Automating the Import of Electronic Timetable Data to EMME/2-Based Public Transport Models

John Armstrong, John Preston
Transportation Research Group, School of Civil Engineering and the Environment,
University of Southampton, Highfield, Southampton SO17 1BJ, UK
e-mail: j.armstrong@soton.ac.uk j.m.preston@soton.ac.uk

Matt Carlson
Arup, 13 Fitzroy Street, London W1T 4BQ, UK, e-mail: matt.carlson@arup.com

Abstract

One of the speakers at RailDelft2005 issued a "plea for standardised models" for the simulation of railway operations. This issue is highly relevant in Britain, where the industry-standard CIF (Common Interface Format) timetable data files are not directly compatible with modelling tools such as RailSys and EMME/2. RMCon has developed a CIF import facility for RailSys, but no equivalent exists for EMME/2, which is used in the PLANET and Railplan multi-modal public transport models employed by the Department for Transport and Transport for London, respectively.

This issue was highlighted during passenger demand forecasting work conducted by Arup for Network Rail (Britain's railway infrastructure owner and operator) for the South West Main Line Route Utilisation Strategy (SWML RUS), a strategy for the utilisation and development of the rail routes between London's Waterloo station and the area to the south-west. The PLANET South model was used for demand forecasting, using timetable data imported from a CIF file supplied by Network Rail.

The issues encountered in importing the CIF data to PLANET led to the development of a Perl-based tool to automate the process, in a collaboration between Arup and the University of Southampton's Transportation Research Group. The success of this tool resulted in Arup being commissioned by Transport for London to develop a similar tool for Railplan, which simulates a less extensive but more detailed network than PLANET. The latter tool is now working and in use, although some aspects of it may yet be developed further.

Keywords
Timetable, Data, Import, Conversion, EMME/2

1   Introduction

This paper describes the issues underlying the import of UK-standard CIF-formatted railway timetable data to EMME/2-based models, and the development of tools to automate the process. Following this introduction, the second section of the paper describes the background to the issue and the reasons underlying the development of the tools. Section 3 then describes the development of the tools. Section 4 outlines areas where there is scope for further improvements, while Section 5 presents some

conclusions, and is followed by a reference list.

2   Background

2.1   EMME/2, PLANET and Railplan

The EMME/2 multi-modal transport modelling software package is well-established and widely-used. It forms the basis of the PLANET and Railplan multi-modal public transport models owned and maintained by the UK's Department for Transport (DfT) and Transport for London (TfL), respectively.   There are three PLANET models: PLANET North, PLANET South and PLANET Strategic, with respective emphases on London and the South-East of England, Northern England and Britain as a whole.

Public transport services are represented in EMME/2 by 'transit lines', which may be either timetable- or frequency-based, although both approaches represent an underlying timetable of services. Service frequencies and times typically change more frequently than does the underlying transport infrastructure, and the updating of transit line data is thus a relatively frequent requirement.   This process is time-consuming (and thus expensive), laborious and error-prone, and is thus well-suited for automation where this is possible.

2.2   Modelling the Brighton and South West Mail Lines in PLANET

Brighton Main Line

Arup was appointed by the (since abolished) Strategic Rail Authority to forecast demand on the Brighton Main Line (BML), using the PLANET South model.   The BML, originally built and operated by the London, Brighton and South Coast Railway, is one of the main (and the most direct) rail routes between London and England's south coast, operating between London's London Bridge and Victoria stations, Brighton and the south coast, and points between, including East Croydon, Gatwick Airport and Haywards Heath.

