



# HeTOP, crosslingual health terminology server: towards pragmatic applications

**Stefan J. Darmoni, MD, PhD & Julien Grosjean, PhD**

***Rouen University Hospital & LIMICS, INSERM U1142***

Email: [Stefan.Darmoni@chu-rouen.fr](mailto:Stefan.Darmoni@chu-rouen.fr) & [Julien.Grosjean@chu-rouen.fr](mailto:Julien.Grosjean@chu-rouen.fr)



# TOR Terminological and Ontological Resources

KOS Knowledge Organization System

Ontology  
Terminology  
Thesaurus  
Nomenclature  
Taxonomy  
Dictionary  
Lexicon  
List





# TOR/KOS

CiSM<sup>e</sup>F

- To represent information & knowledge
  - Store
  - Describe
  - Structure
  - Annotate
- Information retrieval
  - Indexing / coding
- Fact inferencing
  - Rules=> computer-aided decision support systems
  - Reasoning
  - Subsumption (hierarchy)





# Roles of terminology servers

- Provide access to the main TOR/KOS in health
  - To humans
  - To computers (Web services)
- Diffusion of the main health terminologies
  - ICPC, ICD10 (11), CCAM (ICHI)
  - LOINC, ATC
  - MeSH
  - SNOMED CT



# Main (academic) terminology servers

- UMLS (USA)



- BioPortal (USA)\* -> instance francophone LIRMM



- Ontology Lookup Service (UK)



- LexGrid / NCI Term browser (USA)



- HeTOP (France)\* **HeTOP**

\* Grosjean J et coll. An Approach to Compare Bio-Ontologies Portals. Stud Health Technol Inform, 2014;205:1008-1012.

# HeTOP editorial policy

- Mixt approach: content (KOS)/container (server)
- Multi-terminology, with one pivotal terminology (MeSH)
  - For EHR and clinical applications, SNOMED CT = pivotal?
- Crosslingual with two pivotal languages (French/English) : *matricial navigation among languages & terminologies*
  - Systematic integration of KOS in French & in health (BNPC, CCAM, LPP, Cladimed, DRC, NABM, etc.) + French translations performed by others (WHO – ICD10, ICF-WONCA – ICPC2, ATC, LOINC, ICNP, etc.)
- Choice of KOS driven by academic or health information use cases (DRGs, EHR, documentary curation, information retrieval, real-time terminologies, primary care, rare diseases, etc.)
- Only last version
- Semantic enrichments: translations & mappings +++ (main added value of our team)

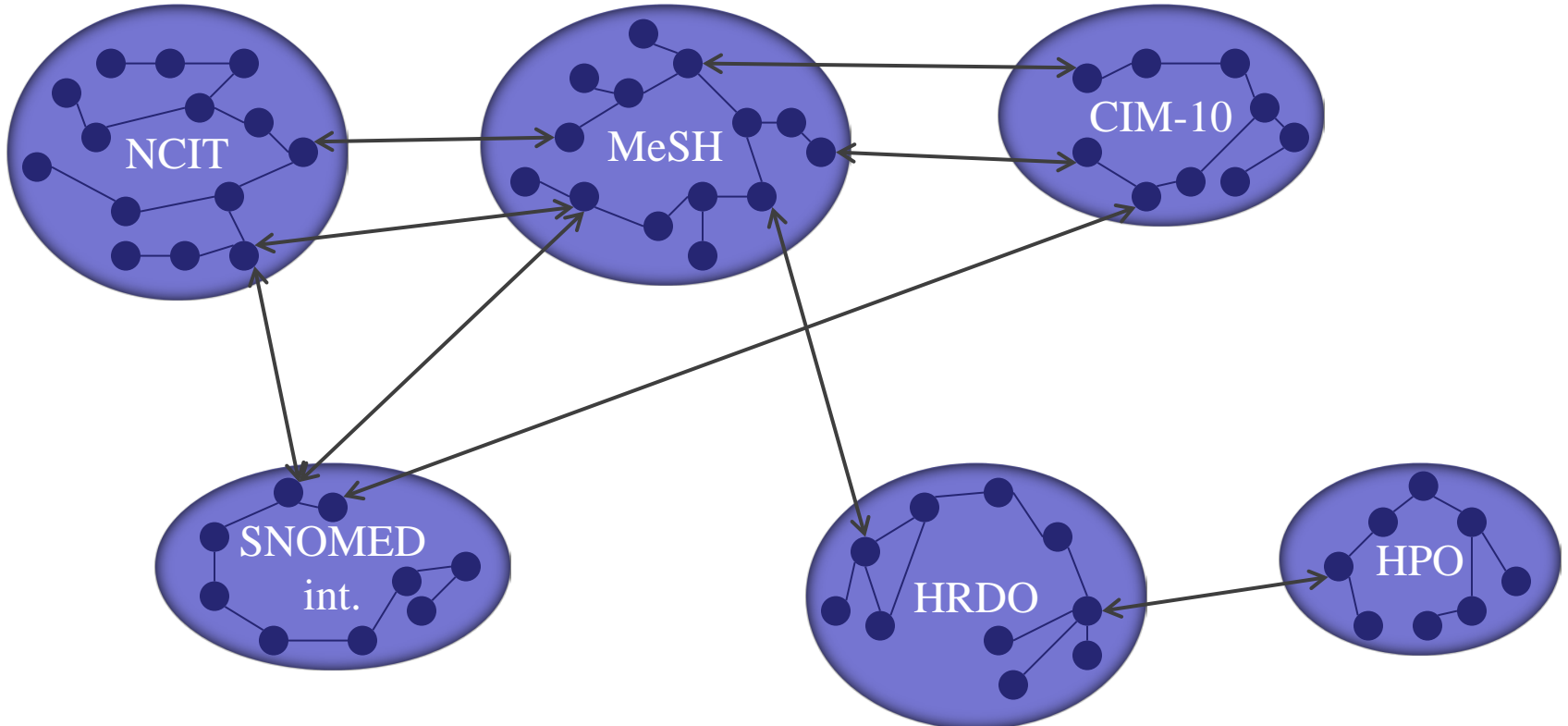
# HeTOP

- Health Terminology/Ontology Portal
- D2IM, CHU de Rouen – LIMICS
- URL : [www.hetop.eu](http://www.hetop.eu)
- 75 KOS, 32 languages
- Partially free (27 KOS); some others after authentication (ICPC2)
- Around 900 uniques machines per working day
- More than 2,800 subscribers
  
