Serveur Terminologique

Recherche préliminaire en rapport avec le développement d'un serveur belge de terminologie dans le secteur de la santé.

Ref. Ctr-n09-03-serveur terminologique-MIM, financé par le SPF Santé Publique

Partie 1/3 : Vocabulaire contrôlé multilingue et multi-classification

Section A : Méthodologie

Version : 1.1

31 octobre 2010

Coordinateur-rédacteur :   Benny Van Bruwaene

Comité rédactionnel : /

Comité d'accompagnement : Etienne De Clercq (président de la MIM),

Luc Nicolas (représentant SPF Santé publique)

Note : cette partie du rapport contient trois sections :

Section A : méthodologie

Section B : Evaluation critique intermédiaire

Section C : (uniquement sous format excel) : vocabulaire contrôlé

Editeur. : MIM

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

1. <u>National controlled medical vocabulary (CMV)</u>

1.1. *Definition*

A national controlled medical vocabulary is a list of medical concepts which are well defined and clinically useful. They are described in the local languages Dutch, French, German and English without any ambiguity.

It is important that the same medical concept is clearly defined by the medical term which must be detailed enough to exclude homonyms (same term with more than one meaning). Eg 'tafel' is not precise enough because it can be 'operatietafel' 'tafel van vermenigvuldiging' '......'.

Synonyms and common abbreviations are not part of a CMV. They can be considered in order to help search engines to find the proper concepts. However, the development of lexica with synonyms and natural language processing ontologies used for search engines of terms/concepts should be independent of the development of the controlled medical vocabulary which contains only one preferred term for one concept.

Term finders are developed by different vendors, e.g. Language & Computing, OBO Foundry, Sosoeme (based on 3BT)...

1.2. *Objective*

The objective of a national CMV is to improve the semantic interoperability: the same medical concept should be understood in the same way by different users and computer systems: care takers, patients, research, billing, decision support systems… and this independent of the language used.

The most common use of standard language is coding of diagnoses and procedures for billing and epidemiology.

To this extent the medical events of a patient are reduced to codes of different classifications (ICD, RIZIV/INAMI). Because codes of one classification don't match with

the codes of another system, the same concepts are to be coded twice. A typical example is the coding of hospital discharge information. Procedures are coded in ICD-9-CM and RIZIV/INAMI and diagnoses are registered in ICD-9-CM for the minimal data set (MKG/RCM) and ICD-10 is used for coding of the registry of deceases. Anatomopathology is coded in ICDO for the cancer registry and ICD-9-CM for the minimal data set (MKG/RCM). Because of the lack of interoperability, the same item is to be coded several times generating loss of time and efficacy.

Epidemiological research in primary care is done with ICPC, another classification system which requires separate coding.

The purpose of a National terminology is to offer physicians, nurses, … precise terms in their own language so that the registration of the term in the patient's record can be used for clinical purposes (e.g. integration in clinical and administrative reports) and gives automatically multiple codes to multiple systems without the further need of secondary coding.

With the increasing number of decision support systems (drug surveillance, clinical paths) the same registration on line will help the physician to guide his working up, to feed resource management systems and to control for potential hazards or mistakes.

'One single entry does it all!' A major saving of time enhancing quality and consistency.

In a Belgian and European context, more and more patients travel abroad and consult physicians in different languages.

The national terminology aims to convert the patient summary from one language to the other. Dutch, French and German are the Belgian official languages. English and Spanish are widely used international languages. The Belgian national terminology should provide preferably these five languages.

For clinical research diagnoses should be translated into English on the Case Report Form.

1.3. Link with classifications

A national controlled medical vocabulary should include different domains:

- Signs, symptoms, diagnoses (available in Snomed CT, ICD-9-CM, ICD-10, ICPC, DSMIV)
- State, personal and familial history (Snomed CT, ICD-9-CM, ICD-10)
- Accidents (Snomed CT, ICD-9-CM, ICD-10)
- Medical acts of primary and/or hospital care and surgery (Snomed CT, ICD-9-CM, ICPC, RIZIV/INAMI)
- Imaging (Snomed CT, ICD-9-CM, RIZIV/INAMI, ICPC)
- Diagnostic and functional tests (Snomed CT, ICD-9-CM, RIZIV/INAMI, ICPC)
- Labtests (Snomed CT, LOINC, ICD-9-CM, RIZIV/INAMI, ICPC)
- Medication (ATC)
- Nursing acts (Snomed CT, NIC)
- Level of functioning ICF
- Adverse events (ICPS, WHOART, COSTART)

Validated cross map tables exist between

- Snomed CT and
  - ICD-9-CM, ICD-10 for diagnoses. Cross map with ICD-10-CM planned (as an upgrade of the ICD-9-CM cross map).
  - UMLS, MeSH
  - ICPC? There seems to be a Snomed CT – ICPC working group (cf dr. Marc Jamoulle)
- ICD-10
  - ICPC for diagnoses

The terms of the national terminology should be independent from classification systems, but should be linked to the relevant classification systems for statistics, financing, decision support.

