source. Rail information is, clearly, a highly competitive market with only NRE being able to provide a data source that can be thought to be the most accurate in terms of disruption information.

As a central resource, Darwin could prove its worth on many levels. However, the data contained within the service should be liberated under an open licence. It is not particularly necessary to release the entire output that Darwin provides, as other companies will undoubtedly fill the gap of aggregating the data sources. It is only vital that the gaps that are left otherwise void, by the differences between reality and what can be provided from other sources, are filled.

## The Code

### PURPOSE OF THE CODE

Darwin is clearly intended to be a one-stop shop which is able to provide information about the operation of the rail network at that moment in time, with the ability to look forward to see expected times within the next few hours. This makes Darwin wholly unsuitable for a number of potential use cases of real-time rail data – but it remains possible for workarounds to be implemented in order to facilitate this use.

ATOCs aim is to facilitate the *enabling [of] new products and services to enter the market*. Darwin is intended for one use, and given the remaining restrictive clauses; it would be infeasible to diverge to a large degree from this. One app for iOS, from Anecdote Software, appears to attempt innovation through providing a *unique ‘minutes until departure’ for each train* by synchronising to the clock provided by Darwin. This seems, to me, as about as far as you can reach by using the Darwin service and abiding by the code. However, it should be pointed out, that it is possible to be more accurate, with respect to this functionality, in the most important areas by providing functionality using the Network Rail feeds.

#### *Material adverse impact*

Both versions of the code<sup>3</sup> state that there is a requirement for the usage of NRE data to not pose a material adverse impact on TOCs, *“whether financially, strategically, operationally or in regards to their reputation”*. Given Darwin’s aim of providing live departure information, it would be difficult to violate this code if meeting the standard use case providing that there are no biased outputs – appearing to be akin to the *impartial retailing* clause contained within the Ticketing & Settlement Agreement.

It is arguable that users intentionally withholding data from applications that build upon the Darwin service could be contributing to a material adverse impact to TOCs. However, some TOCs now provide apps driven from Darwin data. In particular, London Midland’s app restricts data to only show stations that they serve. It is feasible that if a user were to look at an area with ‘group stations’<sup>4</sup> and London Midland does not show all of them, that another TOC could see a financial impact if this information impacted passenger flows, potentially resulting in the change of ORCATS distributions. This clause may be intended to ensure that bias to any particular TOC is not shown – but this raises an important question as to whether the TOCs themselves are allowed to breach the code in this respect.

One clear use case, which is filled partially by Network Rail, is the availability of performance statistics. The general public has long been sceptical of the output from TOCs with regards to the Public Performance Measurement despite a continued audit process being carried out. The use of Darwin data, being named as ‘the one source of the truth’, can therefore be considered as the ideal source of such data to drive independent statistics – the code restrictions clearly imply that this would simply not be possible.

There is a growing market for free-to-use services to generate ‘Delay Repay’ compensation claims. There is an appeal in using real-time data in order to verify these claims prior to generating the forms automatically. It seems likely that NRE may reject any application to use Darwin data in this way, due to the potential of fraudulent claims. However, there is a likelihood of fraudulent refund claims being made, regardless, to TOCs. It is unlikely that TOCs do not have robust systems to combat long term fraudulent claims. Developers have a variety of tools available to them to verify that valid claims are made – such as through tracking users, with permission, and determining a likelihood of the services that they are on.

NRE appear to consider themselves to be beholden to the TOCs, given their position as a subsidiary of the trade association of TOCs. Therefore, it is clear that they would not wish to be involved as a

---

<sup>3</sup> Revised issue as part of the consultation, and NREs release during the consultation

<sup>4</sup> Where more than one station exists for a town

facilitator of bias. Given that TOCs predominantly run services over a railway owned and maintained by Network Rail, it is likely that Network Rail would be a better source of information for statistics. There are, however, issues relating to the liberation of data, which is discussed further in question 2.

#### *Reputable company or person*

This clause should be removed. While NRE has provided clarification as to whether a company is considered reputable for the purposes of the consultation, it seems conceivable that they may use a previously undefined basis on which to reject licence applications.

