Office of Rail and Road and Network Rail

#17190 Independent Reporter Review of CRM-P and FDM-R

Final Report

Issue 2 | May 2021

Job number 278975

Ove Arup & Partners Ltd
Scotstoun House
South Queensferry
Edinburgh EH30 9SE
United Kingdom
www.arup.com

ARUP

Contents

Appendices

Appendix A

Independent Reporter Framework Statement of Works #17190 – Review of CRM-P and FDM-R

Appendix B

Freight Delivery Metric – by Region or Route (Definition and references)

Appendix C

Confidence Grading System

Appendix D

List of files supplied to the Reporter Team

1 Executive Summary

1.1 Purpose

In Control Period 6 (CP6), the Office of Rail and Road (ORR) monitors Network Rail's (NR) delivery of passenger and freight rail performance using a range of indicators. The following two consistent metrics are used to compare performance across regions:

- A consistent region measure for passenger services, known as Consistent Region Measure – (Passenger) Performance (CRM-P); and
- A freight delivery metric for each region known as FDM-R.

CRM-P measures the minutes of NR-attributed delay to all passenger trains from incidents occurring within the region normalised by the train kilometres travelled by passenger trains within that region.

FDM-R measures the percentage of commercial freight services that arrive at planned destination within 15 minutes of their booked arrival time or with less than 15 minutes of delay caused by either NR, non-commercial freight operators or passenger operators.

ORR's Final Determination for CP6 originally set baseline trajectories and regulatory minimum floors for CRM-P and FDM-R for each NR route. Following the "Putting Passengers First"[1] programme, these measures are now monitored at a regional level. ORR monitors delivery against annual targets and regulatory floors for each of the five NR regions.

While the National FDM was audited in 2016 by Arup in its role as the Independent Reporter[2], supported by Winder Phillips Associates (WPA), we understand that neither metric has been independently reviewed in its current guise since introduced. The objective of this review was to measure the system reliability and data accuracy of the reported regional FDM and CRM-P metrics in line with the confidence rating grading system shown in Appendix C. Arup, supported by WPA, have undertaken this audit in the role as Independent Reporter.

1.2 Key Findings and Recommendations

1.2.1 CRM-P

Our audit of the CRM-P metric was based on discussions with the National Performance and Analysis Team (NPAT) who are responsible for calculating and

[1] https://www.networkrail.co.uk/putting-passengers-first/
[2] The 2016 review findings 'review-of-freight-delivery-metric-2016-10-24.pdf can be accessed at: https://www.orr.gov.uk/media/16767.

reporting the metric and the NR Process & Controls team who are responsible for the governance of data held within TRUST[3] and PSS[4].

The CRM-P figures are generated via automated reports from Business Objects Performance System Strategy (BOPSS). The process of converting the data within PSS into the metrics is consistent with the metric definition. A data assurance exercise was undertaken to recalculate the metrics from raw PSS data. No concerns were identified with the conversion calculation.

However, it was observed that the metric definition could be mis-interpreted in terms of those delay minutes which should be included within the calculation, specifically with respect to delays suffered on non-NR networks and to train services that have no contractual performance regime with NR (e.g. the East London Line) that are caused by a NR-attributed incident in a NR region. It is recommended that this definition is tightened to ensure clarity.

It was further observed that no process document exists which outlines this calculation, and it is recommended that this is developed.

CRM-P is reported based on the 'Adjusted Data Series' delay database in PSS, which forecasts where incidents in dispute will be finally attributed. This ensures that the reported metrics will not be suppressed by incidents in dispute (which remain assigned to the Operator until resolved) and reduces the 'swings' in historical reported figures as disputes are resolved.

However, a consequence of this is that the historical data used to calculate CRM-P will continually update until all incidents in a period are resolved. This can lead to different figures being reported in the industry. It is therefore suggested that NR and ORR extract the figures on an agreed date each period to ensure consistency of reporting.

Delay is assigned to regions on the basis of the geographical location that caused the delay, not where the train was delayed. This is done via a geographical code ("NR Manager Code") in TRUST which is assigned to each incident based on its location. Daily checks are in place to identify any discrepancies between the NR Manager Code in TRUST and the incident location to enable them to be rectified in TRUST if there is an error.

The governance of data within TRUST and PSS is controlled via the processes and controls outlined in the Performance Measurement Manual (PMM) and Delay Attribution Principles and Rules (DAPR). These documents provide clear guidance on how to assign delay to incidents and outline the controls in place regarding managing reference data within TRUST and PSS. The PMM outlines a set of mandated and recommended verification checks that NR routes and NR Centre should undertake daily and periodically to ensure the quality and consistency of data within TRUST/PSS.

It should be noted that the accuracy and quality of delay attribution was not within the scope of this audit but will influence the reported figures. Specifically

[3] TRUST is the system used to monitor train movements against their schedule
[4] PSS is the NR data warehouse for storing train running and delay data from TRUST

ensuring that incidents are correctly assigned to NR and delay minutes, particularly reactionary delay, are assigned to the right incident.

The Reporter team has considered the reliability and accuracy of CRM-P using the confidence grading system set out in Appendix C and have determined CRM-P to be B1. The reported metrics accurately reflect the underlying data in PSS with a score of 1 (accuracy between 0.1 and 1%). Recommendations are made to produce a process guide to support user training (recommendation 1), further improve accuracy and consistency (2) and remove the ambiguities in the definition of the metric (3 and 4).

When all CRM-P recommendations in Table 1 (recommendations 1-4) are addressed then the reliability grading for this metric would be an A.

1.2.2 FDM-R

Our audit of the FDM-R metric was based on discussions with the Freight Directorate in NR who are responsible for calculating and reporting the metric.

The FDM-R process is more complex than CRM-P, based on extracts of data from BOPSS, complemented with cancellation data from Schedule 8 claims. Data is consolidated in a set of spreadsheets.

The Reporter Team has reviewed the end-to-end process for calculating this metric based on the latest available period of data (period 8 of 2020/21). The majority of data transfer between files is automated via macros, although there are some manual transfer stages which may import some risk.

This review, accompanied by a series of spot-checks and validation calculations, found no material concerns with this process. While a few very minor discrepancies were picked up in the review, these have been discussed with the Freight Performance Analysis and none have a material impact on the reported figures.

Two minor ambiguities in the definition of the metric were identified in our review, which have led to minor inconsistencies in the calculation approach. It is recommended that these are reviewed and resolved. These were:

- The process for attributing trains between regions potentially double counts trains that run on a region on separate occasions on the same schedule; and
- Some cancelled Class 0 trains are included in the metric in contravention to the definition.

Following the audit of the National FDM in 2016, three recommendations were made to improve the resilience of the process; fully documenting the process, strengthening the number of people trained in the process, and defining a set of verification checks. Good progress has been made against each of these.

Our review has highlighted that further enhancements to the documentation would benefit new users given the complexity of the process, and number of files in the model suite. A process document containing a map and description of each BOPSS query and file would further support the documentation already in place.

The grading of the FDM-R has been determined as B1 on the basis of this review.

When all FDM-R recommendations in Table 1 (recommendations 5-8) are addressed then the reliability grading for this metric would be an A.

1.3 Recommendations

Following our review, we have made the following recommendations.

Table 1: Study Recommendations

Number	Metric	Recommendation to Network Rail	Benefits	Evidence of Implementation	Location in Text
SOW17190-1	CRM-P	Produce a process guide for CRM-P	A documented record of the queries used to generate the report, supporting training of new staff	Production of process documentation	Section 3.3.2
SOW17190-2	CRM-P	Document an agreed date each period for ORR and NR to report the CRM-P metric	Ensure that NR and ORR are reporting consistent CRM-P figures each period	Documentation	Section 3.5.1
SOW17190-3	CRM-P	Update definition of delay minutes used for CRM-P metric in "Definitions of Railway Performance Metrics" document to clarify those off-NR network delays that are included	The current definition could be mis-interpreted to mean that all delay minutes suffered on non-NR networks as a result of a NR incident are included.	Definition updated in "Definitions of Railway Performance Metrics" document	Section 3.2
SOW17190-4	CRM-P	Review and agree which of the trains that operate partly on the NR network and partly on non-NR networks should be included in the CRM-P metric	Confirm intention - currently trains without a contractual framework for train performance with NR are excluded from the metric, even if they operate partly on the NR network. This is the case for services operating on the East London Line and could be	Treatment of delays to these trains clarified in "Definitions of Railway Performance Metrics" document	Section 3.2

Number	Metric	Recommendation to Network Rail	Benefits	Evidence of Implementation	Location in Text
			the case for CrossRail services when they start to operate.		
SOW17190-5	FDM-R	Produce a user-guide for FDM-R supplementing the existing 'update guide', covering a process map and description of each query and file	This will support new users to help understand the files used to calculate the metric, and support future enhancements	Production of an enhanced user guide	Section 4.3.2
SOW17190-6	FDM-R	Review FDM-R definition to confirm treatment of: - Trains which enter a route/region more than once - Class 0 trains	This will remove any ambiguity from the definition, and ensure consistency in approach	Clarification of treatment within process documentation	Section 4.3.2
SOW17190-7	FDM-R	Document completion of mandated FDM-R checks as defined in documentation and any issues identified	This will provide confirmation that mandated checks are being undertaken, and any issues raised are documented	Written evidence of checks undertaken	Section 4.5.1
SOW17190-8	FDM-R	Strengthen Service Variation & Cancellations (SV&C) data collation and processing for FDM-R through: - Requesting further information to be supplied by FOCs in support of Schedule 8 claims; and - Reviewing opportunities for further automation of the process to remove reliance on manual updates	Receiving additional data from FOCs in support of Schedule 8 claims (e.g. Train Service Code) will make it easier to identify the relevant trains within PSS, so improve the validation process Automating more of the SV&C data files, including introducing checks, will remove the risk of error through manual data input	Information supplied by FOCs An updated SV&C process	Section 4.4.1.4 / Section 4.4.2.5

1.4 Acknowledgements

The Independent Reporter Team would like to thank both ORR and Network Rail staff for their assistance with this study.

2 Introduction

2.1 Background

Arup, in its role as Independent Reporter, supported by Winder Phillips Associates (WPA) were appointed by the Office of Rail and Road (ORR) and Network Rail (NR) to undertake an audit of the system reliability and data accuracy of reporting of the following regulatory metrics:

- Consistent Region Measure – (Passenger) Performance (CRM-P); and
- Freight Delivery Metric – Region (FDM-R)

The scope of this study was defined in the Statement of Work (SoW) #17190, a copy of which is included in Appendix A.

2.2 Mandate Aims and Requirements

The objective of this review was to measure the system reliability and data accuracy of the reported CRM-P and FDM-R metrics. These metrics are the core regulatory performance metrics monitored by ORR during Control Period 6 (CP6) and so it is critical that ORR, NR and rail industry stakeholders have assurance of the quality of the data and robustness of the measures. While the national FDM was audited by the Independent Reporter team in 2016[5], neither of these metrics have been previously audited in their current guises.

CRM-P is the key metric used by ORR to assess NR's delivery of passenger rail performance to customers during CP6 and forms part of the CP6 scorecards. This is calculated by the National Performance and Analysis Team (NPAT) within NR.

The FDM was originally introduced as a regulatory measure of freight performance in Control Period 5 (CP5). To reflect devolution, regional (and route) level metrics have been introduced (FDM-R). This is calculated and monitored by the Freight Directorate within NR.

ORR's Final Determination for CP6[6] set baseline trajectories and regulatory minimum floors for CRM-P and FDM-R at a route level. Following the "Putting Passengers First" (PPF) programme[7], these measures are now monitored at a regional level. ORR monitors delivery against annual targets and regulatory floors for each of the five NR regions.

The scope of the project required review, comment and any recommendations on the:

[5] The 2016 review findings 'review-of-freight-delivery-metric-2016-10-24.pdf can be accessed at: https://www.orr.gov.uk/media/16767.
[6] https://www.orr.gov.uk/sites/default/files/om/pr18-final-determination-overview-and-decisions.pdf
[7] https://www.networkrail.co.uk/putting-passengers-first/

1. Governance and methodology for transforming data from TRUST/PSS into the CRM-P and FDM-R outputs, including the methodology employed to attribute delay and mileage data between regions;
2. Reliability, quality, consistency, completeness and accuracy of reported data; and
3. Processes in place to produce, quality assure and provide consistent period-end figures to customers, including ORR.

As part of this review, the Reporter Team were required to provide a confidence grading for each metric; covering both an 'alpha' score (system reliability) and 'numeric' score (data accuracy) based on the most up-to-date dataset available at the time of the commission. The grading system for confidence rating is shown in Appendix C.

In terms of the scope of this audit it was confirmed in the inception meeting:

- the attribution of delay minutes between NR and Train Operators is outside the scope of this review;
- although NR reports the national FDM metric, the focus of this study was on the FDM-R metric only.
- while CRM-P and FDM-R are reported by NR at route (14) and region (5) level, this mandate is focused on auditing the regional metrics only, so reflecting those which ORR monitors delivery against annual targets and regulatory floors. However, it is noted that most findings will be relevant to the route-level metrics also.

2.3 Our Approach

The approach that we adopted for this study was designed to provide an assessment of NR's reporting process, procedures and governance, alongside an audit of the underlying data to review accuracy of reported results. Our approach is summarised in Figure 1.

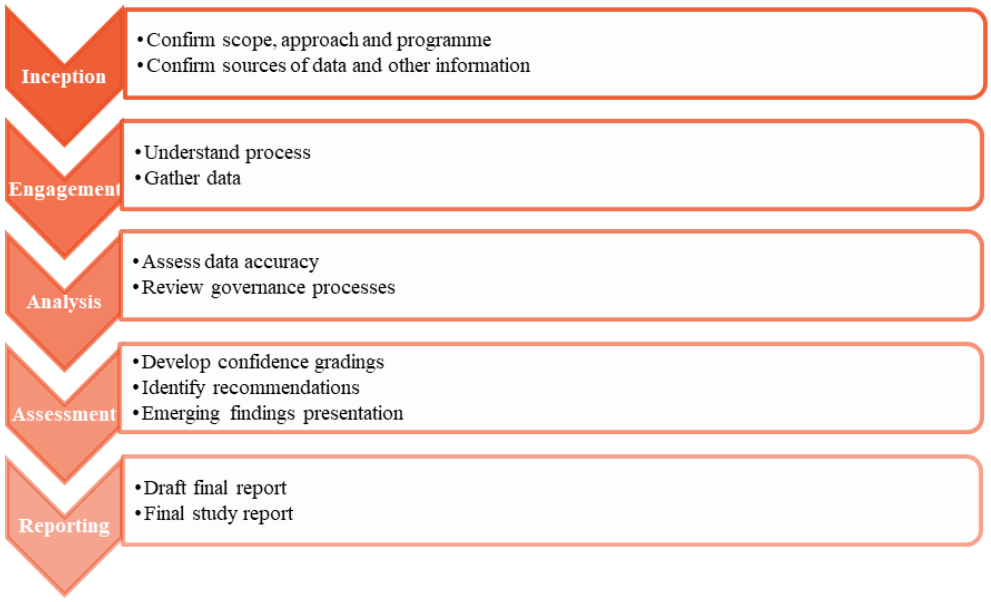


Figure 1: Summary of Audit Approach

During the engagement phase, we held meetings with a number of representatives from NR as summarised in the table below.

