

12 APPENDIX C- User reference documents

12.1 *mHealthSurvey user manual*

12.1.1 Introduction

The Real-Time Biosurveillance Program (RTBP) is an m-Health pilot project aiming to answer the question: “Can software program that analyze health statistics and mobile phone applications that collect and report health information potentially be effective in the early detection, intervention, and prevention of disease outbreaks?” This pilot aims to study the technology, human, and policy predicaments in introducing the RTBP to Sri Lanka and India.

This document’s objective is to explain how to install and handle the Mobile Application “*Application Name*” to the field healthcare workers.

12.1.1.1 Outline

The healthcare worker (ex.VHN or Suwacevo) has to carry the mobile phone with health application installed, to the villages or health facilities at the time of health visit. They have to fill the health survey form and send the data to the RTBP server through GPRS.

The application has two parts:

- Local setting – Application has provision to store the local setting like healthcare workers profile and their working location. First, the lookup values such as the location types, health worker type, disease, symptom, and sign lists must be downloaded from the server through GPRS.
- Health survey – Healthcare worker has to fill the health survey form and sends data to RTBP server. In case of problem in GPRS or signal, the form will store the data in the mobile (persistence storage). Later after getting the signal, the healthcare worker has to resend the stored data to RTBP server.

12.1.1.2 Features

The “*Application Name*” has following features, which enables the easy and efficient way of usage to the healthcare workers.

- The User Interface consists of J2ME GUI components like Forms, Buttons, Canvas, Textbox, Text fields, Alert boxes and Lists.
- User has the options to change the settings both locally and remotely.
- Healthcare workers are provided to create, edit and delete their profiles and locations.
- Single phone has the provision to create and access multiple healthcare workers profiles.
- To save time from typing the values through keypad, popup menus are created.
- Compared to WAP portal, this application takes less downloads and very less bandwidth.

12.1.2 How to use this Application

12.1.2.1 Need to follow

- Before you start survey please check that the mobile has sufficient charge. If not, you have to charge your mobile and then take survey.
- Make sure the signal is available when you start working on “Server connection,” ”Profile,” ”Location.” Without signal, you cannot work on these forms.
- While doing “Download,” do not press the cancel or exit button until you get “Successfully download” message.

12.1.2.2 How to open our “Health Survey Form”

There are different ways to open the “M-Health Survey’.

- Opening the application from Memory Card (→ indicate ‘Choose’)
Menu → Applications → Memory Card → RTBP → Health Survey

12.1.2.3 Types of forms available in “m-Health Survey”

The “M-Health Survey” has following menus,

- Health Survey
- Offline Survey
- Server Connection
- Profile
- Location

Before start working on the Health survey, do the following steps and its details are given below

- Connect the server
- Set your (Healthcare worker) profile(s)
- Download the location(s) you are (Healthcare worker) working on

12.1.3 Server connection

12.1.3.1 Server connection Instruction

It is a simple process. Just select the “Server connection” menu from “m-Health Survey”. Now the list will be downloaded automatically and you will get a status message as “Server connected successfully”

12.1.3.2 Error and exception on Server connection



Figure 77: Main menu of the application

Network Error! Try after some times

Generally, this error occurs when you are not having network coverage or GPRS failed. So please try this process once you have proper network coverage.

12.1.4 Profile

12.1.4.1 Profile Menu Options

- Save – **To save the entered data to server**
- Reset – Clear the textbox, display a new form
- Back – return back from the current form

12.1.4.2 Profile Instruction

Here you have to provide your personal details for the registration process. These instructions will help you to fill up the profile page:

Step 1: Choose the “Profile” menu from the “m-Health Survey”

Step 2: The following fields are available in the profile and mandatory fields are marked with an asterisk (*). If you leave these fields as blank then it will give you warning message as “**Fill the Required Values (*)**”.

- ID: (*) → Here you have to type your Healthcare Worker’s ID.
- Retype ID: (*) → Here you should retype your Healthcare worker’s ID for confirmation.

If the “ID” and “Retype ID” values are not same, it will give you an error message as “**ID is mismatching**”.

- First Name: (*) → Type your first name.
- Last Name: → Type your last name, if any. You can type your initial instead.
- Type: (*) → Here we have three options. i.e VHN, DDHN and Health Inspector from there please choose one option.

To Choose, press the enter key and use down arrow to choose the option and again press the enter key. Now the chosen option will display

Figure 78: Profile registration form

Phone: → Give your contact number. It is not mandatory.

E-Mail: → Enter your email ID. It is also not mandatory.

Step 3: Now to “save” your record please press the left side first key, you will get 3 options

- Save – Save the filled data to server
- Reset – If you want to enter a new profile data press this option now all the profile for fields are cleared and you will get a blank application
- Back – If you want to skip from the current application press this option, your application will be skipped and will display the “main form” window.

Step 4: After you press “save,” you will be asked, “Allow network access?” You should press the enter key for “yes.” If the network access is available at that time, your details will be saved and it displays a message as “Successfully saved.” Otherwise, you will get an error message as “**Network Error! Try after some time.**” It means your data was not saved. In such case, you have to give your information again, once you get the signal.

12.1.5 Error and Exception on Profile

Fill the Required Values ()*

Mandatory fields are marked with an asterisk (*). If you leave those fields as blank, it will give you warning message as “**Fill the Required Values**”. To avoid this error you need to fill all the mandatory fields.

ID is mismatching

If the “ID” and “Retype ID” values are not same then it gives you an error message as “ID is mismatching”. To avoid this error, you need to give same value for ID and Retype ID fields.

Network Error! Try after some times

Generally, this error occurs when you are not having network coverage or GPRS failed. So please try this process once you have proper network coverage.

12.1.6 Location

12.1.6.1 Location Menu options

- Save – To save the entered data to server
- Reset – Clear the textbox, display a new form

12.1.6.2 Location Instruction

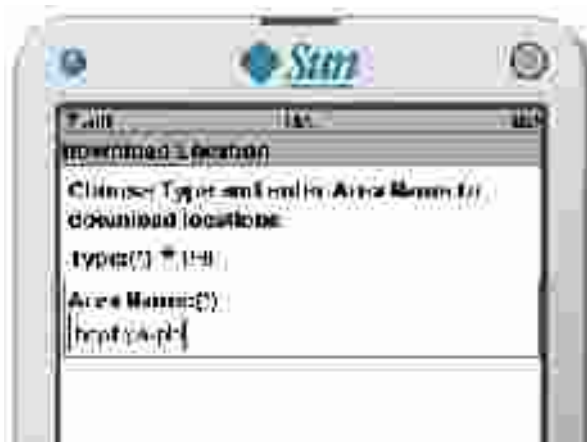


Figure 79: Location(s) retrieval form

Here you have to fill up your location type and location name To download your villages. Eng location type = “PHI” and location name = “bopitiya-phi”

Step 1: Choose the “Location” menu from our “m-Health Survey”

Step 2: It has two fields, there are - Location Type and Location Name

Step 3: Choose location type from the drop down box.

Step 4: Enter the location name

Step 4: Press “send” button to send the information to server.

Step 5: Now it will ask, “Allow network access?” You should press the enter key for “yes.” If the network is available at that time, your details will be saved and display a message as “Location Successfully Downloaded.” Otherwise, you will get an error message as “Network Error! Try after some times.” In such case, you have to download the location once you get the signal.

Error and Exception on location

12.1.6.3 No locations found

If the location type and location name combination is not available in the server then you will get this error.

Fill the Required Values (*)

Mandatory fields are marked with an asterisk (*). If you leave those fields as blank, it will give you warning message as “Fill the Required Values”. To avoid this error you need to fill all the mandatory fields.

Network Error! Try after some times

Generally, this error occurs when you are not having network coverage. So please try this process once you have proper network coverage.

12.1.7 Health Survey

12.1.7.1 Health Survey Menu Option

Next – This option will be available on the first page of health survey and used to go to the next page.

Figure 80: Health Survey Form one (demographics)

Update – Used to save the data to server.

12.1.7.2 Health Survey Instruction

Once the process on Profile, Location and server connection has been completed, you can start working on our health survey. The following picture shows our health survey details:

Fields available:

- Date
- Location
- ID
- Patient First Name
- Last Name
- Notes
- Sex (*)
- Age Group (*)
- Search Disease
- Disease
- Search Symptoms
- Symptoms (*)
- Search Signs
- Signs
- Status



Figure 81: Health Survey Form two (Diagnosis)

Step 1: Date (*) - Here it displays the patient symptom reporting date and time to health worker. By default, it is set to current date and time automatically. If you want to edit the field, press the “enter” button to place the cursor on that field and change the value and then press “save” button.

Step 2: Location (*) - Here it will display the list of villages that you have downloaded already using the location form. Here you cannot edit the village name. To choose a village press the enter button and now it will display the options, using up or down arrow place the cursor on an option and again press the enter button. Now the chosen village will be displayed.

Step 3: ID (*) - It displays the healthworker ID automatically here. You cannot edit it but should select the one relevant to you.

Step 4: Patient’s First Name (*) - Enter the Patient’s first name with their initial. This is optional if you need the health record anonymous.

Step 5: Patient’s Last Name - Enter the patient’s last name here; this field is not mandatory.

Step 6: Note - This can be used to enter additional information about the patient

Step 7: Gender (*) - Select the patient gender from the list. This field is a mandatory field. To choose a Sex press the enter button and now it will display the options, using up or down arrow place the cursor on an option and again press the enter button. Now the chosen Sex will be displayed.

Step 8: Age Group (*) - Select the age group from the drop down list. This field is a mandatory field. To choose an Age Group press the enter button and now it will display the options, using up or down arrow place the cursor on an option and again press the enter button. Now the chosen Age Group will be displayed.

The first page is completed here. To go to the next page please press the option button it will display 'next' button, press that. Now you can see the next page.

Step 9: Search Disease - If you enter the first letter of the disease, it will display a list of disease starts with that letter. You can choose any one from this list. If the disease is not available in the list, you can use the "others" option.

Step 10: Disease - You will get a result of your search here and please choose the disease from the drop down box. To choose disease we have to use up and down scroll button and place the cursor on an option and then press the enter button, then go to Symptoms option.

- **Others ::** If the disease is not available in the list, you can choose the "others" option from the disease box and enter the disease name in the 'search disease' text box .
- **Unknown ::** If you are not able to diagnosis the disease name of the patient, choose this "unknown" option from the disease list box

Step 11: Search Symptoms - When you select a disease, the related symptoms will be displayed here. You can Add, Edit, Delete the symptoms. Each symptom has to be comma separated. If you have not selected a disease from the list, then no symptoms will be displayed here. You have to enter the symptoms.

In this case, when you enter the first letter of the symptom, it will automatically display the list of symptom starts with that letter in the "symptoms dropdown box". You can choose any one from this list. The selected symptom will be displayed in this text box. If the symptom is not available in the list, you can directly type the symptom here.



Figure 82: Search and select symptoms

12.1.7.3 Symptoms (*)

When you enter the first letter of the symptom in the "search symptoms box", it will automatically display the list of symptom starts with that letter. You can choose any one from this list. The selected symptom will appear in the search symptom box. If the symptom is not available in the list, it will not

display any items.

12.1.7.4 Search Signs

When you select a disease, the related signs will be displayed here. You can Add, Edit, Delete the signs. Each sign has to be comma separated. If you have not selected a disease from the list, then no sign will be displayed here. You have to enter the signs.

In this case, when you enter the first letter of the sign, it will automatically display the list of signs starts with that letter in the “signs dropdown box”. You can choose any one from this list. The selected sign will be displayed here. If the sign is not available in the list, you can directly type the sign in this text box.

12.1.7.5 Signs (*)

When you enter the first letter of the sign in the “search sign box”, it will automatically display the list of sign starts with that letter. You can choose any one from this list. The selected sign will appear in the search sign box. If the sign is not available in the list, it will not display any items.

12.1.7.6 Status

This drop down box has three values such as **Treated, Referred and Unknown**. This field is not a mandatory. To choose a status press the enter button and now it will display the options, using up or down arrow place the cursor on an option and again press the enter button. Now the Chosen status will be displayed. If you do not know the case status, you can select ‘Unknown’ from the list.

Once you have filled the required fields, you can save the data. If there is no signal, the data will be stored in the “Offline Survey “form. While doing survey if you want to go back to the first page from the current page press the option button and use back button.

6.3 Error and exception on Health Survey

Fill the Required Values (*)

Mandatory fields are marked with an asterisk (*). If you leave those fields as blank, it will give you warning message as “Fill the Required Values”. To avoid this error you need to fill all the mandatory fields.

12.1.8 Offline Survey

12.1.8.1 Offline Survey Instruction

If you do not have signal when doing survey, the data will be stored here. After getting the signal, just press the “Offline Survey” form. Now the data will send to the server and the offline survey form will be empty.

12.1.8.2 Error and exception on Offline Survey

Network Error! Try after some times

Generally, this error occurs when you are not having network coverage. So please try this process once you have proper network coverage.

12.1.1 Basic Instructions

12.1.1.1 8.1 Need to follow

- Keep your mobile out of the reach of small children.
- Do not use the options on Security code, PIN code and PUK code.
- Keep the device dry, if your device does get wet, remove the battery, and allow the device to dry completely before replacing it.
- Do not drop, knock or shake the device. Rough handling can break internal circuit boards and fine mechanics.
- Do not use harsh chemicals, cleaning solvents or strong detergents to clean the device.
- Do not store the device in hot and dusty areas.
- If the device is not working properly, report to the team immediately.

