# 



DNA Source System Data Integration Interface Specifications  
- DTAC

Table of Contents

[1. Overview 6](#_Toc25144889)

[1.1 Interface Diagram 6](#_Toc25144890)

[1.2 File Feed List 6](#_Toc25144891)

[1.3 File Transfer Mechanism 6](#_Toc25144892)

[1.4 ODS Table Naming convention 7](#_Toc25144893)

[2. Assumptions 8](#_Toc25144894)

[3. Interface Details 9](#_Toc25144895)

[3.1 Details of Source Feeds 9](#_Toc25144896)

[3.1.1. Source Feed Files 9](#_Toc25144901)

[3.1.2. Source Feed wise - Record Format 11](#_Toc25144902)

[3.1.3. Control Feed - Record Format 11](#_Toc25144903)

[3.1.4. Privacy Columns/Business Rules 12](#_Toc25144904)

[4. Interface SLAs & Methods 13](#_Toc25144905)

[4.1 Availability 13](#_Toc25144906)

[4.2 File Validation 13](#_Toc25144907)

[4.2.1 Duplicate File Check 13](#_Toc25144908)

[4.2.2 Missing file from Source 13](#_Toc25144909)

[4.2.3 Missing records or Incomplete File 14](#_Toc25144910)

[4.2.4 Alerts 14](#_Toc25144911)

[4.2.5 Invalid file format 15](#_Toc25144912)

[4.2.6 Erroneous data 15](#_Toc25144913)

[4.3 Source System Changes 15](#_Toc25144914)

[4.4 Unable to Collect or receive files from source 15](#_Toc25144915)

[4.5 Unavailability of Data Collection & Integration Layer 15](#_Toc25144916)

[4.6 Data Quality and Timeliness 16](#_Toc25144917)

[4.7 Exception Handling 16](#_Toc25144918)

[5. Appendix 17](#_Toc25144919)

[5.1. Email Notification for missing file: 17](#_Toc25144925)

**Revision History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Revision Date** | **Author / Reviewer** | **Summary of changes** | **Remarks** |
| 0.1 | 29.07.2019 | Hirak /Chandrima | First Draft Version | For DTAC Review |
| 1.0 | 05.08.2019 | Hirak Sen | 1) DTAC 1set of Review taken dare.  2) Derived fields removed as per DTAC’s Comments.  3) Added the ‘File Feed List’ Table with file length Size | For DTAC Review |
| 1.1 | 21.08.2019 | Hirak Sen | Inclusion of 2nd set of Review comments shared by DTAC | For DTAC Review |
| 1.2 | 06.09.2019 | Hirak Sen | Inclusion of review comments shared by DTAC during Review Session of ISD feeds | For DTAC Review |
| 1.3 | 18.11.2019 | Hirak/Chandrima | Landing server path corrected as per revision done in BUName\_SourceSystemSummary\_DTAC\_V 0.1.1.xlsx | For DTAC Review |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Review and Approval**

|  |  |  |  |
| --- | --- | --- | --- |
| **Person** | **Role** | **Responsibility** | **Date** |
| Vorraluck Sarechuer Vorraluck.sarechuer@dtac.co.th | Delivery Manager - DTAC |  |  |
| Porntip Soponchai PorntipS@dtac.co.th | Sr Manager -DTAC |  |  |
| Rohit Umaraw  rohit.umaraw@wipro.com | Manager- Wipro |  |  |

|  |  |
| --- | --- |
| Project Name | TELENOR - AEP |
| Account | Telenor - DTAC |
| Current Version | 1.3 |
| Author | Chandrima / Hirak |

Referenced Documents

| **#** | | **Document Name** | **Document Description** | **Document Type/Drop#** | **Version** |
| --- | --- | --- | --- | --- | --- |
| 1 | BUName\_SourceSystemSummary\_DTAC\_V 0.0.9.xlsx | | This document captures the source feed technical details and its data structures | Excel File | 0.9 |
| 2 | AEP - Source System Ingest Data.xlsx | | This document contains source system data ingestion path and server for DEV/SIT/PROD | Excel File | - |
| 3 | SourceFeed1.xlsx | | Holds EDW file wise control file mapping per As-IS BI Process | Excel File | - |
| 4 | BUName\_SourceSystemSummary\_DTAC\_V 0.1.1.xlsx | | This document captures the source feed technical details and its data structures | Excel File | 1.1 |
|  |  | |  |  |  |

Terminologies & Acronyms

| **Acronyms Used** | **Description** |
| --- | --- |
| ODS | Operational Data store |
| AEP | Analytics Enablement Program |
| DNA | Device Profile System |
|  |  |

# Overview

This document contains information and specification for Source System (DNA) for Data Integration into ODS layer. The purpose of this document is to describe the design of output interfaces from DNA (Source system) to AEP platform (Data Integration).

