2: National Tuberculosis Surveillance System Data Flow

Definitions

Term	Definition
CDC data warehouse	CDC database used for reporting and analysis. The data stored in the warehouse are uploaded from the operational systems and may pass through an operational data store for additional operations before they are used for reporting.
Content validation	Process used to validate state-generated Health Level 7 (HL7) 2.5 TB case notification messages (e.g., date of birth occurs before the date of death).
Firewall	A computer software program or hardware with a predetermined set of rules that controls the incoming and outgoing network traffic. Data are analyzed to determine whether they should be allowed through or not. A network's firewall builds a bridge between an internal network that is assumed to be secure and trusted, and another network, usually an external network, such as the Internet, that is not assumed to be secure and trusted.
Health Level 7 (HL7) code	A code developed to promote and facilitate use of international healthcare informatics interoperability standards. HL7 code provides a framework (and related standards) for the exchange, integration, sharing, and retrieval of electronic health information.
National Electronic Disease Surveillance System (NEDSS)	A web-based surveillance system with an infrastructure developed by CDC that uses specific Public Health Information Network (PHIN) and NEDSS messaging standards.

Term	Definition
National TB Indicators Project (NTIP)	A monitoring system using standardized definitions, indicators, and calculations to track progress toward attaining national TB program objectives.
National Tuberculosis Surveillance System (NTSS)	The only national repository of TB surveillance data in the United States. NTSS receives data on TB cases from reporting jurisdictions' web-based systems through a standardized data collection form, the Report of Verified Case of Tuberculosis (RVCT).
Public Health Information Network (PHIN) code	A standardized code used by computer programmers to assign TB data to a specified RVCT variables. These variable codes are essential in transmitting data to CDC. A number of data issues have been attributed to errors on data system programming involving PHIN codes. For example, if a code is incorrect, the data can disappear. If the data are all missing, check the PHIN Variable ID.
Public Health Information Network Messaging System (PHINMS)	Local, state, and federal governments use the Public Health Information Network Messaging System, PHINMS, (pronounced "fin em es") to rapidly and securely send messages, which contain sensitive health information, over the internet to other organizations as well as to the CDC. The data are securely encrypted to ensure patient information is protected.
Report of Verified Case of Tuberculosis (RVCT)	The NTSS standardized data collection form. Data are collected by 60 reporting jurisdictions and submitted electronically to CDC. Data are used to monitor national TB trends, identify priority needs, and create the DTBE annual surveillance report, Reported Tuberculosis in the United States.
Structural validation	This is part of a flexible message-quality framework of services and utilities designed to assist public health partners with preparing and communicating quality standard electronic messages, as defined by the applicable messaging standards. This is used to validate standard HL7 messages per defined message specification guidelines.
Surveillance	An on-going systematic collection, analysis, interpretation, and dissemination of data to allow TB programs to target resources and interventions that will provide the most impact in eliminating TB. These data are essential in describing morbidity and mortality, monitoring trends in TB incidence and prevalence, detecting potential outbreaks, and defining high-risk groups.
TB data mart	A subset of the CDC data warehouse that contains RVCT data submitted from the states and used to supply data to the Division of Tuberculosis Elimination.

Term	Definition
Tuberculosis Genotyping Information System (TB GIMS)	A secure web-based system designed to improve access, management, and application of genotyping data at the state and local level. As part of the NTSS, TB GIMS contains tools to detect and prioritize TB outbreaks.
Tuberculosis Information Management System (TIMS)	TIMS was a Windows-based, client-server application that helped health departments and other facilities manage TB patients, conduct TB surveillance activities, and manage TB programs overall. TIMS replaced former DTBE software (SURVS-TB and TBDS) and provided for electronic transmission of TB surveillance data and program management reports. TIMS was replaced by web-based surveillance systems in 2009.

National Tuberculosis Surveillance System Data Flow

Primary Purpose

This section describes the National Tuberculosis Surveillance System's (NTSS) web-based structure and how TB surveillance data flow from reporting jurisdictions to CDC.

