

A Dynamic Medical Terminology Mapping System – MeTMapS

Shao F. Liang, PhD¹, Talya Porat, PhD¹, Archana Tapuria¹, Jean-François Ethier²,
Brendan C. Delaney, MD³, Vasa Curcin, PhD¹

¹King’s College London, London, UK; ² French Institute of Health and Medical Research,
Centre de Recherche des Cordeliers, Paris, France; ³Imperial College London, London,
UK

Abstract

Medical terminology mapping has been widely recognised as a complex yet critical prerequisite for querying and integrating patient data from heterogeneous data sources. Several studies have tried to address this issue but only a few systems have successfully implemented mappings from multiple medical terminologies. While these systems typically provide a dynamic and thorough terminology coverage, they exhibit several limitations in terms of usability and sustainability. To address these limitations, we have developed MeTMapS, a dynamic system based on UMLS, allowing users to select desired terminologies and author their own mappings. Users can also add their project-specific terminologies for mappings if appropriate terminologies are unavailable from the UMLS collection. The system is easy to use, set up and maintain and provides a long-term stabilised service. The search results from users’ queries are organised into tree structures, so as to provide a better view on relationships among terms, their parents, siblings and descendants. The interface allows users to search terms and select terminologies; browse results and pick up mappings from structured trees; remove unwanted mappings and store results for reuse. All the information is displayed on a single page, simplifying navigation and information presentation. MeTMapS was evaluated by a set of GPs, clinical informaticians and IT specialists months and has received positive feedbacks.

Introduction

There are good reasons to use standard terminologies [1] such as SNOMED CT [2], ICD10* and READ Clinical Terminology [3], to integrate legacy data, comply with meaningful requirements and to unleash the next generation of analytics. However, healthcare data are too broad and dynamic to have “one size fit all” terminology for information interoperability [4], and hence mapping between terminologies becomes essential to make data readily available for analysis on national and international scale. The research community is working towards facilitating the process of terminology mapping and to that goal developed systems such as LexEVS[†], BioPortal and UTS[‡].

LexEVS (Enterprise Vocabulary Services)[§] is a central terminology server developed at Mayo Clinic^{**}. Users load data into the LexEVS database where it is transformed into LexEVS data objects enabling users to query the data and create their own mappings. LexEVS has been used as the terminology mapping centre in several projects such as TRANSFoRm [5] and BioPortal

* <http://www.who.int/classifications/icd/en/>

† <https://wiki.nci.nih.gov/display/LexEVS/LexEVS>

‡ <https://uts.nlm.nih.gov/home.html>

§ <https://wiki.nci.nih.gov/display/LexEVS/LexEVS+Documentation+and+Training>

** <http://www.mayoclinic.org/>

[6], however, it has some limitations. Using LexEVS requires prior knowledge relating to mappings and it is complicated to load terminologies. Every change of terminology version needs to be reloaded and the mapping needs to be recreated. In addition, LexEVS APIs do not currently conform to the official Object Management Group (OMG) Clinical Terminology Service Version 2 (CTS2)^{††} specification.

BioPortal^{‡‡} is a joint project between Stanford University, University of Victoria and Mayo Clinic USA. It stores a large collection of ontologies from biomedical domains. Originally BioPortal has its own terminology database and then adapted LexEVS as its second terminology database. However, these two databases were designed by different people and to have different formats. Therefore, transforming them to be visualised in a single system becomes a double work. In order to simplify the process for visualising the contents from the databases, since 2013 BioPortal has been working on replacing the two databases backend with a single RDF database [7]. Furthermore, the recent CTS2 development framework^{§§} developed by Mayo Clinic has adopted BioPortal widgets^{***}, to facilitate the use of BioPortal ontologies. Although BioPortal's widgets have provided useful visualisation of terminologies in tree structures, multiple ontologies cannot be viewed and compared on a single page. In addition, BioPortal does not have a sophisticated browsing facility to filter elements for searching as provided by Unified Medical Language System (UMLS) Terminology Services (UTS).