8.2 Install SIM card and battery

- Always switch the device off and disconnect the charger before removing the battery.
- Don't change the battery. Always use original Nokia batteries.
- The SIM card can be easily damaged by scratches or bending, so be careful when handling, inserting, or removing the card

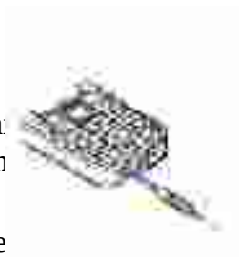
12.1.1.2 Charge the battery

- Check the model of any charger before use with this device. This device is intended for use when supplied with power from the AC-3 or AC-4 charger.
- Do not leave a fully charged battery connected to a charger, since overcharging may shorten its lifetime.
- If the battery is completely discharged, it may take several minutes before the charging indicator appears on the display or before any calls can be made.

12.1.1.3 Warning

User only batteries, chargers and enhancements approved by Nokia for use with th
The use of any other types may invalidate any approval or warranty, and may be dan

- Connect the charger to wall socket
- Connect the lead from the charger to the socket on the bottom of your device



If the battery is completely discharged, it may take a few minutes before the charging indicator appears

on the display or before any calls can be made.

The charging time depends on the charger used. Charging a BL-5c battery with the AC-3 charger takes approximately 2 hours and 45 minutes while the device is in the standby mode.

12.1.1.4 8.5 Antenna

Your device has an internal antenna.

Note: As with any other radio-transmitting device, do not touch the antenna unnecessarily when the device is switched on.



12.1.1.5 Keypad lock

- To prevent accidental key presses, select Menu, and press * within 3.5 seconds to lock the keypad.
- To unlock the keypad, select Unlock, and press * within 15 seconds.
- To answer a call when keyguard is on, press the call key. When you end or reject the call, keypad automatically locks.

12.1.1.6 Time and Date

To change the clock type, time, time zone and date setting, select Menu → Settings → Time and Date → Time → Time, Date, Zone or Auto update of Date and Time

12.1.1.7 Your Device – Nokia 3110C

The device may prompt you to set the time and date. Enter the local time, select the time zone of your location in terms of the time difference with respect to Greenwich Mean Time (GMT), and enter the date.

12.2 TCWI user manual

12.2.1 1. Introduction

This manual is intended as a quick introduction for public health officials and epidemiologists – the end-users of T-Cube Web Interface (TCWI). It is relevant to the specific version of TCWI tailored to the requirements of the Real-Time Biosurveillance Program (RTBP), conducted as a pilot in India and Sri Lanka. The TCWI as well as the underlying data representation and analytic technologies have been developed by the Carnegie Mellon University Auton Lab, in Pittsburgh, Pennsylvania, the United States of America.

TCWI is a front-end for the public health databases collected through RTBP. It allows for visualization, statistical analysis and navigation through data. It uses modern computer science, statistics and machine learning technologies to support public health analysts and epidemiologists in their daily duties, including monitoring for disease outbreaks, outbreak investigations, and reporting.

The underlying technology provides the users of TCWI with unique abilities to very quickly navigate through and mathematically analyze large amounts of data even if it spans highly multidimensional spaces of multiple diseases, locations, symptoms, syndromes, and demographic factors. The users of TCWI benefit from it by becoming able to concurrently monitor many more hypotheses about the status of public health much more thoroughly than it was possible before. The data can be comprehensively mined for statistically significant increases of counts of specific subpopulations reporting specific symptoms, with respect to baselines inferred from historical trends. The results of such automated massive screenings are presented to the users in a form of the list sorted according to their statistical significance. The users can then navigate through the list and easily drill-down into details of the corresponding data for additional hints and explanations. A skilled operator can use TCWI to support maintenance of high levels of awareness of current epidemiological situation and of ongoing processes that affect monitored populations, enabling fast and reliable identification of emerging problems and creating opportunities for implementing focused and effective responses to crises.

The following sections briefly review the essential functionality provided by TCWI. The ideal user of TCWI should be fundamentally computer literate and be familiar with fundamental methods of analysis of public health data and with the objectives of such analyses.

While RTBP would provide their respective users with a web link (URL), other first-time users may access a demonstration version of the TCWI available on the Auton Lab site (<http://www.autonlab.org/T-Cube/>) for the purpose of hands-on experimentation with the features and functions described in this manual.

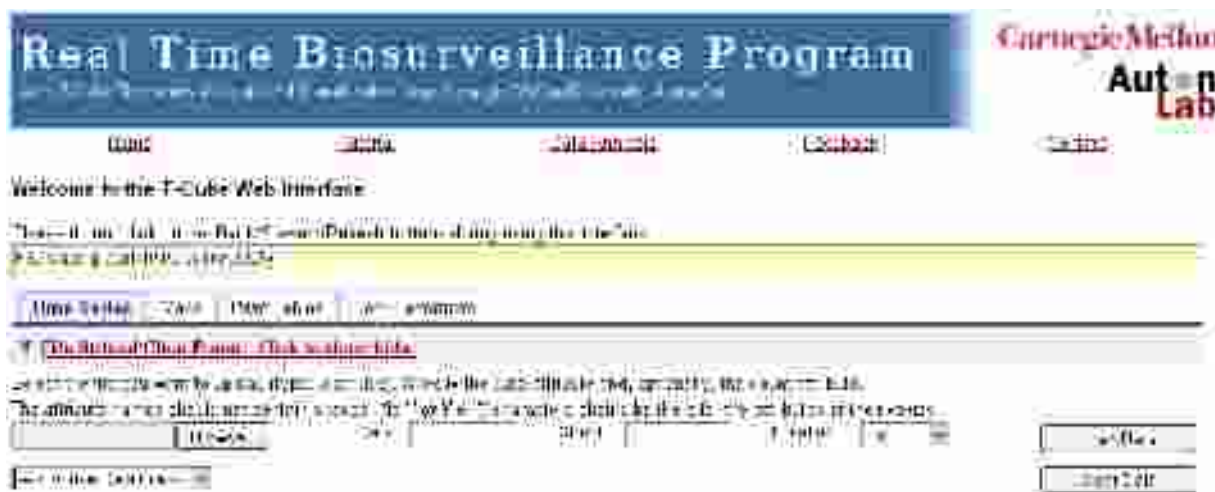


Figure 83: Front panel of the T-Cube Web Interface.

Figure 83 shows the starting panel of the TCWI with the project header and tabs for panels containing specific analytic components: **Time Series**, **Maps**, and **Pivot Tables** (note: tab labeled “Poor Performers” is irrelevant to the RTBP project). Each analytic component is described in a separate section of this manual: Time Series Analysis, Maps (Spatio-Temporal Analysis), and Pivot Tables, respectively.

Immediately below the header, there are links for specific sections of the TCWI. Figure 83 shows the **Data Analysis** section of the interface. In this manual, we will focus on the data analysis section. The **Feedback** section can be used to send direct feedback to the TCWI developers. The **Tutorial** section points to an older version of this manual and it will be updated soon. For now, please refer to this document as the current reference manual.

12.2.2 Time Series Analysis

Time series analysis component of TCWI supports both univariate (Analysis Panel) and multivariate temporal analysis (Massive Screening Panel). Each of the five panels under Time Series Analysis is described below.

12.2.2.1 2.1 File Upload/Clear Panel

This panel allows the users to select data for analysis. The data can be either loaded from disk on the local machine (advanced usage) or it can be preloaded on the server (typical usage).

12.2.2.2 Using Pre-loaded Data (Typical Usage)

To use a pre-loaded data, just click on the **Choose Data File** to expand a drop-down list of available data files and pick one for analysis. This tutorial uses **lk_flat_table** data as an example (at the moment of writing this revision of the manual we did not have access to an equivalent set from India, but the functionality of TCWI will be identical when used on data collected in India). This data contains a total of 69,000 reportable disease cases from multiple regions in the country of Sri Lanka collected starting

December 16, 2006 and ending on July 10, 2009. It spans the following fields: **loc_name**, **age_grp**, **disease**, **gender**, **sign**, and **symptom**. Note that the original data contained weekly aggregates for each combination of **loc_name** and **disease**. We have semi-synthetically converted this data into daily resolution with more demographic attributes to match the scheme and level of detail of data being currently collected in through the RTBP project.

12.2.3 Loading External Data (Advanced Usage)

The users of TCWI can load their own data instead of using the pre-loaded data sets. This mode of usage is not expected or considered typical in the context of the RTBP project, since the relevant data will be prepared for use and frequently updated through automatic processes set up by the maintainers of the system. We explain it here for the sake of completeness.

TCWI accepts external files in the comma-separated values (csv) format. The first line of it contains names of the data dimensions. The subsequent lines contain the actual data, each record denoting either an individual disease case (in the purely transactional format of data), or it represent a cumulative count of disease cases matching the specific combination of values of dimensions represented in the record (the cumulative count format of data). A top fragment of an example of a TCWI-compatible file in the cumulative count format is shown below. It has four dimensions named *date*, *lk_city*, *disease*, and *count*. Each record reveals the number (*count*) of specific reportable *disease* cases reported in Sri Lanka region labeled with its main city (*lk_city*) on a given day (*date*). If the count dimension is not available, TCWI data loader will assume that each record represent a single case (that is that its count value equals 1), and so that the data is provided in the purely transactional format.

```
date, lk_city, disease, count
DEC-16-2006, Colombo, Dengue_fever, 7
DEC-16-2006, Colombo, Dysentery, 1
DEC-16-2006, Kandy, Dengue_fever, 1
DEC-16-2006, Kandy, Dysentery, 2
DEC-16-2006, Matale, Viral_Hepatitis, 1
DEC-16-2006, Nuwara_Eliya, Viral_Hepatitis, 1
DEC-16-2006, Galle, Dengue_fever, 1
DEC-16-2006, Hambantota, Typhus_fever, 1
DEC-16-2006, Matara, Dengue_fever, 3
DEC-16-2006, Matara, Leptospirosis, 1
DEC-16-2006, Vavuniya, Dengue_fever, 1
DEC-16-2006, Kurunegala, Dengue_fever, 2
DEC-16-2006, Kurunegala, Typhus_fever, 1
DEC-16-2006, Puttalam, Encephalitis, 1
```

To load an external data set, use the **Browse** button (cf. Figure 83) to interactively select the file to load from the disk on the local machine. Once the location of the file is specified (for instance, C:/mydata/myfile.csv), the user needs to identify which of its dimensions represents the date dimension and enter its name in the field next to **Date** (in the above example, the user would have entered “date”). Similarly, if the individual records of data represent multiple counts (as in the example above), the user needs to provide the name of the count dimension by entering it in the field next to **Count** (that field is named “count” in the above example). Purely transactional data, in which each record represents a single unique event, does not require the count dimension and therefore the **Count** entry field can be

left empty.

If the data contains a spatial dimension (in the example above, the spatial column was *lk_city*), then the user must specify the type of map to be loaded using **Location** drop-down list.

Once all required information for loading external data is specified, click on the **Load Data** button to load data into TCWI. Depending on the current workload of the TCWI server and the complexity and size of the dataset being loaded, this operation may take some time and patience is advised. The investment in waiting time will be returned many times by the resulting speed of responses to user and algorithm queries against the loaded data.

It is not a requirement, but a good habit, to clear computer memory of data that is no longer in use, before loading a new data set. To accomplish that click on **Clear Data** button at any time, and then proceed to loading new data as needed.

Note that TCWI expects the external data to meet the above described format requirements (specifically, all dimensions should be categorical, except for the count and date dimensions), and it should not contain any missing values. The interface will signal an alert if an attempt is made to load an incompatible data set.

12.3 Query Selection Panel

Query selection panel incorporates drill-down functionality to filter data by selecting subsets of values of individual dimensions. It also enables interactive visualization of multiple time series. Below we briefly introduce two basic modes of its operation. This description is not comprehensive, but the functions not addressed here are intuitive and accessible via obvious user interactions.

12.3.1.1 Filtering Data by Selecting Subsets of Values of Individual Dimensions

Figure 84 shows an example view of the query selection panel involving the kind of data typically used within the RTBP project. Using value filtering, the user can select a subset of values for each dimension of data, and the corresponding time series labeled as **Current Query** will be automatically updated in the time series analysis window, as shown in Figure 87. For convenience, TCWI automatically visualizes the time series resulting from the all-inclusive query called **All Data**. It reveals the temporal distribution (by day) of all disease case records stored in the currently loaded data. Any other, more specific query, will result in a subset of **All Data**.



Figure 84: Starting view of the Query Selection Panel.

There are two modes in which the user can select the subsets of values of individual dimensions: **list view** and **tile view**, accessible by clicking on one of the small icons next to each dimension name

(Figure 84). In the **list view** (see Figure 86) the values of the selected dimensions are presented in the checklist form and the user can select any subset of values. **Tile view** mode (Figure 85) offers more extensive view, in which the users can filter for specific values using regular expressions.



Figure 85: List view of the Query Selection Panel.

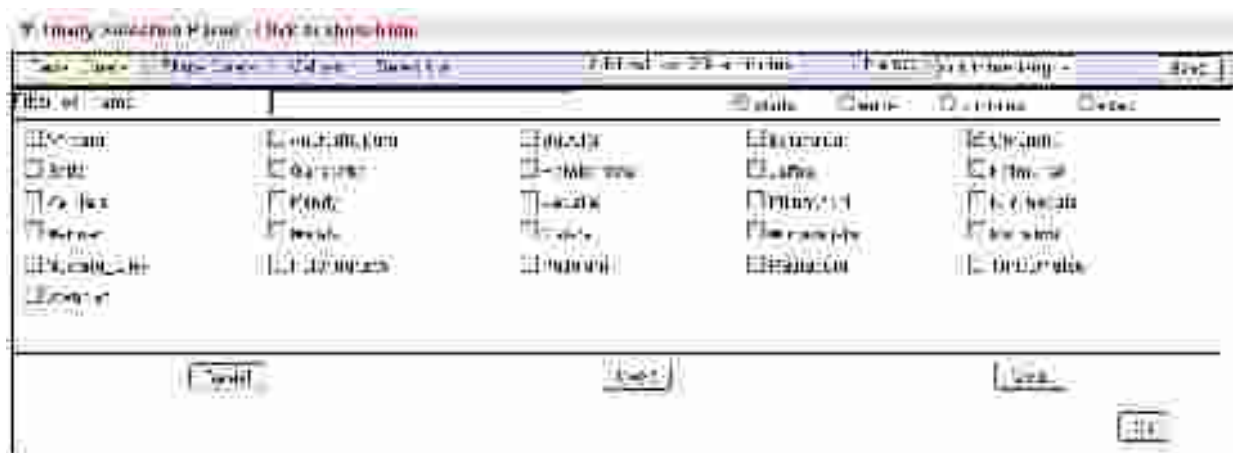


Figure 86: Tile view of Query Selection Panel.

