

RAIL DATA MARKETPLACE

Commercial Research Findings

April 2022

Rail Delivery Group



Private and Confidential

RDM Commercial model: Qualitative interview findings – Q1 2022

A series of interviews with likely publishers and consumers were undertaken to gain insight and feedback on a number of potential commercial models.

Rail Delivery Group

National Rail

Summary 1/5

📚 National Rail

Train Operating Companies:
 TOC's are still upploars

- TOC's are still unclear about their data publishing obligations and await clarity from DfT about what will be required under future Passenger Service Contracts (PSC's).
- TOC's own data but are unsure what is worthwhile publishing.
- Some TOC's lack the right skills in-house to enable them to gather and publish data.
- Commercial sensitivity and competitive advantage.
- Data Ownership and Intellectual Property Rights: Companies may have to seek permission to publish.
- **Costs**: Staffing costs, hosting fees and development costs all need to be considered when making data available.
- Licensing: Licensing restrictions can prevent data publishing.
- Other factors: Include data quality, warranties, SLAs and support.
- Data to improve the customer experience: Station assets, disruption data, delay / repay, train occupancy, station crime data, distressed stock etc. to benefit the customer journey.
- Historical Network Rail Data: HSP Historical Store of Data (Old Network Rail Data) and TRUST data in a user-friendly format to understand current issues and predict future ones.
- Track condition, engineering: Data to assist enhanced maintenance, scheduling & operational efficiencies.
- **Consolidated TOC data:** Travel Management Companies want peer benchmarking and retail customer data combined with TOC data to provide a better customer experience.
- Enhanced CO2 Data: Accurate CO2 data which incorporates rolling stock whether it's electric, hybrid, diesel, actual passenger count and seat turns.
- Environmental: Historical & real time weather data to forecast demand, propensity to travel and operational efficiencies and impact on track conditions such as leaves, heat, flooding along with the potential for disruption.
- Data based around a specific location: Useful when looking at connected/cross-modal travel around stations and local areas.
- **Events Data**: Can provide inspiration and encourage people to travel. Also useful for demand prediction, revenue management, better customer journey planning and planning around engineering works.

Barriers to Publishing for Initial Data Publishers

Data sets of

particular interest

Summary 2/5

Rail Delivery Group



.....

How Data Publishers charged for data

Current charging structures from the participants we spoke to replicated the charging structure proposed by the RDM and therefore highlights that our pricing approach would be familiar and acceptable to users. Charging structures mentioned were:

- Charge by volume
- Standard Charge with bespoke data manipulation charges or one-off fees
- Volume and monthly/annual subscription charges
- Charge per transaction
- Freemium
- Monthly/Annual
- Subscription
- Variable charge depending on the type of data being accessed
- Different pricing depending on the customer. The RDM needs to ensure the platform functionality can facilitate this.

Our Consumer Pricing Structures

How our models were viewed:

1. Open data: Rarely seen as completely open and some form of licence may need to be attached to its use.

2. Pay As You Go: Useful payment method for static/downloadable data or one-off payments.

3. Freemium: Recognised charging model that allows consumers to test and try out data before committing to other charging methods. Can also support innovation by allowing access to different datasets without any cost.

4. Subscription: Common model across the publishers with either monthly or annual billing.

\bigoplus

What would make users pay for data

1 Additional value-added services: Such as modified or cleansed data, and data with robust service levels.

2. Warranties: Guarantees around data quality were seen to be a value-add that was worth paying for.

3. Opportunity to increase revenue or provide a better service: Organisations looking to gain a competitive advantage over their rivals may look to purchase data to provide an enhanced offering to their customers.

Summary 3/5

Rail Delivery Group



た Hosting Fees

- Observations from research point towards this method not aligning with 'Open by Default'.
- Where a publisher may be
 'obligated' to make data public
 there are already costs in
 curating, managing and
 recruiting staff so this would add
 further costs.
- Would deter smaller publishers and new data sets without an existing customer base with no guarantee of income generation.
- Risk of reducing the number of publishers and value of the RDM



Commission Fees and Charging

- Paying commission is standard in a data marketplace with ranges between 5% and 33%:
- The optimal commission rate varies on the perceived value of the RDM and data set maturity.
- Data Publishers seemed happy to pay commission for a route to market particularly for data with a small or no existing customer base.
- The commission rate is important as publishers would need to see a potential return to cover costs and make a profit.
- If too high commission could encourage direct contact to consumers bypassing RDM.
- Flexibility over charging may be necessary with charging adapting to the amount of data used and by combining different charging mechanisms e.g. Standard and volume related charges.

Summary 4/5

Rail Delivery Group





Licensing Findings

Licensing was seen as a complex and challenging area. The costs involved were high so there seems an opportunity to simplify the approach in this area. General acknowledgment that licensing should not be a barrier to data access, however larger organisations may default to existing licensing due to specific clauses important to the organisation and the cost of legal fees.

- **Restricted licence based on competitors:** Users may want to decide who can access their data.
- Off-the shelf licensing: Some organisations have already paid for their licences and therefore have confidence that their liabilities are covered.
- Licensing may be impacted GBR: In terms of data ownership and TOC's publishing obligations.
- Challenges with contract drafting for domestic and international audiences: Non-UK consumers could be a factor and legal advice required around how this would work.
- Longevity of Contracts: Contract lengths ranged from seasonal through to 1 year or 5 years. The longer the contract the more a business could rely on that data for their business.



Aspects of Licencing that are Important

- Warranties and data quality: Allows businesses to have confidence in the data they agree to use.
- Data ownership and licencing: If data is published who has ownership and who benefits commercially. This point was raised several times when companies work together to provide data.
- Intellectual property rights: Who owns the IPR and the implications of aggregated data sets.
- **Different licence based on the customer**: Different licencing conditions depending on who can consume e.g. Academia vs commercial organisations.

