# Introduction to SNOMED CT Translation

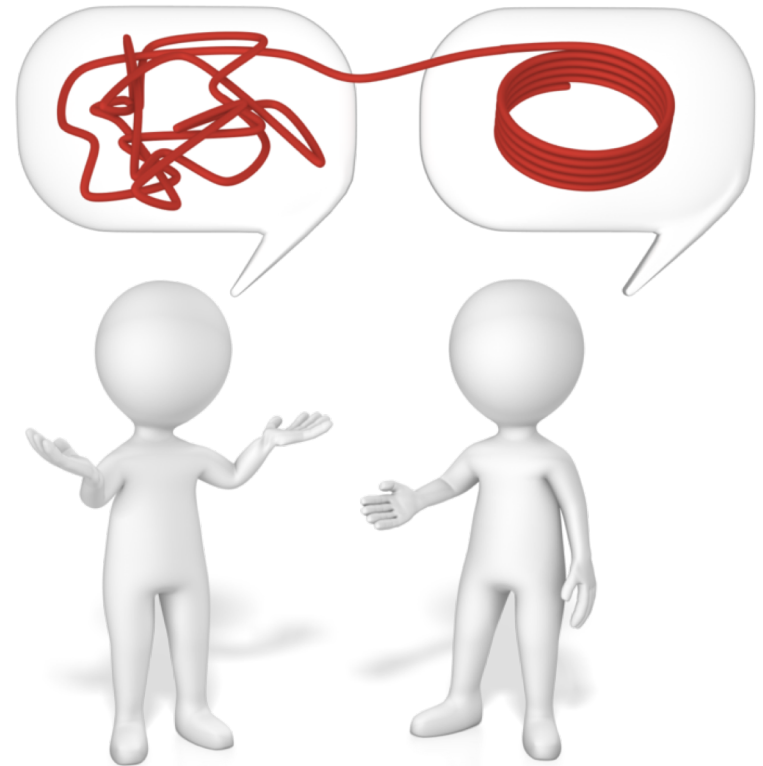*Feikje Hielkema, Maria-Cornelia Wermuth, Ian Green and Anne Randorff Højen*

Expo 2018 Tutorial

# Overview

- Introduction to translation
- SNOMED International translation initiatives and support
- Translation approaches
- Translation efficiency
- Q/A and ending
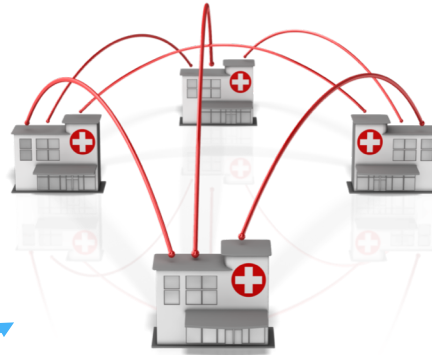
# Introduction to Translation
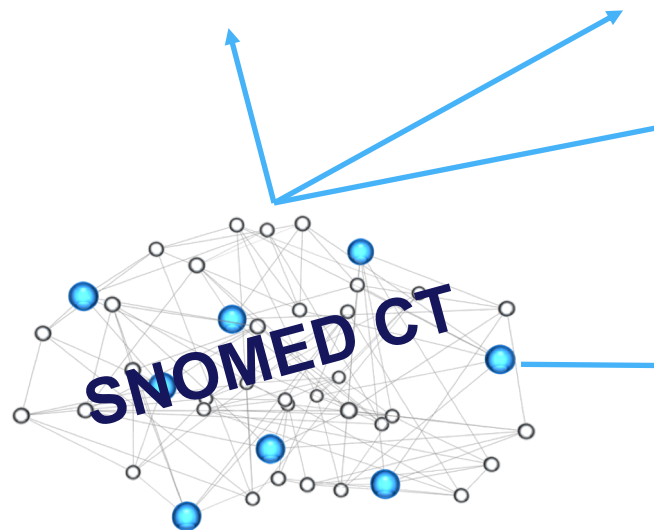
# Why Translate SNOMED CT?

People        Organisations        Countries



SNOMED CT

Translations must remain faithful to terminological and linguistic principles and at the same time be able to produce national terminologies useful for clinicians in their daily work

# SNOMED CT Translation Heuristics



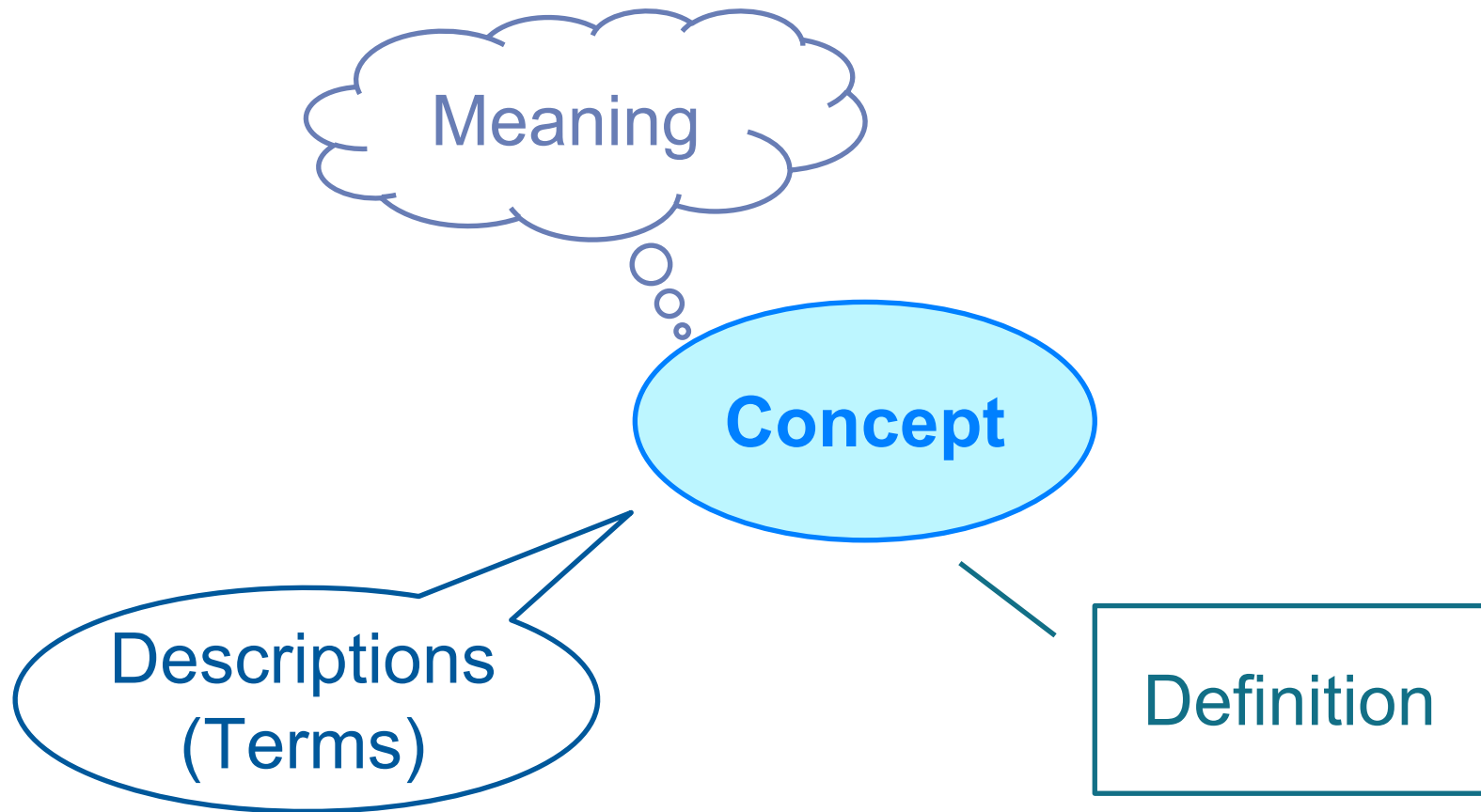Unambiguity (FSN)

