

Meeting: CSTE Data Standardization – Industry and Occupation -Data Element-Specific Team Meeting

Date: October 20, 2023

Attendees: 70

CSTE (4)	PHAs (25)	CDC (31)
Annie Fine Becky Lampkins Cailyn Lingwall Gillian Haney	<u>Co-Chairs</u> Laura Erhart (AZ) Molly Crockett (MA)	Abhilasha Saxena GDIT Allison Miller Andre Berro Angie Agunloye GDIT
JMC (4)	Annalyse Bergman AL Benjamin Schram Cathy Wood MO Cher Levenson WA Collin Morris WI Emily Sparer-Fine MD	Audrey Reichard Darlene Wells Jasmine Varghese Jennifer Cornell Jennita Reefhuis Jenny Couse
Unknown Affiliation (6)	Erik Zabel MN Faith Baldwin AZ Fran Lendacki CDPH Hanna Chakoian MN Isaiah Francis NM Jim Sainsbury IN Karene Thomas-Watts CO Kate McCoy WI Ketki Patel TX Kyra Kofodimos IN Lunden Espinosa IN Margret Watkins WV Maria Jaurretche DC Marija Borjan NJ Misty Johnson WI Vajeera Dorabawila NY	Katherine DeBord Katie Dodd Kaylea Nemechek Kevin Lan Lauren Lambert Layne Dorough Lesliann Helmus Luis Hernandez Maddie Stewart Marie Haring Sweeney Matt Stuckey Myrna del Mar Gonzalez Otto Ike Peter Boersma Robert Pratt Sam Williams

	Ximena Vergara (CA	Sandy Price Sandy Roush Stacey Marovich Sundak Ganesan Walter Alarcon
--	----------------------	---

**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):

Meeting Ground Rules and Updates

- 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.

Data Element-Specific Teams (DEST)

- DEST member expectations – members are not expected to write documents, lead meetings, or prepare slides. The goal is to keep members engaged in conversations around content during and between meetings.
- There is an expectation of active participation in polling, document review, discussion, and other input during and between meetings.
- **[Action Requested]** Please sign up for upcoming data elements on the [DEST Sign-Up here](#).
- After SOGI or Industry and Occupation wrap up, the next data elements will be
 - Exposure/Risk/Underlying Conditions
 - Immunization History

Core Data: In Scope for DSWG

- Focus on core data for surveillance for transmission to CDC within selected data classes.
- We will work to identify standard definitions and value sets within that definition of core data.

Core Data FAQ

- Identification as core does not mean it is required to fill in or send an answer for every core data for every case; it is okay to leave it blank or unknown.
- Data, elements, and data classes should be considered core if they are relevant for many conditions.
- A good way of thinking of this is if we were to create a GenV3 MMG, would it make sense to include that data element here as core?

Industry and Occupation Background

**Here is a link to the [pre-recorded background video](#) and the [transcript/PDF](#) information for reference as needed*

Potential Core Data elements

- Here is a brief overview of many data elements related to industry and occupation used either in CDC NNDSS message mapping guides (MMGs) or established in other standards. In bold, current industry and current occupation are likely to be the most familiar given they are present in the standard template used across many MMGs.
- Industry refers to the setting where someone works, such as restaurant, school, or hospital. Occupation refers to the kind of job someone has, such as pilot, nurse, or teacher.
- Data elements such as "food handler" and "case patient a healthcare worker" is used across multiple MMGs. There are also data elements that exist in other standards but are not used in any MMGs. An example is "Usual industry or occupation," which represents the type of work (paid or unpaid) done by a person for the longest amount of time during his or her life.

Case Notifications to CDC, Message Mapping Guides (MMGs)

- The template of four data elements seen in the table on this slide is currently used across many message mapping guides to collect information on occupation and industry. The same occupation may be found across multiple industries, so it is valuable to have both pieces of information to get a better understanding of occupational risk and exposures. For example, a nurse in a hospital and in a school are likely to have different potential exposures.
- The list on the top of the slide indicates which MMGs this template is currently part of. Jurisdictions can send both narrative text (in a text data element) or a standardized census code (in a coded data element). This template in the NNDSS MMGs currently uses the value sets from the 2010 census codes for industry and occupation.
- It is recommended that jurisdictions accept and send the free text and coded values when they are available but should still accept and send free text if the codes are not available. Jurisdictions can run the free text thru NIOCCS for coding before sending it to CDC, but if this isn't possible, jurisdictions can send the free text values for CDC to code through NIOCCS after receipt.
- NIOCCS stands for the NIOSH Industry and Occupation Computerized Coding System. Briefly, NIOCCS is a free web-based API that converts industry and occupation text into standardized codes and can output codes from the Census coding schemes, or NAICS and SOC. More information on NIOCCS is available in the April 2023 eSHARE call recording, available in the NNDSS Technical Resources Center on CDC's website.

