



## Serializing the UMLS into RDFS

#### Janos G. Hajagos Stony Brook School of Medicine 9/29/2010

### RDFS

<http://www.w3.org/2000/01/rdf-schema#>

- A light weight schema for Resource Description Framework (RDF)
- Class and properties

   rdfs:Class & rdf:Property
- Basic inference
  - rdfs:subClassOf & rdfs:subPropertyOf
  - rdfs:domain & rdfs:range
- Human readable labels
  - rdfs:label & rdfs:comment

## UMLS

## Unified Medical Language System

- A long term (1986) research project of NIH's National Library of Medicine (NLM)
- A metathesaurus connecting different medical/biomedical vocabularies together with a concept unique identifiers (CUI)
- A semantic network of 54 broad types
  - "Neoplastic process" isa "Disease or Syndrome"
- rdfs:seeAlso
  - "Ontologies for Data Integration: A Semantic Web Perspective" for more in depth material on the UMLS
  - Initial publication of UMLS in 2007 in a semantic format.

#### **UMLS 2010AA Sources in Conversion**

Sorted by decreasing order of the total number of terms in source vocabulary

- SNOMED Clinical Terms
- Metathesaurus Names
- MeSH
- RXNORM
- Read Codes
- LOINC
- MEDCIN
- NCBI Taxonomy
- MedDRA
- NCI Thesaurus
- Gene Ontology
- ICD-10-PCS
- Digital Anatomist
- Foundational Model of Anatomy
- National Drug File Reference Terminology
- FDA National Drug Code Directory
- HUGO
- SNOMED Intl 1998
- Multum
- OMIM
- UMDNS
- National Drug File
- ICPC2-ICD10 Thesaurus
- Medical Entities Dictionary
- FDA Structured Product Labels
- Alternative Billing Concepts
- ICD-9-CM
- WHOART
- CRISP Thesaurus
- Alcohol and Other Drug Thesaurus
- CPT in HCPCS
- National Drug Data File
- ICNP
- Nursing Outcomes Classification
- SNOMED 1982
- Clinical Problem Statements
- MedlinePlus

- , PDQ
- Clinical Classifications Software
- HCPCS
- Psychological Index Terms
- Gold Standard Alchemy
- ICPC-2 PLUS
- CPT
- HL7 Version 3.0
- ICD-9-CM Entry Terms (UMLS)
- CPT Hierarchical Terms (UMLS)
- Micromedex
- ICD-10
- Nursing Interventions Classification
- HL7 Version 2.5
- International Classification of Functioning, Disability and Health
- COSTART
- Standard Product Nomenclature
- Read Codes Synth
- Neuronames Brain Hierarchy
- Minimum Data Set, 2.0
- USP Model Guidelines
- Read Codes Am Engl
- International Classification of Functioning, Disability and Health for Children and Youth
- HCPCS Hierarchical Terms (UMLS)
- Minimal Standard Terminology (UMLS)
- Master Drug Data Base
- Minimum Data Set, 3.0
- DXplain
- Clinical Concepts by R A Miller
- Congenital Mental Retardation Syndromes
- Current Dental Terminology in HCPCS
- Beth Israel Problem List
- Outcome and Assessment Information Set
- Patient Care Data Set

- Library of Congress Subject Headings
- Omaha System
- ICPC
- Home Health Care Classification
- ICPC2E
- Authorized Osteopathic Thesaurus
- COSTAR
- Pharmacy Practice Activity Classification
- AI/RHEUM
- DSM-IV
- Patient Health Questionnaire
- PNDS
- Patient Monitoring Guidelines for HIV care and antiretroviral therapy (ART)
- Read Codes Am Synth
- DSM-III-R
- ICD-10 Am Engl
- Source Terminology Names (UMLS)
- Classification of Nursing Diagnoses
- Quick Medical Reference
- ICPC2-ICD10 Thesaurus, 7-bit
- ICPC2-ICD10 Thesaurus, Am Engl
- HL7 Version 2.5, 7-bit equivalents
- Diseases Database

