TRACE Data Visualization Developer Manual

Reference Guide for Using Microsoft Power BI
and ESRI ArcGIS Online for Data Visualization

Table of Contents

Acronyms

AGOL	ArcGIS Online
BI	Business Intelligence
csv	Comma Separated Values File
DAX	Data Analysis Expression Language
DHIS2	District Health Information System
IT	Information Technology
M	Mashup Language
.pbit	File extension for Power BI Template File
.pbix	File extension for Power BI Desktop File
ODK	Open Data Kit
QC	Quality Control
RITA	Recent Infection Testing Algorithm
RTRI	Rapid Test for HIV-1 Recent Infection
TRACE	Tracking with Recency Assays to Control the Epidemic

Definitions

Admin	Refers to Power BI system administrators. Ordinarily this is an IT/Informatics team member with responsibility for maintaining institutional software licenses, system management and user access.
DAX	DAX stands for Data Analysis Expressions and is the formula language used in the front end of Power BI. DAX is also used in PowerPivot (in Excel) and SSAS Tabular. DAX is a functional language, which means the full executed code is contained inside a function.
Developer	Refers to individuals who will be building Power BI reports. Ordinarily this is an IT/Informatics/M&E team member with a solid understanding of relational data models.
End-User	Refers to consumers of Power BI data. Typically, these individuals are responsible for monitoring programmatic outputs and making data driven decisions.
Forums	Refers to the online Power BI community where users can ask questions to and read about problems/questions other users have experienced.
IT administrator	Refers to the individual performing the installation of Power BI and with administrative rights.
M	M stands for Mashup, and is the formula language used in Power Query, in the back end of Power BI. The M language is typically executed through point and click actions.
Visualizations/Visuals	A visualization or visual is a graphic representation of information or data to communicate relationships using visual elements, such as tables, charts and graphs, or maps. The goal of a visual is to present data in a way that provides context and insights, that may be difficult to discern from a raw table of numbers or text.

1. Background

The TRACE Initiative is supporting two critical innovations to improve surveillance of new HIV infections. First and most notably, TRACE is supporting the roll out, in routine HIV testing services, of the rapid test for recent infection (RTRI) or RTRI alone or the Recent Infection Testing Algorithm or RITA, which combines RTRI and baseline viral load results in establishing recent infection status of a client. Secondly, TRACE is supporting the development and maintenance of electronic data systems of RTRI and RITA results along with critical demographic and risk factor information to allow for real-time continuous capture and characterization of new HIV infections. The use of a data visualization interface is integral to establishing near real-time recency surveillance monitoring in routine HIV services to give health officials at all levels (i.e., site, district, and central) access to actionable information about their epidemic.

Microsoft Power BI business intelligence software and ArcGIS Online are used in concert to produce near real-time visualizations of HIV recent infection surveillance data. This data visualization platform provides a template for visualizing data on recent infection to aid in the rapid identification of testing quality issues, demonstrate critical epidemiological trends and geographical clusters and/or sub-populations with high potential of HIV transmission, upon which countries can build additional analysis depending on available data and need.

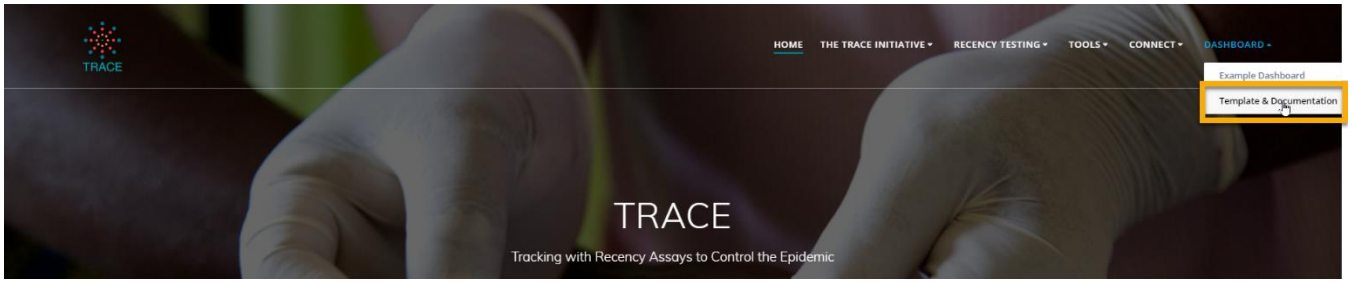
2. Purpose of this Manual

The ***TRACE Data Visualization Developer Manual*** (hereafter, "***manual***") serves as a reference document to guide data managers and analysts of HIV recency surveillance programs in using Power BI and ArcGIS to visualize HIV recent infection surveillance data and help program implementers make informed, data-driven, programmatic decisions. The manual is designed to provide information to data managers on installation and setup of Power BI, configuration of simulated data using the ***Power BI TRACE Template*** (hereafter, "***TRACE Template***"), data linkage to ArcGIS Online, and embedding maps for view within the TRACE dashboards. The document also serves as a guide for data managers to connect in-country teams' recency data with the *TRACE Template*, so that the data visualized within Power BI accurately reflects local HIV recent infection surveillance data. Users can think of this data visualization package as a "starter kit" for using Power BI to visualize recency data.

This guidance aims to provide a starting point for visualizing HIV recent infection surveillance data using Power BI and ArcGIS Online. Following the implementation of this guidance and use of the *TRACE Template*, users can further customize visualizations and reports to meet country-specific project needs.

This manual accompanies the ***Power BI for TRACE: System Administration Guide***, which along with the ***TRACE Template*** and ***Simulated Data Files***, can all be found on the TRACE eLearning hub (https://trace-recency.org/). The complete set of documents and files needed to get starter are:

1) *Power BI for TRACE: System Administration Guide*
2) *TRACE Template (Power BI Desktop File (.pbix) and Power BI Template File (.pbit))*
3) *Simulated Data Files*
    a. *RecencyData*
    b. *IndexTestingData*
    c. *LabQCData*
    d. *FacilityXYData*
4) *Power BI Resource Guide*



Location of Templates and Documentation on website trace-recency.org

3. Using Power BI for TRACE

The TRACE Initiative is harnessing the capabilities of both Power BI and ArcGIS online to monitor the implementation of HIV recent infection surveillance systems. This section gives a brief introduction to Power BI and how it is being used to monitor recency implementation. *Section 7. Embedding Online Maps for TRACE* describes how ArcGIS online is being used and how to link online maps to the Power BI workspace.

3.1. What is Microsoft Power BI?

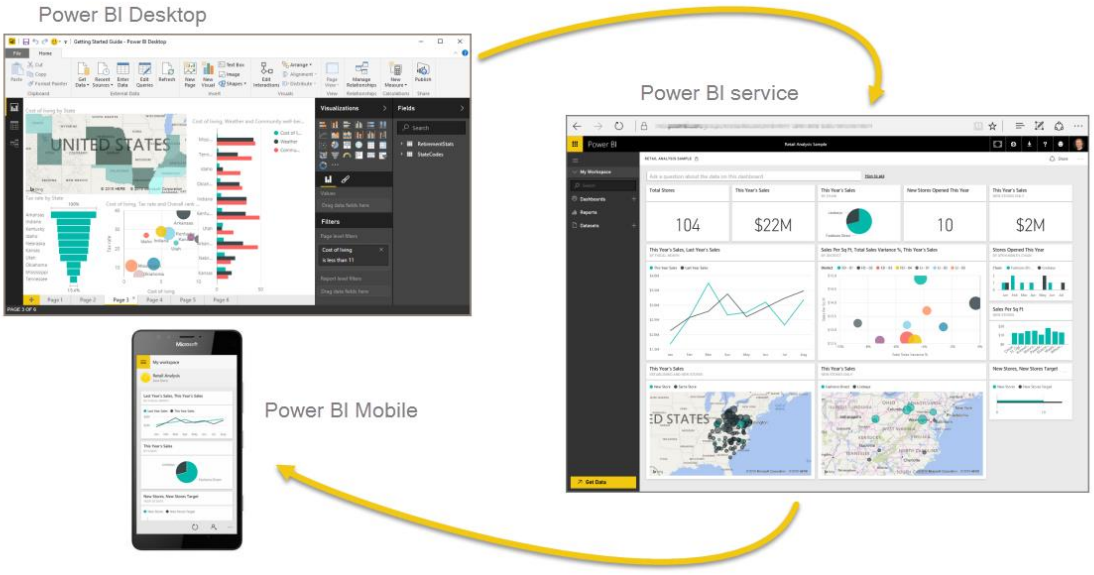
Power BI is a Business Intelligence (BI) tool that lets you visualize your data and share insights across your organization or embed them in your app or website. It has the power to ingest, transform, and visualize data that is stored just about anywhere, either on premises in country (on a Ministry of Health or project server), on a cloud-based server, or in a third-party application. Power BI is not a database and is not intended to be used as a database or data warehouse; it is effective in bringing together and transforming data from multiple sources for access by end users in the form of visualizations.

Power BI allows decision-makers to make data-driven decisions quickly. Country teams can connect, model, and then explore data with visual reports that are collaborative and easy to share. Power BI integrates with other tools, including Microsoft Excel, so you can get up to speed quickly and work seamlessly with existing solutions.

Power BI is a well-supported BI tool with a strong community. It is recommended that users consult the *Power BI Resource Guide* prior to getting started.

3.2. Power BI Basics: What you need to know

Power BI consists of a Windows desktop application called Power BI Desktop, an online service called the Power BI Service, and Power BI Mobile App (available on Windows phones and tablets, as well as for iOS and Android devices) or Power BI Web App.[1]



Power BI Product - Desktop, Service, and Mobile

These elements – the Desktop, Service, Web App and Mobile App – are designed to let you create, share, and visualize data easily, efficiently and rapidly.

A common flow of work in Power BI begins in Power BI Desktop, where a report is created. That report is then published to the Power BI Service, and shared so users of the Power BI Mobile App or Power BI Web App can view the information.

[1] Information contained in this section has been adapted from Microsoft documentation. The full text (along with instructional videos) can be found here.

3.2.1. User Roles

There are three general groups of individuals that will interact with Power BI; administrators, developers and end-users. This guide (and the accompanying Admin guide) refer to individuals and accompanying tasks according to these three categories. Depending on the structure of your organization and/or the number of persons responsible for data visualization, these roles may overlap.

Administrator: Individuals with access to system administration of Power BI. These individuals manage user access, and assign and manage tenant level settings regarding functionality, security and monitoring via the admin portal. This role is typically handled by a member of the organization's IT department or group.

Developers: Individuals with access to the source data who will build a relational Power BI data model and subsequent report. These individuals work on the backend of Power BI, building and/or publishing reports and are typically IT, Informatics, or M&E professionals.

End-users: Individuals that need to use the data represented in the Power BI visuals to make informed decisions. End-users are consumers of the data and interact with Power BI in a web interface. End-users do not need to have an advanced understanding of how to build Power BI reports. Typically, these are surveillance officers or program managers who will be viewing and acting on the data.

3.2.2. Power BI Desktop Key Features

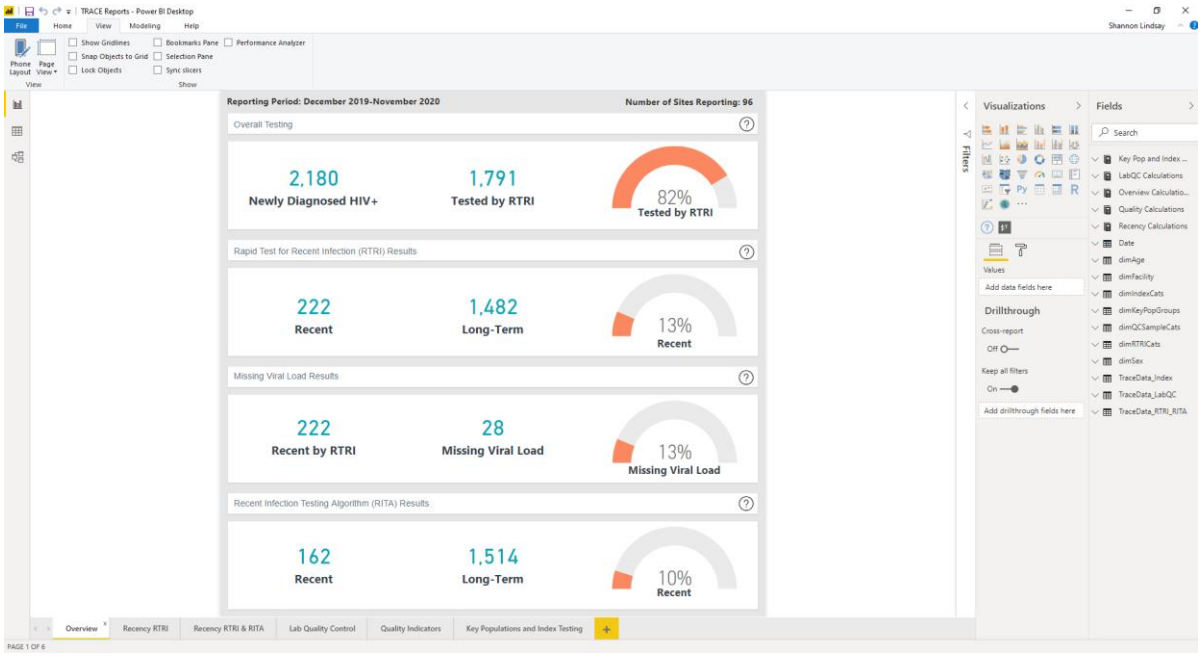
Working with Power BI requires a basic understanding of the key elements used to transform data and create visualizations in the platform. The main building blocks are:

1. Report Editor
    - Report View
        - Filters
    - Data View
    - Model View
2. Power Query Editor
3. Datasets/Data Model
    - Calculated Columns
    - Measures

Report Editor

A Power BI report is a collection of visualizations that appear together on one or more pages. After ingesting and transforming data and defining relationships developers are able to build the report in the report editor. Here, developers can create, edit, format and build reports using

