Long-Term Care Data Cooperative

Technical User Guide

Version 1.5

As of May 15, 2024

For more information, visit ltcdatacooperative.org or email ltcdc@exponent.com.

2009811.000 – 5927

Disclaimer: No guarantees are made regarding the Long-Term Care Data Cooperative data model, which may vary in correctness and completeness by electronic medical record (EMR) vendor and long-term care organization and facility. These data do not constitute a legal medical record and no medical or individual care decisions should be made based on Long-Term Care Data Cooperative data without consultation with source EMRs by an appropriate clinician. No guarantee of availability or quality of data elements described here is intended or implied.

The Long-Term Care Data Cooperative is supported by the National Institute on Aging of the National Institutes of Health under IMPACT Collaboratory contract #3U54AG063546-02S6. The content is solely the responsibility of the Long-Term Care Data Cooperative and does not necessarily represent the official views of the National Institutes of Health.

Table of Contents

History of Modifications

Document Version	Date	Data Model Version	Changes
1.0	2023-06-12	1.4	Initial Release
1.1	2023-09-06	1.5	All primary keys switched from hashed strings to integers. (ref) Assessment concept split into header and result information (Assessment) and question-response data (Assessment Element). (ref) Primary key for the Provider concept renamed. (ref)
1.2	2023-10-20	1.5.1	Organization identifier added to Facility concept. (ref)
1.3	2024-02-09	1.5.2	Data elements with restricted distribution are now shaded in data element lists. (ref) Inactive fields removed: *Episode.readmission_ind, Episode.episode_sequence, Stay.stay_sequence, Medication.medication_reason_code, Medication.medication_reason_code_type, Medication.medication_reason_code_desc*
1.4	2024-03-15	1.6	Allergy concept added. (ref) Removed "Exchange" option from *Stay.payer_type*. (ref)
1.4.1	2024-04-15	1.6	Updated language
1.5	2024-05-15	1.7	Advance Directive concept added. (ref) Hospice payer/plan indicator added to Stay concept. (ref) Added "Private" option to *Stay.payer_type*. (ref)

Purpose

Long-Term Care Data Cooperative

The Long-Term Care (LTC) Data Cooperative, a provider-led coalition of long-term care facilities across the United States, has committed to the following aims: assembling an electronic medical record (EMR) based dataset for the purposes of improving resident care and outcomes; extending the capabilities and potential of real-world data-driven research; enabling timely public health surveillance; and offering a launchpad for effective randomized clinical trials.

The LTC Data Cooperative Dataset

Central to serving these aims is the formation of the LTC Data Cooperative dataset, which normalizes the various data sources being provided to the LTC Data Cooperative and harmonizes the elements to give users a consistent experience when working with the data. These data are expected to provide a foundation of capabilities that offer practical insights for providers and a core set of concepts for researchers. Through feedback from users, data enhancements from sources, and enrichment through targeted exploration, the LTC Data Cooperative dataset will continue to grow to meet the potential and promise of the aims of the LTC Data Cooperative.

This Document

The purpose of this document is to provide users with an understanding of what data comprise the LTC Data Cooperative dataset, how those data are captured and integrated, and how decisions are made on what data to include, harmonize, and present. Information about the data elements and structures are mixed with editorial content describing how these structures came to be so that users are informed on the provenance and limitations of the data when making decisions about what data to use. As the value, content, and userbase of the LTC Data Cooperative dataset grows, the LTC Data Cooperative will continue to work toward creating documentation that tracks the needs of the data and its users.

Data Sources

The EMR data comprising the LTC Data Cooperative dataset are sourced from long-term care facilities with differing EMR systems, organizational policies, and individual workflows. As a result, there are a number of ways in which the data can vary between systems. In this section, we describe some of the differences between data sources; in *Data Integration*, we explain how these differences appear in the dataset itself to allow users to determine fitness for use of data from individual sources.

EMR Vendors / Products

All LTC Data Cooperative dataset EMR data are sourced from EMR vendors, each of which shares data in a distinct format that may differ substantially from that of other vendors.

For example, one EMR vendor provides complete responses to all cognitive tests regardless of the circumstances under which that test was sourced, whereas another provides responses to cognitive tests sourced from MDS-3 data but only provides scores for cognitive tests assessed under other circumstances. As another example, one EMR vendor provides consistent CVX codes for all immunizations but not vaccine lot numbers and expiration dates, while another EMR vendor provides lot numbers and expiration dates when available but cannot guarantee detail to the level of CVX codes in other cases.

Differences also include the amount of historical data provided by the EMR vendor. In some cases, a resident's history of admissions goes back decades, but detailed information about stays is limited to 3-5 years. In other cases, the amount of historical data provided is uniform for a given organization, introducing an artificial "epoch" that must be accounted for. The recency of the data – that is, what is the most recent information available – also varies by EMR vendor based on data refresh schedules.

High-level descriptions of these differences for each data type are included in *Data Domains*.

Long-Term Care Companies

LTC companies are the highest-level organizing structure for health care providers in the LTC space. In most cases, it is the LTC companies that join the LTC Data Cooperative, bringing their facilities into the project as participants.

Differences in organizational policies introduce variation into the data provided to the LTC Data Cooperative. For example, even though they use the same EMR, different companies may use

differently structured assessments, resulting in slightly different wording or numbering of questions, even though the responses and scoring are conceptually the same.

Different companies may also choose to make use of different modules within the same EMR. For example, data from a third-party lab module may be available for some companies, while others may not have that module available. Similarly, some companies may participate in licensing that allows them to present data in specific formats, such as making use of in-EMR pharmacy databases. These differences may be reflected in the presence of such data elements for some residents and not others.

Long-Term Care Facilities

Given the complexity of capturing and sharing data among facilities, companies, and EMRs, some between-facility variation is to be expected.

For example, within a given organization, all facilities may record information about resident falls, including when, where, and how staff responded to the incident. However, facilities may document this information with differing detail and regularity. In some cases, these differences may be due to decisions by departments or preferences of individual care partners. In other cases, facilities may use completely different workflows or data entry forms to capture this information.

These differences are addressed through the data harmonization process – see *Data Integration*.

Patient Privacy and Responsible Use of Data

Information security and data protection are cornerstones of the LTC Data Cooperative. All data are transmitted, manipulated, and stored in secure environments that meet the requirements of multiple industry-standard frameworks. For details, please review the Information Security & Data Privacy Guide on ltcdatacooperative.org.

In addition to employing best-in-class standards for data handling and storage, the LTC Data Cooperative is also committed to the principles of ethical and responsible use of data established in the research community. Approval of investigator research plans and data handling protocols, as well as application of the Minimum Necessary Requirement (ref 45 CFR 164.502(b), 164.514(d)), create a culture of conscientious and respectful use of data for the entire user community.

Data Integration

If the data comprising the LTC Data Cooperative dataset are a mix of structures, operations, decisions, and preferences defined by the parties described in *Data Sources*, the Data Integration process is the mechanism by which those data sources are transformed into a uniform and consistent data model for users. The Data Integration process consists of the following steps:

1. **Normalization** – Once raw EMR data have been retrieved from their original sources, they are transformed into structures that are clinically or operationally meaningful and useful. These structures align with concepts that have been identified as relevant to the LTC Data Cooperative mission – see *Concept Identification* – and allow varying source data to be tied to more standardized or well-understood concepts.
2. **Harmonization** – Normalized data that have been structured along commonly understood concepts may still vary considerably in how those data were originally captured by their source systems. The harmonization process takes into account that variability and applies a range of transformation processes to the data to coerce them into alignment.
3. **Presentation** – The presentation process transforms the concepts and data elements in the LTC Data Cooperative dataset into structures that provide more value and flexibility for users. Examples include: pivoting tables to place variables of interest in individual columns rather than in different rows; creating de-identified versions of tables that can be used in scenarios where protected health information (PHI) is not needed; and presenting data in formats that support interoperability and regulatory requirements.

Although the details behind these stages are beyond the scope of this document, information about key components in this process is included here.

Concept Identification

Concept Identification is the process by which concepts are chosen for representation in the LTC Data Cooperative dataset.

The initial construction of the LTC Data Cooperative dataset was based on input from researcher, provider, and user communities, with a focus on what concepts would create a foundation for further research. With each revision of the LTC Data Cooperative dataset, existing concepts are refined and new concepts are introduced.

The LTC Data Cooperative dataset is now overseen by an Advisory Committee, which reviews the state of the dataset and the foundations of its design and makes recommendations on further development and effort.

Integration Policies

Data Types

Data elements have been assigned data types based on their most common representation in source systems and, occasionally, by common or standard representation (such as encodings used on claims or in standard code sets).

Rules for integration of data types are as follows:

Primary Keys

- Primary keys are generated for each concept, and uniquely identify any given record in the dataset.
- Primary keys and foreign key references are all represented as integers. They are internally consistent within each release but can change between data cycles.
- An additional copy of the primary key in a hash string form (40 characters long) can be found in each concept table. This hash string persists between data cycles as long as the underlying keys are not changed by the EMR vendor or the key generation process for this dataset has not changed. As a result, these keys are likely to persist but cannot be guaranteed.

Dates and Timestamps

- All dates and timestamps are extracted as-is from the original source system.
- In most cases, both dates and timestamps are represented as a timestamp data type.
- Dates represented as timestamps will present a time of “00:00:00.000”. The data do not indicate whether an event timestamp is exactly midnight.
- NULL values are persisted in date and timestamp type fields. Unless explicitly specified otherwise, dates are not imputed or filled in for NULL values. Furthermore, in cases where imputed values were substituted for empty or unknown values by data sources, those values have been reverted to NULL. However, this cannot be guaranteed for all cases and users should use their best judgment with nonsense or extreme dates.

Quantitative Fields

- All fields representing quantitative values, such as results of an observation, are stored in floating-point precision.
- Because some quantitative data elements were sourced from fixed-decimal precision data fields, drift may be evident in certain cases (e.g. “2.1” becomes “2.0000009”).
- Numeric values captured as character strings are converted to numeric fields based on simple pattern matching; at this time, exponentiation is not supported (e.g. “211e-2”).
- NULL values are persisted in quantitative values. Any character string transformed to quantitative values is converted to NULL if it cannot be represented as a number (e.g. “”, “~”, “-”). Unless noted, no attempt is made to identify imputed values in quantitative fields.

Categorical, Nominal, and Coded Fields

- Fields representing categorized concepts, named levels, and standard codes are stored as strings.
- Fields representing ordinal concepts, counts, and internal identifiers are stored as integers.
- Depending on context, NULL values may be represented as NULL or as an appropriate category or level, such as “Unknown”.
- External identifiers, facility CMS Certification Numbers (CCNs) and provider National Provider Identifiers (NPIs), are considered text fields and are not included in this group.
- Native data storage is case-insensitive; users employing case-sensitive analytic tools should be aware of potential differences.