UTS provides both web interfaces and Web Services for users to search and retrieve UMLS data. The UTS web search allows searching in various ways, such as by Term, CUI or Code; release version; different input text type; and source name. The search results can be viewed in three styles: Basic, Report and Raw. However, the UMLS Metathesaurus displays general views and does not allow filtering the results according to desired terminologies. Users are advised to customise the UMLS Metathesaurus for effective use in their own applications.

Following a thorough analysis of the current terminology mapping systems: LexEVS, BioPortal and UMLS, we have developed MeTMapS with the aim to utilise their advantages and address their limitations, by creating an easy to use, efficient and sustainable medical terminology mapping service.

Implementation

MeTMapS's architecture (**Figure 1**) contains four main parts: UTS APIs, BioPortal Widgets, User Interface and the Additional Resource Centre (ARC). The UTS APIs and BioPortal Widgets are the bridges for MeTMapS to send queries and receive responses from UMLS and BioPortal, acting as the backend of the system. ARC is a supportive functionality in the case that resources are not available from both BioPortal and UMLS. ARC relies on mapping files and new ontologies. Each of the mapping files contains the source codes, descriptions, target codes and Concept Unique Identifiers (CUIs), and has been pointed to a newly created ontology. The CUIs need to exist in the UMLS collection and to be linked to either the source or target codes. The paired ARC ontologies are generated offline and need to be uploaded to BioPortal for the tree view purpose. For example, the INFORM project, described below, requires READ Clinical Terminology Version 2 codes to be mapped to both SNOMED CT and READ Version 3 codes. Since the READ V2 is unavailable from both BioPortal and UMLS, a mapping file needed to be produced for READ V2. We used SNOMED CT's CUIs from UMLS, and then used the mapping data among READ V2, READ V3 and SNOMED CT from the UK Technology

^{††} <http://www.omg.org/spec/CTS2/>

^{‡‡} <http://biportal.bioontology.org/>

^{§§} <http://informatics.mayo.edu/cts2/framework/2011/10/version-0-3-0-released/>

^{***} http://www.bioontology.org/wiki/index.php/NCBO_Widgets

Reference data Update Distribution (TRUD)⁺⁺⁺ site to produce the mapping file. OWL APIs^{###} were then used to produce READ Version 2 ontology and upload it to BioPortal. Once the paired files are ready, MeTMapS can perform the mapping from READ V2 to other chosen terminologies.