This particular clause states that any previous user of the service, while an access token requirement was not in place with no evidence of licences being required, can and/or will be prohibited from using the data feeds at any point of time – regardless of whether they have a valid licence agreement in place.

Given my position of having an in-depth interest of the use and role of smartcards and smart ticketing, it could be considered with the level of interest I hold that I would not be a reputable person to hold a licence. I frequently make use of social media and other mediums, as well as going through direct channels, to highlight faults and errors in process of implementation and operation of these services. I have found that using social media is able to result in engagement resulting in rapid solutions to the problems found. This, however, could be considered as something where it damages TOCs reputations and therefore mark myself, and my company, as non-reputable.

This runs the risk of setting a dangerous precedent. Many software developers are not affiliated with large companies—they are individuals, often working alone or in small groups, to develop applications. With social media becoming increasingly common, there is an alarming possibility that opinionated individuals who express unfavourable views about the API, Darwin's design, or, indeed, any aspect of a TOC's or ATOC's business, could be denied a licence as a result.

#### *Number of licences issued*

NRE, in their revised code released during the consultation, detail the number of licence applications rejected. Alex Hewson, helpfully, has provided a web page with some crowd sourced application details<sup>5</sup>. On this page (retrieved 27<sup>th</sup> February 2013, last updated 26<sup>th</sup> January 2011), it states that two licence applications had been rejected for a free web and Android app respectively. NRE state that three licence applications have been rejected, with at least two rejected due to fraud concerns. Therefore, by the process of elimination – at least one of the apps listed on Hewson's page are related to fraud, or NRE have, for whatever reason, not remembered these applications.

Given that Alex Hewson has stated in his submission that his licence application was not rejected for fraudulent purposes, it can be concluded that Andy Botting's submission likely *was*. Andy Botting states, in the details for his 'Tube Chaser' application for Android<sup>6</sup>, that he cannot include any detail regarding London Overground due to *ridiculous licencing costs from both TfL and National Rail Enquiries*. This may be due to NRE refusing to offer free licencing for a free app. Having reviewed the utility of the Tube Chaser application, it seems more likely that an earlier rejection by NRE has been forgotten.

However, this information raises questions as to the validity of their statements.

## **PROCESS**

The process for application of real-time rail information should be simple and straightforward unlike that made by ATOC. Network Rail makes the use of their feeds very easy. There was little in the way

---

<sup>5</sup> [https://mocko.org.uk/ldb/ldb\\_licenses.html](https://mocko.org.uk/ldb/ldb_licenses.html)

<sup>6</sup> <https://play.google.com/store/apps/details?id=com.andybotting.tubechaser>

of documentation with regards to how to consume Network Rail feeds initially, but this is slowly being changed through community contributions.

The application process is archaic in the internet age. The block of text accompanying this portion makes note of the fact that other services are made available by NRE. These services are *not* however part of the code and therefore could be perceived as an irrelevance.

NRE makes note of the fact that it is unable to provide service level agreements to any licensee. This would be in line with that Network Rail also does not make any service level commitments for their open data platform – but it does reduce the apparent value of the service.

## CHARGES

I cover this in a separate section given the importance of the topic.

I believe that NRE are seen, as the dominant player, to be open to anti-competitive practices and have, in the past, shown a continued refusal to move away from this position on their datasets. While it is an absolute right in a free market for NRE to allow themselves to do this, the railways are still considered, by the public, to be a public service. Given that the majority of TOCs are still subsidised, to a large extent, by the British taxpayer, there is substantial merit to this argument: therefore, I firmly believe that the railways should be subject to the UK's open data agenda.

The NRE branded iOS and Android apps are both made available for free on their respective stores. The proposed code declares that NRE expect to charge between £1 and £1.50 per mobile app. In order to compete with NRE's own free app, it would be necessary to run advertisements on the mobile platform and experience from other developers indicate that if the market is flooded, it would be difficult to fund the licencing fee per app regardless of the on-going development costs<sup>7</sup>.