SPG-formatted data, derived from MOIRA (a timetable-based model of passenger rail transport supply and demand, used to predict demand and revenue resulting from timetable changes), was used to generate the transit lines.   An example of this data is shown below:

```
…
T
IVICM0 497  497N6660EMU  100mph
ICLJM0 502  503N6660EMU  100mph
IBALM0 506  506P6660EMU  100mph
ISRSM0 510  510P6660EMU  100mph
IECRM0 513  513N6660EMU  100mph
IPURM0 517  517P6660EMU  100mph
IELDM0 524  524P6660EMU  100mph
IGTWM0 529  530N6660EMU  100mph
ITBDM0 533  533P6660EMU  100mph
IHHEM0 541  545N6660EMU  100mph
IPRPM0 557  557P6660EMU  100mph
IHOVM0 560  561N6660EMU  100mph
ISSEM0 567  567N6660EMU  100mph
IWRHM0 573  574N6660EMU  100mph
```

```
IWWOM0 575 575P6660EMU 100mph
IANGM0 580 580N6660EMU 100mph
IFODM0 586 586P6660EMU 100mph
IBAAM0 590 590N6660EMU 100mph
ICCHM0 598 598N6660EMU 100mph
IHAVM0 609 609N6660EMU 100mph
ICSAM0 616 616N6660EMU 100mph
IFRMM0 624 625N6660EMU 100mph
ISNWM0 631 631N6660EMU 100mph
ISDNM0 643 643P6660EMU 100mph
ISOUM0 651 651N6660EMU 100mph
IHHEV1 541 547N6180EMU 100mph
IWVFV1 551 551N6180EMU 100mph
ILWSV1 562 563N6180EMU 100mph
IPLGV1 575 575N6180EMU 100mph
IEBNV1 582 582N6180EMU 100mph
T
IECRM0 498 498N6610EMU 100mph
…
```

The letter T indicates that the following lines of data (all starting with the letter I) represent a timetabled train. The three letters after the letter I are a station code, and are followed by a two-character code representing a train 'section' (see below). Two numbers of up to four digits follow, describing the arrival and departure times of the train at the current station in terms of the number of minutes after midnight. The letter following the second time value describes the stop type: normal (i.e. the train sets down and picks up passengers at the station in question), pick-up only or set-down only, and the remaining data comprises a MOIRA code and rolling stock characteristics.

The data shown describes a train leaving London Victoria at 08:17 and running to Southampton Central (section M0), with a portion (section V1) splitting from it at Haywards Heath (HHE) and running to Eastbourne. Any sections that subsequently split from the main section would be labelled V2, etc., while subsidiary trains joining a main train have section labels J1, J2, etc.

The regular, column-based SPG format makes such data easy to import to spreadsheet and database applications, and a Microsoft Access-based tool was used to generate transit lines for use in PLANET South, employing a combination of a sequence of complex queries and a VBA macro. This process, while a major improvement over manual transit line coding, was found to be quite complex and still somewhat cumbersome, and prompted consideration of the development of software to process the SPG data and generate the transit lines directly.

South West Main Line
Arup was subsequently appointed by Network Rail to produce forecasts of demand for the South West Main Line (SWML), again using PLANET South. The SWML, originally built and operated by the London and South Western Railway, carries train services between London's Waterloo station and the area to the south-west, including Reading, Woking and Guildford, the south coast between Portsmouth and Weymouth, and Salisbury and Exeter in the south-west of England.

To generate transit lines representing the then-current timetable on the route, a CIF

(Common Interface Format) file was provided by Network Rail. The supplied file contained almost 1.5 million lines of data, representing over 79,000 timetabled trains, meaning that simply identifying and extracting the required data was less than straightforward. Furthermore, the format of CIF data, an example of which is shown below, renders it impractical for direct import to spreadsheet or database applications, so the data was 'pre-processed', using a PHP script to extract and convert the relevant data into a format consistent with that obtained from the SPG files, after which it was again processed using the Access-based tool. While this approach was ultimately successful, it was time-consuming and even more cumbersome than before, and it thus prompted further consideration of the development of a specific software tool for the import of SPG- and CIF-based data to the suite of PLANET models. Following discussions between Arup and the University of Southampton's Transportation Research Group (TRG) and School of Electronics and Computer Science, it was decided to proceed with the development of such a tool, jointly funded by Arup and TRG, using the Perl programming language. The following extract from the supplied CIF file represents a weekday evening peak service from London Waterloo to Southampton Central:

```
...
BSNW67323041213050610111100 PXX1B939293124620204ZEMU
BX          SWY
LOWATRLMN 1748 174811 MFL     TB
LICLPHMJM          1754 000000009  FL
LIWDON             1757H000000007
LINEWMLDN          1759H00000000
LISURBITN          1801H00000000      FL
LIHCRTJN           1802H00000000      FL
LIWOKING  1811 1813      181118134      T
LIWOKINGJ          1814 00000000
LIFRBRMN           1820H00000000
LIBSNGSTK 1831H1833H      183218332       T
LIWRTINGJ          1837 00000000
LIWNCHSTR 1849H1851      18501851       T
LISHAWFDJ          1854 00000000  FL
LIELGH    1858H1900      185919002       T
LISOTPKWY 1903 1904H      19031904       T
LISTDENYS          1910 00000000  FL
LINTHMJN           1911 00000000
LTSOTON   1914 19163B     TF
...
```

