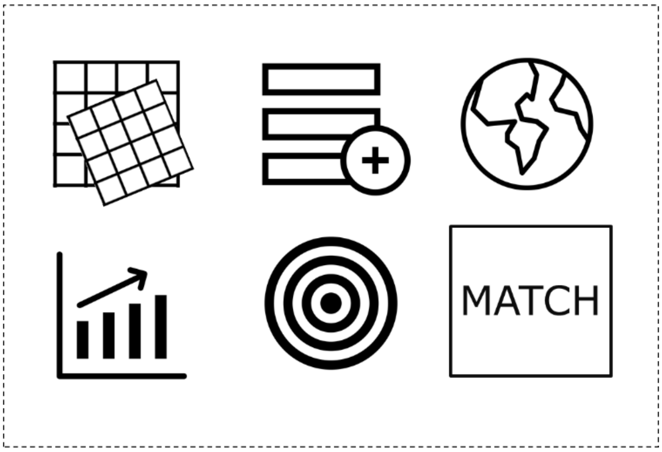
DATA MANAGEMENT PLAN FOR

MATCH: MAPPING AND ANALYSIS FOR TAILORED DISEASE CONTROL AND HEALTH SYSTEM STRENGTHENING



Version 1.0

Amsterdam, October 2017

KIT Royal Tropical Institute

With support of

This document is an initiative of the Stop TB partnership, funded by the Global Fund to Fight AIDS, Tuberculosis, and Malaria, and written by KIT Royal Tropical Institute.

This manual has been prepared by:

Mirjam Bakker
Ente Rood
Christina Mergenthaler
Lucie Blok
Margo van Gurp
Masja Straetemans
Sandra Alba

All documents and annexes referred to in this manual can be found on the KIT Health website:
https://www.kit.nl/health/service/kit-match-approach-enhancing-tb-care-coverage/

Tabel of content

1. Background 1

2. Purpose of this document 2

3. Data description 3

4. Data formats and metadata 5

    Data table standards 5

    Codebooks 5

    Geographic data 5

    Record linkage 6

    Data quality 7

    Metadata content and documentation 9

5. Database model 9

    Database querying 11

6. Roles and responsibilities 12

7. Access and sharing 13

    Access to the database using TCP/IP 13

    Access using QGIS 13

Annex 1 Variable codebooks of data tables included in the database 16

Annex 2 probabilistic record Linkage in R 27

Annex 3 Indicator calculations. 30

Annex 4 SQL query for indicator calculation. 37

1. Background

In the summer of 2016, the Stop TB Partnership and The Global Fund announced a joint initiative to strengthen the use of subnational data for decision-making in TB programs. This initiative comes as a response to the global health emergency that in 2015 approximately one-third of the estimated 10.4 million incident TB cases go unreached, either not being reported, not receiving treatment or not even being properly diagnosed. The rationale for the initiative is to take advantage of the considerable amount of existing, yet underutilized subnational data which more specifically tells the story of who and how many of these missing TB cases there are, where they are located, why they are missed, and how to find and treat them. It is subsequently decided to train NTP staff and stakeholders dedicated to this initiative and to enable them to optimally utilize disaggregated subnational data to make local programmatic decisions to tackle their individual TB epidemic.

The main objectives of the Stop TB Partnership's initiative are to provide guidance and build capacity of the participating countries to collect, collate and analyse sub-nationally disaggregated data to inform strategic and ambitious planning, and to improve country analytic capacity for subnational differentiated responses to TB. With countries routinely collecting and reporting key TB indicators from district or BMU level, required data are often available yet not accessible in a way required to collate and analyse these at a subnational level. This initiatives aims to build long-term capacity to utilize data to support decision makers to target TB interventions and to increase ownership among data users and policy makers. By linking data to policy objectives, combining and simplifying complex data into a more intuitive format and building capacity on data management and analytical techniques described in detail below, our aim is:

1. to strengthen in country capacities to use publically available and user friendly tools to analyse their subnational data effectively to understand and interpret the main programmatic gaps and challenges in order to plan appropriate local interventions that reduce the burden of TB in their communities,
2. to increase critical thinking and appraisal of the available data and analytical results to guide decision making on planning, monitoring and evaluation of TB (data -> information -> knowledge -> policy).
3. to enable routine utilization of data and to determine which steps need to be taken to collect more useful subnational data when it is found to be missing.

Data harmonization and the integration and simplification of data obtained from multiple sources into one coherent, and unified source, is the key to being able to use existing data to identify spatial and temporal trends. Data derived from various sources should be validated using fit-for-purpose criteria (i.e. derived data is assured of its suitability for its end use context) and should be spatially and temporally consistent to allow integration and analysis. Standardized data formats, processing and management procedures will need to be developed.

2. Purpose of this document

The purpose of this document is to provide data managers, epidemiologists and end-users of the final database with a comprehensive description of the data management steps that are required to map subnational factors of the epidemiology, the programmatic responses and risk factors of TB. This data management plan (DMP) follows the steps required to prepare data obtained from various sources form mapping using QGIS Free and open source software. The analysis and mapping of the processed data as described in this document is further described in the MATCH manual which can be feely downloaded from the KIT Royal Tropical Institute website.

The current version and examples provided in this DMP are based on the work done together with the Stop TB Partnership, supported by the Global fund to fight AIDS Tuberculosis and Malaria.

3. Data description

Depending on the countries TB program implementation and reporting system. TB data can be obtained at varying levels of aggregation and for different time series. For the purpose of this initiative, TB data are collected and processed at the lowest possible level of aggregation corresponding to the level of geographical units available in the spatial data. The data collected and collated for the subnational mapping and analysis of TB epidemic and program responses include five broad categories:

1. *TB notification and burden*
   The primary source of data is the National TB program (NTP) which reports data following WHO standard as defined in the *Standards and benchmarks for tuberculosis surveillance and vital registration systems*[1] published by WHO. These reports cover the following categories:

   a. indicators of TB case notification
   b. indicators of laboratory testing
   c. indicators of TB treatment outcome
   d. indicators of HIV testing amongst TB patients
   e. indicators of burden of TB (TB prevalence)
   f. indicators of TB underreporting (inventory studies, capture recapture studies (CRC)

   TB prevalence data, resulting from prevalence surveys, are available for various countries. Such surveys are commonly designed to generate estimates of predefined statistical accuracy (confidence) at a national level. In some cases prevalence estimates are also stratified by geographic zones. The use of these data to monitor prevalence at a sub-national level has been debated as the confidence at lower spatial scales is usually insufficient to monitor spatial trends at a subnational level. Noting this concern, however, these data can provide valuable insight in plausible geographic variations in TB prevalence . In addition, spatial statistics can provide methodological tools to improve confidence in these estimates and should not be ignored when using subnational estimates of the burden of TB.

2. *Population data*
   Population data are usually derived from national population censuses which are commonly conducted at regular time intervals (5-10 years). The primary source for subnational population data are the national bureaus of statistics, yet other repositories can be found online. Where subnational population survey estimates are not available these can be obtained from secondary sources. A commonly used source of population estimates is Worldpop gridded population data of the world. Worldpop provides population estimates on a high resolution (1 sq.km) by disaggregating data obtained from census data and using satellite derived land cover data[2]. Where possible, the following data should be obtained:

   a. Subnational estimates of age and sex stratified population numbers

[1] http://www.who.int/tb/publications/standardsandbenchmarks/en/
[2] www.worldpop.org

3. *Risk factors and key populations*
   For some countries, subnational data on TB risk factors and socio-economic data are available from the Demographic and health Surveys DHS[3]. These data are often collected to estimate health indicators at national, zonal (non-administrative subnational) or province level at a very low spatial resolution (usually 2-3 strata). The following data can be freely downloaded from the DHS website:

   a. Socio-economic factors (age, wealth)
   b. Risk factors (behavioural)
   c. Key populations (occupational, migrants, prisoners)

4. *Access to healthcare*
   Health systems data should be used to assess the delivery and access to health services including TB diagnosis and treatment services. These data are key to differentiate whether geographical variations in TB case detection are a result of variations in the TB epidemic and risk population distributions or whether these reflect variations in health system coverage and access. Various survey tools and data could be used to assess the supply of services. This could range from a simple service-to-population ratio per district to a more sophisticated assessment of health care access taking into account travel distances and other access components (financial, cultural ,etc). Such analysis generally rely on high quality and high resolution data on the location of service providers and the distribution of the population within health facility catchment areas. To allow for this analysis, the following data should be considered:

   a. Indicators of health care access (DHS, Healthcare Utilisation Surveys (HUS))
   b. Diagnostic facilities: type, quality, human resources
   c. Insurance data (testing and treatment) and pharmaceutical prescriptions data
   d. Service Provision Assessment (SPA) surveys (DHS)
   e. Service Availability and Readiness Assessment (SARA) surveys
   f. MICS (Multiple Indicator Cluster Surveys)

5. *Spatial data*
   Spatial data of geographic units corresponding to administrative or operational divisions can be freely downloaded from the GADM or GAUL repositories. In some countries land-surveying departments have established Spatial Data Infrastructures (SDI) where versioned spatial data files can be accessed and downloaded. Spatial data should be obtained for:

   a. Vector polygon files of subnational administrative boundaries
   b. Vector point data of health facilities (country specific files)

[3] http://dhsprogram.com/

4. Data formats and metadata

The tabular format used throughout the process of data collection, processing and storage in the final database are Comma Separated Value files (CSV files). Tabular data (attribute data) shared and extracted from various sources are commonly stored as MS excel workbooks which are converted to singe table CSV files. National TB Programs or other sources of data which have data stored and maintained in a data repository or standardized database can be queried and output data stored as CSV files for further processing.

TB prevalence data generated by prevalence surveys that are conducted in various countries, are commonly available in the form of PDF reports providing formatted tables. Prevalence data should be extracted at the lowest geographic denominator for which data are available and stored in CSV format.

Spatial data representing first, second or third level administrative boundaries of countries obtained from the GADM data portal are initially stored in an ESRI shapefile format. Spatial data files obtained from other sources , such as national bureaus of statistics or humanitarian data portals, can include other formats including GeoJSON and Keyhole Markup Language (KML) files. These files can be loaded directly into QGIS and should be converted to shapefiles before being processed and uploaded to the database.

Finally, spatial data files are uploaded to the PostgreSQL data base using the postGIS 2.4[4] extension and stored as data tables with the geometries stored in a single column as a Well Known Binary (WKB)[5].

Data table standards

Data are processed and stored in different thematic tables according to the categories listed under (section 3). Data tables are formatted as long tables with single records representing the data for one reporting unit which is defined by the geographic unit (e.g. adm2 unit, health facility, zone, etc), and temporal unit (month/quarter/year). Per category, different data elements (e.g variables) are stored as single variables. See *annex 1 -variable codebooks of data tables included in the database-* for an overview of data elements available.

Different variables available for each data category are stored in a single table which are named using the ISO 3361 international three character country codes[6], followed by the data category and subclass: [ISO]_[category]_[subclass]. Table 1 below shows a screenshot of the data schema for Indonesia (IDN)

Codebooks

All variables in the different data tables are described in a codebook (see annex 1) consisting of a list of variables, the type of variable (e.g. character, integer etc.) and a description of that variable. Additional comments can be added to the codebook if necessary.

Geographic data

All geographic data are stored using a multipart-polygon geometry (Areal data) or as point locations (health facility data) using the EPSG:4326 (WGS84) coordinate reference system.