Text Fields

- Fields representing unstructured text are stored as strings or database-appropriate data types.
- This group includes external identifiers, such as facility CCNs and provider NPIs, which should always be stored as strings, as well.
- In cases where de-identification has been applied, redacted values will be replaced with a NULL value.
- Although empty or nonsense fields may be removed under some circumstances, typically text fields will undergo only minimal scrubbing.
- Native data storage is case-insensitive; users employing case-sensitive analytic tools should be aware of potential differences.

Data Element Harmonization

Whenever possible, data elements are harmonized to represent a commonly understood or standardized concept.

Harmonization choices were generally made based on the following series of questions:

1. **Is it a commonly understood concept?** For example, date of birth can be commonly understood as a date representation based on best-available information for that individual.
2. **Is a relevant and approachable standard available?** For example, race and ethnicity are based on a standard representation of race/ethnicity designations in practice in the United States. All race and ethnicity data provided by the EMRs are coerced to one of the appropriate representations.
3. **Can a common vocabulary be formed?** For example, the status of an individual in a skilled nursing facility during a span of time may be represented differently in different sources. A vocabulary of descriptive names, including "In House" and "Hospital Leave" provide sufficient detail to represent the original data from most sources while providing a useful number of categories for the end user.

Data concepts and elements to which the preceding questions all answer negative may still be included in the data model but require more scrutiny to define and integrate.

Master Patient Identifiers

As individuals can be represented repeatedly within a given data source and across multiple sources, Master Patient Identifiers are included with all patient-related core tables (see the Patient concept for more detail). These identifiers are internally consistent for a given dataset and can be used to join together most kinds of patient-centric records; they allow the user to denote which events occurred under which circumstances. Master Patient Identifiers are recalculated with each data refresh and should never be used to store patient identities.

Patient identifiers tied to a specific patient record from the EMR are also made available in all cases. Both identifiers can be used to join together data elements; only the Master Patient Identifier is more likely to track patients over time.

Master Patient Identifiers are generated by a Master Patient Index, a software tool that identifies similar patient profiles based on a combination of demographic and operational features. The operation of the process is out of the scope of this document, but briefly:

1. The Master Patient Index collects data on all patients from all data sources. Patient attributes are normalized and cleaned. Names are normalized to phonic-based fragments, structured values (such as Social Security Numbers (SSNs) and phone numbers) are formatted, and invalid values are removed. Records that are likely test records – either as explicitly marked or based on naming – are excluded.
2. The Master Patient Index attempts to make "primary matches" between records by identifying simple likely-but-not-sufficient clusters, such as name/sex/date of birth (DOB), SSN/DOB, and partial-name/DOB/geographic groupings.
3. The Master Patient Index attempts to substantiate "primary matches" with additional elements. In most cases, one additional matching element is needed to confirm the match. Any matches made in this stage are marked as confirmed matches and assigned a single Master Patient Identifier to represent the patient and their cluster of records.

Data Domains

The following Data Domains are currently stored in the LTC Data Cooperative dataset:

Domain	Description
Patient	Basic demographic information for individuals residing in or rendered care by LTC facilities, including the individual's location, sex, age, race, and ethnicity.
Facility	Facility information, including name, location, identifiers, key characteristics, and rankings/metrics.
Provider	Basic demographic and professional information regarding service providers. This includes physicians, nurses, and clinical and non-clinical staff who render services to patients in a care setting.
Episode	Information about admissions and discharges to the facility, along with admission and discharge timing and circumstances.
Stay	Information about the presence of the patient within and outside the facility, including dates of arrival and departure and status changes.
Advance Directive	Documentation of orders related to advance directives, along with the type of advance directive, timing of the order, and the ordering provider.
Allergy	Information about a report or assessment of patient allergy or intolerance to a substance, including medications, foods, and environmental allergens.
Assessment / Assessment Element	Standard assessment forms as described in the EMR, such as instruments and surveys. Assessments include the MDS-3 (reported by EMRs), Activities of Daily Living (ADL), PHQ-9 and PHQ-9-OV, BIMS, and other physical, behavioral, and cognitive assessments. Assessment Elements are the question-response contents of Assessments, where available.
Condition	Information about the medical state of the patient. Includes diagnoses, as assessed by a provider or clinician, the date/time of the diagnosis, diagnosis codes, and circumstances of assessment.
Observation	Observations are measurements or assertions made about the state of the patient, including vital signs, laboratory tests, and clinical findings.

Domain	Description
Medication	Orders for and administration of medications by facility staff or providers, including prescription and over-the-counter medications and treatments.
Vaccination	Information about vaccines administered to patients, as well as records of historical vaccinations reported by patients, providers, and others. Also includes records of vaccination declinations as reported by patients or their representatives.

Each domain description is accompanied by the following information:

- **Definition** – The definition applied to the data, and the basis for inclusion of data from the EMR into the data domain.
- **Integration** – Specific integration policies applied and case-specific harmonization challenges and data variability encountered during the integration process.
- **Timing** – The unit of time embodied by the record, typically either as a snapshot in time, a specific event in time, or a span of time.
- **Granularity** – The logic by which new records are generated; for example, whether a record is generated per-person, per-event, or some other logic.
- **Data Elements** – A list of the data elements stored in the LTC Data Cooperative dataset for a given record. Elements that are greyed out are restricted and may not be available in all distributed datasets. Descriptions include the following fields:
    - Field – The name of the data element in the LTC Data Cooperative dataset.
    - Type – The data type of the data element. See *Data Types* for descriptions. Internal keys, including Primary Keys (PK) and Foreign Keys (FK) are flagged here.
    - Description – Brief description of the data element.
    - References – Indicator of whether the data element references another concept or an entry in its Terminology list.
- **Terminology** – A list of terminology sources or references, as applicable.
- **Additional Notes** – Additional information specific to integration, representation, or interpretation of the data domain or its attributes.

> ***Note:*** The *Data Elements* subsections are not a data dictionary for any dataset. Data made available to each user will be dictated by user requirements and supporting approvals (e.g., Institutional Review Board/Privacy Board). Such data may not include some of the fields described under a domain or may include alternate representations.

Patient

Definition

This table contains information related to persons who interact with the long-term care facilities in the dataset. The attributes are intended to provide demographic information necessary to support patient characterization. Only persons to whom services are rendered are considered patients; other persons, such as providers, staff, family, care partners, and other associated individuals are not considered patients.

> *Note:* The term "Patient" refers to the residents whose data are the subject of this Dataset. Although "Resident" is the preferred term, "Patient" is used in this data model, and throughout the *Technical User Guide*, to align with other healthcare data ontologies.

Integration

Patient data are typically represented in the EMR as distinct snapshots of a patient's demographics. In some cases, a patient record is only associated with a single facility, whereas in other cases, a patient record may follow an individual across multiple facilities and episodes. The integration process does not treat these cases differently; users who want to track patients by facility should consider using a combination of the Patient and Facility or the Episode, depending on their requirements.

Categorical inputs are represented in different ways depending on the source EMR. In all cases, values were coerced to standard code sets; see *Terminology* for more information.

Although the data model contains personally identifiable information (PII) in the form of patient names, date of birth, and identifiers, de-identification compliant versions of some of these elements have been added to the data object for easy removal of PII from output datasets. For example, a patient's ZIP5 is present in the Patient record, but a ZIP3 field is also present, with '000' ZIP3 logic already built in, in cases where the ZIP5 needs to be redacted.

To allow users to link together records related to the same patient that use different Patient records, a *master_patient_id* is included in the Patient data object. This identifier is described in greater detail in *Master Patient Identifiers.*

Timing

Patient records represent a snapshot in time as recorded in the source system by the provider.

Although the Patient record is a snapshot, some elements such as patient-supplied sex, race, and ethnicity might change over time. While it is possible that such changes would be updated in the Patient record, it is not guaranteed. For that reason, the patient record may include information known to be a true at the most recent update of the record and may not be correct for all times to which the record applies.

Granularity

A single Patient record is generated for each representation of a Patient record in the source EMR. However, different EMRs and even facilities make use of this concept in different ways. Therefore, multiple patient records can represent a single individual.

To get around this issue, a *master_patient_id* field has been added to the Patient table to allow users to track individual patients across multiple facilities and data sources. This identifier is described in greater detail in *Master Patient Identifiers.*

Data Elements

The following data elements are included in the Patient data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	
Patient_id	PK	Patient identifier derived from EMR.	
Sex	Categorical	Patient sex as reported by the patient or representative.	See Terminology
date_of_birth	Date	Patient date of birth.	
Year_of_birth	Integer	Patient year of birth, redacted for persons 90 years of age or older.	
Race	Categorical	Patient race, as reported by the patient or representative. If the race and ethnicity were reported together, all persons listed as a race/ethnicity of Hispanic are assigned a race of Unknown.	See Terminology
ethnicity	Categorical	Ethnicity, as reported by the patient or representative. If the race and ethnicity were reported together, all persons listed as a race/ethnicity other than Hispanic are assigned an ethnicity of Not Hispanic or Latino.	See Terminology
language	Text	Language as described by ISO 639-2. All persons with a reported language not included in this standard are assigned a language of Unknown.	See Terminology
zip5	Text	Patient address* ZIP5 code (first five digits of ZIP postal code).	
Zip3	Text	Patient address* ZIP3 code (first five digits of ZIP postal code), redacted for small population ZIP codes**.	
State	Text	Patient address* state abbreviation.	See Terminology
created_date	Date	Record creation date per source system.	
Updated_date	Date	Record update date per source system.	

** Most recent patient address on file.*

*** As described in HHS documentation on de-identification. For more information, see:*
https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#zip

Terminology

The following terminology is used in the Patient data object:

Field	Type	Definition	Values
sex	Categorical	Sex of the patient.	M, F, U
race	Categorical	Patient race. Defined by OMB (link).	White, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Multiple Races Reported, Unknown
ethnicity	Categorical	Patient ethnicity. Defined by OMB (link).	Hispanic or Latino, Not Hispanic or Latino, Unknown
language	Categorical	Patient language. Language value set is defined in ISO 639.2 (link).	Values enumerated in *English Name of Language* column described in ISO 639.2 (link).
State	Categorical	Patient state based on most recent address on file.	AK, AL, AZ, …

Provider

Definition

This table describes the basic demographic and professional information regarding health care service providers associated with participating long-term care facilities. This includes physicians, nurses, and clinical and non-clinical staff who render services to patients in a care setting.