The first line of data includes the train's unique identifier, information about its dates and days of operation, its type (express passenger) and its constituent rolling stock. The second line provides information about the Train Operating Company, or TOC: 'SW' represents South West Trains. The line beginning with 'LO' describes the train's place and time of origin, in this case Waterloo, leaving at 17:48 from platform 11. The lines beginning with 'LI' describe intermediate points on the train's route: a pair of time values represents an intermediate stop, with the arrival and departure times specified, while a single time value represents a non-stop, passing time ('H' represents a half-minute: e.g. 1757H represents 17:57:30). Finally, the record beginning with 'LT' represents the place

and time of the service's termination, in this case Southampton Central at 19:16, at platform 3B. It can be seen that, whereas the SPG data uses three-letter station codes, the CIF data uses codes of up to 8 characters in length representing Timing Point Locations, or TIPLOCs, which include significant locations other than stations, such as junctions (e.g. Woking Junction and Worting Junction in the example above, represented by WOKINGJ and WRTINGJ, respectively). A more detailed description of the CIF data format is available from the Association of Train Operating Companies' (ATOC's) website (ATOC, 2007).

3   Tool Development

3.1   Developing the PLANET tool

As can be seen from the data examples above, there are separate requirements for processing SPG and CIF data, although both types are converted to a single, common format for input to PLANET. The processes differ also in that SPG data is typically more tailored to the requirements of a particular situation, usually comprising concise train information for a single TOC and date, from which it is normally only necessary to extract trains that operate within a certain time period, e.g. the morning peak between 07:00 and 10:00. CIF data, on the other hand, tends to be more comprehensive, containing timetable information for multiple TOCs and days of the week, thus requiring a greater amount of processing and filtering.

SPG Data Import

When running the program, the user is first asked to specify the file from which the timetable data is to be imported, and then whether or not joining and splitting trains (as described in the preceding section) are to be analysed.

   If an SPG input file is specified, the user is then asked which time period is to be modelled: AM (07:00 – 09:59 inclusive), PM (16:00 – 18:59), Inter-Peak (10:00 – 15:59) or All Day (00:00 – 23:59). The user is then asked to specify the TOC under consideration, although this information is only required later, to identify the relevant 'key stations' for the TOC in question (see below).

   Once these criteria have been input, the program reads the SPG file, assigning a unique, incremental identifying number to each train, and storing each element of each line of data in a temporary, intermediate file. If analysis of joining and splitting trains has been requested, the program creates lists of joining and splitting trains, together with their corresponding 'main trains', for subsequent EMME/2-based processing within PLANET.

   Once this initial SPG-specific process is complete, the subsequent data processing procedure is as for CIF data, and is described following the next, CIF Data Import, section of text.

CIF Data Import

As noted above, the nature of CIF data is such that it requires a greater degree of filtering than does SPG data. When a user has specified a CIF input file, and whether or not a join/split analysis is required, some further information is requested: the user is asked to specify the PLANET model (North, South or Strategic) being used, the time period to be modelled (as for SPG input) and the day of the week and the date for which for which timetable data is required (the day of the week could easily be obtained from the specified date, but it is useful to ensure that the user specifies the correct day, given the significant

variations in timetables from day to day, particularly from weekday to weekend). Finally, the user is asked to specify the TOC(s) for which timetable data is required.

