Using ISO and Semantic Web standards for creating a Multilingual Medical Interface Terminology: A use case for Heart Failure

Elena Cardillo
Fondazione Bruno Kessler

Co-authors :
Maxime Warnier (UCL)
Marc Jamoulle (UCL)
Joseph Roumier (CETIC)
Robert Vander Stichele (UGENT)

UNIVERSITEIT GENT

language and translation technology team

FONDAZIONE BRUNO KESSLER

cetic
Your Connection to ICT Research

UCL Université catholique de Louvain

Outline

- Context
- Objective
- Approach
  - Building a multilingual Reference Terminology in TMF
  - Building unilingual End-user Lexicons in LMF / Lemon
- Implementation of the use case
- Conclusion

Context

- Efficient encoding and retrieval of medical data have become two crucial needs for clinicians
- Classification systems are not well integrated in EHRs and links to them hampered by a translation gap
- Terminological solutions are often naive:
  - Too overwhelming or too restrictive in concepts
  - Not addressing local, cultural differences between vocabularies
- Need for a more efficient interface between the human language and the machine language

Objective

Provide a hybrid healthcare interface terminology capable of

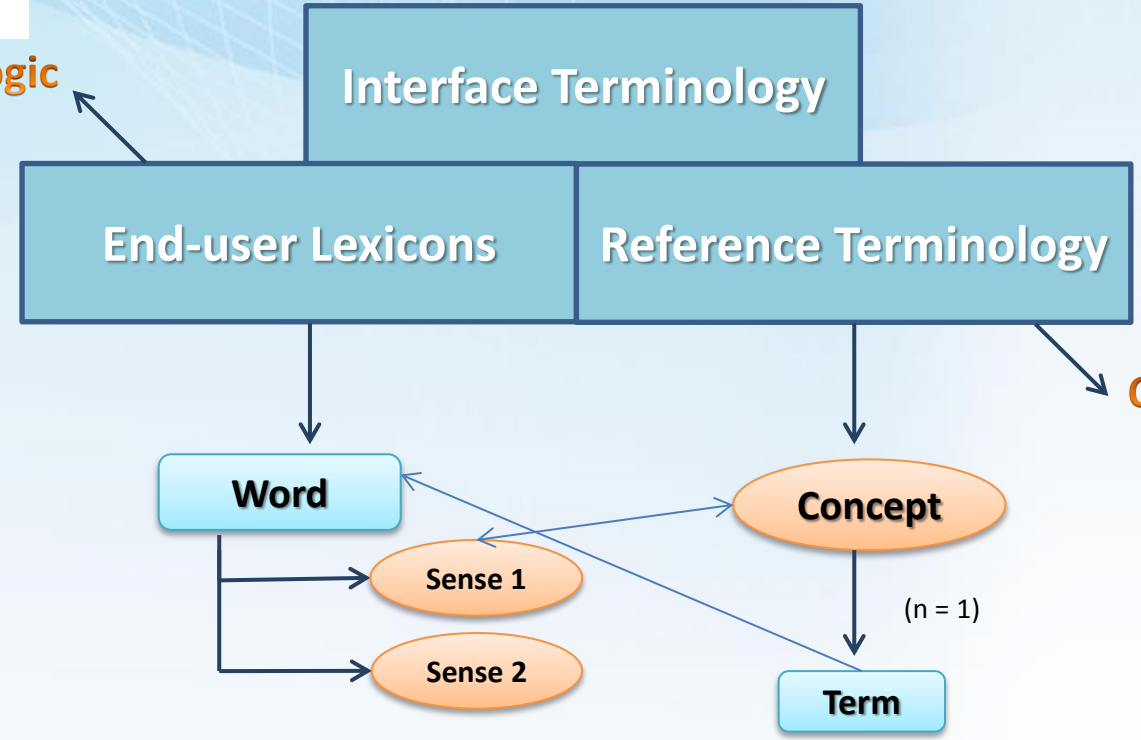
> ‘‘Systematic collection of healthcare-related phrases (terms) to support clinicians’ entries of patient-related information into computer programs …”
>
> *Rosembloom et al, 2006*

That includes:

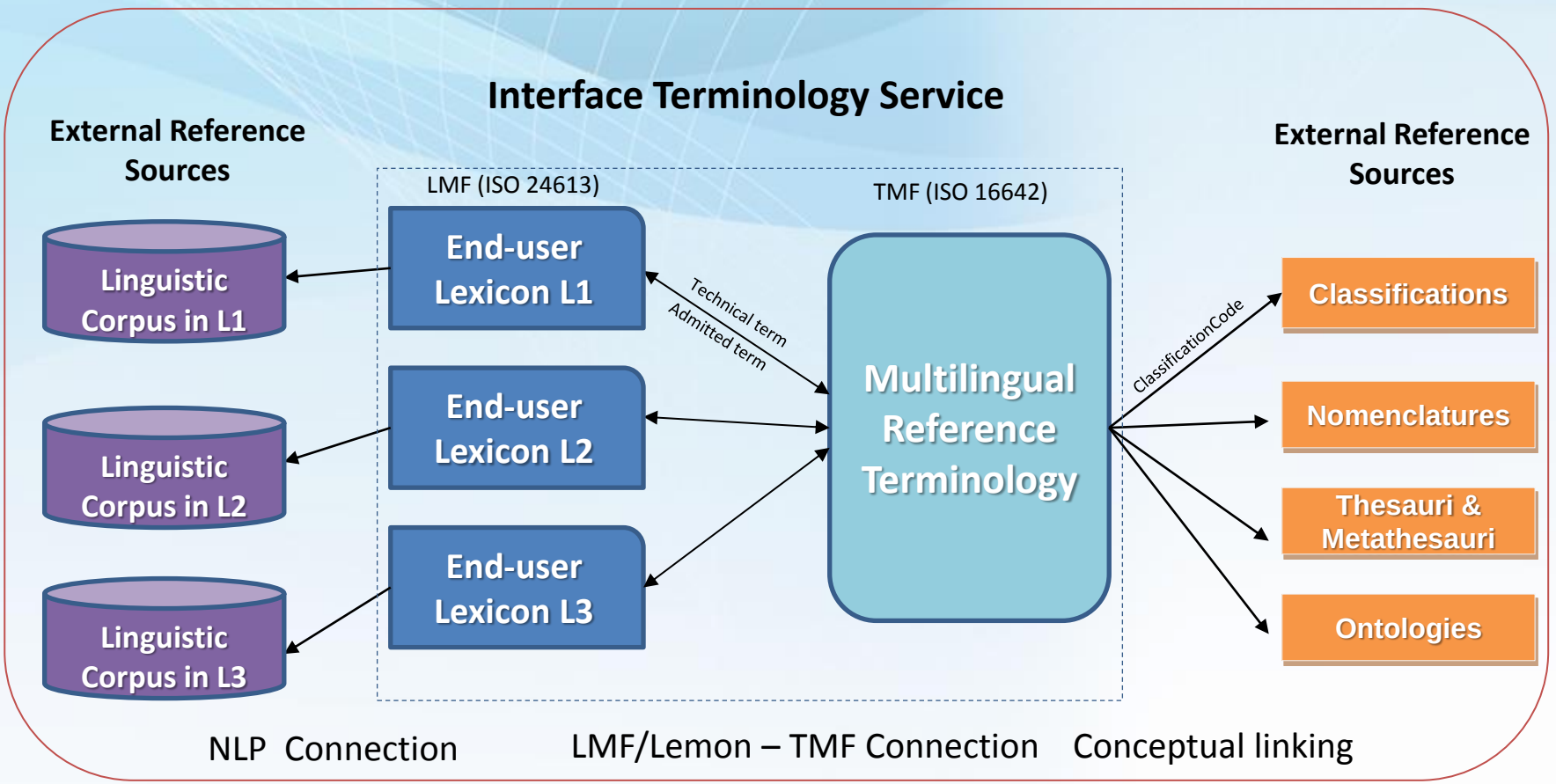
1. A conceptual reference terminology
   - Multidisciplinary, multilingual
   - Restricted to reference concepts
   - Linked to international nomenclatures, thesauri, classifications, ontologies
2. A series of specific end-user lexicons
   - One for each language
   - Bridged to the reference terminology
   - Connected to rich linguistic corpora for NLP