Table 2: Meetings held during review

Date	Who	Purpose
24/11/2020	NR / ORR	Project inception meeting
09/12/2020	NPAT Team	Review CRM-P reporting process
09/12/2020	Freight Directorate	Review FDM-R reporting process and calculation spreadsheets
18/12/2020	Process & Controls Team	Review governance and verification arrangements for data captured in TRUST/PSS
07/01/2021	NPAT Team	Follow-up questions on CRM-P data
18/01/2021	Eastern Region Performance Team	Review use of metrics in region, and how delay is attributed between regions
21/02/2021	Freight Directorate	Follow-up questions on FDM-R calculation approach
28/02/2021	NR / ORR	Emerging findings meeting

Following these meetings, the Reporter Team were supplied with data and information from which to undertake our review. A full list of files supplied is included in Appendix D.

2.4 Report Structure

Section 3 outlines the observations and key findings from the review of CRM-P. This has been structured to answer the three questions posed in the Mandate as outlined in Section 2.2, both in terms of system reliability and data accuracy. This section concludes with confidence gradings for CRM-P.

Section 4 outlines the observations and key findings from the review of FDM-R and is structured identically.

Proposed recommendations from this study are provided in Section 5.

2.5 Glossary of Terms

The table below provides a description of the standard rail industry acronyms and abbreviations that are used in this report.

Table 3: Glossary of Terms

Abbreviation	Description
ADS	"Adjusted Data Series" Universe in PSS
BOPSS	Business Objects Performance Systems Strategy
CP6	Control Period 6 (the 5-year period 2019-2024)
CRM-P	Consistent Regional Measure – Passenger Performance
CVL	Core Valley Lines
DAB	Delay Attribution Board
DAPR	Delay Attribution Principles and Rules
ELL	East London Line
ECS	Empty Coaching Stock
FD	Freight Directorate (within NR)
FDM	Freight Delivery Metric (National measure, not audited here)
FDM-R	Freight Delivery Metric – Regional (note, sometimes termed R-FDM)
FOC	Freight Operating Company
FPRS	Freight Performance Regime Specialist
GRAI	Governance Risk Assurance and Improvement (NR process)
HAL	Heathrow Airport Limited
HS1	High Speed 1
MAA	Moving Annual Average
MFSDD	Management of Freight Services during Disruption
NPAT	Network Rail's National Performance & Analysis Team
NR	Network Rail
ORR	Office of Rail and Road
PDAC	Performance Data Accuracy Code
PfPI	Planning for Performance Improvement
PMM	Performance Measurement Manual
PPF	"Putting Passengers First" NR programme
PSS	Performance Systems Strategy
ROSCO	Rolling Stock Leasing Company
SV&C	Service Variations & Cancellations (for FDM)
TfW	Transport for Wales
TOC	Train Operating Company
TRAA	Reference data held in TRUST
TRUST	Train Running Systems TOPS
VSTP	Very Short Term Planning

3 Findings from CRM-P Metric Audit

3.1 Overview

This section summarises the findings from our review of the process, governance and data accuracy related to the CRM-P metric. A description of the metric is provided, followed by sections outlining our findings and observations related to each of the three questions in the Mandate:

1. Governance and approach for converting TRUST/PSS data into the metric;
2. Reliability, quality, consistency, completeness and accuracy of the underlying data used to calculate the metric; and
3. Processes to produce, quality assure and provide consistent period-end figures to customers, including ORR.

A confidence grading is provided at the end of this section based on the findings of our review.

3.2 CRM-P Metric Definition

3.2.1 Overview

The CRM-P metric is officially defined within the "Definitions of Railway Performance Metrics_v3.03" document. This document is clearly version controlled and dated (September 2020).

The one-line description of CRM-P is:

> *"The amount of attributed delay that a Network Rail route has caused to in service passenger trains per 100 train kms in service passenger trains have travelled on that route."*

While this description refers to "routes", as noted in Section 2.2 the metric is now monitored at a regional level following the PPF programme. For the purposes of the regional metric, "route" can be replaced by "region" in this definition.

As well as a one-line description of the metric, this document includes assumptions of what data should be included within the delay and kilometres figures.

3.2.2 Delay Minutes

The document defines delay minutes as "*The sum of all Attributed Network Rail delay, including sub threshold and attributed unexplained delay, suffered by any in service passenger train which is associated with an incident that occurred within the boundary of the Network Rail Route no matter where the train suffered that delay*".

Delay minutes are assigned to regions based on the location of the incident that caused the delay, regardless of where a train actually suffered the delay. For example, any delay minutes suffered in Eastern region as a result of a NR-caused incident in Southern region would be included in the Southern CRM-P calculation.

3.2.3 Off-NR Network Delay Minutes

Delays as a result of incidents occurring on non-NR infrastructure are excluded. Such infrastructure currently includes the Core Valley Lines (CVL), Heathrow Airport Link (HAL), High Speed 1 (HS1) and any TfL-controlled infrastructure.

From discussion with NR, it was agreed that the final clause in the definition of delay ("no matter where the train suffered that delay") needs to be clarified. While some 'off-NR network' reactionary delay caused by a NR-attributed incident is included in CRM-P, not all of it is.

NR has clarified that the delay used for CRM-P is derived from PfPI minutes that are captured within TRUST and attributed in accordance with the Delay Attribution Guide. Therefore, any delay suffered on networks where this is the case, such as the CVL or HAL, that is caused by a NR-attributed incident would be assigned to the relevant NR region in which the incident occurred.

However, some off-NR network delay is excluded from the metric. Examples include:

- LUL network: only the NR-owned lines are within TRUST so any reactionary delay on the LUL network is excluded;
- Supertram; only Rotherham is within TRUST so any reactionary delay elsewhere on the tram network is excluded; and
- Heathrow Terminal Shuttles; while these trains are included in TRUST, they are not an advertised train service, so delay to these services is excluded.

3.2.4 East London Line Trains

More significantly, reactionary delay minutes suffered by Arriva Rail London (ARL) trains on the East London Line (ELL) network are also excluded. These trains operate under Service Code "22218000" on both the NR network and the TfL-owned network north of New Cross Gate and Queens Road Peckham. While these trains are in TRUST and delay is attributed in accordance with the Delay Attribution Guide, they are not covered within the contractual framework with NR. As such these trains are assigned to a 'non-PfPI' service code and so any delay minutes suffered by these trains are excluded from all NR delay reporting (including CRM-P) – both on the NR network and the TfL-owned network.

This does raise a question as to whether such delay minutes ought to be included within the CRM-P metric. Analysis by NR showed that there were 4,839 delay minutes to ARL services on the ELL as a result of NR-attributed incidents which occurred in Southern region in Periods 1 to 12 of 2020/21 (this represents 0.4% of PfPI delay in that region). If these delay minutes were included, this would have a

very small impact on the CRM-P MAA for Southern region (increasing by 0.006 minutes). That said, it is noted that a similar situation will arise when MTR Elizabeth Line trains commence cross-London operation.

3.2.5 Train Kilometres

The distance data for each region for this metric is defined as that which is measured in TRUST/PSS.

This means that whereas delays to trains operated under Service code "22218000" are excluded from CRM-P, their train kilometres on the NR network are included. NR has stated to us they believe this is appropriate because it reflects the busyness of the network. Also these kilometres only account for a very small proportion of the total; 0.27% of train kilometres on Southern region.

3.2.6 Publicly Advertised Trains

Section 2.3.1 of the definitions document referred to above states: "*All the metrics include the performance of all publicly advertised trains (in line with the conditions for the Applicable Timetable) operated by these companies even if the service operates fully off the Network Rail owned network (for example Dalston Junction to New Cross) unless the service is non-revenue earning and not part of a franchise agreement.*"

During this review NR has clarified that this was written to ensure all publicly reported performance metrics for operators included all passenger trains, irrespective of who owned the network they travel on, to avoid any potential confusion. They further clarified the statement is not necessarily applicable to network owner metrics and not for CRM-P.

3.2.7 Conclusion

On the basis of these observations, it is recommended that the definition of CRM-P is tightened to clarify the delays to which trains and at what locations are to be included within the metric, particularly in relation to the ELL. This will need to take into account how delays are recorded and attributed on non-NR networks.

3.3 Governance and Approach for Converting TRUST/PSS Data into CRM-P

3.3.1 Outline of Metric Calculation Approach

The CRM-P metric for each region is calculated directly within NR's Business Objects – Performance System Strategy (BOPSS) reporting tool. This report is fully automated, based on two standard template queries:

- Extract periodic NR-caused delay by region from the 'Adjusted Data Series' (ADS) Universe within the PSS database; and

- Extract periodic train kilometres ran by region from the 'Mileage Universe' within the PSS database.

Each of these queries was provided to the Reporter Team for review along with extracts of data from PSS for validation.

3.3.1.1 CRM-P Delays Query

There are two databases within PSS containing attributed delay minutes:

- "Attributions Universe": all delay minutes based on current dispute status, where all minutes in dispute between NR and Operators are attributed to the Operator; and
- "Adjusted Data Series (ADS) Universe": applies an algorithm to estimate where disputed minutes are expected to be re-allocated to once settlement is reached, based on historical resolution of such incidents.

NR use the ADS Universe for reported CRM-P delay to ensure it is not under-stated due to disputed minutes, particularly for the most recent periods. This means that changes in historically reported CRM-P as disputes are resolved are significantly smaller than if the Attributions Universe data was used.

The ADS Universe only contains delay minutes associated with commercial passenger and freight services (PfPI minutes). The diagram summarises the filters applied by the query, which are consistent with the definition.

Delay minutes in the ADS Universe are assigned to regions based on the geographical location of the incident that caused the delay. This is done via the "Network Rail Manager Code (Geography)" (hereafter referred to as NR Manager Code) assigned to each incident in TRUST. Each NR Manager Code is uniquely assigned to a region within PSS.

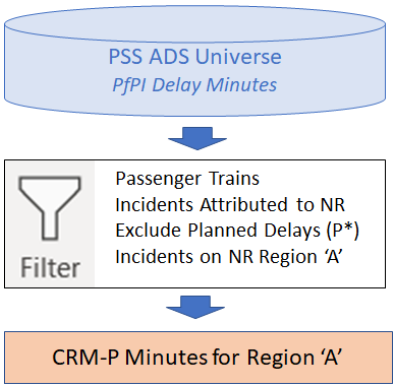


Figure 2: Application of query filters when extracting data from PSS ADS Universe

3.3.1.2 CRM-P Mileage Query

Distance data is held within the 'Mileage Universe' in PSS. The currency of distance data in PSS is miles, so the report is designed to convert to kilometres.

The distance query filters on all passenger trains with a 'PfPI' Train Service Code only. The only exception is to include those ARL passenger services on the ELL which operate under a 'non-PfPI' code but do operate some kilometres on the NR network (~220,000 km per year which represents 0.27% of all kilometres operated on Southern Region). As noted in the previous section, delay minutes suffered by these trains are currently excluded from the CRM-P metric.

We have reviewed the full list of Train Service Codes within the PSS reference data to confirm that no relevant Service Codes are being excluded.

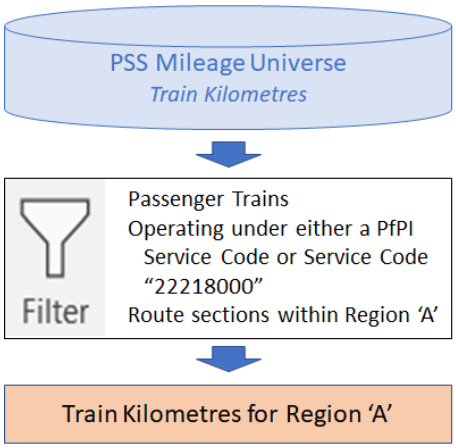


Figure 3: Application of query filters when extracting data from PSS Mileage Universe

3.3.2 Governance of Metric Calculation Approach

The automated CRM-P report is owned by the NPAT team and stored within a 'NPRT Controlled Folder' within Business Objects. A 'Principles and Standards' document covers governance of reports within this folder, although this is in draft format currently, pending approval.

This document covers the approval process for creating and editing reports within this controlled folder to ensure they adhere to required standards and definitions. Specifically, reports must be approved by at least two members of the reporting team. This approval is documented in a reports' register. Evidence of testing is required prior to application of approval, although the required testing process has not yet been included within this document. There are strict controls on 'edit' access within this controlled folder, limited to five named members of the NPAT team.

The CRM-P report is rarely required to be changed. The most recent change was a result of the transfer of the CVL infrastructure from Wales & Western NR region to Transport for Wales (TfW) in February 2020. A new NR route code ("T") was set up for the CVL, so an additional clause was added to the report to ensure delays assigned to NR Manager Codes on this new NR route were excluded.

While the metric is defined in the industry definitions document and there is a documented approval process, we observed that there is no documentation related to this BOPSS report. We recommend that a brief process document is produced which outlines the queries required for the report, the filters that are required to be applied and any appropriate verification checks particularly when circumstances change.

3.3.3 Data Assurance of Metric Calculation Approach

To assure this calculation, we were supplied with raw unfiltered delay and distance data from PSS for period 9 of 2020/21, along with outputs from the CRM-P report. To avoid discrepancies due to resolved disputes, both datasets were extracted from PSS on the same date (7th January 2021). Given the CRM-P report is auto-generated from BOPSS, a deep-dive into a single period would highlight any potential concerns.

Delay Minutes

Delay data was extracted from the ADS Universe by NR Manager Code, Operator Type and Incident Category. This reflects the deepest granularity of data held within this Universe.

From this data, we recalculated CRM-P delay minutes by NR region. This matched exactly the figures generated by the CRM-P report and are shown in the left-side of the table below.

We also reviewed those delay minutes excluded from the CRM-P report to ensure that they had been correctly excluded. Firstly, the review confirmed that all minutes attributed to an Operator code (7**) or to planned delays (8**) had been excluded. Any other delay minutes excluded from the CRM-P report were confirmed to be related to incidents attributed to an "NR" incident category, but which occurred on non-NR infrastructure[10]. This is summarised on the right-side of the table below.

This confirms that the BOPSS report is extracting delay data as intended.

Table 4: Summary of NR-Attributed Delays for Period 9 (2020/21)

NR-Attributed Delay: Included in CRM-P	Delay Minutes	NR-Attributed Delay: Excluded from CRM-P	Delay Minutes
Eastern	120,995	Core Valley Lines (TfW)	3,116
North West & Central	82,258	HS1 Network (HS1)	340
Scotland	47,064	Heathrow Spur (HAL)	167
Southern	119,633	East London Line (RfL)	31
Wales & Western	45,019	Chiltern Met Line (RfL)	19

Kilometres

Distance data was extracted from the Mileage Universe by individual train and NR route (including non-NR infrastructure). From this data, we recalculated the distance by region to confirm it matched exactly the figures produced by the CRM-P report.

Further detailed analysis was undertaken for trains where the total train kilometres did not match the sum of the region kilometres, i.e. some distance was covered on

[10] For incidents occurring outside NR infrastructure, the "NR" incident category codes are used to designate the relevant Infrastructure Manager, as outlined in DAB Process & Guidance Document 22, as published on the DAB website.

non-NR infrastructure. This confirmed that all excluded kilometres were indeed on one of the networks listed in Table 4 above. For example, our analysis showed that trains from Aberdare to Barry Island were correctly shown as operating for 15km in Wales & Western region (included in CRM-P) and 37km on the CVL (excluded from CRM-P). This analysis highlighted no concerns.

CRM-P

We calculated CRM-P for each region from the raw BOPSS data and confirmed that this matched the figures from the CRM-P report exactly.

On this basis, no concerns have been identified with the accuracy of the method for converting PSS data into the reported CRM-P figures.

3.4 Reliability, Quality, Consistency, Completeness and Accuracy of Reported Data for CRM-P

In this section, we focus on the reliability and accuracy of the underlying delay and distance data within TRUST/PSS as used to calculate the CRM-P metrics.

