

**Virgin Trains West Coast**

**PreProcessing Design specification**

PreProcesing Requirements Document

Gary Newsome

Technical Consultant

**July 2018**

z

Contents

[1. Document Management 2](#_Toc510603468)

[2. Document Purpose 4](#_Toc510603469)

[2.1 In Document Scope 4](#_Toc510603470)

[2.2 Out of Document Scope 5](#_Toc510603471)

[3. Project Scope 6](#_Toc510603472)

[3.1. Project Background 6](#_Toc510603473)

[3.2. Objectives 6](#_Toc510603474)

[3.3. Committed Obligations (CO) 7](#_Toc510603475)

[3.4. High Level Success Criteria 7](#_Toc510603476)

[3.4. Scope of Work 8](#_Toc510603477)

[3.5. Project Deliverables, Benefits, New Capabilities 9](#_Toc510603478)

[3.6 Workshops 12](#_Toc510603479)

[4. Out of Scope 14](#_Toc510603480)

[5. Requirements 15](#_Toc510603481)

[6. Data Quality Auditing 16](#_Toc510603482)

[6.1 Auditing Approach 16](#_Toc510603483)

[6.2 Auditing Summary 16](#_Toc510603484)

[7. Data Feed Summary 16](#_Toc510603485)

[7.1 Data Feeds 16](#_Toc510603486)

[8. Constraints 22](#_Toc510603487)

[8.1 Constraints 22](#_Toc510603488)

[8.1.1 Technical 22](#_Toc510603489)

[9.1.2 External Environment 23](#_Toc510603490)

[9.1.3 Industry, Security and Legislative 23](#_Toc510603491)

# Document Management

|  |  |  |  |
| --- | --- | --- | --- |
| **Version Control** | | | |
| **Version** | **Date** | **Author(s)** | **Sections Changed** |
| 0.1 | 01/07/2018 | Gary Newsome | Initial draft document |

| **Distribution List** | | |
| --- | --- | --- |
| **Organisation** | **Name** | **Role** |
| Merkle | Mark Jones | Client Lead / Project Manager |
| Merkle | Jez Cox | IBM Software SME |
| Merkle | Gary Newsome | Technical Consultant |
| Merkle | John Whittome | Business Analyst |
| Merkle | Peter Malherbe | AWS SME |
| Merkle | Mat Lynd | Development |
| Merkle | Juanjo Diaz | Technical consultant |
| Merkle | Steve Forster | Technical Solution Lead |
| Merkle | Avtar Aswell | Tester |

# Document Purpose

The purpose of this ***Preprocessing Design Document i***s to define the scope of how the solution will process data for each feed from the S3 buckets into the preprocessing tables. This document makes the assumption that he data has been transferred from the SFTP site to the S3 bucket for the feed being processed.

Every feed will follow the same processing descripted in this document, and should be used in conjunction with the Interface Dcouments that define the input file layouts and Processing table definitions.

The list of functional requirements that will be documented in this design are covered below in the scoping section.

Each preprocessing feed process will be developed as an individual SSiS package to enable scheduleing at different times depending on the file arrival details.

The load to staging will be dependent on the preprocessing for each feed to complete without error

Some of the feeds detailed in this document are subject to change as we discover feed details around the data and formats that will be received.

This document should be read in conjuction with the Technical Design document and the Business Requirements documents

## 2.1 In Document Scope

The following items are in scope for this document:

* Scheduling of SSiS package to poll for received data on the S3 buckets.
* Assigning of a DataImportLogID for each feed received to enable tracking of the feed load process through to Staging
* Assigning of a DataImportDetailID for each file within a feed to enable tracking at the file level of the load processing through to Staging
* A data flow diagram showing the steps that will be covered as part of the preprocessing
* Logging of the number of records loaded, rejected and loaded into the processing tables
* The details of the metadata required to support the data feed loads

## 2.2 Out of Document Scope

The following items are out of scope for this document and will be covered to separate design documents:

* The SFTP to S3 bucket transfer and logging
* The design to process records through to staging
* Any processing related to SCV
* Feeds that are ingested via an API
* Feeds that are ingested via IBM WCA or Unica processes
* User testing strategy and approach

# High Level Overview

## 3.1. Framework for Tracking Feed Processing

For each feed that is ingested via SFTP a number of key tables are used to manage the feed details and the feed processing, The operations ERD below shows how the tables that manage the feeds and load processes link together. These tables need to be populated as we onboard each feed and will be part of this feed design document



To initialise each feed process you need to call [PreProcessing].[CBE\_Delta\_Initialise] [Operations].[DataImportLog\_Initialise]

These tables need to be populated to track both incoming and outgoing data feeds

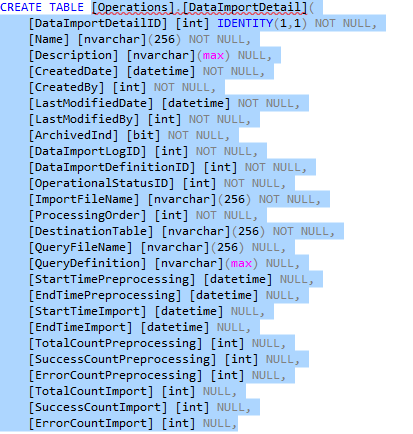
Add in error table to log rejected preprocessing tables