- Technologies :
  - Meta-model
  - BDD NoSQL since 2016



# Interoperability

- Interoperability of KOS & semantic network



- May 2017 : 314 000 supervision mappings, 58 000 manual mappings



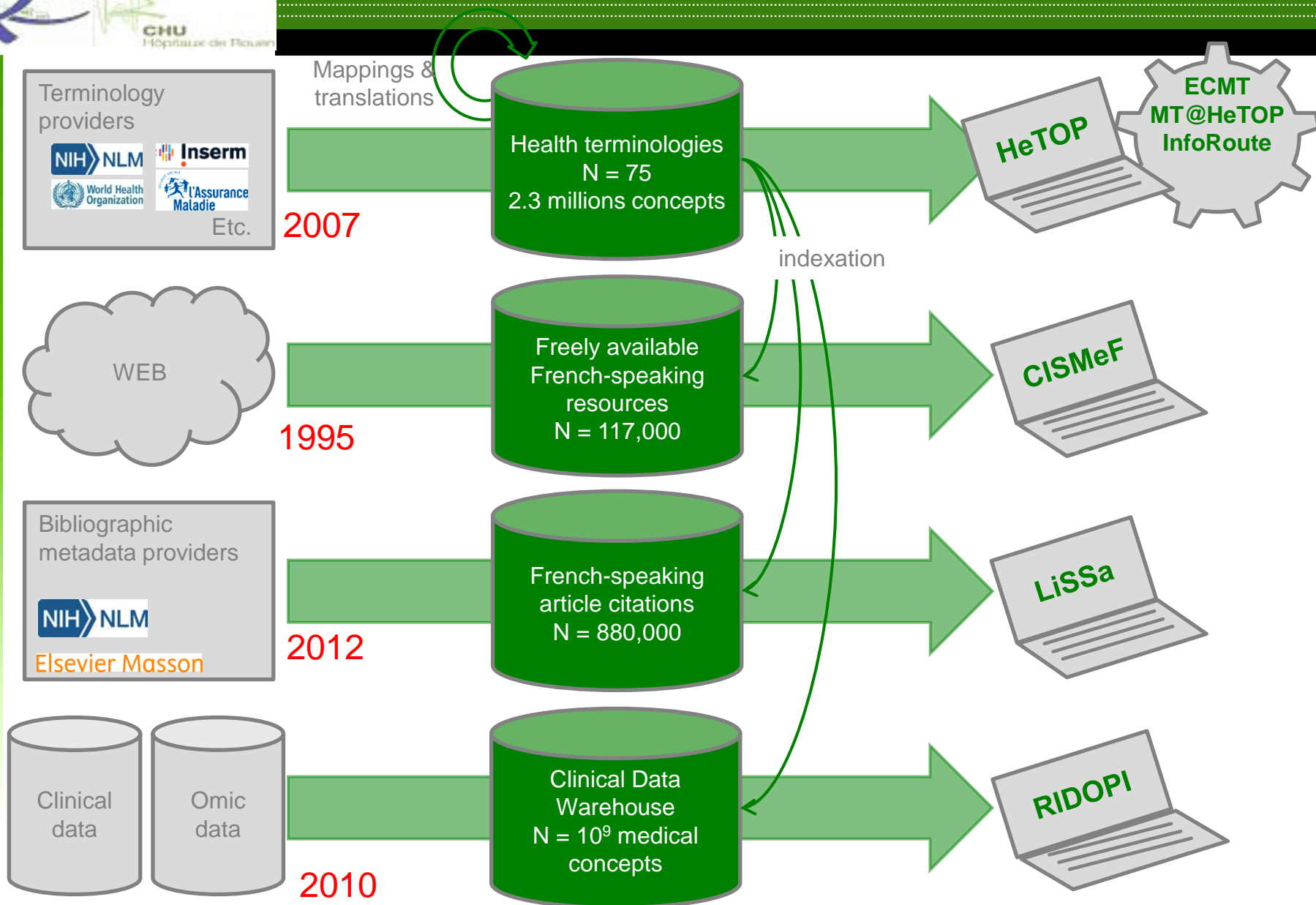
# Content

Terminologies	Concepts	Concepts in English	Concepts in French	Terms	Attributes	Relations
75 (17 UMLS)	2 538 595	1 885 880	1 267 074	9 982 113	15 966 667	9 965 417

- Overall, in HeTOP, the number of UMLS CUI with a least one French translation is  $\approx$  **388,900** vs.  $\approx$  **88,000** in UMLS (x4.42)
- 108 millions de triplets RDF en 2014 => Big data in health

# Semantic Web Plateform

CISMeF





# HeTOP use cases

CiSM<sup>e</sup>F

- Terminology server chosen by the Rouen University Hospital since end of 2016
  - Interface terminologies mapped to reference terminologies (biology, imaging)
- HeTOP « instantiation »
  - Version DRG (simplified) for coding technicians (ICD10, CCAM)
  - Version Emergency to code « passage » (ICD10, SFMU)
- BDD HeTOP: terminological & lexical support to automatic indexing tool (commercialised by Alicante)
- Research projects (TerSan, SYNODOS, RAVEL, SIFADO, PlaIR2.018, C3-Cloud, ...)



