MOHCCN Clinical Data Model v3.1 Guidelines

A. Background

This document provides guidelines to help data curators complete the MOHCCN Clinical Data Model. Click here to view the full data model, including complete lists of permissible values for each field. These guidelines provide some general examples and likely sources of the required information; however, differences in how clinical information is collected, stored and accessed may mean that some aspects of this guideline are not applicable to every site.

MOHCCN data are ingested into and shared through CanDIG[1], which is a federated system that allows data to be queried without leaving its site of origin. Each site may develop its own process for collecting and formatting clinical data to match the MOHCCN Clinical Data Model (e.g. using RedCAP or another system). CanDIG provides a python package, clinical_ETL_code, that helps with transforming data from a site's own data store (e.g. RedCAP) into the json format needed for ingest into CanDIG. Each site will work with its parent consortium to either deploy its own CanDIG instance or access another.

B. General guidelines

Identifiers

Several data elements in this model are identifiers (of the program, donor, specimen, etc.). ID conventions are to be developed within each consortium and program (cohort) to avoid duplications (i.e. the same ID being assigned to two or more donors). **IDs should be anonymized and should not contain any patient-identifying information** (e.g. name, hospital code).

Documents to review

For each patient, curators should review the medical record available at the institution that contributed the case. This should be reviewed from the date of diagnosis through to the current date. Documents to review should include, but are not limited to, pathology reports, CT/MRI/ultrasound/PET imaging reports, and clinical notes from any treating oncologist.

Null and blank values

[1] L. Jonathan Dursi et al., "CanDIG: Federated Network across Canada for Multi-Omic and Health Data Discovery and Analysis," *Cell Genomics* 1, no. 2 (November 10, 2021): 100033.

Most data elements have null options to indicate if the information is unavailable or inaccessible. Curators should conduct a thorough review of the patient's medical record and should only use null options or leave field blanks when absolutely necessary. Null values should be used in the following circumstances:

- "**Not available**" should be used for elements where the data point is not available for the patient and there is no reasonable expectation that it will ever be available.
    - For numeric fields, use "-99" to indicate "Not Available"
- If the data point is not available at the time of curation, but there is a reasonable expectation that it will be available in the future (e.g. after clinical follow-up or after a clinical trial report is released), the field should be left **blank**. This can be used to flag incomplete elements that are intended to be completed at a later date.

Questions, clarifications and requests

Curators can create an issue on the MOHCCN Clinical Data Model GitHub page (hosted by CanDIG) if they have questions about the model or how to complete it. Questions may be answered by other curators or by the CanDIG team as appropriate, or may be brought to the MOHCCN Clinical Data Standards subcommittee for clarification if necessary.

If curators encounter a data element where the list of permissible values is missing the intended value, they should complete the Change Request/Additional Information form on the MOHCCN Policies and Guidelines webpage describing the missing value and proposed change.

Data monitoring

Sites should create their own guidelines for monitoring of data quality and accuracy, including but not limited to periodic data audits.

C. Basic definitions

Clinical notes: Notes in the donor's medical record. A clinical note typically corresponds to a medical visit. For the purposes of this guideline, clinical notes refer to those that relate to the cancer diagnosis associated with the index specimen, plus those that relate to comorbidities and exposures as appropriate.

Data element: Individual component of the data model.

Donor: A living or deceased individual who is the source of the specimen whose sequencing data were submitted to the MOHCCN Gold Cohort.

Gold Cohort: The MOHCCN Gold Cohort is the dataset to which sequencing and clinical data are being contributed. The minimum requirements for a donor's data to be included in the Gold Cohort can be found in the MOHCCN Gold Cohort Policy (the up-to-date policy can be found on this page). A Gold Cohort sample requires, at minimum, tumour and normal whole-genome sequencing data and clinical data according to this model. Tumour transcriptome data should also be included where possible. Accompanying data may also be submitted, for example from tumour models derived from a specimen (e.g. cell lines, xenografts).

Index cancer: The cancer associated with the donor specimen, regardless of whether the specimen was obtained from a primary, recurrent or metastatic tumour. For example, if the specimen was obtained from a recurrent tumour, "index cancer" refers to the ensemble of the related primary and recurrent tumours.

Index specimen: Tumour specimen obtained from the donor and from which samples were derived, with subsequent sequencing data submitted to the MOHCCN Gold Cohort.

Medical record: Entirety of the donor's medical record, including clinical notes, reports, birth/death certificates, etc. The medical record may be divided across multiple institutions, and may not be in one single location.

Permissible values: The allowed options for a data element. In controlled vocabulary fields, the curator must choose a value from the set list. Open text fields are less restrictive, the the best practices included in these guidelines should be followed (e.g. capitalization). This document lists permissible values where the list is short, but complete lists can be found on the model spreadsheet. The most up-to-date spreadsheet can be found on this page.

Specimen: A sample of material (tissue, blood, etc.) collected for analysis. A specimen is usually collected via a standard- or research-indicated medical procedure (e.g. blood draw, biopsy, surgery).

Sample: A unit of biomaterial derived from a single specimen (e.g. DNA, RNA). A sample is generated in a laboratory from a specimen.

Schema: A single CanDIG object or table that contains related data elements. See section D and the entity relationship diagram to view how different schema relate to each other. Each schema is defined in section E.

Source: The original record where clinical data was originally documented. This includes medical records, laboratory reports, imaging results and other documents that provide evidence of clinical care, treatments and outcomes. This document lists likely sources for each data element; however, specifics may differ between sites, and these should not be considered exhaustive lists. Only sites who are health information custodians will have access to the original source, and other sites may need to work with custodians to acquire the necessary data.

Tumour: In this document, the term "tumour" is intended to broadly cover all malignancies, including hematological cancers that do not form a solid tumour (sometimes referred to as "liquid tumours".

D. Data model diagram

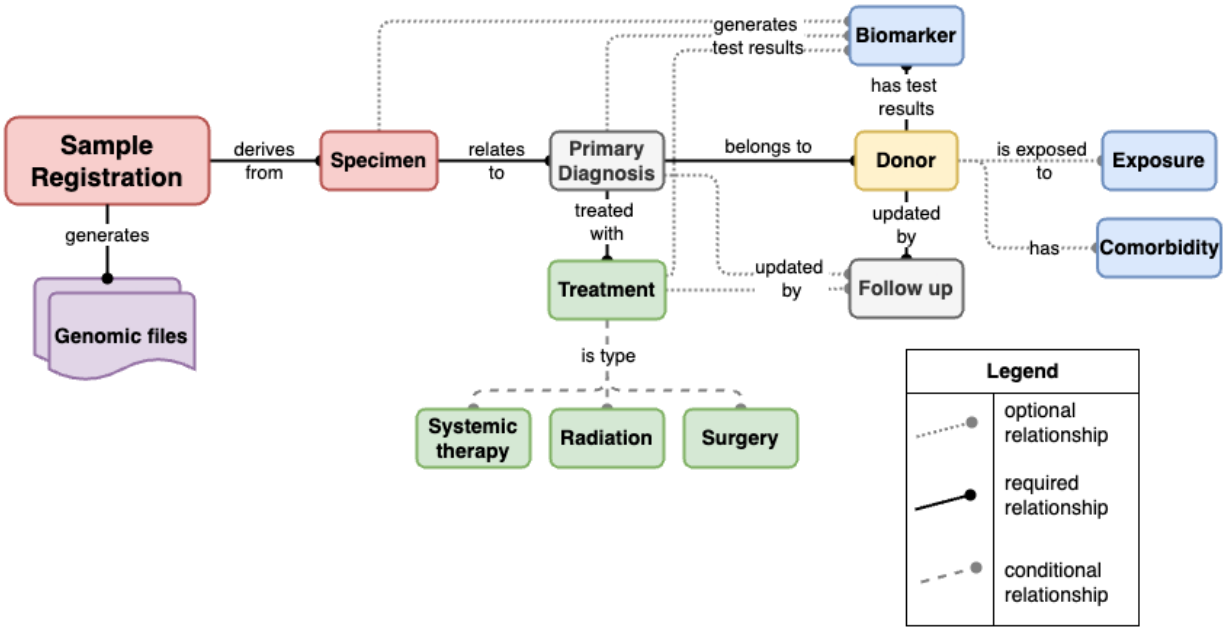


Figure 1. Diagram of the relationships between data objects in the MOHCCN clinical data model.

E. Data elements

Sample registration

Sample registration requires the data submitter to submit a basic set of data about each sample before or concomitant with submitting clinical or molecular data. This is to ensure that relationships between different entities are maintained across all molecular and clinical data submissions.