Transparency

Psychological acceptability

Systematicity and consistency

Linguistic correctness

# Concept-Based Translation

# Concept and Terms

**Concept**

- SNOMED CT concepts represent clinical ideas/meanings
- Every concept has a unique numeric concept identifier

**Term**

- Terms are human-readable representations of concepts
    - A single term may represent one or more concepts
    - Different terms may represent a single concept

# Concept and Terms

**Concept:**
22298006
| Myocardial infarction (disorder) |

- "Myocardial infarction"
- "Infarction of heart"
- "Cardiac infarction"
- "Heart attack"

**Term:**
"cold"

- Cold sensation quality

- Chronic Obstructive Lung Disease

- Common cold

# Translating Concepts instead of Terms

- Concept equivalence is essential
- Concept translation requires clarifying the underlying meaning of the concept by considering the
    - meaning of the source language term
    - hierarchical position of the concept
    - relationships of the concept to other concepts

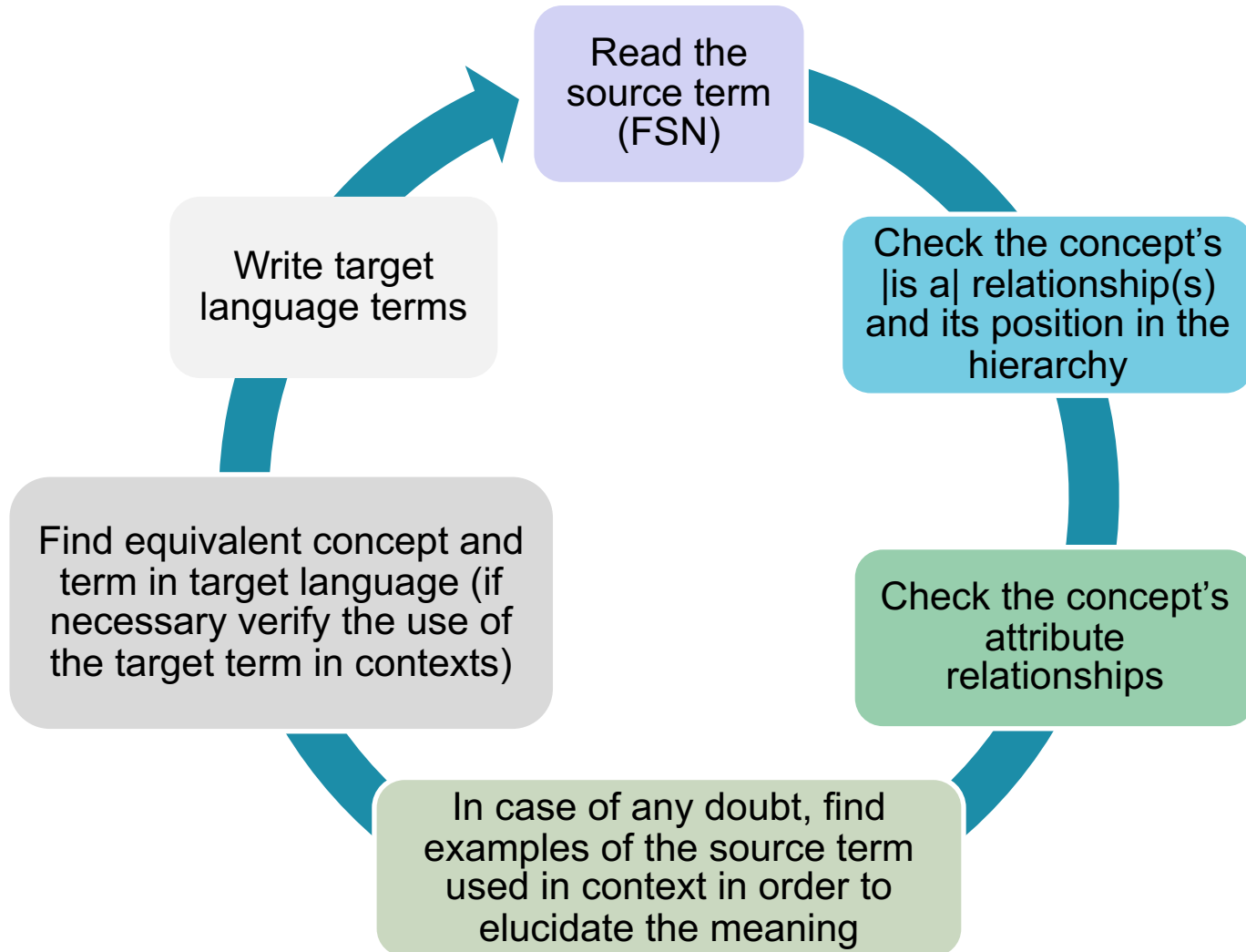| Original term | Actual translation | Correct translation |
|---|---|---|
| Sore bottom | Øm **bund** | Ømt sædeparti |