Each time the user updates their choices in the query selection panel, the TCWI extracts the daily counts of disease cases corresponding to the user-selected set of values, and the resulting time series is shown in the visualization window under the name of **Current Query**. The user can optionally save the time series corresponding to the current query by giving it a unique name by entering it in the **Name** text box. The time series visualization window will then add it under the new name to the list of available series. For example, in Figure 87, the *red* time series corresponds to temporal distribution of counts of *Dengue fever* cases in *Colombo* and is named “colombo-dengue”. It was obtained by manually selecting Colombo and Dengue_fever values of the *loc_name* and *disease* dimensions through the query selection panel. Similarly, the *green* time series corresponds to the total number of patients in all regions and diseases. The user has named this series “all-values”. The **Current Query**, shown in *blue*, represents in this example the daily number of Dengue fever cases collected across all data – this matches the current selection of values and dimensions in the query selection panel.

Note that in the case of the *lk_flat_table* dataset, the dimensions named *sign* and *symptom* are available to the user in a composite form. Each disease case can be associated with multiple signs and symptoms, and therefore each individual value of a sign as well as symptom variable is represented internally as a separate binary dimension. In order to avoid cluttering the TCWI panel with multiple

binary dimensions of data, it collapses their groups together. When the user selects one value of such a dimension for their query, all records for which that value is present will be reflected in the result. If the user selects two distinct values of e.g. *symptom* the response to such query would include all records mentioning both of the two symptoms. This is different than in the case of regular (not composite) dimensions, where selecting two values leads to concatenating the records of data which match either of the selected values, so that the resultant is a sum of the two subsets matching the individual values. On the other hand, two or more values selected out of a composite attribute on the other hand would produce the joint part (the product) of the individual value matching subsets – only records matching both of them will be reported. Currently, only *sign* and *symptom* are represented in the composite form. Future release of TCWI will allow the users to select one of the two types of representations for each qualifying attribute.

12.3.2 Visualization of Time Series

Figure 87 shows the Time Series Analysis panel used for visualization. It shows a couple of time series previously queried for, extracted, and named by the user. Their display status can be toggled on-and-off using the check boxes next to their names in the legend to the left of the graph.

The horizontal axis of the main time series visualization window denotes time at a daily resolution, arranged from the oldest (left) to the newest entries (right). The vertical axis is primarily used to reference the values of the daily counts corresponding to the displayed time series. Its secondary purpose is to reference the magnitude of alerts generated by the statistical algorithms for event detection, discussed later. The scale for daily counts is shown at the left side of the main time series window, while the scale for alerts is on the right. The scaling of counts axis can be controlled via the **Scaling** dropdown menu. The scale of the magnitude of alert signal can be toggled between linear and logarithmic display (note that no alert signal time series are shown in Figure 87, this functionality will be explained later).

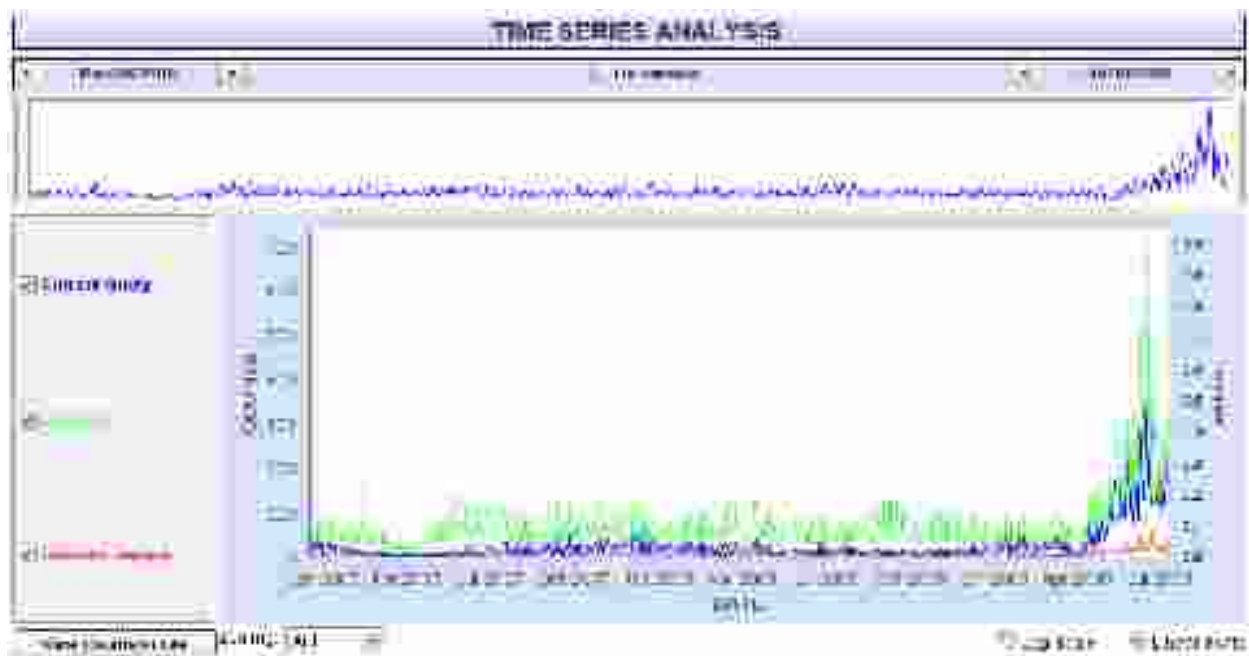


Figure 87: Visualization of time series.

The user can change the temporal range of time series currently being viewed in the main window by either changing the dates on the top of the window (Dec/16/2006 and Jul/10/2009 in Figure 87), or by dragging with the mouse the bars right below the dates. The upper smaller time series window always displays the full temporal range of the cumulative daily counts of all records in the current data. The movable bars in it depict the current selection of the temporal field of view of data being displayed in the main window. The users can fix the current width of the temporal field of view by checking the **Fix window** box. When it is checked, the field of view can be dragged left and right by dragging the right bar with the mouse.

The users can save the time series currently in the field of view to a local csv (comma-separated-values) file using **Save Counts to File** button. Saved files can then be further processed by external software. An upcoming revision of TCWI will enable the users to look up the raw data and save its subset corresponding to a query to a disk file. It will allow performing follow-up analyses outside of TCWI.

The appearance of the individual time series plots can be adjusted by mouse-clicking on their respective names in the legend. The users can change the color, thickness, type, and stroke of each plot (Figure 88).

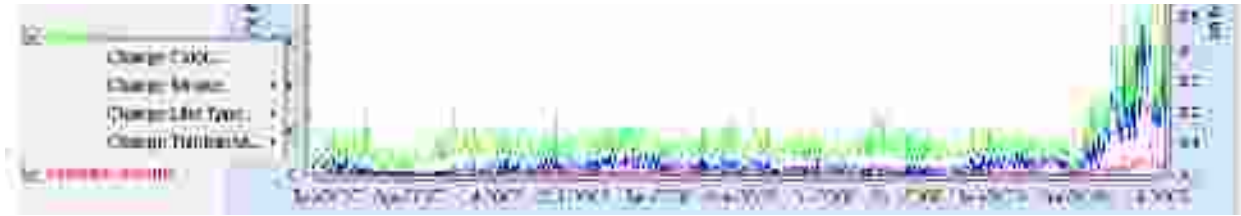


Figure 88: Changing the appearance of the time series plots.

12.4 Analysis Panel

This panel provides an interface with a number of the univariate statistical analysis algorithms that can be applied to any of the currently visualized time series. The currently implemented analytic algorithms are briefly explained in the following subsections.

The **Analysis Panel** can be activated by selecting one of the time series listed by name in a small window under the **Select Target** heading (Figure 89). The panel implements a dialog allowing the users to:

- Select the time series for analysis (from the list under **Select Target** header),
- Select the method of analysis (**Choose Method**),
- Set parameters of the analysis (the set of parameters is specific to each particular algorithm),
- Indicate whether the result should be stored separately, or the analysis is just an update of prior result (**Create new vs. Update**),
- Assign or reassign a name to the result of the analysis (**Series name**),
- Select the option to automatically update the result upon a change to the target time series (**Auto update on target change**).

The Analysis Panel also allows the users to remove the selected time series from the list of currently considered and visualized series by using the **Delete Series** button, or to change its name (**Rename** button and text entry window). Note that the **Current Query** and **All Data** series cannot be deleted or renamed.



Figure 89: Analysis Panel

12.4.1 Time Series Modeling and Forecasting Functions

TCWI currently implements a small set of select univariate forecasting functions. They are useful in smoothing noisy data, and in predicting the counts of disease that should be observed on a given day using historical data. A forecast can be useful under assumption that the modeled processes are predictable and that the selected forecasting algorithm can capture their dynamics. In health surveillance applications, forecasting functions are often used to compute baselines for detection of statistically significant discrepancies between the predicted and actually observed counts of disease cases.

12.4.1.1 Moving Average

Moving average is a simple and popular method of smoothing “bumpy” time series. Given e.g. a time series of daily counts of disease cases selected for smoothing, for a search time step (each day in our case) the moving average algorithm computes the average daily count over the past few days (excluding “today”). The extent of smoothing can be controlled by choosing the number of days to average over (**Window size**). Note that in most practical applications related to public health it is advisable to select window sizes among multipliers of 7 days. That accounts for typically strong day-of-the-week effects present in data.

The **Window size** is the only parameter of the basic moving average algorithm. The basic algorithm is offered as default among other variants and named **Retrospective (Default)** in the **Estimation type** pull-down menu.

Alternatively, the moving averages can be computed for each day using only data observed on the same weekdays from the past (that is, if the analysis is performed for a Tuesday, only data from Tuesdays in the past will be considered in computing the forecast for this Tuesday). That can be accomplished by selecting **Moving Average (day of week)** from the **Estimation type** list. In this context, the **Window size** parameter still determines of how many such days from the past should be taken into consideration in computing the daily moving average estimates. This algorithm is often useful in practice to model longer-than-a-weeklong trends in data that is subjected to a strong day-of-the-week bias.

12.4.1.2 Regression

The forecast produced using the day of week algorithm explained above can be correct if the modeled time series is stationary (that is, if its daily counts distribution does not change over time, except due to random fluctuations). Running the moving average algorithm using **Regression (day_of_week)** estimation type allows to account for linear non-stationarities (trends) in data. For each day, it takes the daily counts from the number of past identical weekdays (specified under **Window size**), fits a linear regression model to this data, and sets the today's value of the resulting time series to the value predicted from that model. Running the moving average algorithm using **Regression (day_of_week, last_week_mean)** extends that approach to take into account the mean count of the last week of data as a separate covariate in the linear model.

Note that both moving average and regression-based algorithms can be used to forecast expected counts of disease cases pertinent to time series selected by the users for monitoring. In turn, these forecasts can be compared against the currently observed actual volumes of disease cases to verify if they may be significantly excessive with respect to expectations, and therefore indicative of a potential disease outbreak. The extent of discrepancy between the actual and forecast daily counts can be easily computed using time series arithmetic operations (Section 2.3.2) by subtracting the original target time series from the time series resulting from application of one of the moving average algorithms.

12.4.1.3 Moving Sum

Moving Sum algorithm is analogous to the above-described moving average, except that daily counts selected for consideration are simply added together instead of being averaged.

12.4.1.4 Linear Trend

Linear trend method fits a linear regression model to the target time series. Users can optionally use day of week feature to compute a set of seven models, each independently dealing with data from one day of the week. The users can also de-trend their time series by computing the linear trend and subtracting it from the target time series. A de-trended time series is then created and it can be used in further processing.

12.4.1.5 Arithmetic Operations

The users can perform fundamental arithmetic operations (add, multiply, divide, and subtract) on any two target time series. The result becomes a new time series added to the Time Series Analysis panel. An example use scenario of this functionality is to compare results of a forecast against the true data. To experiment with that, create a forecast time series using moving average function with an estimation type of your choice. Then, apply **Subtract** operator selecting the forecast series as the **Target** (this can be accomplished by clicking on the forecast time series name on the list under **Select Target** header), and choose the original time series for which you have computed the forecast as **Target 2**. Hitting the **Submit** button will produce a time series of the daily forecast errors.

12.4.2 Temporal Anomaly Detection Functions

12.4.2.1 Temporal Scan

Temporal scan is a bi-variate algorithm for detecting anomalies in time series. It therefore requires the users to specify two time series as inputs: the **target** and the **baseline**. It is sensitive to statistically significant changes in counts of target time series that cannot be statistically explained by the corresponding changes in the baseline.

Fundamental procedure of the temporal scan algorithm executed for one period (day) of analysis aggregates four sets of counts. One set is the sum of target time series counts observed during the period identified as current (user-selectable **Temporal scan window size**). The other set is the sum of target counts corresponding to the period of reference. The remaining two aggregates are the identical counterparts computed for the baseline time series. The results of aggregation can be put in a 2-by-2 contingency table such as the one shown in Table 1.