NRE have previously briefly suggested that they may be able to provide their own advertising framework for apps, but this is not mentioned in the code. I am, therefore, led to conclude that this option is not publicly available. However, this may still remain possible. In this instance, it raises questions as to how NRE propose to give a share of revenue should it be over and above the costs that NRE would demand from the service.

NRE has said that it *takes neither a penny of government subsidy nor makes a penny of profit* (Chris Scoggins, Wired magazine). ATOC indicates in its business plan (2012/13 – 2014/15)<sup>8</sup> that it has built its on-going business model on *charging for ATOC-provided services through commission paid by members (and commercially exploiting the value of our services to third parties)* (ATOC business plan, page 13). It is clear that NRE will, therefore, resist any change to their view of openness to their model given their stated aim of supporting their organisation.

ATOC's business plan states that their income covers their gross costs of £59 million, where the majority comes from a TOC. Given that NRE itself is a subsidiary of ATOC, it is clear that the public will read this that the losses that TISL incur are covered by ATOCs costs. TOCs themselves are known to receive subsidies with Northern Rail receiving 34.9p per passenger mile (2011-12) – as such, it is inconceivable that a public subsidy does not end on the balance sheets of TISL directly. This only goes to further serve the argument that data should be liberated.

The release of real-time data from Network Rail means that it is now possible to conclude that the only items of interest that NRE can provide are manual interventions during service disruption and

---

<sup>7</sup> A search for 'iAd revenue' (iAd is Apple's built-in advertisement publishing framework) indicates that niche apps are more likely to earn significant sums of money – but a free app is equally likely to result in a large number of downloads. Given that any rail departure board app will not be a niche market in the mobile ecosystem and NREs presence, it would be inconceivable to consider that this is a feasible business case.

<sup>8</sup> <http://www.atoc.org/clientfiles/File/ATOC%20Business%20Plan%20-%20Final.pdf>

predictions. Actual real-time running data is available either directly from TRUST or derived through TD. NRE also provide the hosting for the 'pull' service and it can be assumed that this is what the licence fee covers.

The value of predictions is an important area of discussion – it is clear that NRE provide a robust source that attempts to show the same data in all points of contact with the passenger. The reliability of NRE predictions is, however, questionable – there are many examples every day where a train is expected to make up significant amounts of time. If these predictions were to be made into reality, it is likely that trains would have to travel faster than the permitted line speed, faster than is permitted by their traction equipment, and, indeed, faster than would be permitted by the laws of physics. This can, however, be explained through the desire to be 'optimistic' to ensure that passengers are at the station before the train itself arrives. In the era of 'always-on' connectivity with mobile devices, it can be argued that any prediction change to suggest a train run *earlier* than previously advertised would be better served as sending an SMS or push notification to a device.

### *Push/pull licencing*

NRE have recently made their push licencing terms clearer and this is to be commended. However, they make comments that *licensees applying for push services should ensure they have the financial capacity to give indemnities*. These indemnities, following discussion with NRE, have been repeatedly defined as having a requirement of being 'unlimited' in order to not cause impact on TOC revenues and reputation.

As mentioned before, the release by Network Rail of their real-time information means that Darwin could be used to fill, what is otherwise, a void. Only predictions, in this case, are automated and therefore could be considered to be requiring an indemnity. Service updates in disruption are provided through manual direct input to any system available – meaning there is a likelihood of human error. The wording of the code implies that it is possible that a licensee could therefore be at risk whether it was a processing error at the licensee end or an input error at the TOC/NRE end.

An unlimited indemnity insurance policy is likely to cost a substantial sum of money for any developer or company, meaning that access to a push data feed would only be available for high-level corporate firms or persons with substantial external financial backing. NRE specifically state that it is for the licensee to be able to defend themselves from TOCs with relation to portrayed data inaccuracies. As mentioned before, given the prevalence of data from other sources, it could only be assumed that NRE are requiring indemnity against incorrectly portraying their own predictions. These conditions are extraordinarily wide-ranging, and again suggest that access to the push API will mean a high barrier to entry, and that many smaller, individual developers might be 'priced out.'

