Contents

[Knowledgebase – DPPP/ATP comparison spreadsheet 2](#_Toc66052336)

[What are Knowledgebase, DPPPs and ATPs? 2](#_Toc66052337)

[Why compare them? 2](#_Toc66052338)

[This spreadsheet 3](#_Toc66052339)

[Using this Spreadsheet 5](#_Toc66052340)

[Updating the Spreadsheet 7](#_Toc66052341)

[Anomalies and Disclaimers 8](#_Toc66052342)

[Attribution 10](#_Toc66052343)

# Knowledgebase – DPPP/ATP comparison spreadsheet

## What are Knowledgebase, DPPPs and ATPs?

Whilst every UK train station is owned by Network Rail, they are run and managed by some 25 different companies. Such companies are known as Station Operating Companies. (They mainly consist of Train Operating Companies; though Network Rail run the biggest 20 or so direct, and some individual organisations (e.g. airports) operate a small number of stations.)

Station operating companies are required to publish information on the features and accessibility of the stations they manage, in two separate places.

1. **Knowledgebase.** A database maintained by Rail Delivery Group, under their subsidiary National Rail Enquiries. The information in this database is used to supply the pages on the National Rail Enquiries Website, for example, [this station information page](https://www.nationalrail.co.uk/stations/BYA/details.html). It is also used by Train Operating Companies to supply information on their websites, e.g., [this station information](https://www.scotrail.co.uk/plan-your-journey/stations-and-facilities/crr)  [page o n Sco trail’s website](https://www.scotrail.co.uk/plan-your-journey/stations-and-facilities/crr) .

This information is also available for use and download by third parties e.g., me.

1. **Disabled People’s Protection Policies (DPPPs) / Accessible Travel Policies (ATPs).** Station operating companies are obliged to produce and publish a set of two documents making commitments as to what accessibility provision they will make for disabled travellers. They are obliged to get these approved by the Office of Rail and Road and are under licensing and legal obligations to comply with the policies they write in them.

Up until recently, these documents were called DPPPs; from August 2019 they have been called ATPs.

The Office of Rail and Road specifies what information must be included in these documents. Up until September 2020, this included information on accessibility features of each station that the company manages. Many Station Operators still publish this data.

Station Operating Companies are under a duty to keep both sources of information accurate.

## Why compare them?

There are issues with the accuracy of the station accessibility information provided in Knowledgebase and thus provided through National Rail Enquiries, Train Operating Companies and so on. This causes problems for disabled people.

One mechanism for auditing the accuracy of the information is by comparing the information in Knowledgebase with that in DPPPs / ATPs. In as much as they contain the same information for the same stations, the information should of course be the same between the two sources.

So, for example,

* If National Rail Enquiries says that staff help is available for disabled passengers at a particular station, but the ATP / DPPP states that staff help is not available at the same station, then evidently something is awry.
* If National Rail Enquiries says that a station has step-free access to the whole station, but the ATP / DPPP says for the same station that there is step-free access to the Northbound platform, but the Southbound platform can only be accessed using 30 steps over a footbridge, then something is wrong.

As a wheelchair user, that latter inaccuracy is a major problem for me.

## This spreadsheet

I wrote this spreadsheet to compare the station data in Knowledgebase against the data in station operators’ DPPPs/ATPs. It automates the collection, comparison, and analysis of this data for each station, and then produces statistics on the degree of agreement between the two sources.

Structure of the spreadsheet

There are several worksheets, identified by tabs along the bottom of the screen. Going from left to right:

* “Charts”. This workbook shows automatically produced pie charts showing the consistency of information across 10 common/key accessibility attributes. For example, it shows that at the time of writing, for 32% of stations for whom data on the presence of seated areas are indicated in both the DPPP/ATP and Knowledgebase, that information differed between the two sources – i.e., the DPPP/ATP said there is no seated area whereas Knowledgebase says there is or vice versa.

“All Comparable Attributes” only shows “Agree” for stations where there are NO detected differences between the DPPP / ATP and Knowledgebase.

“Charts” also includes the list of station operators and their codes, and this document.

* “Stats”. This sheet shows automatically produced statistics of the congruence of the two data sets (DPPPs / ATPs and Knowledgebase) over a range of different data attributes. NB: it only provides data for stations where accessibility information is present in both the DPPP/ATP and Knowledgebase.

