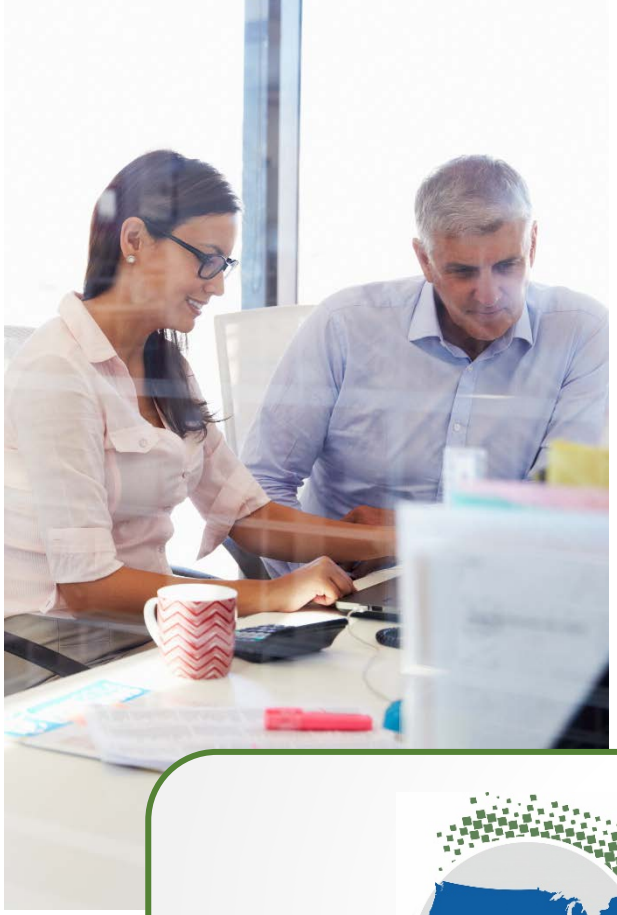
NSSP
National Syndromic Surveillance Program
BioSense Platform

January 2019

DATA DICTIONARY

Data Elements Used in NSSP Data Processing Journey

Contents



NSSP
National Syndromic Surveillance Program
BioSense Platform

The Data Processing Journey

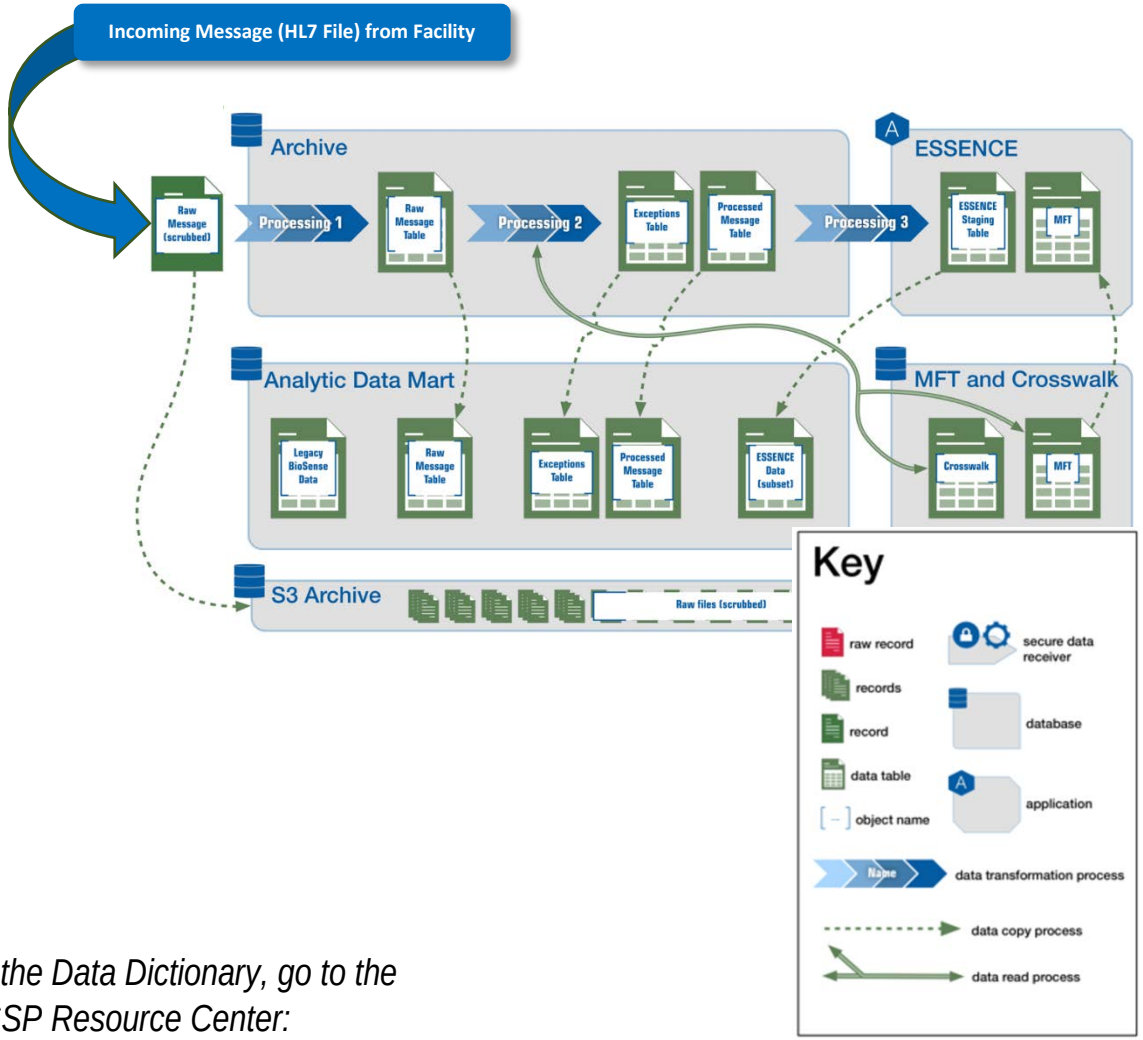
The National Syndromic Surveillance Program (NSSP) makes actionable, timely, and accurate information available to sites. NSSP strives for high-quality data from the moment data are captured throughout the "data processing journey."

The **Data Dictionary** is an Excel guidance document that contains information on data processing steps, tables, and data elements involved in NSSP data processing (see https://www.cdc.gov/nssp/biosense/docs/NSSP-Data-Dictionary.xlsx).

The Archive Database houses various tables involved in NSSP processing. These include:

- Raw Table
- Processed Table
- Exceptions and Except_Reason Table
- ESSENCE ER_Base Table

This document provides an overview of each table, the supporting reference files, and the core data elements associated with each.



To access the Data Dictionary, go to the NSSP Resource Center:

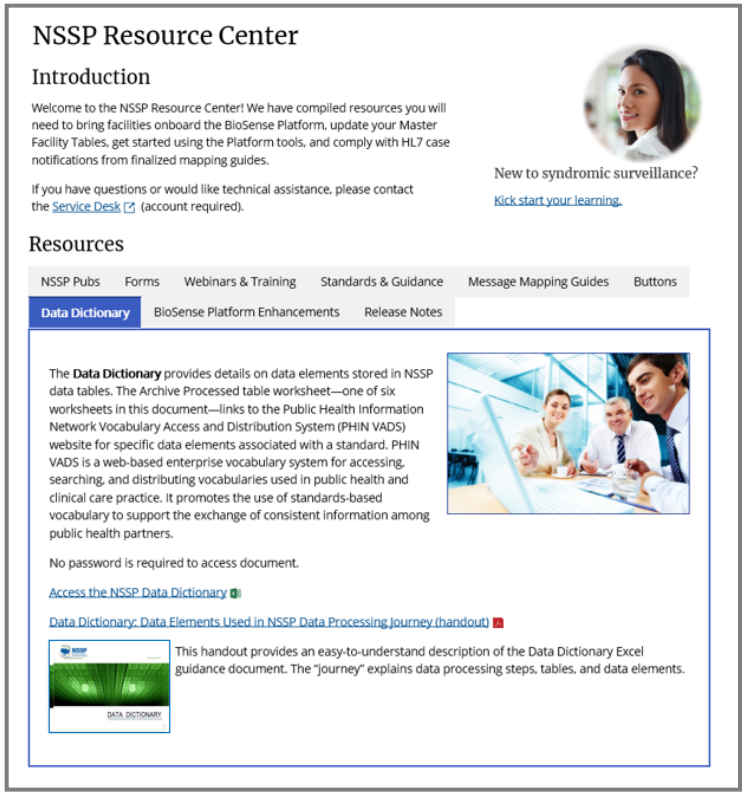
www.cdc.gov/nssp

NSSP Resource Center Web Page

You may access the **Resource Center** from CDC's NSSP Web page. The Resource Center contains several guidance documents, including the Data Dictionary.

The Data Dictionary contains the complete list of NSSP processing data elements starting with the Archive Raw Table to the final ESSENCE Table.

Link to Resource Center:
https://www.cdc.gov/nssp/biosense/publications.html



Raw Table

In the data's journey through NSSP processing, the **Raw Table** is the first step on the BioSense Platform. The Raw Table, which resides in the DataMart, receives and stores syndromic surveillance (SyS) messages that are sent to the BioSense Platform. These SyS messages can be used for various analyses of data flow and data quality.

HL7 Archive Raw Name	HL7 Archive Processing Description	Comments
Message_ID	Contains an Internal Unique Identifier to trace this message as it traverses multiple tables and systems. A Sequencer datatype should be used for this column to ensure uniqueness across the entire platform and all of it's sites.	
	A flag to determine whether this record was processed. Values: * New - No attempted processing has been done. * Read - Message has been successfully read and passed to the stored procedure to be written to the Processed or Exceptions table in the Archive.	

Sample page from Data Dictionary that summarizes data elements available in the BioSense Platform Raw Table.

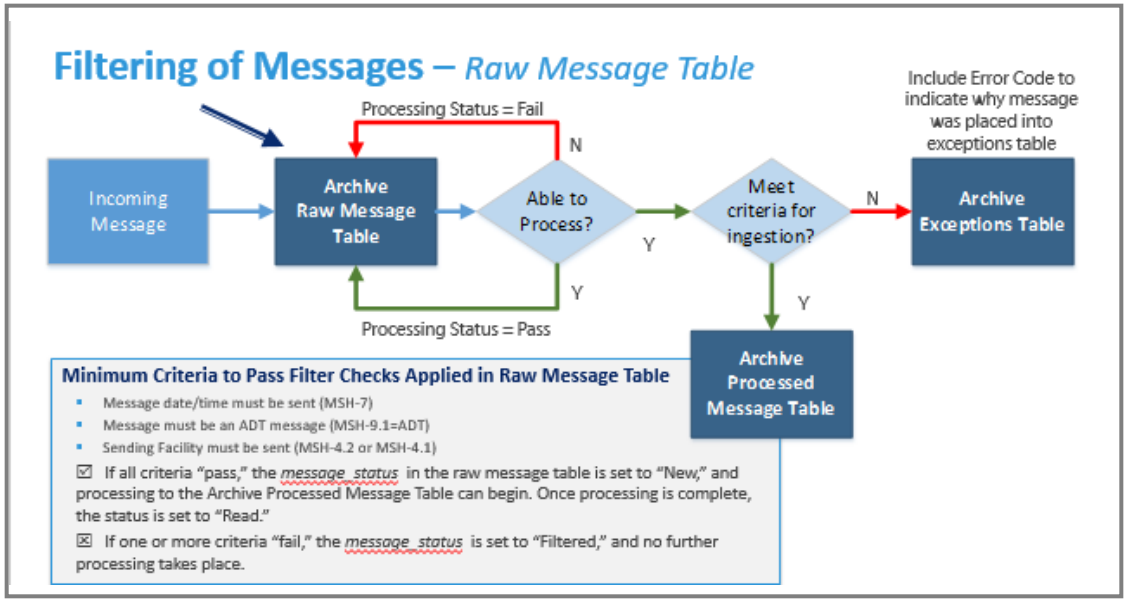
Filter Reasons

The **Filter_Reasons Table** is a reference file that describes the filter reasons codes.

For the SyS messages to continue the processing journey and *not be filtered*, messages must contain three essential features:

1. Message Date Time (i.e., MSH.7 must be present)
2. Message must be an ADT message
3. Facility information must be present (i.e., MSH.4.1 or MSH.4.2 must be present)

If one or more of these features is not present, the message is "Filtered." The *message_status* column in the Raw table is updated to "Filtered" and the *filter_reason_code* column in the Raw Table is updated with the code associated with filtering.



Data journey as messages make their way to the Processed Table.