During sample registration, each Donor, Specimen and Sample entity will be assigned an ID that maps to your program's internal identifier (also referred to as the 'program_id'). Any attempts to submit molecular or clinical data that does not refer to a registered donor, specimen or sample will result in an error. The 'submitter_sample_id' field is used to link the clinical data submission to genomic and other data files generated during sequencing, and care must be taken to keep this ID unique. The 'submitter_sample_id' does not need to match the genomic (or other) data file ID, as long as the two can be linked. For example, see the following script linking a file name ('genomic_file_id'), a sample within that file ('genomic_file_sample_id'), and a sample ID submitted in the clinical data model ('submitter_sample_id'):


```
{
    "program_id": "SYNTH_02",
    "genomic_file_id": "HG00096.cnv.vcf",
    "main": {
      "access_method": "http://s3.us-east-
1.amazonaws.com/1000genomes/release/20130502/ALL.chr22.phase3_shapeit2_mvn
call_integrated_v5a.20130502.genotypes.vcf.gz?public=true",
      "name": "HG00096.cnv.vcf.gz"
    },
    "index": {
      "access_method": "http://s3.us-east-
1.amazonaws.com/1000genomes/release/20130502/ALL.chr22.phase3_shapeit2_mvn
call_integrated_v5a.20130502.genotypes.vcf.gz.tbi?public=true",
      "name": "HG00096.cnv.vcf.gz.tbi"
    },
    "metadata": {
      "sequence_type": "wgs",
      "data_type": "variant",
      "reference": "hg38"
    },
    "samples": [
      {
        "genomic_file_sample_id": "HG00096",
        "submitter_sample_id": "SAMPLE_0017"
      }
    ]
  }
```


Each sample is a single unit containing material derived from one specimen (e.g. DNA, RNA). Most donors will have at least three associated samples: DNA and RNA samples from the

tumour specimen, and a DNA sample from the normal specimen. Tier B cases may only include DNA samples. Additional samples related to the same Primary Diagnosis can be included but are not required (e.g. a protein extraction sample linked to proteomic data).

1. **program_id**

   Requirements: Required

   Description: Unique identifier of the program (cohort). Each cohort must have a unique identifier within the MOHCCN. To ensure this uniqueness, incorporate the consortium code (ACC, MOH-Q, PM2C, PR2C, BC2C) into the ID and ensure that cohort codes are not repeated within the consortium. For Pan-Canadian projects, consider using the lead site's consortium code, followed by the collaborating site's consortium code, then a site code, and finally a project code. This structure will help distinguish each cohort effectively. Spaces are not allowed in the ID.

   Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

   Source: Identified at the consortium level or the cohort level, as long as uniqueness can be ensured.

   Examples:

   - PR-SK-BRCA (PR – Prairies consortium; SK – Saskatchewan-based cohort; BRCA; breast cancer cohort)
   - MoHQ-M-1 (MoHQ- Québec consortium; M – centre; cohort 1)
   - MOHCCN-P# (MOHCCN – program; P# – number assigned to each project)

2. **submitter_donor_id**

   Requirements: Required

   Description: Unique identifier for the donor, assigned by the data provider. This ID only needs to be unique within the program; however, for ease of use, we recommend appending it to the program ID to ensure uniqueness across the entire Network. Spaces are not allowed in the ID.

   Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

   Source: Identified at the consortium level or the cohort level.

   Examples:

- PR-SK-BRCA-0001 (PR – Prairies consortium; SK – Saskatchewan-based cohort; BRCA; breast cancer cohort; 0001 - donor 0001)
  MohQ-M-1-66 (MohQ- Québec consortium; M – centre; 1 – cohort 1; 66 – patient 66)

3. submitter_specimen_id

Requirements: Required

Description: Unique identifier of the specimen from which the sample was derived, assigned by the data provider. Each sample should be linked to a single specimen. Spaces are not allowed in the ID. We recommend that the specimen ID be appended to the donor ID to ensure uniqueness.

Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

Source: Identified at the consortium or cohort level.

Examples:

- PR-SK-BRCA-0001-TP-01 (PR – Prairies consortium; SK – Saskatchewan-based cohort; BRCA; breast cancer cohort; 0001 - donor 0001; TP01 - tumour primary specimen 01)
- PR-SK-BRCA-0001-NB-01 (PR – Prairies consortium; SK – Saskatchewan-based cohort; BRCA; breast cancer cohort; 0001 - donor 0001; TP01 – normal blood specimen 01)
- MohQ-M-1-66-5731 (MohQ- Québec consortium; M – centre; 1 – cohort 1; 66 – patient 66; 5731 – specimen)

4. specimen_tissue_source

Requirements: Required

Description: Type of tissue or bodily fluid that was the source of the specimen. Choose the most specific term that matches the known information about the specimen you are curating. We recommend that standard blood collection for normal blood samples be designated as "Whole blood".

Permissible values: Controlled vocabulary field where permissible values are taken from various ontologies. See model spreadsheet for the full list.

Source: Team / biobank that provided the specimen or sample.

5. tumour_normal_designation

Requirements: Required

Description: Description of specimen's tumour/normal status for data processing.

Permissible values: Controlled vocabulary field with two permissible values, 'Normal' or 'Tumour'.

Source: Team / biobank that provided the specimen or sample.

6. specimen_type

Requirements: Required

Description: Detailed description of the type of tumour or normal tissue from which the specimen was obtained. Choose the most specific term that matches the known information about the specimen you are curating. Core Gold Cohort specimens should all fall under a 'Normal' or 'Tumour' category, but accompanying data may be derived from cell lines or xenografts. The generic 'Tumour - unknown if derived from primary or metastatic tumour' should only be used if no further information about the tumour is available.

Permissible values: Controlled vocabulary field with several values.

Source: Team / biobank that provided the specimen or sample.

7. submitter_sample_id

Requirements: Required

Description: Unique identifier of the sample, assigned by the data provider. Spaces are not allowed in the ID. We recommend that the sample ID be constructed by appending the sample type to the specimen ID as shown in the examples below.

Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1, 64} $. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

Source: Identified at the consortium or cohort level.

Examples:

- PR-SK-BRCA-0001-TP-01-DNA (PR – Prairies consortium; SK – Saskatchewan-based cohort; BRCA; breast cancer cohort; 0001 - donor 0001; TP01 - tumour primary specimen 01; DNA – DNA sample)
- MohQ-CM-1-66-5731-1DT (MohQ- Québec consortium; CM – centre [CHUM]; 1 – cohort 1; 66 – patient 66; 5731 – specimen; 1DT – tumour DNA)
- MOHCCN-P1-0001-01T-01 (MOHCCN – program; P# – assigned project #; 01T – tumour specimen; 01 – first block submitted)

8. **sample_type**

Requirements: Required

Description: The type of molecular material that was extracted from the specimen and submitted for sequencing analysis. The most specific term from the permissible values should be chosen. The value in this field determines which pipeline will be used to process the molecular data (e.g. DNA vs RNA pipeline).

Permissible values: Controlled vocabulary field with several values.

Source: Team / biobank who provided the specimen or sample. May be available from the centre that performed the sequencing if the team / biobank only provided the specimen.

Donor

The 'Donor' schema describes information about the patient donor, such as demographic information and key life events. The 'Donor' object is linked to a 'Program' as its parent through the 'program_id' element. Each donor should have at least one 'Primary Diagnosis' linked as a child object. It can optionally also be linked to 'Exposure', 'Biomarker', 'Comorbidity' and 'Follow up' objects.

9. **program_id**

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

10. **submitter_donor_id**

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

11. **gender**

Requirements: Required

Description: Self-reported gender of the donor. Gender refers to an individual's personal and social identity as a man, woman or non-binary person (a person who is not exclusively a man or a woman; see also Statistics Canada definition). A donor's gender may differ from their sex at birth, and from what is indicated on their medical record. A person's gender may change over time. You should only use a self-identified gender field to curate this data element; if none can be found, the guidelines for null values should be followed (i.e. 'Not available' should be used if this information is not expected to ever be available,

otherwise the field should be left blank – see "Null and blank values" in section B for more information).

Permissible values: Controlled vocabulary field with six permissible values, 'Man', 'Woman', 'Non-binary', 'Other', 'Prefer not to disclose', 'Not available'.

Source: Clinical notes (self-identification needs to be clearly labelled), study-administered survey

12. sex_at_birth

Requirements: Required

Description: Donor's sex assigned at birth. Can be different from a donor's gender. This field is not expected to change after first recorded at enrolment. Note that 'Other' can be used for any sex apart from 'Male' and 'Female', but should not be used for missing data.

Permissible values: Controlled vocabulary field with four permissible values, 'Male', 'Female', 'Other', or 'Not available'.

Source: Clinical notes, birth certificate. May also be obtained from genomic sequencing data through analysis of sex chromosomes.

13. is_deceased

Requirements: Required

Description: Whether or not the donor is deceased at the time of curation. **This field should be revisited whenever a follow-up curation is performed**, meaning that it can be assumed to be accurate as of the latest date associated with a clinical event ('primary_diagnosis', 'treatment', 'follow-up', etc.). If this information cannot be clearly identified in the donor's medical records as of the latest reported clinical event, the field should be left blank.

Permissible values: Controlled vocabulary field with two options, 'Yes' or 'No'.

Source: Clinical notes, death certificate.

14. date_of_birth

Requirements: Required

Description: Donor's date of birth.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Clinical notes, birth certificate

15. date_of_death

Requirements: Conditional – Only required if "is_deceased" = "Yes"

Description: Donor's date of death.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Clinical notes, death certificate

16. date_resolution

Requirements: Required

Description: Whether day or month resolution is used for submitting date fields for this donor. This applies to all date fields submitted and impacts how the dates provided will be converted to intervals for data sharing: if 'month' is indicated here, only monthly intervals will be provided (e.g. 18 months since diagnosis), and if 'day' is provided, intervals will be provided with day resolution (e.g. 232 days since diagnosis). To ensure that dates where the day is known are accurately reported, we recommend indicating 'day' unless all the dates provided only have month resolution. An arbitrary day can be added to dates where only the month is known – using '15' is preferred as it minimizes the possible margin of error to 16 days.

Permissible values: Controlled vocabulary field with two options, 'month' or 'day'.

17. cause_of_death

Requirements: Conditional – Only required if "is_deceased" = "Yes"

Description: Whether donor's cause of death was cancer. The curator may exercise discretion; for example, if complications from the cancer are the listed cause of death, "Died of cancer" should be used. "Not Available" should be used if the cause of death is unknown.