---

[4] http://postgis.net/
[5] https://postgis.net/docs/using_postgis_dbmanagement.html#OpenGISWKBWKT
[6] https://www.iso.org/obp/ui/#search/code/

Table 1: Screenshot of the various data tables stored in the data schema (e.g. data container, similar to a folder) in the final database. For each table a description in given.

Tables (35)	Tables in Schema for IDN (Indonesia)
idn_cnr_alltb	Indonesia case notification rate for all forms of TB
idn_cnr_child04	Indonesia case notification rate for all forms of TB in children under 4 years of age
idn_cnr_child514	Indonesia case notification rate for all forms of TB in children between 5 - 14 years of age
idn_cnr_eptb	Indonesia case notification rate for extra pulmonary TB
idn_cnr_newbact	Indonesia case notification rate of new bacteriologically confirmed TB
idn_cnr_newclindx	Indonesia case notification rate of new clinically confirmed TB
idn_cnr_newrel	Indonesia case notification rate of new and relapse cases of TB
idn_coverage	Indonesia health care coverage
idn_eqa	Indonesia results of external quality assurance
idn_facilitydens	Indonesia facilities per 100.000 population
idn_pop	Indonesia population data
idn_prog_test	Indonesia laboratory testing results
idn_prv_all	Indonesia TB prevalence
idn_risk	Indonesia TB risk factors
idn_smooth_cnr_all	Indonesia spatially smoothed case notification rates per 100.0000 all forms of TB
idn_smooth_cnr_newrel	Indonesia spatially smoothed case notification rates per 100.0000 new and relapse TB
idn_tb_hiv	Indonesia HIV prevalence in TB
idn_treat	Indonesia treatment outcomes
kit_tb_idn_adm1	Indonesia first level administrative units multi-polygon spatial data table
kit_tb_idn_adm1_link	Indonesia first level administrative level record linkage table
kit_tb_idn_adm2	Indonesia second level administrative units multi-polygon spatial data table
kit_tb_idn_adm2_adj	Indonesia second level administrative spatial adjacency table
kit_tb_idn_adm2_adj2	Indonesia second level administrative spatial adjacency table (including self intersections)
kit_tb_idn_adm2_link	Indonesia second level administrative level record linkage table
kit_tb_idn_hf	Indonesia health facilities point locations spatial data table

Record linkage

To map the properties of different spatial features (also called attributes in a GIS), tabular data collected from different sources are linked to the corresponding geographic (spatial) units. As the reporting units encountered in different M&E frameworks, surveys and surveillance networks commonly do not correspond to administrative geographical units in a country, the linking of the two can be subjected to a number of common issues:

A. TB reporting units do not match administrative units contained in spatial data files.
   - Modified area units (e.g. reporting based on administrative units, zones of a survey, basic management units, health facility catchment areas);
   - Disputed boundaries lead to a mismatch between the spatial data boundaries and the actual boundaries of TB operational units (BMUs/ facility catchment areas);

- TB reporting is based on population size, which results in reporting units which represent varying administrative levels (for example: a BMU at a urban adm3 level and another BMU at rural district adm2, whereas a single spatial file showing administrative boundaries generally does not contain mixed levels);
- Temporal mismatch between available geographic data of administrative areas (spatial data file) and reporting units (TB data).

B. Linking based on a common identifier is not possible.
   - Reporting units are named differently than the administrative units from which they are reporting (e.g. hospital name is used in the TB file, rather than the district in which it is based in the spatial data file);
   - Differences in naming and spellings of identical units between reporting units of the TB file, and geographical features in the spatial data file, which leads to mismatching between two independent sources of data.

Issues in a mismatch between the spatial units present in a spatial dataset and the reporting units for which attributes (e.g. indicators) are available need to be addressed on a case by case basis. Expert consideration is needed to confirm or redefine the boundaries of areas which are reporting. Dynamic boundaries (hospital catchment areas might change over time) pose additional challenges when comparing data over consecutive years, limiting the ability for unbiased comparisons over time.

Linkage problems due to non-matching identifiers (e.g. reporting unit names vs geographical unit names) can be solved by manual assessment and linkage. This procedure, however presents a considerable effort with datasets containing more than 100 records. When record linkage is imperfect (for example where there are > 50 mismatches), deterministic linkage using multiple ID variables or probabilistic linkage can be used. Before linkage algorithms can be applied data cleaning and parsing algorithms might need to be applied to remove inconsistencies in the use of capitals, non-alphanumeric characters and excess spaces. See *annex 2 probabilistic record Linkage in R* - for an example syntax using R to match districts in Indonesia.

For probabilistic linkage, the user also needs to determine the level of precision necessary for the software being used to return a match. This level of precision should be set higher or lower depending on how clean the databases are and the types of characters being compared. While matches are often reliable, it is critical for the user or a content expert to review the matches and non-matches, as user judgment - particularly that of local experts - will almost always be needed to generate the most well matched list of reporting units to geographic units.

Once successfully matched, a unique identifier code (UID) is generated and appended to the source (e.g ntp data, population date , etc) as well as the target dataset (e.g spatial data file). The unique identifier code should, where possible, use existing coding systems such as the ISO 3166[7]. Where such coding is not available, a unique coding respecting the hierarchy of various administrative units should be developed.

Data quality

Data quality assessments are performed to ascertain that the data are fit for purpose. This means that the indicators used to assess geographic variation in the TB epidemic and TB response, and the geographic scale and resolution used to make these assessments are appropriate to reflect real world processes and therefore can be used to inform actions. As any data used in this process is a result of

---

[7] https://www.iso.org/standard/63546.html

data collection and processing activities which inherently will contain a certain amount of error it is important to assess the completeness and internal (single source/variable) and external (across sources / variables) consistencies of various data elements used in the analysis. The following main data quality standards are assessed against the quality criteria determined by the purpose of use:

1. ACCURACY: The data are internally consistent. This means that data values fall within the range of expectation and that no apparent outliers are present in the data. Spatial accuracy is assessed by comparing the geographic units in a map to the area covered by a reporting unit . For example, all reporting health facilities should fall within the districts where they are reported.
2. TIMELINESS: The data should be recorded and reported in a timely fashion. For example, predictions for 2017 can only be made when data for the preceding period, including 2016, are available.
3. COMPLETENESS: The data used to calculate indicators and to map outcomes for different geographic areas should be based on a complete set of reports which includes all reporting units contained by that area. Population based notification rates should include the reports of cases across all ages.
4. COHERENCE: The data (spatial and non-spatial) should be coherent between sources. For example the number of individuals tested positive by a laboratory cannot exceed the number notified TB cases within an area.
5. PRECISION: The number of cases notified should reflect the true number of TB patients accessing healthcare. External quality assurance of diagnostic services can be used to assess the expected number of falsely diagnosed cases.
6. CURRENCY: The data should reflect the current situation. For example, when NTP data are mapped per district in a certain year, the district spatial data should reflect the actual administrative divisions for that year (e.g. do not use a district boundaries file for 2000 to map district case notifications in 2017).
7. SCALE: The scale (or resolution) of your outputs should be appropriate to answer operational questions. For example, to show subnational variations of TB case notification rates you would like to use a map of administrative areas which allows you to assess spatial patterns (rule of thumb > 30). NOTE: Using a too high resolution can results in the use of areas with very small populations and instable rate estimations as a result.
8. RELEVANCE: Both the data as well as the maps produced based on certain data should be relevant to serve its intended purpose.

As the standards are assessed against country specific criteria, which allow to assess data quality, it is essential that these are predefined by a TB expert and data manager. For example, the criterion for completeness is that data are available for at least 90% of the reporting units (BMU's or health facilities) before subnational mapping is done. Likewise, at least 30 spatial units should be present in the dataset to be able to map and to assess the existence spatial trends in a meaningful fashion.

An easy and intuitive way to assess the accuracy of a spatial trend observed is to look at the spatial consistency of the pattern. This can be done by comparing the observed pattern to a spatially smoothed pattern. This allows to dissect the spatial trend against the spatial random fluctuation of the pattern. To assess the stability of a spatial pattern over time, the deviation of a single time period from the trend of the preceding periods is calculated and mapped. This shows how much the pattern has been consistent over time or how likely this is to be subject to change.

Metadata content and documentation

The full process of data source identification, data extraction and processing to generate quality assured data which can be used for mapping is multifaceted and therefore prone to error. The process involves numerous processing steps involving various datasets obtained from various sources. Therefore it is key to document and monitor various metadata regarding the source and processing of different data components. This allows to track and trace the status of the data management process and to trace where errors might have occurred when they are encountered.

For each data table the following components are recorded and documented throughout the data processing steps:

1. Source and owner of the data
2. Date on which the data was received or downloaded.
3. Coverage and units
4. Access restrictions which might apply
5. Processing steps conducted

Throughout the data processing each data table is monitored against the following processing standards:

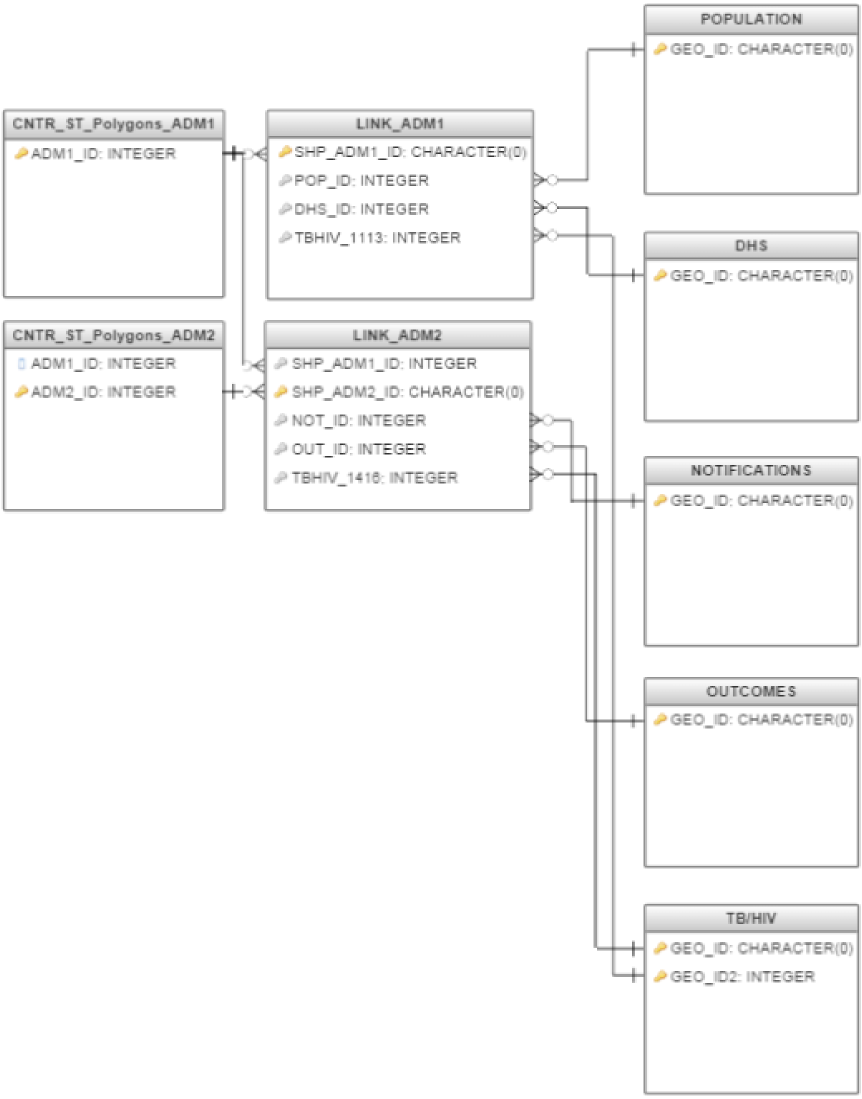
1. The data source has been identified and is accessible.
2. Data has been extracted or received
3. Data has been formatted according to the predefined standard
4. Data tables record linkage has been completed
5. Data quality assurance and revision has been conduced

5. Database model

The processed data including the unique identifier codes are uploaded into a cloud based PostgreSQL vs 9.3 database operated using a pgAdmin 4 vs 1.5. Using a cloud based database has several advantages over the use of a file based storage and archiving system:

1. All data are available from a single location to different users.
2. There are no issues with versioning of data tables generated by various users as these are centrally managed and archived
3. All data are consistently formatted according to predefines standards improving coherence and integrity.
4. The use of a relational database reduces (removes) redundancy of data elements which are used in different data-tables (e.g. single data tables for each data type)
5. The use of schemas within the database allows to manage access rights to different users.

