Network Rail and Office of Rail and Road

## Independent Reporter - Lot 3

Mandate L3 AR 004: Review of
New Performance Metrics

Issue | 18 July 2017

Job number    254086-00

ARUP

# Contents

**Appendices**

**Appendix A**

Mandate

**Appendix B**

Best Practice Guide

**Appendix C**

Confidence grading system

# Executive Summary

## Background

In order to improve industry performance and better reflect passenger needs and priorities, a suite of new performance metrics has been developed by a National Task Force (NTF) sub-group and proposed for use in Control Period 6 (CP6: 2019-24). The measures are intended to complement and potentially supersede the existing PPM and CaSL metrics. As part of the development process, the Office of Rail and Road (ORR) and Network Rail commissioned an Independent Reporter review of some of the metrics included in the suite, to obtain an indication of their current accuracy, reliability and general suitability for use in CP6, and also the extent to which they might usefully be improved.

## Purpose and Objectives of Review

The metrics included in the review are as follows:

- Total Passenger Lateness – Passenger Time Lost (million hours)

- Reliability – Cancellations (%)

- Reliability –Severe Disruption (no. of days with > x% cancelled)

- On time at all recorded Stations – On Time (< 1 minute late) (%)

- On time at all recorded Stations –Time to 15 (< 15 minutes late) (%)

Note that while the review was underway, it was decided by the NTF sub-group that Average rather than Total Passenger Lateness should be reported; this report reflects our findings in respect of the originally-proposed Total Passenger Lateness metric but we do also comment on issues relating to Average Passenger Lateness.

For each metric, the purpose of the review was to review its reliability and accuracy in terms of the standard Independent Reporter confidence grading system (as set out in Appendix 2 to the project mandate), and identify any necessary actions prior to CP6 to ensure its suitability for use as an industry performance metric. Because the metrics are new, and at a relatively early stage of development, it was emphasised that the review should be forward-looking and comparatively indicative, in contrast to the more definitive, retrospective approach typically taken to the review of established metrics and measures. The remit included:

- Review of and commentary on input data used;

- Review of and commentary on accuracy, reliability and general quality of metric outputs;

- Guidance on steps to be taken to improve metric robustness;

- Provision of indicative current confidence grades;

- Identification of potentially achievable confidence grades, and the required effort and associated value for that effort;

- Identification of any serious issues affecting the implementation of the metrics; and

- Provision of a view on the likely effectiveness of the metrics in improving industry behaviours and alignment.

## Approach

Following the project inception meeting, meetings were held with Network Rail staff responsible for producing the metrics and providing the underlying data, and with Route-based Performance staff who will be accountable for the metric outputs for their Routes and for meeting any targets set for the metrics. Meetings were also held with Train Operating Companies' (TOCs') representatives at the Rail Delivery Group (RDG) and with RDG's Communications team, responsible for communicating the purposes and outputs of the metrics both within the railway industry and outside, to its users. One outcome from these meetings was a request for us to consult with TOC performance staff to gain their views and understand any concerns and so, in addition to our original scope, we held telephone conversations with four TOCs.

Samples of metric input data were obtained from Network Rail, together with copies of the spreadsheet tools and process documentation used to generate the metric outputs. Reference was also made to previous Independent Reporter work undertaken in relation to some of the metric data sources. To aid our understanding, we developed process maps for the metrics, indicating the data sources used and the steps taken to process the data to produce the metric outputs.

The information thus gathered was then analysed and reviewed to produce our overall findings, recommendations and observations, which were in turn reviewed to establish the conclusions and confidence grades reported below.

## Conclusions

The metrics have cross-industry support and endorsement, and the consensus from the meetings is that they should have the desired effect of improving those aspects of performance of most importance to railway users. The calculation processes used to produce the metrics are sound, although the process documentation needs some improvement. The coverage and accuracy of some of the underlying performance data, however, needs to be reviewed in particular the berth offsets at station timing points for the On Time metric. The transparency and documentation of the processes for producing Average Minutes Lateness and passenger journeys data are also in need of improvement.

In Table 1 we have summarised our views of the metrics against their objectives and principles that were expressed to us in the various meetings. We have used a 4-scale "RAG" assessment ranging from green (fully meets the objective / principle) to red (fails to meet it). It can be seen that, in our view, the metrics fully meet many of their objectives and principles. We have some minor concerns

| Issue | 18 July 2017

Page 2

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

(yellow) that the individual metrics on their own do not focus on all issues relevant to passenger interest, but this reinforces the message that they should be seen as a basket of metrics rather than focussing on any individual one. This should also be borne in mind when deciding how these metrics should be regulated by the ORR.

We also have minor concerns over the behaviours that will be incentivised by most of the metrics. Some of these concerns were expressed to us in our meetings. We believe that practical details will need to be worked through before CP6 to produce daily, weekly and period targets that can be actively managed by staff in the industry. This will need to be backed up by a comprehensive education campaign within the industry to explain the rationale behind the new metrics and expectations on targets to achieve. More specific interventions may be required on how control strategy or timetable planning may need to evolve.

The one aspect over which we have some major concern is the transparency of the passenger lost hours metric (whether total passenger lateness or average passenger lateness). Many of the people we met did not understand how it is calculated, in particular the calculation of Average Minutes Lateness, and to some degree it is viewed as a 'black box'. In our opinion this is going to make it difficult for the industry to know how best to reduce passenger lateness. Instead, this metric could be viewed as an outcome of the other metrics – if reliability and punctuality of trains are improved then that should improve (i.e. reduce) passenger lateness.

Communicating the metrics outside of the industry will need careful consideration. In our view the most attention will be paid to the On Time metric. The industry faces a particular challenge in explaining why PPM values of close to 90% are being replaced with significantly lower on-time performance measure (~60%). It will therefore be important to reinforce the message that the On Time metric (alongside the other metrics) demonstrates to passengers that 'every minute of lateness counts', and the industry has both ambition and enthusiasm to provide improved performance through its increased transparency in reporting. In contrast most other measures will perform better – the Time to 15 is likely to be in the mid to high 90% range for most TOCs (which is higher than PPM) and cancellations are likely to be nationally about 2% (and lower than CaSL). A cross industry communication strategy will be needed supported by RDG and Network Rail.

| Issue | 18 July 2017

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Page 3

## Table 1: Overall Conclusions

| Performance Metric | Objectives | | | Principles | | |
|---|---|---|---|---|---|---|
| | Provide a level of transparency in respect of the impact of performance on passengers | To focus on issues relevant to passenger interest | Provide right incentives for NR, TOCs and funders | Simple to understand | Future proofed with scope for development | Alignment with contractual relationships / regulation |
| Total Passenger Lateness (TPL) | TPL is the only metric that focuses on passengers rather than trains. The concept is the same as lost hours reported by London Underground. | This metric combines the impact of cancellations and punctuality on overall passenger lateness. | Improvements to the above metrics should improve this metric. But it relies on the calculation of Average Minutes Lateness which is viewed by some as a black box, making it difficult to know directly how to influence the metric | It is not a simple calculation | Eventually this metric could be improved with access to and analysis of passenger movement data | Average MInutes Lateness is taken from that calculated for Schedule 8 payments |
| Cancellations | Measures a direct impact on passengers. Cancellations on high frequency routes will have different impacts than on less frequently served routes (e.g. intercity) which will be shown by reporting by TOC | Transport Focus research concluded that passengers want a reliability metric that is separate to punctuality; note this metric does not take account of the number of passengers affected | An important counter-balance to the On Time metric, i.e. to prevent cancelling trains to preserve Right Time when not necessarily in the best interests of passengers | The definition of full and partial cancellations against the Applicable Timetable are clear | Currently the NR and TOC numbers of cancellations come from PSS and Bugle respectively and are then reconciled. Note that the proposed ITED should provide greater visibility to aid this reconciliation (but not prior to CP6) | The definition is the same as that used in Schedule 7.1 of franchise agreements |
| Severe disruption | This metric sheds light on the number of days when the train service is severely disrupted and provides greater transparency than the aggregate CaSL metric | This metric was suggested by Transport Focus to measure the impact of infrequent but large delays as distinct from more consistent delays. As for cancellations, no account is taken of numbers of passengers impacted | To avoid exceeding the % threshold of trains cancelled, this metric should help to incentivise running trains in times of severe disruption and avoid undue cancellations for the sake of preserving punctuality. Sub-operators with fewer trains might be penalised more often by applying the same threshold as larger operators | The thresholds of 20% (sub-operator) and 5% (national) cancellations are clear | Currently the NR and TOC numbers of cancellations come from PSS and Bugle respectively and are then reconciled. Note that the proposed ITED should remove the need to do this (but not prior to CP6) | The definition of a cancellation is the same as that used in Schedule 7.1 of franchise agreements |
| On time | Measuring Right Time at all recorded station stops is a more detailed measure of punctuality than PPM, and includes the impact on all passengers not just those at the train's destination | Transport Focus' research reportedly showed that passengers want on time running against the 'promise' of the timetable. Note, though, that it averages all recorded station stops and is not weighted by numbers of passengers | Will help incentivise right time running, which in turn will help Time to 15 and APL metrics; will also help preserve train connections for passengers changing trains. Some concern expressed over resources needed to identify root causes of sub-threshold delays | The definition is clear to understand | There is the potential for measuring to the nearest second in future (rather than truncating timings to minutes) | If this metric is going to be regulated then analysis of variation will need to be undertaken to set an appropriate target |
| Time to 15 | Time to 15 is a useful measure of the impact of longer delays on passengers. Note that by measuring timings at all recorded stations along a train's route, it is less useful as a comparison against the air industry | As for On time, this metric is an average of all recorded station stops rather than weighted by passenger numbers | Time to 15 will help incentivise the reduction of significantly late trains. However, by including all recorded stations, a trial suggests the metric will be close to 100% and there is a question over its sensitivity to causes of delay | The definition is clear to understand | There is the potential for measuring to the nearest second in future (rather than truncating timings to minutes) - though will have little impact on this metric | Time to 15 has a broad alignment with new delay-repay mechanisms |

| Colour coding | No aspects met | Some aspects met | Mostly met | Fully met |
|---|---|---|---|---|

| Issue | 18 July 2017

Page 4

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

# Accuracy and reliability of metrics

We have assessed the accuracy and reliability of the metrics with reference to the Independent Reporter Confidence Grading system shown in Appendix C. Table 2 summarises our views on the current grades and those that could be achieved with improvements.

Table 2: Confidence Grades

| Metric | Current Confidence Grade | Potential Confidence Grade with Recommended Improvements |
|---|---|---|
| Total Passenger Lateness – Passenger Time Lost (hours) | C4 | A2 (possibly A1 in the longer term) |
| Reliability - Cancellations | B2 | A2 |
| Reliability – Severe Disruption | B1 | A1* |
| On Time at All Recorded Stations – On Time | C2 | A1-2 |
| On Time at All Recorded Stations – Time to 15 | C1 | A1* |

Improvements in documentation will be needed in order to improve the reliability of the Passenger Time Lost and two On Time at All Recorded Stations metrics. In addition, all three metrics rely on the accuracy and currency of berth offsets. In some Routes these offsets are checked and updated by data quality specialists but in other Routes they are the responsibility of staff with wider roles. Data provided by Network Rail suggests that those Routes without dedicated data quality specialists are less compliant to the annual review of berth offsets, which increases the risk that they do not reflect any recent infrastructure and / or rolling stock changes.

With the new metrics measuring punctuality at more stations (not just at destination required for PPM), this problem is likely to become more acute. Without adequate resourcing to support good data quality, it is difficult to see how the reliability and accuracy of these metrics can be improved. Longer term, a technological solution could be developed. This could include the use of GPS or possibly the use of Radio-frequency identification (RFID) tags.

The accuracy of the calculation of Passenger Time Lost is the lowest for all of the metrics. The passenger journeys file used as an input double counts people changing trains along their journey (it is a count of trains caught rather than journeys made). Assuming the measure is lateness at destination, then we estimate that this means Total Passenger Lateness could be over-stated by about

| Issue | 18 July 2017

Page 5

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

15%, however the impact on Average Passenger Lateness should be fairly minimal. If this is corrected, then the accuracy should be improved, although the method still fails to properly account for connections that are made or lost.

In our opinion, achieving an accuracy grade of 1 (within 1%) for Total Passenger Lateness can only be achieved in the longer term by more accurate tracking of passenger movements through the railway system by means of, for example, the use of smartcard technology. That said, a number of challenges would need to be overcome. Smartcards only record actual times of entry to and exit from the system rather than the planned journey times unless tickets are pre-booked. This means they would not measure the lateness experienced by passengers who, for example, alter their travel times in response to advance knowledge of late running of trains.

## Recommendations and Observations

We have made a number of recommendations and observations which are listed in chapter 7 of this report. In summary the recommendations cover:

- Adding to the documentation of the processes for producing all the metrics to improve transparency and ensure business continuity. This also includes formalising the calculation of Average Passenger Lateness following the decision to report it;

- Recalibrate all monitoring point weights to ensure they are up-to-date in time for the start of CP6 for Total and Average Passenger Lateness (TPL and APL); also review the suitability of the use of PEARS payment rates to determine the peak / off peak weightings in aggregating APL;

- Ensure there is a defined list of stations for monitoring On Time and Time to 15, and that there is a formal process for adding and removing stations in future;

- To help improve the accuracy of On Time, ensure there are sufficient staff in each of the Routes to achieve at least 90% compliance of annual desktop checks of berth offsets for Contractual Monitoring Points and Delay Recording Points by the start of CP6; and to check and update the offsets for Station Timing Points according to a suitably defined process.

- Develop a longer term strategy for use of technology to improve the reliability and accuracy of reporting train arrival and departure times at stations;

- Publish an industry code of practice setting out the rules for responding to the metrics, for example in terms of adding allowances to the timetable or publication of separate arrival and departure times; and

- To allow industry staff to manage train performance, ensure they are briefed on the new metrics, also ORR to decide if and how the metrics will be regulated in CP6.

| Issue | 18 July 2017

Page 6

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

# 1      Introduction

This report presents the findings of the review of additional performance metrics undertaken by Arup and Winder Phillips Associates (WPA) in the course of their roles as Lot 3 Independent Reporter. The metrics were proposed by a National Task Force (NTF) sub-group with representatives from Network Rail, Office of Rail and Road (ORR), Department for Transport (DfT) and Rail Delivery Group (RDG) for potential use in Control Period 6 (CP6) from 2019 to 2024. The review was undertaken in response to Mandate L3 AR 004 (included at Appendix A), dated 12th December 2016 and issued by the ORR and Network Rail on 14th December 2016.

Following this introduction, the objectives of the review and the methodology adopted are described in Chapter 2. Chapter 3 provides some background to each of the new/proposed measures, explaining their purpose and initial development. Chapter 4 describes the process and findings of the review, covering each measure in turn. The possible effects of the metrics upon industry behaviours are considered in Chapter 5 and the Confidence Grades for the metrics are presented in Chapter 6. Finally, our Recommendations and Observations are set out in Chapter 7, followed by the Appendices, containing the project mandate (Appendix A) and our Best Practice Guide for the implementation and development of the metrics (Appendix B).

The mandate for this review contained a number of requirements and in Table 1.1 below we show where each requirement is addressed within this report.