Columns

Record

File

Error reason



[ParsedAddressEmail] [nvarchar](100) NULL,

[ParsedEmailInd] [bit] NULL,

[ParsedEmailScore] [int] NULL,

[ParsedAddressMobile] [nvarchar](50) NULL,

[ParsedMobileInd] [bit] NULL,

[ParsedMobileScore] [int] NULL,

[ParsedAddressPhone] [nvarchar](50) NULL,

[ParsedPhoneInd] [bit] NULL,

[ParsedPhoneScore] [int] NULL,

[CreatedDateETL] [datetime] NOT NULL,

[LastModifiedDateETL] [datetime] NOT NULL,

[ProcessedInd] [bit] NOT NULL,

[DataImportDetailID] [int] NULL,

### InformationSource Table

This table holds the top-level source id of each source feed and is populated by running:

* EXEC [Reference].[InformationSource\_Set]

With the following parameters

Toc plus

|  |  |
| --- | --- |
| Parameter | Value(s) |
| @userid | 0 |
| @name | ‘TOC’,’Beam’ etc. |
| @desc | ‘This is the main feed as supplied by TrainLine’,’ Passengers using onboard entertainment service Beam’ |
| @displayname | ‘TOC’,’Beam’ |
| @typecode | ‘Internal’, ‘External’ |
| @prospectind | 0 or 1 |
| @addinfo | = NULL |
| @returnid | The allocated InformationSourceId will be returned |

This procedure will update an existing row where @name = reference.InformationSource.Name. For this reason all entries must have a unique name.

An identity column is used to create the unique InformationSourceID and is returned from the procedure

### DataImportType

This table holds the group name for a set of data files that are to be loaded. For example, feed ‘TOC’ would be the Informationsource, but we may split the load process into different schedules, the first schedule would load the customer records and a second schedule would load all the remaining files.

Generally, this table would hold one row per load schedule and is populated by running:

* + EXEC [Reference].[DataImportType\_Set]

With the following parameters

|  |  |
| --- | --- |
| Parameter | Value(s) |
| @userid | 0 |
| @name | ‘TOC Regular Import’,’Beam’ |
| @desc | ‘Used for processing information from TOC’,’ Used for processing information from Beam’ |
| @archivedind | 0 or 1, always 0 unless feed is no longer in use |
| @informationsourceid | The InformationSourceId returned when setting up the source feed above or NULL |
| @informationsourcename | This is the ‘Name’ of the InformationSourceId above or NULL |
| @returnid | The allocated InformationSourceId will be returned |

If the InformationSourceName is provided then the procedure will look up the InformationSourceId to populate the DataImportType table. If the informationSourceName is not provided and the informationSourceId is incorrect then the procedure will log an error and finish

This procedure will update an existing row where @name = reference.DataImportType.Name. For this reason, all entries must have a unique name.

An identity column is used to create the unique DataImportTypeId and is returned from the procedure

### DataImportDefinition

This table holds the individual feed file details for loading. For example, an entry will exist on DataImportType for ‘TOC Regular Import’ and the DataImportDefinition table would hold all the details for each file that makes up a ‘TOC Regular Import’, we would have an entry in this table for ‘Customer’, ‘Bookings’, ‘Journeys’ etc.

Generally, this table would hold one row per file received and is populated by running:

* + EXEC [Reference].[DataImporDefinition\_Set]

With the following parameters

|  |  |
| --- | --- |
| Parameter | Value(s) |
| @userid | 0 |
| @name | ‘TOC Customer’,’Beam Customer’ |
| @desc | ‘Used to support processing of Customer data in table TOC customer’,’ Used to support processing of Customer data in table Beam customer’ |
| @archivedind | 0 or 1, always 0 unless feed is no longer in use |
| @dataimporttypeid | The DataImportTypeId returned when setting up the DataImportType above or NULL |
| @dataimporttypename | This is the ‘Name’ of the DataImportType above or NULL |
| @querytemplate | No Longer used – MSD XML template |
| @processingorder | 1,2,3 etc. depending on order of processing files, for reference |
| @maxbatchsize | No longer user – MSD limitation |
| @destinationtable | Preprocessing target table |
| @querydefinition | No longer user – MSD XML |
| @typecode | ‘File’, ‘Table’, ‘API’ |
| @subquerydefinition | No longer used – MSD XML |
| @localcopyind | 0 or 1, Only set to 1 if data exists on a linked server |
| @returnid | The allocated DataImportDefinitionID will be returned |

If the DataImportTypeName is provided then the procedure will look up the DataImportTypeId to populate the DataImportDefinition table. If the DataImportTypeName is not provided and the DataImportTypeId is incorrect then the procedure will log an error and finish

This procedure will update an existing row where @name = reference.DataImportDefinition.Name. For this reason, all entries must have a unique name.

An identity column is used to create the unique DataImportDefinitionId and is returned from the procedure.