**ICPC2E** Am Engl

- Confusion Assessment Method (CAM)
- UltraSTAR
- Braden Scale for Predicting Pressure Sore Risk

Routine Health Outcomes Ltd. (RHO)

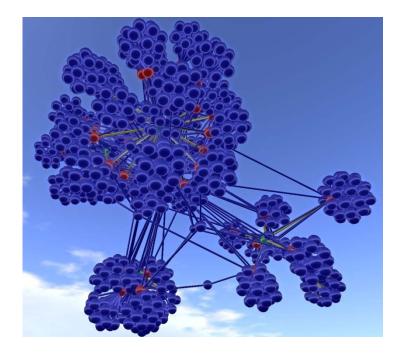
**Glossary of Clinical Epidemiologic Terms** 

4

 Faces, Legs, Activity, Cry, and Consolability (FLACC) Scale

# Relation of RxNorm to UMLS

- RxNorm's data table structure is very close to the UMLS
- RxNorm sources are contained within the UMLS
- RxNorm is updated more frequently so as to better capture new drugs entering the market



RxNorm RDF data of drugs containing Lithium visualized in 3D by Nexus

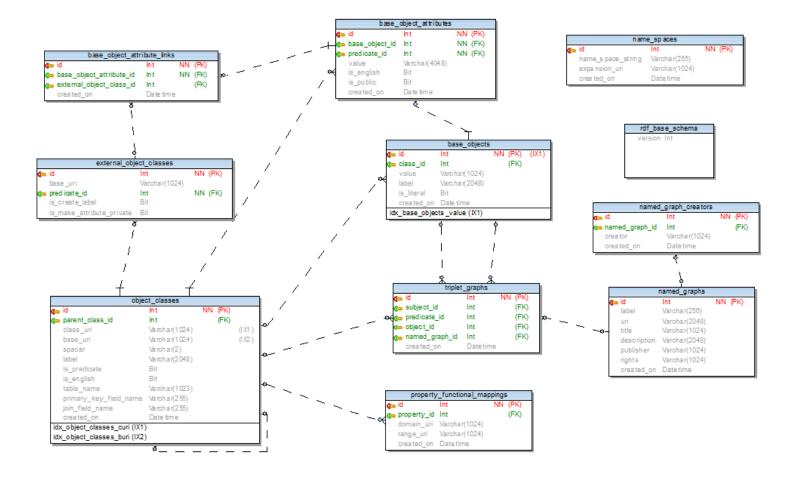
#### Comparison of the UMLS to RxNorm

	UMLS	RxNorm
Developed by	NLM	NLM
Focus	Genes, Drugs, Disease, Anatomy	Drugs
Identifiers	CUI AUI	RXCUI RXAUI
Update frequency	2-4 times a year	Full monthly with partial weekly updates
Release format	.NLM with Java program to generate site RRF	RRF (Rich Release Format)
Database load scripts	MySQL and Oracle	MySQL and Oracle
License	UMLS Metathesaurus Site License Each individual source is licensed separately	UMLS Metathesaurus Site License Each individual source is licensed separately

# RdfBase

- Targets "one time" conversions of relational databases into RDFS
- RDFS is serialized into ntriples format
- Used to publish RxNorm, UMLS, and a data warehouse schema with reference table values
- Python based library using Microsoft SQL Server on the backend