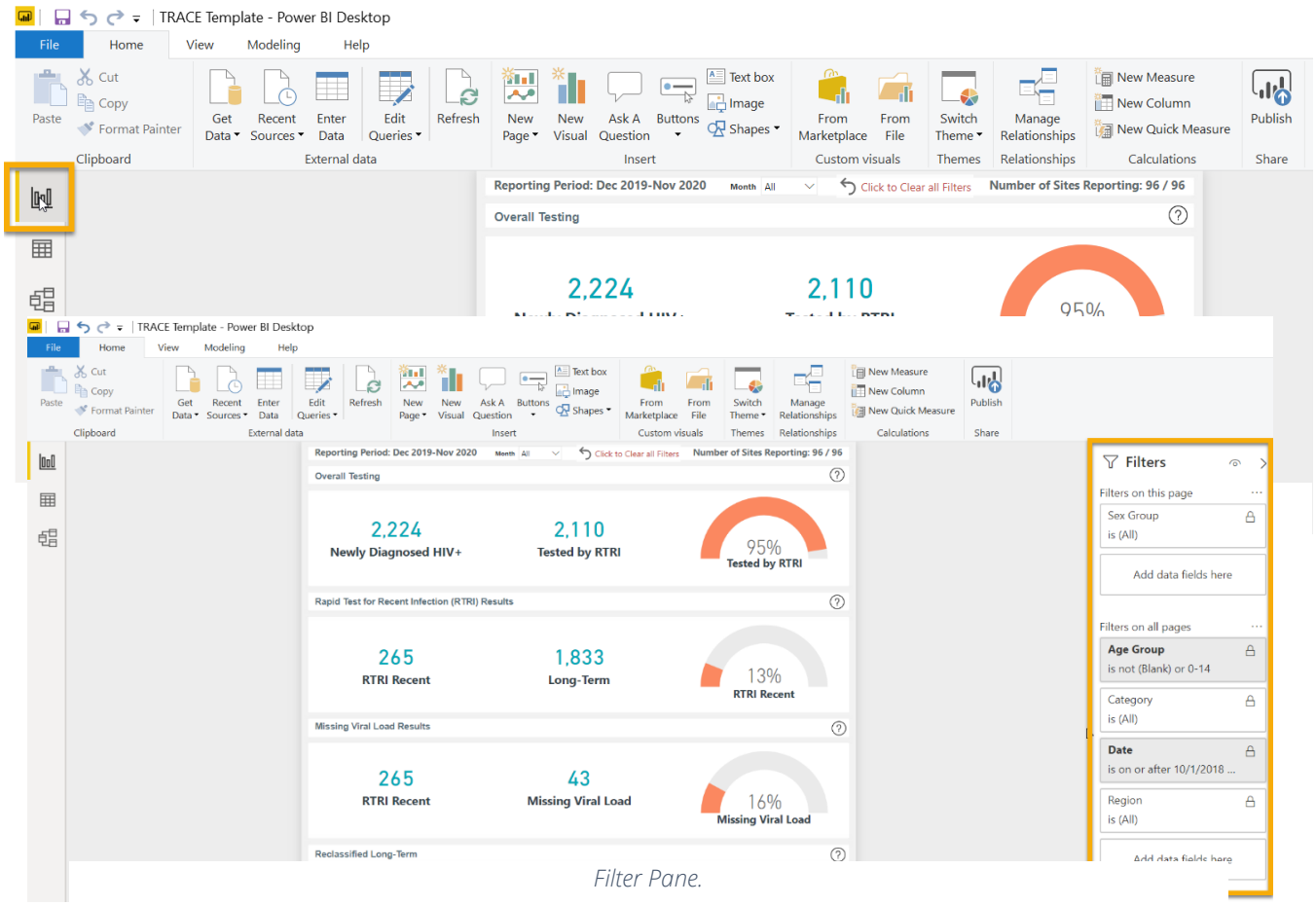
the ‘visualizations’ pane and pulling measures and data from the ‘fields’ pane. (for more information go to [Microsoft Power BI Information: Power View in Power BI Desktop](#)).The TRACE Template is a **report** containing multiple pages, each with a collection of visualizations that have been carefully curated to display key HIV recency data (see *Section 3.3 TRACE Template*).



Sample Power BI Report in the TRACE Template Power BI Desktop Report Editor View with Custom Visuals

Report View

The report editor contains 3 key views, where users can see the **report** canvas (where visuals are created), the **data** view (where users can view each dataset), and the **model** view, where users can visually see each table in their data model and how the tables relate to each other.

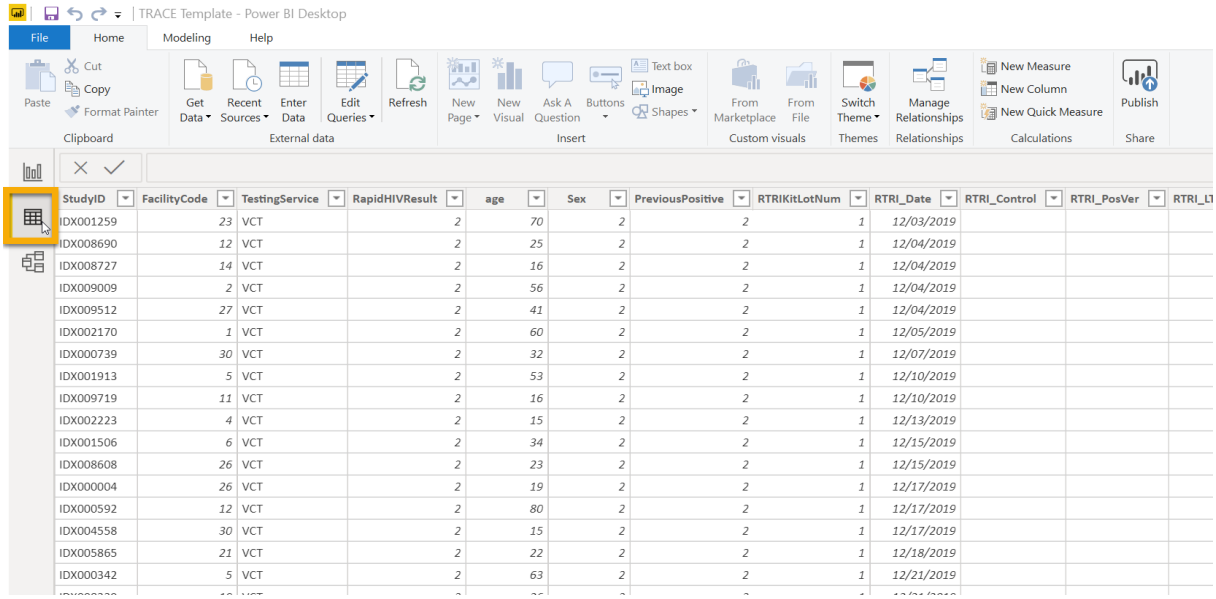


Filter Pane.

Filters: The filter pane enables users to limit the data included in the visuals to only data of interest, excluding any data that users do not wish to view without removing data from the data table. Filters can be applied to individual visuals, pages or the entire report (for more information go to Microsoft Power BI Filter and highlighting in Power BI reports). The TRACE template includes predefined filters to meet data quality standards, excluding records with missing or erroneous data on key variables or data for ineligible clients (see TRACE Template).

Data View

In the data view, users can view, filter, and sort data to explore the row level data. Note that users cannot change data in this view.

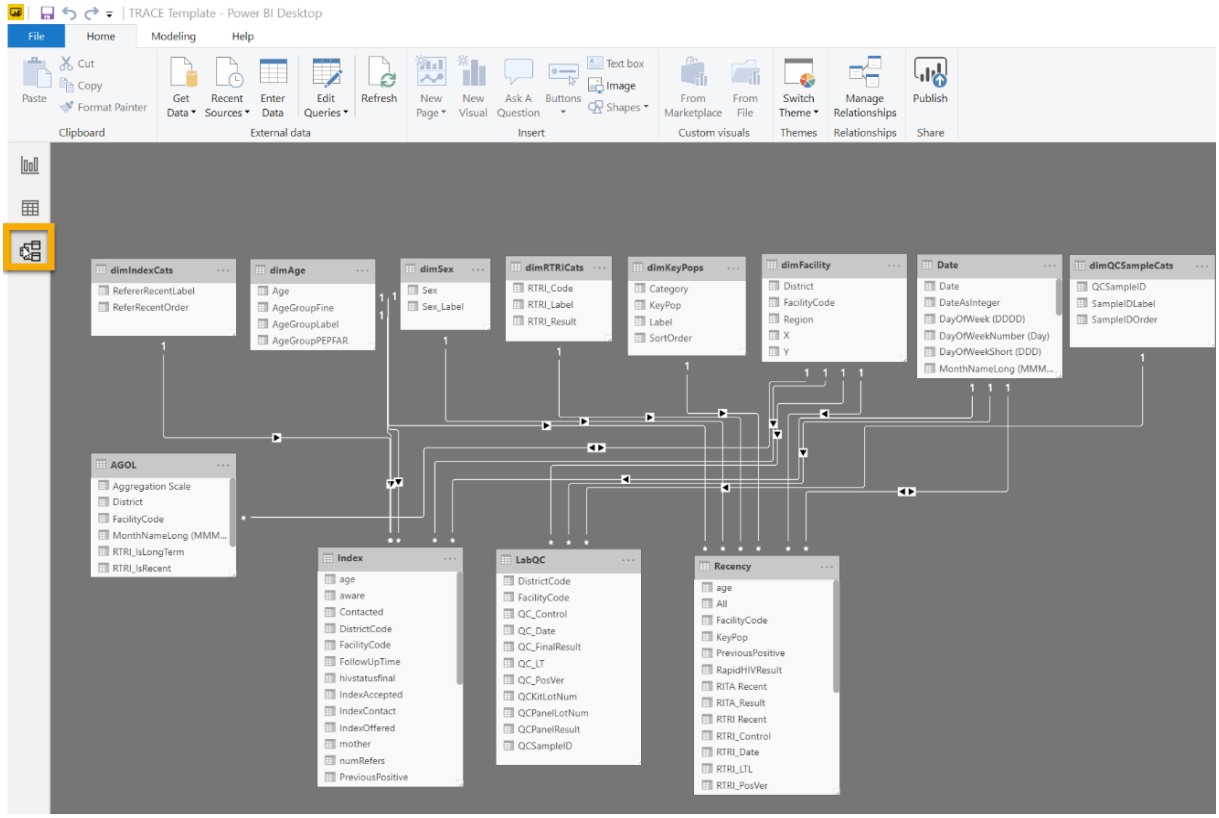


StudyID	FacilityCode	TestingService	RapidHIVResult	age	Sex	PreviousPositive	RTRIKitLotNum	RTRI_Date	RTRI_Control	RTRI_PosVer	RTRI_LT
IDX001259	23	VCT	2	70	2	2	1	12/03/2019			
IDX008690	12	VCT	2	25	2	2	1	12/04/2019			
IDX008727	14	VCT	2	16	2	2	1	12/04/2019			
IDX009009	2	VCT	2	56	2	2	1	12/04/2019			
IDX009512	27	VCT	2	41	2	2	1	12/04/2019			
IDX002170	1	VCT	2	60	2	2	1	12/05/2019			
IDX000739	30	VCT	2	32	2	2	1	12/07/2019			
IDX001913	5	VCT	2	53	2	2	1	12/10/2019			
IDX009719	11	VCT	2	16	2	2	1	12/10/2019			
IDX002223	4	VCT	2	15	2	2	1	12/13/2019			
IDX001506	6	VCT	2	34	2	2	1	12/15/2019			
IDX008608	26	VCT	2	23	2	2	1	12/15/2019			
IDX000004	26	VCT	2	19	2	2	1	12/17/2019			
IDX000592	12	VCT	2	80	2	2	1	12/17/2019			
IDX004558	30	VCT	2	15	2	2	1	12/17/2019			
IDX005865	21	VCT	2	22	2	2	1	12/18/2019			
IDX000342	5	VCT	2	63	2	2	1	12/21/2019			

Data View.

Model View

In the model view, users can see the relational data model, including how each of the tables relate to each other. In the TRACE templates, relationships between the 3 data tables, 1 union table and 8 dimension tables and 3 measure tables have already been predefined for the reports to function (see TRACE Template).



Model View.

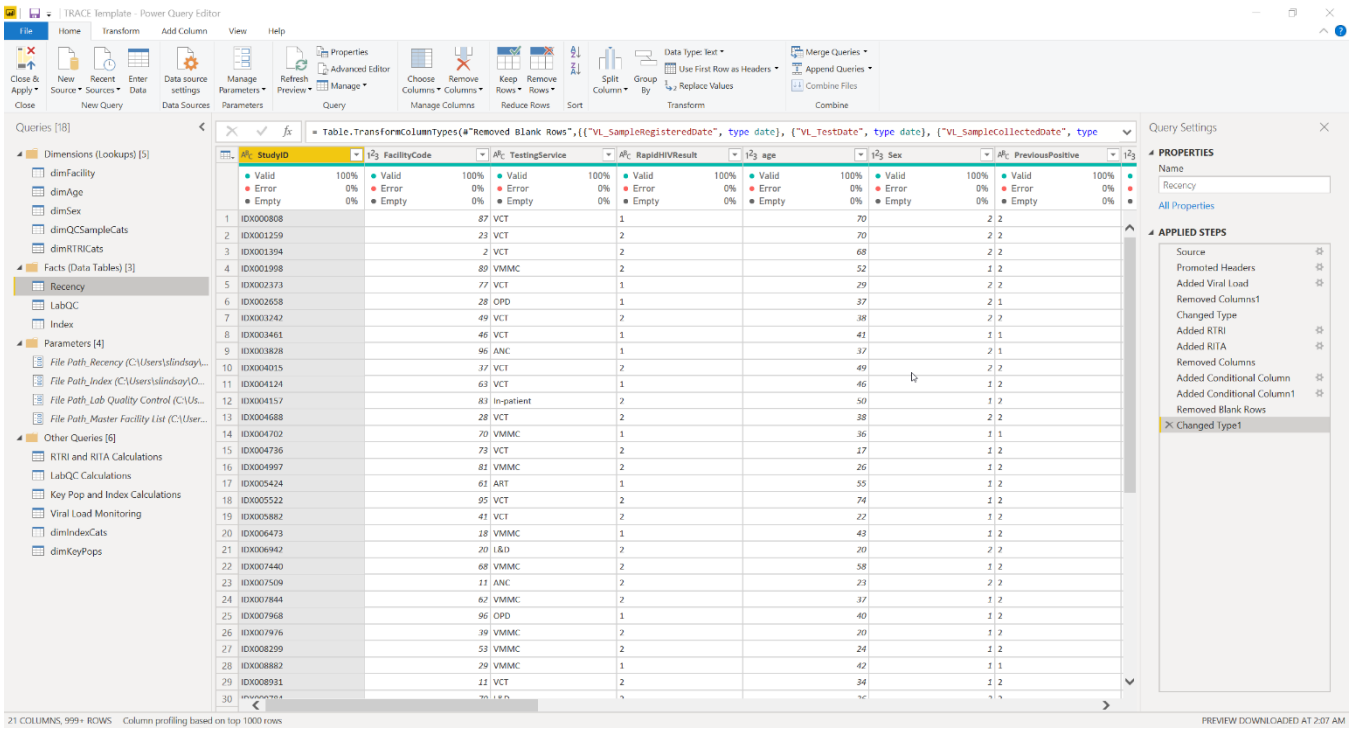
Measures: Measures are the core feature that enable developers to perform analyses on data in data tables in the model. Measures can be built into data tables or created in separate tables in the model (for more information go to [Microsoft Power BI Measures in Power BI Desktop).](#) In the TRACE Template, measure tables have been created to summarize core variables and indicators of interest to track HIV recency programming (see TRACE Template). In the fields pane, a measure is denoted by

Calculated Columns: Using DAX formulae, developers can create new columns and variables in the data tables by creating new values based on row level calculations of the existing data. This allows developers to concatenate, sum, divide or calculate the difference between values in two columns of the data table (for more information go to [Microsoft Power BI Using calculated columns in Power BI)](#). The TRACE template have number of calculated

columns already built in to the tables that allow the visuals in the reports to function (see TRACE Template).

Power Query Editor

After connecting to a data source, the power query editor (henceforth referred to as **query editor**) allows users to transform the data in preparation for creating visuals in the report editor. Here, developers are able to rename, reformat or create new variables in the dataset that are needed for visualizations in the report layer (for more information go to Microsoft Power BI Information Tutorial: Shape and combine data in Power BI Desktop or Query overview in Power BI Desktop). In this space, Power BI keeps a log of data editing steps taken for quality and tracking purposes. The formula language that is being written as users click through different transformation steps is called M (stands for mashup).



Power Query Editor.

There is a feature inside of the query editor called the *advanced editor*. Advanced Power BI Users with high level M proficiency may also edit the data in the advanced editor. It is harder to track edits and may cause irreversible problems with the data while ingesting it into the Power BI Workspace, so we strongly recommend users avoid using this feature.