NRE provides an argument regarding the security of their service being different if taking data through push or pull feeds. Dependant on the context of their statement, this is highly questionable. Security vulnerabilities can exist in any application and it could be considered that a licensee is at fault if vulnerabilities are found and exposed within Darwin through the licensee's access credentials.

There are complexities with handling any level of real-time data and Darwin can be considered as no exception. However, the Darwin 'push feed' can be used with far greater ease than any of the raw feeds from Network Rail. Given the increasingly numerous successful implementations of services from the Network Rail feeds, it is of interest that NRE show a concern about data integrity and ensuring that users of push feeds ensure consistent output.

### *New ideas*

NRE have stated that they are open to considering 'new ideas' and are willing to develop these if a business case can be developed which is mutually beneficial between both parties. I consider this particular paragraph to point towards an unwillingness to release push access to the data.

At the ORR conference, it was stated by Dave Addey (Agant) that Agant had previously requested a

certain feature be implemented. Following this, the certain feature in question was implemented, but only in NRE's own smartphone application; Agant further commented that NRE requested a substantial sum of money in order to repurpose the feature for other applications. NRE maintain and hold any intellectual property rights (IPR) over the Darwin service and it should be assumed that they maintain these for any additions requested by developers. From my own personal experience, NRE aim to increase the size of their IPR portfolio and their future business plan serves to confirm this.

Developers may be pleased by the prospect of a future in this respect, as NRE seem to appease attempts to diversify away from their own ecosystem. However, given the costs that have been supplied to other companies in this area, NRE are not in a position to provide new features cost-effectively, nor (more critically) rapidly. Developers are likely to find that they could develop equivalent functionality, given the same data, at a much lower cost compared to NREs outsourced teams.

## **APPEALS**

Clearly, this is the biggest area of change for the code. The prior code led to any appeals being dealt with within NRE itself with no independent input. It is to be welcomed that an independent arbitrator will now be involved with this process.

It appears that the code remains unchanged up to the point where the NRE (or TISL) board become involved. Should the appeal not be upheld, the independent arbitrator would become involved. Significantly, the code states that the appellant must make the pay the costs of the arbitrator and that NRE are not liable unless the arbitrator finds otherwise.

NRE claim that the independent arbitrators are appointed in order to adjudicate disputes in the rail industry and that this must be the case in order to ensure the level of knowledge required to adequately investigate the case is maintained. This is highly disputable – the code is in relation to the access of data *alone*. Rail arbitrators would normally be expected to discuss issues at a much lower level, such as revenue or delay minute distribution. If the code were to be open and clear, any arbitrator should be able to work on cases relating to Darwin without much difficulty.

The broad-ranging nature of the code leads to the unfortunate level that the odds of a successful appeal are stacked highly against the developer; it is likely that NRE would be able to find reasons that are within the word of the code for rejected applications. This is extremely unfair towards applicants.

## **LEGAL TERMS**

ATOC have released the standard legal terms for any licence agreement involving Darwin. These are not of great concern considering the current arrangements around Darwin. However, there is one core concern that is raised from its contents.

The legal terms contain a gagging clause. As mentioned previously, my interest in smartcards means that I frequently hold TOCs to account publicly, as well as privately. This clause prevents myself, and by proxy, my company, from having any involvement in both areas simultaneously.



**Question 2: We are looking for stakeholder comments on the extent to which Network Rail's data feed represents a viable alternative to Darwin and the uses that these feeds can be put to.**

Network Rail's data feed represents the starting base to create a viable alternative to the Darwin system. I aim to demonstrate through Realtime Trains (RTT). During normal operations (i.e. the majority of the time), there is no discernible major difference between the output from Darwin and the output from RTT.

During disruption conditions, there are still numerous issues surrounding data in disruption. RTT is designed to allow 'disruption change' inputs from multiple sources, whether these are calculated automatically from the train movement feeds or manually entered in a console, but is not able to use all the sources that Darwin does.

The primary missing data sources in order to create a viable alternative to Darwin are those from Tyrell and station CIS. Tyrell is able to provide details of intermediate stop cancellation prior to it occurring (RTT is able to calculate these during the event) and details of short formations. It is also able to provide detail about additional services not entered through the Very Short Term Plan feed straight away, etc.