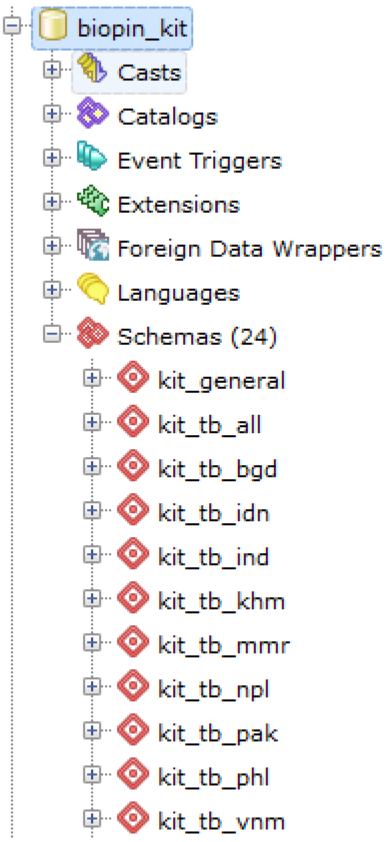
The following schematic Entity Relation Diagram shows the conceptual design of the database relating different sources of data to their respective geographic reporting units:



Relational joins between the units of each table are made based on a unique identifier key representing the geographic unit of reporting. After record linkage, unique spatial identifier codes (UID) are assigned to each table using a standardised coding in the form of XX-XX: Two digits number (e.g including leading zeros) identifying the first level administrative units of a country followed by a dash and a two digits number (e.g including leading zeros) identifying the second level administrative units nested within the first level units.

The UID codes for each geographic unit are generated for each record in the spatial data tables and are appointed as the Primary Key (PK). The same UID's are added to the other non-spatial data tables and assigned as a Foreign Key (FK) which allowed each record to be linked to the corresponding record (e.g unit) in the spatial data table.

Data for each country are uploaded into country specific schema following a standardized table definition. Schemas are names according as follows: kit_tb_[ISO]. Where the [ISO] represents the ISO 3166 three character country codes[8] which are used to distinguish schemas for each respective country.



Database querying

The data reported by the NTP and contained in the final database generally contain absolute numbers reported for different data elements (e.g. number cases notified, number of bacteriologically diagnosed, treated, etc). Where available these should be stratified according to different age classes and sexes. Although the absolute numbers are informative, the absolute numbers cannot be used to compare TB case detection, and programme responses across subnational areas. To do this, these absolute numbers need to be transformed into rates (relative change over a certain denominator population) which can be mapped and used to identify spatial trends. The use of long tables in the SQL database allows to easily select (query) and aggregate data. Subsets of selected data from different tables are joined based on the primary and foreign keys assigned to each table. The resulting output is then used to calculate proportions or rates using data elements stored in different tables.

The outcome tables resulting from the indicator calculation still consist of long-tables in which reports made by different units over time are represented by different records. However, in spatial data tables (and spatial data files like shapefiles) single geographic units are represented by single records and different time intervals are represented by columns (wide-format; see figure below). Before the

---

[8] https://www.iso.org/obp/ui/#search/code/

data are loaded into QGIS the long tables need to transposed into a wide format. The calculation of key indicators (see *annex 3 - indicator calculations)* based on the data element provided in the codebooks (*annex 2*) are provided in *annex 4 - SQL query for indicator calculation*.

LONG TABLE

District	Year	CNR
A	2010	10
A	2011	11
A	2012	13
B	2010	13
B	2011	12
B	2012	11
C	2010	21
C	2011	16
C	2012	18

WIDE TABLE

District	2010	2011	2012
A	10	11	13
B	13	12	11
C	21	16	18

6. Roles and responsibilities

The final database recognises different levels of users, each with their specific responsibilities and uses of the data. In case of the National Tuberculosis this would be the NTP manager or the line manager responsible for the data collection management and disclosure. Other users who require access to the data will require varying access rights depending on the application of the data. The varying user types, their roles and access rights are listed in the following table:

Role	**Description**	**Responsibilities**	**Access Rights**
Data owners	Institutional managers and contributing stakeholders (NTP manager, NGO's, MoH, Bureau of Statistics)	Defining use and access rights Publishing and disclosure permissions	• Set security permissions • Set access permissions • Backup rights • Write access • Read Access
Database managers	Technical staff involved in the process of database design, data collection and storage	Perform all configuration and maintenance activities on the database, and can also drop the database in SQL Server	• Set security permissions • Set access permissions • Backup rights • Write access • Read Access
Data manager	Technical staff involved in the process of database design, data collection and storage from the TB programme	Perform read and write operations on the predefined data schemas and data tables in the database	• Write access • Read Access
M&E Officer	Technical staff involved in routine monitoring and evaluation tasks of the TB programme	Reading data from the database for use in external software	• Read Access
External users	Researchers, NGO's , external stakeholders with interest in operational research	Reading data from the database for use in external software	• Partial Read Access

The data obtained from different sources remain the propriety of the institution who have collected and shared the data.

7. Access and sharing

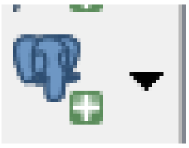
Access to the database using TCP/IP

Remote access to the database is established using an authorised TCP/IP connection. This can be accessed directly from a postgreSQL administration platform like pgAdmin or from QGIS using a TCP/IP network address for the postgreSQL server which consists of two parts: an IP address associated with one or more network cards in a computer, and a TCP port address specific to an instance of the PostgreSQL server. The postgreSQL database instances use TCP port 5432 by default.

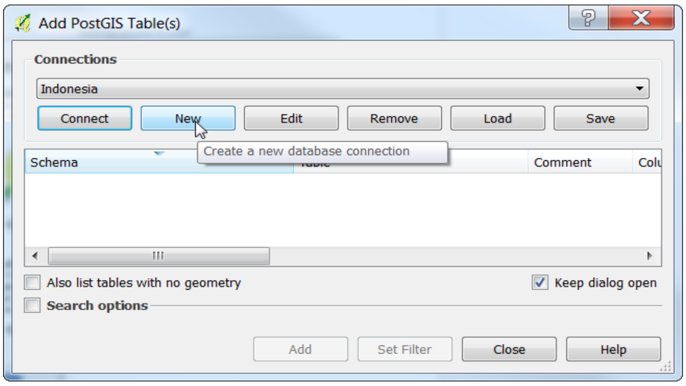
Many companies use a firewall system to isolate their networks from unauthorized access from the Internet. A firewall can be used to restrict access to your network by forwarding only requests targeted at specific TCP/IP addresses in the local network. Requests for all other network addresses are blocked by the firewall. You can allow Internet applications to access an instance of the postgreSQL server in the local network by configuring the firewall to forward network requests that specify the network address of the instance of the database server.

Access using QGIS

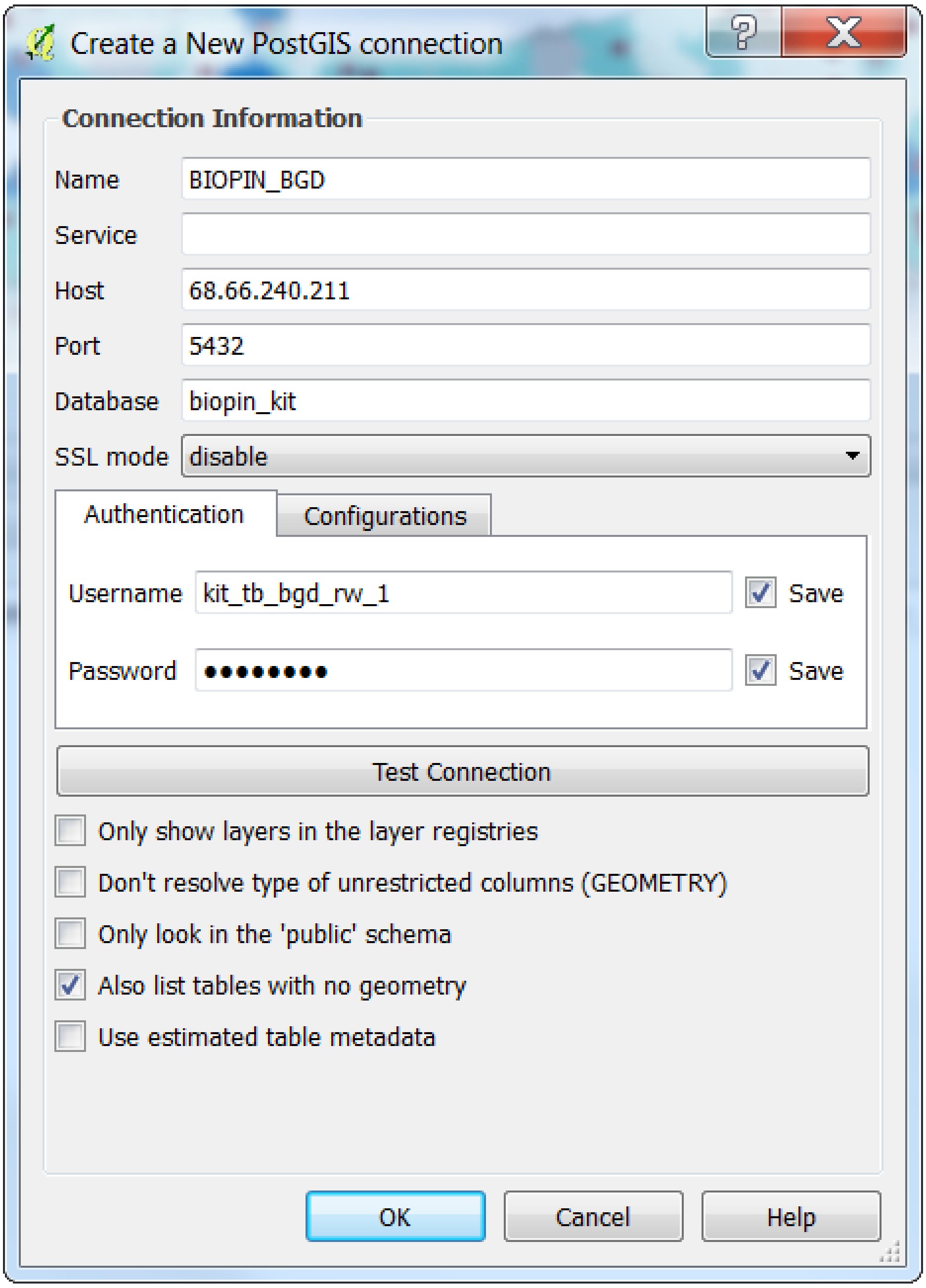
1. To load data from the server into QGIS a connection to this database should be made using a password and username.
2. In QGIS Desktop, click "Add PostGIS Table(s)" on the left side of your screen:



3. In the Add PostGIS table interface, click New. A dialogue box will appear.



4. In the dialogue box the login details can be entered as shown in the example below replacing the username and password details to those supplied by the database manager. This will grant you access rights to the data according to the role assigned to a specific user.