Integration

Information about providers and staff are presented in a similar fashion for all EMR sources in which this information is available. However, the integration process currently generates provider records differently depending on whether a National Provider Identifier (NPI) is associated with the record.

For EMR-based provider records with no associated NPI, a Provider record is generated for each EMR-sourced record. The EMR-sourced primary ID – typically the primary key for the source table – is carried through. No effort is made to match these providers across multiple facilities or companies, nor is any effort made to link external data elements with these provider records.

For EMR-based provider records with an associated NPI, a single Provider record is generated per NPI. That means that all references to that provider, regardless of facility or organization, will refer to a single record. Most data elements for that record are sourced directly from the National Plan and Provider Enumeration System (NPPES) database.

Timing

Provider records represent a snapshot in time as recorded in the source system by the provider.

It is possible for data elements in the Provider record to change over time; changes may include the taxonomy code and profession type described in the record.

For Provider records with an NPI, changes made to the NPPES database will be carried over into the Provider record. (Records without an NPI will not change unless updated in the EMR.) Therefore, it is possible for an event referencing a Provider record to predate changes to that record.

Granularity

As discussed in Integration, for EMR-based records with an NPI, all provider records are merged into a single record with the associated NPI. In all other cases, a provider record is generated for each distinct reference in the dataset. Since the same provider could be listed in two different systems, one with an NPI and one without, it cannot be guaranteed that a record with an NPI is inclusive of all references to that provider.

Data Elements

The following data elements are included in the Provider data object:

Field	Type	Description	References
provider_id	PK	Unique provider identifier.	
Provider_npi	Text	National Provider Identification (NPI) number for provider. This field is null if no NPI was indicated by the EMR.	
Taxonomy_code	Categorical	Taxonomy code associated with the provider. If NPI was provided by the EMR, the primary taxonomy code from NPPES is used.*	See Terminology
taxonomy_classification	Categorical	Taxonomy classification associated with above taxonomy code, if available.**	See Terminology
taxonomy_specialization	Categorical	Taxonomy specialization associated with above taxonomy code, if available.**	See Terminology
last_name	Text	Provider last name.*	
first_name	Text	Provider first name.*	
middle_initial	Text	Provider middle initial.*	
credential	Text	Provider credential, as provided by the EMR or NPPES (e.g MD, OD, APRN).*	
sex	Categorical	Provider sex.*	See Terminology

** Variables pulled preferentially from the National Plan & Provider Enumeration System (NPPES). (link)*

*** Variables pulled from the National Uniform Claim Committee (NUCC). (link)*

Terminology

The following terminology are used in the Provider data object:

Field	Type	Definition	Values
taxonomy_code	Categorical	Taxonomy codes per the National Uniform Claim Committee (NUCC). (link)	
taxonomy_classification	Categorical	Taxonomy classifications per the National Uniform Claim Committee (NUCC). (link)	
taxonomy_specialization	Categorical	Taxonomy specialization categories per the National Uniform Claim Committee (NUCC). (link)	
sex	Categorical	Sex of the provider.	M, F, U

Facility

Definition

A Facility is defined as a space in which health care services are rendered. Providers and staff associated with the facility render health care services to patients in that space and outside it on behalf of the facility's parent organization. Patients are admitted to facilities as reflected in Episodes; outpatient services may also be rendered at facilities to patients who are not admitted at the time.

> *Note:* The term "Facility" is used in this data model, and throughout the *Technical User Guide*, to refer to any long-term care facility or community whose data are included in this Dataset.

In the LTC Data Cooperative dataset, most facilities are skilled nursing facilities (SNFs) with an assigned CMS Certification Number (CCN). It is also possible for non-LTC hospitals with swing beds and other long-term care facilities with a CCN to be included in the dataset. Due to the variability of this classification between companies and localities, as well as over time, facility types are not currently presented in the Facility concept.

Integration

All EMR sources support specification of facilities at a CCN-level; for facilities without CCNs, such as ALFs, facilities may be based on named groups or geographic positioning of buildings. For the purposes of the LTC Data Cooperative dataset, no additional imputation of divisions or organization of facilities is applied.

In some EMRs, a very broad set of facilities are included in the source dataset, ranging from SNFs to ALFs to support entities such as churches and funeral homes. Therefore, the LTC Data Cooperative integration process applies a validation process for facilities confirming that patients have been admitted to a facility during the integration time frame. (The integration time frame varies between sources and facilities but typically defaults to 3-5 years.) Facilities with no history of patient admissions in the dataset are excluded from the final data payload.

Timing

Facility records represent a snapshot in time as recorded in the source system by the provider.

For Facility records with a CCN, changes made to the CMS Nursing Home Provider Information database will carry over into the Facility record. (Records without a CCN will not change unless updated in the EMR.) The CCN also allows the user to link a facility in the LTC Data Cooperative dataset to information about that facility in other public and third-party data.

The information presented in the LTC Data Cooperative dataset represent the most recent version of facility data available from the EMR, along with the most recently available data from CMS. Users interested in pursuing operational or structural changes over time may need to consider changes in facility management (and CCN) beyond what is reflected in the LTC Data Cooperative dataset.

Granularity

A single record is generated for each Facility as listed in the source EMR. Typically, this means that a record is created for each SNF CCN, with additional records representing associated ALFs and other care facilities. However, assignment of facility records is determined by how the information is captured in the source EMR and so some differences in granularity of source record may be expected.

Data Elements

The following data elements are included in the Facility data object:

Field	Type	Description	References
facility_id	PK	Unique facility identifier.	
facility_ccn	Text	Facility CMS Certification Number (CCN).	See Terminology
facility_npi	Text	Facility National Provider Identifier (NPI)*.	See Terminology
facility_name	Text	Facility name*.	
facility_address	Text	Facility street address*.	
facility_city	Text	Facility city*.	
facility_state	Text	Facility state*.	
facility_zip	Text	Facility ZIP Code*.	
facility_county	Text	Facility county name*.	
facility_fips_code	Text	Facility FIPS Code**.	See Terminology
facility_phone	Text	Facility phone number*.	
facility_owner_type	Categorical	Type of the entity that operates the facility, as described by CMS provider information**.	See Terminology
facility_owner_subtype	Categorical	Geographic or legal classification of the entity that operates the facility, as described by CMS provider information**.	
organization_id	Integer	Organization identifier derived from EMR***.	
reported_certified_beds	Integer	Number of Federally certified beds according to CMS records**.	
emr_certified_beds	Integer	Number of certified beds reported by the EMR.	
provider_type	Text	Facility provider type, as described by CMS provider information**.	See Terminology
in_hospital_ind	Text	Indicator of whether the facility is in a hospital, as reported by CMS provider information**.	Y, N
created_date	Date	Record creation date per source system.	
updated_date	Date	Record update date per source system.	

** Variables pulled preferentially from CMS Nursing Home (including rehab services) Provider Information dataset. (link)*

*** Variables pulled exclusively from CMS Nursing Home (including rehab services) Provider Information dataset. (link)*
**** Organization information is subject to change over time and may not correspond with ownership/management at the time of a particular event, or with ownership information available from CMS.*

Terminology

The following terminology are used in the Facility data object:

Field	Type	Definition	Values
facility_ccn	Nominal	Facility CMS Certification Number (CCN).	
facility_npi	Nominal	Facility National Provider Identifier (NPI). Defined and maintained by HHS (link).	
fips_code	Nominal	Facility FIPS code. Defined by U.S. Census. Defined by NIST (link).	
owner_type	Categorical	Facility owner type, as categorized by CMS Nursing Home Provider Information.	Non profit, Government, For profit
owner_subtype	Categorical	Geographic or legal classification of the entity that operates the facility, as described by CMS Nursing Home Provider Information.	Corporation, Limited liability company, County, State
provider_type	Categorical	Facility provider type, as described by CMS Nursing Home Provider Information.	Medicare, Medicaid, Medicare and Medicaid

Episode

Definition

An Episode represents the time between when a patient is first admitted to a facility for a particular reason, until that patient is either discharged from the facility and does not return to the facility as part of that admission, or the patient is reported as deceased.

The episode definition is based on logic published by CMS related to defining the period for which a patient admission to a facility should be evaluated[1]. The period of the episode is defined by one or more stays (see *Stay*), beginning with a starting date (admission) and ending with an ending date (discharge).

For the purposes of the Episode logic in the LTC Data Cooperative dataset, the following definitions apply:

- An episode admission occurs when the patient enters the facility and (a) has never been admitted to the facility before, or (b) has been in the facility previously and was discharged "return not anticipated", or (c) has been in the facility previously, was discharged "return anticipated", and did not return within 30 days of discharge.
- An episode discharge occurs when (a) the patient is discharged from the facility "return not anticipated", or (b) the patient is discharged from the facility "return anticipated" and does not return within 30 days, or (c) the patient is listed as deceased in the EMR.

Based on the above definitions, an Episode is generated for each admission, with a matching discharge identified or generated depending on available data.

Integration

Episodes are extracted from the EMR as a combination of spans labeled as patient visits and patient census data. Given a list of admissions for a given patient to a given facility, an episode is generated for each admission. Those admission dates are compared against that patient's discharges from the same facility, and discharges then assigned to each episode based on the timing of the respective events.

In some cases, no suitable discharge exists to match with an admission. For example, the patient may still be admitted to the facility as of data publication, or at least there is no evidence for their

[1] "MDS 3.0 Episode and Stay Determination Logic" from *MDS 3.0 Quality Measures User's Manual*, RTI International.

discharge at that time, in which case the discharge date is left blank. It is also possible that the patient was discharged but that information was not provided to the EMR. The Episode records do not attempt to infer or impute a discharge in the intervening time, but such logic can be applied in the presented data (see *Additional Notes*).

Once the admission and discharge dates of an episode have been identified, additional metadata regarding the episode from EMR-based census data are attached to the episode. This can include details of the admission (such as admission reason and source facility) and discharge (such as reason and target for discharge). For clarity, this information is formatted using standard concepts similar to those used for claims. However, these attributes are *not* directly derived from claims data; see *Additional Notes* for more detail.

In cases where the dataset history does not extend far enough back in time to cover the episode, the episode will be marked with "Information Not Available" for source attributes. For example, if the episode admission occurred on July 1, 2017, but no stay data are available until June 1, 2018, the details of admission for the episode will be marked with "Information Not Available", with the Stay records starting on the first patient transition after June 1, 2018.

Episode attributes relying on the duration of the episode are calculated on available admission and discharge information. If no discharge information is available, episode duration will be calculated based on the last available date for the data source at the time the dataset was locked. Users should be aware that discharges occurring after the data lock date will not yet be reflected in the dataset and so should be taken into consideration when determining patient episode duration or survival.

