
Office of Rail Regulation
Network Rail
Right Time Performance Measure
Final Report (Redacted)

Issue 5 | 22 January 2013

This report takes into account the particular instructions and requirements of our client.

It is not intended for and should not be relied upon by any third party and no responsibility is undertaken to any third party.

Job number 223767-07

Ove Arup & Partners Ltd
13 Fitzroy Street
London
W1T 4BQ
United Kingdom
www.arup.com

ARUP

Table 5-1: Manual Reporting Locations and train Services by Route

Route	Percentage of Manually Reporting Locations	Percentage of Trains Terminating at Manually Reported Locations
Anglia	4%	3%
Kent	0%	0%
London North Eastern (including Midland)	23%	11%
London North Western	16%	9%
Scotland	20%	13%
Sussex	0%	0%
Wessex	0%	0%
Western (and Wales)	37%	31%
National	23%	9%

These numbers are gradually reducing as re-signalling takes effect, and the pace of change will quicken over Control Period 5 as Network Rail implements its Traffic Management strategy for signalling control. Whilst it is easy to suggest that these manual reporting points are on branch lines where train numbers are low and thus do not contribute significantly to the overall accuracy of RTP this is not always the case particularly on the Western and Wales Routes. In addition, the number of the reporting points in some Routes means that they cannot be ignored as being irrelevant to the overall accuracy of the measure. In the meantime, the apparent inadequate arrangements for ensuring data accuracy and reliability mean that overall quality of the performance data may be being compromised.

5.3 Manual Reporting

The processes associated with the use and monitoring of manual reporting locations are described in Section 5.2.

It is noted that the Margin Books provide no account of the manual reporting locations. It has been determined that in the absence of comparative data it would be necessary to undertaking an on-site assessment of the accuracy of the manual reporting of trains. This data are already in existence from work undertaken by Sheffield University as a direct commission for a number of TOCs as part of their review of train performance. The study team was granted access to such data for specifically nominated TOCs – Arriva Train Wales, First Great Western, First ScotRail, and Northern Trains. Each of these organisations approved the use of their respective data in the development of this study.

The data which were provided covered the sampling of arrival times undertaken by the University at the respective manual reporting locations. This provided some five hundred samples overall.

6 Conclusions

6.1 Introduction

This section uses the findings of the study to develop a series of conclusions. These are presented under what are considered to be the key headings. In turn these lead on to the Confidence Gradings described in Section 7.

The study aimed to cover the process and practicality of the derivation of Right Time Performance data across the country. In doing so, and as described in Section 4 and 5, a number of issues have emerged. The following seeks to capture the key conclusions of the study based on the findings of the various meetings and analysis which has taken place. The practical issues associated with the gathering of the data are presented, not in terms of a single Route's approach, but rather as the process is worked nationally. However, this is clearly influenced by the National Centre and individual Route findings.

6.1.1 Process

In terms of the locations where the RTP measures are based on auto-reporting it is clear that there is a well understood process which is being followed by a group of experienced individuals in the Routes.

The time taken to cycle from the identification of a need to re-survey a site to the point where it is signed-off is considered to be very long for what would appear to be a relatively straight forward procedure. The delay to any update represents a period of time when potentially the offset is wrong. This clearly compromises the integrity of the data particularly if a significant number of sites are in the review process at any one time.

It is concluded that whilst there is a clear understanding of the process and its drawbacks it was felt that it was too early to say that the issues had all been solved. Whilst work has clearly been done to advance the commercial settlement issues, there remain issues associated with obtaining the sign-off for changes to the offsets and the length of time this can take.

6.1.2 Backlog

There has been significant progress in reducing the backlog of updated berth offsets. Network Rail National Centre has set a target of April 2013 for the whole backlog to be cleared. Given the progress to date and in particular the efforts of the National Centre Performance Process and Controls team there is no evidence to suggest that this will not be achieved given the plan and revised processes which are in place and providing there is co-operation from all TOCs.

6.1.3 Resources

The issue of resources surfaced regularly during the Route meetings. It was no surprise that the Route with the dedicated individual looking after the data had the best grasp on the process and was clearly on top of the reviews and the progression of the updates. More commonly however, the work associated with

the berth offsets is part of a wider job description. In addition, because of the less immediate nature of the work to keep the offsets up to date, where there were competing pressures in the role, updating the offsets tended to get pushed down the priorities.

