# 



USSD Source System Data Integration Interface Specifications  
- DTAC

Table of Contents

[1. Overview 7](#_Toc25155914)

[1.1 Interface Diagram 7](#_Toc25155915)

[1.2 File Feed List 7](#_Toc25155916)

[1.3 File Transfer Mechanism 7](#_Toc25155917)

[1.4 ODS Table Naming convention 9](#_Toc25155918)

[2. Assumptions 10](#_Toc25155919)

[3. Interface Details 11](#_Toc25155920)

[3.1 Details of Source Feeds 11](#_Toc25155921)

[3.1.1 Source Feed Files 11](#_Toc25155922)

[3.1.2 Source Feed wise - Record Format 13](#_Toc25155923)

[3.1.3 Control Feed - Record Format 17](#_Toc25155924)

[3.1.4 Privacy Columns/Business Rules 17](#_Toc25155925)

[4. Interface SLAs & Methods 19](#_Toc25155926)

[4.1 Availability 19](#_Toc25155927)

[4.2 File Validation 19](#_Toc25155928)

[4.2.1 Duplicate File Check 19](#_Toc25155929)

[4.2.2 Missing file from Source 19](#_Toc25155930)

[4.2.3 Missing records or Incomplete File 20](#_Toc25155931)

[4.2.4 Alerts 20](#_Toc25155932)

[4.2.5 Invalid file format 21](#_Toc25155933)

[4.2.6 Erroneous data 21](#_Toc25155934)

[4.3 Source System Changes 21](#_Toc25155935)

[4.4 Unable to Collect or receive files from source 21](#_Toc25155936)

[4.5 Unavailability of Data Collection & Integration Layer 22](#_Toc25155937)

[4.6 Data Quality and Timeliness 22](#_Toc25155938)

[4.7 Exception Handling 22](#_Toc25155939)

[5. Appendix 23](#_Toc25155940)

[5.1. Email Notification for missing file: 23](#_Toc25155946)

[5.2. USSD Transaction Log ODS script <<trn\_ussd\_fileid.sh>> 23](#_Toc25155947)

[5.3. USSD CDR ODS script << MoveUssd.sh>> 23](#_Toc25155948)

[5.4. DataStage job << P\_TX\_USSD\_CDR\_D>> for USSD\_CDR processing 23](#_Toc25155949)

**Revision History**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Revision Date** | **Author / Reviewer** | **Summary of changes** | **Remarks** |
| 1.0 | 09.09.2019 | Hirak/Chandrima | First Version Submission to DTAC | For DTAC Review |
| 1.1 | 10.09.2019 | Hirak/Chandrima | DTAC review comments incorporated | For DTAC Review |
| 1.2 | 12.09.2019 | Hirak/Chandrima | DTAC review comments incorporated | For DTAC Review |
| 1.3 | 13.09.2019 | Hirak/Chandrima | DTAC review comments incorporated | For DTAC Review |
| 1.4 | 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@datac.co.th | Delivery Manager -DTAC |  |  |
| Porntip Soponnchai PorntipS@dtac.co.th | Sr Manager - DTAC |  |  |
| Rohit Umaraw  Rohit.umaraw@wipro.com | Manager - Wipro |  |  |

|  |  |
| --- | --- |
| Project Name | TELENOR - AEP |
| Account | Telenor - DTAC |
| Current Version | 1.4 |
| Author | Hirak Sen |

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 |

# Overview

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

USSD

ETL TOOL

ODS- Hadoop

EDW Landing Path(nickel4)