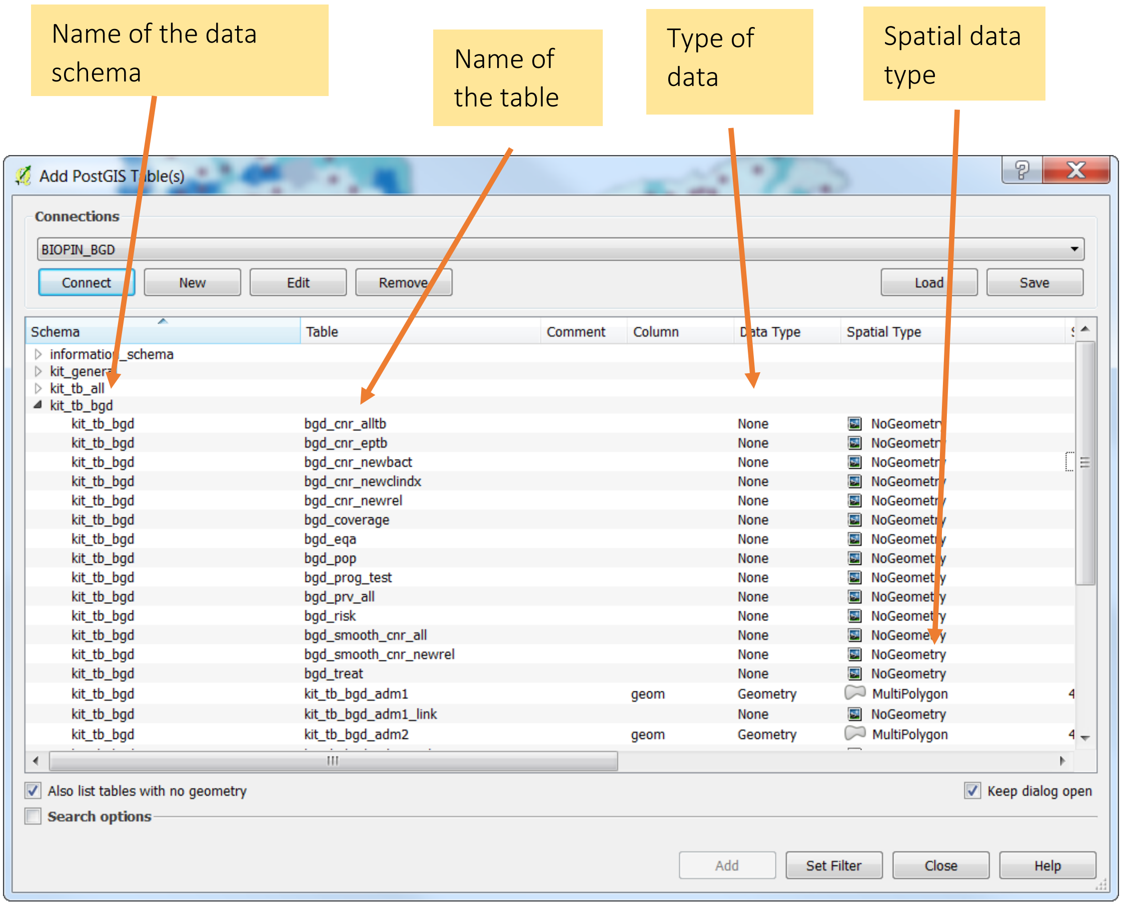
5. Once completed click ***TEST Connection***

6. If the connection was established click **OK**. This will bring you back to the Add PostGIS data dialogue box.

7. Select the database connection you just created.

8. Check the box ***ALSO LIST TABLES WITH NO GEOMETRY*** and click Connect.

If the connection was established the following tables should show (here Bangladesh is shown).



9. To load data to your QGIS project, select the (spatial) data table you require and click the button **Add.**

Annex 1 Variable codebooks of data tables included in the database

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Notification	GEO_UNIT	GEO_UNIT	character(5)	Geographical unit of reporting (ADM1, ADM2, OTHER)
Notification	GEO_ID	GEO_ID	character(9)	Alphanumeric ID of geographical unit (primary key)
Notification	TIME_UNIT	TIME_UNIT	character(1)	Temporal unit of reporting (Y = year, Q= quarter, P = period)
Notification	YEAR	YEAR	Date(YYYY)	Year or report
Notification	QRTR	QRTR	Date(Q)	Quarter of reporting year (ordinal value Q1, Q2, Q3,Q4)
Notification	new_sp	new_labconf	integer(9)	New pulmonary smear-positive cases
Notification	new_sn	new_clindx	integer(9)	New pulmonary smear-negative cases
Notification	new_ep	new_ep	integer(9)	New extra pulmonary cases (bacteriologically confirmed or clinically diagnosed). As of 2013 this also includes extra pulmonary cases with unknown previous TB treatment history.
Notification	new_oth	…	integer(9)	Other new cases
Notification	ret_rel	ret_rel_labconf	integer(9)	Relapse cases
Notification	ret_tad	ret_nrel_labconf	integer(9)	Treatment after default cases
Notification	ret_taf	ret_nrel_labconf	integer(9)	Treatment after failure cases
Notification	ret_oth	ret_nrel_labconf	integer(9)	Other re-treatment cases
Notification	new_sp_m04	…	integer(9)	New pulmonary smear positive cases: males aged 0-4 years
Notification	new_sp_f04	…	integer(9)	New pulmonary smear positive cases: females aged 0-4 years
Notification	new_sp_m514	…	integer(9)	New pulmonary smear positive cases: males aged 5-14 years
Notification	new_sp_f514	…	integer(9)	New pulmonary smear positive cases: females aged 5-14 years
Notification	new_sp_m1524	…	integer(9)	New pulmonary smear positive cases: males aged 15-24 years
Notification	new_sp_f1524	…	integer(9)	New pulmonary smear positive cases: females aged 15-24 years
Notification	new_sp_m2534	…	integer(9)	New pulmonary smear positive cases: males aged 25-34 years
Notification	new_sp_f2534	…	integer(9)	New pulmonary smear positive cases: females aged 25-34 years
Notification	new_sp_m3544	…	integer(9)	New pulmonary smear positive cases: males aged 35-44 years
Notification	new_sp_f3544	…	integer(9)	New pulmonary smear positive cases: females aged 35-44 years
Notification	new_sp_m4554	…	integer(9)	New pulmonary smear positive cases: males aged 45-54 years
Notification	new_sp_f4554	…	integer(9)	New pulmonary smear positive cases: females aged 45-54 years
Notification	new_sp_m5564	…	integer(9)	New pulmonary smear positive cases: males aged 55-64 years
Notification	new_sp_f5564	…	integer(9)	New pulmonary smear positive cases: females aged 55-64 years

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Notification	new_sp_m65	…	integer(9)	New pulmonary smear positive cases: males aged 65 years and over
Notification	new_sp_f65	…	integer(9)	New pulmonary smear positive cases: females aged 65 and over
Notification	new_sn_m04	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 0-4 years
Notification	new_sn_f04	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 0-4 years
Notification	new_sn_m514	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 5-14 years
Notification	new_sn_f514	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 5-14 years
Notification	new_sn_m1524	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 15-24 years
Notification	new_sn_f1524	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 15-24 years
Notification	new_sn_m2534	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 25-34 years
Notification	new_sn_f2534	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 25-34 years
Notification	new_sn_m3544	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 35-44 years
Notification	new_sn_f3544	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 35-44 years
Notification	new_sn_m4554	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 45-54 years
Notification	new_sn_f4554	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 45-54 years
Notification	new_sn_m5564	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 55-64 years
Notification	new_sn_f5564	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 55-64 years
Notification	new_sn_m65	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: males aged 65 years and over
Notification	new_sn_f65	…	integer(9)	New pulmonary smear negative/smear unknown/smear not done cases: females aged 65 years and over
Notification	new_ep_m04	…	integer(9)	New extra pulmonary cases: males aged 0-4 years
Notification	new_ep_f04	…	integer(9)	New extra pulmonary cases: females aged 0-4 years
Notification	new_ep_m514	…	integer(9)	New extra pulmonary cases: males aged 5-14 years
Notification	new_ep_f514	…	integer(9)	New extra pulmonary cases: females aged 5-14 years
Notification	new_ep_m1524	…	integer(9)	New extra pulmonary cases: males aged 15-24 years
Notification	new_ep_f1524	…	integer(9)	New extra pulmonary cases: females aged 15-24 years
Notification	new_ep_m2534	…	integer(9)	New extra pulmonary cases: males aged 25-34 years
Notification	new_ep_f2534	…	integer(9)	New extra pulmonary cases: females aged 25-34 years
Notification	new_ep_m3544	…	integer(9)	New extra pulmonary cases: males aged 35-44 years
Notification	new_ep_f3544	…	integer(9)	New extra pulmonary cases: females aged 35-44 years
Notification	new_ep_m4554	…	integer(9)	New extra pulmonary cases: males aged 45-54 years