Summary 5/5

Rail Delivery Group





Some potential risks emerged around publishers and consumers:

- **Data Consumers circumventing costs:** Consumers could potentially try to obtain specific data elsewhere ultimately compromising data quality.
- **Layer of ownership of the data:** Data may have several layers of ownership. Consideration needs to be given as to how contracts are structured with agreement from relevant parties.
- **Quality assured data:** If quality isn't defined, then consumers have little understanding of what to expect from the data.
- **Pricing:** If the commission is too high, there is a risk that publishers and consumers will connect outside the RDM. Other value-add services such as exposure to new customers, needs to outweigh perceived cost.
- Handling Breach of contracts: Organisations may seek to involve RDM in any disputes around a breach of contract as they purchased data through the RDM.

The findings revealed several opportunities for the RDM:

• **TOC's as consumers of data:** TOCs would be equally likely to consume data when the range of data that might potentially appear on the RDM was mentioned.

Opportunities

- Driving Innovation: Publishers mentioned reduced costs for innovators / startups and were supportive of a Freemium model to make data sets available initially for review and testing.
- **Monetisation / Cost Neutral:** The commission model is common practice in a data marketplace and was well supported which is likely to be a foundation revenue stream for the RDM.
- Helping organisations scale: new consumers could be acquired without the expense of marketing or sales teams. In addition, the self-service nature of the platform and off-the-shelf contracts would make this easy for new customer acquisition.
- **Invoicing and Community:** RDM invoicing, and reconciliation was seen as a value add as well as the community to set challenges, share experience and drive innovation through rich content important, case studies, success stories.

About Commercial Research

- This round of research will be looking to the commercial elements of the RDM.
- The aim is to understand what publishers and consumers require from the RDM to make it a compelling proposition.
- We will be looking to test the following:

Our Assumptions

Test our current assumptions as to what would be a useful commercial proposition.

- Testing whether users fully understand the RDM proposition.
- Testing that Data Publishers will need to add a layer of value for their data to be consumed such as support or warranties.
- Users will pay for data if they understand the potential opportunities of consuming that data.

Users Assumptions

Test user's assumptions about the RDM proposition and the impact of this from a commercial perspective.

These are:

- That the RDM is a data lake/data warehouse that stores data.
- That the RDM will publish data.
- RDM will purchase data.
- RDM will aggregate and manipulate data.
- All data is chargeable.

What Users Do Currently

- Gather insights on currently used commercial models.
- How existing commercial models might be impacted by future changes to the industry e.g. GBR.
- How these existing models could be adapted to the RDM commercial proposition.

What are we hoping to learn?

Rail Delivery Group

📚 National Rail



Method

Rail Delivery Group

📚 National Rail





Our approach will broadly cover the following areas:

- **Perceptions of RDM**: Understand what our participants perceptions with regards to the RDM and introduce the proposition.
- RDM aligns to organisational objectives: How the concept aligns to Data
 Publisher and Data Consumers objectives.
- Attitudes: Towards data consumption and publishing.
- **Cost vs benefit:** How do users weigh up the costs and benefits they gain from data and what are these benefits.
- **Opportunities and gaps:** What data are consumers looking for and potentially pay for. What opportunities do publishers foresee and the potential to monetise.
- **Current Costs to Publishing:** What are costs publishers currently face and what is their general approach to publishing.
- What data do consumers currently pay for: Investigate what types of data consumers currently pay for.
- **RDM Features and Benefits:** Test the RDM features and benefits and do users see value in this (via both quantitative & qualitative surveys).
- Value of providing a service: What services would users be willing to pay for what value does that provide.

Hypotheses

Rail Delivery Group

🔁 National Rail

What are we explicitly testing?



Publisher: Preference for charging

Users have preference of how they would like to charge for data they publish.

Preference for what they pay for

Users have specific criteria that would make them decide to pay for data.

Charging for Data

Our charging and pricing options are options that they find appealing.

Cost Recovery

Users will want to recoup cost of publishing or paying for data and this can be defined in different ways.

Who we spoke to

Rail Delivery Group

National Rail



Findings

Rail Delivery Group

🔁 National Rail

Users understanding of the RDM

Rail Delivery Group

📚 National Rail

During our research we found that there were many assumptions around the RDM that could impact the commercial model.

RDM

User assumptions

Most of the participants in the research had limited knowledge and some pre-conceived assumptions about the RDM.

- The impact of this is that it creates a different perspective of how the commercial element of the RDM would work.
- The proposition had to be explained to most users. There was the expectation that the RDM was a data warehouse, would host data, manipulate and aggregate data and in some cases pay for data.
- Further assumptions were the RDM would replace existing core rail retailing systems. (2 x TIS participants)
- It was also mentioned that not much was known outside of the participants we spoke to.

"Travel management companies don't know much about the RDM. It would be a good opportunity to create awareness to this sector." (Commercial, TOC)

Impact:

The impact of misunderstanding the RDM's remit is that potential users do not appreciate the opportunities and benefits the RDM can facilitate. Areas for clarification include:

- Commercial Offering
- Data Provision
- Data Access
- Security
- Costs
- Obligations
- Open Data

Action: Continue to promote the RDMs offering and it's features and benefits.

What are the barriers to publishing data

Several reasons emerged as to why publishers may not publish data.

- Uncertainty over their data and it's quality.
- Costs of publishing such as hosting costs and staff costs to maintain the data.
- Not knowing what they are obliged to publish Train Operating Companies (TOC's)
- Ownership of the data
- Commercial sensitivity and competitive advantage.



Barriers to Publishing

Train Operating Companies awaiting clarity from the DfT.

All of the Train Operating Companies we spoke to wanted to know what their data obligations were – without exception:

- They have no idea of the scope, commitment or the data sets required.
- They have data available that they deem to be commercially sensitive and other internal data that helps them understand their business to help run it efficiently and improve the passenger experience.
- They aren't certain what data they have that would be of value outside of their organisation.
- Some direction is needed so they can begin planning. They want further clarity over their expectations so they can resource and subsequently account for the costs involved.
- They may seek to negotiate with Department of Transport and Great British Railways around any data provision obligations. If advised to make data open via RDM they will comply but also discuss costs.

6 Train Operating Companies

Rail Delivery Group



Skillsets within Train Operating Companies:

Different TOCs due to size have different infrastructure and data personnel with some not having the skillsets to harness the value from data.