# HeTOP use cases

CiSM<sup>e</sup>F

- Terminology server:
  - multi-terminological indexing (CISMeF, LiSSa)
  - Teaching support (students in medicine, information sciences)
  - SNOMED CT (cooperation with Prof. Lemoine, Hop. Erasme, Bruxells)
  - Support to maintain the French translation of MedDRA (ICH) during two years
  - Creation of KOS de novo
    - Q-Codes (Marc Jamouille, MD, PhD applicant)
    - Ontology in bioinformatics (OMICX startup)
  - ICPC2? (discussion with WICC)

# For practice

- No wildcard search
- Do not search into definitions
- Terminologies selection
- filter translated concepts

## Your queries

59 matches in 0,04 s

### Top terms

- L95 vascularite [angéite] limitée à la peau, non classée ailleurs [ICD-10 category]
- Vasculitis limited to skin, not elsewhere classified [ICD-10 category (who)]
- L95 osteoporosis [ICPC-2 Rubric]
- L950 vascularite livédoïde [ICD-10 Sub-category]
- L951 erythema elevatum diutinum [ICD-10 Sub-category]
- L958 autres vascularites limitées à la peau [ICD-10 Sub-category]
- L959 vascularite limitée à la peau, sans précision [ICD-10 Sub-category]
- Lower Arm and Wrist Tendon, Right [PCS menu]
- Livedoid vasculitis [ICD-10 Sub-category (who)]
- Erythema elevatum diutinum [ICD-10 Sub-category (who)]

### MeSH (17)

### Genes & Proteins (5)

### ICD-10 ATIH (5)

### ICD-10 PCS (19)

### ICD-10 WHO (5)

### ICPC-2 (1)

### NCIt (7)

## L95 osteoporosis (ICPC-2 Rubric)



Description Hierarchies Relations PubMed / Doc'CISMeF



Show all languages  On

Add a metadata

### Origin ID

L95

### Preferred Label

- osteoporosis
- ostéoporse
- Osteoporose Osteoporose Osteoporosis Osteoporosi
- Osteoporosa osteoporosis Osteoporosi 骨粗鬆症
- Osteoporose osteoporose Osteoporose Osteoporosa
- osteoporosa osteoporos Osteoporoz Ocneonopos
- Loãng xương

### ICPC2 definition

- characteristic imaging appearance
- 画像診断上特有の所見を有すること
- sinais imagiológicos característicos
- aspect imagistic caracteristic

### Inclusion

- pathological fracture due to osteoporosis
- pathological fractures due to osteoporosis
- fracture pathologique dues à l'ostéoporse
- patologiska frakturer beroende på osteoporos
- 骨粗鬆症による病理学的骨折
- fraturas patológicas devido a osteoporose
- fracturi patologice datorate osteoporozei

ICPC2 in 20 languages

# Knowledge



HeTOP

en

marfan syndrome

- No wildcard search
- Do not search into definitions

Terminologies selection  filter translated concepts

Your queries

56 matches in 0,02 s

### Top terms

- marfan syndrome [MeSH Descriptor]
- marfan syndrome [MeSH concept]
- 3.1711 - Marfan syndrome [ACR pathology]
- 4.1711 - Marfan syndrome [ACR pathology]
- 5.1971 - Marfan syndrome [ACR pathology]
- 9.1971 - Marfan syndrome [ACR pathology]
- Marfan syndrome [HRDODisease]
- marfan's syndrome [MedDRA Preferred Term]
- marfan syndrome [MedlinePlus Topic]
- Marfan Syndrome [NCIt concept]

MeSH (14)

ACR (4)

Gene Ontology (1)

Genes & Proteins (1)

HRDO (10)

ICD-10 WHO (1)

ICPC-2 (1)

MedDRA (1)

MedlinePlus (1)

NCIt (2)

OMIM (6)

PASCAL (1)

SNOMED CT (10)

SNOMED int. (3)

## Marfan syndrome (MeSH Descriptor)



Description Hierarchies Relations PubMed / Doc'CiSMef

### Full tree

#### MeSH top tree

##### Diseases Category

cardiovascular diseases

cardiovascular abnormalities

heart defects, congenital

**marfan syndrome**

heart diseases

heart defects, congenital

**marfan syndrome**

congenital, hereditary, and neonatal diseases and abnormalities

congenital abnormalities

abnormalities, multiple

**marfan syndrome**

cardiovascular abnormalities

heart defects, congenital

**marfan syndrome**

genetic diseases, inborn

**marfan syndrome**

musculoskeletal diseases

bone diseases

bone diseases, developmental

**marfan syndrome**

skin and connective tissue diseases

connective tissue diseases

**marfan syndrome**

# Knowledge



HeTOP

en

marfan syndrome

- No wildcard search
- Do not search into definitions

- Terminologies selection  filter translated concepts

Your queries

56 matches in 0,02 s

### Top terms

- [marfan syndrome \[MeSH Descriptor\]](#)
- [marfan syndrome \[MeSH concept\]](#)
- [3.1711 - Marfan syndrome \[ACR pathology\]](#)
- [4.1711 - Marfan syndrome \[ACR pathology\]](#)
- [5.1971 - Marfan syndrome \[ACR pathology\]](#)
- [9.1971 - Marfan syndrome \[ACR pathology\]](#)
- [Marfan syndrome \[HRDODisease\]](#)
- [marfan's syndrome \[MedDRA Preferred Term\]](#)
- [marfan syndrome \[MedlinePlus Topic\]](#)
- [Marfan Syndrome \[NCIt concept\]](#)