Table 77: Example of a contingency table obtained for a single test in the Temporal Scan.

<i>Counts</i>	Current	Reference
Target	23	847
Baseline	95	11,550

Table 1 shows example results of aggregation of target and baseline counts for one day of analysis. Looking at the proportion of target and baseline counts observed during the period of reference, and comparing it to the same proportion observed currently, one can notice that currently it is substantially elevated. TCWI uses Fisher’s exact test of significance to quantify the extent of surprise in the observed increase. The result is a p-value – an estimate of the probability that the observed counts can be a result of a random fluctuation of data. The lower the p-value, the slimmer the chance for the observed elevated current counts to be just accidental. The p-value computed for data in Table 1 is close to 8×10^{-8} , a very small number indicating that the observed increase is highly significant and very hardly explainable as an effect of the random chance.

Temporal scan algorithm executes the above-prescribed procedure independently for all days in the scope of analysis. The interface offers a few different methods of computing the reference counts (selectable from the **Estimation type** pull down list). Most **Estimation type** options have been introduced above in Section 2.3.1 in the context of forecasting functions. The **Retrospective (Default)** method uses all the data outside the “Current” period as “Reference”. The **Prospective** method aggregates as reference the counts over the number of days, specified in the **Estimation window size** dialog, that immediately precede the current window (excluding the current window). Forecasting-like estimation types explained in Section 2.3.1 are also allowed as the methods of reference counts estimation in temporal scan. In addition, the **Univariate** variant does not require separate time series to be designated baseline. Instead, temporal scan uses the global average of target series counts as the outside counts of reference, turning the algorithm into a univariate anomaly detector. It may become handy when the baseline events are rare, their frequency not exceeding the counts of the target series, and the attainable reference information may therefore be rendered unreliable.

The users can execute temporal scan aiming at any kind of departure from expected, or strictly constrain the detector to alert for either increased or decreased counts. That is controlled with the **Scan option** through which the users specify whether to use a **Two-sided** test (detects all discrepancies, without differentiating between increases and decreases of activity), or **Upper tail** (focuses on monitoring for increased activity), or **Lower tail** (targets decreases in activity).

Users can also define the threshold of sensitivity of the alerting procedure. It is set to 0.05 by default, but it can be manually adjusted. The days for which the p-value obtained from temporal scan procedure was lower than that threshold are considered days of alert. The users can also apply False-Discovery-Rate (FDR) algorithm to automatically adjust the sensitivity threshold, guarding against the multiple hypothesis effects. The FDR-selected thresholds are usually more conservative than the manually selected ones, leading to fewer alerts being raised. Note that FDR does not affect the individual p-values computed using the Fisher's test. It only selects an alternative threshold to pick potentially fewer of results from the top of the ranked list of sorted from the most to the least statistically unusual.

Figure 90 shows the results of executing the temporal scan algorithm using the familiar *colombo-dengue* as the **target** and *all-values* as the **baseline** time series. We expect it to produce low p-values on days when temporal changes in the number of dengue fever cases reported in Colombo region could not be explained by similar fluctuations in all kinds of disease activity recorded across the country of Sri Lanka (other scenarios may involve comparing Colombo cases of dengue against the counts of all other reportable diseases recorded in Colombo, or the Colombo dengue activity against the temporal distribution of country-wide aggregates of dengue cases).

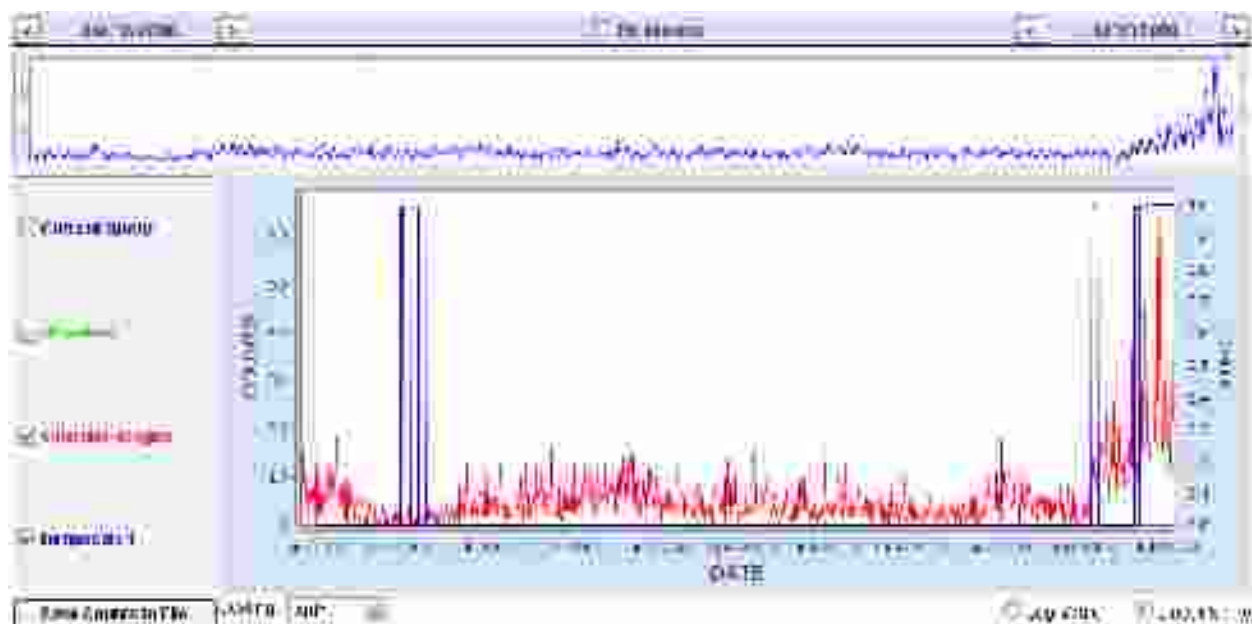


Figure 90: Example result of executing temporal scan on Colombo dengue fever data.

The result of executing temporal scan is a time series of p-values. After completion of the computations, this time series is automatically visualized in the time series visualization window. The plot depicts complement of p-values ($=1.0 - p\text{-value}$) for all days on which the resulting p-value was lower than either the user-selected or, when activated, the FDR-based sensitivity threshold. In Figure

89, days like that are marked with blue spikes, and there is no signal shown on days when the p-values were greater or equal to the threshold. It is often convenient to toggle the display of the p-values to the logarithmic scale to see more minute differences between the individual alerts.

Here are a few recommendations for setting temporal scan parameters. Temporal scan window size should be set to a multiplier of 7 days to remove weekly trends from data. Until substantial amounts of historical data is collected, we recommend using Prospective estimation type with estimation window size set to 35 days (5 weeks) and temporal scan window size set to 7 days. Longer scan windows enable detection of slower growing or generally longer-term problems. Longer estimation window sizes lead to taking into account more historical data when making forecasts. That will be a desirable strategy when the data collection process and therefore the data quality have stabilized. To reduce the false positive rate, always apply FDR correction. Given that the most common use of TCWI analytics would involve detecting unusual increases in disease case counts, we recommend performing primarily the upper tail tests. The lower tail tests would detect unusual decreases in certain disease counts. That may be useful in monitoring effects of responses to public health threats or in evaluating effects of longer-term policy changes. . Also, use the last day scan option to find alerts related specifically to the latest date in the data.

12.4.2.2 CUSUM

CUSUM (Cumulative Sum) control chart is a popular method for monitoring temporal processes primarily for changes in level. It is robust against short-term and presumably random fluctuations. It is being widely used in the practice of disease surveillance. TCWI implements the basic uni-variate CUSUM algorithm. For each consecutive time step, it computes the excess of target time series counts over the expectation. The expectation is computed over the number of past days specified using **Cusum Period (in days)** dialog, as a sum of the mean counts over that period. If the counts observed today exceed the sum of mean and K standard deviations (K is the tolerance multiplier setting specified by the user), the cumulative sum is incremented with the value of the observed difference and the algorithm moves to processing the data from the next time step. Initially, the cumulative sum is set to zero. Whenever the cumulative sum exceeds a user-selected threshold, H standard deviations, and whenever the sum goes below zero, it is set back to zero, and the accumulation process starts over at the next time step. As a result, TCWI currently produces and displays a time series of the cumulative sum. The upcoming version of software will also produce a time series of alerts marking the days when the cumulative sum exceeded the running threshold. The optimal settings of the CUSUM algorithm parameters depend on the amount of variance in data and of desired sensitivity. Greater values of the tolerance multiplier and threshold lead to fewer alerts and therefore to detecting only more spectacular changes in the level of counts of disease. Selecting longer periods of aggregation (via estimation window parameter) typically leads to more stable estimates of variance used to determine the running threshold and tolerance, but that may also lead to undesirable suppression of sensitivity of CUSUM to temporally local fluctuations of variance in data. It is advisable to experiment with a few different sets of parameter values to empirically determine the most suitable set point.

12.4.2.3 Change Scan

The Change Scan method is similar to temporal scan in that it uses the 2-by-2 contingency to

accumulate counts of disease cases and to evaluate statistical significance of the observed departures from the expected. The main difference is that Change Scan compares counts observed before and after the day of analysis, while temporal scan focuses on counts “inside” and “outside” the current period of analysis. Therefore, the change scan is well suited for detecting days of change of the level of the monitored time series, similarly to CUSUM algorithm. However, unlike CUSUM, Change Scan is bivariate – it allows using baseline time series as the reference of comparison against the monitored target. It also allows for estimating reference counts in various ways, accessible through the **Estimation type** selector, and it can be set to focus on detecting either positive or negative changes, or both kinds of them, analogically to temporal scan.

12.4.2.4 Peak Analysis and Range Analysis

Peak analysis method can be used to explain peaks observed in the target time series. User specifies the peak date and the algorithm screens the data for dimensions and their specific values, which jointly contribute to at least 90% of the counts observed in the peak. It is often useful in explaining characteristics of data that might correspond to a disease outbreak.

Range Analysis performs similar computations, but with respect to a range of consecutive dates, instead of a single “peak” day.

12.4.2.5 Massive Screening Panel

The above-described forecasting and anomaly detection methods are useful and applicable when the users know a priori which specific time series queries are of interest. This includes the above examples of monitoring of dengue cases in Colombo. However, when dealing multidimensional data, the number of possible unique projections of it onto subsets of dimensions and subsets of their values (and therefore the number of the extractable time series), can be very large. Analyzing them one-by-one in a sequence could become a quite tedious task.

Massive screening approach applies to exactly such scenarios in which the analysts need to concurrently monitor multiple time series. It exhaustively checks all individual time series resulting from conjunctive queries fitting in the user-selectable scope of search, and reports the findings in the form of a list of time series sorted according to the statistical significance of temporal anomalies found in them. The elements of the resulting list are clickable links to results visualized in the Time Series Analysis window. The suggested usage pattern is to first execute a massive screening of the desired subset of data, and then to inspect results starting from the most statistically surprising, one-by-one, possibly drilling down the data for further explanations using interactive visualization capabilities of the Query Selection Panel, or the explanatory analysis functionality such as Peak or Range Analysis functions described above.

Massive screening provides the analysts with the ability of comprehensive monitoring of multidimensional data without having to make many assumptions regarding the expected impact of anticipated disease outbreaks. It allows for bringing their attention to specific patterns in data such as subpopulations of patients who appear in recent data at unusually high frequencies, which might

otherwise go unnoticed.

Implementation of massive screening in TCWI uses the previously described methods of temporal scan and change analysis as the core anomaly detection methods.

The users select **Start Date** and **End Date** parameters to specify the period of analysis, and **Scan window sizes** to specify temporal granularity of the “Current” windows in the scan. **Pvalue threshold** parameter is used to decide what level of statistical discrepancy warrants an issuance of the alert. The higher the p-value threshold, the higher the sensitivity of event detection procedures, and the more alerts would be generated. The lower the p-value threshold, the more conservative filtering of the results, and the fewer the alerts. **Estimation type** selector allows the users to define the method of estimating “Reference” counts in the 2-by-2 contingency table used to determine p-values. **Scan Option** allows the users to select the type of the significance test (two-sided vs. one-sided upper- or lower-tail). **Last Day Only** flag is used to complete the analysis only for the selected **End date**. This can be used for prospective surveillance when the users are only concerned if there are any alerts today. Current implementation of the massive screening algorithm allows the users to pick up to three dimensions and any subsets of their values for screening. The algorithm will try every query within the scope of this selection, derived as a conjunction of individual values taken from one, two, or three dimensions (if the users selected three attributes for screening).

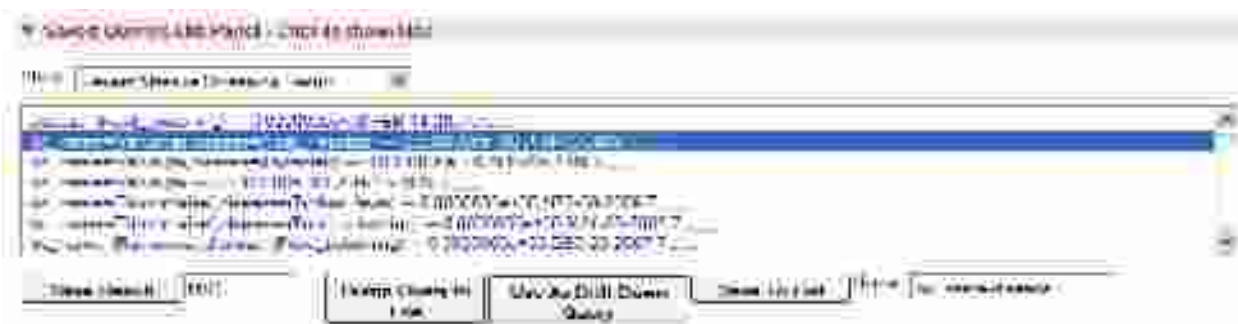


Figure 91: Saved Queries List Panel

There are a few additional parameters that could be applied to control massive screening algorithm, but they are not fully applicable to the current shape and contents of the RTBP datasets. We will keep monitoring their utility as more data becomes available and either include full description in the next revision of this manual, or eliminate the corresponding functionality from TCWI.