6 TOCS, 1 Data Aggregator, 1 Supplier

Understanding what to publish:

TOCs have data at their disposal but are unsure what would be useful to publish. In addition, there are concerns over competitive advantage and commercially sensitive data.

6 TOCS

What are the barriers to publishing data

Rail Delivery Group

📚 National Rail



Barriers to Publishing: Data Publishers

Data ownership & IPR	Cost	Commercially Sensitive	Licensing
Data ownership was a key point raised by participants. Understanding who owns the data and who benefits from it may need to be established as	There is significant cost attached to making data available at the right quality and maintaining that data. Resource will need to manage the data sets and it comes at a cost to	Concerns were raised about commercially sensitive data and competitive advantage – especially from the TOC's. East Coast with 2 open access competitors and 2 airlines were particularly	A complex area, with smaller organisations keen on off-the-shelf licensing but larger enterprise and some others would insist on their own licensing. Recognised as timely and costly to negotiate.
part of contracts. The publisher needs to warrant that they have permission to publish the data sets. Further complexity around aggregated data sets	the organisation. Cost recovery or support for costs when publishing was an obligation was raised.	concerned. TOC's recognised that what is innovation one day becomes BAU going forward.	<i>"We own the data in some sense, but there's licencing restrictions as its tied to the licencing requirements of Darwin."</i> Strategy – Data Aggregator.
5 Data Aggregators, 2 Suppliers, 2 Retailers, 1 Academia	3 suppliers, 3 TOC's	3 TOC's, 2 suppliers	3 Data Aggregator, 2 Retailers

What are the barriers to publishing data







Barriers to Publishing

Data Quality

Data quality and the effort involved to bringing it up to a certain standard that can be a challenge.

"We don't stand behind the quality of the data we're providing. We say it will be whatever we're provided with" Strategy – Retailer

2 Academia,2 Aggregators, 2 Retailers, 1 TOC

Underpinned by RDM support is important as it allows publishers to have confidence that any issues will be addressed and in turn their consumers have a level of confidence.

Support

Up to date documentation and community are also deemed important.

2 Data Aggregator, 2 Supplier, 1 TOC, 1 Supplier

Warranties

Being able to rely on the quality of the data and the service provided around it is critical, particularly when paying for data.

In one particular case a publisher would only warrant data if the data they consumed came with warranties.

"if you're going to charge for it, and use that as the business model for funding of the RDM, you better pass on warranties and ensure there's a clear value add."
Strategy – Aggregator
2 Aggregators, 2 innovator, 1 retailer

Data Availability / Format

Ensuring that data is available and in the right format to consume and subsequently publish can be a challenge. TOC data was mentioned as particularly hard to obtain. NR TRUST data format was mentioned by several participants as not ideal. A clear data pipeline of upcoming data sets would be key here.

"Knowing what data is becoming available and having access to that data is valuable". Data Specialist– Data Aggregator

2 Data Aggregator, 1 TOC, 1 Supplier/ Innovator

Service Level Agreements

Provides confidence that the data can be used and deliver what is expected of it. It was mentioned that it would be difficult to build a business without SLAs.

'If all of a sudden you say we're going to put SLA's on the data that changes the game a little bit in terms of what needs to be in place for people to start building proper commercial products" Data Specialist – Supplier

2 Data Aggregator, 1 TOC, 1 Supplier/ Innovator

Data sets of particular interest

Rail Delivery Group

📚 National Rail

These are data sets that could attract interest in the RDM. This was largely based on:

- Data sets that publishers have the potential to provide with or without an enhanced offering and commercialise or make freely available.
- Data that consumers said they struggled to obtain. Some of these were being offered by other potential publishers we spoke to, highlighting a potential match between Publishers and Consumers.



Data to improve the customer experience

Additional data; station assets, disruption data, delay / repay, train occupancy, station crime data, distressed stock etc. that can be consumed and integrated into Apps that would benefit the customer journey.

Travel Management Companies and TIS providers keen, enabling them to have to have greater ownership of the customer / traveller.

6 TOCs, 1 Data Aggregator, 2 Suppliers, 2 Principles



Historical Network Rail Data

This was the HSP Historical Store of Data (Old Network Rail Data) and TRUST data in a more userfriendly format which would prove useful to understand current issues and predict future ones.

This proved to be popular with Academia who advised it was particularly hard to obtain.

1 Government, 2 Suppliers, 2 TOCs, 2 Data Aggregators, 2 Academia, 2 Suppliers Â

Track condition, engineering

Rail data from multiple sources to develop enhanced maintenance scheduling & operational efficiencies.

Real time data aggregated with other- data so technical track and train data and weather / ground saturation data for example.

3 Suppliers, 2 Academia, 2 Data Aggregators, 1 Innovator

Data sets of particular interest

Rail Delivery Group

📚 National Rail

Enhanced CO2 Data

Accurate CO2 data per passenger. Data today is blended. Needs to consider, train type and how powered, diesel, electric or hybrid, actual passenger count and seat turns.

Travel Management Companies would pay for this as their corporate customers complain about the quality of the current data. Would help drive modal shift from car and plane to train and grow train journeys.

Sustainability is no.2 on business travel agenda according to the survey conducted by the Business Travel Association (Nov. 2021).

1 Aggregator, 1 Innovator, 2 Suppliers, 2 TOCs , 2 TIS, 1 consultant



Consolidated TOC data

The Travel Management Companies wanted peer benchmarking and more data across retail channels about their customers with more TOC data feeds they can pull into their Apps for a better customer journey.

Ideally a consolidated GBR view of all data booked via them – they currently aggregate individual TOC data with TOCs filling in the gaps.

The TOC's themselves also wanted more data - 'a single version of the truth' rather than using multiple data sets to get close to the truth.

6 TOCs, 1 Supplier, 1 consultant

Environmental / weather

Important from 2 key perspectives: propensity to travel & demand forecasting (historical & real time weather data) and operational efficiencies – impact on track conditions, leaves, heat, flooding & potential for disruption.

Ground Saturation/ flooding data was important and not easily obtainable. Has real value aggregated with other data to preempt track conditions & disruption.

1 Supplier, 2 TOCs, 1 Innovator, 1 Principle

Data sets of particular interest

Rail Delivery Group

📚 National Rail

Data based around a specific location

Four organisations stated information based around a specific location would be valuable to enhance the customer journey.