#### **RdfBase Relational DB Schema**



## RdfBase Code Snippet

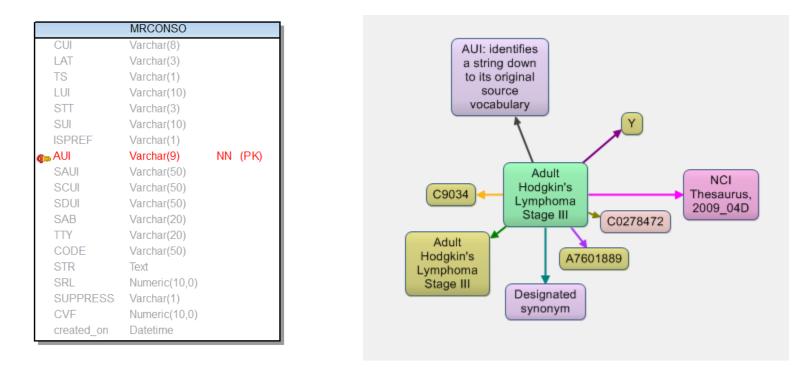
```
86 class m004PublishMRCONSO(MigrationClass):
 87
        def up(self):
 88
 89
            self.cursor.execute("""
 90
                 select CUI, AUI, TTY, CUI, CODE, ISPREF, cast(STR as varchar(4000)) as [STR]
 91
                    into MRCONSO cleansed
 92
                from MRCONSO where SUPRESS = 'N'
 93
            .....
 94
95
            umls obj = RdfBase(self.cursor)
            aui obj = RdfBaseTable("MRCONSO cleansed",umls obj)
 96
97
            aui obj.set class name("umls:AUI", "AUI: identifies a string down to its original source vocabulary", spacer="/")
98
            aui obj.set label field name("STR")
 99
            aui obj.set value field name("AUI")
            aui obj.set primary key field_name("AUI")
100
101
            aui obj.map field name("SAB")
102
            aui obj.map field name("CODE")
103
            aui obj.map field name("AUI")
104
            aui obj.map field name("CUI")
105
            aui obj.map field name("TTY")
106
            aui obj.map field name("CUI")
107
            aui obj.map field name("ISPREF")
108
            aui obj.publish()
109
110
```

def down(self):

111 112

self.cursor.execute("drop table MRCONSO cleansed")

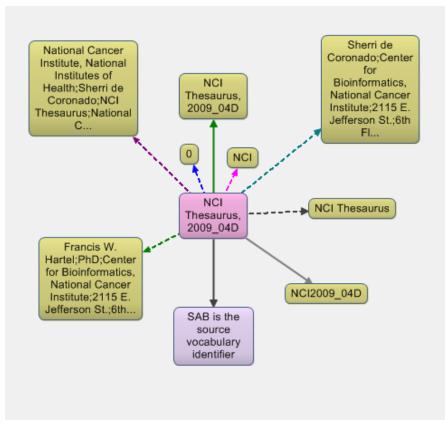
# MRCONSO (objects)



#### <http://link.informatics.stonybrook.edu/umls/AUI/A7601889>

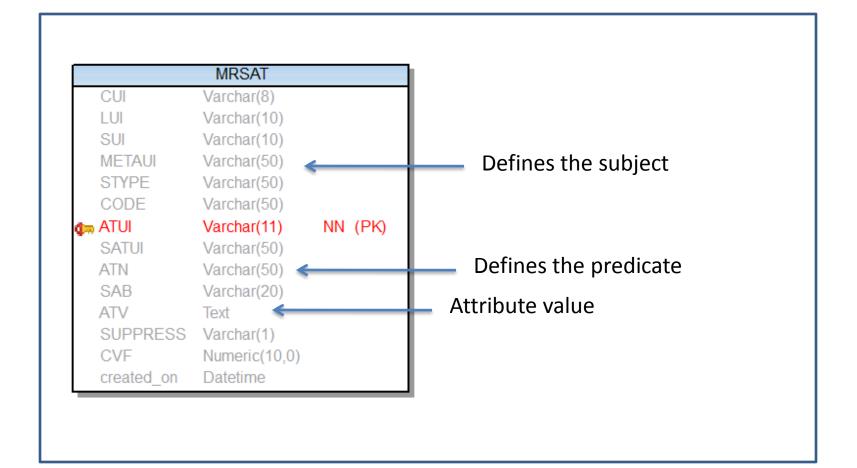
## MRSAB (metadata)

