



In Confidence  
DeltaRail -LD32952/02  
Issue 1

## Station Usage 2006/07

<b>Title</b>	Station Usage
<b>Customer</b>	Office of Rail Regulation
<b>Customer reference</b>	Jay Lindop / Aruna Ramyeed
<b>Confidentiality, copyright and reproduction</b>	Copyright DeltaRail Group Limited Enquiries about copyright and reproduction should be addressed to the Head of Business Risk Services, DeltaRail Group Limited.
<b>File reference</b>	J:\Rail\Consultancy\1_Projects\32952 (ODM and Station Usage 06/07)\LD32952 Station Usage
<b>Report number</b>	LD32952 / 02
<b>Issue number</b>	Issue 1

DeltaRail Group Limited  
Central House  
2<sup>nd</sup> Floor  
London  
WC1H 0JN

DeltaRail is the trading name of DeltaRail Group Limited  
DeltaRail is certified to BS EN ISO9001

	<b>Name</b>	<b>Signature</b>	<b>Date</b>
<b>Author</b>	Georgios Georgiou		23 May 2008
<b>Reviewed by</b>	Brian Ball		23 May 2008
<b>Approved by</b>	Brian Ball		23 May 2008

## Executive Summary

This report explains the information contained within the Station Usage file. The report provides guidance to the methodology we have followed during the process of creating the final information in this file for financial year 2006/7.

Station Usage data consists of estimates of the total numbers of people

- Travelling from the station (entries)
- Travelling to the station (exits)
- Interchanging at the station (interchanges)

Information is given for all the national rail stations around England, Scotland, and Wales based on tickets sales data. These results are the most recent in a series we have supplied since 1997/98. The spreadsheet is in a similar format to those previously provided.

Several major enhancements have been made to the methodology for 2006/7 compared with previous years. These will result in a more comprehensive and accurate dataset. The four main changes that have been made are:

1. We have added additional data representing journeys made on national rail using London Travelcards that were sold at a TfL outlet, or at a TOC office in Zone 1. These data were excluded in previous years since it is not clear which journeys have been made, but this year a process has been employed to estimate the likely journeys being made.
2. We have added information about Gatwick and Stansted Express tickets that are sold outside of the LENNON system.
3. We have included tickets sold for national rail journeys at London stations that have only an LUL, not a TOC, booking office
4. We have changed the methodology for journeys made to and from "Group stations" (including London).

Enhancements 1, 2 and 3 have had the effect of increasing the number of journeys represented in the Station Usage file. Therefore, the increase in the number of exits and entries, up by 20% on last year from 1.6 billion to 1.9 billion, is partly due to these methodology changes.

We have undertaken a thorough programme of checks on the Station Usage data. Results of some checks are listed in the spreadsheet, and further details are given in Appendix 1.

# Contents

- 1 Introduction ..... 1**
- 2 Base LENNON Data ..... 1**
- 3 Summary of Results ..... 2**
  - 3.1 Overview of the Entries and Exits Results ..... 2
  - 3.2 Overview of the Interchanges Results ..... 4
- 4 Station Usage File Definition ..... 5**
- 5 Data Excluded From the Study ..... 6**
- 6 Entries and Exits Methodology ..... 7**
  - 6.1 Journeys added in this year ..... 7
    - 6.1.1 Travelcards Issued by TfL or sold at TOC-run stations in Zone 1 ..... 8
    - 6.1.2 Airport Links ..... 8
  - 6.2 Unknown Destinations ..... 8
- 7 Interchanges Methodology ..... 12**
- 8 Regions, Counties and Districts ..... 12**
- 9 Limitations of LENNON data ..... 13**
  - 9.1 Known Problems of Data Capture ..... 13
  - 9.2 Travelcards ..... 13
  - 9.3 Return and Single Journey Tickets ..... 13
  - 9.4 Multiple Tickets ..... 14
  - 9.5 Rail Staff Passes ..... 14
  - 9.6 Ticketless Travel ..... 14
  - 9.7 Other Rail Systems ..... 14
  - 9.8 Factors Affecting the Data ..... 14
  - 9.9 Journey Factors ..... 15
- Appendices ..... 16**
- Appendix 1 - Checks ..... 16**
  - Check 1: Unequal Entries & Exits ..... 16
  - Check 2: Large Changes in Usage ..... 16
    - New Station ..... 17
    - Gating ..... 17
    - Change in Service Pattern ..... 17
    - Ticket Issuing Facilities Changes or Product Changes ..... 17
    - Engineering Work ..... 17
    - Tourism ..... 17
    - Special Stations ..... 17
    - Trend of Growth or Decline ..... 18
    - Changes in the Sales of Individual Ticket Types ..... 18
  - Check 3: Group Stations & London Termini ..... 18
    - Group Stations ..... 18
- Appendix 2 - Overview of the ORCATS Allocation Process ..... 19**

## 1 Introduction

DeltaRail has provided a series of Station Usage data for the rail industry in previous years. This report accompanies the Station Usage data for 2006/7, provided in the spreadsheet "Station Usage 2006-07.xls". It gives details of the process and outputs in producing the Station Usage file for financial year 2006/7, on behalf of the ORR. The journey data has been taken from the ticket sales recorded in the LENNON database.

