Metathesaurus - Original Release Format (ORF)

[*4*]

Metathesaurus users may select from two relational formats: the Rich Release Format (RRF), first introduced in 2004, and the Original Release Format (ORF). Both are available as output options of MetamorphoSys, the installation and customization program.

Developers are encouraged to use the RRF, which offers significant advantages in source vocabulary transparency (that is, ability to exactly represent the detailed semantics of each source vocabulary); in the ability to generate complete and accurate change sets between versions of the Metathesaurus; and in more convenient representations of concept name, source, and hierarchical context information.

Neither Metathesaurus format is fully normalized. By design, there is duplication of data among different files and within certain files. In particular, relationships between different Metathesaurus concepts appear twice (e.g., from entry A to entry B and from entry B to entry A). Developers will need to make their own decisions about the extent to which this redundancy should be retained, reduced, or increased for their specfic applications.

Note: The preferred and more complete format is described in Chapter 3, the Metathesaurus Rich Release Format (RRF).

All files except MRRANK are sorted by row.

4.1. Data Files

The data in each Metathesaurus entry may be represented in more than 20 different "relations" or files. These files correspond to the four logical groups of data elements described in Section 2.3 - 2.6 and the indexes described in Section 2.7 as follows:

- Metathesaurus concept names and their sources (2.3) = MRCON, MRSO

- Attributes (2.5) = MRSAT, MRDEF, MRSTY

- Relationships between different concept names (2.4) = MRREL, MRCOC, MRATX, MRCXT

- Data about the Metathesaurus (2.6)=MRSAB, MRRANK, AMBIG.LUI, AMBIG.SUI, DELETED.CUI, MERGED.CUI, DELETED.LUI, MERGED.LUI, DELETED.SUI, MRCUI

- Indexes (2.7) = MRXW.BAQ, MRXW.DAN, MRXW.DUT, MRXW.ENG, MRXW.FIN, MRXW.FRE, MRXW.GER, MRXW.HEB, MRXW.HUN, MRXW.ITA, MRXW.NOR, MRXW.POR, MRXW.RUS,MRXW.SPA, MRXW.SWE, MRXNW.ENG, MRXNS.ENG

The AMBIG* files now provide a convenient way to identify all Metathesaurus terms and strings that have more than one meaning in Metathesaurus source vocabularies.

4.2 Columns and Rows

Each relation or named table of data values has by definition a fixed number of columns; the number of rows depends on the content of a particular version of the Metathesaurus.

A column is a sequence of all the values in a given data element or logical sub-element. In general, columns for longer variable length data elements will appear to the right of columns for shorter and/or fixed length data elements. The information for all columns in the ORF files is described on the Columns and Data Elements page of the current release documentation.

A row contains the values for one or more data elements or logical sub-elements for one Metathesaurus entry. Depending on the nature of the data elements involved, each Metathesaurus entry may have one or more rows in a given file. The values for the different data elements or logical sub-elements represented in the row are separated by vertical bars (|). If an optional element is blank, the vertical bars are still used to maintain the correct positioning of the subsequent elements. Each row is terminated by a vertical bar and line termination.

4.3 Descriptions of Each File

The descriptions of the files appear in the following order:

- Key data about the Metathesaurus: Files, columns or data elements
- Concept names and their vocabulary sources
- Attributes
- Relationships
- Other data about the Metathesaurus
- Indexes

4.3.1 Files (File = MRFILES)

There is exactly one row in this file for each physical segment of the files in the relational format. The columns or data elements in the file are as follows:

Col.	Description
FIL	Physical FILENAME
DES	Descriptive name
FMT	Comma separated list of COL, in order
CLS	# of COLUMNS
RWS	# of ROWS
BTS	Size in bytes in this format (ISO/PC or Unix)

Sample Records

MRATX|Associated Expressions|CUI,SAB,REL,ATX|4|7295|442571|
MRCOC|Co-occurring Concepts|CUI1,CUI2,SAB,COT,COF,COA|6|9061980|343331578|
MRCOLS|Attribute Relation|COL,DES,REF,MIN,AV,MAX,FIL, DTY|8|115|5728|

4.3.2 Data Elements (File = MRCOLS)

There is exactly one row in this file for each column or data element in each file in the relational format.

