NCSD DATA BUILD PROCESS

1. Introduction

Coach data is provided by the following operators for England, Scotland and Wales:
- National Express
- Megabus
- Easybus
- NewBharatCoaches
- Terravision

National Express provides a weekly Cif file.

Megabus use an online timetable creation tool, belonging to the current contractor, to create their own data and then provide an updated Cif file every few months.

The Easybus, NewBharatCoaches and Terravison data is created and updated by the NCSD team.

1.1 Supporting Documentation

This document should be read in conjunction with the following document:

NCSD_CIFStructureandOutputFiles_V2

The above document provides information on what is required in respect to Cif data and other supporting files in order to provide the weekly build.

2. Overview of NCSD Data Build Process

2.1 Import latest NaPTAN data into NCSD

The full NaPTAN Version1 database is downloaded from:

http://81.17.70.199/public-transportdatamanagement/

This is currently maintained by Landmark Ltd on a weekly basis and imported into the NCSD.

2.2 Operator Data

Suppliers are requested to provide data to NCSD on a regular basis.

National Express: A ZIP file containing CIF data is supplied by National Express by email on a weekly basis.

Megabus: Stagecoach use an online timetable creator, which belongs to the current contractor, to capture the service timetable information. This then exports a Cif file which is then emailed. This is for their Gold service and normal services.

Easybus / New Bharat / Terravision: For these operators, data is manually maintained by taking information from websites or other marketing material. The current contractor uses their own software to create and manage the timetable data from these operators.

All supplied data is downloaded / copied to the local NCSD build server prior to the weekly build commencing.

2.3 Logical Stop Lookup Table

NCSD maintains a database of National Express 9000 stop codes matched to NaPTAN ATCOCode. National Express provide this lookup by using ZP records in their weekly Cif file (see example below)

ZPN900059008C 149000006048

NCSD maintain this database by updating any existing lookups or adding in new look ups from the data provided each week. Any ZP records which have a blank ATCOCode are ignored.

As part of each weekly build, any National Express stops which do not have a lookup will be reported as errors back to National Express. These are often festival stops or service station stops and not deemed to be relevant to the exported Cif file. National Express 9000 stops that are not found in the look up file **must not** be added to the exported Cif file. Instead, the pre-process file for National Express is updated to remove these stops from that weekly build.

It is the responsibility of National Express to ensure that all relevant ZP records are provided in their source Cif files before they can be used in a build.

2.4 Pre-processing files

Pre-processing files are used if certain stops, operators or services need to be removed from the build process e.g. stops may be removed if the supplied NaPTAN ATCO Code provided in the source route data is not in the latest NaPTAN file. The stop would be entered into the pre-processing file and then this would be excluded in the build but other valid stop data for the route would be retained. Services may be removed if backward timings are present. Pre-processing files, therefore, allow the latest data to be included and remove problematic data without compromising the build data integrity.

For the National Express Cif data, pre-processing files are used to ensure that certain stops, operators and services can be removed from the build. At present, National Express provide non UK services and these need to be excluded during the build process.

NCSD takes the CIF variant of the files provided by National Express. The following process is applied to the National Express data:

- Drop operators not required. Currently these include the following: EN,EP,ER,EU,EZ,HH,QA,QD,QE,QH,QI,QK,QN,QO,QP,QQ,QR,QT,UL,WE
- Remove stop records for stops outside Britain (using 9000AXXXXX format where X is digit and A is alphabetic). Remove non-GB records from services
- Remove stop records for Motorway Service Areas and Festival stops

The following notes are added to Megabus routes:

Route M34: "Note2, For further information please contact Customer Care on 0871 266 3333"

Route MB: "MB1, Passengers should pre-book at megabus.com or call 0900 1600900"

For **all** NewBharat services the following note is added: "Passengers must pre-book tickets to travel on this service"

2.5 Add Route Association and Journey Association records

NCSD maintain an old list of these records but this has not been updated in many years. This list is then automatically added to the build file each week. National Express does now include QG and QJ records but no decision has been made to use these as yet.

2.6 The ATCO - CIF Route Association Record (QX)

The current NCSD does not import any QX records from its current suppliers and so does not export any QX records into the NCSD-atco.cif export.

2.7   The ATCO - CIF Vehicle Association Record (QY)

The current NCSD does not import any QY records from its current suppliers and so does not export any QY records into the NCSD-atco.cif export.

