Sexual Orientation and Gender Identity

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

The collection of sexual orientation and gender identity (SOGI) data in a structured way is essential to public health surveillance in that it allows for the identification of specific health care needs and health disparities affecting the lesbian, gay, bisexual, transgender, queer/questioning (LGBTQ) community. Both gender identity and sexual orientation are concepts that can change over time; in order to maintain their usefulness, it is important that these data elements have the flexibility to adapt to societal changes. This flexibility will allow for the accurate capture of this information as this space continues to evolve.

While these concepts have often been treated in the past as risk factors or potential exposures for various conditions, they are now being considered as key demographic data elements used to better target public health communications to the appropriate audiences.

Several efforts in recent years have sought to define these important, yet complex, concepts for use in public health and surveillance:

- The Gender Harmony Project (GHP)[1] is an initiative consisting of terminology and vocabulary experts, under the umbrella of the HL7 organization, that aims to capture "sex" and "gender" in health models using proven methods. The goal is to develop recommendations for how to model sex and gender with defined data elements, value sets, and code systems for implementation in EHRs. The recommendations received comment and broader buy-in from interested parties before being published in 2022 as implementation guides.
- The National Academies of Science, Engineering, and Medicine (NASEM) evaluated existing measures and methodological issues associated with measuring sex, gender identity, and sexual orientation in various settings. The findings were released in a 2022 report, titled "Measuring Sex, Gender Identity, and Sexual Orientation"[2], that offers recommendations to standardize the collection of this information for surveys and research studies, in administrative settings, and in clinical settings. The report provides suggested standardized language for survey questions that ask about these nuanced terms in order to obtain reliable responses and improve the overall data quality.

[1] Robert C McClure, Caroline L Macumber, Clair Kronk, et al. Gender harmony: improved standards to support affirmative care of gender-marginalized people through inclusive gender and sex representation. Journal of the American Medical Informatics Association, Volume 29, Issue 2, February 2022, Pages 354–363, https://doi.org/10.1093/jamia/ocab196

[2] National Academies of Sciences, Engineering, and Medicine. 2022. Measuring Sex, Gender Identity, and Sexual Orientation. Washington, DC: The National Academies Press. https://doi.org/10.17226/26424.

- The Office of Management and Budget (OMB) also made recommendations[3] on best practices for collecting sexual orientation and gender identity data, specifically as it applies to federal statistical surveys. The focus of this effort was the conduction of primary data collection by the federal government, where the methodology is highly standardized; that is, the questions are ordered and administered in the same way every time. The report does not mandate any particular approach or create any new requirements regarding collection of these data elements, but because SOGI variables are high-profile data elements that undergo additional scrutiny during OMB review, any approach taken by a federal agency must be fully explained and justified to be approved.

Following the release of these pieces of guidance, multiple groups have endeavored to evaluate the data collection methodology surrounding SOGI data elements and to recommend improvements in those processes. Among and within public health jurisdictions, there is variability in the data collected and available for analysis. What is collected can depend on what is acceptable to the jurisdiction's population, what is allowed by its policies and regulations, and what functionality is available in its surveillance systems. For example, some jurisdictions may ask specifically about pronouns or transgender status, while other jurisdictions do not. Other examples of variability among and within jurisdictions can be seen in the use of a coded value set, the values within such a value set, and whether a free text field is used to capture additional responses. Free text fields allow for these nuanced concepts of gender and sexual orientation to be described according to the person's preferences but may make data analysis more difficult. One commonality often seen within case-based surveillance systems is that jurisdictions will typically collect some variation of sex and/or current gender at the subject or person level, while sexual orientation and/or gender identity may be condition-based and collected at the case level.

Sexual orientation and gender concepts are included to varying degrees in the Message Mapping Guides (MMGs) used to transmit case-based surveillance data to CDC's National Notifiable Diseases Surveillance System (NNDSS). Table 1 shows the data elements relevant to the SOGI data class and their current description and value sets in various MMGs.