Approach

Semasiologic

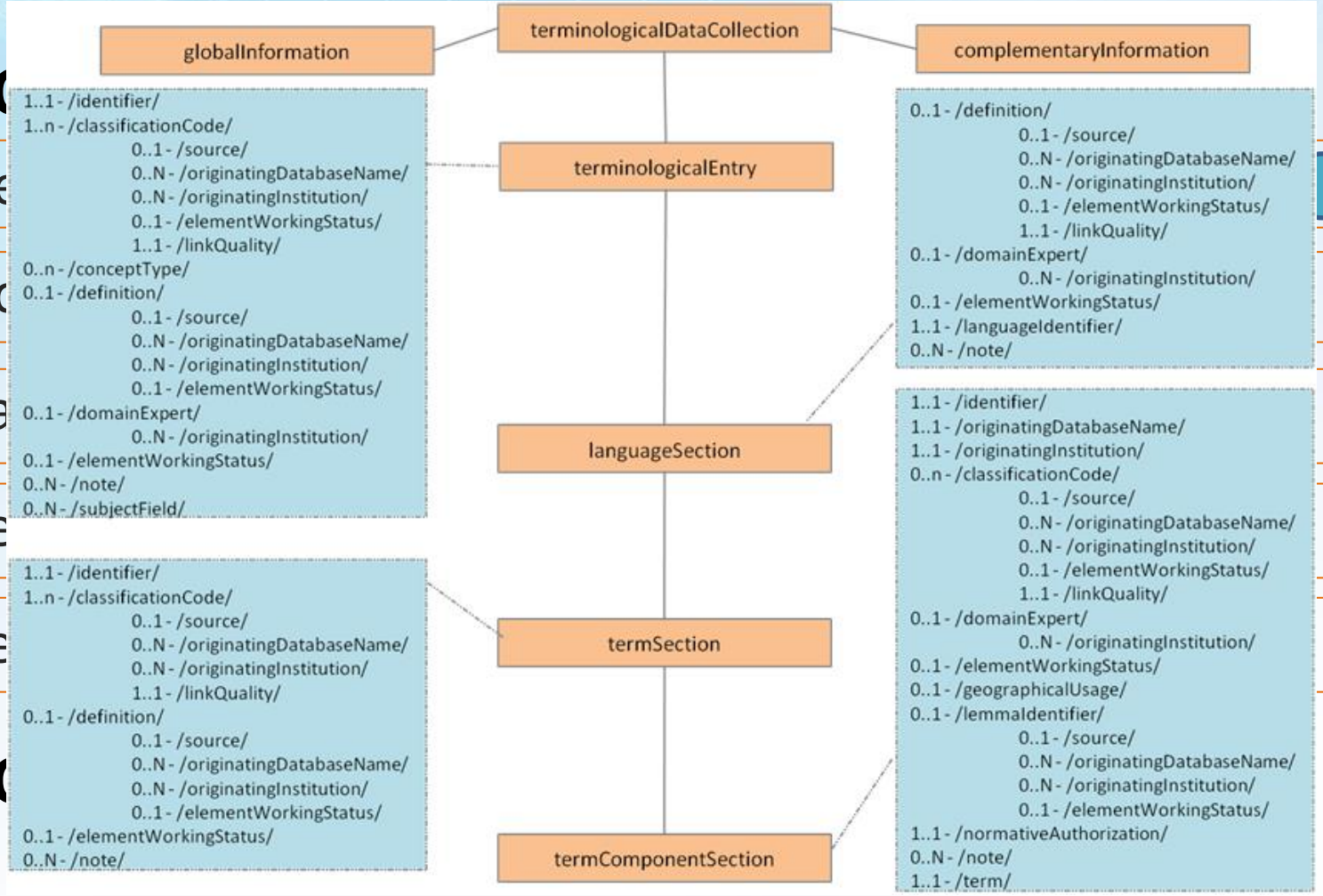


Architecture



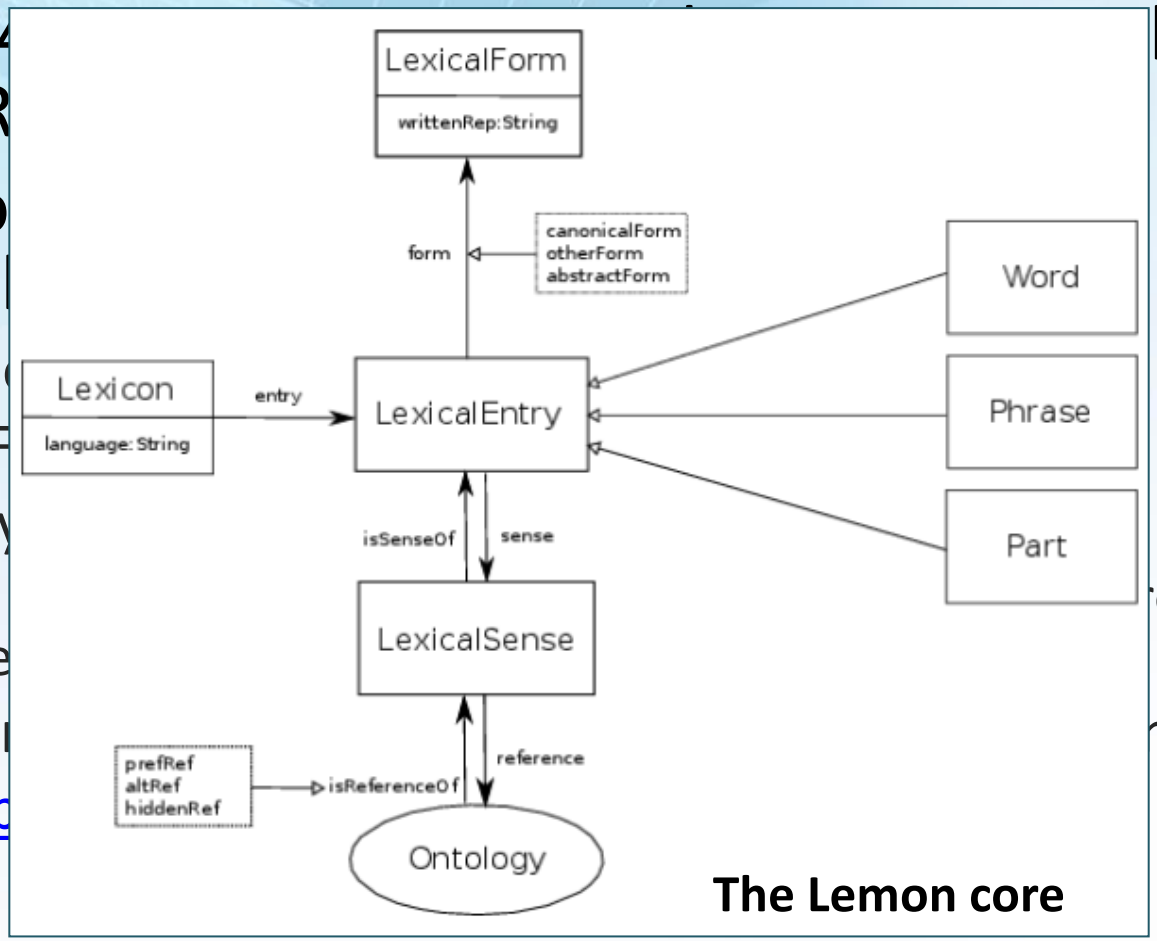
TMF-based Reference Terminology

- ISO
- ISO



Lemon/LMF-based End-user Lexicons

- ISO 24 k (LMF) as a R
- Use o for ONtol
  - Bas
  - RDF
  - Easy
  - The ence to an exte
  - Mo model
  - http



The Lemon core

OWL encoding of the TMF model

- Implementation of the TMF model as a series of OWL-DL ontologies using Protégé:

›Makes links between the first two;

›Adds some restrictions and links between classes and individuals;

›Defines cardinality for each data category;

›Characterizes properties;

›Assign domain values

› Use of *owl:equivalentClass* to link data categories to the corresponding ones on ISOcat.

End-user Lexicon in RDF

- RDF representation of the lexicon with Lemon;
- Links between phrases and their components are provided in the form of an ordered RDF list.
- Properties come from:
  - the lemon model itself, ISOcat; Dublin Core;
- Linguistic properties:
  - POS (only common nouns, adjectives and adverbs in our lexicon);
  - Canonical form (i.e. lemma) with a possible spelling variant;
  - Inflected forms and normative authorization;
  - Grammar gender and number
- Lexical sense/s defined outside the lexicon
  - E.g. in the Reference Terminology (for medical sense);
  - Polysemy and Synonymy can be inferred