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Notification	new_ep_f4554	…	integer(9)	New extra pulmonary cases: females aged 45-54 years
Notification	new_ep_m5564	…	integer(9)	New extra pulmonary cases: males aged 55-64 years
Notification	new_ep_f5564	…	integer(9)	New extra pulmonary cases: females aged 55-64 years
Notification	new_ep_m65	…	integer(9)	New extra pulmonary cases: males aged 65 years and over
Notification	new_ep_f65	…	integer(9)	New extra pulmonary cases: females aged 65 years and over
Notification	…	…	integer(9)	Bacteriologically confirmed cases with unknown treatment history
Notification	…	oth_clindx	integer(9)	Clinically diagnosed cases with unknown treatment history
Notification	…	oth_ep	integer(9)	Extra pulmonary cases with unknown treatment history
Notification	…	ret_rel_clindx	integer(9)	Clinically diagnosed cases retreated after relapse
Notification	…	ret_nrel_clindx	integer(9)	Clinically diagnosed cases retreated after failure, defaults and outcome unknown
Notification	…	ret_rel_ep	integer(9)	Extra pulmonary cases retreated after relapse
Notification	…	ret_nrel_ep	integer(9)	Extra pulmonary cases retreated failure
Notification	…	newrel_04m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_04f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_514m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_514f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_1524m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_1524f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_2534m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_2534f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_3544m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_3544f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_4554m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_4554f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_5564m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Notification	…	newrel_5564f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_65m	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Notification	…	newrel_65f	integer(9)	All forms of new and relapse cases and those with unknown treatment history which are clinically or bacteriologically confirmed
Laboratory		country	Character	country
Laboratory		name_higher_admin	Character	name of the administrative unit above the administrative unit at which data is disaggregated
Laboratory		level_higher_admin	Integer	level of the administrative unit above the administrative unit at which data is disaggregated
Laboratory		name_lower_admin	Character	name of the administrative unit at which data is disaggregated
Laboratory		level_lower_admin	Character	level of the administrative unit at which data is disaggregated
Laboratory		year	DATE(YYY)	year
Laboratory		quarter	DATE(Q)	quarter
Laboratory		population_size	Integer	population size of organizational geographic unit
Laboratory		suspect_test	Integer	number of individuals tested for tuberculosis at diagnosis
Laboratory		suspect_test_smear	Integer	number of individuals tested for tuberculosis by smear microscopy
Laboratory		suspect_test_xpert	Integer	number of individuals tested for tuberculosis by Xpert MTB/Rif
Laboratory		pos_test	Integer	number of tested individuals with a positive tuberculosis test result
Laboratory		pos_test_smear	Integer	number of patients with positive bacteriological results by smear microscopy
Laboratory		pos_test_xpert	Integer	number of patients with positive bacteriological results by Xpert
Laboratory		dx_ind_examined	Integer	number of individuals tested for tuberculosis at diagnosis
Laboratory		dx_test_smear	Integer	number of smears samples tested at diagnosis
Laboratory		dx_high_pos_smear	Integer	number of positive smears with 1+, 2+, or 3+ smear grading at diagnosis
Laboratory		dx_scant_pos_smear	Integer	number of positive smears with scanty (1-9 AFB/100) grading at diagnosis
Laboratory		dx_tot_pos_smear	Integer	total number of positive smears with 1+, 2+, 3+ or scanty (1-9 AFB / 100) grading at diagnosis (or total number of positive smears without an indication of smear grading at diagnosis)
Laboratory		dx_ind_tot_pos_smear	Integer	number of TB patients found TB positive at diagnostic examination
Laboratory		fu_ind_smear	Integer	number of TB patients examined for follow-up smear examination
Laboratory		fu_test_smear	Integer	number of smears samples tested during follow up examination
Laboratory		fu_high_pos_smear	Integer	number of positive smears with 1+, 2+, or 3+ smear grading during follow-up examination
Laboratory		fu_scant_pos_smear	Integer	number of positive smears with scanty (1-9 AFB/100) grading during follow-up examination
Laboratory		fu_tot_pos_smear	Integer	total number of positive smears with 1+, 2+, 3+ or scanty (1-9 AFB / 100) grading during follow-up examination

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
				(or total number of positive smears without an indication of smear grading during follow-up examination)
Laboratory		fu_ind_tot_pos_smear	Integer	number of TB patients found TB positive during follow-up smear examination
Laboratory		tot_test_smear	Integer	total number of smears tested (sum of the variables dx_test_smear and fu_test_smear)
Laboratory		tb_case_pos_rate	Real	tuberculosis case positivity rate
Laboratory		dx_sm_pos_rate	Real	diagnostic smear positivity rate
Laboratory		fu_sm_pos_rate	Real	follow-up smear positivity rate
Laboratory		total_hh_contacts	Integer	total number of household contacts identified
Laboratory		hh_contacts_screened	Integer	total number of household contacts screened
Laboratory		hh_cases_det	Integer	total number of household contact TB cases detected
Laboratory		hfp	Integer	high false positive
Laboratory		hfn	Integer	high false negative
Laboratory		lfp	Integer	low false positive
Laboratory		lfn	Integer	low false negative
Laboratory		qe	Integer	quantification errors
Laboratory		nr_districts	Integer	the number of districts within the province
Laboratory		nr_microscopic_hf	Integer	the number of microscopic health facilities
Laboratory		pc_hf_crosscheck	Real	percentage of health facilities that have carried out cross checks
Laboratory		nr_hf_error	Integer	The number of health facilities with fatal and or a minor error > 3
Laboratory		pc_hf_error	Real	the percentage of health facilities with fatal and or minor error > 3
Laboratory		nr_hf_minor_error	Integer	The number of health facilities with minor error < 3
Laboratory		pc_hf_minor_error	Real	The percentage of health facilities with minor error < 3
Laboratory		nr_hf_no_error	Integer	The number of health facilities with no error
Laboratory		pc_hf_no_error	Real	The percentage of health facilities with no error
Laboratory		nr_hf_eqa	Integer	The number of microscopic laboratory which shows adequate implementation of microscopic external quality assessment (cross check)
Laboratory		pc_hf_eqa	Real	The percentage of microscopic laboratory which shows adequate implementation of microscopic external quality assessment (cross check)
Laboratory		ansv	Integer	annual negative slide volume
Laboratory		spr	Real	slide positivity rate
Laboratory		dx_neg_smear	Integer	number negative slides collected
Laboratory		tot_pos_scn_neg	Integer	total number of slides collected

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Laboratory		fpr	Real	false positivity rate
Laboratory		fnr	Real	false negativity rate
Laboratory		sen	Real	sensitivity
Laboratory		ppv	Real	positive predictive value
Laboratory		fdc	Integer	functional diagnostic centers
Laboratory		ddc	Integer	diagnostic centers checked
Laboratory		cmje	Integer	diagnostic centers with major error
Treatment outcomes		GEO_UNIT	character(5)	Geographical unit of reporting (ADM1, ADM2, OTHER)
Treatment outcomes		GEO_ID	character(9)	Alphanumeric ID of geographical unit (primary key)
Treatment outcomes		TIME_UNIT	character(1)	Temporal unit of reporting (Y = year, Q= quarter, P = period)
Treatment outcomes		YEAR	Date(YYYY)	Year or report
Treatment outcomes		QRTR	Date(Q)	Quarter of reporting year (ordinal value Q1, Q2, Q3,Q4)
Treatment outcomes		new_sp_out_coh	integer(9)	new and relapse sputum smear positive cases in cohort
Treatment outcomes		new_sp_out_cur	integer(9)	new and relapse sputum smear positive cases cured
Treatment outcomes		new_sp_out_cmplt	integer(9)	new and relapse sputum smear positive cases completed treatment
Treatment outcomes		new_sp_out_lost	integer(9)	new and relapse sputum smear positive cases lost to follow up
Treatment outcomes		new_sp_out_fail	integer(9)	new and relapse sputum smear positive cases failed treatment
Treatment outcomes		new_sp_out_neval	integer(9)	new and relapse sputum smear positive cases not evaluated
Treatment outcomes		new_sp_out_died	integer(9)	new and relapse sputum smear positive cases who have died
Treatment outcomes		new_snep_coh	integer(9)	new and relapse extra pulmonary sputum smear positive cases in cohort
Treatment outcomes		new_snep_cmplt	integer(9)	new and relapse extra pulmonary sputum smear positive cases completed treatment
Treatment outcomes		new_snep_lost	integer(9)	new and relapse extra pulmonary sputum smear positive cases lost to follow up
Treatment outcomes		new_snep_fail	integer(9)	new and relapse extra pulmonary sputum smear positive cases failed treatment
Treatment outcomes		new_snep_neval	integer(9)	new and relapse extra pulmonary sputum smear positive cases not evaluated
Treatment outcomes		new_snep_died	integer(9)	new and relapse extra pulmonary sputum smear positive cases who have died
Treatment outcomes		ret_coh	integer(9)	retreatment cases in cohort
Treatment outcomes		ret_cur	integer(9)	retreatment cases cured
Treatment outcomes		ret_cmplt	integer(9)	retreatment cases completed treatment
Treatment outcomes		ret_lost	integer(9)	retreatment cases lost to follow up

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Treatment outcomes		ret_fail	integer(9)	retreatment cases failed treatment
Treatment outcomes		ret_neval	integer(9)	retreatment cases not evaluated
Treatment outcomes		ret_died	integer(9)	retreatment cases who have died
DHS		country	Character	Name of the country
DHS		division	Character	Name of the division/ first administrative level
DHS		year	DATE(YYYY)	Year of DHS report
DHS		stunting_u5	Real	Percentage of stunted children under 5, defined as height-for-age Z-scores -2SD below mean
DHS		sev_stunting_u5	Real	Percentage of severely stunted children under 5, defined as height-for-age Z-scores -3SD below mean
DHS		wasting_u5	Real	Percentage of wasted children under 5, defined as weight-for-height z-scores -2SD below mean
DHS		sev_wasting_u5	Real	Percentage of severely wasted children under 5, defined as weight-for-height z-scores -3SD below mean
DHS		underw_u5	Real	Percentage of underweight children under 5, defined as weight-for-age Z-scores of -2SD below mean
DHS		sev_underw_u5	Real	Percentage of severely underweight children under 5, defined as weight-for-age Z-scores of -3SD below mean
DHS		thinnes_female	Real	Percentage of thinness among ever-married women, defined as a BMI<18.5 kg/m2
DHS		mild_thin_female	Real	Percentage of mild thinness among ever-married women, defined as a 17.0< BMI <18.5 kg/m2
DHS		sev_thin_female	Real	Percentage of moderate and severe thinness among ever-married women, defined as a BMI <17.0 kg/m2
DHS		thinnes_male	Real	Percentage of thinness among men (15-49), defined as a BMI<18.5 kg/m2
DHS		mild_thin_male	Real	Percentage of mild thinness among men (15-49) , defined as a 17.0< BMI <18.5 kg/m2
DHS		sev_thin_male	Real	Percentage of moderate and severe thinness among men (15-49), defined as a BMI <17.0 kg/m2
DHS		mortality_u5	Real	Number of under-5 mortality per 1000 live births
DHS		wq_lowest	Real	Percentage of population in the lowest wealth quintile
DHS		wq_2nd	Real	Percentage of population in the second wealth quintile
DHS		wq_middle	Real	Percentage of population in the middle wealth quintile
DHS		wq_4th	Real	Percentage of population in the fourth wealth quintile
DHS		wq_highest	Real	Percentage of population in the highest wealth quintile
DHS		vaccine_u2	Real	Vaccination coverage for all basic vaccinations of children aged 12-23 months in percentage
DHS		bcg_coverage	Real	BCG vaccination coverage of children aged 12-23 months in percentage
DHS		no_vaccinations	Real	Percentage of children aged 12-23 months that did not receive any vaccinations
DHS		nr_vaccinations	Integer	Number of children aged 12-23 months included in the vaccination coverage survey
DHS		ari_u5	Real	Percentage of children under-five with symptoms of acute respiratory infections