Table 1. SOGI Data Elements Currently in Message Mapping Guides

Data Element	MMGs	Description	Value Set
Subject's Sex	Gen V2	Subject's current sex	1. Female 2. Male 3. Unknown
Birth Sex	Hepatitis	What sex were you assigned at birth, on your original birth certificate?	1. Female 2. Male 3. Unknown
	TB	What was the patient's sex at birth?	
Sexual Orientation	Hepatitis	Patient identified sexual orientation (i.e., an individual's physical and/or emotional attraction to another individual of the same gender, opposite gender, or both genders)	1. Bisexual 2. Straight or heterosexual 3. Lesbian, gay, or homosexual 4. Something else, please describe 5. Don't know 6. Choose not to disclose

[3] Office of the Chief Statistician of the United States. (n.d.). Recommendations on the Best Practices for the Collection of Sexual Orientation and Gender Identity Data on Federal Statistical Surveys. (Washington, D.C.) https://www.whitehouse.gov/wp-content/uploads/2023/01/SOGI-Best-Practices.pdf

	STD		1. Bisexual 2. Straight (not gay or lesbian) 3. Gay or lesbian 4. Other 5. Unknown
Gender Identity	Hepatitis	Patient identified gender identity (i.e., an individual's personal sense of being a man, woman, or other gender, regardless of the sex that person was assigned at birth)	1. Female 2. Male 3. Female-to-male (FTM)/Transgender Male/Trans Man 4. Male-to-Female (MTF)/Transgender Female/Trans Woman 5. Genderqueer, neither exclusively male nor female 6. Additional gender category or other, please specify 7. Choose not to disclose
	STD		1. Cisgender/Not transgender 2. Female-to-male transsexual 3. Male-to-female transsexual 4. Transgender unspecified 5. Unknown

Historically, data element descriptions were often written as questions based on previous recommendations, as seen in the descriptions for Birth Sex. However, CDC is moving towards using non-question descriptions, recognizing that use of a question as a description may imply that was the exact language used when the data were collected. This was one of the recommendations put forth by the Measuring SOGI Research Group, one of the interest groups in the Federal Committee on Statistical Methodology (FCSM). With the understanding that public health is often not the primary collector of the data, the descriptions should better represent the information being collected since it is often unknown exactly how the question was asked.

Core Surveillance Sexual Orientation and Gender Identity Data Elements for Routine Case Notifications

The data elements below are those that the DSWG agreed to define as core surveillance data for case notifications to CDC. For each of these data elements, there is no pre-defined default value. Refer to Tables 2-5 below for definitions and additional information for each data element. Further details on eCR and USCDI as input data sources for SOGI data can be found in Appendix A.

- Sexual Orientation
- Gender Identity
- Sex
- Sex Assigned at Birth

The following data elements were also discussed, but there was no consensus to include them in core surveillance data for routine case notifications to CDC.

- Current Sex

Comparison of the data elements Current Sex and Sex at Birth may be used as a method to identify transgender individuals, however, standardized electronic health record data collection and transmission of these fields is not widely implemented. Routine data identifying an individual's sex is captured in ELR and eCR messages from fields that have no clear point of origin and may not be defined, captured, or transmitted in a consistent manner across healthcare organizations. The DSWG recommends a generic Sex data element as core so that data routinely sent via ELR and eCR can be transmitted as a part of case notifications.

Table 2. Core Surveillance Data Element for Routine CDC Case Notification – Sexual Orientation

Sexual Orientation	
Definition	A person's identification of their emotional, romantic, sexual, or affectional attraction to another person.
Standard Code for Transmission to CDC	LOINC: 76690-7 – Sexual orientation
Value Set or Coding System for Transmission to CDC	1. Asexual 2. Bisexual 3. Pansexual 4. Straight or heterosexual 5. Lesbian, gay, or same-gender loving 6. Queer 7. Not listed, please specify 8. Unsure/questioning 9. Prefer not to answer 10. Unknown
Abstracted/ Derived/ Created	Abstracted
Derived from (examples)	N/A
Recommendations for Input Data Source(s)	Case investigation eCR
Notes	This value set does not limit the collection of a broader set of terms at the jurisdictional level. "Not listed, please specify" implies that there will be a follow-up free-text field to allow for responses outside of the current dataset. The ability to capture new or less common terms is highly important in the SOGI space as it evolves to allow public health to incorporate emerging terminology.
Data Source Details*	
eCR CDA	N/A
eCR FHIR	N/A
ELR	N/A
USCDI v4	Patient Demographics/Information data class: • Sexual Orientation: A person's emotional, romantic, sexual, or affectional attraction to another person.
USCDI+ Public Health: Case Reporting Use Case	N/A

* Additional information on input data sources for SOGI data can be found in Appendix A.