NR has developed a series of processes and governance arrangements for ensuring the quality and completeness of data stored in TRUST and PSS, which are defined in a number of documents.

The Performance Measurement Manual (PMM) outlines a clear set of controls and guidance relating to the management of performance-related data in TRUST and PSS. This document is owned and maintained by the NR Process & Controls team and aims to drive consistency and high standards to performance measurement processes. This is measured via a set of KPIs outlined in Module 27 of this document, supported by a series of mandated and recommended verification checks to be undertaken by both NR routes and NR Centre.

The Delay Attribution Principles and Rules (DAPR) provides a comprehensive guide to creating incidents in TRUST and attributing delay and cancellations to these incidents. This document is owned and maintained by the Delay Attribution Board (DAB), a cross-industry governance body.

NR has developed a set of 'Governance Risk Assurance and Improvement' (GRAI) documents, of which two specifically relate to managing network performance:

- Managing [network performance] data and data reference points; with the objective of "ensuring that the measurement systems are in place to enable comprehensive capture of performance data, and that processes are followed so that the capture of train reporting data meets or exceeds industry standards"
- Managing the delay attribution framework and resolution of disputes; with the objective of ensuring "the accurate identification and allocation of the causes of Minutes Delay, Cancellations, Diversions and other events."

These documents clearly define the process owners and required outcomes, along with core process activities, controls and responsibilities.

3.4.1 Delay Data within TRUST/PSS

There are three critical factors relating to delay data within PSS which may influence the accuracy of the delay figures used for CRM-P calculations:

- Correct attribution of each incident to each region;
- Correct attribution of responsible organisation to each incident; and
- Correct attribution of delay minutes to each incident.

This section is also relevant to the FDM-R metric, which relies on attributed delay data held within TRUST/PSS.

3.4.1.1 Attribution of each incident to each region

Every location in TRUST has a default NR Manager Code, as defined in the TRAA tables in TRUST and PSS reference tables. These are 4-digit alpha-numeric codes which represent a specific geographical area on the network, both within NR regions and on private infrastructure.

When an incident is created in TRUST, the NR Manager Code field is automatically populated based on the geographical location of the incident, so reducing the risk of manual error. This can be amended in TRUST if required, for example if more accurate information is forthcoming on the specific location of the failure that caused the incident. The NR Manager Code is updated via a drop-down menu to ensure that valid codes are entered.

Process and Governance

Clear guidance on defining the NR Manager Code for an incident is provided to Delay Attribution (DA) staff in Section B6 of the DAPR, along with a number of examples including where incidents occur close to the boundary between two management areas.

Module 27 (Key Performance Indicators and Verification Checks) of the PMM mandates NR routes to audit NR Manager Codes assigned to incidents in TRUST daily to ensure they are valid. Compliance reports are required to be provided periodically.

To support this, the NR Process & Controls team produce a daily report flagging any incidents live in TRUST (so for the last 8 days) where the NR Manager Code does not match the default for the incident location. It was noted in discussion with the Process & Controls team that such a discrepancy is not necessarily an error, but this check allows a review to be undertaken. The daily report is provided to NR routes for review and amendment in TRUST if required.

The TRUST TRAA tables contain the mapping of location to NR Manager Code to enable population in this system. A more detailed mapping is contained in the reference tables within PSS to enable reporting, since PSS contains many more locations than TRUST.

The processes for changes to the NR Manager Codes are outlined in Modules 5 and 6 of the PMM. Module 5 provides the documented process for adding,

amending or removing NR Manager Codes within both TRUST and PSS, including the approvals process and who has authority to make changes to systems. Module 6 contains a table documenting the activities that should be undertaken when re-franchising occurs, or changes are applied to geographical boundaries such as in 2020 with the PPF programme.

We were advised by the Process & Controls team that both of these modules are currently being enhanced to provide a more detailed step-by-step guide to support the NR routes when such changes are required.

The reference tables have recently been reviewed in detail and updated to reflect the organisation changes within NR in terms of routes and regions as part of the PPF programme. A spreadsheet summarising the changes required to the TRUST and PSS reference data relating to Phase 2 of this programme for Eastern region was provided for review. This related to changes to the geographical boundary between the North & Eastern route and East Coast route. This provided documentary evidence of how such changes to this data are recorded and was the basis for review and verification of the changes required.

Data Assurance

Based on the controls outlined above, the risk of error in terms of incidents coded to the wrong location is low. Automatic population of default codes, along with daily checks to verify any discrepancies provides a good level of assurance of the process.

An example daily report was provided for 16th December 2020. This showed a total of 80 incidents live in TRUST (so covering the period 9th December to 16th December) where the populated NR Manager Code did not match the default. Just 11 of these showed a discrepancy in terms of the assigned NR region. With over 10,000 incidents live in TRUST on this date, this represents less than 0.1% of incidents. Following review of the report, NR routes may update some of these incidents, so this represents a 'maximum case' of error for this snapshot. Indeed, 7 out of the 11 incidents noted above were on the most recent 2 days of the report.

This process and verification checks provides assurance that the risk of incorrect attribution of incidents to regions is very small.

3.4.1.2 Attribution of Responsible Organisation to Incident

The quality of attribution of incidents between Operators and NR is clearly an important factor on the accuracy of the reported CRM-P figures. While attribution of delays is out of scope for this audit, we provide some observations on delay attribution processes from our review in this section and the subsequent section relating to attribution of minutes to incidents.

The Schedule 8 financial regime is a long-established process which provides a strong incentive for both Operators and NR to ensure incidents are correctly attributed to the right party. Where an Operator disagrees with incidents attributed to them, they will raise a dispute and it will be investigated in further detail.

Module 24 of the PMM outlines the process for disputing and resolving incidents that have been attributed to a NR Responsible Manager, including responsibilities. There are strict time limits for raising a dispute if it believed that the responsible party is an Operator, and so the processes have been developed to reflect this.

Module 27 of the PMM outlines a set of KPIs to monitor the number of incidents in dispute and aim to resolve these as quickly as possible so that data in PSS can be 'finalised' as early as possible.

The Process & Controls Team produce periodic reports on delay attribution data quality to DAB. A KPI which aims to reflect the quality of Level 1 delay attribution is the proportion of incidents for which the incident code remains unchanged after Day 1. This was 92% for the year to Period 9 of 2020/21. As a comparator, this figure was 83% in 2012/13 when the process was last independently audited[11].

Module 27 of the PMM also outlines a set of mandated and recommended verification checks for both the NR routes and the National Centre, as outlined in Section 3.5.2 of this report.

3.4.1.3 Attribution of delay minutes to each incident

Cross-boundary delay can represent a potential risk to the accuracy of the CRM-P figures. This may arise where a train suffers reactionary delay in region A, and this is erroneously assigned to an incident in region A, when the root cause was an incident in region B. The DAPR describes the agreed process for attributing reactionary delay minutes to the cause incident to ensure consistency. However, there inevitably remains some risk of delays not being attributed in line with this guidance.

We were advised that Attribution Managers in the NR routes do undertake a review of the biggest incidents occurring each day, which may pick up some errors in attribution. Similarly, representatives from each Responsible Manager in NR are assigned to check and validate delays assigned to their Responsible Manager codes, which provides the opportunity for challenge. Issues may be picked up if, for example, a higher than usual number of delay minutes has been assigned to a particular incident type.

However, it is much more difficult to measure and monitor this post-attribution with the current systems, since it involves manual investigation of each train delay. We were advised that it can take an hour to review 20-30 individual train delay events. Some larger incidents may have in excess of 1,000 train delay events.

To support this, the Process & Controls team have provided a tool to NR routes to show the chain of delays assigned to incidents. This is based on a tool initially developed by a Freight Operating Company for this purpose and which was shared with DAB to help support such investigations.

[11] The 2013 audit 'AO/039: Review of Performance Measures' can be accessed on the ORR website here.

While it is not possible to quantify the potential risk without a detailed audit of delay attribution, our assessment is that the risk to CRM-P is likely to be relatively low. Cross-regional delay is a relatively low proportion of the total, certainly less than 10%.

3.4.2 Distance Data within PSS

There are two critical factors relating to the underlying distance data within BOPSS which may influence the accuracy of the reported CRM-P figures:

- The completeness of distance between locations captured by the BOPSS queries; and
- The accuracy of the distance between each pair of locations that is recorded within PSS.

The latter issue is out of scope for this audit, since the CRM-P definition explicitly states that the distance data stored in PSS is to be used for normalising delay.

PSS Distance Update in Summer 2020

As part of the Putting Passengers First (PFF) programme, NR created the five regions and restructured the routes to provide greater local focus on decision making. As part of this process, the TRUST and PSS reference data was required to be updated to ensure that default NR Manager Codes for each incident location reflected the new route and regions structure.

As part of this process, it was observed that there were a number of locations within the Mileage Universe with no NR route assigned. When running the BOPSS query to export distance by NR route or region, any kilometres involving these locations were excluded. It was estimated that ~1-2% of kilometres were being missed. To improve data capture, the reference data was updated to ensure all locations were allocated to a NR route. Given this reduced CRM-P slightly, regulatory targets were adjusted to neutralise the effect.

The update and infill of the TRAA tables in TRUST and PSS reference tables was led by the P&C team. Revised reference data was sent to NR routes for verification. It was noted that it would be in the interest of the NR routes to raise any concerns to avoid any delay incorrectly attributed to them. This was implemented, along with revised CRM-P targets, in July 2020.

Governance

As described in Section 3.4.1.1, Module 5 of the PMM provides guidance on changes to TRUST (and PSS) reference data. This includes adding new locations, to ensure that all reference data is updated with all relevant fields. This module is currently being updated to provide clearer guidance on the process.

It was observed by the Process & Controls team that a lot of work had gone into ensuring that all locations had been assigned to a NR route, so they want to ensure clear processes are in place to maintain this. To support this, they are developing a

new set of reports to check for missing mileage in train schedules in TRUST to provide a further back-check.

Data Assurance

As noted in Section 3.3.3, we have assured that the BOPSS queries are extracting the correct data from PSS for the CRM-P report. We have undertaken two further detailed checks to ensure that the data held within PSS is complete, based on period 9 of 2020/21.

Firstly, data was extracted for all in-service passenger trains (individually) from the Mileage Universe with no NR route or region filter applied. This ensures that the full end-to-end train kilometres were extracted, including off-network distance. For each train, this was compared to the sum of the kilometres by region, including off-network. This showed an exact match for every train, confirming that all relevant train kilometres held within the Mileage Universe are being applied to CRM-P.

The second test was to confirm these end-to-end train kilometres for each individual train in the same period reflected the distance data stored in the separate 'PPM Universe' in PSS. This database does not include any reference to NR routes or regions, so distance data is stored in a different way, but the overall values are expected to be consistent.

For the majority of trains, the total kilometres were identical within a tolerance (within 50 metres). One minor exception was identified where, over a two-week period, kilometres between Newark North Gate and Lincoln for a number of trains were shown as zero in the 'Mileage Universe'. The reason for this is currently under investigation within NR. It is noted that this only occurred for a two-week period on a handful of trains, so that the discrepancy is less than 0.1% of total Eastern region kilometres in the period. This had no impact on the reported CRM-P figures for this region.

3.5 Processes to produce consistent, quality assured period-end CRM-P figures to customers

3.5.1 Consistency of Reported Figures

The use of an automated report to generate CRM-P figures each period ensures consistency of approach. The results are stored in a single location, on the NR SharePoint site, so ensuring everyone will have access to the same figures on a given day.

PSS data is refreshed daily to reflect changes, for example as the attribution of incidents in dispute is resolved. The data held within the ADS Universe will update to reflect actual attribution for newly accepted incidents. This means that historically reported CRM-P figures will be revised.

Therefore CRM-P figures for a given period and extracted on different days may differ. We note that the use of the ADS Universe ensures that such changes would be significantly smaller than if the Attributions Universe was used.

This is an important point to note to ensure reporting of the metric is consistent across the industry each period. In discussion with NR and ORR at the Emerging Findings meeting for this audit, it was proposed to agree and document a consistent date each period for both companies to extract the metrics from the SharePoint site for onward reporting.

To provide an indication of the potential magnitude of changes in the CRM-P MAA as disputes are resolved, we have compared the reported figures for period 9 of 2020/21 based on two dates:

- 16 December 2020 when we received the initial data samples from NPAT; and
- 26 January 2021 based on reported figures on NR SharePoint site

The table below shows that changes are very small, typically just over 0.1%, indicating that the process is not materially under-representing the true value.

Table 5: Comparison of CRM-P MAA figures as extracted on 16/12/20 and 26/01/21

CRM-P MAA 2020/21 Period 9	Extracted 16/12/20	Extracted 26/01/21	Change: Minutes	Change: Percentage
Eastern	1.197	1.199	0.002	0.14%
North West & Central	1.362	1.363	0.001	0.10%
Scotland	1.072	1.073	0.001	0.11%
Southern	1.908	1.912	0.004	0.17%
Wales & Western	1.302	1.303	0.001	0.06%

When any changes are required to be made to the TRUST/PSS reference data, for example remapping of locations to NR Manager Codes following NR route boundary adjustments, all historical data in PSS is refreshed to reflect this change. From this point, any figures reported via Business Objects will be based on the updated reference data, and so may be inconsistent with previously reported figures. As previously noted, the process for such changes is outlined in Module 5 of the PMM to ensure there is clear governance.

3.5.2 Quality Assurance

Once the CRM-P report has been approved, there are no further mandated verification checks of the reported figures each period within NPAT.

Within the regions, CRM-P is reported at the Quarterly Business Reviews and Quarterly Route Reviews. Route performance teams will review underlying data to ensure that the observed changes in the metric can be explained. Eastern region, who we engaged with during this project, did not raise any concerns with the reported metrics. They observed that this KPI is useful for understanding performance within the route or region at ‘Exec’ level, but their day-to-day focus remains on monitoring the critical performance indicators of incidents, delay minutes and Time-3 punctuality.

Independent verification checks are undertaken on the CRM-P figures produced for the Annual Return. The documents relating to the 2019/20 Annual Return

were provided for review, which documents the reviewers for each metric within the Annual Return and their findings.

As outlined in Section 3.4, Module 27 of the PMM outlines the KPIs through which NR monitor the quality of performance data capture, including delay attribution. This is supported through a series of mandated and recommended verification checks, both daily and periodic.

3.6 CRM-P Conclusions and Confidence Ratings

Our audit has identified no material concerns with the accuracy of the reported CRM-P metric. Minor concerns with system reliability have been noted with respect to the definition of the delays to be included within the metric, specifically off-NR network delays, and the lack of evidence of process documentation for this metric.

Automatic reporting of the metric ensures a consistent approach to extracting the data. Using the 'ADS Universe' minimises changes to historical reported metrics as disputes are resolved.

The approach for assigning incidents to regions is clear and consistent and well audited by NR in terms of allocation of incidents to the regions. There are clear governance processes and guidance in place to drive consistency and quality in delay attribution, with a number of mandated verification checks outlined in the PMM, although this area was out of scope for this audit.

The extraction of train kilometres from PSS now appears to be comprehensive and in line with the metric definition.

Reflecting the minor concerns identified, the CRM-P metric is awarded a 'B' for overall reliability.

The ambiguities in the definition of the metric lead to a reliability score of B (minor shortcomings). Producing a process guide would also support user training.

The table below summarises a grading for each of the criteria considered for this award, along with a list of the key evidence provided to support the grade.