DeltaRail are providing the ORR with an MS Excel file containing entries, exits and interchanges made at stations throughout England, Scotland and Wales, for the financial year 1 April 2006 to 31 March 2007. For the entries and exits, figures are split into the three main categories of the available ticket products (Full, Reduced, and Season), for the interchanges only totals are available.

## 2 Base LENNON Data

DeltaRail received comprehensive LENNON sales data from ATOS Origin based on the financial year 1<sup>st</sup> April 2006 until 31<sup>st</sup> March 2007. The fields in the base data are shown below:

- Product Code
- Product Description
- Origin NLC
- Destination NLC
- Route Code
- Day of Week
- Issues
- Revenue
- Journeys

### 3 Summary of Results

Compared with the initial LENNON dataset received, around 9% of the total journeys are excluded from the final Station Usage delivered for 2006/7. Details of which records have been excluded are shown in Section 5. The following table gives the total number of entries, exits, and interchanges made over the whole network for 2006/7.

**Table 3-1: Entries, Exits and Interchanges for 2006/7 and 2005/6 (millions)**

Year	Entries	Exits	Entries & Exits	Interchanges
2005/6	800.46	800.63	1,601.09	120.02
2006/7	958.10	958.76	1,916.85	133.19

#### 3.1 Overview of the Entries and Exits Results

In this section we set out a summary of the overall entries and exits results. The spreadsheet contains entries and exits results for 2,520 stations. Around 960 million entries, and a similar number of exits, were made in 2006/7 – an increase of 20% on the year 2005/6.

The big increase in the total entries and exits occurred mainly because of the improvements we made to the process. This was the first year we considered fully journeys made within the London Travelcard area. In particular, we analysed journeys for TfL sold Travelcards and Travelcards issued by a TOC office in Zone 1. These products account for about 150 million journeys.

The table below shows data for the 10 stations with the highest numbers of entries and exits for 2006/7.

**Table 3-2: Top 10 stations based on the entries and exits made for 06/07.**

Station Name	Total Entries & Exits 06/07
London Waterloo	83,993,314
London Victoria	66,749,335
London Liverpool Street	55,265,748
London Bridge	47,576,684
London Charing Cross	34,779,287
London Paddington	27,258,741
London Euston	25,585,113
London Kings Cross	22,503,777
London Cannon Street	21,106,127
Glasgow Central	21,002,296

Nine of the top ten stations have remained the same as last year and nine of the ten are London Terminals. London Cannon Street is the new entry in the top ten list while Manchester Piccadilly is the station that did not make it into the top ten. A major reason for Manchester Piccadilly usage apparently having decreased by 30% is that our revised methodology for Group stations has recognised more journeys going via other Manchester stations, whereas the previous methodology assumed they all went via the main station, i.e. Manchester Piccadilly. For similar reasons, Glasgow Central usage has apparently decreased by 30%, though it did still manage to make it into the top 10.

The total journeys made at one of the top ten stations account for a total of 406 million, 17% more than the journeys made last year. The top ten stations account for 21% of all entries and exits, that is about the same proportional split in 05/06 and 04/05.

The above table shows that the journeys for the top 10 stations have increased compared to last year. These changes, and indeed any comparisons between the data year on year, need to be viewed with the knowledge of significant improvements in the overall methodology for 2006/7 compared to 2005/6.

The main changes affecting total entries and exits are:

- Additional (non-LENNON) flow data included for Gatwick and Stansted has increased the journey figures for the involved station.
- TfL sold Travelcards account for about 150 million journeys were not included in our analysis the previous years.
- Changes to the methodology applied for the group of stations. This change effectively shows a decrease for most of the major station of each group and an increase for the other stations forming part of the group (as described for Glasgow Central and Manchester Piccadilly above).

We need to mention that for stations like Glasgow Central and Manchester Piccadilly, and indeed all Passenger Transport Executive (PTE) areas, the usage figures are underestimated because of the missing data from PTE sold tickets.

For the first time this year, we include in the station usage file information from joint stations that have only an LUL ticket office. These were excluded in previous years because not all sales were captured in LENNON. For this year we confirmed that all the sales that concern any rail travel appear in LENNON so we have included these stations. Another category of joint stations are the stations with only a TOC issuing office. More details on how we analyse these stations can be found in section 6. The following table is a list of the stations that were not included in the station usage file last year, but they do appear into this year file, with the reason why.

**Table 3-3: New Stations in the Station Usage for 2006/7.**

Station Name	Why?	Station Name	Why?
Highbury & Islington	Joint	Liverpool South Parkway	New
Tottenham Hale	Joint	Larkhall	New
Seven Sisters	Joint	Kelvindale	New
Farringdon	Joint	Manchester United FC Halt	New
West Ham	Joint	Reddish South	New
Old Street	Joint	Denton	New
Harrow-on-the-Hill	Joint		
Amersham	Joint		
South Ruislip	Joint		
Greenford	Joint		
Chalfont & Latimer	Joint		
Rickmansworth	Joint		
Barbican	Joint		
Chorleywood	Joint		

### 3.2 Overview of the Interchanges Results

In all, around 133 million interchanges are estimated to have been made among National Rail operated services (interchanges between rail and tube or other modes is excluded). This is an increase of about 11% compared to the 2005/6 results (120 million). Again the main contribution to that big increase was due to the inclusion of TfL issued Travelcards. 46 million interchanges were estimated to have occurred at the top 10 stations, about 35%, a similar proportion to last years results. These ten stations are listed in the table below.