Figure 90 shows the result of running temporal scan using default parameter values, except for the **Scan option** set to **Upper tail**, and selecting “loc_name” and “disease” as dimensions for analysis. The list appears under **Saved Queries List Panel** described below. The top alert is for disease “Food_poisoning” on February 14, 2007. The p-value of the Fisher’s exact test conducted for that day was lesser than the numeric precision of the computer representation. Equally extremely significant score was associated with a few more results shown on the top of the list.

The second result shows an alert about Viral_hepatitis in Vavuniya that would have been issued on June 10 2009. Clicking on it brings up the corresponding data to the time series visualization panel.

Figure 91 shows the screenshot presenting that result as well as the complete set of parameters used to arrive at it. The time series legend has been expanded by adding three items: **MS Baseline** is the temporal distribution of the data corresponding to query used as the baseline in the temporal scan procedure (in this example, we chose **All Data** to serve as the baseline, but the users can select any other baseline query to use in massive screening). The **Result Query** depicts time series of the target that is counts of viral hepatitis cases reported in Vavuniya. The 7-day long temporal scan window corresponding to the alert of June 10 is highlighted with yellow background. Zooming in onto the period of interest, we can see the results in more detail (Figure 92). We can see the number of the hepatitis cases ramping up significantly in early June.

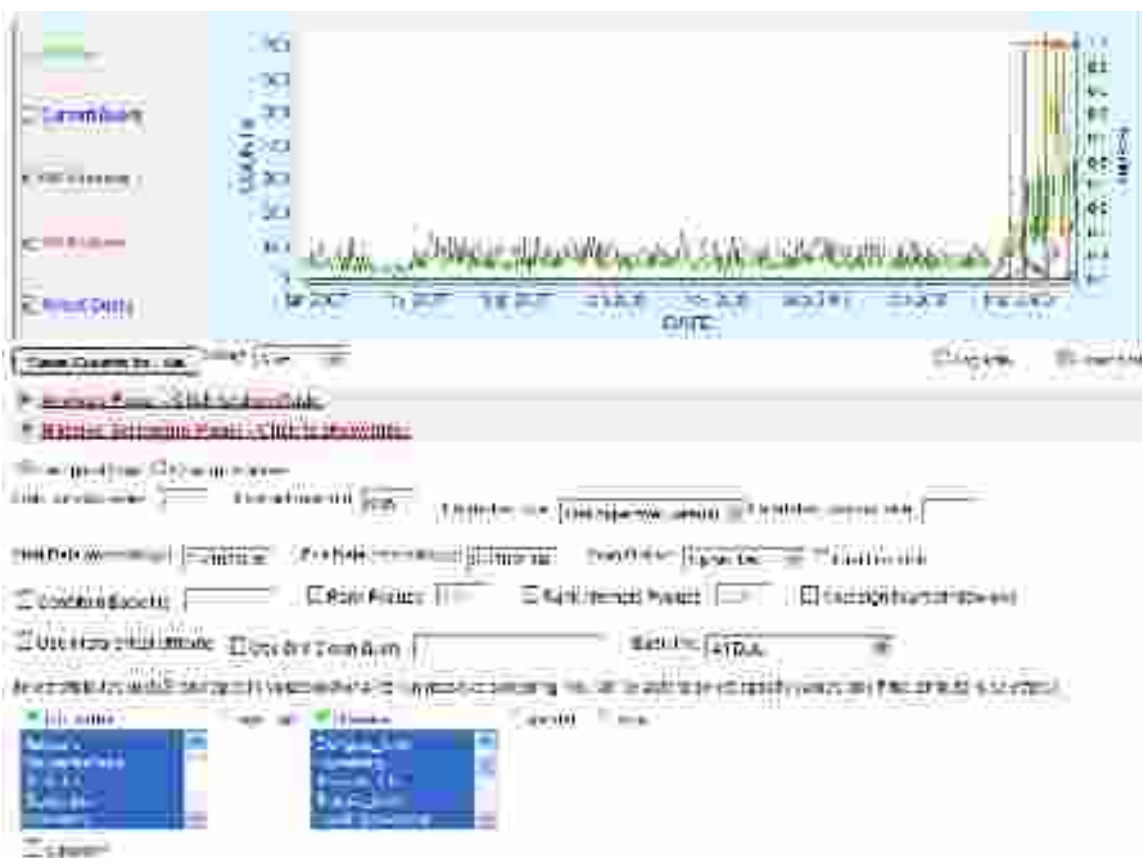


Figure 92: Inspecting one result of the massive screening with temporal scan.

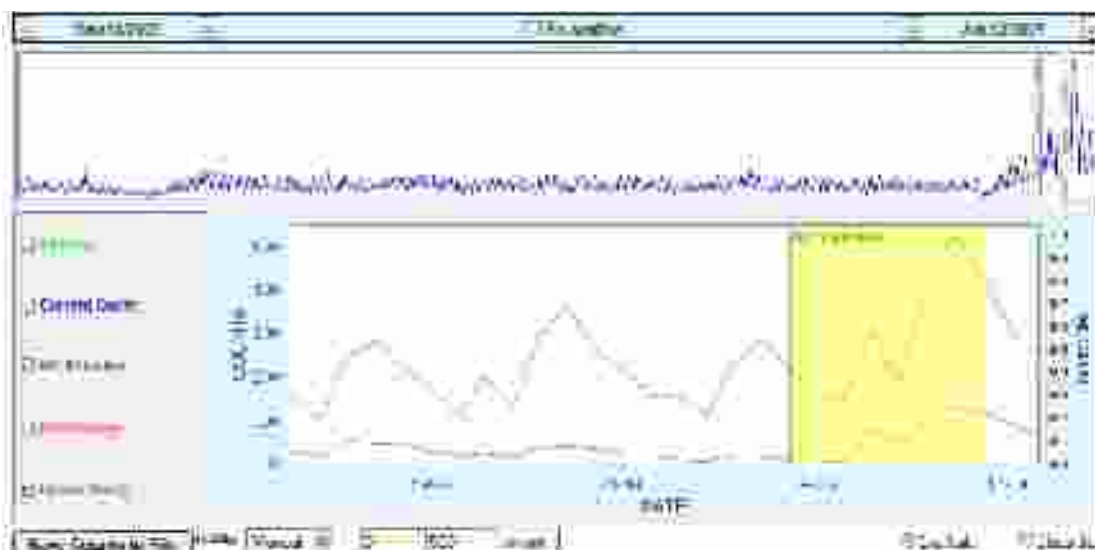


Figure 93: Zooming in on the results from Figure 10.

12.4.3 Saved Queries List Panel

This panel allows manipulation of the results generated by massive screening. **Save To List** button saves the results for future retrieval with a specified name. **Save Query** button saves the query as a .csv file on a local disk. **Use As Drill-Down Query** button allows to stage the next run of massive screening within the scope of data corresponding to the specific item on the finding list. If we do that with our hepatitis in Vavuniya finding, the **Use Drill-Down Query** selector in the Massive Screening Panel will be automatically populated with the query corresponding to our finding. Now, let us unselect the disease and loc_name dimensions from the massive screening list, and instead select age_grp and hit the **Run Screening** button again. Now, the algorithm will screen through all age groups of reported hepatitis cases from Vavuniya and sort them by the most statistically significant increase in the corresponding patient counts. Figure 94 shows the time series of patients over 45 years old, which happens to be the age group most significantly affected in this outbreak of viral hepatitis in Vavuniya region.

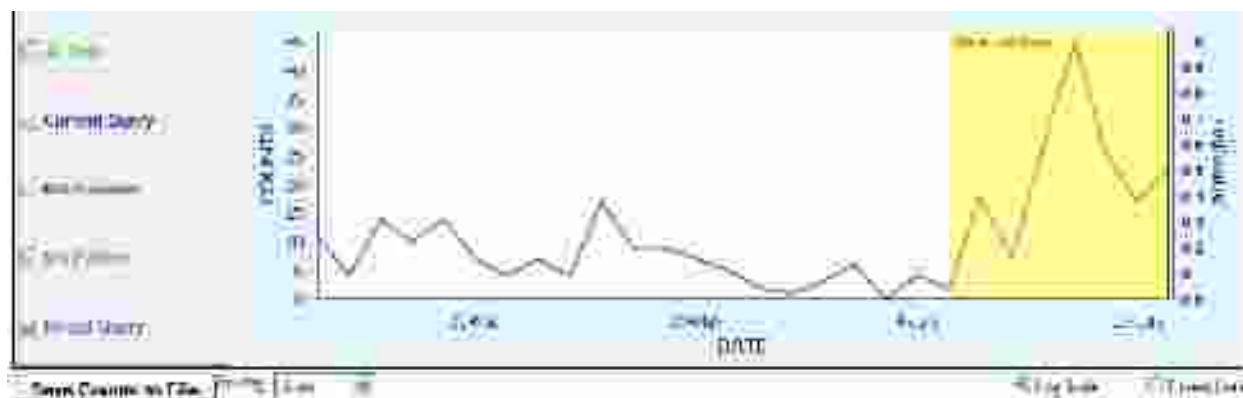


Figure 94: Drill down using another run of massive screening reveals that people older than 45 years seem to be the age group most spectacularly affected by the June 2009 viral hepatitis outbreak in Vavuniya

12.4.4 Spatio-Temporal Analysis

The **Maps** component of TCWI can be used to detect and analyze spatio-temporal patterns in data. **Maps** panel can process data that contains a spatial attribute. In the discussed `lk_flat_table` data, it is named “`loc_name`” and its values correspond to individual regions of Sri Lanka. In order to initialize the **Maps** panel, specify the spatial attribute name and click on **Load Map** button to view the map. Once the map is loaded, **Clear Map** resets the **Maps** panel.

12.4.4.1 Map Visualization

Figure 95 shows the geographic distribution of all disease cases over the complete period of time covered in the `lk_flat_table` data on the background of the Sri Lanka map. Figure 96 presents the corresponding temporal distribution of the visualized disease cases. Center of each circle corresponds to a Sri Lanka region, while its radius indicates the total number of patients for the date range shown in Figure 96. Try changing the date range using slider bars in time series window. Rendering of data on the map will be updated to reflect the volume of disease reports per region corresponding to the adjusted temporal scope. The legend in the bottom left part of the map display window presents the size to volume correspondence used in the map diagram.

12.4.4.2 3.2 Time Series Visualization

Temporal distribution of data is shown right under the map display. Its operation is analogous to the same panel under the Time Series tab. The green line shows the 7-day moving average. Animation is a new function that allows the users to watch spatial and temporal changes in the displayed data. Adjust values of **Scan window size** and **Slide window by** parameters (Figure 96) to adjust the resolution and speed of animation.



Figure 95: Map visualization.

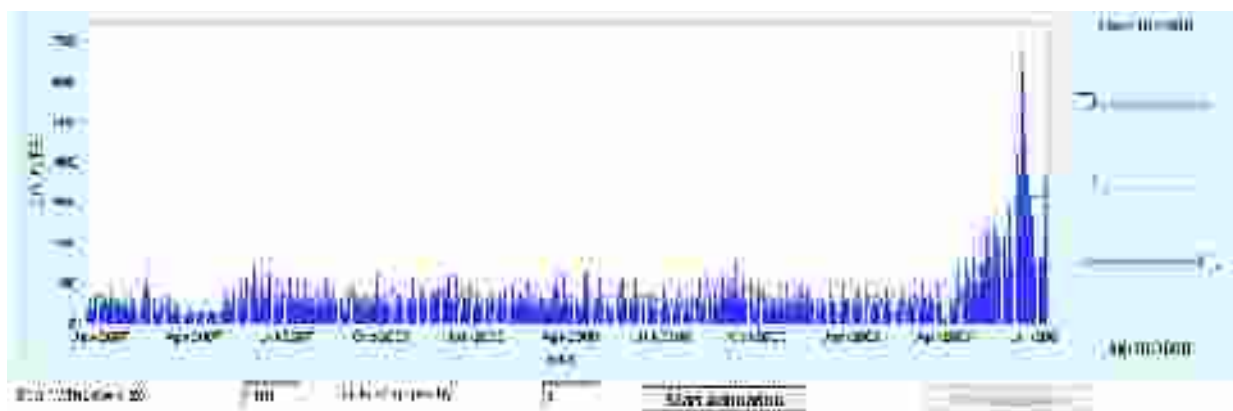


Figure 96: Time series view under the Map tab.

12.4.5 Attribute Selection Panel

The time series in Figure 96 shows the total daily counts of all disease cases. To display the time series corresponding to specific values of specific dimensions of data use the **Attribute Selection Panel**. Functionality of this panel is generally similar to that of **Query Selection Panel** described before. Activating of the **Auto Filter** modifies the dialog of the user-driven dimension-value selection such that only the actually present in data combinations of dimensions and their values are available for selection, given the already selected subsets. Selecting one of the dimensions and hitting the Submit/Reset button leads to splitting the currently visualized data into time series and circles on the map, separately colored for each of the individual values selected for this dimension.

Figure 99 shows how to display time series corresponding to *Dysentery* patients aging up to 5 years old recorded in regions of *Badulla* and *Batticaloa*. Note that after selecting the values, the user must press “Submit/Reset button” to update the time series and map visualizations.