Linking with a granular classification which has cross maps to other classification systems saves cross mapping working on behalf of the Belgian terminology working group.

For diagnoses Snomed CT seems to offer the most granular classification with cross maps to ICD-9-CM, ICD-10, ICPC.

For procedures RIZIV/INAMI seems to be the most practical starting point because the link between a procedure and RIZIV/INAMI is needed for tarification and invoicing. Cross maps between RIZIV/INAMI to ICD-9-CM exist in many hospitals. Cross map to ICPC might exist. The cross map to Snomed CT doesn't exist and might be useful for further mapping to UMLS and MeSH.

2. Existing vocabularies

2.1. 3BT

3BT is a controlled medical vocabulary (CMV) developed by the academic centres of general practice of RUG and ULB commissioned by the Belgian ministry of health.

The existing data structure:

Field Name	Data Type
IBUI	Number
FR_Clinical_Label	Text
NL_Clinical_Label	Text
ICPC_2_Code	Text
ICD_10_Code	Text

The IBUI number was meant to be a unique identifier for a medical concept.

However synonyms appear to have different IBUI.

881 identical Dutch labels and 1.849 identical French labels correspond with different IBUI-numbers.

Example:

IBUI	3BT FR_Clinical_Label	3BT NL_Clinical_Label
10032231	pharyngite à gonocoques	gonokokken pharyngitis
10032232	pharyngite gonococcique	gonokokken pharyngitis

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

Also headings get a specific IBUI.

Example:

IBUI	FR_Clinical_Label	NL_Clinical_Label	ICPC_2_Code
40001207	réparation / fixation / suture / plâtre / prothèse	herstel/fixatie - hechting/gipsspalk/prothese (aanbrengen)	*54

The vocabulary is built from different classification systems:

- ICPC concepts. They are characterized with one concept per ICPC-code (1.554 concepts in total). They include different items of consultations (30x), clinical examination (31x), labtests (31x-38x), function tests (39x), technical examinations (40x-43x), vaccination and prevention (44x), education and information (45x), consultation of paramedics (46x), consultation of specialist (47x), prescription (50x), interventions (51x-59x), administrative acts (62x), type of follow visits (63x), referral to medical aid (66x-67x), referral to public aid

- ICD-10 concepts: signs, symptoms, diagnoses. Many concepts per ICPC-code.

Besides terms which can be used in a clinical context, terms drawn from classifications are too precise for Belgium and can be confusing:

Example:

ICD 10	Term UZB	3BT NL_Clinical_Label
A010	typhoid fever	ileotyfus
A010	typhoid fever	tyfuskoorts
A010	typhoid fever	infectieuze enteritis door Salmonella typhi
A010	typhoid fever	Salmonella typhi infectie
A010	typhoid fever	typhus abdominalis intestinaal ulcus
A010	typhoid fever	typhus abdominalis darmulcus
A010	typhoid fever	tyfus
A010	typhoid fever	buiktyfus
A010	typhoid fever	tyfogastrische koorts
A010	typhoid fever	tyfoperitonitis
A010	typhoid fever	typhus abdominalis
A010	typhoid fever	tyfeus ulcus
A010	typhoid fever	Salmonella typhi lymphadenitis
A010	typhoid fever	posttyfeus abces
A010	typhoid fever	enteritis typhosa
A010	typhoid fever	tyfo-enteritis
A010	typhoid fever	tyfusperforatie

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

ICD 10	Term UZB	3BT NL_Clinical_Label
A010	typhoid fever	tyfusinfectie
A010	typhoid fever	Eberth
A010	typhoid fever	mesenteriale lymphadenitis door Salmonella typhi
A010	typhoid fever	Salmonella typhus abdominalis

Example:

Clinical_Label
bacteriologisch en histologisch bevestigde tuberculeuze bronchopleuromediastinale fistel
bacteriologisch en histologisch bevestigde tuberculeuze bronchocutane fistel
bacteriologisch en histologisch bevestigde tuberculeuze bronchomediastinale fistel
bacteriologisch en histologisch bevestigde tuberculeuze broncho-oesofagale fistel
bacteriologisch en histologisch bevestigde tuberculeuze bronchopleurale fistel
bacteriologisch en histologisch bevestigde tuberculeuze bronchoviscerale fistel
bacteriologisch en histologisch bevestigde glottis tuberculose
bacteriologisch en histologisch bevestigde tuberculeuze laryngitis

Terms can be ambiguous: it can be the one or the other

Example:

NL_Clinical_Label
acute of subacute endocarditis
acute of subacute bacteriële endocarditis
acute of subacute infectieuze endocarditis
acute of subacute reumatische endocarditis
acute of subacute reumatische endocarditis met chorea

ICD-10 doesn't match always:

Example:

3BT

ICD-10	NL_Clinical_Label
Z721	probleem met alcoholgebruik

ICD-10:

ICD-10	Nederlands
Z721	Alcoholgebruik

3BT contains equally a separate ontology with different concepts in Dutch and French and their logical relationships for making an intelligent search engine for terms. This ontology is already integrated in one GP-Electronic Medical Record 'Sosoeme'.