Datasets

A **dataset** is a collection of data that Power BI uses to create its visualizations. Datasets can be comprised of one single table or can be a combination of different tables ingested into the Power BI Desktop workspace (for more information go to Microsoft Power BI Information

QuickStart: Connect to data in Power BI Desktop). In the TRACE Template, there are 4 ingested *Simulated Data Files* (*RecencyData, LabQCData, IndexTestingData* and *FacilityXYData*) that are connected to the Power BI Workspace, 1 union table (AGOL), 8 dimension tables (dimIndexCats, dimAge, dimSex, dimRTRICats, dimKeyPopGroups, dimFacility, Date and dimQCSampleCats) and 4 measure or calculation tables (LabQC Calculations, RTRI and RITA Calculations, Key Pop and Index Calculations, and Viral Load Monitoring Calculations) that are required to populate the visuals (see *Section 3.3 TRACE Template).* Datasets that have been transformed and loaded to the data model can be viewed in the Data View window (for more information go to Microsoft Power BI Information Data View in Power BI Desktop).

Power BI Service Building Blocks

Workspace

The Power BI Workspace is a place to is an online portal that allows authorized users to create, publish and manage app as well as host published reports for collaboration, sharing feedback and editing (for more information start with Microsoft Power BI Understand the Power BI Service). For Power BI service user roles, set-up and administration of the Power BI Service for the TRACE project and templates see *Admin Guide*.

App (Web and Mobile)

Once reports have been reviewed and approved by collaborators, finalized reports are then published to the app for access by end users. Once published, users can view reports either through password protected login to institutional accounts via the Power BI Web App at powerbi.com or Mobile App.

3.3. TRACE Template

The *TRACE Template* exists in both .pbix and .pbit format and consists of a data model that visualizes sample data in a 7-page report. The data model links 3 datasets with the 8 custom built dimension tables and 1 union table. Within the report, measure tables (RTRI & RITA Calculations, Lab QC Calculations and KP and Index Calculations) have predefined analyses informing the visuals in the report view. The report view includes 6 pages of visuals that display the key indicators for HIV recency broken down by common variables such as location, age and sex.

The *TRACE Template* is:

1. Standardized: key recency program metrics are pre-coded and readily usable by countries.
2. Adaptable: chosen data visualization platforms are flexible and easy to use so that country-specific indicators can be easily added to existing reports.
3. Multipurpose: reports display RTRI and RITA results but also collect and analyze monthly RTRI quality control (QC) results as a way to implement post-market surveillance on the test kits as well as testing quality at the site.
4. Action-Oriented: reports are intuitive and display visuals that health officials at all levels can use to easily focus in on important trends and program gaps.

3.4. TRACE Simulated Data

For the purpose of exploring the templates and getting started quickly there are 4 sample data files included for download with the *TRACE Template*. The *simulated data files* used to build the *TRACE Template* file contain simulated and entirely fictitious data. No specific epidemic profile is presented. In addition, the geographic setting, including the location ("Angola") and the associated geospatial location data (i.e., X and Y coordinates) of health facilities used are also simulated. The template and accompanying datasets are meant to be generic and serve as a starting point for countries to further customize and adapt to country needs.

Three of *simulated data files* mimic data countries are collecting as part of recency testing from testing sites for RTRI (Rapid Test for Recent Infection), RITA (Rapid Infection Testing Algorithm) or Quality Control Testing, and act as the source data for the *TRACE Template* .pbix file. The fourth *simulated data file* contains health facility geocoordinates that are needed in addition to testing data visualize recency testing data on maps within both the inbuilt Power BI maps and create tables of aggregated data needed for more detailed maps within the AGOL platform.

The *TRACE Template* is built to ingest data in .csv sample data files attached for demonstration purposes. However, Power BI is capable of ingesting data from over 100 different sources and with some configuration the *TRACE Templates* can be modified to accommodate other data formats (including, but not limited to .xlsx, SQL, MySQL, PostgreSQL, etc.) depending on the countries source data and data management systems. To aid countries in adapting their own data for ingestion into the *TRACE Template*, please go to *Section 5. Ingest Your Own Recency Data into* Power BI of this guide or for more information, please go to Microsoft Documentation Connect to Data and Connect to data in Power BI Desktop.

For more information on elements and definitions for each simulated data table please see *TRACE Template Codebook*.

3.5. TRACE Report Pages

Included in of each of the pages within the *TRACE Template* file are the minimum quality indicators that TRACE recommends using for above site surveillance of recent HIV infections, based on the HTS_RECENT indicators defined by PEPFAR.

Overview: Provides a snapshot of recency testing metrics overall and filterable by month. This page provides a brief indication of coverage of RTRI testing and viral load (if in use), test kit performance and final recency results for the country overall.

Recency RTRI: Provides a more in-depth view of RTRI results for countries not implementing the full algorithm with filters and charts to display data disaggregated by location, sex, age group, health facility or point-of-testing.

Recency RTRI & RITA: Provides a more in-depth view of RTRI & RITA results for countries implementing the full algorithm with filters and charts to display data disaggregated by location, sex, age group, health facility or point-of-testing.

Quality: Provides an overview of recency testing coverage by sex, month and district. For countries implementing the full RITA this page includes coverage of viral load testing.

Lab Quality Control: Provides an overview of how the test kits are performing at laboratory and testing site settings. It includes panel testing quality control by specimen type, panel lot, kit lot, facility and month for facilities doing QC testing, plus quality of testing site kit performance (invalids and negative results) and test kit sensitivity.

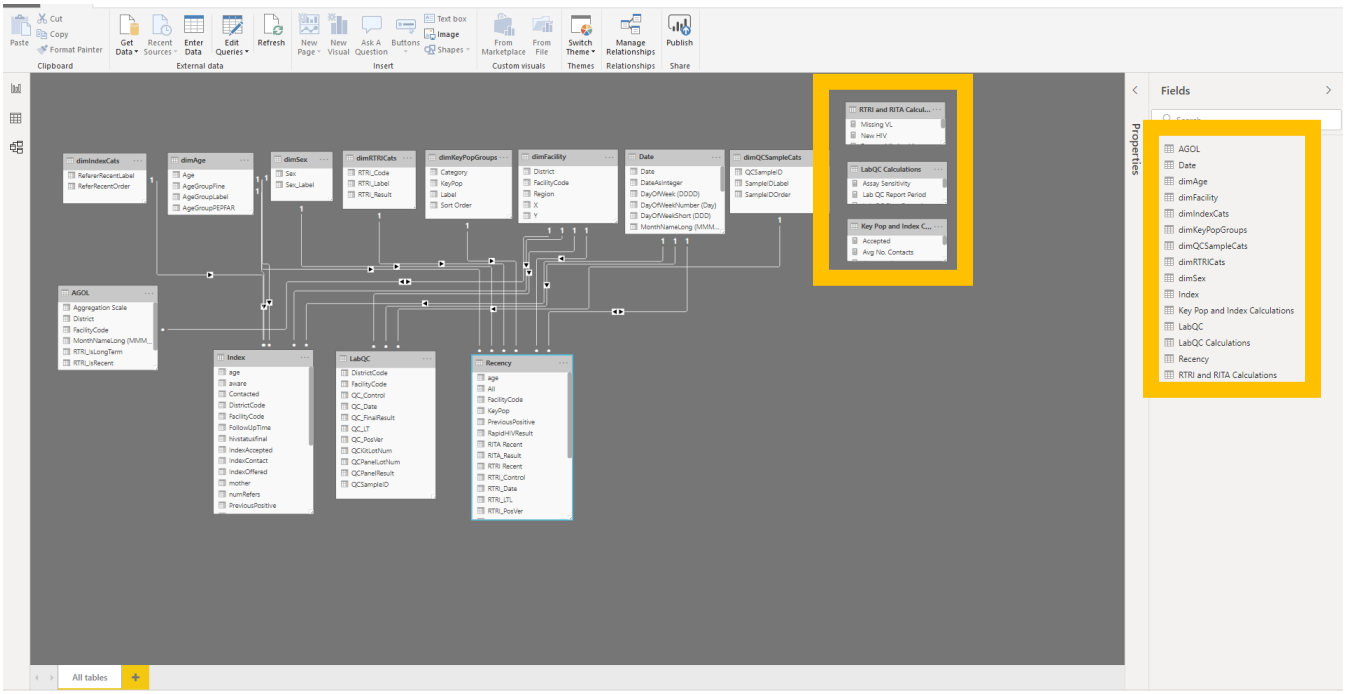
Key Populations and Index Testing: Provides a summary of RTRI testing coverage and results for key populations plus coverage and outcomes of index testing results for countries who are collecting data on key populations and index testing as part of recency testing.

Viral Load Monitoring: Provides more detailed information on coverage of viral load testing from sample collection, through registration at the laboratory, sample analysis and availability of results to identify where blockages may be occurring, as well as tracking specimens that have exceeded 14 day expected turn-around-time.

AGOL: A hidden data tab summarizes RITA long-term and recent results by facility location formatted to ensure successful upload into the AGOL operations dashboard, this page will not be visible to end users.

3.6. TRACE Measures

The ***TRACE Template*** contains several pre-defined measures to visualize key recency indicators based on the HTS_RECENT indicators defined by PEPFAR, plus standardized measures for monitoring implementation and lab quality control indicators. These measures are found in 3 tables in the Model View or Fields menu in the ***TRACE Template***: *Key Pop and Index Calculations*, *Lab QC Calculations*, *RTRI and RITA Calculations, and Viral Load Monitoring Calculations.* Measure names, the corresponding table and DAX expressions can be found in the ***TRACE Template Codebook***.



TRACE Template model view, highlighting measure tables and fields.

The expressions are currently written in reference to the ***simulated data files***. Ingestion, transformation, and visualization of country data through these expressions will require some data transformation steps to be complete first, as described in *Section 5. Ingest Your Own Recency Data into* Power BI.

4. Getting Started

Prior to ingesting one's own data into Power BI and AGOL, the TRACE team recommends a base level of training/guided learning to familiarize oneself with the tools. These resources can be found in the ***Power BI Resource Guide***. Detailed instructions on setting up Power BI and ArcGIS for your institution can be found in the ***Admin Guide***, and should be completed before continuing.

The sections below detail the steps required to use the *TRACE Template* and to then import and visualize *simulated data files*.

Minimum system requirements to run Power BI Desktop are:

- Windows 7 / Windows Server 2008 R2, or later
- .NET 4.5
- Internet Explorer 10 or later
- Memory (RAM): At least 1 GB available, 1.5 GB or more recommended.
- Display: At least 1440x900 or 1600x900 (16:9) recommended. Lower resolutions such as 1024x768 or 1280x800 are not recommended, as certain controls (such as closing the startup screen) display beyond those resolutions.
- Windows Display settings: If your display settings are set to change the size of text, apps, and other items to more than 100%, you may not be able to see certain dialogs that must be closed or responded to in order to proceed using Power BI Desktop. If you encounter this issue, check your Display settings by going to Settings > System > Display in Windows, and use the slider to return display settings to 100%.
- CPU: 1 gigahertz (GHz) or faster x86- or x64-bit processor recommended.

Note that Power BI Desktop does not run on a Mac operating system. For more information go to Microsoft Power BI Documentation: Get Power BI Desktop.

4.1. Ensure you have an active Power BI account

Prior to opening the *TRACE Template* file data managers must first establishing an active account with Microsoft Power BI and complete installation steps found in the *Admin Guide*.

4.2. Install Power BI Desktop

Install the Power BI Desktop as an App from Microsoft Store to ensure you receive monthly software updates automatically. This will require administrative access to your computer. For complete (and alternate) instructions on download and installation of Power BI Desktop, see the *Admin guide*, or go to Microsoft Documentation: Get Power BI Desktop.

4.3. Open the TRACE Template File

The *TRACE Template* file has been provided to you in two file extension types, .pbix and .pbit. We describe here how to open and interact with the templates using both formats.

The TRACE Template.pbit file

The .pbit, or Power BI Template File, guides you through the data import process as soon as you open the file. A .pbit file is a template created by Power BI Desktop, to create reports and

visualizations. It contains queries, visualization settings, data models, reports, and other data added by the user and requires access to *the* ***simulated data files*** tables in order to open.

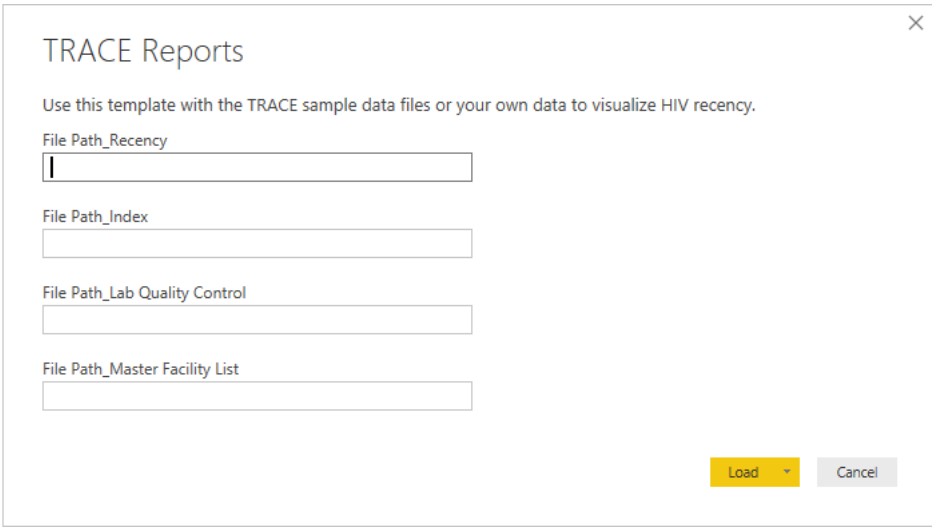
Step 1:

Download the ***TRACE Template.pbit file*** and the sample data files and save locally to your computer

Step 2:

Open the ***TRACE Template.pbit file*** and paste in the file path of the sample data files.

When you open the Template File, a popup will appear to enter the file path of the sample files. This is where you indicate *where* Power BI should look to ingest the sample data.



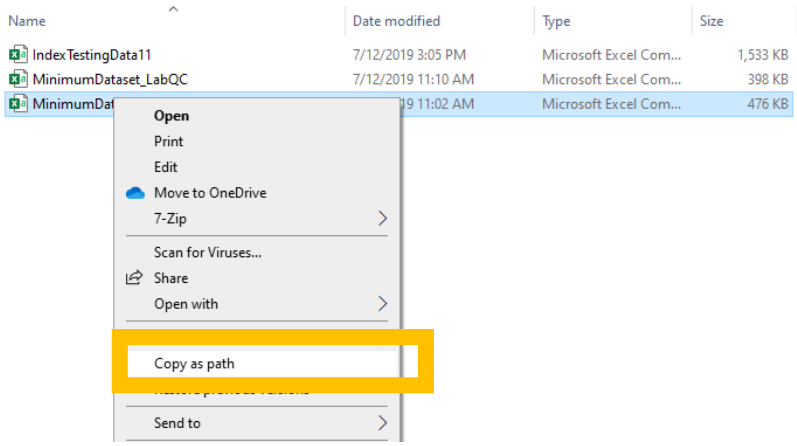
Power BI Template File Popup Window - Enter Source Data Location Here

Step 4:

Enter the local file path for each sample file. For example, if the file is stored on your desktop, the file path would be: "C:\Users\username\Desktop\MinimumDataset.csv". Make sure to delete "" from start and end of the file path.