	MRSAB		
VCUI	Varchar(8)		
RCUI	Varchar(8)		
🚓 VSAB	Varchar(20)	NN	(PK)
RSAB	Varchar(20)		
SON	Text		
SF	Varchar(20)		
SVER	Varchar(20)		
VSTART	Varchar(8)		
VEND	Varchar(8)		
IMETA	Varchar(10)		
RMETA	Varchar(10)		
SLC	Text		
SCC	Text		
SRL	Numeric(10,0)		
TFR	Numeric(10,0)		
CFR	Numeric(10,0)		
CXTY	Varchar(50)		
TTYL	Varchar(300)		
ATNL	Text		
LAT	Varchar(3)		
CENC	Varchar(20)		
CURVER	Varchar(1)		
SABIN	Varchar(1)		
SSN	Text		
SCIT	Text		
created_on	Datetime		

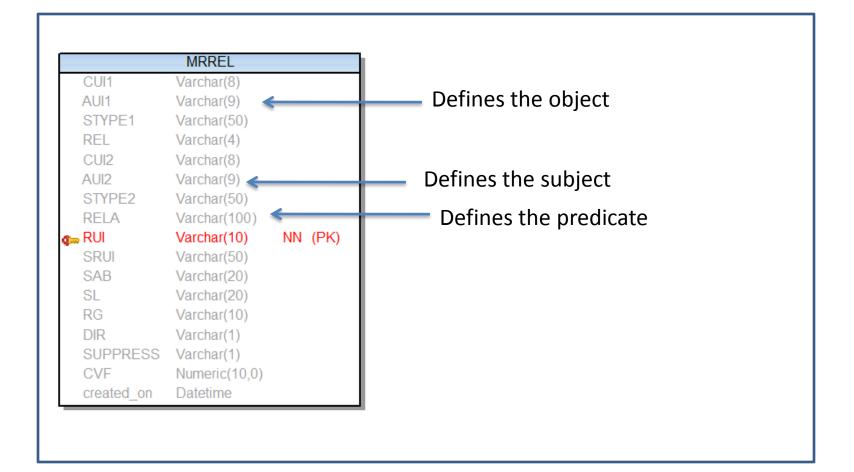


<http://link.informatics.stonybrook.edu/umls/SAB/NCI>

# MRSAT (attributes)



# MRREL (relationships)

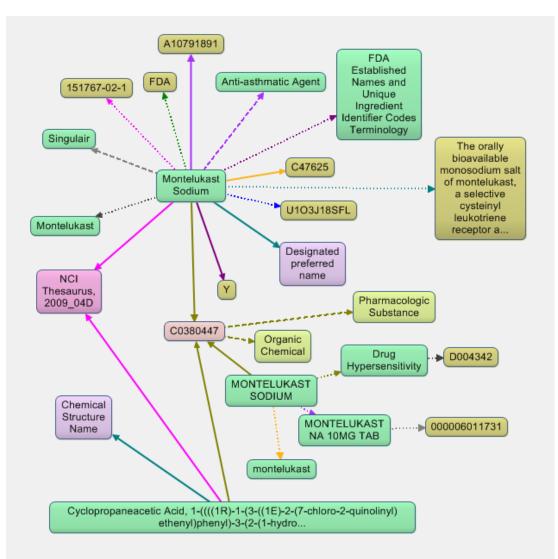


#### Example: Montelukast Sodium in NCI

- NCI Thesaurus is not part of RxNorm
- NCI Thesaurus is well structured in RDFS
  - available separately as OWL
- <u>Allegrograph's Gruff</u> is a tool being used to explore the UMLS
- CUIs allow linking across different source vocabularies