Implementation of the use case

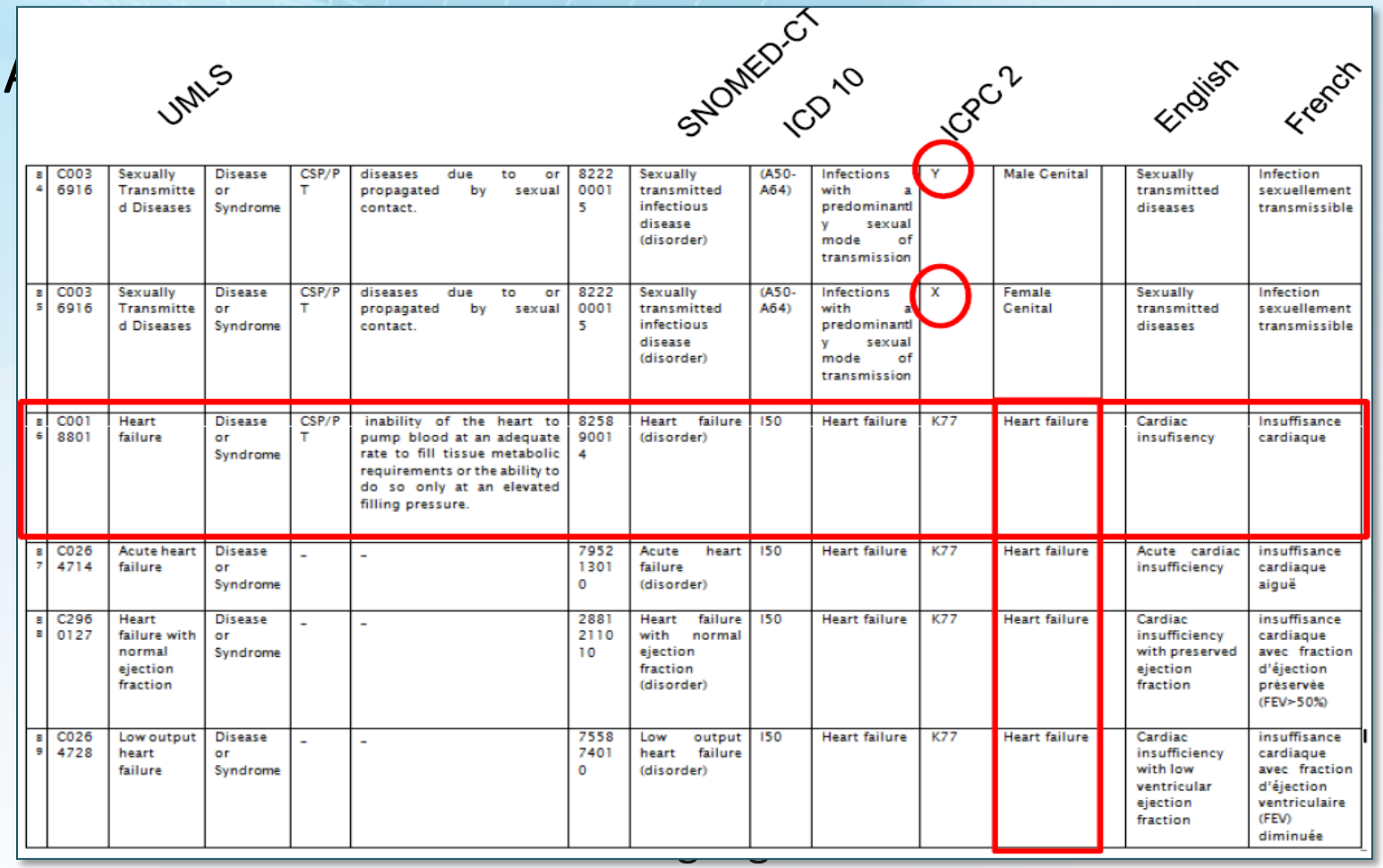
- Extraction of relevant concepts and words or phrases from a Belgian bilingual (Dutch and French) guideline on Heart Failure (Scientific Associations of Primary Care Physicians)
- French version as a starting point
- Two different extraction process:
  1. Manual selection of the relevant concepts by domain experts to populate the Reference Terminology
  2. Automated term extraction by a bilingual term extractor to populate the End-user Lexicon

Populating the Reference Terminology 1/2

- 168 clinical concepts manually selected by a domain expert
- Concepts entered as term entries in the *Term section* together with:
  - a definition in French;
  - a reference to the French original guideline;
  - a preferred term (E.g. "*ascite*");
  - one/more admitted terms (E.g. "*eau dans le ventre*")
  - Italian translation for each preferred and admitted term (in the Italian language section)