You can locate the file path by:

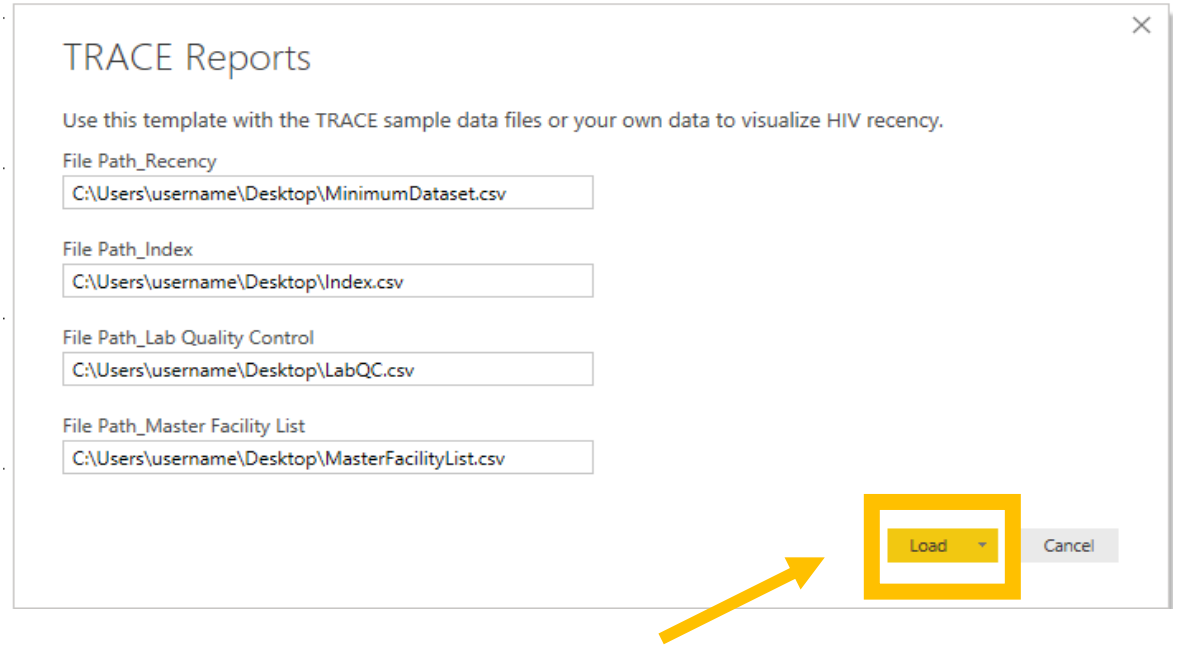
1. Opening a new File Explorer window,
2. Navigating to the file of interest,
3. Hold *Shift* and clicking on the file,
4. Select "Copy as Path" from the menu.



Located the file, hold Shift and right click to reveal this menu. Select Copy as path.

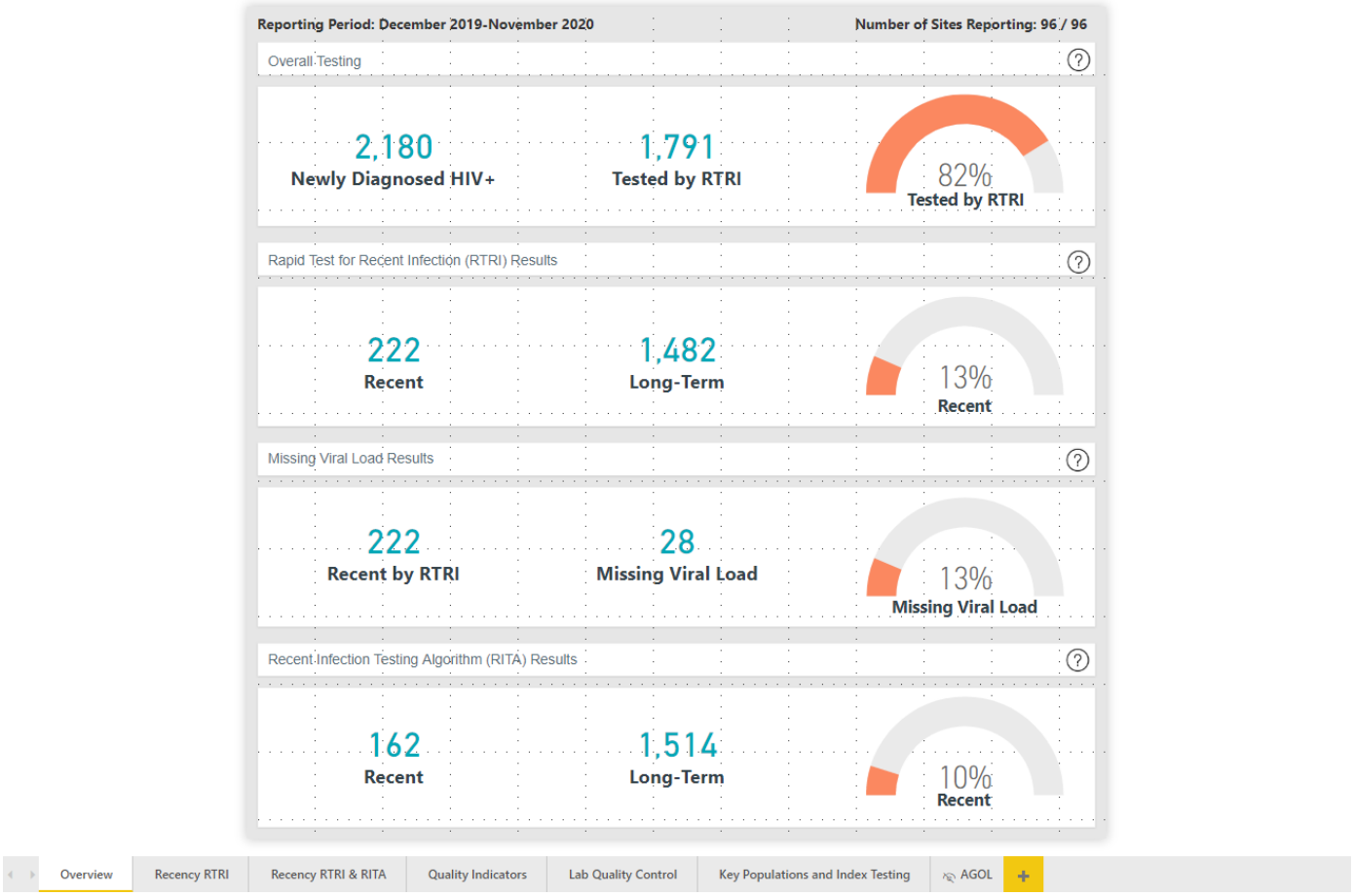
Step 5:

Once all the file names have been updated, click Load and Power BI will attempt to connect to those source files.



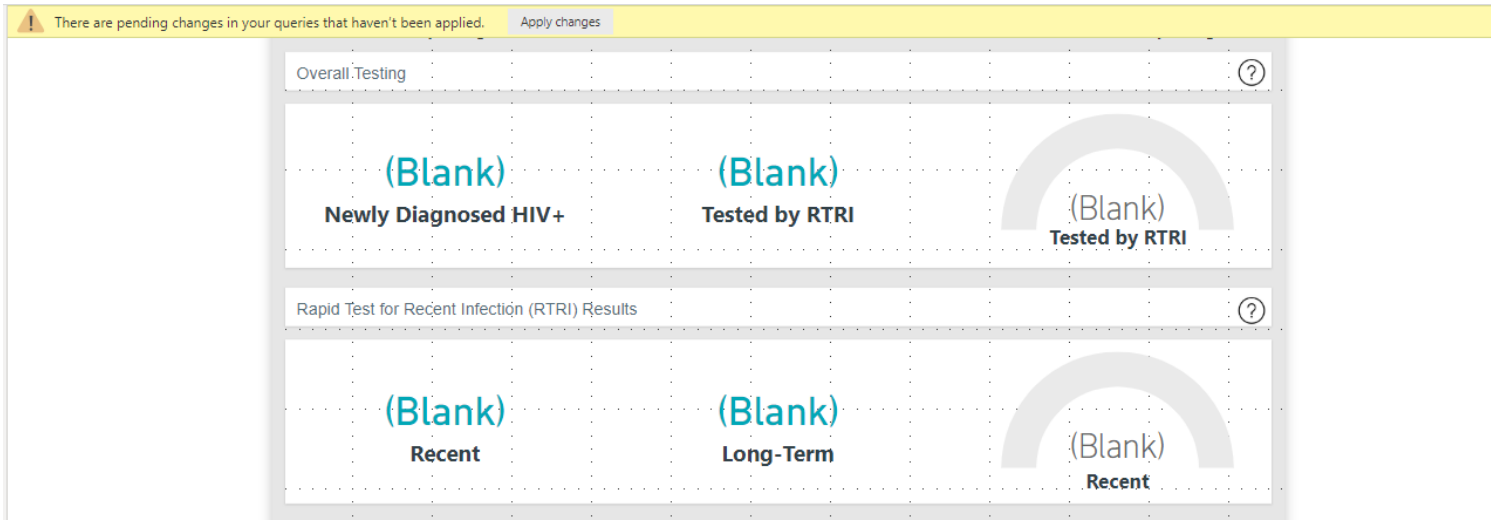
Power BI Template File Popup Window -Enter Source Data Location and Click Load

If the files paths are correct and Power BI successfully connects to the source files, you will see a Power BI report with all the pages and numbers displaying correctly.



Power BI Template File after Successful Data Load

If the files paths are incorrect and/or there is an issue loading the sample data, you will see an error message and a report that appears to contain no data.



Power BI Template File with Unsuccessful Data Load

If you see this error message:

1. Close the file
2. Ensure you know the proper location of the sample files
3. Ensure that the sample files are saved in the proper format (.csv),
4. Start from step 1.

Once you have loaded the ***simulated data file*** into the ***TRACE Template***, save the file to your computer in a location you will remember and you are ready to start exploring the Power BI file.

Note: Steps described here work for the ***simulated data files*** provided with the ***TRACE Template***. In order to link country recency data for use in this template, the data files will need to be structured in the same way as the ***simulated data*** tables, with data in the same format and column names labelled in the same way as the ***simulated data files***. To do this, data files will have to be transformed and data cleaning steps scripted to ensure new data tables align exactly with the ***simulated data*** tables. There are several ways to do this including data transformation in staging databases prior to ingesting data into Power BI Desktop or through transforming using the Query Editor or Advanced Editor, though only advanced M developers should use the Advanced Editor. See *Section 5.2 Transform Your Data* for a starter guide in transforming data for ingestion into the ***TRACE Templates*** using the Query Editor in Power BI Desktop.

The TRACE Template .pbix file

The .pbix, or Power BI Desktop File, contains visualizations that have been created using the ***simulated data files***. This file can be updated to include the ***simulated data files*** as a source, but also allows users to view and modify the report layer without having access to the data files, and connect to the data once within the software through editing the file path parameters. Start with this file if to explore the ***TRACE Template*** report view, model and data view that is disconnected form the source data. Step-by-step instructions can be found in *Section 5. Ingest Your Own Recency Data into Power BI*. For more information on editing parameter setting in Power BI Desktop or Service go to Microsoft Documentation: Edit parameter settings.

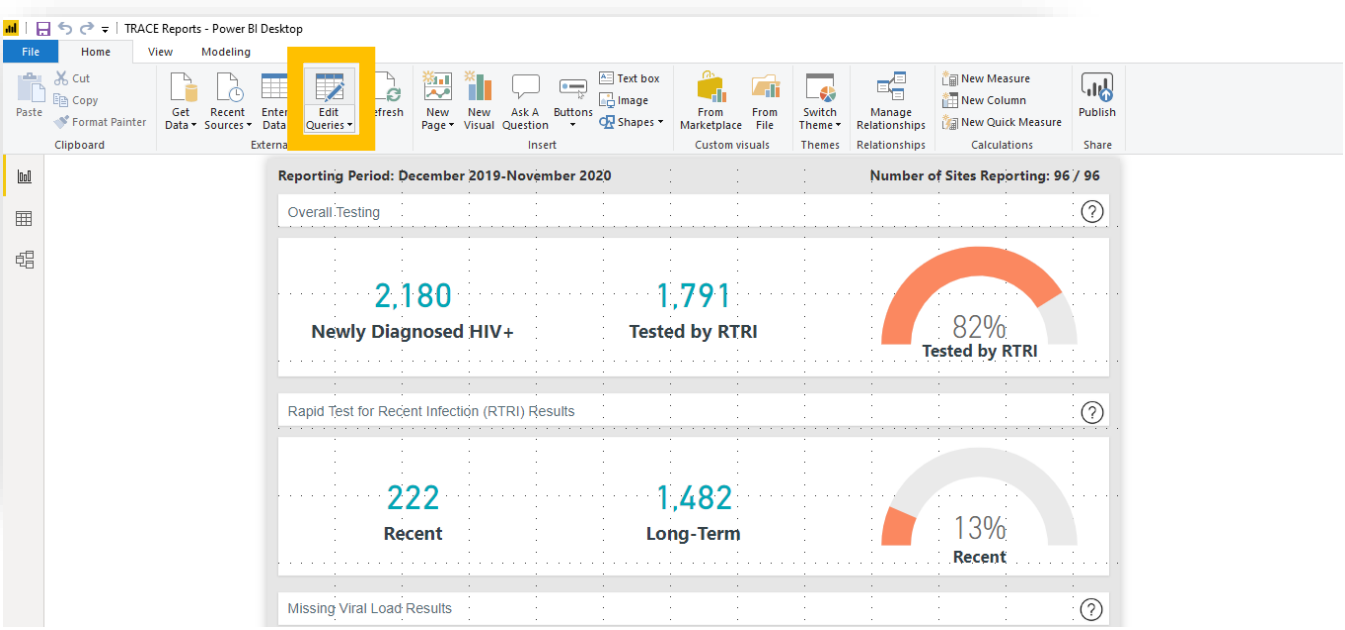
4.4. Explore the TRACE Template File

In order to replace the sample data with your own data in Power BI, it is important that you know your way around Power BI desktop enough to be able to *locate and change the files path, data source and type,* to *view the data* after it is loaded, and to *troubleshoot* if anything goes wrong.

Locate and Change the File Path

Once in Power BI there are two ways to update the file path and data source.

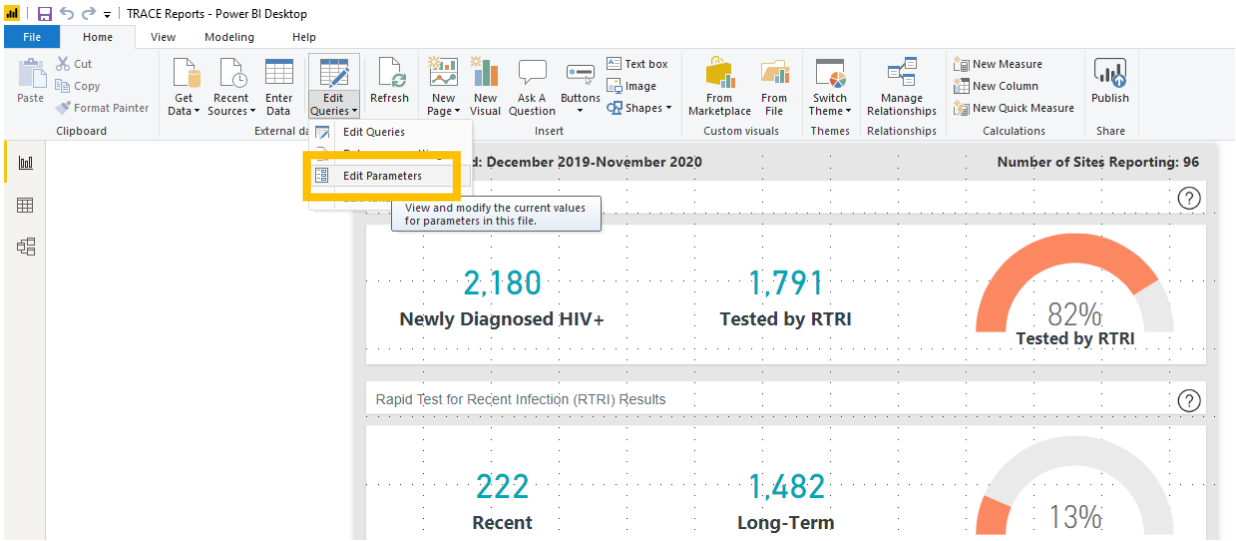
1. Open the *TRACE Template .pbix* file
2. On the home ribbon, click on the down arrow next to the “Edit Queries” button.