Filter_Reason_Code	Description
1	No message date/time for HL7. In formats other than HL7, the field converting to message date/time would trigger the filter. (MSH.7 is present)
2	Message must be an 'ADT' message (Admit/Discharge/Transfer). (MSH.9.1 == 'ADT)'
3	No facility info for HL7. In formats other than HL7, the field converting to facility info would trigger the filter. (MSH.4.1 OR MSH.4.2 is present)

This page in the Data Dictionary provides the reference key for the Filter_Reason_Code and its respective description.

Processed Table

If the SyS messages contain the three essential features needed for the data journey to continue, some additional processing is launched and more data checks are applied. If the data checks are successful, the data land in a record housed in the site-specific Processed Table. The **Processed Table** is the table from which data are pulled for ingestion into ESSENCE.

The Processed Table includes all records that have passed data checks and can include multiple records associated with a single visit. Variables from the SyS messages "land" into corresponding data element columns.

This page in the Data Dictionary lists all data processing elements—both those elements received in the message and those calculated during processing. Calculated variables are distinguishable by the "C_" prefix before the data element name.

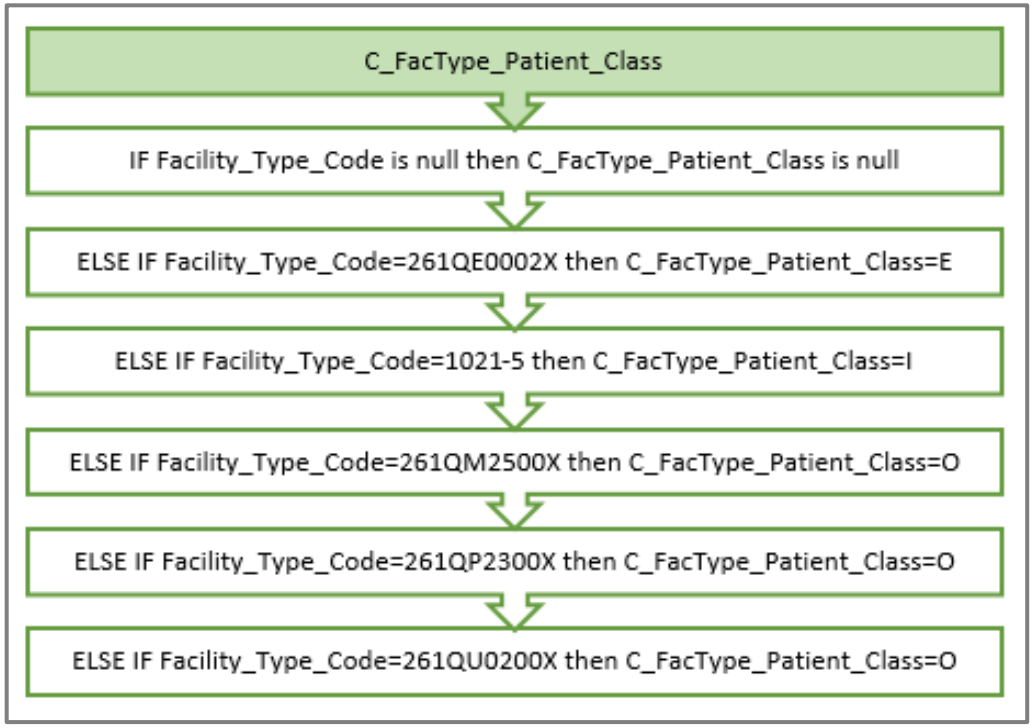
Reference			Ingestion Process - HL7 Archive	
Element Name	Description	HL7 Segment(	HL7 Archive Processed Name	HL7 Archive Processing Description
Message ID (System Generated)	The ID associated with the record in the Archive Raw table that maps to the Archive Processed (or Exceptions) table	N/A	Message_ID	Populated from Mirth; this ID is copied from the Archive Raw table for the message associated with the record in Archive Processed or Archive Exceptions table. This value will be globally unique such that any ID appears one and only one time within the Archive.
Record creation date/time (Arrival Time, pre-processing)	Date message arrived at the BioSense Platform (may differ from date record was created)	N/A	Arrived_Date_Time	Date the message arrived at the BioSense Platform and was picked up for scrubbing
Record creation date (Arrival Date, pre-processing)	Date message arrived at the BioSense Platform (may differ from date record was created)	N/A	Arrived_Date	Date portion of Arrived_Date_Time
Record creation date/time (Arrival Time, pre-processing)	Date message arrived at the BioSense Platform (may differ from date record was created)	N/A	Str_Arrived_Date_Time	Date the message arrived at the BioSense Platform and was picked up for scrubbing Stored as a String in the database
Record creation date/time (Processed Table)	Internal field that captures the date/time the record was written to the processed table in the BioSense Platform Archive	N/A	Create_Processed_Date_Time	Populated with the date/time the record is first written into the Processed Table of the BioSense Platform Archive
Record creation date/time (Raw Table)	Internal field that captures the date/time the record was written to the raw table in the BioSense Platform Archive	N/A	Create_Raw_Date_Time	Populated with the date/time the record is first written into the Raw Table of the BioSense Platform Archive
Record updated date/time	Internal field that captures the date/time the record was last updated in the database	N/A	Update_Processed_Date_Time	If a message is reprocessed, this field should capture the reprocessing date/time. Create_Date_Time should not change.
Feed Name	CDC Feed Name associated with message	N/A	Feed_Name	Mirth will set this value based on the feed associated with the incoming message. * For data converted from the Legacy environment, - if legacy record was sent by phinms feed, we will place "phinms" - if this value is blank, we will place "Legacy_Unknown" - if value is populated in legacy record, we will copy that value over