Pull

## 1.2 File Feed List

The below table captures the list of files which will be integrated from USSD 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 | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) | .txt | NS-ENTERPRISE INFRASTRUCTURE OPERATION TEAM (ES DEPT) BSS-SERVICE OPERATION TEAM (CRM DEPT) <BSS-SERVICEOPERATIONTEAMCRMDEPT@dtac.co.th> | Rangsun Chaicompa <Rangsun@dtac.co.th> |
| 2 | USSD\_CDR | 99XXXXX\_YYYYMMDD.DAT  ( ex. 9911035195\_20190804.DAT ) | .DAT | BSS-BILLING OPERATION TEAM (BS DEPT) <BSS-BILLINGOPERATIONTEAMBSDEPT@dtac.co.th> | Akapon Buachum <AkaponB@dtac.co.th> |
| 3 | UMB103 | UMB103\*<YYYYMMDD>.txt (ex, UMB103\_ussd103nm1\_2019071700.txt) | .txt | SP-RefillPaymentSOTeam@dtac.co.th | ChatpanP@dtac.co.th |

## 1.3 File Transfer Mechanism

USSD will push the data into EDW Landing server-nickel4 , AEP has to pull the data from nickel4 to AEP landing zone. Broadly, the following steps will be carried out as part of the file transfer mechanism:

1. USSD will generate **USSD\_ALL\_TRANS** files with .txt extension and **USSD\_CDR** files with .DAT extension. Following are the lists of files

| **Data File Name** | **Data File Format** | **Control File Name** | **Control File Format** |
| --- | --- | --- | --- |
| USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt, USSDReward\_20190804.txt ) | .txt | No Files | NA |
| 99XXXXX\_YYYYMMDD.DAT ( ex. 9911035195\_20190804.DAT ) | .DAT | No files | NA |
| UMB103\*<YYYYMMDD>.txt (ex, UMB103\_ussd103nm1\_2019071700.txt) | .txt | No files | NA |

1. File will be transferred from EDW Landing area-on nickel4 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:
   1. For Raw Data files path should be

* /<root>/SRC\_DATA/USSD/USSD\_LOG/<YYYYMMDD>/
* /<root>/SRC\_DATA/CDR/MD/USSD\_REFILL/<YYYYMMDD>/
  1. For Control File path should be
* No Control files available.

1. Both the feed has multiple transaction files and as per agreement the same will be Pushed to the EDW landing path from where AEP will pull it inside secured DI Zone for processing.
2. 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.
3. Header & footer available inside the **USSD CDR** feed files while processing both should be ignored.
4. Source system is responsible to provide the correct data.
5. All transaction files will have previous day records of processing day.
6. All .DAT files are fixed length and .txt files are delimited files.
7. 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 | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) | /DISK\_POOL/dskp408/canopus4/fs\_stage/cdr/ussd/data\_input/USSD\*\_<YYYYMMDD>.txt | /SIT/EDW/SRC\_DATA/CDR/USSD/<YYYYMMDD>/USSD\*\_<YYYYMMDD>.txt | /DEV/EDW/SRC\_DATA/CDR/USSD/<YYYYMMDD>/USSD\*\_<YYYYMMDD>.txt |
| 2 | USSD\_CDR | 99XXXXX\_YYYYMMDD.DAT  ( ex. 9911035195\_20190804.DAT ) | /DISK\_POOL/dskp407/MP/APP/USSD\_REFILL/<YYYYMMDD>/99\*\_<YYYYMMDD>.DAT | /SIT/EDW/SRC\_DATA/CDR/USSD/USSD\_RFLL/<YYYYMMDD>/9911006866\_<YYYYMMDD>.DAT | /DEV/EDW/SRC\_DATA/CDR/USSD/USSD\_RFLL/<YYYYMMDD>/9911006866\_<YYYYMMDD>.DAT |
| 3 | UMB103 | UMB103\*<YYYYMMDD>.txt (ex, UMB103\_ussd103nm1\_2019071700.txt) | /DISK\_POOL/dskp408/canopus4/fs\_stage/cdr/ussd/data\_input/ UMB103\*\_<YYYYMMDD>.txt | /SIT/EDW/SRC\_DATA/USSD/UMB/<YYYYMMDD>/ UMB103\*\_<YYYYMMDD>.txt | /DEV/EDW/SRC\_DATA/USSD/UMB/<YYYYMMDD>./ UMB103\*\_<YYYYMMDD>.txt |

