Terminology map implementation Guide

Using terminology concept maps

31 January 2021

Contents

Introduction

The purpose of this document is to provide some implementation guidance for selecting and implementing terminology maps against data.

1 Understanding the scope of your data and use case

Before applying a map, it is important to understand your data and the reason for mapping the data.

Consideration should be taken into

- Where the data was sourced from and what is the scope of the data that you have. For example, is the data from a Procedure code field and only contains surgical procedures, is the data from a Presenting problem field and contains a mix of free text and health problem codes. It may also be important to know what code set version(s) may be.

- Workflow – who recorded the data and for what intent did they record the data. Is all the data coded or does the workflow allow users to break out of the preferred coding system?

- What use case are you mapping for – what is the end result that you are after?

2 Selecting an appropriate map

Maps should have a defined and specific purpose, which:

- Provides context to the map

- Influences decisions/rules made when mapping and how to map when there is discrepancies between the source and target code systems

A map is tightly bound to the maps' purpose, use case and requirements. This means that it is important to understand why you are mapping your data and to be able to select an appropriate map.

When considering if an existing map is appropriate or building a new map, it is important to understand the following:

- What is the main purpose of this map?

- Who is the intended audience?

- What is the scope of content (source and target)?

- Who is responsible for developing and maintaining the map?

- How will the map be implemented?

Where there is an existing map that matches your required use case and scope, then it may be appropriate to use.

If they do not match, a decision needs to be made on whether the map may be a suitable starting point to augment and update, based on your requirements, or whether a new map needs to be built.

3   Preparing your data

Once a map has been built and/or identified as suitable for the required use case, your data (the source data) needs to be prepared appropriately.

3.1   Data considerations

When you are preparing your data for mapping (or processing using the map), consider the following:

- Is your data uniquely identified?

- Is it Patient level data, and are Patient cases uniquely identified by one or more indexing fields? (MRI? Episode number? Visit time/date?, All three?) Have you decided your foreign key?)

- Are you working with principal codes only, OR a principal code + N additional codes?

- Are the fields for all these codes in separate and identified sequential columns or are these arranged in rows per Patient?

- If you are not using Patient level data, but rather using the map as a reference file for later joining back to Patient data, does the data contain duplicates, If yes, remove duplicates leaving only unique identifiers (this means you will be using the id as the foreign key).

- Ensure your encoded data does not have any leading or trailing white space.

- Not all software products that are in use for data management are capable of dealing with terminology encoded data. Microsoft Excel and import/export routines from EMRs or data warehouses often corrupt long identifiers making 12115441000119102 appear as 1.21154E+16; these scientific numbers are not usable by map products. If you are using SNOMED encoded data, or other codes with long identifiers, items should be a TEXT field (not integer) and the text data type needs to be preserved in each analytic step.

- Encoded data needs to match the format of the source codes that your map requires so that it can parse correctly.

> When using SNOMED data, ensure that the correct IDs are present in your data – most SNOMED maps use concept IDs rather than description IDs. You will need to convert all Description ids to concepts ids using an official SNOMED reference file OR obtain the concept id in your specified data extract from the EMR or warehouse.

  - Some systems re-represent codes from official code systems; these are proprietary and will not be recognised by most formal map products. For example: identifiers such as CKI!240299002 for a SNOMED code will not be usable, whereas the SNOMED code 240299002 will be recognised. You will need to parse the identifiers to obtain the official and formal SNOMED identifiers before using the map.

  - Some EMR systems (and jurisdictions) have their own SNOMED namespace and have deployed concepts in their system that are not (yet) part of the official

SNOMED product. These often look like a proper SNOMED identifier but are not shared or available to the entire user community – they are local additions only. These namespace concepts will not be recognised by the formal map.

- Source data also needs to match the scope of the source codes that your map requires. If the source data does not match the scope, then you will not get a map result. A decision needs to be made on whether to exclude this data prior to mapping or wait to receive no map result.
  - o It is not always possible to tell just by looking at the data that it matches the scope of the map; additional reference and look-up files will be required. If you decide to exclude this data before using the map, you will need to firstly and independently QA your data against official code system files (not a map product) and exclude any data that does not meet the scope.
  - o Alternatively, it may be easier to use the map product and then exclude any records that do not have a map. This MAY decrease your data sample size. "No map" results do not necessarily indicate the quality of a map. Often, "no map" results are more likely to indicate poor quality data, invalid data or inconsistencies in data as originally documented and collected.

When using SNOMED data, scope is often defined by which hierarchies the content belongs to e.g. Procedure hierarchy (containing procedure type concepts), Clinical Finding hierarchy (containing problems and diagnoses)

However, terminology-to-data-item binding is often described by EMRs but is poorly implemented or used differently in real life. Your system may specify that the SNOMED hierarchy of Clinical Finding is bound to the Diagnosis data field in the system, but clinical users remain able to enter a Procedure code in this Diagnosis field. System and data specifications themselves give an indication of what the scope of your data SHOULD BE, but the actual data may not be consistent.

3.2   Mapping Considerations

3.2.1   Source data

- Where data has been obtained over a period of time or from multiple implementations, there may be multiple versions of a code system within your data. It is important to note what versions of code system coverage your selected map contains, as there may be a discrepancy. If your map contains only active codes from a future version of the code system compared to your data, you may have a number of inactive codes that will not map even though they may have been valid at the time of recording.

Maps are often built to a single version of a code system. Not all maps contain maps for previous versions of a code system.

3.2.2   Target data

- Ensure that the target data of the map meets your requirements – maps are built on rules and use case requirements that influence the map targets. If your requirements do not match the use case of the map, this may result in unexpected and inappropriate map results for your use.

- If you are using a data set that contains records encoded in different code systems, you will also need to be aware of the map target codes that exist in your data collection, their edition and format.

4   Applying the map

Depending on how the map is supplied, there are different methods of applying the terminology map. Two common formats for maps are a tabular file in the form of a csv or tsv, and the FHIR concept map.

A tabular file format can be loaded with a tool like Excel or PowerBI and source data can be cross referenced to provide the resulting output.

A FHIR concept map can be stored in a FHIR terminology server and an $expand and translate function used.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.