**24 October 2012**

CDR Microbiology

for the

MHS Data Repository (MDR)

(Version 1.00.00)

Current Specification

Revision History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Date | Originator | Para/Tbl/Fig | Description of Change |
| 1.00.00 | 10/24/2012 | C. Kangas |  | Baseline |

# CDR Microbiology

1. Background

This specification describes the process required to create the MDR Microbiology tables based on data received from the Clinical Data Repository (CDR). The scope of the microbiology result data coming from the CDR is a subset of MHS direct care lab orders and results. Separate MDR datasets contain lab result data related to Chemistry and Pathology lab results. The MDR has contained ancillary lab order datasets for many years, generated from extracts received from CHCS. This Microbiology SAS dataset represents the first microbiology result data available in the MDR.

The goal of antimicrobial susceptibility testing is to predict the success or failure of antibiotic therapy. Tests are performed to measure the growth response of an isolated organism to a particular drug. The results of antimicrobial susceptibility testing should be combined with clinical information and experience when selecting the most appropriate antibiotic for a patient. In the CDR data, a microbiology testing event contains three levels of data:

* + Lab Order information. The lab test types (NCIDs) that are ordered are usually fairly general, like “throat culture”.
	+ Microbiology results typically have a lab order ID to follow back upwards to the lab order. Their lab test types are not the same NCIDs as were on the lab order, but are more specific such as “test for Strep B in a throat culture”. Each result has a unique CDR event ID of a range that represents “reporting a lab result event”. The result information at this level consists of several large free-text character strings (the microbiologist’s interpretations).
	+ Susceptibility detail, which can be tracked to a result by the lab result event\_ID (from the same range as the microbiology result event\_ID above). The susceptibility detail information describes the organism’s susceptibility to various types of antibiotics.
1. Sources

The source data files used to create the MDR Microbiology tables are extracted from the AHLTA/CDR. The transfer of the raw source extracts is handled by DHSS for loading into the MDR for further processing. These raw extract files are listed below:

**Table 1. CDR Sources**

| **CDR Source** | **Data Files** | **Purpose** |
| --- | --- | --- |
| CDR Lab Orders Table | LBORD\*.DAT | Collection of records for direct care Laboratory Orders in raw text form, prepared in accordance with the ICD. |
| CDR Microbiology Results Table | LBMTS\*.DAT | Collection of records for direct care Microbiology results in raw text form, prepared in accordance with the ICD. |
| CDR Susceptibility Detail Table | LBMTD\*.DAT | Collection of records for microbiology susceptibility detail 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. The format of these feeds is described in ICD-1300-6445-01.

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

|  |  |
| --- | --- |
| **Source File** | **Frequency** |
| CDR Laboratory Orders Feeds | Weekly |
| CDR Microbiology Results Feeds | Weekly |
| CDR Susceptibility Detail Feeds | 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 according to routine MDR operating procedures.

Output Products: There are two output products from the MDR Microbiology processor: the primary output is the Microbiology Results dataset, and the second is a Susceptibility detail dataset. Both are described below:

The Microbiology Results table is a single FY level SAS dataset. Each record represents a summary-level result for a microbiology testing event, provided in the form of several large free-text character strings. These represent the microbiologist’s interpretation of the test results. The multiple, long free-text character strings cause the width (i.e., observation length) of this SAS dataset to be very large. Consequently, MDR users are strongly encouraged to utilize the SAS compression option (options compress=binary;) when accessing the Microbiology Results dataset to reduce unnecessary disk I/O and to improve performance.

The lab ordering information is physically stored on each record of the Microbiology Results table as well. The processor performs several external merges and many field derivations, and must also apply updates to records across extract cycles. The processor needs to be run once for each FY to be processed. The processor also merges in the first 10 organisms available from the Susceptibility Detail table (described below), when present, for a given Microbiology Result record.

The Susceptibility detail table is a single FY level SAS dataset. This is provided to MDR users who want to perform analysis on the susceptibility of each organism present in the specimen, usually to a set of different antibiotics. The main Microbiology Results dataset physically stores the first 10 organisms from the Susceptibility detail table, however there can be instances when more than 10 organisms were present. This Susceptibility detail table can store an unlimited number of organisms and antibiotics tested for a single event. The key field that can be used to join the Susceptibility detail table to the main Microbiology Results table is the MICRO\_EVENT\_ID field.

