Meeting: CSTE Data Standardization –Reporting Source -Data Element-Specific Team (DEST) Meeting
Date: August 16, 2024
Attendees: 59

CSTE (2)	PHAs (28)	CDC (23)
Gillian Haney	DSWG Co-Chairs	Alejandro Pérez
Meredith Eddy	Laura Erhart | AZ	Angie Agunloye
JMC(5)	Molly Crockett | MA	Austin Brown
Erin Kough	DEST Members	Carmen Pugh
Laura Lane	Andria Apostolou | IHS	Jeanne Ocampo
Natalie Raketich	Beth Isaac | NYC	Aruna Babu
Sam Spoto	Charlotte Picard | IL	Kristie Clarke
Susan Downer	Claire Stump | AK	Kristie Clarke
	Debra Rivera | NM	Kristin Delea
Unknown (1)	Ed Hartwick | MI	Lizzi Torrone
Latasha Terry	EJ Klingler | NYC	Luis Hernandez | GDIT
	Elizabeth Schiffman | MD	Melissa Pagaoa
	Hanna Chakoian | MN	Myrna del Mar González | GDIT
	Isaiah Francis | NM	Nicole Gallaway
	Janet O'Connor | SC	Noreen Kloc
	Jesse Ellis | MS	Peter Boersma
	Jill Krause | NB	Rachelle Oseni
	Kacee Redden | SD	Renee Boehmert
	Kiara Vicente |AK	Robert Pratt
	Kristin Goldesberry | IL	Sandra Roush
	Lunden | IN	Sara Johnston
	Maribeth Hoyle | OR	Shelby Lyons
	Mario Beltran | MI	Tricia Aden
	Elizabeth Lando-King | MN	
	Morrow Toomey | AK	
	Rob Laing | OR	
	Stella Tsai | NJ	
	Stephanie Moberg | NM	
	Steve Eichner | TX	
	Theron Huntamer |NV	

**SOME INDIVIDUALS DIALED IN; THE NAMES ARE UNKNOWN, AND THEY ARE LEFT OFF OF THE ATTENDANCE*

HIGH-LEVEL MEETING NOTES (*REFER TO MEETING SLIDES FOR MORE DETAIL AS NEEDED*):

Reporting Source

Meeting Ground Rules

- We welcome everyone's opinions and active participation from jurisdictions, CDC programs, CSTE membership, etc.
- As we work through data elements/classes, we strive for 80/20 consensus.
- This DSWG Post-Meeting Input Form is available at any time to provide additional input

In-Scope vs Out-of-Scope

- Focus on core data for surveillance and transmission to the CDC within selected data classes.
- We will work to identify standard definitions and value sets within that definition of core data.
- This is not a discussion of how to collect this data, frame questions in surveys or interviews, or implement these data standards. This is also not meant to impose limitations on how jurisdictions collect data.

Reporting Source

Potential Reporting Source Data Elements

- Listed here are a few data elements and concepts related to the reporting source in current Message Mapping Guides (MMGs).
- CDC programs are interested in identifying who is initially reporting cases to public health. This information helps CDC programs understand where cases are being identified and is often included in annual surveillance reports.
- The Reporting Source data element in the Generic v2 MMG asks for the "Type of facility or provider associated with the source of information sent to Public Health."
- Some MMGs also request a detection method to understand how the case was identified. This may include responses such as autopsy, self-report, or routine surveillance.
- In the background recording it was mentioned where data elements that address facility type or address are present in USCDI, eCR, and ELR fields. In the interest of time, we won't be going over that in detail again today but will very briefly show the ELR and eCR slides.

U.S. Core Data for Interoperability (USDCI)

- As a data class, Facility Information was first included in the USCDI v4. Relevant to reporting source, the Facility Information data class provides standards for both facility type and facility address.
- Facility Type is in USCDI Version 4 and describes the facility based on the category of service or resource available. Examples include but are not limited to hospital, laboratory, pharmacy, ambulatory clinic, long-term and post-acute care facility, and food pantry.
- Facility Address is Level 2 in USCDI, meaning it is not part of a USCDI standard yet but has relevant referenced standards and evidence of widespread adoption. It is the physical address for the facility, which includes the zip code.

Electronic Lab Reporting (ELR)

- In ELR, the main field that captures information on who is reporting is the Sending Facility information sent in MSH-4. This identifies the facility that transmitted the HL7 message and establishes the data's origin.