The lowest or deepest part of anything, as distinguished from the top

# Process

Read the source term (FSN)

Check the concept's |is a| relationship(s) and its position in the hierarchy

Check the concept's attribute relationships

In case of any doubt, find examples of the source term used in context in order to elucidate the meaning

Find equivalent concept and term in target language (if necessary verify the use of the target term in contexts)

Write target language terms
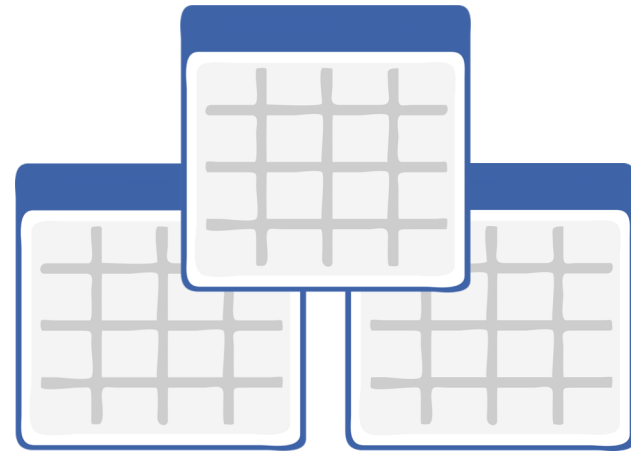
# Representing Translations

# Descriptions and Language Reference Set

## Concept File

Is used to hold the clinical concepts that make up SNOMED CT

## Description File

Is used to represent the association between human-readable phrases (terms) and SNOMED CT concepts
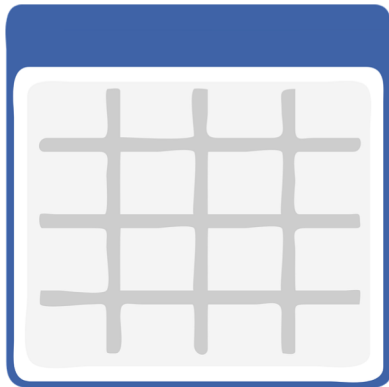
## Language Reference Set

Is used to indicate which descriptions contain terms that are acceptable or preferred in a particular dialect or use case
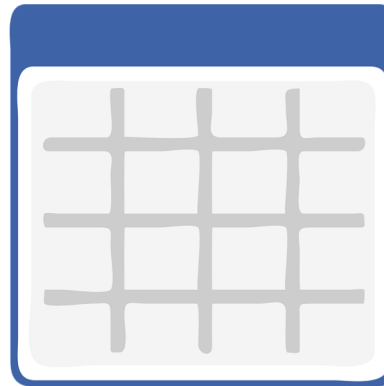
# Descriptions and Language Reference Set

**Concept File**

**Description File**

**Language Reference Set**

| | |
|---|---|
| … | 22298006 |
| … | … |

| | | |
|---|---|---|
| (S) | 37436014 | Myocardial infarction | |
| (S) | 37442013 | Cardiac infarction | |
| (S) | 37443015 | Heart attack | |
| (S) | 37441018 | Infarction of heart | |
| (F) | 751689013 | Myocardial infarction (disorder) | |

| | |
|---|---|
| (★) | 37436014 |
| (✓) | 37442013 |
| (✓) | 37443015 |
| (✓) | 37441018 |
| (★) | 751689013 |

| | | | |
|---|---|---|---|
| (S) Synonym | (F) Fully Specified Name | (✓) Acceptable | (★) Preferred |

# Descriptions and Language Reference Set
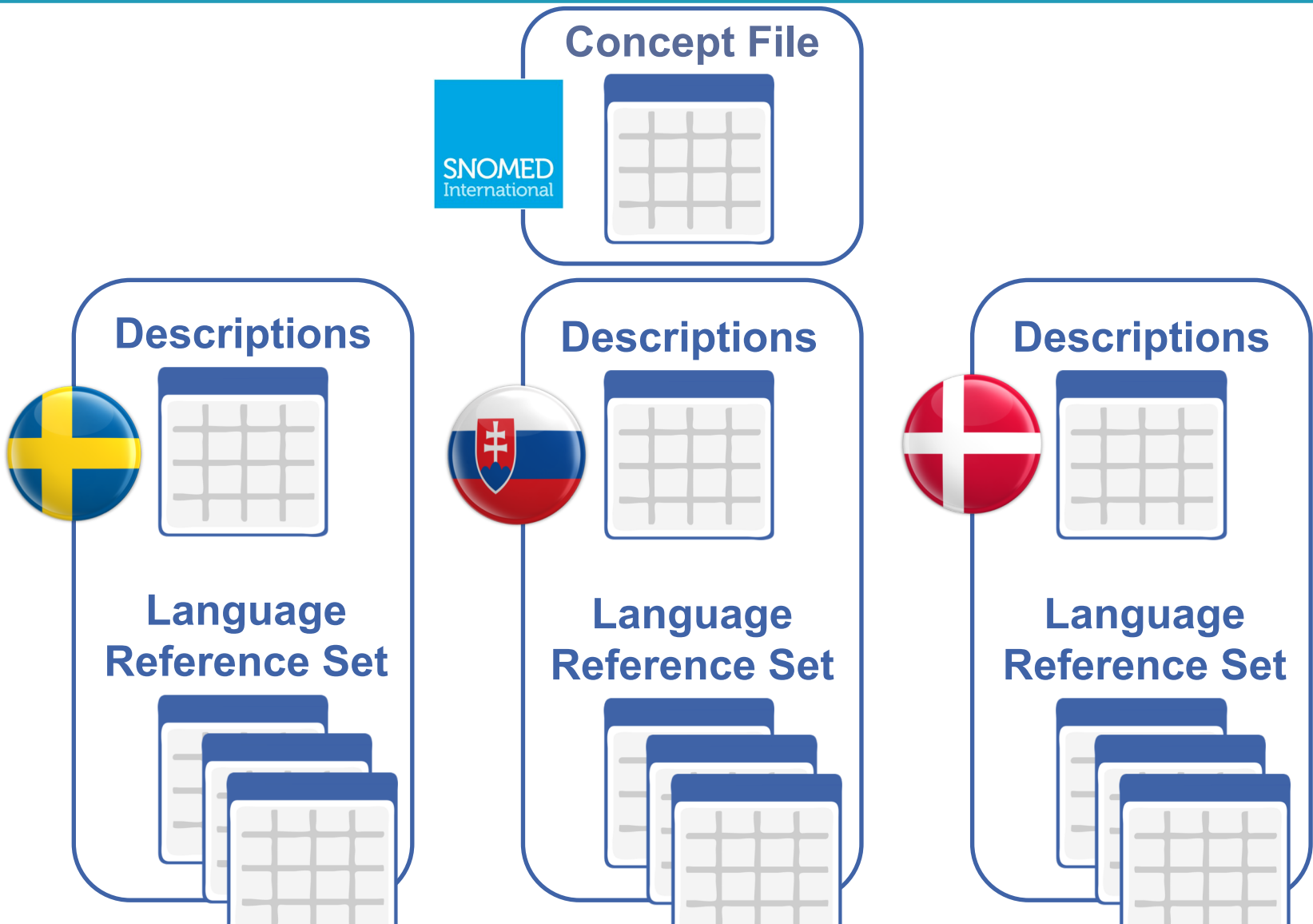
# SNOMED International Translation Initiatives and Support