### MeSH (14)

### ACR (4)

### Gene Ontology (1)

### Genes & Proteins (1)

### HRDO (10)

### ICD-10 WHO (1)

### ICPC-2 (1)

### MedDRA (1)

### MedlinePlus (1)

### NCIt (2)

### OMIM (6)

### PASCAL (1)

### SNOMED CT (10)

### SNOMED int. (3)

## Marfan syndrome (HRDODisease)



Description Hierarchies Relations PubMed / Doc'CiSMef

Intra-terminologic

Inter-terminologic

Add a metadata

Mappings

Semantic type(s) (1)

Related sign(s) (48)

- Abnormal dentition/dental position/implantation/unerupted/dental ankylosis HRDO Sign
- Aortic dissection HRDO Sign
- Aortic root dilatation/dilation/aneurysm HRDO Sign
- Aortic valve anomaly/incompetence/insufficiency/regurgitation/bicuspid HRDO Sign
- Arterial aneurism (excluding aorta) HRDO Sign
- Arterial rupture HRDO Sign
- Articular/joint pain/arthritis HRDO Sign
- Asthenia/fatigue/weakness HRDO Sign
- Autosomal dominant inheritance HRDO Sign
- Cleft palate without cleft lip/submucosal cleft palate/bifid uvula HRDO Sign
- Coxa profunda/acetabular protrusion HRDO Sign

Related MeSH (1)

UMLS correspondences (same concept) (9)

Currated CiSMef NLP mapping (15)

SNOMED CT concept(s) (1)

MedDRA code(s) (1)

OMIM phenotype(s) (2)





# Towards practice with the web of data...

HeTOP

es

evaluación del estado



T/O



julien.grosjean@chu-rouen.fr



- No wildcard search
- Do not search into definitions
- Terminologies selection
- filter translated concepts

[http://www.hetop.eu/hetop/?la=es&rr=CGP\\_QC\\_QD33](http://www.hetop.eu/hetop/?la=es&rr=CGP_QC_QD33)

Su consulta

2321 Recurso(s) encontrado en 0,07 s

Top terms

- Evaluación [CISMeF resources Type]
- tiene evaluación [SNOMED CT concept]
- evaluación del coeficiente intelectual [SNOMED CT concept]
- autoevaluación [CISMeF resources Type]
- evaluación de la piel [SNOMED CT concept]
- Evaluación de Daños [DeCS]
- Evaluación en Salud [DeCS]
- evaluación del dolor [SNOMED CT concept]
- evaluación de salud [SNOMED CT concept]
- QD33 evaluación del estado de salud [Q-code]

MeSH (27)

CISMeF (10)

DeCS (28)

MedDRA (12)

SNOMED CT (2234)

Q-Codes (10)

Q-code (10)

- QD33 evaluación del estado de salud
- QP23 competencia cultural
- QP41 satisfacción del paciente
- QR21 farmacoepidemiología
- QS46 farmacéutico
- QT3 garantía de la calidad
- QT13 evaluación de enseñanza
- QT33 lectura crítica de la literatura
- QT35 acreditación
- QT52 evento adverso

## QD33 evaluación del estado de salud (Q-code)

Descripción Jerarquías Relaciones PubMed / Doc'CISMeF



Show all languages

OFF



Origin ID

QD33

Preferred Label

- evaluación del estado de salud
- health status assessment évaluation de l'état de santé

Category label

- Evaluación del estado de salud
- Health status assesment Evaluation de l'état de santé

Q-Code entry term

- Activities of Daily Living (ADL) functional status Activités de la Vie Journalière (AVJ)

Q-Code acronym

- ADL AVJ

Q-Code definition

- La evaluación objetiva y subjetiva de la capacidad de una persona para realizarse y adaptarse a su entorno durante un período de tiempo determinado. (Woncadic)
- assessment of the ability of a person to perform and adapt to his environment, measured both objectively and subjectively over a stated period of time. (Woncadic)
- évaluation de la capacité d'une personne à fonctionner et s'adapter à son environnement, mesuré à la fois objectivement et subjectivement sur une période déterminée. (Woncadic)



# Contextual queries

- No wildcard search
- Do not search into definitions
- Terminologies selection
- filter translated concepts

## Your queries

56 matches in 0,06 s

### Top terms

- [marfan syndrome \[MeSH Descriptor\]](#)
- [marfan syndrome \[MeSH concept\]](#)
- [3.1711 - Marfan syndrome \[ACR pathology\]](#)
- [4.1711 - Marfan syndrome \[ACR pathology\]](#)
- [5.1971 - Marfan syndrome \[ACR pathology\]](#)
- [9.1971 - Marfan syndrome \[ACR pathology\]](#)
- [Marfan syndrome \[HRDODisease\]](#)
- [marfan's syndrome \[MedDRA Preferred Term\]](#)
- [marfan syndrome \[MedlinePlus Topic\]](#)
- [Marfan Syndrome \[NCIt concept\]](#)

### MeSH (14)

#### MeSH Descriptor (3)

- [Fibrillin-1](#)
- [Loeys-Dietz syndrome](#)
- [marfan syndrome](#)