1. There are Multiples files available as part of USSD Transaction Log **(USSD\_ALL\_TRANS)** source however for these feeds first 22 columns are coming from source rest all columns to be ignored, for details refer to section 3.1.2.1.
2. In case of **USSD\_CDR** source system pushes files near real-time, but the process currently run once a day to populate ODS tables, AEP processing can follow similar lines.
3. On **USSD\_CDR** feeds currently in EDW there is script moving file into date wise folder (refer to section 5.3 for the script file), which pulls data from EDW landing Source Path = /DISK\_POOL/dskp407/MP/APP/USSD\_REFILL (no date folder) and underlying filenames are like 99????????.DAT ( ex. 9911035195.DAT ). The same script renames the file with suffixing <<date>> with feed files then moves into date-wise folder path as mentioned in the point # 10 above (refer to column Source File path – PROD).
4. Once EDW landing zone has been descoped and source feed will be directly pushed to DI Zone, AEP processing could need similar workaround on date wise folder creation and stamping date with files. Now AEP can refer files from existing as is EDW landing zone path and filename that are mentioned in the point # 10 above (refer to column Source File path – PROD).

## 1.4 ODS Table Naming convention

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

# Assumptions

* Master data file will be Pushed to the EDW landing path daily even if there is no new data also.
* For multiple transaction files which are of same feed will be Pulled once in a day from the EDW landing path.
* 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.
* File count validation cannot be done for source feed which are not having any control file.  
  No process currently in DTAC.
* 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 | **USSD** |
| 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 USSD as part of first phase of ODS go-live.

| **#** | **Feed Name** | **File Name** | **File Format** | **File Type** | **File Name Size** | **Frequency** | **Mechanism of File Transfer** | **Expected min no of files EOD** | **DELIMITER** | **Extract Type** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) | .txt | Transaction | 50 | DAILY | Pull | - | | | Increment |
| 2 | USSD\_CDR | 99XXXXX\_YYYYMMDD.DAT  ( ex. 9911035195\_20190804.DAT ) | .DAT | Transaction | 50 | DAILY | Pull | - | Fixed length | Increment |
| 3 | UMB103 | UMB103\*<YYYYMMDD>.txt (ex, UMB103\_ussd103nm1\_2019071700.txt) | .txt | Transaction | 50 | DAILY | Pull | 1 | | | Increment |

#### 3.1.1.1 Interface Characteristics

|  |  |
| --- | --- |
| Source Feed Name | All source feeds |
| Source Feed Description | See Section 3.1.2 on Source feed mapping structure |
| 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 | Fixed length file for USSD\_CDR whereas USSD\_ALL\_TRANS is having ‘|’ (pipe) delimiter being used. |
| Control File available? (Yes / No) | No [refer to Section 3.1.3 for details] |
| Header Record Available? | Yes [ Only for USSD CDR transactions]; USSD\_ALL\_TRANS doesn’t have any Header record |
| Footer Record Available? | Yes [ Only for USSD CDR transactions]; USSD\_ALL\_TRANS doesn’t have any Header record |
| Retention Period at Source/ File Archival policy | 7 days – USSD\_ALL\_TRANS [ File Based]  3 days – USSD\_CDR [ 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 : nickel4  For Data files – USSD\_ALL\_TRANS  /DISK\_POOL/dskp408/canopus4/fs\_stage/cdr/ussd/data\_input/USSD\*\_<YYYYMMDD>.txt  USSD CDR -  /DISK\_POOL/dskp407/MP/APP/USSD\_REFILL/<YYYYMMDD>/99\*\_<YYYYMMDD>.DAT |
| SIT IP Address | TBD |
| SIT Port | TBD |
| SIT User Name | TBD |
| SIT Source Folder | Server : nickel18  For Data files – USSD\_ALL\_TRANS  /SIT/EDW/SRC\_DATA/CDR/USSD/<YYYYMMDD>/USSD\*\_<YYYYMMDD>.txt  USSD CDR-  /SIT/EDW/SRC\_DATA/CDR/USSD/USSD\_RFLL/<YYYYMMDD>/9911006866\_<YYYYMMDD>.DAT |
| Dev IP Address | TBD |
| Dev Port | TBD |
| Dev User Name | TBD |
| Dev Source Folder | Server : nickel14  For Data files – USSD\_ALL\_TRANS  /DEV/EDW/SRC\_DATA/CDR/USSD/<YYYYMMDD>/USSD\*\_<YYYYMMDD>.txt  For USSD CDR  /DEV/EDW/SRC\_DATA/CDR/USSD/USSD\_RFLL/<YYYYMMDD>/9911006866\_<YYYYMMDD>.DAT |