Table 6: Reliability Grading for CRM-P

Element of Process	Criteria to be met	Grade	Evidence Provided
Objectives	Clear and unambiguous description of the purpose and objectives of producing the metric	B	"Definitions of Railway Performance Metrics_v3.03.docx" Clarity required in this document with respect to define which off-NR network delays are included in the CRM-P metric
Requirements	Clear and unambiguous description of the standards required for the data and its collation	A	PMM DAPR

RACI	Clear identification of those Responsible for, Accountable for, Consulted about and Informed about the metric	A	Discussion with NPAT / Eastern region
Source(s)	Description of who or what (system) provides the data	A	Discussion with NPAT
Means and Frequency of Data Provision	Description of how the data is provided, how often, and when	B	Discussion with NPAT; not explicitly documented
Data Format and Expected Values	Definition and description of the format in which the data are to be supplied, and the expected range (if any) of values	A	Discussion with NPAT Expected values (targets) displayed on NR SharePoint site with actual data
Data Quality	Definition and description of the required data quality and accuracy	A	PMM and DAPR Sample TRUST verification checks provided by Process & Controls team
Data Processing	Documentation of processes, sufficiently clear for new users	B	Discussion with NPAT; screen print of BOPSS queries; not explicitly documented
Staff Training	Sufficient availability of trained staff to maintain data and produce the metric	A	Discussion with NPAT; five named individuals have edit-access to report
Data Checking and Verification	Description of mandated and recommended verification checks to ensure accuracy of reported metric	A	PMM Discussion with NPAT / Process & Control team / Eastern region
Managing Change	Description of process for updating underlying data to reflect major changes, e.g. route/region restructuring	A	PMM Discussion with NPAT / Process & Control team Evidence from July 2020 Mileage Universe Update
Internal Review and Audit Procedures	Description of internal review and audit requirements, processes and frequencies; evidence that these are being met	A	Discussion with NPAT 2019/20 Annual Return audits

The CRM-P is awarded a '1' for accuracy

The assessed and potential data inaccuracies with the reported metric falls between a range of 0.1% and 1%, on the basis that delay attribution is being correctly undertaken. The rationale for this grade reflects two factors.

Firstly, the fact that the reported MAA will only be finalised once all historically disputed incidents are resolved. As a result, the CRM-P figures for a given period may differ when extracted on different days. However, the difference between initially-reported figures and 'final' figures will be minimised through use of the ADS delay data. The latest period MAA has been observed to change by just over 0.1% for most regions.

Secondly, as outlined in Section 3.2, there is some ambiguity as to whether delays caused by NR that are suffered by ELL trains operating under Service Code

“22218000” ought to be included within the metric. Analysis by NR shows that the impact of including these delays would increase the CRM-P MAA for Southern region by 0.4% or 0.006 minutes.

4 Findings from FDM-R Metric Audit

4.1 Overview

This section summarises the findings from our review of the process, governance and data accuracy related to the FDM-R metric. As for the CRM-P metric, a description of the metric is provided, followed by sections outlining our findings related to each of the three questions in the Mandate:

1. Governance and approach for converting TRUST/PSS data into the metric;
2. Reliability, quality, consistency, completeness and accuracy of the underlying data used to calculate the metric; and
3. Process to produce, quality assure and provide consistent period-end figures to customers, including ORR.

A confidence grading is provided at the end of this section based on the findings of our review.

While the FDM-R metric has not been independently reviewed since its introduction in 2019, the national FDM was audited by Arup and WPA under the Independent Reporter framework in 2016, and three recommendations were made. We have observed significant progress against each of these, as outlined below.

Table 7: Recommendations from 2016 FDM Audit

Reference	Recommendation	Status
L3 AR 001/01	Get the processes and documentation reviewed, formalised and up to date	Two version-controlled documents produced covering the definition of the metric and a step-by-step guide to updating the metric
L3 AR 001/02	Undertake some internal, structured data checks on both ‘included’ and ‘excluded’ data for FDM	Mandated checks on ‘included’ and ‘excluded’ trains are defined in the documentation.
L3 AR 001/03	Training up / developing other personnel to undertake specialist functions in relation to FDM within the Freight Performance team (for example, cover for the Freight Performance Analyst and Freight Performance Regime Specialist.)	The team now includes two analysts trained in the calculation process and two Freight Performance Regime Specialists to provide cover for each of these roles.

The 2016 FDM review examined the National metric in significant detail and where possible the Reporter team has tried to avoid covering the same ground. This is to ensure that the focus of this review is on the features and processes that are unique to the regional variant of the metric. For further information on the

review of the National FDM metric, the 2016 report can be accessed via the link below[12].

4.2 FDM-R Metric Definition

The FDM-R metric is officially defined within the "*Definition Region and Route Freight Delivery Metric 20201120.docx*" document. This document is version controlled and dated (November 2020).

A full description of the metric definition is provided in the above documentation, however this extract, taken from the 'Metric defined in words' section of the document, provides a succinct summary:

> *"The metric is the percentage of trains which Network Rail has delivered successfully. Failed to deliver is the percentage of commercial freight services that do not reach their destination within 15 minutes of their booked arrival time; and which have either been cancelled, or delayed 15 or more minutes, by Network Rail or another operator that is not a commercial freight operator (FOC)"*

This metric description is supported by a mathematical definition which takes the following form:

$$1 - \left(\frac{\sum \text{FDM Delayed Trains by geography} + \sum \text{NR Assumed Cancelled Trains by geography} + \sum \text{NR caused Service Variations by geography}}{\sum \text{Trains Ran by geography} + \sum \text{NR Assumed Cancelled Trains by geography}}\right)$$

A full description of each term in the formula above is provided in the FDM-R definitions document (see Appendix B). To summarise:

- *Trains Ran*: The number of (FDM-R Qualifying) trains which departed their planned origin and arrived at their planned destination (no matter what route was taken) and no matter how late.
- *FDM Delayed Train*: A (FDM-R Qualifying) train that departs its planned origin and arrives at its planned destination 15 or more minutes late and has 15 or more minutes of delay attributed in total to Network Rail or another operator that is not a commercial FOC.
- *NR Assumed Cancelled Train*: The number of cancelled (FDM-R Qualifying) trains caused by Network Rail or another operator that is not a commercial FOC.
- *NR caused Service Variation:* The number of (FDM-R Qualifying) trains which were subject to a service variation e.g. diversion, partial cancellation, schedule alteration caused by Network Rail or another operator that is not a commercial FOC.

For FDM delayed trains, delay minutes are assigned to each NR region in the same way as for CRM-P (as described in Section 3.2), based on the location of the incident that caused the delay, regardless of where a train actually suffered the delay. Each FDM delayed train is then allocated between regions in proportion to

[12]The 2016 review findings 'review-of-freight-delivery-metric-2016-10-24.pdf can be accessed at: https://www.orr.gov.uk/media/16767.

FDM-attributable delay caused by each region. Using an example of a train that was delayed by 40 minutes, with 10 minutes of the delay caused by Eastern region and 30 minutes of the delay caused by North West & Central region - for this train, Eastern region would be assigned 0.25 FDM delayed trains and North West & Central region would be assigned 0.75 FDM delayed trains.

Any delay minutes caused by incidents on non-NR infrastructure (except where caused by commercial FOCs) are treated as FDM-attributable. When attributing FDM delayed trains between regions for FDM-R, any failures as a result of such incidents are effectively excluded. Using a similar example to above, in this case where a train was delayed by 40 minutes, with 10 minutes of delay caused by Wales & Western region and 30 minutes caused by the TfW CVL network - for this train, Wales & Western region would be assigned 0.25 FDM delayed trains. The remaining 0.75 FDM delayed train is assigned to TfW CVL so not relevant to FDM-R.

The Trains Ran component of the metric, which represents all FDM-R qualifying services (excluding cancellations), is derived from the total number of FDM-qualifying trains attributed in proportion to the train count within each region. The Network Rail FDM-R definition document has been appended to this document in full and is located in Appendix B. It includes an example of the allocation process using values from Period 1 2020/21 (Appendix 3A) but in simple terms, if a train operates across two regions, then for the Trains Ran, this is treated as 0.5 trains in each region. This ensures that the sum of Trains Ran applied to the five regional FDM-R calculations matches the national total of Trains Ran.

We have reviewed the definition document and found it to be a clear and accurate description of the FDM-R metric and includes a set of useful examples to explain the attribution between regions.

4.3 Governance and Approach for Converting TRUST/PSS Data into FDM-R

4.3.1 Outline of Metric Calculation Approach

The FDM-R calculation process is undertaken using data calculated and extracted from NR's PSS database via a set of Business Objects queries and fed through a system of Excel workbooks. The flow of data through these files is via a combination of automated processes (via macros) and manual transfer.

Data relating to cancellations and service variations is not taken from PSS since it is not fully recorded in TRUST. Instead this comes via Schedule 8 performance regime claims submitted by FOCs each week. These claims are validated by the Freight Performance Regime Specialist (FPRS) team, through investigation within TRUST. Through these investigations, they assign the incident location that is deemed to have caused the cancellation, which is used to assign the cancellation to a region. Due to the level of scrutiny applied to validate Schedule 8 claims, this is considered the most accurate and reliable source of cancellation and service variation information for freight services.

4.3.1.1 Process Overview

The diagram below (created by the Reporter team) provides an overview of the process of converting the data from the raw BOPSS extracts and the Schedule 8 claim data into the FDM-R metric. Each of the numbered steps of the process is summarised in the following text. Note, all calculations are done at a NR route and region level; while the text below refers to regions for simplicity, this should be read as both route and region.

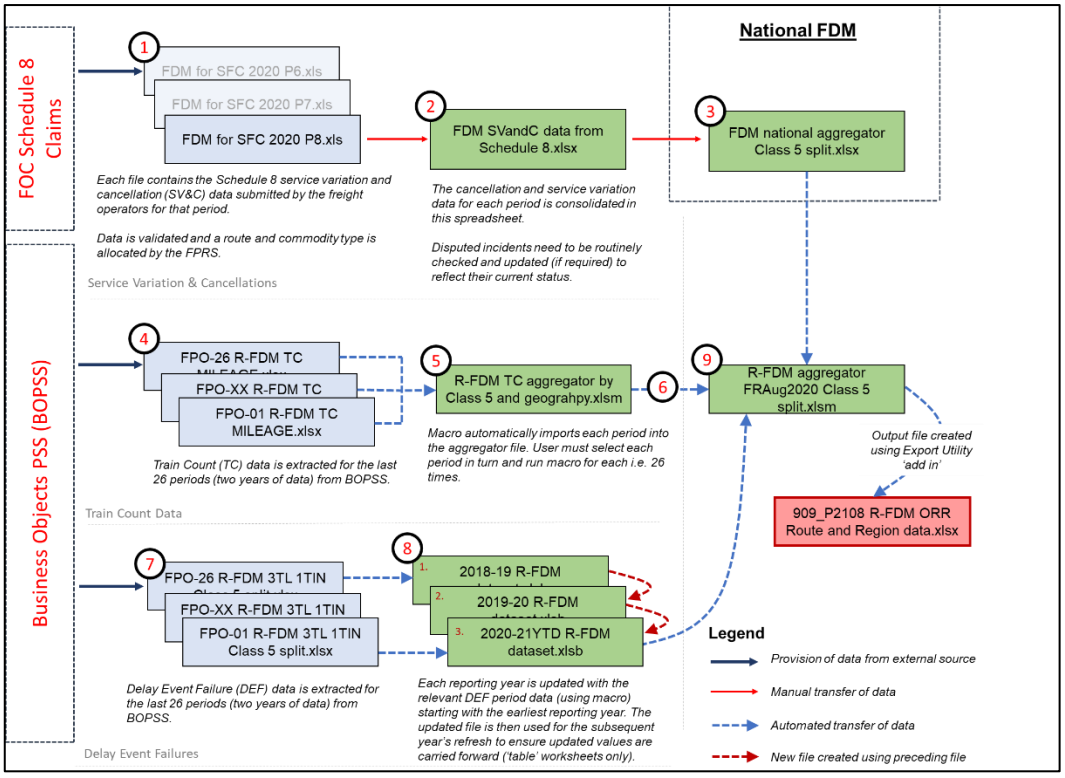


Figure 4: FDM - R calculation process diagram (each **step** in described below)

Service Variation & Cancellation (SV&C)

Objective: To validate and count the number of cancellations/service variations in each period and assign to the responsible NR region based on incident location.

Step 1. The '*FDM for SFC yyyy Pp.xlsx*' file contains the validated Schedule 8 claim data for each Period from the FPRS team. The file contains details of both cancellations and service variations, with responsible region and train commodity type as assigned by the FPRS team.

Step 2. The SV&C data for each period is consolidated in the '*FDM SV&C data from Schedule 8.xlsx'* file (manually copied and pasted), which contains all such historical data since April 2014. This file also contains a list of disputed incidents which need to be routinely checked with the FPRS team and updated as their status changes.

Step 3. The consolidated SV&C datasets are used for the calculation of both the National and regional FDM. The data is first manually copied and pasted into the file used to calculate the National FDM ('*FDM national aggregator Class 5*

split.xlsx'). This is subsequently fed into the file used to calculate FDM-R (*'R-FDM aggregator FRAug2020 Class 5 split'*) (step 9) via a macro. Note that no modification of the dataset is undertaken during this step.

Train Count Data

Objective: To correctly identify and count the number of FDM Qualifying services that ran in each period. This is achieved by allocating the National Train Count to each region in proportion to the number of trains that operate in each geographical area.

Step 4. A BOPSS query is automatically set to run every period to export train count data for the last 26 periods into separate *'FPO [x] R-FDM 3TL 1TIN.xlsx'* files. The query exports a list of FDM Qualifying services that ran and calculates the apportionment by region.

Step 5. A macro is used to transfer and consolidate the National and regional train numbers by period into a single file '*R-FDM TC aggregator by Class 5 and geograhpy.xlsm'*[sic]. Prior to running the macro, the cells in this file relating to the 26 periods to be refreshed must be manually cleared. The macro then needs to run separately for each of the 26 periods.

Step 6. The consolidated region datasets are then automatically transferred to the final FDM-R aggregator file via a copy and paste macro. No modification or manipulation of the dataset is undertaken during this step.

Delay Event Failures

Objective: To correctly identify and count the FDM Qualifying services that incur a Delay Event Failure (based on the definition described in Section 4.2) and attribute each service proportionally to NR regions based on the incidents that caused the qualifying delay.

Step 7. A BOPSS query is automatically set to run every period to export the Delay Event Failure data for the last 26 periods (it was agreed with ORR that the data refresh should cover the last 2 years to capture any changes to delay attribution). Each BOPSS export contains a list of the FDM Qualifying services, whether the train failed the punctuality threshold, and FDM qualifying delay by region. The query also automatically exports the total FDM failures by region, which feeds the FDM-R calculation.

Step 8. The *'[Year] YTD R-FDM dataset.xlsb*' files consolidate the periodic data from Step 7. Separate files are maintained for each year due to their size. A macro is used to automate the import of the data from the BOPSS queries into each file, and the user must ensure the correct periods to import into each file are selected.

Aggregation of Data and Calculation of FDM-R

Step 9. The final aggregator file (*R-FDM aggregator FRAug2020 Class 5 split.xlsm*) combines the above three strands of data to calculate the final FDM-R metric. Data is transferred from each of these three files via a macro. An export 'Add in' function is used to export the regional FDM-R figures into the file which is sent to ORR.

4.3.2 Governance of Metric Calculation Approach

4.3.2.1 Process Documentation

The FDM-R calculation process is owned by the Freight Directorate. Two documents have been produced to describe the process:

- "*Definition Region and Route Freight Delivery Metric 20201120.docx*"; providing the core definition of the metric as outlined above
- "*R-FDM consolidation file for YTD and MAA values.docx";* a step-by step guide to updating the FDM-R calculation each period.