Understanding the structure of the web-based system and how data flow helps staff to examine errors in data transmissions.

Flow of NTSS Data from Jurisdictions to CDC DTBE

The process of how the data flow from reporting systems to CDC includes the following steps (Figure 2.1).

1. Patient data are entered into a Public Health Information Network (PHIN)-compliant system at your state or local health department.
2. Data are converted into PHIN code and formatted into an HL7 TB case notification message to be sent to CDC. PHIN codes are standard vocabulary codes for health care.
3. The TB case notification message is sent to CDC using a PHINMS sender.
4. This message travels through the CDC firewall and is picked up by a PHINMS receiver. The PHINMS receiver is located in the Office of Surveillance, Epidemiology, and Laboratory Services (OSELS) at CDC. The OSELS data brokering team receives not only TB messages, but other disease messages as well.

5. After the TB message is received by CDC, it must go through a validation and cleaning process for
   - Structural validation – The message must be capable of being translated. When a message passes structural validation, it is then sent into the CDC Central Warehouse.
   - Content validation (e.g., date of birth reported as before the date of death) – Each message is tagged with any content validation issue.
6. After the message passes structural validation and completes content validation, the message is placed in the TB Data Mart and combined with the content validations for each message.

Figure 2.1
Tuberculosis Surveillance Data Flow
**from Jurisdictions to CDC/DTBE**
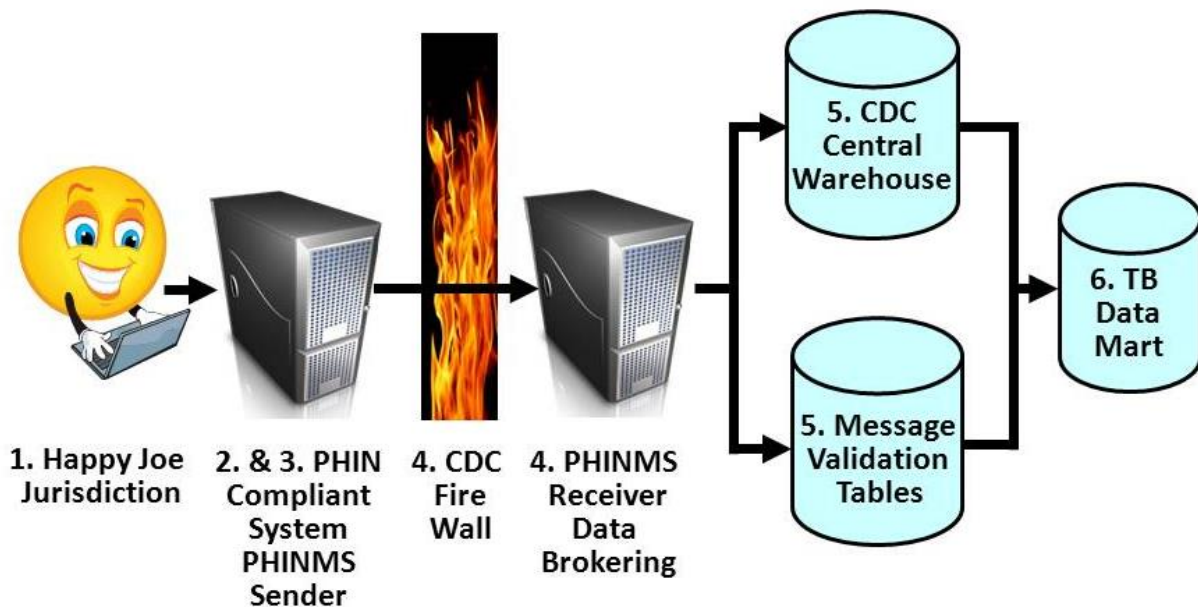
Flow of NTSS Data within CDC DTBE

Once data are received at CDC, they are sorted to ensure they represent the most recent version of each TB case. The data are then moved to create the two sources of data that make up the NTSS warehouse, including data from TIMS and the TB Data Mart. Figure 2.2 provides a description of the steps in the process and highlights the activities that are processed daily (outlined in red) and those processed weekly (outlined in blue).