**Table 3-4: Top 10 stations based on the interchanges made for 06/07.**

Station Name	Total Interchanges 06/07
Clapham Junction	12,898,160
London Bridge	7,768,209
East Croydon	5,122,701
London Waterloo	4,226,542
Birmingham New Street	3,794,980
London Victoria	3,293,684
Leeds	2,836,231
Reading	2,688,110
London Kings Cross	2,100,857
London Euston	1,956,634

Interchanges occurred at 510 stations in 2006/7, similar to the 505 stations in 2005/6. The 5 new interchange stations appearing since 2005/6 are listed below, and all of them are joint stations that are included in the Station Usage file for first time.

**Table 3-5: New Interchange Stations in 2006/7.**

Station Name	Total Interchanges 06/07
Farringdon	214,544
West Ham	138,198
Tottenham Hale	108,386
Highbury & Islington	79,370
Seven Sisters	6,734

We did not identify reasons for changes in the interchanging stations. However, it is important to note that interchanges can change significantly from year to year for a variety of reasons. Factors such as new service patterns and changes in journey times play a part. Each may alter the opportunities to travel in ORCATS and may include an opportunity to interchange at a station, which was not previously feasible. Please refer to Appendix 2 for more information on ORCATS allocation process and the CAF.

## 4 Station Usage File Definition

This spreadsheet lists the entries, exits and interchanges made at stations throughout England, Scotland and Wales in the financial year 2006/7 (1<sup>st</sup> April 2006 to 31<sup>st</sup> March 2007). It also gives details about the entries and exits for different ticket categories. It contains data on entries and exits made at Rail stations by passengers using the rail network.

The fields included in the Station Usage file are:

**Table 4-1: Station Usage File**

Field	Description
Station (Name, NLC, TLC)	Station Name, NLC: National Location Code, TLC: Three Letter Code
District, County and Region for the Station.	Station's geographical location.
Station Facility Owner (SFO)	The company that is the station facility owner (provided by Network Rail). Note that this refers to the current SFO, not the SFO in 2006/7.
Group of Stations	Name of the Group where applicable
Entries (Full, Reduced, Season, Total)	Entries made at the stations split by ticket categories and in total
Exits (Full, Reduced, Season, Total)	Exits made at the stations split by ticket categories and in total
06/07 Entries & Exits	Sum of Entries and Exits for 2006/7
06/07 Entries & Exits Total	Sum of Entries and Exits for 2006/7
06/07 Interchanges	Total Interchanges made for 2006/7
Change	Compare 2006/7 Entries & Exits against 2005/6
Check 1	Fail when either entries or exits individually constituted less than 40% of total entries and exits
Check 1 Reason	Identified reason(s) for failing the check
Check 2	Fail if number of entries & exits combined is more than 20% higher or lower than figure in 2005/6
Check 2 Reason	Identified reason(s) for failing the check

Note: Compared to last year changes have been applied to the tickets classification methodology. As a result there has been a slight shift from ticket category "Full" to ticket category "Reduced" for the entries and exits figures.

## 5 Data Excluded From the Study

Some of the LENNON data was excluded from the analysis, and subsequently from the Station Usage data.

All the products that were classified as miscellaneous tickets were excluded.

These products were:

- Rover Tickets
- BritRail Tickets
- Car Parking,
- Railcards Sales,
- Penalty/Excess Fares,
- Seat Reservations,
- Sleeper Supplements.

Also from the analysis were excluded all the flows that had either Origin or Destination an "I code", usually some form of special pass that cannot be identified with a station, e.g.

- Freedom Passes
- West Midlands Zones
- Merseyrail Zones
- Centrocard
- Gate passes usually used by staff
- Passenger Charter Discounts
- Headquarters Items

Finally flows that have either Origin or Destination a Private Settlement Code were excluded from the analysis, e.g.

- PTE tickets
  - WYPTE Metrocard
  - GMPTE Traincard
  - Anglia Plus
- Buses
  - Canvey Bus Zone,
  - Norwich City Bus Link,
  - St Albans + Bus
- Places of Interest
  - Leeds Castle
  - Alton Towers
  - Bluewater through ticket
- Irish Stations
  - Dublin
  - Belfast
  - Dun Laoghaire

In total the data we have excluded from the analysis account for around 6% of the total revenue and 9% of the total journeys.

## 6 Entries and Exits Methodology

We calculated an estimate of the number of people entering and exiting each of the National Rail stations for the financial year 2006/7. The entries and exits at each station were derived from the journey data recorded in the LENNON database.

Each year DeltaRail receives a file from Atos Origin containing the revenue, number of journeys and ticket issues for each flow on the network, for each product. These revenue and journey figures are unallocated, in other words are based on the actual ticket sales before the allocation to the different Train Operating Companies.