Station CIS is able to augment the scheduling data from Network Rail, as well as the data from Tyrell. Station CIS' predominant usefulness is around the area of advance platform changes. RTT makes use of the Network Rail train describer feed in order to determine service platforms well in advance of scheduled departure times. At present, RTT is using this functionality for the South West Trains network only.

In order to work around these issues, RTT makes use of the Network Rail feeds in unconventional and novel ways. By making use of the most detailed information available (the TD feed), RTT will soon be able to manipulate trains to add cancellations on demand based on the routing of services. This is of most use where trains run on diversions from their planned routing. It is, however, possible to calculate these when trains are travelling along their planned routes should they take different lines (e.g. planned for all stations via the slow line but travelling via the fast line with no platforms). It is not possible to calculate additional calls.

I believe that TOCs should be compelled to directly release Tyrell and station CIS data feeds to the developer community. Darwin is a well defined and mature system that fulfils its aims reliably, it is not possible for Darwin to be able to provide a good grounding and basis to innovate rapidly. However, all of these feeds cannot be taken as being guaranteed to be reliable – for reasons I will explain below.

It is my understanding that Darwin is reliant on input from the NRCC and TOC control centres with relation to disruption conditions and cancellations. However, I have noticed on several occasions that trains which has shown Darwin to show an inaccurate reflection of reality, in particularly on London Midland during the recent heavy snowfall.

London Midland's 'Journeycheck' website, a passenger-centric version of Tyrell, was showing that a train on which I was travelling was cancelled. Accordingly, Darwin, taking in the data from Tyrell, was agreeing with this. London Midland control had not cancelled this train through TRUST and Realtime Trains was showing it as operational with its associated real-time data. It is for this reason that I believe the release of data should be encouraged, but also taken with a health warning.

These situations highlight problems with information flow from both TOCs and NRE. The continued growth of the 'one source of the truth' concept therefore gives cause for concerns. I would imagine that developers in similar situations would be drawn to developing solutions to these problems through making use of all available data sources.

[REDACTED]

**Question 3: We are interested to hear consultees' views on the evidence that we present in Chapter 5 on the number of new licences and apps., and on any reasons why they consider this growth might overstate the health of this market. In particular we welcome stakeholder views on:**  
**(a) The medium-term sustainability (to the extent that this is possible to predict in a fast-moving technology market) of the relatively large number of apps that are currently on the market, including on the feasibility of paid and ad-funded or free-to-download apps coexisting; and**  
**(b) The likelihood of a significantly better range of applications and functionality being made available under a more open data standard.**

I have detailed in question 2 the depth to which I feel that NR feeds provide a viable alternative to NRE's Darwin system. I believe that there will be a point at which the amount of apps based from the Darwin feed become too numerous and the market from these becomes too crowded – to a degree, on the iOS platform this has already begun given the increasing lack of differentiation. As such, I personally am an advocate of stepping towards increased amounts of liberated data.

The market is wholly reliant on NREs contribution at present. As such, there is a distinct single point of failure. It is important to consider that the market is also reliant on the uptime of the Network Rail external services gateway – which faces occasional issues. The most likely problems facing the Darwin service are downtime, data errors or consistency issues.

Where there is a problem with one service, it is important that an alternative exists. I also feel that it is vital that there is an alternative source of data in order to ensure that the data made available to passengers is valid and accurate. An analogy in this instance is the requirement of three distinct systems on aircraft to verify data to ensure continued safe flight – where at least two systems must agree or show a very minor discrepancy. While used for safety critical purposes, this is one potential strategy that app developers could use.

This problem is compounded by nearly every app and website with real-time train information making use of NRE data. Should the Darwin feeds be made open, it will increase the amount of potential innovation and would also give developers a choice of source. In order to create the best choice of source, TOCs should also be compelled to release their data directly.