The microbiology tables are processed weekly, and past fiscal years are processed on a less frequent basis (ex. FY09 is processed twice a year). 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 Microbiology Processor Output Products**

|  |  |  |
| --- | --- | --- |
| **Microbiology Processor Output** | **File Naming Convention** | **Member Name** |
| MDR Microbiology Results File  | /mdr/pub/cdr/micro/fy<yy> | micro\_results.sas7bdat |
| MDR Susceptibility Detail File  | /mdr/pub/cdr/micro/fy<yy> | suscept\_detail.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. In the feeds, there are free text fields that occasionally contain tildes within them, which disrupts the normal ingest function and structure of fields in a delimited file. In these cases, the processor reads in the field created by the inadvertent extra delimiter and drops it. For the "good" field that had the extra tilde within it, this method only keeps the content of the field up to the tilde embedded in it; any content after the tilde in that field is dropped.
3. Only raw records with result dates in FY09 or later are kept.
4. Field Transformations and Deletions for MDR
5. The key fields that link the Lab Orders (LBORD\*) feeds to the Microbiology Results (LBMTS\*) feeds are the HOST\_FACILITY\_ID and LAB\_ORDER\_ID.
6. The key field that links the Microbiology Results (LBMTS\*) feeds to the Susceptibility Detail table (LBMTD\*) feeds is the EVENT\_ID field.
7. During the extraction of the raw lab order records, de-duplication of records, or anytime a lab order key collision occurs between incoming data and existing master data, the processor de-duplicates data by selecting the largest value of the Feed Date (FEEDDATE) for any given order key (HOSTDMIS + LAB\_ORDER\_ID). If multiple records exist with the same FEEDDATE and order key, the record with the largest ROW\_NUMBER is kept.

The final record that remains for a given order is then kept or deleted based on the ACTION field. For records with an ACTION = “D” (delete), the record is removed from the lab orders dataset. For all other records with an ACTION = “I” (insert) or ACTION = “U” (update), the records are kept.

1. During the initial ingest step of the raw microbiology results (LBMTS\*) records, the data are first sorted by EVENT\_ID, FEEDDT, ACTION, FILENUM, GROUP\_TYPE\_NCID, and GROUP\_ORDINAL. If any records exist that have completely duplicate information in these fields, the records are de-duplicated using SAS by-level processing, keeping the last GROUP\_ORDINAL record.
2. During the processing of the raw microbiology result (LBMTS\*) records, and anytime a microbiology event key collision occurs between incoming data and existing master data, the processor de-duplicates data by selecting the set of records associated with the largest value of the Feed Date (FEEDDATE) for a given event (EVENT\_ID). If the same EVENT\_ID appears multiple times in the same batch, the processor de-duplicates these by selecting the set of records associated with the largest value of the ACTION and GROUP\_ORDINAL fields. In the raw feed data, only ACTION values of “INSERT” and “UPDATE” exist, and no “DELETE” records have ever been received from the source.
3. During the extraction of the raw microbiology susceptibility detail (LBMTD\*) records, de-duplication of records, or anytime a lab event key collision occurs between incoming data and existing master data, the processor de-duplicates the data by selecting the set of records associated with the largest value of the Feed Date (FEEDDATE) for any given event key (EVENT\_ID). If multiple records exist with the same FEEDDATE and event key, the records with the largest ACTION and OBSERVATION\_SEQ are kept for the event. In the raw feed data, only ACTION values of “INSERT” and “UPDATE” exist, and no “DELETE” records have ever been received from the source.
4. There are several merges and formats required to prepare the MDR Microbiology Results and Susceptibility Detail tables as described in Table 4.

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

| **Merge** | **Date Matching** | **Additional Matching** |
| --- | --- | --- |
| CDR Patient Table | Most recent CDR Patient Table is used for the Microbiology file. | See CDR Patient Specification. |
| MDR Longitudinal VM File | Date Resulted, with begin and end dates for each changeable demographic segment | EDI\_PN (if available). |
| CHCS Host Format | None | Apply the format to host\_facility\_id, which will return HOSTDMIS. |
| 3M HDD | None | Apply HDD-based formats to several different NCID types as described in Tables 5 and 6. |

1. 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.
2. record layout and content

The MDR Microbiology tables are stored as one SAS dataset per fiscal year. The dataset is prepared according to the derivation rules listed in Table 5.

