Office for Statistics Regulation

UK Statistics Authority

Assessment of compliance with the Code of Practice for Statistics

Estimates of station usage

(produced by the Office of Rail and Road)

Assessment Report 352

November 2020

Office for Statistics Regulation

We provide independent regulation of all official statistics produced in the UK. Statistics are an essential public asset. We aim to enhance public confidence in the trustworthiness, quality and value of statistics produced by government.

We do this by setting the standards they must meet in the Code of Practice for Statistics. We ensure that producers of government statistics uphold these standards by conducting assessments against the Code. Those which meet the standards are given National Statistics status, indicating that they meet the highest standards of trustworthiness, quality and value. We also report publicly on system-wide issues and on the way statistics are being used, celebrating when the standards are upheld and challenging publicly when they are not.

Contents

Executive Summary

Judgement on National Statistics Status

ES.1 Estimates of station usage provide a valuable, trusted source of information about the use of the rail network across Great Britain (GB). They are used by a broad range of people, who appreciate the granularity of the data and the ability to use these statistics for further analysis.

ES.2 We identified several ways in which the Office of Rail and Road (ORR) should develop the statistics in order to meet the highest standards of the Code of Practice for Statistics and achieve National Statistics status. These are described in chapters one to three of this report. As we identified the improvements required during the assessment, ORR has made great progress towards addressing them. ORR has committed to further actions which will be carried out before the next release of the statistics in December 2020.

ES.3 Once ORR demonstrates that these steps have been taken, we will recommend that the UK Statistics Authority designate estimates of station usage as National Statistics.

Key Findings

Quality

ES.4 One of the main strengths of estimates of station usage is that they provide a data series going back to 1997/98. This means that users can explore trends over time and combine the data with their own local knowledge to understand the impact of infrastructure projects or changes to the usage of the rail network.

ES.5 Estimates of station usage are based on appropriate data and methods. The primary source for these statistics is ticket sales data. This is the best available source without fully gated stations or manual passenger counts at every station. ORR works well with its contractor Steer to improve the methodology used to calculate the statistics each year, while balancing the need for consistency to allow users to make comparisons over time.

ES.6 The use of ticket sales data as the primary source for these statistics results in several limitations. Although ORR publishes information about data quality and methods, it should be more detailed, and be available and communicated in a more accessible way to enable people to understand the caveats and use the statistics appropriately. We also consider that existing quality assurance practices carried out during the production process should be developed and put into a structured framework in order to further mitigate the risk of errors.

Public Value

ES.7 Estimates of station usage provide unique information about each of the approximately 2,500 mainline rail stations in GB. This granularity is one of the main reasons that the statistics are of interest to a very broad range to users: they are relevant to anyone no matter where they live. The team at ORR has a good understanding that the statistics are used by a variety of different users. A better

understanding of what people do and how they interact with the statistics will allow ORR to prioritise improvements and provide greater insight for users. We encourage ORR to continue its work to explore the underlying data and to work more closely with stakeholders to make use of their local knowledge. Providing contextual information, for example about infrastructure projects or wider issues such as the impact of COVID-19, will support users to understand the statistics.

Trustworthiness

ES.8 ORR is a well trusted and respected producer of official statistics. Consequently, the ORR website is often the first place that users go to find statistics and data about the rail industry. The analytical team at ORR is well managed, responsive to information requests and knowledgeable on issues which impact the GB rail network. ORR currently publishes several National Statistics which were reviewed by OSR via a compliance check in 2019. We found that there are good practices in place to ensure independent decision making and orderly release of ORR's official statistics.

ES.9 To enhance transparency, in addition to the information on user engagement for its broader set of official statistics, ORR has published an improvement plan for estimates of station usage that covers both short and mid-to-long-term developments. Maintaining this plan will keep users informed of changes, give them the opportunity to shape developments, and allow them to monitor progress against the plan. We would like to see ORR increase direct communication with users outside the rail industry. This will ensure that all users are informed about changes to the publication schedule or revisions to the statistics.

Next Steps

ES.10 ORR is aiming to meet the requirements described in Tables 1, 2 and 3 by its next annual release of the statistics in December 2020. Once ORR reports back to OSR on how it has met the requirements within this report, the UK Statistics Authority will take advice from OSR and decide whether to award the National Statistics designation. We expect to have a decision before the next publication.