### Key Changes

These tables use an identity column to allocation InformationSourceId’s, DatadefinitionId’s etc. This has the potential to create Id’s that are different across environments and while not critical it can cause confusion.

When defining these entries in metadata it is suggested that we modify the \_Set procedures to pass in an ID, rather that the procedure allocating it, and control what feed gets each ID, the procedures will then be scripted and version controlled for release into the next environment.

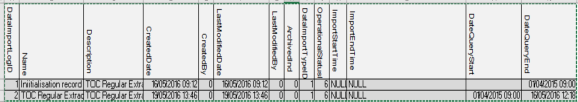
## DataImportLog and DataImportDetail

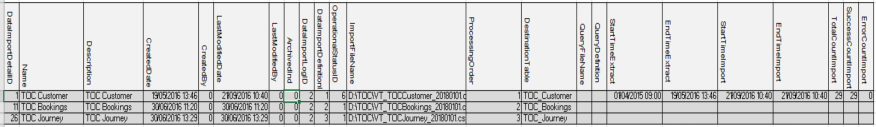
Before processing a feed, the procedure below is called, this populates the DataImportLog table with the feed group that is to be processed. It then populates the DataImportDetailLog with a row for each file within the feed group that is to be processed.

EXEC [PreProcessing].[CRM\_Delta\_Initialise]

@userid = **0**,

@dataimporttype = 'TOC Regular Import'





DataImportLog table records the start and end of the batch run for each source group. It also logs the current state of the processing by logging the OperationalStatusId

|  |  |  |
| --- | --- | --- |
| OperationalStatusID | Name | Description |
| 1 | Pending | Initial status before any processing is initiated |
| 2 | Retrieving | Status when data is being pulled back from an external system |
| 3 | Retrieved | Status when data pull back from an external system has completed successfully |
| 4 | Imported | Status when data from a file has been loaded successfully into the database |
| 5 | Processing | Status when data is being processed in the database |
| 6 | Completed | Final status when all processing of data has completed successfully |
| 7 | Failed | Status when processing has errored |

The procedure PreProcessing.CRM\_Delta\_Initialise checks to see whether a current load is already running for an individual source feed based on the OperationalStatusId above

DataImportDetail table records the start and end of each individual file load within the batch run for a given source group. It logs the current state of the processing by logging the OperationalStatusId so that it’s clear to see where a process hasn’t completed and the current state of the load process

* Pending – This is the initial state for both DataImportLog and DataImportDetail tables
* Retrieving – This is only relevant where a process is pulling from an external source, this was used to manage the XML pull for MSD feeds on VTEC.
* Retrieved – This is the starting state for a feed that is being loaded to the preprocessing tables
* Imported – This is the state for processes that have completed loading into preprocessing and awaiting further processing
* Processing – this is the state when a feed is being processed from preprocessing and loading into the main CRM database staging tables
* Completed – all processing for this feed is complete
* Failed – used to log where a feed has failed to process through to CRM database

The StartTimeExtract and EndTimeExtract track when the data is loaded into the preprocessing tables.

The StartTimeImport and EndTimeImport track when the load to CRM Staging started and ended.

For each individual feed, we will record the path and filename of the file being processed, we will also log the total records, records loaded to staging and rejected records.

Before processing a feed for the first time, we will need to create an initialise row with operationalstatusid = 6 in both DataImportDetail and DataImportLog tables as the current process uses this to create the next row for that source/feed

The DataImportDetail and DataImportLog table should also be used to record data that is being extracted and send to 3rd parties.

Virgin Trains East Coast (VTEC) has now successfully dev/eloped and deployed a Customer Experience Management (CEM) System to deliver both sales and service communications to its customers. The CEM platform has been developed and is operated by Merkle (formerly Comet).

VTWC currently uses the Trainline to provide its core CRM capability; specifically, its customer database, campaign management and reporting capability.

VTWC has now approved a project to replace its current Trainline CRM capability with a new CRM System that will be developed and operated by Merkle. This is known internally to VTWC as Project Phoenix.

One of the objectives of the project is to re-use as much of the IP, designs and System built for VTEC as is appropriate, in order to reduce both timescales and costs for the development of the VTWC System.

The proposed platform is inclusive of campaign management, campaign reporting and underlying marketing database layers.

The project comprises:

* Design and build of environments to meet platform requirement
* Data migration, matching and deduplication of historic data and ongoing feeds into a new VTWC single customer view database
* Migration of existing CRM campaigns/programmes and the creation of new ones
* Migration of a subset of KPI and campaign performance reports from VTEC into Tableau Dashboards
* Provision of customer and transactional data into VTWC BI teams for the purpose of insight and analytics
* Creation of service structure, including definition of service and support model

The service and support model is expected to include:

* CRM database management, maintenance and optimisation
* Implementation and maintenance of system hardware and software stack, including regular software patching
* Creation and management of campaigns/programmes
* Creation of KPI and campaign performance dashboards
* Data management and archiving policy management

## 3.2. Objectives

The overarching objective of the CRM Platform implementation (Project Phoenix) is to deliver the Business Case associated with the project, including the delivery of significant revenue uplift. The details of this are commercially sensitive, but key objectives for the project are outlined below.

