**Representation of inter-source mappings in the UMLS Metathesaurus**

1. **Scope of this document**
2. **Mapping overview**
3. **Representation of mappings in the UMLS Metathesaurus**
   1. **MRCONSO.RRF**
   2. **MRSTY.RRF**
   3. **MRSAT.RRF**
   4. **MRMAP.RRF**
   5. **MRSMAP.RRF**
   6. **MRREL.RRF**
4. **Additional Resources**
5. **Scope of this document:**

This document explains the representation of mappings in MRMAP.RRF and MRSMAP.RRF to enable users to make informed decisions about utilizing this data. Information is also provided on “XM” concepts created to represent map sets, as well as details on attributes attached to these concepts in MRSAT.RRF, and mapping relationships included in MRREL.RRF.

Source-specific details are available on the individual source pages available <here>.

1. **Mapping overview**

Mappings provide links from entities in one terminology (the source terminology) to entities in another terminology (the target terminology). Entities may be terms, concepts, descriptors or expressions. Mappings may be used for a variety of purposes, including:

* using data collected for one purpose for another purpose (e.g. translating clinical information coded with SNOMEDCT to ICD-9-CM for reimbursement purposes)
* retaining the value of data when migrating to newer terminology requirements (e.g. updating from ICD-9-CM to ICD-10-CM)

Given the diversity of mapping applications, it is important to understand that a single, simple mapping between two terminologies may not be relevant for all tasks.

Mappings in the UMLS Metathesaurus may be implied by concept synonymy (i.e. sharing a CUI in MRCONSO.RRF), or by relationships in MRREL.RRF. In some cases, “mapping” or “cross-reference” information is available in MRSAT.RRF (e.g. in MedDRA the MXR attribute contains mappings to COSTART, HARTS, ICD-9, ICD-9-CM. J-ART, and WHOART. See also documentation on hidden sources.

However, some sources explicitly provide inter-terminology mapping data. In these cases, the information is represented in MRMAP.RRF and MRSMAP.RRF, with auxiliary data in MRCONSO.RRF and MRSAT.RRF.

1. **Representation of mappings in the UMLS Metathesaurus:**

Inter-source mapping data is represented using the following specifications (there may be exceptions, e.g. for map sets that have not been updated recently):

1. **MRCONSO.RRF:** For each map set represented in MRMAP.RRF, there is a single “Cross mapping set” concept in MRCONSO.RRF. Note that the CUI changes when the map set is updated from one version to the next.

Field values are assigned as follows:

* + SAB: the source that asserts the mapping information. For example, ICD10PCS provides three kinds of mappings:
    1. ICD9CM\_2009 to ICD10PCS\_2009 Mappings (GEMs)
    2. ICD10PCS\_2009 to ICD9CM\_2009 Mappings (GEMs)
    3. ICD10PCS\_2009 to ICD9CM\_2009 Mappings (Reimbursement)

The SAB for all of these map set atoms is “ICD10PCS.”

* + TTY: XM for all map set atoms
  + STR: The atom name is created as “<VSAB> to <VSAB> Mappings <optional additional information>” For example:
    1. SNOMEDCT\_2010\_01\_31 to ICD9CM\_2010 Mappings
    2. ICD10PCS\_2009 to ICD9CM\_2009 Mappings (Reimbursement)
  + CODE: If an appropriate identifier for the map set is available from the source, it will be used as the CODE. SAUI, SCUI and SDUI may also be populated. If no source-asserted identifier is available, a CODE beginning with “MTHU” will be generated during Metathesaurus production.

Example:

C2733618|ENG|P|L9072757|PF|S11318505|Y|A17317295|100046|||SNOMEDCT|XM|100046|SNOMEDCT\_2010\_01\_31 to ICD9CM\_2010 Mappings|9|N||

1. **MRSTY.RRF:** All map set concepts are assigned an STY of “Intellectual Product”.

Example:

C2733618|T170|A2.4|Intellectual Product|AT121455255||

1. **MRSAT.RRF:** Every map set concept has numerous attributes in MRSAT.RRF which provide additional details. The following attributes can be found in MRSAT.RRF. The attributes are attached using STYPE=CODE.