Since PLANET uses the same three-letter codes (TLCs) for stations as those used in SPG data, no conversion of SPG station identifiers is required; however, since CIF data uses TIPLOCs, it is necessary to convert these to TLCs. A further issue arises from the fact that the three different PLANET models each represent a different set of stations, meaning that the data to be extracted from the CIF file varies according to the PLANET model for which it is intended. A .csv file is used to store the relevant station codes, and their inclusion in or exclusion from the various models, an extract from which is shown below:

```
TLC,TIPLOC,PSM,PSouth,PNorth
AAP,ALEXNDP,1,1,0
AAT,ACHANLT,0,0,0
ABA,ABDARE,0,0,0
ABC,ALTNBRC,0,0,0
ABD,ABRDEEN,1,0,1
ABE,ABER,0,0,0
ABH,ABRE,0,0,0
ABR,ABRCYNN,0,0,0
ABW,ABWD,0,1,0
ABY,ASHBRYS,1,0,1
ACB,ACBG,1,0,1
ACC,ACTNCTL,0,1,0
…
```

For example, it can be seen here that Alexandra Palace (AAP/ALEXNDP) is represented in PLANET Strategic (PSM) and PLANET South (PSouth), but not in PLANET North (PNorth), whereas Aberdeen is represented in PLANET Strategic and PLANET North, but not in PLANET South.

The program reads this .csv file, and, for each line, if the relevant station is represented in the specified version of PLANET, the corresponding TLC and TIPLOC are stored in a list for future reference.

The program then starts reading through the specified CIF file. If a join/split analysis has been requested, joining and splitting trains are identified and recorded. The CIF records this information in 'Association' records, in the format shown below:

```
AANW67166W6717104121305061011111100JJSSOTON    TP
```

In this example, trains W67166 (main train) and W67171 (joining train) join (JJ) at Southampton Central (SOTON), the association lasting 13 December 2004 and 10 June 2005, and being valid on Monday – Friday inclusive (1111100). The program generates a file containing a list of all joining and splitting trains whose association is valid on the specified date and day, recording the main and subsidiary train identifiers, and the TIPLOC at which the join or split occurs. This file is then used, as in the SPG case described above, for subsequent EMME/2-based processing within PLANET.

The timetable data, an extract from which is shown above, is then processed. If the specified date and day are within the ranges listed in the timetable data line beginning with 'BS', and the TOC shown in line beginning with 'BX' is one of those specified by

the user, that information, plus information including the train's unique ID, category, headcode, status, vehicle type and speed are recorded in a 'line header' file listing these characteristics for each train that satisfies the specified parameters.

For each train satisfying these parameters, its detailed stopping pattern and times are then recorded. For each of the following lines beginning with 'LO', 'LI' or 'LT', the specified TIPLOC is checked against the previously-recorded list of TIPLOCs included in the specified PLANET model. If the TIPLOC is represented in the model, it is recorded in a second, 'line segment' file listing detailed train journey information, together with the corresponding unique train ID and stop type (origin; pass; pick-up and set-down, pick-up only or set-down only; destination).

When the CIF file has been read in its entirety, some summary data (total number of trains found, and total number of specified trains and TIPLOCs found) is output to the screen, the file is closed, and the program proceeds to the next stage of the process, shared by both SPG- and CIF-derived data.

Common Data Processing: Train Selection

Once the trains have been selected from a CIF file according to TOC and day/date of operation, or the relevant data has been extracted from an SPG file, the next step is to select those trains that fall within the specified peak, inter-peak or all-day time period. For trains that operate entirely within the relevant time period (i.e. the times at origin and final destination are both within the period), the selection process is straightforward; however, decisions also have to be made about trains that operate only partially within the specified time period.

In PLANET, the concept of 'key stations' is used for this purpose. For each TOC, a list of the key, or most important, stations served by the TOC are stored in descending order of priority in a file. The key station listing for South West Trains is shown below:

```
TLC,Hierarchy,Name
WAT,1,London Waterloo
SOU,2,Southampton Cent
RDG,3,Reading
PMS,4,Portsmouth & S
```

The key station data is used as follows: the stations served by each 'candidate train' are checked against the list of key stations for that TOC. If the highest-hierarchy key station served by the train is served within the specified time period, the train is included in the model; if it is served outside the specified time period, it is excluded. If a train serves none of the key stations, the average of the times at its origin and destination are calculated, and, if this average is within the specified time period, the train is included in the model; otherwise, it is excluded. For example, if a SWT arrives at Waterloo at 07:05 or departs from Waterloo at 09:58, it is included in the AM Peak model, whereas, if it leaves Southampton Central at 08:50, and arrives in Waterloo at 10:05, it is excluded, despite operating mainly within the AM Peak period. The unique ID of each train thus selected to be included in the specified model is stored in a list for subsequent processing.