2.2. *UZB*

The CMV of the UZB is a controlled medical vocabulary, developed by the department of medical registration. It is based on Snomed CT concepts and validated by the heads of the departments of the different specialties.

In order to develop a start up vocabulary with a maximum of useful terms, the university hospital UZB developed a vocabulary from the ICD-9-CM codes for diagnoses registered in the hospital from 2000 to 2006. These codes were mapped with Snomed CT concepts based on the ICD-9-CM cross map tables version 2004. From these ICD-codes the Dutch descriptions of the ICD-coding handbook of the K.U.Leuven were entered on the corresponding Snomed CT concepts. This initial list is validated and completed by the heads of the clinical departments.

Criteria used for validation:

- Terms should be precise and standardized.

- Terms should be usable in medical reports

- No synonyms in order to reduce development and maintenance costs. This means that the CMV can be compared to the preferred terms of Snomed.

- No abbreviations

- No homonyms.

- The absence of another condition is not specified. E.g. 'buil zonder wonde' = 'buil'

- Avoid combined terms E.g. 'hypertensief hartfalen'

- Avoid attributes like 'primary' 'idiopathic'

Termid	Unique random number.
Term	Preferred term for unique medical concept. No synonyms. Full description of concept excluding homonyms.
Snomed CT number	
Snomed ok	'0' means that the term doesn't match in a perfect way with the Snomed-concept.
Type	Indicates the type of the concept based on Snomed CT. 'Procedure', 'Procedure.OR' ; 'Problem'; 'Problem.Finding'; 'Problem.Disorder'; 'Status'; 'Study', 'Heading'

Based on the Snomed CT cross map tables, ICD-9-CM version 2009 and ICD-10 are linked.

The actual UZB CMV is validated for diagnoses, signs and symptoms. The terms are carefully selected on their clinical use.

The actual database contains 10.500 validated terms for diagnoses. 2.500 terms with Snomed CT code were disqualified as being non clinical, ambiguous, synonym or not relevant.

15.000 terms without Snomed CT code are ICD-9-CM descriptions which are aspecific classification terms or diseases which are very rare.

UZB is working at a CMV for surgery with RIZIV/INAMI and ICD-9-CM but the work is not finished and not validated.

The CMV is used in the EMD with the semantic search engine of Language & Computing: The terminology server is built around LinKBase® [14], a very large linguistic knowledge base with a medical ontology consisting of concepts and terms. The concepts are linked to each other using a wide variety of relationship types hence defining the semantics of

the concepts. This forms a large semantic network optimized for making medical terminology understandable by computers.

Some classifications are linked to LinKBase® such as ICD9-CM and SNOMED-CT towards which can be coded. The user of the system does not require knowledge of particular coding systems since input in natural clinical language is permitted. The system handles synonymy (e.g. diabetes mellitus and sugar disease), homonymy (cortical atrophy of the brain or the kidneys), spelling differences and compound words.

The semantic engine parses an expression into terms made up of one or more words. These terms are looked up in a lexicon wherein they are linked to concepts. The most relevant codes are then extracted from an intersection of the different sets of concepts. Ambiguities are automatically detected and resolved. The output consists of a list of codes with corresponding labels of the coding system. For example, suppose a physician enters the phrase, "acute heart infarction." The semantic search is able to determine that this is equivalent to the LinKBase® concept "acute myocardial infarction," even though the phrase "acute heart infarction" doesn't actually appear anywhere in LinKBase®.

To make this work, the concept is stored along with its formal definition (description logic formula), which for this example looks like:

((ACUUT HARTINFARCT) { [IS-A] (INFARCT) } { [HAS-ACTEE] (MYOCARDIUM) }
{ [HAS-WE-STATE] (ACUTE STATE)}) (simplified example).

Every concept in the definition is linked to terms like "acute," "myocardium," "myocardial," "infarct," and "infarction." The semantic engine maps user-entered text to these terms, and then maps this to the appropriate concepts using the description logic definitions. The semantic engine then gives a prioritized list of codes from the desired classification system, based on LinKBase's® internal mappings. In this way semantic engine offers a

multi-dimensional, concept-based approach to code finding which goes well beyond both string searching and hierarchical searching of the coding system.