#### 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 | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) | .txt | ALL records | 6 months |
| 2 | USSD\_CDR | 99XXXXX\_YYYYMMDD.DAT  ( ex. 9911035195\_20190804.DAT ) | .DAT | N | 6 months |
| 3 | UMB103 | UMB103\*<YYYYMMDD>.txt (ex, UMB103\_ussd103nm1\_2019071700.txt) | .txt | N | 6 Months |

#### 3.1.1.4 Header Record Format

Only for USSD CDR transactions header is present and It’s like a string representation

‘10T01201908042019080400142101002’ or ‘10P01201908042019080400142101002’.

No clue about its logic however, length is 32 character as found in sample source feed.

#### Footer Record Format

Only for USSD CDR transactions header is present and It’s like a string representation

‘9000142001312001449’ or ‘9000142001312001249’. No clue about its logic however, length is 19 character as found in sample source feed.

#### File count Check

Refer to Section 3.1.1 on Source feed files for column “Expected min no of files EOD”. In case for any feed the value is blank are exempted from the file count validation check.

#### Feed to Table Mapping

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

| **#** | **Feed Name** | **TIER1 (ODS) Table Name** | **Data Volume Per Day** | **Type** |
| --- | --- | --- | --- | --- |
| 1 | USSD\_ALL\_TRANS | DWO\_USSD\_ALL\_TRANS | 2795064.7 | Transaction |
| 2 | USSD\_CDR | DWO\_USSD\_CDR | 3748373.04 | Transaction |
| 3 | UMB103 | DIM\_USSD\_UMB103 | 172,704.18 | Transaction |

### Source Feed wise - Record Format

##### DWO\_USSD\_ ALL\_TRANS

Feed stores USSD log transaction

| **S.NO** | **Field Name** | **Data Type** | **Sample Values** |
| --- | --- | --- | --- |
| 1 | TIME\_STAMP | DATETIME | 2019-08-04 01:23:53,329 |
| 2 | FUNTION | VARCHAR2(3) | 0 |
| 3 | USER1 | VARCHAR2(12) | SMSGW2 |
| 4 | REQIP | VARCHAR2(15) | 10.117.0.10 |
| 5 | USETIMR | VARCHAR2(10) | 306 |
| 6 | SUBRNUMB | VARCHAR2(20) | 66805600042 |
| 7 | STATUS | VARCHAR2(2) | 0 |
| 8 | ERRCODE | VARCHAR2(4) | 2 |
| 9 | ERRMSG | VARCHAR2(100) | - |
| 10 | RETRY | INTEGER | 0 |
| 11 | OTHER | VARCHAR2(100) | \*191# |
| 12 | OTHER2 | VARCHAR2(100) | 66805600042\_1564856632981 |
| 13 | SALE\_CODE | VARCHAR2(10) |  |
| 14 | RTR\_CODE | VARCHAR2(10) |  |
| 15 | SALE\_MOBL | VARCHAR2(20) |  |
| 16 | PACK\_ALIS | VARCHAR2(100) |  |
| 17 | PACK\_CODE | VARCHAR2(8) |  |
| 18 | APLY\_DATE | DATETIME |  |
| 19 | APLY\_TIME | VARCHAR2(5) |  |
| 20 | SUBR\_TYPE | VARCHAR2(1) |  |
| 21 | COMP\_CODE | VARCHAR2(2) | Dtn |
| 22 | SYS\_NM | VARCHAR2(5) | Cbs |

