Proceedings of the APIII conference on "Anatomic Pathology Informatics and Imaging Support for Translational Medicine", Pittsburgh, Pennsylvania, September 9-12, 2007:Abstract 342

Oncologic Pathology in Biomedical Terminologies: Challenges for Data Integration

Olivier Bodenreider[1], Anita Burgun[2]
[1]U.S. National Library of Medicine, NIH, Bethesda, Maryland, USA
[2]EA 3888, School of Medicine, IFR 140, Rennes University, Avenue Pr Léon Bernard, 35043 Rennes Cedex, France

[1]olivier@nlm.nih.gov, [2]anita.burgun@univ-rennes1.fr

Motivation: The domain of oncologic pathology is represented in several biomedical terminologies developed for epidemiology (e.g., ICD-O), clinical practice (e.g., SNOMED CT) and research (e.g., NCI Thesaurus). ICD-O, SNOMED CT (SNCT) and the NCI Thesaurus (NCIt) are all integrated in the NCI Metathesaurus and pairwise mappings have been created between some of these terminologies. The two major dimensions in the description of neoplasms are topography and morphology. These two dimensions are present in ICD-O and SNCT. The neoplasms themselves are represented in SNCT and NCIt as pre-coordinated terms, but not in ICD-O. The anatomy component (topography) is present in all three terminologies. The objective of this paper is to explore the degree to which the representations provided by the three terminologies are consistent. The consequences of inconsistencies on data integration are discussed.

Examples: *Ovarian gynandroblastoma* (OG) is represented in SNCT with links to the morphology *Gynandroblastoma* and the topography *Ovary*. In ICD-O, *Ovary* and *Gynandroblastoma* are represented explicitly, but *OG* is present only through the coordination of the topography and morphology concepts. In NCIt, *OG* and *Ovary* are present, but not the morphology concept. However, NCIt provides a mapping between *OG* and the morphology concept *Gynandroblastoma* in ICD-O. The concepts for *OG* (disorder) and *Gynandroblastoma* (morphology) are distinct in the NCI Metathesaurus. In contrast, the representation of *Renal cell carcinoma* (RCC) is more problematic as the Metathesaurus rearranges the morphology and disease terms for *RCC* in a way that is inconsistent with the original organization in SNCT (e.g., the terms *renal cell carcinoma (morphologic abnormality)* and *adenocarcinoma of kidney* are in the same concept *C0007134*). Moreover, the topology for RCC (*Structure of parenchyma of kidney*) in SNCT is finer grained than in ICD-O and NCIt (*Kidney*).

Conclusions: Aggregating coded data from epidemiology, clinical and research sources through integrated terminologies remains challenging due to differences in the representation of oncologic pathology in these terminologies and terminology integration resources, such as the Metathesaurus. However, the existence of shared codes between ICD-O and SNCT and mappings between NCIt and ICD-O can help assess consistency and suggest corrections.