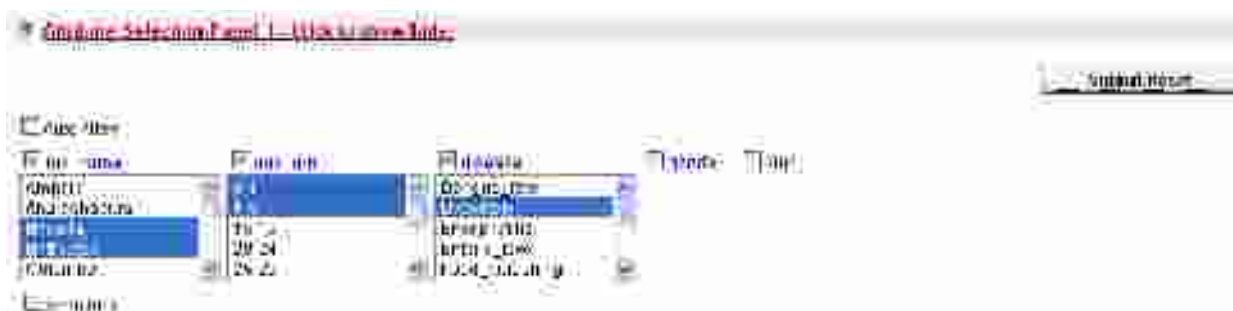


Figure 97: Attribute Selection Panel under the Map tab.

12.5 Spatial Scan

TCWI implements a variant of the Multivariate Bayesian Spatial Scan algorithm for spatio-temporal analysis of public health data. This spatial scan method computes the overall probability of a disease outbreak anywhere in the scope of data selected by the user in the **Attribute Selection Panel** dialog, separately for each day in the scope. It is reported as the Spatial Scan Global score in the upper right

corner of the map display window. The displayed value corresponds to the current last day of analysis. The users can inspect the specific values of the global score for other days by moving the end date slider bar located to the right of the time series window. The daily global scores are visualized with a red line in the upper part of the time series diagram.

The spatial scan algorithm also computes for each day, the probability of an outbreak occurring at each geographic location. The results are visualized on the map as circles filled with a color depending on the value of the respective probability estimate.

Figure 98 shows the dialog for setting the parameters controlling the spatial scan algorithm. **Temporal window size** acts as a smoothing parameter and it must be set to a value lower than seven. This value should reflect the expected disease outbreak duration. **Max group size** is the anticipated maximum number of regions affected by an outbreak. Based on initial experiments performed using *lk_flat_table* data, we recommend setting the **Temporal window size** to 7 and the **Max group size** to the value approximately equal 25% of the total number of distinct locations represented in the data.



Figure 98: Selecting the Spatial Scan parameters

Figures 14 and 15 show results of running the spatial scan against counts of Leptospirosis cases using the end date of August 5, 2008. The global score (0.9612) in the top right of Figure 99 indicates that the chance of a Leptospirosis outbreak anywhere in the country on that day exceeds 96%. The shaded circles centered at each region (heat map legend is placed at the top left side of Figure 99) depict spatial distribution of outbreak probability. The blue circles still indicate the total number of Leptospirosis cases in each region. In Figure 100, the red time series plot depicts the temporal distribution of past global. Once the review of results of the spatial scan run is complete, the users can **Clear Scan Results** using the button shown in Figure 98.



Figure 99: Detecting an outbreak of Leptospirosis with Spatial Scan (map view).

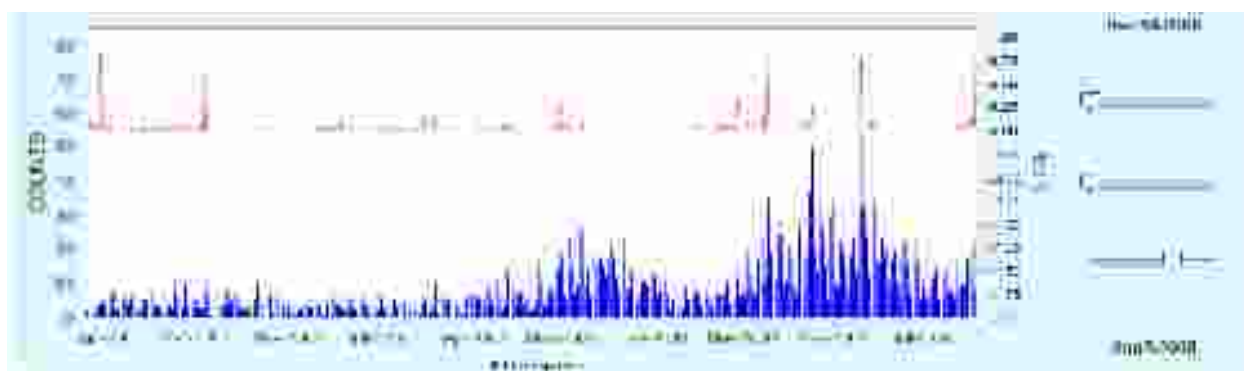


Figure 100: Detecting an outbreak of Leptospirosis with Spatial Scan (time series view).

12.5.1 Summarization of Data with Pivot Tables

Pivot Tables tab offers an interactive data summarization capability. TCWI provides an efficient algorithm for aggregating multidimensional data of counts of events (such as the numbers of reported disease cases) into a two-dimensional matrix view. Figure 101 depicts an example pivot table computed for the `lk_flat_table` data using two dimensions: `disease` and `age_grp`. Each cell in the table shows the total number of recorded cases of specific disease (row) among patients in the specific age group (column). The table also shows the marginal sums (by row, by column, and the total number of cases).

	10/20	10/21	10/22	10/23	10/24	10/25	10/26	Total
Dengue fever	24	142	118	81	23	24	24	316
Typhoid	14	11	10	7	10	11	10	73
Paratyphoid	3	7	2	2	1	1	1	18
Malaria fever	2	22	4	22	4	2	2	57
Food poisoning	2	1	1	1	2	1	1	9
Shistosomiasis	1	1	1	1	1	1	1	7
Leptospirosis	4	22	4	22	4	2	2	58
Typhus fever	1	1	1	1	1	1	1	7
West Nile virus	6	17	10	17	6	7	7	70
Total	57	247	152	157	67	67	67	1027

Figure 101: An example 2-way Pivot Table

Construction and editing of pivot tables can be done by dragging and dropping icons with the individual dimension names between the “Attributes,” “Rows” and “Columns” lists. The dimensions in the “Rows” and “Columns” lists are used to create rows and columns of the table, respectively, while those that remain in the “Attributes” list are ignored. Multiple attributes can be put in “Rows” and “Columns” to create nested tables. To move from table shown in Figure 101 to the one shown in Figure 102, just drag the icon denoting “gender” to the “Row” list. Now, the rows are split by disease and gender, and each cell of the table contains counts of either male or female patients diagnosed with a specific disease, and belonging to a specific age group.

The users can always clean the current table and start over by clicking the **Start over** link (Figure 101). Left clicking on any data cell of the table will produce a pop-up window with a time series plot of the data represented by that cell. The temporal range of data can be adjusted by changing the dates at the bottom of the page. Figure 103 shows the time series for male patients aged 30-34 years old having been diagnosed with typhus fever.

		2011	2012	2013	2014	2015	Total
Income Group	Female	12	18	22	25	28	105
	Male	15	20	24	27	30	116
Employment	Female	10	14	18	21	24	87
	Male	12	16	20	23	26	97
Food possession	Female	8	11	14	17	20	60
	Male	10	13	16	19	22	60
Human capital	Female	5	7	9	11	13	45
	Male	6	8	10	12	14	40
Expenditure	Female	3	4	5	6	7	25
	Male	4	5	6	7	8	30
Total	Female	42	58	70	77	84	231
	Male	47	62	74	81	86	250

Figure 102: Example 3-way pivot table.

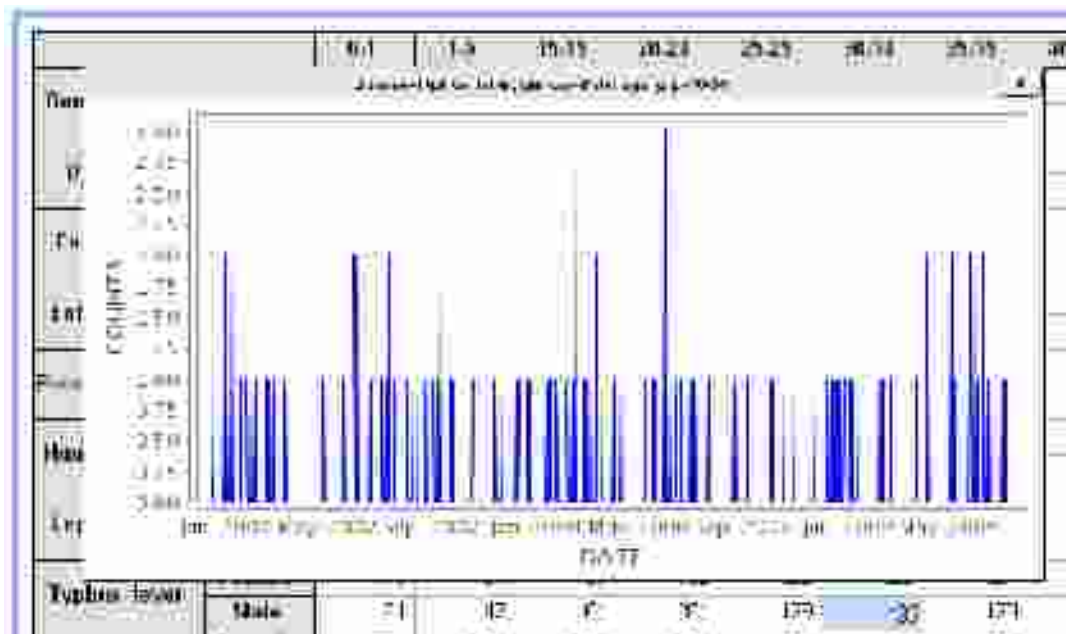


Figure 103: Example Time Series view under Pivot Table tab.

Left-clicking on any row or column name of a pivot table will produce a pie chart visualizing the distribution of data split according to the values of the counterpart row or column dimensions. Figure

104 shows the distribution of age groups for female patients diagnosed with dengue fever.



Figure 104: Example pie chart under the Pivot Table tab.

Each cell in a pivot table reflects the aggregate counts of disease cases computed over a specific period of time. It is shown and it can be edited at the bottom of the page. If either or both of the dates are changed, TCWI will quickly re-compute counts for all cells in the current pivot table.

12.5.2 Future Work

We have mentioned above a few minor extensions of the capabilities of TCWI scheduled for inclusion in the next release of the software (expected in mid-December 2009). The next release is also intended to include the following major extensions:

- Ability to browse the original transactional data. We are working on an interface that will allow users to see the individual original records of disease cases loaded to TCWI, besides the aggregated counts shown currently in the form of time series. This will support detailed drill downs and other basic database operations not easily accessible through TCWI.
- Background screening and alerting capability. Currently, TCWI offers a variety of analytic functions available for interactive use. The background screening and alerting will enable setting up scripts for user-determined analyses to be automatically executed at user-defined periods of time. It is of practical importance to execute screening for emerging patterns at regular intervals in order to maintain situational awareness among stakeholders of RTBP, even if the TCWI operators are not always present at their consoles to invoke the analyses manually. The results of scheduled analyses will be stored for inspection by the users, and the automatically generated alerts will be made available for distribution to the designated recipients.

12.6 Sahana Alerting Broker user manual

12.6.1 Introduction

This guide is intended for users of the Sahana Messaging/Alerting Module, who wish to send and process alerts based on the Common Alerting Protocol (CAP). This guide specifically covers the CAP alerting section. Those of you not familiar with the CAP standard should refer to the “[CAP Cookbook](#)”.

12.6.2 Messaging/Alerting Module

The Messaging/Alerting Module is a Sahana Module that is used for the sending and receiving of messages and/or alerts to recipients. At the time of writing, the module allows for the generic sending and receiving of messages as Short Messages via the Short Messaging Service (SMS), sending messages as Email, conducting SMS based surveys and sending CAP alerts. The CAP alerting section, which falls under the scope of this guide, is accessed via the Alert and Templates subsections.

We shall use an example of a *Government Health Department; specifically the Epidemiology Unit* use of the CAP Messaging/Alerting Module for the purpose of notifying disease situational reports.

Although not mentioned in this guide, it is possible to control the access of modules and functions through the *Sahana Security Administration Module* (but not mandatory and is at the discretion of the implementers) . Please refer to the user guide on the security administration module (abbreviated as secadmin): <http://wiki.sahana.lk/doku.php/doc:secadmin:english> .

The Sahana Administrator can setup user accounts, define roles, and authorization points through the secadmin. For the purpose of messaging/alerting, the security administrator should setup a *super-user* account and a *user* account. The super user's roles are mainly defining the CAP profile, establishing operational guidelines/policies, and setting up CAP message templates. The user's roles are issuing messages either directly or with the aid of a predefined template and monitoring responses.

In our example, health officials would be the main users and super-users of the messaging/alerting module. The alert recipients will be 1) Closed User Group (CUG) of health officials and 2) Public. The super users would be the National and Regional Epidemiologist, per say, and the users would be the epidemiology unit staff and District/Divisional Health Services office staff.

It is possible to issue public alerts but not efficient via the technologies such as SMS/Email that are used in the Sahana messaging/alerting module. Once a channel is established to issue Cell Broadcast (CB) alerts, for example, then authorized users may push public alerts directly through the messaging/alerting module to mobile handsets of the public. For the purpose of our example, we will assume the scenario where a public alert is generated but, first, sent to main-stream-media (like TV and

Radio) and Cellular Network Operators (GSM/CDMA) to rebroadcast the message over their networks to the public.

12.6.3 Templates

The system allows users to create CAP templates and store them in the system. These can then be used when creating CAP alerts/messages, which allows the message to be populated based on the relevant template.