Table 1.1: Addressing Mandate Requirements

| Mandate requirement | Section of report |
|---|---|
| Review and comment on the processes and procedures by which the industry captures other data used in these metrics. | 4.1.1, 4.2.1, 4.3.1 |
| Review and comment on the reliability, quality, consistency, completeness and accuracy of reported data in the trial. | 4.1.2, 4.2.2, 4.3.2 |
| Provide guidance on what steps may be reasonably taken to make the calculation of the metrics more robust, including recommendations on how to ensure route based metrics can be compared across routes fairly[1] | 7, Appendix B |
| Identify an indicative confidence grading for the system reliability and data accuracy of each metric as it stands currently | 6 |
| Identify which confidence grading might be achievable and outline what improvements would be needed (along with associated costings) which could be taken to reach that grading. This should take into account the value for money of any identified improvements.[2] | 6 |
| Identify any serious issues in implementing these measures across the industry. | 5.1 |
| Provide an opinion on whether the proposed metrics of performance are likely to be effective in creating more industry alignment by driving the right behaviours. | 5.2, 5.3 |

[1] The review of route based metrics was subsequently dropped

[2] At the review inception meeting, this was clarified to be a qualitative review

| Issue | 18 July 2017

Page 7

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

# 2 Review Objectives and Methodology

The measures included in the review are listed below:

- Total Passenger Lateness – Passenger Time Lost (million hours)

- Reliability – Cancellations (%)

- Reliability – Severe Disruption (no. of days with > x% cancelled)

- On time at all recorded Stations – On Time (< 1 minute late) (%)

- On time at all recorded Stations –Time to 15 (< 15 minutes late) (%)

While the review was underway, we were informed that Average rather than Total Passenger Lateness should be reported by the industry. We have reviewed the process given to us for calculating Total Passenger Lateness, however we have also commented on issues that we identified relating to Average Passenger Lateness including our understanding of its definition.

In addition to the above, the mandate included the review of two alternative options for the measurement of Passenger Weighted Disruption, partly with a view to enable cross-Route performance comparisons. However, it was subsequently agreed with Network Rail and ORR to exclude Passenger Weighted Disruption from the review.

## 2.1 Objectives

Work has been undertaken by a National Task Force (NTF) sub-group, representing Network Rail, ORR, the DfT and the RDG, to develop performance metrics for potential use in CP6 as alternatives to the existing Public Performance Measure (PPM) and Cancellations and Significant Lateness (CaSL) metrics. The purpose of the proposed metrics is to address some of the perceived limitations of the existing metrics, by providing improved incentives for Network Rail, train operators and system funders, and also to provide improved clarity and transparency of the effects of performance on railway passengers. Importantly, they will encourage the industry to focus more directly on passenger needs and expectations, including, for example, right-time arrivals.

The objectives of this piece of work, as set out in the mandate, are two-fold. The first stage is to review the proposed metrics and provide an indicative assessment of their reliability and accuracy. Recognising these are new metrics, this assessment should be an indicative view rather than the more definitive grading that Reporters carry out for established measures. The second main objective is to identify any shortcomings and the associated necessary actions to ensure that the metrics are sufficiently robust if they are to be used in CP6.

## 2.2 Methodology

The scope, activities and outputs of the review were discussed and refined at the Inception Meeting for the project held on 25th January 2017. One clarification made was that the Total Passenger Lateness metric makes use of the Average

| Issue | 18 July 2017

Page 8

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Minutes Lateness (AML) measure used in the Schedule 8 financial regime, and that its adequacy for the new metric should be reviewed in our work, together with the passenger ticket sales and TRUST performance data used.

A review of background documentation relating to the development of the metrics by NTF was also undertaken, to provide context and improved understanding of the objectives and eventual selection of the metrics.

A series of meetings was held with Network Rail staff (at Milton Keynes and on the Routes) and NTF representatives to obtain industry views of the metrics, the underlying data and their implications for the industry, and to understand and obtain samples of the data and tools used to produce the metrics. Following an initial meeting with RDG representatives to obtain the overall view of the Train Operating Company (TOC) community, it was agreed that additional meetings should be held with the following individual TOCs: Virgin West Coast (VWC), c2c, Northern and Southeastern. These TOCs were selected to provide a representative mix of TOCs across different passenger markets. The schedule of meetings is summarised in Table 2.1.

Table 2.1: Schedule of Meetings

| Date | Organisation | Purpose | Location | Attendees |
|------|--------------|---------|----------|-----------|
| 25-01-17 | Network Rail and ORR | Inception meeting | Arup, 8 Fitzroy St. | Client group |
| 06-02-17 | RDG | Obtain overall RDG view of metrics | RDG, 200 Aldersgate St. | RDG representatives |
| 08-02-17 | Network Rail | Review metric production process, obtain copies of data and tools | Network Rail, Milton Keynes | National Performance Analyst for new performance metrics |
| 15-02-17 | Network Rail | Information on and understanding of use of PEARS AML calculations in TPL metric | Network Rail, Milton Keynes | Performance Analysis Manager |
| 15-02-17 | Network Rail | Information on and understanding of use of TRUST data in metrics | Network Rail, Milton Keynes | Performance Data Quality Specialist; Performance Support Analyst, Process and Controls Team |
| 22-02-17 | VWC TOC | Obtain long-distance TOC view of metrics | Arup, 8 Fitzroy St. | VWC Head of Performance |
| 23-02-17 | RDG, ORR, Network Rail | Information on and understanding of use of passenger ticket sales data in metrics | Arup, 8 Fitzroy St. | RDG representatives, ORR Business Intelligence Manager, Network Rail Performance Analysis Manager |
| 24-02-17 | c2c TOC | Obtain south-eastern commuter TOC view of metrics | By telephone | Network Rail Programme Manager, Anglia Route; |

| Issue | 18 July 2017

Page 9

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

| Date | Organisation | Purpose | Location | Attendees |
|------|-------------|---------|----------|-----------|
| | | | | c2c Performance Compliance Manager |
| 27-02-17 | Network Rail – Wales Route | Obtain Network Rail Route view of new performance metrics and their implications | Network Rail, Cardiff | Network Rail Route Performance Manager, Wales |
| 27-02-17 | Network Rail – LNE Route | Obtain Network Rail Route view of new performance metrics and their implications | Network Rail, York | Network Rail Route Performance Manager, LNE |
| 28-02-17 | RDG Comms | Obtain RDG Comms team's view of communication of metrics within and outside industry | RDG, 200 Aldersgate St. | RDG Head of News; RDG Delivery Manager |
| 01-03-17 | Northern TOC | Obtain regional TOC view of metrics | By telephone | Northern Performance Planning and Regimes Manager |
| 02-03-17 | Southeastern TOC | Obtain south-eastern commuter TOC view of metrics | By telephone | Joint Head of Performance for Network Rail and Southeastern |
| 15-03-17 | Network Rail and ORR | Presentation and discussion of emerging findings | Arup, 8 Fitzroy St. | Client group |

Following the initial meetings with Network Rail, RDG, ORR and individual TOCs and the collection of data, spreadsheet tools and process documentation (where available), a review was undertaken of the accuracy and reliability of the data and metric calculation processes used in the preparation of the metrics.

The emerging findings were presented to ORR and Network Rail at the tripartite meeting held on 15th March 2017. The feedback obtained at the meeting informed the preparation of this final report and accompanying Best Practice Guide, including current and potential confidence grades and recommendations for the development of the metrics.

| Issue | 18 July 2017

Page 10

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

# 3 Background: Objectives, Development and Definitions of the Metrics

To provide some context, within this chapter we summarise the objectives and development of the five metrics under review, and set out their definitions provided to us with the mandate.

## 3.1 Metric Objectives and Development

The original objectives and the development of the metrics are described in a series of NTF working papers and other documents, provided by Network Rail for the purposes of this review and listed below in Table 3.1.

Table 3.1: Background papers to development of performance metrics

| Paper | Date of paper |
|---|---|
| Paper to the National Task Force, Title: What we measure – train service delivery, reference NTFP 151103 Paper D | 19 February 2014 |
| Paper to the National Task Force, Workstream: What we measure? - Measuring Train Service Delivery | 17 February 2015 |
| Paper to the National Task Force, Title: What we measure? - Measuring Train Service Delivery Update, reference NTF 15 0610 Paper X | 03 June 2015 |
| Paper to the National Task Force, Title: What we measure? Passenger Performance Metrics update | 28 July 2015 |
| ORR letter, Title: Performance Measurement: Putting Customers at the Heart of the Industry's Systems | 28 July 2015 |
| Paper to the National Task Force, Title: What we measure? Passenger Performance Metrics update | 30 September 2015 |
| NTF Paper – Performance metrics impact assessment and trial proposal, APPENDIX B, Table 2: Metrics definitions and assessment of their impact | undated |
| Paper to the National Task Force, Title: CP6 Performance Input into the Initial Industry Plan, reference NTFP 161702 Paper B | 10 February 2016 |
| NTF Better Monitoring Workstream, Title: CP6 Passenger Performance Metrics - A Proposal | March 2016 |
| Paper to the National Task Force, Title: What we measure? CP6 Passenger Performance Metrics Proposal, reference NTFP 161603 Paper F | 16 March 2016 |
| Paper to the National Task Force, Title: CP6 performance planning update, reference NTFP 060516 Paper G | 11 May 2016 |
| Paper to the National Task Force, Title: CP6 performance planning update, reference NTFP 160607 Paper G | 30 June 2016 |
| Paper to the National Task Force, Title: CP6 performance planning update, reference NTFP 160308 Paper H (PfN) | 26 July 2016 |
| Paper to the National Task Force, Title: CP6 performance planning update, reference NTFP 211216 Paper D | 14 December 2016 |

Three broad categories for improved, passenger-focussed metrics were identified at the outset of the NTF development process:

- A measure of reliability with a focus on 'Bad Days';

- A Passenger Lateness measure (where every minute and every passenger counts); and

- A measure of on-time punctuality (right time) throughout a train's route.

NTF considered early on the monitoring of connection provision and maintenance, but this was subsequently dropped (monitoring and incentivising on-time performance at intermediate stations should assist with the maintenance of connections, however). Four metric categories were proposed, which can be seen as a significant part of the evolution towards the metrics covered by this review:

- Passenger Experience – Average Passenger Lateness (APL) measure

- Reliability and severe disruption ('bad day') measures – Cancellations, Time to 15 and Cancellations by critical route

- On time measures – All recorded station right time

- Passenger Weighted Route Delay Minutes

The originally-proposed APL metric was subsequently replaced by Total Passenger Lateness (TPL) (although as already noted this has recently reverted to APL). It was noted during the course of the metric development that improvements to the reliability and on-time metrics should also result in an improvement to the 'headline' APL metric.

It was agreed that cross-industry support would be crucial to the successful achievement of full alignment of objectives and targets in order for the metrics and underlying data to:

- Encourage the right behaviours from all parties;

- Be defendable in public;

- Disincentivise perverse behaviours;

- Be fair to all industry partners;

- Allow funders to accurately state what they want; and

- Be easily [and] economically collected.

The development of a 'Metro' measure of actual vs. expected waiting and journey times was proposed for use on high-frequency suburban routes, but was excluded from the remit of this review.

The focus on right-time running throughout trains' journeys moves away from an attitude of 'the train can make up time' and is supported by Transport Focus and DfT, and should help to reduce distrust of performance measures. It was also observed by NTF that the new metrics should improve the focus upon the service received by the large numbers of users of commuter TOCs, putting them on a

| Issue | 18 July 2017

Page 12

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

more equal footing with users of long-distance operators when regulating services during times of disruption.

In September 2015 NTF members endorsed a proposal to carry out an initial trial of six performance measures for five periods. The main aims of the trial were to confirm the metrics reflected performance in a meaningful way, to lock down their definitions and to determine if targets could be set against them. Industry feedback on the outcome of the trial was generally positive. However, there were some concerns expressed about potential loss of capacity as a result of the focus upon on-time running through trains' journeys, the effects of Working Timetable / Public Timetable differentials, the accuracy of TRUST measurements due to rounding making trains appear late that are on time, proposals for the use of on-train GPS instead of expansion of Monitoring Pont network, and possible incentives for early running because of its inclusion as on-time in the proposed metrics. The need for a 'Metro' measure was endorsed, together with the alignment of delay-repay with the T-15 metric.

The current definitions of the metrics follow. They have been taken from the mandate for this review[1].

## 3.2     Total Passenger Lateness - Passenger Time Lost (hours)

### 3.2.1     Definition and Calculation

This metric is described as

*The total lateness experienced by our passengers.*

The Total Passenger Lateness (TPL) for a Train Operating Company's (TOC's) Service Group is the Average Minutes Lateness (AML) multiplied by the number of passenger journeys, and expressed in terms of millions of hours, i.e.

$$TPL = AML*Passenger\ Journeys\ /\ (60*1,000,000)$$

TPL for a TOC is then the sum of TPL for all of the TOC's Service Groups.

Base AML values are calculated for arrival times at Contractual Monitoring Points (CMPs) which are stations where the lateness of trains are monitored for the Schedule 8 performance regime[2]. Typically there are around four to five such monitoring points for each service code (although for short-distance services or very long-distance ones fewer or more would be appropriate). The calculation of AML for each CMP is

$$AML = (sum\ of\ lateness + sum\ of\ cancellations*cancellation\ weight)\ /\ no.\ of\ planned\ arrivals$$

Early-running trains are recorded as being on time, i.e. zero minutes late.

---

[1] Mandate L3 AR004, Appendix 3 – New Performance Metric Definitions
[2] This regime compensates train operators for unplanned service disruption caused by Network Rail and other train operators.

| Issue | 18 July 2017

Page 13

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

In accordance with the calculations within Schedule 8, the cancellation weight is based on 1.5 times the average service interval.

Daily Service Group AML = weighted sum of AML by Monitoring Point

where the Monitoring Point weightings reflect the proportion of passengers in the Service Group alighting at the stations represented by Monitoring Points.

Periodic Service Group AML = weighted sum of daily AML

where the daily weightings within a 4-week period reflect the numbers of trains calling at the Monitoring Points.

Service Group definitions, Monitoring Points, Monitoring Point weightings and cancellation weightings are set out in Schedule 8 of TOCs' Track Access Agreements. Annual passenger journeys per Service Group, based on ticket sales and other revenue data, are provided once a year by ORR and updated at the start of each financial year. Passenger numbers are assumed to be spread evenly across the year, i.e.

Periodic passenger numbers = annual numbers / 13.

However, there is an aspiration in the industry to more accurately reflect periodic passenger number variations, for which the use of the busyness factor profile from the AML calculation has been proposed as an initial proxy. The busyness factor for each service group is a measure of the number of its trains stopping at Monitoring Points within the timetable during the 4-week period.

TPL was seen as the core measure for the public reporting of industry performance, with APL likely to be the metric used within the industry to set targets and compare performance between TOCs. The PEARS AML calculations are to be shared with all interested parties, to enable them to be incorporated in industry systems like BUGLE and ITED.

In the absence of any formal definition seen during our review[3], a possible description of the APL metric is

*The average lateness experienced by our passengers*

At Service Group Level, APL would be the same as AML, while APL at TOC (and National) levels would be:

APL for a TOC = $\sum$ (AML*Passenger Journeys) for each relevant service group

/ TOC passenger journeys

---

[3] We understand there is a definition of Average Passenger Lateness within Schedule 8 documentation but this was not provided to us for this review

| Issue | 18 July 2017

Page 14

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

### 3.2.2     Objectives and Development

With regard to the wider objectives of the metric, this is both

*   A Passenger Lateness measure (where every minute and every passenger counts)

and (to some extent)

*   A measure of on-time (right time) throughout a train's route

As noted in the course of the development of the metrics, TPL provides a genuine reflection of passenger experience, aligns with Schedule 8 regime and is the only measure combining the effects of lateness and cancellations.

## 3.3     Reliability – Cancellations and Severe Disruption

### 3.3.1     Definitions and Calculations

The Reliability - Cancellations metric is described as

> *The percentage of planned trains which either did not run their full planned journey or did not call at all their planned station stops.*

A train is fully cancelled (cancellation weighting = 1) if it runs less than 50% of its planned journey length (i.e. distance). It is classed as part cancelled (cancellation weighting = 0.5) if it runs between 50% (inclusive) and 100% (exclusive) of its journey length or if it completes its full journey length but fails to stop at one or more of the stations at which it is planned to call.

The metric is calculated as the sum of the cancellation weightings into a cancellation score, then expressed as a percentage of the total number of planned trains.

Cancellation % = Cancellation Score / planned train count

The Reliability – Severe Disruption metric is described as

> *The number of days when the service was severely disrupted.*

This is to be measured at sub-operator group and at national levels only. Each TOC has a number of sub-operator groups which correspond to its different market groups and vary in number from one (for example, c2c) to six (Abellio Greater Anglia). Severe disruption is defined as the cancellation of at least 20% of services at sub-group level, and of at least 5% of services at national level (cancellations are calculated in the same manner as for the Reliability – Cancellations metric).

