

PCSI 2024 Bled Slovenia

Healthcare Terminology Re-use: From ICD9 to SNOMED/ICD11

J.M Rodrigues¹, C Koné², M Babri² and B Trombert¹

¹Université Jean Monnet Saint Etienne France

² FHB National Polytechnic Institute Yamoussoukro Cote D'Ivoire

Program

- **Re-use of Casemix information system clinical data ?**
 - **No Casemix based funding without Casemix information system clinical data**
 - **History of healthcare terminologies needed by Casemix systems**
- **ICD11/SNOMED CT**
 - **ICD11 and SNOMED CT**
 - **From Lexical Alignment to Ontology Representation**
 - **Results**
- **Conclusion**

Casemix information system clinical data

- Casemix Information systems are based on a **minimum dataset** for **each** discharge or contact with the healthcare system on:
 - Specific clinical characteristics ,
 - **Diagnosis** principal and secondary,
 - **Medical and Surgical Procedures**
 - Functional health status
 - Activity of daily Leaving
 - resource utilization,
 - Accounting centers
 - **Individual consumption by patient**
 - other relevant factors

3



Only to group patients or Re-use ?

- The target is the classification and **grouping** of patients or episodes of care
- to provide a **standardized way to measure and compare** the complexity and resource needs of different patient populations

OR as well

- **Re-use of costly clinical data for Clinical Management and Internal management**

4



Program

- **Re-use of Casemix information system clinical data ?**
 - **No Casemix based funding without Casemix information system clinical data**
 - **History of healthcare terminologies needed by Casemix systems**

- **ICD11/SNOMED CT**
 - **ICD11 and SNOMED CT**
 - **From Lexical Alignment to Ontology Representation**
 - **Results**

- **Conclusion**

PCSI Summer School JM rodrigues



History since 1967

- **Diagnosis**
 - WHO ICD 9 1975 not specific enough for DRG
 - ICD 9 CM added localization in the US and widely used from 1980
 - ICD 9 XM in some countries
 - ICD 10 XM in most countries using DRG
 - ICD 11 ? (only for mortality)
- **Medical Procedures**
 - ICD 9CM vol 3 beginning based on WHO ICPM
 - Babel Tower in the world
 - ICD 10 CM PCS in the US, XM in most countries
 - ICHI WHO



Program

- Re-use of Casemix information system clinical data ?
 - No Casemix based funding without Casemix information system clinical data
 - History of healthcare terminologies needed by Casemix systems
- ICD11/SNOMED CT
 - ICD11 and SNOMED CT
 - From Lexical Alignment to Ontology Representation
 - Results
- Conclusion

PCSI Summer School JM rodrigues



3 different types of artefacts

- **Clinical language in Interface terminologies**
- **Reference terminologies:**
Compromise between bottom up (Terms and Synonyms) and Top down (Fully Specified Names and Applied Ontology)
- **Statistical classifications or aggregation**
terminologies: Mix of clinical vocabulary, taxonomy, knowledge and coding rules with meaningless codes as “other” or “unspecified”



ICD 11 MMS <https://who.int>

2024/01

- Aggregation (statistical) terminology
 - Mix of
 - clinical vocabulary,
 - taxonomy,
 - knowledge and
 - coding rules with meaningless codes as “other specified” or “unspecified”

And

- Semantic framework (non Formal Ontology)
 - Semantic Categories (13)
 - Textual definitions



9

ICD 11 Content

- The ICD-11 Mortality Morbidity Statistics (ICD 11-MMS) .
- The new ICD version approved by WHO World Health Assembly May 2019 to be used for mortality statistics from 2021.
- 28 chapters (1 to 26) plus
 - codes V for Functioning assessment and
 - codes X for Extension codes.
 - It proposes a Browser, a Coding tool and Special View for Infectious agents (lists of infectious agents, notifiable diseases and vaccinal conditions.)