The step-by-step guide was found to be detailed and comprehensive and is evidence that the process has been formally documented in line with Recommendation "L3 AR 001/01" from the 2016 audit. This document is version controlled and dated (November 2020).

A copy of the BOPSS queries used to extract data from PSS each period were provided to review against the definitions. No concerns were identified with these queries. It was advised that these queries are not altered unless an underlying change to the network requires it.

Our review of the calculation process confirms it aligns with the documentation, and the step-by-step guide ensures a clear path to follow to produce consistent and accurate figures that comply with the formal FDM-R metric. The level of automation (including both the extract of BOPSS data and the transfer of data between spreadsheets) provides reassurance that the process is repeatable and would produce consistent period end figures.

Our review did highlight two minor potential inconsistencies with the approach and definition, which we have discussed with the Freight Directorate and are outlined below. It was agreed that there is an element of ambiguity with each of these, which needs to be reviewed.

Firstly, if a train operates within a route or region on more than one occasion, the train count BOPSS query (Step 4 above) will assume each occasion is a separate train. For example, a service that starts in region A, travels through region B, and then returns to region A would be treated as 2 trains in region A and 1 train in region B. How such trains are counted is not explicitly defined in the documentation, but the definition could be interpreted to mean that a train should only be counted as operating once in any region, regardless of how many times it re-enters that region. The appropriate treatment of such trains should be reviewed and either clarified in the documentation or updated in the BOPSS query.

The second issue involves treatment of Class 0 services (light loco movements). The definition document states that these should be excluded from the FDM-R metric. Our review found that several cancelled Class 0 services have been included in the SV&C data due to their status as a commercial service and inclusion in the Schedule 8 regime. In contrast, all Class 0 services that ran are excluded by the BOPSS query, regardless of status. This reflects an inconsistency both with the documented definition and within the calculation. The appropriate

treatment of these services should be reviewed and either clarified in the documentation or updated in the BOPSS query or SV&C data.

We have identified a potential enhancement to the process documentation which we have discussed with the Freight Performance Analyst. The existing documentation covers the metric definition and a step-by-step guide to updating the metric. However, there is no intermediary 'process guide' to provide an overview of the process and describe the queries and spreadsheet suite, which would assist new users. Without the assistance of the Freight Performance Analyst providing a thorough walk-through of the files and processes, it would have been more challenging to audit the process.

We therefore recommend producing a process guide, which could supplement the existing step-by-step guide. To assist new users, this could include a process map (such as the version displayed in Section 4.3 of this report), a description of the BOPSS queries and an overview of what each of the spreadsheets is designed to do. We observe that the spreadsheet suite has a number of additional calculations and reports requested by regions, and not directly relevant to our audit. These do add to the complexity for new users to navigate, which such a guide may support.

4.3.2.2 Process Resilience

The team within the Freight Directorate includes two analysts who are trained in the calculation process and two Freight Performance Regime Specialists to validate Schedule 8 claims. This represents notable progress against Recommendation "L3 AR 001/02" from the 2016 audit and provides cover in case of staff absence. However, we do observe that training further resources would provide additional resilience to the team.

Each period, a 'back-up' copy of the aggregator spreadsheet is maintained while the periodic updates are being undertaken. Once this update has been finalised, the back-up file is deleted so there is only one 'master' version. From discussions with NR it is understood that this process is undertaken but is not currently included in the update procedure documentation. The documentation should be amended to reflect this.

We also suggest it would be prudent to retain copies of files from at least the previous period as a back-up, to guard against any risk of error with the 'master' copy, such as file corruption.

The structure of the spreadsheets is stable unless a change is applied to the NR organisation, such as the PPF re-organisation in 2020. The process of updating the spreadsheet suite to reflect the new NR routes was a material task, and the tasks that need to be undertaken to make such a change do not appear to be documented.

It is acknowledged that pre-empting and documenting future changes to the organisation and assessing its potential impact on the FDM-R process/spreadsheets would be a difficult and time-consuming task. It is therefore proposed that any significant NR organisation change that affects the source datasets e.g. geographical changes, should trigger a risk review of the process.

This should focus on understanding the likely changes to the source data and any associated workbooks.

4.3.3 Data Assurance of Metric Calculation Approach

We have quantified how much effect the following two identified potential inconsistencies might have on the FDM-R to confirm that they are not material issues.

Quantified impact of train count ambiguity

The table below provides a quantification of the impact of this inconsistency on the 2020/21 period 8 metric results.

Table 8: Impact on FDM-R of Train Count Attribution Inconsistency

Description	Region	% Difference
Discrepancy in the Train Count due to double counting of services.	Eastern (ZEST)	-0.008% (<0.1%)
	North West and Central (ZNWC)	-0.010% (<0.1%)

As this issue is related to the standard query used to extract the train count data from BOPSS it is likely to affect the train count data every period. However, the quantification of the impact shows that the materiality is insignificant.

Quantified impact of Class 0 issue

We have calculated the impact of removing all Class 0 services that have been included in the cancellation count over the last 13 periods on the FDM-R MAA for each region. The results of this are presented in the table below.

Table 9: Impact of Removing Cancelled Class 0 Services from FDM-R MAA

Description	Region	% Difference
Metric includes cancelled services that are defined as Class 0 (light loco which contradicts the formal definition).	Eastern (ZEST)	0.002% (<0.1%)
	North West and Central (ZNWC)	0.062% (<0.1%)
	Southern (ZSTH)	0.030% (<0.1%)

The quantification of the impact on the sample period provides confidence that this issue is unlikely to have had a material impact on the FDM-R metric.

4.4 Reliability, Quality, Consistency, Completeness and Accuracy of Reported Data for FDM-R

In this section, we focus on the quality, completeness and accuracy of the underlying delay used to calculate the FDM-R metric. This includes the following datasets and the associated extraction process via BOPSS queries:

- Punctuality data from PSS;
- Train count data from PSS;

- Delay data from PSS; and
- Cancellations and Service Variation data from Schedule 8 claims.

4.4.1 Processes and Governance

4.4.1.1 Punctuality Data

The arrival time of each commercial freight train at scheduled destination is captured in TRUST. The majority of train timings are fed directly from the signalling system, but a number of freight trains terminate in yards or depots which are off the NR-controlled network. As a result, their arrival times must be reported through manual entry into TRUST. Where a time is not manually input, it is interpolated by TRUST based on the final time a train auto-reported on the NR infrastructure.

The Performance Data Accuracy Code (PDAC) was updated in September 2020 to include a provision that manually-input timings may not be retrospectively amended to change train lateness or delay, including off NR infrastructure, unless in line with an agreed process or evidence can be produced to justify the change. This was introduced to provide tighter controls over this information.

The NR Process & Controls team keep records of any such retrospective changes. An example was provided for period 9 (2020/21) showing for 3% of freight trains that terminated at a manually reported location terminus, the time was retrospectively updated in TRUST.

4.4.1.2 Train Count Data

The FDM-R metric is based on services operated by commercial Freight Operating Companies (FOCs) as 'commercial' freight services. Therefore this excludes any non-commercial or infrastructure trains operated by a commercial FOC. The metric does, however, include loco-hauled Empty Coaching Stock (ECS) moving from works to a passenger operator depot and operated by a commercial freight company under a commercial contract with a TOC or ROSCO. It does exclude normal ECS movements by TOCs between depots and stations.

The information on train movements is captured in TRUST and fed into PSS. A train is defined as an FDM-Qualifying train if it meets each of the following three rules, as outlined in the BOPSS query.

- It must be a commercial freight service, based on the 8-digit Service Code
- It must not operate as a Class 0
- It must not have been cancelled or part cancelled

The definitions document outlines a mandated check of this process every six months by extracting all non-passenger services from PSS for a period and reviewing all exclusions to ensure consistency with the defined process. In discussion with the Freight Performance Analyst, it was confirmed that this check

is undertaken more regularly than every 6 months at present due to an increased number of ECS moves falling within the metric. However, details of these checks are not regularly documented and retained, so were unable to be audited.

4.4.1.3 Delay Data

Delay data is captured in TRUST, and the attribution of delay minutes to regions is on the same basis of NR Manager Code as per CRM-P.

One core difference to the CRM-P approach is that delay data is extracted from the "Attributions Universe" in PSS, rather than the ADS data. For FDM-R, calculations are done by individual train and this granularity of data is not stored within the ADS Universe. This means incidents in dispute between the Operator and NR will be initially attributed to the Operator, and so runs the risk of initially under-reporting the number of FDM failures.

The definition document mandates that the impact of incidents in dispute is monitored to ensure that they are not expected to have a material impact on the outputs (see section 4.5.1 for further information on the checks that are specified in the documentation).

4.4.1.4 Service Variation and Cancellation Data

As previously noted, SV&C data is derived from Schedule 8 claims submitted by FOCs. This data is all manually input into an Excel file following investigations within TRUST, with no observed consistency checks. The file is accepted as accurate when received from the FPRS and fed into the calculation.

It was observed that researching and validating each claim is time-consuming, and made harder by the limited information currently required to be supplied by FOCs for each claim (effectively the date and train headcode). This process could be made more efficient if NR can mandate FOCs to provide additional information with their claims, such as Train Service Code, Origin and Destination.

In our previous audit, we noted the national Management of Freight Services during Disruption (MFSDD) initiative to standardise management response to freight services during significant disruption. One of the observations was that a train which had its planned journey changed as the result of implementation of MFSDD would be treated as an FDM failure. If a supplementary service was operated to complete the journey, e.g. under VSTP the following day, and arrived late enough to trigger an FDM failure, then this train is effectively double-counted as a failure. While it was confirmed that the same issue can still arise, it is difficult and time-consuming to identify in the data, and in practice there remain very few of these so will not have a material impact on the metric. Just 0.18% of trains in the latest year were Service Variations and only a small proportion of these are likely to have been affected by this issue.

4.4.2 Data Assurance

This section of the report describes the analysis that was undertaken to support the formulation of a view on accuracy of the reported FDM-R figures.

4.4.2.1 Overview

The basis of the data assurance review included a series of validation checks on the 'Aggregator Spreadsheets' and an assessment of a sample of raw data from PSS. This included:

- A review of raw data from PSS to ensure the correct trains have been included in the metric.
- A series of computational checks to ensure the calculations undertaken by BOPSS and within the Excel workbooks are accurate.
- A review of the SV&C data to ensure manually entered data is accurate and verify the correct trains have been 'included' and 'excluded'.
- A review of the data at each stage of the process to ensure the data transfer and data flow process is robust and complete.
- A review of the Aggregator spreadsheet outputs against the published metrics on the ORR website.

To support the data assurance review, the Freight Directorate supplied the suite of Excel workbooks used to calculate the metric in period 8 of 2020/21, along with raw input data for the same period.

4.4.2.2 Trains included in Metric

To assist with the review of the FDM Qualifying Trains list, NR provided screenshots of the BOPSS queries used to extract the list as well as an additional dataset that was requested by the Reporter team. The additional dataset ('*P2108 FDM A2F ORR Reporter All Trains list.xlsx'*) included a list of all non-passenger services in PSS in the period which allowed the team to apply the rules (as per the metric definition) to ensure the same result was achieved.

In period 8 2020/21 there were a total of 43,303 non-passenger trains within PSS, of which 12,186 were designated as FDM-qualifying trains in the BOPSS query (FDM Trains Ran). The 31,117 excluded trains are summarised below. In each row, the train numbers exclude those which have already appeared above for another reason (e.g. the number of light loco moves' excludes Class 0 trains with a non-commercial or unknown operator type) to avoid double-counting.

Table 10: Non-passenger trains in PSS not measured under FDM-R in period 8 (2020/21)

Reason	Number of Trains Excluded	Notes
Operator Type: 'non-commercial'	11,970	Exclude operator types that are non-commercial (i.e. Ballast and Yellow Plant).
Operator: 'unknown'	0	Invalid service code entered when schedule created, so unknown operator in PSS (marked "XX"). These are not included within FDM and NR confirmed there are usually very few.
Train Class Code: '0'	6,530	Class 0 (Light loco movements) are not included in the FDM-R calculation.

Actual Origin or Destination Time: 'Blank'	12,198	These are either fully cancelled or partially cancelled so not relevant. Relevant cancelled trains will be captured in the SV&C data.
Actual Train Miles = Planned Train Miles: 'False'	146	Trains that have not run planned distance and all terminated at a different location to the planned schedule, so part cancelled (see above).
TSC PfPI Flag: 'N'	260	Train Service Code (TSC) Process for Performance Improvement (PfPI) Flag indicates whether a service is operating under a non-commercial service code. There are a number of commercial operators which run a small number of trains under a non-commercial service code which need to be excluded.
Planned Origin or Destination Time: 'Blank'	11	Train does not have a valid schedule to measure against, so excluded from FDM-R.
Headcode: '1Z99'	2	Breakdown/emergency trains with headcode '1Z99' are excluded from the metric.
TOTAL	**31,117**	

By applying these filters to the full non-passenger train list, we were able to recreate the FDM Trains Ran dataset exactly. This confirms that the data extracted by the BOPSS query is accurate and provides assurance that the query filters have been applied in a manner that is consistent with metric definition.

One point to highlight is that the data extract showed no non-passenger services being allocated an incorrect service code ("Operator: 'unknown'" in above table). For comparison, during the previous FDM review it was found that 100 services in the sample period were excluded for this reason. This reflects improvements in train planning processes / systems which now only allow valid service codes to be entered when schedules are created.

4.4.2.3 Trains Ran which fail FDM (due to Delay Event Failures)

The Reporter team subsequently checked how many of the 12,186 FDM Qualifying services which ran in period 8 were reported to fail the metric due to a Delay Event Failure (DEF). The calculations that flag this are undertaken within BOPSS, however the appended delay minutes and incident data in BOPSS extract '*FPO-01 R-FDM 3TL 1TIN Class 5 split'* enabled the reporter team to undertake an independent check to confirm this process was accurate. As a reminder a DEF is defined as:

- A train that did not reach its planned destination within 15 minutes of schedule time; AND
- Incurred 15 or more minutes of eligible attributed delay (caused by Network Rail or another operator that is not a commercial FOC).

In the sample period, 575 services out of 12,186 were reported to fail the metric due to a DEF. Of these services, our review identified one train whose 'Total Lateness at Destination' was one minute, so had therefore been flagged as a DEF in error. The table below provides a quantification of the impact of this error on the metric results for period 8.

Table 11: Impact on FDM-R of Train incorrectly identified as FDM failure

Description	Region	% Difference
SQL query in BOPSS flags one service as an FDM failure that arrives only 1 minute late at its destination.	Eastern (ZEST)	0.022% (<0.1%)
	Wales & Western (ZWST)	0.002% (<0.1%)

A review of the last 13 periods (P2009-P2108) show 12 separate instances of this issue which indicates that this is a recurring error but has not had a material impact on the metric results. This has been raised with the Freight Performance Analyst for further investigation.

While undertaking this check it was identified that several fields in the BOPSS extracts included blank cells, which should have been populated. This suggests an issue with the SQL query used to extract the data. None of these blank cells relate to data used to calculate the FDM-R, but it does reduce the level of transparency and completeness when checking the extracts. This issue was also observed in the checks undertaken in the following section.

No further issues were identified in relation to the accuracy of trains listed as Delay Event Failures.

4.4.2.4 Computational Checks of BOPSS and Aggregator Spreadsheets

The reporter team has undertaken detailed checks of the calculations undertaken within BOPSS and has also reviewed the Excel formulae used within the spreadsheets supplied by NR.

