**15 February 2019**

CDR Radiology Results

for the

MHS Data Repository (MDR)

(Version 1.03.00)

Current Specification

Revision History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Date | Originator | Para/Tbl/Fig | Description of Change |
| 1.00.00 | 11/27/2013 | J. MacLeod |  | Baseline |
| 1.01.00 | 3/31/2016 | J. MacLeod | Table 5  Table 4 | Modified the line for the “Appointment Record Id”.  Fixed the source to be MDR Ancillary data not CDR Appointment. |
| 1.02.00 | 3/28/2018 | J. MacLeod | Table 5 | Adjusted for NDAA-related changes in the LVM.  Dropped ACV. |
| 1.03.00 | 2/15/2019 | J. MacLeod | Table 5 | Adjusted the logic for the ACV Group to have it be set to blank for dates on or after 1/1/2019. |

# CDR Radiology Results

1. Background

This specification describes the process required to create the MDR Radiology Results data sets based on data received from the Clinical Data Repository (CDR). This Radiology Results SAS dataset represents the first radiology result data available in the MDR.

1. Sources

The source data files used to create the MDR Radiology Result data sets are extracted from the AHLTA/CDR. Transfer of the raw source extracts is handled by DHSS for loading into the MDR for further processing. The raw extract file is listed below:

**Table 1: CDR Sources**

| **CDR Source** | **Data Files** | **Purpose** |
| --- | --- | --- |
| CDR Radiology Results | RORST\*.DAT | Collection of records for direct care Radiology results in raw text form, prepared in accordance with the ICD. |

1. Transmission (Format and Frequency)

Source files are provided according to the frequency described in the Table 2.

**Table 2: Frequency of CDR Source Files**

|  |  |
| --- | --- |
| **Source File** | **Frequency** |
| CDR Radiology Results Table | Weekly |

1. Organization and batching

Source Data: The first step in MDR processing is to batch records received from CDR. Raw data batches are stored in /mdr/raw/cdr according to routine MDR operating procedures.

Output Products: There is one output product from the Radiology Results data processor, the Radiology Results dataset. It is described below:

* The Radiology Results table is a SAS dataset containing all FYs of data for FY09 and later. Each record represents a single result for a radiology test/status combination, in the form of a free-text 4,000 byte character string (the ‘Result Text’). The processor performs several external merges and many field derivations, and must also apply updates to records across CDR raw feed extract cycles.

The Radiology Results table is processed weekly. Table 3 contains the location and name of the output products. The preparation of them is described in subsequent sections of this document.

**Table 3: MDR Radiology Results Processor Output Products**

|  |  |  |
| --- | --- | --- |
| **Processor Output** | **File Naming Convention** | **Member Name** |
| MDR Radiology Results File | /mdr/pub/cdr/rad/results | rad\_results.sas7bdat |

Archival of files is also required, so that corresponding “apub” and other processing files (i.e. log, aprod, etc) are also loaded into the MDR according to routine operating procedures.

1. Receiving Filters
2. The feed files sent to the MDR are tilde (~) delimited files. The raw feeds are reviewed to ensure that the correct number of delimiters appear in each record.
3. Raw records on which both the result date and the order date are prior to FY09 are dropped.
4. When the raw data are ingested, if multiple records are found for a single record key & feed date combination, the record with the highest ROW\_NUMBER is kept. (Row Number is a field on the raw data feed but is not kept on the final MDR Radiology Results data set.) The record key is the combination of the following three variables: Event Id, Status Id, and Result Text. If multiple raw data feeds are ingested at the same time, for each record key, keep the record with the most recent FEEDDATE.
5. Update Processes
6. The raw feeds contain insert and update records. The type of record is identified by the ACTION field. The content of the action field shall be reviewed with each processing cycle. If delete records ever show up in the data, this specification and the processor it governs will need to be modified.
7. The master data set shall be updated by combining the raw data feed(s) with the master data set and keeping one record per key. Where there is more than one record per key, the record with the most recent FEEDDATE shall be retained.
8. Field Transformations and File Types

This section of the functional specification describes the data merges that are necessary to append some of the fields in the MDR Radiology Results data. There are several merges, described in Table 4, required to prepare the MDR CDR Radiology Results data set.

