Integrating consumer-oriented vocabularies with selected UMLS ones through ICPC2 and Semantic Web Technologies

Elena Cardillo

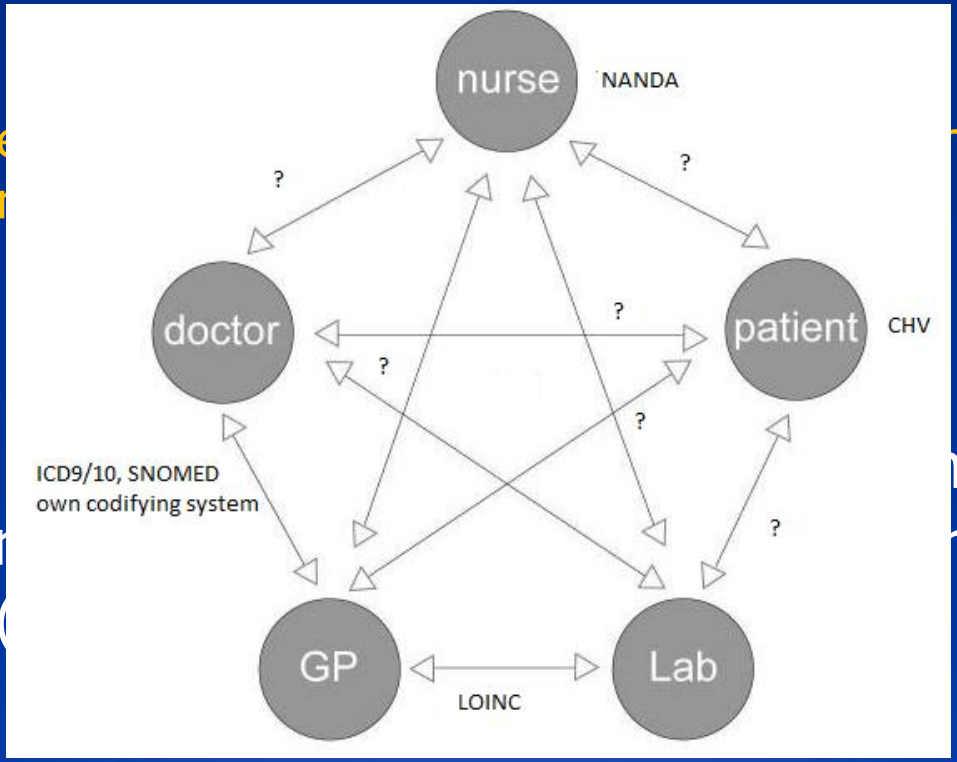
40° WICC Annual Meeting
Ravello, October 29 - November 02, 2012

Introduction

- Huge effort in integrating medical terminologies by creating mappings between them
- Use of Semantic Web Technologies
- More emphasis on the patient perspective
  - Personal Health Records accessible from the web
  - Active role played by consumers
  - Development of consumer-oriented vocabularies

Critical Issues

- Interoperability is still a significant problem
- Medical li
  - Difference nd profession
    - Epistaxis
    - Dyspnea
- Need for al medical vocabular rmation systems (



Objectives

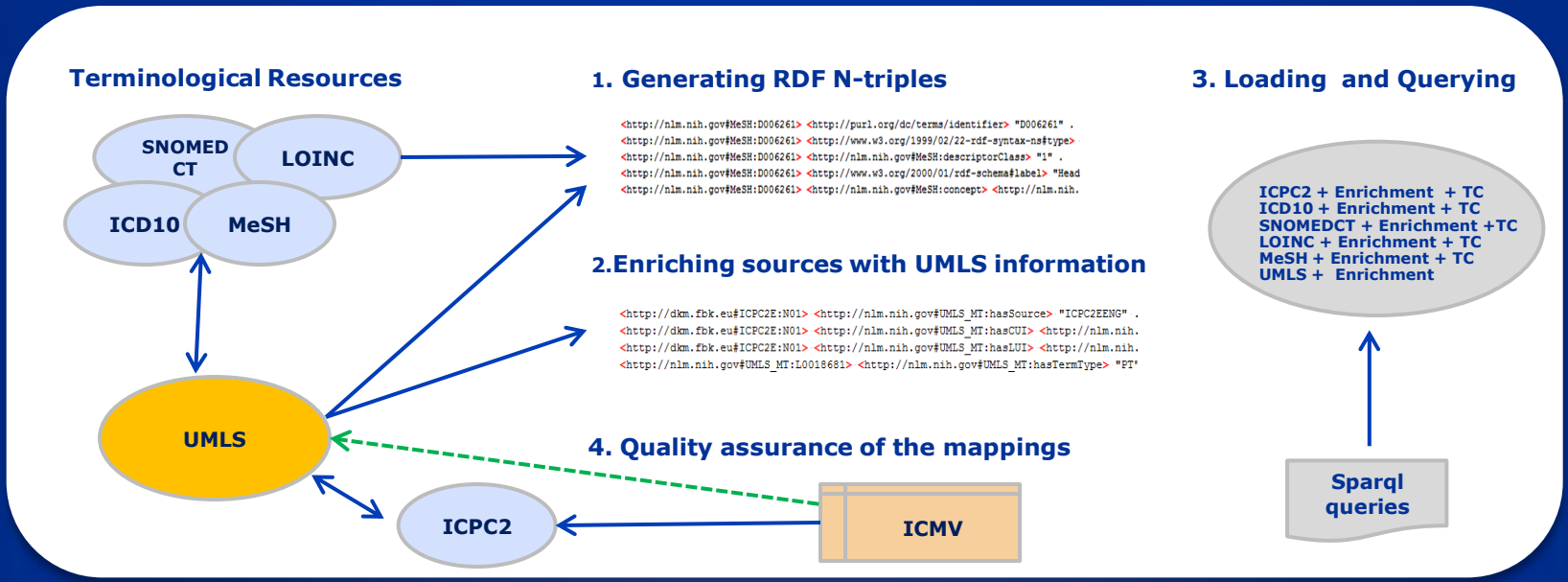
- To create an Integration Framework for the General Practice domain
  - Map consumer-oriented vocabularies to standard professional medical terminologies
- Why:
  - Mitigate the linguistic gap between lay and professional resources
  - Facilitate querying and searching of healthcare information
  - Improve consumer-oriented healthcare information systems
- How:
  - Using UMLS as a source of mappings between medical terminologies
  - Using Semantic Web technologies for integration purposes