Montelukast Sodium         Entropy         Show All Traject           Property         Value         Click the righthand column to visit that resource in the table view AND add the right and column to Visit that resource in the table. Control-click to ONLY visit the resource in the table. Control-click to ONLY vise due non-control visit the table. Control-click to	ile View Add Link Remove Layout Select Inclusion Options Layo	ut Options Drawing Options Table & Query Options Help		
Property         Sulu         Is the grap invession. Shift-click the right hand column to DNLY add the node to the autor of the source in the table. Control-shift-click as a source in table. Control-shift-click as a source in the table. Control-shift-click as a source in table. Control-shift-click as a	Montelukast Sodium	Revisit 🔶 Show All Triples		
CAS REGISTRY15/76/2-1CODEC4763CommentThe orally bioexalialis monosodum sait of monteukust, a selective y and competitively bucks the cystery leukotrien a function of autoination a divides. Monteukust, a selective y and competitively bucks the cystery leukotrien a function of autoination a divides. Monteukust, a selective y and competitively bucks the cystery leukotrien a function of autoination a divides. Monteukust, a selective y and competitively bucks the cystery leukotrien a function of autoinaphilis, and bronchoctalism, and cystery leukotrien a function of a selectively and competitively bucks the cystery leukotrien a function of a selectively and competitively bucks the cystery leukotrien a function of a selectively and competitively bucks the cystery leukotrien a function of a selectively and neutophilis, and bronchoctarition. The Cyst 11 receptor is found in a number of tissues including advises. Monteukust a selectively and competitively bucks the cystery leukotrien a function of a selectively leukotrien a function of a selectively leukotrien a function. The Cyst 11 receptor is found in a number of tissues including advises. Monteukocytes to vascular andotecytes on vascular andotecytes and averaged and restrophilis, and bronchoctaritich. The Cyst 11 receptor is found in a number of tissues including advises. Monteukocytes to vascular andotecytes and averaged and restrophilis, and bronchoctaritich. The Cyst 11 receptor is found in a number of tissues including advises. Monteukocytes to vascular andotecytes and averaged and restrophilis, and bronchoctaritich. The Cyst 11 receptor is found in a number of tissues including advises. Monteukocytes to vascular andotecytes and averaged and restrophilis, and bronchoctaritich. The Cyst 11 receptor is found in a number of tissues including advises. Monteukocytes to vascular andotecytes to vascular andotecytes. The Cyst 11 receptor is found in a number of tissues including advises. Mont	Property	Value to the graph view. Shift-click the righthand column to ONLY add the node to the graph. Control-click to ONLY visit the resource in the table. Control-shift-click a a		
CODE         C47625           Comment         The orally bioavailable monosodium salt of monteukast, a selective cysteinyl leukotrien receptor antagonist with anti- inflammatory and bronchodisting activities. Monteukast selective dynatocyste by cysteinyl jeukotrien i CSULT1) receptor, preventing binding of the inflammatory mediator leukotriene D4 (LT04). Inhibition of leukotrylise andesion of leukotrylise monosodium, monocyste and neutrophil aggregation; increased airway dema: increased capilary permeability; and bronchoostiriction. The CysLT1 receptor is found in a number of issues including concept in subset           Concept in subset         A12794052           Contractive SOURCE         FDA           FDA UNI CODE         U103/185FL           Has GUI         C0380447           Has SAB         NCI           Has readoname         A12794052           Ispace         PT           Has radoname         A1070412           Has radoname         A1070412           Ispace         S9302           PDQ OPEN TRAL SEARCH ID         S9302           PDQ OPEN TRAL SEARCH ID         S9302           PREFERED NAME         Monteukast Sodium           SUPPRESS         N           Statis form of         A12815994           Subset Includes concept of         A7664376	AUI	A10791891		
