MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI230929

Promoting Learning Health System Cycles by Optimizing EHR Data Clinical Concept Encoding Processes

Piper RANALLO[a], Bronwyn SOUTHWELL[b], Christopher TIGNANELLI[b], Steven G. JOHNSON [b], Richard KRUEGER[a], Tess SEVAREID-GROTH[a], Adam CARVEL[a] and Genevieve B. MELTON [a,b,1]

[a] *Fairview Health Services, Minneapolis, MN USA*

[b] *University of Minnesota, Minneapolis, MN USA*

ORCiD ID: Svensson-Ranallo https://orcid.org/0000-0003-2532-0804, Melton https://orcid.org/0000-0001-5193-1663

Abstract. Electronic health records (EHRs) and other real-world data (RWD) are critical to accelerating and scaling care improvement and transformation. To efficiently leverage it for secondary uses, EHR/RWD should be optimally managed and mapped to industry standard concepts (ISCs). Inherent challenges in concept encoding usually result in inefficient and costly workflows and resultant metadata representation structures outside the EHR. Using three related projects to map data to ISCs, we describe the development of standard, repeatable processes for precisely and unambiguously representing EHR data using appropriate ISCs within the EHR platform lifecycle and mappings specific to SNOMED-CT for Demographics, Specialty and Services. Mappings in these 3 areas resulted in ISC mappings of 779 data elements requiring 90 new concept requests to SNOMED-CT and 738 new ISCs mapped into the workflow within an accessible, enterprise-wide EHR resource with supporting processes.

Keywords. Semantic interoperability, terminology, OMOP, process improvement

1. Introduction

Research data sharing networks and other multi-institutional initiatives often result in pooled clinical electronic health record (EHR) and other real-world data (RWD), which have the potential to dramatically improve point of care clinical decision-making and secondary uses such as research and quality improvement [1-3]. EHR/RWD must precisely and accurately represent clinical information with industry-standard concepts (ISCs). EHR data transformation into standard codes is also often a prerequisite for clinical decision-support (CDS) tools [4]. As a result, there is a trend towards use of common data models (CDMs) such as the Observational Medical Outcomes Partnership (OMOP) [5], Informatics for Integrating Biology & the Bedside (i2b2) [6], and the National Patient-Centered Clinical Research Network (PCORnet) [7].

[1] Corresponding author: Genevieve B. Melton, Mayo Mail Code 450, 420 Delaware Street SE, Minneapolis, MN 55455, USA, email: gmelton@umn.edu.

The process of transforming data into a widely consumable CDM formats with ISCs as opposed to local proprietary codes, however, is difficult and laborious and often does not occur. Moreover, organizations undertaking this time-consuming transformation often perform this activity in such a way that the outputs become unavailable for re-use. Concept mapping often occurs outside of the EHR, persisting mappings in files, tables, or entirely external schemas and also often do not become part of the organization's centralized inventory of enterprise data assets. Consequently, the same mapping process may be performed multiple times, by multiple groups for each downstream use (e.g., CDS, analytics or reporting, research, data transmission). Also, concept mapping is often performed by technical and clinical experts without in-depth clinical terminology expertise and concept mapping best practices across teams with combined clinical, technical, and concept representation expertise.

To address these potential inefficiencies, we launched an initiative aimed to optimize data encoding to CDMs and ISCs. Our guiding principle was that EHR data required for all clinical, operational, and research use needed to not only be encoded using ISCs but that encoding must occur leveraging the EHR platform lifecycle for the benefit of other downstream processes in the context of three organizational informatics projects.

2. Methods

This initiative was launched at an integrated hospital system with ten hospitals and both academic and community sites in greater Minnesota and western Wisconsin, United States (Institutional Review Board approval STUDY00014481, Not Human Subjects). The methods used were developed in the context of the organization's simultaneous implementation of three key knowledge discovery and decision support solutions and grew iteratively out of routine work performed as part of these projects. One project was the implementation of the OMOP CDM, a research repository used to standardize the structure and content of EHR and other observational data [3], making it amenable to analysis using standard, open-source tools. For this project, we identified, extracted, and mapped EHR data and loaded it into the OMOP CDM. Two additional projects included: (1) the external validation of an artificial intelligence (AI) and machine learning (ML) risk model with OMOP mappings [8] and (2) development of an interoperative CDS tool using fast healthcare interoperability resources (FHIR) for evidence-based anti-coagulation in the setting of traumatic brain injury and its implementation as a multi-site pragmatic clinical trial [9]. All three projects required the same time-consuming identification of relevant source data and mappings to industry-standard concepts, which at baseline were not mapped or were not precisely enough encoded for our needs.