- **TIMS Data** – Legacy data entered into the Tuberculosis Information Management System (TIMS).

  o **TIMS data** were extracted and translated using **Rhapsody**, which is an integration engine tool that translated the **TIMS data** into **TIMS tables** with standardized PHIN codes and NEDSS structure.

- **TB Data Mart** – Data submitted by states as TB cases not as messages.

  o **Data from the TB Data Mart** are placed into **National Electronic Disease Surveillance System (NEDSS) tables**. These data are matched by a unique identifier to avoid duplicate cases and to determine the most recent message for each case.

- **Standardize Description** – This process takes the PHIN codes provided and creates standardized descriptions for variables that are consistent regardless of the source or type of surveillance system.

- **Match Cases –** This process looks for matching state case numbers in TIMS and NEDSS to avoid duplicate cases. During this process, the NEDSS data are considered to be the most recent version for matching state case numbers.

- **Calculate Derived Variables** – This process uses the data that the jurisdiction has entered into its system and sent to CDC and creates useful derived or calculated variables. For example, the jurisdiction may confirm that it has a case of TB and that the verification of this case is based on a positive culture. CDC will calculate the case using the case verification algorithm and compare the results with the variables that the state provided in the message. If the verification results differ, this information is reflected in a report so that the reporting jurisdiction is aware of the different case categorizations. DTBE will classify the case according to the definition calculated from the data provided to CDC.

- **Combined Tables** – Any updated data to legacy TIMS cases will be received via NEDSS messages. These data will take precedence over and replace other data for the TIMS case. The data in the combined tables are used to create the reports in the NTSS application. The reports reflect exactly how the data were sent in the message or how they were provided from TIMS. These tables reflect the most recent version of all TB cases within the United States since 1993.

- **Transform and Pivot** – The data are now in a relational database; this format is **not** the preferred format for analyzing data. Therefore, the PHIN codes that are received in the message are translated into more readable codes and multi-select questions are pivoted to present one row for each case.

- **Clean Data** – Data are being received from many different systems throughout the country and systematic issues sometime arise. The cleaning process consists of removing data that are incomplete, improperly formatted, or do not meet required validation quality rules.

- **SAS Table** – These data are pushed into a staging area and finally pushed to the National Tuberculosis Indicators Project (NTIP) and the Tuberculosis Genotyping Information Management System (TB GIMS) applications. The data are also shared with the CDC Surveillance Team to produce the TB annual report and to conduct epidemiologic analyses.