Areas of particular focus included:

- Replicating (in Excel) the calculations that are undertaken in BOPSS to allocate FDM Qualified trains to regions.
- Replicating (in Excel) the Period totals that are calculated in BOPSS e.g. Total FDM failures, Train Count by region etc.
- Spot checks of the formulae implemented throughout the FDM-R calculation process.

The computational checks identified no issues or inaccuracies that could impact the FDM-R metric results.

4.4.2.5 Accuracy review of Service Variation and Cancellation (SV&C) data

The SV&C data is sourced from the FOC Schedule 8 claims which is validated by the Freight Performance Regime Specialists (FPRS) using information from TRUST. The FPRS append several columns to the data submitted by the FOCs to assist with the validation process which is subsequently used during the FDM-R process to determine the number of cancellations and the region these should be attributed to. This data is entered manually by the FPRS.

Whilst a check of the accuracy of the allocation of cancelled services to incidents and responsible regions based on TRUST is outside of scope of this review, several checks were possible to ensure that:

- The trains included in the list of cancellations and service variations were consistent with the Metric definition.
- The data manually entered by FPRS was consistent across the different fields.

The submission for Period 8 2020/21 by the FPRS team '*FDM for SFC 2020 P8.xls*' as well as the consolidated historic data going back to Period 1 2014/15 (in the aggregator files) was used as the basis of the check.

The review found no material issues, although two minor inconsistencies were identified.

Firstly, there was an inconsistency observed for a small number of trains between the 'Commodity Type' and 'Cancellation flag'. Only cancellations of loaded freight trains can be claimed under Schedule 8. Certain trains (Intermodal / Royal Mail / MOD services) are expected to be loaded in each direction while the remainder are assumed to travel loaded in one direction and empty the other. The 'Cancellation Flag' is used to determine how many cancelled trains each Schedule 8 claim relates to, so set to 1 for the train types listed above, and 2 for all others.

Our review identified a small number of trains with commodity type set to "Intermodal" but allocated as 2 trains in the "Cancellation Flag" field. The Freight Directorate confirmed that the Cancellation Flag field in the dataset for these services is most likely correct and the commodity type had been assigned in error. However looking at historical data, the team acknowledged at least two trains that had been incorrectly assigned a '2' cancellation flag.

We only identified 10 intermodal services in the period since 2014/15 that had been allocated a '2' cancellation flag and so the impact on FDM-R of this discrepancy will be significantly lower than 0.1%.

The second issue identified related to the transfer of CVL infrastructure ownership from NR to TfW in February 2020. We observed that historical cancellations attributed to locations on the CVL prior to February 2020 remain assigned to NR Wales & Western rather than having been updated. There were very few of these trains, so will not have a material impact on the FDM-R figures. The impact of these trains is now receding from the MAA and will shortly drop out altogether.

Despite the low risks, these examples highlight the issues associated with manually inputted data and the inaccuracies that can arise as a result as well as a reduction in traceability and transparency. Therefore, it is recommended that the SV&C process is strengthened with an aspiration of further automation and less reliance on user inputted data/manual interventions.

Requesting FOCs to provide additional information in support of Schedule 8 claims would support this and enhance the process as outlined in Section 4.4.1.4. For example the provision of the Train Service Code with the FOC would make it

possible to automatically map each service to a commodity type based on the Track Access Billing System (TABS).

4.4.2.6 Review of data consistency/integrity

In addition to the computational checks, the Reporter team also reviewed the key datasets at each stage of the process to validate the accuracy of the data transfer process and ensure data had not been accidentally/unintentionally lost or modified. A check which compared the final aggregator FDM-R metric results against those published on the ORR data portal was also undertaken.

The data consistency checks identified no issues or inaccuracies that could impact the FDM-R metric results.

4.5 Processes to produce consistent, quality assured period-end FDM-R figures to customers

4.5.1 Governance

The FDM-R queries are fixed and automatically run on the Wednesday after period end. Therefore, the data will be extracted in the same form, via the same calculation every period. There is a clearly documented update process to follow each period. This provides clear instructions on who to send the final metrics to, via what means, and by when. This information is disseminated to key stakeholders on the same day, so ensuring consistent metrics reported across the industry.

Each period, historical data is refreshed to ensure any resolved disputes are reflected. This does mean that historical periodic and MAA data may change each period. Since incidents in dispute are assigned to the Operator until resolved, this indicates that the metric will increase as such changes are applied.

The documentation specifies recommended checks:

- To verify that all FDM-Qualifying trains are included in the calculation every six periods;
- To check the national train count against the sum of regional FDM-R trains; and
- To review the potential impact of incidents in dispute.

The spreadsheets also contain a number of ad-hoc verification checks on the data. Each spreadsheet also logs the date and time that each period of data has been updated when done via a macro. This provides assurance that the macro has been run for all relevant periods at the same time. A table summarising the key information regarding the latest import (typically on the IMPORT DATA tab) is used as a method for sense checking the import data (does each sheet contain the correct the number of Periods for example).

We were advised that the recommended checks are not documented as a matter of course. We would recommend that details of these reviews are recorded where possible.

4.5.2 Data Assurance

As part of the audit of the flow of data through the spreadsheets, we have undertaken a number of additional verification checks. This has highlighted some areas where such verification checks would be beneficial to incorporate into the process to flag up any potential issues. These include:

- In the "*FPO-01 R-FDM 3TL 1TIN Class 5 split*" file, aggregated and total data is populated by the PSS query (for example, delay by train for Eastern region, plus a total delay value). A calculation check to confirm that the sum of the values in each column matches the total would pick up the 'blank' cells identified in Section 4.4.2.3.
- A simple replication of the FDM failure calculation for each train based on lateness and qualifying delay would verify the correct failure trains are included.
- Additional controls in the SV&C file would prevent the discrepancies identified, e.g. a warning flag if the inter-modal flag is incompatible with the commodity type

4.6 Further Observations from FDM-R Review

This section has been included to make two observations on the defined approach for calculating the FDM-R metric. We recognise that this is beyond the scope of this audit, but they reflect observations that were made during the review that we felt would be helpful to capture, and may feed into considerations for the freight performance metric for Control Period 7 (CP7). These observations are not reflected in our conclusions or grading.

Attribution of FDM Qualifying Trains between Regions

Section 4.2 outlines how the FDM Qualifying Trains are attributed between regions in proportion to the number of regions they operate in. We observed that that this may not fully reflect the interactions that freight services have with each region. As an example, a train from Acton to Dagenham Dock runs for 59km, 0.8km on Wales & Western region and 58km on Anglia region. Under the current approach, this train is attributed as 0.5 trains on each region, despite travelling 98% of its journey on Anglia region.

An alternative approach would be to attribute the FDM-Qualifying trains in proportion to distance travelled on each region. We understand that the current approach taken has been adopted due to historical limitations of the distance data held within PSS with respect to freight services. We recognise that such limitations may still exist, although the recent improvements and updates to the data held within the Mileage Universe within PSS are likely to support this suggestion.

An attributed trains count for FDM Qualifying trains is already calculated based on mileage data each period by the Freight Directorate for reference purposes. As an example, the table below shows the impact of using mileage for this calculation for each region in terms of attributed FDM-Qualifying Trains. This shows some material differences.

Table 12: FDM-R Qualifying Trains by Region based on Trains Operated vs Distance

FDM Qualifying Trains	Attributed by Trains Operated	Attributed by Distance Travelled	Difference
Eastern	4,594	4,894	+300
NW & Central	3,649	4,062	+413
Scotland	540	662	+122
Southern	1,154	740	-414
Wales & Western	2,249	1,828	-421
National	**12,186**	**12,186**	**0**

Impact of ECS on FDM-R

We observed that, with the wide-spread programme of new train introduction across the network, the number of such FDM-Qualifying ECS moves has increased recently, as shown in the table below. While this is consistent with the definition, it is an area of consideration for ORR / NR to ensure that the metric continues to measure as intended. We observe that the FDM-R spreadsheets contain shadow calculations with ECS trains excluded, following request from at least one region so they could understand the impact.

Table 13: Proportion of FDM-R Qualifying Trains that are ECS Movements

% FDM-Qualifying Trains Run as ECS	MAA at 2017/18 Period 13	MAA at 2020/21 Period 08	Change
Eastern	3.7%	7.2%	+3.5%
NW & Central	3.2%	5.2%	+2.0%
Scotland	12.2%	11.8%	-0.4%
Southern	2.7%	3.8%	+1.1%
Wales & Western	4.1%	6.1%	+2.0%

4.7 FDM-R Conclusions and Confidence Ratings

Our audit has identified no material concerns with the system reliability or accuracy of the reported CRM-P metric. The recommendations from the previous audit have been acted upon, with strengthened resources and documentation. Our report has highlighted where further enhancement would support the process further.

The FDM-R is awarded an overall "B" grade for reliability

The table below summarises a grading for each of the criteria considered for this award, along with a list of the key evidence provided to support the grade.

Table 14: Reliability Grading for FDM-R

Element of Process	Criteria to be met	Grade	Evidence Provided

Objectives	Clear and unambiguous description of the purpose and objectives of producing the metric	A	Definition document[13] User Guide for updating metric[14]
Requirements	Clear and unambiguous description of the standards required for the data and its collation	B	PMM; DAPR Slight ambiguities observed in process (Section 4.3.2)
RACI	Clear identification of those Responsible for, Accountable for, Consulted about and Informed about the metric	A	Discussion with Freight Directorate (FD) / Eastern region
Source(s)	Description of who or what (system) provides the data	A	User Guide for updating metric Discussion with FD
Means and Frequency of Data Provision	Description of how the data is provided, how often, and when	A	Definition document User Guide for updating metric
Data Format and Expected Values	Definition and description of the format in which the data are to be supplied; expected range (if any) of values	A	User Guide for updating metric Targets displayed in Aggregator spreadsheet
Data Quality	Definition and description of the required data quality and accuracy	A	PMM; DAPR Definition document SV&C data requires manual validation and update of data supplied by FOCs in line with Schedule 8 Agreement
Data Processing	Documentation of processes, sufficiently clear for new users	B	Definition document User Guide for updating metric Discussion with FD Additional documentation of process would support new users (see Section 4.3.2)
Staff Training	Sufficient availability of trained staff to maintain data and produce the metric	B	Discussion with FD Additional trained resource would provide greater resilience (see Section 4.3.2)
Data Checking and Verification	Description of mandated and recommended verification checks to ensure accuracy of reported metric	B	Definition document User Guide for updating metric Discussion with FD No written evidence captured when check undertaken; Additional verification checks within spreadsheets would support resilience
Managing Change	Description of process for updating underlying data to reflect major changes, e.g. route/region restructuring	B	PMM in relation to TRUST/PSS data Process for updating spreadsheet suite to reflect such changes is not documented
Internal Review and Audit Procedures	Description of internal review and audit requirements, processes and frequencies; evidence that these are being met	A	Definition document User Guide for updating metric

[13] "Definition Region and Route Freight Delivery Metric 20201120.doc"
[14] "R-FDM consolidation file for YTD and MAA values.doc"

The FDM-R is awarded an overall "1" grade for accuracy

This is because the assessed and potential data inaccuracy falls between a range of 0.1% and 1%.

We have observed a number of minor issues which could compromise the reported figures, although none of these were identified as material on their own.

- Cancellation data manually calculated from sources external to TRUST; some inconsistencies observed
- MSFDD trains may be double counted
- Some trains erroneously identified as failures
- Class 0 trains included in cancellations in contravention to the definition
- Some ambiguity over the process for counting trains passing through the same region twice
- Arrival times at termini off NR infrastructure are manually entered

With the exception of the final point, we have quantified the potential impact of these issues on the reported metric.

5 Study Recommendations

Based on this audit, we have identified the following recommendations.

Table 15: Study Recommendations

Number	Metric	Recommendation to Network Rail	Benefits	Evidence of Implementation	Location in Text
SOW17190-1	CRM-P	Produce a process guide for CRM-P	A documented record of the queries used to generate the report, supporting training of new staff	Production of process documentation	Section 3.3.2
SOW17190-2	CRM-P	Document an agreed date each period for ORR and NR to report the CRM-P metric	Ensure that NR and ORR are reporting consistent CRM-P figures each period	Documentation	Section 3.5.1
SOW17190-3	CRM-P	Update definition of delay minutes used for CRM-P metric in "Definitions of Railway Performance Metrics" document to clarify those off-NR network delays that are included	The current definition could be mis-interpreted to mean that all delay minutes suffered on non-NR networks as a result of a NR incident are included.	Definition updated in "Definitions of Railway Performance Metrics" document	Section 3.2
SOW17190-4	CRM-P	Review and agree which of the trains that operate partly on the NR network and partly on non-NR networks should be included in the CRM-P metric	Confirm intention - currently trains without a contractual framework for train performance with NR are excluded from the metric, even if they operate partly on the NR network. This is the case for services operating on the East London Line and could be the case for CrossRail services when they start to operate.	Treatment of delays to these trains clarified in "Definitions of Railway Performance Metrics" document	Section 3.2

Number	Metric	Recommendation to Network Rail	Benefits	Evidence of Implementation	Location in Text
SOW17190-5	FDM-R	Produce a user-guide for FDM-R supplementing the existing 'update guide', covering a process map and description of each query and file	This will support new users to help understand the files used to calculate the metric, and support future enhancements	Production of an enhanced user guide	Section 4.3.2
SOW17190-6	FDM-R	Review FDM-R definition to confirm treatment of: - Trains which enter a route/region more than once - Class 0 trains	This will remove any ambiguity from the definition, and ensure consistency in approach	Clarification of treatment within process documentation	Section 4.3.2
SOW17190-7	FDM-R	Document completion of mandated FDM-R checks as defined in documentation and any issues identified	This will provide confirmation that mandated checks are being undertaken, and any issues raised are documented	Written evidence of checks undertaken	Section 4.5.1
SOW17190-8	FDM-R	Strengthen Service Variation & Cancellations (SV&C) data collation and processing process for FDM-R through: - Requesting further information to be supplied by FOCs in support of Schedule 8 claims; and - Reviewing opportunities for further automation of the process to remove reliance on manual updates	Receiving additional data from FOCs in support of Schedule 8 claims (e.g. Train Service Code) will make it easier to identify the relevant trains within PSS, so improve the validation process Automating more of the SV&C data files, including introducing checks, will remove the risk of error through manual data input	Information supplied by FOCs An updated SV&C process	Section 4.4.1.4 / Section 4.4.2.5

Appendix A

Independent Reporter
Framework Statement of Works
#17190 – Review of CRM-P and FDM-R

Independent Reporter Framework

Statement of Works

1.0 COMMISSION INFORMATION	
Project Name:	Review of CRM-P and FDM-R
Bravo Sourcing Request Number:	#17190
Network Rail Contact:	Matthew Blackwell
Network Rail Department:	Planning & Regulation
SoW Number:	0006
Network Rail PO Number:	[insert NR PO# when available]
Commission Value:	[insert the SoW value after this has been agreed with the supplier]
Supplier Name:	[insert the name of the selected supplier after appointment]
Main Supplier Contact:	[name and email address of the main supplier contact]

This Statement of Work (SoW) is the contractual vehicle for defining, authorising and commissioning a piece of work to be undertaken under the Independent Reporter Framework. The SOW has six sections:

1 *Commission Information*
2 *Commission Overview*
3 *Scope of Services and Deliverables*
4 *Knowledge Transfer*
5 *Resource & Commercial Details*
6 *Invoicing*

This SoW is entered into under and in accordance with the terms of the Independent Reporter Framework dated 1 February 2020 between Network Rail, the Office of Rail and Road, and the Supplier and includes and incorporates any special Terms and Conditions and any other amendments captured in this SoW.

Any dispute surrounding this SoW will be resolved in accordance with the Terms and Conditions outlined in the Framework Agreement.