Without exception, the individuals carrying out these roles had been in post for several years and are knowledgeable on the subject. There was very little evidence of any form of succession planning or holiday cover in the majority of the Routes. This combined with competing pressures, means that there is a risk that keeping up to date may be a start-stop affair. Further, the reliance on one specialist means that the individual Route's compliance is vulnerable to staff moving on, or going on long-term sick leave. This increases the risk of the data being unreliable at any given time.

6.1.4 Manual Reporting

The processes associated with the recording of arrivals in manually reporting areas appear to be very poor. This is because the majority of the PDQS could not demonstrate any process, and were not able to confirm the accuracy of the means of measuring arrivals as they had not been checked for some time.

In one Route the study team saw copies of documentation covering checks which had taken place on an annual basis and this gave confidence that the relevant timings were as accurate as the process would allow. Even in this Route there was no evidence of documentation to support the means by which the reporting should take place including, for example, written instructions for the signallers. This was felt to be a significant omission leading to a conclusion that the rules governing the recording of the train arrivals appeared somewhat ad-hoc.

The removal of the need to audit the manual recording offsets has also effectively allowed the PDQSs to ignore the five yearly reviews or address any changes to the infrastructure, or train set formations. This cannot be good for data accuracy. The view that the manual reporting locations account for only a very small proportion of the train movements is not valid since on some routes they account for up to 37% of the CMP locations and up to 31% of the train movements. Clearly if these sites are misreporting then the data for the Route and TOCs operating over it are significantly compromised.

The use of data from Sheffield University and with the agreement of the respective TOCs it has been possible to come to a view on the accuracy of the manually reported locations. The analysis which has been undertaken is described in Section 7.3.2.

6.2 Summary

In summary the conclusions of the study are:

- For automatic reporting there is a documented process to develop the berth offsets but this is compromised by the length of time taken to include updates (although it is acknowledged that this has reduced significantly of late);
- There is a high risk that once the current National Centre push to clear the backlog of berthing offset updates is concluded the system could revert to type;

- The staff resourcing to maintain data quality are susceptible to competing pressures with little spare coverage; and
- Manual reporting would appear to be largely ill defined in terms of process and accuracy.

7 Confidence Grades

7.1 Introduction

There is a requirement under the Mandate covering the study to provide a Confidence Grading covering the system reliability and data accuracy associated with the generation of Right Time Performance data. This Section provides a commentary on the Confidence Grades which has been awarded and the justification for the grading.

The definition of the respective grades in the Confidence Grades is included in Appendix A.

7.2 System Reliability

7.2.1 Auto Reporting

The key factor in the determination of whether there is accurate data available on which to base RTP for auto reporting locations is the berth offset. Specifically it is the addition of the offset which determines the accuracy or otherwise of the arrival data.

As described in Section 4 there are three key issues associated with the data at present. These are:

- the delay in updating the data;
- the uncertainty surrounding the manual reporting point accuracy; and
- abnormal or disrupted working close to the final signal berth.

All these are considered to be a function of the currently defined processes. In the first case, the delay means that out of date offsets are being used when new data are available. There is also the process to determine the offset which is based on surveys and then averaged findings. Finally, in the case of abnormal working, this will invariably extend the time that trains take to reach their termination point and may result in trains being reported as 'right time' when they are actually one or two minutes late.

The significance of these issues means that the system is currently classified as Reliability Grade 'B'. This is as a result of the fact that whilst there is a strong well documented process in place there was evidence that there was a backlog of lapsed measures which were shown to be contributing albeit in a relatively minor way to inaccuracies in the reporting.

7.2.2 Manual Reporting

As described earlier the processes surrounding the manual recording of arrival times appear to be somewhat vague or missing. No evidence was produced during the course of the study to demonstrate the presence of documented processes covering manual reporting. In addition, the removal of the requirement to validate these measures, understood to be driven by the resource implications and practicality, has further clouded the governance of the processes. In the

absence of evidence to the contrary the manual reporting regime is currently classified as Reliability Grade 'D'.

7.2.3 Overall Reporting

In order to develop an overall assessment for the processes present to support RTP it is necessary to combine the grades awarded to the automatic and the manual reporting. The significant number of manual reporting locations (23%) which between them represent 9% of train arrivals means that the poor grading for the manual reporting processes cannot be ignored. However, it is clear that the overwhelming majority of the arrivals rely on the auto reporting systems. Nevertheless it has been decided to award an overall Reliability Grading of 'C'.

7.3 Accuracy Grading System

7.3.1 Automatic Reporting

The analysis which has been undertaken on the data is described in Sections 4.4 to 4.6 of the report.