- Data to facilitate connected/cross-modal travel around stations and local areas – Smarter Cities.
- Interoperability and single payment for total trip is seen as attractive and would require multiple data sets and integrated journey payment.
- Busy stations best time to travel based on known events would enhance the customer journey and well-being giving them the opportunity to make informed choices.
- Real time disruption data with intelligent alternative travel options.

1 TOC, 1 Data Aggregator, 1 Innovator (3)



"Hybrid working is here to stay and the legacy commuter cash cow needs to be supplemented by yielding up on leisure travel" Strategy – Data Aggregator

Inspiration: Events data can provide inspiration for travellers to encourage people to travel. Train Operating Companies can promote events to passengers to encourage them to travel by rail.

Demand Prediction: Enables TOC's to predict where large numbers of people are travelling by locality enabling better staff rotas, timetabling and pricing / yield management.

Wellbeing / better customer journey: enables travellers to avoid busy services when events are taking place.

Planning around Engineering Works: Engineering works can be factored in around events and expected demand to reduce the impact on the customer experience.

1 TOC, 1 Data Aggregator



Other Data Sets Mentioned

- Robust Seat Maps
- Booked seats vs Actual seats
- Safety and Wellbeing: recommend to customers the best services that are less crowded
- Real time train running
- Delays and disruption data
- Delay Attribution Data
- Train Geo Location Data
- GPS train location dashboard
- Digital twins
- Digitised signalling / points
- Track design / geometry
- Vehicle / rolling stock data
- Car park availability
- Croydon Trams
- Onward transport: scooters, bikes, taxis
- Track layout
 - Real time fare availability without having to select fulfilment
- Gateline data
- IDMS
- Delay/repay

What would make users pay for data

Rail Delivery Group

📚 National Rail

Warranties for chargeable data

Data that carried warranties such as data quality was seen to provide value that was worth paying for as it provides reassurance and credibility of the data being consumed.

Warranties can also be passed on to data consumers as part of their licence, encouraging further data consumption.

Insight: Could RDM accredit data or could is come with an ISO standard? This could potentially be a value add but at a high cost and potential liabilities.

2 Suppliers, 1 SME



Providing additional services was said to be one way to encourage potential data consumers to pay for data.

This can range from:

- Modified or cleansed data
- Providing an extra service e.g. monitoring the data and informing the consumer of updates/problems/errors
- Providing extra data at no cost or minimal cost
- Service wrapper
- Robust service levels

1 TOC, 3 Suppliers, 1 Data Aggregator



Potential to increase revenue or provide a better service

Organisations looking to gain a competitive advantage over their rivals may look to purchase data to provide an enhanced offering to their customers.

Both Travel Management Companies and TIS organisations are looking for better quality data to pass on to their clients – accurate CO2 data was mentioned as was more ownership of clients (TMC's) to manage delay / repay and disruption communication.

2 Suppliers

Reviewing Proposed Consumer Data Charging Structures

Rail Delivery Group

📚 National Rail



Data was rarely seen as completely open and there would need to be some licence attached to its use.

Free data could be made available e.g. organisations that currently made their data free and unrestricted, such as government organisations.

Data could also be free but apply a cost for its transfer.

"The data itself might be free, but this is the transfer cost associated with it." Strategy – Data Aggregator

"Free to use within the context of the Licence" Strategy - Supplier

Pay As You Go

Pay as you go proved to be a useful payment method for more static/downloadable data.

Also was mentioned for one off payments for a bespoke data set or a specific data manipulation request.



Was a recognised charging model and allows consumers to test and try out data before committing to other charging methods. Freemium can support innovation by allowing access to different datasets without any cost.

This was popular with publishers who saw the benefits of exposing some of their data, making available latent rather than real-time data or throttling it to gain interest and drive innovation and ultimately lead to subscription.

"We have a similar model as well in terms of a freemium approach. Our API is completely free to access. You can get all sorts of data queries from there, but you can't get images, and you're limited to the number of queries you can make, but certainly for a developer to go and play there's easily enough."

Strategy – Data Aggregator

Subscription

Subscription was a common model across the publishers either monthly or annual.

Payment was either in advance or retrospectively.

In some cases subscription was for a set amount of API calls and over and above the contracted calls further charges were invoked.

Reviewing RDM Data Publisher charging models

Rail Delivery Group

🔁 National Rail

We tested our proposed charging models with users to see how well they resonated with them



Hosting Fees

Doesn't align with 'Open by Default'.

Where a publisher may be 'obligated' to make data public there are already costs in curating, managing and people so this would add further costs.

It would deter smaller publishers and new data sets without an existing customer base with no guarantee of income generation.

Would likely reduce the number of publishers on the RDM and the overall success of the project.



Commission Fees

Paying commission is standard in a data marketplace with ranges between 5% and 33%:

The commission 'sweet spot' varies on the perceived value of the RDM and data set maturity.

Paying commission to reach a larger market: Data Publishers seemed happy to pay commission for a route to market particularly for data with a small or no existing customer base.

RISK: if commission is perceived as high the RDM could become a free shop window with publishers and consumers bypassing the RDM. Can this be in part mitigated through contract drafting? *"If you provide a route to market, then I think you have something that's certainly of value certainly to a business like ourselves and I would expect to pay a commission for that access" Strategy - Supplier*

"5% would be too low to be able to deal with your risk margin. So, 10% seems about right."

Strategy – Data Aggregator

"For data sets with existing customers with potential to attract new customers the value of the RDM is not as high as a new data set with no customers. Commission at 20% is high – 10-15% would be acceptable but I would readily pay 33% for a new data set where I have no customers." Strategy – Data Aggregator

Other Factors Highlighted with Regards to Charging

Rail Delivery Group

📚 National Rail

Flexibility over charging:

One of the themes to emerge is that Data Publishers and Consumers will need some flexibility over charging.

