

Kaiser Permanente's Convergent Medical Terminology

Robert H. Dolin, John E. Mattison, Simon Cohn, Keith E. Campbell, Andrew M. Wiesenthal, Brad Hochhalter, Diane LaBerge, Rita Barsoum, James Shalaby, Alan Abilla, Robert J. Clements, Carol M. Correia, Diane Esteva, John M. Fedack, Bruce J. Goldberg, Sridhar Gopalarao, Eza Hafeza, Peter Hendler, Enrique Hernandez, Ron Kamangar, Rafique A. Khan, Georgina Kurtovich, Gerry Lazzareschi, Moon H. Lee, Tracy Lee, David Levy, Jonathan Y. Lukoff, Cyndie Lundberg, Michael P. Madden, Trongtu L. Ngo, Ben T. Nguyen, Nihilkumar P. Patel, Jim Resneck, David E. Ross, Kathleen M. Schwarz, Charles C. Selhorst, Aaron Snyder, Mohamed I. Umarji, Max Vilner, Roy Zer-Chen, Chris Zingo

Kaiser Permanente CMT Team, Kaiser Permanente HealthConnect

Abstract

This paper describes Kaiser Permanente's (KP) enterprise-wide medical terminology solution, referred to as our Convergent Medical Terminology (CMT). Initially developed to serve the needs of a regional electronic health record, CMT has evolved into a core KP asset, serving as the common terminology across all applications. CMT serves as the definitive source of concept definitions for the organization, provides a consistent structure and access method to all codes used by the organization, and is KP's language of interoperability, with cross-mappings to regional ancillary systems and administrative billing codes.

The core of CMT is comprised of SNOMED CT, laboratory LOINC, and First DataBank drug terminology. These are integrated into a single poly-hierarchically structured knowledge base. Cross map sets provide bi-directional translations between CMT and ancillary applications and administrative billing codes. Context sets provide subsets of CMT for use in specific contexts.

Our experience with CMT has lead us to conclude that a successful terminology solution requires that: (1) usability considerations are an organizational priority; (2) "interface" terminology is differentiated from "reference" terminology; (3) it be easy for clinicians to find the concepts they need; (4) the immediate value of coded data be apparent to clinician user; (5) there be a well defined approach to terminology extensions.

Over the past several years, there has been substantial progress made in the domain coverage and standardization of medical terminology. KP has learned to exploit that terminology in ways that are clinician-acceptable and that provide powerful options for data analysis and reporting.

Keywords:

Terminology, Semantics, Vocabulary, SNOMED, LOINC

Introduction

This paper describes Kaiser Permanente's enterprise-wide medical terminology solution, referred to as our Convergent Medical Terminology (CMT)¹. We'll provide an overview of CMT, the underlying technical components, lessons learned during the de-

velopment and deployment, and areas where we feel terminology standards need further evolution.

Kaiser Permanente (KP) is the largest not-for-profit health maintenance organization in the United States, serving 8.4 million members across the country. An integrated health delivery system, KP organizes and provides or coordinates members' care, including preventive care such as well-baby and prenatal care, immunizations, and screening diagnostics; hospital and medical services; and pharmacy services. Over 11,000 Permanente Medical Group physicians and tens of thousands of allied health professionals contribute to KP's care delivery.

The Kaiser Permanente HealthConnect program is a multi-year initiative focused on the national deployment of a highly sophisticated and integrated information management and delivery system. The goal of KP HealthConnect is to deliver on the KP Promise of quality, caring, convenient and affordable care for our members. KP HealthConnect integrates the clinical record with appointments, registration and billing to deliver improvements in care delivery and cost savings across the organization. Patients, physicians and other authorized health care staff have immediate secure access to complete, up-to-the-minute medical records, including test/lab results, prescriptions, and other ancillary data regardless of time or location. Immediate access to best medical practice recommendations within the proper contexts in the applications helps streamline patient care processes and improve health outcomes. KP HealthConnect is rolling out nationally to all KP facilities over the next several years.

