Data

CIF - Common Interface Format

stp stands for Short Term Planning

- C - Planned cancellation: the schedule does not apply on this date, and the train will not run. Typically seen on public holidays when an alternate schedule applies, or on Christmas Day.
- N - STP schedule: similar to a permanent schedule, but planned through the Short Term Planning process and not capable of being overlaid
- O - Overlay schedule: an alteration to a permanent schedule
- P - Permanent schedule: a schedule planned through the Long Term Planning process

CIF_train_id: Every train has unique train_id on one given day.

operating_charateristics
- list in CIF user guide

Schedule_days_runs is a bit field 0000010 (runs Saturday only)
Signaling_id: This field represents what the train is known as
CIF_power_type: an example is EMU (electric)

Schedule_location records: These records contain several values:

location_type (LO is origin, LI in the intermediate and LT stands for terminates)
tiploc_type is location code where it departing from
tuploc_code: FRAT - coach service depot

line: if there are more than 2 lines, this is the piece of the line where the train is **expected** to go through

Pathing_allowance: Time to cross over a junction

H stands for half a minute in time fields

Engineering_allowance: if there is known engineering work (because there will be a slight delay)
performance_allowance is used when the train transports people to signify time lost by people getting on the train

The date of a train is the origin date of a train (if it runs over midnight, it is still called saturday only train)

signalling_id: first letter represents a class (0 is bus, 1 is train class, 2 is other passenger

train, 3 expedited empty coaching stock, 5 is ECS)

List of tiploc code is on the wiki

DO NOT WORRY about association (SCHEDULE) since it does not happen most of the time
Tiploc - time in point location
tiploc_code is the train planners' code used to describe the location

scr_code: Seat reservation

TRAIN MOVEMENTS

There are 2 kinds of reports in this stream:

A signal controls a fixed set of a track (This set of the track is called a berth?)
TD monitors in which part the trains are according to the signals

(If it passes a signal, it changes its current location) So it will say that it passed XYZ

Some places do not have the signal boxes - Manually says that x just passed XYZ (unreliable)
auto_expected says that an automatic report was *expected* at a location

You have an **offset** because the signals are not very accurate. These offsets can have some 'interesting effects' for example it can be a few minutes off

event_type (arrival or departure) does not know if the train has stopped so it does not know if it is an arrival or a departure -> It needs further interpretation

gbtt_timestampt: Great Britain timetable (public time)
line_ind: the line that the train actually took -> Takes very little relation to the planned line

stanox is the location code that the operational systems use
loc_stanox is the location code

offroute_ind means that the train is heading in a different direction than where it is meant to go

schedule_type (P permanent, O overlap (STP - short time), C cancellation)

timetable_variation is the number of minutes early or late (The sign is in the variation state field)

toc_id is the train operating company's numerical id (List of all companies should be in the data folder on Drive)

train_id is composed of multiple fields, including the origin, destination and some characters that make it unique

train_uid: when there's a match to the time table it is the field on the timetable

DIAGRAM

- Created by the operators and it says what a piece of rolling stock should do during a day.
- We get the train_id and the schedule of a particular rolling stock
- set_id directly refers to a piece of rolling stock
- Pos is position where you have 2 or more sets together. So for example it goes on the back of a separate piece of rolling stock (so 2 means that it is behind the first set)
- Fuel represents the milage since you have last fueled

For some operators it is important in which direction the train arrives. This can be important for crowd control

GPS don't work very well in tunnels because the satellites are near the horizon.
The satellites also have problems re-acquiring signals after a train exits out of tunnel at 60mph. Finally there might also be problems with GPS data in stations where lots of steel work