This year, we tried to improve the accuracy and usefulness of the results shown in the Station Usage file. This was achieved by applying new procedures on the way we treated journeys with unknown origin and/or destination and by including journeys that were previously excluded from the file or do not appear into LENNON sales data. In summary, the main changes were:

- Adding in previously missing journeys, e.g. TfL sold Travelcards, and some airport link tickets
- Estimating the split of records for station groups, including London BR, into the constituent individual stations
- Dealing with Travelcard “Joint Stations” in a more helpful way, i.e. those stations where the ticket office serves both LUL and national rail stations.

After completing checks on the data, the data was transformed into an origin/destination file, where each record reflected an estimate of the actual journeys undertaken. The file contained the number of journeys for each flow, where a flow consisted of a unique origin, destination and route code combination.

The number of entries and exits was calculated for a particular station by summing all journeys starting at the station, and all journeys terminating at the station.

### 6.1 Journeys added in this year

The largest source of missing journey information for the Station Usage 2005/6 file was the London Travelcards. TfL issued London Travelcards appear in LENNON but usually as a lump sum using codes for Headquarters Items to a travelcard destination. These products were excluded from our LENNON derived dataset in previous years. For the 06/07 file, we analysed these sales using the Travelcard Diary Survey (TDS). The same analysis using the TDS was applied to Travelcards sold at TOC-run stations in Zone 1.

Another missing source of journeys are the Airport Links. Some sales made directly by airlines are not appearing in LENNON, but for these 2006/7 results such journeys have been added in to the data.

### **6.1.1 Travelcards Issued by TfL or sold at TOC-run stations in Zone 1**

For this year the largest improvement we applied to the Station Usage process was the analysis on travelcards issued by TfL or sold at TOC-run stations in Zone 1. The “Infill” matrix is consistent between Station Usage & the O-D Matrix, also produced for ORR.

We firstly identified from the LENNON derived dataset the products and the sales that were of our interest (Travelcards sold by TfL, or at TOC stations in Zone 1). After obtaining the sales figures we then analysed the revenue based on the Travelcard Diary Survey (TDS), to split into TOC and Zonal validity. In the next step, we used LATS 2001 to split the TOC / Zonal validity pots of revenue into flows, giving a percentage of revenue on specific flows. Then we applied these percentages to the aggregate revenue data for such categories of Travelcard, from the original LENNON derived dataset, to give revenue infill.

At the last step we used TDS to derive the revenue per journey for each TOC type and zonal validity combination and apply them to revenue infill figures to give the journeys infill figures.

### **6.1.2 Airport Links**

Rail links to airports are increasingly important in terms of passenger numbers. Rail links to Stansted, Gatwick and Heathrow airports carry significant numbers of passengers. However, not all tickets for rail links to airports are captured in LENNON and so need to be specially added in to the Station Usage file. Tickets not issued at rail stations are not captured, so we had to approach the operators of the three rail links to Heathrow, Gatwick and Stansted to obtain details of the tickets issued.

We finally received data from the two of them, Stansted and Gatwick Express. This enabled us to add journeys made for the flows Stansted Airport to/from London Liverpool Street and Gatwick Airport to/from London Victoria. However, it should be noted that Heathrow Express is excluded from all Station Usage figures, due to the data not being made available.

## **6.2 Unknown Destinations**

Ticket sales do not always tell us where a passenger is travelling. Ticket sales can be divided into the seven categories listed in table below. We have converted the ticket sales data from LENNON into an estimate of the actual stations that passengers are travelling to. This methodology was already employed in the creation of the Station Usage information in previous years, but for the 2006/7 file we applied some changes to improve the accuracy of the information.

**Table 6-1: Categorisation of Ticket Sales in LENNON**

<b>Flow Category</b>	<b>Description</b>
Category 1	Origin and Destination Stations Known
Category 2	Origin or Destination a Group Station (excl. London Terminals)
Category 3	Origin or Destination is London Terminals
Category 4	Origin or Destination a London Travelcard including Zone 1
Category 5	Origin or Destination a London Travelcard excluding Zone 1
Category 6	Origin or Destination a London Travelcard Boundary Zone
Category 7	Non-National Rail Stations

Category 1 – Origin and Destination Stations Known

Both the origin and destination were known stations so no further work was needed for such flows.

Category 2 – Origin and/or Destination a Group Station (excluding London BR)

For this category, the methodology has been changed compared to last year. In last year’s Station Usage information, all origins or destinations that were a group station (with the exception of London BR) were changed to the major station within the group. For example, all ticket sales to or from Manchester BR were recoded to Manchester Piccadilly. However, the previous methodology does not always reflect the actual passenger movements. With the current methodology we believe that the information provided is much improved. This method is more accurate than the basic set of assumptions used in previous years.

For 2006/7 Station Usage, the methodology is based on the split of ticket sales from the individual stations within a group. Generally for destinations, LENNON records ticket sales as being to the station group. However, where the station group would be the origin, tickets sold at one of the individual stations will appear with that specific station as origin. By assuming similar patterns of travel to and from that town or city, we can infer the likely split of station group data. For each flow involving a group station we split the number of journeys to the group station in line with the number of journeys made to the stations within the group. We assume that passengers travelling **to** the stations in a group will act in the same way as passengers travelling **from** the stations in a group. We believe that this is a valid assumption to make, and no bias will be introduced into the journey figures.