Col.	Description
COL	Column or data element name
DES	Descriptive name
REF	Documentation section number
MIN	Minimum length, characters
AV	Average length
MAX	Maximum length, characters
FIL	Physical FILENAME in which this field occurs
DTY	SQL-92 data type for this column

Sample Records

ATN|Attribute name||2|3.15|7|MRSAT|varchar(20)|
ATV|Attribute value||1|9.71|3634|MRSAT|varchar(4000)|
ATX|Associated expression||5|35.89|242|MRATX|varchar(300)|

4.3.3 Concept Names (File = MRCON)

There is exactly one row in this file for each meaning of each unique string in the Metathesaurus, i.e., there is exactly one row for each unique CUI-SUI combination in the Metathesaurus. Any difference in upper-lower case, word order, etc., creates a different unique string.

Col.	Description
CUI	Unique identifier for concept
LAT	Language of term
TS	Term status
LUI	Unique identifier for term
STT	String type
SUI	Unique identifier for string
STR	String
LRL	Least restriction level

Sample Records

C0002871|ENG|P|L0002871|PF|S0013742|Anemia|0|
C0002871|ENG|P|L0002871|VP|S0013787|Anemias|0|
C0002871|ENG|P|L0002871|VC|S0352787|ANEMIA|0|
C0002871|ENG|P|L0002871|VC|S0414880|anemia|0|
C0002871|ENG|P|L0002871|VO|S0470197|Anemia, NOS|3|
C0002871|ENG|S|L0280031|PF|S0803242|Anaemia|3|

4.3.4 Vocabulary Sources (File = MRSO)

This file contains the vocabulary source(s) for a concept, term, and string.

There is exactly one row in this file for each source of each string in the Metathesaurus. All Metathesaurus concepts have entries in this file.

Col.	Description
CUI	Unique identifier for concept
LUI	Unique identifier for term
SUI	Unique identifier for string
SAB	Abbreviated source name (SAB) for source vocabulary. Maximum field length is 20 alphanumeric characters. Two source abbreviations are assigned: • Root Source Abbreviation (RSAB) — short form, no version information, for example, AI/RHEUM, 1993, has an RSAB of "AIR" • Versioned Source Abbreviation (VSAB) — includes version information, for example, AI/RHEUM, 1993, has an VSAB of "AIR93" Official source names, RSABs, and VSABs are included on the Source Vocabularies page.
TTY	Abbreviation for term type in source vocabulary, for example PN (Metathesaurus Preferred Name) or CD (Clinical Drug). Possible values are listed on the Abbreviations Used in Data Elements page.
CODE	Unique identifier or code for string in that source
SRL	Source restriction level

Sample Records

C0002871|L0002871|S0013742|CCS|MD|4.1|0|
C0002871|L0002871|S0013742|ICPCPAE|PT|B82005|3|
C0002871|L0002871|S0013742|LCH|PT|U000235|0|
C0002871|L0002871|S0013742|MSH|MH|D000740|0|
C0002871|L0002871|S0013742|MTH|PT|U000161|0|
C0002871|L0002871|S0013742|MTH|PT|U000164|0|
C0002871|L0002871|S0013742|PSY|PT|02450|3|
C0002871|L0002871|S0013742|RCDAE|PT|XM05A|3|

The information in MRSO can be used in combination with MRCON to determine whether a particular concept, name, or code is present in a particular source, and in what form it appears.

Note: In the RRF, the concept name and vocabulary source information appear in a single file, MRCONSO.RRF.

4.3.5 Simple Concept and String Attributes (File = MRSAT)

There is exactly one row in this table for each concept, term and string attribute that does not have a sub-element structure. All Metathesaurus concepts have entries in this file.

Col.	Description
CUI	Unique identifier for concept
LUI	Unique identifier for term (optional)
SUI	Unique identifier for string (optional)
CODE	Unique identifier or code for entry in the source of the attribute, e.g., for all attributes derived from MeSH, the MeSH unique identifier (optional).
ATN	Attribute name. Possible values are all described in Columns and Data Elements page.

SAB	Abbreviated source name (SAB). Maximum field length is 20 alphanumeric characters. Two source abbreviations are assigned: • Root Source Abbreviation (RSAB) — short form, no version information, for example, AI/RHEUM, 1993, has an RSAB of "AIR" • Versioned Source Abbreviation (VSAB) — includes version information, for example, AI/RHEUM, 1993, has an VSAB of "AIR93" Official source names, RSABs, and VSABs are included on the Source Vocabularies page.
ATV	Attribute value described under specific attribute name in Columns and Data Elements page. A few attribute values exceed 1,000 characters.

