

## Duplication in Timetable data

A description of the different types of duplication in the data and how to identify them

### Duplicated Labels

#### Line name duplication

Services have the same line name, possibly in the same area, and by the same operator. eg Stagecoach 1

Unlikely that there will be any NaPTAN points in common.

Handled within a Traveline region by Service Codes which are unique.

#### Cross boundary duplication

Line names normally the same. NOC is used to match local operator names between Traveline regions.

NaPTAN points and timing of journeys, are similar but may not be identical.

Different conventions in exporting TXC, by number of files, or time horizons for the exports mean that there may not be a one to one match of services.

There may also be variations in methods of interpolation or inclusion of stops or short journeys in the remote areas.

## Established definitions in ITO

#### Journey Duplicate

Report of duplicated journeys. A pair of journeys within an area are regarded as duplicate if they start at exactly the same time, follow exactly the same timing pattern, and there is at least one day of their operating

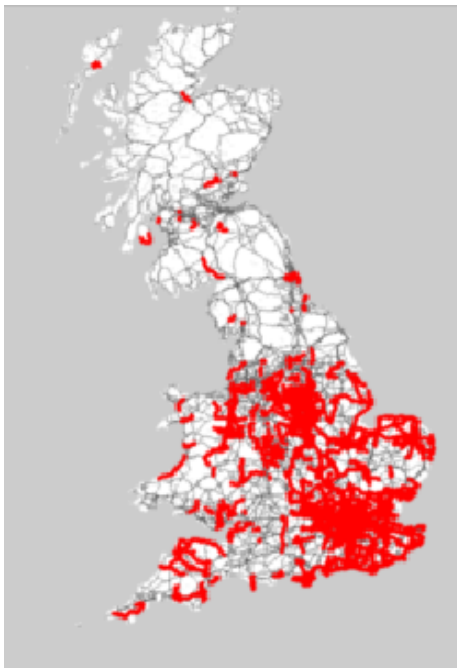
period in which they both run.

### Journey Partial Duplicate

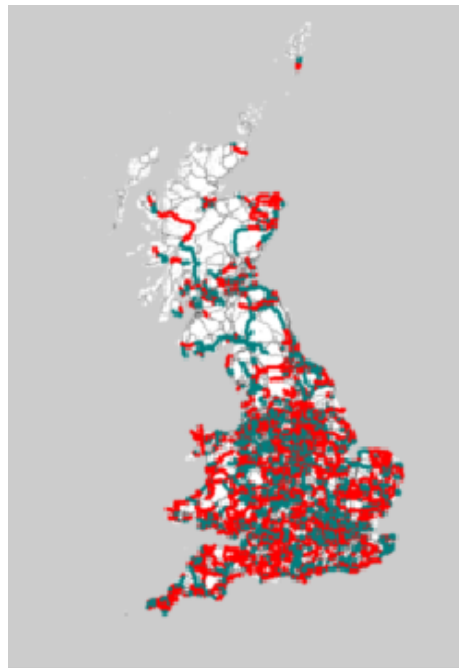
Report of journeys that are partial duplicates. A pair of journeys within an area are considered to be partial duplicates if they arrive at at least 10 of the stops at exactly the same time, and there is at least one day of their operating period in which they both run.

Maps from ITOQuality:

Total duplicates



Partial duplicates



## A further breakdown of types of duplicate

### Complete Duplicate

Same as ITO journey duplicate but where the period a journey runs is also the same. (ie not just a one day or more overlap)

One of the files can be suppressed without any loss of information.

### Partial duplicate (calendar)

Same as ITO Journey duplicate but where running on the same day is because of a dating overlap. ie operating period or special days.

### Partial duplicate (DayType)

Same as ITO journey duplicate but where running on the same day is

because of a DayType overlap ie day of week, Bank Holiday or School day conflict.

Partial duplicate (routing)

arrive at at least 10 of the stops in exactly the same time, and there is at least one day in their operating period in which they both run.

## Data Source Codes

The DataSource codes file is work in progress but is helping to interpret the ServiceCode information and/or filenames that are used in the data supplied to TNDS. This will facilitate reporting on the data and feedback to the right people to resolve duplication issues.