We illustrate the methodology with an example. In 2006/7 the journeys to or from Manchester stations (Manchester BR is the group station) and Crewe are shown by the column “jnys” in the table below. We estimated the split of journeys (column “ BR portion”) to/from Manchester BR (the group station) based on the split of journeys from Manchester Oxford Road, Manchester Piccadilly, Manchester Victoria, and Deansgate to Crewe (column “%split”). So the column “Total jnys” shows how the journeys figures will be after re-allocating the two flows Crewe to/from Manchester BR.

**Table 6-2: Example of how we split journeys to/from a BR group of stations**

Orig	Dest	Origin Name	Destination Name	Jnys	%split	BR portion	Total Jnys
2963	1243	DEANS GATE	CREWE	83	0.32%	85	168
2966	1243	MANCH OXF RD	CREWE	5,464	21.03%	5580	11,044
2968	1243	MANCH PICC	CREWE	19,733	75.95%	20152	39,885
2970	1243	MANCH VICT	CREWE	700	2.69%	714	1,414
0438	1243	MANCH BR	CREWE	26,533	<b>Remove</b>		
1243	2963	CREWE	DEANS GATE	207		1478	1,685
1243	2966	CREWE	MANCH OXF RD	2,262		97287	99,549
1243	2968	CREWE	MANCH PICC	8,017		351349	359,366
1243	2970	CREWE	MANCH VICT	343		12464	12,807
1243	0438	CREWE	MANCH BR	462,578	<b>Remove</b>		

For all flows with more than 1,000 journeys in total leaving the Group stations (not including the BR Group NLC to the destination figures), the above methodology has been employed. For all other flows an average split between the stations within the group has been calculated based on the flows with more than 1,000 journeys.

#### Category 3 – Origin or Destination was London BR

This category contained all flows that had London BR as either the origin or destination. In order to assign an appropriate London station on flows where either the origin or destination is London BR (NLC=1072), we analysed responses from the 2001 LATS survey. For journeys from any given station, we found out what percentage of passengers use each London terminus.

For example, if the flow was from Ashford International to London BR, we used our pre generated table showing the percentage split between the alternative London termini for passengers starting at Ashford International. From this we apportioned the exits between London Bridge, Charing Cross, Victoria and other London termini.

Stations with small sample sizes were removed from the 2001 LATS data. Where there was insufficient data in the 2001 LATS to generate the split for a particular station, a similar process with the Non London Groups methodology was applied. Firstly for all the flows with more than 1000 journeys leaving London BR and having as a destination the particular station we used split factors as above. However, if the sum of journeys was less than 1000 we assigned to the flow the top origin from the London BR stations. This is different to the methodology applied the previous years, where the TOC operator at the station was the factor used to determine which London terminal the passenger start or end their journey.

#### Category 4 – Origin or Destination a London Travelcard including Zone 1

All origins and destinations that were London Travelcard Zones that include Zone 1 were converted to 'London BR' under the assumption that they will travel to the same stations as point-to-point passengers and then transfer to another mode. The methodology set out above for Category 3 was then applied.

Before applying the above methodology, we also used MOIRA<sup>1</sup> assumptions for the joint stations with only a TOC issuing office, in order to split the flows into a rail and a LUL portion. This part was necessary since for these stations, LENNON shows all the sales, while only a proportion of those passengers will be travelling with National Rail.

So if a flow was showing Finsbury Park to Z1234, then we firstly split the journeys, based on MOIRA assumption into Rail and LUL portion. Then we continued for the rail part of the flows applying the above methodology, as detailed for category 3. The LUL portions of these flows have been left out of the Station Usage figures.

#### Category 5 – Origin or Destination a London Travelcard excluding Zone 1

This category contained all Travelcards that did not include Zone 1, for example Zone R2345 London.

For flows with origin or destination a London Travelcard (excluding zone 1) we use a set of assumptions used in MOIRA to convert the Travelcard into a destination station. These assumptions have been created using survey responses from the 2001 LATS. They use the starting station to work out which stations it is possible for the passenger to be travelling to, and also give the proportion of passengers travelling to each of these stations. This is based on the assumption that a passenger holding a Zones 2-6 Travelcard would travel as far as Zone 2. Also for this category when the origin was a joint station with only a TOC issuing office we firstly split the journeys between Rail and LUL, before applying the MOIRA assumption to split the journeys

For example, if a passenger is travelling from Walthamstow Central to London Z2356, the passenger can travel as far a zone 2. The MOIRA assumptions state that a proportion of travellers will use the Rail services and a proportion will use the LUL services depending on what ticket type (full, reduced, season) they hold. Having obtained the Rail proportion, we then use again MOIRA assumptions based on LATS 2001 to assign a rail station as a destination, instead of the travelcard destination. For the Walthamstow Central to London Z2356 example, MOIRA suggests that passengers can alight at Hackney Downs and Bethnal Green.