10

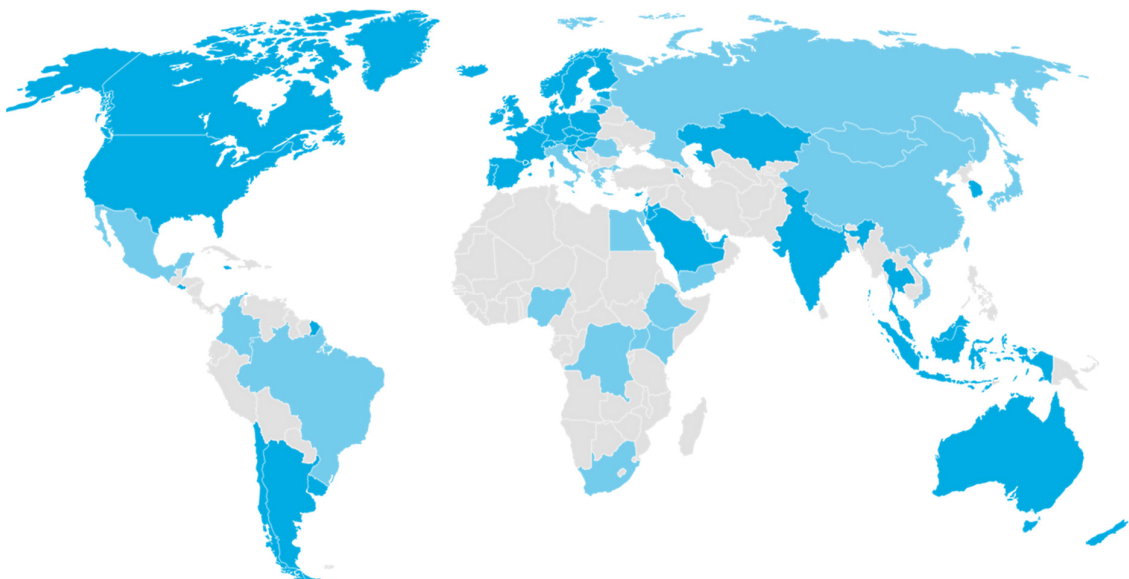
The PhD Dissertation : ICD11 Chapter 1

- 1027 ICD 11 codes
- 376 **excluded**
 - Needed only to insure the exclusivity and exhaustivity of any Classification system or Aggregation terminology as ICD
 - The excluded codes contain the words “other specified” or “unspecified”
 - These are necessary but have no possible “FORMAL Ontology” meaning
 - The number measures the gap between the ICD standardized clinical information and a full clinical description as in SNOMED
- 651 **selected** :
- 474 with textual definitions

11



SNOMED International country members



12



SNOMED CT 2 Components

- SNOMED CT
A *terminology* – as constituted by concepts (entities of lexical meaning), related terms of different types (Fully Specified Names, Preferred Terms, and Synonyms).
- B *ontology* constituted by classes, individuals and formal relations expressed as axioms in "Compositional Grammar" equivalent to EL⁺⁺/OWL-EL – what SNOMED names the "Concept Model", which can be processed by description logics reasoners

PCSI nov 15 -2019



SNOMED CT Overview

• **SNOMED CT (SCT)** <http://browser.ihtsdotools.org/>: 49 countries, 21 versions, 19 languages or dialects

Reference Clinical Terminology AND Ontology .