**Required Attributes:**

|  |  |  |  |
| --- | --- | --- | --- |
| **ATN** | **Description** | **Valid Values, if applicable** | **Examples** |
| FROMRSAB | Root source abbreviation for the "from" identifiers of a map set | range=MRSAB.RSAB | SNOMEDCT |
| FROMVSAB | Versioned source abbreviation for the "from" identifiers of a map set | range=MRSAB.VSAB | SNOMEDCT\_2010\_01\_31 |
| MAPSETRSAB | Root source abbreviation for a map set. In general, the same as the value for FROMRSAB. | range=MRSAB.RSAB | SNOMEDCT |
| MAPSETVERSION | Version of the map set. |  | 20100131 |
| MAPSETVSAB | Versioned source abbreviation for the provider of a map set | range=MRSAB.VSAB | SNOMEDCT\_2009\_01\_31 |
| TORSAB | Root source abbreviation for the "to" identifiers of a map set | range=MRSAB.RSAB | ICD9CM |
| TOVSAB | Versioned source abbreviation for the "to" identifiers of a map set | range=MRSAB.VSAB | ICD9CM\_2010 |

**Optional Attributes:** In general, these attributes are extracted directly from source provided data and may have a diverse range of values and formats. ATNs for attributes created during Metathesaurus source processing begin with “MTH\_”.

|  |  |  |
| --- | --- | --- |
| **ATN** | **Description** | **Valid values and/or examples** |
| MAPSETNAME | Official name of a map set | ICD-9-CM |
| MAPSETREALMID | Identifier of a "Realm" to which a source is mapped, within which this cross mapping table is applicable. Used in cases where Realm specific business rules or guidelines alter the acceptable mappings. Realm is the same as used in SNOMED CT subsets. It includes a four character ISO6523 identifier followed by an optional series of concatenated subdivision codes defined by the registered organization. | ??? |
| MAPSETRULETYPE | Indicates the types of rules used in a map set and cross map targets to which a source is mapped. | ??? |
| MAPSETSCHEMEID | Standard identifier for the scheme to which a map set belongs. This may be an International Coding Scheme Identifier (ISO7826) or an Object Identifier (OID) used as specified by HL7. | 2.16.840.1.113883.6.5.2.1 |
| MAPSETSCHEMENAME | Full name of the target scheme in a map set. | International Classification of Diseases and Related Health Problems, 9th Revision, Clinical Modifications. |
| MAPSETSCHEMEVERSION | Version number of the target scheme (as published by the issuing organization) in a map set. | 2009 |
| MAPSETSEPARATORCODE | XML entity code (for example, "&#x7c;" to represent the vertical-bar character) for the character used as a separator between the individual codes in the target codes field in a map set. | &#x7C; |
| MAPSETSID | Source asserted identifier for a map set. If present, matches the CODE in MRCONSO.RRF. | 100046 |
| MAPSETTYPE | Indicates the nature of a map set. Its value is map set specific. It can be used to indicate the inclusion of one to one, one to many and choices of maps. | 2  ICD-10-PCS to ICD-9-CM Mappings |
| MAPSETXRTARGETID | Map set target identifier used for XR mappings. Only used for map sets that explicitly map source codes to “nothing.” | 100051  NoPCS |
| MTH\_UMLSMAPSETSEPARATOR | The character used in the UMLS Metathesaurus as a separator between the individual codes in the target codes field of the cross map targets to which a source is mapped. | , |
| MTH\_MAPFROMCOMPLEXITY | Two-part value indicating the complexity of "from" expressions used in a map set. Valid values can be combined in a comma-separated list | Valid values:  Part 1  SINGLE  LIST  BOOLEAN\_EXPRESSION  Part 2  AUI  CODE  CUI  LUI  SAUI  SCUI  SDUI  SUI  STR  Example:  SINGLE CODE |
| MTH\_MAPFROMEXHAUSTIVE | Indicates whether or not the "from" source of a map set is completely mapped | Valid values:  Y/N |
| MTH\_MAPSETCOMPLEXITY | Indicates the overall complexity of a map set.  This field is computed in the following manner:   1. Compute FROMEXPR cardinality (left hand side) based on whether >1 FROMEXPR exists for same TOEXPR OR MTH\_MAPTOCOMPLEXITY indicates MULTIPLE. 2. Compute TOEXPR cardinality (right hand side) based on whether >1 TOEXPR exists for same FROMEXPR OR MTH\_MAPFROMCOMPLEXITY indicates MULTIPLE. 3. RULE\_BASED if >1 non-null distinct MAPSUBSETID | Valid values:  N\_TO\_N  N\_TO\_ONE  ONE\_TO\_N  ONE\_TO\_ONE  RULE\_BASED |
| MTH\_MAPTOCOMPLEXITY | Two-part value indicating the complexity of "to" expressions used in a map set. Valid values can be combined in a comma-separated list | Valid values:  Part 1  SINGLE  LIST  BOOLEAN\_EXPRESSION  Part 2  AUI  CODE  CUI  LUI  SAUI  SCUI  SDUI  SUI  STR  Example:  SINGLE CODE |
| MTH\_MAPTOEXHAUSTIVE | Indicates whether or not the "to" source is completely mapped | Valid values: Y/N |
| SOS | Scope statement | Example: This set maps ICD10 codes to ICPC2EENG codes. The data was extracted from the ICPC to ICD10 Thesaurus. |
| TARGETSCHEMEID | Identifier for the target scheme in the map set. This may be an International Coding Scheme Identifier (ISO7826) or an Object Identifier (OID) used as specified by HL7. | Example:  2.16.840.1.113883.6.5.2.1 |
| MAPSETGRAMMAR | Grammar used by expressions in FROMEXPR or TOEXPR fields | ATX ::= expr; expr ::= disj; disj ::= conj , conj "OR" conj; conj ::= unary , unary "AND" unary; unary ::= neg , pos; neg ::= "NOT" pos; pos ::= "(" expr ")" , slash , atom; slash ::= atom , atom "/" atom; atom ::= "<" (any non ">" character) ">"; |

