OCCUPANCY FORECAST – PROFILE FOR SWITZERLAND

Systemaufgaben Kundeninformation (SKI) – Team SKI+
https://öv-info.ch/de/datenmanagement/ski/standards-der-ski

Status	Draft
Version	0.9
Date	2023-12-14
Authors	Markus Meier (SBB SKI+), Matthias Günter (SBB SKI+)
Copyright	CC BY 4.0

Document information

Description	This document contains a specification and profile of a two data formats for occupancy forecasts in public transport in Switzerland. The forecasts are supposed to provide a rough forecast (in 3 stages, symbolized with 3 manikins) about how many passengers are expected on a given train, bus or other mobility service.
Target audience	People who use or want to use occupancy forecasts of Swiss public transport.
Electronic documentation	https://www.öv-info.ch/de/datenmanagement/ski/standards-der-ski

Change History

Version	Status	Changes	Authors	Date
0.1	Draft	Initial draft	M. Meier	2023-03-20
0.8	Draft	Completed JSON flavour	M. Meier	2023-08-22
0.9	Draft	Added SIRI flavour	M. Meier	2023-12-14

Table of Content

1. What is it all about?

This document is intended to lay the foundation for an efficient, standardized exchange of occupancy forecasts of vehicles (trains, buses, etc.) in public transport, and possibly also for private mobility operators (on demand services, car sharing, etc.) in Switzerland.

The profile should serve for the following purposes:

- provide guidance for future IT systems providing occupancy forecasts in Switzerland,
- allow for a quick and easy implementation of applications consuming the forecasts,
- lay the foundation for an effective and efficient data exchange infrastructure in this field.

The profile specifies two data formats ("flavours") for the delivery of occupancy forecasts:

- A standards-based, fully compliant variant based on CEN SIRI ET version 2.1.
- A lightweight, independent, JSON-based data format.

2. Description and Context

The term "occupancy forecast" may be misleading and requires a proper definition and disambiguation first:

An occupancy forecast is a prediction (prognosis, forecast) of the occupancy rate of a given transport option (train, bus, etc.) expected at some date and time in the future. However, it is ***not*** a measurement or count of current or past occupancies.[1]

The main use cases of occupancy forecasts will be in the field of customer information and guidance. The forecasts can provide guidance in trip planner applications (timetables), helping to find comfortable, not too crowded trains and buses.

Overall, this should turn out to the benefit of all passengers and of transport operators alike, helping to reduce over-crowded options and encouraging to use options with low frequencies.

With this profile, we intend to define a simple data format for exchanging occupancy forecasts in Switzerland. The format should serve as a lightweight, temporary solution that may be replaced with little effort later, once an adequate standard will be available.

3. Who is responsible?

The team SKI+, which is responsible for the open data platform mobility Switzerland (ODMCH), defines this profile and will setup and provide the first data feeds.

[1] In SIRI version 2.1, "occupancy" is defined as "Passenger load status of a VEHICLE" (siri_reference.xsd), or "An approximate figure of how occupied or full a VEHICLE and its parts are, e.g., 'manySeatsAvailable' or 'standingRoomOnly'." (siri_journey.xsd).

4. Important Abbreviations and Links

Abbreviation	Description	Link
CAPRE	Capacity and Reservations, an SBB system.	SBB Capre
BO	Business Organisations, a list of Swiss transport companies.	https://opentransportdata.swiss/en/dataset/goch
DIDOK	Dienststellendokumentation (location documentation).	https://opentransportdata.swiss/en/dataset/didok
FOT	Swiss Federal Office of Transport.	https://www.bav.admin.ch/bav/en/home.html
ODMCH	Open Data Platform Mobility Switzerland.	https://opentransportdata.swiss
SBB	Swiss Federal Railways, Switzerland's main train operator.	https://www.sbb.ch
SBB timetable	SBB's online timetable for the entire Swiss public transport.	https://www.sbb.ch/en/timetable.html
SIRI	Service interface for real-time information relating to public transport operations, a European CEN standard. GitHub provides some open parts of the SIRI specification, including XML examples and XML schema definitions.	https://github.com/SIRI-CEN/SIRI
SKI	SKI agency / Transport Data Management Standards.	https://transportdatamanagement.ch/de/standards
SKI+	SKI sub-team, supporting FOT in the setup of a Swiss National Access Point.	https://opentransportdata.swiss/en/
UIC	International Union of Railways.	https://uic.org/