Case Notifications to CDC, Message Mapping Guides (MMGs) (continued)

- Several NNDSS message mapping guides ask specific questions about the case patient working in a particular occupation or industry. The most common asks whether the case patient was a healthcare worker. This is used in 6 guides.

Prior Work on Industry and Occupation

- In general, NIOSH recommends use of the HL7 templates provided for v2.9, CDA, and FHIR data exchange, discussed previously.
- Important things to remember when collecting industry and occupation data include the following:
 - Industry and occupation are related, but different. It is important to collect both.
 - It is important to collect free text descriptions and not use drop down lists – limiting the response options also limits what we can learn. Drop downs make it difficult to capture emerging occupations/industries.
 - It is best to collect detailed information. For instance, "Engineer" is okay, "Civil Engineer" is better.

- Each occupation and industry have a specific standardized code assigned to it, and using these codes assists in data analysis, integration, and interpretation.
- There are also important things to remember when coding industry and occupation data:
 - It is important to decide which classification systems to use before coding the data. Descriptions of the code systems can be found on NIOSH's website.
 - Critically, use an autocoder, such as NIOCCS, to code the data.
 - Software applications, or “autocoders,” assign industry and occupation codes to free-text descriptions using the classification system you choose (NAICS and SOC, or Census).
 - There are a number of benefits to using an autocoder:
 - Autocoding is much faster than manual coding.
 - There is greater consistency in codes from one record to another.
 - Autocoding does not require the extensive training and experience needed for accuracy with manual coding.
 - Finally, data can be coded as it is collected, or as a batch after collection is complete.

Prior Work on Industry and Occupation (continued)

- NIOSH also recommends the use of NAICS and SOC over use of the Census codes for a number of reasons:
 - NAICS/SOC are more broadly used than the Census codes as standards for coding industry and occupation data.
 - They are hierarchical while the Census codes are not, and this hierarchy adds useful information about relationships.
 - NAICS/SOC provide more detailed information than the Census codes, which are intentionally less detailed to protect privacy.
 - NAICS/SOC are unrelated, independent classification systems – so coding is clearer in cases when only industry or only occupation is available.
 - And finally, mapping (or cross-walking) NAICS/SOC to Census codes is a simple process. NAICS/SOC are the foundation of the Census codes. Mapping from the more detailed NAICS/SOC to the less detailed Census codes can be done when it is helpful.

Industry and Occupation for Discussion

Potential Core Data Elements

- The first two elements in this table for discussion should look familiar as they are what was in the standardized template in many of the message mapping guides the current industry and current occupation. A final template I wanted us to talk about is maybe an idea of setting up a data element to not necessarily a core data element for surveillance in terms of it would be generic for all conditions. But is there a way we can format a question to ask about specific occupations, in specific settings, for specific cases, where, if a question about occupation is going to be asked, here's the format that it should be asked in.

Potential Core Data Elements (continued)

- Currently, these first 2 data elements of current industry and current occupation in the message mapping guides are defined just as the cases current industry or cases current occupation in previous work this group defined a calculated case counting date, which I'll explain on the next slide.
- One discussion point for this group would be if it's better to center the definition of current on the CCCD, which is a specific date associated with the case. For value sets for these first 2 currently utilized are the census coding schemes, and we're going to open it up for discussion later, if an adjustment should be made to use NAICS/SOC as recommended by NIOSH. Then, after that discussion, we'll engage in more discussion about this potential wording for a template for use in collecting condition-specific information on a particular occupation, such as food handler or healthcare worker.

CCCD

- Before we open the discussion. I wanted to go over what the CCCD is. It's something that in previous work the data Standardization Work group has defined a calculated case counting date for use by jurisdictions. It's defined as the earliest available of 6 component dates which you can see listed on the right side of the screen. Here those would be the symptoms, onset date, the clinical diagnosis date, the earliest specimen collection date associated with a positive lab result, the earliest result date of a positive lab result, the date first received by a public health agency, and the date entered, or record initiated.
- The important feature of the CCCD is that it's possible to calculate it for all cases, even if all components aren't available for each case. The first draft of the pregnancy standardization efforts recently received comments from members of that data element specific team. And there was agreement in utilizing CCCD to define the time of event for pregnancy status.