Days when a sub-operator group is planned to run fewer than 20 trains are excluded, with the exception of Caledonian Sleepers (and other operators running small numbers of trains) which will be assessed separately from the larger operators (no details of the planned assessment regime were provided).

### 3.3.2 Objectives and Development

The Reliability - Cancellations metric provides

- A measure of Reliability and a focus on 'Bad Days'

It provides a 'pure reliability' metric (i.e. it excludes significant lateness, in contrast to CaSL) that can be used for both Network Rail and TOCs on a day-to-day basis. It is easy for staff and the public to understand, and provides a true measure of service delivery, while the attribution of cancellations makes it easy to identify the causes of failure. A train running 50% or more, but less than 100%, of its journey, or completing its journey but failing to call at one or more of its scheduled stops, forms a part cancellation, which is counted as half a cancellation in the metric calculation, and aligns with DfT's Schedule 7.1 cancellation measure. In one of the papers to NTF, it is noted that "*DfT is at present carrying out work to see if they can combine the cancellation and short formation aspects of Schedule 7.1 and the Task and Finish Group has asked to be kept informed of this work.*"[4]

## 3.4 On Time at All Recorded Stations – On Time and Time to 15

### 3.4.1 Definitions and Calculations

The On Time at All Recorded Stations – On Time metric is described as

> *The percentage of recorded station stops called at on time.*

A train is considered as being on time at a station if the actual time is less than one minute later than (i.e. within 59 seconds of) the time shown in the public timetable (GBTT). Trains running early are considered as being on time. The metric is calculated as the number of recorded station stops made less than one minute late as a percentage of the total number of recorded station stops.

A recorded station stop is defined as having both a planned GBTT date time and an actual date time recorded in TRUST. Cancelled services are excluded from the measure, since they are captured in the Reliability metrics (see above). Pick-up only and set-down only stops are included in the measure, since they have both planned and recorded actual times; request stops are included only when they are made.

Train lateness is recorded on departure from the origin, and on arrival at intermediate and terminal stations; not all stations are included in TRUST as recording points, but it is anticipated that more stations will be included over time.

Using the same approach, the On Time at All Recorded Stations – Time to 15 metric is described as

> *The percentage of recorded station stops called at within 15 minutes of the planned time.*

---

[4] CP6 performance planning update, reference NTFP 160607 Paper G, 30 June 2016

| Issue | 18 July 2017

Page 16

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

A train is considered as achieving the measure at a station if the actual time is less than 15 minutes later than (i.e. within 14 minutes and 59 seconds of) the time shown in the GBTT. Trains running early are considered as being within 15 minutes. The metric is calculated as the number of recorded at station stops made less than 15 minutes late as a percentage of the total number of recorded station stops.

## 3.4.2 Objectives and Development

The On Time at All Recorded Stations – On Time metric provides

- A measure of train lateness (where every minute counts)

and

- A measure of on-time (right time) throughout a train's route

Every minute of lateness counts for this metric (and improvements will feed into TPL). It can be disaggregated and presented by any grouping of stations, Service Groups, TOCs or Routes, and by using any point of the lateness distribution and any time period. It reflects how well the timetable is being delivered, and is easy for staff and public to understand. Items for consideration include TRUST coverage, data quality, and the differentials between public and working timetables.

The On Time at All Recorded Stations – Time to 15 metric additionally provides

- A measure of Reliability and a focus on 'Bad Days'

| Issue | 18 July 2017

Page 17

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

# 4    Review of Metrics

This chapter focuses on the first two requirements of the mandate, namely:

- Review and comment on the processes and procedures by which the industry captures other data used in these metrics; and

- Review and comment on the reliability, quality, consistency, completeness and accuracy of reported data in the trial.

Regarding the first requirement, and as agreed at the inception meeting with Network Rail and ORR, our review does not extend to checking the source data itself, rather it acknowledges known issues with the accuracy of the source data and identifies their implications on the metrics.

## 4.1    Total Passenger Lateness - Passenger Time Lost (hours)

As noted above, in the course of the review, it was decided by the NTF sub-group to report Average instead of Total Passenger Lateness, but no formal definitions, updated tools or output data were available for APL, and our findings are based upon the review undertaken of TPL.

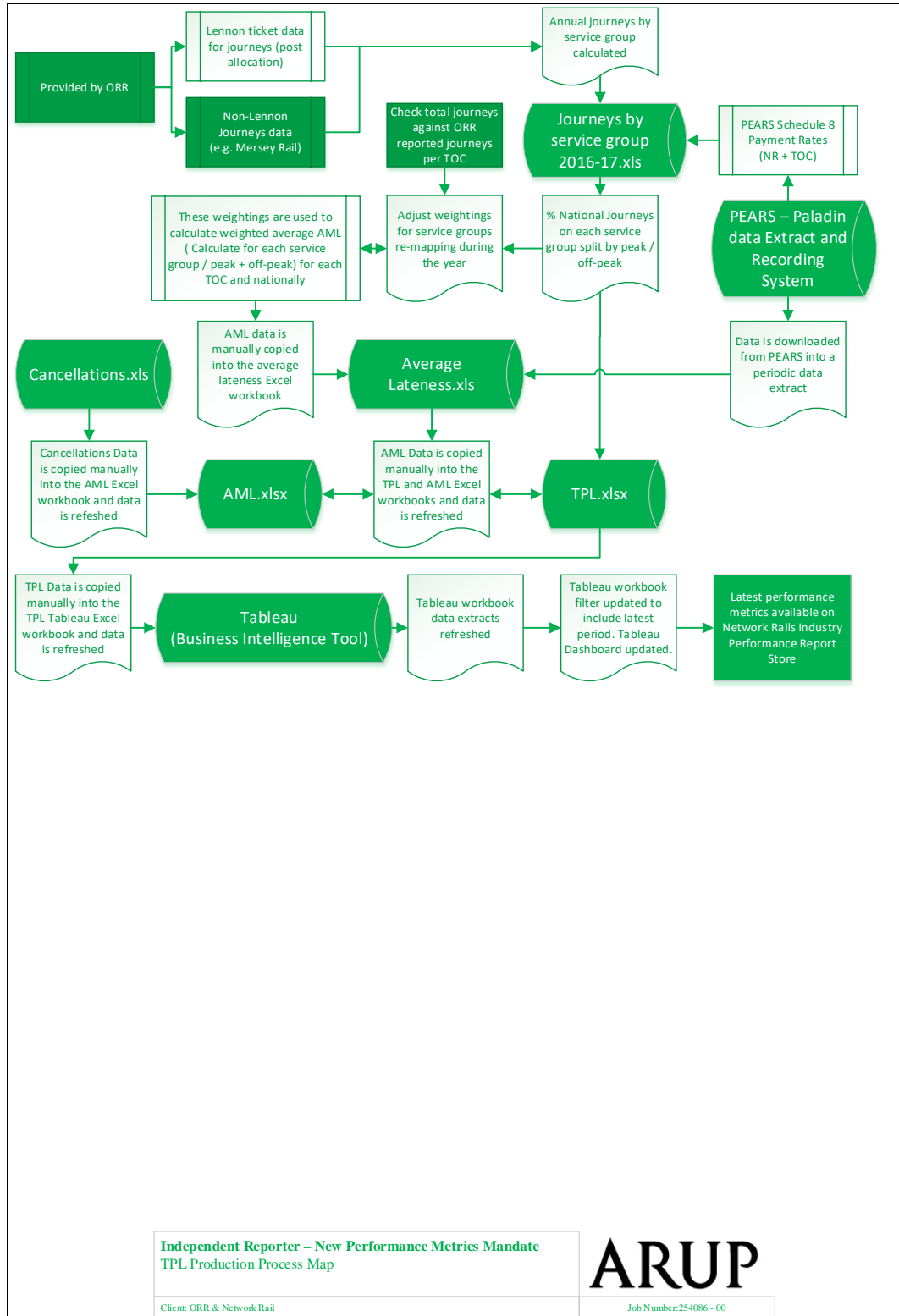### 4.1.1    Data Sources, Collection and Collation

The data sources and the collation and calculation processes used for the calculation of TPL are summarised in the Process Map shown in Figure 4.1 (we anticipate that the data sources and collation process will remain largely unchanged for APL, and that only relatively minor changes will be made to the calculation processes).

In contrast to the other metrics TPL uses two sets of input data, none of which is derived from Business Objects – PSS (BO–PSS): Average Minutes Lateness (AML) data, derived from TRUST via PALADIN and PEARS, and passenger ticket sales data provided by ORR.

The spreadsheet-based calculation processes used to produce the TPL metric are described in an internal Network Rail document ("*Process note for cp6 reporting to ntf*", undated) that was provided for the purposes of the review.

The AML and passenger ticket sales data inputs are both relatively 'black box' in nature, in that the data sources, calculations and processes are not clearly described for or understood by many of the TPL metric producers and users.

| Issue | 18 July 2017

Page 18

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Figure 4.1: Process Map for the 'Total Passenger Lateness' Metric

## AML Data

The AML data is calculated as part of the process for determining Schedule 8 compensation payments between Network Rail and the TOCs for delays to trains. The calculation is carried out in PEARS which is based on Access 2000 software that is no longer supported, and therefore cannot easily be updated and is due for replacement. Network Rail has an aspiration to implement the AML calculations in BO-PSS as part of the metric calculation process, and possibly as a replacement for PEARS.

One aspect of the calculation of AML is the monitoring point weights reflecting the proportion of passengers alighting at each station. We understand that the need to update monitoring point weights is considered at the start of each five-year control period in a recalibration exercise, since it is viewed as important that the Schedule 8 regime gives the correct connection to the drivers of TOC revenue. It is normally left to the TOC to decide if the monitoring points and their weights need to be adjusted.

According to Network Rail, the CP5 national recalibration did not review monitoring point weights unless by exception. This suggests that the last major review of monitoring point weights was for CP4 in 2009 and raises concerns about their currency for use in the calculation of TPL.

Using the Schedule 8 process to calculate AML for the TPL metric has the advantage of being consistent with an established performance-related contractual regime which is accepted by both Network Rail and TOCs. As such, the calculation of AML is seen as accurate enough to compensate for lost income from passengers. However, there would appear to be possible conflicting priorities whereby for Schedule 8 TOCs are unlikely to want the risk of fluctuating compensation payments as a result of regular updates to monitoring point weights. This conservatism is likely to make AML less accurate as a true measure of passenger lateness.

We recommend that all monitoring point weights should be recalibrated for the start of CP6 and reviewed to see how much they have varied since the last recalibration. A view can then be taken on their likely impact on the accuracy of TPL and whether a process is required for more regular reviews for this metric.

## Passenger Ticket Sales Data from ORR

The number of passengers is calculated from data provided by the ORR in a file that shows the annual numbers by service code. This file is derived from the LENNON (Latest Earnings Network Nationally Over Night) ticketing and revenue database. It is recognised that LENNON is primarily a ticket sales database and that the number of journeys is an estimate based on a number of assumptions. Some of these assumptions described to the Reporter team in a meeting with RDG, ORR and Network Rail representatives held on 23rd February 2017 include:

- Journeys on season tickets are spread evenly across the period / year according to the assumed number of journeys made; and

| Issue | 18 July 2017

Page 20

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

- Anytime returns – return journeys are assumed to be on the day of sale, but in practice could be a later date.

ORR supplements these journeys with those from TOC products that are not captured by LENNON; for example, Stagecoach offer tickets through their Megatrain website. ORR contacts the TOCs every quarter to provide the volume of passenger kilometres and passenger journeys (every six months for a few products). Nationally, these additional journeys only represent about 2% of all journeys, though are more significant for some service groups (for example, TfL Rail and Merseyrail where they can account for up to approximately 50% of journeys made).

The combined file is then handed over to Network Rail for the process to calculate TPL. The process for producing the file is described in a document entitled "Passenger journeys by train operator" provided to us by ORR.

The file taken from LENNON is the post-allocated file which apportions the ticket revenue and journeys between the different service codes that could be used by passengers (from the ORCATS[5] allocations). In this process, journeys that require a change of trains are shown separately on each journey leg. So, for example, a journey from A to C with an interchange at B would be shown as two journey legs, one on A to B and one on B to C.

This is an important distinction as it effectively double counts passengers making a connection from one train to another on their journey. The station usage data produced by the ORR estimates that for the 1,463 million passengers in 2015/16 (i.e. station entries and exits), there were 222 million interchanges[6]. This would suggest that the 'double counting' of passengers in the file produced for the TPL is about 15%.

This is a concern for TPL and results in the national lost hours figure potentially being over-stated by 15%. It is much less of a concern for calculating APL where the journeys are only used to calculate relative weights between service groups.

One area of uncertainty during the review was whether interchanges between trains of the same service code are double counted or counted as a single journey. RDG and Worldline (who operate the LENNON system), have confirmed that ORCATS does not record two journey legs within the same service code. There is therefore no double counting within a service code.

## Manipulation of Passenger Journeys Data for TPL

Network Rail processes the ORR journeys data to produce relative weights to apply to the AML figures produced by PEARS. AML is calculated for the Schedule 8 payment regime for each service group and peak type. The journey weights therefore have to be at the same level of disaggregation.

Network Rail provided us with the spreadsheet that produces the relative weightings for matching with AML (v006 - Journeys by service group 2016-

---

[5] Operational Research Computerised Allocation of Tickets to Services
[6] Estimates-of-station-usage-2015-16.xlsx, National Data Trends Data Portal

| Issue | 18 July 2017

Page 21

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

17.xls). The spreadsheet aggregates the journeys from service code to service group using a lookup list. Some manual changes are made and highlighted, which we understand have been agreed with the ORR and result in an additional 2% of journeys for those not captured in the file provided by ORR. The service group journeys are then split by peak / off peak / all day to match the AML figures, using the corresponding payments rates within PEARS as a proxy. We have not reviewed in detail the derivation of these payment rates, their assumptions or the level of approximation made (and there is no documentation to explain this with the spreadsheet). The underlying assumption is that they are accurate enough for the Schedule 8 payment regime, however we would recommend carrying out some sensitivity tests to see how changing them impacts TPL and APL to determine whether it is worth a more detailed review of their accuracy.

Relative weightings are then produced for each service group and peak type, summing to 1.0 nationally. Adjustments are made where there has been re-mapping of service groups with franchise changes during the year and / or where new train services have been introduced.

The spreadsheet contains some documentation in the sheets, but it would help to have a description of the overview of the calculation included within a user guide for calculating TPL which sets out how to make and agree the manual adjustments to the journeys data and weightings, a Record of Assumptions document and for the spreadsheet to follow best practice guidelines (see Appendix B).

### 4.1.2 Reported Data: Reliability, Quality, Consistency, Completeness and Accuracy

The TPL.xlsx spreadsheet which calculates the metric and contains outputs from 2014-15 Period 01 to 2016-17 Period 10 was provided for review in the 1st February 2017. The Excel workbook comprises five worksheets:

- The first worksheet, 'AML', is populated by pasting data for each TOC's AML periodically and as a Moving Annual Average (MAA) from the Average Lateness data file.

- The second worksheet, 'Journeys', includes pasted Journeys data by TOC for each quarter provided by ORR in the 'v006 - Journeys by service group 2016-17.xls' spreadsheet (or annual equivalent), but is also linked to the spreadsheet for data for the current year (no reference is made to the spreadsheet or link in the process note provided). The calculations within the worksheet use only simple sum and average formulae to calculate periodic averages.

- The third sheet 'TPL(MILLION HOURS)' is a sheet for calculating the TPL and MAA AML for each TOC and national region by period.

- 'Lookup' is a reference sheet for each TOCs average estimated journey time and all of the periods used in the workbook.

- The final sheet 'OUTPUT' is an output of TPL, average lateness and lateness as a percentage of average journey time for each TOC and nationally. The sheet is driven by a drop down list containing every period which updates the performance metrics displayed. The calculations use 'VLOOKUP' and

| Issue | 18 July 2017

Page 22

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

'MATCH' functions which were checked and found to be correct: the formulae used are constructed correctly and display correct values.