Interpretation Notes

Having started with a more limited value set, DSWG discussion resulted in the elimination of outdated and potentially offensive terminology in favor of a value set with a more inclusive set of terms and a variety of answers for people unable or unwilling to answer this question, or unsure of an answer. The value set for this data element is more granular than value sets included in larger standardization efforts, and some jurisdictions may be limited to collection of a more limited value set based on local policies or regulations. These factors may affect interpretation of the data at a national level.

Table 3. Core Surveillance Data Element for Routine CDC Case Notification – Gender Identity

Gender Identity	
Definition	Person-identified gender identity (i.e., an individual's personal sense of being a man, woman, or another gender, regardless of the sex that person was assigned at birth).
Standard Code for Transmission to CDC	LOINC: 76691-5 – Gender identity
Value Set or Coding System for Transmission to CDC	1. Woman/Girl 2. Man/Boy 3. Transgender 4. Non-binary or genderqueer 5. Two Spirit 6. Another gender, please specify 7. Choose not to disclose 8. Unknown
Abstracted/ Derived/ Created	Abstracted
Derived from (examples)	N/A
Recommendations for Input Data Source(s)	Case investigation eCR
Notes	This data element is intended to allow for multi-select responses, enabling many variations which might describe a person's gender identity. Implementation guidance should encourage this strategy. An additional free text field should accompany "Another gender, please specify" to allow new or novel terms to be transmitted. These transmissions allow emerging gender identity terms to be captured and assessed for inclusion into future value set updates.
Data Source Details*	
eCR CDA	Gender Identity Observation • Observation/code = LOINC | 76691-5 | Gender identity • Observation/effectiveTime • Observation/value = Value set: Gender Identity USCDI core (https://vsac.nlm.nih.gov/valueset/2.16.840.1.113762.1.4.1021.101/expansion) AND Value set: Asked but Unknown and Other (https://vsac.nlm.nih.gov/valueset/2.16.840.1.113762.1.4.1114.17/expansion)
eCR FHIR	US Public Health Gender Identity Extension • Extension.extension:value = Value set: Gender Identity (https://vsac.nlm.nih.gov/valueset/2.16.840.1.113762.1.4.1021.32/expansion) • Extension.extension:period • Extension.extension:comment
ELR	N/A
USCDI v4	Patient Demographics/Information data class: • Gender Identity: A person's internal sense of being a man, woman, both, or neither.
USCDI+ Public Health: Case Reporting Use Case	Gender Identity: • Definition: A person's internal sense of being a man, woman, both, or neither. • Part of Patient Demographics data class

* Additional information on input data sources for SOGI data can be found in Appendix A.

Interpretation Notes

Gender Identity is a concept that has changed much in recent years, rendering relatively recent value sets incomplete and out of date. The value set decided upon here attempts to strike a balance to accommodate established concepts while also eliminating offensive language and allowing capture a person's full sense of identity. Two Spirit, a Native American umbrella term, was included as a value to ensure the value set remains inclusive of important cultural concepts; however, there is ongoing concern about cultural appropriation of this term. Ways to shield this variable from such appropriation should be incorporated into implementation guidelines.

