

# Identifiers

CSTE DSWG Data Element Specific Team  
7/19/2024



## What does it mean for an identifier to be core?



A core identifier should be something that would be capable of being transmitted across conditions when needed.

- All cases require a unique ID to be identified as a unique case. (~~Local Record ID~~ Case ID)
- Something like an "Outbreak ID" would rarely be present, but if core then a jurisdiction should be capable of capturing/sending that field across conditions.
- Core = Will be prepared to utilize it across conditions as needed. Minimum value set/set of elements for what states are ready to send.

Council of State and Territorial Epidemiologists

Starting with a reminder of the scope of our work for identifiers, a core identifier should be something that would be capable of being transmitted across conditions when needed. For instance, it's very clear that each case requires a unique ID to be identified as a unique case, so this value will always be present.

However, something like an "outbreak ID" would rarely be present, but if core then a jurisdiction should be capable of capturing/sending that field. When we define an identifier as core, we are saying that jurisdictions and CDC will be prepared to transmit and receive that identifier across conditions when applicable. We can think about what we're defining as a minimum value set, or minimum set of identifiers, for what jurisdictions are prepared to send.

There may be additional identifiers that are needed on a condition specific basis, and our minimum set of identifiers would not prevent the potential repeating group structure from including those identifiers. The conversation here, both in May and in July, is going to focus on what should be consistently implemented across many conditions.

## Identifier Structure & Minimum Set



## Potential Identifier Structure



- After the last call, the structure has been modified to remove "identifier assigner".
  - Could create the "parent-child-grandchild" scenario in some use cases.
  - Given our current minimum value set for identifiers, it is not needed for core.

| Data Element (DE) Name | DE Identifier Sent in HL7 Message | Data Element Description  | Data Type |
|------------------------|-----------------------------------|---------------------------|-----------|
| Identifier Type        | TBD                               | Description of identifier | Coded     |
| Identifier Value       | TBD                               | Unique identifier value   | Text      |

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

Council of State and Territorial Epidemiologists

In our May call we talked about utilizing a repeating group structure for Identifiers, with data elements that captured the type of identifier, the unique value, and the assigner. After discussion on the last call, and given what is included in our current minimum value set for identifiers, the proposed structure for the identifiers repeating group has been simplified to only include the Identifier Type and Identifier Value.

- Identifier Type would have a defined value set. For core, this would include our "Minimum Value Set" for identifiers.

- The second element called "Identifier value" captures the actual ID value.

Previously the third data element was "Identifier assigner", a way to denote who assigned the ID, however this could potentially create the "parent-child-grandchild" structure scenario in some use cases, may be complex to implement with a structured data type like an OID, and may have limited utility for analysis if free text. For these reasons we'd like to talk on our July call about this new simplified structure that only contains the first two data elements. Last month 64% of our group preferred the repeating group structure for identifiers versus 36% preferred sending as discrete data elements, and this month we would like to finalize the data elements within this structure and their minimum value set.

A reminder that this approach would exclude local record ID and local subject ID, or as we renamed them on the June DSWG call the Case ID and Person ID, as those are already required to be included in every message in their own specified HL7 fields.

## Minimum Set of Identifiers



| Identifier Type                     | Data Element Description                                                                                                   |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Legacy Case Identifier              | <del>Links current case notifications to case notifications submitted by a previous system.</del>                          |
| CDC-assigned Case ID                | A CDC-assigned ID for use in situations such as an emergency response.<br>(Like DCIPHER ID for mpox.)                      |
| Outbreak ID – Jurisdiction-assigned | A jurisdiction-assigned ID for an identified outbreak.<br>(Replacement of Case Outbreak Name.)                             |
| NORS ID                             | The outbreak ID from CDC's National Outbreak Reporting System (NORS) for the outbreak this case report is associated with. |