Timing

Episode records represent a span of time. The initial admission date is the start of that time and the discharge date, if present, is the end of that time. If no discharge data were provided by the EMR or established by the CMS admission/discharge rules described above, the discharge date will be left blank.

A patient should be present in a given facility in at most one Episode record for any given point in time. If data gaps from the EMR indicate that a patient had two admissions to the same facility with no intervening leave or discharge, the first episode will be terminated based on the admission date and a new episode record created starting with the second admission date.

Granularity

A single record is generated for each distinct Episode for each patient for each facility to which they were admitted.

If a patient is transferred from one LTC facility to a second LTC facility without being discharged from the first facility, a new episode will be created for the second facility without terminating the first facility. Therefore, although a patient may be in only one Episode record at a given point in time for a given facility, a patient may be in multiple Episode records at a given point in time provided that those records describe different facilities.

Additional Notes

Disposition Codes – In cases where patient disposition is described by the source EMR, the normalization process attempts to coerce the values to standard codes, such as the UB-04 admission source codes for the *source_type* field. While these codes are typically used for coding health care claims, they are being used in a descriptive form in our context. Users requiring adjudicated claims information must refer to claims-linked data.

Cumulative Days in Facility (CDIF) – CDIF is calculated based on rules described in CMS's MDS-3 Quality Measure User's Manual[2]. For Stays in which the patient is "In House," the number of days of the Stay, including the first but not the last day, are summed across Stays to determine the number of days that the patient was in facility. Stays that are in progress as of the data cutoff use the data cutoff date as the last day of the Stay. Admissions in which the patient was admitted/returned and was discharged on the same day are also considered one day. The CDIF calculation uses EMR-supplied census information and does not reference MDS records. CDIF is only calculated for episodes that have continuous stay coverage; that is, if an episode began in June 2012 but Stay data are only available for that facility starting in June 2018, the *episode_cdif* field will be left blank.

[2] "QM Sample and Record Selection Methodology" from *MDS 3.0 Quality Measures User's Manual*, RTI International.

Data Elements

The following data elements are included in the Episode data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	Patient
episode_id	PK	Unique episode identifier.	
patient_id	FK	Patient identifier derived from EMR.	Patient
facility_id	FK	Facility identifier derived from EMR.	Facility
admit_date	Date	Date of admission.	
discharge_date	Date	Date of discharge. If no discharge was recorded by the EMR, this field remains blank.	
episode_cdif	Integer	Cumulative Days in Facility (CDIF) for this episode as of discharge or data cutoff date. For details, see *Additional Notes.*	
admit_source	Text	Source of the admission. Descriptions may vary based on data sources.	EMR-Based
admit_type	Text	Type of admission. Codes presented are based on UB-04 Admission Type Code definitions, which are either provided by the EMR or mapped based on descriptions provided by the EMR.	See Terminology
admit_diagnosis_ind	Text	Indicates if a diagnosis code is present for this admission. Y = Yes, Dx present; N = No, Dx not present.	
admit_diagnosis_code	Text	Diagnosis code for primary diagnosis on admission, if present.	See Terminology
admit_diagnosis_code_type	Text	Diagnosis code type for primary diagnosis on admission, if present. Options include “ICD9” and “ICD10”.	See Terminology
admit_diagnosis_code_desc	Text	Diagnosis code description for primary diagnosis on admission, if present.	See Terminology
discharge_disposition_code	Text	Discharge disposition code, if applicable.	See Terminology
discharge_disposition_desc	Text	Discharge disposition description, if applicable.	See Terminology
discharge_diagnosis_ind	Text	Indicates if a diagnosis code is present for this discharge.	
discharge_diagnosis_code	Text	Diagnosis code for primary diagnosis on discharge, if present.	See Terminology

Field	Type	Description	References
discharge_diagnosis_code_type	Text	Diagnosis code type for primary diagnosis on discharge, if present.	See Terminology
discharge_diagnosis_code_desc	Text	Diagnosis code description for primary diagnosis on discharge, if present.	See Terminology
created_date	Date	Record creation date per source system.	EMR-Based
updated_date	Date	Record update date per source system.	EMR-Based

Terminology

The following terminology are used in the Episode data object:

Field	Type	Definition	Values
admit_type	Categorical	Type of admission, modeled on UB-04 Type of Admission / Visit Code and as reported by EMR.	Emergency, Urgent, Elective, Newborn, Trauma Center, Information Not Available.
admit_diagnosis_code	Codeset	Diagnosis code for primary diagnosis on admission. Valid codesets include ICD-9-CM, ICD-10-CM and SNOMED-CT.	
admit_diagnosis_code_type	Categorical	Codeset for the *admit_diagnosis_code*, as reported by the EMR.	ICD9, ICD10, SNOMED
admit_diagnosis_code_desc	Text	Description of the *admit_diagnosis_code*, as reported by the EMR.	
discharge_disposition_code	Categorical	Discharge disposition code, describing the nature and destination of the discharge, based on UB-04 Patient Discharge Status Code.	
discharge_disposition_code	Categorical	Discharge disposition description, describing the nature and destination of the discharge, based on UB-04 Patient Discharge Status Code.	

Field	Type	Definition	Values
discharge_diagnosis_code	Codeset	Diagnosis code for primary diagnosis on discharge. Valid codesets include ICD-9-CM, ICD-10-CM and SNOMED-CT.	
discharge_diagnosis_code_type	Categorical	Codeset for the *discharge_diagnosis_code*, as reported by the EMR.	ICD9, ICD10, SNOMED
discharge_diagnosis_code_desc	Text	Description of the *discharge_diagnosis_code*, as reported by the EMR.	

Stay

Definition

A stay represents a span of time in which a patient is in a certain status within (or outside) a facility. That status is determined based on a set of data elements; if any of those data elements change, the existing stay ends and either a new stay begins or an episode ends.

One or more stays comprise an episode. At any given time during an episode, a patient should always be associated with one and only one stay that is associated with that episode. Not every stay record indicates that a patient is present in the facility, however. Patient status can include any of the following:

- In House – The patient is present in the stay facility.
- Hospital Leave – The patient has been transferred to a hospital but is still considered admitted to the stay facility.
- Therapeutic Leave – The patient has been transferred to or is attending a clinical setting outside the stay facility but is still considered admitted to the stay facility.
- Outpatient – The patient has been transferred to or is attending an outpatient visit but is still considered admitted to the stay facility.
- Other – The patient has been transferred to an unspecified setting outside the stay facility but is still considered admitted to the stay facility.

Although the definition of stay status varies by EMR vendor and facility, the following factors generally define the patient's status during a stay:

- Episode – Any change in episode status, such as a change in facility or discharge, triggers an end to the current stay.
- Status – Changes in the patient status triggers a new stay.
- Payer – For most facilities, a change in payer triggers a change in stay.
- Room – In cases where room information is provided, a change in room triggers a change in stay.

Integration

The concept of a Stay is represented similarly across all integrated systems. That is, all source systems provide a representation of patient movements during their episode at a facility. Some mapping has been required to translate the descriptions used in one EMR to more general language. However, the detail may differ by source.

For example, some systems include numerous records indicating room changes, payer changes, and even internal organizational or workflow changes that are not necessarily evident from the information provided by the EMR. The reason for these types of changes may simply be labeled as “Information Change”. In contrast, other EMRs provide less detail, resulting in fewer Stay records describing a similar Episode.

Payer type (*payer_type*) is also reported inconsistently between systems. At time of writing, keyword matches are used to identify the payer type; for example, the Medicare payer type is associated with keywords like “Medicare”, “MCR”, and some region-specific variants.

The hospice flag (*hospice_ind*) is set when the term “hospice” or other related identifier is found.

Timing

Stay records represent a span of time. The start and end date fields are the boundaries of that span. Although those fields are timestamp data types, not all start and end dates are recorded with the same precision.

As with Episodes, a patient should be present in a given facility in at most one Stay record at any point in time. Concurrent Stay records may be present for a given patient if, for example, they are on hospital leave from one facility in the LTC Data Cooperative dataset and have been admitted to another facility in the LTC Data Cooperative dataset during that time. Those Stays will be tied to different Facilities and different Episodes so the presence of the patient will be unambiguous.

Granularity

A single Stay record is generated for every census interval described by the source system.

As described above, this can result in a much larger number of “Information Change” stay transitions for some sources compared to others. This integration policy has been established to maintain the fidelity and traceability of the original patient data. For most analytic purposes, filtering the data for the desired patient state or transition reason should exclude most unneeded references.

Data Elements

The following data elements are included in the Stay data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources	Patient
stay_id	PK	Unique identifier for set of contiguous days spent by the patient in the facility.	
patient_id	FK	Patient identifier derived from EMR.	Patient
facility_id	FK	Facility identifier derived from EMR.	Facility
episode_id	FK	Identifier for the episode of which this stay is a part.	Episode
start_date	Date	Timestamp of start of stay.	
end_date	Date	Timestamp of end of stay.	
payer_type	Categorical	Type of payer associated with this stay.	See Terminology
payer_name	Text	Name of the payer associated with this stay.	EMR-based
hospice_ind	Categorical	Indicator of whether the payer (or plan) associated with this stay is marked as hospice.	See Terminology
patient_status	Categorical	Status of the patient during this stay, particularly as to whether or not they are in the facility.	See Terminology
start_reason	Categorical	Reason for the start of this stay or change in status from a previous stay.	See Terminology
end_reason	Categorical	Reason for the end of this stay or change in status.	See Terminology
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

Terminology

The following terminology are used in the Stay data object:

Field	Type	Definition	Values
payer_type	Categorical	Type of payer associated with this stay.	Medicare, Medicaid, Commercial, Private
hospice_ind	Categorical	Indicator of whether the payer (or plan) associated with this stay is marked as hospice.	Y, N
patient_status	Categorical	Status of the patient during this stay.	In House, Hospital Leave, Therapeutic Leave, Outpatient, Other
start_reason	Categorical	Reason for the start of this stay or change in status from a previous stay.	Admission, Return, Information Change, Internal Transfer, Leave, Outpatient
end_reason	Categorical	Reason for the end of this stay or change in status.	Return, Information Change, Internal Transfer, Leave, Discharge, Outpatient, Outpatient Discharge, Expired

Advance Directive

Definition

An advance directive represents documentation of intentions related to care in the event that they are unable to represent their interests themselves, previously elected by the resident or their healthcare proxy.

A record of an advance directive can represent different events and actions, including medical orders (e.g., Do Not Resuscitate (DNR) or Full Code), establishing power of attorney or a living will, and completion of Physician Orders for Life-Sustaining Treatment (POLST). Advance directives can also document communication with residents or their representatives (for example, autopsy requests).