Current state processes were transformed to future state processes through the following processes (Figure 1) starting with understanding current state of: (1) current mapping process documentation; (2) identifying activities for identifying and mapping content and (3) inputs and outputs, as well as core resources for each activity. Then, two improvement steps which were: (4) standardize and then (5) optimize activities into processes for standard, repeatable, and maintainable steps. We also developed a multi-step process over the course of the three projects and tested the process with mappings, their maintenance, and interactions with Standard Development Organizations (SDOs).

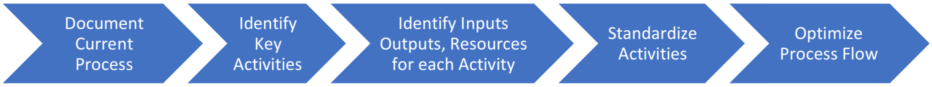

Figure 1: Stepwise development of optimized EHR data encoding processes.

For this case report, we provide our findings specific to SNOMED-specific mappings for Demographics, and Services (including surgical services), and Specialty. We explored processes to leverage native EHR functionality for maintaining mappings including functionality to store, view and extract, and maintain mappings for both interactively and programmatically (i.e., via bulk load) assigning concept codes to EHR components (e.g., flowsheets, note templates) and data elements. ISC mappings were characterized, including need for SDO submissions to SNOMED, current mappings, and resultant mappings to EHR workflow from these initiatives.

3. Results

Analysis of current state demonstrated key opportunities by standardizing activities. First, there was a significant amount of duplicative work across projects as each worked to identify the location of the same relevant EHR clinical content and mappings. This was likely to recur in future projects, as project mappings were typically not fed back into the source system or common resource as part of new EHR build, but instead were siloed in project-specific data mappings. To address this, as concepts were mapped, each was stored in a shared repository (here with native EHR tools, where possible).

We observed lags or process gaps when information required to identify or map clinical content was needed from subject matter experts (SMEs) external to the project, and the time at which these SMEs were able to provide the needed information. In addition to these delays or process defects, this also created disruptions in workflow for both concept mappers and SMEs. To address this, we worked to identify and build SMEs into our project teams upfront when mapping existing build, and – for new build - to make data encoding an integral part of the EHR build process itself, where possible. There were also inefficiencies in selecting which concepts and sometimes which standard terminologies were optimal for representing certain clinical data elements. On smaller projects we previously often relied on individuals without specialized knowledge in semantic interoperability and best practices for concept representation. The process of selecting the correct concept(s) or distinguishing between nuanced concepts was often inefficient and sometimes created unreliable mappings, resulting in re-work. This informed our decision to increase terminology and interoperability expertise, ensuring at least one resource with deep terminology expertise was included on all projects.

For encoding EHR content, a five-step process was developed in our process optimization (Figure 2). First, the source concept is determined and - if the content has already been encoded (natively within the EHR) - verified. In some cases, the content is encoded using non precise or inaccurate concepts (e.g., non-specific CPT billing codes or ICD10CM codes). In others, encoding is inaccurate or outdated (i.e., inactivated concept replaced by a new concept in the source terminology). In some cases, the precise concept does not exist in the terminology. In these cases, a new concept is defined and submitted in the format required by the SDO. If the concept exists in the terminology, but appears to be defined incorrectly or incompletely, errors or missing features of the definition are identified and a request to add or modify the concept is submitted to the appropriate SDO. Finally, concept codes are added, where possible, directly into the EHR using the tools provided by the vendor to make them available for clinical and secondary (including future) uses of the data.

Figure 2: Standard process for encoding clinical information in EHR for *existing* build.

Overall, for 779 data elements, we submitted 90 requests to SNOMED for gaps in concepts or inaccuracies in concept definitions for Demographics, Specialty, and Services (Table 1). All requests were accepted and completed (i.e., available in the terminology) within 7 months (range, 4 to 7 months). Requests to correct the concept definition were submitted to SNOMED using their content request system.

Table 1. EHR Information Elements Identified and Mapped.

Clinical Domain	Relevant Data Elements Identified	Previously in Enterprise Resource	Concepts into Enterprise Resource	SNOMED Concepts Requested
Demographics	395	9	355	24
Religion	77	0	72	15
Language	232	0	201	1
Marital Status	10	0	8	0
Race	11	6	10	2
Ethnicity	65	3	64	6
Specialty – Healthcare Entities	310	0	310	15
Services - Healthcare Entities	74	0	74	51
Overall	779	9	739	90

An example is a concept for a procedure that was incorrectly defined as a type of nerve destruction procedure, when it is a procedure that simply blocks the transmission of a signal from the nerve). Similarly, we discovered that SNOMED had no concepts for robotic-assisted surgical procedure services. Our terminologists submitted requests to SNOMED with the publicly available US Content Request System (US-CRS), making the concepts available to others (Figure 3).