**Table 4: External File Merges and Formats**

| **Merge** | **Date Matching** | **Additional Matching** |
| --- | --- | --- |
| CDR Patient Table | Most recent CDR Patient Table is used for the radiology results processing. | See CDR Patient Specification. |
| MDR Longitudinal VM File | Radiology data: If the result date is populated, use that. Otherwise, if the order date is after 01 Oct 2003, use the order date.  LVM data: Use the LVM begin and end dates associated with the appropriate changeable demographic segment | EDI\_PN |
| CHCS Host Format | None | Apply the format to host\_facility\_id, which will return HOSTDMIS. |
| Health Data Dictionary | None | Status\_id to derive the status field  Rad\_type\_id to derive the rad\_type variable. |
| MDR Ancillary Data | None | HOST DMIS ID and Exam Number |

Business rules for each of the derived and appended fields that result from file merges and formats are described in the body of Tables 5 and 6.

1. record layout and content

The MDR Radiology Result data are stored in one SAS dataset that contains all FYs of data. The dataset is prepared according to the derivation rules listed in Table 5.

**Table 5. MDR Radiology Results SAS Data Set**

| **Variable Name** | **SAS Field Name** | **Format** | **Source Feeds** | **Related Source Field** | **Transformation Rule** |
| --- | --- | --- | --- | --- | --- |
| Radiology Event ID | EVENT\_ID | 8 | RORST | EVENT\_ID | No transformation. |
| Radiology Exam Number | EXAM\_NUMBER | 8 | RORST | RAD\_ORDER\_ID | No transformation. |
| Host Facility ID | HOST\_FACILITY\_ID | 8 | RORST | HOST\_FACILITY\_ID | No transformation. |
| CDR Patient ID | CDR\_PATIENT\_ID | $20 | RORST | UNIT\_NUMBER | No transformation. |
| Radiology Type ID | RAD\_TYPE\_ID | 8 | RORST | RAD\_TYPE\_ID | No transformation. |
| Status ID | STATUS\_ID | 8 | RORST | STATUS\_ID | No transformation. |
| Date and Time Ordered | DATE\_ORDERED | $20 | RORST | DATE\_ORDERED | No transformation. |
| Date and Time Resulted | DATE\_RESULTED | $20 | RORST | DATE\_RESULTED | No transformation. |
| Result Code | RESULT\_CODE | $3 | RORST | RESULT\_CODE | No transformation. |
| CDR Provider ID | PROVIDER\_ID | 8 | RORST | PROVIDER\_ID | No transformation. |
| Radiology Result | RESULT\_TEXT | $4000 | RORST | RESULT\_TEXT | No transformation. |
| Document ID | DOCUMENT\_ID | 8 | RORST | DOCUMENT\_ID | No transformation. |
| Document System NCID | DOCUMENT\_SYSTEM\_NCID | 8 | RORST | DOCUMENT\_SYSTEM\_NCID | No transformation. |
| Event GM Time | EVENT\_GMTIME | $20 | RORST | EVENT\_GMTIME | No transformation. |
| Transcription GM Time | TRANSCRIPTION\_GMTIME | $20 | RORST | TRANSCRIPTION\_GMTIME | No transformation. |
| **Health Data Dictionary** | | | | | |
| Result Status | STATUS | $70 | N/A | STATUS\_ID | Derive using the status lookup format. The format is produced from the HDD. |
| Radiology Type | RAD\_TYPE | $150 | N/A | RAD\_TYPE\_ID | Derive using the radiology type lookup format. The format is produced from the HDD. |
| **Internally-Derived Variables** | | | | | |
| Radiology Feed Date | FEEDDT | $8 | N/A | N/A | Derived from the file name of the feed.  if substr(file\_info,length(file\_info)-2,3) = ".gz" then feeddt = substr(file\_info,length(file\_info)-23,8);  else if substr(file\_info,length(file\_info)-2,3) = "DAT" then feeddt = substr(file\_info,length(file\_info)-20,8); |
| Date Ordered | ORDER\_DATE | 8 | RORST | DATE\_ORDERED | Date component (positions 1-8) of the DATE\_ORDERED field. |
| Time Ordered | ORDER\_TIME | $6 | RORST | DATE\_ORDERED | Time component (positions 9-15) of the DATE\_ORDERED field. |
| Date Resulted | RESULT\_DATE | 8 | RORST | DATE\_RESULTED | Date component (positions 1-8) of the DATE\_RESULTED field. |
| Time Resulted | RESULT\_TIME | $6 | RORST | DATE\_RESULTED | Time component (positions 9-15) of the DATE\_RESULTED field. |
| Fiscal Year Ordered | FY\_ORDERED | $4 | RORST | DATE\_ORDERED | Fiscal year of the DATE\_ORDERED field. |
| Fiscal Month Ordered | FM\_ORDERED | $2 | RORST | DATE\_ORDERED | Fiscal month of the DATE\_ORDERED field. |
| Fiscal Year Resulted | FY\_RESULTED | $4 | RORST | DATE\_RESULTED | Fiscal year of the DATE\_ RESULTED field. |
| Fiscal Month Resulted | FM\_RESULTED | $2 | RORST | DATE\_RESULTED | Fiscal month of the DATE\_ RESULTED field. |
| **CHCS Host Format** | | | | | |
| HOST DMIS ID | HOST\_DMIS\_ID | $4 | RORST | HOST\_FACILITY\_ID | Derive using the Host DMISID format. The format is generated using the HDD. |
| **MDR Ancillary Table Merge** | | | | | |
| Appt Record Id | APPTNO | $20 | N/A | APPTNO | From the MDR Ancillary data:  drop records where the first digit of the MPERSCD is A  keep only rectype=R  keep the first appointment id for each unique combination of Host DMIS ID & Exam Number. |
| **CDR Patient Table Merge** | | | | | |
| Universal Patient Identifier | UPID | $14 | N/A | N/A | Derived from the Patient table merge based on CDR\_PATIENT\_ID |