The specification document will capture all the source feed details, their frequencies, naming conventions and their corresponding rules like exception handling, transformation rule, filtration rule if any, surrogate key and encryption logic.

## 1.1 Interface Diagram

TIER - 1

ODS – Vertica Vertica

AEP Landing Path

DNA

ETL TOOL

ODS- Hadoop

Pull

EDW Landing Path(pladius22)

## 1.2 File Feed List

The below table captures the list of files which will be integrated from DNA as part of first phase of ODS go-live.

| **#** | **Feed Name** | **File Name** | **File Format** | **Source System Notification Mail\_Group** | **Source System Contact Point** |
| --- | --- | --- | --- | --- | --- |
| 1 | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT | .DAT | DNA DEVICE | [Atthakorn@dtac.co.th](mailto:Atthakorn@dtac.co.th) |

## 1.3 File Transfer Mechanism

DNA source system will push the data into EDW Landing server-pladius22, AEP has to pull the data from pladius22 to AEP landing zone. Broadly, the following steps will be carried out as part of the file transfer mechanism:

1. DNA will generate two files one with .DAT extension and other with .CTL extension for the control files for each feed per day. Following are the lists of files

| **Data File Name** | **Data File Format** | **Control File Name** | **Control File Format** |
| --- | --- | --- | --- |
| <YYYYMMDD>\_DEVICE\_PROFILE.DAT | .DAT | <YYYYMMDD>\_DEVICE\_PROFILE.CTL | .CTL |

1. File will be transferred from EDW Landing area-on pladius22 using FTP protocol in uncompressed form.
2. There will be separate folder created for each day where the files will be Pulled inside AEP- Landing zone. Folder structure as below:

* For Raw Data files path should be /<root>/SRC\_DATA/DNA/<YYYYMMDD>/
* For Control File path should be /<root>/SRC\_DATA/DNA/<YYYYMMDD>/

1. As soon as file gets picked for processing, it will be moved to processed folder. Once the file is processed successfully, the original file will be moved to ‘Archive’ folder where the file will be retained as per data retention period. Archive folder will contain all the raw files.
2. No Header & footer available inside the feed files.
3. Source system is responsible to provide the correct data.
4. All transaction files will have transaction data for the previous day.
5. Available Source file path @ EDW Landing zone is mentioned in the below table-

| **#** | **Feed Name** | **File Name** | **Source File path - PROD** | **Source File path - SIT** | **Source File path-DEV** |
| --- | --- | --- | --- | --- | --- |
| 1 | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT | /EDW/nfsedw105/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_DEVICE\_PROFILE.DAT | /SIT/EDW/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_DEVICE\_PROFILE.DAT | /DEV/EDW/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_DEVICE\_PROFILE.DAT |

## 1.4 ODS Table Naming convention

* Transaction Table : DWO\_<Application\_System>\_<feed\_name>
* Dimension / Master Table : DIM\_<Application\_System>\_<feed\_name>

# Assumptions

1. Master data file will be Pushed to the EDW landing path on a daily basis even if there is no new data also.
2. For multiple transaction files which are of same feed will be available once in a day to the EDW landing path from there it will be pulled to AEP landing.
3. Files are required to be send to existing platform and new platform concurrently until AEP is stabilize. However Existing file formats remains the same in both the platform.
4. File count validation cannot be done for source feed which are not having any control file. No Process in As-is DTAC application.
5. This Source data server path details are considered as is with current BI system, any change in setup will have impact on the data ingestion configuration mechanism.

# Interface Details

## 3.1 Details of Source Feeds

|  |  |
| --- | --- |
| Interface Number |  |
| Interface Name | **DNA** |
| Interface Owner | Refer to section 1.2 on feed wise Ownership details |



## Source Feed Files

The below table captures the list of files which will be integrated from DNA as part of first phase of ODS go-live.

| **#** | **Feed Name** | **File Name** | **File Format** | **File Name Length** | **File Type** | **Frequency** | **Mechanism of File Transfer** | **No. of files EOD** | **DELIMITER** | **Extract type** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT | .DAT | 50 | Transaction | DAILY | Pull | 1 | | | Incremental |

#### 3.1.1.1 Interface Characteristic

|  |  |
| --- | --- |
| Source Feed Name | All source feeds |
| Source Feed Description | See Section 3.1.2 on Source feed mapping |
| File Type  (Transaction / Reference) | Transaction |
| Collection Mechanism (Push / Pull) | Pull |
| Extraction Criteria | Daily Full Dump/ Incremental [ Check section 3.1.1 on individual feed wise details] |
| Collection Protocol | **SFTP** |
| File frequency /Collection Frequency | Refer to Section 4.1 for details. |
| File Format | “|” delimited File.  File extension is .DAT |
| Control File available? (Yes / No) | Yes [refer to Section 3.1.3 for detail]] |
| Header Record Available? | No |
| Footer Record Available? | No |
| Retention Period at Source/ File Archival policy | 7 days [ File Based] |