Materials

Name	Version used	Statistics
ICMV - Italian Consumer Medical Vocabulary	v.2011 Tables and RDF version	2,348 Italian lay terms 1,521 terms mapped to ICPC2
SNOMED CT	RDF v. July 2009 (created at NLM)	308,000 active concepts 951,000 relationships 792,000 descriptions
MeSH Thesaurus	RDF v. 2009 (created at NLM)	25,186 descriptors (main headings) 83 qualifiers 180, 682 supplementary concepts records
ICD-10	OWL v. 2008	40,869 concepts (classified in 21 Chapters)
ICPC-2	OWL v. 2008	722 concepts (classified in 3 Components and 17 Chapters)
LOINC	v. 2.27 2009 RDF version	50,809 tests and observations 44,314 "part" concepts
UMLS	Partial RDF v. 20009AB (created at NLM)	10 million names 2.2 million concepts 10 million relations 100 families of biomedical vocabularies

Approach

- ICPC2 serves as a pivot between ICMV and other vocabularies in the UMLS
- UMLS Metathesaurus to provide mappings between ICPC2 and SNOMED CT, MeSH, LOINC and ICD10



Step 1.
Generating RDF N-triples

- Medical terms and their inter-relations represented by RDF N-Triples
  - SNOMED CT and MeSH already converted to RDF

    <SNOMEDCT:37796009> <hasLabel> "Migraine" .
    <MeSH:MeSH:D008881> <hasConcept> <MeSH:M0013864> .
  - OWL resources (ICPC2 and ICD10) serialized in RDF

    <ICPC2E:N89> <hasICD10Correspondent> <ICD10#G43> .
    <ICD10:G43> <SubClassOf> <ICD10:G40-G47> .
  - Java program to create RDF triples for LOINC from data in the UMLS Metathesaurus

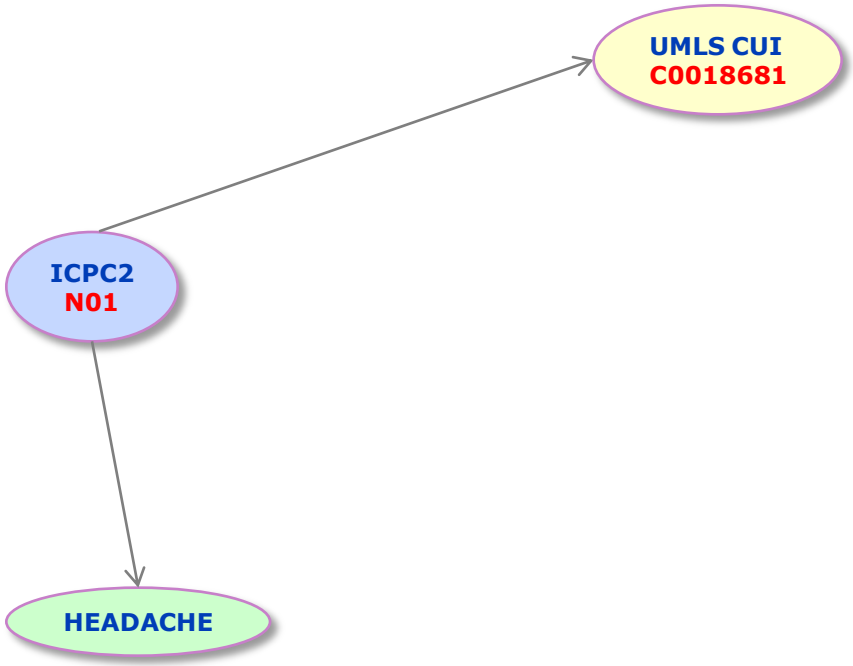
    <LOINC:LP74908-2> <hasLabel> "Headache" .
    <LOINC:1575-0> <hasComponent> <LOINC:LP15292-3>
  - Java program to encode ICMV in RDF from Excel tables

    <ICMV:Cuore in gola> <hasICMVCode> "ICMV559" .
    <ICMV:Cuore in gola> <rdfs:SubClassOf> <ICMV:Sintomo>
- Creation of 6 RDF graphs for a total of 2.1M RDF triples