* Leverage the learnings, designs and campaigns/reports from the VTEC CEM system wherever appropriate to reduce costs/timescales for VTWC
* Enable VTWC to build more sophisticated and dynamic campaigns with higher conversion through multiple communication channels
* Enable VTWC to lower costs by leveraging economies of scale by working more closely with VTEC
* Deliver the first stage of migration from Trainline which will de-risk potential future investments in non-Trainline capability
* Future proof the VTWC CRM System and enrich the VTWC data to build a single customer view
* Improve the facilitation of train disruption related communications to passengers in-journey

## 3.3. Committed Obligations (CO)

There are no CO’s associated with Project Phoenix although an expectation has been set with the Department of Transport that it will deliver an improvement of messaging to customers during disruption.

The specific parameters that have been referenced are that customers affected by disruption will receive a relevant communication within an hour of the disruption occurring.

## 3.4. High Level Success Criteria

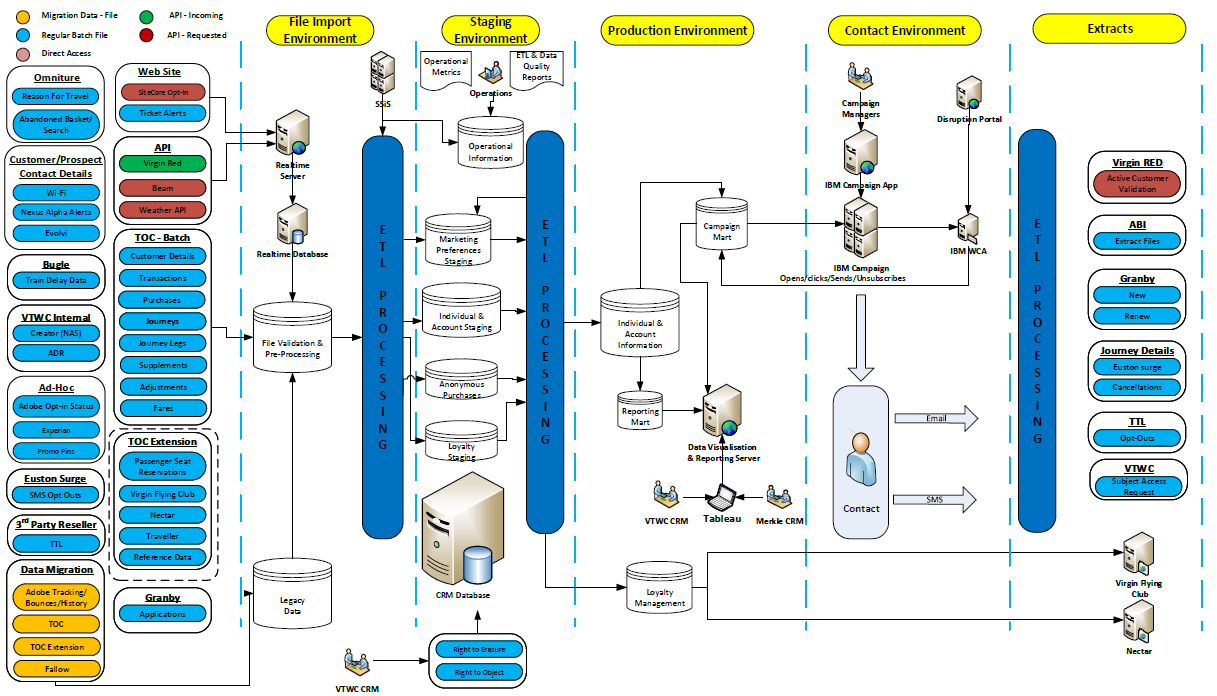
The System will be considered a success if the following success criteria are met:

|  |  |  |  |
| --- | --- | --- | --- |
| **Ref** | **Description** | **Measure** | **Evidenced By…** |
| S-01 | The System maintains the integrity of the data | * Validate and reject data where appropriate * Handles data exceptions * Identifies missing scheduled data files * Provides reporting on reject, exception and variance | Feed Monitoring Logs |
| S-02 | Data is staged and stored efficiently | * Data is maintained and kept up to date * Data is stored in an efficient manor * Data is secure and auditable * Data is relevant and available | Technical Specification Document |
| S-03 | Single Customer View created and specification documented | * Data is recorded in an interface specification which details data ingested by the System at field level * Logical system design is defined * Data linage is detailed (source to target mapping) | Technical Specification Document |
| S-04 | Campaigns delivered | * Ability to deliver communications via email and SMS * Campaigns/Programmes can be defined, mapped, scheduled and delivered * Campaign automation capability/ability to deliver campaigns based on customer behavioural triggers * Emergency campaigns can be delivered by VTWC with a response time of 1 hour, as and when required 24/7 | IBM Watson Marketing Performance Reports |
| S-05 | Business Intelligence Reporting delivered | * Standard reports are developed and delivered for BAU\* * Tableau is configured to display agreed dashboards and ad-hoc and scheduled reports to be generated   *\*See VTWC Report Migration Template for in-scope reports* | Presence of Tableau Dashboards |
| S-06 | History Migration | * Applicable campaign history is migrated to the CRM Database contact history table (dependent on TTLs provision of this) * Applicable booking history is migrated into CRM Database * History is usable in the campaign management process | Reporting Data Mart, IBM Watson Marketing |
| S-07 | Migration of campaigns from TTL to Merkle | * Service up and running * In scope campaigns defined in the Campaign Migration report are live * Campaign logic preserved or improved across the migration * Unsubscribes honoured in both directions * De-commissioning of TTL service | Presence of campaign segments, workflows, creative in IBM Watson Marketing |
| S-08 | Transition to BAU | * Database management, campaign and reporting activities are effectively transitioned to Business As Usual operations | Ongoing monitoring of BAU operations |
| S-09 | System Compliancy | * The solution is deemed by VTWC to be compliant from an information security perspective, taking into account any risks\* accepted by VTWC as part of the Project * The solution is deemed by VTWC to be GDPR compliant, taking into account any risks\* accepted by VTWC as part of the Project   *\*Please see ‘VT – DPIA Template for Phoenix Merkle\_20180321’ for details* | Technical Specification Document, Master Services Agreement |