1. **MRMAP.RRF:** MRMAP.RRF contains information on what entities are mapped to what and the source responsible for the mapping. MRMAP.RRF represents complex information, including source-asserted rules and other potentially powerful data. Note that mapping expressions are not resolved in the MRMAP and MRSMAP representation: the original expression as designated by the source is included, it is not converted to UMLS identifiers such as CUIs or AUIs.

|  |  |  |
| --- | --- | --- |
|  |  | **MRMAP.RRF** |
| **#** | **Name** | **Description** |
| 1 | MAPSETCUI | Unique identifier for the UMLS concept which represents the whole map set. |
| 2 | MAPSETSAB | Source abbreviation (SAB) for the provider of the map set. |
| 3 | MAPSUBSETID | Map subset identifier used to identify a subset of related mappings within a map set. This is used for cases where the FROMEXPR may have more than one potential mapping. There is an implied “OR” between multiple lines with the same MAPSETCUI, FROMEXPR and MAPSUBSETID (including null or 0). (optional). |
| 4 | MAPRANK | Order in which mappings in a subset should be applied. Used only where MAPSUBSETID is used (optional). |
| 5 | MAPID | Unique identifier for this individual mapping. Primary key of this table to identify a particular row. |
| 6 | MAPSID | Source asserted identifier for this mapping (optional). |
| 7 | FROMID | Identifier for the entity being mapped from. This is an internal UMLS identifier used to point to an external entity in a source vocabulary (represented by the FROMEXPR). When the source provides such an identifier, it is reused here. Otherwise, it is generated by NLM. The FROMID is only unique within a map set. It is not a pointer to UMLS entities like atoms or concepts. There is a one-to-one correlation between FROMID and a unique set of values in FROMSID, FROMEXPR, FROMTYPE, FROMRULE, and FROMRES within a map set. |
| 8 | FROMSID | Source asserted identifier for the entity being mapped from (optional). |
| 9 | FROMEXPR | Entity being mapped from - can be a single code/identifier /concept name or a complex expression involving multiple codes/identifiers/concept names, Boolean operators and/or punctuation. |
| 10 | FROMTYPE | Type of entity being mapped from. Valid values: Field name from MRCONSO.RRF, including AUI, CODE, CUI, LUI, SAUI, SCUI, SDUI, STR, or BOOLEAN\_EXPRESSION\_<field\_name>. |
| 11 | FROMRULE | Machine processable rule applicable to the entity being mapped from (optional). |
| 12 | FROMRES | Restriction applicable to the entity being mapped from (optional). |
| 13 | REL | Relationship of the entity being mapped from to the entity being mapped to. |
| 14 | RELA | Additional relationship label (optional). |
| 15 | TOID | Identifier for the entity being mapped to. This is an internal identifier used to point to an external entity in a source vocabulary (represented by the TOEXPR). When the source provides such an identifier, it is reused here. Otherwise, it is generated by NLM. The TOID is only unique within a map set. It is not a pointer to UMLS entities like atoms or concepts. There is a one-to-one correlation between TOID and a unique set of values in TOSID, TOEXPR, TOTYPE, TORULE, TORES within a map set. |
| 16 | TOSID | Source asserted identifier for the entity being mapped to (optional). |
| 17 | TOEXPR | Entity being mapped to - can be a single code/identifier/concept name or a complex expression involving multiple codes /identifiers/ concept names, Boolean operators and/or punctuation. |
| 18 | TOTYPE | Type of entity being mapped to. Valid values (may be null for XR relationships):  Field name from MRCONSO.RRF, including: AUI, CODE, CUI, LUI, SAUI, SCUI, SDUI, STR, or BOOLEAN\_EXPRESSION\_<field\_name>. |
| 19 | TORULE | Machine processable rule applicable to the entity being mapped to (optional). |
| 20 | TORES | Restriction applicable to the entity being mapped to (optional). |
| 21 | MAPRULE | Machine processable rule applicable to this mapping (optional). |
| 22 | MAPRES | Restriction applicable to this mapping (optional). |
| 23 | MAPTYPE | Type of mapping (optional). |
| 24 | MAPATN | The name of the attribute associated with this mapping [not yet in use]. |
| 25 | MAPATV | The value of the attribute associated with this mapping [not yet in use]. |
| 26 | CVF | The Content View Flag is a bit field used to indicate membership in a content view. |