Chapter 1: Quality

Introduction

1.1 Quality means that the statistics represent the best available estimate of what they aim to measure, are based on appropriate data and methods, and are not materially misleading.

1.2 Quality requires skilled professional judgement about the collection, preparation, analysis and dissemination of data and statistics in ways that meet the needs of the people who want to use them.

Findings

Communicating data quality issues and methods

1.3 Estimates of station usage are published by ORR once a year. The statistics include estimates of the number of entries, exits and interchanges at each station in GB. Statistics are presented for the most recent reporting period (financial year) and as time series data going back to 1997/98.

1.4 Statistics on station usage in GB must be estimated because there is not a fully gated rail network and therefore a complete record of passenger flows through stations cannot be obtained. With over 2,500 stations across GB it would not be practical or proportionate to carry out manual counts of passengers at every station. Therefore, ORR's estimates of station usage are derived primarily from ticket sales data using the rail industry's ticketing and revenue database, LENNON (Latest Earnings Network Nationally Over Night). These data are supplemented with other available data sources, such as manual passenger counts at individual stations, and adjusted to more accurately represent passenger movements across the rail network.

1.5 In the absence of a fully gated rail network or comprehensive national passenger count data, the use of ticket sales data is the best approach available for calculating estimates of station usage. Nonetheless, this approach does result in limitations which users need to be aware of. For example, ticketless travel is not captured using this data source and assumptions must be made to account for tickets which do not have specified start and end locations (such as tickets used in travelcard areas). Users need to understand how these limitations impact interpretation and appropriate use of the statistics.

1.6 Estimates of station usage are produced for ORR by Steer, a business consultancy specialising in infrastructure and transport. ORR works collaboratively with Steer to review the methodology for adjusting ticket sales data each year. Changes are implemented to address known issues or make use of new data sources as they become available.

1.7 The main source of information about quality and methods for users is the methodology report, which is produced by Steer for ORR and published alongside the statistics on the ORR data portal (the statistics section of the ORR website). There is also a short section in the factsheet and at the start of the Excel data tables which explains the data source for the statistics and makes users aware that

there are limitations associated with using ticket sales data. Some of these limitations are described in the data tables, but not in the factsheet.

1.8 Several of the more technical users we spoke to, such as analysts or academic researchers, told us that they consider the methodology report to be highly informative. The report answers many of the questions they have about methods and data quality, such as which ticket types are not included in ticket sales data or recent methodology changes. However, not all users are aware of the quality of the statistics and how they should or should not be used. For example, figures for individual stations should not be aggregated to create regional or train operator totals. Other statistics published by ORR, such as regional and passenger rail usage figures, provide a more appropriate measure for these groupings as they are derived from a different cut of the ticket sales database.

1.9 Many users, including those who are more technical, go straight to the data tables, factsheet or recently developed dashboard without looking at the methodology report. While the data tables and factsheet do contain some quality and methodology notes, there is no information about these in the dashboard. Users who only visit the dashboard would be unaware of the limitations associated with the statistics and any data quality issues associated with specific stations. The dashboard guidance, published in May 2020 by the Office for National Statistics' Best Practice and Impact Division, will support ORR to continue to develop this dashboard in line with best practice.

1.10 Some users have read a previous version of the methodology report and assume the quality issues remain the same each year. However, data quality issues can vary from year to year, for example as a result of methodology changes or required revisions. Although ORR and Steer have worked together to reduce the length of the methodology report by splitting information about historical methodology changes into a separate document, some users told us that the methodology report is still too long or that the language is not very accessible.

1.11 Finally, some users find the use of the word 'estimate' confusing, without the statistics being accompanied with a form of confidence limits. Providing such a range around the estimates is not possible given the methodological approach taken to calculate the statistics. ORR does provide users with some information in the tables about the likely cause of large changes in the data. However, more could be done to explain to users what impact the assumptions and adjustments have on the statistics and whether changes over time are likely to be real or a result of methodology choices.