Examples of core data elements.

Core Calculated Variables Page

An NSSP-calculated variable is defined as *a variable computed via formula during the processing from the Raw Table into the Processed Table*; a calculated variable is not included in the raw message.

On the **Core Calculated Variables Page**, users can find information on eight commonly referenced NSSP calculated variables, including visual diagrams detailing how each variable is calculated. (Note: This is not a complete list of NSSP calculated variables. To view all NSSP calculated variables, please refer to the Processed Table explanation.)



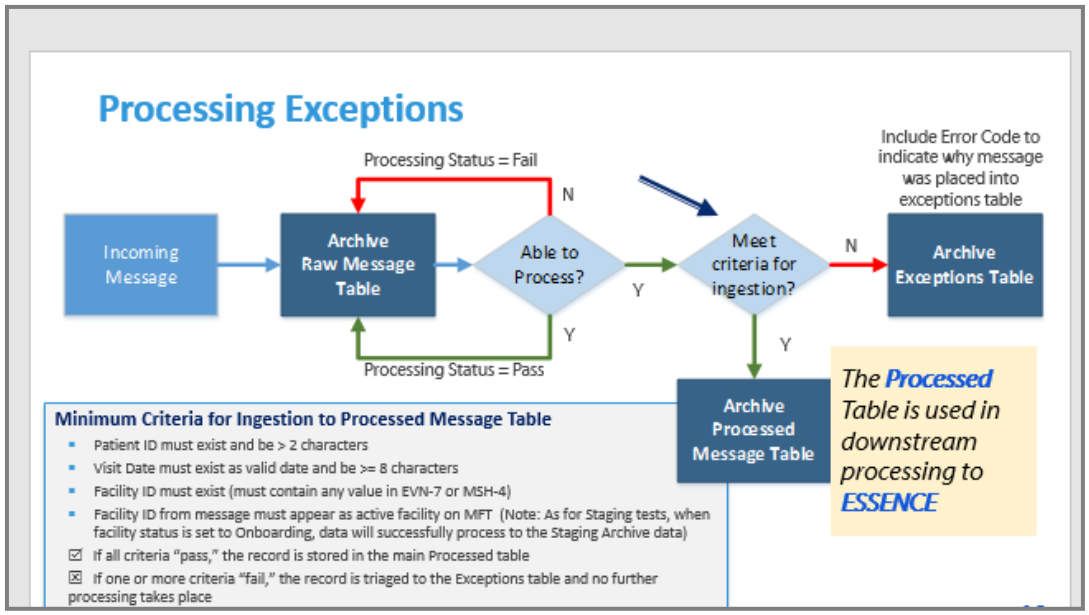
Example of calculated variable diagram.

Exceptions Reason

As SyS messages progress on the data journey, additional data checks are applied. And, if these checks are successful, data are processed into the Processed Table. However, if one or more data checks fail, these data are triaged to a separate table called the **Exceptions Table**.