Permissible values: Controlled vocabulary field with three options: 'Died of cancer', 'Died of other reasons', or 'Not available'.

Source: Clinical notes, death certificate

18. lost_to_followup_after_clinical_event_identifier

Requirements: Optional/Conditional – This field can only be submitted if 'is_deceased' = 'No'

Description: If the donor became lost to follow up, indicate the identifier of the clinical event (e.g. 'submitter_primary_diagnosis_id', 'submitter_treatment_id' or 'submitter_follow_up_id') after which this occurred. Since each clinical event is associated with a date, this will indicate when the donor was last known to be alive. The 'is_deceased' field must be 'No' if the donor is lost to follow-up (i.e. if the donor is known to be deceased, then they cannot be lost to follow-up.

Examples:

1. Donor was lost to follow-up after their last treatment:

Donor

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event_identifier
DN_10	No	TR_01

Primary Diagnosis

submiter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
DN_10	PD_01	2017-09	C15

Treatment

submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
DN_10	TR_01	PD_01	Surgery	2017-10	2017-10

Follow Up

submitter_donor_id	submitter_follow_up_id	date_of_followup	disease_status_at_followup

Donor was lost to follow-up after their last treatment (TR_01). This means the donor was last known to be alive in October 2017.

2. Donor is found to be alive after they had been indicated as lost to follow-up:

Donor

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event_identifier
DN_10	No	FL22

Primary Diagnosis

submiter_donor_id	submitter_primary_diagnosis_id	date_of_diagnosis	cancer_type_code
DN_10	PD_01	2017-09	C15

Treatment

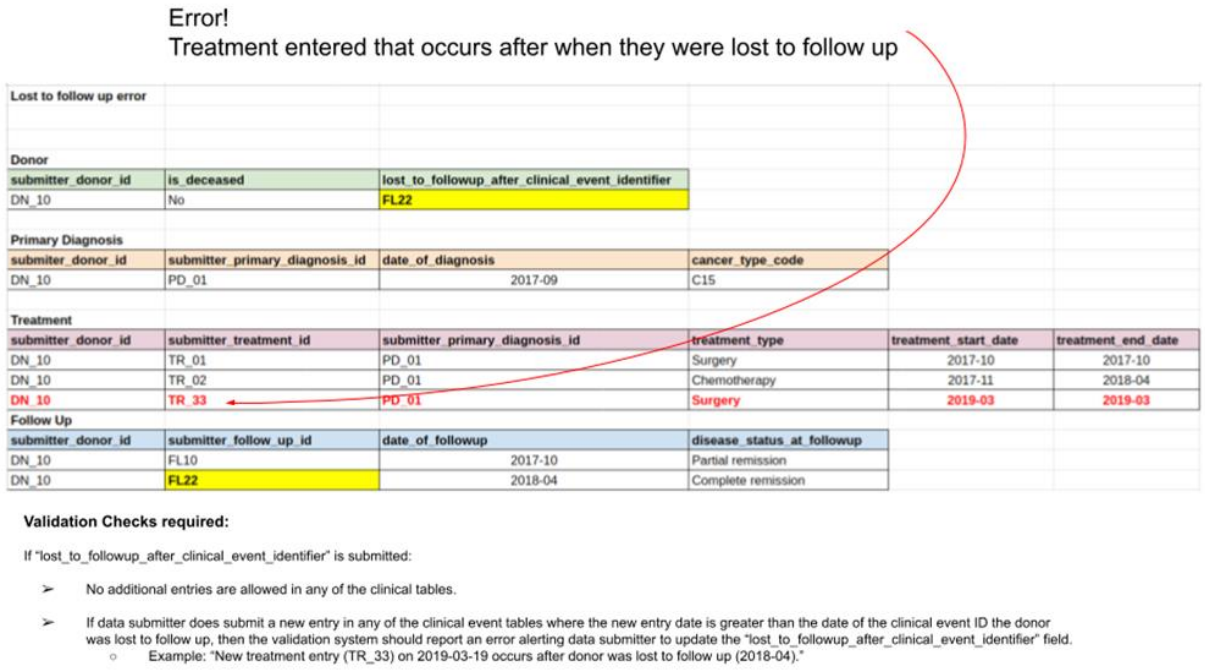
submitter_donor_id	submitter_treatment_id	submitter_primary_diagnosis_id	treatment_type	treatment_start_date	treatment_end_date
DN_10	TR_01	PD_01	Surgery	2017-10	2017-10
DN_10	TR_02	PD_01	Chemotherapy	2017-11	2018-04

Follow Up

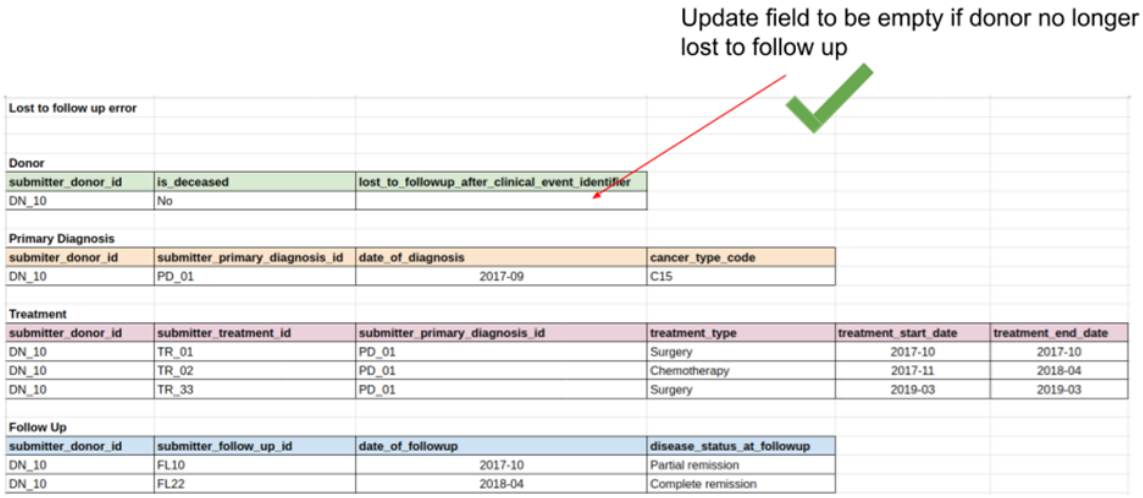
submitter_donor_id	submitter_follow_up_id	date_of_followup	disease_status_at_followup
DN_10	FL10	2017-10	Partial remission
DN_10	FL22	2018-04	Complete remission

Donor was lost to follow-up after their last follow up appointment ("FL22") in April 2018, indicated by submitting 'lost_to_followup_after_clinical_event_identifier' = "FL22".

Donor was found again a year later when they had surgery in March 2019. The surgery is added as a new row in the Treatment table ('submitter_treatment_id' = "TR33"). In the above example, since the data submitter did not update the 'lost_to_followup_after_clinical_event_identifier' field (i.e. update to be empty), the clinical validation will fail since the new treatment event has a newer date.

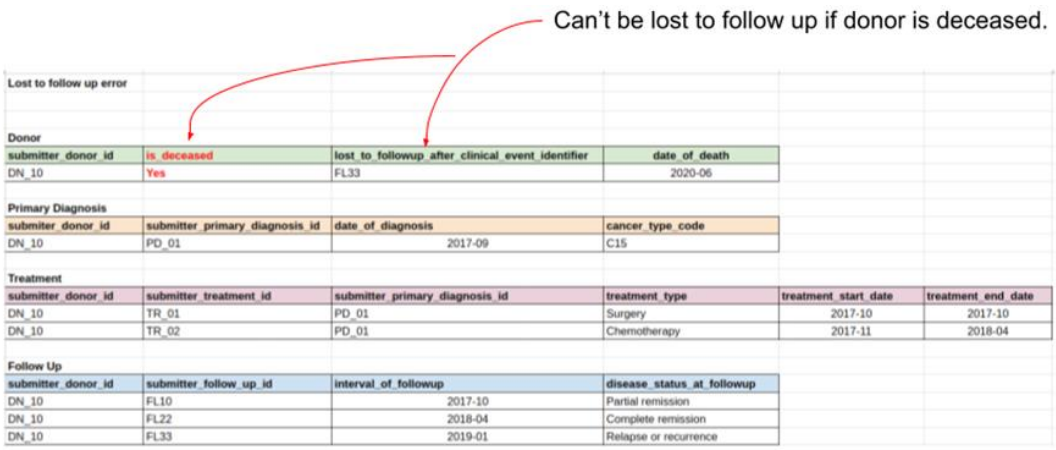


To correct the submission, the data submitter will need to update the 'lost_to_followup_after_clinical_event_identifier' field to be empty, since the donor is no longer lost to follow-up.



3. Donor was found to have died 6 months after they were lost to follow up.

If a donor is discovered to be deceased after they were lost to follow up (i.e. 'is_deceased' is updated to "Yes"), then the 'lost_to_followup_after_clinical_event_identifier' field should be updated to be empty. The date of death should also be submitted in the 'date_of_death' field in the Donor table.



Validation Checks Required:

"lost_to_followup_after_clinical_event_identifier":

- Cannot be submitted if "is_deceased" = "Yes"
- Can only be submitted if "is_deceased" is not "Yes"

19. lost_to_followup_reason

Requirements: Optional/Conditional – This field should only be submitted if 'lost_to_followup_after_clinical_event' was submitted

Description: The reason why the donor was lost to follow-up. Choose the most accurate reason from the permissible values.

Permissible values: Controlled vocabulary field with six options: 'Completed study', 'Discharged from follow-up', 'Discharged to palliative care', 'Lost contact', 'Not available', or 'Withdrew from study'

Source: Clinical notes or study records indicating discharge or study completion.

20. date_alive_after_lost_to_followup

Requirements: Optional/Conditional – If this field is submitted, then "lost_to_followup_after_clinical_event_identifier" must be submitted to indicate when the donor was lost to follow up.

Description: Date when donor was discovered to be alive after they had been indicated as lost to follow-up, but their cancer status is unknown (i.e. there is not enough information to submit a new Follow-up table). If this field is completed, the other 'lost_to_followup' fields should also be completed, that is, 'lost_to_followup_after_clinical_event_identifier' and 'lost_to_followup_reason'.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Clinical notes

Example:

Donor is found to be alive after they had been indicated as lost to follow-up, but their cancer status is unknown (i.e. denoting cancer-specific survival vs overall survival).

Donor had a follow-up appointment at the cancer clinic in August 2015 where their disease status was found to be complete remission. The donor became lost to follow up after that August 2015 follow-up appointment. Two years later, the cancer clinic found out that the donor attended an ophthalmology appointment in January 2017, but they don't have any information about the donor's cancer disease status.

The data submitter must indicate when the donor became lost to follow up in their cancer-related timeline. In this example, the donor became lost to follow up after their last appointment at the cancer clinic in August 2015, meaning that 'submitter_follow_up_id' = "SF0012".

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event	date_alive_after_lost_to_followup
Donor123	No	SF0012	

submitter_follow_up_id	submitter_donor_id	date_of_followup	disease_status_at_followup
SF0012	Donor123	2015-08	Complete remission

Clinic discovers the donor attended an ophthalmology appointment 2 years later in January 2017, but they don't have any cancer disease status information on the donor, they only know the donor is still alive as of January 2017.

submitter_donor_id	is_deceased	lost_to_followup_after_clinical_event	date_alive_after_lost_to_followup
Donor123	No	SF0012	2017-01

submitter_follow_up_id	submitter_donor_id	date_of_followup	disease_status_at_followup
SF0012	Donor123	2015-08	Complete remission

Specimen

The 'Specimen' schema describes information about the specimen from which the sample(s) was/were obtained. The specimen is linked to a Primary Diagnosis as its parent and should have at least one sample registration that derives from it. It may be linked to a Treatment if the specimen was collected as part of a treatment event, such as a Surgery. It can also be linked to a Biomarker element if clinical tests were performed on the specimen.

21. program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

22. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

23. submitter_specimen_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

24. submitter_primary_diagnosis_id

Requirements: Required

Description: Unique identifier of the primary diagnosis event related to the specimen, assigned by the data provider. Spaces are not allowed in the ID.

Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

Source: Identified at the consortium or cohort level.

25. submitter_treatment_id

Requirements: Optional – to be completed if the specimen was obtained through surgery

Description: Unique identifier of the treatment related to this specimen's acquisition, assigned by the data provider. Spaces are not allowed in the ID. If completed, this identifier should also be associated with a treatment table (see #59 for more information).

Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

Source: Identified at the consortium or cohort level.

26. specimen_collection_date

Requirements: Required

Description: Date when the specimen was collected from donor.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Team / biobank that collected the specimen.

27. specimen_anatomic_location

Requirements: Required

Description: ICD-O-3 topography code for the anatomic location from where the specimen was collected. Refer to the guidelines provided in the ICD-O-3 manual.

Permissible values: Must be an ICD-O-3 topography code and meet the regular expression ^[C][0-9]{2}(.[0-9]{1})?$ or be 'Not available'.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

* The curator may need to extrapolate the code from these sources.

28. tumour_histological_type

Requirements: Required/Conditional – This field is only required if the specimen is a tumour.

Description: Code to represent the histology (morphology) of neoplasms, according to the International Classification of Diseases for Oncology, 3rd Edition (WHO ICD-O-3). Refer to the ICD-O-3 manual for guidelines at https://apps.who.int/iris/handle/10665/42344.

Permissible values: Must be an ICD-O-3 morphology code and meet the regular expression ^[8,9]{1}[0-9]{3}/[0,1,2,3,6,9]{1}[1-9]{0,1}$ or be 'Not available'.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

* The curator may need to extrapolate the code from these sources.

29.reference_pathology_confirmed_diagnosis

Requirements: Required/Conditional – This field is only required if the specimen is a tumour.

Description: Whether the pathological diagnosis was confirmed by a pathologist. 'No' should be used to indicate cases where a review was done but the diagnosis was not confirmed, and 'Not done' should be used to indicate cases where a review was not performed.

Permissible values: Controlled vocabulary field with four values: 'Yes', 'No', 'Not done', or 'Not available'.

Source: Team / biobank that collected the specimen.

30.reference_pathology_confirmed_tumour_presence

Requirements: Required/Conditional – This field is only required if the specimen is a tumour.

Description: Whether a pathologist confirmed the presence of tumour in the specimen. 'No' should be used to indicate cases where a review was done and tumour presence was not confirmed, and 'Not done' should be used to indicate cases where a review was not performed.

Permissible values: Controlled vocabulary field with four values: 'Yes', 'No', 'Not done', or 'Not available'.

Source: Team / biobank that collected the specimen.

31.tumour_grading_system

Requirements: Required/Conditional – This field is only required if the specimen is from a tumour.

Description: Tumour grading system used to evaluate tumour histology.

Permissible values: Controlled vocabulary field with multiple values consistent with ICGC-ARGO, plus 'Not available'.

Source: Pathology report(s)

32.tumour_grade

Requirements: Required/Conditional – This field depends on the selected 'tumour_grading_system', and is only required if the specimen is from a tumour.

Description: Grade of the tumour as assigned by the reporting 'tumour_grading_system'. This field will be cross-validated against the 'tumour_grading_system' field. Refer to ICGC ARGO's Tumour Grading Classifications documentation to ensure the correct tumour grade term is submitted.

Permissible values: Controlled vocabulary field with multiple values consistent with ICGC-ARGO, plus 'Not available'.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

33.percent_tumour_cells_range

Requirements: Required/Conditional – This field is only required if the specimen is from a tumour.

Description: Range representing the percent of tumour cells in a specimen. This can be obtained from a different sample of the specimen to the one submitted for sequencing (e.g. if 'percent_tumour_cells_measurement_method' is not 'Genomics').

Permissible values: Controlled vocabulary field with four values: '0-19%', '20-50%', '51-100%', or 'Not available'.

Source: Team / biobank that provided the specimen, pathology report (may be reported as "tumour cellularity")

34.percent_tumour_cells_measurement_method

Requirements: Required/Conditional – This field is only required if the specimen is from a tumour.

Description: Method used to measure 'percent_tumour_cells'.

Permissible values: Controlled vocabulary field with five values: 'Genomics', 'Image analysis', 'Pathology estimate by percent nuclei', 'Other', or 'Not available'.

Source: Team / biobank that provided the specimen, pathology report

35.specimen_storage

Requirements: Optional

Description: Method of long-term storage for the specimen if the sample was not extracted freshly. Use 'Not Applicable' if the specimen/sample was not stored (i.e. extracted freshly, e.g. from blood), and leave blank or use 'Not Available' if storage conditions cannot be confirmed (see Section B for more details).

Permissible values: Controlled vocabulary field with multiple values.

Source: Team / biobank that collected the specimen and/or processed the sample.

36.specimen_processing

Requirements: Optional

Description: Technique used to process the specimen for storage. If the sample was directly extracted from fresh tissue, then select "Fresh".

Permissible values: Controlled vocabulary field with multiple values.

Source: Team / biobank that collected the specimen and/or processed the sample.

37.specimen_laterality

Requirements: Optional

Description: For cancer in a paired organ, the side on which the specimen was obtained. 'Not applicable' should be used if the cancer is not in a paired organ. If the specimen was obtained from a brain tumour, this field should be used to indicate the hemisphere(s) from which the specimen was obtained.

Permissible values: Controlled vocabulary field with four values: 'Left', 'Right', 'Not applicable', or 'Not available'.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

Primary diagnosis

The 'Primary Diagnosis' schema describes information about the primary diagnosis related to the specimen. If the specimen was obtained from a recurrent tumour or metastasis, the information here should be obtained from the related primary diagnosis. Any other cancer diagnoses should be indicated in the 'Comorbidity' schema.

The Primary Diagnosis is linked to a Donor as its parent. The Primary Diagnosis should have at least one Specimen and one Treatment linked as children. It may also be linked to a Biomarker element if clinical tests were performed as part of the diagnosis. If there was a follow-up to the Primary Diagnosis, it can be linked as a parent to a Follow up data element.

38.program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

39.submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

40.submitter_primary_diagnosis_id

Unique identifier of the primary diagnosis, assigned by the data provider. See #25 for more information.

41.date_of_diagnosis

Requirements: Required

Description: Date that the donor was first diagnosed with the primary cancer associated with the index specimen. This date should be based on the *earliest* diagnosis (e.g. a donor can be given neoadjuvant treatment before the definitive diagnosis, so the date of earliest diagnosis should be submitted). As this date is used as an index for calculating all date intervals in a patient's dataset, it is critical that this field be completed accurately.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

42.cancer_type_code

Requirements: Required

Description: WHO ICD-10 code used to represent the cancer type associated with the primary diagnosis. This should be the most accurate/definitive diagnosis available, not necessarily the initial one (i.e. if the diagnosis was refined after initial diagnosis).

Permissible values: ICD-10 code (see https://icd.who.int/browse10/2019/en)

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

* The curator may need to extrapolate the code from these sources.

43. primary_site

Requirements: Required

Description: Primary site of the cancer related to the index specimen. If the index specimen is from a metastasis, this field should indicate the primary site of the tumour. If the donor has multiple primary cancer diagnoses, only the primary site of the cancer related to the index specimen should be included in this field.

Permissible values: Controlled vocabulary field where sites are taken from the WHO ICD-O. See the 'primary_site' tab of the MoH Clinical Data Model spreadsheet for the full list of options. Only one primary site can be specified per Primary Diagnosis.

Source: Clinical notes, diagnostic notes

44. basis_of_diagnosis

Requirements: Required

Description: Most valid basis used to identify the primary diagnosis.

Permissible values: Controlled vocabulary field with multiple values ordered in hierarchical order according to IACR guidelines, with the term at the top of the list ("Histology of primary tumour") being the most valid basis. If more than one diagnosis technique was used, select the most valid basis.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

45. laterality

Requirements: Optional

Description: For cancer in a paired organ, the side of the body on which the primary tumour or cancer first developed at the time of primary diagnosis. 'Not a paired site' should be used if the cancer is not in a paired organ. For brain tumours, this field should be used to indicate whether the hemisphere(s) in which the tumour is located.

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging report(s)
3. Clinical notes

46. clinical_tumour_staging_system

Requirements: Required/Conditional - This field is only required if the specimen is a tumour and if pathological staging is not submitted. If both pathological and clinical staging are available, both should be submitted.

Description: Tumour staging system used to stage the cancer at the time of primary diagnosis (prior to treatment).

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging report(s)
3. Clinical notes

47. clinical_t_category

Requirements: Required/Conditional - This field is required only if 'clinical_tumour_staging_system' is submitted and the selected value 'AJCC cancer staging system'

Description: Code representing the extent of the primary tumour (T) based on evidence obtained from clinical assessment parameters determined at the time of primary diagnosis (prior to treatment), according to criteria from the AJCC's Cancer Staging Manual (any edition).

Permissible values: Controlled vocabulary field with AJCC T category values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging report(s)
3. Clinical notes

48.clinical_n_category

Requirements: Required/Conditional - This field is required only if 'clinical_tumour_staging_system' is submitted and the selected value is 'AJCC cancer staging system'

Description: Code representing the extent of regional lymph node (N) involvement for the cancer based on evidence obtained from clinical assessment parameters determined at the time of primary diagnosis (prior to treatment), according to criteria from the AJCC's Cancer Staging Manual (any edition).

Permissible values: Controlled vocabulary field with AJCC N category values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging report(s)
3. Clinical notes

49.clinical_m_category

Requirements: Required/Conditional - This field is required only if 'clinical_tumour_staging_system' is submitted and the selected value is 'AJCC cancer staging system'

Description: Code representing the extent of distant metastasis (M) for the cancer based on evidence obtained from clinical assessment parameters determined at the time of primary diagnosis (prior to treatment), according to criteria from the AJCC's Cancer Staging Manual (any edition).

Permissible values: Controlled vocabulary field with AJCC M category values. MX is not a valid category and cannot be assigned.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging report(s)
3. Clinical notes

50.clinical_stage_group

Requirements: Required/Conditional - This field is required only if 'clinical_tumour_staging_system' is submitted and is dependent on the selected value. Refer to the ICGC ARGO documentation for Tumour Staging Classifications.

Description: Stage group of the tumour that indicates the overall prognostic tumour stage (ie. Stage I, Stage II, Stage III etc.), based on the selected 'clinical_tumour_staging_system'.

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging report(s)
3. Clinical notes

51. pathological_tumour_staging_system

Requirements: Required/Conditional – This field is only required if the specimen is a tumour and if clinical staging is not submitted. If both pathological and clinical staging are available, both should be submitted.

Description: Tumour staging system used to assess the cancer associated with the primary diagnosis. Pathological classification is based on the clinical stage information (acquired before treatment) and supplemented/modified by operative findings and pathological evaluation of the resected specimen. Pathological classification is considered the gold standard for staging, and should be submitted if available. This element refers to the cancer associated with the primary diagnosis and therefore is not necessarily tied to the specimen; for example, if the specimen is from a recurrent cancer, this element should be used to indicate the tumour staging system used to assess the primary cancer tied to the recurrence. Similarly, if the specimen is from a biopsy and pathological staging was performed at a later date, this element should be used to indicate the tumour staging system used.

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

52. pathological_t_category

Requirements: Required/Conditional – This field is required only if 'pathological_tumour_staging_system' is submitted and the selected value is 'AJCC cancer staging system'

Description: Code to represent the stage of cancer defined by the size or contiguous extension of the primary tumour (T), according to criteria from the AJCC's Cancer Staging Manual (any edition). See 'pathological_tumour_staging_system' description for further guidelines.

Permissible values: Controlled vocabulary field with AJCC T category values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

53.pathological_n_category

Requirements: Required/Conditional – This field is required only if 'pathological_tumour_staging_system' is submitted and the selected value is 'AJCC cancer staging system'

Description: Code to represent the stage of cancer defined by whether or not the cancer has reached nearby lymph nodes (N), according to criteria from the AJCC's Cancer Staging Manual (any edition). See 'pathological_tumour_staging_system' description for further guidelines.

Permissible values: Controlled vocabulary field with AJCC N category values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

54.pathological_m_category

Requirements: Required/Conditional – This field is required only if 'pathological_tumour_staging_system' is submitted and the selected value is 'AJCC cancer staging system'.

Description: Code to represent the stage of cancer defined by whether there are distant metastases (M), meaning spread of cancer to other parts of the body, according to criteria from the AJCC's Cancer Staging Manual (any edition). See 'pathological_tumour_staging_system' description for further guidelines.

Permissible values: Controlled vocabulary field with AJCC M category values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

55.pathological_stage_group

Requirements: Required/Conditional – This field is only required if the specimen is a tumour and if clinical staging was not submitted.

Description: Tumour stage assessed at time the tumour was surgically resected, based on the selected 'pathological_tumour_staging_system'.

Permissible values: Controlled vocabulary field with multiple values.

Treatment

The 'Treatment' schema describes one or more treatments administered to the donor in relation to the index cancer. Each Treatment table is linked to a Primary Diagnosis table as its parent. If the 'treatment_type' field contains one or more of the treatment types represented in downstream schema (i.e. 'Surgery', 'Radiation', and/or 'Systemic therapy'), the relevant schema(s) should be linked as (a) child element(s).

Multiple treatment modalities can only be included in the same Treatment schema if they are part of the same treatment regimen and share the same start and end dates (e.g. if a systemic therapy and targeted molecular therapy are administered at the same time); otherwise, separate Treatment tables must be generated for each treatment modality.

For a treatment regimen that includes multiple systemic therapies, a single Treatment table should be completed with the overarching start and end dates and linked to individual Systemic therapy tables each containing the relevant drug information (e.g. if drugs A and B are started at the same time but A is discontinued six months later than B, a single Treatment table with 'treatment_type' = 'Systemic therapy' should be completed and linked to two child Systemic therapy tables, one describing drug A and the other drug B with their relevant dates).

56. program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

57. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

58. submitter_treatment_id

Requirements: Required

Description: Unique identifier of the treatment, assigned by the data provider. Spaces are not allowed in the ID.

Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

Source: Identified at the consortium or cohort level.

59. submitter_primary_diagnosis_id

Requirements: Required

Description: Unique identifier of the primary diagnosis event this treatment is linked to. Spaces are not allowed in the ID.

Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

Source: Identified at the consortium or cohort level.

60. treatment_type

Requirements: Required

Description: Type of treatment regimen that the donor completed. This field accepts multiple values; if a treatment regimen consists of two or more treatments given at the same time, these should be submitted together, separated by a pipe delimiter ("|") within your file. Treatments administered at different times should be added as separate treatment tables. For each treatment indicated in this field, one or more additional supplementary table(s) need to be completed, and clinical validation will throw a warning if it (they) is (are) missing; for example, if 'treatment_type' includes 'Systemic therapy', at least one supplementary 'Systemic therapy' table is required.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes (the curator may need to extrapolate the treatment type from e.g. a drug name)

61. is_primary_treatment

Requirements: Required

Description: Whether the treatment was the primary treatment following the initial diagnosis, i.e. the first treatment given for the disease. See NCI definition for further

details. This field can be 'Yes' for more than one treatment table; for example, if the primary treatment consists of neoadjuvant chemotherapy and surgery, these should be submitted separately since they have different start and end dates but 'is_primary_treatment' should be 'Yes' for both.

Permissible values: Controlled vocabulary field with three values: 'Yes', 'No', or 'Not available'.

Source: Clinical notes

62. treatment_start_date

Requirements: Required

Description: Date the treatment was initiated.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Clinical notes

63. treatment_end_date

Requirements: Required

Description: Date the treatment ended.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information). If the treatment is ongoing, this field should be left blank.

Source: Clinical notes

64. treatment_intent

Requirements: Required

Description: Intended disease outcome for which the treatment was given.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes

65. response_to_treatment_criteria_method

Requirements: Optional

Description: Criteria used to assess the donor's overall response to the applied treatment regimen.

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Clinical notes
2. Imaging report(s)

66.response_to_treatment

Requirements: Optional

Description: Donor's response to the applied treatment regimen, as assigned by the reported 'response_to_treatment_criteria_method'. Curators can populate this field to the best of their ability, but we recommend that it be reviewed by the cohort lead and/or affiliated clinican.

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

3. Clinical notes
4. Imaging report(s)

67.status_of_treatment

Requirements: Optional

Description: Status of the prescribed treatment at time of curation.

Permissible values: Controlled vocabulary field with multiple values

Source: Clinical notes

Systemic therapy

The 'Systemic therapy' schema describes systemic therapies administered to the donor in relation to the index cancer. As a supplementary table, it is only required if the 'treatment_type' field in the parent Treatment table includes 'Systemic therapy'. Separate Systemic therapy tables should be completed for each individual systemic therapy administered (e.g. if a treatment course includes two or more drugs, separate tables should be completed for each drug).

68.program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

69. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

70. submitter_treatment_id

Unique identifier of the treatment this systemic therapy was a part of. See #59 for further information. The curator should ensure that the linked treatment has 'Systemic therapy' as one of the values in the 'treatment_type' field.

71. systemic_therapy_type

Requirements: Required

Description: Type of systemic therapy administered to donor.

Permissible values: Controlled vocabulary field with three values: 'Chemotherapy', 'Hormone therapy', or 'Immunotherapy'

Source: Clinical notes

72. start_date

Requirements: Required

Description: Date that the selected systemic therapy type was initiated.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Clinical notes

73. end_date

Requirements: Required

Description: Date that the selected systemic therapy ended.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Clinical notes

74. drug_reference_database

Requirements: Required

Description: Chosen drug reference database where drug name is found. Must be one of the three options provided, or leave blank if the drug is not found in any of the databases (as may be the case for experimental drugs).

Permissible values: Controlled vocabulary field with four values: 'RxNorm', 'PubChem', 'NCI Thesaurus', or 'Not available'

Source: Clinical notes (curator may need to choose the database to reference the drug described in the record)

75. drug_reference_identifier

Requirements: Required

Description: Unique identifier assigned to the treatment regimen drug by the 'drug_reference_database'. If the drug is not found in any of the databases (as may be the case for experimental drugs), the field should be left blank.

Permissible values: Text field. Identifier should be identical to what is found in the 'drug_reference_database' (e.g. capitalization).

Source: Clinical notes. Curator may need to choose identifier based on name/information provided in the medical record using the following approach:

1. Navigate to the browser of the selected reference database ('drug_reference_database') browser and search for the name or description of the surgery.
2. Review the list of results to find the specific term that matches the type of drug administered.
3. Record the unique code or identifier listed.

76. drug_name

Requirements: Required

Description: Name of agent or drug administered to patient as part of the treatment regimen, as it exists in the 'drug_reference_database' and linked to the 'drug_reference_identifier'. In cases where the drug is not listed in any public databases (i.e. 'drug_reference_database' and 'drug_reference_identifier' are blank), which may be the case with experimental drugs, this field should still be completed to record the drug's name. If the drug's name is not known (e.g. the donor is participating in a clinical trial and results have not yet been released), this field should be left blank.

Permissible values: Text field. Identifier should be identical to what is found in the 'drug_reference_database' (e.g. capitalization)

Source: Clinical notes (curator may need to extrapolate following steps outlined under 'drug_reference_identifier').

77. drug_dose_units

Requirements: Required if cumulative drug dose is submitted.

Description: Units used to record drug dose.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes

78. days_per_cycle

Requirements: Optional

Description: Number of days in a treatment cycle.

Permissible values: Integer field.

Source: Clinical notes

79. number_of_cycles

Requirements: Optional

Description: Number of treatment cycles.

Permissible values: Integer field.

Source: Clinical notes

80. prescribed_cumulative_drug_dose

Requirements: Optional

Description: Total prescribed cumulative drug dose, in the same units specified in 'chemotherapy_drug_dose_units'.

Permissible values: Number field.

Source: Clinical notes

81. actual_cumulative_drug_dose

Requirements: Optional

Description: Total actual cumulative drug dose, in the same units specified in 'chemotherapy_drug_dose_units'.

Permissible values: Number field.

Source: Clinical notes

Radiation

The 'Radiation' schema describes radiation therapy courses administered to the donor in relation to the index cancer. As a supplementary table, it is only required if the 'treatment_type' field in the parent Treatment table includes 'Radiation'. If multiple radiation therapy courses were administered, separate Radiation tables should be completed for each course (e.g. if external radiation therapy and brachytherapy are both performed, separate tables should be completed for each course).

82. program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

83. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

84. submitter_treatment_id

Unique identifier of the treatment this radiation was a part of. See #5 9 for further information. The curator should ensure that the linked treatment has 'Radiation therapy' as one of the values in the 'treatment_type' field.

85.radiation_therapy_modality

Requirements: Required

Description: Method of radiation treatment or modality.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes

86.radiation_therapy_type

Requirements: Required

Description: Type of radiation therapy administered.

Permissible values: Controlled vocabulary field with three values: 'External', 'Internal', or 'Not available'.

Source: Clinical notes

87.radiation_therapy_fractions

Requirements: Required

Description: Total number of fractions delivered as part of treatment.

Permissible values: Integer field.

Source: Clinical notes

88.radiation_therapy_dosage

Requirements: Required

Description: Total dose given in units of Gray (Gy).

Permissible values: Integer field.

Source: Clinical notes

89.anatomical_site_irradiated

Requirements: Required

Description: Anatomical site where radiation therapy was administered.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes

90.radiation_boost

Requirements: Optional

Description: Whether this radiation therapy was a radiation boost, i.e. an extra radiation treatment targeted at the tumor bed given after the regular sessions of radiation are complete.

Permissible values: Controlled vocabulary field with three values: 'Yes', 'No', or 'Not available'.

Source: Clinical notes

91.reference_radiation_treatment_id

Requirements: Optional/Conditional - Only required if 'radiation_boost' = 'Yes'

Description: If a radiation boost was given, the 'submitter_treatment_id' of the primary radiation treatment to which the radiation boost treatment is linked.

Permissible values: Text field. Needs to match an existing 'submitter_treatment_id' linked to radiation therapy.

Surgery

The 'Surgery' schema describes surgeries administered to the donor in relation to the index cancer. As a supplementary table, it is only required if the 'treatment_type' field in the parent Treatment table includes 'Surgery'. Separate Surgery tables should be completed for each individual surgery performed.

92.program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

93.submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

94.submitter_treatment_id

Unique identifier of the treatment this surgery was a part of. See #59 for further information. The curator should ensure that the linked treatment has 'Surgery' as one of the values in the 'treatment_type' field.

95.surgery_reference_database

Requirements: Required

Description: Chosen reference surgery database where surgery type is found.

Permissible values: Controlled vocabulary field with three values: 'SNOMED', 'NCIt', 'UMLS', or 'CCI'.

Source: Surgical report (curator may need to choose the database to reference the surgery type described in the record)

96.surgery_reference_identifier

Requirements: Required

Description: Unique identifier assigned to the surgery procedure by the 'surgery_reference_database'.

Permissible values: Text field. Identifier should be identical to what is found in the 'surgery_reference_database' (e.g. capitalization).

Source: Surgical report, clinical notes. Curator may need to find identifier based on name/information provided following these steps:

1. Navigate to the browser of the selected reference database ('surgery_reference_database') and search for the name or description of the surgery.
2. Review the list of results to find the specific term that matches the type of surgery done.
3. Record the unique code or identifier listed, along with the term.

97.surgery_type

Requirements: Required

Description: Type of surgical procedure that was performed, as it exists in the 'surgery_reference_database'.

Permissible values: Text field. Identifier should be identical to what is found in the 'drug_reference_database' (e.g. capitalization).

Source: Surgical report (name may be formatted differently than in the database, curator will need to confirm)

98. surgery_site

Requirements: Required/Conditional - This field is not required if a specimen was resected during surgery since the 'specimen_type' is collected in the Specimen table.

Description: Code for the anatomic site where the surgical procedure was performed, according to the International Classification of Diseases for Oncology, 3rd Edition (WHO ICD-O-3).

Permissible values: Text field. Must be an ICD-O-3 code.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Diagnosis record

* The curator may need to extrapolate the code from these sources.

99. surgery_location

Requirements: Required/Conditional - This field is not required if a specimen was resected during surgery since 'specimen_type' is collected in the Specimen table.

Description: Whether the surgical procedure was done at the primary, local recurrence or metastatic location.

Permissible values: Controlled vocabulary field with three values: 'Local recurrence', 'Metastatic', or 'Primary'.

Source: Surgical report

100. tumour_length

Requirements: Optional

Description: Length of the tumour, in millimetres (mm).

Permissible values: Number field.

Source: Surgical pathology report (gross description section)

101. tumour_width

Requirements: Optional

Description: Width of the tumour, in millimetres (mm).

Permissible values: Number field.

Source: Surgical pathology report (gross description section)

102. greatest_dimension_tumour

Requirements: Optional

Description: Measurement of the greatest dimension (length, width, or depth) of the tumour, in millimetres (mm).

Permissible values: Number field.

Source: Surgical pathology report

103. tumour_focality

Requirements: Optional

Description: Whether the tumour is unifocal or multifocal.

Permissible values: Controlled vocabulary field with five values: 'Unifocal', 'Multifocal', 'Not applicable', 'Cannot be assessed', or 'Not available'.

Source: Surgical pathology report (usually provided in gross description section and confirmed by microscopic examination in the findings section)

104. residual_tumour_classification

Requirements: Optional

Description: Whether there is residual tumour at the margins of the resected tissue following surgery. R0 indicates no residual tumor, R1 indicates microscopic residual tumor, R2 indicates macroscopic residual tumor, and RX indicates that an assessment of residual tumour was not possible. (Reference: AJCC 8th ed.)

Permissible values: Controlled vocabulary field with six values: 'RX', 'R0', 'R1', 'R2', 'Not applicable', or 'Not available'.

Source: Surgical pathology report (microscopic findings and final diagnosis sections)

105. margin_types_involved

Requirements: Optional

Description: Margin type(s) involved in the surgery. Multiple values can be submitted, separated by a pipe delimiter ('|') within your file.

Permissible values: Controlled vocabulary field with multiple values.

Source: Surgical pathology report (microscopic examination section)

106. margin_types_not_involved

Requirements: Optional

Description: Margin type(s) not involved in the surgery. Multiple values can be submitted, separated by a pipe delimiter ('|') within your file.

Permissible values: Controlled vocabulary field with multiple values.

Source: Surgical pathology report (microscopic examination section)

107. margin_types_not_assessed

Requirements: Optional

Description: Margin type(s) that could not be assessed. Multiple values can be submitted, separated by a pipe delimiter ('|') within your file.

Permissible values: Controlled vocabulary field with multiple values.

Source: Surgical pathology report (microscopic examination section)

108. lymphovascular_invasion

Requirements: Optional

Description: Whether the tumour has spread in the lymphatic system or blood vessels, as assessed by pathology examination following surgery. (Reference: AJCC 8th ed.)

Permissible values: Controlled vocabulary field with multiple values.

Source: Surgical pathology report (microscopic examination section)

109. perineural_invasion

Requirements: Optional

Description: Whether the tumour has spread along and infiltrated nerve fibers, as assessed by pathology examination following surgery.

Permissible values: Controlled vocabulary field with multiple values.

Source: Surgical pathology report (microscopic examination section)

Follow-up

The follow-up schema is intended to provide information about events following the donor patient's initial diagnosis related to the index specimen. Follow-ups are only meant to include events related to the index specimen; for example, a recurrence or progression should be captured, but a separate secondary malignancy would not.

For prospective cases, the follow-up table should be filled at yearly intervals following the initial curation event. The timing of this yearly interval will be chosen at the site or consortium level.

For retrospective cases, the initial curation event should include all relevant information up to completion of the primary treatment; events after this date should be considered follow-ups. The curator should canvas all relevant clinical notes (imaging reports, clinical notes, lab reports, etc.) and look for oncological events (i.e. anything other than stable or no evidence of disease). Each oncological event (e.g. relapse) should be captured in a follow-up table. If the donor is stable or disease-free, then the yearly follow-up table should be used to indicate that.

Examples where the yearly interval is chosen as one calendar year:

Retrospective case example:

- A patient was diagnosed in April of 2018. Their primary treatment ended in March of 2019 and the curation is being performed in August of 2024.
- An initial curation event is performed that includes information about the donor, primary diagnosis, primary treatment, etc. with relevant dates.
- If the donor has been disease-free since completion of their primary treatment, a follow-up table is filled for each calendar year starting with the one during which the treatment ended (i.e. 2019 in this example) indicating disease-free status (i.e. December 31 2019, December 31 2020, and so on).
- If the donor relapsed in September 2021, a follow-up table is completed describing the event, with the actual date of the event. An end-of-year follow-up table is also completed for 2021.

Prospective case example:

- A patient was diagnosed in June of 2024 and the initial curation is being performed in August of 2024. The initial curation event includes all available information to date.
- Going forward, a curator reviews the donor's medical record at a chosen yearly interval. For example, if yearly curation is performed in February, the next curation

event would be performed in February 2025 for the calendar year 2024. If the patient was stable on treatment through to the end of 2024, a follow-up table should be completed indicating this.
- If the patient progressed on treatment in 2024, this should be indicated in the 'response_to_treatment' element and a follow-up table should also be completed.

The yearly interval of curation and the yearly curation event should be relatively close (1-2 months) to provide up-to-date information during the yearly curation events.

The following elements from the Donor and Treatment tables should also be reviewed and updated during annual follow-ups:

- 'is_deceased' and 'lost_to_followup' elements (Donor schema)
- 'treatment_end_date', 'response_to_treatment', 'status_of_treatment', and other relevant end dates and treatment details if treatment was ongoing during the last curation event

A Follow-up element is always linked to a specific Donor as its parent, and may also be linked to **either** a Primary Diagnosis or a Treatment if the follow-up was specific to one of these events.

110. program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

111. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

112. submitter_follow_up_id

Requirements: Required

Description: Unique identifier of a follow-up event in a donor's clinical record, assigned by the data provider. Spaces are not allowed in the ID.

Permissible values: Must meet the regular expression ^[A-Za-z0-9\-\._]{1,64}$. This means that it can be up to 64 characters and composed of letters, numbers or the symbols '-', '.', '_'.

Source: Identified at the consortium level or the cohort level.

113. submitter_primary_diagnosis_id

Unique identifier of the primary diagnosis this follow-up relates to. Can only be filled out if 'submitter_treatment_id' is blank in this schema. See #24 for more information.

114. submitter_treatment_id

Unique identifier of the treatment this follow-up relates to. Can only be filled out if 'submitter_primary_diagnosis_id' is blank in this schema, and should only be populated if the follow-up event directly relates to a treatment. See #58 for more information.

115. date_of_followup

Requirements: Required

Description: Date of the clinical follow-up event related to the index cancer.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging or lab report(s)
3. Clinical notes

116. disease_status_at_followup

Requirements: Required

Description: Donor's disease status at time of follow-up. This should only relate to the index cancer and the most recent treatment received.

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Imaging or lab report(s)
2. Clinical notes

117. relapse_type

Requirements: Required/Conditional – This field is required if 'disease_status_at_followup' indicates a state of progression, relapse, or recurrence.

Description: Type of relapse.

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Imaging or lab report(s)
2. Clinical notes

118. date_of_relapse

Requirements: Required/Conditional – This field is required if 'disease_status_at_followup' indicates a state of progression, relapse, or recurrence.

Description: Date of donor's relapse.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

119. method_of_progression_status

Requirements: Required/Conditional – This field is required if 'disease_status_at_followup' indicates a state of progression, relapse or recurrence.

Description: Method(s) used to confirm the donor's progression, relapse, or recurrence disease status (e.g. biopsy, imaging, blood work). Multiple values can be submitted, separated by a pipe delimiter ('|') within your file. Permissible values are defined as follows:

- Imaging: If any method that uses a visual display of structural or functional patterns of organs or tissues was used for diagnostic evaluation.
- Histopathology test: If any histopathologic test (i.e. the microscopic study of characteristic tissue abnormalities, with the use of various cytochemical and immunocytochemical stains) was used to confirm progression, relapse or recurrence.
- Assessment of symptom control: If the treating oncologist deemed that the index cancer progressed, relapsed or recurred based on an evaluation of the patient's symptoms.
- Physical examination procedure: If the treating oncologist deemed that the index cancer progressed, relapsed or recurred based on a physical examination of the patient.
- Tumour marker measurement: If a test was performed measuring a marker of tumour activity (e.g. biomarker blood test, circulating tumour DNA assay). Note that most tumour marker measurement tests will fall under "laboratory data interpretation", and both values can be submitted if appropriate.

- Laboratory data interpretation: If a laboratory test was performed and its data used to determine progression, relapse or recurrence (e.g. bodily fluid tests).

Permissible values: Controlled vocabulary field with multiple values.

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Imaging or lab report(s)
3. Clinical notes

120. anatomic_site_progression_or_recurrence

Requirements: Required/Conditional – This field is required if 'disease_status_at_followup' indicates a state of progression, relapse, or recurrence. This field is not required if 'relapse_type' = 'Biochemical progression'

Description: ICD-O-3 topography code(s) for the anatomic site where disease progression, relapse, or recurrence occurred. Refer to the guidelines provided in the ICD-O-3 manual. Multiple values can be submitted, separated by a pipe delimiter ('|') within your file.

Permissible values: Controlled vocabulary field where the allowed values are ICD-O-3 topography codes.

Source (in order of reliability, if multiple available):

1. Imaging report(s)
2. Clinical notes

* The curator may need to extrapolate the code from these sources.

Biomarker

The Biomarker schema represents clinical tests that were performed on the donor patient. It can be linked to one other element, depending on what clinical event the tests were linked to. It can be linked to either a Specimen, Primary Diagnosis, Treatment or Follow-up. A separate Biomarker table should be completed for each individual test.

121. program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

122. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

Each row should include one or more biomarker test(s) associated with a particular clinical event, so only one of 'submitter_specimen_id', 'submitter_primary_diagnosis_id', 'submitter_treatment_id' or 'submitter_follow_up_id' fields is required. If the biomarker test is not associated with a particular clinical event, then indicate the date when the biomarker test was performed ('test_date' field).

123. submitter_specimen_id

Unique identifier of the specimen from which the sample was derived, assigned by the data provider. See #3 for more information.

Requirements: Required if the biomarker test is associated with the specimen acquisition event.

124. submitter_primary_diagnosis_id

Unique identifier of the primary diagnosis event related to the specimen, assigned by the data provider. See #25 for more information.

Requirements: Required if the biomarker test is associated with the primary diagnosis event.

125. submitter_treatment_id

Unique identifier of the treatment, assigned to the data provider. See #59 for more information.

Requirements: Required if the biomarker test is associated with a treatment event.

126. submitter_follow_up_id

Unique identifier of a follow-up event in the donor's clinical record, assigned by the data provider. See #113 for more information.

Requirements: Required if the biomarker test is associated with a follow-up event.

127. test_date

Requirements: Required if the biomarker test is not associated with a particular clinical event.

Description: If the biomarker test was not associated with a specific specimen, primary diagnosis, treatment or follow-up event, then indicate the date on which the biomarker test was performed.

Permissible values: YYYY-MM-DD format. If only the month is known, add an arbitrary DD date (using '15' is preferred as it minimizes the possible margin of error to 16 days – see 'date_resolution' for more information).

Source: Clinical notes

128. psa_level

Requirements: Optional

Description: Amount of prostate-specific antigen (PSA) present in a blood serum sample, in ng/mL, as measured by a laboratory test.

Permissible values: Integer field.

Source: Clinical notes

129. ca125

Requirements: Optional

Description: Amount of cancer antigen 125 present in a blood serum sample, in U/mL, as measured by a laboratory test.

Permissible values: Integer field.

Source: Clinical notes

130. cea

Requirements: Optional

Description: Amount of carcinoembryonic antigen present in a blood serum sample, in ng/mL, as measured by a laboratory test.

Permissible values: Integer field.

Source: Clinical notes

131. er_status

Requirements: Optional

Description: Status of estrogen receptor (ER) expression. As per NAACCR guidelines, if less than 1% of cells stain positive the status is considered negative. 'Cannot be determined' should be used when a test was performed but produced an inconclusive result.

Permissible values: Controlled vocabulary field with five values: 'Negative', 'Positive', 'Cannot be determined', 'Not applicable', or 'Not available'.

Source: Pathology report, clinical notes

132. er_percent_positive

Requirements: Optional

Description: Percent of cells that stain ER positive by immunohistochemistry (IHC). The percentage should be indicated as a decimal value (e.g. 60% = 0.6).

Permissible values: Number field.

Source: Pathology report

133. pr_status

Requirements: Optional

Description: Status of estrogen receptor (ER) expression. As per NAACCR guidelines, if less than 1% of cells stain positive the status is considered negative. 'Cannot be determined' should be used when a test was performed but produced an inconclusive result.

Permissible values: Controlled vocabulary field with five values: 'Negative', 'Positive', 'Cannot be determined', 'Not applicable', or 'Not available'.

Source: Pathology report, clinical notes

134. pr_percent_positive

Requirements: Optional

Description: Percent of cells that stain ER positive by immunohistochemistry (IHC). The percentage should be indicated as a decimal value (e.g. 60% = 0.6).

Permissible values: Number field.

Source: Pathology report

135. her2_ihc_status

Requirements: Optional

Description: Status of human epidermal growth factor receptor-2 (HER2) expression, as assessed by immunohistochemistry (IHC). (Reference: AJCC 8th Edition, Chapter 48). Status values are defined as follows:

- Negative: 0 or 1+ staining
- Equivocal: 2+ staining
- Positive: 3+ staining

Permissible values: Controlled vocabulary field with six values: 'Positive', 'Negative', 'Equivocal', 'Cannot be determined', 'Not applicable' or 'Not available'.

Source: Pathology report, clinical notes

136. her2_ish_status

Requirements: Optional

Description: Status of human epidermal growth factor receptor-2 (HER2) gene copy number, as assessed by *in situ* hybridization (ISH). (Reference: AJCC 8th Edition, Chapter 48). Status values are defined as follows:

- For single-probe ISH:
  - Negative: ≤ 3 HER2 copies
  - Equivocal: 4 or 5 HER2 copies
  - Positive: ≥ 6 HER2 copies
- For dual-probe ISH:
  - Negative: HER2/CEP17 ratio < 2.0 AND HER2 copy number < 4
  - Equivocal (requires alternative ISH test to confirm equivocal): HER2/CEP17 ratio < 2.0 AND HER2 copy number ≥ 4 but < 6
  - Positive: HER2/CEP17 ratio ≥ 2 OR HER2 copy number ≥ 6 regardless of ratio

Permissible values: Controlled vocabulary field with six values: 'Positive', 'Negative', 'Equivocal', 'Cannot be determined', 'Not applicable' or 'Not available'.

Source: Pathology report, clinical notes

137. hpv_ihc_status

Requirements: Optional

Description: Human papillomavirus (HPV) p16 status, as assessed by immunohistochemistry (IHC).

Permissible values: Controlled vocabulary field with five values: 'Positive', 'Negative', 'Cannot be determined', 'Not applicable' or 'Not available'.

Source: Pathology report, clinical notes

138. hpv_pcr_status

Requirements: Optional

Description: Description: Human papillomavirus (HPV) status, as assessed by a polymerase chain reaction (PCR) assay.

Permissible values: Controlled vocabulary field with five values: 'Positive', 'Negative', 'Cannot be determined', 'Not applicable' or 'Not available'.

Source: Clinical notes

139. hpv_strain

Requirements: Optional/Conditional - This field can only be submitted if "hpv_pcr_status" = "Positive".

Description: If HPV status is positive by PCR, which strain(s) was (were) identified.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes

Comorbidity

The Comorbidity schema describes any other known diseases or conditions that are relevant to the index cancer, such as prior malignancies. It is linked directly to the Donor element as a parent.

140. program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

141. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

142. prior_malignancy

Requirements: Optional

Description: Whether the donor was affected by another malignancy prior to the one related to the index specimen.

Permissible values: Controlled vocabulary field with three values: 'Yes', 'No' or 'Not available'

Source: Clinical notes

143. laterality_of_prior_malignancy

Requirements: Optional/ Conditional – This field should only be submitted if 'prior_malignancy' = 'Yes'

Description: Laterality of previous malignancy.

Permissible values: Controlled vocabulary field with multiple values

Source: Clinical notes

144. age_at_comorbidity_diagnosis

Requirements: Optional/ Conditional

Description: Donor's age at diagnosis of comorbidity, in years.

Permissible values: Integer field

Source: Clinical notes

145. comorbidity_type_code

Requirements: Required/Conditional – This field is marked 'Conditional' because it depends on the value of the 'prior_malignancy' field.

Description: WHO ICD-10 code for the comorbidity (WHO ICD-10 classification). If 'prior_malignancy' = 'Yes', then an ICD-10 code related to cancer is expected in this field; if 'prior_malignancy' = 'No', then an ICD-10 code related to a non-cancer condition is expected.

Permissible values: ICD-10 code that meets the regular expression ^[A-Z][0-9]{2}(.[0-9]{1,3}[A-Z]{0,1})?$

Source (in order of reliability, if multiple available):

1. Pathology report(s)
2. Clinical notes

* The curator may need to extrapolate the code from these sources.

146. comorbidity_treatment_status

Requirements: Optional/Conditional

Description: Whether the donor has been or is being treated for the comorbidity, including a prior or concurrent malignancy.

Permissible values: Controlled vocabulary field with three values: 'Yes', 'No' or 'Not available'.

Source: Clinical notes

147. comorbidity_treatment

Requirements: Optional/Conditional

Description: Treatment details for the comorbidity, including a prior malignancy.

Permissible values: Text field.

Source: Clinical notes

Exposure

148. program_id

Unique identifier of the program (cohort), assigned by the data provider. See #1 for more information.

149. submitter_donor_id

Unique identifier for the donor, assigned by the data provider. See #2 for more information.

150. tobacco_smoking_status

Requirements: Optional

Description: Donor's self-reported smoking status and history.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes, study-administered survey

151. tobacco_type

Requirements: Optional/Conditional – This field cannot be submitted if 'tobacco_smoking_status' = 'Not applicable', 'Smoking history not documented' or 'Lifelong non-smoker (<100 cigarettes smoked in lifetime)'

Description: Type(s) of tobacco used by donor.

Permissible values: Controlled vocabulary field with multiple values.

Source: Clinical notes, study-administered survey

152. pack_years_smoked

Requirements: Optional/Conditional – This field cannot be submitted if 'tobacco_smoking_status' = 'Not applicable', 'Smoking history not documented' or 'Lifelong non-smoker (<100 cigarettes smoked in lifetime)'

Description: Smoking intensity in Pack Years, where the number of pack years is defined by the number of cigarettes smoked per day multiplied by the number of years smoked, divided by 20. This field only applies to cigarettes.

Permissible values: Number field.

Source: Clinical notes, study-administered survey

Document Revision History

Developed by	Reviewed by	Endorsed by	Effective Date	Policy Version	Summary of revisions
Clinical Data Model Guidelines Subgroup	Steering Committee	Network Council	3-Oct-24	V1	n/a

Authors

Name	**Institution**	**Title**
Véronique LeBlanc (Co-Chair)	TFRI	Network Program Manager, Scientific Writer
Emily Van de Laar (Co-Chair)	UHN	MOHCCN Project Manager
Parsa Bagherzadeh	McGill University	Postdoctoral Fellow
Christine Caron	CHUM	Project Manager
Juan Duran	McGill University	
Aileigh Kay	UHN	MOHCCN Project Manager
Josiane Lafleur	Jewish General Hospital	
Vanessa Pinatto Gaspar	McGill University	Project Manager
Carolyn-Ann Robinson	University of Calgary	Senior Research Associate
Marion Shadbolt	UHN	
Lincoln Stein	UofT, OICR	Professor, Head – Adaptive Oncology
Linda Xiang	OICR	