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
DHS		ari_care	Real	Percentage of children under five with ARI symptoms for whom advice or treatment was sought from a health facility or provider
DHS		tobacco_fem	Real	Percentage of women using any kind of tobacco
DHS		tobacco_male	Real	Percentage of men using any kind of tobacco
DHS		smoke_fem	Real	Percentage of women that smoke cigarettes or bidis
DHS		smoke_male	Real	Percentage of men that smoke cigarettes or bidis
DHS		smoke_pipe_female	Real	Percentage of women that smoke pipe
DHS		smoke_pipe_male	Real	Percentage of men that smoke pipe
DHS		tobacco_other_female	Real	Percentage of women that use tobacco products other than cigarettes or pipe
DHS		tobacco_other_male	Real	Percentage of men that use tobacco products other than cigarettes or pipe
DHS		no_tobacco_female	Real	Percentage of women that do not use tobacco products
DHS		no_tobacco_male	Real	Percentage of men that do not use tobacco products
DHS		drinking_fem	Real	Percentage of women that drink alcohol
DHS		drinking_male	Real	Percentage of men that drink alcohol
DHS		insurance	Real	Percentage of households covered by a health scheme of health insurance
DHS		diabetes_fem	Integer	Number of women per 100,000 with diabetes
DHS		diabetes_male	Integer	Number of men per 100,000 with diabetes
DHS		tb_prevalence	Integer	Number of persons per 100,000 suffering from tuberculosis
DHS		tb_treatment	Integer	Number of persons per 100,000 who are receiving medical treatment for tuberculosis
DHS		hf_number	Integer	Number of health facilities
DHS		hf_distribution	Real	Distribution of health facilities as percentage of total number of health facilities
DHS		xray_tb	Real	Percentage of health facilities with equipment for diagnostic imaging with an X-ray machine (linked with TB)
DHS		microscopy_tb	Real	Percentage of health facilities with equipment for TB microscopy testing
DHS		tb_referral	Real	Percentage of facilities that refers clients for TB diagnosis
DHS		tb_diagnostics	Real	Percentage of facilities that offer TB diagnostic services by using either: sputum smear and/or X-ray or clinical symptoms. Or refer clients outside facility for TB diagnosis
DHS		tb_treatment_fu	Real	Percentage of facilities that offer TB treatment and/or follow-up services
DHS		tb_services	Real	Percentage of facilities that offer some form of TB services (diagnosis, treatment and/or follow-up)
DHS		tb_smear	Real	Percentage of facilities offering any form of TB services that offer TB smear microscopy
DHS		tb_culture	Real	Percentage of facilities offering any form of TB services that offer TB culture medium
DHS		tb_rdt	Real	Percentage of facilities offering any form of TB services that offer TB rapid diagnostic test kits

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
DHS		tb_xray	Real	Percentage of facilities offering any form of TB services that offer TB x-ray
DHS		tb_hiv	Real	Percentage of facilities offering any form of TB services that offer systems for diagnosing HIV among TB clients
DHS		tb_flmed	Real	Percentage of facilities offering any form of TB services that have first line treatment medicines for TB available
Census		country	Character	name of country
Census		division	Character	Name of division (ADM1)
Census		district	Character	name of district (ADM2)
Census		year	Date(YYYY)	year of census report
Census		population	Integer	total population in absolute numbers
Census		urban	Integer	urban population total
Census		rural	Integer	total rural population
Census		elderly	Real	Percentage of population over 'elderly age'
Census		male	Integer	total male population in absolute numbers
Census		female	Integer	Total female population in absolute numbers
Census		density	Real	Population density in persons/square meter
Census		illiteracy	Real	Percentage/number of population above a certain age who cannot read or write
Census		literacy_read	Real	Percentage/number of population above a certain age who can read but not write
Census		literacy	Real	Percentage/Number of population above a certain age that can both read and write
Census		household	Real	Average size of household
Key populations		country	Character	Name of country
Key populations		district	Character	Name of district (adm2)
Key populations		poor_extreme	Real	incidence of extreme poverty
Key populations		poor	Real	incidence of poverty
Key populations		country	Character	Name of country
Key populations		province	Character	Name of district (adm1)
Key populations		poverty incidence	Real	The proportion of individuals who are in a household with an average or per capita expenditure below the poverty line
Key populations		migrants	Integer	Number of migrants
Key populations		miners	Integer	Number of persons engaged with mining/quarrying
Key populations		mines	Integer	Number of mines
Key populations		country	Character	Name of country

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Key populations		region	Character	Name of district (adm1)
Key populations		nr_immigrant_male_2015	Integer	Number of male foreign residents
Key populations		nr_immigrant_fem_2015	Integer	Number of female foreign residents
Key populations		nr_immigrant_tot_2015	Integer	total number of foreign residents
Key populations		prison_convict_2015	Integer	Number of people convicted
Key populations		prison_undertrial_2015	Integer	Number of people under trial
Key populations		prison_tot_2015	Integer	Total number of people in prison
Key populations		mines_metal_2015	Integer	Number of metal mines
Key populations		mines_nonmetal_2015	Integer	Number of non-metal mines
Key populations		mines_total_2015	Integer	Total number of mines
Key populations		urban_poor_2010	Real	Poverty incidence urban areas
Key populations		rural_poor_2010	Real	Poverty incidence rural areas
Key populations		poor_total_2010	Real	Poverty incidence all areas
Key populations		country	Character	Name of country
Key populations		province	Character	Name of district (adm1)
Key populations		migrant	Real	Percentage population migrants
Key populations		migrant_distribution	Real	Distribution of migrant population
Key populations		mines	Integer	Number of mining/quarrying establishments
Key populations		miners	Integer	Average daily employment (Nos)
Key populations		poor_urban	Real	Percentage of urban population below poverty line
Key populations		country	Character	Name of country
Key populations		region	Character	Name of district (adm1)
Key populations		nr_immigrant_tot	Integer	Total number of immigrants from a foreign country
Key populations		immigrant_domestic_short	Integer	Total number of within state/region immigrants
Key populations		immigrants_domestic_long	Integer	Total number of inter-state/region immigrants
Key populations		prison_tot	Integer	Number of prisoners
Key populations		miners	Integer	Number of miners
Key populations		country	Character	Name of country
Key populations		region	Character	Name of district (adm1)

DATA	Data element	Alternative name	Type(length)	DESCRIPTION
Key populations		province	Character	Name of district (adm2)
Key populations		migration_in_2014	Real	Migration rate into the province
Key populations		migration_out_2014	Real	Migration rate out of the province
Key populations		migration_net_2014	Real	Net migration rate of the province
Key populations		new_hiv_2015	Integer	Newly infected HIV cases
Key populations		acc_hiv_2015	Integer	Accumulated HIV infected
Key populations		poverty_2015	Real	Poverty rate
Key populations		country	Character	Name of country
Key populations		development region	Character	Name of district (adm1)
Key populations		migrant_male	Integer	Number of male migrants internal
Key populations		migrant_femal	Integer	Number of female migrants internal
Key populations		migrant_distribution	Real	Distribution of total migrant population in percentage
Key populations		perc_miners	Real	Percentage of working population in mining and quarrying industry

Annex 2 probabilistic record Linkage in R

The following syntax provides an example of how R can be used to perform an fuzzy matching algorithm to link records from a NTP reporting units to the names of spatial data units (e.g. districts)

```
#install.packages('rvest')
#install.packages('stringr')
#install.packages('tidyr')
#install.packages('stringdist')

library(rvest)
library(stringr)
library(tidyr)
library(stringdist)

list.files("../LINK Table/")

kabkot <- read.csv("../LINK Table/Indo_Kab_Kot.csv")

NTP <- read.csv("../LINK Table/Indo_LAB.csv")

# Rename NTP data as S1 and spatial data attrubutes to S2
s1<-as.data.frame(NTP)
s2<-as.data.frame(kabkot)


# Set unit names to lower cases

s1$name<- as.character(tolower(NTP[,1]))
s2$name<- as.character(tolower(kabkot[,4]))

# Data cleaning: remove prefixes and suffixes relating to administrative
units or geographic units.

s1$name <- sub("kota", "", s1$name, ignore.case = TRUE, perl = FALSE,
               fixed = FALSE, useBytes = FALSE)

s1$name <- sub("kabupaten", "", s1$name, ignore.case = TRUE, perl = FALSE,
               fixed = FALSE, useBytes = FALSE)

s1$name <- sub("kepulauan", "", s1$name, ignore.case = TRUE, perl = FALSE,
               fixed = FALSE, useBytes = FALSE)

s2$name <- sub("kota", "", s2$name, ignore.case = TRUE, perl = FALSE,
               fixed = FALSE, useBytes = FALSE)

s2$name <- sub("kabupaten", "", s2$name, ignore.case = TRUE, perl = FALSE,
               fixed = FALSE, useBytes = FALSE)

s2$name <- sub("kepulauan", "", s2$name, ignore.case = TRUE, perl = FALSE,
               fixed = FALSE, useBytes = FALSE)


# Removing trailing and leading spaces in unit names.
trim <- function (x) gsub("^\\s+|\\s+$|\\s+", "", x)

s1$name <- trim(s1$name)
s2$name <- trim(s2$name)
```

```
#### Match function ####
# Create a matrix with the Standard Levenshtein distance between the name
fields of both sources
dist.name <- stringdistmatrix(s1$name,s2$name, method='jw', p=0.1)

# We now take the pairs with the minimum distance for each row in the
distance matrix
min.name <-apply(dist.name, 1, min)

# Create and empty object
match.s1.s2<<-NULL

# Iterate over each row in the distance matrix and return the first name
which matches the minimum distance found
for(i in 1:nrow(dist.name))
{
  s2.i<-match(min.name[i],dist.name[i,])
  s1.i<-i
  match.s1.s2<<-rbind(data.frame(s2.i=s2.i,s1.i=s1.i,s2name=s2[s2.i,]$name,
s1name=s1[s1.i,]$name, adist=min.name[i]),match.s1.s2)
}

# Add a field to the spatial data table (S2) based on the row index
s2$s2.i <- as.integer(row.names(s2))

# Merge the match table based on the row index

match.s1.s2 <- merge(match.s1.s2, s2, by= "s2.i")

# Add a field to the NTP data table (S2) based on the row index

s1$s1.i <- as.integer(row.names(s1))

# Merge the match table based on the row index

match.s1.s2 <- merge(match.s1.s2, s1, by= "s1.i")

# Create a field called duplicates with values "0"

match.s1.s2$s1.dplc <- 0

# Label units with duplicate matches as in the matching results with values
"1"

match.s1.s2$s1.dplc[duplicated(match.s1.s2$name, fromLast = FALSE)] <- 1

match.s1.s2$s1.dplc[duplicated(match.s1.s2$name, fromLast = TRUE)] <- 1


# Create a field called s1.nomatch with values "0"

match.s1.s2$s1.nomatch <- 0

# label all imperfect matches with a value of "1"

match.s1.s2$s1.nomatch[match.s1.s2$adist > 0.06] <- 1

# Tabulate the results

with(match.s1.s2, table(PROVINSI, s1.nomatch))
```

```
with(match.s1.s2[match.s1.s2$s1.dplc == 1,], table(PROVINSI, s1.nomatch))

match.s1.s2[which(match.s1.s2$adist > 0.0),]

match.s1.s2[which(match.s1.s2$s1.dplc == 1),]
# Write the final results to a csv table
write.csv(match.s1.s2, XXX_match.csv")
```

Annex 3 Indicator calculations.

The following calculations are used to calculate key TB indicators based on NTP reports and population data.