1. The type of data being published and how that will be consumed.	2. Growth of the business	3. Ability to charge by consumer - not data set	4. Reduction in commission based on data maturity.	5. Invoicing
Insight: If for example, static data was downloaded one or more times a one-off charge model may be appropriate. If real-time data was being consumed, then possibly a monthly subscription charge may apply.	Insight: Charges may need to be flexible based on publisher businesses. If a publishers business grows and they incur more costs they may seek to reduce the cost of their commission fee. <i>"I get about 60,000 hits a month…that</i> <i>theoretically will double if not triple. With</i> <i>200,000 hits a month some of those are quite</i> <i>discreet and small calls for data." Strategy -</i> <i>Supplier</i>	Insight: There may be the need to charge a smaller amount to SME's or Academia but a larger amount to bigger organisations. The platform will need to be facilitate negotiation between publisher and consumer.	Insight: Commission rates may be dependent on how mature data sets are. The more mature a data set is the less commission may be charged.	Insight: RDM invoicing, and reconciling was seen as a value-add service that publishers find attractive.

Reviewing participants charging models

Rail Delivery Group

🔁 National Rail

Current Data Publishers charging structures



Rail Delivery Group

Reviewing participants charging models

📚 National Rail



Licencing Feedback

Rail Delivery Group



Licences varied across organisations and depended on the data set and how it was consumed. In general licencing was seen as a complex area but many respondents acknowledged that licensing was costly and that it should not be a barrier to consumers and innovation.



Off-the shelf licensing would work for some organisation but not for others

For some publishers this would be a value-add benefit but not for others.

Certain publishers would default to their own licence because:

- 1. They already paid for their licence.
- 2. They are comfortable that they have their own liabilities covered off through their own licencing.

There's a potential risk that off-the-shelf licencing wouldn't add value to some more mature / enterprise / regulated organisations.

3 Data Aggregators



Restricted licence based on competitors

Two organisations we spoke to said they used licencing that included commercial restrictions around competitor organisations.

"Commercially, we require that the data is not shared with anyone who is a competitor. There is a requirement probably for us to do an analysis of users just to make sure that they're not a competitor." Strategy – Data Aggregator

Insight: Users in some cases will want to decide who can access their data. Consideration needs to be given as to how this handled in the platform interface as well as from a licensing perspective.

1 Data Aggregator

Aspects of Licencing that are important

Rail Delivery Group

📚 National Rail



Data ownership and licencing

The question of data ownership was raised and recognised as not always straightforward particularly in the rail industry where several organisations may be responsible for the data. If data is published who has ownership of the data and if commercialised who benefits?

Insight:

Contracts may need to establish the data sources prior to publishing and cover points such as revenue and liability. Some existing contracts may not cover the publishing elements and the role and responsibilities of the RDM.

Publishers must provide clarity on who has ownership of the data and will need to confirm in a contract with the RDM that they have the right to make the data available whether its commercialised or not.



Providing a warranty around data quality allows businesses to have some level of reassurance that they can rely on the data to build a business around.

If organisations are paying for data the expectation is that it will come with a higher level of warranties

"Data that is charged for has to carry warranties with data quality wrapped around it and this is important. Otherwise, people can't depend on the data." Strategy – Data Aggregator

1 Data Aggregator, 1 Retailer

 -×

Intellectual property rights

IPR was seen as important and there were questions about this when data sets were aggregated.

This would affect who can access the data and if it was commercialised who benefits from the revenue?

Insights: Licensing would need to consider intellectual property to avoid disputes and provide security around data provision.



Different licence based on the customer

Different licencing conditions depending on who the consumer is and what they intend to use the data for.

Licensing may stipulate data is free to use for academic / research purposes with no limitations and a chargeable with limitations on use for larger organisations.

1 Academia, 1 Supplier

2 Suppliers

3 Data Aggregators, 2 Academia, 1 Retail

Licencing

Rail Delivery Group

📚 National Rail



Licensing may be impacted by GBR

Licensing may be impacted by GBR

Legal contracts may be dependent on the new Great British Rail Contracts from 2 perspectives:

- Data ownership
- TOC publishing obligations

Insight:

Further information would be needed in this area and there is a current workstream within GBRTT around data which should clarify ownership / availability.

New PSC TOC contracts are in development

2 Suppliers, 1 Data Aggregator, 1 Retailer, 1 Government



Challenges with domestic and international audiences and contract drafting

- Consideration may need to be given around security issues
- Legal issues and whether UK Law is the prevailing law
- Whether extra legal costs involved with international licenses.
- Are contracts extensive enough to cover international audiences and could this be a value-added service where extra commission could apply.

This point came about through one Data Publisher working with Network Rail investigating the possibility of a 'Free to Use' licence that would cover data usage internationally.

The cost of this type of licence was extremely high and required a potential rethink of the approach.

Insight: The RDM potentially attracting an international audience came up in the last round of research with companies looking to use UK data. Consideration may need to be given to the audience of the data and if that would be UK or International. If a reduction of costs can be gained from providing international licences this could be an attractive feature.

1 Data Aggregator



Longevity of Contracts

Largely depends on how integral the data is to a particular business.

Many publishers referred to annual contracts.

Others saw longer term contracts as benefits with one stating that 5 years would be the ideal time span to rely on data of a certain quality.

Conversely some had data that was seasonal so may only be accessed for 3 months over the Autumn months.

"We charge seasonally based on track condition when leaves are on the line." Strategy – Supplier

"We normally have five-year contracts. We need to be sure that we can have access to the core data over that sort of timescale" Strategy - Retail