#### MeSH Supplementary Concept (6)

#### MeSH concept (5)

### ACR (4)

### Gene Ontology (1)

### Genes & Proteins (1)

### HRDO (10)

### ICD-10 WHO (1)

### ICPC-2 (1)

### MedDRA (1)

### MedlinePlus (1)

## Marfan syndrome (MeSH Descriptor)



- Description
- Hierarchies
- Relations
- PubMed / Doc'CISMeF

### 1. Allowed qualifier(s) for this descriptor:

#### Display the qualifiers hierarchical list

- |   |   |   |
|---|---|---|
| <input type="checkbox"/> blood              | <input type="checkbox"/> cerebrospinal fluid    | <input type="checkbox"/> chemically induced |
| <input type="checkbox"/> classification     | <input type="checkbox"/> complications          | <input type="checkbox"/> diagnosis          |
| <input type="checkbox"/> diagnostic imaging | <input type="checkbox"/> diet therapy           | <input type="checkbox"/> drug therapy       |
| <input type="checkbox"/> economics          | <input type="checkbox"/> embryology             | <input type="checkbox"/> enzymology         |
| <input type="checkbox"/> epidemiology       | <input type="checkbox"/> ethnology              | <input type="checkbox"/> etiology           |
| <input type="checkbox"/> genetics           | <input type="checkbox"/> history                | <input type="checkbox"/> immunology         |
| <input type="checkbox"/> metabolism         | <input type="checkbox"/> microbiology           | <input type="checkbox"/> mortality          |
| <input type="checkbox"/> nursing            | <input type="checkbox"/> parasitology           | <input type="checkbox"/> pathology          |
| <input type="checkbox"/> physiopathology    | <input type="checkbox"/> prevention and control | <input type="checkbox"/> psychology         |
| <input type="checkbox"/> radiotherapy       | <input type="checkbox"/> rehabilitation         | <input type="checkbox"/> surgery            |
| <input type="checkbox"/> therapy            | <input type="checkbox"/> urine                  | <input type="checkbox"/> veterinary         |
| <input type="checkbox"/> virology           |   |   |

### 2. Options

only the main ones

without explosion

all types

### 3. Queries:



You may also use the Query Builder:



# Limits

- Terminological model: no inferencing (non ontological)
  - Do it before: use Protégé
- Currently, no versionning: in development
- Licences?





# Thank you

**CiSM<sup>e</sup>F**

## Questions

### **Email:**

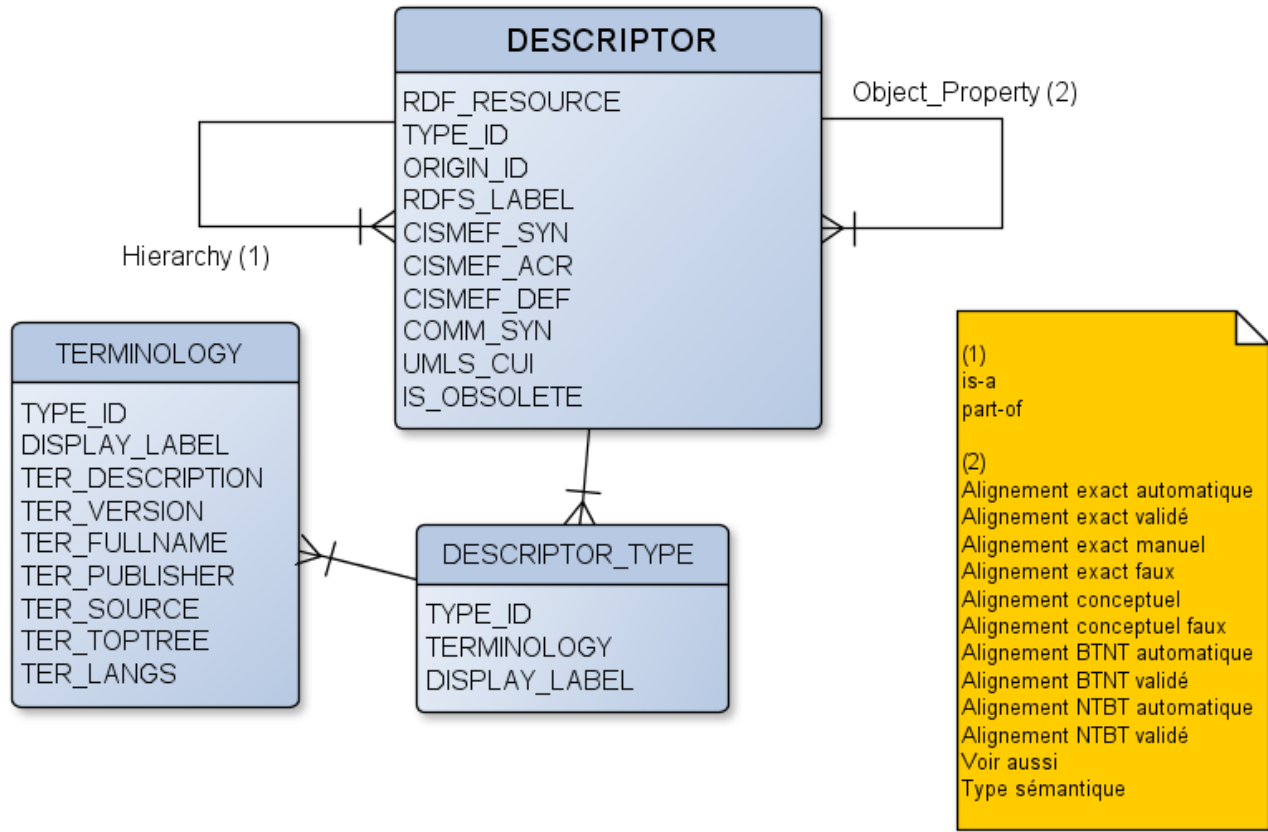
[Stefan.Darmoni@chu-rouen.fr](mailto:Stefan.Darmoni@chu-rouen.fr)

[Julien.Grosjean@chu-rouen.fr](mailto:Julien.Grosjean@chu-rouen.fr)