The spreadsheet would be improved by following best practice guidelines in Appendix B. Specific suggestions include a front cover sheet explaining where all information and data is sourced from (including the Journeys spreadsheet link) and who manages the workbook. It would also be helpful to distinguish pasted data cells from formulated calculations cells by the use of different colour fills.

The process of the calculation of the TPL metric is a relatively straightforward spreadsheet-based one but with some copying and pasting of data which requires care and attention. Introducing more linking to input spreadsheet would provide a better audit trail. Longer term there is the potential for full implementation of the process in BO – PSS (should the benefits be considered worthwhile) before the new metrics go 'live'. A clear Record of Assumptions is required with improved documentation.

As indicated above, the reliability of the processes used to generate the underlying AML and passenger journeys data is more questionable, as they are not comprehensively documented, transparent or well understood within the Performance community. There is scope to improve the transparency of the AML calculations by implementing them in BO – PSS.

The metric outputs for the latest 13 Periods were checked and found to be consistent, complete and correct relative to the input datasets, indicating that the accuracy of the calculation of the metric is high. Equivalent checks were not undertaken for the AML and passenger journeys data; since the AML calculations form the basis of the Schedule 8 financial regime, they are already accepted by industry, but improved documentation and transparency of the processes used would provide further reassurance in this regard.

## 4.2 Reliability – Cancellations and Severe Disruption

These metrics are train- rather than passenger-focussed, in contrast to the TPL metric.

### 4.2.1 Data Sources, Collection and Collation

The data source and the collation and calculation processes used for the calculation of the reliability metrics are summarised in the Process Map shown in Figure 4.2. In contrast to TPL, the reliability metrics are based solely on TRUST data, which relies on input (e.g. 'fail-to-stop' incidents) from and verification by Operators, prior to processing in BO – PSS.

As for the TPL metric, the spreadsheet-based calculation processes used to produce the metrics are described in the internal Network Rail document that was provided for the purposes of the review. The documentation for the BO – PSS querying process is not yet fully developed, but screenshots of the querying processes used to extract the data were provided to enable the generation of the Process Map and the replication of the calculations used as part of the checking process.

| Issue | 18 July 2017

Page 23

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Figure 4.2: Process Map for the 'Reliability – Cancellations and Severe Disruption' Metrics



Independent Reporter – New Performance Metrics Mandate
Cancellations Production Process Map

ARUP

Client: ORR & Network Rail

Job Number: 254086 - 00
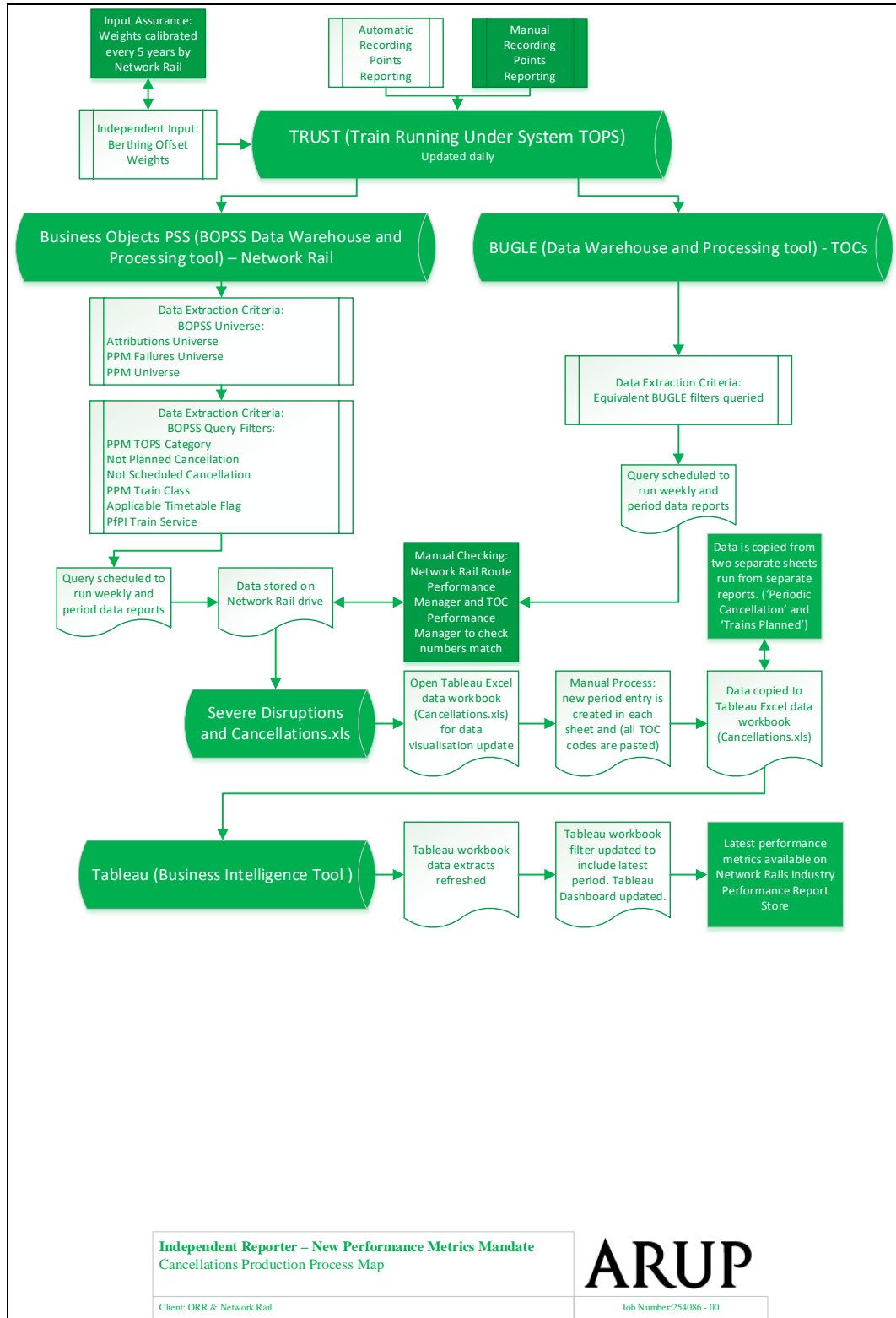
### 4.2.2 Reported Data: Reliability, Quality, Consistency, Completeness and Accuracy

The severe disruption and cancellations calculation spreadsheet (Severe Disruption and Cancellations.xlsx) was provided by Network Rail on the 1st February 2017 for review and evaluation. The Excel workbook consists of five worksheets:

- The first worksheet, 'Charts', is a data visualisation sheet which shows trains planned, cancellations, percentage cancelled per period, MAA of percentage cancelled per period, severely disrupted days, cancellation splits by cause, and severely disrupted day splits by cause. The sheet is driven by a drop-down list containing every period which updates the performance metrics displayed. The formulae used are all lookups, so no calculations are undertaken within the worksheet. The first three rows of the worksheet use different lookup references to the list of cells below; although everything was found to have been calculated correctly, there is scope for incorrect 'copying down' of formulae here, and the use of judicious colour-coding would help to distinguish between the different formulae and reduce the potential for error.

- The following three sheets, 'TOC Data', 'SubToc Detail' and 'Aggregated Data' are all data dumps from BO – PSS.

- The final worksheet, 'Aggregated Data', contains a summation of the 'TOC Data' outputs, aggregating the metrics at the national level. It uses 'SUMIFS' 'SUMIF' and 'IF' functions which were checked found to be used consistently and correctly, producing accurate outputs.

Again, the spreadsheet could be further improved by the provision of a front cover sheet explaining where all information and data is sourced from, and who manages the workbook. Cells containing pasted data should be distinguished from cells containing formulae by means of the use of different colour fills. See Appendix B for more guidelines on best practice.

In addition to the review of the metric calculation process, the data manipulations undertaken in BO – PSS were checked for the latest 13 Periods (1612 to 1711 inclusive) by using screenshots of the BO – PSS querying process and source data provided by Network Rail to undertake the equivalent calculations in Excel and compare the results with the BO – PSS outputs. For the items checked it confirmed that the calculations are being undertaken correctly.

The reliability of the reporting process for the output metric is reasonable, being a reasonably straightforward spreadsheet-based process. It does involve some copying and pasting of data and has scope for further automation, but could be improved by further application of spreadsheet best practice. The documentation could be improved with a clear Record of Assumptions.

The metric outputs for the latest 13 Periods were checked and found to be consistent, complete and correct relative to the input datasets, indicating that the accuracy of the metric calculations is high.

| Issue | 18 July 2017

Page 25

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

## 4.3 On Time at All Recorded Stations – On Time and Time to 15

These metrics report punctuality at all reported stations, as part of a 'punctuality spectrum' ranging from on-time to 30 minutes late.

### 4.3.1 Data Sources, Collection and Collation

The data source and the collation and calculation processes used for the calculation of the punctuality metrics are summarised in the Process Map shown in Figure 4.3. The punctuality metrics are again based solely on TRUST data and, in contrast to the reliability metrics, do not require Operator input.

Figure 4.3: Process Map for the 'On Time at All Recorded Stations – On Time and Time to 15' Metrics
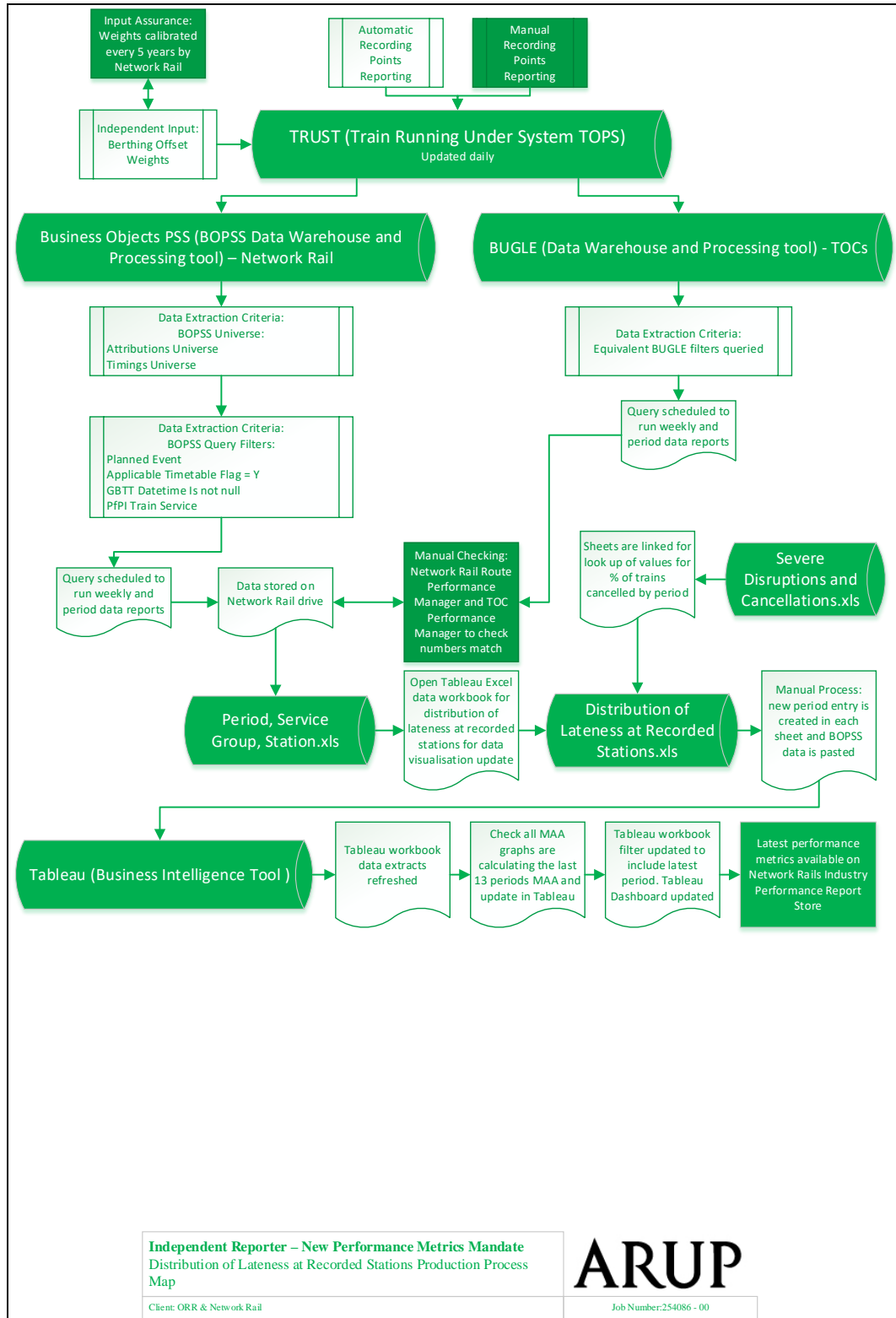


**Independent Reporter – New Performance Metrics Mandate**
Distribution of Lateness at Recorded Stations Production Process Map

ARUP

Client: ORR & Network Rail

Job Number:254086 - 00

## TRUST Data

Train arrival times at stations (and at other timing points) are recorded in the TRUST (Train Running System on TOPS) system. Of the 2,618 stations on the national network, 876 are Contractual Monitoring Points (CMPs), where lateness is monitored for Schedule 8, or Delay Reporting Points (DRPs); the remaining 1,742 are Station Timing Points (STPs).

The TRUST system captures timing information for each train over the course of its journey. This automated input occurs when a train activates a signalling detection system (for example track-circuit or axle counter) as it passes throughout the network. When this occurs the information is linked to the identification of the train (from the train describer) and is then passed into the SMART system. This aggregates the data from all such connected train describer systems and passes it into downstream information systems. Where the details correspond to a TRUST reporting point, SMART will flag the activity accordingly and the event will be processed by TRUST.

It is important to note that SMART records accurate strike-in times to the second. TRUST, however, only records whole minutes, truncating the seconds' element of the recorded time. Thus, a strike-in which aggregates to two minutes and fifty seconds will be recorded in TRUST as two minutes, as will an aggregate of two minutes and two seconds. This explains the rationale behind the On Time definition – a scheduled arrival time of 14:22 is actually achieved if the strike-in (plus the berth offset time – see below) aggregates to no more than 14.22:59. In this case it will be recorded in TRUST as 14:22.

There are a number of locations in manually signalled areas where automatic reporting equipment is not available. At these locations reporting information is still required and thus an alternative reporting method must be employed. These are termed manual reporting locations. Generally they rely on the signaller inputting the arrival time of the train directly into TRUST. In some locations, where the signaller may be unsighted, traincrew or other Network Rail staff will advise the signaller of the train's arrival time.

Train arrival times at the majority of CMPs and DRPs, and at over half of STPs are reported to TRUST automatically by means of Automatic Train Reporting (ATR). Arrival times at the remaining CMPs and DRPs, and at 50 STPs are recorded manually. Of the remaining STPs, 166 are partially reported by means of GPS (i.e. they are reported when GPS-equipped trains are in use), and 608 are not currently reported.