# Translation Initiatives and Support

- Translation is responsibility of Members, SNOMED International does <u>not</u> undertake translation
- SNOMED International is committed to supporting Members with translation activities, via:
  - Translation policy
    - Incentivizing new and existing members to undertake translation by providing some financial support
  - Publication of translation guidance documentation and resources, developed by the Member community
  - Encourage discussion between Members on translation issues
    - to assist in developing a greater understanding of the management of translations at a national level, and the process of translation

# Translation Policy
## **Policy Assumptions**

- IHTSDO will not assume managerial responsibility for any translations other than the existing Spanish translation
- The IP of any derivative work belongs to IHTSDO
- The IP of any translated work belongs to IHTSDO and is to be shared with all Members
- IHTSDO encourages collaboration and will work to make national translation available to all Members
- Minimum core set should be translated for adoption purposes in non-English speaking Member countries

# Translation Policy
## Financial Support

SNOMED International will cover:

- 100% of a reasonable Member cost for the translation of up to 5,000 concepts in the Member's language
- the cost of translation of one implementation document per Member
- the cost of translation of one of the SNOMED CT courses as chosen by the Member
- Funding support is up to $150,000

# Translation User Group

- Forum to
  - share expertise
  - identify best practice
  - provide advice and guidance for all activities related to translation
- Led by Members and supported by SNOMED International
- Aim to promote and support the use of SNOMED CT in non-English speaking countries globally
- Confluence based discussion forum –
  - https://confluence.ihtsdotools.org/display/TRANSLATIONUSERGROUP/Translation+User+Group+Home
  - Confluence site is open to all viewers, but requires a Confluence account to partake in discussions

# SNOMED International Resources

- Translation Guidance documentation
  - Guidelines for the Translation of SNOMED CT
  - Guidelines for the Management of Translations of SNOMED CT
  - Developing a Methodology and Toolkit for Evaluating SCT Translation Quality (Background)
  - Developing a Methodology and Toolkit for Evaluating SCT Translation Quality
- Translations Confluence site
  - https://confluence.ihtsdotools.org/display/TRAN/Translations+Home
- Initial Translation Set – French, German, Chinese and Japanese
  - 6,000 terms based on usage
    - NOT a SNOMED CT Starter set

# Translation (for more information)



**Translations Home**

Created by Juliet Gole Krarup, last modified by Ian Green on 2017-Sep-19

## Welcome to the Translations space!

Here you can find information about translations of IHTSDO products, including SNOMED CT and its derivatives and product support documentation.

The global dissemination of SNOMED CT® increases the need to provide the terminology in many different languages. SNOMED CT® has a built-in framework to manage different languages and dialects. Currently, entire or partial translations of SNOMED CT® are available in US English, UK English, UK Australian, Spanish, Danish, Dutch, Lithuanian, Swedish and Canadian French, translations in other SNOMED International Member countries and elsewhere are also in progress.

| INFORMATION | SUPPORTING MATERIALS | RESOURCES |
|---|---|---|
| Introduction to translation | SNOMED International Translation Policy | Initial Translation sets |
| Guidelines for Translation of SCT | Refset/Translation tool | Member translations |
| Guidelines for the Management of Translations of SCT | IHTSDO Quality Assurance Framework | Third party translation artefacts |
| Developing a Methodology and Toolkit for Evaluating SCT Translation Quality (Background) | Automated translation | ICD-11 Translation |
| Developing a Methodology and Toolkit for Evaluating SCT Translation Quality | Useful articles | |

https://confluence.ihtsdotools.org/display/TRAN/Translations+Home

# Translation Activities

- Translation activities of members focused on SNOMED CT content and documentation
- Most important document to translate is the SNOMED CT Starter Guide
  - Supports community engagement and national/local discussions
  - Number of translations already completed
- Foundation Course
  - Spanish translation, and support provided to the course in Spanish, available online
- Good practice
  - French speaking Members and SNOMED CT Users have developed a focused group looking at French translation issues