**Notes:**

* There are Multiples files available as part of USSD\_ALL\_TRANS source however all these feeds first 22 columns are coming from source rest all columns to be ignored. As-is process only consider the first 22 columns and add another derived column named as “USSD\_CODE”.
* The below table depicts respective feed file wise additional number of columns available under sample data. Columns grey-ed out means not exists in respective feeds. The one which is having values marked only has 25 columns. Rest all are having 23 columns.

|  |  |  |  |
| --- | --- | --- | --- |
| Column Count | 23 | 24 | 25 |
| **USSD\_LOG\_File\_name** | **Filler1** | **Filler2** | **Filler3** |
| USSDReward\_20190804.txt | P |  |  |
| USSD19X\_20190804.txt | P |  |  |
| USSD\_20190804.txt | P | NULL | NULL |
| USSDFsp\_20190804.txt | P |  |  |

* Refer to the STM document for the derived column population logic. The existing script is attached in **the section 5.2 USSD Transaction Log ODS script <<trn\_ussd\_fileid.sh>>** which does the merging and stamping of USSD codes (Derived Column) depending on different <USSD\_<<FileName>>.txt>.

##### DWO\_USSD\_CDR

Feed stores USSD CDR transaction

| **S.NO** | **Field Name** | **Data Type** | **Sample Values** |
| --- | --- | --- | --- |
| 1 | RCRD\_TYPE | VARCHAR2(2) | 40 |
| 2 | SUBR\_NUMB | VARCHAR2(12) | 66947817056 |
| 3 | IMSI\_NUMB | VARCHAR2(15) | 520050317251184 |
| 4 | IMEI\_NUMB | VARCHAR2(20) | 359498096530560 |
| 5 | ARTM\_TYPE | INTEGER | 0 |
| 6 | APPL\_ID | INTEGER | 0 |
| 7 | ORGT\_SWID | VARCHAR2(15) | 520050301119910 |
| 8 | ORGT\_CELL\_ID | VARCHAR2(13) | 19910 |
| 9 | TERM\_SWID | VARCHAR2(15) | 3011 |
| 10 | TERM\_CELL\_ID | VARCHAR2(13) | 0 |
| 11 | DIAL\_DIGT | VARCHAR2(24) | 104 |
| 12 | CALL\_STRT\_DATE | DATE | 20190804 |
| 13 | CALL\_STRT\_TIME | INTEGER | 1312 |
| 14 | MD\_DURN | INTEGER | 0 |
| 15 | DURN | INTEGER | 0 |
| 16 | DROP\_CALL\_FLAG | CHAR(1) | N |
| 17 | PIN\_FLAG | CHAR(1) | N |
| 18 | DRTN | CHAR(1) | 0 |
| 19 | CALL\_FRWD\_RESN | CHAR(2) |  |
| 20 | CDR\_CLSE\_RESN | CHAR(2) | 0 |
| 21 | TERM\_OPER | CHAR(2) |  |
| 22 | TERM\_AREA | CHAR(1) |  |
| 23 | CDR\_STTS | CHAR(1) |  |
| 24 | FILE\_NAME | VARCHAR2(10) | 9911035213 |
| 25 | FILE\_RCRD\_SEQN | VARCHAR2(5) | 1 |

**Notes:**

* AS-is DataStage job **<< P\_TX\_USSD\_CDR\_D>>** for USSD\_CDR processing, there are conditions for checking record length- 197. AEP processing should also consider the same. The outline of such job has been placed in the appendix section under 5.4.
* File wise Fixed Length column position for record length- 197 are as follows for necessary classification.