VIEW NAME	COLUMNS	CALCULATION Before2014	TYPE
(XXX = Country ISO3 code)	These are the column names as they can be defined in the VIEW or TABLE		

XXX_CNR_alltb	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_all_2013	100000 * (new_sp + ret_rel + ret_taf + ret_tad + ret_oth + new_sn + new_oth + new_su + new_ep) / population_YY	DECIMAL(10,3)
	cnr_all_2014	100000 * (new_labconf + ret_rel_labconf + ret_nrel_labconf + new_clindx + new_ep) / population_YY	DECIMAL(10,3)
	cnr_all_2015		DECIMAL(10,3)
	cnr_all_2016		DECIMAL(10,3)
	cn_all_2013	new_sp + ret_rel + ret_taf + ret_tad + ret_oth + new_sn + new_oth + new_su + new_ep	DECIMAL(10,3)
	cn_all_2014	new_labconf + ret_rel_labconf + ret_nrel_labconf + new_clindx + new_ep	DECIMAL(10,3)
	cn_all_2015		DECIMAL(10,3)
	cn_all_2016		DECIMAL(10,3)
XXX_SMOOTH_CNR_alltb	geo_id	SHP_ADM2_UID	CHAR(6)
	lagincl_cnr_all_2013	spatially lagged average of VIEW XXX_CNR_allform, including the focal area (see worksheet sql for syntax)	DECIMAL(10,3)
	lagincl_cnr_all_2014		DECIMAL(10,3)
	lagincl_cnr_all_2015		DECIMAL(10,3)
	lagincl_cnr_all_2016		DECIMAL(10,3)
	lagexcl_cnr_all_2013	spatially lagged average of VIEW XXX_CNR_allform, excluding the focal area (see worksheet sql for syntax)	DECIMAL(10,3)
	lagexcl_cnr_all_2014		DECIMAL(10,3)
	lagexcl_cnr_all_2015		DECIMAL(10,3)
	lagexcl_cnr_all_2016		DECIMAL(10,3)
XXX_CNR_newrel	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_newrel_2013	100000 * (new_sp + ret_rel + new_sn + new_oth + new_su + new_ep) / population_YY	DECIMAL(10,3)
	cnr_newrel_2014	100000 * (new_labconf + ret_rel_labconf + new_clindx + new_ep) / population_YY	DECIMAL(10,3)
	cnr_newrel_2015		DECIMAL(10,3)

VIEW NAME	COLUMNS	CALCULATION Before2014	TYPE
(XXX = Country ISO3 code)	These are the column names as they can be defined in the VIEW or TABLE		
	cnr_newrel_2016		DECIMAL(10,3)
	cn_newrel_2013	new_sp + ret_rel + new_sn + new_oth + new_su + new_ep	DECIMAL(10,3)
	cn_newrel_2014	new_labconf + ret_rel_labconf + new_clindx + new_ep	DECIMAL(10,3)
	cn_newrel_2015		DECIMAL(10,3)
	cn_newrel_2016		DECIMAL(10,3)
XXX_SMOOTH_CNR_newrel	geo_id	SHP_ADM2_UID	CHAR(6)
	lagincl_cnr_all_2013	spatially lagged average of VIEW XXX_CNR_newrel including the focal area (see worksheet sql for syntax)	DECIMAL(10,3)
	lagincl_cnr_all_2014		DECIMAL(10,3)
	lagincl_cnr_all_2015		DECIMAL(10,3)
	lagincl_cnr_all_2016		DECIMAL(10,3)
	lagexcl_cnr_all_2013	spatially lagged average of VIEW XXX_CNR_newrel excluding the focal area (see worksheet sql for syntax)	DECIMAL(10,3)
	lagexcl_cnr_all_2014		DECIMAL(10,3)
	lagexcl_cnr_all_2015		DECIMAL(10,3)
	lagexcl_cnr_all_2016		DECIMAL(10,3)
XXX_CNR_newbact	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_newbact_2013	100000 * (new_sp) / population_YY	DECIMAL(10,3)
	cnr_newbact_2014	100000 * (new_labconf) / population_YY	DECIMAL(10,3)
	cnr_newbact_2015		DECIMAL(10,3)
	cnr_newbact_2016		DECIMAL(10,3)
	cn_newbact_2013	new_sp	DECIMAL(10,3)
	cn_newbact_2014	new_labconf	DECIMAL(10,3)
	cn_newbact_2015		DECIMAL(10,3)
	cn_newbact_2016		DECIMAL(10,3)
XXX_CNR_newclindx	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_newbact_2013	100000 * (new_sn + new_oth + new_su) / population_YY	DECIMAL(10,3)
	cnr_newbact_2014	100000 * (new_clindx) / population_YY	DECIMAL(10,3)

VIEW NAME	COLUMNS	CALCULATION Before2014	TYPE
(XXX = Country ISO3 code)	These are the column names as they can be defined in the VIEW or TABLE		
	cnr_newbact_2015		DECIMAL(10,3)
	cnr_newbact_2016		DECIMAL(10,3)
	cn_newbact_2013	new_sn + new_oth + new_su	DECIMAL(10,3)
	cn_newbact_2014	new_clindx	DECIMAL(10,3)
	cn_newbact_2015		DECIMAL(10,3)
	cn_newbact_2016		DECIMAL(10,3)
XXX_CNR_eptb	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_eptb_2013	100000 * new_ep / population_YY	DECIMAL(10,3)
	cnr_eptb_2014		DECIMAL(10,3)
	cnr_eptb_2015		DECIMAL(10,3)
	cnr_eptb_2016		DECIMAL(10,3)
	cn_eptb_2013	new_ep	DECIMAL(10,3)
	cn_eptb_2014		DECIMAL(10,3)
	cn_eptb_2015		DECIMAL(10,3)
	cn_eptb_2016		DECIMAL(10,3)
XXX_CNR_child04	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_child04_2013	100000 * (new_sp_m04 + new_sp_f04 + new_oth_m04 + new_oth_f04 + new_sn_m04 + new_sn_f04 + new_ep_m04 + new_en_f04 +/ population04_YY	DECIMAL(10,3)
	cnr_child04_2014		DECIMAL(10,3)
	cnr_child04_2015		DECIMAL(10,3)
	cnr_child04_2016		DECIMAL(10,3)
XXX_CNR_child514	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_child514_2013	100000 * (new_sp_m514 + new_sp_f514 + new_oth_m514 + new_oth_f514 + new_sn_m514 + new_sn_f514 + new_ep_m514 + new_en_f514 +/ population514_YY	DECIMAL(10,3)
	cnr_child514_2014		DECIMAL(10,3)
	cnr_child514_2015		DECIMAL(10,3)
	cnr_child514_2016		DECIMAL(10,3)

VIEW NAME	COLUMNS	CALCULATION Before2014	TYPE
(XXX = Country ISO3 code)	These are the column names as they can be defined in the VIEW or TABLE		
XXX_CNR_male	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_male_2014	100000 * (newrel_04m + newrel_514m + newrel_1524m + newrel_2534m + newrel_3544m + newrel_4554m + newrel_5564m + newrel_65m) / population_YYmale	DECIMAL(10,3)
	cnr_male_2015		DECIMAL(10,3)
	cnr_male_2016		DECIMAL(10,3)
XXX_CNR_female	geo_id	SHP_ADM2_UID	CHAR(6)
	cnr_female_2014	100000 * (newrel_04f + newrel_514f + newrel_1524f + newrel_2534f + newrel_3544f + newrel_4554f + newrel_5564f + newrel_65f) / population_YYfemale	DECIMAL(10,3)
	cnr_female_2015		DECIMAL(10,3)
	cnr_female_2016		DECIMAL(10,3)
XXX_PRV_all	geo_id	SHP_ADM2_UID	CHAR(6)
	TB_prev_YYYY	100000 * cases / population_YY	DECIMAL(6,3)
XXX_PROG_test	geo_id	SHP_ADM2_UID	CHAR(6)
	testrate_2013	100000 * suspect_test / population_YY	DECIMAL(6,3)
	testrate_2014		DECIMAL(6,3)
	testrate_2015		DECIMAL(6,3)
	testrate_2016		DECIMAL(6,3)
	posrate_2013	pos_test / suspect_test	DECIMAL(6,3)
	posrate_2014		DECIMAL(6,3)
	posrate_2015		DECIMAL(6,3)
	posrate_2016		DECIMAL(6,3)
	facilitydens	population_YY / "Number of facilities per geographic ADM2 unit"	DECIMAL(10,3)
XXX_EQA	geo_id	SHP_ADM2_UID	CHAR(6)
	workload	tot_test_smear	INTEGER(10)
	error_all_2013	(hfp + hfn + lfp +lfn)	DECIMAL(6,3)
	error_all_2014		DECIMAL(6,3)

VIEW NAME	COLUMNS	CALCULATION Before2014	TYPE
(XXX = Country ISO3 code)	These are the column names as they can be defined in the VIEW or TABLE		
	error_all_2015		DECIMAL(6,3)
	error_all_2016		DECIMAL(6,3)
	error_hfp_2013	hfp	DECIMAL(6,3)
	error_hfp_2014		DECIMAL(6,3)
	error_hfp_2015		DECIMAL(6,3)
	error_hfp_2016		DECIMAL(6,3)
	error_hfn_2013	hfn	DECIMAL(6,3)
	error_hfn_2014		DECIMAL(6,3)
	error_hfn_2015		DECIMAL(6,3)
	error_hfn_2016		DECIMAL(6,3)
XXX_RISK	geo_id	SHP_ADM2_UID	CHAR(6)
	HIV_prev_YYYY	100000 * cases / population_YY	DECIMAL(6,3)
	DIAB_prev_YYYY	reported rate	DECIMAL(6,3)
	SMOKE_prev_YYYY	reported rate	DECIMAL(6,3)
	WEALTH_Q1_YYYY	reported rate	DECIMAL(6,3)
	WEALTH_Q2_YYYY	reported rate	DECIMAL(6,3)
	WEALTH_Q3_YYYY	reported rate	DECIMAL(6,3)
	WEALTH_Q4_YYYY	reported rate	DECIMAL(6,3)
	STUNT	% of stunted children under 5	DECIMAL(6,3)
	WASTE	% of wasted children under 5	DECIMAL(6,3)
	ARI	% of children under 5 with symptoms of acute respiratory infections	DECIMAL(6,3)
	POPDENS	Population density	DECIMAL(6,3)
	URBAN	% of population living in urban areas	DECIMAL(6,3)
	ELDERLY	% of population defined as elderly	DECIMAL(6,3)
	LITRATE	literacy rate (% of population above certain age who can read and write)	DECIMAL(6,3)

VIEW NAME	COLUMNS	CALCULATION Before2014	TYPE
(XXX = Country ISO3 code)	These are the column names as they can be defined in the VIEW or TABLE		

XXX_TB_HIV (only from 2014-2016)	geo_id	SHP_ADM2_UID	CHAR(6)
	hiv_test	hiv_test	INTEGER(10)
	hiv_pos	hiv_pos	DECIMAL(6,3)
	hiv_art	hiv_art	DECIMAL(6,3)
	hiv_cpt	hiv_cpt	DECIMAL(6,3)
	hiv_testrate	hiv_test / (new_labconf + ret_rel_labconf + new_clindx + new_ep)	DECIMAL(6,3)
	hiv_pos_newrel	hiv_pos / (new_labconf + ret_rel_labconf + new_clindx + new_ep)	DECIMAL(6,3)
	hiv_pos_test	hiv_pos / hiv_test	DECIMAL(6,3)
	hiv_art_newrel	hiv_art / (new_labconf + ret_rel_labconf + new_clindx + new_ep)	DECIMAL(6,3)
	hiv_cpt_newrel	hiv_cpt / (new_labconf + ret_rel_labconf + new_clindx + new_ep)	DECIMAL(6,3)

XXX_ACF

XXX_Coverage	geo_id	SHP_ADM2_UID	CHAR(6)
	U5MORT_YYYY	Number of under 5 mortality per 1000 live births	DECIMAL(6,3)
	VACCOV_YYYY	Vaccination coverage (%) for all basic vaccinations of children aged 12-23m	DECIMAL(6,3)
	ARIADV_YYYY	% of under 5 children with ARI symptoms for whom advice was sought	DECIMAL(6,3)
	INSUR_YYYY	% of households covered by a health scheme/insurance	DECIMAL(6,3)

XXX_TREAT	geo_id	SHP_ADM2_UID	CHAR(6)
	TSR_2014	(new_sp_out_cur + new_sp_out_cmplt + new_snep_cmplt) / (new_sp_out_coh +new_snep_coh)	DECIMAL(6,3)
	TSR_2015	(new_sp_out_cur + new_sp_out_cmplt + new_snep_cmplt) / (new_sp_out_coh +new_snep_coh)	DECIMAL(6,3)
	TSR_2016	(new_sp_out_cur + new_sp_out_cmplt + new_snep_cmplt) / (new_sp_out_coh +new_snep_coh)	DECIMAL(6,3)
	DFLR_2014	new_sp_out_lost /new_sp_out_coh	DECIMAL(6,3)

VIEW NAME	COLUMNS	CALCULATION Before2014	TYPE
(XXX = Country ISO3 code)	These are the column names as they can be defined in the VIEW or TABLE		
	DFLR_2015	new_sp_out_lost /new_sp_out_coh	DECIMAL(6,3)
	DFLR_2016	new_sp_out_lost /new_sp_out_coh	DECIMAL(6,3)
	FLR_2014	new_sp_out_fail /new_sp_out_coh	DECIMAL(6,3)
	FLR_2015	new_sp_out_fail /new_sp_out_coh	DECIMAL(6,3)
	FLR_2016	new_sp_out_fail /new_sp_out_coh	DECIMAL(6,3)
	DTHR_2014	new_sp_out_died /new_sp_out_coh	DECIMAL(6,3)
	DTHR_2015	new_sp_out_died /new_sp_out_coh	DECIMAL(6,3)
	DTHR_2016	new_sp_out_died /new_sp_out_coh	DECIMAL(6,3)

Annex 4 SQL query for indicator calculation.