# Translation Approaches

*Experiences from the Netherlands*

# Content

**Different translation approaches:**

- Use-case driven

- Large-scale translation

- Pragmatic approach

# Use-case driven translation

| Omschrijving↓ | SNOMED-CT FSN | Id |
|---|---|---|
| 'blow out'-orbitafractuur | 'blow out'-orbitafractuur (aandoening) | 263167007 |
| 'bull's eye'-maculopathie | 'bull's eye'-maculopathie (aandoening) | 232050001 |
| 'floppy eyelid'-syndroom | 'floppy eyelid'-syndroom (aandoening) | 231812008 |
| 'paving stone'-retinadegeneratie | 'paving stone'-retinadegeneratie (aandoening) | 69134001 |
| 'snail track'-retinadegeneratie | 'snail track'-retinadegeneratie (aandoening) | 232001007 |
| A-patroonvergentie | A-patroonvergentie (bevinding) | 246695005 |
| aandachtstekortstoornis met hyperactiviteit, overwegend onoplettend type | aandachtstekortstoornis met hyperactiviteit, overwegend onoplettend type (aandoening) | 35253001 |
| aandoening van benige orbita | aandoening van benige orbita (aandoening) | 371522000 |
| aandoening van choroidea | aandoening van choroidea (aandoening) | 128468007 |
| aandoening van conjunctiva | aandoening van conjunctiva (aandoening) | 59698003 |
| aandoening van cornea | aandoening van cornea (aandoening) | 15250008 |
| aandoening van corpus ciliare | aandoening van corpus ciliare (aandoening) | 68575007 |
| aandoening van corpus vitreum | aandoening van corpus vitreum (aandoening) | 76682005 |
| aandoening van discus nervi optici | aandoening van discus nervi optici (aandoening) | 128331005 |
| aandoening van episclera | aandoening van episclera (aandoening) | 421866006 |
| aandoening van filterblaas | aandoening van filterblaas (aandoening) | 410724009 |
| aandoening van iris | aandoening van iris (aandoening) | 85478004 |
| aandoening van lens | aandoening van lens (aandoening) | 10810001 |
| aandoening van nervus opticus | aandoening van nervus opticus (aandoening) | 77157004 |
| aandoening van ooglid | aandoening van ooglid (aandoening) | 60113004 |
| aandoening van oogspier | aandoening van oogspier (aandoening) | 128603005 |
| aandoening van pupil | aandoening van pupil (aandoening) | 68633000 |
| aandoening van retina | aandoening van retina (aandoening) | 29555009 |
| aandoening van sclera | aandoening van sclera (aandoening) | 33064008 |
| aandoening van traanapparaat | aandoening van traanapparaat (aandoening) | 31053003 |
| aandoening van voorste oogkamer | aandoening van voorste oogkamer (aandoening) | 302217008 |
| abrasie van cornea | abrasie van cornea (aandoening) | 85848002 |
| accommodatieparese | accommodatieparese (aandoening) | 68158006 |
| accommodatiespasme | accommodatiespasme (aandoening) | 30069002 |
| accommodatieve esotropie | accommodatieve esotropie (aandoening) | 419494007 |
| accommodatieve insufficiëntie | accommodatieve insufficiëntie (aandoening) | 232140004 |

# Use-case driven translation

**Requires:**

- Early-adopting domain experts
- Semantic reviewer

**Steps:**

- Determine the type of concept needed and the branch(es) these are in
- Create a simple reference set for your use-case
- Translate the reference set

**Short-term pro's:**

- Efficient in the short term
- Only translate concepts you recognize
- Adopters already involved

**Long-term con's:**

- Inefficient in the long term
- Inconsistency

# Use-case driven translation

**Alternative: mapping a local terminology to Snomed**

• Use local terms as SNOMED CT translations

• Create new SNOMED CT-concepts for any missing terms

**Pro's:**

• Result is familiar & recognizable

• Translated set is easy to adopt for users of the old, local terminology

• Presumably covers most frequently used terms

**Con's:**

• Inconsistency

• Not-quite-right translations

• Very time-consuming per concept

• Risk of creating duplicate concepts

**In the long term we have regretted going down this road...**

# Large-scale translation of entire hierarchies

**Requires:**

• Professional translators

• Semantic and medical reviewers

• Translation tooling with collaboration support

**Steps:**

• Choose your hierarchy, excluding any branches you are guaranteed never to need

• Translate

• Report errors back to Snomed International

**Pro's:**

• Efficient *per concept*

• Consistent translations

• Removes a major obstacle for future SNOMED CT adopters

**Con's:**

• Very time-consuming due to bulk

• Translate many concepts you will never use

• What if no-one adopts?

# Large-scale translation workflow
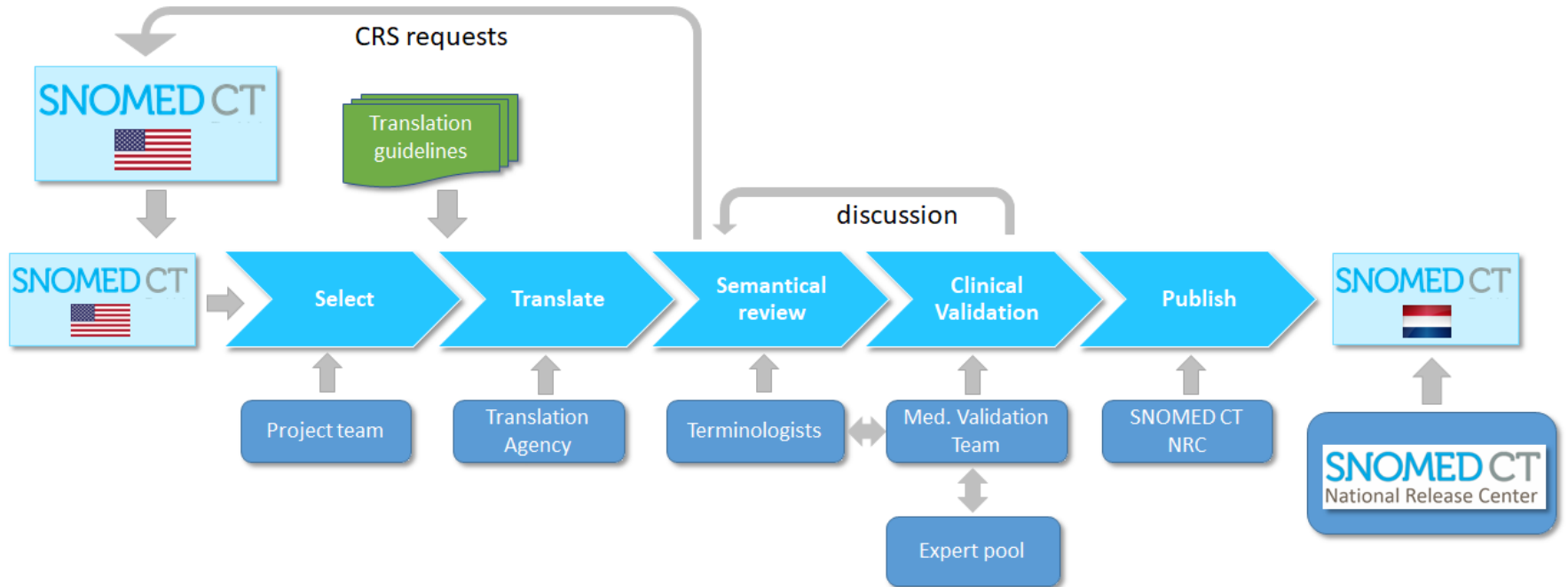
# Pragmatic translation: semi-automated

**Optimize the order in which you translate to maximize re-use**

**Requires:**

- Semantic and medical reviewers
- Translation of defining concepts (body structures, qualifier values)
- Computer-assisted translation (CAT) algorithm:
  - Generates descriptions in the target language
  - Based on translations of the defining concepts

