

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 case ID that identifies the case and the person ID that identifies the 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 (DE) Name	DE 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 (DE) Name	DE 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.

All of the above identifiers in MMGs 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 does not allow for any flexibility in the data element being used to collect 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, a process that can take months to complete and would defeat the goal of a timely emergency response.

The workgroup aimed to define both a data structure with which to capture identifier information and to enumerate which identifiers should be defined as a core surveillance data element for routine case notification to CDC. Such an identifier would be one 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 to be a core data element. The specific discussion 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 support 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 identifiers Local Subject ID and Local Record ID are used to identify the person as a unique individual and the case (event) as a unique case, respectively. These are essential for transmission of case notifications for CDC, but the current names and definitions of these data elements have caused confusion. Table 3 gives updated labels and descriptions for these identifiers.

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

Data Element (DE) Name	DE 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 DSWG recommends be defined as core surveillance data for case notifications to CDC. For each of these data elements, 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 instance, 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 condition but are necessary to be able to transmit across conditions. Identifier types included in this minimum value set are:

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

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

Identifier Type	
Definition	Description 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 locally defined codes 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 – Value Set Descriptions

Value Label	Value Description
CDC-assigned Case ID	A CDC-assigned ID for use in situations such as an emergency response.
Outbreak ID – Jurisdiction-assigned	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 for a condition as needed. The minimum value set for Identifier Type defined here does not exclude the addition of additional values to the value set as needed for certain conditions. 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.

Table 7. Possible Additional Values for Identifier Type

Identifier Type	Code	Data Element Description	DSWG Conclusion
Legacy Case Identifier	77997-5	Links current case notifications to case notifications submitted by a previous system.	Important only for jurisdictions undergoing a system transition.
Birthing Parent's Case ID	TBD	The local case ID used for reporting the birthing parent's case of this condition.	Repeating group can support this when needed, but it may not be the best option for all conditions.*
Birthing Parent's National Reporting Jurisdiction	TBD	The national reporting jurisdiction used for reporting birthing parent's case of this condition.	Repeating group can support this when needed, but it may not be the best option for all conditions.*
CDC Specimen ID	TBD	CDC specimen ID number.	Repeating group should support this when needed for a condition.
Public Health Specimen ID	TBD	Public health laboratory specimen ID.	Repeating group should support this when needed for a condition.

*There are two methods of transmitting information about the birthing parent within an infant's case notification – either providing the birthing parent's case ID and national reporting jurisdiction to link the two cases, or resending key data elements about the birthing parent within the infant's case notification. The first option has the benefit of sending only the case ID rather than numerous data elements; however, some considerations are that if the birthing parent's case was reported in another jurisdiction, it is unlikely that the infant's jurisdiction will know the parental case ID. In this situation, it is often easier to ask the parent for the information in the interview rather than get the case ID from another jurisdiction. A second consideration is that the information needed may not necessarily be populated in the parental case, or the parent may not meet a case definition themselves. The second option of sending the individual data elements means the infant case notification will have more data elements, but ensures the information needed is present. Populating these data elements in both case notifications will require duplicate data entry, which may result in data entry errors. Furthermore, while reentering case data may not be burdensome for low volume-conditions like congenital rubella, it may represent a significant increase in workload for high-volume conditions. In addition, with this method, any information included in the parental case notification but not requested to be reentered in the infant case will not be available for analysis.

A final consideration brought up on the call was that if both methods are used simultaneously, it may be difficult to resolve any discrepancies that occur between what is sent in the birthing parent's case and the infant's case. Due to these numerous considerations for or against either method, the DSWG did not come to consensus on a preferred method for linking infant and birthing parent cases; however there was a slight preference (57%) to utilize the birthing parent's case ID and national reporting jurisdiction.