

# Using SNOMED CT subset for pathology data retrieval



醫院管理局  
HOSPITAL  
AUTHORITY

Barry FONG, Kent TSUI, John MOK, Vicky FUNG

Information Technology & Health Informatics Division, Head Office, Hospital Authority, HKSAR

## Introduction

The need for secondary use of pathology data is increasing because cancer is a leading cause of death in Hong Kong. The standardization of anatomical pathology data enhances advanced informatics program to support patient care. SNOMED CT subsets have been defined to facilitate use of pathology data for cancer case notification, decision support and data analytics.

## Background

In Hong Kong Hospital Authority, there are 11 anatomical pathology laboratories. Their laboratory information system has an SNOMED III auto-encoding function by matching keywords in the diagnosis result of a pathology report. Encoded SNOMED III data were used to support various cancer related programs. However, SNOMED III will be retired in 2017 and the highly localized SNOMED III codes were lack of maintenance. Therefore, data retrieval program will be migrated to using SNOMED CT.

Keywords for diagnosis and localized SNOMED III were consolidated as a corporate SNOMED III list. The mapping of SNOMED CT concept to corporate SNOMED III list has been endorsed by Anatomical Pathology Working Group (APWG) in 2015. With the support and governance from APWG, SNOMED CT subsets have been defined for retrieval programs to support quick, comprehensive and accurate data retrieval.

## Design

- Apply program SNOMED III queries on corporate SNOMED III list.
- Group retrieved terms with reference to queries.
- Define subset nodes according to program criteria. e.g. topology, definite malignancy.
- Include SNOMED CT concepts which mapped to retrieved terms as subset node members.
- A subset was defined with multiple subset nodes.

## Result

- SNOMED CT subsets represent retrieval programs.
- Subset nodes represent retrieval criteria.
- Retrieval with SNOMED CT subset includes all terms retrieved with original program queries.
- Terms previously encoded with inaccurate SNOMED III can be retrieved based on SNOMED CT mapping.
- Retrieval result by SNOMED CT subsets is more complete.

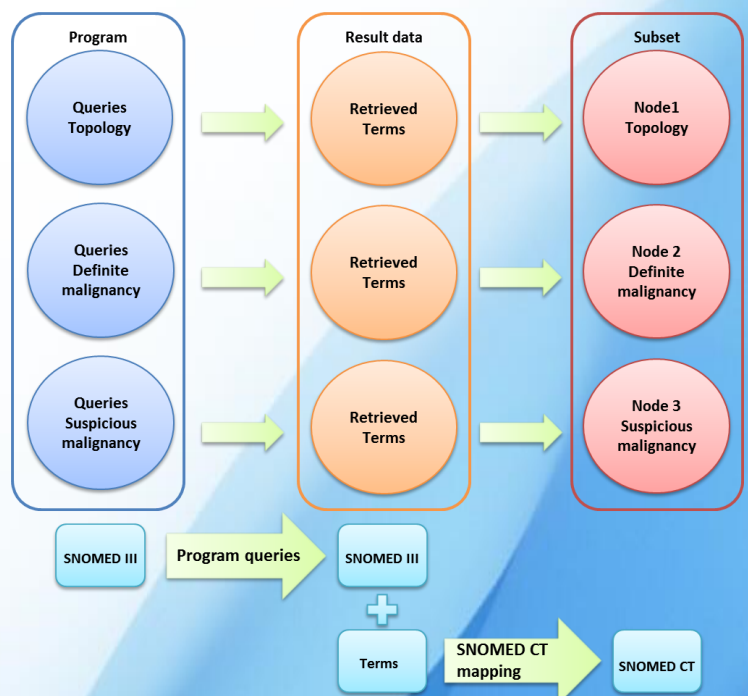


Fig 1. An overview of building subset and subset nodes.

## Conclusion

SNOMED CT has become the primary key for data retrieval and subsets will be used for management of program retrieval criteria. SNOMED III code will be retired and become local code.

Maintenance of highly variable local SNOMED III code is not necessary because any new terms mapped with subset concepts can be automatically included. Each retrieval program has a corresponding subset. Subset nodes facilitate flexible queries by inclusion or exclusion of nodes. e.g. To exclude 'suspicious malignancy'.

The poly-hierarchy structure of SNOMED CT provides knowledge for data analysis and quality assurance. e.g. To check if retrieved terms are definite malignant with SNOMED CT subsumption testing. Hence, accurate data can provided to improve health care services.

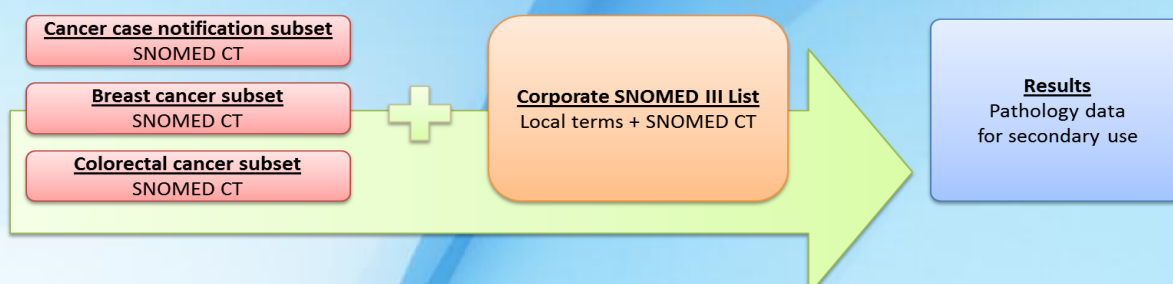


Fig 2. Example of using different SNOMED CT subsets for pathology data retrieval.