2.8   The ATCO - CIF Journey Date Running Records (QE)

There is a fundamental difference between the QE records as defined by ATCO – CIF and how they are used by the TXC Importer.

The Journey Date Running Records are used to define date ranges *outside of the normal start and end date of the service* upon which the vehicle will or will not run. Within these dates the vehicle *will still run on the days of the week defined by the service.*

As an example: if a QS defines a vehicle running from the beginning of March to the end of August - Monday to Friday only - and is then followed by a QE defining a two week period in October in which the vehicle runs additionally, then within that two week period the vehicle will still only run Monday to Friday.

The TXC importer uses the QE records (with operational days set to 1 –known as QE1) to define the bank holiday dates on which the vehicle will run additionally. For example: there may be a Saturday only service which runs additionally on Good Friday and Easter Monday. It is obvious that under the conditions described previously the TXC method will not work. As the vehicle will be defined to run on Saturday only, the QE1 records for Friday and Monday will simply be ignored.

The way around this will be to first convert one of the QE1 records into a secondary QS record. This will be a copy of the QS record that the QE1 is associated with but with the DaysOfOperation set to MondayToSunday and the Start and End Dates set to the Start and End Dates of the QE1 (the Start and End dates will actually be the same date - the date of the bank holiday). The other QE1 record will then be removed from the old QS and added to the new QS.

3.   Intermediate Processing

Before the NCSD CIF data is exported, a number of intermediate processes are performed.

3.1   National Express clusters.

This process clusters all National Express Logical and Physical stop codes with their NaPTAN equivalents.

3.1.1   The 500m Clustering Rule

Coach Exchange Points (CEP) downloaded from

http://81.17.70.199/public-transportdatamanagement/DataUser_Home.aspx

are used as a basis to create clusters.   For each weekly build the existing StopsInClusterReport.CSV can be updated with new stops and service data if any new stops fall within 500m of the CEP coordinates for each cluster and stops are not listed in the StopsNotToCluster file described below.

3.1.2   "Stops Not To Cluster" File

To remove the undesired stops from being automatically clustered, a file of the unwanted stops will be loaded into the NCSD database. This will be referred to as the StopsNotToCluster file, which is manually maintained.

The Stops in Cluster Removal Criteria

The working criteria for allowing stops to remain in a cluster are outlined below in order of priority. These criteria can, however, be overridden following a DfT written instruction.

- Priority 1, the stop closest to the cluster centre, with more than one service associated to it and which has a timing point for one or all services, will be selected.
- Priority 2, will be a stop closest to the cluster centre, with just one service associated to it with a timing point for that one service.
- Priority 3 will be a stop closest to the centre, which has more than one service associated to it but does not have a Timing Point for any of the services.
- Priority 4 will be a stop which will be the closest to the cluster centre, with just one service associated to it and which does not have a timing point for that one service.

3.2   Fixed Links

3.2.1   Fixed walk links (QW records)

A file was created with fixed walk links (QW records) for any clusters which are within 3000 metres (crow flies) of another cluster, with a time calculated using a walk speed of 4km/hr and assuming the on-the-road distance is 1.4 times the crow flies distance. This file is used to add the QW records for each build.

3.2.2   Barrier Walk Link File

For some cluster geographical locations, the fixed walk link time is too short to circumnavigate barriers, such as rivers for example, and then board a connecting service to continue a journey.

To enable the current process to ignore fixed walk link records a file, known as the CurrentBarriers.csv, will contain unwanted cluster walk link records.

The CurrentBarriers.csv file will be referenced during the making of the fixed walk links (QW records) to ensure that the unwanted fixed walk links, which have come from the CurrentBarriers.csv, are not included in the NCSD-atco.cif file. This file will be maintained by the DfT.

3.2.3   NCSD Cluster Walk links Report

Although this data is in the NCSD-atco.cif, this report will include cluster names and coordinates to help locate the clusters geographically. This comma separated variable (csv) report can be created and called ClusterWalkLinks.csv.

3.2.4   Stop and Cluster interchange times

Previously, an operator has been able to define interchange times using QG (stop) and QJ (cluster) records. The ATCO.cif format does not allow these interchange times to be related to that specific operator so the values provided by one operator affect all operators at a given location.

Some data received by NCSD includes extremely high interchange times, effectively prohibiting the interchange between services at the specified locations. Increasingly, as more operators are added to the NCSD, these high interchange times are prohibiting acceptable, inter-operator interchanges.