1. **MRSMAP.RRF:** This file provides a simpler representation of most of the mappings in MRMAP.RRF to serve applications which do not require the full richness of the MRMAP.RRF data structure. Mappings that support rule-based processing need the additional fields of MRMAP.RRF and are not included in MRSMAP.RRF. Specifically, all mappings with meaningful (non-null, non-zero) values for MAPSUBSETID are excluded from MRSMAP.RRF. Map sets are represented either in their entirety or not at all, i.e. no partial map sets are included.

|  |  |  |
| --- | --- | --- |
|  |  | **MRSMAP.RRF** |
| **#** | **Name** | **Description** |
| 1 | MAPSETCUI | Unique identifier for the UMLS concept which represents the whole map set. |
| 2 | MAPSETSAB | Source abbreviation (SAB) for the provider of the map set. |
| 3 | MAPID | Unique identifier for this individual mapping. Primary key to identify a particular row. |
| 4 | MAPSID | Source asserted identifier for this mapping (optional). |
| 5 | FROMEXPR | Entity being mapped from - can be a single code/identifier /concept name or a complex expression with multiple codes/identifiers/concept names, Booleans and/or punctuation. |
| 6 | FROMTYPE | Type of entity being mapped from. Valid values: Field name from MRCONSO.RRF, including AUI, CODE, CUI, LUI, SAUI, SCUI, SDUI, STR, or BOOLEAN\_EXPRESSION\_<field\_name>. |
| 7 | REL | Relationship of the entity being mapped from to the entity being mapped to. |
| 8 | RELA | Additional relationship label (optional). |
| 9 | TOEXPR | Entity being mapped to - can be a single code/identifier /concept name or a complex expression with multiple codes/identifiers/concept names, Booleans and/or punctuation. |
| 10 | TOTYPE | Type of entity being mapped to. Valid values (may be null for XR relationships):  Field name from MRCONSO.RRF, including AUI, CODE, CUI, LUI, SAUI, SCUI, SDUI, STR, or BOOLEAN\_EXPRESSION\_<field\_name>. |
| 11 | CVF | The Content View Flag is a bit field used to indicate membership in a content view. |

1. **MRREL.RRF:** A subset of mappings is redundantly represented as relationships in MRREL.RRF, based on the following guidelines:
   * FROMEXPR and TOEXPR are simple expressions
   * Map set is not rule-based
   * REL is not XR
   * Partial map sets may be represented in MRREL

Notes:

* There is currently no simple way of identifying cross-source mappings in MRREL.RRF. The RELAs for these relationships currently include “mapped\_to/from,” “same\_as,” “classified\_as/classified\_by” and the null RELA. All of these RELAs are also used for within-source RELAs. To identify cross-source mapping relationships, users will have to find MRREL.RRF cases where the AUI1 and AUI2 in MRCONSO.RRF have different SAB values and neither STYPE1 or STYPE2 is CUI.

1. **Additional Resources**

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=16779043>

<http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_022744.hcsp?dDocName=bok1_022744>

<http://www.cms.hhs.gov/MLNProducts/downloads/ICD-10_GEM_factsheet.pdf>

pcs\_gemguide\_2009.pdf (available at <http://www.cms.hhs.gov/ICD10/01m_2009_ICD10PCS.asp#TopOfPage>, under the “2009 Mapping” link)