5. Basic Technologies and Standards

We take a "dual approach" by providing a standard-compliant SIRI PT (version 2.1) variant and a lightweight JSON-based alterative, respectively:

- The formats are thus based on XML (Extensible Markup Language) and JSON (JavaScript Object Notation), respectively.
- The data is provided as a collection of files in folders and subfolders (file system).
- ZIP archive file format is used for (lossless) data compression and for bundling all files/folders in one single file.

6. Usage

In Switzerland, the Swiss Federal Railways (SBB) have pioneered occupancy forecasts in their online timetable and platform displays since 2018. SBB has introduced symbolic "manikin" icons (1, 2 or 3 passenger icons) to predict how crowded their first- and second-class coaches will be. The forecasts are computed by the SBB system CAPRE.

In some passenger information systems and for some selected trains, SBB forecasts are provided for every coach of the train, to guide passengers to those coaches with lower frequencies.

In this profile, however, we only define a data format for forecasts per class (not per coach).

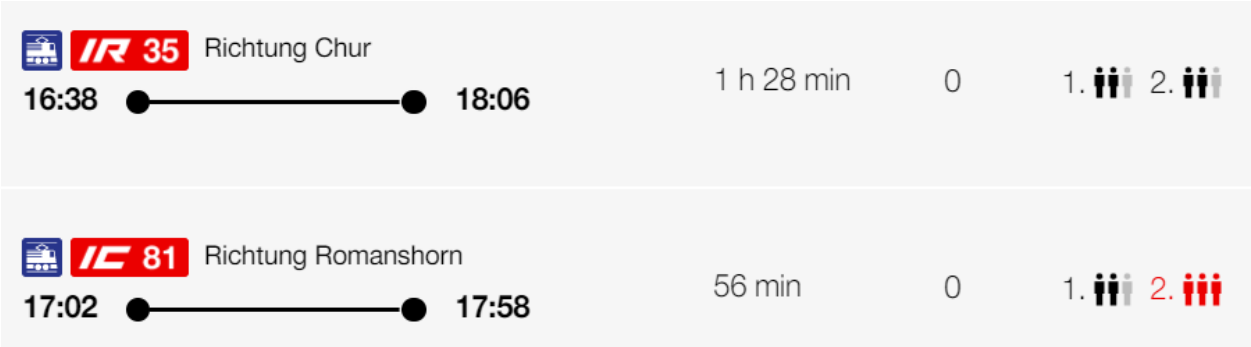

Figure 1: Screenshot of the SBB web timetable with occupancy forecasts per class, symbolized by three manikins.

The German main railway operator Deutsche Bahn introduced similar forecasts in 2019 (Figure 2). They do not distinguish between first and second class.