NCSD has therefore decided to replace any interchange time as defined by one operator (or group of operators) where another operator uses the same location with a default value of 15 minutes.

4.   Exported data and Reports

4.1   NCSD ATCO.cif

The most significant file exported by the NCSD process is the ATCO.cif file. The details of the contents are defined in "*NCSDOutput_V2.Doc*".   The ATCO.cif file is created for valid operator's / services. Services that have expired are not exported.

4.2   Weekly Quality Check

An SQL script is run on the database each week. This script checks a number of parameters:

Title of test	Reason for inclusion	User Action
Date of report	Are you looking at the latest data?	Check the date is recent
NaPTAN details	Confirm that the NaPTAN database used in the build is current	Check the date is less than one week old and number of stops is within reasonable limits (380,000+)
Services, Clusters, Vehicle Journeys	Indicators of NCSD contents	Check the values in this build are consistent with the benchmark values
Services by supplier	Shows the number of routes in the current and last build and the differences	Check that there are no significant differences between two builds.
Incorrect Stop Types	Indicators of NaPTAN stop codes used	All stops types should be "BCS" or "BCT" or "BCQ" or "FER". Others are flagged.
Time Running Backwards	Indicator of timetables with departure time prior to arrival time.	This should not occur. Suppliers to be asked to correct data at source.
DELeted or PENding NaPTAN codes	Indicators of NaPTAN stop codes used	All stops should be ACTive. Others are flagged.
Stops assigned to more than one cluster	Possible issues with clustering	Any stop in this list is assigned to multiple clusters.
Services added / removed since last week	A list of data source / operator name / service numbers added or removed since the last build	Information only. Data providers will be contacted to confirm services removed are intentional.
Duplicated Services	A list of details where the service number is supplied from more than one source	Information only. A check to ensure a service is not duplicated. Duplicates will not be removed.
Summary of operators and services	The number of services by operator and data source.	Information only.
Detailed list of services	A list of each service number, operator and data source in the NCSD.	Information only.

Services Removed	Services entirely removed from a cluster	Cluster maintenance. Unless deliberate, replace the service
Services Added	New services added since last build	Information only.
Journeys with over 25 stops	A listing of routes for review as they have a large number of stops, which are included in the NCSD.	Information only.

4.3   Supporting files

The following additional files will also be uploaded to the ftp site on a weekly basis.

NaPTANLookUp.csv
Contains a look up of the NatExA and equivalent NaPTAN codes used by stops in the NCSD (including historical stops). Other fields from NaPTAN are also exported.

NCSDStopNames.csv
A simpler version of NaPTANLookUp.csv. Contains only the ATCOCode, ATCOName and National Express Physical StopID

CoachReferences.csv
Contains the NPTG v2 data derived from the NCSD

NCSDStops.csv (NaPTAN extract)
This file contains the NaPTAN v1 fields for all the stops in the current build of the NCSD.
Note that this shows the status of stops as in NaPTAN - including DEL or PEN stops which have been used in the NCSD data build. Downstream systems are advised to import all stops in this file as if they are ACTive - and to rely on local editors to update their data in subsequent weeks.

Archive yyyymmdd.ZIP
A single ZIP file containing all the files from that week's build.

NCSDStopsInCluster.csv
Contains stops and their services for each cluster

UnusedCEPs.csv
List of NPTG Coach Exchange Points which are not used in the current NCSD build.

ClusterWalkLinkReport.csv
List of Walk Links in the current NCSD build.

CurrentBarriers.csv
Cluster walk links to discard.

5.   Completing the weekly build

The WeeklyQualityCheck file is reviewed after each run to create a new build.

Entries in the file which would cause further investigation are:

- o Significant difference in the number of operators
- o Significant difference in the number of routes, clusters or vehicle journeys
- o Significant number of stops removed from the last build (a further indication of services removed or updated StopsNotToCluster file)

Any suspected issues will be analysed by looking at the main build log produced during the processing of the NCSD data and the appropriate action taken. This action may be to:

- fix the data and rerun the build
- use a previous data file provided by that supplier, and rerun the build
- to continue and notify the NCSD community of the issues.

If there are no significant issues, the NCSD is uploaded to the ftp site and the NCSD build process is deemed to be complete. The comparators used in the checking processes are reviewed and updated regularly to ensure that the checks remain appropriate in all cases.