The layout of the Exceptions Table is *identical* to that of the Processed Table. The Exceptions Table contains identical data elements as well. Once messages are sent to the Exceptions Table, the system tags the message with exception codes that are stored in the *exceptions_reason_code* column under a separate table named Except_Reason. The codes provide **Exceptions Reasons** for why the message failed further processing.

The Data Dictionary's Exceptions Reason code page lists all codes for which a message could be exceptioned in the BioSense Platform Archive Processed Table. (Note: The Exceptions table can be joined with the Except_Reason table using *message_id*.) The table layout and contents (data elements) are identical to that of the Processed Table. Messages sent to this Exceptions Table are tagged with exception codes that are stored in the *exceptions_reason_code* column in a separate table named Except_Reason. The codes provide information on why the message failed further processing. The Data Dictionary's Exceptions Reason Code page lists all codes for which a message could be exceptioned in the BioSense Platform Archive Processed Table. (Note: the Exceptions table can be joined with the Except_Reason table using *Message_ID.*)



Data flow for processing the exceptions.

Exceptions_Reason_Code	Description
01	Message did not include valid C_Unique_Patient_ID
02	Message did not include valid C_Visit_Date_Time
03	Message did not include C_Facility_ID
04	C_Facility_ID was not found on the Operational Crosswalk
05	C_Facility_ID was not an active facility at time of message processing
06	Site_ID is null
07	<Site> Raw tablename Parameter is Null
08	Message ArrivedDatetime is null or Invalid
09	Create_Raw_Date_Time is null or Invalid
10	Invalid Message ID
11	C_Visit_Date_Time is set in the future. We currently allow a 12 hour grace period for messages to come in. This means if the current time is 1pm and a message comes in with a C_Visit_Date_Time of 4pm this message would be successfully processed.
12	Invalid <Site> Raw tablename Parameter to Stored Procedure
13	Missing MSH4_Sending_Facility (Legacy only)
98	Site confirmed facility exclusion (Legacy only)
99	PHIN MS Site confirmed facility exclusion (Legacy only)

Example of Exceptions Reason Code.

ESSENCE

The last step in the NSSP data processing journey is to translate data elements from the NSSP Archive Processed Table to the ESSENCE Table.

During processing, data from the Processed Table are collapsed into a single record per visit. This page summarizes the data elements available in ESSENCE and maps Archive Processed variable names to ESSENCE variable names.

ESSENCE Element Nam	Processing (Archive Processed to ER Import Staging)	Processing (ER Import Staging to ER Base)
C_Patient_Class_List	Placeholder for information derived during later processing.	Alphabetic, de-duplicated list of all Calculated Patient Class (c_patient_class) values present across messages with the same ESSENCE ID. Examples: E I O EO EI EV IP EIO Note: A formatted description version of this field is used in ESSENCE screen display, applies only to values that have various combinations of E, I, and/or O. All other combinations are displayed as stored in the field. Examples: Emergency Inpatient Outpatient Emergency, Outpatient Emergency, Inpatient EV IP Emergency, Inpatient, Outpatient Note: This is referred to as the "Calcualted Patient Class History" in the
HasBeenE	Calculated	Set to 1 if 'E' is found in the calculated patient class history list. (c_patient_class_list) Otherwise set to 0 Note, formatted descriptive data are displayed in ESSENCE screens as follows: 1 is displayed as "Yes" 0 is displayed as "No"
C_Patient_Class_Updates	Placeholder for information derived during later processing.	All Calculated Patient Class (C_Patient_Class) values present across messages, stored in a consecutively deduplicated list. Examples: {1};O;{2};E;{3};I;{4};E; Note: this is referred to as "Computed Patient Class History" within the ESSENCE application.

Example of ESSENCE data elements.

LegacyToProcessed

This section of the Data Dictionary explains some common business rules applied to convert legacy BioSense 2.0 SQL tables into the new NSSP data structures supported by Processed and Exceptions tables (information includes variables used from the "old table" and how those data land into variables in the "new tables").