Ownership and use of any Intellectual Property Rights shall be in accordance with the Framework Agreement Terms and Conditions.

Change control procedures are to be applied as set out in the Terms and Conditions of the Framework Agreement.

2.0 COMMISSION OVERVIEW	
2.1 Background	Consistent Region Measure – (Passenger) Performance (CRM-P) is a key metric in ORR assessing Network Rail's delivery of passenger rail performance to customers during Control Period 6 (CP6), forming part of the CP6 Scorecards. CRM-P measures the minutes of Network Rail attributed delay to all passenger trains from incidents occurring within the Region normalised by the train kilometres travelled by passenger trains within that Region. The Freight Delivery Metric (FDM) was introduced as the regulatory performance output for freight in Control Period 5 (CP5). It was formally assessed in 2016 and scored a good level of confidence grading at a national level (A2). To reflect devolution, Network Rail developed the Freight Delivery Measure at route level to monitor a route's contribution to freight performance. The Freight Delivery Metric (FDM) measures the percentage of commercial freight services that arrive at planned destination within 15 minutes of their booked arrival time or with less than 15 minutes of Network Rail or passenger operator delay. The Freight Delivery Metric at a regional level (FDM-R) assigns 'FDM failures' to Network Rail Regions based on the Delay, Cancellation or Service Variation events affecting each qualifying train. ORR's Final Determination set baseline trajectories and regulatory minimum floors for CRM-P and FDM-R at a route level. Following the Putting Passengers First programme, these measures are now monitored at a Regional level. ORR monitors delivery against annual targets and regulatory floors set for each of Network Rail's five Regions, measuring a Region's contribution to performance and the relative performance between Regions. The rail industry uses both National FDM and FDM-R.This review will focus only on the FDM-R element
2.2 Business Objectives and Priorities	The objective of the Independent Reporter review is to measure the system reliability and data accuracy of CRM-P and FDM-R. As new measures, and a key focus of ORR's monitoring of train performance, it is critical that Network Rail, ORR and rail industry stakeholders have assurance of the quality of the data and robustness of the measure.

3 .0 SCOPE OF SERVICE AND DELIVERABLES	
3.1 Key requirements	The Reporter will be required to provide confidence grading; both an alpha (system reliability) and numeric (data accuracy) grading based on the most up to date dataset available during the commission (see Appendix 1). The Reporter will be required to review, comment and make recommendations on the: governance and methodology for transforming data from TRUST/PSS into CRM-P/FDM-R outputs, including the methodology employed to attribute delay and mileage data between Regions; reliability, quality, consistency, completeness and accuracy of reported data; and

	processes in place to produce, quality assure and provide consistent period end figures to customers including ORR. The CRM-P and FDM-R are calculated centrally by Network Rail's National Performance Analysis Team (NPAT) and FNPO Performance Team respectively. Therefore, we expect this commission to require working primarily with these teams as well as speaking to one of the Regions on the delay attribution process between Regions. To pre-empt any potential recommendations and minimise the risk of duplicating work, the Reporter should work with Network Rail and ORR to understand any known issues / existing work streams that could impact on potential gradings. The Reporter should familiarise themselves with the background to these metrics including reviewing the PR18 train performance trajectories assessment and Review of Freight Delivery Metric. Additionally, the Reporter should review previous Independent Reporter reviews that have covered accuracy of data out of TRUST, including: • Review of new performance metrics (2017) • Right Time Performance review (2013) • Review of performance measures (2013) NB. Network Rail is in the process of making improvements to methodology for calculating train kilometres within the performance data warehouse (PSS). It is anticipated this work will be completed in time to feed into 2020-21 Period 4 reporting. Therefore, the Reporter is required to carry out a data quality assessment on data from the post-implementation phase.
3.2 Key deliverables	- Project inception meeting - Project update meetings - A joint feedback / interim findings session - Production of a draft report - Production of a final report
3.3 Proposed approach	[Demonstrate and detail the proposed approach for the project, covering all areas of the projects scope and clearly state the requirement(s)] *[Insert at contract award stage]*
3.4 Schedule & timings	Contract Start Date: 19/10/2020 Contract End Date: 31/01/2021 *These are indicative dates and will be agreed once the contract has been awarded and the PO has been approved.

	[Insert details pertaining to the commission's intended start and end date, as well as a commission schedule e.g., a Gantt chart with tasks and attributive start/end dates]
3.5 Relationship applicable for performing the duties under this statement of works contract	Data Controller and Data Processor. The only processing that the Supplier is authorised to do is listed as in Appendix 1 and may not be determined by the Supplier

4.0 KNOWLEDGE TRANSFER	
4.1 Knowledge Transfer	[Explain and detail how knowledge transfer is to be enabled throughout the commission and how the final output will be delivered and presented to Network Rail and ORR.] *[Insert at contract award stage]*

5.0 RESOURCE & COMMERCIAL DETAILS	
5.1 Supplier Resource	[Key personnel which will be engaged in the commission, along with their responsibilities. Details should include sub-contractors, if sub-contractors are being utilised for the delivery of this contract commission] *[Insert at contract award stage]* In the event of "key personnel" becoming unavailable the supplier agrees to provide a replacement of equal standard and status within 48 hours of notice.
5.2 Pricing Schedule	**This contract is based on a FIXED PRICE contract commission payable on completion** All prices detailed are exclusive of VAT which will be charged at the prevailing rate.
5.3 Payment Milestones	n/a

5.4 Place of work	Due to the current COVID-19 situation most of Reporter's work will be conducted from their own office or remotely.
5.5 Expenses	For the purpose of this contract, business travel expenses to any of Network Rail's offices [if this becomes necessary] may be claimed in accordance with Network Rail's Business Travel and Expenses policy.
5.6 Contract Variations	Variations to this Statement of Work contract may be permitted in accordance with Clause 88 of the Utilities Contract Regulations (modification of contracts during their term). All variations to this Statement of Work contract must be agreed in writing under a restated statement of works document, duly signed by all parties

6.0 INVOICING	
6.1 Invoice Details	Network Rail operates a strict "NO PO – NO PAYMENT" policy. Invoices are to be raised on completion of the contract or in accordance with the milestone payments [where applicable] set out in this SOW. Invoices should contain the following information as a minimum: • Purchase Order number • SOW number as detailed in Section 1.0 • Project Title and description Business expenses should be invoiced as a separate line and supported with receipts, as described in terms and conditions of the framework agreement and the Network Rail Business Expenses Policy. Please be aware that failure to provide the information above may potentially cause a delay in processing the invoice. Our preference wherever possible, is for invoices to be submitted via EDI. Alternatively, invoices may be submitted By email - invoices@networkrail.co.uk By post – Network Rail Accounts Payable, PO Box 4145, Manchester M60 7WZ

This Statement of Work will be executed as per the Terms and Conditions agreed in the Independent Reporter Services Framework Agreement.

[supplier name to be completed at contract award]

Signed:……..

Name (CAPS):……..

Position:……..

Date:……..

NETWORK RAIL

Signed:……..

Name (CAPS):……..

Position:……..

Date:……..

[This SOW does not require further contract signatures from the ORR]

ANNEX 1 – Protection of Personal Data

Where Data Controller and Data Processor applies

The Supplier shall only process personal data as detailed below:

Description	Details
Data Protection Officers	**Network Rail**: Fiona McConachie, The Quadrant, Elder Gate, Milton Keynes, Buckinghamshire, MK9 1EN **Supplier**: TBC at contract award stage
Subject matter of the processing	The processing is needed to ensure that the Processor can effectively deliver the services under the management Consultancy framework contract.
Duration of the processing	The duration of processing refers to the duration of the contract, as specified in the call-off contract
Nature and purposes of the processing	The nature of the processing means any operation such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction of data (whether or not by automated means). The purpose might include (but not limited to): statutory obligation, arranging Stakeholder meetings, data research and analysis and compliance with Network Rail's Business Travel and Expenses policy.
Type of Personal Data being Processed	This may include (but is not limited to): name, address, job title, location, email address, telephone number, images, cost centre number biometric data.
Categories of Data Subject	Examples include (but is not limited to): staff (including sub-contractors, volunteers, agents), customers/ clients, suppliers, students, apprentices, members of the public, users of a particular website.
Plan for return and destruction of the data once the processing is complete UNLESS requirement under union or member state law to preserve that type of data	On completion of the processing (interpreted as being contract expiry) the supplier shall cease to use the personal data and shall arrange for it's prompt and safe return to Network Rail, or destruction if instructed by Network Rail, of all Personal Data.


OFFICE OF RAIL AND ROAD


NetworkRail

Appendix B

Freight Delivery Metric – by Region or Route (Definition and references)

Freight Delivery Metric – by Region or Route (R-FDM)

Definition and references

Document Control

Version	Date	Description of Change	Author
1.0	07/03/2018	First issue of Document	Scott Provan
1.1	20/11/2020	Region amendments and additions	Scott Provan

Contents

Freight Delivery Metric - Route definition and outline process

Metric defined in words

The metric is the percentage of trains which Network Rail has delivered successfully. Failed to deliver is the percentage of commercial freight services that do not reach their destination within 15 minutes of their booked arrival time; and which have either been cancelled, or delayed 15 or more minutes, by Network Rail or another operator that is not a commercial freight operator (FOC).

The Region / Route split is derived from the standard allocation of a delay minute incident location to a location through the STANOX codes of the TRUST delay section. The reference data table for the allocation of the STANOX is part of the TRAA dataset in TRUST, Table B. Appendix 1A shows an extract of the TRUST TRAA tables.

Each qualifying train has all of the qualifying FDM delay proportionally split across the responsible Route(s). Some trains may only be delayed on a single Route. Qualifying cancelled trains are allocated to a Route based on the same incident location scenario, although some cancelled trains may not have TRUST incidents assigned to them. Both Delay Event Failures (DEF) and qualifying cancellations for the Regional values are an aggregation of the relevant Routes in that Region. Appendix 1B is a table showing the hierarchy of Route to Region mapping.

Metric 'sum'

The number of commercial freight trains Network Rail failed to deliver to satisfaction, divided by the Number of trains Network Rail could have delivered to satisfaction, expressed as a percentage point value

Metric formulae

$$1 - \left(\frac{\sum \text{FDM Delayed Trains by geography} + \sum \text{NR Assumed Cancelled Trains by geogrpahy} + \sum \text{NR caused Service Variations by geogrpahy}}{\sum \text{Trains Ran by geography} + \sum \text{NR Assumed Cancelled Trains by geogrpahy}} \right)$$

The term 'by geography' means either 'by Region' or 'by Route'. The grouping remains the same for all metric components i.e. Region and Route values will not be mixed.

Definitions

Trains:	Unless otherwise stated a train is any commercial Freight service which is not a Class 0 (light loco movement). This includes Very Short-Term Plan (VSTP) movements and unloaded services. A commercial Freight service is defined by their train service code, through exclusion.
Trains Run:	The number of trains which departed their planned origin and arrived at their planned destination (no matter what route was taken) and no matter how late. In March 2018 approximately 70% of all commercial freight services run over more than one Route. The Trains Run by Region or Route values are derived from the proportional split of the national value of Trains Run, pro-rata to the across all geography Trains Run value. Appendix 3A gives a sample period of a table showing the proportional split of Trains Run across the Regions; Appendix 3B shows Trains Run values by Route.
Network Rail Assumed Cancellation:	The number of non-commercial FOC caused cancellations as captured in the freight Schedule 8 performance regime but adjusted to capture the fact that cancelled unloaded trains are not included in Schedule 8. To keep things simple all intermodal cancellations will count as 1 failure, all non-intermodal cancellations will count as 2 failures (thereby assuming the following empty train movement would have also been cancelled). The count by Route is derived from the incident location that caused the Cancellation. Region values are the aggregation of the Route values within that Region.
Network Rail Delayed Train:	A train that departs its planned origin and arrives at its planned destination but arrives at its planned destination 15 or more minutes late and has 15 or more minutes of delay attributed in total to Network Rail plus other Passenger and other non-commercial Freight operator responsible managers (i.e. excluding all FOC commercial causes, not just FOC on Self, and planned delays). For R-FDM the delay for each affecting Route incident is proportioned across the total FDM delay for the train. Regional values are the aggregation of the Route values within the Region. Appendix 2 shows a worked example of the calculation. Appendix 1B shows the Route to Region mapping.

Definitions (continued)

Network Rail caused Service Variation:	The number of trains which were subject to a service variation with non-commercial Freight caused service variation minutes in the Schedule 8 performance regime. The count by Route is derived from the incident location that caused the Service Variation. Regional values are aggregated from the Route values within that Region.
Region	One of five (as of April 2019) geographical Network Rail Region based business units. Within the five Regions there are 14 Route business units. Appendix 1B is a table showing the Region and Route mapping.
Route	One of 14 (as of April 2019) geographical Network Rail Route based business units. This includes Route Scotland and Route High Speed One. Route High Speed One is not regulated infrastructure and therefore any R-FDM failures within that Route are only captured nationally and are excluded from R-FDM values. Route Scotland has the same geographical boundaries as the Region, therefore, to avoid a misleading duplicate set of values for Scotland, Route-FDM Scotland values are not calculated.

Granularity and Reporting

A Region or Route value will be produced each period, with the aggregation of 13 reporting periods to a rolling annual value. The R-FDM figure will be produced on the Thursday of week one to allow reporting in to the Regional and Route Scorecards. All R-FDM results will be expressed as a percentage point value to an accuracy of one decimal place.

Handling disputed minutes and cancellations

Network Rail delayed trains will be calculated using data extracts from Business Objects PSS based on the delay attribution within the system at that time. The potential impact of incidents in dispute is monitored. Any potential Network Rail assumed cancellations that are logged as disputed between Network Rail and the Operator will be assumed to be the responsibility of Network Rail as this presents R-FDM results as a ‘worst-case’ scenario.

Data collation – train running and associated delay data

The current Network Rail reporting tool (Business Objects PSS) has several template documents that are used to extract train count and delay attribution data for qualifying freight services from imported TRUST data. A rolling historic 26 period data extract is run to cover any potential attribution changes in that time.

Reference data for the extract is governed by CORPUS and TRUST TRAA tables, so changes to the National Reference Database are captured. Data for all holders of a freight operating license is captured for the extract period, even those that have ceased operation.

Data collation – Schedule 8 Service Variation and Cancellation (SV&C) data

For the Assumed Cancellation and Service Variation element of R-FDM, data is provided by the Freight Regime Specialist role, based within the Freight and National Passenger Operators Directorate. The values seen are derived from claims submitted from the Operator(s) and are validated through the Schedule 8 performance regime. The Freight Regime Specialist allocates Service Variation and Cancellation claims to a Route at this time which are aggregated to provide Regional values. Cancelation events may occur on infrastructure not controlled by Network Rail, e.g. Cardiff Valley Lines. Any validated Cancellation events on this type of off-network infrastructure are only counted towards the National value of FDM.

FDM calculation – data consolidation

Several MS Excel workbooks are used to collate all of the elements of R-FDM (Assumed Cancellations, Delayed Trains, Trains Run etc.) and produces results (by formulae) for Regional and Route values of R-FDM for period, year-to-date and a rolling annual time-frame.

Data quality checks

Not every 'freight' service on the network is monitored for R-FDM – see Trains Run above: National Supply Chain and Yellow Plant movements, for example, are amongst the types of train excluded. Every six months a full train list of potential qualifying freight services is extracted for a period, with 'flags' added to show why the exclusions are applied. This is done to monitor potential planning errors and if anomalies are spotted, other periods of data will be reviewed. The extract file used for the ORR Reporter Audit in summer 2016 contained 42,617 potential services, with 11,908 qualifying as R-FDM monitored trains.