**Steps:**

- Select a set of well-defined, uncomplicated concepts
- Translate by adopting CAT suggestions, modifying if necessary to increase recognizability
- Report incidental errors to Snomed Int. – caused by incomplete definition
- Fix any structural errors asap by updating the CAT algorithm

# Computer-automated translation generation
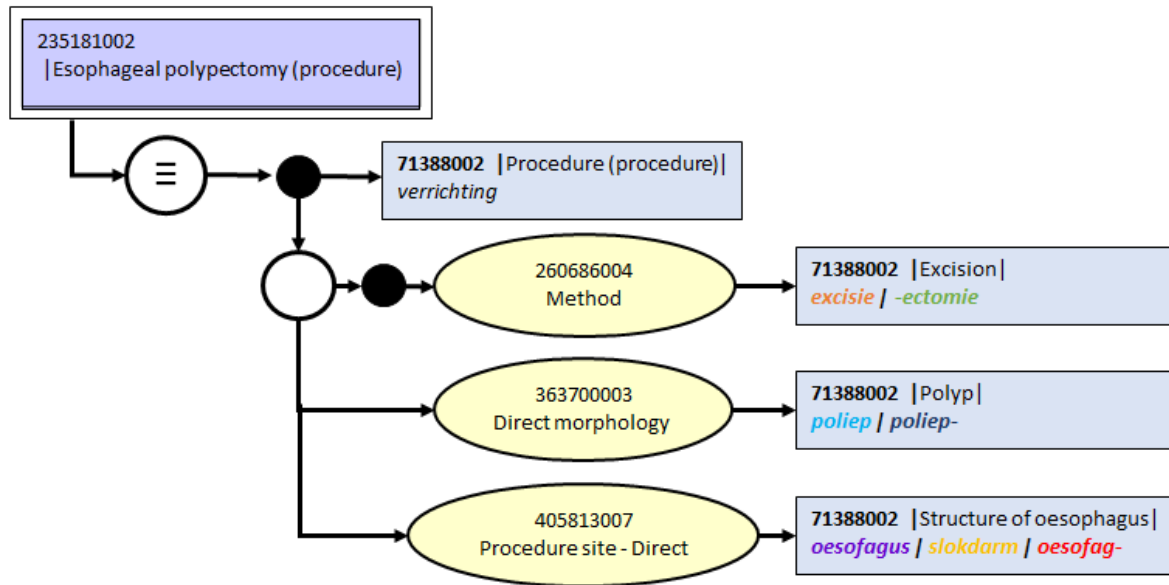


For details, come see me at the poster stand

# Pragmatic translation

**Pro's:**

- Most efficient
- Consistent results
- Well-defined concepts are rarely obscure, obsolete, erroneous, veterinary or plain confusing
- Semantic review is trivial

**Con's:**

- Inadequate for complex concepts
- Result is piecemeal, insufficient for any practical use-case
- Beware of structural errors in the CAT algorithm
- Stilted, less natural translations – medical review is essential
- Requires translating more than just disorders & procedures