## 3.4. Scope of Work

The following High-Level Design diagram shows the scope of the System, as defined during the discovery process.

Note that the proposed High-Level Design featured below is for illustrative purposes. Full and confirmed technical details of the final solution will be contained in the Technical Specification Document delivered as part of the Design phase.



## 3.5. Project Deliverables, Benefits, New Capabilities

The following shows what will be delivered by Project Phoenix, with an indication of either the new capability or business benefits that it will provide to VTWC.

|  |  |  |
| --- | --- | --- |
| **Category** | **Deliverable** | **VTWC Benefit/New Capability** |
| **Data Integration** | Customer data integrations with CRM System via batch or API (Inbound and Extract/Outbound feeds) | * Automation and centralisation of marketing information to power an enhanced VTWC customer experience, Including Trainline and WiFi data\* * Migration of legacy customer & transaction data from TTL (including Beam), and onboard and station Wi-Fi data * Consumption of customer registration details captured via webform to facilitate ticket alert campaigns (as per VTEC solution) * Consumption of Bugle data to support pre-departure comms and also insight into the impact of delays on customer behavior (as per VTEC solution)   *\*Please see the Data Feeds Category within the* [Business Requirements Document in Section](#_Requirements) 5 *for full details of in-scope data feeds.* |
| 3rd Party Booker data integration | * Integration of 3rd Party Booker data\* with the CRM System, to facilitate 3rd party booker specific:   + Communications   + Reporting   + Fallow groups   + Opt-in preference updates   + GDPR compliance fulfilment (SARs, Right to Erasure, Right to Object etc)   *\*Please see the ‘3rd Party Booker Assumptions’ tab within the* [Business Requirements Document in Section](#_Requirements) 5 *for known 3rd Party Booker Data Assumptions* |
| **Data Management** | Master Data Management | * Standardisation and verification of customer contact information * A single Customer View database (SCV) that can be used to facilitate segmentation and communication at different audience levels * Increased accuracy of customer contact targeting * Automated audit reporting of customer data integrity and data feed failures. This will enable proactive management in this area * A data model configuration enabling the partitioning of 3rd party booker data from 1st party |
| CRM Database and data accessibility | * Flexible and scalable industry specific data model leveraging IP Merkle developed for VTEC * Key customer marketing aggregates that are optimised for fast use/accessibility in the campaign management System - Travel information, customer value, existing VTWC customer segmentation and other behaviour based segmentation * Centralised contact and response history |
| **Marketing Preferences Management** | Customer preference data management | * Multi-channel and nuance marketing preference master data management (opt-in preferences, customer volunteered information) * Multiple channel entity and audience levels supported (e.g. email only, 1 per account etc) |
| **CRM Toolset** | Campaign Management including Outbound Campaign distribution (via IBM Watson Marketing tool) | * Best in class CRM tool * Email and SMS channel management and deployment capability (service based comms only for SMS) * Barcodes in emails * Increased marketing automation for programmes and campaigns * Event/Trigger based marketing capability * Campaign execution success monitoring and alerts * Send time optimisation * Facilitation of near real-time follow up capability, allowing for ad-hoc comms delivery (see below) * Capture of Customer Volunteered Information via landing pages |
| Self-Service campaigns – ‘Disruption Portal’ | * Capability for VTWC staff to access SMS and Email templates to deploy high priority messages around service disruption (the ‘Disruption Portal’) to customers 24/7 without reliance on supplier process and availability * The VTWC target is to contact customers within one hour (subject to the agreed business process) |
| Promotions capability | * Centralised process/capability for managing the import, allocation & recording of TTL generated PINS * Ability to insert PINS into both promotional email campaigns and automated campaigns |
| **Campaigns and Communications** | Core Campaign set-up | The following programme types are in scope for the development phase, with the specific campaigns to be developed and timings agreed during the design phase:   * Welcome / Nursery * Abandoned Basket/Search * Pre-Departure * Reactivation * NAS Follow-ups * Value Emails I&R * Anniversary * Retention * IP Warming   New campaign development approaches will be adopted within CRM to both increase the amount of automation surrounding campaign execution, and to reduce the time/effort required to create/maintain new campaigns. |
| **Measurement Capability** | Measurement Strategy | VTWC will implement VTEC’s measurement strategy as part of Project Phoenix, including:   * Sales Attribution business rules * Fallow and control strategy and business rules |
| **Reporting and Data visualisation** | Reports & dashboards (via Tableau tool) | KPI and Campaign Performance reports will be created in Tableau based on VTEC’s existing set of Dashboards  Please see the ‘In Scope Reports + Amends’ tab within the [Business Requirements Catalogue](#_Requirements) for full details |
| **Loyalty Currency Calculation** | Loyalty Currency management | * Capability to calculate Virgin Flying Club or Nectar currency from bookings * Automated daily feed of points information to VFC and Nectar * Audit trail of points and reconciliation of spend capability in CRM System to identify value add to VTWC * Incentivise individual customers with bonus points/miles for specific activity e.g. buying a ticket in the next 5 days, purchasing an upgrade, selecting a digital ticket etc. |
| **Operational Business Process** | CRM operating model | * Efficient end-to-end campaign delivery process defined and documented with creative agency buy-in * Centralised Data feed audit and management function * Application and issue management support * CRM System governance and change management best practice |
| **Non-Functional Requirements** | NFRs | * System designed to be GDPR compliant in line with the criteria defined in the Master Services Agreement and DPIA * Secure cloud based hardware that is easily scalable * Software application support from the team that installed and configured it * Proactive performance tuning as the System scales |