So, for example, at the time of writing, DPPPs/ATPs, and Knowledgebase both had Staff Help Availability information listed, for 617 of the UK’s 2,585 stations. Of those stations, 385 (62%) of the “Staff Help” information agreed between that in the DPPP/ATP and in Knowledgebase (i.e., either both Knowledgebase and the DPPP/ATP said “*staff help available at this station*”, or they both said “*staff help not available at this station*”.) 232 stations (38%) differed between the two sources – i.e. the DPPP/ATP said “*staff help is available at this station*” but Knowledgebase said, for the same station, that “*staff help is not available*”, or vice versa. It also lists the total number of discrepancies I’ve found (6,139, not including omissions), the average errors per station (2.4), the number / proportion of stations for which there were no discrepancies (396/15), and the total proportion of compared station attributes that differ between the two sources. (20%)

* “Combined”. This combines the data from the Knowledgebase worksheet and the UberDPPP worksheet into one table and shows the results for each station.
* “Knowledgebase” worksheet. This contains the information from National Rail Enquiries’ Knowledgebase database.

I download a snapshot of the Knowledgebase database as a .xml file. This is interpreted by a power query, which extracts and formats the information for every UK station and inserts it into the table in this worksheet.

With a couple of exceptions, my query only interprets and uses the data that is amenable to automated comparison.

* “UberDPPP”. This combines the accessibility information from all 25 Station Operating Companies’ DPPPs / ATPs into one large table. It should contain information on all mainline stations in the UK, give or take a couple of anomalies (for which see below). This data is automatically combined using a power query.

NB: in general, I have only kept the data that is also present in Knowledgebase and that is amenable to automated comparison (e.g., no free text)

* 25 sheets named AW, CC, CH …. XS. These contain the station information I have manually imported from each station operator’s ATP / DPPP; with one sheet per ATP/DPPP. The code letters refer to different station operating companies as detailed in the legends on the Charts sheet – for example, AW refers to Transport for Wales (originally stood for Arriva Trains Wales); LT refers to London Underground (who manage a small number of mainline stations.)

Some companies manage more than one station operating franchise. So, for example, Govia Thameslink Railway manages Great Northern, Gatwick Express, Southern and Thameslink franchises, and has therefore produced one combined DPPP / ATP covering all four franchises’ stations. It, therefore, has one worksheet with this combined information on it. The first line of each sheet shows where I got the information from, and when said information was published. For example, LO (London Overground) states that the data is taken from their DPPP of October 2018.

Each worksheet consists primarily of one table, featuring every station managed by the station operator and the information I have manually imported from their DPPP.

## Using this Spreadsheet

“Charts” and “Stats” both have a “slicer” at the top of the page, which looks like this:

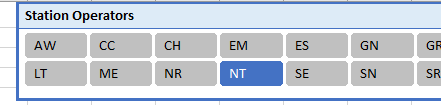


or this:



Both these do the same thing. They can be used to change the Stats and Charts worksheets to show statistics for stations run by individual station operators, combinations of operators or all station operators.

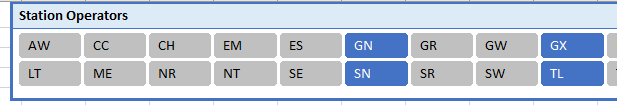
For example, if I want to see information just for Northern Rail, I click on the “NT” button, so it looks like this:



I can now see that of the 478 stations run by Northern, for 128 (27%) of these stations their ATP says that there are audio announcements whereas Knowledgebase says there is no such facility, or vice versa.

Where the DPPP / ATP doesn’t include information on a particular attribute, it simply isn’t displayed - for example, Northern doesn’t provide any information in their DPPP as to the staffing level of stations or whether they have waiting rooms, so the pie chart doesn’t display and the equivalent table in Stats says simply “N/A”.

It is possible to select more than one station operator by the “Ctrl” button. For example, to select the stations run by the four franchises owned by GTR, I would click “GN” then hold down “Ctrl” and click on “GX”, “SN” and “TL”, ending up with this:



The pie charts and statistics then automatically update to show the data for all the selected station operators.

One can return to showing data for all stations in the UK by clicking the “filter” icon in the top right of the “Station Operators” box. This re-selects all station operators and thus shows all stations.

The data in “Combined” can be filtered by e.g., train operating company, using the buttons in the column headings with downward-facing arrows/triangles. The discrepancies between each station’s accessibility features as listed in Knowledgebase / the DPPP are highlighted through the colour in the ATP / DPPP column; solid Green means they agree, solid Red means that they do not.

## Updating the Spreadsheet

The various queries in the spreadsheet take the raw data in the station operators’ worksheets (AW, CC etc.) and collate it. It is possible to change data in these individual worksheets and to refresh relevant queries so that the station data in UberDPPP and Combined are updated, and the statistics in Stats and Charts similarly.