Click on "Edit Queries" to Open the Query Editor in Power BI

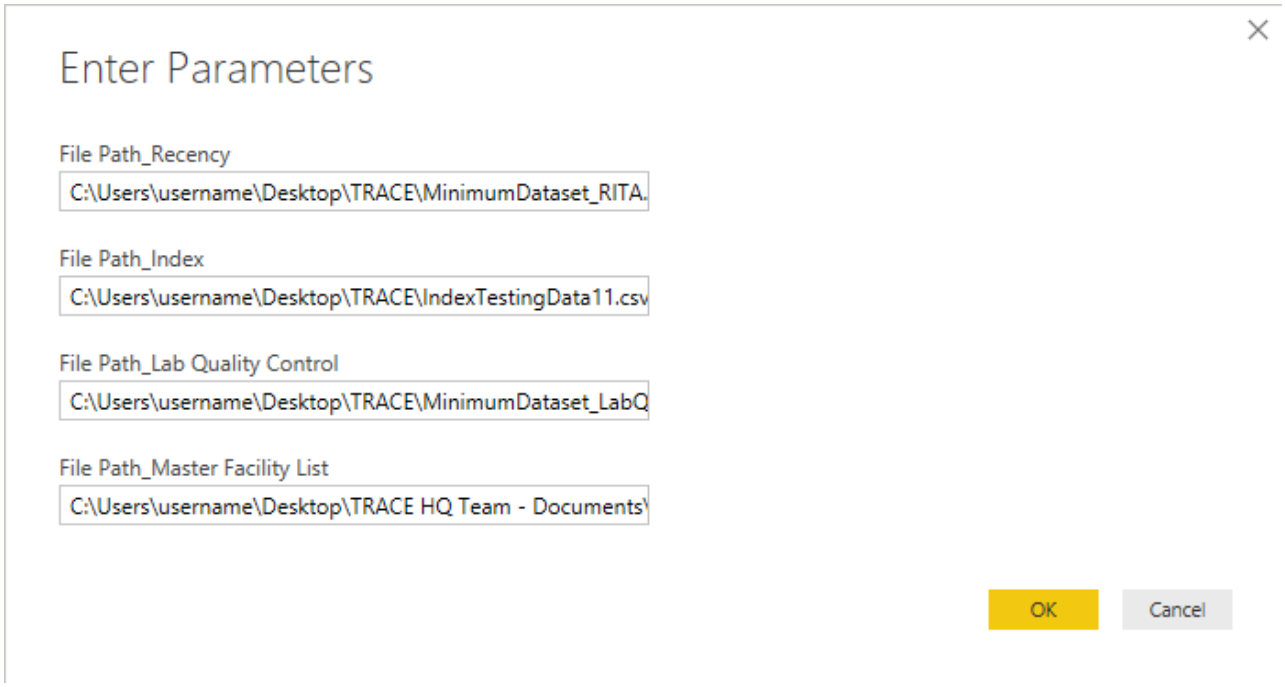
3. After clicking on Edit Queries, a popup window (the Power BI Query Editor) will appear.

4. This will open a dropdown menu. From here, select “Edit Parameters.”



TRACE Template Edit Query and Edit Parameter Function to Change the Data Source

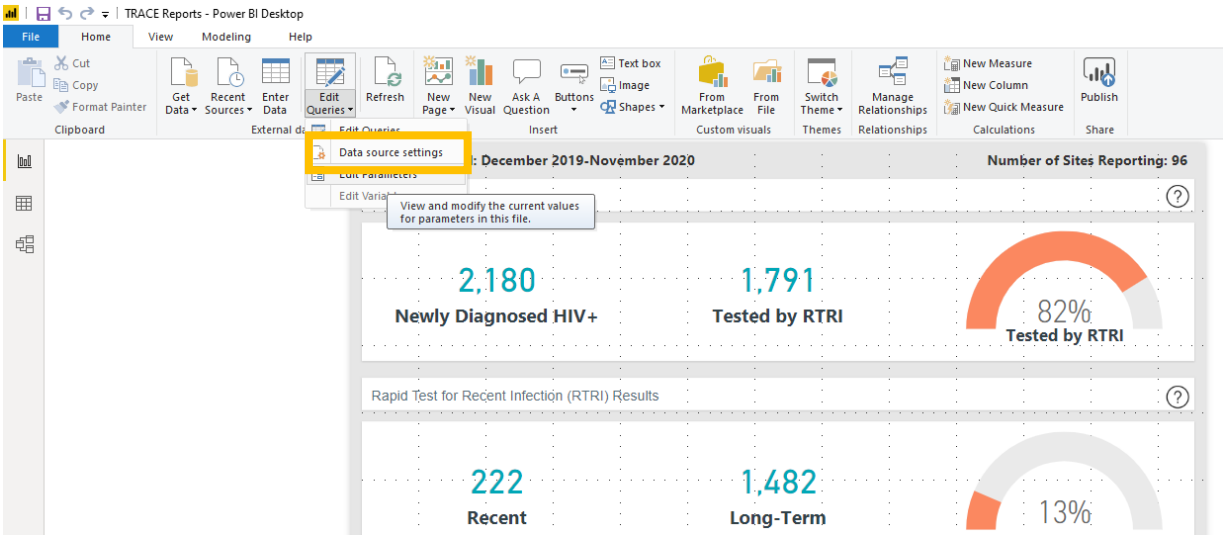
5. Once you have the file path, enter it into the File Path Parameters as you did following the steps outlined in *Section 4.3. Open the TRACE Template File:*



Copy and paste file names into the File Path Parameters fields.

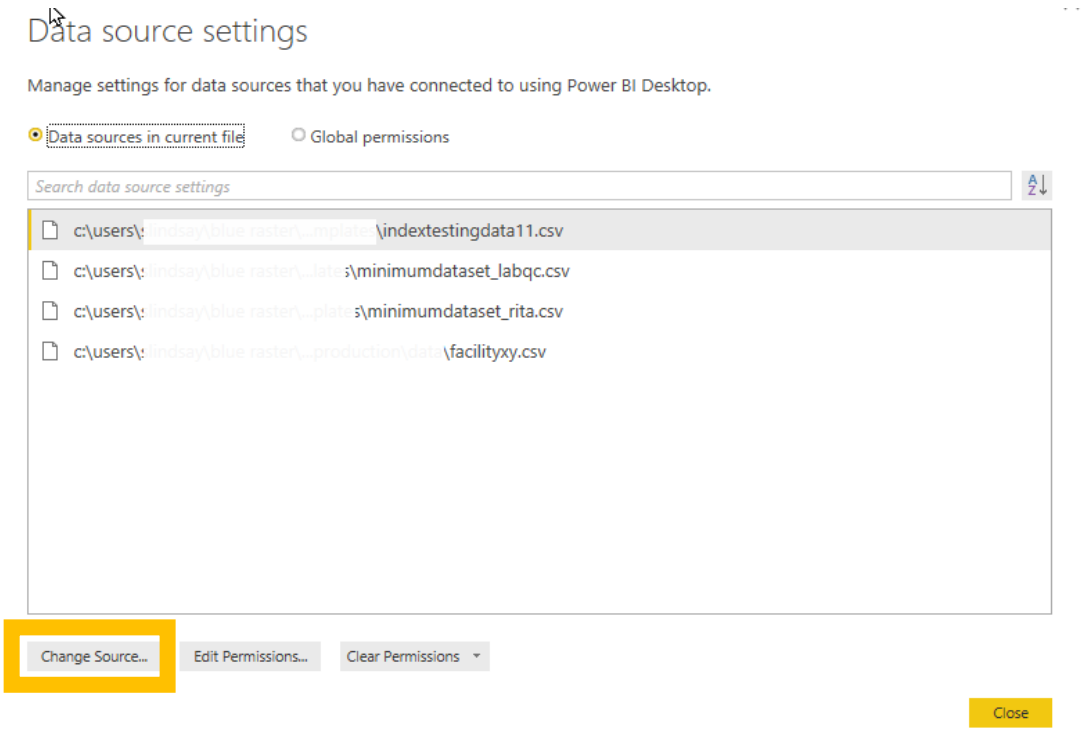
Alternatively, change the file path under Data source settings

1. Select Edit Queries menu on the Home tab of the ribbon.



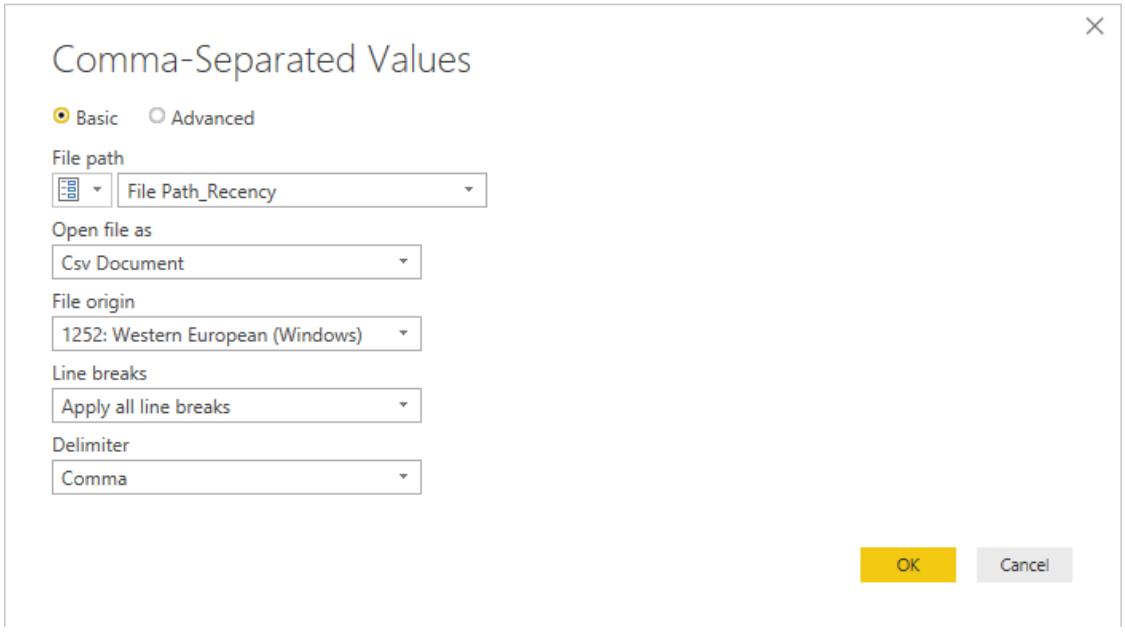
Change source data through Edit Queries in the Home ribbon of report, data or model view.

2. Click on change source and amend the file path that you would like to load into the template.

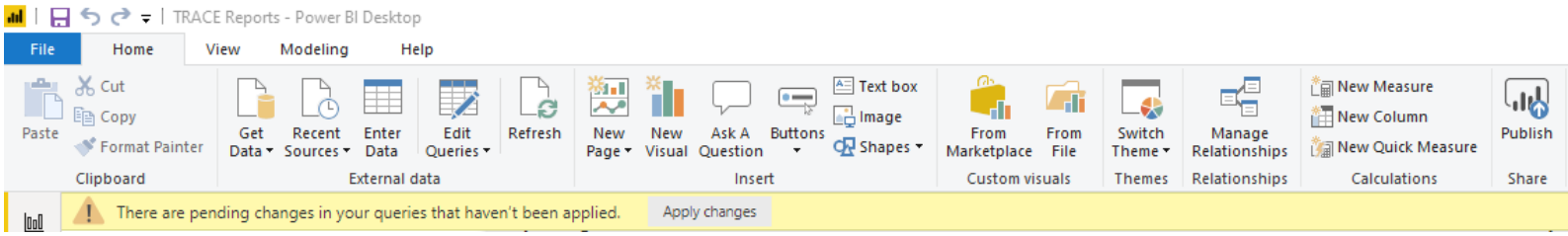


Change the Source of the Data in the Data Source Settings Window

3. Here you can locate and update the file paths, change the data file type (depending on the type of file in the file path selected) and edit other parameters of the source data (this will be important for ingesting your own data later).

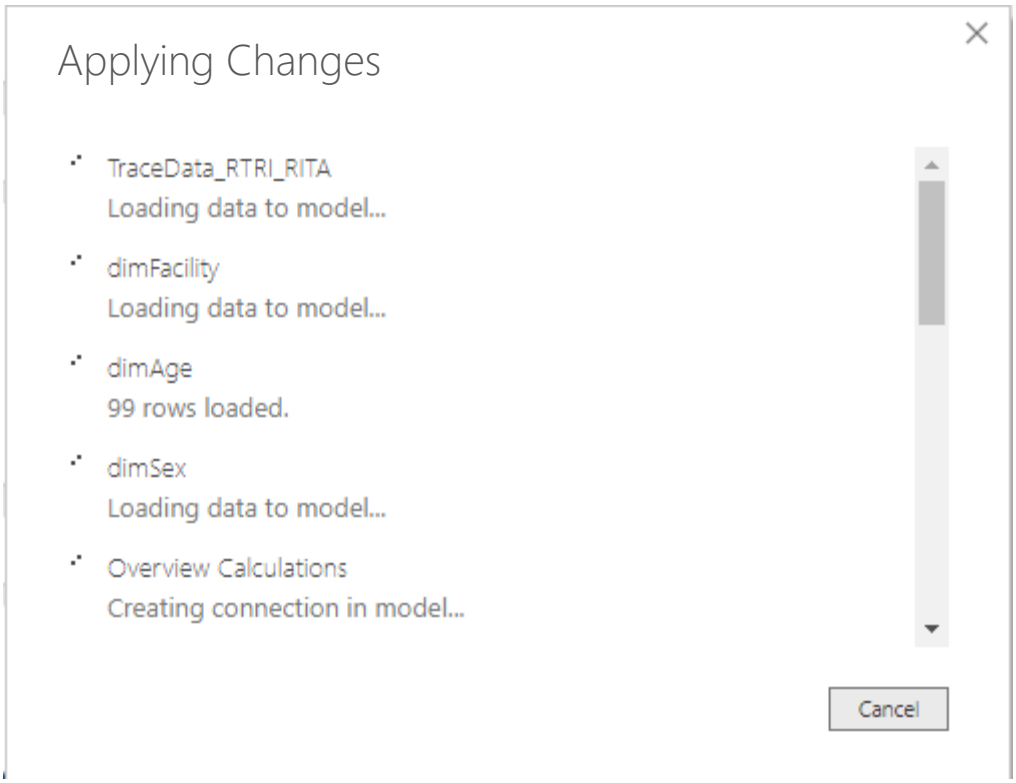


After clicking OK, Power BI will load your data files into the model. While this is happening, you will see a yellow bar across the top of the window. If the yellow bar does not automatically disappear, you can click "Apply changes"



Click "Apply Changes" if this yellow bar doesn't disappear automatically.