CMT is evolving to become KP's central terminology resource, and it serves as a foundational component of KP HealthConnect. It originally evolved in two somewhat parallel tracks - as a collaborative project with the College of American Pathologists as a means of enabling KP's participation in the enhancement of SNOMED, and as a terminology database to serve the needs of the KP Colorado Clinical Information System. It has evolved into a core KP asset, serving as the common terminology across all applications. CMT serves as the definitive source of concept definitions and terminology cross-maps for all codes used within the organization, and it provides a consistent structure and access method to all codes used by the organization. CMT is KP's

"lingua franca" of interoperability: CMT concepts are mapped to lab, radiology, immunization, pharmacy, and EKG order and result codes within each KP region - enabling data comparison across regions and across applications; CMT concepts are mapped to HL7 vocabulary concepts - enabling exchange of standard HL7 messages; CMT concepts are mapped to administrative billing codes - enabling a semi-automated translation from clinician documentation.

Many pockets of local terminology continue to exist within the organization. As applications are integrated into KP HealthConnect, their local terminology is mapped and integrated into CMT.

CMT Objectives and Guiding Principles

A set of high-level overarching objectives and guiding principles has been developed by a national consensus process within the organization.

CMT Objectives

- Support the integration of "clinical content" into KP HealthConnect. (Clinical content is the form in which rules, guidelines, and evidence-based medicine are surfaced to clinicians within an application, and includes data entry templates and questionnaires; orderable items such as lab, pharmacy, and radiology; patient instructions; selection lists; synonyms and preference lists; health maintenance reminders; automated alerts such as for drug-drug and drug-allergy interactions.)
- Provide a common reference terminology for ancillary interfaces.
- Serve as the reference terminology against which KP will analyze operational data stores.
- Support the requirements for accurate and complete administrative coding and regulatory reporting.
- Analyze and develop processes (and if necessary, tools) which maintain enterprise consistency and synchronization of terminology as well as the integration of terminology in clinical content as the source terminologies are updated.

CMT Guiding Principles

- CMT is a KP enterprise asset, intended to serve the needs of the organization while remaining standards-compliant. KP can extend CMT where needed to meet business requirements, and will typically work to get our KP extensions formally incorporated into the standard terminologies underlying CMT.
- Evidence-based medicine is a primary driver of decision support, which is a primary driver for the use of consistent documentation with coded terminology. Clinical content and coded terminology should be created and used to support the integration of evidence-based medicine into KP HealthConnect.
- Where there is interregional consensus to use standard codes, the primary repository of those codes will be the CMT.

- The strategy for deploying CMT in KP will be a interregional consensus-driven process. The CMT national team will take the lead in preparing and refining the strategy, subject to iterative refinement.
- There is a natural and expected update cycle to each of the source terminologies underlying CMT. Updates to the CMT are translated to all products that rely on it (such as clinical content).
- There is a need to preserve usability while working towards rich and unambiguous data capture.
- Coded terminology should be used where it can serve as a hook to trigger decision support (such as in drug-drug and drug-allergy interaction checking).
- Areas where there is consistent documentation and where addition of CMT imposes minimal user impact should use CMT (e.g. immunization site). Areas where there is minimal impact on users with CMT, and there is potential benefit - should use CMT (e.g. antimicrobial susceptibilities, and much of the coded data that originates in ancillary systems).
- Areas where we should carefully consider the use of CMT, and weigh the benefits against the potential adverse usability impacts include:
 - Data entry is inconsistent, and the variability in the use of coded terminology and free text is so high that we can't demonstrate the ability to effectively query the data any better than we can query free text.
 - Where an application is in the process of enhancing the underlying data model, we may choose to time the CMT deployment with the model enhancement.
 - Where the volume of requests would be enormous, without clear business case (e.g. for all concepts in all narrative throughout KP HealthConnect).
 - Domains within a source terminology that are highly evolving and likely to undergo major refinement. We would like to minimize our backward compatibility / retired concept issues.

