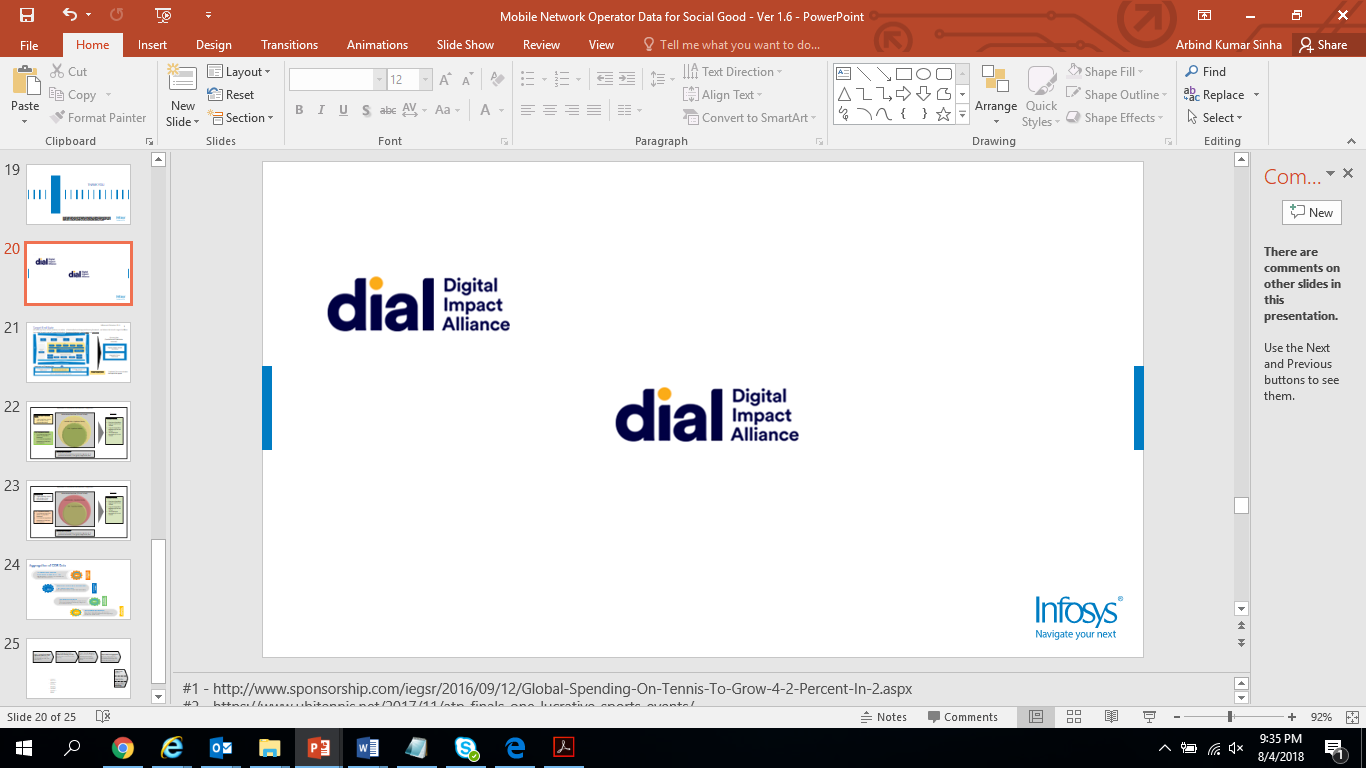
**Malawi Use Case**

**Data Dictionary**

**Version – 1 0**





Document Information and Revision History

| **Version** | **Date** | **Author** | **Reviewer** | **Remarks** |
| --- | --- | --- | --- | --- |
| 1 0 | 14/09/2018 | Infosys | Infosys | Draft version |
|  |  |  |  |  |

**Table of Contents**

[1.0 Introduction 4](#_Toc524723860)

[2.0 Data and Data Sources 4](#_Toc524723861)

[3.0 Data – Call Detail Records (CDR) 4](#_Toc524723862)

[3.1 About the data 4](#_Toc524723863)

[3.2 Schema 4](#_Toc524723864)