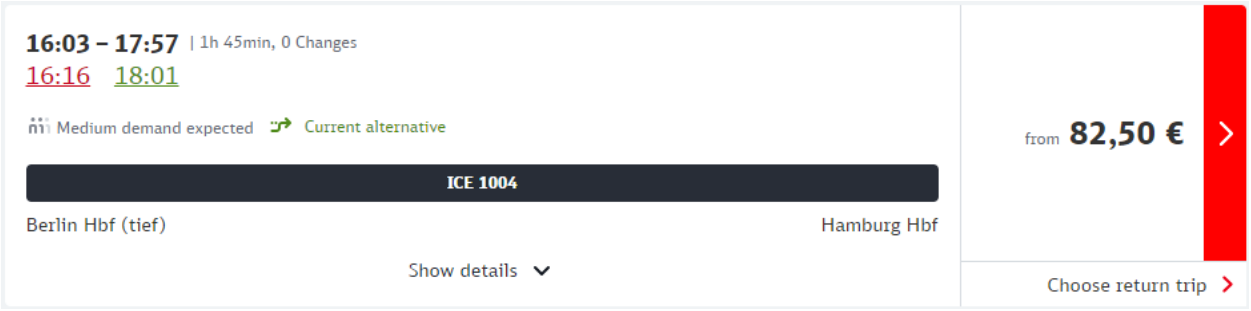

Figure 2: Screenshot of the Deutsche Bahn (bahn.de) web timetable with an occupancy forecast, displayed with three manikins and a text ("Medium demand expected").

7. Datasets/Services Switzerland

The first adoption will be the provisioning of SBB occupancy forecasts per class, for SBB trains and for the trains of some other operators, e. g. BLS, depending on the availability of the data and the consent of the data owners.

The data will be provided on www.opentransportdata.swiss (ODMCH).

8. Assessment

The following table shows an assessment of the proposed format by the SKI+ team:

P1 international	X	Currently not harmonized/aligned internationally, except for some alignment with SIRI.
P2 open	+++	Everything related to this profile will be completely open.
P3 simple	+++	Compact, simple formats, both JSON and SIRI flavours.
P4 established	X	Not established so far.

P5 evolutionary	++	Aiming at "good enough" rather than perfect quality.
P6 of high quality	++	SIRI overall is at a high quality level; however, SIRI leaves (too) much room for interpretation, hindering interoperability, requiring the definition of complementary profiles such as the present document.
P7 compliant	+++	In the SIRI variant, we aim at being fully compliant and equivalent to the JSON variant, which is not standards-based.
P8 interpretation-free	++	Same as P6.

9. Conclusions

We provide a specification of a data format in two flavours (variants SIRI/XML and JSON) for a data format that should serve as a standard and profile for an efficient exchange of occupancy forecasts in Switzerland. In a future version of this profile, the formats may evolve further, according to the evolution of the international standards and markets.

10. Specification of Swiss Occupancy Forecast Profile

10.1. Purpose of this Profile

1. The profile defines two different data formats, surnamed “flavours”, that we will use for our first delivery of occupancy forecast data of four Swiss rail operators (SBB, BLS, Thurbo and SOB), computed by the SBB system CAPRE, by courtesy of SBB.
2. We would like to fuel the discussion with stakeholders, including those who will use the data, and get feedback and proposals for the further evolution of this profile.
3. In future versions, this profile should provide guidance and guardrails for occupancy forecast data feeds by other data providers in Switzerland.

10.2. Two Flavours

In this current version, we define two different formats (flavours), nicknamed “SIRI” and “JSON”:

Nickname	What?	Why?
SIRI	Fully standard-compliant format, based on the SIRI ET (Estimated Timetable Service) in its upcoming version 2.1. As such, based on XML and XML schema and XML namespaces.	To provide a fully compliant version for those adopters who prefer XML data formats and/or strive for best-possible alignment with CEN standards.
JSON	An independent, JSON-based format which is using mostly the terminology and data structures of the original data source (CAPRE).	To provide a lightweight approach for those adopters who prefer simple, JSON-based formats and do not set a high value on standards compliance.

10.3. General Considerations

10.3.1. Provide a Minimal, Add-on Data Set

Our data feed does not aim to provide a comprehensive, self-contained timetable data set. Rather, its goal is to provide extra, additional data about occupancy forecasts, which are currently not available through other channels. It thus forms an “add-on” or “side car” to a timetable data set.

Therefore, we do not strive to add as much as possible data; instead, we follow a minimalistic approach:

- Provide only the forecast data.
- Include only minimal extra information needed to identify/match data elements in other data sets (mainly, timetables): train numbers, departure/arrivals stations ids, etc.
- Add some extra fields helpful for debugging purposes (e.g., station names).
- In the SIRI format, leave away all non-mandatory elements otherwise.