The Knowledgebase query loads and interprets the stations' data from stations.xml as downloaded directly from Knowledgebase / Rail Delivery Group via the [National Rail Data Portal.](https://opendata.nationalrail.co.uk/) Instructions on how to download this file are on the [Open Rail Data Wiki](https://wiki.openraildata.com/index.php/KnowledgeBase).

It is non-trivial to download this information unless you’re a programmer. It may be best simply to ask me (Doug) to supply you with an updated spreadsheet. (I have straightforward mechanisms in place for obtaining and updating such information in the sheet.)

In any case, I may have supplied you with a spreadsheet in which all the mechanisms for updating such data have been disabled, i.e., with all queries deleted. This is because spreadsheets with power queries in them can spark warnings of potential risk to third parties when they receive and open such spreadsheets.

## Anomalies and Disclaimers

* + I can only compare information where the same information is provided, for the same station(s), across both sources (Knowledgebase and a DPPP/ATP). Station operators are left to their own decisions as to what station accessibility information they include in their DPPP/ATP; with the result that there is a wide variance between the operators. Also, station operators’ information provided in Knowledgebase is sporadic and patchy. For many stations, it is not possible to compare many access attributes for that reason.
  + The data are all out of date to one degree or another. The DPPP/ATP data is generally refreshed periodically and may be between a few months to a few years old (as shown by the date on the various pages / in the station rows). As of September 2020, station operators are no longer obliged to publish station accessibility information in this format anyway. Knowledgebase is supposed to be kept up to date by station operators, but they are very… variable in their approach and commitment to doing so. Therefore, the apparent discrepancy between accessibility may feature described in the ATP/DPPP and that in Knowledgebase could be because the information in Knowledgebase is simply more up to date, or indeed older, than that in the DPPP/ATP. (Though I don’t think that this accounts for the number of such discrepancies.)
  + There are some “odd one out” stations that don’t fit the system. For example, some new stations have opened, and some old stations have closed since relevant DPPPs / ATPs were published. Horden station opened on 29th June 2020, whereas Northern’s most recent DPPP was published in June 2017. Horden, therefore, doesn’t feature in any DPPP / ATP and the information for it can’t be compared.

Knowledgebase includes a few anomalous “stations”, for example, Dublin Ferryport (which isn’t a railway station in Great Britain) and three Underground / DLR stations that have no mainline train service. I don’t know why these are included, but they don’t feature in any DPPP or ATP so don’t affect the statistics.

“West Midlands Trains” (code LM) is still listed in Knowledgebase, though the company hasn’t run a franchise for some time; and it still listed as running one station – “Norton Bridge”. This station is permanently closed.

Doubtless, there are others.

* + Some years ago, Rail Delivery Group made the arbitrary decision to only permit stations’ StepFreeAccess Knowledgebase field to contain “*wholeStation*” or “*noPartOfStation*” – even though the [Knowledgebase specification](https://www.nationalrail.co.uk/static/documents/UserGuides/Stations%20XML%20Feed.pdf) (last updated May 2020) specifically allows options “*partialStation*” and “*allPlatforms*” – at one stroke, making the information provided by this field utterly unreliable and useless. Go figure.
  + Knowledgebase can store details of more than one car park at each station, but my spreadsheet only parses the data for the first one (if there is one), due to technical limitations. If the ATP / DPPP quotes more spaces / blue badge spaces than Knowledgebase, then this MAY be because the station has more than one car park.
  + My code for interpreting the Knowledgebase database is long and complex. So was the process of extracting all the stations’ details from 23 DPPP/ATP PDFs. There will therefore undoubtedly be errors in my code and my transfer of data. I have been very thorough and careful, but in such a big task this is inevitable. I hope that there are no errors big enough to significantly affect the resultant statistics, however; and my errors are certainly dwarfed by the erroneous recorded data provided by the operators!

## Attribution

Information from Knowledgebase has been accessed from the National Rail Data Portal and is provided by [National Rail Enquiries,](https://www.nationalrail.co.uk/) a trading name of [Rail Delivery Group / ATOC Ltd](https://www.raildeliverygroup.com/).



Information from ATPs / DPPPs has been taken from said documents published by each of the 25 Station Operating Companies concerned. They may hold the copyright to this data; I am claiming fair use.

Any of this spreadsheet (and this document) that does not comprise the above two data sources, is my copyright – Doug Paulley – and I license anybody to use it under Creative Commons:



This work is licensed under a [Creative Commons Attribution 4.0](http://creativecommons.org/licenses/by/4.0/)  [International License.](http://creativecommons.org/licenses/by/4.0/)

I can be contacted via my website – [Do ug P aulley’s website](https://www.kingqueen.org.uk/) .

My thanks to Jeff Harvey for his assistance in providing a better way to access the Knowledgebase feed without Windows mangling some of the character encoding.