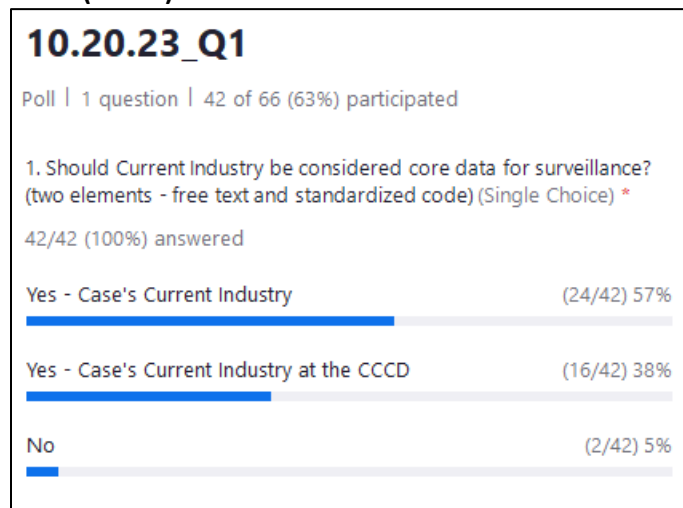
Current I/O

- Some of the first topics to focus on for discussion are the current industry and current occupation data elements. So, is that standard template present across MMGs working for everyone? Does this definition need a time element concerning the case, such as CCCD, we're going to hold off on talking about value sets until the next slide. So, first, focus on these data elements and their definitions.
- **Discussion Summary:**
 - There were several comments for an agreement that when using the term "current," it will need a time context associated, whether that is CCCD or another benchmark for a date.
 - The group must agree on what "current" means, which was acknowledged to be more challenging for longer-term conditions.
 - There was a note made that the word "current" could also be removed.
 - CDC commented that it's important to distinguish that we're trying to distinguish which occupation is relevant to the disease event.
 - Additional comments included the distinction between "ever," "current," or "both" for occupation and industry.
 - JMC offered a reminder that this is specific to how the group has defined a core data element for surveillance and should be similar across all conditions. Jurisdictions will still be able to capture additional information within their systems as needed, and we are focused on standardizing those data elements that could be submitted to CDC.
 - A note that even though we are not discussing implementation, individuals are encouraged to consider what that might look like for a repeating format and how your surveillance systems are working to collect this information.

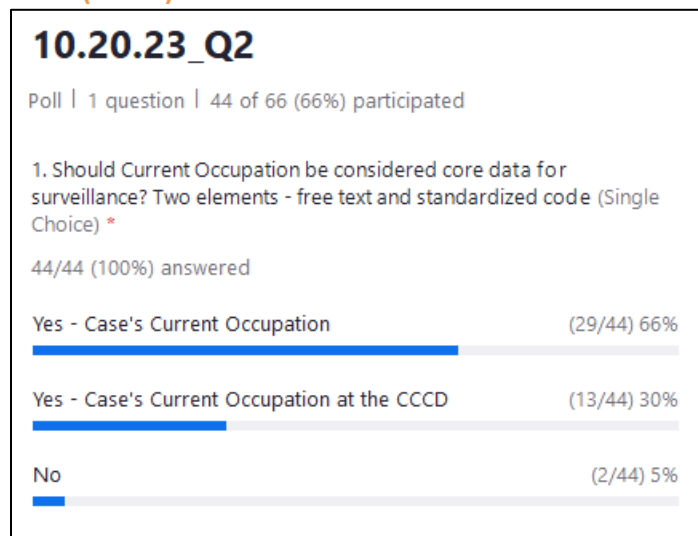
- CDC commented that NCIRD is excited to help meet implementation for implementation of inclusion of industry and occupation in a standardized, harmonized way for MMGs for COVID and other diseases. The template was the same across each of the pathogens. Also, agreement that having a date associated with this repeating group would be good.
- There were several questions on whether states have implemented some of the MMGs, how many states, and the percentage of data completeness in this field for eCR or case investigation.
 - JMC noted that responses were varied from the in-between meeting poll received before this meeting. It varied by jurisdiction whether it was collected for all conditions as a free text field, whether it was only for some conditions, whether it was collected only for some and as a drop-down field, or not collected at all. The variety of responses received indicate collection of these elements likely varies widely across all jurisdictions.
 - CSTE noted that the work of the DSWG is intended to be disease agnostic, so the answers for data completeness and total use by states are challenging when looking at a broad spectrum of diseases. This will vary significantly by jurisdiction and disease. This information is naturally more relevant for some diseases than others.
 - CDC also noted that we are trying to standardize certain data elements we should be collecting. The goal is to consider social determinants of health and other factors that affect why people get a disease. The most critical issue here is capturing data more consistently and generalizing it to capture it better when considering disease transmission and risk factors.
- Another jurisdiction commented that they use repeating blocks across different diseases in their surveillance system, which works well overall. The data is not always completed, but it's good to have a standardized format and place to put the information when it's available.
- There was a general discussion on the advantages of using codes for occupation and industry.
- A comment was made that industry and occupation are also relevant for conditions outside of infectious disease, such as environmental exposures to pesticides which have been linked to industry and occupation. Most relevant for pesticide exposure might be the job they worked currently when the exposure occurred, or for chronic exposure in someone who may have multiple jobs, what are all the jobs they work.
- JMC noted the definition of "Current" Industry/Occupation being aligned with a specific date, like CCCD, is not intended to capture "did someone work at that job on that date", but rather "what occupations was a case actively part of as of that date".
- The discussion shifted back to the use of CCCD as a time element within the definition of "current" industry/occupation with concern that utilizing CCCD may be too precise of a metric for date, and that the CCCD may also not be ideal because that date can change over the course of case investigation. Using wording like "time of event" may allow more flexible capture of this element.
- It was brought up that maybe all occupations ever should be captured, rather than just current, as some conditions like TB can have exposure occur long ago rather than acutely associated with the condition. JMC and CDC reminded the group that this effort is not to standardize all industry and occupation information asked for any condition, but rather to standardize the data elements that should be collected in every condition. The work here does not preclude the ability to ask for condition-specific industry/occupation information for a particular condition when needed.