- 300.000 concepts in 19 domains or hierarchies: **TERMINOLOGY**
 - Is_a relation subsuming concepts
- 8 hierarchies or sub hierarchies (150 semantic tags) and only 186 121 concepts (international version release january 2024) have a so called « **concept model** »: **ONTOLOGY**
 - Definitional attributes of a DL concept model: coherence can be tested only with a DL reasoner.
 - **Clinical finding 150 000 codable concepts=values**
 - vs ICD11 17 000 codes and 120 000 codable terms
 - ICD10 CM Diagnosis 69000
 - **Procedure 55 423 concepts=values vs ICD10 CM PCS Procedures 70 000**
 - **Body structure 31 039 concepts=values**
 - Situation with explicit concept 4251
 - Event 3611
 - Pharmaceutical /biological Products 17291
 - **Physical object. 14 625 concepts=values**
 - Specimen 1634



Program

- Re-use of Casemix information system clinical data ?
 - No Casemix based funding without Casemix information system clinical data
 - History of healthcare terminologies needed by Casemix systems
- ICD11/SNOMED CT
 - ICD11 and SNOMED CT
 - From Lexical Alignment to Ontology Representation
 - Results
- Conclusion

PCSI Summer School JM rodrigues



Lexical Alignment ICD11 SNOMED CT

- This step was to map the **ICD-11 titles to the Fully Specified Name (FSN) or synonyms of SCT.**
- Two tools were used in this lexical step,
 - MetaMap [19] for automatic alignment and
 - SNOMED browser [5] for manually validation.
- MetaMap implements processing based on linguistic and algorithmic principles [20].
- It maps any terminology entered as input to any terminology included in the UMLS Metathesaurus. **SCT is one** of the terminologies included in the **Metathesaurus.**
 - MetaMap proposes an option to activate the processing of conjunctions during the analysis of the input terminology; in our work, ICD-11 codes titles or labels.
 - MetaMap provides two types of alignment results. Either an alignment with an expression corresponding to a single concept, known as a **pre-coordinated concept**, or an alignment with an expression corresponding to several concepts, known as **post-coordinated concepts.**
- After this automatic alignment phase using MetaMap, all post-coordinated alignments were manually validated in the SCT browser and showed some more pre-coordinated alignment.



Ontology Representation of ICD11

1 It is defined as the analysis of the representation of **ICD-11 textual definitions** or descriptions by the ontology representations of the SCT concepts aligned lexically in the previous step.

- On one hand a **text in English language** describing the health problem by its main characteristics: clinical signs, anatomy localization, transmission and etiology, proof by biological tests identifying the cause which in chapter 1 is mainly an infectious agent: bacteria, virus, parasite etc.
- On the other hand, a **representation in EL language** and by colored diagram available on SCT browser as for example « 111839008 | Intestinal infection caused by Escherichia coli (disorder) »

2 To be fully defined in ontology the representation of a concept shall contain all the **characteristics necessary and sufficient** to represent the code.

3 For the **necessary condition**, we have considered that if all the characteristics edited in the ICD-11 code label are present in the conceptual model of the SCT codes lexically aligned with the ICD-11 code the necessary condition was fulfilled

4 For the **sufficient condition**, we have decided to consider that the criteria were to assess if these ICD-11 textual definitions characteristics were **ALWAYS present**.

- If Yes these characteristics shall be considered in an **ontology representation**.
- If Not for example the characteristics following words as often, sometimes, is possible, **MAY BE** present, they are **not** to be considered in an ontology representation.

17



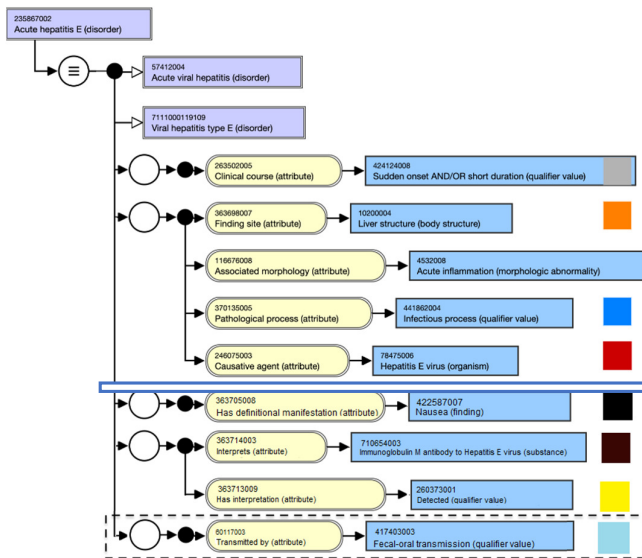
Exemple : ICD11 code by SNOMED CT Ontology