10.3.2 No Guaranteed Coverage and Quality

The data set is generated by a given system (CAPRE) and packaged and delivered by our SKI+ team to the best of our knowledge, as is, without systematic checks for completeness or data quality.

The same will hold true for any future data feed by any other party.

Thus, in any implementation scenario, the developers of consumer applications should implement their software in such a way that they can tolerate missing, non-matching or flawed records. They should

implement search or look-up algorithms which try to find matching forecasts based on the given identifiers: operatorRef, operation date, train number, departure stop and time, destination stop). If a matching section or call is found, the occupancy forecast data can be integrated in the given target application; otherwise, forecasts should be ignored.

10.4. General, Overlapping Aspects of Both Flavours

10.4.1. Delivery File Format

Our data feeds are delivered in one single, compressed file per flavour, named OccupancyForecastSIRI.zip and OccupandyForecaseJSON, respectively. The files are provided in the common ZIP file-format (produced by the algorithm of Java 17 Open JDK java.util.zip package).

10.4.2. Basic Folder and File Organisation

After unzipping, the base folder contains some 92 sub-directories for the respective operation days, starting from to “today” and covering the upcoming three months.

In order to organize the data, we introduce “operation-day folders” for all files. The folders are simply named by their respective date in ISO-8601-format. *Example: “2023-12-08".*

10.4.3. Operation Day

“operation day” in public transport parlance is the calendar day on which most of the traffic takes place. It may comprise transport services (trains, etc.) which run over midnight and thus, their actual date and time is one day later, meaning a “day shift” of +1. In rare cases, there may also be a day shift of –1.

Example: Train 21993 of SBB (operator 11), regional train S9 from Luzern to Lenzburg:

- *Departs from Birrwil on Dec. 15, 2023, 23:57 hours: operation day = 2023-12-15, day shift = 0.*
- *Arrives at Beinwil am See on Dec. 16, 2023, 0:01 hours: operation day = 2023-12-15, day shift = +1.*

10.4.4. Operator Files

In order to further organize the data in “reasonably sized” files, the occupancies are provided in separate per operator. Each operator file comprises all trains of this operator.

The file names contain the operator Ids commonly used in Swiss public transport (currently: 11: SBB, 33: BLS, 65: Thurbo AG, 82: Schweizerische Südostbahn SOB). They are named “operator-<operatorRef>.xml” and “operator-<operatorRef>.json”, respectively, for the SIRI (XML) and JSON flavours.

Examples: operator-33.xml (SIRI file for BLS), operator-82.json (JSON file for SOB).

10.4.5. Call Perspective versus Section Perspective

A journey of a train (or other public transport service) usually follows a given line, with a given number of calls (stops to allow passengers to board or alight) at given stations (stops).

When representing a journey in a data set, there are two different approaches, or “perspectives” to look at it:

- The call perspective: a journey consists of a sequence of calls; each call being a stop at a given stop place with an arrival and departure time.

- The section perspective: a journey consists of a sequence of sections; each section being the journey between two adjacent stops, with a departure and a destination stop (station), plus respective departure and arrival times.

In our context, we are using both perspectives:

- Out data source (the SBB system CAPRE for occupancy forecasts) uses a section perspective.
- In our JSON file formats, we use the section perspective as well.
- In our SIRI-based format, we use a call perspective, as prescribed by SIRI.