When each 'candidate train' has been tested for inclusion in the specified PLANET model, some further summary data (total number of candidate trains included, and total number rejected) is output to the screen, and the program proceeds to the generation of the EMME/2 transit line files for input to PLANET, a process again shared by both SPG- and CIF-derived data.

Common Processing: EMME/2 Input File Generation
EMME/2 transit line data is generated for each train included in the list of unique train IDs generated as described in the preceding paragraph. For each unique ID, the program reads, first the 'line header' data file, converting the header data for the train into EMME/2 format and writing it to the EMME/2 input file, and then populates the EMME/2 segment data (i.e. station-to-station data) by searching the 'line segment' file for the relevant unique ID, and then converting and writing the corresponding data to the EMME/2 format and input file. The resulting transit line data for the two SPG-formatted Brighton Main Line trains above are shown below, followed by that for the CIF-formatted South West Main Line service. In each case, the first two lines of data define various overall characteristics of the service, including unique ID, mode, headway and service description. The subsequent lines list the services' stopping points, the inter-station journey times and the stopping (i.e. pick-up and set-down) characteristics, with the time at each station included at the end as a 'comment'.

   Brighton Main Line trains:

```
a '_10323' t 1 180 999 'VIC-SOU (6660) 0817' 0 0 999
 path=yes  dwt=#.00 us2=1 us3=1 ttfl=2 ttft=0
  VIC   us1=5.5   tus2=0   tus3=0                 / 08:17:00
  CLJ   us1=3.5   tus2=0   tus3=0    tdwt=+0.0  / 08:22:30
  BAL   us1=4     tus2=0   tus3=0    tdwt=+0.0  / 08:26:00
  SRS   us1=3     tus2=0   tus3=0    tdwt=+0.0  / 08:30:00
  ECR   us1=4     tus2=0   tus3=0    tdwt=+0.0  / 08:33:00
  PUR   us1=7     tus2=0   tus3=0    tdwt=+0.0  / 08:37:00
  ELD   us1=5.5   tus2=0   tus3=0    tdwt=+0.0  / 08:44:00
  GTW   us1=3.5   tus2=0   tus3=0    tdwt=+0.0  / 08:49:30
  TBD   us1=10    tus2=0   tus3=0    tdwt=+0.0  / 08:53:00
  HHE   us1=14    tus2=0   tus3=0    tdwt=+0.0  / 09:03:00
  PRP   us1=3.5   tus2=0   tus3=0    tdwt=+0.0  / 09:17:00
  HOV   us1=6.5   tus2=0   tus3=0    tdwt=+0.0  / 09:20:30
  SSE   us1=6.5   tus2=0   tus3=0    tdwt=+0.0  / 09:27:00
  WRH   us1=1.5   tus2=0   tus3=0    tdwt=+0.0  / 09:33:30
  WWO   us1=5     tus2=0   tus3=0    tdwt=+0.0  / 09:35:00
  ANG   us1=6     tus2=0   tus3=0    tdwt=+0.0  / 09:40:00
  FOD   us1=4     tus2=0   tus3=0    tdwt=+0.0  / 09:46:00
  BAA   us1=8     tus2=0   tus3=0    tdwt=+0.0  / 09:50:00
  CCH   us1=11    tus2=0   tus3=0    tdwt=+0.0  / 09:58:00
  HAV   us1=7     tus2=0   tus3=0    tdwt=+0.0  / 10:09:00
  CSA   us1=8.5   tus2=0   tus3=0    tdwt=+0.0  / 10:16:00
  FRM   us1=6.5   tus2=0   tus3=0    tdwt=+0.0  / 10:24:30
  SNW   us1=12    tus2=0   tus3=0    tdwt=+0.0  / 10:31:00
  SDN   us1=8     tus2=0   tus3=0    tdwt=+0.0  / 10:43:00
  SOU   us1=0     tus2=0   tus3=0    tdwt=+0.0  / 10:51:00
c -------------------------------------------------
a '_10324' t 1 180 999 'HHE-EBN (6180) 0904' 0 0 999
 path=yes  dwt=#.00 us2=1 us3=1 ttfl=2 ttft=0
  HHE   us1=7     tus2=0   tus3=0                 / 09:04:00
  WVF   us1=11.5  tus2=0   tus3=0    tdwt=+0.0  / 09:11:00
  LWS   us1=12.5  tus2=0   tus3=0    tdwt=+0.0  / 09:22:30
```

```
   PLG    us1=7      tus2=0  tus3=0    tdwt=+0.0   / 09:35:00
   EBN    us1=0      tus2=0  tus3=0    tdwt=+0.0   / 09:42:00
```