Table 7.1 summarises the study's findings with regard to the various categories required in the project brief. This analysis only covers the auto-reporting locations. It is based on the data from the RPCR forms showing the sampling process and variations in offsets at the time of change married to the actual train arrival data for Period 5 2012/13. Combining these data sets allows the assessment of the trains which are on the cusp of right time which could be impacted by either the offset determination process or those offsets which are in a 'lapsed' state.

Table 7-1: Automatic Recording Accuracy Grading by Measure

Measure	% Error	Accuracy Grading
NATIONAL		
Great Britain	2.00%	2
England and Wales	1.94%	2
Scotland		
SECTOR		
Long Distance	1.84%	2
London & the South East	2.03%	2
Regional	1.72%	2
Scotland		
FRANCHISED TOC		
First TransPennine Express		
Greater Anglia		
Northern Rail		
Heathrow Connect		
First Great Western		
First Capital Connect		
CrossCountry		
London Midland		
London Overground		
East Midland Trains		
First ScotRail		
East Coast		
Merseyrail		
Virgin Trains		
Arriva Trains Wales		
Chiltern		
c2c		
Southeastern		
Southern		
South Western Trains		
Franchised	2.00%	2
NON FRANCHISED TOC		
Grand Central		
Heathrow Express		
Hull Trains		
Non-Franchised	1.56%	2

7.3.2 Manual Reporting

Based on the analysis of the data obtained from Sheffield University, taking an overall average of those results, and applying these to the Period 5 2012/13 arrival data population (as used in the auto reporting review) the accuracy gradings shown in Table 7.2 have been derived.

Table 7-2: Manual Recording Accuracy Grading by Measure

Measure	% Error	Accuracy Grading
NATIONAL		
Great Britain	4.93%	2
England and Wales	4.37%	2
Scotland		
SECTOR		
Long Distance	4.41%	2
London & the South East	4.91%	2
Regional	4.19%	2
Scotland		
FRANCHISED TOC		
First TransPennine Express		
Greater Anglia		
Northern Rail		
Heathrow Connect		
First Great Western		
First Capital Connect		
CrossCountry		
London Midland		
London Overground		
East Midland Trains		
First ScotRail		
East Coast		
Merseyrail		
Virgin Trains		
Arriva Trains Wales		
Chiltern		
c2c		
Southeastern		
Southern		
South Western Trains		
Franchised	4.63%	2
NON FRANCHISED TOC		
Grand Central		
Heathrow Express		
Hull Trains		
Non-Franchised	3.20%	2

7.3.3 Combined Accuracy

Based on the findings of the separate auto and manual reporting accuracy analysis Table 7-3 has been produced which shows the overall accuracy score. The combination of the results in Tables 7-1 and 7-2 takes account of the respective numbers of trains reported by each process.

Table 7-3: Combined Accuracy Grading by Measure

Measure	% Error	Accuracy Grading
NATIONAL		
Great Britain	2.23%	2
England and Wales	2.14%	2
Scotland		
SECTOR		
Long Distance	2.20%	2
London & the South East	2.11%	2
Regional	2.18%	2
Scotland		
FRANCHISED TOC		
First TransPennine Express		
Greater Anglia		
Northern Rail		
Heathrow Connect		
First Great Western		
First Capital Connect		
CrossCountry		
London Midland		
London Overground		
East Midland Trains		
First ScotRail		
East Coast		
Merseyrail		
Virgin Trains		
Arriva Trains Wales		
Chiltern		
c2c		
Southeastern		
Southern		
South Western Trains		
Franchised	2.23%	2
NON FRANCHISED TOC		
Grand Central		
Heathrow Express		
Hull Trains		
Non-Franchised	1.68%	2

7.4 Wider Implications

This study has focused on the current processes and accuracy of the arrangements associated with the reporting of RTP. It is clear however that some of the findings may have implications on the processes associated with other performance measures. These are likely to be the subject of further Mandates and have not been considered as part of this commission.

7.5 Comparison with Benchmark

Table 7-4 provides a summary of the current benchmarks and attained grading based on the current study.