10.5. SIRI Flavour Data Format

10.5.1. Example of the Basic Data Structure in SIRI ET (XML)

The following data excerpt illustrates the JSON flavour. Since the data refers to operator 11 (SBB), it would be contained in a file named operator-11.xml.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- (C) Copyright 2005-2012 CEN SIRI -->
<Siri xmlns="http://www.siri.org.uk/siri"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="2.1"
xsi:schemaLocation="http://www.siri.org.uk/siri ../../xsd/siri.xsd">
  <ServiceDelivery>
    <ResponseTimestamp>2023-12-15T09:57:22+01:00</ResponseTimestamp>
    <ProducerRef>OdmchOccupancyForecast</ProducerRef>
    <EstimatedTimetableDelivery version="2.1">
      <ResponseTimestamp>2023-12-01T09:57:22+01:00</ResponseTimestamp>
      <EstimatedJourneyVersionFrame>
        <RecordedAtTime>2023-12-01T09:57:22+01:00</RecordedAtTime>
        <EstimatedVehicleJourney>
          <LineRef>null</LineRef>
          <DirectionRef>ch:1:Direction:H</DirectionRef>
          <FramedVehicleJourneyRef>
            <DataFrameRef>2023-12-04</DataFrameRef>
            <DatedVehicleJourneyRef>null</DatedVehicleJourneyRef>
          </FramedVehicleJourneyRef>
          <OperatorRef>11</OperatorRef>
          <TrainNumbers>
            <TrainNumberRef>1009</TrainNumberRef>
          </TrainNumbers>
          <EstimatedCalls>
            <EstimatedCall>
              <StopPointRef>8503424</StopPointRef>
              <StopPointName>Schaffhausen</StopPointName>
              <AimedDepartureTime>2023-12-04T06:47:00+01:00</AimedDepartureTime>
              <ExpectedDepartureOccupancy>
                <FareClass>firstClass</FareClass>
                <OccupancyLevel>fewSeatsAvailable</OccupancyLevel>
              </ExpectedDepartureOccupancy>
```