Potential Risks

Rail Delivery Group

🔁 National Rail

Several potential risks emerged as part of the discussion and centred around participants experiences and knowledge of the industry. **—** ×-Handling Breach of **Data Consumers** circumventing costs contracts Layer of ownership of the data Pricing **Quality assured data** Data may have several layers of Organisations may seek to involve There is a risk that the RDM If quality isn't defined, then consumers have It was mentioned that ownership. Some consideration needs to RDM in any disputes around a becomes a free shop window. little understanding of what to expect from when data is charged for be given as to how contracts are breach of contract as they the data. It was also mentioned that allowing data consumers may try structured with agreement from relevant purchased data thorough the If the commission level is too publishers and consumers to review data to circumvent costs and parties. RDM. high, there is a risk that obtain the data could prove problematic. publishers and consumers will elsewhere ultimately This will need to be addressed in "If I'm IBM, and I've been buying connect outside the RDM. This compromising data *"Everyone has a different view. If you let* data through the RDM and it's publisher contractual obligations. needs to be factored into the quality. people assess quality, you're going to be in being managed by RDG and licensing. trouble. You can talk about everything around "There's a question of ownership and I will there's a breach of contract I'm *"When you charge for* it around such as quality of provision, ease of going to seek legal action - you say it's not purely down to us to decide if The value of off-the-shelf data, people go to great use, ISO standards - but only for people who we want to open up access to it. It would need robust contracts.." lengths to circumvent licensing, community and paid." Strategy – Aggregator be up to RDG." Strategy- Data Aggregator that charge" Strategy invoice management, Strategy – Retailer Data Aggregator exposure to new customers, "You need quality assurance with datasets that all underpinned by robust all parties in the industry agree to, to ensure SLA's needs to outweigh its suitable for use in in work and underpin perceived cost. decision making" Data Specialist - Academia. 3 Data Aggregators, 2 Academia, 1 Retail **3** Supplier, 1 innovator 2 Data Aggregators **1** Data aggregator 1 Aggregator, 1 Academia

Opportunities

Rail Delivery Group

📚 National Rail

Several opportunities emerged in the research that could be leveraged in terms of opportunities for the RDM, Publisher and Consumers.



TOCs as consumers of data

It became apparent that Train Operating Companies would be equally likely to consume data when the range of data we were looking to provide on the RDM was mentioned e.g. Events, Grounds Saturation data etc.

One TOC thought there was real value in regionalised data to help with planning and forecasting and yield management.

"If we can utilise events data on our flows in advance it's going to enable a more scientific approach to forecasting, planning and yield management." Data Specialist - TOC

6 Train Operating Companies



Helping organisations scale

Some smaller/medium size organisations may not have internal resource to scale their business.

The RDM would expose their data to new consumers without the expense of marketing or sales teams. In addition, the self-service nature of the platform and offthe-shelf contracts would make this easy for new customer acquisition

"Being a small company, we don't have the bandwidth to sign-up smaller customers, it's very difficult for us to actually make time for them as our focus is on our larger enterprise clients. This would enable us to sign-up new SME customers." Strategy – Data Aggregator

3 Data Aggregator



Driving Innovation

There was resounding support of the RDM and recognition that it could drive innovation back into the rail industry.

Publishers mentioned reduced pricing for innovators / start-ups and were supportive of a Freemium model to make data sets available initially for review and testing.

Digital Twins & Smart Cities projects, which require a broad range of data sets can be further surfaced and supported via the RDM Community and Blogs to further drive innovation and participation.

"We would look to invest or 'seed fund' in people or organisations who doing something innovative with our data." Strategy - Supplier



Monetisation / Cost Neutral

There is a demand from a broad range of publishers to make available monetised data sets.

The commission model is common practice in a data marketplace and was well supported which is a foundation revenue stream for the RDM.

This gives a level of confidence that the RDM will generate revenues which should grow as more data sets are made available.

MVP already has one specified commercialised data set.

Prototype Feedback

Rail Delivery Group

National Rail

Prototype: Known data quality

Rail Delivery Group





Data Accuracy: Several points were raised around data accuracy.

Accuracy of data could be more defined as the accuracy can be established in different ways e.g.

- The degree to which data has been modified to make it accurate.
- The inherent accuracy of the data e.g. delayed data could have inaccuracies based on recent factors so data can be considered accurate within certain constraints.
- Accuracy can depend on how regularly the data is updated or be accurate at the time of data collection: *"For example a company that provides a service makes up part of our data set may cease trading 3 days before we publish. We may not know about that so the data would be inaccurate"*

<u>Prototype Insight</u>: Guidance may need to provide more clarity around what accuracy is referring to.

<u>Uniqueness</u>: User correctly assumed that this was the degree to which data might be duplicated. <u>Prototype Insight</u>: Could be more explicit about what is meant here. User suggested 'Degree of duplication'.

Validity: "Format expected by whom? As promised by the spec that the data provider submits. Is it the degree of validation rather than validity or the range and format as promised by the spec....or range and format as promised by the data specification... if you set a field in the specification document as being always populated and it isn't it could this could throw the system."

Prototype Insight: Could be more explicit about what is meant by validity.

Quality score: Potential users should be able to see the written text by publishers about the quality of the data. *"If it said only 82% for accuracy and validation I'd want to know why"*

<u>Prototype Insight:</u> The quality scores need some sort of reference to enable understanding, Providing commentary from the publisher could help with this but it also needs to be useful when comparing one data set against another data set.

Prototype: licencing – Who can use this data?

Rail Delivery Group

🔁 National Rail

PUBLISH AN API K Back to previous step Licensing Type of user Who can use the data? Anyone Only certain groups Only consumers approved by me Please provide any further comments about the level of restrictions needed: Enter text here Confirm

Protection of data: Participant suggested most people would choose *'only certain groups'* and *'only consumers approved by me'*, to protect data publishers from having their data repackaged and sold.

"If it's charged for and anyone can use it, there's the danger that they may resell or repurpose that information."

Simple licencing: User stated, "a good licence agreement is one that is brief, covers the main points and written in plain English otherwise people won't read it or understand it."

Selecting multiple options: There may be a need to select multiple options so certain groups and also consumers approved by publishers.

Prototype: licencing – Type of usage

Rail Delivery Group





Non-commercial vs Commercial and non-commercial use: User stated he would struggle to know what *'commercial and non-commercial'* means in this context and that essentially it boils down to whether the data is being resold or not.

e.g. "if we provide data to a company to use in their app is that commercial or non-commercial regardless of whether a cost is involved."

Prototype Insight:

 Provide more clarity around what noncommercial mean in this context.

Prototype: licencing – Limitations on use

Rail Delivery Group

📚 National Rail



Enter text here

Confirm

Limitations on use

'Commercial gain': User again wanted to know what was meant by commercial gain.

"Does this mean selling or reselling the data? Trainline use our data but they don't make any money out of it. This would be direct and indirect commercial gain."

 User would select 'no limitations' and suggested this would also depend on who data is licenced to.