[4.0 Data – District / Site Lookup Files 5](#_Toc524723865)

[4.1 About the data 5](#_Toc524723866)

[4.2 Schema 5](#_Toc524723867)

[4.3 Key Considerations 6](#_Toc524723868)

[5.0 Data – CH District Data Summary 6](#_Toc524723869)

[5.1 About the data 6](#_Toc524723870)

[5.2 Schema 6](#_Toc524723871)

[5.3 Key Considerations 7](#_Toc524723872)

[6.0 Data – Airtel MW\_Database 8](#_Toc524723873)

[6.1 About the data 8](#_Toc524723874)

[6.2 Schema 8](#_Toc524723875)

[6.3 Key Considerations 8](#_Toc524723876)

[7.0 Data – Geography\_Master 8](#_Toc524723877)

[7.1 About the data 8](#_Toc524723878)

[7.2 Schema 8](#_Toc524723879)

[7.3 Key Considerations 9](#_Toc524723880)

[8.0 Data – Health Post Catchment Population Data 9](#_Toc524723881)

[8.1 About the data 9](#_Toc524723882)

[8.2 Schema 9](#_Toc524723883)

[8.3 Key Considerations 10](#_Toc524723884)

[9.0 Data – Facebook High Resolution Settlement Layer 10](#_Toc524723885)

[9.1 About the data 10](#_Toc524723886)

[9.2 Schema 10](#_Toc524723887)

[9.3 Key Considerations 10](#_Toc524723888)

[10.0 Data – Administrative Boundary shape files 10](#_Toc524723889)

[10.1 About the data 10](#_Toc524723890)

[10.2 Schema 11](#_Toc524723891)

[10.3 Key Considerations 12](#_Toc524723892)

[11.0 Data – Malawi Burden of Disease Data - Annual by Facility 12](#_Toc524723893)

[11.1 About the data 12](#_Toc524723894)

[11.2 Schema 12](#_Toc524723895)

[11.3 Key Considerations 12](#_Toc524723896)

[12.0 Data – Population Projection Data 12](#_Toc524723897)

[12.1 About the data 12](#_Toc524723898)

[12.2 Schema 13](#_Toc524723899)

[12.3 Key Considerations 13](#_Toc524723900)

# Introduction

This document covers details of data, data sources and associated data dictionaries.

# Data and Data Sources

Data requirements for the use case –

|  |  |
| --- | --- |
| Data | Source of data |
| CDR data | AirTel Malawi |
| Administrative shapefiles (TA and GVHs) | [www.diva-gis.org/gdata](http://www.diva-gis.org/gdata) (GIS data) |
| Facebook High Resolution Settlement Layer | Facebook |
| Population projection [2008-2030] | Cooper / Smith |
| CH District data summary [2017] | Cooper / Smith |
| Catchment population | Cooper / Smith |
| Malawi Burden of Disease Data | Cooper / Smith |
| Geography\_Master | Cooper / Smith |
| Airtel MW\_Database | AirTel Malawi |
| District / Site Lookup Files (*covers Tower locations (e.g. LAT/LONG, district, regions, etc.))* | AirTel Malawi |

# Data – Call Detail Records (CDR)

## About the data

This data covers details like call descriptions, call initiation mobile #, call termination mobile #, LAC Id, Cell Id and call duration for SMS and voice.

## Schema

| **File** | **Columns** | **Data Type** | **Description** | **Remarks** |
| --- | --- | --- | --- | --- |
| MNO - SMS and Voice | USAGE\_CALL\_DESCRIPTION | Text | Type of call made-SMS, voice call, roaming, etc. |  |
| CALL\_ORIGINATING\_NUMBER | Alpha Numeric | Mobile number from which call was made (sender) |  |
| CALL\_TERMINATING\_NUMBER | Alpha Numeric | Mobile number to which call was made(receiver) |  |
| CALL\_START\_DATE | Date (mm/dd/yyyy) | Date the call was made |  |
| CALL\_START\_TIME | Time (hh: min: sec) | Time of the call |  |
| LAC\_ID | Numeric | Location Area code from where the call was initiated |  |
| CELL\_ID | Numeric | Unique Number used to identify each Base Transceiver Station (BTS) |  |
| CALL\_DURATION | Numeric | Duration of the call |  |

# Data – District / Site Lookup Files

## About the data

These files have details of Tower locations (e.g. LAT/LONG, district, regions, etc.)