```
            <ExpectedDepartureOccupancy>
              <FareClass>secondClass</FareClass>
              <OccupancyLevel>standingRoomOnly</OccupancyLevel>
            </ExpectedDepartureOccupancy>
          </EstimatedCall>
          <EstimatedCall>
            <StopPointRef>8503000</StopPointRef>
            <StopPointName>Zürich HB</StopPointName>
          </EstimatedCall>
        </EstimatedCalls>
      </EstimatedVehicleJourney>
    </EstimatedJourneyVersionFrame>
  </EstimatedTimetableDelivery>
 </ServiceDelivery>
</Siri>
```

The data excerpt shows a real example (a train from Schaffhausen to Zürich HB) and is well-formed and valid SIRI ET version 2.1 file. For the sake of brevity, however, it contains only one train with train number 1009, which has only one section, i.e., a non-stop ride from Schaffhausen to Zürich HB. In the real delivery files, the "trains" array contains many trains, and most trains have more than one section.

10.5.2. Specification and Discussion of the Data Fields

In the subsequent sections, we specify and discuss all the relevant fields (elements, attributes), in the sense that we provide details and explanations where needed for disambiguation and clarity, but trying to remain as brief as possible and avoid evident, self-understood information. We intend to provide adopters (software developers) quick start guides and assistance, rather than giving formal or machine-readable specifications.

The SIRI ET 2.1 specification (schema files) may be consulted for extra information, such a basic description of the semantics, element hierarchy, cardinalities, element types, etc.. Often, SIRI defines complex type hierarchies, but the details of these add no extra value in our context.

Example: Various identifiers come with their own data types, which are ultimately based on xsd:NMTOKEN, thus, essentially a string without whitespaces. We will refer to this simply as "white-space-free string".

We provide data sets that are compliant with the specs, in the sense that they are well-formed and valid (as verified by XMLSpy).

10.5.3. Definition of the Keys in an Occupancy Forecast Record in SIRI Flavour

Field name (bold=mandatory, minOccurs>0)	**type**	**description**	**example value**
Response Timestamp (various occurencies)	xsd:dateTime	This timestamp should give a hint about when the data set was produced. Usually, all occurrences of this element should contain the same value.	2023-12-15T09:57:22+01:00

Field name (bold=mandatory, minOccurs>0)	type	description	example value
		In our data set, this will be the time when the data set is packaged (not the time when the forecast was computed).	
ProducerRef	white-space-free string	Identifier of the producer of the data set. In our data set, we use OdmchOccupancyForecast. Although not mandatory, we strongly recommend to provide this value for debugging purposes.	OdmchOccupancy Forecast
RecordedAt	xsd:dateTime	In our context, this element is of little added value. In our data set, we use the same value as ResponseTimestamp.	2023-12-15T09:57:22+01:00
EstimatedVehicle Journey	complex type	The SIRI element for a given train (or other means of transport) on a given date and at a given time. In other words, this is a container for the data of one train. For each train, one such element is appended.	Train 1009 on Dec. 4, 2023, departing at 6:47 CET from Schaffhausen.
LineRef	white-space-free string	Identifier for the line (service) of this train. In our data set, we do not have this information, thus provide “null”.	S3 ch:1:slnid:123456
DirectionRef	white-space-free string	A white-space-free string identifier, used for the direction of this train service. In our data set, we do not have this information, we provide a constant ch:1:Direction:H value.	ch:1:Direction:H
DataFrameRef	white-space-free string	Identifier of a reference date. The SIRI documentation suggests that this will “often” be the operation day. We require it to be the operation day.	2023-12-04
DatedVehicle JourneyRef	white-space-free string	A white-space-free string identifier, used for the journey. In our data set, we do not have this information.	null
OperatorRef	white-space-free string	Identifier of the operator. Although not mandatory and already part of the file name, we recommend to include this	11

Field name (bold=mandatory, minOccurs>0)	type	description	example value
		element at this point as well, for maximum clarity. In our data set, this is the operator code commonly used in public tranpsort (see section above).	
TrainNumberRef	white-space-free string	Identifier used by the operator to identify the given transport option (e. g. a train or bus). In Transmodel terminology, this corresponds to a DatedVehicleJourneyRef. In our data set, we provide the train numbers used in the Swiss public transport timetable published on ODMCH.	1009
EstimatedCall	complex type	SIRI uses a “call perspective” (see section above). The EstimateCall element is the container for a given call. We only provide departure times. The last call (terminal station of the train) has no departure, thus no departure time element and no occupancy forecasts.	
StopPointRef	white-space-free string	The identifier of the station (stop) of this call. In our data set, we use the stop identifier (UIC code) from DIDOK.	8503424
StopPointName	string	The name of the stop (station). Not a mandatory field; we recommend providing it to facilitate debugging.	Schaffhausen
AimedDeparture Time	xsd:dateTime	The date and time (with time zone) of the (planned) departure from this call. Although not mandatory, we strongly recommend that this element be provided and used for matching or data quality checks. If this time does not match with the time from another source (timetable data set), the forecasts should be skipped.	2023-12-04T06:47:00+01:00
Expected Departure Occupancy	Complex type	This element is not mandatory, but necessary for the main purpose of our data set. According to the SIRI documentation, this is “an approximate	

Field name (bold=mandatory, minOccurs>0)	type	description	example value
		figure of how occupied the journey is after departing from a given stop". To be meaningful, this element must contain the two child elements FareClass and OccupancyLevel.	
FareClass	FareClass Enumeration	SIRI provides various options (Enum values) here. In most Swiss trains, firstClass and secondClass are the two actually available fare classes. In our data set, we provide occupancy forecasts for these two classes; hence, we include two elements ExpectedDepartureOccupancy, one with FareClass firstClass and secondClass, respectively.	<FareClass>firstClass </FareClass>
OccupancyLevel	Occupancy Enumeration	SIRI defines various enum values. We consider the best matches as follows: 🧍 (1) = manySeatsAvailable (SIRI: "The vehicle has a large percentage of seats available." 🧍🧍 (2) = fewSeatsAvailable (SIRI: "The vehicle has a small percentage of seats available." 🧍🧍🧍 (3) = standingRoomOnly (SIRI: "The vehicle can currently accommodate only standing passengers." ❓ (0) = unknown. In our data set, we obtain level "indicators" (0, 1, 2, 3) directly from the source system (CAPRE) and package them accordingly.	<OccupancyLevel> fewSeatsAvailable </OccupancyLevel>

10.6.JSON Flavour Data Format

This section defines our JSON-based data format, tagged as version 0.9, as of December 14, 2023.

10.6.1. Example of the Basic Data Structure in JSON

The following data excerpt illustrates the JSON flavour. The example shows the same data as in the SIRI XML flavour example above. It would be delivered in a file named operator-11.json.