Sample Records

C0002871|L0002871|S0013742|D000740|MMR|MSH|19960610|
C0002871|L0002871|S0013742|D000740|MN|MSH|C15.378.71|
C0002871|L0002871|S0013742|D000740|TH|MSH|POPLINE (1994)|
C0002871|L0002871|S0414880|208/04453|SOS|PDQ|secondary related condition|
C0002871|L0002871|S0470197|DC-10010|SIC|SNMI|285.9|

4.3.6 Definitions (File = MRDEF)

There is exactly one row in this file for each definition in the Metathesaurus. A few definitions approach 3,000 characters in length.

Col.	Description
CUI	Unique identifier for concept
SAB	Abbreviated source name (SAB) of the source of the definition. Maximum field length is 20 alphanumeric characters. Two source abbreviations are assigned: • Root Source Abbreviation (RSAB) — short form, no version information, for example, AI/RHEUM, 1993, has an RSAB of "AIR" • Versioned Source Abbreviation (VSAB) — includes version information, for example, AI/RHEUM, 1993, has an VSAB of "AIR93" Official source names, RSABs, and VSABs are included on the Source Vocabularies page.
DEF	Definition

Sample Records

C0002871|MSH|A reduction in the number of circulating erythrocytes or in the quantity of hemoglobin.|

4.3.7 Semantic Types (File = MRSTY)

There is exactly one row in this file for each semantic type assigned to each concept. All Metathesaurus concepts have at least one entry in this file. Many have more than one entry.

Col.	Description
CUI	Unique identifier of concept
TUI	Unique identifier of Semantic Type
STY	Semantic Type. The valid values are defined in the Semantic Network.

Sample Record

C0002871|T047|Disease or Syndrome|

4.3.8 Locators (File = MRLO)

This file has been deleted from the Metathesaurus effective with the 2004AB release. Some of the information was outdated, some duplicated information contained in other Metathesaurus files, and some was easily obtained from other publicly available sources, e.g., PubMed.

4.3.9 Related Concepts (File = MRREL)

There is one row in this table for each relationship between Metathesaurus concepts known to the Metathesaurus, with the following exceptions found in other files: co-occurrences found in MRCOC and Associated Expressions found in MRATX.

Note that for asymmetrical relationships there is one row for each direction of the relationship. Note also the direction of REL - the relationship which the SECOND concept (with Concept Unique Identifier CUI2) HAS TO the FIRST concept (with Concept Unique Identifier CUI1).

Col.	Description
CUI1	Unique identifier of first concept
REL	Relationship of SECOND to first concept
CUI2	Unique identifier of second concept
RELA	Relationship attribute
SAB	Abbreviated source name (SAB) of the source of relationship. Maximum field length is 20 alphanumeric characters. Two source abbreviations are assigned: • Root Source Abbreviation (RSAB) — short form, no version information, for example, AI/RHEUM, 1993, has an RSAB of "AIR" • Versioned Source Abbreviation (VSAB) — includes version information, for example, AI/RHEUM, 1993, has an VSAB of "AIR93" Official source names, RSABs, and VSABs are included on the Source Vocabularies page.
SL	Source of relationship labels
MG	Machine-generated and unverified indicator (optional). G indicates 'machine generated'

Sample Records

C0002871|CHD|C0002891|isa|MSH|MSH||

Anemia, Neonatal (C0002891)

has CHILD REL and isa RELA

to Anemia (C0002871)

C0002871|RB|C0221016||MTH|MTH||

[Red blood cell disorder, NOS (C0221016)

has broader REL

to Anemia (C0002871)]

C0002871|RL|C0002886|mapped_to|SNMI|SNMI||

[Anemia, Macrocytic (C0002886)

has like relationship

to Anemia (C0002871)]

C0002871|RO|C0002886|clinically_associated_with|CCPSS|CCPSS||

[Megaloblastic anemia due to folate deficiency, NOS (C0151482)

has clinically_associated_with relationship

to Anemia (C0002871)]

4.3.10 Co-occurring Concepts (File = MRCOC)

There are two rows in this table for each pair of concepts that co-occur in each information source represented one for each direction of the relationship. (Note that the COA data may be different for each direction of the relationship). Many Metathesaurus concepts have no entries in this file. Due to the very large number of co-occurrence relationships, they are distributed in a separate file.