In order for a station to be fully compliant with the reporting requirements for the new performance metrics, Network Rail's view is that it should be capable of reporting the times of all train arrivals in every direction. By this standard, 871 (i.e. all but five) CMPs and DRPs have the potential to achieve compliance, as do 1,074 STPs. The situation is summarised in Table 4.1.
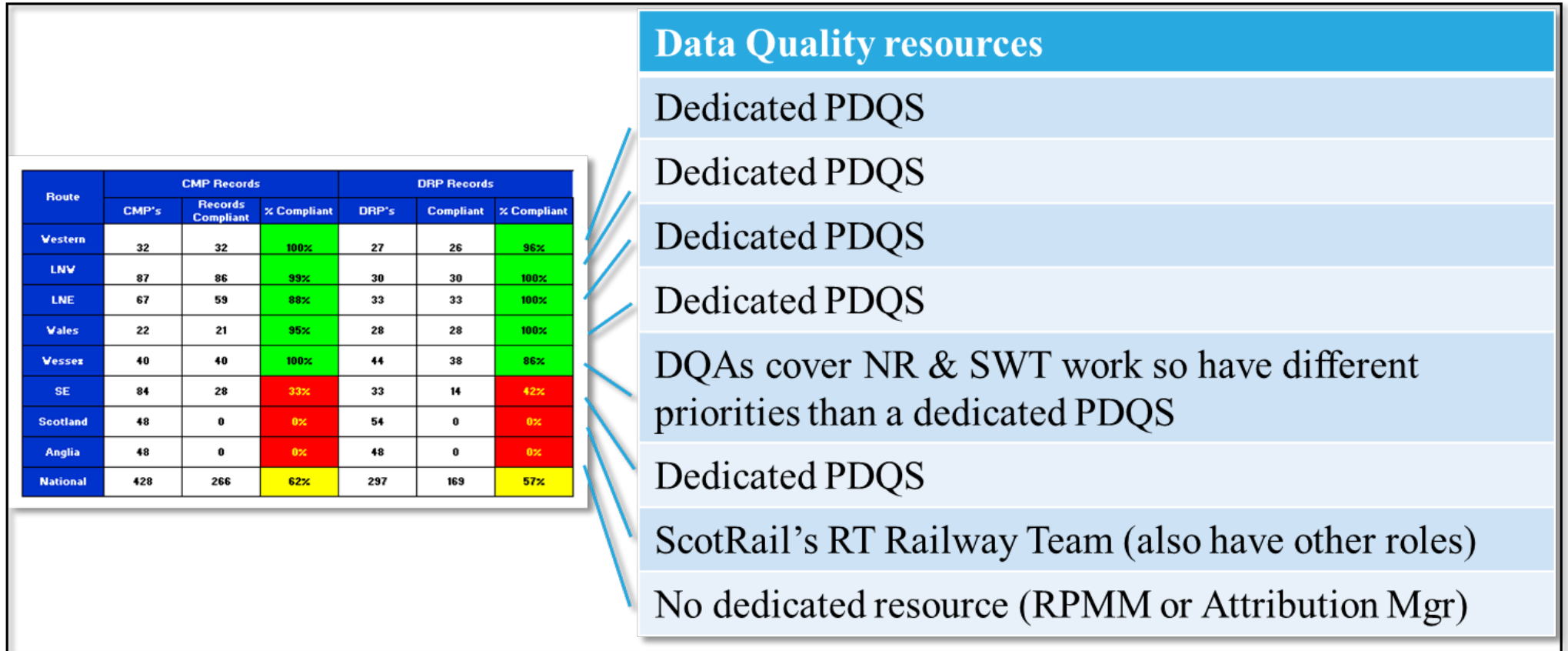
| Issue | 18 July 2017

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Page 28

Table 4.1: Station Reporting

| Station Reporting Types | Station Counts | Reporting Method | | | | Potentially Fully compliant |
|---|---|---|---|---|---|---|
| | | ATR | Manual | GPS (Partial) | None | |
| CMPs/DRPs | 876 | 694 | 192 | 0 | 0 | 871 (99%) |
| STPs | 1,742 | 918 | 50 | 166 | 608 | 1074 (62%) |
| Total | 2,618 | 2,618 | 242 | 166 | 608 | 1,945 (74%) |

Of the 668 STPs that are not potentially fully compliant, Network Rail advise that approximately 430 have the potential for part-compliance to detect some train arrivals, with the remainder currently having no potential for compliance due to the restrictions of the signalling.

ATR is not simply the time recorded at the final signal berth prior to stopping at a station platform. An allowance of time, the berth offset, needs to be added for the time taken from passing the final signal berth to the wheels of the train stopping at the platform. The berth offset must be calculated for each permutation of approach route and destination platform that is used by services on a planned basis. This is derived by undertaking a site survey of actual train times covering various combinations of rolling stock and calculating an average value. A record of every offset for every route is maintained in a Route Margin Book.

The offsets need to be regularly monitored and kept up to date to reflect operational and other changes. This task has been made more difficult in recent years as a result of the loss and non-replacement on some Routes of the Data Quality Specialist staff. This can be seen in Figure 4.4 where the compliance of annually checking if CMPs and DRPs need to be reviewed is compared against the available staffing resources in each Route. There is an apparent correlation between available resources and level of compliance, and raises concerns about the consistency of accuracies across the Routes. Note that failure to carry out this annual check does not necessarily mean that the CMPs or DRPs are out-of-date, just that there is a risk of this following any infrastructure or rolling stock changes.

| Issue | 18 July 2017

Page 29

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Figure 4.4: Compliance of annual desktop review of CMP and DRP currency compared against available Performance Data Quality Specialists (PDQS)



| Route | CMP Records | | | DRP Records | | |
|---|---|---|---|---|---|---|
| | CMP's | Records Compliant | % Compliant | DRP's | Compliant | % Compliant |
| Western | 32 | 32 | 100% | 27 | 26 | 96% |
| LNW | 87 | 86 | 99% | 30 | 30 | 100% |
| LNE | 67 | 59 | 88% | 33 | 33 | 100% |
| Wales | 22 | 21 | 95% | 28 | 28 | 100% |
| Wessex | 40 | 40 | 100% | 44 | 38 | 86% |
| SE | 84 | 28 | 33% | 33 | 14 | 42% |
| Scotland | 48 | 0 | 0% | 54 | 0 | 0% |
| Anglia | 48 | 0 | 0% | 48 | 0 | 0% |
| National | 428 | 266 | 62% | 297 | 169 | 57% |

**Data Quality resources**

Dedicated PDQS

Dedicated PDQS

Dedicated PDQS

Dedicated PDQS

DQAs cover NR & SWT work so have different priorities than a dedicated PDQS

Dedicated PDQS

ScotRail's RT Railway Team (also have other roles)

No dedicated resource (RPMM or Attribution Mgr)

*Source of compliance table: Network Rail*

| Issue | 18 July 2017

Page 30

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

The currency of STPs is, understandably, worse than CMPs and DRPs given there is no current requirement to check their accuracy. They will be needed for the new metrics. Approximately 45% of ATR STPs have not been checked since 2011 or earlier as shown in Figure 4.5 below.

Figure 4.5: Year that the offsets for automatically reported STPs were last checked
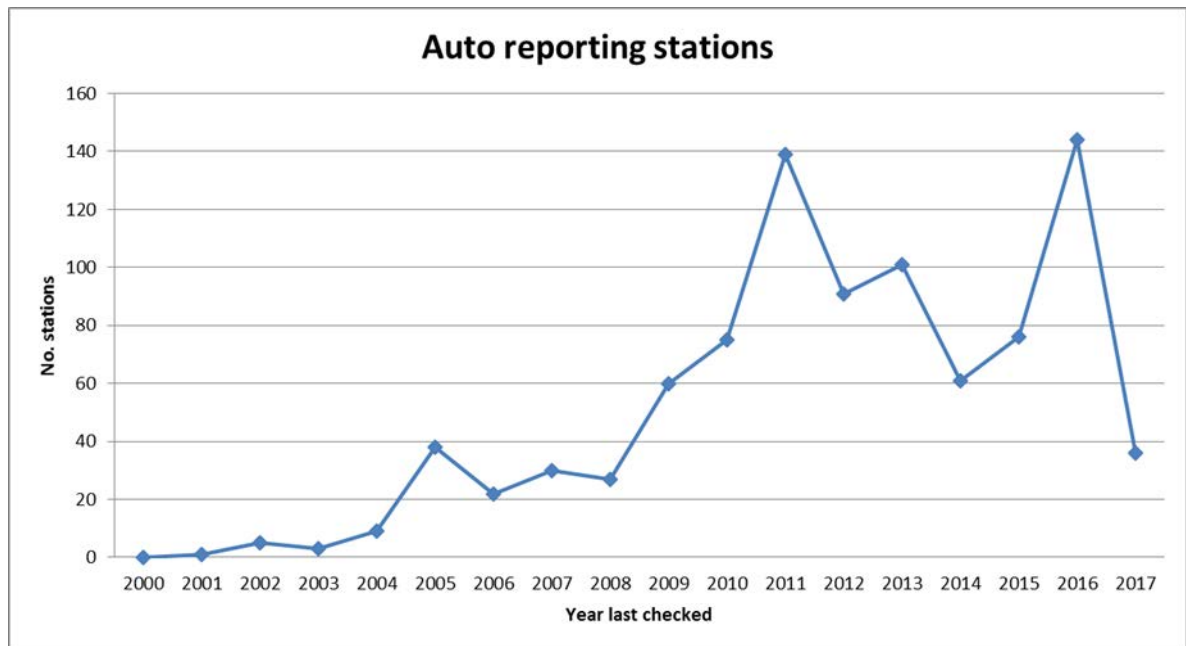


*(Source: Network Rail)*

There are thus significant concerns in relation to (i) the reliability of the underlying TRUST data, where it is recorded manually or only partially, and (ii) the accuracy of ATR data in general, given the 'one size fits all' nature of berth offsets, and, in particular, the currency of berth offsets for those STPs for which monitoring is not up-to-date. The resourcing challenges involved would be exacerbated by the extension to STPs, since they would then also need to be monitored and updated as necessary, increasing the existing workload further.

## Technology solutions to measurement of train arrival times

Given the additional work required for the new metrics to check and update berth offsets, the current lack of available staff resources and the inherent inaccuracies of offsets as a one-size-fits-all, we feel it is worth considering if a technological solution could be found as an alternative.

Network Rail's Train Location Services Programme is looking at the fitment of GPS to rolling stock fleets with the aim of providing more accurate train location data. The vision is to produce a single data warehouse that takes in feeds from GPS, TRUST and other Network Rail and Operator data so that train performance analysis can be undertaken from a single source. Phase 1 of the programme piloted the approach in 2016, and Phase 2, the GPS Gateway, is looking to deliver a national system in the spring of 2018.

Conceptually, GPS data might be able to validate or replace some or all of the berth offsets, although our understanding is that it may not cover all routes and /

or trains. A possible alternative approach for replacing the need for berth offsets might be the use of Radio-frequency identification (RFID) tags, for which there are two potential implementations:

### 1. Trackside

It could be possible to use some form of passive train identification system based on RFID tags. These could be fitted to the side of the cab ends on each train. Readers located below each platform edge would then read the identity of the passing train. The reader could be located at a point on the platform where it could be guaranteed that the train would pass. In some cases multiple readers may be necessary. A local controller with some control logic could manage multiple readers in order to monitor the output of each reader in order to ensure that the passage of trains was not double counted and that reversing trains or newly coupled units were properly handled. The controller would timestamp the train arrival and departure. The schema requires that each train set of IDs is centrally managed so that tag IDs can be correctly associated with a train.

Depending on the tag technology used it may be possible to fit the readers about the platform and so avoid the need for track access. This would greatly reduce the installation costs.

### 2. Trainborne

Readers could be fitted to the trains to read tags located on the sides of station platforms. This would involve a train fitment programme and the need to provide a radio communications link to communicate the reader data back to a central database and system. The train reader would timestamp the passage of Tag IDs. For main stations, the link could make use of existing WiFi connectivity. In stations where WiFi is not available then this would need to be separately provided or possibly skipped (dependent on the granularity of the reporting required). The schema requires that each set of station IDs is centrally managed so that tag IDs can be correctly associated with a station and its platform. As with the trackside solution, some control logic is needed in order to ensure that the passage of the same platform was not double counted and that reversing trains or newly coupled units were properly handled.

Both solutions could probably make of use of Commercial Off The Shelf (COTS) tagging technology although this has not been properly explored. In addition there are alternative technologies which have been used in light rail systems such as mass detectors e.g. Hanning & Kahl equipment. These detect the presence of a train using changes in inductive fields when a large metallic mass is brought into close proximity. Coupling the operation of these with the operation of train describer berth occupancies could be sufficient to provide the necessary timing granularity.

Generally, there are a variety of sensor technologies which could be used including: optical, inductive, near field radio and microwave (radar) technologies. GPS technologies have not been considered here because of the problem with station canopies and building structures.

| Issue | 18 July 2017

Page 32

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Which technology is best will depend on:

- System installation costs

- Sensors/reader technology cost

- Ease of connection to the existing train reporting systems and/or comms infrastructure

- Where the readers are train based or trackside based

A further study would be needed to explore the feasibility of these to determine which offers the lowest CAPEX cost whilst still providing the necessary functionality.

## Stations for measuring train arrival times

However the timings are measured, it will be important to define which stations can provide suitably accurate timings and so be included in the On Time metrics. It will also be important to control the addition or removal of stations from this list.

### 4.3.2    Reported Data: Reliability, Quality, Consistency, Completeness and Accuracy

The spreadsheet used for calculating the distribution of lateness at recorded stations (Distribution of Lateness at Recorded Stations.xlsx) was provided by Network Rail on the 1st February 2016 for checking and evaluation by the Independent Reporter team.

- The first worksheet, 'Data Input', is a data dump from BO – PSS of train timing performance. It uses some simple formulae to calculate the percentage values of trains falling within a range of punctuality bands, and a lookup to display data from the linked Severe Disruptions data spreadsheet.

- The second worksheet, 'Summary National', calculates the sum of timing metrics by performing 'SUMIFS' calculations on the contents of the 'Data Input' worksheet.

- The third worksheet, 'National MAA', calculates the moving annual average of the sum of timing metrics from the 'Summary National' worksheet by performing range sums.

- The last calculation worksheet, 'Operator MAAs', uses 'SUMIFS' formulae to produce equivalent MAA metrics from the 'Data Input' worksheet.

- The final worksheet within the workbook, 'ChartData', is driven by a drop-down list of all Periods which uses a 'SUMIFS' function to calculate a summary of all metrics by TOC. The rest of the worksheets within the workbook are used for displaying charts.

Calculation checks were undertaken and demonstrated that the formulae used are constructed correctly and produce correct values. Again, suggestions for improvement include a front cover sheet explaining where all information and

| Issue | 18 July 2017                                                                 Page 33

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

data is sourced from, and who manages the workbook, the use of colour coding to distinguish pasted data cells from cells containing formulae, and other best practice guidelines in Appendix B.

Network Rail also provided data dumps for the latest 13 periods (from 1612 to 1711 inclusive) for distribution of lateness. Data in these files was processed independently in Excel, based on the BO – PSS querying screenshots provided, to verify the calculations used, and the results were found to be consistent.

The reliability of the reporting process for the output metric is reasonable with an easy-to-follow spreadsheet process. Some copying and pasting of data is involved, providing scope for further automation, although this may add limited value to the process. A documented Record of Assumptions would also be beneficial.

The metric outputs for the latest 13 Periods were checked and found to be consistent, complete and correct relative to the input datasets, indicating that the accuracy of the metric calculations is high.

| Issue | 18 July 2017

Page 34

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

# 5 Potential Impact on Behaviours and Industry Alignment

In addition to providing broad-based, passenger-focussed measures of performance, the new metrics are intended to encourage and facilitate improved behaviours and alignment across the railway industry.

## 5.1 Cross-Industry Implementation Issues

There appears to be widespread acceptance of and enthusiasm for the metrics within Network Rail and across the TOC community, but the industry does face some challenges in terms of their consistent implementation, particularly if they are to be used to compare performance across Routes and between Operators.

Chief among these challenges is the variable coverage and accuracy of TRUST across the network. Although the punctuality metrics are based only upon stations at which arrival times are recorded and reported, reliance upon manual recording will be greater on some parts of the network than on others, and there is also considerable uncertainty about the accuracy of berth offsets at Station Timing Points that are not Contractual Monitoring Points or Delay Reporting Points. This variability and uncertainty will need to be taken into account for the purposes of monitoring and comparing performance and setting targets.

Another issue to address is how train performance will be regulated in CP6. Currently PPM and CaSL are Regulated Outputs with national and TOC targets set by the ORR. From our meetings, we are unclear on whether PPM and CaSL will continue to be regulated in CP6. It has been suggested that those TOCs with targets in their franchise agreements will continue to be regulated against them.

In addition, Network Rail Routes and corresponding TOCs have jointly developed their own performance scorecards. These are based on a common framework but have the flexibility of choosing metrics and targets that are best suited to the markets they serve. For example, Anglia Route's scorecard in future might include Time to 3 for c2c to align to their franchise obligations. Further, the Welsh Government is considering setting targets for passenger lateness in the forthcoming Wales & Borders franchise.