Current I/O: (continued)

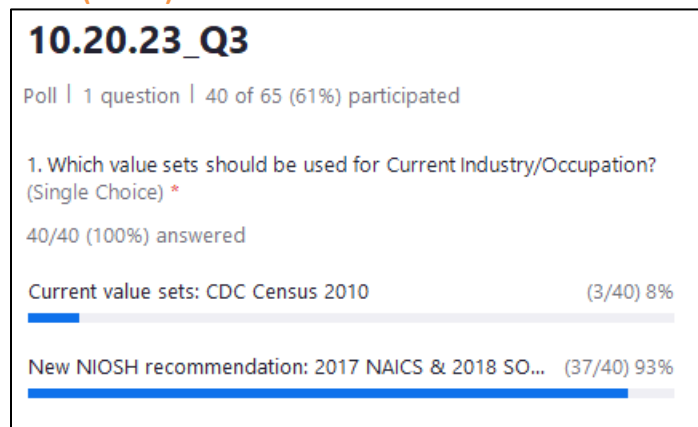
- Here, the conversation moved on to discussions around value sets. Currently, the MMG definition uses census value sets for industry and occupation. NIOSH recommends using the NAICS/SOC value sets. In NAICS, the autocoder of both census, NAICS, and SOC codes are available. JMC wanted to know whether we should switch the value set to the recommended NAICS/SOC value sets. Or would there be a purpose to stick with the census codes currently utilized in the standard template?
- Discussion Summary:**
 - Several comments were endorsing the use of NAICS/SOC
 - It was noted that one of the main reasons jurisdictions prefer this value set is that it is hierarchical and easy to aggregate upwards. Using this set makes it easier if you are dealing with small cell sizes and the ability to get into more detail. The standard 6-digit codes give a lot of specificity and provide more flexibility over the census value sets.
 - CDC commented that the MMG coding is moving towards using NAICS/SOC as well and this is currently in work.

Polling Results**Poll 1 (n = 42)**

Poll 2 (n = 44)



Poll 3 (n = 40)



Next Steps

- We will hold another Industry and Occupation meeting on November 17 and discuss.

- The definitions of “Current” Industry/Occupation, and whether the definition should include reference to a specific date associated with the case, like CCCD, or include a less defined time period such as “time of the event”.
- Potentially creating a template for a data element that can be utilized by specific conditions to capture condition-specific occupation information. This would capture information such as if someone is a healthcare worker, food handler, or childcare worker during a specific point of exposure or potential transmission.
- The next full DSWG meeting is scheduled for December 8.
- Feel free to sign up for other data element-specific teams on the [interest form here](#)
- Additional input is always welcomed from DSWG members on the [form located here](#).