Col.	Description
CUI1	Unique identifier of first concept
CUI2	Unique identifier of second concept Note: Where COT is MeSH topical qualifier (LQ) and CUI2 is not present, the count of citations of CUI1 with no MeSH qualifiers is reported.
SOC	Abbreviation of the source of co-occurrence information if applicable
COT	Type of co-occurrence
COF	Frequency of co-occurrence, if applicable
COA	Attributes of co-occurrence, if applicable

Sample Records

C0002871||MED|LQ|1||
C0002871|C0000530|MBD|L|2|CI=1,EN=1,ME=1,PA=1|
C0002871|C0000727|MBD|L|1|BL=1,ET=1|
C0002871|C0000737|MBD|L|1|ET=1|
C0002871|C0000772|MBD|L|2|CN=2|

Co-occurrences are concepts that occur together in the same entries in some information source. The relationships represented here are obtained from machine-manipulation of the information source. Co-occurrence relationships may exist between similar concepts (e.g., Atrial Fibrillation and Arrhythmia) or between very different concepts that nevertheless have some important connection in the field of biomedicine (e.g., Atrial Fibrillation and Digoxin), or between a primary concept and a qualifier (e.g., Lithotripsy and instrumentation). A co-occurrence relationship can exist between two concepts that have no other apparent relationship, although the frequency of such co-occurrences will be small.

In the current Metathesaurus, there are three sources of co-occurrence data: MEDLINE, AI/ RHEUM, and CCPSS. From MEDLINE, co-occurrence data was computed for concepts that were designated as principal or main points in the same journal article i.e., the co-occurrence counts do not include articles in which either or both of the concepts were present and indexed in MEDLINE but not designated as main points. (A concept is considered to be a main point if the * is attached to the main heading or any of its subheadings.)

Two overall frequencies of MEDLINE co-occurrence are provided: one for recent MEDLINE data (MED) and one for MEDLINE data from a preceeding block of years (MBD). Separate counts are provided for the frequencies with which the first concept was qualified by different MeSH qualifiers or by no qualifier at all when it co-occurred with the second concept. There are separate entries for each direction of the co-occurrence relationship. The related subheading occurrence information in each entry belongs to the first concept in the entry and is therefore different for each direction of the relationship.

In addition to the specific qualifier information associated with two co-occurring concepts, in entries with LQ and LQB values for type of co-occurrence, this element also includes totals for the number of times each main concept was qualified by a specific subheading or by no subheading.

The AI/RHEUM co-occurrence data represent the co-occurrence of diseases and findings in the AI/RHEUM knowledge base, i.e., the diseases that co-occur with a particular finding and the findings that co-occur with a particular disease. Each disease/finding pair can co-occur only once in the AI/RHEUM knowledge base.

In CCPSS, the co-occurrence data is extracted from patient records and includes problem-problem co-occurrences within a patient record as well as problem-modifier co-occurrences.

4.3.11 Concept contexts (File = MRCXT)

This file is no longer distributed. To create the MRCXT file (Table 1), use the new MRCXT Builder application, accessible from the MetamorphoSys Welcome screen. Information on the MRCXT Builder can be found at http://www.nlm.nih.gov/research/umls/ implementation_resources/metamorphosys/MRCXT_Builder.html. The information below describes the content of the file when produced by the MRCXT Builder.

There are rows in this file for each occurrence of a concept in a hierarchy in any of the UMLS source vocabularies - a "context" in this discussion. Many Metathesaurus concepts have multiple contexts while others may have none. The number of rows per context differs depending on the number of ancestor, sibling, or child terms the concept has in that context. Because some concepts have multiple contexts in the same source (e.g., MeSH), a context number (CXN - e.g., 1, 2, 3) is used to identify all members of the same context. The CXNs are not global but are created as required for each concept. Since some concepts have multiple contexts in the same vocabulary with the same SUI, each distinct context can be retrieved with a CUI-SUI-SAB-CXN key. The "distance-1 relationships" i.e., the immediate parent, immediate child, and sibling relationships, represented in this file are also present in the MRREL file.

Sample Records

```
C0002871|S0013742|MSH|D000740|1|ANC|1|MeSH|C0220876||||
 C0002871|S0013742|MSH|D000740|1|ANC|2|Diseases (MeSH Category)|C0012674|C|||
 C0002871|S0013742|MSH|D000740|1|ANC|3|Hemic and Lymphatic Diseases|C0018981|
C15|||
 C0002871|S0013742|MSH|D000740|1|ANC|4|Hematologic Diseases|C0018939|C15.378|
isa||
 C0002871|S0013742|MSH|D000740|1|CCP||Anemia|C0002871|C15.378.71|isa|+|
 C0002871|S0013742|MSH|D000740|1|CHD||Anemia, Aplastic|C0002874|C15.378.71.85|isa|
+|
 C0002871|S0013742|MSH|D000740|1|SIB||Blood Protein Disorders|C0005830|C15.378.147|
isa|+|
```

C0002871|S0013742|MSH|D000740|1|CHD||Anemia, Hemolytic|C0002878|
C15.378.71.141|isa|+|

4.3.12 Associated Expressions (File = MRATX)

There is one row in this table for each vocabulary expression (i.e., combination of terms from a specific Metathesaurus source vocabulary) identified as having a relationship to a concept in the Metathesaurus. The majority of Metathesaurus entries have no entries in this table.