The syntactic engine parses a given expression into tokens. Each token is matched against a list of tokens derived from the CMV lexicon. Mismatches are spellchecked meaning that misspelled words are corrected and character groups (for fast query entering) are extended at a certain cost. Tokens of a given query that match the token lexicon undergo permutations and are matched against the CMV term lexicon. The permutations leading to complete or partial token overlap with the terms of the CMV are scored in function of the token overlap and corrective actions made. Another feature of the syntactic parser is that permutations are concatenated in order to find compound words that are frequent in Dutch (e.g. "lon" and "oed" => longoedeem).

3. <u>Merge CMV 3BT/UZB</u>

In order to create a start up terminology which is suitable for primary care and hospitals, the existing terminologies '3BT' and the 'CMV UZB' are merged with a direct link to Snomed CT, ICD-10, ICD-9-CM and ICPC.
All concepts have a Dutch label and a Snomed CT concept.
Based on the ICD-10 cross map of Snomed, both existing dictionaries were linked and the right 3BT term was chosen generating a French label and ICPC-code.
Concepts without French label were connected to the French translation of Snomed concepts made in Canada. The validation of the French translation was not done.

3.1. *Mapping'3BT' with 'CMV UZB'*

The ICD-10 codes of 3BT are linked with Snomed CT based on the ICD-10 cross map. The UZB term is listed if available and all the 3BT labels.

Two medical doctors (dr. Kurt Mies kurt.mies@telenet.be and dr. Jean-Luc Mommaerts jean.luc.mommaerts@pandora.be) perform each the mapping of 50 % of the records.

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

For every Snomed CT concept the <u>matching</u> 3BT/UZB concept is indicated with '1' in the field '3BT ok'. For every single Snomed CT concept, only one single label is selected.

Example:

ICD-10	Snomed CT	3BT ok	UZB Term	NL_Clinical_Label	FR_Clinical_Label
A010	Typhoid fever (disorder)		typhoid fever	posttyfeus abces	abcès post-typhique
A010	Typhoid fever (disorder)		typhoid fever	mesenteriale lymphadenitis door Salmonella typhi	lymphadénite mésentérique à salmonella typhique
A010	Typhoid fever (disorder)		typhoid fever	Salmonella typhi infectie	infection à salmonella typhique
A010	Typhoid fever (disorder)		typhoid fever	tyfoperitonitis	typhopéritonite
A010	Typhoid fever (disorder)		typhoid fever	buiktyfus	typhoïde abdominal
A010	Typhoid fever (disorder)		typhoid fever	tyfeus ulcus	ulcération typhique
A010	Typhoid fever (disorder)		typhoid fever	ileotyfus	typhus intestinal
A010	Typhoid fever (disorder)		typhoid fever	Salmonella typhi lymphadenitis	lymphadénite à Salmonella typhi
A010	Typhoid fever (disorder)		typhoid fever	Salmonella typhus abdominalis	typhus abdominalis à Salmonella
A010	Typhoid fever (disorder)		typhoid fever	tyfus	typhus
A010	Typhoid fever (disorder)		typhoid fever	tyfusperforatie	perforation typhoïdique
A010	Typhoid fever (disorder)	1	typhoid fever	tyfuskoorts	fièvre typhoïde
A010	Typhoid fever (disorder)		typhoid fever	Eberth	Eberth
A010	Typhoid fever (disorder)		typhoid fever	typhus abdominalis	typhus abdominalis
A010	Typhoid fever (disorder)		typhoid fever	infectieuze enteritis door Salmonella typhi	entérite infectieuse à Salmonella typhi
A010	Typhoid fever (disorder)		typhoid fever	enteritis typhosa	entérite typhique
A010	Typhoid fever (disorder)		typhoid fever	tyfo-enteritis	entérite tiphoïdique
A010	Typhoid fever (disorder)		typhoid fever	tyfusinfectie	infection typhoïde
A010	Typhoid fever (disorder)		typhoid fever	typhus abdominalis intestinaal ulcus	typhus abdominalis avec ulcère intestinal
A010	Typhoid fever (disorder)		typhoid fever	typhus abdominalis darmulcus	ulcération intestinale à Salmonella typhi
A010	Typhoid fever (disorder)		typhoid fever	tyfogastrische koorts	fièvre typhoïde avec atteinte gastrique

There are 531.310 records corresponding with the found combinations.

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

When the 3BT term matches, but the concept is judged to be very specific, a 'S' is recorded.

In the merged database the UZB label has priority over the Dutch 3BT label unless the Dutch 3BT label is considered to be better. In that case '3BT ok' ='3'.

When the UZB term matches with the Snomed CT concept but no 3BT label, the UZB term will be appended afterwards automatically. '3BT ok' = Blank.

When the reviewer doubts about the use of the 3BT concept, he puts '?'. These items will be reviewed by the coordination team and/or the heads of the departments in every specialty.