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

| **Variable Name** | **SAS Field Name** | **Format** | **Source Feeds** | **Related Source Field** | **Transformation Rule** |
| --- | --- | --- | --- | --- | --- |
| CHCS Host DMIS ID | HOSTDMIS | $4. | LBORD, LBMTS | HOST\_FACILITY\_ID | Derived from application of the CDR Host DMISID format:hostdmis = put(host\_facility\_id,hostdmis.);See Appendix A for the hostdmis format. |
| Lab Order ID | LAB\_ORDER\_ID | $12. | LBORD, LBMTS | LAB\_ORDER\_ID | No transformation. This field is part of the key (along with HOSTDMIS) to join the LBORD and LBMTS feeds together. |
| Ordering DMISID | DMISID | $4. | LBORD | CLINIC\_ID | Derived from application of the CDR Clinic to DMISID format:DMISID = put(clinic\_id,cdr\_dmis.); |
| Ordered Test NCID | ORDER\_TEST\_NCID | 8. | LBORD | LAB\_TEST\_ID | No transformation. |
| Microbiology Test NCID | MICRO\_TEST\_NCID | 8. | LBMTS | LAB\_TEST\_ID | No transformation. |
| Microbiology Test Name | MICRO\_TEST | $80. | LBMTS | LAB\_TEST\_ID | Derived from application of HDD lab test name format; MICRO\_TEST = put(lab\_test\_id,lab\_test\_name.) |
| Microbiology Result Event ID | MICRO\_EVENT\_ID | 8. | LBMTD, LBMTS | EVENT\_ID | No transformation. This field is the key field to join the LBMTS and LBMTD feeds together. |
| Ordered Test Name | TEST\_NAME | $30. | LBORD | UNIVERSAL\_SERVICE\_TEXT | No transformation. |
| Order Date | ORDER\_DATE | $8. | LBORD | ORDER\_DATE | No transformation. |
| Bacteriology Results | BAC\_RESULT | $2000. | LBMTS | TEXT\_COMMENT | If the first 20 characters of the source field TEXT\_COMMENT are “BACTERIOLOGY RESULT:”, this field is derived by capturing the rest of the field after the semi-colon. If the free text result stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| Gram Stain Results | GRAM\_STAIN | $2000. | LBMTS | TEXT\_COMMENT | If the first 11 characters of the source field TEXT\_COMMENT are “GRAM STAIN:”, this field is derived by capturing the rest of the field after the semi-colon. If the free text result stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| Mycology Results | MYC\_RESULT | $2000. | LBMTS | TEXT\_COMMENT | If the first 16 characters of the source field TEXT\_COMMENT are “MYCOLOGY RESULT:”, this field is derived by capturing the rest of the field after the semi-colon. If the free text result stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| Smear Preparation | SMEAR\_PREP | $2000. | LBMTS | TEXT\_COMMENT | If the first 11 characters of the source field TEXT\_COMMENT are “SMEAR/PREP:”, this field is derived by capturing the rest of the field after the semi-colon. If the free text result stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| Virology Result | VIR\_RESULT | $2000. | LBMTS | TEXT\_COMMENT | If the first 16 characters of the source field TEXT\_COMMENT are “VIROLOGY RESULT:”, this field is derived by capturing the rest of the field after the semi-colon. If the free text result stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| Other Result | OTHER\_RESULT | $2000. | LBMTS | TEXT\_COMMENT | If the TEXT\_COMMENT field does not indicate a Bacteriology, Gram Stain, Mycology, Smear Prep, or Virology Result, and is not an Amendment History comment or a Result Comment, and is not part of the ordering “header” lines (ex. Ordering Provider), and does contain a semi-colon in the TEXT\_COMMENT value, then this OTHER\_RESULT field is derived by capturing the entire TEXT\_COMMENT value. If the free text result stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| Other Result Type 1-3 | OTHER\_RESULT\_TYPE\_(3) | $70. | LBMTS | TEXT\_COMMENT | Derived by capturing the value of OTHER\_RESULT to the left of the semi-colon. |
| Amendment History | AMEND\_HIST | $4000. | LBMTS | TEXT\_COMMENT | If the first 10 characters of the source field TEXT\_COMMENT are “AMENDHIST:”, this field is derived by capturing the rest of the field after the semi-colon. If the free text history stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| LOINC Code | LOINC | $9. | LBMTS | LAB\_TEST\_ID | Derived from application of HDD LOINC format; LOINC = put(micro\_test\_ncid,loinc.); |
| LOINC Name | LOINC\_NAME | $70. | LBMTS | LAB\_TEST\_ID | Derived from application of HDD LOINC Name format; LOINC\_NAME = put(micro\_test\_ncid,loinc\_name.); |
| Organism 1-Organism 10 | ORGANISM\_(10) | $100. | LBMTD | OBSERVATION\_NCID, VALUE\_NCID | See Table 6 for the derivation of the Organism field on the Susceptibility Detail table. These 10 Organism fields are obtained via a merge to the Susceptibility Detail table by MICRO\_EVENT\_ID and stored here on the Micro Results table. |
| Organism Quantity 1-Organism Quantity 10 | QTY\_(10) | $20. | LBMTS | TEXT\_COMMENT | If the TEXT\_COMMENT field does not indicate a Bacteriology, Gram Stain, Mycology, Smear Prep, Virology, or Other Result, and is not an Amendment History comment or a Result Comment, and is not part of the ordering “header” lines (ex. Ordering Provider), then this QTY field is derived by capturing the entire TEXT\_COMMENT value. |