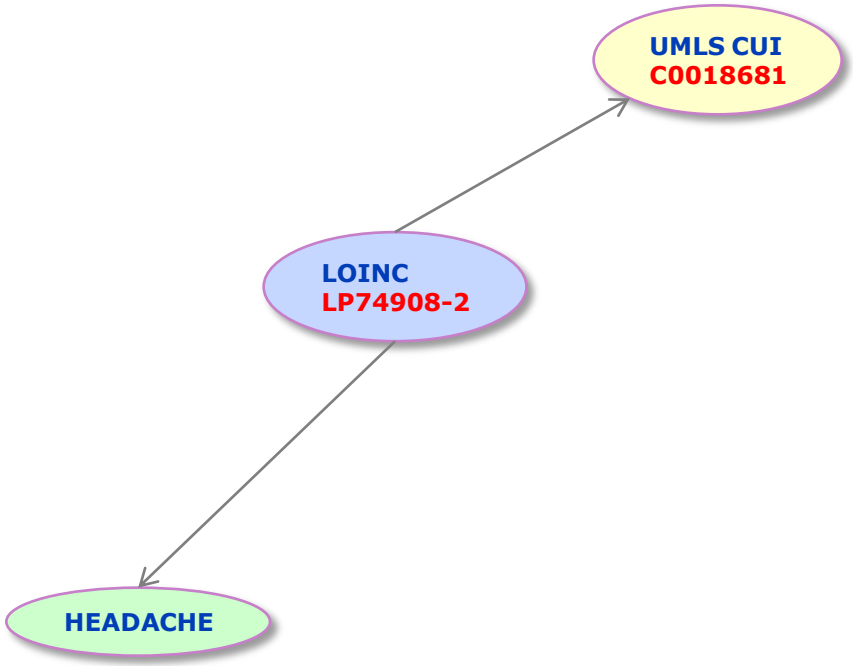
Step 2.
Enriching sources with UMLS attributes

- Objective
  - Enriching each terminology with UMLS attributes to facilitate term comparisons among vocabularies
- UMLS Attributes
  - Concept unique identifier (CUI)
  - Identifier for normalized strings (LUI)
- Automated N-triples creation and extraction of CUIs and LUIs from the MRCONSO table