| EDIPN | EDIPN | $10 | N/A | N/A | Derived from the Patient table merge based on CDR\_PATIENT\_ID |
| Patient SSN | PATSSN | $9 | N/A | N/A | Derived from the Patient table merge based on CDR\_PATIENT\_ID |
| Sponsor SSN | SPONSSN | $9 | N/A | N/A | Derived from the Patient table merge based on CDR\_PATIENT\_ID |
| Patient Date of Birth | PATDOB | 8 | N/A | N/A | Derived from the Patient table merge based on CDR\_PATIENT\_ID |
| Patient Age | PATAGE | 8 | N/A | N/A | Derived using aprod/util macro by subtracting PATDOB from Date Start |
| Patient Category | PATCAT | $3 | N/A | N/A | Derived from the Patient table merge based on CDR\_PATIENT\_ID |
| Age Group Code | AGEGRP | $1 | N/A | N/A | Derived using aprod/util macro. A = 0-17, B = 18-24, etc. |
| **LVM Table Merge** | | | | | |
| DEERS Gender | GENDER | $1 | N/A | N/A | Fill with gender from LVM based on EDIPN. If the gender is blank or U, set to “Z”. |
| DEERS Enrollment DMIS ID | DENRSITE | $4 | N/A | N/A | Fill with enrollment DMISID from LVM based on EDIPN, if the result date is between the begin date and the end date associated with the enrollment site.  See VM6 Specification, Exhibits G-18 and 19 for segment and field positions. |
| DEERS Beneficiary Category | BENCAT | $3 | N/A | N/A | Fill with DEERS beneficiary category from LVM based on EDIPN, if the result date is between the begin date and the end date associated with the DEERS beneficiary category. If no match for the person or the bencat is Z, set to “UNK”.  See VM6 Specification, Exhibits G-18 and 19 for segment and field positions. |
| DEERS Common Beneficiary Category | COMBEN | $1 | N/A | N/A | Derived from DEERS Beneficiary Category during LVM merge using MDR utility programs.  See VM6 Specification, section A.1.12 for derivation.  If no match for the person, set to “3”. |
| DEERS Sponsor Service | DSPONSVC | $1 | N/A | N/A | Fill with DEERS sponsor service from LVM based on EDIPN, if the result date is between the begin date and the end date associated with the DEERS sponsor service. If no match for the person, set to “Z”.  See VM6 Specification, Exhibits G-18 and 19 for segment and field positions. |
| DEERS Sponsor Service Aggregate | DSVCAGG | $1 | N/A | N/A | Fill with DEERS sponsor service (aggregate) from LVM based on EDIPN, if the result date is between the begin date and the end date associated with the DEERS sponsor service (aggregate). If no match for the person, set to “Z”.  See VM6 Specification, Exhibits G-18 and 19 for segment and field positions. |
| DEERS Relationship to Sponsor | RELATIONSHIP | $1 | N/A | N/A | Fill with DEERS Relationship from the LVM based on EDIPN and SPONSSN. If Relationship not found in LVM merge, then derive from FMP from Patient Table Merge. |
| DEERS Health Care Delivery Program Code | HCDP | $3 | N/A | N/A | Fill with DEERS HCDP code from LVM based on EDIPN, if the result date is between the begin date and the end date associated with the DEERS HCDP code.  See VM6 Specification, Exhibits G-18 and 19 for segment and field positions. |
| DEERS ZIP Code | DEERSZIP | $5. | N/A | N/A | Fill with DEERS ZIP code from LVM based on EDIPN, if the result date Is between the begin date and the end date associated with the DEERS ZIP code.  See VM6 Specification, Exhibits G-18 and 19 for segment and field positions. |
| DEERS Eligibility Group | ELG\_GRP | $2 | N/A | N/A | Fill with Eligibility Group from LVM if the date of processing is between the begin date and the end date of the associated segment. If no match for the person, set to “Z”.  See VM6 Beneficiary Specification, Exhibit G19. |
| DEERS Enrollment Group | ENR\_GRP | $2 | N/A | N/A | Fill with Enrollment Group from LVM if the date of processing is between the begin date and the end date of the associated segment. If no match for the person, set to “Z”.  See VM6 Beneficiary Specification, Exhibit G19. |
| DEERS PCM Type | PCM\_TYPE | $1 | N/A | N/A | Fill with PCM Type from LVM if the date of processing is between the begin date and the end date of the associated segment. If no match for the person, set to “Z”.  See VM6 Beneficiary Specification, Exhibit G19. |
| DEERS Assigned Health Care Delivery Program Code | HCDP\_ASSGN | $3 | N/A | N/A | Fill with Assigned HCDP Code from LVM if the date of processing is between the begin date and the end date of the associated segment.  See VM6 Beneficiary Specification, Exhibit G19. |
| DEERS ACV Group | ACVGROUP | $2 | N/A | N/A | For dates on or after 1/19/2019:  Set to blank.  For dates prior to 1/19/2019:  Derived by the MDR utilities during LVM merge based on Enrollment Group, PCM Type, Eligibility Group, and Common Beneficiary Category or ACV and Common Beneficiary Category depending on whether date of processing is before or after 1/1/18. If no match for the person, set to “O”.  See VM6 Beneficiary Specification, Section G.3 for details. |

