

# Architecture for ICD 11 and SNOMED CT Harmonization

Kristina Brand-Persson, James R Campbell, Christopher G Chute, Monica Harry, Sukil Kim, Vincenzo Della Mea, Alan Rector, Molly Meri Robinson Nicol, Jean Marie Rodrigues, Stefan Schulz, Harold Solbrig, Kent Spackman, Jane Millar, Bedirhan Üstün

*Common Ontology Working Group of  
the IHTSDO – WHO Joint Advisory Group (JAG)*

Based on agreed principles for a Common Ontology for ICD 11 and SNOMED CT, an architecture has been elaborated. It is being used to help characterize what is ICD-11 and owned by WHO, what is SNOMED CT and owned by IHTSDO, and what are the Collaborative Work Products, which are jointly owned. The architecture distinguishes the components available for the end user (linearizations) from their background repository (foundation). It furthermore distinguishes the representational status of different artefacts: from context-free, pure ontological content of a SNOMED CT subset to purpose-specific monohierarchical arrangements of codes. It is finally supplemented by medical knowledge provided by the content model and by a rich collection of multilingual names, definitions and interface terms.

**SNOMED CT** is a standardized terminology for health records, based on principles of formal ontology using description logics. Here, a subset is extracted, the **Common Ontology**, mostly consisting of concepts and axioms from the highly polyhierarchical "Finding / Disorder" branch. They denote "situations", i.e. life periods of a patient having a given clinical condition. This common ontology will provide most of the entities of meaning that are necessary to represent the content of ICD-11.

**External sources** cover content that is outside the scope of SNOMED CT but considered necessary for the common ontology, e.g. gene names or new content needed for ICD-11 but still in the submission process for SNOMED CT.

The **ICD Foundation Ontology** contains the ontological content, mostly SNOMED CT, apart from the external sources. It is a multi-hierarchical taxonomy and takes additional description logics axioms from SNOMED CT. Its entities of meaning describe what is universally true for the concepts covered.

The **ICD Content model** provides multilingual names, sets of interface terms (**value sets**) and definitions, together with supportive knowledge about the ICD classes to be represented in the linearizations, e.g. diagnostic criteria, body sites, causal mechanisms, all of them linked to the common ontology. Furthermore the content model provides rules that guide the building of linearization classes, such as exclusions, which ensure the disjointness of linearization classes.

A collection of **Non-Description Logics Entities** constitutes the repository for all linearizations. They are linked via queries to the

concepts in the common ontology. These queries represent the numerous exclusion rules in linearizations and define non-ontological groupers (headings).

**Linearizations** are those releases of ICD-11 which address specific use cases like mortality, morbidity, primary care, reimbursement or classifications for medical specialties. They are familiar to the user, as they incorporate the classical classification principles (single hierarchy, non-overlapping classes, exhaustive partitions). They are expressed as queries on the common ontology, and incorporate additional knowledge from the ICD-11 content model. Residuals (NEC – not elsewhere classified, NOS – not otherwise specified) are automatically generated at all hierarchical levels. That linearizations are expressed by queries highlights their status as a special kind of terminological artefacts, which are not ontologies but whose content can be traced back to a principled ontology, viz. the **Foundation Ontology**. The hierarchical makeup may differ between linearizations, as they reflect pragmatic preferences in the arrangement of classification codes. Linearizations can also be nested.