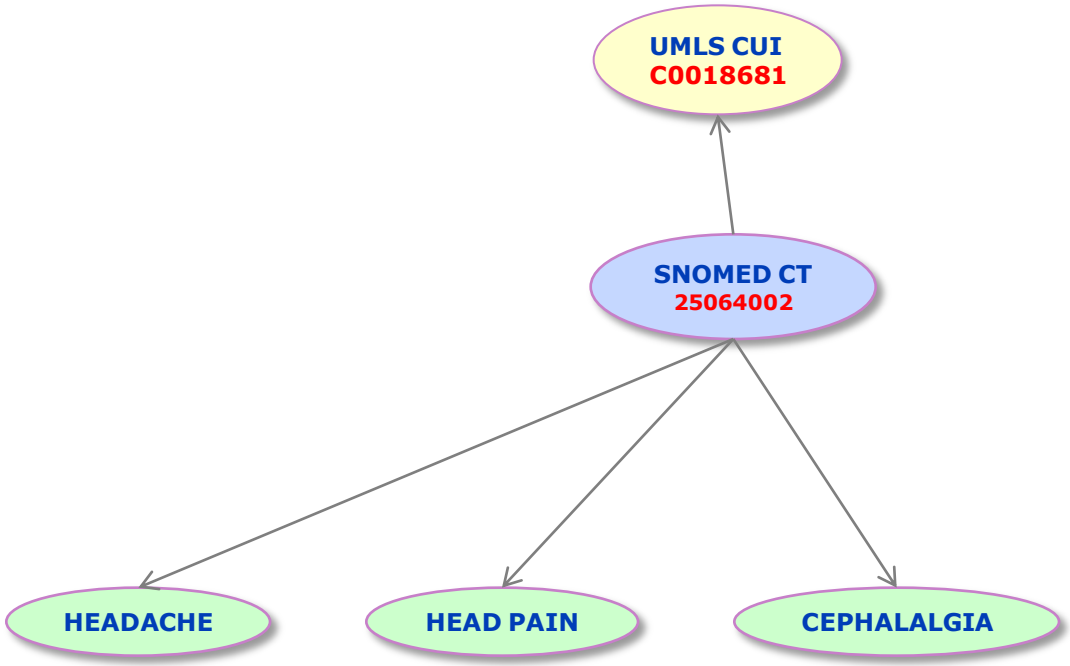
ICPC2



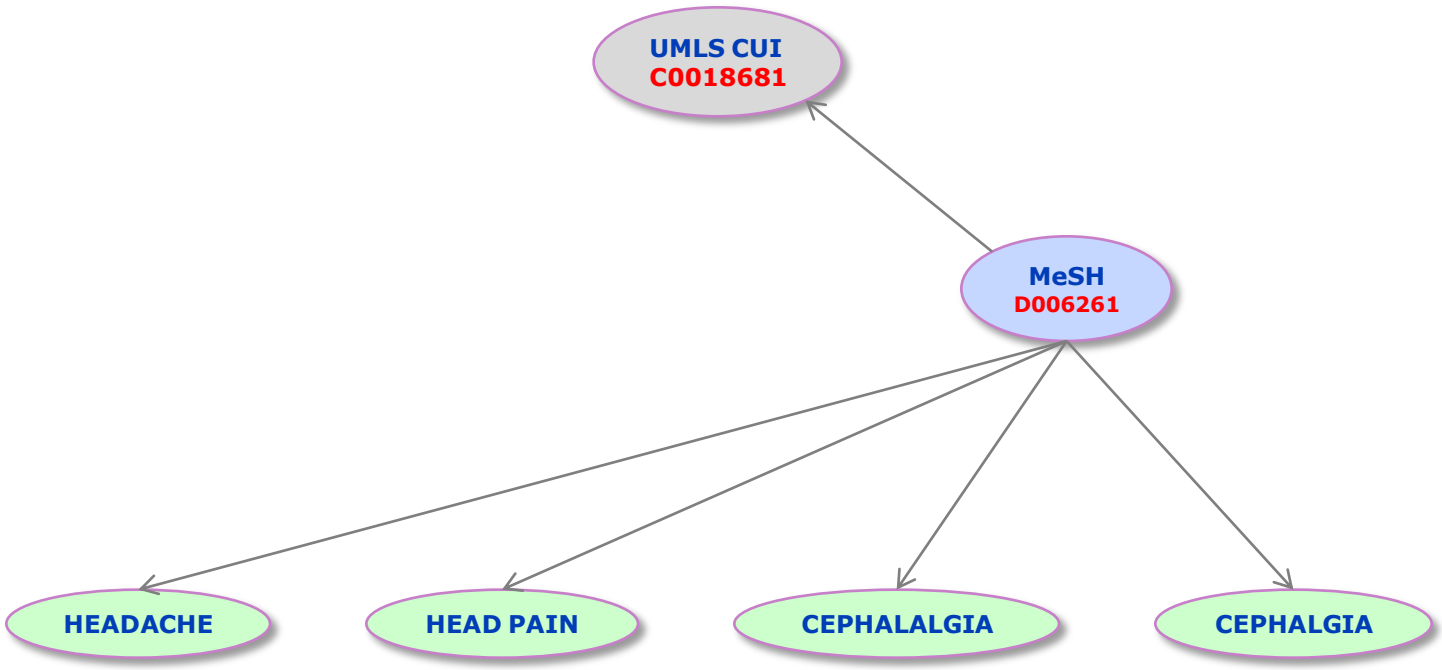
LOINC



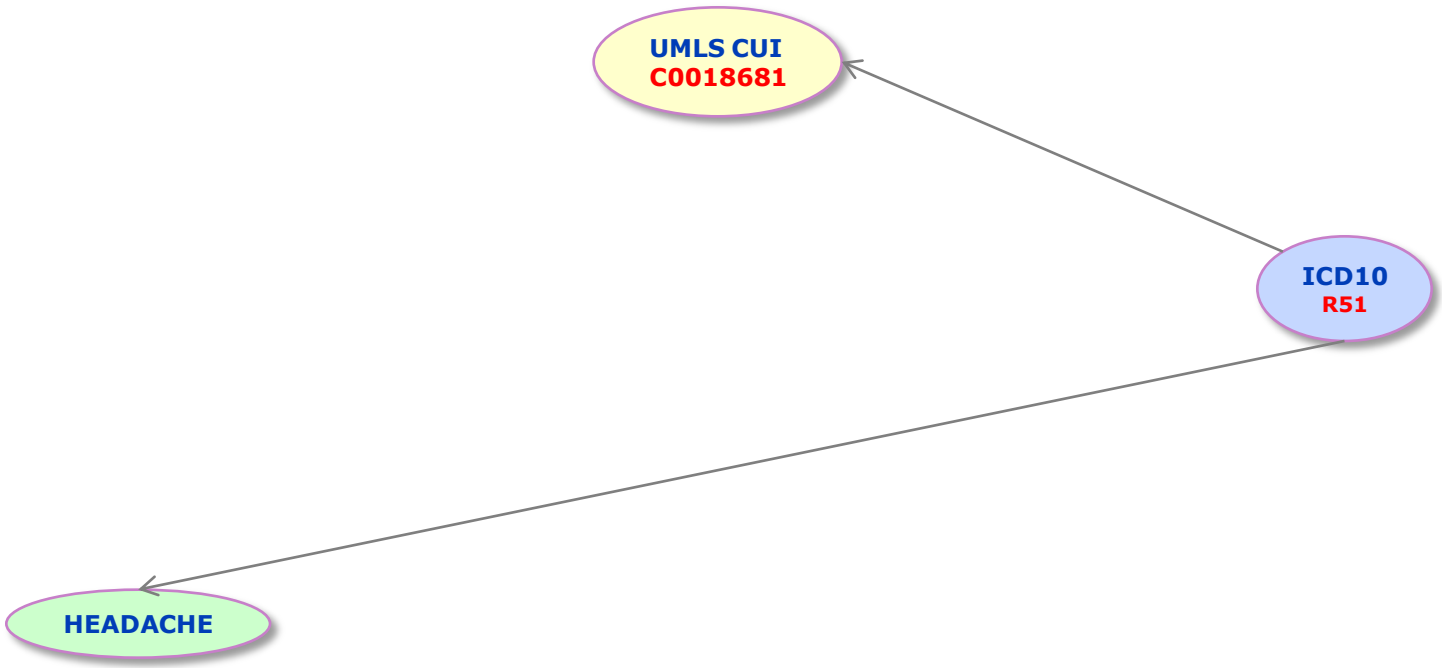
SNOMED CT


UMLS CUI
C0018681
SNOMED CT
25064002
HEADACHE
HEAD PAIN
CEPHALALGIA

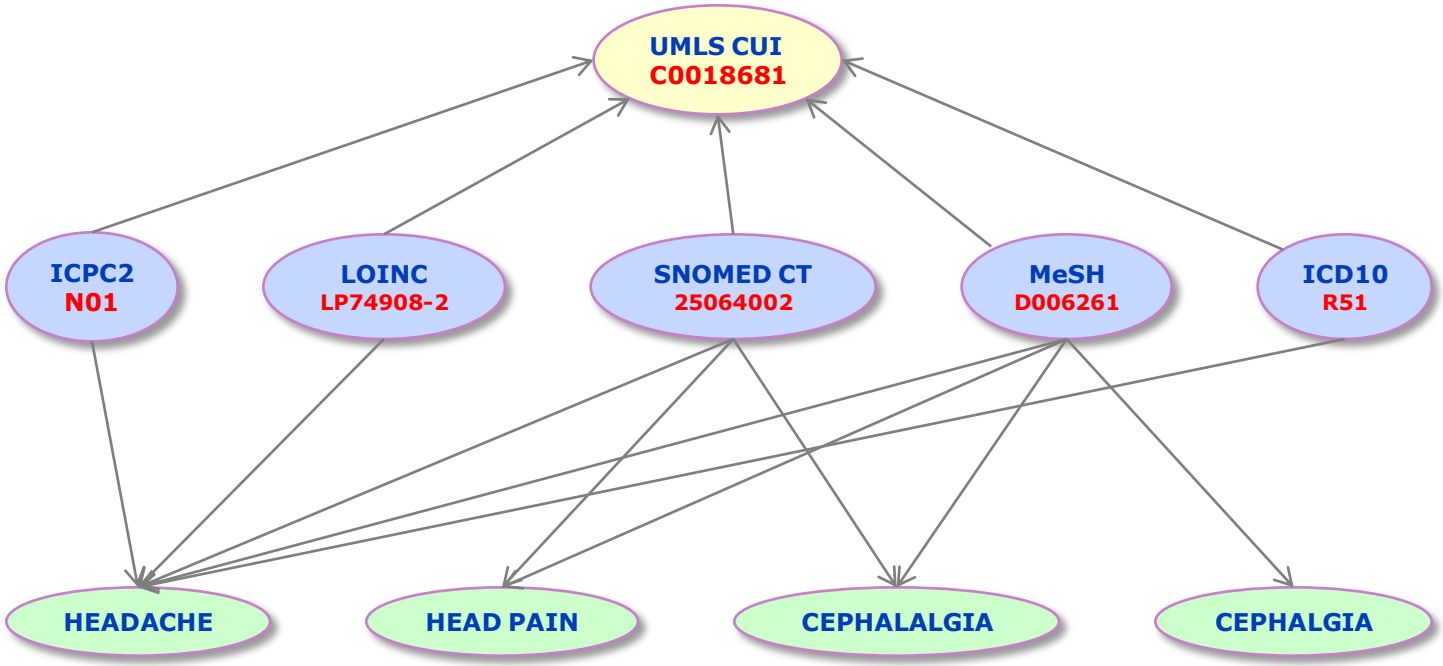
MeSH


UMLS CUI
C0018681
MeSH
D006261
HEADACHE
HEAD PAIN
CEPHALALGIA
CEPHALGIA

ICD10



All sources


UMLS CUI
C0018681
ICPC2
N01
LOINC
LP74908-2
SNOMED CT
25064002
MeSH
D006261
ICD10
R51