12.6.3.1 Create Template



Figure 105: Create template

The user can create a new template, by clicking the Template-New menu item in the left menu bar of the Messaging Module. In the Basic Information section, a user can enter the name of template and save it via the 'Save' button. The user can then update the template by going to the Update section.

12.6.3.2 View Template List

The user can go to the View Template List section, by selecting Template->View in the menubar. This will take the user to a list of templates available in the system



Figure 106: View list of templates

When the user clicks on a template, it will show a 'View' screen for the selected template

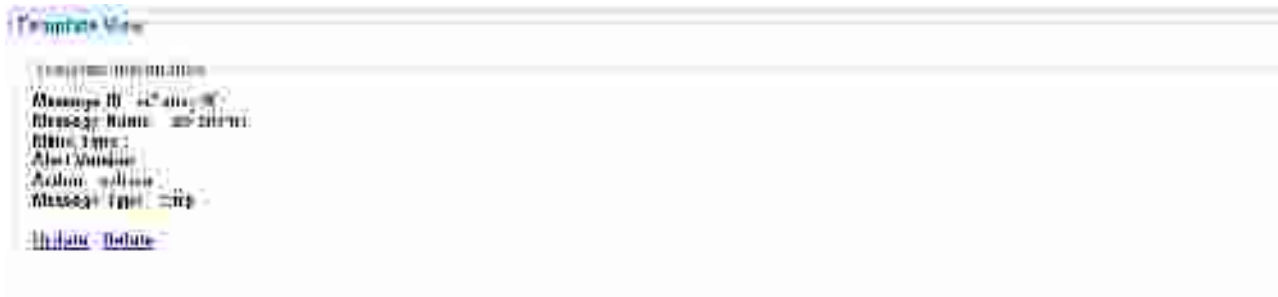


Figure 107: View template

In our example, we shall create two templates: 1) CUG health officials 2) public. Super-user should use steps 3.1 to generate two templates with the names: “CUG-health-dis-outbreak” and “Pub-health-dis-outbreak.” These would be listed in two separate rows in the table shown in Figure 106. If one was to click on the name “CUG-health-dis-outbreak” the next screen (Figure 107) would show messages message-id with a system generated unique identifier comprising a sequence of alphanumeric character, message name as CUG-health-dis-outbreak, and message type as “temp.” At this point if the super-user wants to change the name of the template, then the super-user should click delete to discard that template and repeat the step in section 3.1 to create a new template with desired new name.

12.6.3.3 Update Templates

The Update Template section can be viewed by clicking on the 'Update' link in the View Template screen, as described in section 3.2. This will take the user to a set of CAP fields, which can be filled in by the user. This is similar to the description of filling in a CAP alert, described in section 4.1 below. Once the user is done filling in the CAP Alert Template, he/she can save the Template by clicking on the Update button below.

The template can then be used to pre-populate fields when creating a New Alert, as shown in section 4.1 below. The difference between a message template and an alert message is that most of the CAP fields are not populated and some are set to indicate that it is a template such as the message status being set to “draft.” The unpopulated fields are specific to the issued message and must be populated at the time of issuing the message; for instance, the message identifier that refers to the particular message issued is auto populated at the time of creating a message from a template. There is no concrete rule etched in stone as to how implementers should define the set of elements that can be pre-populated in a template. Therefore, the software has opened all fields leaving it up to the implementers to set their own policies and standard operating procedures.

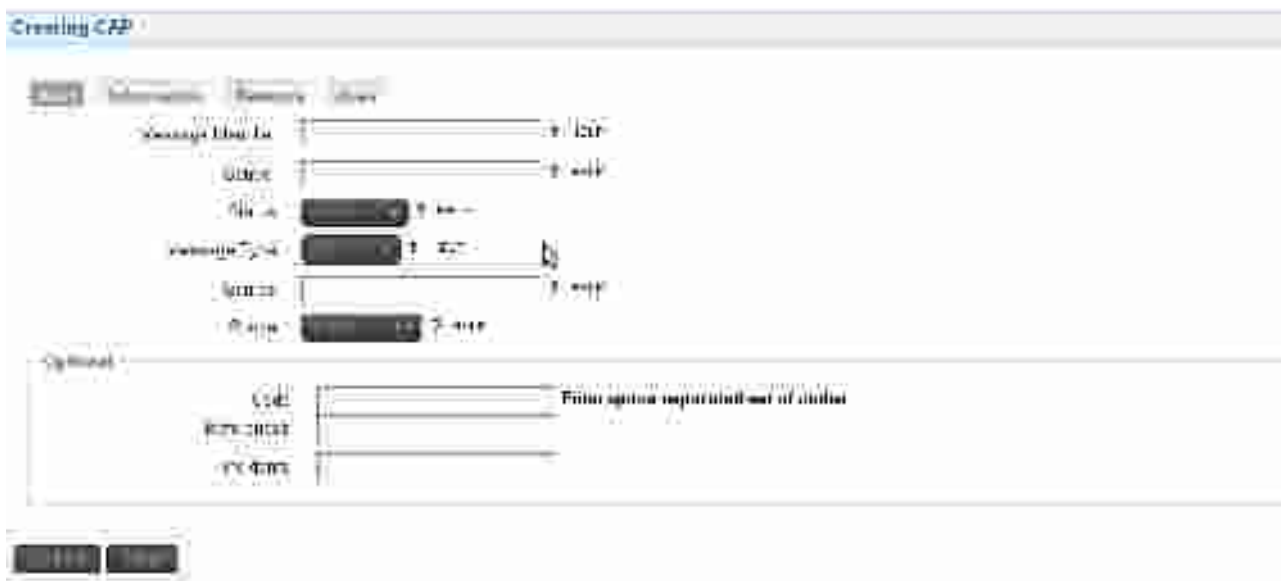


Figure 108: Upload template message

Follow the steps described in section 4.0 to enter the CAP template values and save. Here we will give you the values that would be entered in the templates for the two examples (see Table 1 and table 2). We leave all fields blank or to be populated at the time of issuing the messages and only enter the list of values given below

Table 78: “CUG-health-dis-outbreak”:

Alert : Status = “ Draft ”	Alert : Message Type = “Alert”
Alert : Source = “Epidemiology Unit”	Alert : Scope = “Restricted”
Alert : Restriction = “Health Officials/Workers”	
Info : Language = “English”	Info : Category = “Health”
Info : Event = “Disease Outbreak”	Info : Audience = “Health officials/workers”
Info : Description = “A disease outbreak has been issued for <areaDesc> by <senderName>. Health Officials and Health Workers in these areas are encouraged to <responseType> responsive actions and <instruction>. This event is rated with <Priority> priority and responsive actions should be taken <urgency>. For more information about this event visit <URI> or call <contact>.”	
Info : Instruction = “Contact the respective Regional Epidemiology Unit”.	Info : Headline = “A disease outbreak is in effect”
Info : Web = “ http://www.epid.gov.lk/restricted/alert/c ”	Info : Contact = “+941125551212”

ap”	
Resource : Resource Description = “restricted website”	Resource : url = “ http://www.epid.gov.lk/restricted/alert/cap/ ”

Table 79: “PUB-health-dis-outbreak”:

Alert : Status = “Draft”	Alert : Message Type = “Alert”
Alert : Source = “Epidemiology Unit”	Alert : Scope = “Public”
Info : Language = “English”	Info : Category = “Health”
Info : Event = “Disease Outbreak”	
Info : Description = “A <disease name> disease outbreak has been issued for <Area Description> by <Sender Name>. Public in these areas are encouraged to <Response Type> responsive actions and <Instruction>. This event is rated with <Priority> priority and responsive actions should be taken <Urgency>. For more information about this event tune in to your local TV and Radio stations or visit <URI> or call <contact>.”	
Info : Instruction = “Report any suspected cases to the nearest hospital or clinic”	Info : Headline = “A disease outbreak is in effect”
Info : Web = “ http://www.epid.gov.lk/public/alert/cap/ ”	Info : Contact = “+941125551212”
Resource : Resource Description = “public website”	Resource : URI = “ http://www.epid.gov.lk/public/alert/cap/ ”

12.6.4 Alerts

This module allows the user to create and view new alerts, and send them via SMS or Email.

12.6.4.1 4.1 New Alert

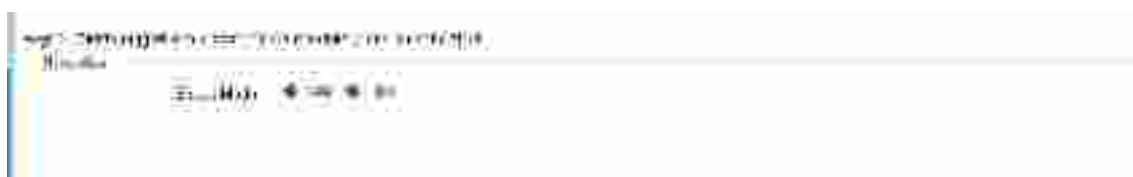


Figure 109: New alert: select mode

Selecting Alert->New allows the user to create a new alert. Alerts can be of type CAP or EDXL – currently CAP is supported. The type of the alert can be selected via the Select Mode radio button:

Once the mode is selected, the User can enter the name of the alert she wants to create, in the Name field (Figure 110).



Figure 110: New alert: enter name and type

Along with the name, the user can select to create a new alert from scratch, or use an existing template that was created in the Templates section. The user can do this by selecting either the New Alert or Existing Alert Template radio button respectively.

We may choose to use a naming convention that helps us identify the message in the future for any audits. Let us assume we want to send a public alert on dengue outbreak. We may use the following name: “pub_health_dengue_10_08_2009”. If more than one alert on the same dengue event is issued on the same day, one may choose to separate the names by add in sequence number 001, 002 to the end; e.g. pub_health_dengue_10_08_2009_001. This is not essential since the message identifier will vary between the two messages and one may distinguish them apart from message identifier field. The main purpose of the alert name is to help the user locate the message from the data base in the future; hence, a general name may be preferred.

For both types of alerts selected above, the User is given a set of Metadata fields that describe the alert, along with a unique system-generated Alert ID. The user can fill in the required fields and proceed by clicking on Next.



Figure 111: Alert metadata

The user is then taken to a Tab-separated New CAP alert Form. The form is separated into four tabs that have various form elements to build a CAP message. Form elements marked with a red asterisk are compulsory/mandatory elements, whilst all others are optional. Once the user is done filling in the fields, they would click the 'Update' button to save the CAP message, or the 'Clear' button to clear the filled elements.

If the user selected the Existing Alert Template button, then upon completing the fields in Figure 111 and clicking Next, the user will be taken to a screen listing all the template names, same as in Figure 106 – List View Templates. Once the user selects the desired template by clicking the appropriate name, they will be navigated to a screen shown in Figure 112 but with pre-populated values that were originally entered at the time of creating the template.

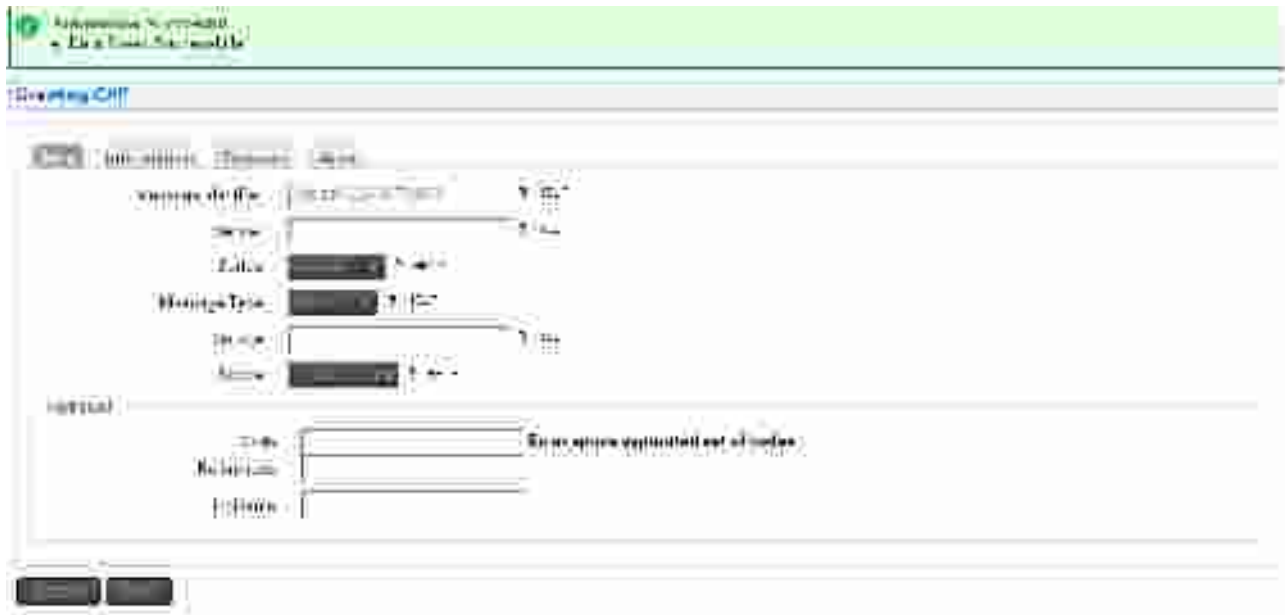


Figure 112: Create CAP alert tab

Some elements such as Status, Message Type and Scope, shown above, have a drop down list with pre-populated values.



Figure 113: Create CAP - Information tab

The Information Tab, show above, captures shows the Information related fields of the CAP message. The Priority drop down contains pre-selected values for the Urgency, Severity and Certainty fields, and will fill them in based on its selection. If the Priority field is set to Unknown, it will then allow users to select custom values for the Urgency, Severity and Certainty fields, from a list of pre-populated values available for each of them.