1.12 ORR should develop ways to communicate the main data quality issues succinctly. This information should be in easily accessible places that users won't miss and should include information about:

- the limitations associated with using tickets sales data
- quality issues which impact the most recent release of data
- what can and cannot be done with the data
- why these statistics are described as 'estimates'

Approach to quality management

1.13 Quality assurance of these statistics is carried out by both ORR and Steer. In the methodology report, the quality checks carried out by Steer are described. For example, Steer investigates large changes for individual stations and makes comparisons with other data sources such as ORR passenger journey statistics, manual passenger count data collected by Department for Transport and Network Rail's gateline data for its managed stations. However, during 2020 two revisions to the statistics were carried out. The first of these was caused when an error was made in a manual update in Steer's process. The second error was caused by a ticket type being allocated the incorrect number of journeys in the original source data (the LENNON ticketing database). Some users we spoke to were concerned that similar errors would be found again in future. ORR currently publishes a revisions log for all its statistics detailing the reasons for each error and the impact of revisions. However, there is no information about the actions taken to mitigate the risk of future errors reoccurring.

1.14 We consider that ORR should have a more detailed understanding of the process carried out by Steer, including Steer's quality management approach. ORR needs to have a high level of assurance that there are not errors in Steer's part of the process and that quality issues in the data source are identified. ORR should work with Steer to develop a plan for how they can collectively reduce the risk of errors happening in future. This could involve commissioning a process map to help understand high risk points or requiring Steer to carry out either a one-off or periodic review of the production process.

1.15 We found that the statisticians at ORR carry out several quality assurance checks after they receive the data from Steer. For example, a considerable amount of time is dedicated to checking the data for each station. The data are split by region across the team and large changes over time within each region are investigated. The statisticians use their industry knowledge to sense-check the figures and carry out several checks on the data tables, for example looking for negative values or decimal points that should not be there. However, we found that there was no systematic process or guidance document to oversee this quality assurance process within ORR, for example by following a checklist for quality assurance checks or using a risk log to identify and manage high risk points in the process. Having such processes in place would provide assurances both to ORR that all the required checks are carried out for each release, and to users about the effectiveness of quality management. Furthermore, ORR has not published any information about what checks it carries out or what arrangements are in place with its data supplier, as we would expect to see accompanying a National Statistics release.

1.16 We found that users such as local transport authorities are keen to engage with ORR before the statistics are published in order to share their local knowledge. This would help to supplement ORR's quality assurance processes by identifying potential issues for further investigation.

1.17 While Steer and statisticians at ORR carry out a range of quality assurance checks on the data, it would be better to manage these using a more formal framework. The statisticians at ORR should also explain to users how they assure themselves of the quality of the statistics and how they manage the risk of errors. For example, by expanding the information provided in the revisions log and providing information about their internal quality assurance checks. Given the level of public interest in

these statistics, we encourage ORR to consider the level of assurance required based on the guidance set out within the Quality Assurance of Administrative Data framework. Referring to the practice area on quality assurance investigations and documentations will help ensure that proportionate quality assurance practices are carried out and that users are provided with sufficient detail to assure them about the quality of the statistics.

Table 1: Quality – Findings and Requirements

Findings	Requirement		Progress to date
The use of ticket sales data is the best approach available for calculating estimates of station usage, although this does result in limitations which users should be aware of. Many users find the methodology report produced by Steer highly informative. However, not all users are aware of the quality of the statistics and how they should or should not be used.	1	To ensure that all users are aware of data quality issues and how they impact the way the statistics should be used and interpreted, ORR should improve its published quality information.	ORR has shared with us a new version of the data table which helps to address this requirement by including new information about the source and the quality of the estimates. ORR is exploring whether the same information can be included in the dashboard. ORR is developing a new quality and methodology report which will be published in addition to Steer's report. ORR has also committed to expanding the information on strengths and limitations in the factsheet.
Although several quality assurance checks are carried out by both Steer and ORR, some users are concerned by recent revisions to the data. ORR should have a more detailed understanding of the process carried out by Steer, including its quality management approach, and should ensure that there is a systemic approach to its own quality assurance.	2	To mitigate the risk of errors in the statistics, ORR should develop a structured framework for quality assurance of the full production process and communicate this with users.	ORR is developing a quality assurance plan which will include a quality assurance checklist, a risk log and a process map of Steer's part of the production of the statistics. ORR has updated its internal guidance on completing revision logs to specify that information on why the error occurred and what steps have been taken to reduce the risk of repeated errors should be included.

Chapter 2: Public Value

Introduction

2.1 Value means that the statistics and data are useful, easy to access, remain relevant and help people to understand important issues and answer key questions.

