

Identifiers

This document was developed by the Data Standardization Work Group (DSWG) of the Council of State and Territorial Epidemiologists (CSTE) Surveillance Practice and Implementation Subcommittee (SPIS) as a part of a larger data harmonization effort. It will be published and maintained in a repository on the CSTE website.

More information about the DSWG and data harmonization can be found here: [\[link to DSWG Core Surveillance Project Brief \(pending\)\]](#).

Background

Unique identifiers are linked to various aspects of any case, including the Local Record ID that identifies the case (event) and the Local Subject ID that identifies the person (individual). Table 1 below shows the identifier data elements present in the Generic V2 message mapping guide (MMG).

Table 1. Identifier Data Elements from Generic V2 MMG

Data Element Name	Data Element Identifier Sent in HL7 Message	Data Element Description
Local Subject ID	N/A: PID-3	The local ID of the subject/entity
Local Record ID	N/A: OBR-3	Sending system-assigned local ID of the case investigation with which the subject is associated.
State Case Identifier	77993-4	States use this identifier to link NEDSS investigations back to their own state investigations.
Legacy Case Identifier	77997-5	CDC uses this identifier to link current case notifications to case notifications submitted by a previous system (NETSS, STD-MIS, etc.).

The types of identifiers that may be connected to a case vary, as some identifier types are only applicable to certain conditions. Throughout the condition-specific MMGs, various identifier data elements are used to capture and transmit specific information that pertains to the condition. Table 2 below shows some examples of identifiers from condition-specific MMGs.

Table 2. Identifier Data Elements from Condition-Specific MMGs

Data Element Name	Data Element Identifier Sent in HL7 Message	Data Element Description
Vaccination Record Identifier	VAC102	Vaccination record identifier
State Outbreak Identification Number	FDD_Q_1129	State outbreak identification number

WGS ID Number	INV949	NCBI SRA accession number (SRR ID): The accession number generated by NCBI's sequence read run when sequence data are uploaded to NCBI. It is most granular for the isolate. This provides both the sequence data and metadata on how the sample was sequenced.
CDC 2019-nCoV ID	94659-0	CDC 2019-nCoV ID: This ID will be used to track information about the case-patient in CDC data systems and must be provided on all forms or specimens related to this individual. The format is the standard alpha code representing the reporting jurisdiction followed by the reporting jurisdiction local ID (6 to 16 numeric characters) which may be the NNDSS local record ID (EXAMPLE AL198754), but is assigned by the reporting jurisdiction and MUST be unique for each case.

The MMG-specific identifiers listed above are currently captured as discrete data elements in which the data element name and description clearly indicate what type of information is represented by that data element. However, this approach is static, restricting data collection to the defined element and prohibiting the collection of any other information. For example, if a new identifier is needed in an emergency response or other non-routine situation, the addition would require approval from the Office of Management and Budget (OMB) and a version update to the MMG. This process that can take months to complete and would defeat the goal of a timely emergency response.

The workgroup aimed to define a data structure that can capture identifier information and enumerate which identifiers should be defined as a core surveillance data element for routine case notification to CDC. Identifiers defined as core surveillance are data elements that jurisdictions should be prepared to transmit to CDC and that CDC should be prepared to receive across conditions. Identifiers that are applicable only on a condition-specific basis would not be considered core data elements. The specific topics for the workgroup discussions were as follows:

- What structure would jurisdictions prefer for sending identifier data: a repeating group or discrete data elements?
- Is it important that the structure supports routine IDs used across conditions, ad-hoc IDs, or both?
- What routine IDs used across conditions should be considered core data elements?

Core Surveillance Identifiers Data Elements for Routine Case Notifications

The data elements “Local Subject ID” and “Local Record ID” have been renamed “Person ID” and “Case ID” to clarify their roles in identifying individual persons and unique cases (events), respectively. This change aims to reduce confusion by using more typical nomenclature for these two concepts, ensuring accurate transmission of case notifications to CDC. Table 3 gives updated labels and descriptions for these required identifiers. There is no change to the content currently sent in these HL7 fields, and these data elements will remain separate from a repeating group structure.

Table 3. Updates to “Local Subject ID” and “Local Record ID” from Generic V2 MMG

Data Element Name	Data Element Identifier Sent in HL7 Message	Data Element Description
Person ID	N/A: PID-3	The jurisdictional identifier that uniquely identifies the person.
Case ID	N/A: OBR-3	The jurisdictional identifier that uniquely identifies the case (event ID).

The data elements below are those that the workgroup recommended defining as core surveillance data elements for case notifications to CDC. For each, there is no pre-defined default value for what should be transmitted when the data element has not been populated. Refer to Tables 4-6 below for definitions and additional information for each data element.