Supplemental data checks are also conducted using the National FDM results as a base, compared to the aggregated Region or Route R-FDM values. Appendix 4 shows a comparison data table between the sum of all Regional and Route R-FDM and National values of FDM.

End

Appendix 1A: Sample of a TRUST TRAA data table

TCTRD77 TRUST NETWORK DISPLAY - DELAY LOCATIONS

COMMAND:
:DELAY LOCATIONS:

ATTR REQUEST
ATR POINT : 30120 R15
IN SERVICE : Y LOCATION TYPE : S REPORTS FOR :

NO	NAME	STNX	TYP	ATTR REQUEST
2	FRNGTNCJN	30164	S	R15
2	FRNGTNCJN	36001	S	R17
2	FRNGTNCJN	30139	S	R17
1	GARSTNG&C	11727	S	R15
4	PRSTNDKST	30117	P	R15
3	PSTN FYJN	30044	S	R17

TCTRD78 TRUST NETWORK - RESP MANAGERS AND DELAY COPY

COMMAND:

RESPONSIBILITY MANAGERS FOR INITIAL DEL CODE CHAR

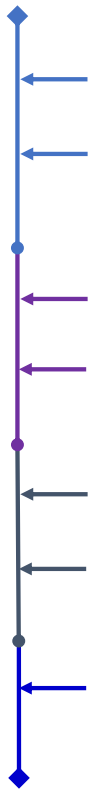
DEL COPY		I	O	P	Q	X
ATR PT : 30120		IQFP	OQFP	PQFR	QQA0	XQFO
RTE	FWD-RTE DEL COPY					
2	30164	IQFP	OQFP	PQFR	QQA0	XQFO
2	36001	IQFP	OQFP	PQFR	QQA0	XQFO
2	30139	IQFP	OQFP	PQFR	QQA0	XQFO
1	11727	IQFP	OQFP	PQFR	QQA0	XQFO
4	30117	IQFP	OQFP	PQFR	QQA0	XQFO
3	30044	IQFP	OQFP	PQFR	QQA0	XQFO

In the above table, the third character of the Responsible Manager Code gives the Route location e.g. IQFP where F is the reference letter for Route North West.

Appendix 1B: Route to Region mapping with BOPSS abbreviations and codes

Region Name	*Region Code*	*Region Abbreviation*	*Route Name*	*Route Code*	*Route Abbreviation*
Eastern	J	ZEST	Anglia	H	REAN
			East Coast	I	REAC
			East Midlands	V	REMD
			North & East	G	RNEA
North West & Central	R	ZNWC	Central	E	RCTL
			North West	F	RNWE
			West Coast South	N	RWCS
Scotland	S	ZSCO	Scotland	L	RSCO
Southern	K	ZSTH	Kent	M	RKNT
			Sussex	B	RSUS
			Wessex	C	RWEX
Wales & Western	U	ZWST	Wales	W	RCYM
			Western	D	RWES

Appendix 2: Example of single train for proportional R-FDM calculation



Origin	Felixstowe, 2 minute late start
Delay Event (1)	Point failure at Colchester 16 minutes primary delay
Delay Event (2)	Point failure at Colchester 14 minutes reactionary delay at Stratford 30 minutes total delay to REAN / ZEST as NR on FOC delay
Region & Route Boundary	Anglia Route / Eastern Region to West Coast South Route / North West & Central Region
Delay Event (3)	Failed LNWR train at Northampton 10 minutes primary delay to RWCS / ZNWC as TOC on FOC delay
Delay Event (4)	Freight (different FOC) train consist issues off network 5 minutes reactionary delay through regulation - NOT R-FDM delay as FOC on FOC
Route Boundary	West Coast South Route to North West Route
Delay Event (5)	Signal Failure at Winsford 12 minutes primary delay to RNWE / ZNWC as NR on FOC delay
Delay Event (6)	Driver change at Preston 15 minutes primary delay - NOT R-FDM delay as FOC on Self
Region & Route Boundary	North West Route / North West & Central Region to Scotland
Delay Event (7)	Caution for animal incusion 8 minutes primary delay to RSCO / ZSCO as NR on FOC delay
Destination	Coatbridge, terminated 75 minutes late

Summary

Train is a national FDM failure as it has terminated above the 15 minute lateness threshold and recevied 15 or more minutes qualifying FDM delay
Train also received 20 non-FDM delay minutes: Delay event (4) 5 FOC on FOC and Delay Event (6) 15 minutes FOC on Self
Train Count: one nationally; three across the Regions and four across the PPF Routes

R-FDM Failures

Eastern Region / Anglia Route
0.500 for 30 out of 60 FDM qualifying delay minutes - Delay Event (1) and Delay Event (2)

North West & Central Region / West Coast South Route
0.167 for 10 out of 60 FDM qualifying delay minutes - Delay Event (3)
North West & Central Region / North West Route
0.200 for 12 out of 60 FDM qualifying delay minutes - Delay Event (5)
North West & Central Region caused 0.367 FDM failures overall

Scotland
0.133 for 8 out of 60 FDM qualifying delay minutes - Delay Event (7)

Appendix 3A: R-FDM Trains Run values across Regions

Financial Year & Period	Financial Year	Financial Period Number	Region Abbrev	Region Name	Base Train Count across geography	Total Train Count across geography	% of Total Train Count for geography	National FDM Train Count	New R-FDM Train Count for geography (from TC%)
2020/21_P01	2020/21	1	ZEST	Eastern	6,595	16,741	39.4%	10,566	4,162
2020/21_P01	2020/21	1	ZNWC	North West & Central	5,020	16,741	30.0%	10,566	3,168
2020/21_P01	2020/21	1	ZSCO	Scotland	908	16,741	5.4%	10,566	573
2020/21_P01	2020/21	1	ZSTH	Southern	1,221	16,741	7.3%	10,566	771
2020/21_P01	2020/21	1	ZWST	Wales & Western	2,997	16,741	17.9%	10,566	1,892

Appendix 3B: R-FDM Trains Run values across Routes

Financial Year & Period	Financial Year	Financial Period Number	Route Abbrev	Route Name	Base Train Count across geography	Total Train Count across geography	% of Total Train Count for geography	National FDM Train Count	New R-FDM Train Count for geography (from TC%)
2020/21_P01	2020/21	1	RCTL	Central	1,974	25,319	7.8%	10,566	824
2020/21_P01	2020/21	1	RCYM	Wales	1,123	25,319	4.4%	10,566	469
2020/21_P01	2020/21	1	REAC	East Coast	2,982	25,319	11.8%	10,566	1,244
2020/21_P01	2020/21	1	REAN	Anglia	2,782	25,319	11.0%	10,566	1,161
2020/21_P01	2020/21	1	REMD	East Midlands	1,874	25,319	7.4%	10,566	782
2020/21_P01	2020/21	1	RKNT	Kent	349	25,319	1.4%	10,566	146
2020/21_P01	2020/21	1	RNEA	North & East	4,187	25,319	16.5%	10,566	1,747
2020/21_P01	2020/21	1	RNWE	North West	2,892	25,319	11.4%	10,566	1,207
2020/21_P01	2020/21	1	RSCO	Scotland	908	25,319	3.6%	10,566	379
2020/21_P01	2020/21	1	RSUS	Sussex	306	25,319	1.2%	10,566	128
2020/21_P01	2020/21	1	RWCS	West Coast Mainline South	2,533	25,319	10.0%	10,566	1,057
2020/21_P01	2020/21	1	RWES	Western	2,432	25,319	9.6%	10,566	1,015
2020/21_P01	2020/21	1	RWEX	Wessex	977	25,319	3.9%	10,566	408

Appendix 4: National FDM to R-FDM data comparison

Financial Year	Financial Period	Proportional R-FDM delayed trains	National Aggregator Delayed trains	Difference in period Delayed Trains (National FDM from aggregator as base)	Variance (National FDM as Base)	Proportional R-FDM train count	National Aggregator train cout	Difference in period Delayed Trains (National FDM from aggregator as base)
2019/20	P09	1,098	1,095	-3	0.3%	13,322	13,322	0
	P10	532	530	-2	0.4%	9,473	9,473	0
	P11	763	762	-1	0.1%	13,029	13,029	0
	P12	970	970	0	0.0%	12,740	12,740	0
	P13	731	731	0	0.0%	13,496	13,496	0
2020/21	P01	149	148	-1	0.7%	10,566	10,566	0
	P02	228	226	-2	0.9%	9,712	9,712	0
	P03	257	256	-1	0.4%	10,799	10,799	0
	P04	277	275	-2	0.7%	11,476	11,476	0
	P05	425	424	-1	0.2%	11,471	11,471	0
	P06	477	472	-5	1.1%	11,107	11,107	0
	P07	477	472	-5	1.1%	11,878	11,878	0
	P08	575	573	-2	0.3%	12,186	12,186	0
13 Period Totals (for MAA)		6,959	6,934	-25	0.4%	151,255	151,255	0

Appendix C

Confidence Grading System

C1 System reliability grading system

Table 16: System reliability grading system

System reliability band	Description
A	Sound textual records, procedures, investigations or analysis properly documented and recognised as the best method of assessment.
B	As A but with minor shortcomings. Examples include old assessment, some missing documentation, some reliance on unconfirmed reports, some use of extrapolation.
C	Extrapolation from limited sample for which Grade A or B data is available.
D	Unconfirmed verbal reports, cursory inspections or analysis.

Notes:

1. System reliability is a measure of the overall reliability, quality, robustness and integrity of the system that produces the data.

2. Some examples of the potential shortcomings include old assessment, missing documentation, insufficient internal verification and undocumented reliance on third-party data.

C2 Accuracy Grading System

Table 17: Accuracy grading system

Accuracy Band	Description
1*	Data used to calculate the measure is accurate to within 0.1%
1	Data used to calculate the measure is accurate to within 1%
2	Data used to calculate the measure is accurate to within 5%
3	Data used to calculate the measure is accurate to within 10%
4	Data used to calculate the measure is accurate to within 25%
5	Data used to calculate the measure is accurate to within 50%
6	Data used to calculate the measure is inaccurate by more than 50%
X	Data accuracy cannot be measured

Notes:

1. Accuracy is a measure of the closeness of the data used in the system to the true values.

2. Accuracy is defined at the 95% confidence level - i.e. the true value of 95% of the data points will be in the accuracy bands defined above.

Appendix D

List of files supplied to the Reporter Team

D1 List of files supplied to the Reporter Team

Table 18: List of files supplied to the Reporter Team

File Name	Type	From
2019-20 R-FDM dataset.xlsb	Excel Workbook	Freight Directorate (NR)
2020-21YTD R-FDM dataset.xlsb	Excel Workbook	Freight Directorate (NR)
FPO-01 R-FDM 3TL 1TIN Class 5 split.xlsx	Excel Workbook	Freight Directorate (NR)
R-FDM aggregator FRAug2020 Class 5 split.xlsm	Excel Workbook	Freight Directorate (NR)
FPO-01 R-FDM TC MILEAGE.xlsx	Excel Workbook	Freight Directorate (NR)
R-FDM TC aggregator by Class 5 and geograhpy.xlsm	Excel Workbook	Freight Directorate (NR)
FDM SVandC data from Schedule 8.xlsx	Excel Workbook	Freight Directorate (NR)
FDM for SFC 2020 P8.xls	Excel Workbook	Freight Directorate (NR)
909_P2108 R-FDM ORR Route and Region data.xlsx	Excel Workbook	Freight Directorate (NR)
Definition Region and Route Freight Delivery Metric 20201120.docx	Word Document	Freight Directorate (NR)
R-FDM consolidation file for YTD and MAA values.docx	Word Document	Freight Directorate (NR)
FDM national aggregator Class 5 split.xlsx	Excel Workbook	Freight Directorate (NR)
model-freight-contract-august-2016.pdf	PDF	Freight Directorate (NR)
ORR Reporter FDM-R BOPSS query screenshots 20201211.docx	Word Document	Freight Directorate (NR)
P2108 FDM A2F ORR Reporter All Trains list.xlsx	Excel Workbook	Freight Directorate (NR)
Sch8 Capes and SV flowchart 20180308.pdf	PDF	Freight Directorate (NR)
356 FDM-R format changes for loading.msg	Email	Freight Directorate (NR)
ORR Reporter FDM-R BOPSS query screenshots 20201211.docx (Updated)	Word Document	Freight Directorate (NR)
Definitions of Railway Performance Metrics_v3.03.docx	Word Document	NPAT Team (NR)
CRM-P Delays Query.docx	Word Document	NPAT Team (NR)
CRM-P Mileage Query.docx	Word Document	NPAT Team (NR)
CRM-P Latest Output.xlsx	Excel Workbook	NPAT Team (NR)
CRM-P Raw Data.xlsx	Excel Workbook	NPAT Team (NR)
Annual Return (2019-20)	Multiple Excel Files	NPAT Team (NR)
crm-p change one slide.pptx	Powerpoint	NPAT Team (NR)
FW CRM-P - Impact of changed Mileage Data.msg	Email	NPAT Team (NR)
FW CRMP Files.msg	Email	NPAT Team (NR)
Impact of Change.xlsx	Excel Workbook	NPAT Team (NR)
Summary of CRMP trajectory Change.docx	Word Document	NPAT Team (NR)

File Name	Type	From
CRM-P Raw Data - Delays amended.xlsx	Excel Workbook	NPAT Team (NR)
CRM-P Raw Data - Mileages amened v1.xlsx	Excel Workbook	NPAT Team (NR)
CRM-P Raw Data - Mileages ammended v2.xlsx	Excel Workbook	NPAT Team (NR)
CRM-P Raw Data - refreshed.xlsx	Excel Workbook	NPAT Team (NR)
Region, Route and Area Listing.xlsx	Excel Workbook	NPAT Team (NR)
TSC Codes.xlsx	Excel Workbook	NPAT Team (NR)
PPM Mileages.xlsx	Excel Workbook	NPAT Team (NR)
Business Objects Controlled Folder Standards.docx	Word Document	NPAT Team (NR)
1 - TRUST v PSS Daily audit.xlsx	Excel Workbook	Process & Controls Team (NR)
05 - Changes to TRUST reference data.pdf	PDF	Process & Controls Team (NR)
06 - Re-franchising geography boundary changes.pdf	PDF	Process & Controls Team (NR)
24 - Internal Dispute Process.pdf	PDF	Process & Controls Team (NR)
EESIC January 2020.doc	Word Document	Process & Controls Team (NR)
IDA22 Infrastructure APRIL 2020.pdf	PDF	Process & Controls Team (NR)
Incidents not using the default.xlsx	Excel Workbook	Process & Controls Team (NR)
Internal disputes - all (004).xlsx	Excel Workbook	Process & Controls Team (NR)
KPI DAB 2021 P08.pdf	PDF	Process & Controls Team (NR)
PDAC November 2020 (clean).pdf	PDF	Process & Controls Team (NR)
PGD22 Other Infrastructure Manager Networks.pdf	PDF	Process & Controls Team (NR)
RE-ORG REQUIREMENTS DOCUMENT Master v1.xlsx	Excel Workbook	Process & Controls Team (NR)
Revenue freight revisions.xlsx	Excel Workbook	Process & Controls Team (NR)
September 2020 DAPR.pdf	PDF	Process & Controls Team (NR)
27 - Key Performance Indicators (KPI's) and Verification Checks.pdf	PDF	Process & Controls Team (NR)
Manage Data and Data Reference Points v0.4_.xlsx	Excel Workbook	Process & Controls Team (NR)
Manage the Delay Attribution Framework Resolve Disputes v0.4_.xlsx	Excel Workbook	Process & Controls Team (NR)