Figure 2.2
Tuberculosis Surveillance Data Flow
within CDC/DTBE with Daily and Weekly Processes
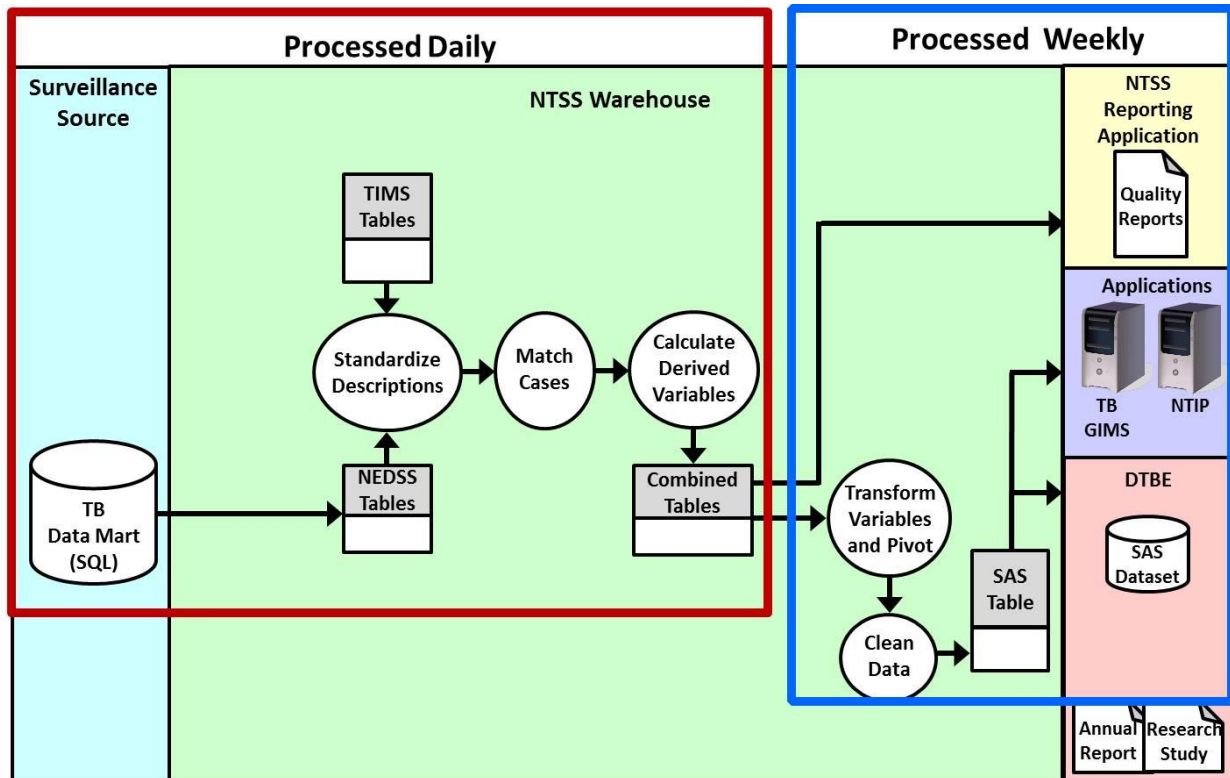
Daily processes include bringing the data from the TB Data Mart to standardized description, match cases, and calculate derived variables so that the QA team gets a first look at what is provided to CDC on a daily basis.

Weekly processes include pivoting these data into analyzable data sets. These data are sent to TB GIMS, NTIP, and NTSS for review and analysis by CDC and reporting jurisdiction personnel.

Schedule for Availability of NTSS Data to Jurisdictions

The schedule for when data are available from CDC is illustrated in (Figure 2.3). Each week, the data goes through the following schedule:

1. **Sunday** data are received at CDC.
2. **Monday morning** data are received in the NTSS Warehouse.
3. **Wednesday** data are available to those individuals that have been granted access to the applications. Keep in mind that these are the data provided to CDC as of Sunday evening.