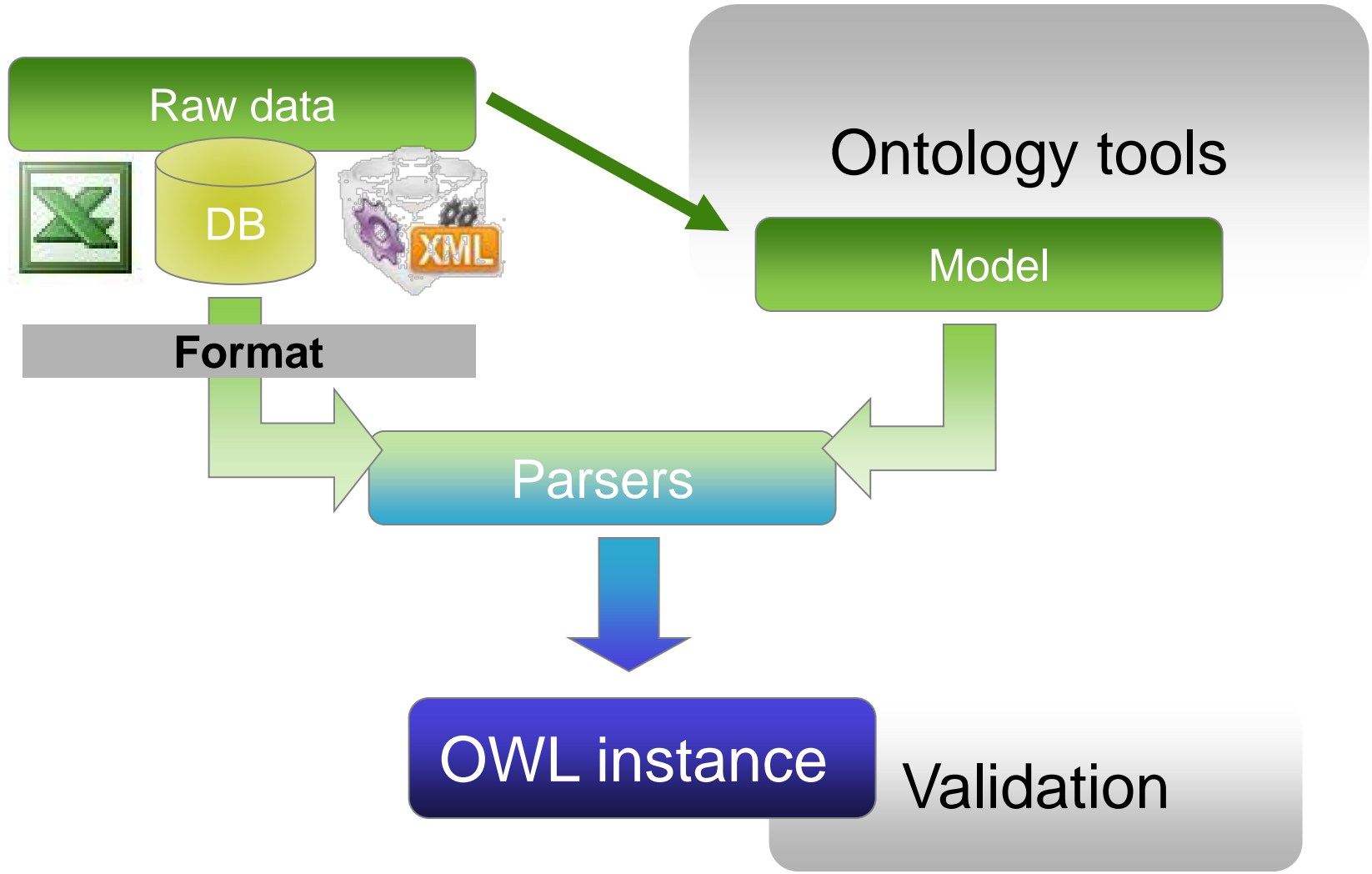
# Modèle générique HeTOP

CISMef



- Compatible avec les recommandations de Tao et al. (2013) et norme ISO 25964 (thésaurus)

# Méthodologie d'intégration





# SIBM 2016

## Department of Biomedical Informatics

**CiSMF**

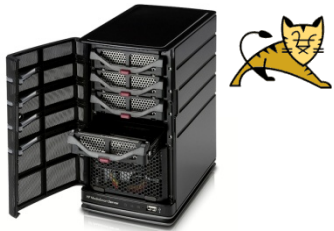
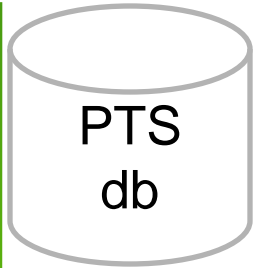
MDs	Engineers	University	Librarians	Secretaries
Stéfan <b>DARMONI</b> (Prof.), head	Badisse <b>DAHAMNA</b>	<b>Lina Soualmia (Senior lecturer, computer science)</b>	<b>Benoit THIRION (Head of the Medical Library)</b>	<b>Annie-Claude LANCELEVEE</b>
Jean-Philippe <b>LEROY</b>	Ivan <b>KERGOURLAY</b>	<b>Chloé CABOT (PhD student)</b>	<b>Catherine LETORD (Pharmacist, librarian)</b>	<b>Sandrine VOURIOT</b>
<b>Nicolas GRIFFON</b>	Julien <b>GROSJEAN</b> (PostDoc)	Wiem <b>CHEBIL</b> (PhD student)	Gaétan <b>KERDELHUE</b> (librarian)	
Matthieu <b>SCHUERS</b> (GP, PhD student)	Romain <b>LELONG</b> (PhD student)	Melissa Mary (PhD student, Grant CIFRE BioMérieux)	Léa <b>SEGAS</b> (librarian)	
<b>2 residents in Public Health</b>				
<b>1 ou 2 residents in GP</b>				
Philippe <b>MASSARI</b> (retired)				
	<b>Financement CHU</b>	<b>Financement Université Rouen</b>	<b>Financement Conseil Régional Normandie</b>	Financement sur projets de recherche

# HeTOP content

- HeTOP provides the usual data for each concept: preferred terms, original code, synonyms, definitions and other attributes, relations and hierarchies.
- Double (matricial) navigation:
  - among T/O
  - among languages
- Time consuming task > 20 man-years (to develop) + 2 man-years per year to maintain (integration & maintenance of T/O + mappings)
- Time consuming task to translate terminologies +++
- Several services on demand
  - access to other resources on the Internet (PubMed, CISMeF, etc.) through a French InfoButton (InfoRoute)
  - access to mappings tools (integrated in a beta version)
  - access to automatic indexing tool (ECMT)