Technical overview of CMT

A high-level graphical depiction of CMT is shown in Figure 1. CMT is built upon industry standard terminologies. SNOMED CT², laboratory LOINC³, and First DataBank⁴ drug terminology form the core of CMT. Core terminologies are integrated into a single poly-hierarchically structured knowledge base. The core also includes concepts created by KP in response to user requests. The majority of these KP-created concepts are submitted to SNOMED or LOINC for formal inclusion in a subsequent release.

The core of CMT is referred to as a "knowledge base" rather than a "database" because concepts have logic-based definitions, imported from the source terminologies⁵. A classifier organizes the CMT concepts into a poly-hierarchy, based on their definitions. The act of classifying helps identify synonymous concepts, and maintains quality and consistency across the some 400,000 concepts. Logic-based definitions allow a computer to compare var-

ious representations and determine whether or not they mean the same thing⁶. For example, when looking for all patient that have had an operation on the pituitary gland, CMT enables one to retrieve those records where the provider entered codes for: "hypophysectomy", "brain excision" + "pituitary gland", "partial excision of pituitary gland by transfrontal approach", or "brain incision" + "pituitary posterior lobe".

All CMT concepts surfaced to the organization are drawn from CMT core or CMT extensions. CMT extensions are those concepts that have not been integrated into the core. Extensions potentially have an adverse impact on data analysis and they potentially introduce synonymy with core concepts, so their use is tightly restricted. Requirements that lead to the need for a formal extension mechanism included: the need to provide concepts that don't fit into the logical core structure; the need for a temporary holding area for concepts that are needed in an application but have not yet been added to the core; the need to accommodate highly localized concepts to support particular applications or deployments; the desire to let other groups within the organization control small code sets that don't overlap with concepts in core (such as HL7 message types); and the need to accommodate requests by various groups to have experimental concepts that they can test prior to formally requesting.

Updates to CMT are done in a distributed fashion^{7,8}. CMT modelers, working on copies of the knowledge-base, perform independent edits and additions during a modeling cycle. These edits are then consolidated into a central work-flow environment where conflicts are identified, QA checks are performed, and batch updates (such as updates from First DataBank or incorporation of a new release of SNOMED CT) are applied. A new cycle then begins with the distribution of a versioned knowledge-base.

Applications can directly access CMT via a provided interface and/or CMT can provide applications with cross map sets and context sets, both of which are patterned after the SNOMED CT model. Cross mapping represents perhaps the biggest cost of enterprise-wide terminology deployment. There is a large initial investment as interfaces are built. Cross map sets are used to store mappings between CMT concepts and other coding schemes (such as ICD9, CPT4, HCPCS, NANDA nursing diagnosis codes, NIC nursing intervention codes, etc). Because the relationship between two concepts can differ depending on the use case, it is possible that different cross map sets will contain the same source and target concept, but with a different relationship. Each cross map set has a defined use case and a steward with overall responsibility for the set. The mapping method and quality assurance process also varies between cross map sets. Order entry and results reporting mappings, which involve a bi-directional translation between CMT and the local codes used by the particular ancillary application, require the highest degree of scrutiny, as they can have major clinical implications if incorrect. These mappings require manual review by at least two domain experts who know both source and target coding schemes and the context where the mapping will be used.

Mappings to administrative codes (ICD9, CPT4, HCPCS) are likewise use case dependent, and also require significant quality assurance to ensure accuracy and to prevent fraudulent billing.

These mappings are qualified with one of the "mapping category" values shown in Table 1. All of these mappings are manually reviewed by at least two certified coders that understand the context in which the mappings will be used.

Context sets are CMT subsets used within a particular context. Contexts can include a particular drop-down list or vocabulary table in an application, a field in an HL7 message, or any other CMT subset needed within the organization. Every context set has a steward with overall responsibility for the set, and a defined use case. Context sets have proven to be quite helpful in tracking and maintaining CMT subsets, and provide a mechanism whereby discrepancies in context sets used by different applications can be readily identified. Within a particular context of use, a given concept can be given a different display name so that the concept is unambiguous and intuitive within that context. Tooling support enables stewards to manage their own context sets while ensuring that concepts are drawn from CMT.