We understand that the ORR is considering how to monitor or regulate the new metrics. We view this as a matter of urgency so that appropriate time series of outputs can be produced and analysed to set reasonable targets, and so that the industry can prioritise implementation of the metrics in time for the start of CP6.

## 5.2 Anticipated Effectiveness in Improving Performance

The TPL (now APL) metric, in common with the others, has cross-industry support, and is acknowledged as providing welcome increased focus on the interests of passengers. There are challenges involved in the regulation of trains to minimise passenger lateness in the absence of perfect information on passenger

| Issue | 18 July 2017

Page 35

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

loadings and numbers, but this appears already to be the *de facto* industry objective, and so does not entail any significant degree of behavioural change, and provides incentives and opportunities for improvement as better train loading information and traffic management tools become available.

By separating cancellations from significant lateness (in contrast to CaSL) the proposed reliability metrics should enable enhanced focus on cancellations, their causes and possible remedies.

The punctuality metrics at all recorded stations should also be good for passengers at those stations, with enhanced focus on punctuality across the network. In the absence of a specific metric for the maintenance of connections between services (as considered during the course of the development of the metrics), it also provides a proxy for this, in that enhanced on-time performance across the network helps to ensure that connections are maintained. This will be of particular importance if the industry should move towards an increased reliance upon the use of connections in preference to the current provision of through services between a wide range of origins and destinations (as in the case of CrossCountry, for example).

## 5.3    Anticipated Effects on Industry Behaviours and Alignment

The intention of the measures is to provide a basket of Key Performance Indicators (KPIs) that across them lead to a balanced approach to measuring performance that drives behaviours to deliver the best deal to passengers.

In our discussions with Network Rail at both Route and HQ level and with TOCs this is clearly seen as an important goal but turning these from KPIs into effective management drivers is still at an early stage. For example, developing revised train regulation policies for issue to signallers and controllers is still an outstanding activity and on Routes with mixed traffic types and operators this is seen as a challenge.

An area for careful consideration is how the measures are translated into daily, weekly and period targets. The industry has developed a very active real time approach to managing PPM (and to a lesser extent CaSL) over the years with the use of real time PPM boards and instant peak reports. These drive Control rooms to be very aware of the service impact of their decisions and drives real time behaviours in managing the service.

Of the new measures, On Time, Time to 15 and Cancellations can all be easily managed in a similar way with decisions tailored to drive these numbers. However, APL/TPL is very different and measuring on a daily basis is of less value. Without careful thought it is therefore possible that the aim of ensuring every minute counts becomes lost by too much emphasis on the On Time measure, i.e. signallers/controllers further delay a late service to deliver another one on time despite the fact this could create more overall delay.

Influencing APL/TPL is a much longer term issue, given the nature of the measure. Individual improvement schemes will only have a small effect on APL

| Issue | 18 July 2017

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

Page 36

and therefore enthusing managers and frontline staff to deliver it should be considered. Use of old targets such as delay minutes will still be very important in converting the measure to a currency which is meaningful when planning and monitoring progress at individual initiative level. Where it becomes more effective is in looking at the reliability of the train plan and the ability to develop ways of measuring the impact on APL of different timetable options should be a valuable tool in future improvements in the quality of the plan.

Related to timetable planning is the impact of the On Time measure on the design of timetables. Often allowances are used, alongside public book differentials to deliver major improvements in PPM delivery at destination. The new measure will reduce the value of this and drive a need for better On Time delivery at all (recorded) stations. This is a positive measure and will start to address the whole journey robustness of train paths. However, there is a risk that this could lead to too much padding in timetables, or extended journey times, which would reduce capacity and not be in the interest of passengers. Similarly, station stops could be shown in public timetables only as the departure time meaning that arrivals are measured against a later time. A transparent code of practice should be considered against which passengers and bodies such as Transport Focus can judge industry compliance.

The use of these measure in setting management and staff objectives will be a key factor in driving behaviours. Joint scorecards will require the full basket to be used to ensure balance but breaking these down to individual manager level will be important.

A comprehensive industry education campaign will be required to ensure staff and passengers fully understand what will be reported and why. In both cases this will need to include an explanation of the reasons and set expectations on the headline numbers. This can be tailored to individual TOCs but should be set in a national context by RDG.

## 5.4    Observations on the metric outputs

The ORR has usefully carried out a trial of On Time, Time to 15 and Cancellations and reviewed them against PPM and CaSL in a paper entitled "CP6 Metrics".

In its review of **On Time**, it shows that as PPM has fallen from the middle of 2015/16, On Time fell at a faster rate. No reason is given in the paper, but it is perhaps not surprising and indicates that On Time is more sensitive to delays than PPM (an issue which will need to be taken into account when setting targets).

On Time at destination was compared with On Time across all timing points (i.e. the new metric). Nationally the new metric in 2015/16 showed On Time as 63.1% whereas the On Time at destination only was slightly higher at 65.3%. However, at TOC level the difference was widely variable. For Arriva Trains Wales, Chilterns and Northern, the new metric was between 10 and 20% lower than On Time at destination. The reason stated was that this reflected the reduced effect of generous timings in the final leg of the journey. This supports the view that in future such TOCs might change their timings (and allowances) in the timetable.

| Issue | 18 July 2017

Page 37

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

At the other end of the spectrum, London based TOCs tended to have higher %s for the new metric reflecting perhaps the delays on approach to congested London termini.

The new metric **Time to 15** at all timing points was also compared with at destination only. For all TOCs the new metric was higher and nationally it was shown to be 98.6% in 2015/16 compared with 95.6% at destination. Further, all TOCs had a high % for the new metric, ranging from 99.8% for c2c and Merseyrail to 94.1% for Virgin Trains East Coast and West Coast[7]. It is our view that this metric will be relatively insensitive to changes in underlying delays and will have limited impact on behaviours in the industry.

**Cancellations** has been compared against the currently reported CaSL. Nationally cancellations were 2.0% in 2015/16 compared against the corresponding CaSL score of 3.1%. All TOCs have a lower cancellation score which range from c2c (0.9%) to GTR (3.8%).

Given that Time to 15 has a high % score and Cancellations has a low % score, it is likely that most public attention will be on the On Time metric. Careful consideration will need to be given to how this is communicated to the public to avoid a negative perception that the railway has gone from performing at close to 90% (PPM) to more like 60% (On Time).

---

[7] Ignoring Caledonian Sleeper at 91.4%

| Issue | 18 July 2017

Page 38

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

# 6 Confidence Grades

he current and potential confidence grades for each of the five metrics are summarised in Table 6.1, based upon the review and analysis described above. We also comment on the effort required to achieve the improved grades.

Table 6.2: Summary of Confidence Grades by Metric

| Metric | Current Confidence Grade | Potential Confidence Grade with Recommended Improvements | Anticipated Value for Money/Effort of Improvement |
|---|---|---|---|
| Total / Average Passenger Lateness – Passenger Time Lost (hours) | C4<br><br>Reliability<br><br>No clear definition of APL, and unclear if / how TPL will be used. Note this metric focuses on train lateness, weighted by passenger numbers rather than direct passenger lateness (an important distinction for passengers changing trains on their journeys – see below).<br><br>Lack of complete documentation for:<br><br>• the creation of annual passenger journeys file from LENNON (particularly the infill journeys from Travelcards) led to uncertainty about the definition of journeys or journey legs;<br><br>• manipulation of journeys data to produce weights for applying to AML;<br><br>• no clear explanation of how PEARS data is used (e.g. why the payment rates are used as a proxy to peak and off peak splits of journeys).<br><br>The process for updating monitoring point weights is not tight enough for this metric (risks to Schedule 8 payments could deter regular updates).<br><br>Lack of data quality specialists in some Routes – updating berth offsets for new rolling stock or infrastructure could be at risk<br><br>No clear process for signing off the values (by Route or by Centre?).<br><br>Accuracy<br><br>Approximately 15% of passengers change trains on their journey and the method fails to measure the impact of missing their connection. (For TPL, the method results in | A2 (possibly A1 longer term)<br><br>Reliability<br><br>Documentation of definition, process and sign off should be straightforward to achieve.<br><br>Define a process for updating the monitoring point weights that is appropriate for this metric (which should involve a one-off exercise to produce a baseline for the start of CP6). However, producing a process consistent with Schedule 8 might be difficult to achieve given the possible reluctance to change weightings.<br><br>Recruit sufficient data quality specialists in all Routes to keep berth offsets up-to-date. Or, longer term, identify and deliver a technology solution.<br><br>Accuracy<br><br>Without specifically tracking the journeys of passengers, an accuracy of 2 is likely to be the highest achievable.<br><br>Tracking passenger journeys, e.g. from smartcard ticketing would provide accurate passenger numbers and lateness at destination and would not require detailed train location timings. In this case, an accuracy score of 1 would seem achievable, but is unlikely to be feasible for CP6. | Improving the accuracy of berth offsets would also benefit the On Time and Time to 15 metrics. The simplest solution would be to appoint more data specialists so that all Routes are sufficiently staffed.<br><br>However, an offset is a one-size-fits-all timing and by its nature will remain an estimate. Fitting GPS to trains is one solution currently being implemented through the Train Location Services Programme and proposed creation of the Industry Train Event Data (ITED) project. Clear protocols would need to be established for the control of the use of this data from TOCs and how it could be switched on and off.<br><br>An alternative technical solution might be a possibility (e.g. RFID tags) but this will need further investigation.<br><br>Tracking passenger journeys is unlikely to be feasible in the short term. However, the advent of smartcards would make this a more practical proposition. |

| Issue | 18 July 2017

Page 40

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

| Metric | Current Confidence Grade | Potential Confidence Grade with Recommended Improvements | Anticipated Value for Money/Effort of Improvement |
|---|---|---|---|
| | counting each leg once and hence over-states the numbers of passengers by approximately 15%). Some monitoring point weights might be out of date. Some berth offsets might be out of date. The infill journeys from Travelcards is approximate. The peak / off peak split is approximate by using the Schedule 8 payment rates. The impact of cancellations is inaccurate for services where an alternative service requiring an interchange is available (an extreme example is Kings Cross to Aberdeen where the next opportunity would be to change trains at Edinburgh rather than waiting for the next direct trains). | | |
| Reliability - Cancellations | B2 <u>Reliability</u> The process for producing cancellations is very similar to that currently in use for CaSL. It does rely on reconciling full / partial cancellations and skip-stopping recorded in TRUST with figures recorded by TOCs. In previous Independent Reporter reviews (last one reported in July 2013 under mandate AO/039) this was found to be robust with an A grading. The process for calculating the metric from the source data was found to be clear although needs to be fully documented. The process for signing off the figures needs to be documented (by Centre and/or Routes?). On a detailed point, the definition of the metric should state if it is the average % over all days in a period (rather than, for example, the average of individual daily %s). This could be important if weekend numbers of trains are very different to weekday numbers. We do support the chosen method of calculation. | A2 <u>Reliability</u> Provision of outstanding documentation. <u>Accuracy</u> Improved reporting of part cancellations and "fails to stop" are needed to improve accuracy. | Improving the accuracy of part cancellations and 'fail to stops' does not have an obvious solution. The proposed ITED would still rely on data provided by the TOC which would need to be reconciled with Network Rail data. |

| Issue | 18 July 2017

Page 41

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

| Metric | Current Confidence Grade | Potential Confidence Grade with Recommended Improvements | Anticipated Value for Money/Effort of Improvement |
|---|---|---|---|
| | Accuracy<br><br>No errors were found in the spreadsheet calculating the metric. Therefore, the accuracy reported in the last Independent Reporter review of CaSL is assumed. | | |
| Reliability – Severe Disruption | B1<br>Reliability<br><br>This metric relies on the process for measuring cancellations.<br><br>The process for calculating the metric from the source data was found to be clear although needs to be fully documented.<br><br>The process for signing off the figures needs to be documented (by Centre and/or Routes?).<br>Accuracy<br><br>Given the accuracy of the measured cancellations, the risk of wrongly identifying days with large numbers of cancellations will be small. | A1*<br>Reliability<br>Provision of outstanding documentation.<br><br><br>Accuracy<br>Improved reporting of part cancellations and "fails to stop" are needed to improve accuracy. | Relies on improving the measurement of part cancellations and "fails to stop" as described above. |
| On Time at All Recorded Stations – On Time | C2<br>Reliability<br><br>It is clearly stated that this metric is calculated at all recorded station stops. To aid transparency, it would help to have a list of such stations.<br><br>Lack of data quality specialists in some Routes is a significant concern. Updating berth offsets for new rolling stock or infrastructure could be at risk. Also the number of stations requiring up-to-date offsets will increase from those needed currently for PPM (destination only); initial analysis by Network Rail states that 1,074 of 1,134 intermediate station timing points could be made compliant. | A1-2<br>Reliability<br>Documentation of definition, process and sign off should be straightforward.<br><br>Recruit sufficient data quality specialists in all Routes to keep berth offsets up-to-date. Or identify and deliver a technology solution, for example GPS, to automatically calculate more accurate arrival times.<br><br><br>Accuracy<br>Subject to sampling the accuracy of updated offsets and comparing their variability for different rolling stock types, a grade of 1 or 2 should be achievable if all offsets | See TPL for effort required to improve the accuracy of berth offsets. |

| Issue | 18 July 2017
\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Page 42

| Metric | Current Confidence Grade | Potential Confidence Grade with Recommended Improvements | Anticipated Value for Money/Effort of Improvement |
|---|---|---|---|
| | On a detailed point, the definition of this metric needs to be tightened to state that it is the average over all trains in a 4-week period (rather than the average of daily %s).<br><br>The spreadsheet for calculating the metric is clear but not fully documented.<br><br>Accuracy<br><br>45% of non-Contractual Monitoring Points (CMPs) / Delay Recording Points (DRPs), essentially intermediate stations, have offsets that are at least 5 years' old. Approximately 40% of CMP/DRPs have not had their annual desktop review in the last year. Note this does not mean these offsets are inaccurate, just that they could be.<br><br>Truncating timings in TRUST should not affect the accuracy, given the definition of On Time being within 59 seconds.<br><br>Previous Independent Reporter reviews (the last one being in July 2013 under mandate AO/039) suggested that On Time accuracy at destination would be +/-0.2% nationally *if all berth offsets were up-to-date*. Reporting at TOC level would have wider confidence limits. | are kept up-to-date for new infrastructure and/or rolling stock. | |
| On Time at All Recorded Stations – Time to 15 | C1<br><br>Reliability<br><br>As for On Time<br><br>Accuracy<br><br>The impact of errors to berth offsets will be limited to the small number of trains about 15 minutes late, affecting whether they are classified as more of less than 15 minutes late. Averaging at all recorded stations will further reduce the impact of errors. | A1*<br><br>Reliability<br><br>As for On Time<br><br>Accuracy<br><br>Ensuring all offsets are up-to-date should ensure this metric being very accurate (similar to PPM). | See TPL for effort required to improve the accuracy of berth offsets. |

| Issue | 18 July 2017

Page 43

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

# 7 Recommendations and Observations

The recommendations arising from the work undertaken are set out below in Table 7.1. Some general suggestions for increasing the usefulness of the metrics then follow in Table 7.2.