## Schema

| **File** | **Columns** | **Data Type** | **Description** | **Remarks** |
| --- | --- | --- | --- | --- |
| Lookup Files used:   1. MW LTE Classification file 10072018 update, 2. 2G Classification file\_14072018 (003) and 3. 3G MW Site Database 15072018 | Siteid | Alpha Numeric | Site ID where the tower is located | Referenced as Siteid, SITE, SiteID in the 3 lookup files |
| SiteName | Text | Site Name of the tower location | Referenced as Sitename, Site Name, SITE\_NAME in the 3 lookup files |
| LAC\_ID | Numeric | Location Area code from where the call was initiated | Referenced as TAC,LAC ID,LAC in the 3 lookup files |
| CELL\_ID | Numeric | Unique Number used to identify each Base Transceiver Station (BTS) | Referenced as CI, CELL\_ID, UCell in the 3 lookup files |
| DISTRICT | Text | District of tower location | Referenced as CITYID, District in the 3 lookup files |
| LONG | Longitude | Longitude of tower location |  |
| LAT | Latitude | Latitude of tower location |  |

## Key Considerations

Only those columns are considered from look up files which can be helpful in connecting MNO data with Tower Id and LAT/Long.

# Data – CH District Data Summary

## About the data

This data can be used to understand community health system as per district population (covered/un-covered), HSA data, Bicycle data and Health Indicators.

## Schema

| **File** | **Sheet Name** | **Column Name** | **Data Type** | **Description** |
| --- | --- | --- | --- | --- |
| CH District Data Summary 2017 | District Summary | District | Text | District Name |
| Investment Priority | Numeric | Investment Priority |
| Total population | Numeric | Total population of the District |
| Uncovered Population | Numeric | Population not covered under Govt. Health Scheme |
| % uncovered Population | Numeric | %Population not covered under Govt. Health Scheme |
| Proposed Health Posts | Numeric | Number of Health Posts proposed to be setup |
| Target Health Posts | Numeric | Number Health Posts targeted to be setup |
| HTRA | Numeric | hard to reach - areas that are more than 5km away from a health facility |
| Total HSAs | Numeric | HSAs + SHSAs |
| HSAs | Numeric | Health Surveillance Assistants in June 2017 |
| HSA to Population ratio | Numeric | Population/HSAs |
| SHSAs | Numeric | Senior HSAs |
| HAS to SHSA ratio | Numeric | SHSAs/HSAs |
| Total HSAs | Numeric | Health Surveillance Assistants in Sept 2017 |
| Number without bicycles | Numeric | Number of HSAs without Bicycles |
| % without bicycles | Numeric | % of HSAs without Bicycles |
| Neonatal Mortality rate | Numeric | neonatal mortality rate in the year 2016 |
| % of under 1 fully immunized | Numeric | % of under 1 fully immunized in the year 2016 |
| Vitamin A Coverage | Numeric | Vitamin A Coverage in the year 2016 |
| Health Posts | District | Numeric | District Name |
| Health Posts by year | Numeric | Health Posts to be setup in 5 years |
| Total | Numeric | Manual adjustment of unedited total |
| Unedited Total | Numeric | Total number of Health Posts |
| DHIS 2 | Organization Unit | Numeric | Name of the Health Facility |
| Neonatal Mortality rate | Numeric | From 2012-2016 |
| Penta III Coverage | Numeric | From 2012-2016 |
| % of under 1 fully immunized | Numeric | From 2012-2016 |
| Vitamin A Coverage | Numeric | From 2012-2016 |
| Bicycles by District | District | Numeric | District Name |
| Total HSAS | Numeric | HSAs in Sept 2017 |
| Number with bicycles | Numeric | Number of HSAs with Bicycles |
| Number without bicycles | Numeric | Number of HSAs without Bicycles |
| HR Data | Row labels | Numeric | Different Council names in each District |
| HSAs | Numeric | HSAs in June 2017 |
| SHSAs | Numeric | SHSAs in June 2017 |

## Key Considerations

We have requested CH District Data Summary data for the year 2016. But we did not get this data.