Col.	Description
CUI	Unique identifier of concept to which the expression is related
SAB	Abbreviated source name (SAB) of source of terms in expression. Maximum field length is 20 alphanumeric characters. Two source abbreviations are assigned: • Root Source Abbreviation (RSAB) — short form, no version information, for example, AI/RHEUM, 1993, has an RSAB of "AIR" • Versioned Source Abbreviation (VSAB) — includes version information, for example, AI/RHEUM, 1993, has an VSAB of "AIR93" Official source names, RSABs, and VSABs are included on the Source Vocabularies page.
REL	Relationship of meaning of expression to main concept
ATX	Associated expression

Sample Records

C0001207|MSH|S|<Acromegaly> AND <Gigantism>|
C0001296|LCH|U|<Insurance>/<Statistics>|
C0001355|MSH|S|<Kidney Failure, Acute> AND <Kidney Papillary Necrosis>|

4.3.13 Source Information (File = MRSAB)

The Metathesaurus has "versionless" or "root" Source Abbreviations (SABs) in the data files. MRSAB (Table 2) connects the root SAB to fully specified version information for the current release. For example, the released SAB for MeSH is now simply "MSH". In MRSAB, you will find the current versioned SAB, e.g., MSH2003_2002_10_24. MetamorphoSys can produce files with either the root or versioned SABs so that either form can be utilized by a user.

There is one row in this file for every version of every source in the current Metathesaurus; when complete, there will also be historical information with a row for each version of each source that has appeared in any Metathesaurus release. Note that the field CURVER has the value Y to identify the version in this Metathesaurus release. Future releases of MRSAB will also contain historical version information in rows with CURVER value N.

MRSAB allows all other Metathesaurus files to use versionless source abbreviations, so that rows with no data change between versions also remain unchanged.

Sources with contexts have "full" contexts, i.e., all levels of terms may have Ancestors, Parents, Children and Siblings. A full context may also be further designated as Multiple, Nosib (No siblings) or both Multiple and Nosib.

Multiple indicates that a single concept in this source may have multiple hierarchical positions.

No siblings (Nosib) indicates that siblings have not been computed for this source.

The Source Vocabularies page of the current release documentation lists each source in the Metathesaurus and includes information about the type of context, if any, for each source.

Sample Record

C1371270|C1140284|RXNORM_04AB|RXNORM|RXNORM Project, META2004AB |
RXNORM | 04AB | 2004_05_17 | | 2004AB | | Stuart Nelson, M.D., Head, MeSH Section; e-
mail: nelson@nlm.nih.gov | Stuart Nelson, M.D., Head, MeSH Section; e-mail:
nelson@nlm.nih.gov | 0 138005 | 110403 | |
BN,IN,OBD,OCD,SBD,SBDF,SCD,SCDC,SCDF,SY | ORIG_CODE,ORIG_SOURCE |
ENG | UTF-8| Y | Y |

4.3.14 Concept Name Ranking (File = MRRANK)

There is exactly one row for each concept name type from each Metathesaurus source
vocabulary (each SAB-TTY combination). The RANK and SUPPRESS values in the
distributed file are those used in Metathesaurus production. Users are free to change these
values to suit their needs and preferences, then change the naming precedence and
suppressibility (TS in MRCON) by using MetamorphoSys to create a customized
Metathesaurus.

Col.	Description
RANK	Numeric order of precedence, higher value wins
SAB	Abbreviated source name (SAB). Maximum field length is 20 alphanumeric characters. Two source abbreviations are assigned: • Root Source Abbreviation (RSAB) — short form, no version information, for example, AI/RHEUM, 1993, has an RSAB of "AIR" • Versioned Source Abbreviation (VSAB) — includes version information, for example, AI/RHEUM, 1993, has an VSAB of "AIR93" Official source names, RSABs, and VSABs are included on the Source Vocabularies page.
TTY	Abbreviation for term type in source vocabulary, for example PN (Metathesaurus Preferred Name) or CD (Clinical Drug). Possible values are listed on the Abbreviations Used in Data Elements page.
SUPPRESS	Flag indicating that this SAB and TTY will create a TS=s MRCON entry; see TS

Sample Records

0210|AIR|SY|N|
0209|ULT|PT|N|
0208|CPT|PT|N|

4.3.15 Ambiguous Term Identifiers (File = AMBIG.LUI)

In the instance that a Lexical Unique Identifier (LUI) is linked to multiple Concept Unique
Identifiers (CUIs), there is one row in this table for each LUI-CUIs pair. This file identifies
those lexical variant classes which have multiple meanings in the Metathesaurus.