| Result Comment | RESULT\_COMMENT | $2000. | LBMTS | TEXT\_COMMENT | If the first 18 characters of the source field TEXT\_COMMENT are “RESULT COMMENT(S):”, this field is derived by capturing the rest of the field after the semi-colon. If the free text comment stretches across multiple records, each successive record is concatenated to build the full string. New records are separated by a caret (^) symbol. |
| Sample Collect Date | DATE\_COLLECTED | $8. | LBMTS | DATE\_COLLECTED | No transformation. |
| Result Date | DATE\_RESULTED | $8. | LBMTS | TEXT\_COMMENT | If the first 14 characters of the source field TEXT\_COMMENT are “DATE\_RESULTED:”, this field is derived by capturing the next 8 characters. |
| Order Start Date | DATE\_START | $8. | LBORD | DATE\_START | No transformation. |
| Order End Date | DATE\_END | $8. | LBORD | DATE\_END | No transformation. |
| Fiscal Year | FY | $4. | LBMTS | DATE\_RESULTED | Fiscal year equivalent of calendar year of DATE\_RESULTED. |
| Fiscal Month | FM | $2. | LBMTS | DATE\_RESULTED | Fiscal month equivalent of calendar month of DATE\_RESULTED. |
| Result Status | STATUS | $1. | LBMTS | STATUS\_ID | Derived from application of HDD Lab Status format:status = substr(put(status\_id,lab\_status.),1,1) |
| Sample Collected | COLLECT\_SAMPLE | $40. | LBMTS | TEXT\_COMMENT | If the first 15 characters of the source field TEXT\_COMMENT are “COLLECT\_SAMPLE:”, this field is derived by capturing the rest of the field after the colon. |
| Specimen | SPECIMEN | $30. | LBORD | SPECIMEN\_TEXT | No transformation. |
| Specimen Code | SPECIMEN\_CODE | 8. | LBORD | SPECIMEN | No transformation. |
| Priority (Result) | PRIORITY\_RESULT | $20. | LBMTS | TEXT\_COMMENT | If the first 14 characters of the source field TEXT\_COMMENT are “Priority:”, this field is derived by capturing the rest of the field after the colon. |
| Lab Order Grouping ID | GROUP\_ID | 8. | LBORD | GROUP\_ID | No transformation. |
| Ordering Provider | ORDERING\_PROVIDER | $50. | LBMTS | TEXT\_COMMENT | If the first 18 characters of the source field TEXT\_COMMENT are “Ordering Provider:”, this field is derived by capturing the rest of the field after the colon. |
| Order Comment | ORDER\_COMMENT | $20. | LBORD | ORDER\_COMMENT | Derived by keeping first 20 characters of the source field ORDER\_COMMENT. |
| Order Status | ORDER\_STATUS | $2. | LBORD | ORDER\_STATUS\_CODE | No transformation. |
| Specimen Collection Method | COLLECT\_METHOD | $1. | LBORD | COLLECT\_METHOD | No transformation. |
| CDR Patient ID | CDR\_PATIENT\_ID | $20. | LBMTS | UNIT\_NUMBER | No transformation. |
| CDR Appointment ID | CDR\_APPT\_ID | 8. | LBORD | APPT\_ID | No transformation. |
| CDR Host ID | CDR\_HOST\_ID | 8. | LBORD, LBMTS | HOST\_FACILITY\_ID | No transformation. This is part of the key (along with LAB\_ORDER\_ID) to join the LBORD and LBMTS feeds together. |
| CDR Clinic ID | CDR\_CLINIC\_ID | 8. | LBORD | CLINIC\_ID | No transformation. |
| CDR Provider ID (Ordering) | CDR\_PROVIDER\_ID | 8. | LBORD | PROVIDER\_ID | No transformation. |
| 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); |
| **CDR Appointment Table Merge** |
| CHCS Appt IEN | APPTIDNO | $10. | N/A | APPT\_ID | Derived from the Appointment table merge based on CDR\_APPT\_ID |
| **CDR Patient Table Merge** |
| Universal Patient ID | UPID | $14. | N/A | N/A | Derived from the Patient table merge based on CDR\_PATIENT\_ID |
| EDIPN | EDI\_PN | $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. |
| DEERS Enrollment DMISID | DENRSITE | $4. | N/A | N/A | Fill with enrollment DMISID from LVM based on EDIPN, if the result date is between the begin and end date associated with the enrollment site. |
| 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 and end date associated with the DEERS beneficiary category. If no match for the person, set to “UNK”. |
| DEERS Common Beneficiary Category | COMBEN | $1. | N/A | N/A | Fill with DEERS common beneficiary category from LVM based on EDIPN, if the result date is between the begin and end date associated with the DEERS common beneficiary category. 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 and end date associated with the DEERS sponsor service. |
| 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 and end date associated with the DEERS sponsor service (aggregate). |
| DEERS Alternate Care Value | ACV | $1. | N/A | N/A | Fill with ACV from LVM based on EDIPN, if the result date is between the begin and end date associated with the ACV, else if ACV is blank after LVM merge and bencat is ACT or GRD then set ACV to M, otherwise set to blank. |
| DEERS Relationship | 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 HCDP | HCDP | $3. | N/A | N/A | Fill with DEERS HCDP code from LVM based on EDIPN, if the result date is between the begin and end date associated with the DEERS HCDP code. |
| 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 and end date associated with the DEERS ZIP code. |