# Data – Airtel MW\_Database

## About the data

This data can be used to understand Cell Tower information with longitude and latitude.

## Schema

| **File** | **Sheet Name** | **Column Name** | **Data Type** | **Description** |
| --- | --- | --- | --- | --- |
| Airtel MW\_Database\_18062018 | Cell Info | Site Id | Text | Unique Id of the site |
| Cell id | Numeric | Unique Id of the Cell |
| Town | Text | Name of the town |
| Coordinates Data | 2G Site ID | Text | Unique Id of 2G site |
| Lat | Latitude | Latitude of 2G site Id |
| Long | Longitude | Longitude of 2G site Id |
| Site Name | Text | Site name |

## Key Considerations

We have Cell Tower information with longitude and latitude shared.

# Data – Geography\_Master

## About the data

This data can be used to provide details of Cell Id, Lac Id as per Local government accountability (LGA) for all states of Malawi.

## Schema

| **File** | **Column Name** | **Data Type** | **Description** |
| --- | --- | --- | --- |
| Geography\_Master | COUNTRY\_NAME | Text | Country name |
| REGION\_DESCRIPTION | Text | Region descriptions |
| ZONE\_CODE | Text | Zone code of the region |
| STATE\_CODE | Text | State code of Malawi |
| STATE\_DESCRIPTION | Text | State name of Malawi |
| LGA\_ID | Numeric | Local government accountability Id |
| LGA\_CODE | Text | Local government accountability code |
| LGA\_DESCRIPTION | Text | Local government accountability |
| SITE\_CODE | Text | Site code of LGA |
| SITE\_NAME | Text | Site name of LGA |
| SITE\_DESCRIPTION | Text | Site name of LGA |
| CELL\_CODE | Text | Cell code |
| CELL\_NAME | Text | Cell name |
| CELL\_DESCRIPTION | Text | Cell description |
| LAC\_ID | Numeric | Location Area Code |
| CI\_ID | Numeric | Cell Id details |

## Key Considerations

None.

# Data – Health Post Catchment Population Data

## About the data

This data provides master list of health facility with the catchment population. Catchment population is available for facility type, dispensary, health center, health post and hospital.

## Schema

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **File Name** | **Column Name** | **Data Type** | **Description** | **Remarks** |
| Master List Free Service Health Facilities in Malawi 20 July 2018 | Facility Name | Text | Facility Name |  |
| Village | Text | Village name |  |
| GVH | Text | Group village head |  |
| TA | Text | Traditional authorities |  |
| District | Text | District name |  |
| Region | Text | Region name |  |
| Eastings | Number | Eastings |  |
| Northings | Number | Northings |  |
| Datum | Text | Datum |  |
| Facility Type | Text | Type of facility |  |
| Controlling Agency | Text | Name of controlling agency |  |
| Name | Text | Name of person |  |
| Phone | Number | Phone number |  |
| Email | Email | Email address |  |
| Cluster | Text | Name of cluster |  |
| Status | Text | Functional status |  |
| Date | Date | Date |  |
| Facility Code | Number | Facility code |  |
| 2018 Catchment | Number | Catchment population for 2018 |  |
| 2017 Catchment | Number | Catchment population for 2017 |  |
| 2016 Catchment | Number | Catchment population for 2016 |  |
| 2015 Catchment | Number | Catchment population for 2015 |  |

## Key Considerations

This covers population coverage within 5 Kms radius of a given health facility.

# Data – Facebook High Resolution Settlement Layer

## About the data

These files provide estimates of population distribution across Malawi.

## Schema

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **File Name** | **Band Category** | **Data Type** | **Description** | **Remarks** |
| High Resolution Settlement Layer-Facebook data-hrsl\_mwi.tif | Band 1 / POP | Numeric  (raw file-Image) | Number of people estimated to live in each pixel |  |
| Band 2 / Settlement | Numeric  (raw file-Image) | A mask representing pixels within which one or more buildings was detected with a value of 1 |  |
| Band3 / Clouds | Numeric  (raw file-Image) | A mask representing pixels that contained clouds in the source imagery with a value of 1 |  |

## Key Considerations

To come up with these estimates convolutional neural network and computer vision techniques were used.