At time of writing, the following advance directive orders are supported:

- Full Code
- Do Not Resuscitate (DNR)
- Do Not Hospitalize (DNH)
- Do Not Intubate (DNI)
- Feeding Restrictions

Integration

While all EMRs capture and store advance directives, the information can be presented in different ways. The integration policies for each of the supported advance directive types is as follows:

Orders

Advance directive records are identified as an "Order" if they are specifically labeled in the EMR or the reported dataset as an order. Some timing information for the order must be present, and it is generally expected that the ordering provider be documented, as well.

Orders are classified as an advance directive record if the order description contains language unambiguously indicating that the order is associated with one of the advance directive types. In cases where EMRs use multiple choice templates to identify the type of advance directive in the order, logic may include specific inclusion parameters. For example, if the text reads "__X__ FULL CODE _____ DNR", transform logic will detect that only Full Code was selected. If the choice was ambiguous, the record will not be classified.

Only the text of the order itself is used to categorize the order type. References within the order to other documents in the EMR, such as a POLST/POST/MOLST on record, can be treated as additional documentation but are not used to specifically categorize the order.

Based on the descriptive text or on categorization by the EMR, a record may be assigned to one or more of the following categories:

Full Code – Full Code orders instruct the care providers to take all appropriate measures to sustain life, including resuscitation, compressions, and intubation. Any orders including a "Full Code" label (or similar) are assigned this type. However, it is still possible for additional directives to specifically restrict individual actions, such as nutrition or medication. A full code order will still be presented even in the presence of other directives.

Do Not Resuscitate (DNR) – DNR orders instruct health care providers not to do cardiopulmonary resuscitation (CPR). Any record unambiguously labeled as DNR is assigned this type. Orders for Comfort Care Arrest (DNR-CCA) and Comfort Care (DNR-CC) are also labeled as DNR. Records indicating a choice between Full Code and DNR are classified on a case-by-case basis.

Do Not Hospitalize (DNH) – DNH orders instruct health care providers not to transfer patients to a hospital setting. Orders stating, "do not hospitalize", "do not transfer", "DNH", or equivalent phrasing are assigned this type. Orders prohibiting hospitalization "unless comfort needs cannot be met" (or similar phrasing) are also assigned DNH.

Do Not Intubate (DNI) – DNI orders instruct health care providers not to intubate the patient. In the absence of a DNR, this would mean that the patient could be given chest compressions and/or cardiac drugs during CPR but could not place a breathing tube. Orders stating, "do not intubate", "no breathing tube", "DNI", or equivalent phrasing are also assigned DNI.

Feeding Restrictions – Feeding restrictions instruct health care providers on how to feed the patient. Instructions like "no artificial nutrition", "no gtube", and "no feeding tube" are included in this category. Some equivalent language is included, but orders only for hydration restriction or for nutritional changes (such as changes in diet or caloric intake) are not included in this type.

Timing

Timing for advance directive records differs depending on the type of record. Orders are generated at a point in time and are intended to be in effect until the order is updated or the patient is discharged.

Granularity

An advance directive record is generated every time an order relates to the topics discussed above.

For patients who are admitted for multiple episodes at the same facility, it is possible that advance directive records may be carried over from one episode to the next. That is, if a DNR was ordered for a patient for their first visit, a DNR may be ordered again for their second visit, or the existing order may be left in place. It is up to the user to determine the appropriate interpretation across multiple episodes.

Additional Notes

Status Registries – Some EMRs allow advance directive information to be presented as a set of status flags. As an example, the EMR may maintain a registry containing data on all active patients. Since the timing and provenance of individual orders and documents are not available through this mechanism, advance directive registry data are not currently included in the Dataset.

Data Elements

The following data elements are included in the Advance Directive data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	Patient
advance_directive_id	PK	Unique identifier for a record describing an advance directive event or documentation.	
patient_id	FK	Patient identifier derived from EMR.	Patient
provider_id	FK	Provider identifier associated with the advance directive record.	Provider
facility_id	FK	Facility identifier derived from EMR.	Facility
episode_id	FK	Episode identifier associated with advance directive record.	Episode
advance_directive_record_type	Categorical	Type of advance directive record.	See Terminology
advance_directive_type	Categorical	Type of advance directive event or documentation described.	See Terminology
advance_directive_description	Text	Description of an advance directive event or documentation provided by the source system.	
advance_directive_order_date	Date	Date when the advance directive record was generated.	
advance_directive_start_date	Date	Start date of advance directive event or documentation.	
advance_directive_end_date	Date	End or discontinuation date of advance directive event or documentation.	
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

Terminology

The following terminology are used in the Advance Directive data object:

Field	Type	Definition	Values
advance_directive_record_type	Categorical	Type of advance directive record.	Order
advance_directive_type	Categorical	Type of advance directive event or documentation.	Order Record Type: - Full Code - DNR - DNH - DNI - Feeding Restrictions

Allergy

Definition

An allergy represents an assessment of or a report by a patient of a current or historical allergy or intolerance to a foreign substance, such as a food, medication, or environmental factor.

A record of an allergy typically includes at least the identity of the patient, the approximate date of onset of the allergy, and the nature of the allergen (or lack of allergies reported). Depending on source data, an allergy record may also include the assessing provider and details about the nature and severity of the reaction. Allergies that are resolved or no longer present may include a resolved date, as well.

Integration

Most EMRs treat records of patient allergies as a distinct clinical concept. The main differences between data sources are in the presence and consistency of information related to how the information was documented (and by whom) and whether allergies are considered resolved.

Where possible, standardized codes have been used to describe reactions and severities. Most EMRs use the SNOMED-CT codeset to report the nature and severity of the reaction; in some cases, descriptions have been coerced to appropriate SNOMED-CT codes for consistency. Allergens are not always reported using standardized codesets and allergen types may not be documented properly. As a result, the dataset includes allergen descriptions as possible, but may be redacted in cases where PII/PHI could be exposed by sharing the full description.

In addition to records describing specific allergens, “No Known Allergy” (NKA) and “No Known Drug Allergy” (NKDA) records indicate that no allergies have been identified for the individual. These records are flagged with a category label of “NKA” or “NKDA,” even if descriptive text is not available. It is still possible for an active allergy record to overlap with an NKA/NKDA record; the user must determine how to resolve such situations. NKA/NKDA records can also be excluded from the data configuration on request.

Timing

Allergy records represent a span of time, in that they include a start (onset) date and an end (resolved) date. Due to frequent omission of end dates in allergy records, absence of an end date may be due to continuation of the allergy or lack of a documentation.

Granularity

An allergy record is generated for each unambiguously continuous span for which a specific allergen/reaction/severity combination is documented. New allergy records can be generated if the allergen or reaction or severity change; multiple allergy records are be generated for allergens with multiple documented reactions. A new allergy record can also be generated if a resident has no documented allergies on admission, or if existing allergies are resolved or closed.

Although some data sources attempt to manage allergy records across episodes, users should assume that allergy records reporting the same or similar allergies may be repeated over time and even within an episode. Since allergy records could be individual assessments by the provider or may be autogenerated by the EMR (e.g., on a repeat visit), users should not assume that repeated allergy records constitute repeated assessments of the allergy.

Additional Notes

Allergen Descriptions – Allergen descriptions are extracted directly from EMR sources. Except for redaction due to potential presence of PHI, allergen descriptions appear as they were entered in their original EMR. While most EMRs present one allergen per record, some permit multiple allergens to be represented in a single record. Sometimes, this is done using a comma-separated list, while in other cases the language is more free-form. Future enhancements may include splitting multiple allergens into multiple records. As of writing, users should be aware that a single record may describe multiple allergens.

Episode Assignment – The timing around when an allergy record was generated and to what time period it applies may vary by EMR. In some cases, documentation of an allergy is unambiguously associated with a specific episode. In other cases, allergies are only presented as spans in time and may not be clearly associated with a specific episode. To maintain consistency with other clinical events and documentation, we attempt to identify a most likely episode during which the allergy was documented, such as by using timestamps within the EMR or identifiers used only during certain episodes. When this is not possible, an episode may be assigned to an allergy record based on temporal proximity to the span documented. Given this variability, we advise that users aligning allergies with specific episodes consider all factors relevant to their needs throughout this process.

Data Elements

The following data elements are included in the Allergy data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	Patient
allergy_id	PK	Unique identifier for a record describing a specific allergen reported or assessed, optionally linked to a specific reaction.	
patient_id	FK	Patient identifier derived from EMR.	Patient
facility_id	FK	Facility identifier derived from EMR.	Facility
provider_id	FK	Provider identifier associated with the allergy record.	Provider
episode_id	FK	Episode identifier associated with the allergy record. See *Additional Notes*.	Episode
allergy_status	Categorical	Status of the allergy at time of last update.	See Terminology
allergy_start_date	Date	Documented date of allergy onset.	
allergy_stop_date	Date	Documented date of allergy resolution, if available.	
allergen_category	Categorical	Type of allergen, if available.	See Terminology
allergen_description	Text	Description of the allergen, as documented in the EMR. See *Additional Notes*.	EMR-based
allergen_code	Text	Standardized code describing the allergen (if available).	See Terminology
allergen_code_type	Text	Allergen code type (if applicable).	See Terminology
allergen_code_desc	Text	Allergen code description (if applicable).	See Terminology
reaction_description	Text	Description of the allergy reaction, as documented in the EMR.	EMR-based
reaction_code	Text	Standardized code describing the allergy reaction (if available).	See Terminology
reaction_code_type	Text	Allergy reaction code type (if applicable).	See Terminology
reaction_code_desc	Text	Allergy reaction code description (if applicable).	See Terminology
severity_description	Text	Description of the allergy severity, as documented in the EMR.	EMR-based
severity_code	Text	Standardized code describing the allergy severity (if available).	See Terminology
severity_code_type	Text	Allergy severity code type (if applicable).	See Terminology

Field	Type	Description	References
severity_code_desc	Text	Allergy severity code description (if applicable).	See Terminology
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

Terminology

The following terminology are used in the Stay data object:

Field	Type	Definition	Values
allergen_category	Categorical	Type of allergen, if available.	Food, Drug, Environmental, Substance, Other, NKA ["No Known Allergies"], NKDA ["No Known Drug Allergies"]
allergen_code	Text	Structured code describing the allergen, as reported by the EMR. Codes currently in use include SNOMED-CT and First DataBank (FDB).	
allergen_code_type	Text	Codeset for the *allergen_code*, as reported by the EMR.	FDB, SNOMED-CT
allergen_code_desc	Text	Description of the *allergen_code,* as reported by the EMR.	
reaction_code	Text	Structured code describing the allergy reaction. Currently, only SNOMED-CT codes.	
reaction_code_type	Text	Codeset for the *reaction_code*.	SNOMED-CT
reaction_code_desc	Text	Description of the *reaction_code*.	
severity_code	Text	Structured code describing the allergy severity. Currently, only SNOMED-CT codes.	
severity_code_type	Text	Codeset for the *severity_code*.	SNOMED-CT
severity_code_desc	Text	Description of the *severity_code*.	