| **Column Name** | **Starting Position** | **No of character** |
| --- | --- | --- |
| RCRD\_TYPE | 1 | 2 |
| SUBR\_NUMB | 3 | 12 |
| IMSI\_NUMB | 15 | 15 |
| IMEI\_NUMB | 30 | 20 |
| ARTM\_TYPE | 50 | 5 |
| APPL\_ID | 55 | 5 |
| ORGT\_SWID | 60 | 15 |
| ORGT\_CELL\_ID | 75 | 13 |
| TERM\_SWID | 88 | 15 |
| TERM\_CELL\_ID | 103 | 13 |
| DIAL\_DIGT | 116 | 24 |
| CALL\_STRT\_DATE | 140 | 8 |
| CALL\_STRT\_TIME | 148 | 6 |
| MD\_DURN | 154 | 6 |
| DURN | 160 | 12 |
| DROP\_CALL\_FLAG | 172 | 1 |
| PIN\_FLAG | 173 | 1 |
| DRTN | 174 | 1 |
| CALL\_FRWD\_RESN | 175 | 2 |
| CDR\_CLSE\_RESN | 177 | 2 |
| TERM\_OPER | 179 | 2 |
| TERM\_AREA | 181 | 1 |
| CDR\_STTS | 182 | 1 |
| FILE\_NAME | 183 | 10 |
| FILE\_RCRD\_SEQN | 193 | 5 |

* AS-is DataStage job **<< P\_TX\_USSD\_CDR\_D>>** for USSD\_CDR processing, there are conditions for checking record length- 194. AEP processing should also consider the same. The outline of such job has been placed in the appendix section under 5.4.
* File wise Fixed Length column position for record length- 194 are as follows for necessary classification.

| **Column Name** | **Starting Position** | **No of character** |
| --- | --- | --- |
| RCRD\_TYPE | 1 | 2 |
| SUBR\_NUMB | 3 | 12 |
| IMSI\_NUMB | 15 | 15 |
| IMEI\_NUMB | 30 | 20 |
| ARTM\_TYPE | 50 | 5 |
| APPL\_ID | 55 | 5 |
| ORGT\_SWID | 60 | 15 |
| ORGT\_CELL\_ID | 75 | 13 |
| TERM\_SWID | 88 | 15 |
| TERM\_CELL\_ID | 103 | 13 |
| DIAL\_DIGT | 116 | 24 |
| CALL\_STRT\_DATE | 140 | 8 |
| CALL\_STRT\_TIME | 148 | 6 |
| MD\_DURN | 154 | 6 |
| DURN | 160 | 12 |
| DROP\_CALL\_FLAG | 172 | 1 |
| PIN\_FLAG | 173 | 1 |
| DRTN | 174 | 1 |
| CALL\_FRWD\_RESN | 175 | 2 |
| CDR\_CLSE\_RESN | 177 | 2 |
| \*\*TERM\_OPER | - | - |
| \*\*TERM\_AREA | - | - |
| CDR\_STTS | 179 | 1 |
| FILE\_NAME | 180 | 10 |
| FILE\_RCRD\_SEQN | 190 | 5 |