# Data – Administrative Boundary shape files

## About the data

These shapefiles contain the administrative boundaries detail of Malawi as per traditional authorities (TA) / group village heads (GVHs) level.

## Schema

| **File Name** | **Columns** | **Data Type** | **Description** | **Remarks** |
| --- | --- | --- | --- | --- |
| MWI\_adm1.shp | ID\_0 | Categorical | Country ID |  |
| ISO | text | Country code |  |
| NAME\_0 | text | Country Name |  |
| ID\_1 | Categorical | District ID |  |
| NAME\_1 | text | District name |  |
| TYPE\_1 | text | Administrative level type |  |
| ENGTYPE\_1 |  |  |  |
| NL\_NAME\_1 |  |  |  |
| VARNAME\_1 | text | Alternate name for district |  |
| MWI\_adm2.shp | ID\_0 | Categorical | Country ID |  |
| ISO | text | Country code |  |
| NAME\_0 | text | Country Name |  |
| ID\_1 | Categorical | District ID |  |
| NAME\_1 | text | District name |  |
| ID\_2 | Categorical | Administrative second level ID |  |
| NAME\_2 | text | Administrative second level name |  |
| TYPE\_2 | text | Administrative second level type |  |
| ENGTYPE\_2 |  |  |  |
| NL\_NAME\_2 |  |  |  |
| VARNAME\_2 | text | Alternate name for district |  |
| MWI\_adm3.shp | ID\_0 | Categorical | Country ID |  |
| ISO | text | Country code |  |
| NAME\_0 | text | Country Name |  |
| ID\_1 | Categorical | District ID |  |
| NAME\_1 | text | District name |  |
| ID\_2 | Categorical | Administrative second level ID |  |
| NAME\_2 | text | Administrative second level name |  |
| ID\_3 | Categorical | Administrative third level ID |  |
| NAME\_3 | text(alphanumeric) | Administrative third level name |  |
| TYPE\_3 | text | Administrative third level type |  |
| ENGTYPE\_3 |  |  |  |
| NL\_NAME\_3 |  |  |  |
| VARNAME\_3 | text | Names of each individual GVH | Only unknown enteries |

## Key Considerations

None

# Data – Malawi Burden of Disease Data - Annual by Facility

## About the data

This file provides details of # of patients treated / services provided for the year 2012-2016 disease-wise at health centers.

## Schema

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **File Name** | **Columns** | **Data Type** | **Description** | **Remarks** |
| Malawi Burden of Disease Data - Annual by Facility | Organization Unit | Text | Name of the health center Organization Unit |  |
| Columns ‘B’ to ‘IQ’ | Number | # of patients treated / service provided |  |

## Key Considerations

This covers # of patients treated / services provided for the year 2012-2016 disease-wise at health centers.

# Data – Population Projection Data

## About the data

This file provides population estimates for the year 2008 to 2030 district-wise.

## Schema

| **File Name** | **Columns** | **Data Type** | **Description** | **Remarks** |
| --- | --- | --- | --- | --- |
| Population by SNU 2008-2030 : Sheet1 | YEAR | Numeric | Year |  |
|  | INDICATOR | Integer | ID for Indicator label |  |
|  | INDICATOR\_LABEL | Text | Type of Indicator (here Population) |  |
|  | ADMINISTRATIVE LEVEL 2 | Integer | ID for Administrative level 2 |  |
|  | ADMINISTRATIVE LEVEL 2\_LABEL | Text | Name of the Administrative level 2(District Name) |  |
|  | VALUE | Numeric | Population value |  |
|  | FLAG | Logical gate | highlight a district of interest |  |
|  | FLAG\_LABEL | Logical gate | highlight a district of interest |  |
|  | UNIT | Integer | ID for the unit used to describe VALUE column |  |
|  | UNIT\_LABEL | Text | Units in which VALUE column is described |  |

## Key Considerations

Source of this data is National Statistical Office, Malawi.