SIRI data summary

Received data

The SIRI-VR data that we are currently receiving essentially consists of a series of `<VehicleActivity>` blocks like this:

```
<VehicleActivity>
  <RecordedAtTime>2017-09-03T23:59:51+01:00</RecordedAtTime>
  <ValidUntilTime>2017-09-03T23:59:51+01:00</ValidUntilTime>
  <VehicleMonitoringRef>SCCM-54307</VehicleMonitoringRef>
  <MonitoredVehicleJourney>
    <LineRef>X5</LineRef>
    <DirectionRef>OUTBOUND</DirectionRef>
    <FramedVehicleJourneyRef>
      <DataFrameRef>1</DataFrameRef>
      <DatedVehicleJourneyRef>467</DatedVehicleJourneyRef>
    </FramedVehicleJourneyRef>
    <PublishedLineName>X5</PublishedLineName>
    <OperatorRef>SCCM</OperatorRef>
    <VehicleFeatureRef>lowFloor</VehicleFeatureRef>
    <OriginRef>0500CCITY476</OriginRef>
    <OriginName>Parkside Bay 16</OriginName>
    <DestinationRef>0500HSTNS064</DestinationRef>
    <DestinationName>Market Sq Stop D</DestinationName>
    <OriginAimedDepartureTime>2017-09-03T23:30:00+01:00</OriginAimedDepartureTime>
    <Monitored>true</Monitored>
    <InPanic>0</InPanic>
    <VehicleLocation>
      <Longitude>-0.2354520</Longitude>
      <Latitude>52.2262192</Latitude>
    </VehicleLocation>
    <Bearing>288</Bearing>
    <Delay>PT35S</Delay>
    <VehicleRef>SCCM-54307</VehicleRef>
  </MonitoredVehicleJourney>
</VehicleActivity>
```

From a review of actual data on three weekdays 2017-04-26, 2017-08-30 and 2017-09-04 the following appear to be true:

RecordedAtTime

A plausible timestamp for the event. Generally a few seconds in the past relative to time of receipt, very occasionally up to 75 minutes in the past or up to 60 seconds in the future.

ValidUntilTime

Always the same as RecordedAtTime.

VehicleMonitoringRef

Looks plausibly to be a vehicle identifier qualified by it's operator. Always the same as VehicleRef. Appears to match a field on a Whippet ticket issued on the Universal.

LineRef

Looks to be an identifier for the *Line* (e.g. Timetable) to which this journey relates. Probably needs to be qua;lified by OperatorRef for uniqueness. Always the same as PublishedLineName

DirectionRef

Always 'INBOUND" or 'OUTBOUND'.

DataFrameRef

Always '1'

DatedVehicleJourneyRef

Integers from 1 to about 10,000, occasionally appearing with one or more leading zeros -- unclear if they should be interpreted as numbers or strings.

These seem to indicate vehicle journey in some sense, and increase throughout the day resetting to 1 at midnight. They are however not unique even in one day, empirically with low numbers occurring more often than higher ones.

For any one day, most combinations of DatedVehicleJourneyRef and VehicleMonitoringRef have a 1:1 relationship with the combination of OriginRef and OriginAimedDepartureTime (which we understand to represent a 'Journey'), but even this breaks down with ~100 examples per day of DatedVehicleJourneyRef/VehicleMonitoringRef corresponding to 2, 3 or 4 separate instances of OriginRef/OriginAimedDepartureTime (and vice versa).

PublishedLineName

See LineRef

OperatorRef

One of

ATS CBLE FECS GP SCCM SCNH WP ZSIN

VehicleFeatureRef

Present in only 16% of records. If present, only ever 'lowFloor'.

OriginRef, OriginName, DestinationRef, DestinationName

One of about 420 Naptan stops

OriginAimedDepartureTime

Monitored

Always 'true'

InPanic

Always '0'

Longitude

-0.755235 to 0.63038 (Milton Kenes/Kettering/Oakam to somewhere between Newmarket and Bury St Edmunds)

Latitude

52.0085564 to 52.8346291 (Milton Kenes/Safron Walden to Spalding)

Bearing

0.0 to 354.0

Delay

A time delta in ISO format. Positive and negative.

VehicleRef

See VehicleMonitoringRef

Extracting trips

The raw data gives bus positions. Sorting by some combination of Origin, Departure, Origin Departure Time, Direction, Operator, Line and Vehicle Ref Might be expected to result in individual timetabled 'trips'. In practice it seems necessary to sort by *all* of these to produce anything useful.

The resulting paths are by and large consistent with expected timetable trips, with the following occasional mis-feastures:

- Position dropouts during trips. These may be more common in some places than others, perhaps reflecting poor GPS and/or mobile coverage. They may also be more common on fast route sections (busway, A14, A428, A1307), though even random dropouts will be more obvious at speed.
- [Often related to dropouts] Very-occasional widely off-route points (e.g. a point north of Ely for a bus clearly going from Fulbourn to the City Centre)
- Paths starting before the Origin stop or ending after the Destination. In most cases these look to be vehicles travelling to the origin or onward from the destination. In many cases these additional trips start from or end at recognisable bus depots or known layover locations.
- Paths starting after the Origin stop or ending before the Destination. These could be caused by position dropouts, or by the driver being late or early updating the bus's route information. For example it seems common for Universal buses to switch from 'To Addenbrokes' to 'To Edington' on approach to the hospital even though the route officially terminates/starts outside outpatients. This does mean that trips frequently appear not to reach their destinations.
- Paths that look as if they may be partial concatenations of more than one trip.