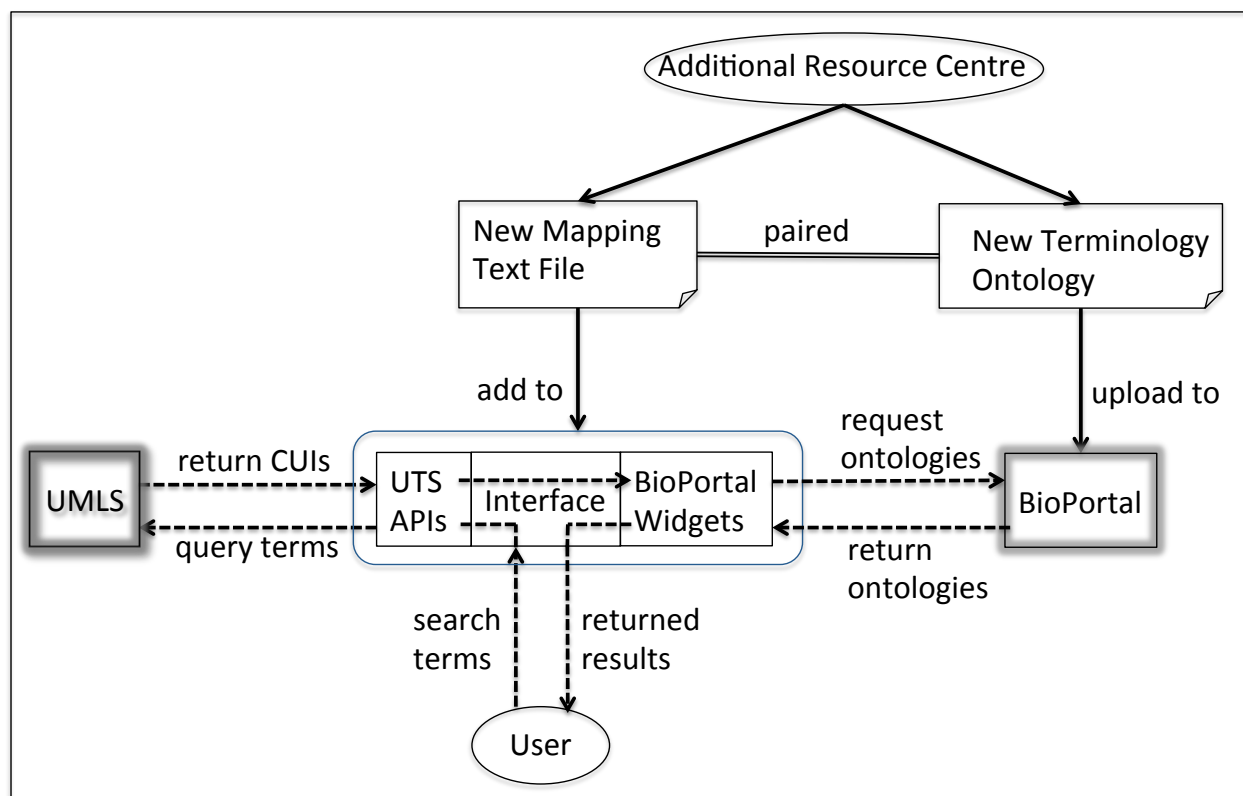


Figure 1 MeTMapS accepts users' search terms and send requests to UMLS for CUIs. After the CUIs have been identified, MeTMapS will extract terminology information from BioPortal and using tree widget to display the returned results to users. To deal with unavailable resources, new terminologies need to be created and uploaded into BioPortal. Each new ontology requires a paired file containing CUIs, term descriptions and codes to be created and stored into the Additional Resource Centre so MeTMapS can work well with the new resources.

The MeTMapS user interface comprises three main sections: 1) Search terms and select coding systems; 2) browse results and select mappings from structured trees; and 3) remove unwanted mappings and store results for further use. MeTMapS gives users full control, enabling them to select only the terminologies they need. The search results are organised into tree structures, providing users a better view on relationships among the selected terms, their parents, siblings and descendants. To minimise the setup needed and the learning curve for the users, the system is designed as a web application with a step-by-step workflow guide, with short and simple instructions for users to follow. All the relevant information is displayed and responded on one page, facilitating navigation and information processing. Auto-completion is also provided to help users quickly locate their desired terms.

Figure 2 shows the first section of the MeTMapS interface, where the user searches for *Echocardiography* from READ CTV2 (RCTV2) and V3 (RCD). Although MeTMapS has been designed to contain three sessions, users can only see the first session at the beginning. This is because we intend to make the user interface as simple as possible.

⁺⁺⁺ <https://isd.hscic.gov.uk/trud3/user/guest/group/2/home>

^{###} <http://owlapi.sourceforge.net/documentation.html>

Therefore, the next session will only appear when the current session is completed. On the top of each session, there is an instruction to instruct users and the instruction is displayed at the top of each section after the bolded section number.

MeTMapS - Medical Terminology Mapping System

1 Type in or select a search term and its desired terminologies:

Search Term:

Terminology: Read CTV2 Read V3 (RCD) ICD10 SNOMED CT ICP2C Plus

Figure 2 MeTMapS Section 1 - selecting a search term and its terminologies

After the user click on the “*Search*” button, a spinner will start spinning to indicate the system is in progressing on finding results. The search results from RCTV2 and RCD will be returned to the user in section 2 after the spinner disappeared as shown in **Figure 3**. Each item in the result set can be expanded to provide additional information.