UZB term is not valid and no 3BT term corresponds to the Snomed CT concept. Put 'd' = to be deleted.

UZB term is valid but doesn't match with the Snomed concept and no 3BT term corresponds to the Snomed CT concept. Put 'e' = 'error' or the exact Snomed CT number if known.

Legend:

Field '3BT ok'	description
1	Dutch label CMV UZB and French 3BT label is ok
?	to be reviewed
3	Dutch 3BT label is better than CMV UZB
S	Dutch 3BT label is correct, however the concept is judged to be too specific
D	UZB term is not valid
E or enter the exact Snomed CT number	UZB term is valid but doesn't match with Snomed concept

The Snomed CT concepts linked with a UZB term but not by any 3BT term, are appended automatically (3.226 concepts).

For the concepts with missing French label, the French label is drawn automatically from the French Canadian Snomed CT 3.5.

The validation between the Belgian French label and the Canadian one is not part of the merge process by UZB.

This implies that the final CMV will contain a Dutch label and a Snomed CT concept but not always a French label. Snomed CT offers the preferred term in English and Spanish as it gives the cross maps towards ICD-9-CM, ICD-10, ICPC, UMLS and MeSH.

3.2.   *Assembly final terminology*

- Correct in CMV if [3BT ok]='e' or ='snomed code' of '?'
- If '3BT ok' = Snomed code -> copy right code in Snomed Merge and change '3BT ok' in '1' or '3'
- Group in 'Merge' Snomed concepts (check doubles)
- Join with CMV UZB (last version)
    - o   Out of join (Term is null): Take N and Fr + all other BT fields including ICPC if [3BT ok]='1' or 's' or '3'
    - o   Inner join (Term not null)
        - ▪   Take N and Fr if [3BT ok]='3' + all other BT fields
        - ▪   Take Term and Fr if [3BT ok]='1' or 's' ('clinical'='1' and 'Snomed ok' is not '0')
- Add Snomed concepts from CMV UZB ('clinical'='1' and 'Snomed ok' is not '0') which were out of join with list of selected terms/concepts: take Term+icd10 code.
- Eliminate duplicates for Snomed CT concept, Dutch and French label.
- If 'F' is empty, propose 'Canada 3.5'
- Complete ICD-9-CM and ICD-10 on Snomed cross map

The appropriate licences of the Snomed CT and the French Canadian Snomed CT 3.5 are needed from the Belgian ministry of health for the start of this development.

3.3.   *Pilot validation*

A first validation was done early in the process in order to prevent structural errors early in the process. The validation helped to statndardise the further coding process.

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

Two hundred fifteen concepts were coded by dr. Philippe Vandenbergh (philippe.vandenbergh@hecamax.be) and compared with the coding of the original coders.

Coding was identical in 75 % of concepts.

Dr. Benny Van Bruwaene reviewed the discrepancies (54 records):

- One puts '1' the other puts 'S' (n=17):

FULLYSPECIFIEDNAME	Term UZB	NL_Clinical_Label
Secondary malignant neoplasm of pleura (disorder)	metastase pleura	maligne metastase in pleura
Secondary malignant neoplasm of large intestine (disorder)	metastase dikke darm	maligne metastase in dikke darm
Secondary malignant neoplasm of the peritoneum (disorder)	metastase peritoneum	maligne metastase in peritoneum
Secondary malignant neoplasm of liver (disorder)	levermetastase	maligne metastase in lever
Secondary malignant neoplasm of spleen (disorder)	metastase milt	maligne metastase in milt
Secondary malignant neoplasm of kidney (disorder)	niermetastase	maligne metastase in nier
Secondary malignant neoplasm of skin (disorder)	huidmetastase	maligne metastasen in huid
Secondary malignant neoplasm of brain (disorder)	hersenmetastase	maligne metastase in hersenen
Secondary malignant neoplasm of bone of skull (disorder)	botmetastase schedel	maligne botmetastase in schedel
Intestinal infection due to Escherichia coli (disorder)	gastro-enteritis door Escherichia coli	intestinale Escherichia coli infectie
Enteric campylobacteriosis (disorder)	gastro-enteritis door Campylobacter	infectieuze enteritis door Campylobacter

This difference is of limited importance. In both cases the term and translation are mapped and preserved. 'S' means that the concept shouldn't be integrated in the standard dictionary because the term is too specific. The discrepancy between coders shows that the difference between '1' and 'S' is subjective. No formal rules can be set to the definition 'S'.

- One chooses the UZB term and refutes the 3BT translation, the other prefers the 3BT term as translation of the concept (n=4). Again the discrepancy will be corrected afterwards in a way that the concept is preserved and the French

translation will be missing eventually but can be recovered afterwards.