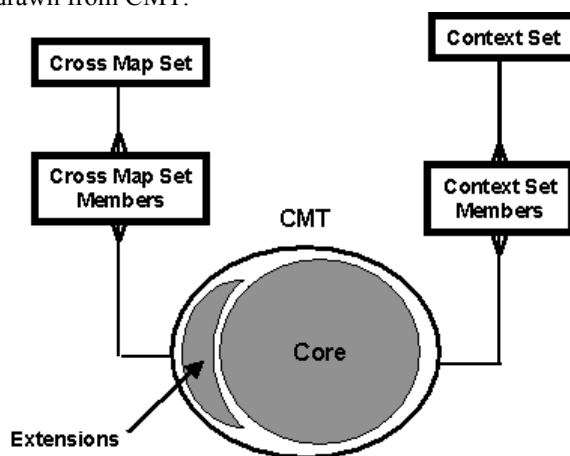


Figure 1 - High-level graphical depiction of Kaiser Permanente's Convergent Medical Terminology

Lessons Learned

Several factors have enabled us to more effectively deploy CMT and meet user expectations:

- **Establish usability as an organizational priority ::** CMT balances many often competing objectives. The need for usability and easy clinician documentation needs to be weighed against the need for data analysis and external reporting. We give higher weight to usability, with the thought that you first have to capture the data before you're able to use it⁹. One of our first steps was to define a set of objectives and guiding principles that clearly articulate the need to support usability, while balancing it against the needs of data analysis and external reporting.
- **Differentiate "interface" from "reference" terminology ::** A level of indirection between an application and CMT can more effectively enable users to capture coded data than if the users directly document by selecting concepts from the whole of CMT. This level of indirection

Table 1: Mapping categories used to qualify the relationship between a CMT concept and an administrative billing concept

Mapping Category	Description
Mapping Category 0 ("No Map")	There is no map between the CMT concept and a concept in the target terminology.
Mapping Category 1 ("Synonymous Map")	This is a 1:1 mapping between a CMT concept and the target concept. The mapped concepts are functionally synonymous, meaning that for the intended use cases, the concepts can be treated as synonyms.
Mapping Category 2 ("Narrow to Broad, CMT to target")	The CMT concept is more specific than the target concept. In a hierarchy, the CMT concept is a child of the target concept.
Mapping Category 3 ("Broad to Narrow, CMT to target")	The CMT concept is less specific than the target concept. In a hierarchy, the CMT concept is the parent of the target concept. Additional patient information, rules, or provider input are necessary to select an appropriate mapping.

is supported by the context set mechanism, and allows us to: (1) Identify and track the specific subsets of the terminology that are relevant in a particular context; (2) Alter concept display names and synonyms depending on the context; (3) Insulate users from changes (such as additions and inactivations) to the underlying source terminologies; and (4) Speed up access to allowable values, by providing smaller subsets of the entire CMT.

- **Make it easy for users to find the concepts they need** :: To facilitate usability, we: (1) Create search keywords; (2) Develop style guidelines for concept display names and abbreviations; (3) Work with user-groups on naming conventions and search strategies; (4) Use the context set mechanism described above to limit the number of concepts that appear in a particular context; (5) Tailor the search strategy to the application serving up the terminology.
- **Demonstrate immediate value of coded data to the user** :: Where we've successfully triggered decision support off coded data, we find the users more willing to use it. Clinicians interacting with KP HealthConnect capture coded data for allergies, diagnoses, and orders almost uniformly, and we often capture coded chief complaints. These codes are used to trigger drug-allergy alerts, data entry templates, and other forms of decision support that users value.
- **Have a well defined approach to terminology extensions** :: There is a trade-off between the speed at which user requests can be accommodated and the quality by which those requests are fulfilled. It can be tempting at times to use an extension mechanism to rapidly meet a large volume of requests, but once an application is functional, there is a tendency to give lower priority to harvesting extensions for incorporation back into the core terminology standards. A strategy for submitting extensions to and for maintaining synchrony with source terminologies can help balance the need for content with the benefits of standardization.