MeTMapS - Medical Terminology Mapping System

1 Type in or select a search term and its desired terminologies:

Search Term:

Terminology: Read CTV2 Read V3 (RCD) ICD10 SNOMED CT ICP2C Plus

2 Select mapping from each ontology

- Echocardiography (RCD: X77c1)
- Cardiac US scan (RCD: X77c1)
- Cardiac echo (RCD: X77c1)
- Echocardiogram (RCD: X77c1)
- US scan of heart (RCD: X77c1)
- US heart scan (RCD: X77c1)
- US heart scan NOS (RCD: 5853z)
- U-S heart scan (RCTV2: 5853.00)
- Echocardiogram (RCTV2: 5853.11)
- U-S heart scan NOS (RCTV2: 5853z00)
- Diagnostic echocardiography (RCTV2: 7P0H.00)
- OS diagnostic echocardiography (RCTV2: 7P0Hy00)
- Diagnostic echocardiograp NOS (RCTV2: 7P0Hz00)

Figure 3 MeTMapS Section 2 - search results

Once an item is clicked from the search results, its sub set of the terminology will be displayed in a tree view, from which the user can select desired terms, which relate to the siblings or children of the selected item. After all desired items are checked in section 2, the user can click on the “*Add*” button to add all selected items to the mapping list in section 3 as shown in **Figure 4**. The users can also remove unwanted items by clicking on the red-cross symbol if they decide to exclude those items from the list. The final mapping can be saved into the MetMapS database for reuse by clicking on the “*save mapping*” button.

MeTMapS has a single screen interface, with all operations provided in the main view. In a wizard-like fashion, each section is displayed only after the previous section was completed to avoid confusion and information overload. The architecture of MeTMapS focuses on building a dynamic system, reflecting the changes in the underlying mappings. We utilised open sources from BioPortal and UTS in order for our system to be updated simultaneously each time BioPortal and UMLS are updated. The backend of the system is set to connect with UMLS and BioPortal. For connecting with UMLS, we have employed UTS APIs into our system. The main reason to include UTS APIs into our work is the final decision after comparing the meta-resources and functionality between UMLS and BioPortal. We have found that although BioPortal provide nice hierarchical tree displays which meet our purpose, the meta-data resources from UMLS is more powerful.

MeTMapS - Medical Terminology Mapping System

1 Type in or select a search term and its desired terminologies:

Search Term:

Terminology: Read CTV2 Read V3 (RCD) ICD10 SNOMED CT ICPC2 Plus

2 Select mapping from each ontology

>>>Add>>

3 Final mapping

Preferred Name	Terminology Code	Remove
Echocardiography RCD	X77c1	X
U-S heart scan	RCTV2 5853.00	X

Search results for Echocardiography (RCD: X77c1):

- Echocardiography(X77c1)
- Contrast echocardiography(X77c8)
- Epicardial echocardiography(X77c7)
- Exercise stress echocardiography(X77c9)
- Fetal echocardiography(X70mc)
- Intracavitary echocardiography(X77c6)
- Intravascular echocardiography(X77c5)
- Transoesophageal echocardiography(X77c4)
- Transthoracic echocardiography(X77c2)

Other results:

- Cardiac US scan (RCD: X77c1)
- Cardiac echo (RCD: X77c1)
- Echocardiogram (RCD: X77c1)
- US scan of heart (RCD: X77c1)
- US heart scan (RCD: X77c1)
- US heart scan NOS (RCD: 5853z)
- U-S heart scan (RCTV2: 5853.00)
- Echocardiogram (RCTV2: 5853.11)

Figure 4 MetMapS interface

For example, the body structure class in Galen ontology is mapped to two ontologies: SNOMED CT and Bilingual Ontology of Alzheimer’s Disease and Related Diseases, but the same class in SNOMED CT is mapped to five ontologies including Galen. In theory, viewing the body structure class from Galen or SNOMED CT should result in the same five mappings, but this is not the case in BioPortal. In addition, BioPortal supports fewer non-English language mappings than UMLS.

Providing an easy and simple to use interface required a complex backend design. We have used WSDLs (technical components) provided by UTS for MeTMapS’s connector to work with UMLS data and used Java Server Page design to organise all the information in one page and in a wizard-like display to facilitate navigation.

