INTERNATIONAL HEALTH TERMINOLOGY
STANDARDS DEVELOPMENT ORGANISATION

Semantic approaches to subset management. Lessons learned from clinical template research

Presenter: Kirstine Rosenbeck Gøeg, Department of Health Science and Technology, Aalborg University.
Anne Randorff Højen, Department of Health Science and Technology, Aalborg University

Audience
Professionals interested in information management, and those who are designing and re-using subsets.

Objectives
To draw experiences from research on template management into the ongoing debate about how to support sharing of reference sets. The presentation will focus on how to utilize visualization and semantic overlap techniques to get overview of and insight into of the semantic content of subsets.

Abstract
Subsets can be used to express the permitted values for specific documentation purposes, and can e.g. contain medication permitted in a country or diagnoses used at a heart surgery department. Subsets are valuable to share because similar subsets used at multiple locations improve the chance of reaching semantic interoperability. Two important prerequisites of sharing subsets are a model for expressing subsets and metadata. IHTSDO already have a subset model i.e. RF2 makes it possible to express subsets using reference sets (refsets). Moreover, refset sharing including metadata is a work item for IHTSDO in 2014. To support the existing initiatives, we propose utilizing the IS-A relations between subset members to give a semantic overview of single subsets. In addition, similar techniques can be used to compare and contrast multiple subsets. We base our proposals on our existing research within template management for EHR systems, where overview of large amounts of templates is a similar use case. The research relevant for subset sharing is the following:

Visualization of sets of terms using the IS-A relations. In our original research [1], we discovered that visualization of sets of terms related to a single template would be valuable for template designers e.g. for improving coding consistency. Visualizations will clarify if terms are drawn from appropriate subtype hierarchies and whether coding consistency has been prioritized. This would be valuable for subsets as well, because it would be easier for new users to get an overview of existing subsets when trying to determine rightness for purpose. Moreover, visualizing the terms from two or more subsets would allow users to determine overlaps and differences.

Semantic overlap using IS-A relations. The common semantic properties between two terms can be expressed in multiple ways, but so far we have only based our research on IS-A relations. We have explored expressing semantic overlap by counting full matches and partial matches. Moreover, we currently work on quantifying semantic overlap using semantic similarity measures. Using semantic overlap analysis could help group similar subsets, and match newly designed subsets with relevant existing subsets.

Both visualization and semantic overlap analysis could support refset re-use because the techniques provide overview of and insight into the SNOMED CT content of subsets. The presentation aims at presenting the techniques and giving examples of use cases for subset sharing and management.

References
1. Højen AR, Sundvall E, Gøeg KR. Methods and applications for visualization of SNOMED CT concept sets. Applied Clinical Informatics 2014;5(1):127-152.