2.2 Value is a product of the interface between the statistics and those who use them.

Findings

Understanding and meeting user needs

2.3 Estimates of station usage have a varied user base. Users range from analysts and policymakers in the rail industry and government who use the statistics to implement and evaluate infrastructure projects, to members of the public with an interest in their local station or the GB rail network.

2.4 The granularity of the data provides huge value to users. The availability of data for each station in GB allows users to carry out the local analysis they need to. It also means that the statistics have relevance to everyone, no matter where they live. Users also value the unique nature of these statistics: for most users, specifically those outside the rail industry, this is the only source of information about every station in GB published on a consistent basis.

2.5 A further strength of the statistics is the availability of data going back to 1997/98, allowing users to explore changes in station usage over time. Although methodology changes are implemented each year to improve data quality, the consistency of the main data source and broad approach to producing the estimates still allows for comparisons across the years. Where methodology changes result in reduced consistency for individual stations, this is explained in the methodology report. A comprehensive list of the changes implemented each year since 2006/07 is published in a supplementary report on historical methodology changes.

2.6 Estimates of station usage are freely available through the ORR data portal. The statistics are presented in several different ways, including a factsheet, Excel data tables and an interactive dashboard. The statistics are also accompanied by a methodology report and a Frequently Asked Questions document. Presenting the statistics in a variety of forms reflects the broad user base: different users value different parts of the publication. For example, the factsheet is of a lot of interest to members of the public and to non-expert users (for example, it is used by analysts in some organisations to share high-level information with policy or communications colleagues). The data tables are also heavily used, mostly by people who are carrying out their own analysis.

2.7 The statisticians at ORR have a good understanding of who their users are and are aware that the statistics are used by a variety of people. They regularly engage with users from the rail industry and government about their statistics via the Rail Statistics Management Group. ORR has developed innovative ways to engage with the public and raise awareness about these statistics. For example, by running an "Ask The Expert" session on Twitter or using animated charts to show the headline findings. However, we found that the team do not always know what users need the

statistics for and therefore what improvements should be prioritised to better meet these needs. For example, we found that users had several requests for additional information which are not currently included in the publication to support their use of the statistics, such as including information about how many services call at each station in the data tables.

2.8 We received feedback from users on how the layout of the data tables could be improved to make them more user friendly. We also found that while the methodology report is highly valued by some users, there are many who never open it and go straight to the data tables or factsheet. While we are pleased to see that ORR is striving to make the data accessible to a wider range of users through the development of an interactive dashboard, it is not clear from our research what purpose it serves. The users we spoke to told us they rarely or never use the dashboard, and those who did told us they found it hard to use. Although it could be the case that the dashboard is used by casual users, it is important to consider user need before, during and after developments to statistics to ensure that these developments add value for users.

2.9 We were pleased to see that ORR carried out a user survey earlier this year to gather feedback on its broader range of official statistics. However, we consider that ORR should be more proactive in seeking feedback on estimates of station usage and the way they are presented. ORR should have a better awareness of user needs for additional information and strive to meet them where possible. Where this is not possible, ORR should be open about which user needs can't be met and why.

2.10 ORR should also develop its understanding of how users interact with the various outputs from these statistics. This will enable ORR to make changes and improvements to the statistics which reflect user experiences and needs. This could be achieved, for example, by carrying out targeted user testing on the outputs or interviewing individual users about their needs. ORR could also develop its use of Google Analytics, which currently can only tell ORR about total page views. Since users do not have to register to access the statistics, making use of such tools would help ORR to understand how people use different parts of the publication including the dashboard and data tables.

Maximising insight and value

2.11 We commend the work that ORR has done since our compliance checks last year to ensure that plain English is used in its publications and encourage ORR to continue to do this in any new outputs that are developed. Removing any language barrier is an important enabler to achieve the goal of government and the rail industry to increase public confidence in rail.

2.12 The estimates of station usage factsheet provides an overview of the main messages from the statistics, for example stations with the most and least entries and exits. People who use the factsheet like the short form which makes it accessible and easy to understand. They also like the heavy use of infographics, rather than text, which makes the content engaging, and the inclusion of interesting facts like the most used seaside town stations. Users of the factsheet would like to see more content, for example the stations which have had the biggest change in use or information about current issues that impact the rail network in general, such as new stations. The current factsheet is quite busy so could benefit from being extended in size, though we encourage ORR to retain the short form and visually

engaging style. Understanding user needs and experiences will be crucial for ORR to balance the competing demands on the limited space available.