# There is no best approach
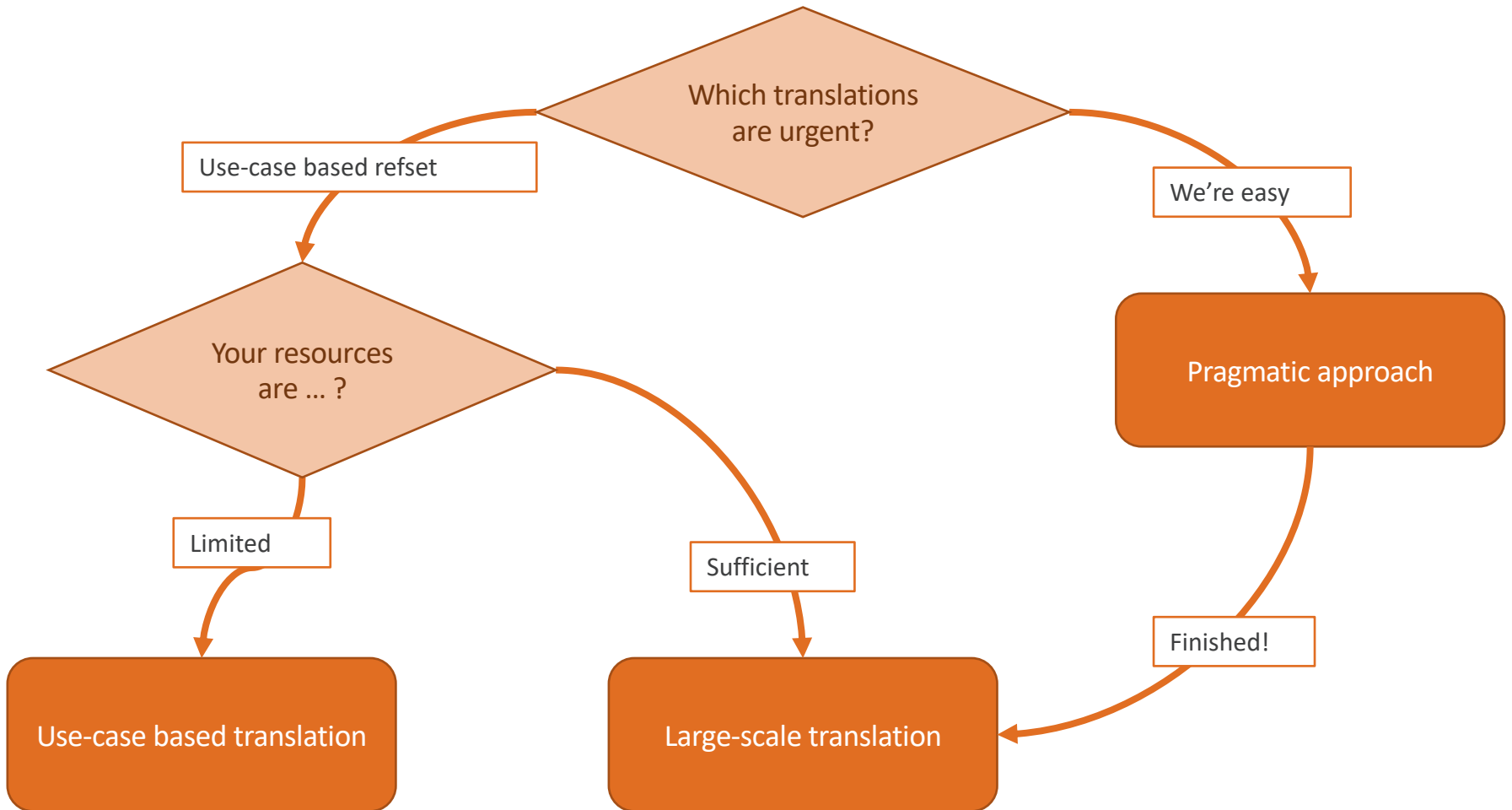
# Translation Efficiency

*Experiences from Belgium*

# Efficiency in translating SNOMED CT: Approaches, Processes, Tools

Nathalie De Sutter, Language Technology Consultancy

Katrien Scheerlinck, Federal Public Service (FPS)

Tom Van Renterghem, Federal Public Service (FPS)

Dr. Maria-Cornelia Wermuth, KU Leuven - University

National
Release
Center
SNOMED CT

# BACKGROUND

# Belgian extension: First steps in 2016

- Launch of Belgian pilot project:
  - Commissioning of cloud-based translation platform
  - Organization of workshops on concept-based translation and use of cloud platform
  - Creation of first refset *Hospital Core (*approx. 6000 concepts based on Kaiser Permanente)
    - Reason: pre-classification of concepts allows for efficient assignment to appropriate domain specialist

.be

# Belgian extension: BEMARC18

- **BEMARC18:**

  - 26.814 translated international concepts (SNOMED CT® January 2018 International Edition), including:
    - **French and Dutch preferred terms for each concept**
    - **1.945 French synonyms and 1.373 Dutch synonyms**
    - **2 new concepts**
    - **3 concepts specific to the Belgian release:**
      - *Belgian module (core metadata concept)*
      - *French and Dutch language refset (foundation metadata concept)*

# Belgian extension: BESEPT18

- **BESEPT18:**

    – Extension with 9241 translated concepts (French and Dutch)
    – Translated concepts belong to the following hierarchies:
        - Disorder
        - Medicinal product
        - Organism
        - Product
        - Substance

# TRANSLATION APPROACH

# Translation approach

- **Main objective:**
- Optimization of the translation process at the lowest possible cost

  – Concept-based
  – Use-case driven  } approach

# Translation approach

- **2 Proofs of Concept (PoC):**

- PoC 1 based on:
  - Crowdsourcing techniques
  - Kaiser Permanente
  - XTM Cloud
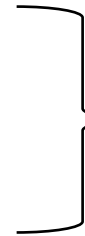
  usage-based approach

- PoC 2 based on:
  - NLP
  - Recombination techniques

  pragmatic approach

.be
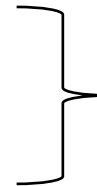
# Translation approach

- **Reasons:**

  - Low budget
  - No internal resources
  - Outsourcing to translation agency still requires validation by domain experts

# TOOLS

# PoC1: Requirements

- **Need for a**
  - Web-based
  - Robust
  - Scalable system
- **Offering**
  - Filters to cope with SNOMED-CT specific formats
  - Full traceability of all tasks performed in the system
  - Transparency through extensive reporting
- **Allowing**
  - Multiple users to work at the same time and in an optimal way
  - Adoption of flexible workflow depending on the refset/subset
  - The resulting translations to meet the highest standards

# Online interface

## Automatic notification of new assignment

# Fuzzy matching

| 16 | Closed fracture carpal bone | | Closed fracture carpal bone | | 1 U |
|---|---|---|---|---|---|
| | ID: 9468002 | | | | |
| | Comment: ⬇ | | | | |
| 17 | Closed fracture of wrist | | Closed fracture of wrist | | 1 M |
| | ID: 9468002 | | | | |
| | | Fuzzy match 79% - Approved | | | |
| | Source: | Closed fracture of ~~lunate bone of~~ wrist | fracture fermée de l'os semi-lunaire | ✓ | |
| | | Fuzzy match 79% - Approved | | | |
| | Source: | Closed fracture of ~~hamate bone of~~ wrist | fracture fermée de l'os crochu | ✓ | |
| | | Fuzzy match 76% - Approved | | | |
| | Source: | Closed ~~reduction of~~ fracture of wrist | réduction fermée de fracture du poignet | ✓ | |
| | Comment: ⬇ | | | | |
| 18 | Arucana chicken | | Arucana chicken | | 1 U |
| | ID: 10643004 | | | | |
| | Comment: ⬇ | | | | |

The translation of terms with lexical similarities is shown for reference in the interface

→ higher consistency

# Terminology



Frequently occurring words are extracted and recognized within the terms

→ higher consistency

# Communication and online collaboration



- Possibility to add comments, e.g. on ambiguous meaning or translations of which one is not sure.
- "Processing status" of each individual concept is indicated by means of a colour code (green, orange, red).

# PoC2: NLP-based - Google

# Additional tools: Machine Translation

# Additional tools: Machine Translation

# Resources

- Translation memories
  - validated RefSet specific translation memory
  - ICPN
  - ICPC

  not for both languages and not 100% reliable in quality

- Reference databases
  - ATC (Anatomical Therapeutic Chemical Classification System)
  - Legacy (non-validated) translations
  - French Starterset
  - NICTIZ
  - MedDRA
  - SAM V2 ("source authentique des médicaments"):
    - Automatic translation of concepts based on SAMv2 reference databank

# THE PEOPLE AND PROCESS

# Multidisciplinary approach

Rationale behind this approach:

" [...] **interdisciplinary collaboration** is crucial in terminology work."
(Infoterm 2005)

**Domain experts with clinical experience**

Semantically validated & usage-based translations

**Linguist**

**Terminologist**

Linguistic revision

.be

# Multi-step translation process

- **Step 1: Translation**
  - Domain specialist 1: proposes translation
- **Step 2: Revision**
  - Domain specialist 2: reviews proposed translation
- **Step 3: Validation**
  - Domain specialist 3 (Refset co-ordinator): validates translation
- **Step 4: Linguistic revision**
  - Validated terms are checked for linguistic accuracy in accordance with the basic principles of spelling and grammar, punctuation and choice of preferred term versus acceptable synonym as defined in the Belgian guidelines for translation into Dutch and French.
  - This document is constantly being supplemented and also contains specific instructions regarding certain hierarchies.