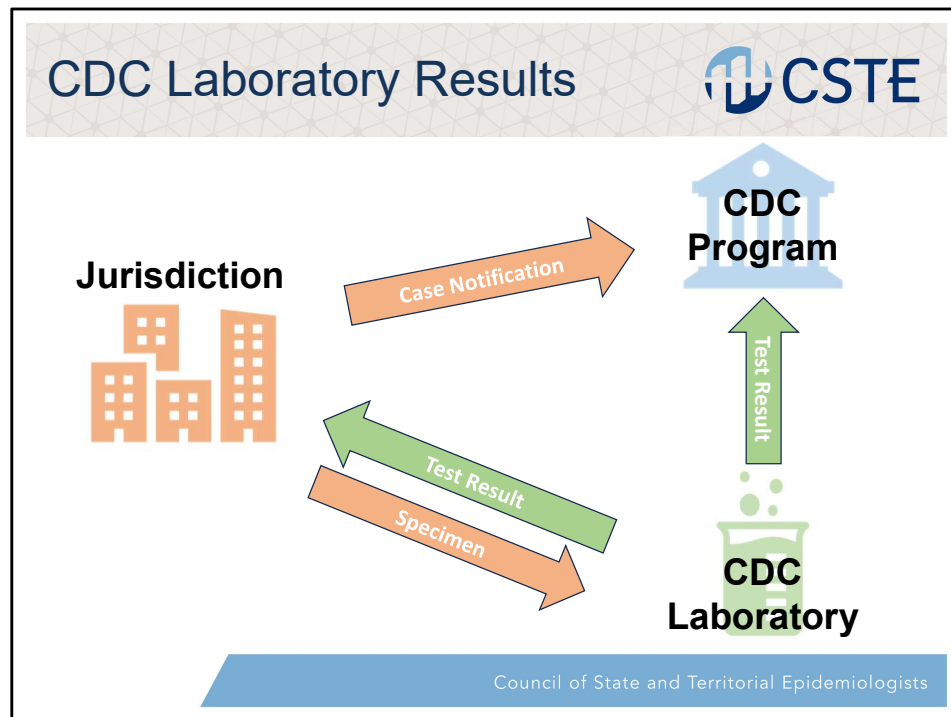
Legacy Case ID will be noted as an ID CDC needs to support, but only used by jurisdictions undergoing transitions.

Council of State and Territorial Epidemiologists

In May we polled on these identifiers as our minimum set of identifiers. We had 54% of our group vote that yes these are core, and 43% vote that only some are core. The one identifier that came up in discussion as problematic for that group is the Legacy case identifier. Now that we're solidifying the identifier structure on this call, we'd also like to re-poll on our minimum set of identifiers with the Legacy Case ID removed. That ID will be noted as something that CDC needs to support as it's important for jurisdictions undergoing a surveillance system transition, but otherwise it is not core for all jurisdictions or conditions.

Specimen ID





Jurisdictions currently send specimens to CDC labs 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. This siloing of information means:

1. CDC cannot easily determine how many cases are associated with the specimens it receives.
2. CDC cannot use this information analytically to better understand the cases.

**A high-level overview of how CDC receives these two sets of information is shown on this slide:**

- Specimens are sent to CDC from various sources, primarily public health labs and commercial labs.
- After CDC conducts the testing, results are always sent back to the jurisdictions, regardless of who submitted the specimen. The results may be sent to the epidemiology program, the public health lab, or both, and they may be delivered through different routes and in different formats.
- The CDC lab also provides the test results to the related program at CDC.
- If the tested condition is notifiable and meets a surveillance case definition, the jurisdiction then sends a case notification to CDC.

When the CDC program receives the test result from the CDC lab and the case notification from the jurisdiction, they often have difficulty matching the test result to the case notification due to differing identifiers. Many programs struggle with this issue. During the mpox response, the team used fuzzy matching, but a more straightforward strategy would be helpful.

We will now look at identifiers that get shared at each of these steps.

## Case notification – Specimen linkage



Specimens are sent to CDC for testing with information in 50.34 form, Global File Accessioning Template (GFAT), or CDC Specimen Test Order and Reporting (CSTOR)

- Data elements include
  - Patient Name
  - Public Health Patient ID, Specimen ID
  - Original Submitter Patient ID, Specimen ID
  - Intermediate Submitter Patient ID, Specimen ID

Council of State and Territorial Epidemiologists

Specimens that are sent to CDC, either from public health labs or commercial labs, are accompanied by either a 50.34 form, a Global File Accessioning Template form, or a CDC Specimen Test Order and Reporting (CSTOR) entry.

The fields across all of these forms are consistent and typically include the patient's name and possibly some demographic information. Depending on the source of the specimen, the following additional information may be included:

- Public health patient ID and public health specimen ID if the specimen is from a public health lab.
- Original submitter patient and specimen ID if the specimen is from a commercial lab, or if a commercial lab first processed the sample and passed it to a public health lab.
- Intermediate submitter information if the specimen was handled by another entity before reaching CDC.

## Example: CDC Laboratory Result (PDF)





**Centers for Disease Control and Prevention**  
Antimicrobial Resistance and Characterization Laboratory

Patient Name: **None Provided**  
Sex: Birthdate: Age: Date of Onset:

**Original Submitter IDs**  
Patient ID: **Direct Observation** Alt. Patient ID:  
Specimen ID: **DO-01** Alt. Specimen ID:

CDC Specimen ID: [REDACTED] CDC Unique ID: [REDACTED]

Suspected Agent: Gram-negative rod

| <b>Xpert Carba-R RT-PCR Assay</b>     | <u>Interpretation</u> |
|---------------------------------------|-----------------------|
| blaIMP detection by Xpert PCR*        | Negative              |
| blaVIM detection by Xpert PCR         | Negative              |
| blaNDM detection by Xpert PCR         | Negative              |
| blaKPC detection by Xpert PCR         | Positive              |
| blaOXA-48-like detection by Xpert PCR | Negative              |

\* Not all blaIMP-variants have been evaluated with the Xpert Carba-R Assay. The Xpert Carba-R Assay can detect blaIMP-1, -2, -4, -6, -10, and -11. Please see package insert for details ([www.cepheid.com](http://www.cepheid.com)).

Council of State and Territorial Epidemiologists

Regardless of who sends the specimen, CDC receives and tests the specimen and then returns results to jurisdictions in one of two ways – as an ELR message or as a PDF that can be accessed in CSTOR. Again, these results may go to the jurisdictional epi program, the public health lab, or both.

Despite the different methods of specimen submission to CDC, the two result delivery methods (ELR and PDF) contain the same information and have been cross-walked.

The results will always include the CDC specimen ID, a unique 10-digit identifier assigned by CDC. If the specimen was sent with a public health lab specimen ID, that ID will also be included in the results.

In the PDF, the IDs the sample was sent with are found on the first page (which is not shown here), and the CDC specimen ID is found in the second box shown on this screen.

In the ELR, the CDC specimen ID can be found in ORC-3.1, OBR-3.1, and SPM-2.2. If a public health lab specimen ID was sent to CDC with the sample, it will be returned in ORC-2.1, OBR-2.1, and SPM-2.1. Specimen IDs from commercial labs or intermediate labs will not be returned in the ELR.

## Case Notifications to CDC



Current Malaria v1.0 MMG:

| Data Element (DE) Name                      | Data Element Description                                                                             | Data Type |
|---------------------------------------------|------------------------------------------------------------------------------------------------------|-----------|
| Specimen Sent to CDC                        | Was specimen(s) sent to CDC?                                                                         | Coded     |
| <b>START: Specimen Type Repeating Group</b> |                                                                                                      |           |
| Specimen Type(s) Sent to CDC                | Type(s) of specimen sent to CDC                                                                      | Coded     |
| CSID                                        | CDC specimen ID number from the 50.34 submission form. Example format (10-digit number): 3000123456. | Text      |

Council of State and Territorial Epidemiologists

On the case notification side, all notifications sent to CDC via MMG include the Local Subject ID and the Local Record ID, which we have renamed Person ID and Case ID, which are person-level and case-level identifiers.

Some condition-specific message mapping guides request additional information to aid in matching, but there is inconsistency in what is requested and whether it is requested at all.

For example:

- The Malaria MMG asks if a specimen was sent to CDC and, if so, requests the type of specimen and the CDC Specimen ID.
- The Mumps, Pertussis, RIBD, Varicella, and FDD MMGs all inquire if a specimen was sent to CDC for testing, with this question preceding the epi lab repeating block.

As a result, typically, the case notification cannot be matched to the specimen:

- NNDSS cannot receive Personally Identifiable Information, so it cannot match on the name, even if the name was submitted with the specimen.
- Specimen IDs are not consistently received in case notifications, preventing matching on the CDC specimen ID.
- The person-level ID in a case notification usually differs from the person ID sent with a specimen, as the case notification is sent by a surveillance system while the specimen was sent to CDC from a lab system.

On this call we will consider capturing information more consistently to support this matching process.

## Specimen ID



Will be discussed in the Lab DEST to be part of lab template/epi lab block.

| Date Element Name | Date Element Description                                                        | Implementation Note                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Specimen ID       | A laboratory generated number that identifies the specimen related to this test | <p>If test conducted at CDC, populate this field with the CDC specimen ID. The CDC specimen ID is a 10-digit number (Example: 3000123456) and can be found in ORC -3.1.1, OBR-3.1.1, or SPM-2.2.1 of CDC's ELR message. It can also be found on page 2 of the PDF result returned in CSTOR.</p> <p>If the CDC Specimen ID is not available and the specimen was sent to CDC from the public health lab, populate this field with the public health lab specimen ID, which can be found in ORC-2.1.1, OBR-2.1.1, or SPM-2.1.1 of CDC's ELR message. It can also be found on page 1 of the PDF result returned in CSTOR.</p> |

Council of State and Territorial Epidemiologists

One method of capturing the CDC Specimen ID could be to include the specimen ID as a data element in case notifications, both in the lab template and in any epi lab block. While the specimen ID is already present in the laboratory template and in some epi lab repeating groups, the Lab DEST will talk about including it consistently. The goal of this is to leverage existing workflows rather than create new ones. However, like I said, this will be discussed in more detail during the Lab DEST in August. Our call about identifiers in July will not talk about the lab template or epi lab block, but I wanted to provide everyone with the context that this Specimen ID is being discussed in more than one data class.

## Specimen ID



Potentially can also be captured in Identifiers:

| Identifier Type                                                                        | Identifier Type Description                                                                                                   | Implementation Note                       |
|----------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|
| <b>If specimens were sent to CDC for testing, populate one of the two below fields</b> |                                                                                                                               |                                           |
| CDC Specimen ID (CSID)<br>(Code TBD)                                                   | CDC specimen ID number.<br><br>Identifier Value Example format (10-digit number):<br>3000123456.                              | Located in ORC-3.1.1 of CDC's ELR message |
| Public Health Specimen ID<br>(Code TBD)                                                | Public health laboratory specimen ID. Only use the public health specimen ID if this was sent to CDC along with the specimen. | Located in ORC-2.1.1 of CDC's ELR message |

- Would exist outside of the present lab workflow.
- Potential for inputting specimen ID specifically for CDC linkage if not transmitting through Lab.

Council of State and Territorial Epidemiologists

Separate from the lab template and epi lab block, the CDC Specimen ID and Public Health Specimen ID may also have a place in our identifiers value set, and that is what we'd like to talk about on the July call. This wouldn't be taking advantage of existing lab workflows, but would be a place for someone to type in a CDC specimen ID if they're holding a paper lab, or if a jurisdiction can't send specimen ID from their automated workflow. Would the group prefer for these IDs to exist in both places but not be required in both places? Are there times when this would need to be sent, but lab data would not be sent, so it would be beneficial for this to function more as a stand alone ID as part of the identifiers repeating group rather than the lab template and epi lab block?

## Poll 3 - Specimen ID



In Identifiers, should these specimen IDs be represented by one or two "Identifier Type" options?

- One
- Two

| Identifier Type                  | Data Element Description                                                               |
|----------------------------------|----------------------------------------------------------------------------------------|
| CDC or Public Health Specimen ID | CDC specimen ID number, or if not available, the public health laboratory specimen ID. |

| Identifier Type           | Data Element Description              |
|---------------------------|---------------------------------------|
| CDC Specimen ID           | CDC specimen ID number.               |
| Public Health Specimen ID | Public health laboratory specimen ID. |

Council of State and Territorial Epidemiologists

First we will talk about and then vote on if this should be one identifier type or two identifier types, where the CDC and public health specimen IDs are separate.

## Poll 4 – Specimen ID Core



Should these specimen ID value(s) be part of the minimum value set for Identifiers?

- Yes.
- No - The repeating group should support them when needed, not core for all conditions.
- No - This repeating group should not support them.

| Identifier Type                  | Data Element Description                                                               |
|----------------------------------|----------------------------------------------------------------------------------------|
| CDC Specimen ID                  | CDC specimen ID number.                                                                |
| Public Health Specimen ID        | Public health laboratory specimen ID.                                                  |
| CDC or Public Health Specimen ID | CDC specimen ID number, or if not available, the public health laboratory specimen ID. |

Council of State and Territorial Epidemiologists

Then we will discuss and vote on if the one or two identifiers we decide on in the previous poll should be considered "core" and part of our minimum value set for identifiers.

## Next Steps



**Join our second Identifiers call on July 19th, 1-2pm.**

Future DEST calls:

- **8/9 1-2pm ET – Laboratory**
- **8/16 1-2pm ET – Reporting Source**

The next large DSWG call will be August 30th, 1-2pm.

Council of State and Territorial Epidemiologists