Assessment / Assessment Element

Definition

An Assessment is any instrument that is used by a provider or staff to assess and report the status of a patient or an event. An Assessment Element is a record of a specific question-response pair associated with an Assessment.

Assessments include many of the commonly assessed instruments used in the long-term care setting, such as MDS, ADL and other cognitive assessments. In some cases, a result value for the assessment is provided by the source system, which is included in the Assessment record. Typically, Assessment Element records contain responses to specific questions or prompts, although in some cases the result of the entire assessment may be included as one of the question-response pairs.

Many users may be familiar with Assessment information presented in "wide-table" format with a variable per column. Although the LTC Data Cooperative data model stores Assessment Elements in a narrow, "normalized" form, most users will make use of standardized presentations of Assessments resembling the more traditional wide-table format.

Integration

The integration policy for Assessments and Assessment Elements is to retain as much detail as possible about the assessment process while also harmonizing the assessment structures to as great a degree as possible given available information.

Considerable variation exists between EMR systems, companies, and facilities in how assessment information is stored. In addition, changes in workflows over time result in shifts in the structure of assessments within facilities. For example, some source systems provide complete responses to all cognitive tests regardless of the circumstances, whereas other source systems only provide assessment scores for those tests.

For this reason, the integration process for Assessments and Assessment Elements is as follows:

1. Identify appropriate assessments. In some cases, such as the MDS-3, assessments are clearly labeled using a uniform structure. In other cases, some exploration is needed. For example, some systems clearly denote PHQ-9 exams both within and outside of MDS-3 assessments, while other systems label PHQ-9 exams in facility-dependent ways.

2. In cases where an assessment uses a standard format – for example, the MDS-3, PHQ-9, and BIMS assessments all have standard questions – the questions for each assessment form are confirmed to match those of the standard assessment. If they do not, a decision is made as to whether the assessment is similar enough to the standard.
3. In cases where an assessment does not use a standard format – for example, descriptions of falls differ between facilities and workflows – a set of key attributes are identified for inclusion and those attributes are manually identified for each assessment. They can then be represented in that form for users, with an understanding that there may be inconsistency in the presence and context of some of the data elements.

As new data sources are integrated, variations in incorporation of assessments arise. Therefore, the recency for assessments may differ by facility in cases where integration of new assessment structures lags behind availability of all other EMR data.

In general, assessment data are captured "as-is". Assessment elements considered PHI or PII (such as detailed resident information found in MDS-3 assessments) are restricted for distribution unless a specific research protocol has been approved to work with these restricted data. Unless noted otherwise, no additional effort is made to alter data or filter results.

Timing

The timing of an assessment is specific to the assessment and cannot be generalized here.

For example, a quarterly MDS-3 assessment represents a span of time defined by the elements within the assessment. In contrast, an individual PHQ-9 assessment represents the state of the patient at the time the assessment was conducted.

It is up to the user's discretion whether the *assessment_date* of the assessment or one or more dates in the assessment responses best characterize the timing of an assessment.

Granularity

An Assessment record is generated for every unique instance of an assessment provided by the EMR, regardless of the presence of a result value or the number of question-response pairs.

An Assessment Element record is generated for each question-response pair provided by the EMR, except in cases where PII or PHI may be exposed. If no question-response pairs are provided for an assessment, no Assessment Element records are generated. Since assessment results are captured by Assessment records, no new Assessment Element records are generated to store assessment results.

Data Elements

The following data elements are included in the Assessment data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	Patient
assessment_id	FK	Record identifier associated with this assessment.	
patient_id	FK	Patient identifier derived from EMR.	Patient
facility_id	FK	Facility identifier derived from EMR.	Facility
provider_id	FK	Assessing provider identifier.	Provider
episode_id	FK	Episode identifier associated with this assessment.	Episode
assessment_date	Date	Date the assessment was completed/documented.	
assessment_type	Categorical	Type of assessment administered. Options include MDS-3 responses (MDS3), assessments of function and cognition (e.g. ADL, BIMS), and behavioral or psychosocial assessments (PHQ-9, PHQ-9-OV).	See Terminology
assessment_result	Text	Result of the assessment, if provided. The nature of the result will depend on the nature of the assessment.	
assessment_result_desc	Text	Description of the result of the assessment, if provided.	
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

The following data elements are included in the Assessment Element data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	Patient
assessment_element_id	PK	Record identifier associated with this record's question-response pair.	
assessment_id	FK	Assessment identifier providing a unique identifier for each assessment.	Assessment
patient_id	FK	Patient identifier derived from EMR.	Patient

Field	Type	Description	References
facility_id	FK	Facility identifier derived from EMR.	Facility
provider_id	FK	Assessing provider identifier.	Provider
episode_id	FK	Episode identifier associated with this assessment.	Episode
assessment_date	Date	Date the assessment was completed/documented.	
assessment_type	Categorical	Type of assessment administered. Options include MDS-3 responses (MDS3), assessments of function and cognition (e.g. ADL, BIMS), and behavioral or psychosocial assessments (PHQ-9, PHQ-9-OV).	See Terminology
assessment_section	Categorical	Category or section of the assessment.	EMR-based
assessment_question	Text	Assessment record question description.	EMR-based
assessment_question_code	Text	Assessment record question code, if applicable. When an assessment is based on a standard instrument (e.g. MDS-3, PHQ-9, BIMS), the assessment question code should correspond with an appropriate question number or code in the instrument.	EMR-based
assessment_response	Text	Response value for the assessment record.	
assessment_response_code	Text	Response code for the assessment record, if applicable.	
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

Terminology

The following terminology are used in the Assessment data object:

Field	Type	Definition	Values
assessment_type	Categorical	Type of assessment administered.	MDS3, CPS, BIMS, CAM, ADL, FS, PHQ-9, PHQ-9-OV

Condition

Definition

A Condition is any clinical condition, diagnosis, problem, or issue relevant to the clinical status of the patient.

In the LTC Data Cooperative dataset, Conditions can include diagnoses, allergies, and other clinical events[3]. They include the date and time of the diagnosis or observation, the diagnosing or observing provider, and any relevant codes (such as ICD-10 codes), and context within the episode. Additional attributes may indicate whether the diagnosis or observation was made on admission, upon discharge, during the patient's stay, or as part of a historical review.

Integration

Conditions are harvested from a wide variety of sources within the EMRs, depending on the source.

All source systems at least provide a list of diagnoses, by patient, date, and diagnosis code, although some provide minimal additional context.

Some source systems provide additional diagnostic data associated with admissions and discharges. For example, some systems include discharge diagnoses with other patient discharge data, whereas others never provide this information. The sources of diagnoses have been marked on a per-record basis accordingly.

Most source systems also have the concept of a Primary Diagnosis. However, the context behind that designation – for example, *Is the Condition primary to that specific event or the patient's stay in the facility?* – is not generally provided. Furthermore, multiple diagnoses can be flagged as primary throughout a patient's stay, or even during a specific clinical assessment. Therefore, no guarantees are made about the utility or clinical meaningfulness of any Condition record marked with the primary diagnosis flag.

[3] At time of writing, only diagnoses are supported in the Condition data object. Allergies and non-diagnostic clinical events are not yet included.

Timing

The timing reflected in Condition data depend on the type of record being inspected.

Records documented at admission are typically a “snapshot” of patient health at that time. A mix of acute and chronic conditions provide an overall view of the health of the patient on admission. The history of the condition may not be included since the purpose of the documentation is to indicate the health of the patient at time of admission.

Following admission, Conditions may be reported to indicate changes in health, such as new injuries or worsening in chronic conditions, as well as to indicate the status quo. Users interested in the specific timing of onset and/or resolution of conditions, and especially chronic conditions, may need to assess the context of those records on a case-by-case basis.

Granularity

A record is generated for each mention of a condition in the EMR.

For example, a patient with diabetes may have that information recorded on admission. Records provided to the EMR may indicate historical assessments predating admission, as well. And, throughout the patient’s stay, the patient’s diabetes diagnoses may be noted as a part of routine exams.

Frequency and consistency of repeated diagnoses can depend on a variety of factors, including differing workflows between companies and facilities.

Data Elements

The following data elements are included in the Condition data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	
condition_id	PK	Unique condition record identifier.	
patient_id	FK	Patient identifier derived from EMR.	Patient
facility_id	FK	Facility identifier derived from EMR.	Facility
provider_id	FK	Diagnosing or assessing provider identifier.	Provider
episode_id	FK	Episode identifier associated with this diagnosis (if applicable).	Episode
condition_type	Categorical	Type of condition documented. Only supported option is: - **Diagnosis**: Diagnosis assessed by a qualified provider or relayed by patient history.	
condition_event	Categorical	Circumstances under which the condition was documented, as reported by the EMR.	See Terminology
condition_code	Text	Condition code (if applicable).	See Terminology
condition_code_type	Text	Condition code type (if applicable).	See Terminology
condition_code_desc	Text	Condition code description (if applicable).	See Terminology
condition_status	Categorical	Condition status (as of reporting date)*.	See Terminology
condition_date	Date	Date at which the condition was assessed.	
onset_date	Date	Condition onset date**.	
resolved_date	Date	Condition resolved date**.	
primary_dx_ind	Text	Indicator of whether condition is a primary diagnosis (Y/N).	
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

** Condition status is populated if the record is documenting a historical condition (*condition_event='History') *or the EMR explicitly indicates onset/resolution dates. Otherwise, condition status is left NULL.*

*** Condition onset/resolved dates are populated if the record is documenting a historical condition (condition_event='History') or the EMR explicitly indicates onset/resolution dates. If dates are not known, these fields will be left NULL.*

Terminology

The following terminology are used in the Condition data object:

Field	Type	Definition	Values
condition_event	Categorical	Circumstances under which the condition was documented, as reported by the EMR.	Admission, Discharge, Stay, History
condition_code	Codeset	Structured code describing the condition, as reported by the EMR. Acceptable codes include ICD-9-CM, ICD-10-CM and SNOMED-CT.	
condition_code_type	Categorical	Codeset for the *condition_code*, as reported by the EMR.	ICD9, ICD10, SNOMED
condition_code_desc	Text	Description of the *condition_code*, as reported by the EMR.	
condition_status	Categorical	Status of the Condition, as reported by the EMR.	Active, Resolved

Observation

Definition

Observations are measurements or assertions made about the state of the patient. These include vitals readings, laboratory tests, and clinical findings.