**Table 6. MDR Susceptibility Detail SAS Data Set**

The MDR Susceptibility Detail table is stored as one SAS dataset for all-time. The dataset is prepared according to the derivation rules listed in Table 6.

| **Variable Name** | **SAS Field Name** | **Format** | **Source Feeds** | **Related Source Field** | **Transformation Rule** |
| --- | --- | --- | --- | --- | --- |
| Microbiology Result ID | MICRO\_EVENT\_ID | 8. | LBMTD | EVENT\_ID | No transformation. This field is the key field to join this table to the Microbiology Results table. |
| Organism | ORGANISM | $50. | LBMTD | OBSERVATION\_NCID, VALUE\_NCID | Derived from application of HDD Organism format:organism = put(organism\_ncid,org\_name.); |
| Antibiotic | ANTIBIOTIC | $70. | LBMTD | OBSERVATION\_NCID, VALUE\_NCID | Derived from application of HDD Antibiotic format:antibiotic = put(antibiotic\_ncid,med\_name.); |
| Operator | OPERATOR | $2. | LBMTD | SENSITIVITY\_OPERATOR\_ID | If OBSERVATION\_NCID = 119279 then do;operator\_full = put(operator\_ncid,operator.);if length(operator\_full) le 2 then operator = operator\_full;else if operator\_full = 'EQUAL' then operator = '=';else if operator\_full = 'GREATER THAN' then operator = '>';else if operator\_full = 'LESS THAN' then operator = '<';else operator = ''; End; |
| Concentration | CONCENTRATION | 8. | LBMTD | VALUE\_NUMBER | If OBSERVATION\_NCID = 119279, then concentration = VALUE\_NUMBER. |
| Susceptibility | SUSCEPTIBILITY | $20. | LBMTD | OBSERVATION\_NCID, VALUE\_NCID | If OBSERVATION\_NCID = 119277, then susceptibility\_ncid = VALUE\_NCID. susceptibility = put(susceptibility\_ncid,sus\_result.); |
| Method | METHOD | $3. | LBMTD | OBSERVATION\_NCID | If OBSERVATION\_NCID = 119269 then method = “KB”. Else if OBSERVATION\_NCID = 119270 then method = “MIC” |
| Organism NCID | ORGANISM\_NCID | 8. | LBMTD | OBSERVATION\_NCID, VALUE\_NCID | If OBSERVATION\_NCID = 26770, then organism\_ncid = VALUE\_NCID |
| Antibiotic NCID | ANTIBIOTIC\_NCID | 8. | LBMTD | OBSERVATION\_NCID, VALUE\_NCID | If OBSERVATION\_NCID in(119273, 119274) then antibiotic\_ncid = VALUE\_NCID |