FULLYSPECIFIEDNAME	Term UZB	NL_Clinical_Label	FR_Clinical_Label
Cryoglobulinemic vasculitis (disorder)	cryoglobulinemische vasculitis	cryoglobulinemische vasculitis	vasculite cryoglobulinémique
Primary malignant neoplasm of cauda equina (disorder)	maligne neoplasma van de cauda equina	maligne neoplasma cauda equina	tumeur maligne de la queue de cheval
Primary malignant neoplasm of urethra (disorder)	maligne neoplasma van de urethra	maligne neoplasma urethra	tumeur maligne de l'urètre
Primary malignant neoplasm of spinal cord (disorder)	maligne ruggenmergneoplasma	maligne neoplasma ruggenmerg	tumeur maligne de la moelle épinière

- Relevant concepts not coded in 8 cases (half by the original coder, half by the reviewer). Half of the concepts are rare and should rather be coded as 'S'. One concept is too general.

FULLYSPECIFIEDNAME	NL_Clinical_Label	FR_Clinical_Label
Infection due to Mycobacterium xenopi (disorder)	Mycobacterium xenopi infectie	infection à Mycobacterium xénopi
Cryofibrinogenemia (disorder)	cryoglobulinemie	cryoglobulinémie
Disorder of immune system (navigational concept)	iimmuunmechanisme stoornis	trouble de la fonction immunitaire
Cystitis with amebiasis (disorder)	amoeben cystitis	cystite amibienne
Tuberculosis of epididymis (disorder)	epididymis tuberculose	tuberculose de l'épididyme
Pneumonia in pertussis (disorder)	kinkhoest met pneumonie	coqueluche avec pneumonie
Polyclonal gammopathy (disorder)	polyclonale gammopathie	gammopathie polyclonale
Group D streptococcal septicemia (disorder)	Streptococcus groep D sepsis	septicémie à Streptococcus du groupe D

- Preserved translation UZB Term not exact in 5 cases:

FULLYSPECIFIEDNAME	Term UZB
Intestinal infection due to Morganella morganii (disorder)	gastro-enteritis door Morganella morganii
Pulmonary disease due to Mycobacteria (disorder)	pneumopathie door Mycobacterium chelonae
Tuberculosis of brain (disorder)	tuberculose van het centraal zenuwstelsel
Congenital hypergammaglobulinemia (disorder)	hypergammaglobulinemie
Sepsis syndrome (disorder)	sepsis met orgaanfalen

3.4. *Validation through double coding*

A validation was done using double coding on final work (after tailoring methodology based on pilot validation and with complete assembly of terminology as specified in 3.2).

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

Thousand six hundred twenty concepts were coded twice (or 32.000 records with ICD-10 code beginning with 'I') by the two physicians.

The terms were assembled as specified in 3.2 for both physicians. The result of both physicians was compared.

- Mismatches

Three mismatches were found of which only 1 error was due to human error:

ICD-10	Snomed Id	Snomed description	Wrong translation N	Wrong translation F	Right translation N	Right translation F
I151	18443005	Rupture of chordae tendineae (disorder)	secundaire hypertensie door nieraandoeningen	hypertension secondaire à une affection rénale	ruptuur van chordae tendineae	
I471	195072008	Paroxysmal nodal tachycardia (disorder)	atrioventriculaire nodale re-entry tachycardie		paroxysmale nodale tachycardie	tachycardie paroxystique nodale
I802	17513007	Thrombophlebitis of popliteal vein (disorder)	iliacale trombose	thrombose iliaque	tromboflebitis van vena poplitea	

- Missing and superfluous concepts

Sixty three concepts were coded by physician1 and not by physician2.

Fifty three concepts were undue (13 ambiguous concepts, six too general, twenty three synonyms of existing term, one combination concept and ten rare concepts).

Ten concepts could be considered as an enhancement:

ICD_10_Code	Snomed CT number	FULLYSPECIFIEDNAME	NL_Clinical_Label	FR_Clinical_Label	IBUI
I62.1	397809001	Nontraumatic extradural hemorrhage (disorder)	niet-traumatische extradurale bloeding	hémorragie extradurale non traumatique	10013280
I49.0	111288001	Ventricular flutter (disorder)	Ventrikelflutter	flutter ventriculaire	10033465
I89.8	67710009	Rupture of thoracic duct (disorder)	ductus thoracicus ruptuur	rupture du canal thoracique	10116290
I78.0	21877004	Osler hemorrhagic telangiectasia syndrome (disorder)	ziekte van Rendu-Osler-Weber	maladie de Rendu-Osler-Weber	10086381
I88.0	44897000	Mesenteric lymphadenitis (disorder)	mesenteriale lymphadenitis	lymphadénite mésentérique	10056346
I25.3	275511009	Mural cardiac aneurysm (disorder)	muraal aneurysma	anévrisme mural	10062123
I89.8	234104002	Escape of lymph	Lymforroe	lymphorhée	10056894

ICD_10_Code	Snomed CT number	FULLYSPECIFIEDNAME	NL_Clinical_Label	FR_Clinical_Label	IBUI
		(disorder)			
I74.3	52156004	Femoral artery thrombosis (disorder)	femorale arteriële trombose	thrombose artérielle fémorale	10099666
I89.8	191394000	Calcified lymph nodes (disorder)	lymfeklierverkalking	calcification d'un ganglion lymphatique	10056475
I86.0	20656007	Sublingual varices (disorder)	sublinguale varix	varices sublinguales	10094432

- Hundred thirty six concepts were coded by physician2 and not by physician1.