2.13 Users find the comments provided in the data tables which explain some of the changes in station usage from the previous year incredibly useful. We are pleased to see that links to external sources are included in order to provide greater context for comments in the data tables. Almost all users we spoke to would like more information about changes over time, in terms of the number of stations that comments are provided for and the level of detail in the comments themselves. While it would not be possible to provide detailed information about each of the approximately 2,500 stations without significant additional resource, ORR could do more to provide context and explanation for users. For example, we were told that sometimes the comments rely on the user having local knowledge of an area in order to understand them. Users would also like to see comments included for a longer period than just changes since the previous year, as some changes may be caused by something that happened more than a year before. Contextual information should be noted beside the station it is relevant to, for example infrastructure projects, or explained in the factsheet in the case of wider issues such as COVID-19. Some users also requested more consistent commentary in the data tables, which they felt can sometimes appear inconsistent.

2.14 As discussed above, we found that users such as local transport authorities are keen to engage with ORR before the statistics are published in order to share their local knowledge. As well as improving ORR's quality assurance processes, this would allow ORR to increase the insight provided in the data tables by expanding the number and level of detail of comments about specific stations.

2.15 One of the strengths of these statistics is that users can combine them with other data sets to carry out further analyses, ultimately increasing their value. Some of these data sets are held by the users themselves, while others are published by ORR and others. Although there is information about related statistics at the end of the Frequently Asked Questions document, it's not very easy to find and some users we spoke to were not aware of related sources of information. Some users found the relationship between ORR estimates of station usage and regional rail usage statistics confusing. To aid people in being able to maximise the value of the statistics by combining them with related information, ORR should link to other available sources in the main outputs such as the factsheet or data tables. These include ORR's other official statistics, as well as station timetables, which allow users to calculate the number of services per station, or Network Rail footfall figures.

Improving data quality

2.16 The statisticians at ORR work well with Steer to improve the quality of the data by making use of newly available data sources such as Transport for London's Oyster card information. ORR is also working to obtain data for the rail stations at Heathrow, which are currently unavailable but would be useful for users. The statisticians at ORR have a good understanding of rail infrastructure issues that will impact the data in the future, such as the Crossrail development in London. They are currently considering how to deal with the Crossrail development once it is implemented. ORR is also in the process of analysing the impact of COVID-19 on estimates of station usage. COVID-19 will have a large impact on station usage during 2020 and potentially beyond, so we are pleased that ORR is already starting

to understand this impact on the data and data quality. This work should also help ORR to support users' needs for insights, for example comparability over time.

2.17 We commend ORR's work with Steer to understand and improve the quality of the data over time. We encourage ORR to continue to find and make use of new data as they become available, for example electronic tickets or mobile phone data. As discussed above, we would like to see ORR work with stakeholders such as local transport authorities to continue to improve its understanding of the data and identify potential data quality issues.

Supporting reuse of the data

2.18 One of the main ways that these statistics are used is for people to carry out their own research or analysis. In order to carry out analysis, the Excel data tables are heavily used. Generally, people are able to use the data tables easily. In particular, the reference data and metadata included alongside the statistics add a lot of value for users. However, both the data tables and dashboard could be improved further. For example, the main columns of interest (entries to and exits from stations) are far off to the right of the spreadsheet, so users need to scroll along or hide numerous columns every time they open it. ORR's recent user survey also found that people would like extra functionality in the data tables and dashboards (for example by adding the ability to download selected data from the dashboard). While some users carry out their analysis using Excel, others import the data to a programming software such as R. To make reuse of the data as easy as possible for these types of users, it is recommended that data should be provided in machine-readable, re-usable formats which require minimal or no manual manipulation by the user.

2.19 To enhance their support for technical users who wish to carry out their own analysis, ORR should review the data tables and dashboard and implement improvements to make them more user friendly. We also recommend that ORR publish a machine-readable version of the data to support users who will use a programming software for their analysis (for example, a CSV file which follows tabular data standards). A better understanding of how users interact with the outputs would support ORR to meet this requirement.

Planning and sharing developments

2.20 We were pleased that ORR recently carried out a user survey about all its official statistics. ORR has developed and published an implementation plan based on the user feedback. This includes commitments to improve the content of its quality and methodology reports and to explore ways to increase the usability of its dashboards.