- There are other fields that collect information about the ordering facility or performing organization, but they should not be used as the facility who is reporting to public health.

Electronic Case Reporting (eCR)

- Information about the facility providing care is included in the electronic case report (eCR) in both the CDA and FHIR formats. When a patient's encounter triggers an eCR, it includes details such as the facility's name, type, and address to show who initiated the report.

Case Notifications to CDC Message Mapping Guides

- The generic v2 MMG has two related reporting source data elements.
- The first, “Reporting Source Type Code”, is described as the “Type of facility or provider associated with the source of information sent to Public Health”.
    - The value set has 35 concepts including hospitals, emergency departments, and laboratory.
- The second data element, “Reporting Source ZIP Code” is used to capture the zip code of the reporting source.
    - In 2023, Reported Source was populated 42% of the time and Reporting Source zip code was populated 15% of the time. CDC programs do not use reporting source zip code as it is rarely populated, and that variable is not one we're going to discuss as a potential core data element.

Top Reporting Source Types for Case Reporting in 2023

- Here is a summary of the top responses received for Reporting Source Type in 2023.
- Laboratory and Unknown make up three quarters of the responses, with hospital lab, other, hospital, and private physician's group office being the next group of most frequent responses.

STD NETSS Facility Type

- STD is one of the main programs at CDC using these data. They analyze it to understand where patients are being seen and who is reporting, which helps guide prevention activities, such as screening recommendations.
    - The NETSS Implementation Plan used by STD provides a data element description of: “Setting or health care facility where a person first received diagnosis, treatment or testing for STD or associated syndrome reported in this case report.” This focuses on the setting where the individual was diagnosed, which differs from the Generic v2 MMG’s focus on the facility that submitted the report.
    - This distinction is important as the reporting source and the diagnosing facility can be the same but may also differ.
- Other CDC programs are not as often utilizing the information in reporting source but have indicated during CDC's background research that it would be of more interest of them to know where a case is first being diagnosed or tested, as that information is more valuable for prevention activities.

Source Type Discussion

- The first item discussed was the “reporting source type” data element.
    - The original definition asks for the type of facility associated with the source of information sent to public health, which could mean the reporting facility as it is often interpreted, but is also being used to collect the diagnosing facility as used by STD.

  - Adding more specificity to reporting source, like the updated definition we have here of "Type of facility or entity reporting the initial case to public health authorities," would standardize the interpretation of this data element and still allow the potential for mapping from the previously mentioned ELR or eCR fields.
- However, to collect information on where the person first received diagnosis/testing/treatment, which is what is of interest to CDC programs, a separate data element would need to be used. Here a proposed data element of "Case Identification Source Type" with a definition of the type of facility or entity where the person first received care (meaning diagnosis or testing or treatment, whichever applies first for the case) for the condition, has been created
- The discussion was opened regarding the important of collecting and transmitting to CDC, both who reported the information as well as where the case first received testing/treatment/diagnosis, or if only one of these data elements is sufficient? It was noted that the discussion will be eventually voting on if either or both should be core. Also, the next slide focused on discussing whether these should allow for single versus multiple responses, so at this time participants were asked to ignore the nuance of capturing "who reported the initial case" versus "anyone who reported the case" in the definition. This first discussion focused only on if it's important to capture and transmit both who reported the case to public health and where the case received diagnosis/testing/treatment or one versus the other or neither.

  Discussion
  - Steve Eichner noted that it is possible for someone to visit a primary care provider (PCP) and be referred to a laboratory for sample collection and testing. The test results might not come directly from the lab but rather through the PCP. He suggested adding more clarity to the definition, particularly in distinguishing where the test is conducted versus where the results are reported. The focus would be on who placed the order rather than who processed it. It might make more sense to consider the entity type rather than the individual.
  - Beth Issac requested clarity because it was mentioned that the CDC doesn't use the reporting source type much. If that's the case, should we even be sending this data element if it's not being utilized? Can anyone clarify this?
    - Angie Agunloye noted that CDC is keeping the reporting source because it's easier to obtain that information from various sources. If it's something that's simple to send, they would want to keep that option open for states. However, the primary interest remains in where the case was identified.
    - Beth Issac commented that, speaking from the NYC perspective, specifically for VPDs, we've routinely collected and reported the reporting source. However, we aren't currently set up to collect and easily report on the case identification source type. If that's what's needed, we'd need to adjust on our end. It's helpful to know exactly what CDC needs so we don't provide unnecessary data.