- The ICD-11 code and title « **1E50.4 Acute hepatitis E** » has the textual definition
- « **Liver** disease caused by acute **infection** with the **hepatitis E virus**. This disease is characterized by **nausea**. **Transmission is generally via the fecal-oral route**. Confirmation is by **detection** of anti-hepatitis E virus IgM antibodies in an individual's serum »
- is lexically aligned to the concept « SCT 235867002 | Acute hepatitis E (disorder) | »

18



Ontology Representation ICD11 code by SNOMED CT Ontology



Step 1 authorized attributes and values

- For the symptom nausea, the attribute relation 363705008 |Has definitional manifestation|.
- The notion of confirmation is by detection of anti-hepatitis E virus IgM antibodies
- the attribute-value pairs :
 - Interpret | 363714003 |
 - Has interpretation | 363713009 |

Step 2 unauthorized but existing attributes and values

- The notion of transmission by the fecal-oral way is materialized by the attribute
- |Transmitted by| 60117003 with
- the concept value |Fecal-oral transmission |(qualifier value) 417403003.

Program

- Re-use of Casemix information system clinical data ?
 - No Casemix based funding without Casemix information system clinical data
 - History of healthcare terminologies needed by Casemix systems

- **ICD11/SNOMED CT**

- ICD11 and SNOMED CT
- From Lexical Alignment to Ontology Representation
- Results

- Conclusion



Lexical Alignments and Ontology Representation

Lexical Alignments		number	%
ICD1 TO SCT			
Pre-coordinated concepts	SCT	615	94
Post-coordinated Concepts	SCT	36	6
Total		651	100

Ontology Representation	num ber	%
ICD11 codes by SCT FD concepts models		
Complete with validated attributes and values	260	53
Incomplete and complete with validated or unvalidated attributes and values	231	47
Incomplete	0	0
Total	491	100

Ontology Representation	num ber	%
ICD11 codes by SCT NON FD concepts models		
Complete with validated attributes and values	160	100
Incomplete and complete with validated or unvalidated attributes and values	160	100
Incomplete	0	0
Total	160	100

21



Program

- Re-use of Casemix information system clinical data ?
 - No Casemix based funding without Casemix information system clinical data
 - History of healthcare terminologies needed by Casemix systems
- ICD11/SNOMED CT
 - ICD11 and SNOMED CT
 - From Lexical Alignment to Ontology Representation
 - Results
- Conclusion



Conclusion

- 1 We have shown in this work that a formal ontological representation of a health terminology proposing a detailed textual description of its titles as ICD-11 is possible by using the **lexical alignment and ontology representation tools** of other terminology systems (UMLS and SCT).
- 2 The work measures **the gap** between a lexical alignment and a meaning alignment between different health terminologies.
- In this example the ICD11 textual definition is **sometimes more precise** than the SNOMED concept model representation.
- As consequence there is a need **to extend the use and validation** of SNOMED attributes and values presently not authorized by SNOMED
- The work opens a way to a **Formal Ontology** comparison between ICD-11 and SCT critical point before a joint development of a **Common Ontology between ICD-11 and SNOMED CT** final goal for a seamless connection between the 2 worldwide health terminology systems.

23



END rodriguesjeanmarie2@gmail.com

- THANK YOU
- **Vielen Dank für Ihre Aufmerksamkeit**
- **Merci**
- **Gracie**
- **Gracias**
- **Obrigado**
- **Efcharisto**
- **Kessenem**
- **Tag**
- **Arigato Gozaimasu**
- **Terima Kasih**
- **Shkran**