2.21 During our assessment, we recommended that ORR develop a specific improvement plan for estimates of station usage. We are delighted that ORR worked quickly on an action plan, which it has published on its website and promoted via its newly developed newsletter. The proposed developments will be of interest to users, particularly where there are significant changes planned to the statistics, for example the plan to include data for rail stations at Heathrow in future.

Table 2: Value – Findings and Requirements

Findings	Requirement		Progress to date
Estimates of station usage have a varied user base. ORR does not always know what users need the statistics for or what parts of the statistics they use, and therefore what improvements should be prioritised to better meet user needs. ORR should be proactive in seeking feedback on the statistics.	3	To meet the needs of the broad range of people who use these statistics, ORR should develop its understanding of what users do with the statistics and how they interact with the outputs.	ORR is currently engaging with some users who contributed to this review in order to understand their specific needs and gain feedback on planned developments. ORR has committed to a proactive approach to requesting feedback on the statistics from users via their existing channels, such as the RSMG or the Transport Statistics User Group. ORR is also making use of its newly developed newsletter to seek feedback. In the longer term, ORR plans to redesign the web page in order to gain more detailed insight into use of the statistics using Google Analytics.

Findings	Requirement		Progress to date
The provision of comments in the tables to explain some of the changes in the data and the use of plain English support users in understanding what the statistics can tell them. Users are also able to combine the statistics with other data sources to carry out further analysis. Users told us they would like more information about changes in the data and some users were not aware of related data sources.	4	To maximise insight and increase the value of these statistics for users, ORR should: a. Work with stakeholders who can provide local knowledge about stations which can be shared with users b. Ensure that users are aware of important contextual information, such as infrastructure changes, and the impact of this on the statistics c. Improve coherence with other related information and data sources	For the next release of the statistics, ORR is currently engaging with local transport authorities to obtain additional insight at station level which can be included in the data tables. As part of the development of the factsheet into a statistical release format, ORR has committed to including contextual information relevant to each release of data. For example, new or closed stations that year, or the impact of COVID-19 on the data. ORR has also committed to including information in the factsheet and the new quality and methodology report about how these statistics relate to other ORR statistics, as well as links to other related sources.
ORR works well with Steer to improve the quality of the data each year by making use of newly available data sources.	5	To further improve the quality of the data used to produce these statistics, ORR should: a. continue to investigate and use new data sources b. engage with stakeholders who can provide local knowledge about stations	As above, ORR is in the process of engaging with local transport authorities for the next release of the statistics. ORR continues to explore the addition of Heathrow stations for the next release of the statistics.

Findings	Requirement		Progress to date
One of the main uses of these statistics is people carrying out their own research or analysis. In order to do this, the Excel data tables are heavily used. While people find that the data tables are easy to use, both the data tables and dashboard could be improved further. There are also some users who prefer to import the data into a programming software, rather than use Excel.	6	To enhance its support for users to analyse and reuse the data, ORR should: a. review the data tables and dashboard and implement improvements to make them more user friendly b. ensure that the data is available in an accessible, machine-readable format	ORR is currently reviewing its data tables. Proposed changes to the layout and structure have been shared with some users for feedback. ORR is developing a CSV file which could be easily reused by people for further analysis in statistical software. Feedback on this has also been sought from users.

Chapter 3: Trustworthiness

Introduction

3.1 Trustworthiness means that the statistics are produced free from vested interest and are based on the best professional judgement of statisticians and other analysts.

3.2 Trustworthiness is a product of the people, systems and processes within organisations that enable and support the production of statistics.

Findings

Demonstrating trustworthiness

3.3 The trustworthiness of these statistics is vital given that they are the only available source of information about station usage for many users who heavily use and rely on them. We found that people trust ORR as an organisation and consider it the first port of call for statistics and data about the rail industry in GB. Users have confidence in ORR as producers of official statistics and, as a result, in the quality of the statistics themselves. ORR is included as a case study in OSR's Code of Practice pages under the Honesty and Integrity principle. We received excellent feedback about the analytical team, who users told us are very helpful in responding to queries and information requests.

3.4 We commend ORR's innovative approach to engaging with users of these statistics by running an "Ask The Expert" session on Twitter on the day the statistics are released. Users, particularly members of the public, value this opportunity to virtually speak to one of the statisticians. We encourage the statisticians at ORR to continue to make use of their social media platforms to raise awareness of the statistics and interact with their users.