HEADACHE
HEAD PAIN
CEPHALALGIA
CEPHALGIA

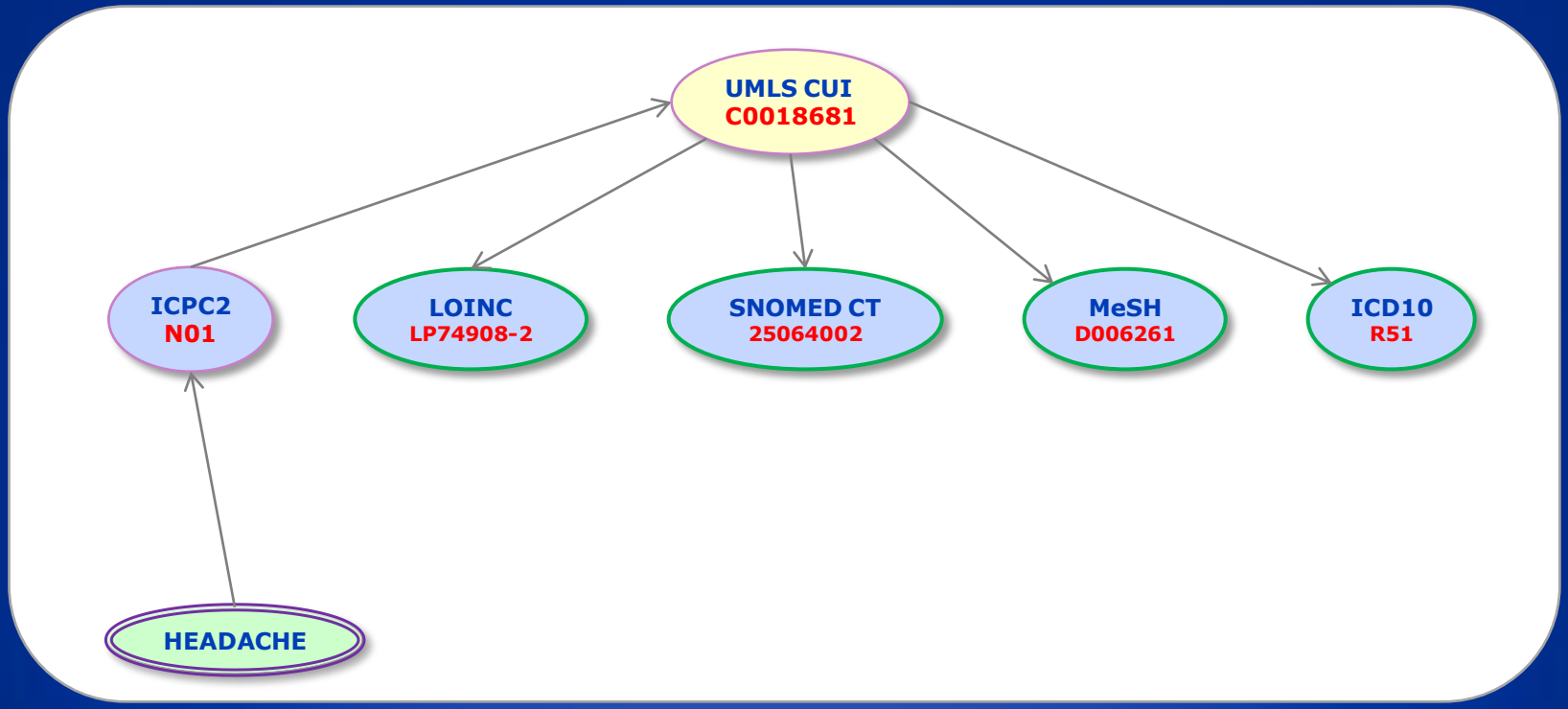
Step 3.
Querying Process

- Use of Openlink Virtuoso (v. 6.0) as RDF triple store
  - 18 graphs loaded
- 3 types of queries:
  1. Find concepts corresponding to ICPC2 concepts, using CUIs
  2. Find synonyms/new names corresponding to ICPC2 concepts, using CUIs, LUIs
- Use of SPARQL as query language
- Automated submission of batch queries to Virtuoso for extracting mappings