## 3.6 Workshops

The below table catalogues the workshops and stakeholder engagement sessions undertaken and the attendees of each workshop.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref.** | **Event** | **Date** | **Detail** | **Attendees** |
| WS-1 | Kick-off Session | 26/02/18 | Project overview, ways of working, introductions | Nick Dodd – Project Director  Matt Hey – CRM Consultant Mark Murphy – Head of CRM Fiona Osborne – Project Manager Mark Jones – Client Lead/Project Manager  John Whittome – Business Analyst |
| WS-2 | GDPR Session 1 | 02/03/18 | Initial session to discuss high level GDPR requirements, ascertain GDPR analysis work undertaken by VTWC | Callum McCormick – GDPR and InfoSec Fiona Osborne – Project Manager Mark Jones – Client Lead/Project Manager  John Whittome – Business Analyst |
| WS-3 | GDPR Session 2 | 02/03/18 | Initial session to discuss high level GDPR requirements, ascertain GDPR analysis work undertaken by VTWC | Callum McCormick – GDPR and InfoSec  Stephen Montgomery – Solution Architect Fiona Osborne – Project Manager Mark Jones – Client Lead/Project Manager  John Whittome – Business Analyst |
| WS-4 | Information Security | 20/03/18 | Session for Merkle to respond to VTWC information security questions | Shaun Roberts - TBC  Steven Montgomery – Data Architect Jason Burch – Director, Engineering & Architecture  Peter Malherbe – Lead Data Architect  Fiona Osborne – Project Manager John Whittome – Business Analyst |
| WS-5 | GDPR Session 3 | 21/03/18 | Follow up session to discuss high GDPR requirements, permissions analysis work undertaken by VTWC | Callum McCormick – GDPR and InfoSec  Stephen Montgomery – Solution Architect Fiona Osborne – Project Manager John Whittome – Business Analyst |
| WS-5 | GDPR Session 4 | 28/03/18 | Follow up session to go through outstanding GDPR related points | Callum McCormick – GDPR and InfoSec  Stephen Montgomery – Solution Architect Fiona Osborne – Project Manager John Whittome – Business Analyst |

# Out of Scope

The following high-level requirements have been discussed with the relevant business stakeholders, and deemed to be out of scope of Merkle’s System delivery specifically:

|  |  |  |
| --- | --- | --- |
| **OOS Ref.** | **Out of Scope Item** | **Description** |
| OOS-1 | Preference Centre Support | The creation or support of an online preference/permissions centre |
| OOS-2 | VTWC Manual Opt-in Updates | The facility for Merkle and VTWC resources to be able to change channel opt-ins for individual customers when requested to do so |
| OOS-3 | IBM Interact (Inbound Marketing) | Real-time Decisioning Software and capability to send personalised offers to VTWC customers on demand (Omni-channel) |
| OOS-4 | Push Messaging | Using IBM to send Push messages |
| OOS-5 | Operational emails | Booking confirmation, seat reservation etc. |
| OOS-6 | Provision of CRM content for website | API’s the provide customer preferences, permissions or any other details |
| OOS-7 | Staff Mobile App / Spotlight | API Incoming |
| OOS-8 | Social Media data integration | Data Feed |
| OOS-9 | Journey Data to support delay repay | API call |
| OOS-10 | “NRS” or Capacity Information | Batch feed |
| OOS-11 | Analysis | Provision of analysis or aggregated analytical data (as opposed to reporting/dashboards) platform |
| OOS-12 | Creation of TRACS Reference Data | The creation of TRACS reference data by Merkle is OOS. Assumed to be provided ready to use by VTWC |
| OOS-13 | Non- website sourced Loyalty Currency processing | The CRM System will not process Loyalty Currency derived from any other sources but the VT.com website |
| 00S- 14 | Amazon Alexa integration | Alexa integration is a pilot due in May 2018 that allows customers to buy their train tickets through their Alexa unit. This is OOS for this project. |
| OOS-15 | Contact Centre Data Integration | Integration of Contact Centre data with the CRM System is OOS |
| OOS-16 | Euston Surge SMS | The CRM System will not deliver Euston Surge SMS comms |
| OOS-17 | Processing of ADR payments | The CRM System will not process Automated Delay Repayment (ADR) payments |
| OOS-18 | Real-time disruption data sources | The CRM System will not integrate with Darwin or any other real-time disruption data source |
| OOS-19 | VTWC and VTEC customer matching | The CRM System will not match customers across VTEC and VTWC |
| OOS-20 | Provision of data to support re-targeting | The CRM System will not generate any data to support retargeting activity |
| OOS-21 | NAS Surveys | The send of NAS surveys to customers and the data capture of their responses is out of scope (this will remain with Creator) |
| OOs-22 | Rich Messaging Service (enhanced SMS) | Any type of SMS message other than basic text with the ability to support links to webpages |

# Requirements

The following sections contain reference to the requirements gathered from the business users during workshops and interviews with VTWC and Merkle.

**Please see the embedded Excel spreadsheet below for the full catalogue of detailed requirements.**



Whilst subject to change, it is expected that those developments required to achieve BAU Run Phase are included in the Design, Build, and Test phases following Discovery.

The requirements identified as out of scope will **not** be taken forward into the Design stage, unless explicitly requested by VTWC. Where this occurs, the Change Request process will be used to assess the impact of any amended requirements, and estimates will then be provided.

# Data Quality Auditing

## 6.1 Auditing Approach

The primary objective of Data Quality Auditing is to ‘sanity check’ a typical sample of the data that will eventually be loaded into the System.

* Checks the supplied data is valid
* Ensures the required data is available
* Measures the referential integrity of the supplied data
* Confirms the supplied data can be used to populate the proposed marketing data model

The Data Quality Auditing also shows the effect of applying the various proposed customer matching rules – as well as showing various other alternative rules that could also be applied.

## Auditing Summary

The auditing summary is provided as a supplemental document, ‘VTWC Data Feeds and Findings’.

**Note:** The Data Quality Auditing approach does **not** include the categorising of data into Personally Identifiable Information (PII), Personal Data, Sensitive Personal Data etc. This General Data Protection Regulation (GDPR) requirement will be addressed during the Design stage.

# 7. Data Feed Summary

## 7.1 Data Feeds

The following list is a list of in scope data feeds based on current understanding, and may be subject to edits based on lower level detail determined during the Design Phase.

Full requirement details are contained within the [Business Requirements Document](#_Requirements).