1. Refresh Frequency

Frequency of updates (based on Microbiology 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 Microbiology 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:

* 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 (ex. DMISID, ORDER\_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, PATH\_RESULT\_ID, and DMISID should be tracked across monthly updates.
* When reading in the Microbiology 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 Microbiology datasets do. In other words, all inserts, updates, and deletions should be tracked and explained in the data flow worksheet.

**Appendix A: Description of HOSTDMIS format**

The raw CDR feeds contain a CDR unique ID (HOST\_FACILITY\_ID) for the CHCS host, which is not common to any other data table within the MDR. Therefore a SAS format was created to translate the CDR host facility ID to a DMIS ID, a field commonly used in the MDR.

The application of the hostdmis format to translate the host\_facility\_id into hostdmis is done with the following statement: hostdmis = put(host\_facility\_id,hostdmis.);

Below is the proc format code that is used to develop the hostdmis SAS format:

**proc** **format**;

value hostdmis

**76313**='0364'

**76318**='0109'

**76323**='0128'

**76328**='0118'

**76333**='0110'

**76338**='0062'

**76810**='1170'

**1046961**='0052'

**1048021**='0090'

**1049621**='0124'

**1059821**='0089'

**1067401**='0125'

**1074201**='0091'

**1097342**='0248'

**1097429**='0018'

**1097561**='0013'

**1097861**='0055'

**1098981**='0338'

**1099041**='0114'

**1099139**='0098'

**1099332**='0096'

**1099822**='0113'

**1100881**='0108'

**1101099**='0029'

**1104242**='0097'

**1104381**='0112'

**1104541**='0014'

**1105841**='0064'

**1106441**='0028'

**1106901**='0131'

**1107161**='0019'

**1107201**='0024'

**1112813**='0057'

**1113124**='0049'

**1113704**='0048'

**1120878**='0047'

**1130428**='0060'

**1132134**='0038'

**1132684**='0039'

**1134172**='0032'

**1135465**='0103'

**1137626**='0101'

**1138685**='0053'

**1138927**='0056'

**1143097**='0073'

**1144654**='0009'

**1145022**='0045'

**1145350**='0067'

**1177297**='0008'

**1178200**='0001'

**1178583**='0330'

**1180847**='0003'

**1181105**='0058'

**1181588**='0607'

**1185029**='0061'

**1187857**='0075'

**1195255**='0035'

**1200322**='0005'

**1208940**='0612'

**1209517**='0086'

**1214474**='0615'

**1214671**='0010'

**1214914**='0129'

**1215101**='0616'

**1215502**='0620'

**1216727**='0621'

**1217255**='0624'

**1217474**='0618'

**1217695**='0084'

**1217869**='0077'

**1217983**='0085'

**1218117**='0006'

**1218586**='0050'

**1218870**='0119'

**1219060**='0617'

**1219293**='0059'

**1219472**='0808'

**1219659**='0310'

**1224255**='0079'

**1224847**='0093'

**1224981**='0004'

**1225163**='0076'

**1225324**='0078'

**1225841**='0106'

**1226061**='0635'

**1226215**='0637'

**1226261**='0633'

**1226659**='0083'

**1226824**='0623'

**1226983**='0638'

**1227261**='0074'

**1227781**='0094'

**1228014**='0042'

**1228561**='0051'

**1228789**='0043'

**1229006**='0639'

**1229178**='0622'

**1229704**='0629'

**1259764**='0046'

**1267414**='0036'

**1272512**='0326'

**1302739**='0095'

other = ' ';

**run**;