| Issue | 18 July 2017

Page 44

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX

Table 7.1: Recommendations

| Reference | Recommendation | Benefit | Report Ref | Owner | Suggested completion date | Potential Confidence Grade on Completion |
|---|---|---|---|---|---|---|
| **2017NPM01** | Enhance the documentation for the calculation of all metrics, produce a Record of Assumptions, and apply best practice guidelines to the spreadsheets. The documentation should include the processes used to generate source data in BO-PSS for relevant metrics. | Improve transparency and reduce risk of reporting error, ensure business continuity | Sections 4.1.2, 4.2.2, 4.3.2 | Network Rail | September 2017 | Requirement to deliver A grades |
| **2017NPM02** | Enhance the documentation for the generation of passenger journeys data. | Improve transparency and reduce risk of reporting error, ensure business continuity | Sections 4.1.2 | ORR | September 2017 | Requirement to deliver A grades |
| **2017NPM03** | Prepare a metric definition for 'Average Passenger Lateness'. | Improve transparency and reduce risk of reporting error | Section 4.1 | Network Rail | July 2017 | As for TPL currently |
| **2017NPM04** | Update the metric calculation processes and the corresponding documentation used to generate Average Passenger Lateness outputs in addition to Total Passenger Lateness. | Produce agreed, amended metric output | Section 4.1 | Network Rail | July 2017 | As for TPL currently |
| **2017NPM05** | Recalibrate all monitoring point weights for the start of CP6 and review how much they have varied since the last recalibration. Review their likely impact on the accuracy of TPL / APL and, if deemed necessary, determine a process for more regular reviews. | Improve accuracy of TPL / APL metric and a process (if needed) for ensuring this element of its ongoing accuracy | Section 4.1.1 | RDG | March 2019 | Dependent on findings from recalibration exercise |
| **2017NPM06** | Review the method for determining the peak / off peak weights in aggregating AML. Initially, carry out sensitivity tests on the use of PEARS payment rates to determine their importance and consider if an alternative and more accurate method would be worth investigating. | Improve accuracy of TPL / APL metric | Section 4.1.1 | Network Rail | July 2017 | Dependent on findings |

| Reference | Recommendation | Benefit | Report Ref | Owner | Suggested completion date | Potential Confidence Grade on Completion |
|-----------|----------------|---------|------------|-------|---------------------------|------------------------------------------|
| 2017NPM07 | Ensure that there is a defined list of reporting stations for On Time and that there is a formal process for the addition or removal of stations. | Ensure that the On Time measure cannot be manipulated by changes to reporting stations | Section 4.3.1 | Network Rail | September 2017 | A requirement to achieve A grade |
| 2017NPM08 | Ensure a nominated Data Quality Specialist is in place in every Route competent in the management of the data requirements driven by the new measures and capable of dealing with the increased workload from checking all timing points. | Ensure that all the base TRUST data is correctly managed | Section 4.3.1 | Network Rail | September 2017 | Requirement to deliver A grade for APL/TPL and On Time and improve accuracy grade |
| 2017NPM09 | Improve current level of national compliance of annual desktop berthing offset checks from 62% for CMPs and 57% for DRPs to at least 90%. | Ensure train reporting at stations is reliable | Section 4.3.1 | Network Rail | March 2018 | Requirement to deliver A grade for APL/TPL and On Time and improve accuracy grade |
| 2017NPM10 | Develop and implement a process for checking the accuracy of timings at station timing points, justifying any difference to the checking of CMPs and DRPs by demonstrating its impact on the overall accuracy of On Time and T-15 metrics. | Ensure train reporting at stations is reliable | Section 4.3.1 | Network Rail | March 2019 | Requirement to deliver A grade for APL/TPL and On Time and improve accuracy grade |
| 2017NPM11 | Working with RDG, develop and coordinate a strategy for the use of GPS or other technologies such as RFID tags to improve the accuracy and reliability of reporting at stations, including reporting from those that do not currently report. | Improved reporting | Section 4.3.1 | Network Rail | March 2019 | Support the delivery of an A1 grade for On Time and APL/TPL |
| 2017NPM12 | Publish a code of practice setting out high level rules for setting the base for measurement, e.g. publication of station arrival and departure times. | Improved transparency and trust | Section 5.3 | RDG | April 2018 | Set a clear framework for judgement of compliance. |
| 2017NPM13 | Ensure Network Rail staff are briefed on the new measures and the reason and impact of the change on them is fully explained. | Increased staff awareness and buy in to measure | Section 5.3 | Network Rail | March 2019 | Underpin long term delivery |

| Issue | 18 July 2017

Page 46

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

| Reference | Recommendation | Benefit | Report Ref | Owner | Suggested completion date | Potential Confidence Grade on Completion |
|---|---|---|---|---|---|---|
| **2017NPM14** | Ensure TOC staff are briefed on the new measures and the reason and impact of the change on them is fully explained. | Increased staff awareness and buy in to measure | Section 5.3 | RDG | March 2019 | Underpin long term delivery |
| **2017NPM15** | Determine if and how the new metrics will be regulated in CP6 and whether PPM and CaSL will also be regulated. | Clarity to the industry to prioritise preparations for the start of CP6 and allow data to be analysed to set appropriate targets | Section 5.1 | ORR | July 2017 | Underpin long term delivery |

| Issue | 18 July 2017

Page 47

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS REPORT FINAL ISSUE 180717.DOCX

Table 7.2: Suggestions for Improving the Usefulness of the Metrics

| Reference | Observation/Suggestion | Benefit | Owner | Suggested completion date | Potential Confidence Grade on Completion |
|---|---|---|---|---|---|
| **2017NPMObs01** | Consider the use of more accurate timings, e.g. to 15 seconds (or to the second) in line with the development of the digital railway programme | Improved accuracy of On Time metrics and understanding of where and when time is lost | Network Rail | N/A | N/A |
| **2017NPMObs02** | Build the delivery of the new metrics into the Timetable Reliability programme to ensure more accurate SRTs, headways etc. underpin the timetable | Help improve the delivery of On Time | Network Rail | N/A | N/A |
| **2017NPMObs03** | Develop a consistent approach to daily, weekly and periodic reporting that underpins the key industry objectives set out for the new measures | Ensure that the basket of measures is managed to drive consistent behaviours | Network Rail/TOCs | March 2019 | N/A |
| **2017NPMObs04** | Develop a communications strategy to external parties that clearly explains why moving from PPM to On Time will shift reported performance from approximately 90% to 60% level | A positive message of the industry's commitment to On Time operation | RDG | July 2017 | N/A |

# Appendix A

Mandate

# Mandate for Independent Reporter Lot 3

| | |
|---|---|
| *Title* | Review of proposed new performance metrics |
| *Unique Mandate Reference Number* | L3AR004 |
| *Date* | 12 December 2016 |
| *ORR Lot Lead* | Peter Moran |
| *ORR lead for this inquiry* | Lynn Armstrong |
| *Network Rail Lot Lead* | Jon Haskins |
| *Network Rail lead for this inquiry* | Stephen Draper |

## Background

A National Task Force (NTF) sub-group with representatives from Network Rail, Office of Rail and Road, Department for Transport and Rail Delivery Group has been working to develop a number of metrics that could be used in CP6 as potential alternatives or supplementary metrics to the regulated performance outputs in CP5, namely Public Performance Measure (PPM) and Cancellations and Significant Lateness (CaSL).

The task force has focussed on producing a suite of performance metrics that address some of the limitations with the current performance measures. The aim of the suite of new metrics is to not only provide the right incentives for Network Rail, train operators and funders but also provide a level of transparency in respect of the impact of performance on passengers and to focus on issues that are relevant to passenger interest (e.g. right time arrivals).

## Purpose

The purpose of this review is for the Independent Reporter to review the reliability and accuracy of five defined measures presented to NTF in March, and two methodologies being considered to represent Passenger Weighted Disruption (as set out below) that could be used in CP6 either as alternatives or supplementary metrics to the CP5 regulated performance outputs. Consequently the Independent Reporter should provide an indicative guide to where the metrics currently stand in terms of reliability and accuracy and identify what needs to be done between now and the start of CP6 to provide assurance to ORR, Network Rail and wider industry stakeholders that these measures are adequately robust to be used as industry performance metrics.

- Total passenger lateness – Passenger time lost (hrs)
- Reliability - Cancellations
- Reliability - Severe disruption
- On time at all recorded stations – Right Time
- On time at all recorded stations – Time to 15
- Passenger Weighted Disruption (of which there are currently two methodologies under consideration)

The findings of the review should identify any gaps in system reliability and data accuracy and assist NTF and Network Rail in developing the robustness of these measures.

## Scope

Under this mandate the reporter should assess the system reliability and data accuracy for the following metrics, as they are defined in **Appendix 3** (new performance metric definitions) and **Appendix 4** (Passenger Weighted Disruption working definitions), based on the last 13 complete periods of data. Both of these appendices are attached as separate documents.

- Total passenger lateness – Passenger time lost (hrs)
- Reliability - Cancellations
- Reliability - Severe disruption
- On time at all recorded stations – Right Time
- On time at all recorded stations – Time to 15
- Passenger Weighted Disruption (of which there are currently two methodologies under consideration)

The work should:
- Review and comment on the processes and procedures by which the industry captures other data used in these metrics.
- Review and comment on the reliability, quality, consistency, completeness and accuracy of reported data in the trial.
- Provide guidance on what steps may be reasonably taken to make the calculation of the metrics more robust, including recommendations on how to ensure route based metrics can be compared across routes fairly;
- Identify an indicative confidence grading for the system reliability and data accuracy of each metric as it stands currently[1];
- Identify which confidence grading might be achievable and outline what improvements would be needed (along with associated costings) which could be taken to reach that grading. This should take into account the value for money of any identified improvements.
- Identify any serious issues in implementing these measures across the industry.
- Provide an opinion on whether the proposed metrics of performance are likely to be effective in creating more industry alignment by driving the right behaviours.

## Methodology

As part of this review the reporter will undertake the following activities:

1. Attend a kick-off meeting with ORR and Network Rail to confirm the methodology and programme;
2. Engage with Network Rail and NTF representatives with responsibility for each of the measures covered by this review to understand the current work being undertaken to improve the metrics;

---

[1] For the avoidance of doubt, the Independent Reporter is not being asked to 'award' a confidence grading with associated statistically significant sample sizes. This will be an indicative view as to what each measure may currently achieve.

3. Building on past assessments of existing measures[2], review the data being used to underpin the new measures and comment on their quality and fitness for purpose;

4. Review and assess the accuracy of the outturn data for each measure, based on a sample from the most recent 13 periods of data;

5. Establish a best practice guide for system reliability and data accuracy that takes into account the limitations of the current system and is aligned to the confidence grading system in **Appendix 2**;

6. Working with Network Rail and the industry, make recommendations on the improvements that could be made to improve the measure to achieve the grade in the best practice guide;

7. Prepare and submit draft and final reports, setting out the main observations and conclusions and recommendations arising from the review process.

## Timescales and deliverables

The formal deliverables for this project are:

1. Minutes of meetings to be provided with the draft and final reports.
2. Interim review meeting;
3. Draft report; and
4. Final report.

The key milestones for the project are as follows:

- Initial tripartite meeting – January 2017
- Network Rail meetings – January 2017
- Interim review meeting – February 2017
- Draft report and tripartite meeting – March 2017
- Final report – March 2017

## Independent Reporter proposal

The Reporter shall prepare a proposal for review by the ORR and Network Rail on the basis of this mandate. ORR and Network Rail will review the proposal with reference to the criteria for selection – see attached guidance document.

The final approved proposal will form part of the mandate and shall be attached to this document.

The proposal will detail methodology, tasks, programme, deliverables, resources and costs.

## Appendix 1 – Joint ORR and Network Rail guidance to Reporters

1. The purpose of this document is to describe the trilateral relationship between ORR, Network Rail and each Reporter. It sets out in a practical context what both ORR and Network Rail expect from Reporters, and seeks to encourage best practice. This will help Reporters to deliver work in

---

[2] http://orr.gov.uk/__data/assets/pdf_file/0013/2722/arup-performance-measures-2013-07-02.pdf
http://orr.gov.uk/__data/assets/pdf_file/0015/2355/right-time-performance-2013-01-22.pdf

a way which meets these expectations and requirements. These requirements will be taken into account as part of the Reporter Framework (as provided to Reporters).

2. This guidance is owned and updated as necessary jointly by ORR and Network Rail. In the event of any discrepancy between this document and the Reporter contract, the latter will prevail. This guidance does not provide an exhaustive list of responsibilities and should Reporters wish to discuss these guidelines further they should contact the following for a trilateral discussion:

- Tim Ward for ORR; and

- Jonathan Haskins for Network Rail.

## The trilateral relationship

3. Licence Condition 13 (LC13) of Network Rail network licence states:

- "The role of the Reporter is to provide ORR with independent, professional opinions and advice relating to Network Rail's provision or contemplated provision of railway services, with a view to ORR relying on those opinions or advice in the discharge by ORR of its functions under, or in consequence of, the Act. Where appropriate, ORR shall give the licence holder an opportunity to make representations on those opinions or advice before relying on them."

4. Reporters should be familiar with the obligations as set out in LC13 and the terms of the contract.

5. For the avoidance of doubt, in delivering this role, ORR and Network Rail expect that Reporters will also add value to Network Rail in helping it to improve its performance and business as provider of railway services, wherever possible. However, it is recognised that this is not the primary purpose of the Reporter under the Licence and that this may not always be possible to deliver each mandate.

## Role & duties of the reporters

6. Reporters must provide an independent view and remain impartial throughout the review.

For example:

- information should be shared equally and at the same time with both clients. Any correspondence or clarifications sought by Reporters should also be dealt with in the same way; and

- communication between all three parties should be open e.g. both ORR and Network Rail should be invited to or made aware of meetings or discussions even if the meeting is more appropriate with only one client.

## Identifying Reporter work

7. ORR will identify instances where there is a requirement to engage a Reporter. In practical terms, this is likely to arise from on-going discussions with Network Rail and in most cases (except urgent or exceptional cases) the potential for engagement of Reporters will have been identified in advance.

## Mandates – Reporter Proposals

8. Clause 4 of the contract sets out the key requirements around provision of services. Requirements for reporter work normally arise from the day to day discussion of issues between ORR and Network Rail.

9. ORR will prepare a draft mandate for each piece of work and will in most cases agree this with Network Rail.

10. Mandates will be presented in a standard format for consistency and will clearly set out:

- the purpose;
- the scope;
- why the review is necessary;
- what it will achieve;
- the expected outputs; and
- timescales for providing reports.

11. Once agreed with Network Rail, ORR will email the mandate to the relevant Reporter(s), asking for comments and a proposal for the work, which should include costs and CVs for the proposed Reporter team. The Reporter has seven working days to respond with a proposal or such other timescale as determined by ORR. Every proposal must include:

- costs;
- resources;
- CVs of the proposed mandate team – when providing proposals, Reporters should make the most efficient use of their resources including the most appropriate make-up of the review team;
- methodology for delivering the aims of the mandate;
- timescales;
- framework of meetings, including a tripartite findings meeting before issue of the draft report;
- expected deliverables and a concise explanation of how the aims of the mandate will be met; and
- for larger scale reporter studies, the project management approach and project plans should be made explicit

12. Where there are multiple Reporters on a Lot, the ORR and Network Rail will use the following criteria to determine which Reporter they will select to conduct the work:

**Procedure for Call Off under the Framework Agreements**

Where more than one Contractor has been selected for any particular lot, ORR and Network Rail will allocate mandates on the basis of the following criteria:

1. The expertise required is only available from one source. This may be due to ownership of exclusive design rights or patents.

2. Where the mandate constitutes follow up work, which is directly related to a recently completed study.

3. The Contractor which demonstrates the greatest expertise in the subject matter of the mandate or the approach required.

4. The Contractor's performance against the performance framework

5. An overall assessment of value for money based on cost and complexity of work.

If the ORR and Network Rail cannot determine the most appropriate Contractor for a mandate using the above criteria, ORR and Network Rail will conduct a mini-tender with the Contractors who have been awarded the relevant lot using the following criteria in order to determine the most economically advantageous proposal:

1. The Contractor demonstrates sufficient knowledge of subject matter and possesses the technical skills, resource and competencies required for the work.
2. Contractor Costs.
3. The Contractor demonstrates innovation and value for money in its proposal.
4. The Contractor's performance against the performance framework.

13. Prior to conducting such a mini-tender, ORR and Network Rail will inform Contractors of the relative weighting of the above criteria and of any additional sub-criteria applicable in the context of a particular mandate.

14. ORR and Network Rail will endeavour to discuss the proposals received and to confirm by e-mail within **five working days** that the proposal is acceptable (or otherwise). There may be circumstances where ORR and Network Rail need longer to respond.

15. ORR will then formally instruct the reporter to start work, and the reporter will arrange a start-up meeting with key representatives from both ORR and Network Rail.