Comment         The only bioavailable monosodium salt of montalukast, a selective cysteinyl leukotriene receptor antagonist with anti- inflammatory and bronchodiating activities. Montelukast, a selective y and competitively biocks the cysteinyl leukotriene 1 (Cys.LT) receptor, preventing binding of the inflammatory mediative leukotriene 04 (LTD4). Inhibiton 1LTD4 activity (Cys.LT) receptor is funding of the inflammatory mediative leukotriene 04 (LTD4).           Concept in subset         A12794052           CONTRIBUTING SOURCE         FDA           FDA UNI CODE         U103.118FL           Has Grad or base form         A12794052           A12794052         CONTRIBUTING SOURCE           FDA UNI CODE         U103.118FL           Has Grad or base form         A1281594           Has Grad or base form         A12744052           Has Free acid or base form         A1271402           Has Free acid or base form         A12744052           FDA UNI CODE         PT           Has Free acid or base form         A1070412           Isa         A7664876           SIDE COLSED TRIAL SEARCH ID         S9302           PDO COLSED TRIAL SEARCH ID         S9302           POR CLOSED TRIAL SEARCH ID         S9302           POR FERERED NAME         Nontelukast Sodium           SUPPRESS         N           Type         A12815994	CAS REGISTRY	151767-02-1		
Inflammatory antorhoditating activities. Montelukast selectively and competitively blocks the cysteriny leukotriene subservers including: migration doesinophils and neutrophil gragation: increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is found in a number of tissues increased capitary permeability: and bronchocenstridion. The Cyst.T1 receptor is fo	CODE	C47625		
CONTRIBUTING SOURCEFAAFOA UNI CODEU03/18FLFAG UNI CODE038047Has CUIA128594Has free aid or base formA1281594Has free aid or base formNCIHas Tage Tage Tage Tage Tage Tage Tage Tage	Comment	inflammatory and bronchodilating activities. Montelukast selectively and competitively blocks the cysteinyl leukotriene 1 (CysLT1) receptor, preventing binding of the inflammatory mediator leukotriene D4 (LTD4). Inhibition of LTD4 activity results in inhibition of leukotriene-mediated inflammatory events including: migration of eosinophils and neutrophils; adhesion of leukocytes to vascular endothelium, monocyte and neutrophil aggregation; increased airway edema;		
FAUNICODEI03018FLHas CUC030447Has CullA181594Has fae acid or base formK12Has SABNCIHas Tarm TypePTHas tandenameA1070112Isa San CullA164676ISPREFYandenameLabelS0302POC OSED TRIAL SEARCH IDS0302POC POST TRIAL SEARCH IDS0302SUPPRESNatureTypeS0404SuppressNatureSuppressNatureSuppressNatureSuppressS15994Suppress for formS15994Suppress for formS15994 <tr< td=""><td>Concept in subset</td><td>A12794052</td></tr<>	Concept in subset	A12794052		
HacUlC038047Has ColumbricationA1281594Has AshNCIHas AshNCIHas AshA1070412Has AshA768476IsananNonutrationIsa AshNonutrationIsa AshNonutration	CONTRIBUTING SOURCE	FDA		
Ha fee acid or base formAt281594Has Fee acid or base formK281594Ha sa form TypeFTHa tradenameAt070412Isa form TypeAr66876IspREFYLabelMonelukas SodiumPOQ CLOSED TRIAL SEARCH IDS9302POQ FUNTRIAL SEARCH IDS9302SUPPRESSNonelukas SodiumSUPPRESSNonelukas SodiumTypeAllis Insert ofAllis Insert ofAllis Insert ofAlfogais Insert ofAlfogais Insert ofAlfogais Insert ofAlfogais Insert ofAlfogais Unsert ofAlfogais Insert ofAlfogais Insert ofAlfogais Unsert of Insert ofAlfogais Unsert of Insert ofAlfogais Unsert of Insert ofAlfogais Unsert of Insert of In	FDA UNII CODE	U103J18SFL		
Has SABNCIHas Tam TypePTHas tadenameA10770412Isa Can Can Can Can Can Can Can Can Can Ca	Has CUI	C0380447		
Har ToppPHas Tom TypePHas todenameA10770412IsaAr664876IspREFVLabdMonlukas SodiumPDC LOSED TRIAL SEARCH ID93502PDC PENTRIAL SEARCH ID93502PREFRRED NAMEMonlukas SodiumSUPPRESSNTypeJ01Ishasal form ofJ28594Ishasal form ofA18594Ishusen searchJ264876Ishusen searchJ264876Ishusen searchJ264876Ishusen searchJ264876Ishusen searchJ264876Ishusen searchJ264876Ishusen searchJ274052	Has free acid or base form	A12815994		