In the Metathesaurus, the LUI links all strings within the English language that are identified
as lexical variants of each other by the luinorm program found in the UMLS SPECIALIST
Lexicon and Lexical Tools. LUIs are assigned irrespective of the meaning of each string. This
table may be useful to system developers who wish to make use of the lexical programs in their
applications to identify and disambiguate ambiguous terms.

Col.	Description
LUI	Lexical Unique Identifier
CUI	Concept Unique Identifier

Sample Records

L0000003|C0010504|
L0000003|C0917995|
L0000032|C0010206|

4.3.16 Ambiguous String Identifiers (File = AMBIG.SUI)

In the instance that a String Unique Identifier (SUI) is linked to multiple Concept Unique Identifiers (CUIs), there is one row in this table for each SUI-CUIs pair.

This file resides in the META directory. In the Metathesaurus, there is only one SUI for each unique string within each language, even if the string has multiple meanings. This table is only of interest to system developers who make use of the SUI in their applications or in local data files.

Col.	Description
SUI	String Unique Identifier
CUI	Concept Unique Identifier

Sample Records

S0063890|C0026667|
S0063890|C1135584|
S0229413|C0008802|

4.3.17 Metathesaurus Change Files

There are six files or relations that identify key differences between entries in the previous and the current edition of the Metathesaurus. Developers can use these special files to determine whether there have been changes that affect their applications.

The usefulness of individual files will depend on how data from the Metathesaurus have been linked or incorporated in a particular application.

Each relation or named table of data has a fixed number of columns and variable number of rows. A column is a sequence of all the values in a given data element. A row contains the values for two or more data elements for one entry. The values for the different data elements in the row are separated by vertical bars (|). Each row ends with a vertical bar and line termination.

4.3.17.1 Deleted Concepts (File = DELETED.CUI)

Concepts whose meaning is no longer present in the Metathesaurus are reported in this file. There is a row for each concept that existed in the previous release and is not present in the current release. If the meaning exists in the current release, i.e., the missing concept was merged with another current concept, it is reported in the MERGEDCUI file (Section 3.3.19.2) and not in this file.

Col.	Description
CUI	Concept unique identifier in the previous Metathesaurus
STR	Preferred name of this concept in the previous Metathesaurus

4.3.17.2 Merged Concepts (File = MERGED.CUI)

There is exactly one row in this table for each released concept in the previous Metathesaurus (CUI1) that was merged into another released concept from the previous Metathesaurus (CUI2). When this merge occurs, the first CUI (CUI1) was retired; this table shows the CUI (CUI2) for the merged concept in this Metathesaurus.

Entries in this file represent concepts pairs that were considered to have different meanings in the previous edition, but which are now identified as synonyms

Col.	Description
CUI1	Concept unique identifier in the previous Metathesaurus
CUI2	Concept unique identifier in this Metathesaurus in format C#######

4.3.17.3 Deleted Terms (File = DELETED.LUI)

There is exactly one row in this table for each Lexical Unique Identifier (LUI) that appeared in the previous version of the Metathesaurus, but does not appear in this version.

LUIs are assigned by the luinorm program, part of the lvg program in the UMLS SPECIALIST Lexicon and Lexical Tools.

These entries represent the cases where LUIs identified by the previous release's luinorm program, when used to identify lexical variants in the previous Metathesaurus, are no longer found with this release's luinorm on this release's Metathesaurus. This does not necessarily imply the deletion of a string or a concept from the Metathesaurus.

Col.	Description
LUI	Concept unique identifier in the previous Metathesaurus
STR	Preferred name of Term in the previous Metathesaurus

4.3.17.4 Merged Terms (File = MERGED.LUI)

There is exactly one row in this file for each case in which strings had different LUIs in the previous Metathesaurus yet share the same LUI in this Metathesaurus; a LUI present in the previous Metathesaurus is therefore absent from this Metathesaurus.

LUIs are assigned by the luinorm program, part of the lvg program in the UMLS SPECIALIST Lexicon and Lexical Tools.

These entries represent the cases where separate lexical variants as identified by the previous release's luinorm program version are a single lexical variant as identified by this release's luinorm.

Col.	Description
LUI1	Lexical unique identifier in the previous Metathesaurus but not present in this Metathesaurus
LUI2	Lexical unique identifier into which it was merged in this Metathesaurus

4.3.17.5 Deleted Strings (File = DELETED.SUI)

There is exactly one row in this file for each string in each language that was present in an entry in the previous Metathesaurus and does not appear in this Metathesaurus.