Populating the Reference Terminology 2/2

- A



	UMLS					SNOMED-CT		ICD 10	ICPC 2		English	French
84	C0036916	Sexually Transmitted Diseases	Disease or Syndrome	CSP/PT	diseases due to or propagated by sexual contact.	82220001 5	Sexually transmitted infectious disease (disorder)	(A50-A64)	Infections with a predominantly sexual mode of transmission	Y	Male Genital	Sexually transmitted diseases	Infection sexuellement transmissible
85	C0036916	Sexually Transmitted Diseases	Disease or Syndrome	CSP/PT	diseases due to or propagated by sexual contact.	82220001 5	Sexually transmitted infectious disease (disorder)	(A50-A64)	Infections with a predominantly sexual mode of transmission	X	Female Genital	Sexually transmitted diseases	Infection sexuellement transmissible
86	C0018801	Heart failure	Disease or Syndrome	CSP/PT	inability of the heart to pump blood at an adequate rate to fill tissue metabolic requirements or the ability to do so only at an elevated filling pressure.	82589001 4	Heart failure (disorder)	I50	Heart failure	K77	Heart failure	Cardiac insufisency	Insuffisance cardiaque
87	C0264714	Acute heart failure	Disease or Syndrome	-	-	79521301 0	Acute heart failure (disorder)	I50	Heart failure	K77	Heart failure	Acute cardiac insufficiency	insuffisance cardiaque aiguë
88	C2960127	Heart failure with normal ejection fraction	Disease or Syndrome	-	-	28812110 10	Heart failure with normal ejection fraction (disorder)	I50	Heart failure	K77	Heart failure	Cardiac insufficiency with preserved ejection fraction	insuffisance cardiaque avec fraction d'éjection préservée (FEV>50%)
89	C0264728	Low output heart failure	Disease or Syndrome	-	-	75587401 0	Low output heart failure (disorder)	I50	Heart failure	K77	Heart failure	Cardiac insufficiency with low ventricular ejection fraction	insuffisance cardiaque avec fraction d'éjection ventriculaire (FEV) diminuée

Manual mapping issues

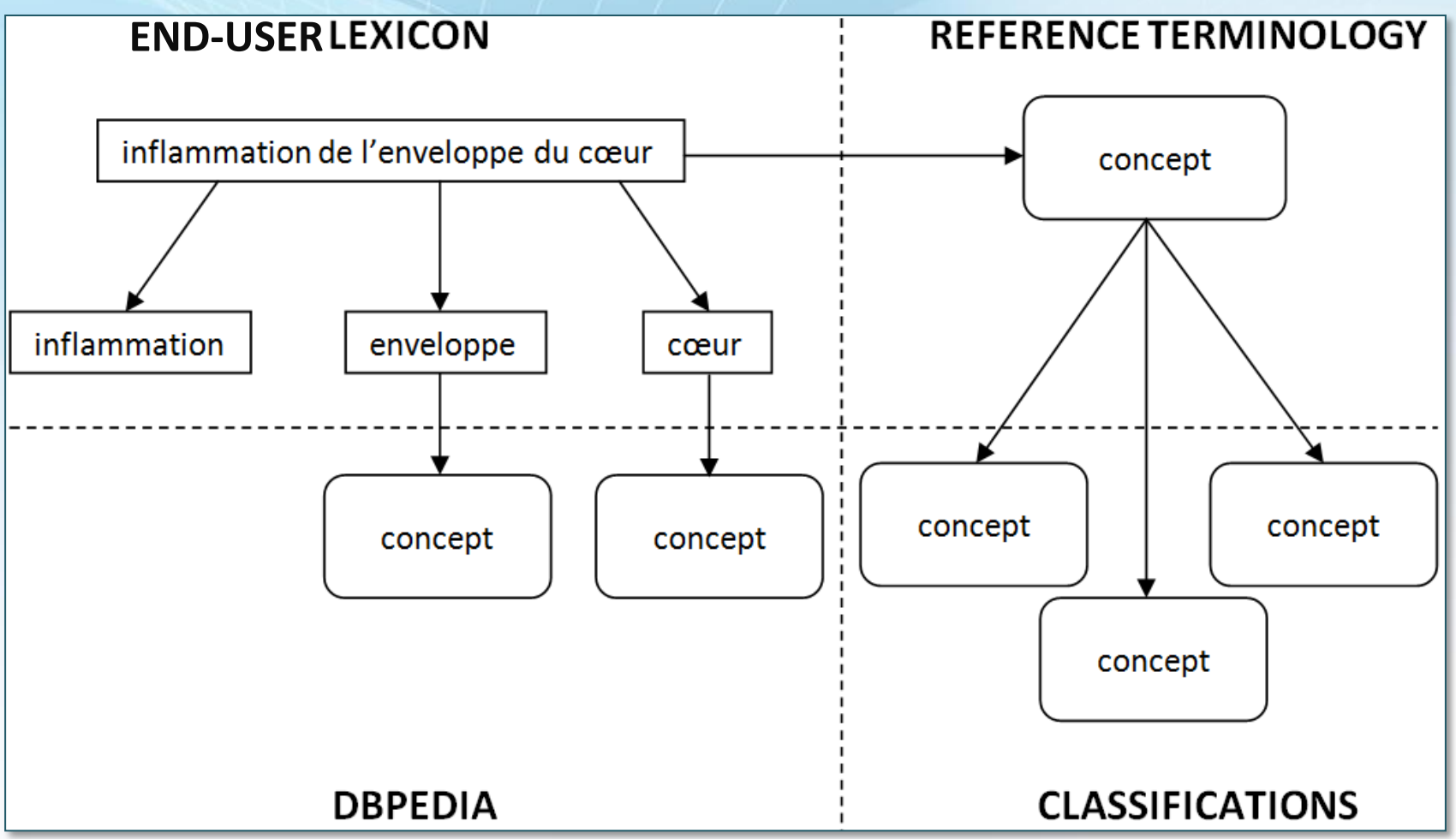
- Some concepts are too general
- No «Symptom and Complaints» in SNOMED-CT:
  - E.g. drug side effect *«dry mouth»*;
- Semantic types mismatching:
  - E.g. UMLS Finding ≠ SNOMED-CT Findings;
- Many *1:m* mapping to ICPC
  - E.g. *«Cardiomiopathy»* opened to 3 different ICPC rubrics;

