

# ***Rail Delivery Group***

---



## **Darwin 'Publish and Subscribe' via the Rail Data Marketplace**

**Subject Ref: RSPS5053  
Version: P-01-00**

## Documentation Management

This documentation is published via the ASSIST website only.

The Version Control and Release Management of this documentation is managed by the Rail Delivery Group.

## Copyright

The copyright in this work is vested in Rail Settlement Plan Limited and the information contained herein is confidential. This work (either in whole or in part) must not be modified, reproduced, disclosed or disseminated to others or used for purposes other than that for which it is supplied, without the prior written permission of Rail Settlement Plan Limited. If this work or any part hereof is furnished to a third party by virtue of a contract with that party, use of this work by such party shall be governed by the express contractual terms between Rail Settlement Plan Limited which is a party to that contract and the said party. © 2024

## Version History

Version	Comments
01-00	Issued.

## Contents

<b>1.</b>	<b>Introduction .....</b>	<b>6</b>
1.1	Overview .....	6
1.2	Purpose .....	6
1.3	Scope .....	6
1.4	Related Subjects and references .....	6
<b>2.</b>	<b>Information on the Darwin PubSub data product.....</b>	<b>7</b>
<b>3.</b>	<b>Consuming data from the Darwin PubSub data product.....</b>	<b>7</b>
3.1	Getting credentials in RDM .....	7
3.2	Using a Kafka client to consume the data .....	7

## Terms and Definitions

Term	Definition
Certificate Authority (CA)	A company or entity that is responsible for attesting to the identity of users, computers, and organisations.
Darwin	Real time train running and related information.
GitHub	A platform for hosting code that has version control and allows for collaboration.
JSON	JavaScript Object Notation; a lightweight data-interchange format.
Kafka	Event streaming platform.
NRDP	National Rail Data Portal.
Offset Explorer	GUI application for managing and using Apache Kafka ® clusters.
PubSub	Publish and Subscribe.
RDM	Rail Data Marketplace.
SASL	Simple Authentication and Security Layer.
SSL	Secure Sockets Layer (provides secure internet communication).
XML	eXtensible Markup Language.
XSD	XML Schema file.

## 1. Introduction

### 1.1 Overview

- 1.1.1 This documentation describes how to interact with the Darwin 'Publish and Subscribe' (PubSub) mechanism via the 'Rail Data Marketplace' (RDM).
- 1.1.2 The Darwin PubSub mechanism differs from the offering via the National Rail Data Portal (NRDP).

### 1.2 Purpose

- 1.2.1 The documentation aims to provide information that enables users to consume messages from the Darwin PubSub data product in RDM.

### 1.3 Scope

- 1.3.1 This document:
  - Provides a description of the contents of the data.
  - Highlights differences between the PubSub product in RDM and the Darwin Push Port itself.
  - Details how to get credentials to consume the data in RDM.
  - Specifies how to consume data from the PubSub data product in RDM using a Kafka client.

### 1.4 Related Subjects and references

RDG Ref	Title	Usage
RSPS5051	Darwin Interface Specifications	Describes the interface specifications for obtaining and interpreting Darwin data via the Push Ports service and Live Departure Boards.

## 2. Information on the Darwin PubSub data product

- 2.1 The push port data contains data on the creation of, and changes to, train schedule records, together with train running predictions made by Darwin.
- 2.2 There are some differences between accessing the Darwin Push Port directly and via the RDM. For example, in order to get the timetable and reference data, as well as snapshots, these are provided as a separate file data product in the catalogue.
- 2.3 The Darwin Timetable Files that can be used in conjunction with the PubSub data product can be found in the RDM, under the name 'Darwin Timetable Files'.

## 3. Consuming data from the Darwin PubSub data product

### 3.1 Getting credentials in RDM

- 3.1.1 To view the 'access credentials' needed to consume the data, you must subscribe to the Darwin PubSub data product in RDM.
- 3.1.2 After subscribing to the data product, you will be able to see a set of access credentials consisting of a consumer key and consumer secret. The consumer key and consumer secret are synonymous with a username and password.
- 3.1.3 These access credentials are shown under the 'Pub/Sub' tab of the data product in RDM and are the same for all 'topics'.

### 3.2 Using a Kafka client to consume the data

- 3.2.1 The PubSub data product is using Kafka, an event streaming platform which in this instance is being used as a 'Publish and Subscribe' based messaging system.
- 3.2.2 There are currently three topics available, these are:
- The base topic.
  - A JSON topic.
  - An Avro topic.

From these topics, messages may be viewed as string, JSON, or Avro serialized messages.

- 3.2.3 A Kafka client will be needed to consume data in the topics, there are clients available in many languages. For more information on this, please visit: <https://docs.confluent.io/kafka-client/overview.html>
- 3.2.4 Along with the credentials from the RDM, when a Kafka consumer is created, the configuration details shown in the data product, such as the 'bootstrap server', the 'SASL mechanism', and 'security protocol' will be needed.
- 3.2.5 If a Java Kafka client is being used, then there is also a need to configure the SASL config in the consumer's configuration.
- 3.2.6 It is important to use the 'bootstrap server', the 'SASL mechanism', and 'security protocol' exactly as they are shown in the data product in RDM. Whilst other settings may work, they are not supported and use of them may not succeed in the future.

- 3.2.7 As the security protocol of the Kafka broker uses SSL, a CA certificate is required. In many operating systems there will be a default CA bundle in the operating system's certificate store. This location is dependent on your operating system but many resources regarding this are available online.
- 3.2.8 There may be a need to specify the location of the CA bundle in your file system for the 'ssl.ca.location'.
- 3.2.9 A code snippet is provided showing how to consume the data in the Darwin PubSub data product using the confluent-kafka client in Python. This will be updated as required so check the data product at the time of reading and regularly thereafter.
- 3.2.10 The code snippet can be found in the GitHub repository for the Rail Data Marketplace organisation here: <https://github.com/raildatamarketplace/rdm-darwin-kafka-client>.
- 3.2.11 There are also JSON schemas for the output of the data that can be found within the 'Reference Material' section of the data product.
- End.