6. Subsequent Actions

6.1 Creation of new Coach Exchange Points

1. DfT to ask Landmark to create a new CEP, and notify the NCSD contractor of this.
2. The new CEP coordinate information will provided when they have updated and provided the new CoachExchanges.CSV file.
3. The NCSD contractor will then update their CoachExchanges.CSV file used in the cluster processing.
4. New cluster records are then automatically generated during the cluster processing.
5. The newly generated cluster records then need to be checked internally before they are passed on to the DfT for final approval.
6. The approved cluster data is then provided with the next build.

6.2 Cluster Walk Link file

DfT will supply to NCSD the edited ClusterWalkLinks.csv. This will be fed into the NCSD database to remove the unwanted fixed links or cluster walk links

6.3 Weekly Timetable

Thursday
Download NaPTAN and NPTG
Receive the Barriers Cluster Walk links file update by Noon

Friday
Receive data from data providers, by 11:00 latest
Start build process
Review Error Logs and weekly quality report and address errors as practically possible
Complete build process
Upload files to ftp site by 20:00 and notify users that data is available, providing information on any known issues

Weekly
Address any further data issues not already covered

Fortnightly
Review Coach Exchanges
Review Clusters
Review websites of operators whose data is manually added to NCSD

7. Procedure for weekly build process

7.1 Build Clusters

Copy the NCSD clustering folder from the NCSD projects directory on NCSD server to the D:\dft_NCSD\Clustering\ directory that has been created locally. In the local Source data ensure that you have the files listed below: -

Coach Exchanges.csv	22/05/2012 06:32	Microsoft Excel C...	150 KB
NaptanLookup.csv	25/05/2012 14:19	Microsoft Excel C...	66 KB
NCSD-atco.cif	25/05/2012 14:06	CIF File	63,796 KB
NCSDStops.csv	25/05/2012 14:18	Microsoft Excel C...	4,377 KB
operators.cif	20/01/2012 10:50	CIF File	31 KB
StopsInClusterReport.csv	24/02/2012 15:44	Microsoft Excel C...	1,092 KB
StopsNotToCluster.csv	18/05/2012 17:30	Microsoft Excel C...	321 KB
UnusedCEPs.csv	18/05/2012 17:23	Microsoft Excel C...	1 KB

Ensure you have updated the relevant files with the ones created in the latest NCSD build.

After the files have been updated browse to the Code folder. In this folder there will be an MBX that is a program that will run the cluster program. After this has run it will populate the Export data folder with the following files: -

CoachReferences.csv	31/05/2012 14:07	Microsoft Excel C...
StopsInClusterReport.csv	31/05/2012 14:07	Microsoft Excel C...
StopsNotToCluster.csv	31/05/2012 14:07	Microsoft Excel C...
UnusedCEPs.csv	31/05/2012 14:07	Microsoft Excel C...

The StopsinClusterReport and Coach References should be placed in the AssociatedData of this week's build. All of these files are also placed in the weekly build directory.

7.2   Download the latest NaPTAN data:

Purpose: To provide the stop QB and QL record data and ensure that the latest data from the regions do not have any missing NaPTAN data. Data is provided by logging into:

http://81.17.70.199/public-transportdatamanagement/DataUser_Login.aspx

Use the one off download option and download the NaPTAN Version1 data in CSV format. Copy this data to the AssociatedData directory in the build directory with file name "NaptanStops".

7.3   Download the latest NPTG file

Purpose: To provide the latest Coach Exchange data for use in the clustering process. Download the NPTG data from:

http://81.17.70.199/public-transportdatamanagement/DataUser_Login.aspx

Extract only the CoachExchanges.CSV file and place this in the "SourceData" Directory of the clustering software.

7.4 Check weekly uploads

Any data that has not been uploaded from the data suppliers needs to be chased up. If suppliers fail to provide the data by 2pm on the Friday, their last previously upload data will be used. The email that informs users about a build will contain information of any suppliers' data that was not refreshed since the last build and will show the date of the last data that was provided to NCSD.

7.5 Copy the Template Folder

Make a copy of the template folder; this will contain all the pre-processing files that are required to make do the run. After copying re-name the new folder LatestBuildDDMMYYYY. The filepath is: - \\TNDSServer\D\Projects\NCSD

7.5.1 Update the NaPTAN Stops

Take the downloaded NaPTAN Stops file and re-name it to NaPTANStops and copy into the latest builds AssociatedData folder.