  Ninety nine concepts were undue (20 ambiguous concepts, 35 too general, 17 synonyms of existing term, 7 combination concepts and 16 rare concepts).

  Thirty seven concepts could be considered as an enhancement:

ICD_10_Code	Snomed CT number	FULLYSPECIFIEDNAME	NL_Clinical_Label	FR_Clinical_Label
I46.0	233927002	Cardiac arrest with successful resuscitation (disorder)	hartstilstand met geslaagde reanimatie	arrêt cardiaque, réanimé avec succès
I23.5	194867006	Rupture of papillary muscle as current complication following acute myocardial infarction (disorder)	papillairspierruptuur na acuut myocardinfarct	rupture du muscle papillaire après un infarctus aigu du myocarde
I23.4	194866002	Rupture of chordae tendinae as current complication following acute myocardial infarction (disorder)	chordae tendineaeruptuur als actuele complicatie na acuut myocardinfarct	rupture de cordages tendineux comme complication récente d'un infarctus aigu du myocarde
I23.1	194863005	Atrial septal defect as current complication following acute myocardial infarction (disorder)	atriumseptumdefect als actuele complicatie na acuut infarct	défect de la cloison interauriculaire comme complication récente d'un infarctus aigu
I23.0	194862000	Hemopericardium as current complication following acute myocardial infarction (disorder)	hemopericard na acuut myocardinfarct	hémopéricarde après infarctus aigu du myocarde
I71.0	308546005	Dissection of aorta (disorder)	aortadissectie	dissection de l'aorte

ICD_10_Code	Snomed CT number	FULLYSPECIFIEDNAME	NL_Clinical_Label	FR_Clinical_Label
I60.9	32728005	Hemorrhage due to ruptured congenital cerebral aneurysm (disorder)	geruptureerd hersenaneurysma	rupture d'un anévrisme cérébral congénital
I89.8	234110002	Lymph cyst (disorder)	chyleuze cyste	kyste chyleux
I62.0	281864001	Non-traumatic intracranial subdural hematoma (disorder)	niet-traumatisch subduraal hematoom	hématome sous-dural non traumatique
I60.8	277299009	Ruptured cerebral arteriovenous malformation (disorder)	arterioveneuze hersenmalformatie ruptuur	rupture d'une malformation artérioveineuse cérébrale
I60.9	270907008	Spontaneous subarachnoid hemorrhage (disorder)	niet-traumatische subarachnoïdale bloeding	hémorragie non traumatique sous-arachnoïdienne
I80.1	234044007	Ileofemoral deep vein thrombosis (disorder)	iliofemorale trombose	thrombose ilio-fémorale
I80.0	234039000	Saphenous vein phlebitis (disorder)	v.saphena flebitis	phlébite de la veine saphéenne
I74.0	233976008	Aortic bifurcation thrombosis (disorder)	aorta bifurcatietrombose	thrombose de la bifurcation de l'aorte
I30.1	17079009	Acute bacterial pericarditis (disorder)	bacteriële pericarditis	péricardite bactérienne
I47.1	233891009	Sinoatrial node tachycardia (disorder)	sino-auriculaire tachycardie	tachycardie sino-auriculaire
I73.0	195295006	Raynaud's disease (disorder)	syndroom van Raynaud	syndrome de Raynaud
I66.2	195234007	Occlusion and stenosis of posterior cerebral artery (disorder)	a.cerebri posterior afsluiting	occlusion de l'artère cérébrale postérieure
I66.0	195232006	Occlusion and stenosis of middle cerebral artery (disorder)	a.cerebri media afsluiting	occlusion de l'artère cérébrale moyenne
I62.0	195176009	Subdural hemorrhage - nontraumatic (disorder)	niet-traumatisch subduraal hematoom	hématome sous-dural non traumatique
I47.0	195105007	Re-entry ventricular arrhythmia (disorder)	ventriculaire re-entry aritmie	arythmie ventriculaire de ré-entrée
I67.1	128609009	Intracranial aneurysm (disorder)	hersenaneurysma	anévrisme cérébral
I46.1	95281009	Sudden cardiac death (disorder)	plotse hartdood	mort subite cardiaque