3.5 The statistics are currently published approximately nine months after the end of the reporting period. Although this does not present a significant issue for users, some told us that they would appreciate more timely information. We are pleased that ORR is working to bring the publishing date forward for future releases.

Communicating with users

3.6 ORR regularly communicates with users from the rail industry and government via the Rail Statistics Management Group. These users have a good awareness of issues such as revisions or delays to the statistics. However, we found that some users outside this group, for example academics or members of the public, do not have enough information about these issues. We also found that these types of users tended not to know how to contact someone at ORR if they had any queries.

3.7 We would like to see ORR improve communication with user groups who it does not currently have direct communication with. This could be achieved by making use of the mailing list that ORR has been developing for these statistics to facilitate better communication with users.

Table 3: Trustworthiness – Findings and Requirements

Findings	Requirement		Progress to date
ORR communicates well with users in the rail industry and government, who are aware of significant changes to the statistics. However, some users outside these groups, for example members of the public, do not have enough information about issues such as revisions or delays.	7	To ensure that all users are aware of revisions and changes to the release schedule, ORR should improve its communication with users outside the rail industry.	ORR has developed a newsletter which is used to share updates about the statistics with users and encourage feedback from users. ORR has also engaged with the Transport Statistics User Group to ensure that links to recent ORR statistical releases are included in its monthly newsletter.

Annex 1: About the Statistics

The Statistics

A1.1 Estimates of station usage are published by ORR once a year. The statistics include estimates of the number of entries, exits and interchanges at each station in GB. The estimates are produced by Steer, a business consultancy firm, on behalf of ORR. Statistics are presented for the most recent reporting period (financial year) and as time series data going back to 1997/98.

Data Sources and Methods

A1.2 Estimates of station usage are primarily derived from ticket sales data using the rail industry's ticketing and revenue database, LENNON (Latest Earnings Network Nationally Over Night). The LENNON database contains information on almost all national rail tickets bought in GB. Its primary purpose is to allocate the revenue from ticket sales between train operators.

A1.3 Since there are known omissions from the LENNON database, this dataset is supplemented by other ticket sale information. For example, for Transport for London tickets or tickets without a geographical destination (zonal tickets). Several assumptions are also made about the ticket sales data: for example, in order to decide how many journeys a multi-journey ticket such as a season ticket should represent; or when allocating tickets for groups of stations (for example London or Manchester stations) to individual stations.

A1.4 Data from ticket sales are then supplemented with data from other sources, such as manual passenger counts at group stations, and adjusted to provide more-accurate estimates of station usage.

Uses and Users

A1.5 Estimates of station usage are the only source of published information on the use of all stations in Great Britain.

A1.6 The information published is used extensively by government departments, local transport authorities and community groups, and transport watchdogs. The statistics are also of interest to journalists, academic researchers and members of the public.

A1.7 User engagement during this assessment identified the following uses of estimates of station usage:

- Monitoring station use to understand capacity and identify potential issues
- Informing business cases for station or service developments
- Planning, monitoring and evaluating infrastructure projects
- Used in the production of other official statistics, such as those from the National Rail Passenger Survey
- National and local media outputs that inform public interest
- To facilitate academic research

- Personal interest

Annex 2: The Assessment Process

A2.1 This Assessment was conducted from April 2020 to October 2020.

A2.2 This report was prepared by the Office for Statistics Regulation and approved by the Regulation Committee on behalf of the Board of the UK Statistics Authority, based on the advice of the Director General for Regulation.

A2.3 The regulatory team – Anna Price, Lewis Jack and Mary Cowan – agreed the scope of and timetable for this assessment with ORR in April 2020. Documentary evidence for the assessment was provided by ORR in May 2020. Following a review of this evidence and after carrying out user engagement, the regulatory team met with ORR in July 2020 to discuss compliance with the Code of Practice for Statistics.

A2.4 A key part of the assessment was talking to people who use the statistics, to help us to understand the current value of the statistics, and where there is the potential to increase this. We approached known and potential users of these statistics and conducted 14 interviews. We also received written feedback on the statistics from several users. Users we engaged with included people who work in government, for local transport authorities and travel watchdogs, as well as journalists, researchers and members of the public with an interest in this topic. We also spoke to Steer in its capacity as data providers.