```
{
  "operatorRef": "11",
  "opDate": "2023-12-04",
  "lastUpdated": "2023-12-01T09:57:22+01:00",
  "timeToLive": "86400",
  "dataSource": "https://opentransportdata.swiss/search?q=occupancy",
  "version": "0.9",
  "trains": [
    {
      "trainNumber": "1009",
      "journeyRef": "null",
      "lineRef": "null",
      "sections": [
        {
          "departureDayShift": 0,
          "departureStationId": "8503424",
          "departureStationName": "Schaffhausen",
          "departureTime": "06:47:00",
          "destinationStationId": "8503000",
          "destinationStationName": "Zürich HB",
          "expectedDepartureOccupancy": [
            {
              "fareClass": "firstClass",
              "occupancyLevel": "fewSeatsAvailable"
            },
            {
              "fareClass": "secondClass",
              "occupancyLevel": "standingRoomOnly"
            }
          ]
        }
      ]
    }
  ]
}
```


10.6.2. Definition of Keys in an Occupancy Forecast Record in JSON Flavour

In the following table, we point out and discuss differences and extras with respect to the XML format. To avoid redundancy, we do not copy-paste any statements, but refer to the respective fields in the SIRI flavour sections above.

key name (bold=mandatory)	Contained in	type	description	example value
operatorRef	root object	string	See “OperatorRef” in SIRI flavour.	11
opDate	root object	string	See “DateFrameRef” in SIRI flavour.	2023-12-04

key name (bold=mandatory)	Contained in	type	description	example value
lastUpdated	root object	string	See “ResponseTimestamp” in SIRI flavour.	2023-12-01T09:57:22+01:00
timeToLive	root object	number	This field is supposed to give a hint about when new data sets will be available. In our data set, this will be a constant 86400 (seconds, 1 day).	86400
version	root object	string	The version of this data format (applying semantic versioning).	0.9
trains	root object	array	Array of train objects. See “EstimatedVehicleJourney” in SIRI flavour.	
trainNumber	train object	string	See “TrainNumberRef” in SIRI flavour.	1009
journeyRef	train object	string	See “DatedVehicleJourneyRef” in SIRI flavour.	
lineRef	train object	string	See “LineRef” in SIRI flavour.	
sections	train object	array	Array of sections of this train. Whereas SIRI uses a call perspective, we use a section perspective in this JSON format, in line with the data source (CAPRE). See the section on perspectives above.	
departure DayShift	section object	number	The shift of the departure with respect to the operation day. Value is -1, 0 or +1: • -1: departure is before midnight of the actual operation day, e.g., the day before the operation day. • 0: departure is on the actual operation day. Most common case. • +1: the departure is after midnight, e.g., the day after the operation day. Whereas SIRI works with absolute timestamps (date and times), we adopt the appraoch of the data source (CAPRE) here, using opDate, departureDayShift and departureTime.	-1, 0, 1
departure StationId	section object	string	See “StopPointRef” in SIRI flavour.	8504300
departure StationName	section object	string	See “StopPointName” in SIRI flavour.	Schaffhausen
departure Time	section object	string	The departure time (local Swiss time, CET or CEST) in hh:mm:ss format.	05:14:00

key name (bold=mandatory)	Contained in	type	description	example value
			Must be used in conjunction with the departureDayShift and the operation day to refer to a given date and time.	
destination StationId	section object	string	Same as departureStationId, but for the end of the section.	8503000
destination StationName	section object	string	Same as departureStationName, but for the end of the section.	Zürich HB
expected Departure Occupancies	section object	array	An array of expectedDepartureOccupancy objects. Corresponds to the two ExpectedDepartureOccupancy elements in SIRI flavour.	
fareClass	expected Departure Occupancy object	string	See “FareClass” in SIRI flavour.	firstClass
occupancy Level	expected Departure Occupancy object	string	See “StopPointName” in SIRI flavour.	fewSeats Available