External Resources	Reference Term. Concepts
ICPC	153
ICD	131
SNOMED-CT	161
UMLS CUI	116

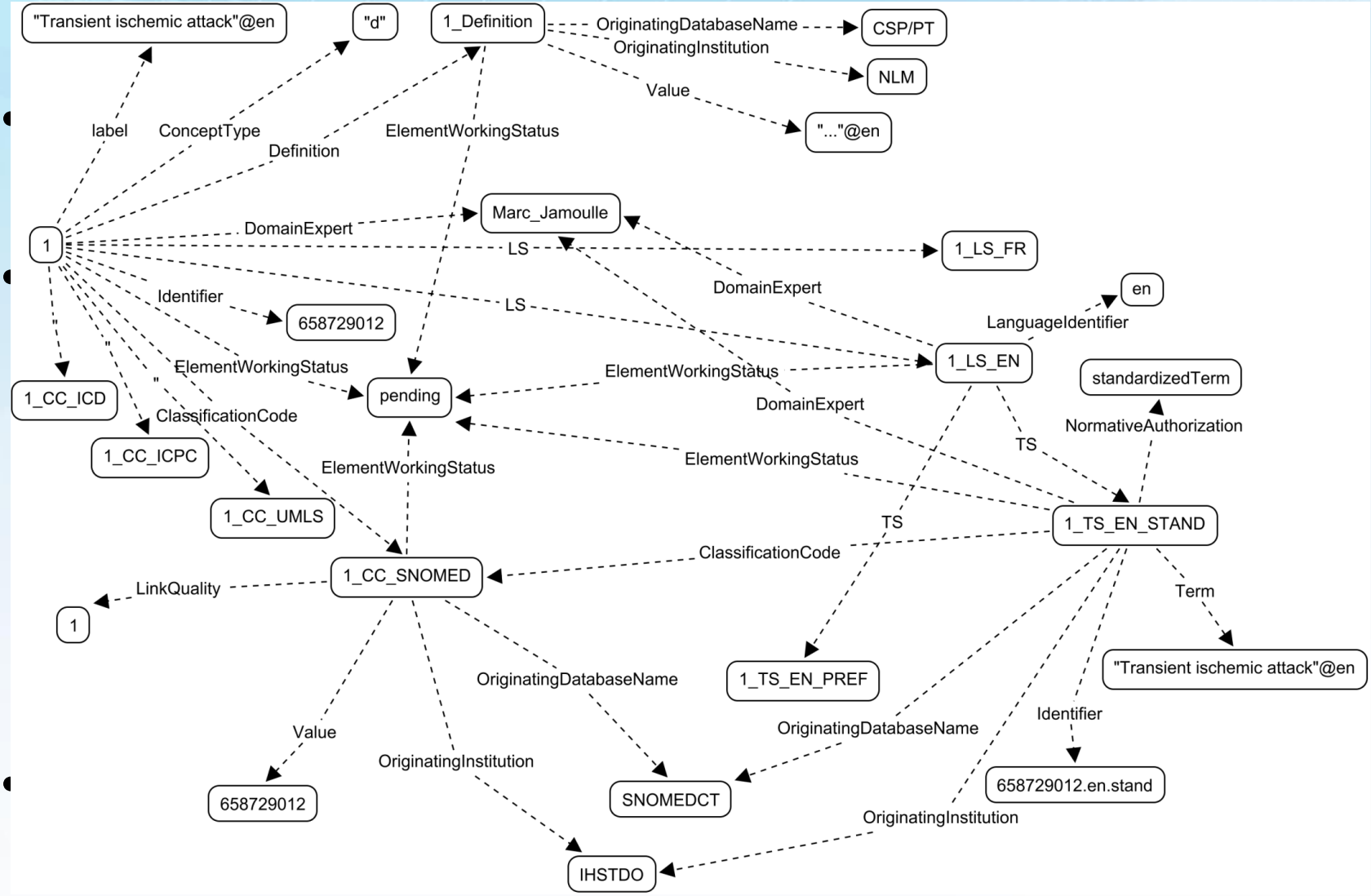
Populating the End-user Lexicon

- Automatic extraction using TexSIS (http://lt3.hogent.be/):
  - French/Dutch glossary of 774 words and phrases
- French terms matched to the French preferred terms in the RT
- 77/168 matched entries entered in the French End-user Lexicon in Lemon
  - Linked to the sense(s):
    - In the Reference Terminology (to the corresponding concept ID)
    - In external resources for other senses (E.g. French DBPedia)
- Admitted terms were added to extend the End-user Lexicon
  - Increase to 138 lexical entries (114/138 phrases)
- A total of 298 lexical entries after decomposition of phrases into single words

Link between the two resources



Reference Terminology in LOD



End-user Lexicon in LOD

- A total of 3.400 RDF triples to represent the French End-user Lexicon
- Published on the Data Hub as Linked Data and available also at:
  - http://meriterm.org/heartfailure/lexicons.rdf
- Linked to the Reference Terminology, to ISOcat and to the French DBPedia

Statistics on the data

	English	French	Italian
Reference Terminology concepts		168	
Preferred terms		168	168
Admitted terms		168	193
Standardized terms (Links to Snomed-CT)	161	161	161
UMLS definitions	116		
Links to ICD-10	131		
Links to ICPC-2	153		
End-user Lexicon entries		298	
Reference Terminology Triples	16,636		
End-user Lexicon Triples	3,400		

Conclusions

- Use of a hybrid approach to create a multilingual Interface Terminology in the healthcare domain
- Use of standards models for developing the terminology
  - ISO TMF for the multilingual Reference Terminology
  - Lemon / ISO LMF for unilingual End-user Lexicons
- ISOcat for the choice and labeling of the data categories
- Publication of these resources in the Semantic Web
  - Concepts and terms have a URI in the web and can be easily retrieved, integrated, linked and reused
- Useful for facilitating:
  - Medical data registration and searching of medical information by physicians
  - Translation of technical terms into lay terms for healthcare consumers
  - Translation of medical terms in other languages

Work in progress + Future works

- Use of a Semantic MediaWiki to maintain and extend our terminologies
- Create End-user Lexicon also for English, Italian, and Dutch
- Improvement of the concept/term extraction selection method
- Working on a new use case about Drugs Contraindications
- Links to other multilingual lexi-ontological resources on the web
  - E.g. Babelnet (http://babelnet.org/)

Thanks for your attention …

Any questions?

cardillo@fbk.eu

www.fbk.eu

meriterm@googlegroups.com