# Methods & technologies (2)

CISMeF



- Oracle 11.1g (optimizations & partitioning) => NoSQL in 2015

- Java J2EE
- CISMeF APIs

- Apache Tomcat
- Infinispan cache layer



- Cross-browser (Vaadin framework)  
=> new framework in 2017 (INSA Rouen Engineering School)





# Croslingual Health Multi-Terminology/Ontology Portal

- First version before HeTOP (French & English)
- URL: <http://pts.chu-rouen.fr/>
- Access for humans and computers (Web services)
  - Since September 2010, daily used by CISMeF team to index manually and automatically Web resources
  - Since January 2011, MeSH is freely available (500 unique users per working day)
  - Teaching tool: Rouen Medical School (since Sept. 2010) to teach anatomy and rare diseases
  - Terminology auditing: HPO/Orphanet
  - T/O translations into French: FMA, HPO, SNOMED CT, MEDLINEplus
- Restricted access to the other terminologies (2,250 registred)
- Cooperation with BioPortal: Clement Jonquet & Mark Musen (ANR Jeunes Chercheurs: project SIFR)

# Three main terminology servers in health

## UMLS

- NIH, Bethesda (USA)
- Plus de 150 T/O
- Essentiellement en anglais
- La référence internationale pour la diffusion, mais pas pour la consultation

## BioPortal\*

- NCBO, Stanford (USA)
- Plus de 400 T/O (beaucoup en biologie, avec peu de concepts)
- Essentiellement en anglais (silo pour les autres langues)
- La référence pour poster une ontologie

## HeTOP\*

- SIBM, Rouen (France)
- 70 T/O en 32 langues
- La référence inter-lingue (navigation entre les langues) et dans le monde francophone

\* Grosjean J et coll. An Approach to Compare Bio-Ontologies Portals. **Stud Health Technol Inform**, 2014;205:1008-1012.



# HeTOP: main figures

CISMÉ

May 2010

Terminologies & ontologies	Concepts	Synonymes	Définitions	Relations & hiérarchies
<b>25</b>	<b>&gt; 580 000</b>	<b>&gt; 840 000</b>	<b>&gt; 220 000</b>	<b>&gt; 1 200 000</b>

May 2011

Terminologies	Concepts	Synonymes	Définitions	Relations
<b>32</b>	<b>&gt; 980 000</b>	<b>&gt; 2 300 000</b>	<b>&gt; 220 000</b>	<b>&gt; 4 000 000</b>

April 2013

Terminologies	Concepts	Synonymes	Définitions	Relations
<b>45</b>	<b>≈ 1 620 000</b>	<b>≈ 3 700 000</b>	<b>≈ 220 000</b>	<b>≈ 5 500 000</b>

April 2016

Terminologies	Concepts in English	Concepts in French	Terms	Attributes	Relations
<b>75 (17 UMLS)</b>	<b>2,358,544</b>	<b>1,171,340</b>	<b>9,134,636</b>	<b>15,348,985</b>	<b>9,697,000</b>



# Main figures

**CiSM<sup>e</sup>F**

<b>Registered users</b>	<b>&gt; 2 600</b>
<b>traffic</b>	<b>15 000 hits/day (6 to 700 users per working day)</b>



# Terminologies in French AND in UMLS (2015AA)

CISMef

Source vocabulary (Translator) [N lang.]	Number of Strings (PT + syn. + acro.)	Number of CUIs (% Translation in French)
MeSH Fr* [14]	105,758	41,229
MedDRA Fr [9]	73,860	73,608
WHO-ART Fr [5]	3,631	3,091
MTHMST Fr	1,833	1,636
ICPC2 [19]	702	722 (100%)
*MeSH Fr HeTOP		
Descriptors (INSERM) [16]	105,274	27,747 (100%)
Supplementary Concepts (CISMeF) [2]	58,514	48,190 (20.13%)
Concepts (INSERM & CISMeF) [2]	108,594	100,661 (27.52%)

# Terminologies in UMLS NOT in French



Source vocabulary (Translator) [N lang.]	Number of Strings	Number of CUIs (% of translation Fr)
ICD10 Fr (WHO) [12]	26,337	12,143 (100%)
ICD10 PCS (CISMeF) [2]	7,297	7,297 (5%)
ICD9 Fr (WHO) [1]	10,716	7,356 (100%)
ICDO Fr (WHO+CISMeF) [3]	1,462	1,362 (100%)
ICF (WHO) [2]	1,496	1,495
FMA Fr (U. of Washington)	4,564	4,452
FMA Fr in HeTOP (CISMeF) [7]	16,631	16,488 (20.35%)
ICNP [2]	2,811	1,158 (92%)
SNOMED Int. [2]	139,792	96,756 (94.51%)
ATC (WHO) [3]	5,834	5,757 (100%)
SNOMED CT		174,754 (~50%)

# Terminologies in UMLS NOT in French

Source vocabulary (Translator) [N lang.]	Number of Strings Fr/En/% of translations	Number of CUIs via CISMeF mappings
WHO-ICPS [2]	424	105
RADLEX (CISMeF) [2]	9,576/42,313/22.61	240
LOINC (APHP & SFIL) [2]	58,950	58,500 (60.71%)
MEDLINEplus Fr (CISMeF & LIMSI) [2]	849	846 (100%)
NCIT Fr (CISMeF) [2]	76,204	68,600 (73.04%)