Observations can take a number of different forms, including numerical, ordinal, and nominal results. Depending on the data source (and user permission), some observations may also include unstructured text. At this time, images and attached documents are not included as observations, although summaries generated by providers or the EMR of images and documents are sometimes present.

Integration

Observations are divided into two distinct concepts, each of which is represented in all sources.

1. Vital signs are represented similarly in all EMR sources. They are represented as individual values – e.g. respiration rate, temperature, weight – recorded at a specific time for a patient by a specific individual. The exception is blood pressure: some sources present systolic and diastolic values together, whereas others store these values separately. In all cases, vitals values are stored as individual readings. Users can then choose to view these readings individually – e.g. temperature readings for every day in a three-month span – or combined – i.e. a record for each patient for each day representing the composite of all vitals readings.

   The following vitals are considered "standard" for the purposes of the LTC Data Cooperative dataset: Temperature, weight, height, diastolic blood pressure, systolic blood pressure, respiration, pulse, pain[4], and blood glucose measurement[5]. All these measures are supported by all sources, although values are not supplied uniformly by all facilities for all patients. Units are also provided directly by sources and are not imputed.

2. Lab Results are represented differently in different EMR sources, with differing consistency, coverage, and reliability.

[4] Representation of pain measures vary by source system; observation name detail and units indicate the specifics around this representation.

[5] Point-of-care blood glucose ("Blood Sugar") measurements are stored as Vitals in contrast to lab-based blood glucose measurements to distinguish between the testing methodologies and workflows.

Source types include the following:

- Qualitative lab results by name with little reference to ordering and test specifics. For example, a record describing a COVID-19 test result may include the patient, the date/time of testing, and the result, listed as Positive, Negative, or Inconclusive. Qualitative results of this type have been standardized across sources, but little additional information is available.
- Quantitative lab results by name with little reference to test specifics beyond the type. For example, an order for a complete blood count (CBC) may be listed, along with the dozen resulting values and the names of the test results (such as "WBC"). However, no LOINC[6] code or collecting provider is available.
- Standard lab results by name and code, along with additional information provided by the lab, typically a third party. For example, the CBC test described above might include a LOINC code for the initial order, with detailed information about each of the specific test results.

The coverage of the above lab types varies across sources. Some sources only support certain types of the above labs. Availability of third-party lab data vary by facility, with some providing substantial third-party lab data and others offering none. Some sources filter lab data availability to certain test groups, resulting in gaps of coverage.

The integration process has attempted to harmonize lab results based on data availability. However, the user is responsible for determining if a specific lab test is available from a particular facility.

Timing

Observations represent the timing around the process of performing and reporting on a measurement.

The *observation_order_date*, if provided, indicates when the observation was ordered by a provider.

The *observation_event_date* typically indicates when the measurement itself was performed. For vitals, this would be when the patient was checked for the specific vital value. For lab results, this

[6] The Logical Observation Identifiers Names and Codes (LOINC) codeset is maintained by the Regenstrief Institute, Inc. and the Logical Observation Identifiers Names and Codes (LOINC) Committee. More details at loinc.org.

would be when a specimen of some kind was “collected” for the purposes of performing the measurement.

Although timing of analysis and reporting are also sometimes provided by the EMR, those data are not yet included in the Observation data domain.

Granularity

An Observation record is generated for every qualitative or quantitative observation reported by the EMR.

Three types of Observations are supported in the LTC Data Cooperative dataset:

- Vitals – Records reflecting a vitals measurement.
- Lab – Records reflecting a lab result measurement or calculation.
- Order – Records only reflecting an order for vitals measurement(s) or labs.

Single-measure observations, such as weight or HbA1c, can be presented in a single record. Multiple-measure events are spread across multiple records. Blood pressure is stored in two records, one for systolic blood pressure and one for diastolic blood pressure. Dozens of records can accompany a CBC lab test, with each individual test (e.g. RBC count) taking up one Observation record.

Data Elements

The following data elements are included in the Observation data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	
observation_id	PK	Unique observation record identifier.	
patient_id	FK	Patient identifier derived from EMR.	Patient
facility_id	FK	Facility identifier derived from EMR for this observation.	Facility
episode_id	FK	Episode identifier associated with this observation.	Episode
observation_record_type	Categorical	Type of observation event described by this record. Options are: - **Order**: Describes the observation order as listed by the data source. - **Vitals**: Describes a vitals result, with or without accompanying order information. - **Lab**: Describes a lab result, with or without accompanying order information.	See Granularity
observation_name	Categorical	Summary description of the observation described by this record. For Vitals and Lab records, the observation may include multiple results or tests, each described in a separate record. See Observation Events Supported table.	See Terminology
observation_name_detail	Categorical	Additional detail for the observation described by this record. See Observation Events Supported table.	See Terminology
observation_order_id	FK	Order identifier associated with this observation. For Order records, this will be the same as the *observation_id*. For Vitals and Lab records, this will be the *observation_id* for the associated order, or NULL if no order record is present.	Observation
observation_order_date	Date	Date that the observation was ordered if an order record is present. NULL if no order record is present.	
observation_order_provider_id	FK	Provider identifier for ordering provider if an order record is present. NULL if no order record is present or no ordering provider is listed.	

Field	Type	Description	References
observation_event_date	Date	Date that the observation occurred. For Vitals and Lab records, this may reflect the date on which an assessment or specimen collection occurred. For Order records, this will only be included if the observation event date is included with the order.	
observation_event_provider_id	FK	Provider identifier for provider listed as performing the action associated with the observation. For Vitals and Lab records, this may be the provider who conducted an assessment or collected a specimen. For Order records, this will only be included if the responsible provider is included with the order.	Provider
observation_event_code	Text	Standard code used to describe observation event, if available. For observations with multiple tests (e.g. panels), this code describes the top-level event.	See Terminology
observation_event_code_type	Text	Type of standard code used to describe observation event, if available.	See Terminology
observation_event_code_desc	Text	Description of standard code describing observation event, if available.	See Terminology
observation_result_name	Categorical	Type or name of result described by this record. For Vitals, this will be the most specific name available. For Labs, this is the specific test name. For Orders, all result fields are NULL.	
observation_result_value	Text	Result value of the Observation, described without units. This result may be numeric or non-numeric, depending on the context.	
observation_result_value_numeric	Float	Result value of the Observation, described without units, if the value is purely numeric in form.	
observation_result_value_categorical	Categorical	Result value of the Observation, described as a categorical value, such as "Positive" or "Negative". Categorical values supported differ based on the type of Observation.	EMR-based
observation_result_units	Text	Units to apply to the observation result value.	EMR-based
observation_result_code	Text	Standard code used to describe observation result, if available.	See Terminology
observation_result_code_type	Text	Type of standard code used to describe observation result, if available.	See Terminology

Field	Type	Description	References
observation_result_code_desc	Text	Description of standard code describing observation value, if available.	See Terminology
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

Terminology

The following terminology are used in the Observation data object:

Field	Type	Definition	Values
observation_name	Categorical	High-level description of the observation.	For **Vitals** records: Blood Pressure, Blood Sugar, Height, Pain, Pulse, O2 Saturation, Respirations, Temperature, Weight For **Labs** records, names are supplied by the EMR.
observation_name_detail	Categorical	Additional detail or subtype of the observation name.	Dependent on the *observation_name*.
observation_event_code	Codeset	Structured code describing the observation order/panel, as reported by the EMR. Acceptable codesets include LOINC[7] and SNOMED-CT.	
observation_event_code_type	Categorical	Codeset for the *observation_event_code*, as reported by the EMR.	LOINC, SNOMED
observation_event_code_desc	Text	Description of the *observation_event_code*, as reported by the EMR.	

[7] LOINC is copyright © 1995-2023, Regenstrief Institute, Inc. and the Logical Observation Identifiers Names and Codes (LOINC) Committee. See disclaimer for full details.

Field	Type	Definition	Values
observation_result_code	Codeset	Structured code describing the specific observation result, as reported by the EMR. Acceptable codesets include LOINC and SNOMED-CT.	
observation_result_code_type	Categorical	Codeset for the *observation_result_code*, as reported by the EMR.	LOINC, SNOMED
observation_result_code_desc	Text	Description of the o*bservation_result_code*, as reported by the EMR.	

Medication

Definition

This table contains information related to orders for, administrations of, and historical use of medications by patients. Here, medications are defined as any substance that has a biochemical effect on patients when consumed. They can include prescription as well as over-the-counter medications. This table does not include non-pharmacological interventions such as radiation, surgery, or implantable devices, although medications administered by implantable devices can be included.

Three kinds of Medication records are currently supported:

1. Medication Order records – i.e. prescriptions – describe an order by a qualified provider for a patient to receive a medication, including the details of the medication, how it is to be administered, how frequently, and for how long.
2. Medication Administration records describe each administration of a medication by provider or staff member to the patient, including the date and time of the administration and which specific medications were administered.
3. Medication History records describe a patient's medication history, either provided by the patient, a current provider, or another health care provider, indicating which kinds of medications have been or are currently being used by the patient and approximate spans of time during which the medication was prescribed or taken.

Although vaccines do count as medications, all records of vaccinations can be found in the Vaccination data object rather than the Medication data object.

Integration

Due to the critical nature of medication management, all sources provide a standard set of attributes related to medication orders and administrations.

Naming conventions do differ somewhat by sources. For example, some sources separate generic and brand names into different data elements, whereas others offer only one medication name data element, which can include generic names, brand names, or both. The harmonization process has attempted to separate out the generic medication name into a separate data element when possible. However, users are encouraged to search for generic medication names in both the *medication_name* and *medication_generic_name* fields.

Data availability varies by EMR and facility. Sources of variability include:

- History of administration data – Some sources provide several years of administration data, whereas others only provide more recent (and ongoing) administration data.
- Presence of medication codes – Some sources provide product-level National Drug Codes (NDCs) for all medications, while others only offer names or medication classes.
- Use of medication classes – Different EMRs use different medication classification schemes. At this time, LTC Data Cooperative data present these classification data as-is, without any attempts to harmonize between medication classes.

In all cases, data elements are included when available, but consistency is not guaranteed.

Timing

Medication records represent events occurring at a particular time, although they have information that can be interpreted to represent encompassing spans of time.

The actual events reflected by the record, such as creation of a prescription or the administration of a medication, are events at a point in time, as reflected by the *medication_order_date* or *medication_event_date* fields. In either case, a date or a complete timestamp may be provided; the accuracy of administration timestamps, particularly, are subject to the timing of actual administration timing versus the timing and precision of data entry.