Where there was insufficient data in the 2001 LATS to generate the split for a particular station we keep the flows as they appeared in the LENNON sales data.

#### Category 6 – Origin or Destination a Boundary Zone

For all flows where the origin or destination was a boundary zone in London, we do not know where the passenger started or ended their journey, therefore such records are excluded from the Station usage file.

Passengers purchasing a ticket starting at a boundary zone will already be in possession of a ticket that is valid out to that boundary zone. We were unable to tell at which station they started their journey, as they used the ticket that they already owned to travel from their starting station to the boundary zone.

---

<sup>1</sup> MOIRA: A model, supported by DeltaRail, to forecast revenue changes due to timetable changes.

For example, a passenger with a zone 1 and 2 Travelcard who wanted to travel from London to Portsmouth would purchase a ticket from boundary zone 2 to Portsmouth. However, we had no way of knowing at which station they started their journey. It could have been any station within Travelcard zones 1 and 2. There are five different boundary zones: Boundary Zone 2, 3, 4, 5 and 6.

#### Category 7 – Non-National Rail Stations and Invalid Flows

This final category contains all those flows that appeared to be invalid, and therefore for which the journeys are excluded from the entries and exits figures. This includes:

- Invalid flows, in particular flows where the origin station equalled the destination station.
- Flows involving invalid codes that were not proper stations.

## **7 Interchanges Methodology**

We obtained an estimate of the number of people interchanging at each station by combining the number of journeys made on each flow from Lennon (from our origin/destination file) with the information in the Central Allocations File (CAF).

The CAF is an output of the ORCATS system used to determine the allocation of passenger revenue between TOCs. Since ORCATS is a model, the CAF contains estimates rather than actual journeys. However, it is used throughout the rail industry, so it is an appropriate source of data to use for this purpose. Since CAFs are updated with the timetable, not with financial years, no CAF will match the LENNON data exactly. We had proposed to use the December 2006 CAF in this study, but it was not available. Therefore we used an older source of data, the December 2005 CAF which relates to the timetable in operation for substantial portion of the 2006/7 financial year covered by the LENNON data.

The CAF contains:

- Origin and destination
- Route alternatives for each origin and destination
- Ticket type data
- For each flow, the proportion of passengers who choose to travel on each route alternative as calculated by the ORCATS model

This methodology was the same as that which we employed to calculate the number of passengers interchanging at each of the stations in 2004/5 and 2005/6. An overview of the ORCATS allocation process can be found in Appendix 2.

## **8 Regions, Counties and Districts**

For all the proper rail stations, we received a file from ATOC, showing the District or the Unitary Authority where the station is located. This file was also cross-checked with another file that ORR provided us. After having that information for each station, we also download data from the ONS<sup>2</sup> website, in order to match each district to a county and a region (GOR).

---

<sup>2</sup> [http://www.statistics.gov.uk/geography/geographic\\_area\\_listings/administrative.asp#04](http://www.statistics.gov.uk/geography/geographic_area_listings/administrative.asp#04)

## **9 Limitations of LENNON data**

The LENNON database captures ticket sales for the entire national rail network from many different input machines. It is as a consequence a very large data set. With all large data sources there will always be input errors resulting in a certain amount of invalid data. Generally such errors will be small, and are more likely to occur in the journeys rather than revenue fields.

We perform checks on the data, but due to the size and complexity of the dataset we are not able to validate each and every entry.

We have used similar information extensively in the last ten years or more, and have found the data to be reliable, particularly when examining the data at an aggregated level.

There are a number of areas of known weaknesses of LENNON when used for the purpose of estimating passenger journeys actually made. These areas are expanded upon below.

### **9.1 Known Problems of Data Capture**

The data in LENNON from which the O-D Matrix is derived is based on ticket transactions. In order for the data to be included in the O-D Matrix it must include an origin station and a destination station. However if this is not the case then the data will automatically be excluded.

Human error at the point the ticket sale is entered into the input machines will produce invalid data in LENNON.

### **9.2 Travelcards**

As Travelcards are for multi-modal travel they allow the purchaser to make journeys on the rail system and on other modes. Equally, tickets purchased elsewhere on the local transport system will be valid for rail travel. Therefore LENNON and hence the Station Usage give only a partial picture of the rail travel in conurbation areas. Such areas include: London, Birmingham, Glasgow, Leeds, Liverpool, Manchester, Newcastle and Sheffield.

Most Travelcards apply to a broad zonal system. This means that it is not possible to define the specific origin or destination station for the data that is included in LENNON.

In summary, this year's Station Usage contains reasonably robust estimates of journeys within London, but not for other conurbation areas where travelcards are widely used. We are investigating how such information could be improved in the future.

### **9.3 Return and Single Journey Tickets**

It is possible that on certain routes the cost of a return ticket could be only slightly more than (or in some cases, lower than) a single ticket. This leads to the cheaper return ticket

being purchased even though the passenger has no intention of making the return journey by rail. This results in two journeys being recorded instead of one.