Figure 2.3
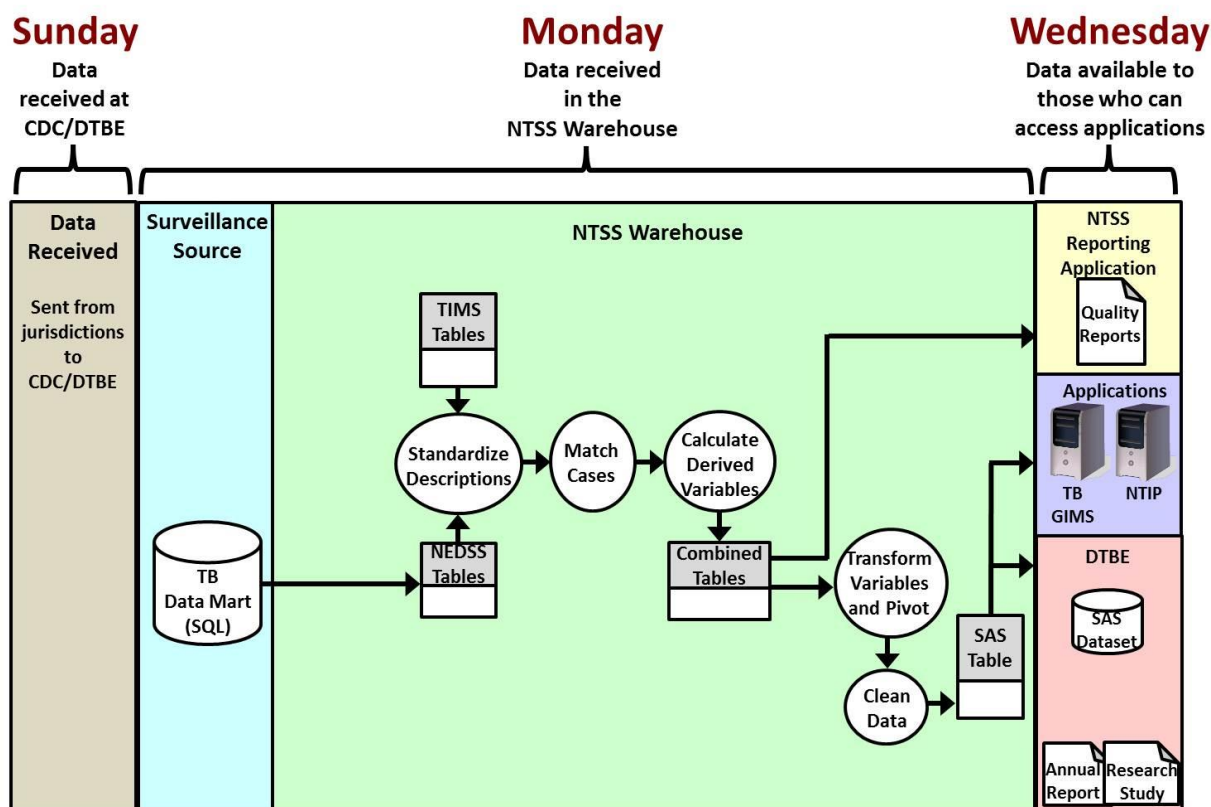**Schedule for Availability of TB Surveillance Data**

Table 2.1 provides a schedule for how the process occurs in a typical week. Red highlights provide information on when the data are staged for NTIP, NTSS, and TB GIMS and when the data are available for jurisdictions to access. This table is also Timeliness Tool–5.

Table 2.1
Timeliness Tool–5

Typical Weekly CDC TB Surveillance Data Availability Chart

Day	Activities	Data Available
MONDAY	• **Collect data received through SUNDAY** • **Process data through warehouse** • **Analyze and QA data** • **Create SAS table** • **Push data to Staging for NTIP, NTSS, TB GIMS**	**WEDNESDAY of current week**
TUESDAY	• **Collect data received through MONDAY (data received on TUESDAY will be available on WEDNESDAY of the next week)** • **Process data through warehouse** • **Analyze and QA data**	**WEDNESDAY of the next week**
WEDNESDAY	• **Applications pick up data from Staging** • **New data available in NTIP, NTSS, TB GIMS** • **Collect data received through TUESDAY** • **Process data through warehouse** • **Analyze and QA data**	**WEDNESDAY of the next week**
THURSDAY	• **Collect data received through WEDNESDAY** • **Process data through warehouse** • **Analyze and QA data**	**WEDNESDAY of the next week**
FRIDAY	• **Collect data received through THURSDAY** • **Process data through warehouse** • **Analyze and QA data**	**WEDNESDAY of the next week**

Additional Information

Contact: Tuberculosis Applications Support (TAPS)
Phone: 678-460-7828
Phone: 404-639-8444

ntss@cdc.gov

For more information about NTIP and TB GIMS see Chapter 9: QA Cross-cutting Systems and Process (NTIP, TB GIMS, and Cohort Review).