Table 4. Core Surveillance Data Element for Routine CDC Case Notification – Sex Assigned at Birth

Sex Assigned at Birth	
Definition	Sex a person was assigned at birth, for example on their original birth certificate.
Standard Code for Transmission to CDC	LOINC: 76689-9 – Sex assigned at birth
Value Set or Coding System for Transmission to CDC	1. Female 2. Male 3. Intersex/Non-binary marker/X 4. Unknown
Abstracted/ Derived/ Created	Abstracted
Derived from (examples)	N/A
Recommendations for Input Data Source(s)	Vital records eCR
Notes	This data element is distinct from a separate "Sex" data element, and should only be abstracted from official documents, such as a birth certificate, or from a record field which indicates the value is specific to the person's sex assigned at birth. Values from a more generic "Sex" field within an electronic message, such as PID-8 in HL7 v2, should be captured within the "Sex" data element in Table 5.
Data Source Details*	
eCR CDA	Birth Sex Observation • Observation/code = LOINC | 76689-9 | Sex assigned at birth • Observation/value = Value set: ONC Administrative Sex (https://vsac.nlm.nih.gov/valueset/2.16.840.1.113762.1.4.1/expansion) AND nullFlavor | UNK
eCR FHIR	US Core Birth Sex Extension • http://hl7.org/fhir/us/core/StructureDefinition/us-core-birthsex • Extension.valueCode = Value set: Birth Sex (http://hl7.org/fhir/us/core/STU6.1/ValueSet-birthsex.html)
ELR	N/A
USCDI v4	N/A
USCDI+ Public Health: Case Reporting Use Case	N/A

* Additional information on input data sources for SOGI data can be found in Appendix A.

Table 5. Core Surveillance Data Element for Routine CDC Case Notification – Sex

Sex	
Definition	Sex, from any data source.
Standard Code for Transmission to CDC	LOINC: 46098-0 – Sex
Value Set or Coding System for Transmission to CDC	1. Female 2. Male 3. Intersex/Non-binary marker/X 4. Unknown
Abstracted/ Derived/ Created	Abstracted
Derived from (examples)	N/A
Recommendations for Input Data Source(s)	Case investigation eCR ELR
Notes	The definition of the Sex data element is intentionally ambiguous to allow for a data element that captures patient sex identifiers in electronic health records, including eCR and ELR data. For example, the data transmitted via PID-8 in ELR HL7 v2 messages or the HL7 Administrative Gender value in the TargetRecord section in eICRs often have no clear point of origin and may not be defined, captured, or transmitted in a consistent manner across healthcare organizations. In many instances, this may be the only information available to represent the sex of a person.
Data Source Details*	
eCR CDA	ClinicalDocument/recordTarget/patientRole/patient/administrativeGenderCode • Value Set: Administrative Gender (HL7 V3)
eCR FHIR	US Public Health Patient profile • AdministrativeGender: Patient.contact.gender
ELR	HL7 v2: Administrative Sex (PID-8)
USCDI v4	Patient Demographics/Information data class: • Sex: Documentation of a specific instance of sex and/or gender information.
USCDI+ Public Health: Case Reporting Use Case	Sex: • Definition: Documentation of a specific instance of sex and/or gender information. • Part of Patient Demographics data class

* Additional information on input data sources for SOGI data can be found in Appendix A.

Interpretation Notes

The definition of the Sex data element is intentionally ambiguous in order to allow for a data element that captures patient sex identifiers in electronic health records, including eCR and ELR. For example, the data transmitted via PID-8 in ELR HL7 v2 messages or the HL7 Administrative Gender value in the TargetRecord section in eCRs often have no clear point of origin and may not be defined, captured, or transmitted in a consistent manner across healthcare organizations. In many instances, this may be the only information available to represent the Sex of a person.

Appendix A. Input Data Sources for Sexual Orientation and Gender Identity Data

Electronic Case Reporting (eCR)

In the eICR document, the recordTarget section includes the administrative and demographic data of the patient and must contain one code from the HL7 Administrative Gender value set. This value set includes the following codes: F=Female, M=Male, and U=Undifferentiated.

The Birth Sex Observation represents the sex of the patient at birth and is not appropriate for recording patient gender. The allowable codes would be either "F" or "M" (from the ONC Administrative Sex value set) or "UNK" as a nullFlavor. The US Core Birth Extension in FHIR aligns with the CDA Birth Sex Observation both in the definition of birth sex and in the associated value set.

The Gender Identity Observation has the following definition in the CDA format: "One's basic sense of being male, female, or other gender (for example, transgender or gender queer). Gender identity can be congruent or incongruent with one's sex assigned at birth based on the appearance of external genitalia." On the other hand, the US Public Health Gender Identity Extension in FHIR "represents an individual's sense of being a man, woman, boy, girl, nonbinary, or something else, ascertained by asking them what that identity is." Although the specific wording differs, the idea that gender identity is a self-identified concept remains consistent. Regarding the associated value set, the FHIR extension aligns with the CDA Gender Identity Observation.