	A	B	C	D	E	F	G
		Ingestion Process - HL7 Archive				Reprocessing of	Legacy BioSense Data
	HL7 Archive Processed Name	HL7 Archive Processing Description	Processing Rationale	Processed?	Required	Replace separator?	Processing Description
	Administrative_Sex	Direct input from HL7 message PID-8.1	PHIN Messaging Guide for Syndromic Surveillance, Release 2.0 (April 2015)	Yes	No	No	Direct input from the Legacy Field: **PID_8_1_Patient_Gender**
	Admission_Type	Direct input from HL7 message PV1-4.1	PHIN Messaging Guide for Syndromic Surveillance, Release 2.0 (April 2015)	Yes	No	No	Direct input from the Legacy Field: **PV1_4_1_Admission_Type**
	Admit_Date_Time	Direct input from HL7 message PV1-44.1	PHIN Messaging Guide for Syndromic Surveillance, Release 2.0 (April 2015)	Yes	No	No	Direct input from the Legacy Field: **PV1_44_1_Admit_Date_Time**
	Admit_Reason_Code	Direct input from HL7 message: * PV2-3.1 * PV2-3.4 Select first non-null value	PHIN Messaging Guide for Syndromic Surveillance, Release 2.0 (April 2015)	Yes	No	Yes (:)	Direct input from the Legacy Field: **PV2_3_1_Admit_Reason_ID**
	Admit_Reason_Combo	Store Admit_Reason_Code and Admit_Reason_Description in this field, accounting for null values. For Repeating Codes store the values using the following methodology: Code Description;Code Description;... Note the space between Code and Description	PHIN Messaging Guide for Syndromic Surveillance, Release 2.0 (April 2015)	No	No	No	Combo fields will not contain legacy data.
	Admit_Reason_Description	Direct input from HL7 message: * PV2-3.2 * PV2-3.5 Select first non-null value and concatenate if repeating	PHIN Messaging Guide for Syndromic Surveillance, Release 2.0 (April 2015)	Yes	No	Yes (:)	Select first non-null value from: **PV2_3_2_Admit_Reason_Text** OR **PV2_3_5_Admit_Reason_Alt_Text** *Note that if both legacy fields contain data, the data from the alt_text column will not be ingested into the Archive during reprocessing

Example of LegacyToProcessed data elements.

Reference Page

Toward the back of the Data Dictionary is the **Reference Page**, which provides the overarching reference information and definitions including usage, cardinality, data types, and other terms from the *Public Health Information Network (PHIN) Guide for Syndromic Surveillance: Emergency Department, Urgent Care, Inpatient and Ambulatory Care Settings*.

Usage	
R	Required, must always be populated
RE	Required, but may be empty. If the Sender has data, it must be sent. The receiver must be capa
C	Conditionally Required
CE	Conditionally Required but may be empty.
X	Not supported
O	Optional
Cardinality	
[0.. 1]	Component may be omitted and can have, at most, one occurrence
[1.. 1]	Component must have exactly one occurrence
[0.. *]	Component may be omitted or repeat an unlimited number of times
[1.. *]	Component must appear at least once and may repeat unlimited number of times

Familiarize yourself with the Reference Page definitions.

Indexes

Indexes are used to expedite database searches. When executing a query against a table that contains no indexes, the query must search the entire table in a linear fashion. However, when an index is added, the query will instead search only a subset of the data, which increases efficiency. Indexes appear by select columns to assist in data analysis.

Column Name	Production		Staging
	Processe	Raw	Processe
Arrived_Date_Time	Yes	Yes	Yes
Arrived_Date	Yes	Yes	Yes
Create_Raw_Date_Time	No	Yes	No
Feed_Name	Yes	Yes	Yes
Message_ID	Yes	Yes	Yes
Message_Status	No	Yes	No
Updated_Processed	No	Yes	No
C_Visit_Date_Time	Yes	No	Yes
C_Visit_Date	Yes	No	Yes

Subset of indexes available on the BioSense Platform.

Disposition Category Mapping

As data are processed from ESSENCE ER_Base to ESSENCE Web, a reference table will map DischargeDisposition to a Disposition Category. The underlying data remain stored in the format initially received (usually disposition codes). Users will see visual reports on the ESSENCE front end that display the disposition category. The **DispositionCategoryMappingTable** is the reference table used to perform this mapping.

Disposition	Category
9	ADMIT
30	ADMIT
31	ADMIT
32	ADMIT
33	ADMIT
34	ADMIT
35	ADMIT
36	ADMIT

Example mapping of disposition codes to disposition categories.

Members of the NSSP Analytic Data Management team are available to answer questions and discuss data quality reports.

To schedule a one-on-one discussion, please contact the NSSP Service Desk.