#### **9.4 Multiple Tickets**

It is possible to buy special cheaper tickets between certain stations for example under a promotion by one of the train companies. In these cases a local ticket may be bought to gain access to a main station and a second ticket bought for the rest of the journey. This results in 2 entries and 2 exits being recorded in the Station Usage, instead of 1 entry, 1 exit and 1 interchange, therefore will not accurately represent the journey undertaken.

#### **9.5 Rail Staff Passes**

Prior to the privatisation of the rail network, British Rail employees and their families were eligible to various levels of free or reduced rate rail travel. When the various rail companies were converted to private companies, this benefit often continued.

If you consider the network as a whole, the effect of staff passes is unlikely to be significant. However, it may be significant on certain routes, for example on routes out of Derby due to large concentration of companies in Derby relating to British Rail both pre and post privatisation.

#### **9.6 Ticketless Travel**

On every route on the network there will always be passengers who travel without purchasing a ticket. This is referred to as ticketless travel. As LENNON data is derived from ticket transactions it cannot reflect this travel.

#### **9.7 Other Rail Systems**

There are a number of rail systems in operation in the country that are not covered by LENNON. For Heathrow Express and Eurostar revenue and journeys data were not available.

In addition some tickets of franchised operators may be sold outside LENNON. However, note that for the main two such cases, Gatwick Express and Stansted Express, data has been included this year.

#### **9.8 Factors Affecting the Data**

Although not relevant for this year, there are factors worth taking into account when considering generic annual data:

- Years may have been affected by industrial action such as 1994/95.
- Major incidents affecting services such as Southall, Ladbroke Grove and Hatfield.
- Adverse weather.
- Infrastructure changes e.g. ticket gating significantly increases revenue - more gates have been installed in recent years which will affect the data but which does not represent higher passenger numbers.

### 9.9 Journey Factors

Ticket transactions are converted into an estimate of the number of journeys made by applying a series of ticket type journey factors. Single and return tickets unambiguously translate into one and two journeys respectively. For season tickets, the factors used represent a rough historic estimate as follows:

**Table 9-1: Journey Factors used in LENNON**

Description	Number of Journeys Per Issue
Single Journey Ticket	1
Return Journey Ticket	2
Return Journey 2 Persons	4
3 Day Return/ 6 Single Journeys	6
4 Day Return/ 8 Single Journeys	8
5 Day Return/ 10 Single Journeys	10
6 Day Return	12
5 Day Single	5
1.5 Journeys	1.5
Weekly Ticket	10.3
10 Day Return/ 20 Single Journeys	20
2 Weekly Ticket	22
Seasons-Variable Periods	***
Monthly Ticket	45
Not Used	0
3 Monthly Tickets	135
Not Used	0
6 Monthly Tickets	270
Summary Group Codes	***
Annual Ticket	480
8 Day Ticket	22
22 Day Ticket	44
14 Day Ticket	30
50 Journeys	50
10 Weeks	103
No Journeys	0

Ticket periods of other lengths are converted to a number of journeys using a proportion of the monthly journey factor.

Therefore the journeys data in the Station Usage represents an assumed number of journeys made, based on the ticket type sold and the above journey factors. In particular it should be noted that the journeys data has not been cross-checked against other data sources of the actual number of journeys made on the network.

These journey factors have been used within the LENNON system for a number of years at their current values. The source of the factors is unclear. It can therefore be argued that these journey factors do not provide an accurate estimate of the number of journeys that result on the rail system at present. They do however, have the advantage that they have been applied consistently within LENNON (and the previous system CAPRI) over the years.

## **Appendices**

### **Appendix 1 - Checks**

Three specific checks on the data have been carried out. All the stations flagged up by the first two checks are marked in the accompanying spreadsheet, together with explanations, where possible. In this section, we explain these checks in more detail. We also performed a series of other checks to help validate the results.

#### **Check 1: Unequal Entries & Exits**

The first check highlights all the stations at which the number of entries differed significantly from the number of exits. A station was considered to fail this check if either its entries or its exits individually constituted less than 40% of its total entries and exits.

In 2006/7 36 stations failed this test. This number accounts to 1.5% of all the stations, with the total number of journeys being only 0.02% of the total journeys. 11 out of the 36 stations had more than 15,000 entries and exits for the year, and only one of them had more than 50,000 entries and exits. This was the Prestwick Airport, where in such special stations we expect that this may happen. For the other stations with over than 15,000 we checked why this happened, and it appeared to be that more single journeys were made for the one direction.

#### **Check 2: Large Changes in Usage**

The second check identified all stations at which usage has changed significantly since the previous year available, 2005/6. A station is deemed to fail this test if its number of entries and exits combined is more than 20% higher or lower than its figure in 2005/6. The 20% takes into account a possible continued growth or decline over the previous year. Only stations whose total entries and exits exceeded 15,000 were considered in this check.

In total 476 stations failed that check. That is 22% of the stations whose total entries and exits exceeded 15,000. For this year we were expecting many London stations to fail that check because of the addition of TfL issued travelcards into the final data, which adds around 200 million journeys. Also we were expecting changes in the stations entries and exits because of the different methodology we applied for the group of stations and the joint stations. If we exclude those stations for which we expected changes because of the above reasons, then 143 stations that failed the check. This is about 7% of the stations with more than 15,000 entries and exits, and is similar to the previous year figure, when 181 stations (7%) failed the check.