The following SQL syntax provides an example of how long tables can be used to calculate case notification rates and transform the output into a wide table for GIS mapping

```
-- -----------------------------------------------------------------------------
----
-- Make consolidated tables for notification. outcome & laboratory
-- -----------------------------------------------------------------------------
----

drop table if exists temp.kit_tb_bgd_notification_summary cascade;
create table temp.kit_tb_bgd_notification_summary as
   select geo_id, year, sum(new_sp) as new_sp, sum(ret_rel) as ret_rel,
sum(ret_taf) as  ret_taf,sum(ret_tad) as  ret_tad,sum(ret_oth) as
ret_oth,sum(new_sn) as  new_sn, sum(new_ep) as new_ep,
          sum(new_sp_m04) as new_sp_m04, sum(new_sp_f04) as  new_sp_f04 ,
sum(new_sn_m04) as new_sn_m04, sum(new_sn_f04) as new_sn_f04 ,
sum(new_ep_m04) as new_ep_m04, sum(new_ep_f04) as new_ep_f04,
          sum(new_sp_m514) as new_sp_m514, sum(new_sp_f514) as new_sp_f514,
sum(new_sn_m514) as new_sn_m514, sum(new_sn_f514) as new_sn_f514,
sum(new_ep_m514) as new_ep_m514, sum(new_ep_f514) as new_ep_f514,
          sum(new_labconf) as new_labconf, sum(ret_rel_labconf) as
ret_rel_labconf, sum(ret_nrel_labconf) as ret_nrel_labconf, sum(new_clindx)
as new_clindx
   from kit_tb_bgd.kit_tb_bgd_notification
   group by geo_id, year
  ;

drop table if exists temp.kit_tb_bgd_outcomes_summary cascade;
create table temp.kit_tb_bgd_outcomes_summary as
   select geo_id, year, sum(new_sp_out_cur) as
new_sp_out_cur,sum(new_sp_out_cmplt) as new_sp_out_cmplt,
sum(new_sp_out_lost) as new_sp_out_lost, sum(new_sp_out_fail) as
new_sp_out_fail,sum(new_sp_out_died) as  new_sp_out_died,
sum(new_sp_out_coh) as new_sp_out_coh
   from kit_tb_bgd.kit_tb_bgd_outcomes
   group by geo_id, year
  ;

drop table if exists temp.kit_tb_bgd_laboratory_summary cascade;
create table temp.kit_tb_bgd_laboratory_summary as
   select geo_id, year, sum(hfp) as hfp, sum(hfn) as hfn, sum(lfp) as
lfp,sum(lfn) as  lfn, sum(suspect_test) as suspect_test, sum(pos_test) as
pos_test, sum(tot_test_smear) as tot_test_smear
   from kit_tb_bgd.kit_tb_bgd_laboratory
   group by geo_id, year
  ;

drop table if exists temp.kit_tb_bgd_population_summary cascade;
create table temp.kit_tb_bgd_population_summary as
   select geo_id, year, sum(e_pop_num) as e_pop_num
   from kit_tb_bgd.kit_tb_bgd_population
   group by geo_id, year
  ;


-- -----------------------------------------------------------------------------
----
-- bgd_CNR_alltb
```

```
-- ------------------------------------------------------------------------
----
   drop table if exists kit_tb_bgd.bgd_cnr_alltb ;
           create table kit_tb_bgd.bgd_cnr_alltb as
         select cast (a.geo_id as char(6))
               ,cast(100000 * (coalesce(not_2013.new_sp,0) +
coalesce(not_2013.ret_rel,0) + coalesce(not_2013.ret_taf,0) +
coalesce(not_2013.ret_tad,0) + coalesce(not_2013.ret_oth,0) +
coalesce(not_2013.new_sn,0) + coalesce(not_2013.new_ep,0) +
coalesce(not_2013.new_labconf,0) + coalesce(not_2013.ret_rel_labconf,0) +
coalesce(not_2013.ret_nrel_labconf,0) +
coalesce(not_2013.new_clindx,0))/pop_2013.e_pop_num  as decimal(15,3)) as
cnr_all_2013
               ,cast(100000 * (coalesce(not_2014.new_sp,0) +
coalesce(not_2014.ret_rel,0) + coalesce(not_2014.ret_taf,0) +
coalesce(not_2014.ret_tad,0) + coalesce(not_2014.ret_oth,0) +
coalesce(not_2014.new_sn,0) + coalesce(not_2014.new_ep,0) +
coalesce(not_2014.new_labconf,0) + coalesce(not_2014.ret_rel_labconf,0) +
coalesce(not_2014.ret_nrel_labconf,0) +
coalesce(not_2014.new_clindx,0))/pop_2014.e_pop_num  as decimal(15,3)) as
cnr_all_2014
               ,cast(100000 * (coalesce(not_2015.new_sp,0) +
coalesce(not_2015.ret_rel,0) + coalesce(not_2015.ret_taf,0) +
coalesce(not_2015.ret_tad,0) + coalesce(not_2015.ret_oth,0) +
coalesce(not_2015.new_sn,0) + coalesce(not_2015.new_ep,0) +
coalesce(not_2015.new_labconf,0) + coalesce(not_2015.ret_rel_labconf,0) +
coalesce(not_2015.ret_nrel_labconf,0) +
coalesce(not_2015.new_clindx,0))/pop_2015.e_pop_num  as decimal(15,3)) as
cnr_all_2015
               ,cast(100000 * (coalesce(not_2016.new_sp,0) +
coalesce(not_2016.ret_rel,0) + coalesce(not_2016.ret_taf,0) +
coalesce(not_2016.ret_tad,0) + coalesce(not_2016.ret_oth,0) +
coalesce(not_2016.new_sn,0) + coalesce(not_2016.new_ep,0) +
coalesce(not_2016.new_labconf,0) + coalesce(not_2016.ret_rel_labconf,0) +
coalesce(not_2016.ret_nrel_labconf,0) +
coalesce(not_2016.new_clindx,0))/pop_2016.e_pop_num  as decimal(15,3)) as
cnr_all_2016
               ,cast(          coalesce(not_2013.new_sp,0) +
coalesce(not_2013.ret_rel,0) + coalesce(not_2013.ret_taf,0) +
coalesce(not_2013.ret_tad,0) + coalesce(not_2013.ret_oth,0) +
coalesce(not_2013.new_sn,0) + coalesce(not_2013.new_ep,0) +
coalesce(not_2013.new_labconf,0) + coalesce(not_2013.ret_rel_labconf,0) +
coalesce(not_2013.ret_nrel_labconf,0) + coalesce(not_2013.new_clindx,0)
as decimal(15,3)) as cn_all_2013
               ,cast(          coalesce(not_2014.new_sp,0) +
coalesce(not_2014.ret_rel,0) + coalesce(not_2014.ret_taf,0) +
coalesce(not_2014.ret_tad,0) + coalesce(not_2014.ret_oth,0) +
coalesce(not_2014.new_sn,0) + coalesce(not_2014.new_ep,0) +
coalesce(not_2014.new_labconf,0) + coalesce(not_2014.ret_rel_labconf,0) +
coalesce(not_2014.ret_nrel_labconf,0) + coalesce(not_2014.new_clindx,0)
as decimal(15,3)) as cn_all_2014
               ,cast(          coalesce(not_2015.new_sp,0) +
coalesce(not_2015.ret_rel,0) + coalesce(not_2015.ret_taf,0) +
coalesce(not_2015.ret_tad,0) + coalesce(not_2015.ret_oth,0) +
coalesce(not_2015.new_sn,0) + coalesce(not_2015.new_ep,0) +
coalesce(not_2015.new_labconf,0) + coalesce(not_2015.ret_rel_labconf,0) +
coalesce(not_2015.ret_nrel_labconf,0) + coalesce(not_2015.new_clindx,0)
as decimal(15,3)) as cn_all_2015
               ,cast(          coalesce(not_2016.new_sp,0) +
coalesce(not_2016.ret_rel,0) + coalesce(not_2016.ret_taf,0) +
coalesce(not_2016.ret_tad,0) + coalesce(not_2016.ret_oth,0) +
```

```
coalesce(not_2016.new_sn,0) + coalesce(not_2016.new_ep,0) +
coalesce(not_2016.new_labconf,0) + coalesce(not_2016.ret_rel_labconf,0) +
coalesce(not_2016.ret_nrel_labconf,0) + coalesce(not_2016.new_clindx,0)
as decimal(15,3)) as cn_all_2016
from (select distinct geo_id from temp.kit_tb_bgd_notification_summary) a
     left join (select * from temp.kit_tb_bgd_notification_summary where
year = 2013) as not_2013
           on a.geo_id = not_2013.geo_id
     left join (select * from temp.kit_tb_bgd_notification_summary where
year = 2014) as not_2014
           on a.geo_id = not_2014.geo_id
     left join (select * from temp.kit_tb_bgd_notification_summary where
year = 2015) as not_2015
           on a.geo_id = not_2015.geo_id
     left join (select * from temp.kit_tb_bgd_notification_summary where
year = 2016) as not_2016
           on a.geo_id = not_2016.geo_id
     left join (select * from temp.kit_tb_bgd_population_summary where year
= 2013) as pop_2013
           on a.geo_id = pop_2013.geo_id
     left join (select * from temp.kit_tb_bgd_population_summary where year
= 2014) as pop_2014
           on a.geo_id = pop_2014.geo_id
     left join (select * from temp.kit_tb_bgd_population_summary where year
= 2015) as pop_2015
           on a.geo_id = pop_2015.geo_id
     left join (select * from temp.kit_tb_bgd_population_summary where year
= 2016) as pop_2016
           on a.geo_id = pop_2016.geo_id
;

   alter table kit_tb_bgd.bgd_cnr_alltb add constraint pk_bgd_cnr_alltb
primary key (geo_id);
   create unique index idx_bgd_cnr_alltb on kit_tb_bgd.bgd_cnr_alltb
(geo_id);
```