Note that this does not necessarily imply the deletion of a term (LUI) or a concept (CUI) from the Metathesaurus. A string deleted in one language may still appear in the Metathesaurus in another language.

Col.	Description
SUI	String unique identifier in the previous Metathesaurus that is not present in this Metathesaurus
LAT	Three-character abbreviation of language of string that has been deleted
STR	Preferred name of term in the previous Metathesaurus that is not present in this Metathesaurus

4.3.17.6 Retired CUI Mapping (File = MRCUI)

There are one or more rows in this file for each Concept Unique Identifier (CUI) that existed in any prior release but is not present in the current release. The file includes mappings to current CUIs as synonymous or to one or more related current CUI where possible. If a synonymous mapping cannot be found, other relationships between the CUIs can be created. These relationships can be Broader (RB), Narrower (RN), Other Related (RO), Deleted (DEL) or Removed from Subset (SUBX). Rows with the SUBX relationship are added to MRCUI by MetamorphoSys for each CUI that met the exclusion criteria and was consequently removed from the subset. Some CUIs may be mapped to more than one other CUI using these relationships.

CUIs may be retired when (1) two released concepts are found to be synonyms and so are merged, retiring one CUI; (2) the concept no longer appears in any source vocabulary and is not 'rescued' by NLM; or (3) the concept is an acknowledged error in a source vocabulary or determined to be a Metathesaurus production error.

See the META/CHANGE files, especially MERGED.CUI and DELETED.CUI, for the changes from the last release only, without mappings.

Col.	Description
CUI1	Retired CUI - was present in some prior release, but is currently missing
VER	The last release version in which CUI1 was a valid CUI
CREL	The relationship CUI2 has to CUI1, if present, or DEL if CUI2 is not present. Valid values currently are SY, DEL, RO, RN, RB.
CUI2	The current CUI that CUI1 most closely maps to
MAPIN	Is this map in current subset? Values of Y, N, or null. MetamorphoSys generates the Y or N to indicate whether the CUI2 concept is or is not present in the subset. The null value is for rows where the CUI1 was not present to begin with (i.e., REL=DEL).

Sample Records

```
C0079138|2001AA|DEL||Y|
C0079138|2001AA|RO|C0037440|Y|
C0079151|1993AA|DEL||N|
C0079158|1997AA|SY|C0009081||
C0079167|1997AA|SY|C0010042|N|
```

4.3.18 Word Index (File = MRXW.BAQ, MRXW.DAN, MRXW.DUT, MRXW.ENG, MRXW.FIN, MRXW.FRE, MRXW.GER, MRXW.HEB, MRXW.HUN, MRXW.ITA, MRXW.NOR, MRXW.POR, MRXW.RUS, MRXW.SPA, MRXW.SWE)

There is one row in these tables for each word found in each unique Metathesaurus string (ignoring upper-lower case). All Metathesaurus entries have entries in the word index. The entries are sorted in ASCII order.

Col.	Description
LAT	Abbreviation of language of the string in which the word appears
WD	Word in lowercase
CUI	Concept identifier
LUI	Term identifier
SUI	String identifier

Sample Records from MRXW.ENG

ENG|anaemia|C0002871|L0280031|S0352688|
ENG|anemia|C0002871|L0002871|S0013742|
ENG|anemias|C0002871|L0002871|S0013787|
ENG|blood|C0002871|L0376533|S0500659|

Sample Records from MRXW.FRE

FRE|ANEMIE|C0002871|L0162748|S0227229|

4.3.19 Normalized Word Index (File = MRXNW.ENG)

There is one row in this table for each normalized word found in each unique English-language Metathesaurus string. All English-language Metathesaurus entries have entries in the normalized word index. There are no normalized string indexes for other languages in this edition of the Metathesaurus.

Col.	Description
LAT	Abbreviation of language of the string in which the word appears (always ENG in this edition of the Metathesaurus)
NWD	Normalized word in lowercase (described in Section 2.7.2.1)
CUI	Concept identifier
LUI	Term identifier
SUI	String identifier

Sample Records

ENG|anaemia|C0002871|L0280031|S0352688|
ENG|anemia|C0002871|L0002871|S0013742|
ENG|anemia|C0002871|L0002871|S0013787|
ENG|blood|C0002871|L0376533|S0500659|

4.3.20 Normalized String Index (File = MRXNS.ENG)

There is one row in this table for each normalized string found in each unique English-language Metathesaurus string (ignoring upper-lower case). All English-language Metathesaurus entries have entries in the normalized string index. There are no normalized word indexes for other languages in this edition of the Metathesaurus.