1. Refresh Frequency

Frequency of updates (based on Radiology Result Date):

* Weekly.
* Retrofits: On an as needed basis when data corrections or updates are

required.

1. Data Quality

It is expected that when the processor is run each week, that basic quality checks are performed throughout the process. It is recommended that the DHSS vendor develop a spreadsheet which tracks key characteristics of the data across processing cycles; making it relatively easy to understand how the data should generally look. DHSS vendors need to review these statistics each month prior to releasing the data. DHCAPE (the functional proponent and the specification author) should be contacted immediately should any quality issues arise. These checks, at a minimum, should include:

* The following variables have been blank on all records on the data received to date. Continue to check whether they become populated.
  + Abnormal Flag NCID
  + Dictated for Clinician Name
  + Dictated for Clinician NCID
  + Dictating Clinician Name
  + Dictating Clinician NCID
  + Dictation GM TIME
  + Entered by Name
  + Entered by NCID
  + Point of Care Location NCID
  + Point of Care Terminal NCID
  + Reporting Flag NCID
  + Location NCID
  + Update Application NCID
* Total record counts in the data feed should have a relatively stable distribution across the Result Date, accounting for weekends and holidays. Any anomalies should immediately be investigated.
* The number of records that match when doing the CDR Patient table merge should be consistent.
* The distribution of all categorical fields (e.g., DMISID, STATUS) should be consistent. The results of proc freq analyses will verify this.
* The number of null values for important fields such as CDR\_PATIENT\_ID and EXAM\_NUMBER should be tracked across monthly updates.
* When reading in the data feeds, a small number of records should be proc printed and manually inspected to ensure they have read in properly.
* Cross tabulations should be reviewed on derived elements to ensure the derivation logic works.
* A data flow tracker should be built to ensure that all records that are intended to make it into the final datasets do. In other words, all inserts and updates should be tracked and explained in the data flow worksheet.