|  |  |  |  |
| --- | --- | --- | --- |
| **Ref** | **Name** | **Description** | **Type** |
| F-01 | TOC | This feed contains all of the primary information from the booking engine from the previous day and includes travel details and contact details for customers. This file originates from TRACS and will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-02 | TOC Extension | In addition to the TOC feed, a TOC Extension feed is required to capture the necessary fields that were previously supplied via TOC+ (soon to be deprecated). This is expected to contain data including Seat Reservations, Virgin Flying Club, Nectar, Traveller, and reference data. Full details are TBC during the Design Phase. | Inbound Feed |
| F-03 | Adobe | Captures the Adobe delivered campaign customer activity post new CRM System live date. This will contain contact history - opens, clicks, opt-outs, bounces. This feed will be live for a maximum of 1 month following the last Adobe deployed campaign. | Inbound Feed |
| F-04 | Omniture | This feed contains Reason for Travel and website data which is used to facilitate campaigns including Abandoned Search & Abandon Basket. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-05 | Virgin Red (API) | This is used to verify the email address of customers who sign-up to the Virgin Red app. This API also sends additional behavioural segmentation (RFM) to Virgin Red for verified customers. Data is transferred in real time. | API |
| F-06 | Evolvi (Railblazers) | This contains sign-ups for the Business SME booking engine via the Rail Blazers platform provided by Evolvi. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-07 | Nectar PAF | This will inform Nectar which customers have booked a qualifying journey and need to receive Nectar points. Feed frequency and full details TBC. | Outbound Feed |
| F-08 | Euston Surge (SMS) | To facilitate the sending of platform notifications in advance to customers at Euston Station. This file contains the contact details for qualifying customers who should receive the text, along with additional information such as origin, destination, seat number, coach number, name and how many people are travelling. Data will be transferred via SFTP every 24 hours. | Outbound Feed |
| F-09 | Euston Surge (SMS) | Inbound Euston Surge data feed featuring SMS opt out preference and the mobile phone number associated with the preference. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-10 | Virgin Flying Club | Informs VFC which customers have booked a qualifying journey and are due to receive VAFC points. Data will be transferred via SFTP every 24 hours (TBC). | Outbound Feed |
| F-11 | Onboard Customer Wi-Fi Data (New Supplier) | New Wi-Fi data feed for both Onboard Wi-Fi. This feed is expected to be live from May/June 2018. Data will include:   o Customer name / email o Marketing permission o Time / Date of login o Train identifier  o Carriage / Class identifier o login/usage | Inbound Feed |
| F12 | In-station Customer Wi-Fi Data (New Supplier) | New Wi-Fi data feed for In-Station Wi-Fi. This feed is expected to be live from May/June 2018. Data will include:   o Customer name / email o Marketing permission o Time / Date of login o RSID  o login/usage | Inbound Feed |
| F-13 | Creator (NAS) | inbound NAS Survey feed. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-14 | Granby Traveller | Granby Traveller data tells Granby (agency who manages Traveller) who has met the qualifying criteria and should be invited to the Traveller club. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-15 | Granby Traveller | This will contain data New and Renew customer data to tell The Trainline who has actively taken up the offer of a membership and successfully signed up to Traveller. Data will be transferred via SFTP every 24 hours. | Outbound Feed |
| F-16 | SiteCore Opt-In API | This API contains marketing opt-in information from the VTWC homepage. The data refreshes in real time. | API |
| F-17 | Nexus Alpha Journey Check Alerts | This is data from a third-party system which facilitates communications to passengers to let them know if their train is on time, delayed or cancelled. Nexus Alpha send a file to Trainline, Trainline create a new account for any passengers who are not on the VT database. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-18 | ADR | ADR gives automatic refunds to customers who have been delayed for more than 30 minutes. Refunds take between 5 - 7 working days to be processed by Trainline. This feed facilitates an automated email which tells customers that they ae eligible. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-19 | BEAM (Go Media) | This contains details of customers who have watched BEAM on a VT service. If customers are not on the VT database they are stored and fall under the Marketing Prospect Category i.e. they receive Beam Welcome, Nursery, etc. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-20 | Promotion PINS | PINS are generated by Trainline and allow VT to discount Advanced tickets on a VT service. This feed will be sent on an ad-hoc basis. | Inbound Feed |
| F-21 | Tableau (ABI) | Implement an outbound feed of Tableau data to include anonymised customer / transaction / contact history / response history for the purposes of facilitating analysis via ABI. Data will be transferred via SFTP every 24 hours. | Outbound Feed |
| F-22 | Experian Feed (to support segmentation) | This feed is needed to support segment creation. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-23 | Weather API | This is used in Pre-departure campaigns to display the weather of the destination for travellers. | API |
| F-24 | Cancellations | Used for the facilitation of train cancellation SMS communications. | Outbound Feed |
| F-25 | Cancellations | Cancellations data feed featuring SMS opt out preference and the mobile phone number associated with the preference. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-26 | Ticket Alerts | This contains data for customers who have signed up for ticket alerts and offers. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-27 | Train Delay Data (Bugle) | Bugle Train Delay Data. Data will be transferred via SFTP every 24 hours. | Inbound Feed |
| F-28 | TTL Campaign Data | This is for Merkle to provide Opt-outs and bounce data to The Trainline from IBM WCA on an ongoing basis. Full format and frequency details to be confirmed during the Design phase. | Outbound Feed |
| F-29 | 3rd Party booker data | This will be a single, daily batch file sent from The Trainline containing all the relevant details for a customers booking. This will be provided via SFTP. | Inbound Feed |

### 8. Constraints

This section of the BRD collates the various constraints and assumptions imposed on the System.

## 8.1 Constraints

### 8.1.1 Technical

1. Merkle will define the technical stack of third party software and data that will be used to build the System. This will be selected to meet the requirements of the System and development/test skillset.
2. Where possible, the System will be designed to be standalone, meaning that the System should be:
   1. Extensible, so that new data can be easily accommodated.
   2. Portable, with the Intellectual Property (IP) rights for the System residing with the Franchise.

Merkle will provide the source code for all non-proprietary aspects of the System, together with the corresponding suite of technical documentation that fully describes the design.

### 9.1.2 External Environment

1. All feed files will be supplied to Merkle via SFTP file transfer or API
2. All feed files will adhere to the agreed schedule, format and content, as documented in the Feed File Specification.
3. With the exception of access to Tableau, VTWC (or its appointed representatives) will only directly access the System as necessary, for auditing and compliance validation purposes.

### 9.1.3 Industry, Security and Legislative

1. The System will comply with the guidance, principles and obligations of the Data Protection Act 1998, (<https://www.gov.uk/data-protection/the-data-protection-act>), in particular only processing data as specified and directed by VTWC.
2. The System will be demonstrably compliant with all obligations for the Data Processor in the forthcoming General Data Protection Regulations. (The regulations are due to come into force on 25/05/18 and the UK Government recently confirmed the UK would be signing up to this legislation).
3. Merkle will use GDPR Privacy by Design principles wherever appropriate, in order to protect customer data and to ensure the privacy of Data Subjects is managed in accordance with their expectations.