Table 7-4: Comparison of Grading with Benchmark

Measure	Benchmark Grade	Attained Grade
National RTP	B1	C2
Sector RTP	B1	C2
TOC RTP	B2	C2

8 Recommendations

Based on the findings of the study the following recommendations are proposed:

Table 8-1: Reporter Recommendations

Number	Recommendation to Network Rail	Location in Text	Responsible Party	Due Date
2012RTP01	Limitations to the length of time permitted for TOC sign-off of the RPCR should be investigated based on deemed acceptance if no response is received after a specified duration taking account of the practicality of gaining such agreement	4.3.5	Industry Wide Acceptance Required	June '13
2012RTP02	Processes should be put in place to ensure that named posts in the Network Rail organisation are accountable for the accurate reporting of train arrival times where manual reporting is required	5.2.1	Stephen Draper	May '13
2012RTP03	The means of recording manual reporting arrivals should be appropriately documented	5.2.1	Stephen Draper	May '13
2012RTP04	The Routes should identify suitable coverage for the PDQS post to maintain the RTP data in times of absence. Consideration should be given to succession planning.	4.3.7	Stephen Draper	March '13
2012RTP05	There are variations in the way in which the Routes undertake the work associated with RTP data. It is recommended that the individual PDQS meet to share and exchange views on a six monthly basis and to check the applicability of the current guidance in the PMM.	4.3.4 5.2.1	Stephen Draper	March '13

9 Glossary of Terms / Abbreviations

CMP	Contract Monitoring Point
DAB	Delay Attribution Board
DRP	Delay Recording Points
FOC	Freight Operating Company
PDAC	Performance Data Accuracy Code
PDQS	Performance Data Quality Specialist
PMM	Performance Management Manual
PPM	Public Performance Measure
RPCR	Recording Point Change Request
RTP	Right Time Performance
SMART	Signal Monitoring and Reporting of Trains System
TOC	Train Operating Company
TRUST	Train Reporting System

Appendix A

Study Scope

Mandate for Independent Reporter Part A – Review of right time performance data

Audit Title:	Review of right time performance data
Mandate Ref:	AO/033
Document version:	Draft
Date:	11 July 2012
Draft prepared by:	Chris Fieldsend
Remit prepared by:	Chris Fieldsend
Network Rail reviewer:	Stephen Draper

Authorisation to proceed

ORR	Chris Fieldsend	
Network Rail	Stephen Draper	

Purpose

This mandate sets out the scope of work for the Part A Independent Reporter (Arup) to review right time performance (RTP) data. ORR has made a commitment to publish RTP data and it is therefore critical that the quality of the data can be reported. This will enable stakeholders and the public to make informed decisions based on the accuracy and reliability of the data.

Background

The ORR believes that RTP data provides the industry and the public with one of the most transparent measures of performance. ORR made a commitment in its 2012-2013 business plan to publish RTP data, and DfT support the move to disseminate the data as soon as possible.

ORR recognises that RTP data has not been independently quality assured, and believe the accuracy and reliability of disaggregated data should be known before it is published. This view is shared by the National Task Force and the Performance Delivery Group.

Scope

This review will assess the accuracy and reliability of the RTP measure, expressed as a moving annual average. RTP measures the performance of individual trains against the published planned timetable for the day, and shows the percentage of trains ‘right time’ compared to the total number of trains planned. A train is defined as ‘right time’ if it arrives at its planned destination station early or within one minute (i.e. 59 seconds or less) of the planned arrival time against the public timetable and does not miss any of its booked stops.

The review should assess the quality of the RTP measure at the following levels:

1. National (Great Britain, England & Wales and Scotland)
2. Sector (London & South East, Long Distance, Regional and Scotland)
3. Train Operating Companies (TOC; Franchised and Non-Franchised)

The review should:

- comment on the reliability, quality, consistency, completeness and accuracy of the reported data;
- present a confidence grade for the RTP measure at each of the three levels listed above (national, sector and TOC);
- make a judgement for each TOC on a likely confidence range of reported data; and
- make appropriate recommendations for the betterment of the RTP measure which could be undertaken at minimal cost.

Methodology

The Reporter should:

- outline their proposed methodology to assess the specific requirements listed above;
- meet with relevant Network Rail employees in each route and at Milton Keynes, to understand the process for collation, storage and computation of the RTP measure, along with the frequency of which it is reviewed;
- carry out a quick review of all relevant documentation and systems (including TRUST, margins books and berthing offsets), and comment upon their quality and fitness for purpose;
- identify the likely accuracy of reporting right time at a location dependent on how the data is captured – report on contributing factors which effect the variability of the accuracy of data;
- propose a statistically robust method of reviewing the RTP measure at TOC level;
- meet with representatives from all TOCs to understand any differences in collation methods;
- review any data quality issues arising from different signalling (manual and automatic) methods across the network; and
- state the confidence that ORR / NR can have in the findings, given their proposed methodology.