Name	Date modified	Type	Size
AdditionalQs	24/07/2014 20:23	File folder	
AssociatedData	24/07/2014 20:23	File folder	
EA	24/07/2014 20:23	File folder	
Easybus	05/03/2015 10:27	File folder	
EM	24/07/2014 20:23	File folder	
Greyhound	24/07/2014 20:24	File folder	
Megabus	05/12/2014 10:44	File folder	
National Express	24/07/2014 20:24	File folder	
NE	24/07/2014 20:24	File folder	
NewBharatCoaches	24/07/2014 20:24	File folder	
NW	24/07/2014 20:24	File folder	
Scotland	24/07/2014 20:24	File folder	
SELTA	24/07/2014 20:54	File folder	
SW	24/07/2014 20:24	File folder	
Terravision	05/03/2015 10:08	File folder	
Thandi	17/12/2014 10:49	File folder	
Wales	24/07/2014 20:24	File folder	
WM	24/07/2014 20:24	File folder	
YK	24/07/2014 20:24	File folder	
140811 NCSD Guidance v2.1.pdf	11/08/2014 13:02	PDF File	17 KB

Copy the latest data across from the NCSD upload server into the relevant folders in the latest build folders.

7.6 Run NCSD Application

Open up the NCSD application on the desktop on NCSD Server. Press RUN. You will be prompted to select the folder where the latest data folder is located along with the location of the last build.
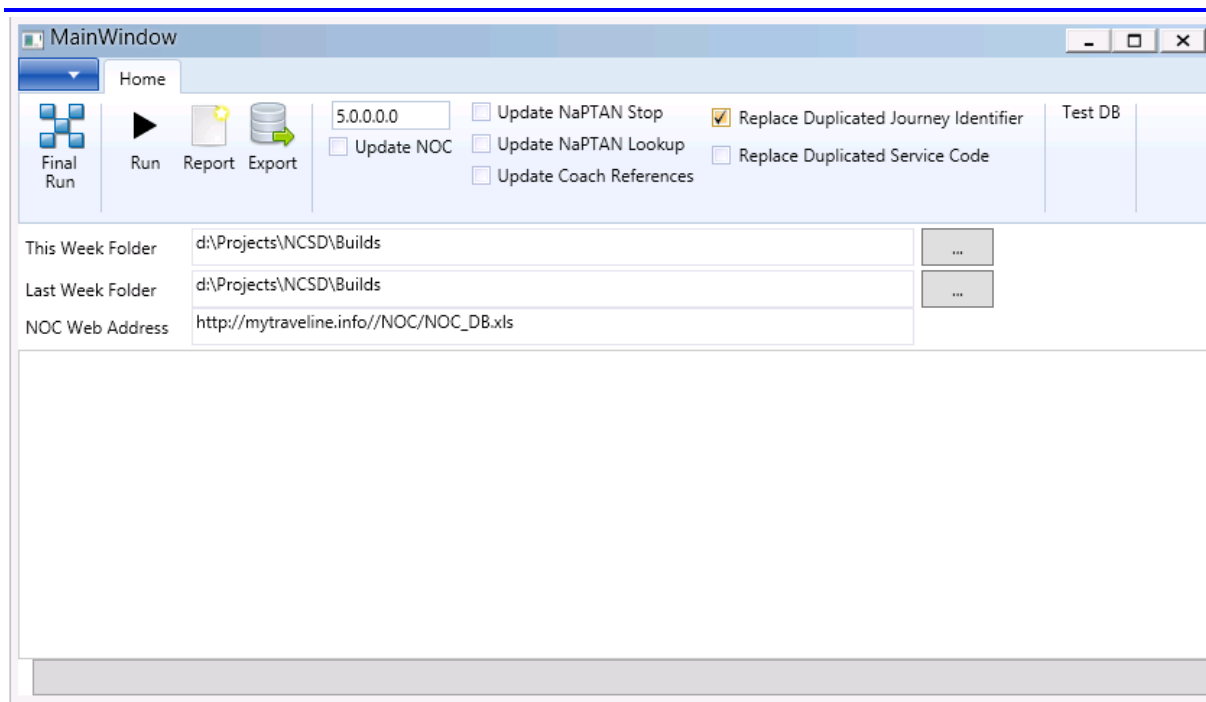
For the first run, please tick
- Update NOC (automated process which downloads and uses latest NOC file)
- Update NaPTAN Stops (uses the latest NaPTAN stop file in the AssociatedData directory)

After these are selected press RUN.