South West Main Line train:

```
a 'W67323' t 1 180 999 'WAT-SOU (1B93) 1748' 0 0 999
 path=yes   dwt=#.00 us2=1 us3=1 ttfl=2 ttft=0
  WAT    us1=24     tus2=0  tus3=0                  / 17:48:00
  WOK    us1=20     tus2=0  tus3=0    tdwt=+0.0   / 18:12:00
  BSK    us1=18     tus2=0  tus3=0    tdwt=+0.0   / 18:32:00
  WIN    us1=9      tus2=0  tus3=0    tdwt=+0.0   / 18:50:00
  ESL    us1=4      tus2=0  tus3=0    tdwt=+0.0   / 18:59:00
  SOA    us1=11     tus2=0  tus3=0    tdwt=+0.0   / 19:03:00
  SOU    us1=0      tus2=0  tus3=0    tdwt=+0.0   / 19:14:00
c -----------------------------------------------------
```

In the CIF-derived output, it can clearly be seen how the TIPLOCs at which the train does not stop have been excluded from the EMME/2 input.

3.2 Modifying and Extending the PLANET Tool for Use with Railplan

Background
Although the PLANET tool was originally developed primarily for Arup's benefit and in-house use, presentations were given to TfL's Railplan Panel, describing the tool's development and application. TfL was sufficiently impressed by the tool to commission a version for use with Railplan, as part of the ongoing development of Railplan: the immediate requirement was the generation of updated heavy rail transit lines for the AM Peak and Inter-Peak time periods, an otherwise laborious and time-consuming process, but it was also recognised that the development of such a tool would greatly facilitate future transit line updates, and thus provide longer-term benefits.

The development of this version of the tool was simplified slightly by the fact that it was to use only CIF-based data, meaning that no facility for importing SPG-formatted data was required. However, it was complicated considerably by the fact that, unlike PLANET, which uses a single EMME/2 node to represent each TIPLOC represented in the models, many TIPLOCs are represented by two Railplan nodes, one for each direction of travel, and Railplan uses up to eight EMME/2 nodes to model some individual TIPLOCs, particularly large terminal and junction stations, and stations used by multiple TOCs.

Another, relatively minor, contrast between PLANET and Railplan is that, whereas PLANET uses three-letter station codes that, particularly with experience, enable users to identify train routes with relative ease, Railplan uses numeric codes, which make stations much less easy to identify. It was therefore stipulated that station names should be included as 'comments' at the ends of the relevant lines of the EMME/2 input files.

Two additional requirements were that multiple trains of the same TOC with identical stopping patterns should be aggregated together into single EMME/2 transit lines, and that these aggregated transit lines should be assigned Railplan codes representing their route and direction of travel, in place of the unique train IDs used in the PLANET transit lines.

Modifications of the PLANET tool for Railplan Use

The initial analysis of joining and splitting trains (where specified), the processing of the CIF file, and the 'key station' analysis to establish which trains should be included in the Railplan input file are all substantially unchanged from the processes employed in the PLANET tool.