# Terminologies in French that are not included in UMLS

CISMeF



Source vocabulary (Translator) [N lang.]	Number of Strings	Number of CUIs
OMIM Fr (CISMeF) [2]	7,770	6,904 (88.84%)
HRDO (Orphanet) [2]	13,535/id/100	4,943
HPO (CISMeF) [2]	11,127/11,908/93.44	1,541
CCAM (procedure) [1]	10,121	0
BNPC (toxicology) [2]	93,857	11,261
Q-Codes	184	
... including interface terminologies	in biology and imaging	

Overall, number of distinct CUI with at least one French translation in HeTOP  
 ≈ **386,530** vs. ≈ **88,000** in UMLS (**x4.39**)

**108 millions** of RDF triplets in 2014





# HeTOP relationships (examples & numbers)

**CISMeF**

	Source Term (Terminology)	Target Term (Terminology)	Number of relations in HeTOP
<b>UMLS</b> <sub>alignment</sub>	<i>Myocardial Infarction</i> <b>(MeSH)</b>	<i>Myocardial infarction, NOS</i> <b>(SNOMED Int)</b>	537,978
<b>CISMeF</b> <sub>manual</sub>	<i>Riedel thyroiditis</i> <b>(HRDO)</b>	<i>Riedel's thyroiditis</i> <b>(MedDRA)</b>	58,260
<b>CISMeF</b> <sub>exact</sub>	<i>appetite stimulants</i> <b>(ATC)</b>	<i>Appetite stimulated</i> <b>(WHOART)</b>	811,085
<b>CISMeF</b> <sub>Supervised</sub>	<i>Gonadotropin releasing hormone</i> <b>(MeSH)</b>	<i>Luteotropin-releasing factor</i> <b>(FMA)</b>	316,991



# HeTOP limits

- Formal representation of complex clinical data structures = none
- Formal representation of physiological models = none
- Temporal relations = none
- Data quality = based on T/O quality and point of view
- Formalism & reasoning capabilities = none
- Collaborative editing/searching/sharing tools = collaboration with BioPortal to share tools (Clement Jonquet)
- T/O versioning = not yet provided by HeTOP
- Semantic resources distribution/dissemination processes = 56 T/O available in OWL format (latest version)/SKOS/RDF in several languages

# Other tools integrated HeTOP

- ECMT Extracteur de Concepts Multi Terminologiques
  - Able to extract health concepts from any text; e.g. discharge summary in ½ second (NoSQL)
  - Valorization with Alicante SME
  - Used in daily practice in the Catholic University Hospital of Lille, France; Dr. Arnaud Hansske; around one million discharge summaries indexed with ECMT
- InfoRoute, a French InfoButton
  - URL: [inforoute.chu-rouen.fr](http://inforoute.chu-rouen.fr)
  - Access to a contextualized knowledge based on semantic expansion based on manual & supervised mapping among terminologies
- MT@HeTOP, tool to perform automatic mappings & translations
- Generic semantic search engine
  - Doc'CISMeF (URL: [doccismef.chu-rouen.fr](http://doccismef.chu-rouen.fr)) on grey literature about health in French on the Internet ( $10^5$  resources)
  - LISSA (URL : [www.lissa.fr](http://www.lissa.fr)), a PubMed in French ( $0,7 \times 10^6$  citations d'articles)
  - RIDOPI, search engine in EHR ( $8 \times 10^6$  discharge summaries in Rouen; around  $10^9$  health concepts in these summaries;  $10^8$  numerical data in Rouen)

# Semantic harmonization: mapping, alignment



Three methods employed

URL: [http://cispro.chu-rouen.fr/MT\\_EHTOP/](http://cispro.chu-rouen.fr/MT_EHTOP/)

- Conceptual
  - Same CUI
  - Other relations: close match, BT-NT, NT-BT (SKOS)
  - On UMLS (n=17 included in HeTOP)
- NLP
  - More or less same algorithm of automatic indexing
  - Bag of words
  - on  $n*(n-1)/2$  T/O (included in the HeTOP)
- Statistical
  - Co-occurrence matrix
  - CCAM-ICD10; CCAM-LPP

asthma

examples: asthme, asthma, D001249.nu., asth,

Terminologies selection

Your queries (1)

Results

▼ MeSH (10)

▼ MeSH Descriptor (7)

- anti-asthmatic agents
- asthma**
- asthma, Aspirin-Induced
- asthma, exercise-induced
- asthma, occupational
- conyza
- status asthmaticus

▼ MeSH Supplementary Concept (3)

- asthmalgine
- asthmasedine
- NPSR1 protein, human

► Record concept(s) (1) ✕

▼ Automatic exact mapping(s) (from CISMeF team) (3) ✕

Asthma	NCIt concept
Asthma (disorder)	SNOMED CT concept
Asthma finding (finding)	SNOMED CT concept

▼ Validated automatic narrower mappings (3) ✕

493.9 - asthme, sai	T_DESC_CIM9CODE
Asthma NOS	MedDRA Preferred Term
asthme bronchique	TUV Concept

▼ Curated CISMeF NLP mapping (13) ✕

.0804493 - asthme	T_DESC_CIM9CODE
asthma	ICPC-2 Descriptor
asthma	MedDRA Preferred Term
asthma	MedlinePlus Topic
Asthma	ICD-10 category
asthma	WHO-ART Preferred Term
Asthma	HPO term
Asthma (disorder)	SNOMED CT concept
asthma, nos	SNOMED Notion
Asthme	TUV Term
asthme	TUV Concept
asthme	DRC Consultation result
asthme	DRC RCE

NLP  
Bag of words

► False automatic mappings (3) ✕

▼ UMLS correspondence (same concept) (9) ✕

asthma	ICPC-2 Descriptor
Asthma	ICD-10 category
asthma	MedDRA Preferred Term
asthma	WHO-ART Preferred Term
asthma	MedlinePlus Topic
Asthma	NCIt concept
Asthma (disorder)	SNOMED CT concept
asthma, nos	SNOMED Notion
Asthma, unspecified	ICD-10 Sub-category

Conceptual  
Same CUI