In addition to those events, medication records may also include date ranges for the span of the associated order, potentially providing information about previous and expected future administration events. These spans are based on the most recent information provided by the EMR at time of dataset generation and can be altered depending on the timing of order discontinuation or patient discharge.

Granularity

A Medication record is generated for each medication administration event. A Medication record is also generated for each Order record provided by the EMR. Users will typically interact with separate Order and Administration tables and should specify if a differing presentation is desired.

Data Elements

The following data elements are included in the Medication data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources	Patient
medication_id	PK	Unique medication record identifier	
patient_id	FK	Patient identifier derived from EMR	Patient
episode_id	FK	Episode identifier associated with this medication record.	Episode
medication_record_type	Categorical	Type of medication event described by this record. Options are: - **Order**: Describes the order or prescription as listed by the data source. - **Administration**: Describes a medication administration event as listed by the data source. Multiple administration records may exist for a single order record.	
medication_name	Text	Name of the medication as listed by the EMR. Where a generic and brand name are provided, the brand name is listed here.	EMR-based
medication_generic_name	Text	Generic name of the medication as listed by the EMR, when available.	EMR-based
medication_order_id	FK	Order identifier for this medication record. For Order records, this will be the same as the medication_id. For Administration records, this will be the medication_id for the associated Order record if one exists.	Medication
medication_order_provider_id	FK	Provider identifier for ordering provider if an Order record is present. NULL if no Order record is present or no ordering provider is listed.	Provider
medication_order_date	Date	Date on which the medication was ordered if an Order record is present. NULL if no associated Order record is present.	
medication_order_start_date	Text	Start date of the medication order, if applicable.	
medication_order_end_date	Text	End date of the medication order, if applicable.	

Field	Type	Description	References
medication_frequency	Text	Description of the frequency with which the medication should be administered. For Order records, this is the frequency listed on the order, if present. For Administration records, this is the frequency listed on the administration record if one is provided; otherwise, if a frequency is listed on an associated Order record, that frequency will be presented.	EMR-based
medication_event_date	Date	Date on which the medication was administered. NULL for Order records.	
medication_event_provider_id	FK	Provider identifier for the medication event. For Administration records, this is the provider who administered the medication. For Order records, this field is NULL (the ordering provider is listed in the medication_order_provider_id field).	Provider
medication_code	Text	Standard code for the medication, if applicable.	See Terminology
medication_code_type	Text	Code type of the standard code for the medication, if applicable.	See Terminology
medication_code_desc	Text	Description of the standard code for the medication, if applicable.	See Terminology
medication_class_code	Text	Classification code for the medication, as provided by the EMR.	See Terminology
medication_class_code_type	Text	Classification code type for the medication, as provided by the EMR, when applicable.	See Terminology
medication_class_code_desc	Text	Classification code description for the medication, as provided by the EMR, when applicable. Class description may be provided in lieu of class code when the latter is not available.	See Terminology
medication_form	Text	Medication form, as described by the EMR, if applicable.	EMR-based
medication_strength	Text	Medication strength, as described by the EMR, if applicable.	EMR-based
medication_dose	Text	Medication amount, as described by the EMR, if applicable.	EMR-based
medication_route	Text	Medication route, as described by the EMR, if applicable.	EMR-based
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

Terminology

The following terminology are used in the Medication data object:

Field	Type	Definition	Values
medication_code	Codeset	Standard code for the medication, if applicable. Acceptable codesets include NDC, SNOMED-CT, and RXNORM.	
medication_code_type	Category	Code type of the standard code for the medication, if applicable.	NDC, SNOMED, RXNORM
medication_code_desc	Text	Description of the standard code for the medication, if applicable.	
medication_class_code	Codeset	Classification code for the medication, as provided by the EMR. Availability of class codes may be subject to licensing restrictions.	
medication_class_code_type	Category	Classification code type for the medication, as provided by the EMR, when applicable.	MMSL (Multum MediSource Lexicon), GPI (Medi-Span Generic Product Indicator)
medication_class_code_desc	Text	Classification code description for the medication, as provided by the EMR, when applicable.	

Vaccination

Definition

This table describes events related to the administration of vaccines to a patient and records describing their vaccination history, as reported by the patient, a provider, or other data sources.

As discussed in the section on Medications, Vaccinations also fit the definition of a Medication. However, due to the differing requirements of Vaccination data by users, all Vaccination events are separated into the Vaccination object. Some examples may still exist in the Medication data object but for analysis purposes, users should specifically refer to the Vaccination table when analyzing Vaccination administrations and histories.

Note that the term "Immunization" is used in similar contexts to "Vaccination", depending on the data source. While there are practical and clinical distinctions between the terms, for the purposes of this dataset the terms are considered interchangeable and all references to vaccinations or immunizations will be considered Vaccinations.

Integration

At a high level, vaccination data are represented similarly from all sources; descriptions of the patient, the date of vaccination, and against what they are being vaccinated are all described.

However, considerable variation exists in what details are available between EMRs, companies, and facilities, generally falling into the following categories:

1. Characterization of Vaccine Product– Some sources provide detailed information about vaccines, including detailed descriptions, manufacturers, and NDCs. Especially in cases where multiple manufacturers and products vaccinate against the same condition, this can determine whether it is possible to know which specific product was used in any given case.
2. Characterizations of Vaccine Administered – Some sources provide additional information about the specific vaccine, such as lot number and expiration date. Although less critical in many cases than (1), such attributes are sometimes needed for health care operations research[8].

[8] In the special case of COVID-19 vaccinations, lot numbers can also be used to link vaccine lot numbers to more detailed information about the vaccine product.

3. Vaccine Administration Operations – Sources differ in the information they provide around health care operations related to vaccination. This includes information about orders, declinations, medical exclusions, rendering providers/staff, and patient consent and education.

Timing

Vaccination records indicate an event at a point in time, although the interpretation of that time may differ based on the record.

For Administration and History records, the *vaccination_event_date* indicates the date on which the vaccination occurred or is believed to have occurred. It is possible for multiple records to exist for a single vaccination event, and imprecision in data entry may result in those records being listed at different times or even different dates. Users should consider de-duplication needs when using these records.

For Decline records, the *vaccination_decline_date* indicates the date on which the vaccination declination was documented. Some workflows may not create a new record of refusal or declination once the patient has done so once, while others may document every refusal distinctly. Users should consider potential variation by facility and workflow when making use of these records.

Granularity

A record is generated for every Vaccination event, whether it is listed as administered by the facility, a historical record of a vaccination, or a declination or refusal of that vaccination by the patient.

Data Elements

The following data elements are included in the Vaccination data object:

Field	Type	Description	References
master_patient_id	FK	Uniform patient identifier across all sources.	Patient
vaccination_id	PK	Unique vaccination record identifier.	
patient_id	FK	Patient identifier derived from EMR.	Patient
facility_id	FK	Facility identifier derived from EMR.	Facility
order_id	FK	Order identifier for the prescription on which this administration is based. Note: Order records not yet available.	Vaccination
provider_id	FK	Prescribing provider identifier.	Provider
episode_id	FK	Episode identifier.	Episode
vaccination_event_type	Categorical	Type of vaccination event record. **Administration**: Vaccinations given by the facility or during the patient episode, regardless of the setting where the vaccine was administered. **History**: Vaccinations for which documentation was provided to the facility but not administered by the facility or during the patient episode. **Decline**: Vaccination was declined or refused by the patient or their representative.	
vaccination_event_date	Date	Date on which the vaccination event occurred.	
vaccination_consent_date	Date	Date on which the patient was provided educational materials and/or gave consent for vaccination.	
vaccination_decline_date	Date	Date on which the vaccination was declined.	
vaccine_type	Categorical	Type of vaccine used during the vaccination event, based on CDC-published vaccine groups**	See Terminology
vaccine_sequence	Text	Description of the vaccination in sequence if provided by the EMR. May be numeric or qualitative, depending on vaccination sequence guidelines.	

Field	Type	Description	References
vaccine_name	Text	Name of the vaccine as listed by the EMR. Description may be standard (i.e. from CDC) or vary depending on data entry method*.	EMR-based
vaccine_desc	Text	Description of the vaccine as listed by the EMR, if available.	EMR-based
vaccine_code	Text	Standard code for the vaccine, if available from the EMR or derived from lot number records*.	See Terminology
vaccine_code_type	Text	Code type of the standard code for the vaccine, if available.	See Terminology
vaccine_code_desc	Text	Description of the standard code for the vaccine, if available.	See Terminology
vaccine_cvx_code	int	The CVX code for the vaccine, as provided by the EMR or determined from other vaccine identifiers.	See Terminology
vaccine_cvx_desc	Text	The CVX code description for the vaccine.	See Terminology
vaccine_manufacturer	Text	Name of the vaccine manufacturer as listed by CDC.**	See Terminology
vaccine_lotnumber	Text	Lot number for the vaccine administered as recorded in the EMR, if available.	See Terminology
vaccine_expiration_date	Date	Expiration date for the vaccination, as recorded in the EMR or from lot number records.*	
vaccine_form	Text	Vaccine form, as described by the EMR, if available. For example, “suspension”.	EMR-based
vaccine_strength	Text	Vaccine strength, as described by the EMR, if available. For example, "50 mcg/0.5 mL".	EMR-based
vaccine_route	Text	Vaccine route administered, as described by the EMR, if available.	EMR-based
vaccine_dose	Text	Vaccine dose administered, as described by the EMR, if available. For example, "0.5 mL".	EMR-based
created_date	Date	Record creation date per source system.	EMR-based
updated_date	Date	Record update date per source system.	EMR-based

** Variables pulled preferentially from CDC records.*

*** Variables pulled exclusively from CDC records.*

For more information, see: https://www.cdc.gov/vaccines/programs/iis/

Terminology

The following terminology are used in the Vaccination data object:

Field	Type	Definition	Values
vaccine_type	Category	Type of vaccine used during the vaccination event, based on CDC-published vaccine groups. (link)	
vaccine_code	Codeset	Standard code for the vaccine, if available from the EMR or derived from lot number records. Currently, only NDC and HCPCS codes are accepted.	
vaccine_code_type	Category	Code type of the standard code for the vaccine, if available.	NDC, HCPCS
vaccine_code_desc	Text	Description of the standard code for the vaccine, if available	
vaccine_cvx_code	Codeset	The CVX code for the vaccine, as provided by the EMR or determined from other vaccine identifiers. (link)	
vaccine_cvx_desc	Text	The CVX code description for the vaccine. (link)	
vaccine_manufacturer	Text	Name of the vaccine manufacturer as listed by CDC. (link)	