You will then see a popup window showing you that the data are being loaded to the model

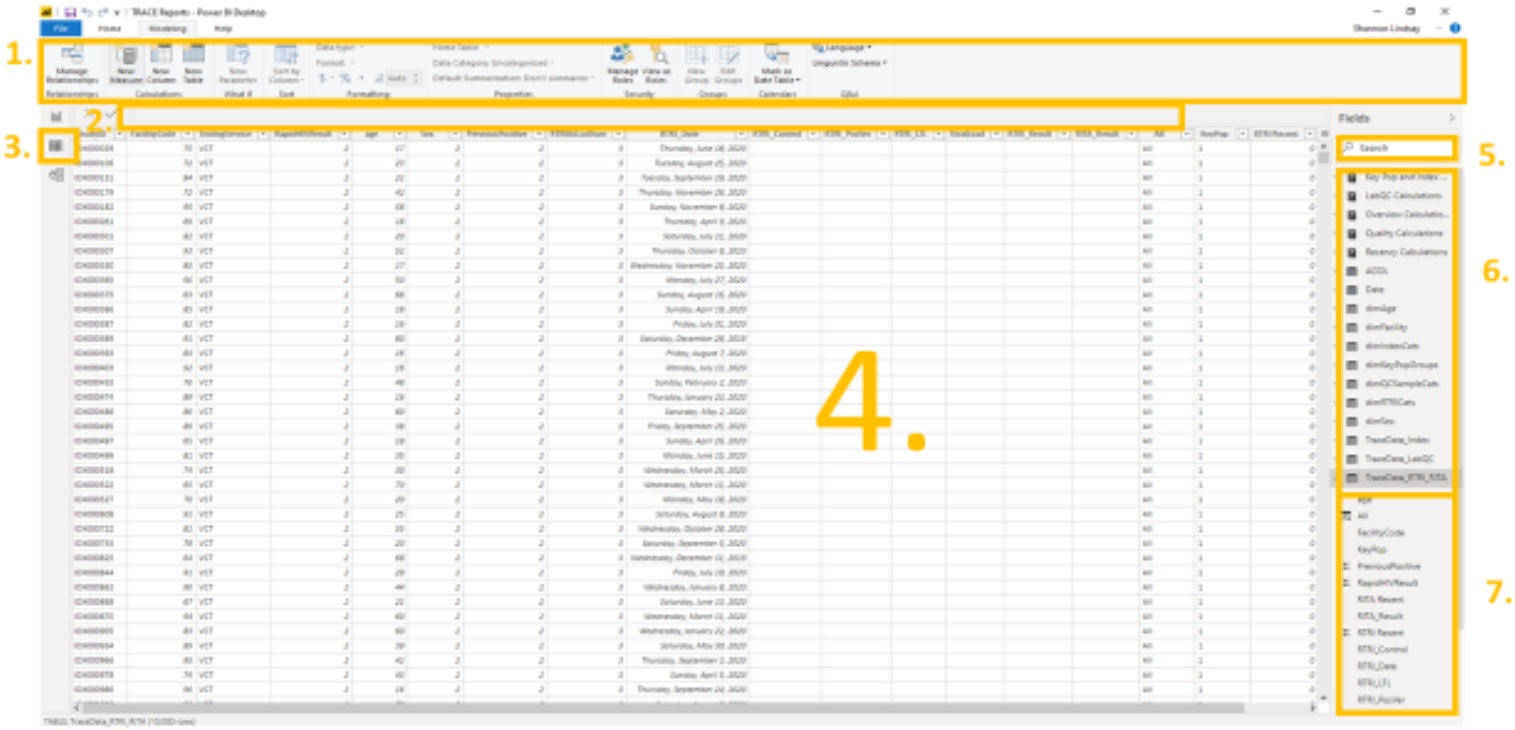


TRACE Template Window while data is being loaded into the model.

Check the visual layer to see if your data is represented in the visualizations.

View Data

The data view in Power BI desktop helps users explore and understand their data without being able to change the data model in any way. When working with your data in Power BI, it may be helpful to see what information is in a row or a column, particularly when creating new calculations.

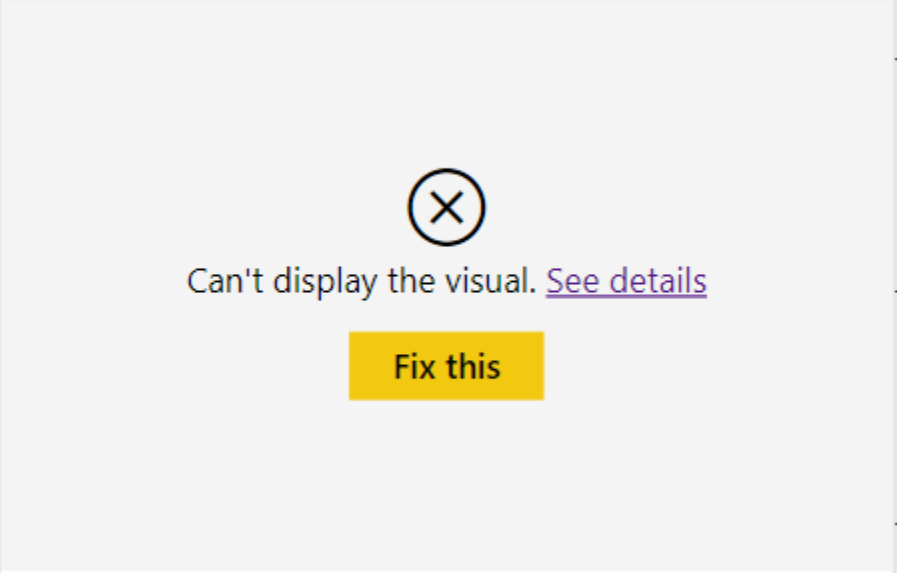


Data View in Power BI

1. Modeling ribbon – Here you can manage relationships, open the formula bar to create calculations, change data type, format, data category for a column.
2. Formula bar – Enter DAX formulas for Measures and Calculated columns.
3. Data View Icon – Select this icon to enter Data View.
4. Data View – This shows the selected table and all columns and rows in it. Columns hidden from Report View are greyed out. You can right-click on a column for options.
5. Search – Search for a table or column in your model.
6. Fields list – Select a table or column to view in the data grid.
7. Expanded Table List – View and select columns or measures you wish to review.

Troubleshoot

If you have opened the *TRACE Template* file and are seeing (Blank) or errors in the visualizations, Power BI is struggling to load the source data. There are a few places where this can be investigated.



Error message when the selected visual cannot display the data feeding in.

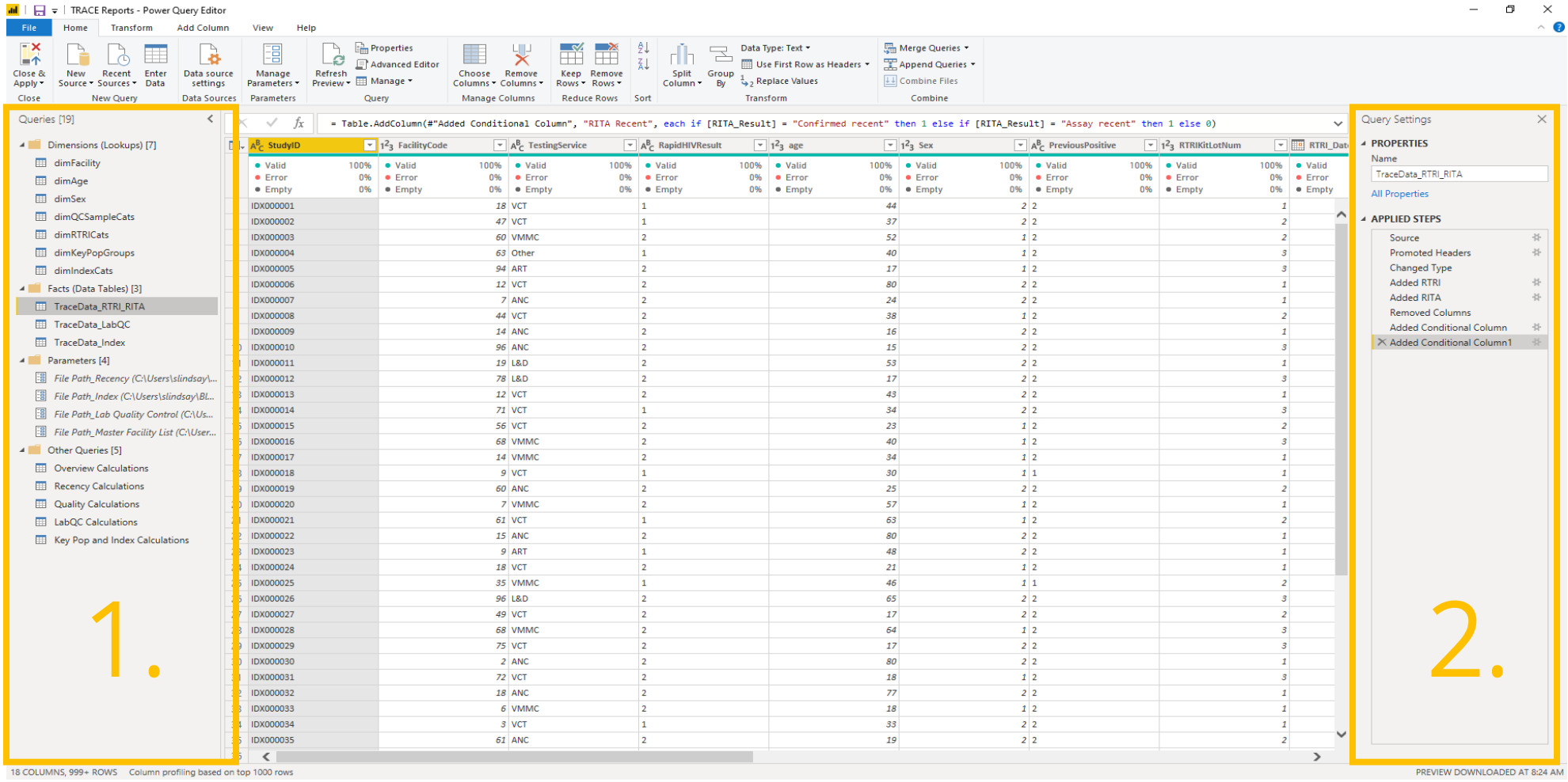
If you follow the instructions above and change the data source settings and still are not seeing any data, the next step will be to enter into the Query Editor to step through the different applied steps in the query and notifications on where the issue might be occurring. Note: this requires an intermediate level of familiarity with Power BI and Power Query. Please consult the *Power BI Resource Guide* for information on free training materials.

Some commonly occurring issues might be:

- The file type in the data source does not match the file type in your folder (eg. .csv and not .xlsx)
- Variable names in the source data are incorrect or misspelt,
- Columns in the source data may be in the wrong order,
- Columns may have been removed from the source data set,
- Response options may not appear in the dimension tables,
- Data may be in the wrong format (i.e. contain text values rather than numerical).

If no solution is found, try downloading the *simulated data files* again and go back to step 1.

Query Editor View See the ***Power BI Resource Guide*** for detailed information about working in the query editor.



Query Editor Window - Use for Troubleshooting

1. The left pane shows the queries that exist in your data model. You can see in this example that the three ***simulated data files*** are located in the "Facts" folder, while the Master Facility list is located in the "Dimensions" folder. (For more info on Facts and Dimensions, see Microsoft Documentation on Dimensional Modeling for Power BI).
2. The Query Settings pane on the right shows a complete list of the steps that have been applied to your data. For the purposes of troubleshooting the data source, you will want to place your cursor over the "source" applied step and read the error message that appears.

5. Ingest Your Own Recency Data into Power BI

Advanced users may be able to live link data from a variety of sources to the *TRACE Templates*, but for the purpose of this guide we will describe linking static .csv files that developers will be required to routinely update with new data. This section goes through the data transformation and connecting steps required to link recency data to the *TRACE Templates*.

5.1. Identify Appropriate Dataset

After looking through the data dictionary, data managers will be able to identify which data sources are necessary to import into the TRACE Template.

Data Managers will note that there are seven pages available in the Power BI template file:

1. Overview
2. Recency Results – RTRI
3. Recency Results – RTRI and RITA
4. Quality Indicators
5. Lab Quality Control
6. Viral load monitoring
7. Key Populations and Index Testing

These pages are displaying data derived from three distinct data sets, in addition to a Master Facility List. The data sets necessary to have on hand for ingestion into the Power BI template file are:

TRACE Template Table	Requirements from Source Data
RecencyData	- Merged dataset of patient level data from recency testing sites and laboratory information systems (LIS) with VL data (if applicable) - Minimum recommended variables required can be found in the *TRACE Template Codebook* - Source data cleaned and transformed
LabQCData	- Lab QC data from health facilities and laboratories performing routine quality control (QC) testing of recent rapid test kits. - Row level data for every QC test performed. - Minimum recommended variables required can be found in the *TRACE Template Codebook* - Source data cleaned and transformed
IndexTestingData	- Merged dataset from sites capturing index testing data as part of recency testing - Row level data for every index client listed for contact tracing.

	- Minimum recommended variables required can be found in the *TRACE Template Codebook* - Source data cleaned and transformed

The list of variables in the *TRACE Template Codebook* is the minimum set of variables needed to enable the visualizations in the template to be generated. Additional variables of interest, such as key population indicators or sexual behavior information, can be added in Power BI, but to make data ingestion easier start by including only the minimum number of variables required to fill the variables in the *TRACE Template Codebook*.

Master Facility List (containing facility GIS coordinates)

The map visuals within Power BI and ArcGIS Online require a reference to the Latitude (Y-coordinate) and Longitude (X-coordinate) of each health facility in a Master Facility List, linked to the health facility name or code, exists in the patient-level data (*RecencyData, LabQCData* or *IndexTestingData*) to enable the facility coordinates to be linked to test results. This will ensure that data can be plotted on maps in both Power BI and AGOL. Ideally, a Master Facility List should contain geographic hierarchies (e.g. a region, district, and facility), a facility ID (to assist with mapping data from multiple sources), and facility codes. Often, facilities have slightly different names in different systems, for example, "Site A" in DHIS2 vs "SiteA" in a patient level database, either the use of a facility code will avoid issues with different spelling, or all possible spellings of the names should appear in the Master Facility List.

The Master Facility List is essential for populating the maps in the *TRACE template* file. If you do not have geocoordinates for health facilities the map visuals in Power BI will not work until they can be added when they become available.

When entering the longitude and latitude, ensure that those are entered under a column labelled as X and Y respectively and that they format is decimal numbers. For the Facility Code and District Code, the two are whole numbers. Please refer to the below table:

Field Name	Data Type	Notes
FacilityCode	Whole Number	Maps to the Health Facility Name or Code in the *RecencyData, LabQCData* and *IndexTestingData*.
District	Whole Number	Maps to the District Name or Code in the *RecencyData, LabQCData* and *IndexTestingData*.
X (Longitude)	Decimal Number	
Y (Latitude)	Decimal Number	

5.2. Transform Your Data

Of critical importance to using this template file is ensuring that your source data file matches the structure and layout of the ***simulated data files***. This section details the necessary structure of your data, steps for data element mapping, identification of appropriate source data, renaming variables, and managing data types.

De-identification and De-duplication

All data cleaning steps to de-identify and de-duplicate the data should occur prior to ingesting into Power BI Desktop to ensure security, privacy and integrity of the data being displayed.