Has radenameA10770412Has tradenameA7664876IsaA7664876ISPREFYLabelMontelukast SodiumPDQ CLOSED TRIAL SEARCH ID593502PDQ OPEN TRIAL SEARCH ID593502PREFERED NAMEMontelukast SodiumSUPPRESSNTypeAUIis Has salt form ofA12815994is Inverse isa ofA7664876is Subst includes concept ofA12794052	Has SAB	NCI		
IsaAf66876ISPREFYLablMontelukas SodiumPDQ CLOSED TRIAL SEARCH ID93502PDQ OPEN TRIAL SEARCH ID93502PREFERED NAMEMontelukas SodiumSUPPRESSNTypeUIis Has salt form ofAf28594is Has salt form ofAf6876is Upset son of of the off son	Has Term Type	PT		
iSPREF         Y           Label         Montelukast Sodium           PDQ CLOSED TRIAL SEARCH ID         99302           PDQ OPEN TRIAL SEARCH ID         99302           PREFERED NAME         Montelukast Sodium           SUPPRESS         N           Type         AUI           is Has salt form of         A26859           is Inverse isa of         A768476           is Supper Solution         A768476           is Upper Solution         A768476           is Job solution         A768476	Has tradename	A10770412		
Label         Montelukast Sodium           PDQ CLOSED TRIAL SEARCH ID         593502           PDQ OPEN TRIAL SEARCH ID         593502           PREFERED NAME         Montelukast Sodium           SUPPRESS         N           Type         AUI           is Has salt form of         A2815994           is Inverse isa of         A766476           is Subset includes concept of         A1294052	Isa	A7664876		
PDQ CLOSED TRIAL SEARCH ID         593502           PDQ OPEN TRIAL SEARCH ID         593502           PREFERED NAME         Montelukast Sodium           SUPPRESS         N           Type         AUI           is Has salt form of         A264876           is Inverse isa of         A7664876           is Subset includes concept of         A1294052	ISPREF	Y		
PAG         Second state           PDQ OPEN TRIAL SEARCH ID         593502           PREFERED NAME         Montelukast Sodium           SUPPRESS         N           Type         AUI           is Has salt form of         A264876           is Inverse isa of         A7664876           is Subset includes concept of         A1294052	Label	Montelukast Sodium		
PREFERED NAME         Montelukast Sodium           SUPPRESS         N           Type         AUI           is Has salt form of         A28594           is Inverse isa of         A7664876           is Subset includes concept of         A12974052	PDQ CLOSED TRIAL SEARCH ID	593502		
SUPRESS         N           Type         AUI           is Has salt form of         A12815994           is Inverse isa of         A7664876           is Subset includes concept of         A12974052	PDQ OPEN TRIAL SEARCH ID	593502		
Type         All           is Has salt form of         Al2815994           is Inverse isa of         Ar664876           is Subset includes concept of         Al29052	PREFERRED NAME	Montelukast Sodium		
A12815994           is Has salt form of         A12815994           is Inverse isa of         A7664876           is Subset includes concept of         A1294052	SUPPRESS	Ν		
is Inverse isa of         A7664876           is Subset includes concept of         A12794052	Туре	AUI		
is Subset includes concept of A12794052	is Has salt form of	A12815994		
	is Inverse isa of	A7664876		
is Tradename of of A10770412	is Subset includes concept of	A12794052		
	is Tradename of of	A10770412		

AUI		
CAS REGISTRY		
CODE		
Comment		
Concept in subset		
Contraindicated drug		
CONTRIBUTING SOURCE		
FDA UNII CODE		
Has CUI		
Has free acid or base form		
Has ingredient		
Has SAB		
Has STN		
Has Term Type		
Has tradename		
Isa		
ISPREF		
MESH UI		
NDC		
Va product component of		
AUI: identifies a string down to its original source vocabu		
CUI: is a unique concept identifier it ties concepts across		
SAB is the source vocabulary identifier		
Semantic Type		
(TTY (term type))		
Literal		



## Future Steps: Integrating with RxNorm and LODD

- Enhance RxNorm RDFS conversion
  - links to UMLS CUIs
  - Semantic types
- Define subset of the UMLS that can be published as part of the LODD
  - 110,169,095 triples in internal draft publication but includes non releasable sources
- Links from UMLS and RxNorm into other datasets in the LODD cloud

## Acknowledgements

- Supreet Padhi and Jakub Pezacki
  - Worked on utilizing the UMLS for clinical rule extraction
- Erich Bremer development of RDF visualization tool