Step 3.
Querying Process

Example

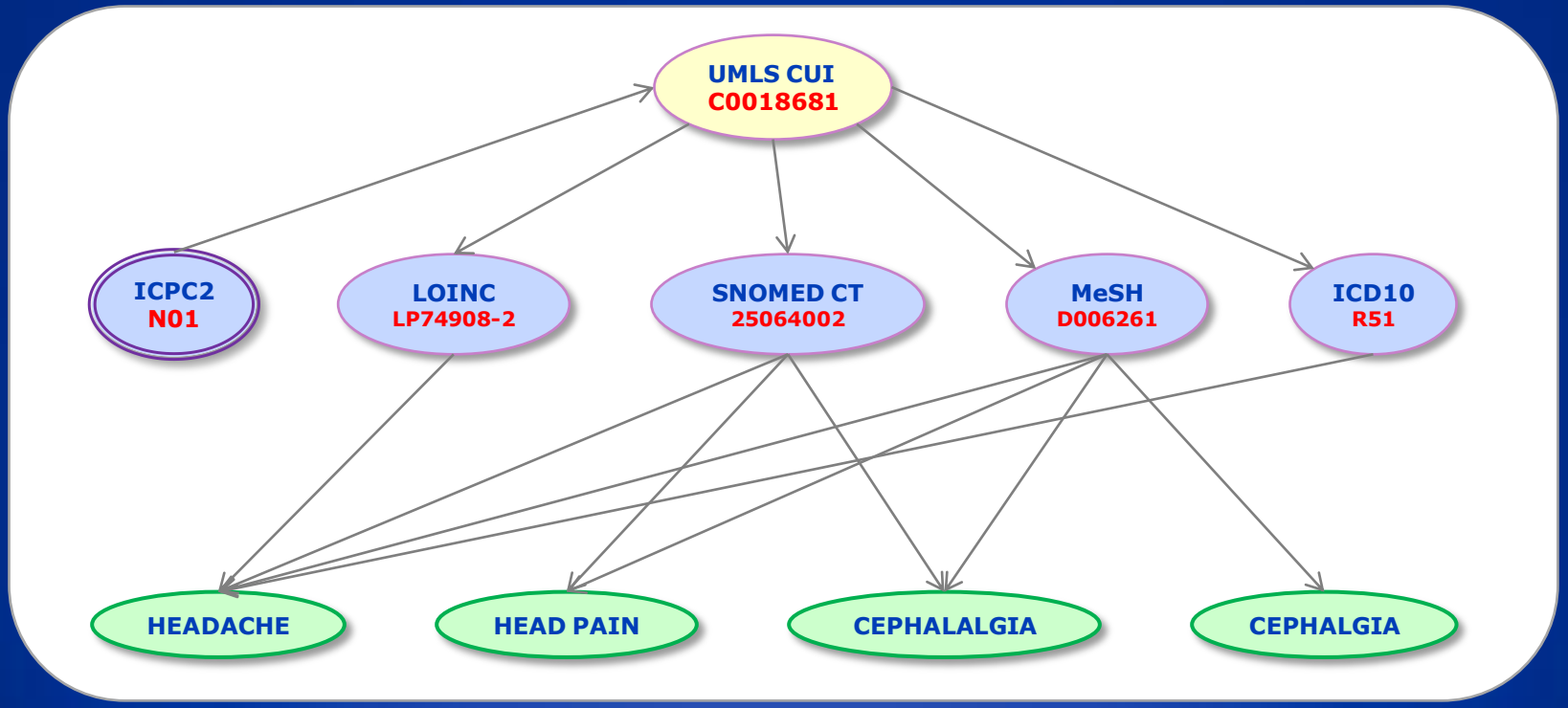
- Find all concepts in UMLS corresponding to ICPC2 Headache (N01):



Step 3.
Querying Process

Example

- Find all synonyms in UMLS corresponding to ICPC2 Headache (N01):



Step 3.
Querying Process

Example

- Find all concepts in UMLS corresponding to ICPC2 Headache (N01):

```
SPARQL
PREFIX ICPC2E: <http://dkm.fbk.eu#ICPC2E:>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX UMLS_MT: <http://nlm.nih.gov#UMLS_MT:>
SELECT ?icpc2_id ?label ?cui ?code
from <http://nlm.nih.gov/ICPC2E_to_UMLS_Enrichment>
from <http://dkm.fbk.eu/ICPC2E>
from <http://nlm.nih.gov/SNOMEDCT_to_UMLS_Enrichment>
from <http://nlm.nih.gov/LOINC_to_UMLS_Enrichment>
from <http://nlm.nih.gov/ICD10_to_UMLS_Enrichment>
from <http://nlm.nih.gov/MeSH_Enrichment>
WHERE
{
?icpc2_id rdfs:label ?label .
?icpc2_id UMLS_MT:hasCUI ?cui .
?code UMLS_MT:hasCUI ?cui .
filter(?icpc2_id = ICPC2E:N01)
};
```