There are many other reasons why stations can fail this check, in particular that demand may have increased or decreased significantly at individual stations. Reasons causing stations to fail the check, including those provided in previous reports, have been listed below for completeness. In addition, where possible a reason was assigned to the stations failing the check. A reason was not identified in every case.

### **New Station**

Stations that are newly opened, or have been opened within the past few years, can reasonably show significant growth. Such an example is Glasshoughton, which opened in February 2005, and for which the 2006/7 station usage is more than double 2005/6.

### **Gating**

Installation of ticket gates can significantly affect not only the usage figures at that station, but also those at neighbouring stations. The gates help to ensure that customers purchase tickets, but customers may also alter their travel patterns to avoid gated stations. We would expect travel patterns to be most affected in the months following the installation of the gates.

### **Change in Service Pattern**

Alterations in service frequency or stopping pattern would be expected to alter station usage figures. This is particularly apparent where a group of consecutive stations show similar increases or decreases. Again, this can be a long-term trend.

### **Ticket Issuing Facilities Changes or Product Changes**

Some London stations have both underground and National Rail trains operating. Lennon does not capture tickets sold by London Underground, only those sold by TOCs. Changes in ticket facilities provided by TOCs, for example the provision of ticket machines, can therefore increase the ticket sales captured by the system.

Product changes can have an affect on passengers' purchasing patterns at rail outlets thus affecting station usage data. For example, the introduction of Oyster cards at rail outlets can affect stations inside the Travelcard boundary in the London area.

### **Engineering Work**

Significant engineering work can alter customers' travel patterns.

### **Tourism**

Stations near to tourist attractions may show significant changes in usage as a result of weather, promotions or other factors, which affect tourists' journeys.

### **Special Stations**

Some stations serve a particular activity or business. Some fluctuation in usage of such stations is reasonable. Such activities include:

- Racecourses
- Sports Events
- Exhibition Centre Glasgow
- Airports

### **Trend of Growth or Decline**

For stations with a history of growth or decline, it is reasonable to expect this trend to continue. There are many possible reasons for these trends, such as demographic and employment changes.

### **Changes in the Sales of Individual Ticket Types**

Miscoding of ticket information entered into Lennon can alter station usage results, although this would not be reflecting an actual change in customers' journeys.

### **Check 3: Group Stations & London Termini**

The third check involved identifying all stations that are part of a group station

#### **Group Stations**

All stations that are members of group stations were identified. Our processes will have assigned these passengers to a specific station within the station group. They are marked in the spreadsheet with the name of the group station. The user of this data may wish to filter on 'Group Station NLC' column, or create pivot tables, to investigate the results at a group station level. For these groups of stations, as mentioned earlier, we were expecting to have a lot of changes compared to last year. That is because of the new methodology we applied to them in order to improve the accuracy of our results. For most of the group of stations, we are confident that the new methodology has improved accuracy in reflecting actual passenger movements. However, there are three particular group stations for which the results should be treated with caution.

These are Gainsborough stations, Pontefract stations, and Tyndrum stations. All three groups of stations are non-staffed stations that lead to the problem that very few tickets were issued showing from which stations passengers are leaving. As a result the results appearing in the station usage, based on our standard methodology, may not reflect how passengers travel. These are, however, only a small proportion of the group stations, and are all low volume stations. For these particular stations, we would recommend using total journeys among the group, rather than relying on individual station results.

## **Appendix 2 - Overview of the ORCATS Allocation Process**

This section gives an outline of the Central Allocations File (CAF), which is used in producing the interchange figures, and the ORCATS process which is used to create the CAF.

Most of the train tickets that are sold are interavailable – the customer has a choice of routes and operators. For example, when a customer buys a ticket to travel from Leicester to Leeds, that customer may travel on various combinations of Midland Mainline, GNER and Virgin Trains, and may interchange at Doncaster, Sheffield or Derby. Lennon captures the sale of the ticket, but unless the ticket has stringent route restrictions, the route actually taken by the customer is not recorded.

The route taken by any particular customer may never be known, but some route options are more attractive than others. The customer is more likely to choose a faster, more frequent service than a slower, less frequent one. This likelihood can be translated into the proportions of customers choosing each route option, on a particular flow. (A 'flow' represents all journeys from a given origin station to a given destination station, irrespective of the route taken.) The revenue received from all customers on that flow should be split between different operators to reflect the proportion of customers which each operator carried.

ORCATS was developed to model the choice made by the customers, and to allow revenue to be split between operators. It applies passenger choice modelling to the train timetable, to determine the relative attractiveness of different route alternatives. It then weights the results by journey mileage.

For any given timetable, ORCATS works out the possible routes between each origin and destination, and calculates the percentage of the passengers that are expected to choose each route based on the services in that timetable.

The output from ORCATS is the Central Allocations File (CAF). This file lists the proportion of journeys on each flow (or origin-destination pair) estimated to be made by each route alternative. For journeys involving interchanges, each leg of the journey is listed. By combining this information with Lennon data, which contains actual ticket sales figures for all flows, we have estimated the number of interchanges occurring at individual stations.