The absence of different versions of Railplan (as opposed to the three versions of PLANET) simplifies the import of data for the purposes of mapping CIF TIPLOCs to Railplan nodes; however, as noted above, many TIPLOCs are mapped to multiple nodes. As part of the means of addressing this issue, the Railplan network is also imported, as a list of (i-node, j-node) pairs.

The node mapping process proved to be the most difficult aspect of the modifications of the PLANET tool for use with Railplan. Where a straightforward one-to-one mapping occurs (i.e. a TIPLOC is represented by a single Railplan node), the process is straightforward, and more-or-less identical to that used for the PLANET tool   For situations where a single TIPLOC is represented by multiple Railplan nodes, initial plans for a standardised node look-up process proved difficult to implement, given the sheer variety of criteria that needed consideration in order to uniquely identify some Railplan nodes.

In simpler areas of the Railplan network, the mapping process has been partially automated, with the program using the network link data to 'find its way' from Railplan node to Railplan node. However, such a process is unlikely to be applicable to the entire network, particularly in major stations where Railplan node numbers relate to platform numbers, and no unique 'Railplan path' can be identified between adjacent TIPLOCs. The result of this is that a considerable element of the node-mapping process has for now had to be 'hard-coded' into the tool, and is thus potentially vulnerable to future changes in train routeings and/or the structure of the network model, and is therefore a priority for further improvements.

Once the node mapping is complete, the station sequence listing with corresponding stop types (i.e. pick-up only, etc.) is determined for each train, together with information on its joining/splitting characteristics. These listings are then used to aggregate trains with identical stopping sequences and join/split characteristics (some further work is required on variations in overall and intermediate travel times).

Two EMME/2 transit line files are then generated for input to Railplan, one of disaggregate transit lines, and one with aggregated transit lines. While the first is not required for Railplan input purposes, it forms a useful part of the 'audit trail' of the conversion process. A file of aggregated joining and splitting transit line information is also generated, listing the joining and splitting transit lines, and the Railplan nodes at which they join or split. Examples of the two Railplan input files are shown below, based on the following CIF extract:

```
BSNW67248041213050610111110000    PXX1B529252124620204    EMU
BX          SWY
LOSOTON    1455 14553B          TB
LINTHMJN              1457H00000000    FL
LISTDENYS             1459 00000000
LISOTPKWY 1501H1503        15021503         T
LIELGH    1506 1513        150615131  FL    T
LISHAWFD             1518H00000000
LIWNCHSTR 1521H1523        15221523         T
```

```
LIWRTINGJ               1537H00000000   SL
LIBSNGSTK 1540 1542       154015424          T
LIFLEET   1552H1553H      15531553           T
LIFRBRMN  1558 1559       15581559   FL      T
LIWOKINGJ               1606H00000000
LIWOKING                1607 000000002
LIHCRTJN                1614H00000000       FL
LISURBITN               1615H000000002      FL
LINEWMLDN               1617H00000000
LIWDON                  1619H000000006
LICLPHMJM 1624 1625H      162416257   MFL    T
LTWATRLMN 1634 163413    TF
```