Reporting Source Discussion

- The discussion moved on to whether the responses should allow for a singular value or multiple values.
  - For Reporting Source this would be a distinction between transmitting who reported the initial case versus anyone who reported the case.
  - For Case Identification Source this would be a distinction between where the case first received care versus anywhere the case received care.

Discussion

- Ed Hartwick noted that, aside from Michigan, we've generally considered the first person or entity responsible for reporting as the key focus. We haven't tracked everyone involved in reporting to the CDC, which would require changes to our processes. For us, it might be easier to use a single value rather than a repeating one. Clarification would be helpful—are we talking about the first to report or the first to diagnose? Could this be updated if new information emerges? Also, some data elements are hard to track, especially with initial reports from reference labs, followed by updates from other labs or ECR. These factors can cause changes over time.
    - Sam Spoto commented that in the chat there was a comment that reporting sources are easier to identify than the identification source type. Determining who initially reported may require more digging into the reporting form, especially if the case was reported by multiple sources. Oregon agrees that it would take significant effort to include all possible reporters in a repeating element. It sounds like focusing on the initial report might be easier for implementation, or simply noting where the case was first seen.
- Morrow Toomey had a question regarding elements that might be easier to extract from ELRs, like the case identification source type. For example, while the ordering facility isn't the reporting source, it's something we have access to and might be easier to map and send. Morrow was recently in a conversation about cases identified in a department of corrections, where many reports come from a different lab, which isn't helpful in identifying where testing or diagnosis occurred. We rely on the ordering facility for this. They wondered if focusing on the ordering facility, as Steve mentioned, might be more helpful and easier for other jurisdictions to collect.
    - Sam Spoto noted that in the chat, EJ mentioned that in New York City, for STIs, they also use the ordering facility to indicate where a person was diagnosed, tested, or treated. EJ confirmed that this is how they currently fill in that information. It's good to know that this approach is already being used for this type of question.
- Sam Spoto called out that in the chat, there was a question about what information is useful. Ordering provider or source can often be more reasonable and relevant than the reporting facility, and it's easier to obtain than the case identification source type. With the ordering provider, we're assuming there's an ELR linked to the case, right? But if someone isn't lab-tested, we're trying to cover that with the case identification source type. The discussion was expanded to hear thoughts on how to balance using something that can be obtained electronically in cases where there is an ELR, which is most of these cases, while still making this field valuable when there isn't a lab test or an ELR.
    - Rob Laing commented that they've heard from the HAI MDRO team that obtaining the ordering provider is crucial, as it's often a long-term care facility where these infections occur. However, it's been challenging for us to reliably provide that information because it doesn't always get captured well in our ELRs
    - Ed Hartwick noted that from the perspective in Michigan, when sending data to the CDC, we try to populate as much information about the ordering provider as possible from the ELR. If the exact match isn't available, we may use a substitute that is approved by our programs. However, this information can still be variable and subject to updates. While it's often included in the ELR by default, it can change during follow-up.
- Beth Issac commented on the comparison between the ordering provider and the case identification source type. There can be some nuance in how she interprets this. For example, someone might see a provider for symptoms but only get tested on a subsequent visit.

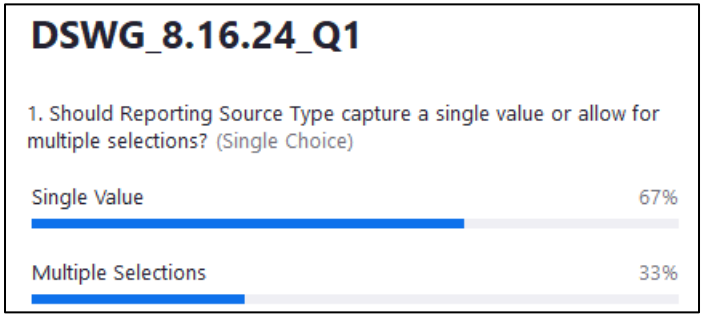
This means they were initially seen for the condition earlier, but the testing was reported later, complicating data reporting. If the focus is really on the reporting provider, then phrasing it that way might make it clearer and more helpful.