.be

# Workflow
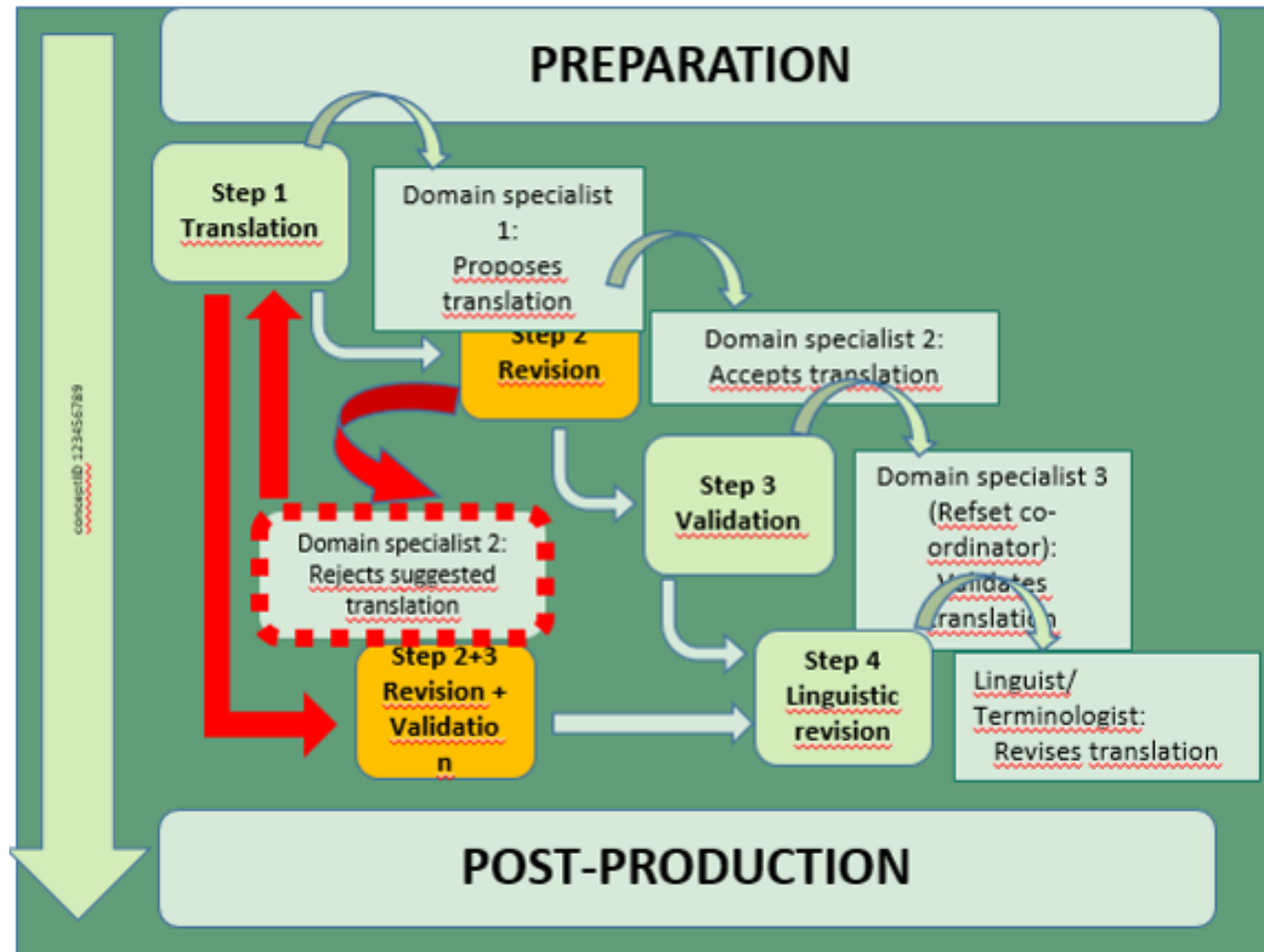
# Resources for translators

- Guidelines for Translation of SNOMED CT (NRC Belgium):
    - Practical guide for translating SNOMED CT into Dutch (Flemish variant) and French with feedback on:
        - Translation techniques
        - Spelling and grammar
        - Loanwords / bastard words
        - Specific guidelines for certain hierarchies
- Based on:
    - ISO Standards: ISO 704: 2009 and ISO 10241 – 1: 2011
    - Starter guide (SNOMED International)
    - Editorial Guide (SNOMED International)
    - Guidelines for the Management of Translations of SNOMED CT
    - Snomed International Quality Assurance Framework
    - Richtlijnen voor het vertalen van SNOMED CT (Nictiz)
    - Guide des règles éditoriales et terminologiques (Infoway)

.be

# Experienced pros

- Adopters involved from the very beginning
- Efficient approach:
  - Conceptual correctness of terms
  - Acceptability of terms in clinical practice
  - Identification of preferred term and synonyms directly by experts
  - Linguistic correctness is guaranteed by linguists (step 4) → linguistic idiosyncrasies are remedied

# Experienced cons

- Translations based on clinical/personal usage/medical school
- Disagreement on interpretation of English source concept and appropriate translation
- **But:** consensus on difficult issues can be reached through dialogue

- No unlimited recourse to voluntary work

# Summary

- Concept-based translation is essential to ensure a high quality translation
- Members need to share experiences and work together as a community
- Different approaches can be taken to undertake concept-based translations
  - Requirements vs. resources
  - Long-term plan (consistency and reuse)
- Efficient translation processes require
  - Appropriate tools
  - Active participation of domain experts
    - Translators, semantic validators

# Questions?

# Links to Further Information

- SNOMED CT Worldwide
  - http://www.snomed.org/snomed-ct/snomed-ct-worldwide
- Translations Home
  - Includes Information sheets, supporting materials and resources
  - https://confluence.ihtsdotools.org/display/TRAN/