Step 4.
Quality assurance of the mappings

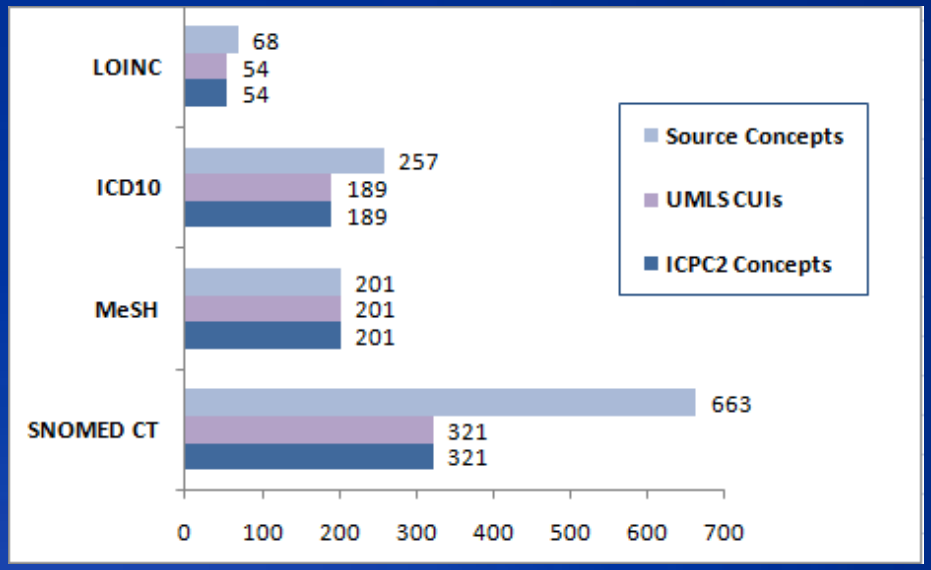
- Direct mapping of the ICMV "lay" terms to the Italian concepts in the UMLS Metathesaurus
  - Exact matching by using UMLSKS application programming interface
    - Output: ICMV term ⟶ UMLS CUI + Preferred Term + Source + Code
- Compare direct mappings through UMLS to the mapping through ICPC2 created by experts
- Compare the mappings to other sources
  - SNOMED CT, MeSH, ICD10, LOINC

Results

Finding mappings between ICPC2 and other UMLS resources

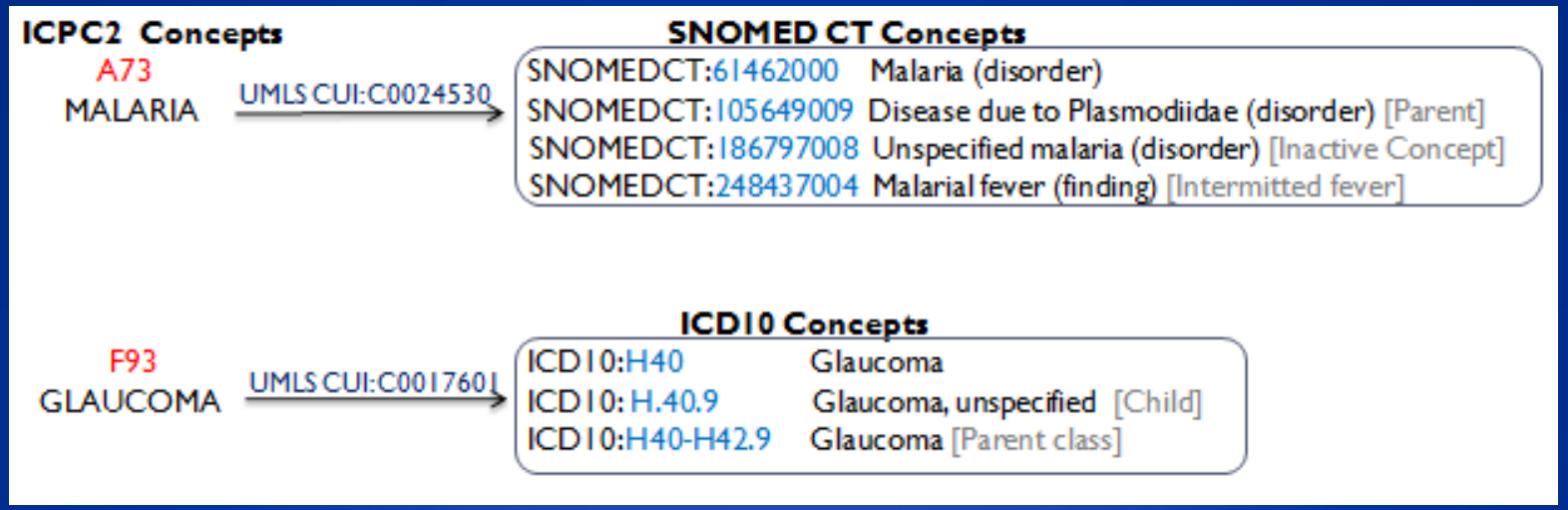
Starting from 760 ICPC2 Concepts

- 587 (77%) present in UMLS
- 251/587 (43%) Specific to ICPC2
- 336/587 (57%) mapped to the other resource
- 1189 unique source concepts