- Identifier Type
- Identifier Value

A repeating group structure for identifiers with two data elements of “Identifier Type” and “Identifier Value” was supported with weak consensus (64%) from the group. Identifier Type conveys the type of identifier to be sent and would be limited to a value set of identifier types. Identifier Value will transmit the actual value of the identifier for the case. For example, transmission of these data elements for a case with a CDC National Outbreak Reporting System (NORS) ID of “EIP1234567” would look like this:

```
OBX|nn|CWE|TBD^Identifier Type^PHINQUESTION|1|TBD^NORS ID^PHINQUESTION||||F
OBX|nn|ST|TBD^Identifier Value^PHINQUESTION|1|EIP1234567||||F
```

The group voted with consensus on a core minimum value set for Identifier Type. This minimum value set represents identifier types that may not always be present for an individual case but are necessary to be able to transmit across conditions. This minimum value set does not exclude additional identifier types from being added to the value set as needed for specific circumstances. Identifier types included in this minimum value set are:

- CDC-assigned Case ID
- Jurisdiction-assigned Outbreak ID
- NORS ID

Table 4. Core Surveillance Data Element for Routine CDC Case Notification – Identifier Type

Identifier Type	
Definition	Descriptor of the identifier transmitted in Identifier Value
Standard Code for Transmission to CDC	TBD
Value Set or Coding System for Transmission to CDC	Table 5 shows a minimum value set that defines specific identifier types defined as core data elements. Additional values may be added to this value set as needed for certain conditions. This value set will utilize the LOINC code system when possible, and the CDC PHINQuestion code system as needed.
Abstracted/ Derived/ Created	N/A
Derived from (examples)	N/A
Recommendations for Input Data Source(s)	N/A
Notes	Part of a repeating group structure with “Identifier Value”.

Table 5. Identifier Type Data Element – Minimum Value Set

Concept Name	Concept Description
CDC-assigned Case ID	A CDC-assigned ID for use in situations such as an emergency response.
Jurisdiction-assigned Outbreak ID	A jurisdiction-assigned ID for an identified outbreak.
NORS ID	The outbreak ID from CDC's National Outbreak Reporting System (NORS) for the outbreak this case report is associated with.

Table 6. Core Surveillance Data Element for Routine CDC Case Notification – Identifier Value

Identifier Value	
Definition	Unique identifier value corresponding to the type transmitted by Identifier Type
Standard Code for Transmission to CDC	TBD
Value Set or Coding System for Transmission to CDC	Text
Abstracted/ Derived/ Created	Created
Derived from (examples)	N/A
Recommendations for Input Data Source(s)	N/A
Notes	Part of repeating group structure with “Identifier Type”.

The repeating group structure allows for flexibility in capturing additional identifier types across conditions as needed. The minimum value set for Identifier Type defined in Table 5 does not exclude the addition of additional concepts to the value set as needed for certain conditions. Some examples of additional identifier types not included in the core minimum value set but that may also be transmitted by this repeating group are seen in Table 7; these examples are not comprehensive.

Table 7. Identifier Type Data Element – Possible Additional Concepts to be Included in Value Set

Concept Name	Code	Concept Description
CDC Specimen ID*	TBD	CDC specimen ID number.
Public Health Specimen ID*	TBD	Public health laboratory specimen ID.
Legacy Case Identifier†	77997-5	Links current case notifications to case notifications submitted by a previous system.
Birth Parent's Case ID‡	TBD	The local case ID used for reporting the birth parent's case of this condition.
Birth Parent's National Reporting Jurisdiction‡	TBD	The national reporting jurisdiction used for reporting birth parent's case of this condition.

***Specimen IDs:** Jurisdictions currently send specimens to CDC laboratories for testing and submit case notifications for disease surveillance. Linking CDC lab results to their associated cases is valuable for CDC programs but is challenging due to the lack of consistent identifiers across specimens and case notifications. Transmitting the CDC Specimen ID or Public Health Specimen ID within a case notification would allow more efficient linking of specimens received by CDC with case notifications. The DSWG is also developing recommendations for sending laboratory result information within case notifications, and it is possible for a "Specimen ID" field within this laboratory block to fulfill the purpose of these identifier types. However, if this information could not be transmitted within the laboratory block, the Identifiers repeating group could be used.

†Legacy Case Identifier: This identifier is presently included as a standalone element in Gen V2 for linking changing Case IDs when transitioning from one surveillance system to another within a jurisdiction, resulting in different Case IDs assigned in different systems for the same case. The workgroup agreed that this is rarely used, but necessary during times of jurisdictional system transition. As such, this identifier was removed as a standalone data element to instead be transmitted as a part of the newly created identifiers repeating group when needed (Table 7).

‡Birth Parent-Infant Data Linkages: The transmission of paired birth parent and infant case information for certain conditions such as Zika virus, congenital rubella, and others is important for situational awareness, analysis of risk factors, and monitoring of long-term effects. There are two methods of transmitting information about the birth parent within an infant's case notification:

1. Providing the birthing parent's case ID and national reporting jurisdiction to link the two cases.

This method is concise and reduces the burden of duplicate data entry; however, it is unlikely that the infant's jurisdiction will know the parental case ID if the birthing parent's case was reported in another jurisdiction. In this situation, parental case information could be more easily collected during an interview or from a medical record during the infant case investigation, where applicable.

2. Resending key data elements about the birthing parent within the infant's case notification.

Sending parental data elements with the infant case ensures the information needed is present, but increases the number of data elements in the infant case notification. Populating these data elements in both case notifications will require duplicate data entry, which may result in data entry errors or contribute to an increase in workload for high-volume conditions. Furthermore, any information included in the parental case notification but not in the infant case notification will not be available for analysis. In situations where both methods are used, it may be difficult to resolve any discrepancies that occur between the case records.

Due to these numerous considerations, the DSWG did not come to clear consensus on a preferred method for linking infant and birthing parent cases. There was a slight preference (57%) to utilize the birthing parent's case ID and national reporting jurisdiction. These data elements can be transmitted as part of the identifiers repeating group when needed (Table 7).