The choice that could be created is simple – developers would gain the option of choosing between processed (Darwin) and raw (Network Rail & TOCs) sources. Those developers who are wishing to create services and apps rapidly are more likely to choose the processed Darwin dataset, while those looking for a greater depth will likely move towards the raw sources.

Open data results in innovation of previously unthought-of capabilities or value. The excellent Loco2 rail hack day led to a number of potential services and apps being created<sup>9</sup> – some with value and some for amusement. Some examples are *Scenic Railways* – an app which tells you what you can see if you look out of the window<sup>10</sup>, an attempt to create rules for journey planners as to whether a bike can be taken on the train<sup>11</sup>, *Realtime Dutch Trains* – using the NS Rail real-time information API<sup>12</sup> and *The Dank Spangle Memorial Train Timeliness Reckoning* – a search of current train movements to rank by lateness<sup>13</sup>. RTT gained a new feature during that hack, creating a workaround for a problem with the data feeds on that day.

Without the presence of some level of openness in rail data, it would have either rendered all of the hacks made on the day extremely difficult or impossible.

---

<sup>9</sup> <http://loco2.com/blog/2012/10/off-the-rails-the-hacks/>

<sup>10</sup> <http://www.scenicrailways.org.uk/>

<sup>11</sup> [http://nrodwiki.rockshore.net/index.php/Can\\_I\\_take\\_a\\_cycle\\_on\\_this\\_train%3f](http://nrodwiki.rockshore.net/index.php/Can_I_take_a_cycle_on_this_train%3f)

<sup>12</sup> <http://dutch.trains.im>

<sup>13</sup> Source code at <https://github.com/icerunner/trainhack>

**Question 4: We ask consultees for views on whether an open data approach, if adopted, would lead to change in the market for RTTI products and services and if so: (a) what this change might look like; and (b) whether it would be desirable.**

A change towards increased liberation of data in an open licence is highly desirable – my answers to the other questions I believe provide sufficient evidence for this.

In terms of real-time data, the most important set of data to be released is service alteration data for disruption. This will allow a complete picture of the rail network to be created from wholly open data sources. I believe that TOCs should be compelled to release this as part of their network operating licence obligations.

Should NRE be receptive to an open release of their data I would propose that, at a minimum, the ‘pull’ feed be made available on a licence similar to, if not the same as, the Open Government Licence. Network Rail release their data on this licencing arrangement. I believe that the release of predictive data – Darwin’s apparent significant cost – is not necessary as these can be made using data available elsewhere.

Darwin’s push data feed would add little to the ecosystem other than the addition of predictions, mentioned above, and TOC-owned data relating to service changes.

ATOC’s other data releases, through RSP, of timetabling, fares and ‘London Terminals’ routing data are under a Creative Commons licence text. The use of Creative Commons licencing goes against the general open data ethos – the licencing terms of the ATOC data feeds have a clause stating that a user may not *sublicence the Work*. This prevents any user of the existing data feeds from releasing **any** output from any service they may build on top of said feeds. This prevents data aggregators, such as an RTT API, from working with ATOC data.

In addition, ATOC has not shown itself to be able to manage these releases well, with issues on the first publication. Creative Commons (CC) states that the licence text must remain unaltered in order to be called a CC licence. The release of the timetable data resulted in a clause being added to ensure that TOCs were not unfairly treated or given bias through that release – this was quickly altered back to the normal licence.

ATOCs releases have also rarely been ‘complete’ – in attempts to ensure that they can maintain licencing fees on frequent data releases. As an example, Network Rail releases daily timetable updates for free – to industry partners directly and through its open data portal for others. Given this knowledge, ATOC therefore charges over £22,000 for a daily update feed of just Retail Service IDs (RSIDs) – the only need for RSIDs being to create a ticket issuing system.

In terms of the resultant ecosystem, I see it being rich with applications and services which all are able to provide a distinct experience. The ability for innovation is greatly enhanced through the direct open release of datasets. It's impossible to foresee what products may enter the market, and what particular applications might come be invented. Whatever happens, however, it is sure to be exciting, and I am sure that the innovations made possible by open data will revolutionise the future of railway travel in the UK.

Rail data needs to be open.