- Sam Spoto asked if this report would be different than the first one?
- Beth Issac replied that yes, she thinks the distinction might be more about the first source. She noted she was a bit confused about the difference between the reporting source type and the ordering provider. The ordering provider is associated with the lab, whereas the reporting source might be the provider who reported the case.
- Angie Agunloye added that it might be useful to provide an example for the reporting source. Sometimes, intermediate agencies report on behalf of entities. For instance, multiple outpatient settings might use a single system to send information to a public health agency. In this case, the reporting source is the entity that submits the information to public health, not necessarily the one who placed the order or conducted the testing. This is one of the nuances that can arise with the reporting source, as it can be any entity reporting on behalf of others.

- Morrow Toomey noted that with our Department of Corrections, they might send lab requests to LabCorp or a public health lab, which are the entities reporting on behalf of those tested. The Department of Corrections is the ordering facility. They work in infectious disease, but STD reporting is handled by a separate section within epidemiology. That section is more familiar with using this field. They've been considering how we handle case identification source types. For example, an ELR might show 'XYZ Correctional Center,' but we need to map that to a type of facility. This adds an extra step, as we must translate names like 'Providence Hospital' into 'hospital.' Currently, people select from a dropdown based on the ordering facility, but they're looking for ways to automate this process. Many of our cases are identified from labs, and while manual entry for early cases is manageable, automating the mapping for other cases would be more efficient. Any ideas on how to streamline this would be helpful.
- Ed Hartwick commented that he agreed with what Angie mentioned. Additionally, it's worth noting that, like most places, we primarily receive notifications from laboratories when positive results come in. We refer to the ordering provider, but they are often not listed in the reporting sources because they typically aren't the ones who report unless the situation is severe. This might be another point to consider.
- EJ Klingler wanted to follow up on the challenge of finding the provider type. As mentioned earlier, we use the ordering facility for case identification, like how SDI collects reporting sources. However, mapping and maintaining this information is quite difficult. We report it by finding a type and keeping it updated, but with new facilities and aggregators who report for multiple entities—such as within community hospital systems where standalone clinics are included—it's a complex task. Even though we use this approach, it's not perfect and finding a provider type for these ordering facilities remains challenging.
- Sam Spoto took a moment to summarize the discussion: Mapping a facility type from an ELR is likely difficult because it involves translating facility names into types. Many people are already using the ordering facility for case identification source type, which works well with ELRs but may need updating during further case investigation. We didn't delve deeply into the single versus multiple responses, but it seems there's a general preference for a single response, such as identifying the initial report or where the person was first diagnosed, tested, or treated.

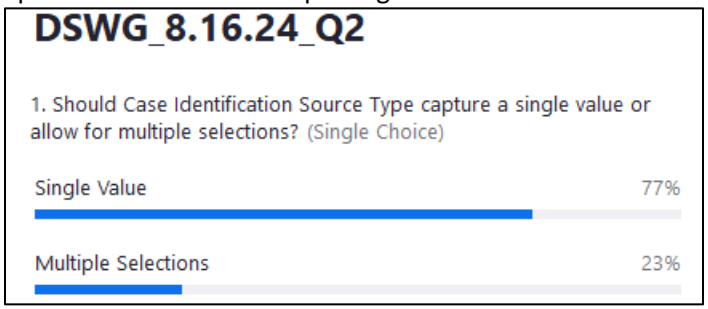
Poll 1 (n= 21) – RS Single or Multi-Select

- The first poll was focused on if these definitions should capture single values or multiple values.



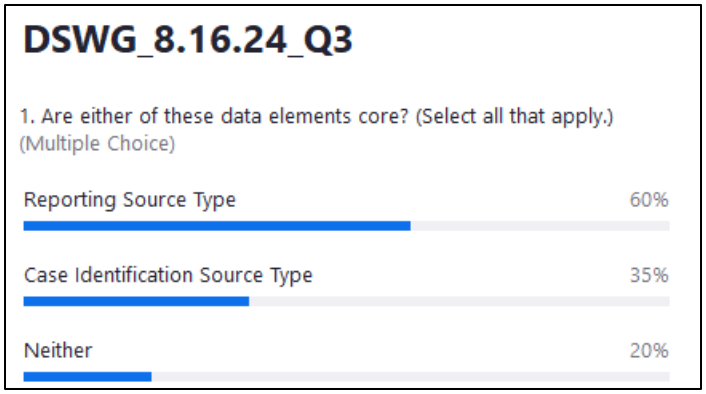
Poll 2 (n=22) CIS Single or Multi-select

- This poll was focused on the case identification source type, a single value would capture where the case first received care, while a multi-select option would allow for reporting all locations where the case received care.