UMLS® Reference Manual

UMLS® Reference Manual

UMLS® Reference Manual

Col.	Description
LAT	Abbreviation of language of the string (always ENG in this edition of the Metathesaurus)
NSTR	Normalized string in lowercase (described in Section 2.7.3.1)
CUI	Concept identifier
LUI	Term identifier
SUI	String identifier

Sample Records

ENG|anaemia|C0002871|L0280031|S0352688|
ENG|anaemia unspecified|C0002871|L0696700|S0803315|
ENG|anemia|C0002871|L0002871|S0013787|

Table 1. Concept contexts (File = MRCXT)

Col.	Description
CUI	Unique identifier of concept
SUI	Unique identifier of string used in this context
SAB	Abbreviated source name (SAB). Maximum field length is 20 alphanumeric characters. Two source abbreviations are assigned: • Root Source Abbreviation (RSAB) — short form, no version information, for example, AI/RHEUM, 1993, has an RSAB of "AIR" • Versioned Source Abbreviation (VSAB) — includes version information, for example, AI/RHEUM, 1993, has an VSAB of "AIR93" Official source names, RSABs, and VSABs are included on the Source Vocabularies page.
CODE	Unique identifier or code for string in that source.
CXN	The context number (to distinguish multiple contexts in the same source with the same SUI)
CXL	Context member label, i.e., ANC for ancestor of this concept, CCP for concept, SIB for sibling of this concept, CHD for child of this concept
RNK	For rows with a CXL value of ANC, the rank of the ancestors (e.g., a value of 1 denotes the most remote ancestor in the hierarchy)
CXS	String for context member
CUI2	Unique concept identifier of context member (may be empty if context member is not yet in the Metathesaurus)
HCD	Hierarchical number or code of context member in this source (optional)
RELA	Relationship attribute providing further categorization of the CXL, if applicable and known. Allowed values are listed on the Abbreviations Used in Data Elements page.
XC	A plus (+) sign indicates that the CUI2 for this row has children in this context. If this field is empty, the CUI2 does not have children in this context.

UMLS® Reference Manual

Table 2. Source Information (File = MRSAB)

Field	Full Name	Description
VCUI	CUI	CUI of the versioned SRC concept for a source
RCUI	Root CUI	CUI of the root SRC concept for a source
VSAB	Versioned Source Abbreviation	The versioned source abbreviation for a source e.g., MSH2003_2002_10_24
RSAB	Root Source Abbreviation	The root source abbreviation for a source e.g., MSH
SON	Official Name	The official name for a source
SF	Source Family	The source family for a source
SVER	Version	The source version e.g., 2001
VSTART	Valid Start Date For A Source	The date a source became active, e.g., 2004_04_03
VEND	Valid End Date For A Source	The date a source ceased to be active, e.g., 2003_05_10
IMETA	Meta Insert Version	The version of the Metathesaurus in which a source first appeared, e.g., 2001AB
RMETA	Meta Remove Version	The version of the Metathesaurus in which a source last appeared, e.g., 2001AC
SLC	Source License Contact	The source license contact information
SCC	Source Content Contact	The source content contact information
SRL	Source Restriction Level	0, 1, 2, 3, 4 – explained in the License Agreement
TFR	Term Frequency	The number of terms for this source in MRCON/MRSO, e.g., 12343
CFR	CUI Frequency	The number of CUIs associated with this source, e.g., 10234
CXTY	Context Type	The type of contexts for this source. Values are FULL, FULL-MULTIPLE, FULL-NOSIB, FULL-NOSIB-MULTIPLE, FULL-MULTIPLE-NOSIB-RELA, null.
TTYL	Term Type List	Term type list from source, e.g., MH, EN, PM, TQ
ATNL	Attribute Name List	The attribute name list (from MRSAT), e.g., MUI, RN, TH
LAT	Language	The language of the source
CENC	Character Encoding	All UMLS content is provided in Unicode, encoded in UTF-8. MetamorphoSys will allow exclusion of extended characters with some loss of information. Transliteration to other character encodings is possible but not supported buy NLM; for further information, see http://www.unicode.org.
CURVER	Current Version	A Y or N flag indicating whether or not this row corresponds to the current version of the named source
SABIN	Source in Subset	A Y or N flag indicating whether or not this row is represented in the current MetamorphoSys subset. Initially always Y where CURVER is Y, but later is recomputed by MetamorphoSys.