\*\* marked columns will be not there in the above feed file structure

##### DWO\_USSD\_UMB103

Feed stores USSD CDR transaction

| **S.NO** | **Field Name** | **Data Type** | **Sample Values** |
| --- | --- | --- | --- |
| 1 | TRNS\_DTTM | VARCHAR2(30) |  |
| 2 | SUBR\_NUMB | VARCHAR2(20) |  |
| 3 | TRNS\_ID | VARCHAR2(50) |  |
| 4 | ACCUM\_ELAPSED\_TM | NUMBER(18,0) |  |
| 5 | ELAPSED\_TM | NUMBER(18,0) |  |
| 6 | COMMAND | VARCHAR2(200) |  |
| 7 | OFFER\_CD | INTEGER |  |
| 8 | ACCPT\_FLG | VARCHAR2(1) |  |
| 9 | THIRD\_PARTY\_CD | VARCHAR2(1) |  |
| 10 | TRNS\_STTS | SMALLINT |  |
| 11 | LOG\_CD | INTEGER |  |
| 12 | SALE\_CD | VARCHAR2(30) |  |
| 13 | RTR\_CD | VARCHAR2(30) |  |
| 14 | SALE\_MOBL\_NUMB | VARCHAR2(20) |  |
| 15 | PKG\_CD | VARCHAR2(20) |  |
| 16 | MENU\_NM | VARCHAR2(100) |  |
| 17 | TELP\_TYPE | VARCHAR2(5) |  |
| 18 | USER\_NM | VARCHAR2(100) |  |
| 19 | TRNS\_DTTM | VARCHAR2(30) |  |
| 20 | SUBR\_NUMB | VARCHAR2(20) |  |

### Control Feed - Record Format

USSD\_ALL\_TRANS transactions and USSD\_CDR transactions doesn’t have any control file associated.

| # | Feed Name | Data File Name | Data File Format | Control File Name | File Format | Control File Format |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) | .txt | No Control files | - | - |
| 2 | USSD\_CDR | 99XXXXX\_YYYYMMDD.DAT  ( ex. 9911035195\_20190804.DAT ) | .DAT | No Control files | - | - |
| 3 | UMB103 | UMB103\*<YYYYMMDD>.txt (ex, UMB103\_ussd103nm1\_2019071700.txt) | .txt | No Control files | - | - |

### 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 USSD 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

USSD related transaction data will be pulled daily once into AEP landing server from EDW- Landing server - nickel4 using SFTP protocol. AEP must pull from EDW Landing path the files.

| **#** | **Feed Name** | **File Name** | **Arrival Time** | **AEP Landing path - Data files** | **AEP Landing path - Control files** |
| --- | --- | --- | --- | --- | --- |
| 1 | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) | 3:00:00 AM | /<root>/SRC\_DATA/USSD/USSD\_LOG/<YYYYMMDD>/ | NA |
| 2 | USSD\_CDR | 99XXXXX\_YYYYMMDD.DAT  ( ex. 9911035195\_20190804.DAT ) | 3:00:00 AM | /<root>/SRC\_DATA/CDR/MD/USSD\_REFILL/<YYYYMMDD>/ | NA |
| 3 | umb103 | UMB103\*<YYYYMMDD>.txt (ex, UMB103\_ussd103nm1\_2019071700.txt) | NA | /<root>/SRC\_DATA/USSD/UMB/<YYYYMMDD>/ | NA |

## 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 (Not Applicable for USSD)
8. Record Length Checking (USSD CDR feed files only)

* In case of missing / unavailability of control file for any feed we should be checking file size changes before pull the data from respective source system. For the detail method refer to Design document.
* Wherever “Expected min no of files EOD” column in Section 3.1 is blank– the process will trigger Alert mail to BU owner as per owner/team marked in section 1.2.

### 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 sent 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

As header and trailer records are provided for reconciliation, reconciliation can be done with source system. Reconciliation will be done only based on source file 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 |
| USSD | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt, USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) |

**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 |
| USSD | USSD\_ALL\_TRANS | USSD\*\_<YYYYMMDD>.txt ( ex. USSD19X\_20190804.txt,  USSDFsp\_20190804.txt,  USSDReward\_20190804.txt ) | 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 |
| USSD | USSD\_ALL\_TRANS | 38 | 34 |

**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 VARCHAR2acter 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 USSD 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 USSD 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:



### USSD Transaction Log ODS script <<trn\_ussd\_fileid.sh>>



### USSD CDR ODS script << MoveUssd.sh>>



### DataStage job << P\_TX\_USSD\_CDR\_D>> for USSD\_CDR processing