Prototype Insight:

- Clarity about what is meant by commercial gain
- Is there a correlation between who a user may licence data to and the limitations of use?. For example a company may decide to limit use for certain companies but not others

Prototype: licencing – Type of Usage

Rail Delivery Group

🔁 National Rail

Charging

Do you plan to charge access some or all of this data?

O No charge

O Pay-as-you-go

Freemium (Charge once cap/throttle reached)

◯ Subscription

Subscription and transaction volume

Please provide any further comments about your proposed charging structure:

Enter text here
Confirm

For Charging user would select:

'Subscription and transaction volume' because the subscription amount would cover the minimum number times a user could access their data up to the maximum usage amount.

Anything over this would be based on a transaction amount. (User wants a mixture of a standard subscription and volume related charges)

Prototype Insight

- Pay as you go: More clarity to describe what form pay as you go takes.
- Also users may also want a mixture of charging methods.
- How do user selections get implemented?

Prototype: licencing – Select licence

Rail Delivery Group

📚 National Rail

Select license

Do you plan to charge access some or all of this data?

Shared without license

Shared under non-conditional license

Shared under conditional open license

Shared under restricted conditional license

○ Shared under bi-lateral contract

O Not shared

Please provide any further comments about your proposed charging structure:

Enter text here

Confirm

Select licence:

Options could provide more information about what they mean:

"I would need more information on Shared under bilateral contract"

User said it would be difficult to pick an option here.

Prototype Insight

Our proposed way of approaching is that the publisher will default to a contract based on the answers they give through the licensing flow

User Insights

- They wanted to see more detail around the licences so a preview option could have been useful.
- Lack of understanding around the different licence types so possibly an explanation of the different licence types would be helpful.

RDM Benefits: Quantitative survey findings – April 2022

A survey was conducted to gain feedback on the benefits of the RDM, as perceived by both publishers and consumers.

Rail Delivery Group

National Rail

The value RDM brings to you: Data publishers





Private & Confidential

The value RDM brings to you: Data consumers





Survey Summary: Comments

Rail Delivery Group

📚 National Rail

Consumer Features and Benefits	What value users expect from of accessing paid data	What data users would like to consume	What else users need to support your objectives
 (No Responses from publishers) Academia would like more visibility of projects funded by RSSB, Government. Data should remain free, open with non-restrictive licence. If it is chargeable then a modest fee. Promote better data standards Ability to add value to existing datasets 	 Data improves an organisations product offering Assist with ticketing and green tourism Quality levels of service Consideration to academia and data costs – lack of budget so rely on grants/goodwill Being able to add value to data Up to date documentation Robust access to data with high quality feeds Ability to market applications based on high quality data feeds Data should be free of charge Only the cost of supplying data should be chargeable 	 (no data provided on provision of data) RCS and IDMS data (fares data) Data covering rail services for UK and Europe Locomotive/unit/rolling stock allocation data Precise location of signal berths Train positioning data through signalling and GPS Departures, Arrivals, Fastest services (inc. Bus) Gateline data from Cubic and S&B TRUST movements, TD data, the plan (plus VSTP feed), TSRs and PSRs, network topography and connectivity, vehicle/set allocations 	 Data should stay open as developers have created products on that basis Existing users should still be able to access for free with new users only being charged APIs should stay the same without needing technical changes Format of the data should stay the same
6 Respondents	10 Respondents	10 Respondent	1 Respondent

National Rail



What is your role?

We wanted feedback on our features and benefits and gain further insight into what users need from the RDM.

Survey ran for approximately 3 weeks

We received 37 responses in total for the survey.

Results came from a mixture of roles with **Developers** being the largest group at 67.6% and **Strategy and Planning** at 2.7% the smallest.

Those who selected 'Other' stated they were a:

- Founder
- Managing Director
- Architect

What is your organisation?

Rail Delivery Group

📚 National Rail



Rail Delivery Group

📚 National Rail

What is your organisations role with data?



With regards to the organisations role with data users were:

57% Data Consumers

43 % Both Data Publishers and Consumers.

How important are the following features and benefits to you as a Data Publisher?

Rail Delivery Group

🔁 National Rail

How important are the following features and benefits to you as a Data Publisher?

Encouraging the use of data standards to aid interoperability	14 %
Ability to provide up-to-date documentation for your users	12 %
Simple licensing using off the shelf agreements and e-signatures	10 %
Manage your data publishing needs in one place	10 %
Ability to make your data openly available alongside other industry	
data	9 %
Ability to access data sharing agreements	9 %
The ability to see how your data is being used and their performance	8 %
Monetising the value of your data	8 %
A community forum to exchange ideas and provide support	8 %
Support in making your data sources available	7 %
Access to consultancy services	4 %

The top three benefits for Data Publishers were for the RDM to:

1. Encouraging the use of data standards to aid interoperability – highlighting a need for data standards in the rail industry.

2. Ability to provide up-to-date documentation for your users

3. Simple licensing using off the shelf agreements and e-signatures

4. Managing data publishing needs in one place

The least important were

- Community forum
- Support in making data sources available
- Access to consultancy services

How important are the following features and benefits to you as a Data Consumer?

R	a	Ľ	D	el	IV	/e	ry		G	r	0	ų	ľ
---	---	---	---	----	----	----	----	--	---	---	---	---	---

🔁 National Rail

How important are the following features and benefits to you as a Data Consumer?				
The ability to gain access to data sources quickly through simple licensing	12 %			
Up-to-date documentation for data that you consume	12 %			
Confidence in the availability of APIs underpinned by robust Service Level Agreements	12 %			
Encouraging data standards to assist with interoperability	10 %			
Accessing a wide range of data from across the rail industry via a simple search	9 %			
Understanding data and how it can be used with example use cases	9 %			
The ability to see Publishers and Consumers views around data quality	9 %			
Ability to access data sharing agreements	8 %			
Trusted data collection providers	7 %			
A community forum to encourage ideas and access support	7 %			
Ability to transition from Data Consumer to Data Publisher and monetise your data	5 %			

The top three benefits for Data Consumers were:

1. The ability to gain access to data sources quickly through simple licensing

2. Up-to-date documentation for data that you consume

3. Confidence in the availability of APIs underpinned by robust Service Level Agreements

The least important were

- Trusted data collection providers
- A community forum to encourage ideas and access support
- Ability to transition from Data Consumer to Data Publisher and monetise your data

"We have been doing some research projects, funded by the RSSB and others, with an aim of using all the rail related data to analyse the performance of the trains and then develop some AI techniques for predicting delays and preventing delays. I think you have covered almost all the features we would like to have. But it may be better to include a category of the projects funded by the Rail Industry(e.g. RSSB, network Rail etc.) and the government(e.g. UK innovation)."	"By far the most important thing is the existing rail data continues to be Free of Charge, and openly available with a non restrictive license (like it is at the moment). If it is "chargeable" it is no longer open data."	"Top Priority: much more focus on the quality of data at the point of generation, TOCs caring about accuracy and best information made available for passengers, and not just whether their own website has correct details. 2nd: overhaul of data schemas and more professional design (e.g. get rid of RCS)."	 Consumer Features and Benefits Academia would like more visibility of projects funded by RSSB, Government. Data should remain free, open with non-restrictive licence. If it is chargeable then a modest fee. Promote better data standards Ability to add value to existing datasets 	
"Free access to high volume data if the publication of that data is NOT monetised (free app). Reasonable tariffs if monetised that take the small developer into account (modest means and in that circumstance modest app prices)."	"The ability to process and add value on existing datasets through data analysis."	"It is exceptionally important that this data is given out as freely as possible, and with as little 'licensing' needs as possible."		

If there are any other features and benefits you would like to see as a Data Consumer

Rail Delivery Group

📚 National Rail

How important are the following to you

Rail Delivery Group

📚 National Rail



Access to paid-for rail consultancy	
services	37 %
Access to paid-for data consultancy	
services	36 %
Access to paid-for development	
services	27 %

Access to consultancy services was rated lowest at 4%

Access to paid for Rail and Consultancy service were most popular however consultancy services in general was seen as the least popular feature and benefit for Data Publishers.

What value would you expect from data you paid to get access to?

Rail Delivery Group

📚 National Rail

"Using rail data to underpin rail ticketing in new green tourism web-app - data must support build of search, booking and ticketing."

"I could answer this more specifically.

Because, as an academia, unless we have some grants, which we don't always have, we then do not have any budget to pay for the access to the data. There should be some consideration or special rates for research. "

"We wouldn't. Data is free, it is not anyone's property to sell as they wish. SMEs like us are adding value to YOU, not the other way round!!!"

"A reliable high quality feed of data with few to no restrictions on use and the ability to market applications or datasets based on that ability." "Data that improves the product that we offer, especially rail industry data from the passenger and freight side. Reliability of data service."

"If I am being paid to access 'open' data, then I'd expect nothing more than 100% 5 star service at every time."

"I do not want to pay for data, I want to add value in its free distribution."

"Up to date documentation. Data provided via robust service endpoints capable of high throughput with low latency."

"That the cost would only cover the cost of supplying it. It is in TOCs' interests to make better data available for free to enable better passenger-facing services to be developed than those the TOCs provide themselves."

What value users expect from of accessing paid data

- Data improves an organisations product offering
- Assist with ticketing and green tourism
- Quality levels of service
- Consideration to academia and data costs lack of budget so rely on grants/good will
- Being able to add value to data
- Up to date documentation
- Robust access to data with high quality feeds
- Ability to market applications based on high quality data feeds
- Data should be free of charge
- Only cost of supplying data should be chargeable

"None of these are important to me. It is only important that the existing data continues to be free of charge. I am very concerned about a move by RDG towards monetisation. I would consider paying for additional data sources above and beyond what is already available if there was commercial interest in products containing the extra chargeable data, which I am currently unsure about."

What data sources would you seek to provide or consume on the Rail Data Marketplace?

Rail Delivery Group

📚 National Rail

"As a researcher, we don't generate any raw data, but would like to consume all the data available from the RDM, if we can afford them."

"RCS and IDMS data. I really don't know why these aren't freely available along with the rest of the fares data, as not having them causes confusion about which fares are actually saleable. It feels like the whole open data initiative ground to a halt when Dennis Rocks left RDG."

"Data covering all rail services in the UK (and possibly Europe if there was scope to expand the project beyond UK data), with timetables and ticketing information."

"Look to consume locomotive/unit allocation data as this can be used in a variety of ways to improve the offering to both general public and enthusiasts." "Consume all data existing free sources, plus possibly seek to consume rolling stock allocations and live carriage loading data. It would be useful to have data with precise geographical location of all signal berths."

"Train positioning data through signalling systems. Onboard GPS vehicle position data where available."

"Departures, Arrivals, Fastest services (inc Bus)"

"I would like to be able to consume gateline data from Cubic and S&B"

"Would depend on what's available"

"All rail data - TRUST movements, TD data, the plan (plus VSTP feed), TSRs and PSRs, network topography and connectivity, vehicle/set allocations."

What data users would like to consume (no data provided on provision of data)

- RCS and IDMS data (fares data)
- Data covering rail services for UK and Europe
- Locomotive/unit/rolling stock allocation data
- Precise location of signal berths
- Train positioning data through signalling and GPS
- Departures, Arrivals, Fastest services (inc Bus
- Gateline data from Cubic and S&B
- TRUST movements, TD data, the plan (plus VSTP feed), TSRs and PSRs, network topography and connectivity, vehicle/set allocations

What else would you need from the Rail Data Marketplace to support your objectives?

Rail Delivery Group

📚 National Rail

"At this stage, we don't know what data you can provide yet, so we cannot say for sure what support we need." "The open rail data has been communicated as being "open" as in free of charge and developers have created their products on that basis. The existing APIs should stay exactly the same to use and access without needing any technical changes on the client side. The format of the data of the existing data feeds should stay exactly the same. It's crucially important that the existing data continues to be free of charge. If this is not possible, at least make it free of charge for existing users, and only charge new users."

- Data should stay open as developers have created products on that basis
- Existing users should still be able to access for free with new users only being charged
- APIs should stay the same without needing technical changes
- Format of the data should stay the same