Merge CMV 3BT/UZB Ctr-n09-03 -MIM

ICD_10_Code	Snomed CT number	FULLYSPECIFIEDNAME	NL_Clinical_Label	FR_Clinical_Label
I83.9	72866009	Varicose veins of lower extremity (disorder)	varices benen	varices des membres inférieurs
I65.2	64586002	Carotid artery stenosis (disorder)	stenose arteria carotis	sténose d'une artère carotide
I36.2	194992008	Non-rheumatic tricuspid valve stenosis with insufficiency (disorder)	niet-reumatische tricuspidale hartklepstenose met - insufficiëntie	sténose non rhumatismale avec insuffisance de la valvule tricuspide
I07.2	194741006	Rheumatic tricuspid stenosis and insufficiency (disorder)	reumatische tricuspidale hartklepstenose met insufficiëntie of regurgitatie	sténose tricuspidienne rhumatismale avec insuffisance ou régurgitation
I37.2	195000004	Pulmonary valve stenosis with insufficiency (disorder)	pulmonale hartklepstenose met insufficiëntie	sténose de la valvule pulmonaire avec insuffisance
I37.0	194997002	Pulmonary stenosis, non-rheumatic (disorder)	pulmonale hartklepstenose buiten RA	sténose pulmonaire, sauf RAA
I86.8	111294009	Distended umbilical veins (disorder)	caput medusae	caput medusae
I88.9	3502005	Cervical lymphadenitis (disorder)	cervicale adenitis	adénite cervicale
I51.2	5919001	Rupture of papillary muscle (disorder)	papillairspierruptuur	rupture du muscle papillaire
I49.4	75532003	Ventricular escape beat (disorder)	escaped beats	Extrasystole
I25.4	75473000	Acquired arteriovenous fistula of heart (disorder)	coronaire arterioveneuze fistel	fistule artérioveineuse coronaire
I60.3	277319006	Ruptured aneurysm of posterior cerebral artery (disorder)	subarachnoïdale bloeding uit a.cerebri posterior	hémorragie sous-arachnoïdienne de l'artère cérébrale postérieure
I60.4	276284000	Subarachnoid hemorrhage from basilar artery aneurysm (disorder)	subarachnoïdale bloeding uit a.basilaris	hémorragie sous-arachnoïdienne de l'artère basilaire
I60.1	276280009	Subarachnoid hemorrhage from middle cerebral artery aneurysm (disorder)	subarachnoïdale bloeding uit a.cerebri media	hémorragie sous-arachnoïdienne de l'artère cérébrale moyenne

This means that 1,5 % more relevant terms could be picked up with double coding and 4,7 % terms could be eliminated.

3.5. *Final terminology*

The final terminology consists of 15.742 terms of which 9.724 exist in 3BT and 6.018 come from the CMV UZB.

8.586 terms were common with the concepts in the Snomed thesaurus of Canada.

For 2.815 concepts the French translation is missing (concept not in 3BT and not in the Canada thesaurus).

ICD-9-CM was missing for 277 concepts, ICD-10 for 1.157 concepts and ICPC_2 for 7.147.

All concepts are linked with one Snomed CT concept.

There are no duplicates of Snomed CT concepts, Dutch or French label.

The link with Snomed concepts was based on the following premises:

1. A fracture without precision that it was 'open' was linked to the concept 'closed fracture'

2. Combination concepts were avoided. This means that the complication should be coded separately. The main concept is general and doesn't include the notion 'with' or 'without complication'. E.g. 'Insect bite' is related to the general concept 'Insect bite' and not to the concept 'nonvenomous insect bite without infection'

3. Concepts with notion 'primary' 'essential' were avoided because the difference between 'primary' and 'secondary' is often not clear.

4. Problems related to pregnancy were linked to the general concept and not to the concept '… delivered'

5. Certain cancers exist in Snomed as a 'disorder' and as a 'morphology'. These cancers were linked to the disorder concept.

1.364 are very specific and rare concepts labeled with 'S'. They are of limited use. The definition of 'specific' was not standardized and differed between coders.

3.6. *Conclusion*

The actual terminology covers a high number of common diagnoses in daily practice.

Terms are carefully selected and are restricted to common and precise concepts.

Of the coded concepts the agreement between coders was almost perfect.

Differences occurred in the selection of appropriate concepts to be coded.

The resulting terminology seems to be a good start but should be updated in a continuous way. A terminology system and centre should allow to add new terms, to correct and to delete existing terms.

3.7.   *Further work to do*

French translation is needed for 2.815 concepts.

The terminology should be tested in daily practice (primary care and hospitals).

To this end the terminology should be integrated in existing software for electronic medical records and /or MKG/RCM, ideally with a NLP program interface.

This testing should allow the users to identify and request missing concepts and synonyms into a terminology management system.

_____