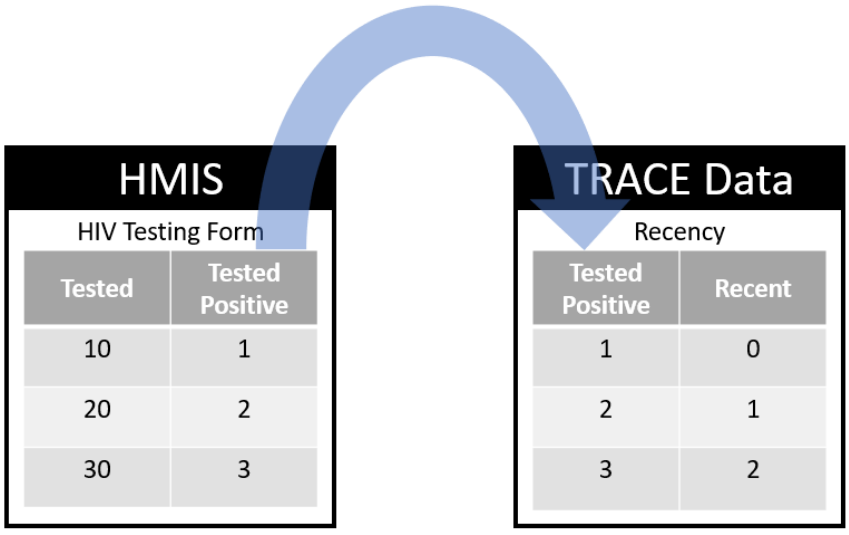
TRACE Template Codebook

The *TRACE Template Codebook* (available for download with the *TRACE Template* and ***simulated data files***) serves as a guide to map data from local source files to the appropriate format to feed into the *TRACE Template*.

The *TRACE Template Codebook* clarifies data types, variable names, DAX calculations included in the *TRACE Template*, and provides data managers the final variables, formats and rules to guide transformation steps needed to adapt the *TRACE Template* from simulated data to real recency data.

Map your data to the TRACE simulated data

Data mapping is the process of mapping data from one information system to another. For example, if a source HMIS has HIV testing data in column 2 of an HIV testing form, we would need to note that data maps to column 1 of the ***simulated dataset***.



Example of Data Mapping from HMIS to TRACE Data

Of critical importance to mapping local data to the TRACE Templates will be formatting source data to match the ***simulated data files*** that feed the ***TRACE Template***. What this means is that the structure of local source data must match the structure of the sample data files exactly. This means:

1. Column headers must be identical
2. Column order must be identical
3. The file type must match exactly (.csv, or .xlsx)
4. Variable types must match (numerical, text and date variables but all match exactly to the TRACE Template Codebook for the corresponding variable).
5. Value coding must also match. For example, if sex is coded as a character variable with values of "M"/"F" you will need to generate a numeric variable with values of 1/2 for the template to use this variable correctly.
6. Response options must match (eg. Sex responses converted to either 1, 2, 3 or 99 for Male, Female, Unspecified or Missing respectively).

The ***TRACE Template Codebook*** contains the required data structure, format and type for the minimum variables used in the ***TRACE Templates.***

There are many ways to map your source data to the sample data files, including Power Query in Excel, the V Lookup function in Excel, the Index/Match function in Excel, or the new X Lookup function in Excel. This step of shaping and transforming your data is labor intensive and is critical to the success of importing your own data into the TRACE Template.

Save you data file as .csv.

5.3. Connecting Your Data Files to the Power BI TRACE Templates

Once the data has been transformed and matched the ***simulated data files*** it is ready for ingesting into the ***TRACE Templates***. Save the files in a reliable location and update the file parameters following the steps outlined in *Section 4.3 Open the TRACE Template File.*, but linking to your own transformed data.

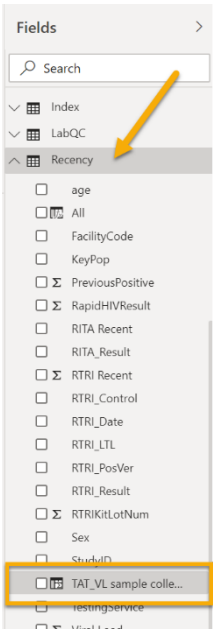
You can select different date types (Access, csv, excel, JSON, text, HTML or XML tables) based on what your source data is, but this will take additional configuration that is beyond the scope of this guide, so for now we recommend saving data as .csv files.

Date Customization within Power BI

In the Viral Load Monitoring page, the Overdue results are calculated based on the different between 2 dates. This should be the date a sample was registered at the laboratory, and today's

date for an accurate measure. As the simulated data occurs in the future date for "today" has been set to November 26, 2020. This date must be changed with the following steps:

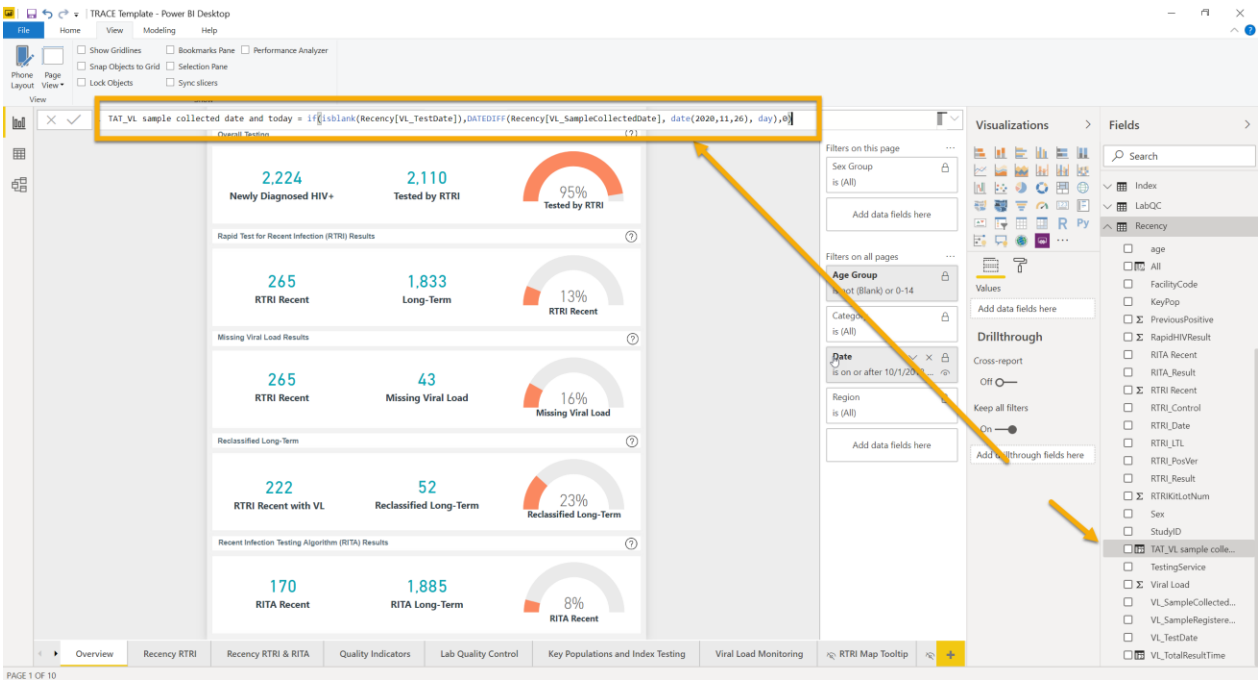
1. In either the Report, Data or Model view, located the Recency Table in the Field menu
2. Expand the Recency table to expose the columns of interest



Expand the Recency table to locate the calculated column

3. Click in the center of the calculated column labeled TAT_VL sample collected date and today.

4. The DAX Formula will appear under the top ribbon



Click in the center of the field labeled "TAT_VL sample collected date and today"

5. The Calculation in the column reads as follows:

TAT_VL sample collected date and today =

IF (

ISBLANK (Recency[VL_TestDate]),

DATEDIFF (Recency[VL_SampleCollectedDate], DATE (2020, 11, 26), DAY),

0

)

6. Replace the entire section by copy/pasting the below formula.

TAT_VL sample collected date and today =

IF (

ISBLANK (Recency[VL_TestDate]),

DATEDIFF (Recency[VL_SampleCollectedDate], TODAY( ), DAY),

0

)

The field should appear as follows:

```
1 TAT_VL sample collected date and today =
2 IF (
3     ISBLANK ( Recency[VL_TestDate] ),
4     DATEDIFF ( Recency[VL_SampleCollectedDate], TODAY(), DAY ),
5     0
6 )
```

***for more information about the DAX language and DAX functions, refer to the Microsoft DAX Function Reference Guide

You should now be able to see your own data in the TRACE template.

What do I do if I cannot see my data?

If you cannot see your data, it means that there has been an issue in mapping your source data to the sample data sets. Take a step back and compare these two.

1) File locations must be exact
    - Check your parameter name for: extra spaces, the presence of unneccessary quotation marks, any slashes that are facing the wrong direction (forward slash vs. back slash)
2) File type should be the same
3) Column header names must match
    - Check each of your source files against both the code book and the simulated data to ensure:
        - Column headers names have no leading or trailing spaces
        - Files contain the same number of columns
        - Data contained in each column matches what the source file expects (for example, if Power BI is expecting an integer and you have text in a column, you will get an error)
4) If you are able to see some but not all of your data, for example, visuals on the Lab Quality Control tab are loading but visuals on the Key Populations and Index Testing tab are not:
    - Open the query editor to see which of your queries is failing. That will help you identify which dataset may have a problem.
    - If you determine that the dataset with a problem is correct, meaning that there are no issues with the structure of the source data, you will need to troubleshoot by stepping through the applied steps in the query editor.

- Troubleshooting in Power Query can be very simple, but it can also become very complex very quickly. Please refer to the ***Power BI Resource Guide*** for more information.

6. Data Management

Power BI can be configured to automatically refresh (or pull in the latest data) at a frequency that you specify. In the short term, refreshes will most likely be on a weekly/monthly/quarterly basis and will involve some manual processes. The long-term goal of a Business Intelligence system is that ingestion, manipulation, and reporting of data is completed automatically. Data management will remain standard regardless of licensing.

6.1. Short-term: manual data refresh

- Developers will manually refresh Power BI reports on an as-needed (or TRACE-specified) basis

- What this will likely look like is:
    - Data are collected in established data collection systems (ODK, DHIS2, RedCap, etc.)
    - Data are pushed to a central server
    - If there is any de-identification or deduplication of data that needs to happen prior to bringing data into Power BI, this is the step where that needs to happen, in the source database
    - Data are extracted from central server in the form of csv or Excel files and dropped into a local folder
    - Power BI will pull data from that folder. An in-country data manager will ensure that files are where they are supposed to be, and at the correct frequency, and will then open a .pbix file and manually hit the "refresh" and "publish" buttons

6.2. Longer-term: automation

- In order to use the *TRACE Templates*, source data must be formatted exactly as the *simulated data files*. This is going to be a starting point for visualizing local data using Power BI. Once data managers/developers are more familiar with the use of Power BI, they will be able to change the data source in the back end of the *TRACE Template*. For example, if you are using a custom mobile application to collect recency data and data resides on a local server, developers will be able to ingest data into Power BI directly from that server rather than feeding in static .csv files.
- Developers will set up automated refreshes in Power BI, meaning that there are no manual processes that need to occur in between the source data and the Power BI report
    - Source data can be refreshed in Power BI up to 8 times daily with Pro licensing, and 48 times daily with Premium licensing. Data refreshes are set up in the Power BI service on the dataset of interest within a workspace. For more information refer to the Power BI documentation on data refresh.
- What this will look like is:
    - Data are collected in established data collection systems (ODK, DHIS2, RedCap, etc.)
    - Data are pushed to a central server
    - If there is any de-identification or deduplication of data that needs to happen prior to bringing data into Power BI, this is the step where that needs to happen, in the source or intermediary database

- The Power BI service will be connected directly to the source database, and will pull data in on a specified routine basis
- Power BI reports will refresh in the service and the system will be entirely hands off

6.3. Data Cleaning

Power BI is also capable of applying a sequenced workflow to your data, meaning that data cleaning steps can be applied to a dataset, and will then be repeated when data are refreshed and new data is pulled in. This method of data cleaning can be used to eliminate, ignore, or systematically correct erroneous records in the Query Editor. Your Power BI report can also be designed to highlight or flag erroneous data, which would alert teams to go back to the source of the data and make corrections there.

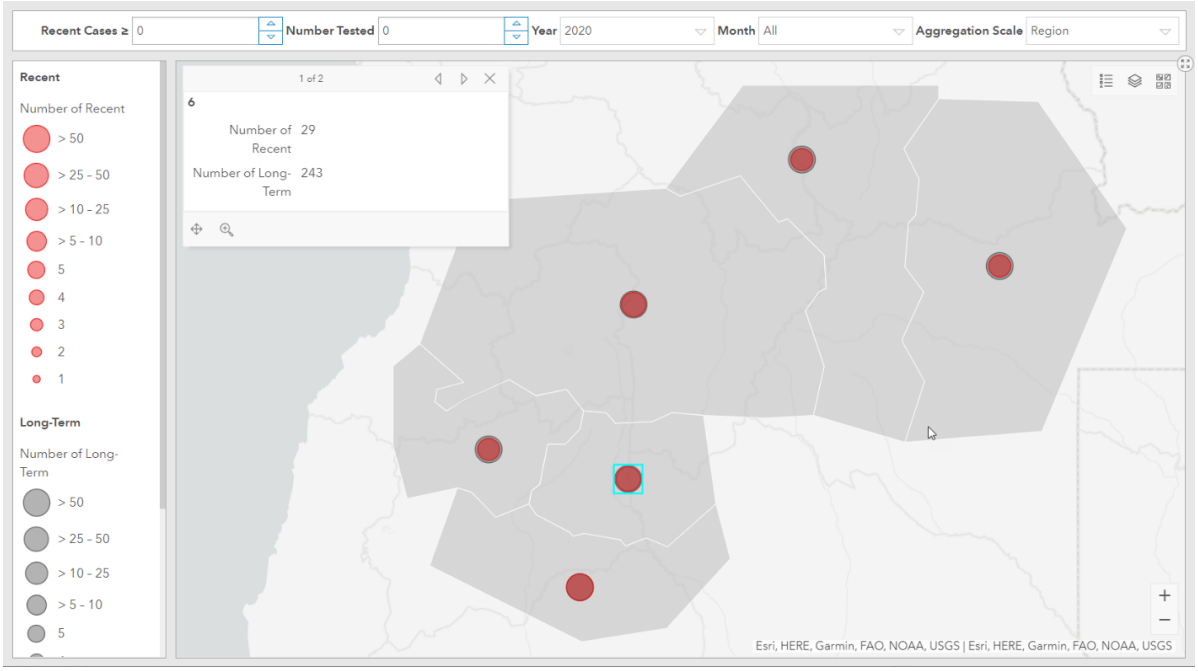
A few common data cleaning items that are handled nicely in the Query Editor are:

- Converting text-based data into numerical types, or the opposite, converting numeric data into text
- Formatting text values – standardizing spelling or removing unnecessary prefixes/suffixes
- Removing bad characters – leading spaces or unnecessary characters

Note that data cleaning that is handled in the Query Editor will be applied to each row in the same way and will then be loaded to the data model. This method is best for changing data types and formatting. Data cleaning steps that are applied to the dataset are tracked in the Applied Steps of the Query Editor.

7. Embedding Online Maps for TRACE

One feature of the Power BI Service is the ability to embed external content into the App. For TRACE, we recommend using ArcGIS Online (AGOL) to build maps displaying the locations and number of recent and long-term infections by health facility geolocation, with the option to add layers displaying number of recent infections by place of residence, viral load suppression or other programmatic data and can be configured according to available data and program needs. While other types of mapping platforms can be embedded into Power BI App, AGOL provides the TRACE project a secure platform to display and analyze recency on customizable mapping applications. For this guide, we provide brief orientation on using AGOL to visualize recency data and embedded in the Power BI App.