#### 3.1.1.2 Environment Details & Access Details

|  |  |
| --- | --- |
| Production IP Address | TBD |
| Production Port | TBD |
| Production User Name | TBD |
| Production Source Folder | Server : Pladius22  Data file Path : /EDW/nfsedw105/SRC\_DATA/DNA/YYYYMMDD/<YYYYMMDD>\_ DEVICE\_PROFILE.DAT  Control file Path : /EDW/nfsedw105/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_  DEVICE\_PROFILE.CTL |
| SIT IP Address | TBD |
| SIT Port | TBD |
| SIT User Name | TBD |
| SIT Source Folder | Server : nickel18  Data file Path : /SIT/EDW/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_DEVICE\_PROFILE.DAT  Control file Path : /SIT/EDW/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_DEVICE\_PROFILE.CTL |
| Dev IP Address | TBD |
| Dev Port | TBD |
| Dev User Name | TBD |
| Dev Source Folder | Server : nickel17  Data file Path : /DEV/EDW/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_DEVICE\_PROFILE.DAT  Control file Path : /DEV/EDW/SRC\_DATA/DNA/<YYYYMMDD>/<YYYYMMDD>\_DEVICE\_PROFILE.CTL |

#### 3.1.1.3 Information required for first time loading & data transfer

For Transaction feeds the data from ODS will start from date of GO-LIVE of ODS.

Historical data loading and ODS Data retention strategy captured in the below table

| **#** | **Feed Name** | **File Name** | **File Format** | **ODS Historical Load** | **ODS Data Retention** |
| --- | --- | --- | --- | --- | --- |
| 1 | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT | .DAT | All | No Purge |

#### 3.1.1.4 Header Record Format

Header not available for this feed.

#### Footer Record Format

Footer not available for this Feed.

#### File count Check

Refer to Section 3.1.1 on Source feed files

#### Feed to Table Mapping

Following table depicts the proposed DNA feeds mapping to TIER 1 (ODS) target table.

| **#** | **Feed Name** | **TIER1 (ODS) Table Name** | **Data Volume Per Day** | **Type** |
| --- | --- | --- | --- | --- |
| 1 | DEVICE\_PROFILE | DWO\_DNA\_DEVICE\_PROFILE | 11850944 | Transaction |

## Source Feed wise - Record Format

#### DWO\_DNA\_DEVICE\_PROFILE

The feed file will hold current Device (IMEI) for each mobile number

| **S.NO** | **Field Name** | **Data Type** | **Sample Values** |
| --- | --- | --- | --- |
| 1 | MSISDN | VARCHAR(15) | 66897766923 |
| 2 | IMSI | VARCHAR(15) | 520050390929126 |
| 3 | IMEI | VARCHAR(20) | 866988046639174 |

## Control Feed - Record Format

Following are the mapping of Raw data file and Feed control file associated with each.

| **#** | **Feed Name** | **Data File Name** | **Control File Name** | **File Format** | **Control File Format** |
| --- | --- | --- | --- | --- | --- |
| 1 | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT | <YYYYMMDD>\_DEVICE\_PROFILE.CTL | .CTL | 7 |

Following is the structure of the DNA feed wise (Type # 7) control files-

|  |  |  |
| --- | --- | --- |
| **Filename** | **Record Count** | **File Checksum** |
| 20190806\_DEVICE\_PROFILE.DAT | 16418854 | 95c199ba35ce5d38bb003c31691ecc3e |

## Privacy Columns/Business Rules

No data filtration and no exception handling for any columns. Direct one to one mapping and loading to ODS. Surrogate key will be generated only for MSISDN.

Only following type of fields will be encrypted while loading to ODS tables. Attached excel contains DNA feed wise consolidations

* Group 1 - MSISDN / Subscriber Number /Phone no /Fax No
* Group 2 - IMSI\_IMEI
* Group 3 - ADDRESS
* Group 4 - Name
* Group 5 – ID Key
* Group 6 - Email
* Group 7 - Credit Card Number
* Group 8 - Bank Account ID



# Interface SLAs & Methods

## 4.1 Availability

DNA related transaction data will be pulled daily once into AEP landing server from EDW- Landing server - pladius22 using SFTP protocol. AEP has to pull from EDW Landing path the files.

| **#** | **Feed Name** | **File Name** | **Arrival Time** | **AEP Landing path - Data files** | **AEP Landing path - Control files** |
| --- | --- | --- | --- | --- | --- |
| 1 | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT | 1:00:00 AM | /<root>/SRC\_DATA/DNA/<YYYYMMDD>/ | /<root>/SRC\_DATA/DNA/<YYYYMMDD>/ |

## 4.2 File Validation

Following are the high-level validations being performed on master and transaction data feeds -

1. Orphan File check
2. Duplicate file check
3. Zero byte check
4. Header and Trailer
5. Record Count
6. Delimiter check
7. MD5 file Checksum (Applicable for DNA feeds)

### 4.2.1 Duplicate File Check

If the same file is received at landing area for the same day, the process has to mark it duplicate (FileName\_Duplicate.DAT) and reject the file.