Poll 3 (n = 20) Source Type Core

- On this slide the single select definitions were written. This poll is a select all that apply poll, and its asking if either of these data elements should be considered core data for case-based surveillance. This means it should be transmitted for all conditions, but could be blank or unknown if not available or applicable.



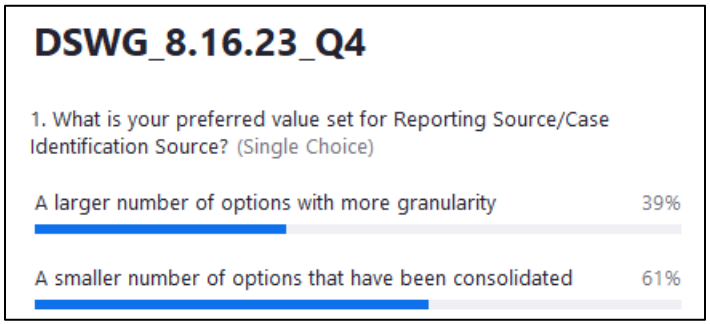
Current Reporting Source Value Set (35 values)

- Finally, we will discuss the value set for Reporting Source or Case Identification Source. The current value set is shown on this slide and encompasses 35 values. We can come back to this slide if we need.

Potential Updated Value Set (13 values)

- This is a proposed consolidated value set that is comprised of 13 values. The sub-bullets on this slide intend to show where some of the values from the previous value set that are missing may be categorized. In the interest of time, we will not be working to agree upon what this value set is exactly, but we are interested if there is a preference for a larger, more granular value set like what currently exists versus a simplified version such as what's shown on this slide.
- The poll was launched to gauge participants initial thoughts on the value set.

Poll 4 (n= XX) RS/CIS Value Set

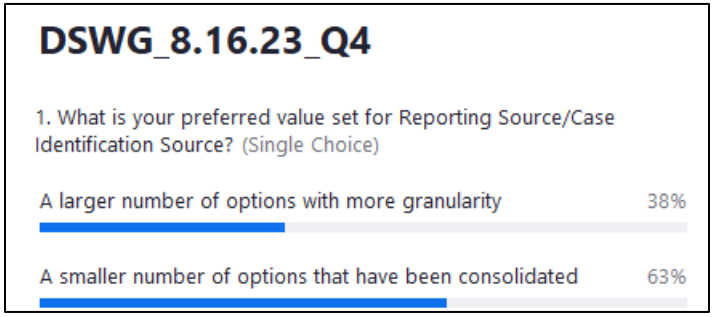


Discussion

- Sam Spoto noted a suggestion in the chat about including the Federal Bureau of Prisons in the report. Sam requested that when the draft document is shared for participants to please review it and provide feedback on specific values in these sets. Let us know if you have opinions on additions, removals, or any other changes.

- Steve Eichner noted that in looking at the consolidated value set, it seems there might be significant overlap. The concern is distinguishing between a more consolidated list and a more specific list. Also, there are overlapping elements because not everything on this list is at the same tier. He noted there could be confusion if we mix elements from different tiers; it's an important balance to consider if we move forward with a more consolidated list.
    - Sam Spoto agreed on all accounts and noted that there is a lot of overlap when these items are consolidated into broader categories.
- EJ Klingler commented that she voted for the more consolidated list because it's challenging to categorize the many reporting facilities accurately. For example, the outpatient category is well-organized in the consolidated list but is missing from the 35-option list. This makes it difficult to understand the primary role of a provider, especially if they have multiple locations with different functions but report together. Overall, she believes the consolidated list is more manageable and will provide more useful information.
- Kristie Clarke noted understanding of the principle behind the consolidated value set—it's intended to be easier to use and more straightforward. However, it needs to be implemented very thoughtfully. For example, combining military and Indian Health Service into the same category can be problematic, as they are quite different despite both being federal agencies. This illustrates the challenge of consolidation. Additionally, colleagues working with people experiencing homelessness were surprised to see that Federally Qualified Health Centers and healthcare for the homeless clinics were not included, even in a more detailed list. Striking the right balance is challenging, and I respect the effort involved.
- Elizabeth Schiffman was concerned that blood banks were categorized under 'other.' For vector-borne diseases like West Nile virus, we rely heavily on blood bank reports of viremic donors. She would appreciate it if blood banks could be reclassified, but generally supports the consolidated list. If keeping blood banks in 'other' is necessary for consolidation, she can accept that.
- Steve Eichner noticed that in the consolidated list, Indian Tribal governments are categorized under 'other.' We've put significant effort into recognizing tribal governments and partners over the past few years, so it might not be ideal to place them in the 'other' category. It seems we might need to find a balance between the current list of 18 and the 35-option list. I'm not sure what the ideal number is, but it might be worth considering a middle ground.

- ***After the discussion the poll was launched a second time***



Detection Method

Case Notifications to CDC Message Mapping Guides

- In addition to the Reporting Source information in GenV2, specific MMGs include the "Detection Method" data element. There are eight MMGs that ask which process was used to identify the case, using similar value sets.
    - COVID and STD MMGs ask about the process by which a case was first identified.
    - RIBD and VPD MMGs ask how the case was identified without specifying if it was the first identification.
- Only the COVID MMG allows for multiple entries for Detection Method.

Detection Method Value Sets

- For comparison, here are the value sets for detection method for the eight condition-specific MMGs.
- There are eight MMGs but only four value sets between them.
- This includes methods such as routine surveillance, diagnostic testing, epidemiological or cluster investigation, and screening.

Detection Method Discussion

- CDC would like to standardize how detection methods are transmitted across conditions, whether it is considered a core data element. The discussion was opened up with similar questions to reporting source- should detection method allow for a single response that represents the process by which the case was FIRST identified by public health, or "select all that apply" to represent any process by which the case was identified by public health? Also, should this be core or condition specific?

  Discussion
    - Elizabeth Schiffman noted that this ties into the earlier discussion about whether to use single or multiple responses. If we're trying to determine the initial process or source, it seems like it should be a single option. Multiple responses might not be necessary if we're aiming to identify the first instance. However, depending on the context, a broader approach might also be useful.
    - Gillian Haney wondered how these data for detection methods are used. This seems like a data element carried over from previous years when it was crucial to understand where cases were identified. However, she was unclear on its current utility.
        - Sandra Roush replied that it's important to recognize that there are differences in the level of authority among various detection methods—such as serology, culture, or PCR from serum, blood, or spinal fluid. The choice of method affects our confidence in meeting the case definition criteria. For instance, different methods used in VPDs and CIRD impact our assessment of the case. The detection method value sets shown on the screen include various methods, such as those used for COVID-19, which are more extensive. If autopsy is associated with the detection method, it relates to IMV 290 and 291, which pertain to test type and specimen source, offering a different level of authority compared to routine physical exams. This usage is based on contextual cues and, of course, should be aligned with the preferences of the wider group.
        - Elizabeth Schiffman noted that uncertainty of understanding why the CDC needs this level of granularity. From her perspective, this detailed information is useful at the local or state level, but it's unclear how it benefits the CDC given the variation across jurisdictions.
        - Sandra Roush clarified that her input wasn't about insisting that this level of detail is necessary or crucial. She was addressing how the CDC might use the data if it is available. For pathogens and NCIRD, having this data is helpful, but not essential. The

real question is how the CDC utilizes it if provided. This is different from data elements like vaccine history, which we consider critical for our work.
    - Gillian Haney noted she was trying to determine whether this should be considered a core surveillance element. A key concern is the burden of collecting specific data elements and ensuring that if they are sent to the CDC, they are used and critical for their work. It might be worth evaluating whether this data meets the threshold for being essential.
    - Molly Crockett commented that since some of these options overlap with what we discussed earlier, we might first need to finalize the value set. Once that's settled, we could then focus on identifying any specific data needs that apply as core values or core questions and address those in more detail.

Next Steps

- Next steps for the Reporting Source DEST are pending, but the Social Determinants of Health DEST will begin in September, during this time slot.
- Upcoming team meetings
    - 9/13, 1-2 ET, Laboratory Reporting Call #2
    - 9/20, 1-2 ET, first Social Determinants of Health DEST call
        - Residence Type (as related to homelessness/incarceration)
    - Link to DSWG: Data-element Specific Team Interest Sign-Up
- The next large DSWG call will be on August 30, 2024.