After the completion of the previous task you need to run the Report. Select Run report from the GUI. You will then need to select this week's build directory and last week's build directory. Reports are automatically saved in this week's build directory.

7.7   Check Report

The report will be created (WeeklyQualityCheck.txt) – this will compare this week's data with last week. A summary of Routes and Vehicle trips will be given, if there is a small difference in the results then there is no need to worry.

Check the missing operators/stops. If any of this is completed then this needs to be fed back to the data supplier/region. If it is only a temporary issue then a piece of text can be added into the pre-processing file and then rerun the NCSD build.

7.8   Run EXPORT function

On the NCSD interface the Export function needs to be run. This will then export the final files that are required for the NCSD build. You need to select the latest build of data (NCSD-atco.cif) and then the folder where you want this data saved.

7.9   Upload the data to the FTP

The next stage involves taking the data and uploading it to the NCSD FTP. The following files will be appended: -

IncludedServices.csv
NCSDStopNames.csv
NCSDStops.csv
NCSD-atco.cif
WeeklyQualityCheck.txt

Name	Size	Type	Date modified
_NCSD ftp site - guide to contents.pdf	37 KB	Adobe Acrobat Document	13/05/2014 01:00
140811 NCSD Guidance v2.1.pdf	17 KB	Adobe Acrobat Document	11/08/2014 01:00
ARCHIVE20150402.zip	9,249 KB	WinRAR ZIP archive	02/04/2015 17:54
ClusterWalkLinkReport.csv	10 KB	Microsoft Excel Comma S...	26/09/2014 01:00
CoachReferences.csv	48 KB	Microsoft Excel Comma S...	02/04/2015 16:06
CurrentBarriers.csv	43 KB	Microsoft Excel Comma S...	20/02/2015 00:00
IncludedServices.csv	76 KB	Microsoft Excel Comma S...	02/04/2015 16:25
NaPTANLookup.csv	364 KB	Microsoft Excel Comma S...	02/04/2015 16:10
NCSD-atco.cif	79,523 KB	CIF File	02/04/2015 16:09
NCSD-atco.zip	7,852 KB	WinRAR ZIP archive	02/04/2015 17:54
NCSDStopNames.csv	1,211 KB	Microsoft Excel Comma S...	02/04/2015 16:10
NCSDStops.csv	4,335 KB	Microsoft Excel Comma S...	02/04/2015 16:10
StopsInClusterReport.csv	911 KB	Microsoft Excel Comma S...	02/04/2015 16:04
StopsNotToCluster.csv	333 KB	Microsoft Excel Comma S...	02/04/2015 16:04
UnusedCEPs.csv	3 KB	Microsoft Excel Comma S...	02/04/2015 16:04
WeeklyQualityCheck.txt	377 KB	Text Document	02/04/2015 16:25

And the additional files will be added from the Clustering process.

Currentbarriers.csv and ClusterwalklinksReport.csv will not need updating unless they have been changed.

7.10   Create TXC Files

After the build has been completed select FINAL RUN function which will then convert the Cif file to TXC.

When completed, ZIP the two TXC folders, along with the IncludedServices.csv and NCSDGuidance.pdf and call it NCSD.ZIP. This is then copied into a directory on the website.

This also needs to be copied to the TNDS FTP site.

The Atco CIF file will be compressed as a ZIP file and placed next to the original. Finally the collection of all the files will be collated and compressed into a file called ArchiveYYYYMMDD. Send Mailout using MailChimp

Copy the build from the previous week, change the text to comment on whos data hasn't been included and what the date of the old version used.

8. Additional Data Checking Processes

8.1 Weekly build reports

The weekly build report provides important information about the build and how this compared with the previous weeks build. The number of services, departures and also information on missed operators, route that were in this week but not last week and routes that were in the previous build but now not included.

8.2 Additional checks

The weekly Cif build is imported into the NCSD contractor's software. This provides the following:

1. The software has powerful Cif file checking procedures so will indicate any issues with the file structure and also where there are missing NaPTAN stops. Reports will be provided to allow the user to clearly identify any issues on import.
2. The software allows the user to visualise the Coach time table data
3. Routes can be analysed at random to allow the users to see any obvious errors