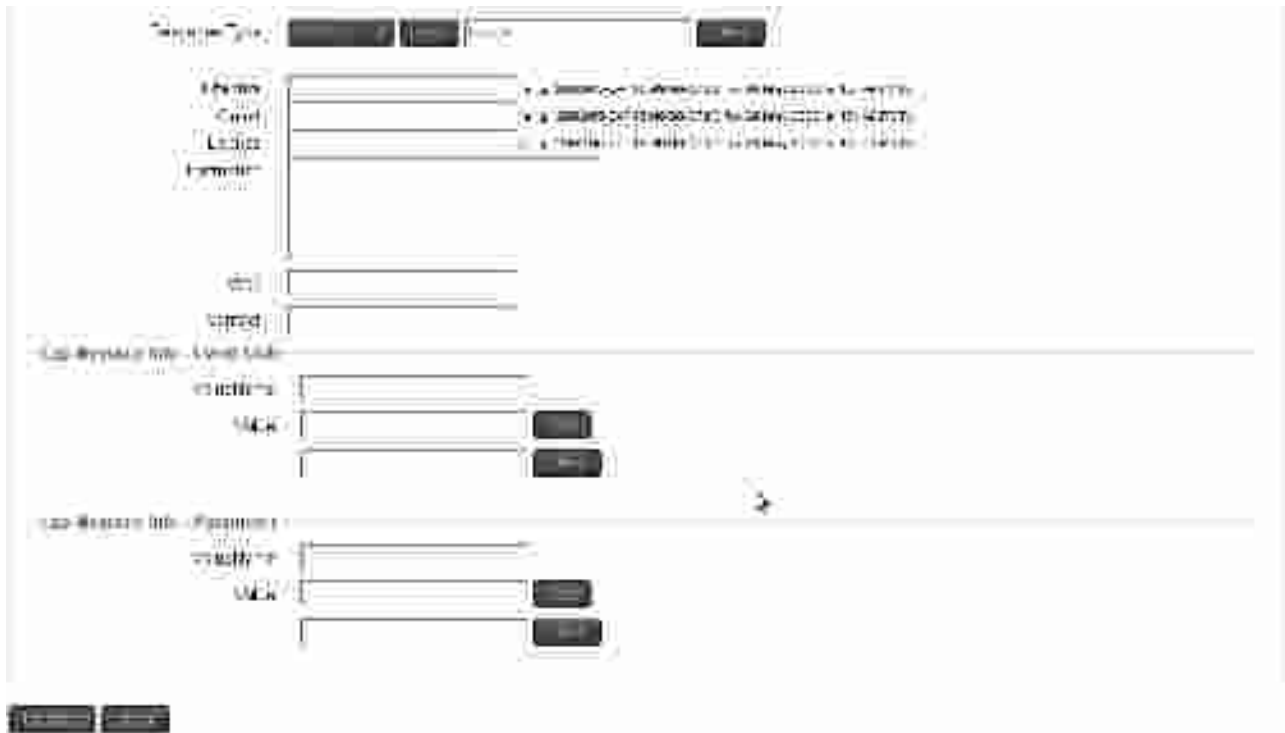


Illustration 114: Create CAP – Information tab showing message information

The Area page allows the user to add Area and Geographic information to a CAP message. When the user starts entering a location in the Area Description field, the system provides an in-place lookup of available locations already available in the system – the user can then select one of these locations to fill the field.

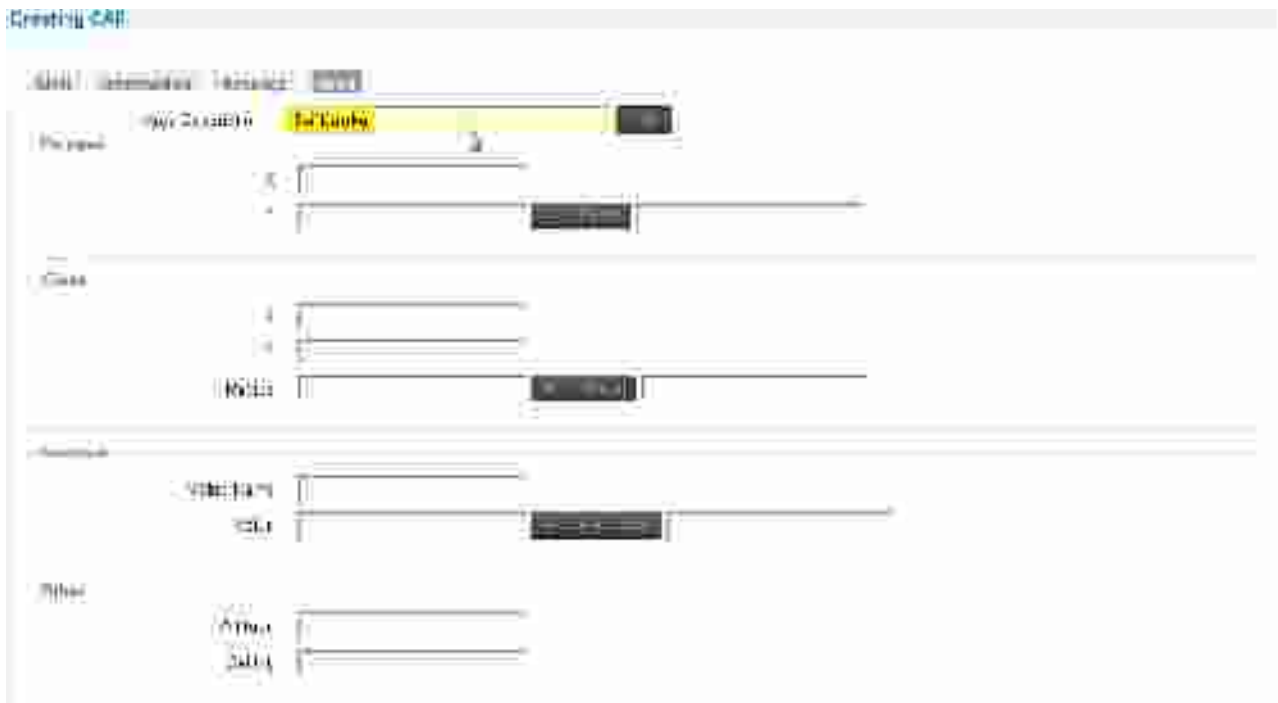


Figure 115: Create CAP - location lookup

If the location a user requires is not entered in the system already, then the user can add a new location by clicking the 'Add' button to the right of the Area Description box. This would allow the user to enter a location, which would then be added simultaneously to the Area Description field and the system location table as well.



Figure 116: Create CAP - add new location

Once the user is done filling in the fields available in various tabs, the user can click the 'Update' button, which creates the CAP alert.

Now to complete the message we would enter the following values. In Table 3 and Table 4 we specify the new values and the changes made to values in Table 1 and Table 2 such as completing the Info : Description field, while the other pre-populated values in the template will remain the same; hence, refer to Table 1 and Table 2 for those values. Fields such as the Message Identifier and Sent (submitted time and date) are generated by the system and the user is discouraged from changing those fields. Once again, this is up to the implemented operating procedures. Such values that are system generated are not listed in Table 3 and Table 4. Thus, Table 3 and Table 4 shows values that are entered or changed by the user.

Table 3 illustrates a message sent to all the Medical Officer of Health in Kurunegala District of Sri Lanka and Table 4 illustrates a message sent to the public in Kurunegala District of Sri Lanka, which will be given the names: MOH-health-dengue-10-08-2009 and pub-health-dengue-10-08-2009, respectively. Reader may wonder why we use “health” as part of the naming string, this is because other departments, such as the irrigation department for flood warnings, may use the same messaging/alerting module implementation; hence, the work “health” helps identify the category of the alert.

Table 80: “MOH-health-dengue-10-08-2009”:

Alert : status = “Actual”	Alert : Sender = “ratnayake@epid.gov.lk”
Alert : Incident = “health-20090810104000”	Info : Response Type = “Execute”
Info : Priority = “Urgent”	Info : Urgency = “Immediate”
Info : Severity = “Extreme”	Info : Certainty = “Observed”
Info : Sender Name = “Dr. Ratnayake, Provincial Director of Health”	
Info : description = “A DENGUE disease outbreak has been issued for Kurunegala District by Dr. Ratnayake, Provincial Director of Health. Health Officials and Health Workers in these areas are encouraged to Execute responsive actions and consult with their respective Regional Epidemiology Unit. This event is rated with URGENT priority and responsive actions should be taken Immediate. For more information about this event visit http://www.epid.gov.lk/restricted/alert/cap or call +941125551212.”	
Info : Effective = “2009-08-10T16:00:00+05:30”	Info : Expires = “2009-08-20T23:30:00+05:30”
Area : Area Description = “Kurunegala District”	

Table 81: “pub-health-dengue-10-08-2009”:

Alert : status = “Actual”	Alert : Sender = “ratnayake@epid.gov.lk”
Alert : Incident = “health-20090810104000”	Info : Response Type = “Monitor”
Info : Priority = “Urgent”	Info : Urgency = “Immediate”
Info : Severity = “Extreme”	Info : Certainty = “Observed”
Info : Sender Name = “Dr. Ratnayake, Provincial Director”	
Info : description = “A DENGUE disease outbreak has been issued for Kurunegala District by Dr. Ratnayake, Provincial Director of Health. Public in these areas are encouraged to Monitor and Report any suspected cases to the nearest hospital or clinic. This event is rated with Urgent priority and responsive actions should be taken Immediate. For more information about this event tune in to your local TV and Radio stations or visit http://www.epid.gov.lk/pub/alert/cap or call +941125551212.”	
Info : Effective = “2009-08-10T16:00:00+05:30”	Info : Expires = “2009-08-20T23:30:00+05:30”
Area : Area Description = “Kurunegala District”	

12.7 4.2 View Alerts

The view alerts menu item under Alerts, is the central point to View alerts and Send them. When a user clicks on the View Alerts menu item, he/she is presented with a list of Alerts in the system



Figure 119: Select contact

The user can pre-select multiple contacts from the Contacts tree - the selected contacts are shown in the Recipient List box. Contacts that are added via the Messaging Module's Add Contact will be visible in the Contacts tree. Alternatively, the user can enter multiple mobile numbers and email addresses directly into the Recipient List box, separated by a 'comma (,)'. Once this is done, the user can proceed by clicking the 'Next->Alert Type' button.

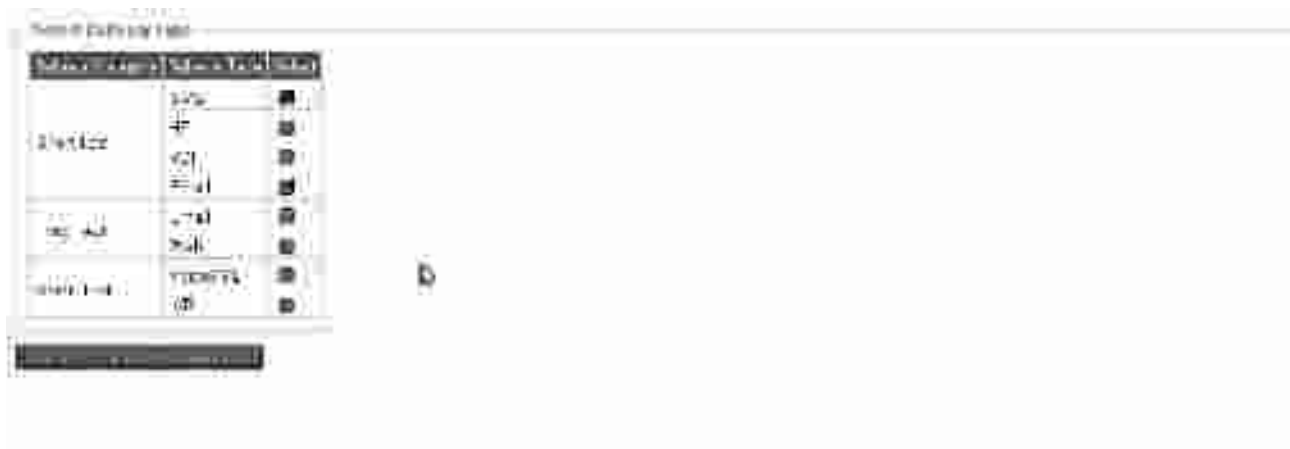


Figure 120: Select delivery type

This allows the user to select the Delivery Type of the alert. Currently the system supports Short-Text SMS, Short-text Email and Long-text Email as the delivery mechanisms (refer to the [software requirement specifications](#) to learn of the different delivery types). The user has the option of selecting one or more of these delivery types. The user can then proceed by clicking the “Next->Transform Message' button.



Figure 121: View converted message and send

In the final screen, as show above, the user is shown the transformed CAP message relevant to the selected delivery types. The user has the option here of changing any values of the shown message, just before it is sent out. When the user clicks the 'Send Message' button, the message is delivered via the selected delivery medium, provided the various servers and configuration have been made by the system administrator.

12.7.1.2 Update Alerts

From the View Alerts screen, described in section 4.2, the users can select to update an existing alert. This can be done by selecting the 'Update' link, found at the bottom of a View Alert section. The Update alert will open the list of fields for the specified alert, with populated fields filled in – the user can change values or enter new values as necessary, navigating via the given tabs – this is similar to the New Alerts section, described in section 4.3. When the user is done with the form, selecting 'Update' will save the changes to the Alert.

12.7.2 Resources

- CAP Cookbook: http://www.incident.com/cookbook/index.php/Main_Page
- Gow, G. and Waidyanatha, N. (2009). CAP Alerting for Sahana Messaging Module: real-time biosurveillance program software requirement specifications. Web link - <http://limeasia.net/wp-content/uploads/2009/06/sahana-cap-msg-mod-v02.pdf>