* Audit table will record name of each file. Compare the file name with previous 30 days filename (configurable), to identify duplicate files.
* Alerts for Duplicate file name to Source system owner over the email once a day (configurable).
* Move duplicate files in reject folder

### 4.2.2 Missing file from Source

An alert will be send to source owner if file is not received in given time frame.

In case of files which are coming in sequence and there is a sequence missing in file, an alert will be sent for missing file sequence.

Email Notification for missing file has been shared by DTAC as per below template



PS: Please see Section – 5 Appendix for email attachment.

### 4.2.3 Missing records or Incomplete File

Reconciliation can be done with source system as able to validate between feed file and control file available with every feed. Reconciliation can also be done based on source file wise count and the data loaded to ODS.

### 4.2.4 Alerts

An alert mechanism will be in place to notify source and business owners in case of deviation from processes.

An Email will be triggered to source owners when

* Less number of Files received in expected time.
* When Files will be pushed to rejected folder.

\*\*A Governance process will be set to address any issues in the file transfer process including manual intervention when necessary.

**Action to be taken by source owner after alert - Once Notification alert is sent, Source owner should check and resend the correct data within 4 hours after the receipt of alert.**

**Sample Email Content for Missing file :-**

|  |  |  |
| --- | --- | --- |
| Source System | Source Feed | File Name |
| DNA | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT |

**For respective list of Feed wise file name refer to section 3.1.1**

**Sample Email Content for Reject file :-**

|  |  |  |  |
| --- | --- | --- | --- |
| Source System | Source Feed | File Name | Reject Reason |
| DNA | DEVICE\_PROFILE | <YYYYMMDD>\_DEVICE\_PROFILE.DAT | Name of file is not correct |

**For respective list of Feed wise file name refer to section 3.1.1**

**Sample Email for Less number of files.**

|  |  |  |  |
| --- | --- | --- | --- |
| Source System | Source Feed | Previous Day File Count | Today’s File Count |
| DNA | DEVICE\_PROFILE | 1 | 0 |

**For respective list of Feed wise file name refer to section 3.1.1**

### 4.2.5 Invalid file format

Files which don’t follow standards of file naming convention will be rejected. For example

* Prefix of file needs to be checked e.g. for all source file
* File should always have the .DAT extension. Rest all other files with different extension will be rejected.

### 4.2.6 Erroneous data

In case of any mandatory field missing, datatype mismatch, record separator is not proper, junk VARCHAR coming in any field or file name is wrong, then file will not be processed. In such a case notification to be sent for error file and source system need to correct the file and push the correct file into landing path.

## 4.3 Source System Changes

In case of any changes in file format or data format, Source team will inform team well in advance. Any new column addition or changes in existing feed format will be considered as Change Request.

Guidelines for adding new columns are as follows:

**Addition/modification/deletion of columns in file:**

1. When new columns added after registered\* columns in the file. File handling process will ignore the newly added columns (until the columns are registered in the metadata table for that file ). File will not be rejected by the processes; however, these columns will be ignored for processing.
2. If columns are added in between registered\* columns the file will be rejected.
3. If number of columns received are lesser than registered\* columns the file will be rejected.

## 4.4 Unable to Collect or receive files from source

If the files are not received then after one hour<configurable> auto email alert will be sent to the source system owner.

## 4.5 Unavailability of Data Collection & Integration Layer

In case of unavailability of Data Collection and Integration layer, landing server will keep the files until Data Collection and Integration layer is restored.

However, in case of capacity issues DTAC will notify source owners, to stop pushing files to landing area and manage the files at source end until normalcy is restored.

## 4.6 Data Quality and Timeliness

It would be responsibility of DNA team to have some files/data level checks before files are made available Platform Consumption. Source system team will inform in case there is any issues in DNA provided files.

## 4.7 Exception Handling

Exception handling is subject to scenarios. Different scenarios will be handled differently:

All such files those are required to be reprocessed will be kept in the same source folder from where files are to be collected.

For source files which do not have date-timestamp, duplicate file identification will not be possible. In such case if source push the file twice then the previous file will get overridden and latest overridden file will be processed.  
  
For re-processing of a file which has already been processed and data loaded into tables in such a case we will delete / unload the data from tables by identifying the data based on FILE\_ID and LOAD\_DATE.  
  
In case the file gets pushed with different name then notification via email / or on call will be sent so that the file can be considered for re-processing.

# Appendix



## Email Notification for missing file:

