**Instructions and Attestation**

*This Check List is designed to facilitate the completion of items required for each data curation submission. Carefully review the Work Plan and Check List and sign the attestation prior to submitting a Data Curation response. Changes in Cycle 16 are highlighted in yellow. Submissions which do not conform to these requirements will not be approved.*

**Preparing for Data Curation**

1. Revise the ETL programs as needed to adhere to the current [CDM v6.1 specifications](https://pcornet.imeetcentral.com/p/aQAAAAAFCP6H) (including the specifications and the 2023\_04\_03 parseable file) and [PCORnet Data Checks v16](https://pcornet.imeetcentral.com/pcornetnew/folder/WzIwLDE0NTM2MDYwXQ/WzIsODU4MDIzNTBd/) (the changes to the Data Checks for this cycle are described in this document). For fields with undefined field lengths (e.g. PATID), harmonize lengths across all tables and use the minimum length needed to store the data in order to save storage space and reduce query processing time. Field lengths for non-RAW fields should not exceed 256 characters unless this is necessary.
2. **IMPORTANT:** Run the [CDM Diagnostic Query v6.10](https://pcornet.imeetcentral.com/p/ZgAAAAAA_PW8) to ensure table-level and field-level conformance to CDM v6.1, including harmonization of field lengths, before proceeding. Failure to conform to the CDM v6.1 structure will result in exceptions to required data checks which will necessitate a re-running of the DC query.
3. If your data is stored in SAS datasets, consider using the [Create Indexes v2.1](https://pcornet.imeetcentral.com/p/aQAAAAAE5Pob) program to improve run times.
4. Include the most recent data available, ideally through 3 months prior to the query run date or later. Failure to include recent data will result in an exception to Investigative Data Checks 3.07, 3.11, and/or 3.14.
5. Include new data for as many tables as possible and ETL changes as needed.
6. Populate optional fields, especially those examined by Investigative Data Checks 3.03, 3.08, and 3.09, as completely and accurately as possible.
7. Verify the accuracy of **all** fields in the HARVEST table. These data are used extensively by the Coordinating Center to characterize and analyze DataMart-level attributes. The DataMart and Network IDs and Names are listed in the [HARVEST Reference Table.](https://pcornet.imeetcentral.com/p/aQAAAAAFB8pe) Failure to populate these fields accurately will require a re-submission. Confirm that the REFRESH\_DATES are the most recent date on which data were extracted, transformed and loaded into the CDM. Failure to populate the REFRESH\_DATES will result in an exception to Required Data Check 1.18.
8. Data are expected to persist across refreshes. Selected data persistence issues may result in an exception to Investigative Data Checks 4.01, 4.02 and/or 4.03.
9. **IMPORTANT:** Run the self-service [CDM Value Set Conformance Query v6.10](https://pcornet.imeetcentral.com/p/ZgAAAAAA_PW9) to identify and resolve value set conformance errors prior to running the data curation query. This query examines the values populated in the 142 fields (21 tables) which have defined value sets (i.e., those listed in the parseable file’s Fields tab where the Valueset column is not “No”), including 17 fields which are required to be populated (i.e. the fields which have defined value sets and are also listed in the parseable file’s Constraints sheets as “required, not null”). The query allows users to designate which table(s) to check, and produces a PDF file reporting field-level evaluation results and SAS datasets and CSV files of the non-conforming records if any are present.
10. Consider running the self-service [Potential Code Errors Program v12](https://pcornet.imeetcentral.com/p/aQAAAAAFHekS) to identify exceptions to the expected formats for selected codes and LOINC codes that are in the incorrect CDM table.
11. Verify that all records conform to the CDM specifications for the following fields: ENCOUNTER.FACILITY\_LOCATION, LDS\_ADDRESS\_HISTORY.ADDRESS\_ZIP5 and LDS\_ADDRESS\_HISTORY.ADDRESS\_ZIP9. Invalid values will result in an exception to Required Data Check 1.17, but they will not be detected by the Value Set Conformance Query since these fields do not have defined value sets.
12. Create a static copy of the SAS version of your DataMart.
13. Decide on which option to use for running the Data Curation query package (modular or non-modular). The query has the capability of splitting the output into 2 parts. In order to use this functionality, the query must be run modularly. See the diagrams below for further instruction on how to set up the SAS programs for either approach. For additional detail, see Section IX: Query Workflow Diagrams of the Work Plan.





**Actions to take after running the Potential Code Errors Query**

1. Review the log file for warnings and errors. Refer to the [Query Package Issues Tracker](https://pcornet.imeetcentral.com/p/aQAAAAAE5RS7) to see if the log message is a known issue. You must remediate any errors and/or email [drnoc@pcornet.org](mailto:drnoc@pcornet.org) to determine if warnings are acceptable.*Note*: statements with the word \_ERROR\_ or the phrase “set the ERROR detection macro variable” are acceptable.
2. Review the Potential Code Errors report. These results are used for Required Data Check 1.13 and Investigative Data Check 1.16.

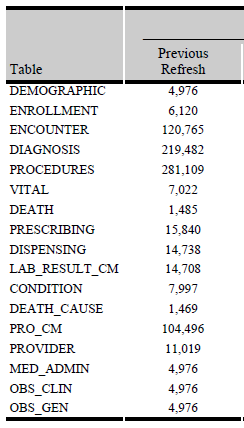
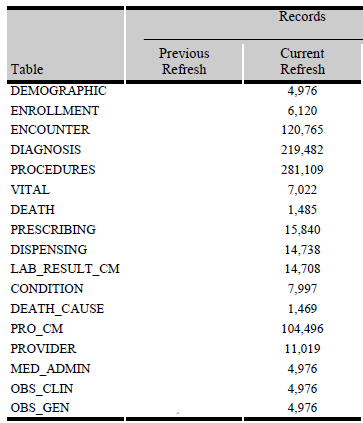
**Actions to take after running the Data Curation Queries**

1. Review the log file(s) for warnings and errors and remediate them as described previously.
2. Review the data curation query output tables. At a minimum, review the following tables to ensure that they accurately represent your source data.

* ***Basic metadata*:** DATAMART\_ALL, XTBL\_L3\_METADATA, XTBL\_L3\_DATES
* ***Code type frequencies****:* DIA\_L3\_DXTYPE, COND\_L3\_TYPE, MEDADM\_L3\_TYPE, OBSCLIN\_L3\_TYPE, OBSGEN\_L3\_TYPE, PROCM\_L3\_TYPE, PRO\_L3\_PXTYPE, IMMUNE\_L3\_CODETYPE
* ***Highest volume codes*:** DIA\_L3\_DX, DISP\_L3\_NDC, LAB\_L3\_LOINC, LAB\_L3\_SNOMED, LABHIST\_L3\_LOINC, MEDADM\_L3\_CODE\_TYPE, OBSCLIN\_L3\_CODE\_TYPE, OBSGEN\_L3\_CODE\_TYPE, PRES\_L3\_RXCUI, PRO\_L3\_PX, IMMUNE\_CODE\_CODETYPE
* ***Dates and times:*** XTBL\_L3\_TIMES and XTBL\_L3\_METADATA. In Cycle 16, variables were added to flag values which have invalid values (e.g. a date/time value) or implausible dates (dates prior to 1800). Due to the lookback date restrictions, implausible dates can only occur in the DEMOGRAPHIC, DEATH, DEATH\_CAUSE, PROVIDER, LDS\_ADDRESS\_HISTORY, HASH\_TOKEN, and LAB\_HISTORY tables.

**Actions to take after running the Empirical Data Curation Report**

1. Review the log file for warnings and errors and remediate them as described previously.
2. Review the entire Empirical Data Curation (EDC) report. Verify that all tables and charts that are relevant for your DataMart are present and that all information is accurate.
3. Address the following items in the EDC report.
   1. **Date obfuscation:** Exceptions to the CRN Statement of Work expectations regarding date obfuscation are highlighted in orange on Page 1 and in Table IF.
   2. **CDM Version**: Exceptions to the value set for the HARVEST.CDM\_VERSION field are highlighted in redon Page 1 and must be corrected before returning results.
   3. **Data Check exception summary**: Exceptions to required data checks are highlighted in red and must be corrected before returning results. Exceptions to investigative data checks are highlighted in blue and must be investigated and explained in the ETL ADD. To understand how each data check is calculated, use the information provided in the EDC report, the Empirical Data Curation specifications, the Data Curation Specifications for the source table(s), and the reference datasets as described in sections VII and X of the Work Plan.
   4. **Data Persistence Checks**: For Data Check 4.01, 4.02 and 4.03, confirm that the data in the Previous Refresh column matches the ‘Current Refresh’ in the previous refresh’s EDC report Table VA, Table VB and Table VC as illustrated below. Discrepancies will result in the need for a resubmission.



Previous EDC Report Table VA

Current EDC Report Table VA

**Actions to take prior to submitting results**

1. Ensure that the additional approval criteria are met. The approval criteria and the exemption request form are located on [iMeet](https://pcornet.imeetcentral.com/p/aQAAAAAE6tZU) and are included in the Zip file distributed with the query request. If a criterion is not met, please remediate this issue. In the rare circumstance that this criterion cannot be met, you must email a completed exemption request form to [drnoc@pcornet.org](mailto:drnoc@pcornet.org) prior to submitting your data curation package. The DRN OC will review the request and inform you if the exemption is approved. Please DO NOT submit results until the DRN OC makes a determination.
2. Update all sections of the ETL ADD survey (link distributed via email) with information about the current DataMart refresh. The ETL ADD must include the name and email of the person who assessed the implications of changes between refreshes and Investigative Data Check exceptions and approved the submission. **A submission is NOT considered complete without an accompanying ETL ADD response**.
3. Sign and date this document and return it with the other query documents as instructed in the Work Plan.

*I confirm that I have followed the instructions and met the requirements defined in the Work Plan and Check List.*

*Name Date*