Future directions

CMT is based on industry standard coding schemes, and while we continue to find new ways to deploy CMT in various applications, a major component of CMT is working to enhance the underlying source terminologies upon which CMT is based. With respect to SNOMED CT, we see value in the further development of SNOMED CT qualifiers as a way of constraining allowable post-coordination. Standardizing various SNOMED CT subsets – e.g. for allergies or for HL7 fields - will foster broader interoperability than is currently possible with today's HL7 messages, where each implementation draws upon a potentially different set of codes or draws from a different coding scheme. Along these lines, it will be important to refine the SNOMED / HL7 Reference Information Model overlap and develop guidelines and/or a formal mechanism for identifying redundancy⁶. It will be important to further enhance the SNOMED CT term request process. Finally, there is a need to foster a broader understanding and demonstrations of the power of the description-logic underpinnings of SNOMED CT.

With respect to laboratory LOINC, we see value in applying additional scrutiny to the values used to populate the component axes (such as the set of measured components), providing unambiguous definitions and hierarchical relationships. A cross-map between component values and SNOMED CT would facilitate a complete integration of these terminologies¹⁰.

With respect to drug terminology, we see value in the continuing evolution of standards, such as the National Library of Medicine and United States Food and Drug Administration's RxNorm activities. Such a standard needs to be specified down to the clinically relevant level of granularity.

Over the past several years, there has been substantial progress made in the domain coverage and standardization of medical terminology. KP has learned to exploit that terminology in ways that are clinician-acceptable and that provide powerful options for data analysis and reporting.

Acknowledgements

The integration of CMT into KP HealthConnect represents the dedicated and conscientious efforts of a tremendous number of people within Kaiser Permanente. Our sincere gratitude goes out to all those people,

who turn the KP Promise into reality on a daily basis through their actions.

References

- [1] Dolin RH. Kaiser Permanente's Convergent Medical Terminology. Testimony to the National Committee on Vital and Health Statistics, Subcommittee on Standards and Security. May 22, 2003. [<http://nevhs.hhs.gov/030522sstr.htm>]
- [2] SNOMED Clinical Terms. College of American Pathologists. [<http://www.snomed.org/>]
- [3] Logical Observation Identifiers Names and Codes (LOINC). [<http://www.loinc.org/>]
- [4] First DataBank. [<http://www.firstdatabank.com/>]
- [5] Campbell KE, Das AK, Musen MA. A logical foundation for representation of clinical data. *JAMIA* 1994; 1(3): 218-32.
- [6] Dolin RH, Spackman KA, Markwell D. Selective retrieval of pre- and post-coordinated SNOMED concepts. *JAMIA Fall Symposium Supplement* 2002;210-4.
- [7] Levy DH, Dolin RH, Mattison JE, Spackman KA, Campbell KE. Computer-facilitated collaboration: Experiences building SNOMED RT. *JAMIA Fall Symposium Supplement* 1998: 870-4.
- [8] Campbell KE, Cohn SP, Chute CG, Shortliffe EH, Rennels G. Scalable Methodologies for Distributed Development of Logic-Based Convergent Medical Terminology. In C.G Chute, Ed., *IMIA Conference on Natural Language and Medical Concept Representation*, Jacksonville, FL, 243-256. 1997.
- [9] Krall MA, Chin H, Dworkin L, Gabriel K, Wong R. Improving clinician acceptance and use of computerized documentation of coded diagnosis. *Am J Managed Care* 1997;3(4):597-601.
- [10] Dolin RH, Huff SM, Rocha RA, Spackman KA, Campbell KE. Evaluation of a "Lexically Assign, Logically Refine" Strategy for Semi-automated Integration of Overlapping Terminologies. *JAMIA* 1998;5:203-213.

Address for correspondence

Robert H. Dolin, MD
Kaiser Permanente
5 Centerpointe Drive
La Palma, California 90623
Robert.H.Dolin@kp.org