The Reporter should draw on (and not duplicate) work previously undertaken in its review of PPM and CaSL. Network Rail will share existing data with the Reporter on the variability of individual train performance. The Reporter should utilise this data to avoid any unnecessary data collation.

Deliverables

The Reporter should provide a publishable report, including findings, conclusions and recommendations. The report should be prepared in draft form and sent electronically to Network Rail and ORR, at the same time. The Reporter should facilitate feedback (via a tripartite feedback session if appropriate) and provide a revised report with track changes. This should be followed by a final report for publication on ORR's website.

Timescales

A fully costed proposal for this work is required by 19 July 2012. Work is expected to commence shortly after following approval by NR and ORR. A draft report is required by 14 September 2012 and a final report is required by 12 October 2012.

Independent Reporter Remit Proposal

The Independent Reporter shall prepare a fully costed proposal for review and approval by NR and ORR on the basis of this mandate. The approved remit will form part of the mandate and shall be attached to this document.

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

Confidence Grades

The Independent Reporter shall provide a confidence grade for the RTP measure at national, sector and TOC level. The confidence grading system in Annex A should be used. For each measure, the Independent Reporter should include the:

- confidence grade for this review;
- commentary on the grade against ORR's benchmark; and
- an indication of the highest achievable grade at each level.

Annex A: Confidence Grading System

System Reliability Grading System

System Reliability Band	Description
A	<p>Appropriate, auditable, properly documented, well-defined and written records, reporting arrangements, procedures, investigations and analysis shall be maintained, and consistently applied across Network Rail. Where appropriate the systems used to collect and analyse the data will be automated. The system is regularly reviewed and updated by Network Rail’s senior management so that it remains fit for purpose. This includes identifying potential risks that could materially affect the reliability of the system or the accuracy of the data and identifying ways that these risks can be mitigated.</p> <p>The system that is used is recognised as representing best practice and is an effective method of data collation and analysis. If necessary, it also uses appropriate algorithms.</p> <p>The system is resourced by appropriate numbers of effective people who have been appropriately trained. Appropriate contingency plans will also be in place to ensure that if the system fails there is an alternative way of sourcing and processing data to produce appropriate outputs.</p> <p>Appropriate internal verification of the data and the data processing system is carried out and appropriate control systems and governance arrangements are in place.</p> <p>The outputs and any analysis produced by the system are subject to management analysis and challenge. This includes being able to adequately explain variances between expected and actual results, time-series data, targets etc.</p> <p>There may be some negligible shortcomings in the system that would only have a negligible affect on the reliability of the system.</p>
B	<p>As A, but with minor shortcomings in the system. The minor shortcomings would only have a minor effect on the reliability of the system.</p>
C	<p>As A, but with some significant shortcomings in the system. The significant shortcomings would have a significant effect on the reliability of the system.</p>
D	<p>As A, but with some highly significant shortcomings in the system. The highly significant shortcomings would have a highly significant effect on the reliability of the system.</p>

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.

Benchmark Grades

As agreed with Network Rail, from Q3 2011-2012 data assurance reviews have been using this new confidence grading system. A characteristic of the new system is the introduction of a benchmark grade; the grade at which ORR believes the measure should be, given what we know about the processes and level of subjectivity in deriving it. It should be noted that the derivation and application of benchmark grades has recently been introduced, and all parties should decide how useful this element is throughout the review. The table below provides ORR's benchmark grades for the 2011-2012 data assurance review of RTP data.

Measure	Benchmark Grade
National RTP	B1
Sector RTP	B1
TOC RTP	B2

Appendix B

Sampling Impact Determination Methodology

The following described the process applied to the determination of the impact the data sampling has in the setting of the offsets.

1. For the berth changes where sample data is available we have looked at the distribution of survey results, to determine the 5% and 95% confidence levels – measured in seconds
2. For each location, made up of one or more samples, we have pulled out the **mean** confidence level, and also the **maximum** confidence level – again these are in seconds. In a few cases a single berth offset (potentially one which is not used that frequently) was significantly higher, hence taking the mean as well as the maximum for comparison.
3. Averaging these across all locations gave a mean of
 - a. +/- 7 seconds (mean value of confidence levels)
 - b. +/- 13 seconds (max value of confidence levels)
4. Using the Period 5 2012/13 arrival time distribution we then looked at the trains arriving either side of the right time plus one minute mark (in second intervals)
5. Based on this we then worked out the proportion arriving late/on-time which should have been classified differently in the 7s and 13s cases. This was done for each TOC.
6. These results were then added (taking account of train numbers) to create the Sector and national results.
7. In the tabulations the 95% confidence level figures are used.