**Functional Specifications – Data Format, Data Masking, Data Access**

1. Data Format

A sample data file has been created to understand the data at a granular level from the data dictionary provided. Each record is believed to consist of Header, Detail and Trailer.

**Header**

* The Header will have a descriptor of HDR to effectively identify the Header record.
* It also outlines the date on which the data has been either created or generated.
* Also it would be good to include the Service Provider and location information. For E.g.: Airtel and Malawi in this regard.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| HDR | 1/12/2018 | AIRTEL | MALAWI |

**Detail**

* The detail section portrays the inbuilt data dictionary and provided information
* Device Id/ Tel # will service as a primary key attribute for our analysis.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Record ID | Event Type | | Calling Party/Device ID/Tel # | | Tower ID | | Date of Event (Call) |
| 1 | Voice Call | | xztysh10001 | | A1004 | | 2/4/2017 |
| 2 | Voice Call | | xztysh10002 | | C1008 | | 2/4/2017 |
| Time of Event(Call) | | Originating Latitude | | Originating Longitude | |
| 2:30:51 | | 40.466 | | 79.982 | |
| 4:51:08 | | 41.568 | | 80.485 | |

**Trailer**

* Trailer will have a descriptor of FTR to effectively identify the Footer record.
* Trailer record is expected to share the total no of records that the file contains.

|  |  |
| --- | --- |
| FTR | Total no of Records:60 |

Please find below the sample data describing the above three sections.



View of file data:



The above file is in ‘xls’ format which can be converted to a csv with the below instructions.

1. Click File > Save as
2. Select Comma Separated Values from the drop down – File Type  
     
   Geo LOOK UP TABLE  
   RESTORATION TO LIVE APPLICATION



View of file data:



The data file is converted to a UTF format file for additional ease.



1. Data Masking

In order to enhance the data security a data masking algorithm is triggered on the actual data to anonymize/mask the data.

* The Masking algorithm doesn’t alter the functional aspects of the raw data.
* Only sensitive attributes will be masked to maintain the integrity.

**Masking Strategy**

* Substitution: Suitable values will be substituted against the existing values.
* Null Values: More sensitive data will be nullified so that the visibility is controlled.
* Variance: An additional variance will updated to the existing data to effectively mask the data which makes the reverse engineering feasible.

3. Data Access

* A FTP Environment will be created to effectively access the data with more focus on Data Security and Integrity.
* Following are the methods under consideration.

**Primary Considerations (favorable):**

1. **Access from AIRTEL**

* DIAL will create a secure application layer to pull data from **AIRTEL** environment.
* The data will be maintained by **AIRTEL** and DIAL will access data on an adhoc basis for Analytical needs.

1. **Create a FTP server in Infosys Ltd**

* DIAL will create a FTP server on its own while the ports open out to **AIRTEL**
* DIAL will be responsible for the maintenance and storage of Sensitive data.

**Secondary Considerations (less favorable):**

1. **Engage a Third Party Data Service**

* The third party vendor will be completely responsible for Data Masking and Storage.
* DIAL will pull the data for processing from the landing area of the third party vendor.

T

1. **Migration to 3rd party file sharing site (i.e. Dropbox)**

* MNO data will be migrated to 3rd party site / location, post masking.
* DIAL will pull the data on an adhoc basis from the 3rd party site for analytical needs.