Sexual orientation is not included in the eCR implementation guide.

Electronic Laboratory Reporting (ELR)

In HL7 v2, the Administrative Sex (PID-8) specifies a patient's sex for administrative purposes but does not necessarily represent any physiology, biology, or identity. The interpretation of this field may vary depending on the specific use case and the practices of the healthcare system sending the data. The associated HL7 value set for this field includes the following codes: A=Ambiguous, F=Female, M=Male, N=Not applicable, O=Other, and U=Unknown. HL7 v2.9 introduced an additional concept, X=Non-Binary, which allows for "for situations where the gender or sex representation of the individual is not strictly male, or strictly female, and is driven by jurisdictional requirements, personal needs, or legal boundaries".

With the exception of the patient's sex at the time of report as transmitted in PID-8, ELR does not typically include SOGI data. More recently, HL7 released the Sexual Orientation and Gender Identity Data Exchange Profile in September 2023 that includes details on how to transmit sexual orientation, gender identity, and pronouns as OBX statements.

USCDI and USCDI+

A Birth Sex data element has been part of USCDI since version 1, but it has evolved over time to Sex (Assigned at Birth) in version 2 and finally to Sex in versions 3 and 4. Only the current iteration of the Sex variable has a stated definition (see Table 6), but there is no defined value set beyond a note that the applicable vocabulary standard is SNOMED CT.

Gender Identity and Sexual Orientation have been included in USCDI since version 2, each with relatively consistent definitions and value sets. Two data elements not yet in USCDI but being considered for a

future version are Pronouns, currently in Level 2, and Variation in Sex Characteristics, currently in Level 1.

Table 6. SOGI Data Elements Included in or Being Considered for USCDI

Data Element	Definition and Value Set	V1	V2	V3	V4
Birth Sex	No definition Codes from HL7 V3 value set for AdministrativeGender and NullFlavor: F, M, UNK	✓			
Sex (Assigned at Birth)	No definition Codes from HL7 V3 value set for AdministrativeGender and NullFlavor: F, M, UNK		✓		
Sex	Documentation of a specific instance of sex and/or gender information. Includes codes from SNOMED CT			✓	✓
Gender Identity	A person's internal sense of being a man, woman, both, or neither. Includes codes from SNOMED CT and HL7 V3 value set for AdministrativeGender and NullFlavor: • Male. 446151000124109 • Female. 446141000124107 • Female-to-Male (FTM)/Transgender Male/Trans Man. 407377005 • Male-to-Female (MTF)/Transgender Female/Trans Woman. 407376001 • Genderqueer, neither exclusively male nor female. 446131000124102 • Additional gender category or other, please specify. nullFlavor OTH • Choose not to disclose. nullFlavor ASKU		✓	✓	✓
Sexual Orientation	A person's emotional, romantic, sexual, or affectional attraction to another person. Includes codes from SNOMED CT and HL7 V3 value set for AdministrativeGender and NullFlavor: • Lesbian, gay or homosexual. 38628009 • Straight or heterosexual. 20430005 • Bisexual. 42035005 • Something else, please describe. nullFlavor OTH • Don't know. nullFlavor UNK • Choose not to disclose. nullFlavor ASKU		✓	✓	✓
Pronouns (Level 2)	The pronoun(s) specified by the patient to use when referring to the patient in speech, in clinical notes, and in written instructions to caregivers.	-	-	-	-
Variation in Sex Characteristics (Level 1)	Were you born with a variation in your physical sex characteristics? (This is sometimes called being intersex or having a Difference in Sex Development, or DSD.) Proposed value set includes Y, N, I don't know	-	-	-	-

Data elements for Sex and Gender Identity are part of the Patient Demographics data class in USCDI+. In addition, there is also a Sexual Orientation data element, but it is present only in the Cancer and Maternal Health domains and not in the Public Health domain. Each of these data elements maintains the same definition as is found in USCDI v4.