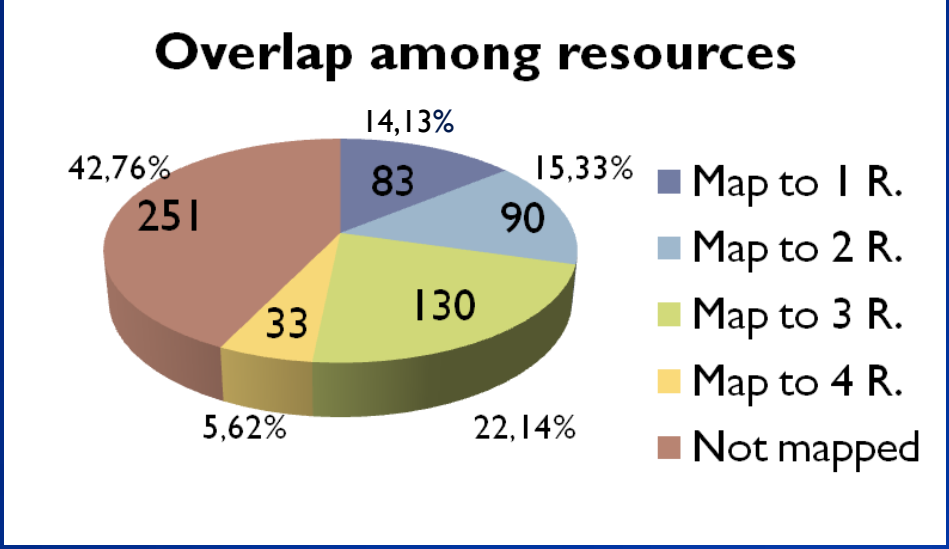


Results

- Large number of multiple mappings
  - 40% map to at least three SNOMED CT concepts and two ICD10 concepts
    - Pairs of SNOMED CT concepts are collapsed in the same UMLS CUI.



Results



Examples

- Map to 4 terminologies:
  A03 - Fever, F93 - Glaucoma
- Map only to SNOMED CT:
  A18 - Concern about appearance
- Map only to MeSH:
  N19 - Speech disorder
- Map only to ICD10:
  H77 – Sprain/strain of ankle

- Among the 83 mapped to only one terminology:
  - 74 map only SNOMED
  - 4 map only to MeSH
  - 5 map only to ICD10

Results

Finding synonyms in the other resources

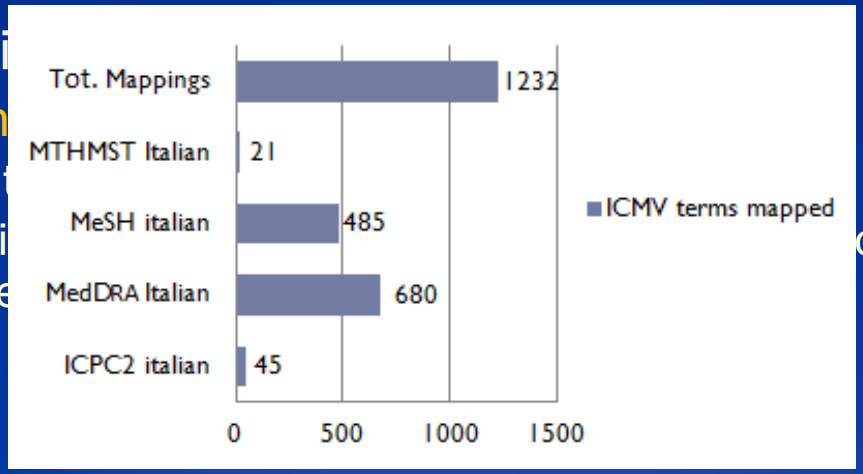
- Na… C2 co…
- A t…
  - 1…
  - 3…
  - 1…
  - 6…
- Syn…
  - A to… concepts
    - E.g. 14 synonyms for the ICPC2 concept "U04" – Incontinence Urine
  - 739 additional Lexical Variants for the ICPC2 concepts

ICPC2 Concept	LUIs	New Synonyms	SNOMED CT
U04 - Incontinence Urine /PT			
	L0042024	*Urinary Incontinence*	
	L0005685	Absence of bladder continence	
	L0308837	Involuntary urination	
	L0527266	Unable to control bladder	
	L0527384	Weak bladder	
	L0583725	Unable to hold urine	
	L0005692	Bladder incontinence	
	L0527301	Unable to prevent bladder emptying	**165232002**
	L0574730	Unable to hold fluids	
	L0586619	UI - Urinary incontinence	
	L0748747	Bladder: incontinent	
	L0042024	Incontinence, urinary	
	L0527264	Lack of bladder control	
	L0527265	Loss of bladder control	
	L0590897	Leaking of urine	

Results

Quality assurance of the mapping:ICMV-2-UMLS

- Mapping ICMV terms to UMLS Italian concepts
  - 655 ICMV terms mapped to 690 unique UMLS CUIs
  - 1,232 total mappings to the UMLS Italian concepts
- Ambiguity i
  - Concept n
    - E.g. the t
    - Anticorpi ophil Cytoplasmic Antibodie



Results

Quality assurance of the mapping:ICMV-2-UMLS

- Reuse of the CUIs mapped to ICMV for enriching the ICMV graph
- ICMV graph loaded into Virtuoso and queried among other resources (both in English and Italian)
  - Mappings extended to 1,990 shared concepts in the other terminologies:

Type of Mapping	UMLS CUIs	ICMV terms	ICPC2 concepts	Other sources
ICMV-2-UMLS Italian	690	655	45	1187
ICMV-2-ICPC2 Manual	0	1521	572	0
ICMV-2-UMLS via ICPC2RDF	336	523	587	1773
ICMVRDF-2-ICPC2 via UMLS	570	559	90	1903

    - 200 derived from LOINC

Conclusions

- ICPC2 integrated with SNOMEDCT, ICD10, MeSH, LOINC using RDF and SPARQL queries
  - 50% of ICPC2 concepts mapped to at least one other terminology
  - Many multiple mappings, that is "ambiguity"
- New mappings btw ICMV and professional terminologies
- Comparing Manual vs Automated mapping:
  - Use of UMLS as a bridge considered as the best way to integrate ICMV with other medical terminologies
  - Manual mapping performed by physicians most profitable for mapping ICMV to ICPC2

Thanks for you for attention

Elena Cardillo
cardillo@fbk.eu
http://ehealth.fbk.eu