## Mandates – During Delivery

16. The following sets out some key points regarding conduct of any inquiry. Reporters must provide an independent view and remain impartial throughout the inquiry. They should expect to discuss their progress and findings trilaterally with ORR and Network Rail and for some challenge to be given – particularly in relation to the factual accuracy of the findings.

### Costs and expenses

17. If additional funds are required to deliver a mandate beyond those agreed at the outset, a timely proposal and justification must be given to ORR and Network Rail (as soon as the issue arises). The Reporter should notify ORR and Network Rail who will discuss and respond in a reasonable timescale. Additional work (and cost) must not proceed without approval.

18. Any reasonably incurred expenses will be reimbursed by Network Rail. Only expenses that have been incurred in accordance with Network Rail's expenses policy will be paid.

19. All invoices should be sent to Matthew Blackwell (Matthew.Blackwell@networkrail.co.uk) at Network Rail prior to being sent to Network Rail Accounts Payable.

### Amendment to mandates

20. For practical reasons it may be necessary for a mandate to be revised once work has commenced or awarded. For the avoidance of doubt this will not lead to the ORR and Network Rail seeking to re-run the award of the mandate unless ORR and Network Rail agree that the revision constitutes a material change to the original mandate.

21. Unless otherwise directed, all key meetings must be trilateral and both parties should be made aware of any other meetings taking place.

22. The Reporter should take minutes of meetings, which should be provided to all parties within 7 working days.

## *Issues or concerns*

23. Should a situation arise whereby either ORR or Network Rail is dissatisfied with the quality of a piece of work, we will explain clearly our reasons, gain approval from the other client and then, if we deem appropriate, may request the Reporter to re-do that part of work at no additional cost.

24. Should the Reporter encounter any issues with an inquiry (review) the Reporter should notify:

- Tim Ward for ORR

- Jonathan Haskins for Network Rail

# Reports

## *The report document*

25. **All** Reports must include an 'Executive Summary' which should be written clearly, concisely and highlight key findings and key recommendations.

26. The full reports should also be written concisely in plain English, and should provide a brief 'Introduction' outlining the aims of the mandate and how these have been met. They should provide further detail on what is mentioned in the Executive Summary and there should not be any material points raised in the main report which have not already been mentioned in the Executive Summary.

27. Where there is commercially sensitive information in the report, the Executive Summary will be published on ORR's website, with any necessary redactions, instead of the full report. Otherwise, usually the full report will be published unless any redactions are appropriate due to a Freedom of Information Act exemption.

## *Recommendations*

28. A recommendation is a specific action that the Reporter considers, following its analysis, should be undertaken by either Network Rail, or any other party. While the majority of recommendations are likely to be for Network Rail, not all need to be.

29. Reporters should make all recommendations SMART (Specific, Measureable, Achievable, Realistic and Timebound). The Reporter should:

- provide a clear description of the recommendation and the benefit that implementation will deliver;

- outline the evidence which is required in order for the recommendation to be closed out; and

- discuss and agree a target date for completion of the recommendation with ORR and Network Rail.

30. Recommendations should only be included in the report if they actually add value to either ORR or Network Rail or another industry party and the benefits are sufficient to justify implementation.  It is acceptable for a report not to include recommendations, as long as key requirements of the mandate have been met (e.g. if an inquiry finds that Network Rail is fully compliant with its requirements).  A smaller number of well-targeted and SMART recommendations which will deliver tangible improvements is preferable to a large number of general recommendations.

31. In order to add further value, the report may also include observations on areas for improvement which do not need to be captured in a formal Recommendation if they are not central to delivery of the mandate requirements.

32. Recommendations will be tracked by the Reporter which generated them.

## Payment

33. Reporters must include the purchase order number, and unique mandate reference (UMR) number for work when invoicing Network Rail for payment.

34. The clients can query invoices and have the right to check timesheets (and expenses) and investigate work before payment is agreed.

## Post-mandate review

35. The clients will provide feedback on the work carried out, having assessed performance using the Performance Framework on a per mandate basis.  This will reflect any issues or concerns raised with the Reporter during delivery of the mandate.

36. The clients will also hold formal feedback sessions with each Reporter every six months to review progress.

# Appendix 2: Confidence grading system

## 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.

## 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 B

Best Practice Guide

Network Rail and Office of Rail and Road

<span style="color:green">**Independent Reporter - Lot 3**</span>

Mandate L3 AR 004: Best Practice Guide for implementation of proposed new performance metrics

Issue | 2 May 2017

Job number    254086-00

ARUP

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new
performance metrics

# Contents

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new
performance metrics

# 1 Introduction

This Best Practice Guide for the implementation and development of the new performance metrics was developed in the course of the review of the metrics undertaken for Independent Reporter mandate L3 AR 004. The metrics reviewed are as follows:

- Total Passenger Lateness – Passenger Time Lost (million hours)

- Reliability – Cancellations (%)

- Reliability –Severe Disruption (no. of days with > x% cancelled)

- On time at all recorded Stations –Right Time (< 1 minute late) (%)

- On time at all recorded Stations –Time to 15 (< 15 minutes late) (%)

Note that while the review was underway, it was decided by the NTF sub-group that Average rather than Total Passenger Lateness should be reported.

As set out in the original proposal for the review, the scope of the Guide is as follows:

- Assessment of how well the metrics meet the business requirements.

- For each metric, indication of the current accuracy and reliability confidence grade, and setting out of the grade that can reasonably be achieved given the limitations of the current systems within the industry.

- Identification of the steps needed to achieve the improved grades.

- Indication of the impact of the current industry limitations and consideration of how they might be alleviated in the longer term (e.g. by emerging technology).

- Consideration of any gaps relative to business requirements and whether any changes to or additional metrics might be required.

| Issue | 2 May 2017

Page 2

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\NEW PERFORMANCE METRICS BEST PRACTICE GUIDE ISSUE.DOCX

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new
performance metrics

# 2 Assessment of Metrics against Business Requirements

The aim of the new metrics is to complement or replace the existing Public Performance Measure (PPM) and Cancellations and Significant Lateness (CaSL) metrics, and, in so doing, to do the following:

- Provide the right incentives for Network Rail, Train Operating Companies (TOCs) and funders

- Provide a level of transparency in respect of the impact of performance on passengers

- Focus on issues that are relevant to passenger interest (e.g. right-time arrivals)

There is widespread cross-industry support for the new metrics, and they provide suitable incentives for the provision of on-time services at all reported stations across the network (in contrast to PPM). This focus on punctuality across the network will also assist with the identification of underlying problems such as inaccurate sectional running times or inappropriate aspects of the Timetable Planning Rules, helping to improve performance further in the longer term, and to make better use of network capacity. Similarly, the separation of cancellations from significant lateness will enable increased focus on and improved understanding of the causes of full and partial cancellations. While TPL (and APL) are quite abstract and difficult to influence directly, improvements to the punctuality and reliability metrics should feed through to improvements in TPL and APL, as intended. The use of APL instead of TPL will also remove the potential perverse consequence of increasing passenger numbers (an industry success) causing an increase in total recorded delay, while average delay is constant or declining.

The focus on punctuality across the system, on cancellations and on average passenger lateness also provide a much clearer representation than PPM and CaSL of the impact of performance on passengers, and enables and encourages a focus on the issues affecting that performance. Again, APL provides a better and more understandable representation than TPL of individual passenger experiences.

Additionally, the explicit measurement of punctuality, cancellations and lateness, and the associated potential for improved understanding of the underlying causes, is helpful in the quantification and development of business cases for system improvements.

In summary, we believe that the metrics, taken together and with careful implementation, will achieve the above objectives.

| Issue | 2 May 2017

Page 3

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\NEW PERFORMANCE METRICS BEST PRACTICE
GUIDE ISSUE.DOCX

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new
performance metrics

# 3      Road map for implementation of metrics

In the main report we have provided an assessment of the likely reliability and accuracy of the new metrics. Using the Independent Reporter confidence grading, we have indicated the grade that would be achieved with current reporting and what could be achieved in future with identified recommended improvements.

We have summarised these grades and recommendations in the following table. The table shows the milestone dates by when the recommendations should be implemented prior to the start of CP6. We also indicate some longer term considerations to further improve confidence, through the use of technology.

For example, there is considerable scope for improving the coverage and accuracy of the train location reporting system, making judicious use of technologies such as GPS or RFID detection (particularly in cases where GPS coverage is restricted by surrounding buildings or structures, or by covered stations like London Euston or Birmingham New Street). These and other, emerging technologies could be used to fill gaps in the current system, in parallel with the expansion of existing ATR technology, but also to supplement and potentially replace it where appropriate, and thus to improve accuracy levels to '1' for Right Time.

| Issue | 2 May 2017

Page 4

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\NEW PERFORMANCE METRICS BEST PRACTICE GUIDE ISSUE.DOCX

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new performance metrics

Road Map for improving reliability and accuracy of performance metrics

| Metric | Improvement in (indicative) Confidence Grading | July 17 | | Sep 17 | | Mar 18 | Apr 18 | Mar 19 | | | Longer term |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Passenger Time Lost | C4 to A2 (A1 in long term) | Define APL and amend calculation model accordingly. Review sensitivity of APL to PEARS payment rates | Determine if and how the new metrics will be regulated in CP6 | Improve user guide, Record of Assumptions, apply spreadsheet best practice guidelines | Adequate Data Quality Specialists in Routes | Improve current level of national compliance against berthing offset checks of CMPs and DRPs to at least 90% | Publish a code of practice setting out high level rules for setting the base for measuring, e.g. publication of station arrival and departure times. | Recalibrate all monitoring point weights, determine a process for more regular reviews (if found to be necessary) | Develop a strategy for the use of GPS or other technologies to improve the accuracy and reliability of train location reporting | Ensure staff are briefed on the new measures and the reason and impact of the change on them is fully explained | Tracking of passengers e.g. by smartcard technology. Implement GPS or alternatives to improve train location reporting |
| Reliability - Cancellations - Severe Disruption | B2 to A2 B1 to A1* | | | | | | | | | | |
| On Time as All Stations - Right Time - Time to 15 | C2 to A1-2 C1 to A1* | | | | Adequate Data Quality Specialists in Routes | Improve current level of national compliance against berthing offset checks of CMPs and DRPs to at least 90% | | Introduce checks at all non CMP / DRP station timing points with a national target to hit 75% compliance | | | |

| Issue | 2 May 2017

Page 5

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\NEW PERFORMANCE METRICS BEST PRACTICE GUIDE ISSUE.DOCX

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new
performance metrics

# 4 Best Practice Guidelines for Spreadsheet Models

Given the early state of development of the spreadsheets that calculate the new metrics, one of the first recommendations for implementation is the use of best practice guidelines for spreadsheet models. In this section we provide some guidelines that we have developed based on our experience for models that are subject to audit and review.

Our spreadsheet modelling philosophy is to employ explicit modularisation by splitting the task into as many components (and sub-components) as possible. Links between modules should be clear, simple and efficient. In cases where this cannot be achieved, it is inappropriate (by definition) for these modules to be distinct. Separate modules can be created, modified and verified independently, helping to reduce the complexity of the task and consequently minimising the risk of errors. At one level, this philosophy leads to the creation of separate Excel workbooks but it can also apply within individual files, for example by grouping worksheets together using colour-coding and/or naming conventions.

Our philosophy for spreadsheets is that a model should be the simplest amongst those that offer the desired capability. There should be no unused inputs or unnecessary calculations, and formulae should be concise and easily interpretable, with as few unique formulae as possible. We note that these objectives can actually often be contradictory; there is usually a trade-off between having many simple formulae and fewer complicated formulae. Such situations are to be assessed on a case-by-case merit in which other factors such as file-size, memory requirements and readability will also be considered.

Our advice on Best Practice guidance includes the following:

- Clear separation of Inputs, Calculations, and Outputs: Each worksheet of a model should be classified within this context, and that 'Input' sheets should not contain calculations or vice versa. There should be no unreferenced inputs or calculations. We recognise that, in many cases, inputs require substantial amounts of pre-processing in order to be in a useful format for the model; in such cases we suggest explicitly separating these calculations as a sub-category of 'Calculations' labelled 'Pre-Calculations'. An example of this would be the conversion of input monetary values from one unit or price base to another. Units should be consistent throughout the model's main calculations.

- Consistent formatting of spreadsheet models: This should be simple and clear, and should allow users to distinguish between different types of cells, e.g. user inputs, parameter inputs, other-model inputs, pre-calculations, calculations, check-cells and outputs. Annotations should be clear but not excessive. In addition, a cover sheet, QA history, and a version control record will assist new users and auditors.

- Linearity: spreadsheet models should be structured and presented in a linear style. This means that calculations within sheets should run from top-left to

| Issue | 2 May 2017

Page 6

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\NEW PERFORMANCE METRICS BEST PRACTICE
GUIDE ISSUE.DOCX

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new
performance metrics

bottom-right and worksheets should be ordered so that data flow from left to right.

- No hidden components: There should be no hidden cells or worksheets, or calculations placed in unusual locations e.g. a high numbered row or column. If, for presentational reasons, it would be useful to 'hide' some areas of a worksheet, the relevant rows/columns should be 'shrunk' to a nominal height/width; this allows a reviewer to clearly identify such areas. In cases where many components of the modelling require indexing by years, it helps if the years align throughout the model, i.e. the same year is in the same column on each sheet.

- Formulae: Should be consistent in 'blocks', or at least within rows/columns. There should be no hard-coded values within formulae except for universal constants (e.g. 60 when converting units of time). Functions that are difficult to audit, such as OFFSET and INDIRECT, should not be used. If named ranges are used, it helps to devise a user-friendly naming convention, and to take care not to retain excessive numbers of such ranges.

The correct transfer of data between the models within the 'Modelling Suite' is imperative, and in our experience is often an Achilles Heel of complex modelling suites. Risk of error can be minimised in three distinct ways:

- Do not use Excel's external references. This is because we believe each sub-model should be self-contained enough to work independently, and external references are notoriously difficult to work with.

- Minimise the amount of data that are required for transfer, and use consistent templates between sub-models. To illustrate the data transfer method: for example a 'Capital Expenditure and Operating Cost Model' will have one Output worksheet labelled "Out_to_BusinessCase". To transfer data, users will copy/paste-values the contents of this sheet into the 'Business Case Model', specifically into a sheet labelled "In_from_OCM". This clear and simple mechanism minimises the risk of human error and increases auditability. We note that in some circumstances it may be useful to retain multiple versions of inputs for different scenarios. This will allow simpler 'switching' within those models, and would facilitate easier comparison between scenarios (as opposed to repeatedly copying data which carries the risk of human error).

- To ensure that the correct versions of the outputs are used throughout the 'Modelling Suite', create and maintain a central Version-Log. This information will also be clearly noted in the models themselves, for example an input sheet will have a header specifying the source of the data (filename and version), time/date of data transfer, and a simple description of the scenario. These will be verifiable against the central Version-Log.

We note that it is not possible to foresee every circumstance and there may be rare occasions where an exception to some of the above is required. Any such situation would be clearly annotated within the relevant sub-model.

Documentation should consist of the following:

| Issue | 2 May 2017

Page 7

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\NEW PERFORMANCE METRICS BEST PRACTICE GUIDE ISSUE.DOCX

Network Rail and Office of Rail and Road

Independent Reporter - Lot 3
Mandate L3 AR 004: Best Practice Guide for implementation of proposed new
performance metrics

- A short note is added to the top of each sheet to explain what it does;

- Model user guide provides a simple set of instructions on how to use the model; and

- Record of Assumptions contains all assumptions within the models.

# Appendix C

## Confidence grading system

The confidence grading used to assess the reliability and accuracy of metrics has been defined by Network Rail and ORR for use by the Independent Reporters. This is shown below and is taken from the mandate for this review.

## 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.

## 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.

| Issue | 18 July 2017

Page C1

\\GLOBAL.ARUP.COM\LONDON\PTG\ICL-JOBS\254000\254086\4 INTERNAL PROJECT DATA\4-05 ARUP REPORTS\REPORT\REVIEW OF NEW PERFORMANCE METRICS
REPORT FINAL ISSUE 180717.DOCX