Results

The MeTMapS system was evaluated, using a common user-centered design method for evaluating the usability of a system, by seven academic and clinical research users including GPs, clinical informaticians and IT specialists at King’s College London over the period of two months. Testing with five users is thought to be optimally cost-effective [8] as the method quickly identifies areas in which navigation and layout needs improving. Participants were asked to think aloud (verbalise their thoughts) while using and interacting with the system, which was followed by semi-structured interviews with a topic guide that explored their experience and collected suggestions for improvements. A total of 54 sessions were completed, where the system was used by the researchers. Participants were asked to map the required clinical terms in 'Hypertension' clinical domain either from ICD10 to Read CTV2 or from ICD10 to Read CTV2 and V3. Participants also tested a list of clinical terms extracted from the study definitions including the inclusion and exclusion criteria of the INFORM clinical trial protocol [9]. The clinical terms listed were mapped to the clinical terminologies Read CTV2 and V3 that are used by most of the GP systems recruiting patients for the trial. A screen shot of searching for *Echocardiography* and the mapped results from Read CTV2 and Read V3 is shown in Figure 3 (the listed results contain the same CUI as Echocardiography from UMLS). A list of other clinical terms was also tried to test the various ‘search’ attributes of the system. The correct

term always appears first with the most relevant results being at the top, and suggestions of terms are provided while typing. The search terms are not order- or case-sensitive. It handles terms separated by hyphen. (For example, search term ‘sugarfree’ includes ‘sugar-free’ and search term ‘breast-cancer’ includes ‘breast cancer’.) It handles exact term matches for different concepts such as ‘Fundus coloboma’ and ‘Gastric fundus structure’ and ‘Fundus of gallbladder’. It handles abbreviations such as ‘CAT scan’ for ‘computerised axial tomography image’. Medical syndrome names are represented either by generic descriptive names (e.g., ‘Acute febrile mucocutaneous lymph node syndrome’) or with eponyms (e.g., ‘Kawasaki disease’ for the same syndrome). When searched with the eponym, MeTMaps suggests the generic descriptive name as well, but when search phrase is the generic descriptive name, it does not suggest the eponym. Multiple terminologies can be simultaneously selected for mapping. The researchers were generally satisfied with the clinical concept coverage, term accuracy and efficiency of the system to use. We have found that MeTMapS could be improved by providing suggestions on misspelled terms, and also enhanced on handling partial words such as ‘Myo inf’ as opposed to ‘Myocardial infarction’. Its search can also be expanded to handle known synonyms such as ‘kidney’ and ‘renal’.

Conclusions

MeTMapS was developed as a simplified tool for performing code mappings, with the aim of easily combining local coding systems with publicly available standards. It requires no previous training from the users and little system maintenance. At present, the performance of MeTMapS can be affected by the Internet connection and the server availabilities of both BioPortal and UTS. The current version is designed as a standalone web-based tool, and in the next phase we plan to integrate it with an EHR-enabled clinical trial management system and a query workbench style interface for observational studies.

Acknowledgment

This work is supported by the National Institute for Health Research (NIHR) Bio-medical Research Centre at Guy’s and St Thomas’ NHS Foundation Trust and King’s College London. The work presented here has also received funding from the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care programme in South London, based at King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

References

- 1 Sachdeva S, Bhalla S. Semantic interoperability in standardized electronic health record databases. *Journal of Data and Information Quality*. 2012;**3**(1):1-37.
- 2 Spackman KA, Campbell KE. Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. *Journal of the American Medical Informatics Association*. 1998:740-4.
- 3 Read JD, Benson TJR. Comprehensive coding. *Br J Healthcare Computing*. 1986;**3**:622-5.
- 4 Rector A. Clinical Terminology: Why is it so hard. *Methods Inf Med*. 1999;**38**(4):239-52.
- 5 Ethier J-Fo, Dameron O, Curcin V, et al. A unified structural/terminological interoperability framework based on LexEVS: application to TRANSFoRm. *Journal of the American Medical Informatics Association : JAMIA*. 2013;**0**:1-9.
- 6 Salvadores M, Alexander PR, Musen MA, Noya NF. BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. *Semantic Web*. 2013;**4**(3):277-84.
- 7 Whetzel PL. NCBO Technology: Powering semantically aware applications. *Journal of Biomedical Semantics*. 2013;**4**(Suppl 1):S8.
- 8 Nielsen J. Why you only need to test with 5 users. 2000.
- 9 Wilkinson I. Randomised cross-over trial in a multi-ethnic cohort: AIM HY – INFORM. Work Strand 3. 2016. [Available from: <http://www.aimhy.org.uk/our-research/ws3/>]