Disaggregate Transit Line:

```
a 'W67248' t 785 180 999 '  SOTON-WATRLMN 1B52' 24620204
   1455 -1
 path=yes  dwt=#.00 us1=1 us2=1 ttfl=2 ttft=0
  902466    us3=14     tus1=0   tus2=0                      /
   14:55:00 SOUTHAMPTON SOUTH WEST (BOTH)
  902566    us3=13     tus1=0   tus2=0     tdwt=+0.0  /
   15:09:00 EASTLEIGH SOUTH WEST (BOTH)
  902066    us3=19     tus1=0   tus2=0     tdwt=+0.0  /
   15:22:00 WINCHESTER SOUTH WEST (BOTH)
  901566    us3=12     tus1=0   tus2=0     tdwt=+0.0  /
   15:41:00 BASINGSTOKE S. WEST (DWN/BTH)
  901666    us3=14     tus1=0   tus2=0     tdwt=+0.0  /
   15:53:00 FLEET SOUTH WEST (DOWN/BOTH)
  910166    us3=8      tus1=1   tus2=1     tdwt=+0.0  /
   16:07:00 WOKING SOUTH WEST (DWN/BTH)
  720367    us3=2      tus1=1   tus2=1     tdwt=+0.0  /
   16:15:00 SURBITON SOUTH WEST (UP)
  720567    us3=2      tus1=1   tus2=1     tdwt=+0.0  /
   16:17:00 NEW MALDEN SOUTH WEST (UP)
  740267    us3=5      tus1=1   tus2=1     tdwt=+0.0  /
   16:19:00 WIMBLEDON SOUTH WEST (UP)
  290367    us3=10     tus1=0   tus2=0     tdwt=+0.0  /
   16:24:00 CLAPHAM JUNCTION S.W. (MAIN) (UP)
  310185    us3=0      tus1=0   tus2=0     tdwt=+0.0  /
   16:34:00 WATERLOO S.WEST (MID-NUMBERS) (UP)
c --------------------------------------------------
```
Aggregate Transit Line:

```
a 'WT113U' t 785 60 999 '  SOTON-WATRLMN 1B52' 24620204
  1455 -1
 path=yes  dwt=#.00 us1=1 us2=1 ttfl=2 ttft=0
  902466    us3=14     tus1=0   tus2=0                      /
   14:55:00 SOUTHAMPTON SOUTH WEST (BOTH)
  902566    us3=13     tus1=0   tus2=0     tdwt=+0.0  /
```

```
       15:09:00 EASTLEIGH SOUTH WEST (BOTH)
    902066    us3=19      tus1=0   tus2=0      tdwt=+0.0   /
     15:22:00 WINCHESTER SOUTH WEST (BOTH)
    901566    us3=12      tus1=0   tus2=0      tdwt=+0.0   /
     15:41:00 BASINGSTOKE S. WEST (DWN/BTH)
    901666    us3=14      tus1=0   tus2=0      tdwt=+0.0   /
     15:53:00 FLEET SOUTH WEST (DOWN/BOTH)
    910166    us3=8       tus1=1   tus2=1      tdwt=+0.0   /
     16:07:00 WOKING SOUTH WEST (DWN/BTH)
    720367    us3=2       tus1=1   tus2=1      tdwt=+0.0   /
     16:15:00 SURBITON SOUTH WEST (UP)
    720567    us3=2       tus1=1   tus2=1      tdwt=+0.0   /
     16:17:00 NEW MALDEN SOUTH WEST (UP)
    740267    us3=5       tus1=1   tus2=1      tdwt=+0.0   /
     16:19:00 WIMBLEDON SOUTH WEST (UP)
    290367    us3=10      tus1=0   tus2=0      tdwt=+0.0   /
     16:24:00 CLAPHAM JUNCTION S.W. (MAIN) (UP)
    310185    us3=0       tus1=0   tus2=0      tdwt=+0.0   /
     16:34:00 WATERLOO S.WEST (MID-NUMBERS) (UP)
  c -------------------------------------------------
```

The major difference between the two transit lines can be seen in the fifth elements of the first lines of data. In the disaggregate case, the number 180 indicates a service headway of 180 minutes, i.e. that the service runs once in every three-hour evening peak period. The equivalent value in the aggregate case is 60, indicating a headway of 60 minutes, and thus that three identical trains have been identified and combined in a single transit line.

4   Outstanding Issues

There are many areas in which the tools could be further improved, but the main issues related to the CIF-to-Railplan node mapping process, which requires considerable further attention. The interface is also a prime candidate for improvement: while the current, command line interface works quite well, a standard GUI-type interface would make the tools considerably more versatile and user-friendly. It would also be preferable to complete all the join/split processing within the Perl tools, rather than generating files which require further manipulation with PLANET or Railplan.

5   Conclusions

Both tools have proved to be practicable and useful, and a significant improvement over manual processing or 'interim' solutions using Excel, Access, etc., although both have considerable scope for further development and improvement, both of their interfaces and of the underlying algorithms and final output.

References

[1]   ATOC (2007), *Common Interface File: End User Specification* [online], Network

Rail. Available from http://www.atoc.org/rsp/_downloads/RJIS/20040601.pdf [Accessed 19 February 2007].