ArcGIS Online Operations Dashboard

7.1. Web Mapping with ArcGIS Online

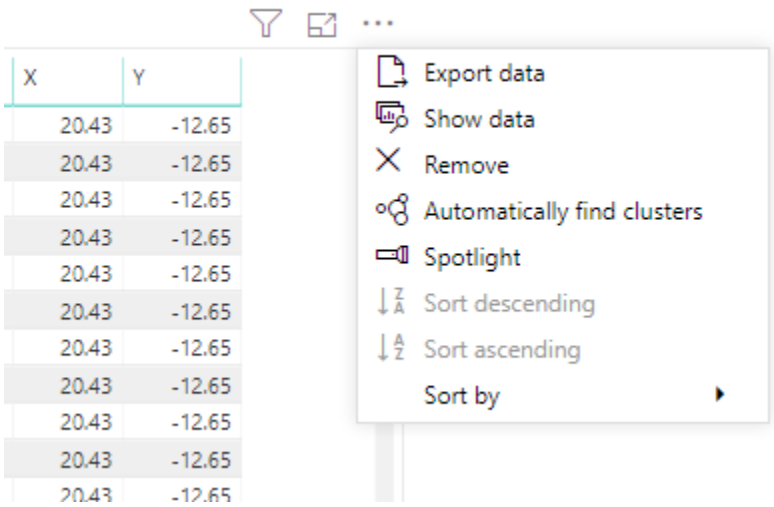
ArcGIS Online (AGOL) is Esri's cloud-based geospatial analysis and visualization platform that connects users with data and locations using interactive maps. ArcGIS Online specializes in hosting spatial data such as point records with latitudes/longitudes or polygon features such as administrative regions. Users can use AGOL to make maps and analyze data in order to gain insight into their geospatial data. AGOL stores data and maps in a secure and private infrastructure and allows you to share maps with specified users via the web or embed in other applications such as Power BI apps. The robust mapping capability of AGOL allows TRACE project viewers to interact with maps containing multiple layers and data points, which enables near real time geographical tracking of recent infections. AGOL is being used in addition to Power BI for its rich mapping capabilities, which are limited within Power BI itself.

To get set up with procuring ArcGIS Online, contact your local Esri office to figure out the appropriate number of license seats. Individuals will need to have a Creator or similar level user type that has access to the Essential Apps Bundle.

7.2. Exporting Data from Power BI Desktop

The recency data in Power BI needs to be summarized and then exported to ArcGIS Online for map visualizations. In the generic TRACE template, there is a preconfigured hidden page in the report called "AGOL". This hidden page does not get published but is used to aggregate data into a required format for the ArcGIS Online Dashboards that includes latitude and longitude coordinates for each record. The count of recent and long-term cases is aggregated using a Power BI table visualization by several variables including aggregation scale (facility, region, district) and monthly time step. For area level aggregations such as by region or district, the point coordinates should be the centroid of the feature.

To export the table, activate the table visualization to bring up the more options button (three dots in the upper right corner). Click the button and select Export data. A dialogue box will open prompting you to save the file as a csv (comma separated value) file.



Export Table Visualization to CSV

7.3. Uploading Data to ArcGIS Online

The csv file that was exported from Power BI can be added to ArcGIS Online as points. ArcGIS Online uses the latitude and longitude of the points to geocode the records. TRACE has created template maps in the TRACE Operations Dashboard within an institutional AGOL account and can support your country to do the same should you want to include this feature in your Power BI Workspace. Once set up, use the following steps to update the data in your AGOL maps.

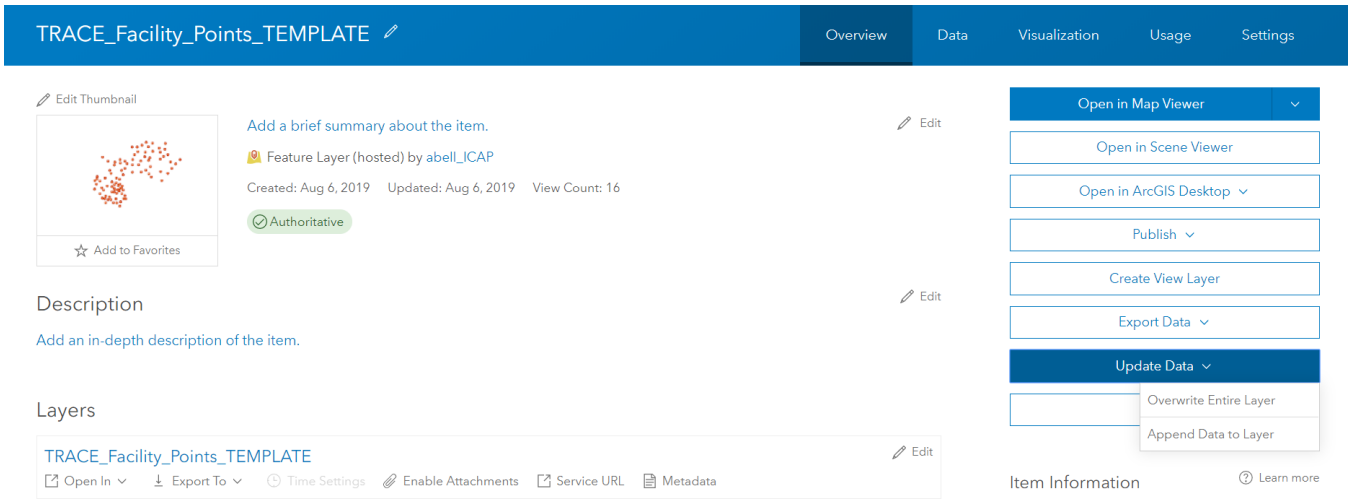
Step 1. The file initially can be uploaded to ArcGIS Online by using the *Add Item button* in the content page.

Step 2. The path to the new data source will need to be updated for any preconfigured web maps. For data updates, the item on ArcGIS will be the same and the entire data set will be overwritten.

Step 3. Open the item details in ArcGIS Online and select the Update Data menu.

Step 4. Pick the option to Overwrite the Entire Layer and upload the csv with the new dataset.

Please note that the file name of the file must have the same name as the original file and all field names must be the exact same. The user must be the owner of the item or an administrator to update the data by overwriting the layer

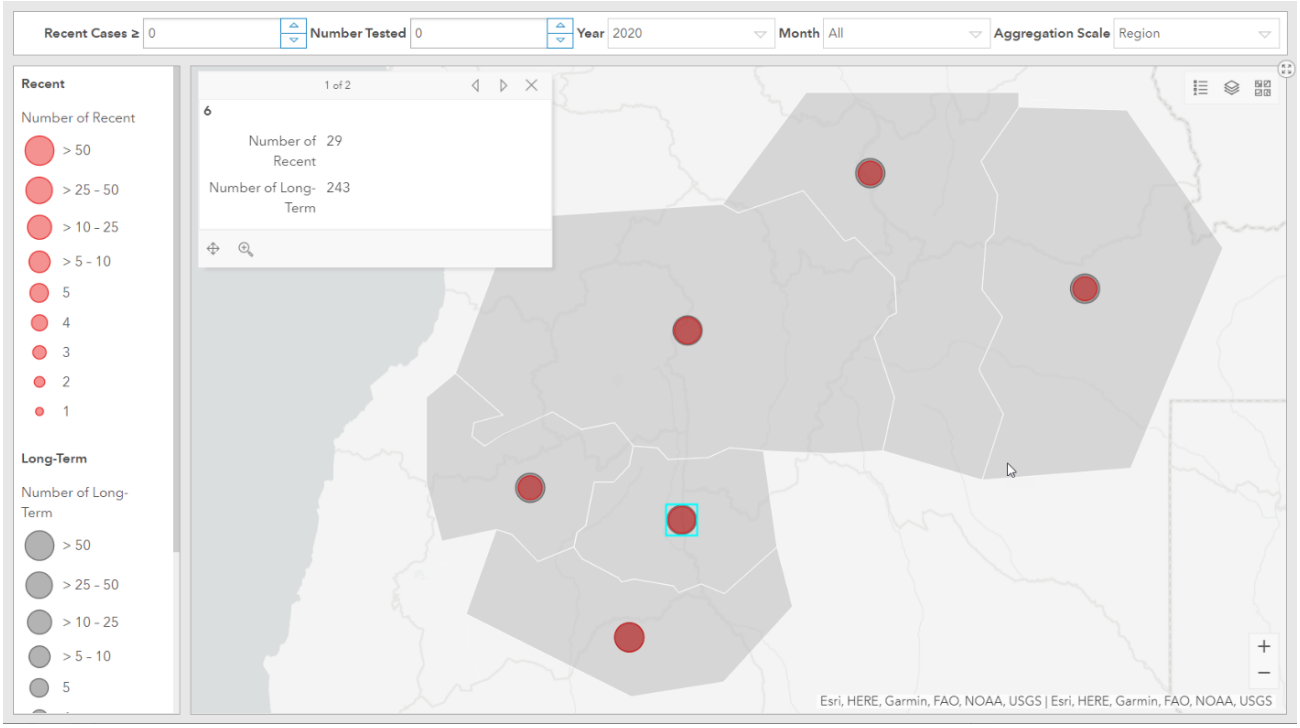


Data update in ArcGIS Online: Overwrite Entire Layer

7.4. TRACE Operations Dashboard Map in AGOL

ArcGIS Online Operations Dashboards are published apps that are designed to provide a view of geographic information to help monitor events and activities. Dashboards share preconfigured maps and visualizations with a select group of AGOL users and/or the general public. For TRACE, operations dashboards are used to provide an user-interface for the recency maps. This allows the user to have dropdown menus to filter and examine different facets of the data on the map.

The default TRACE operations dashboard shows the number of recent and long-term cases at various spatial and temporal scales. The user can select monthly snapshots to view trends over time or see the number of cases reported for the year-to-date. The total number of cases can be aggregated at various spatial scales to provide summaries at the facility, sub-regional or regional scale. Records can be filtered by thresholds including the number of test cases. The dashboard can be customized for needs of each country.

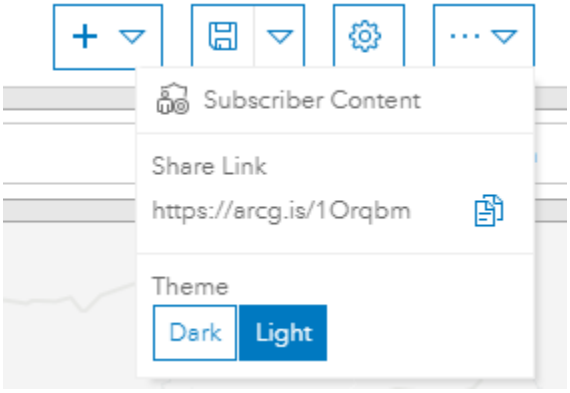


TRACE Operations Dashboard in AGOL

7.5. Embed in Power BI App

The operations dashboard from ArcGIS Online can be embedded into the Power BI App.

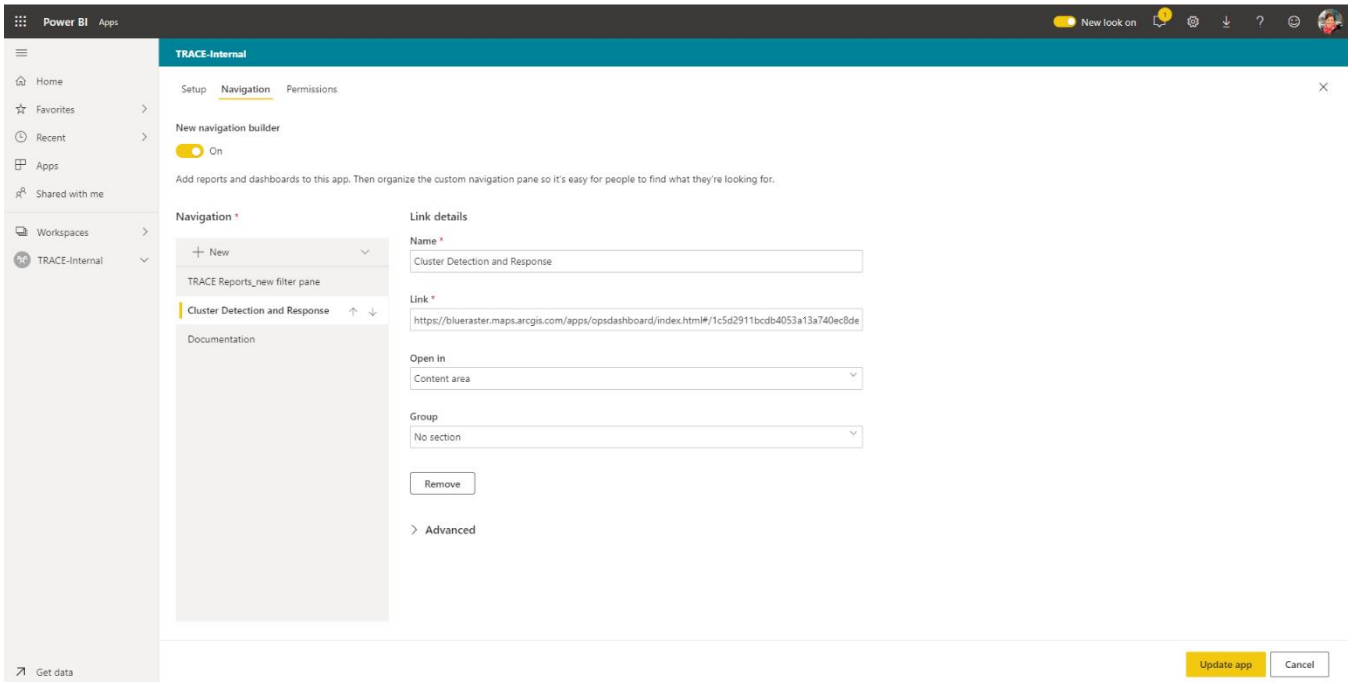
Step 1. To get the URL to share, open the dashboard in edit mode and copy the Share Link under the more options menu. Please also verify that the application and all underlying files (maps, data sets, etc) are shared with the appropriate ArcGIS Online Group.



Operations Dashboard Share Link

Step 2. Once the dashboard URL is obtained, the map can be embedded into the app by logging into the Power BI workspace and clicking the Update App button.

Step 3. Under navigation select the page for the content and replace the Link URL with new or updated URL for the Operations Dashboard and click Update App.



Update Dashboard Link URL in Power BI App

8. Additional Documentation and Resources

In addition to this manual, users can find other documentation available for download on the TRACE Learning Hub website. Documentation available includes:

- *Power BI Resource Guide*
- *TRACE Template Codebook*
- *Power BI Admin Guide*

8.1. Power BI Resource Guide

Power BI is a tool that was designed for the citizen developer, meaning that individuals who wish to learn and implement the solution do not need to have a background in development/code. The number of resources that exist for Power BI is infinite, and as such, the TRACE team has compiled a list of highest quality, free resources that will guide new users down the road to success.

8.2. TRACE Template Codebook

The TRACE Template Codebook contains a list of the data elements contained in the sample data files and will need to be consulted during the data mapping process, in which country data points are mapped to the sample data points. The information contained includes:

- Variable name and description
- Data Type
- Data Format

This document also contains a complete list of all the calculations that have been made by the TRACE team within the TRACE Template, in both the form of calculated measures and columns.

8.3. Power BI Admin Guide

The Power BI Admin Guide provides more detailed information about the nuts and bolts of administering a Power BI Instance; everything from details on configuring Power BI in the web to licensing. This guide is intended for IT/Informatics staff and provides guidance on the bigger picture of Power BI as a software.