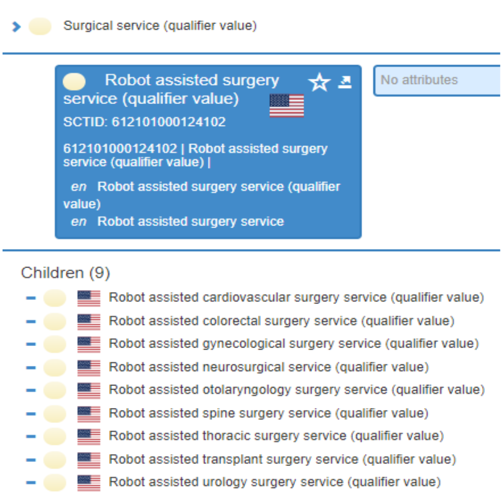

Figure 3: New surgical service lines added to surgical service sub-hierarchy in SNOMED.

4. Discussion

While many organizations map EHR content to ISCs for CDS, exchanging data, analytics, and populating research repositories, this work is often performed on a project-by-project basis with inherent inefficiencies. By using a systems-based approach, our work was able to benefit both point of care functionality (bedside) and analytics, with the end result of accelerating future mapping work and the process of translating data. We also observed the importance of educating those in key roles about the importance of centralizing standard representations and terminologies and developing terminology expertise in clinical, informatics, and IT professionals in the organization.

Our core learnings focus on people, process, and technology. While EHR platforms vary in the robustness of features for encoding ISCs, virtually all provide mechanisms for performing these tasks for EHR content, including base application and reporting and analytics which are invaluable for optimizing concept mapping processes. We also observed the importance of working closely with SDOs and the essential role of applying RWD to improve the quality of terminologies in terms of completeness and precision. The accuracy of terminology content similarly depends on those using the terminology having an adequate understanding of how the terminology defines concepts, and validating definitions (versus selecting 'closest' available concept). Finally, by closing the loop using ISC mappings to encode data natively with EHR tools, resultant processes allow encoding to be performed once and easily reused. This is arguably the most valuable activity we found in the overall process as it not only prevents duplicative work but also ensures that resources invested in concept mapping and terminology development will benefit future initiatives.

We also noted some limitations with our approach. These processes often do introduce some inefficiencies for projects, including the need for maintenance of new mappings performed, possibly reloading terminologies more often (as codes are submitted) and the effort for consensus with enterprise stakeholders for mappings. Future work includes a more in-depth evaluation of the costs and benefits of comprehensive encoding of relevant EHR content with this approach for IT and informatics teams and the effect of these approaches on AI/ML methods across organizations.

5. Conclusions

To address inefficiencies in current approaches of data identification and mapping, we launched an initiative to develop a standard, repeatable process for precisely, accurately, and unambiguously encoding relevant EHR data using appropriate industry standard concepts and codes. This approach leveraged EHR functionality, the SDO process and a system approach to encoding these concepts consistently in the data lifecycle.

Acknowledgements

This work was supported by AHRQ R18HS028583, University of Minnesota Center for Learning Health System Sciences - a partnership between the University of Minnesota Medical School and the School of Public Health, and Fairview Health Services.

References

[1] Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. Nat Rev Clin Oncol. 2019 May;16(5):312-25, doi: 10.1038/s41571-019-0167-7.
[2] Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA. 2018 Sep;320(9):867-8, doi: 10.1001/jama.2018.10136.
[3] Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM. Real-world evidence - what is it and what can it tell us? N Engl J Med. 2016 Dec;375(23):2293-7, doi: 10.1056/NEJMsb1609216.
[4] Delvaux N, Vaes B, Aertgeerts B, Van de Velde S, Vander Stichele R, Nyberg P, Vermandere M. Coding systems for clinical decision support: theoretical and real-world comparative analysis. JMIR Form Res. 2020 Oct;4(10):e16094, doi: 10.2196/16094.
[5] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB. Observational health data sciences and informatics (ohdsi): opportunities for observational researchers. Stud Health Technol Inform. 2015;216:574-8, doi: 10.3233/978-1-61499-564-7-574.
[6] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc. 2010 Mar-Apr;17(2):124-30, doi: 10.1136/jamia.2009.000893.
[7] Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. J Am Med Inform Assoc. 2014 Jul-Aug;21(4):578-82, doi: 10.1136/amiajnl-2014-002747.
[8] Bihorac A, Ozrazgat-Baslanti T, Ebadi A, Motaei A, Madkour M, Pardalos PM, Lipori G, Hogan WR, Efron PA, Moore F, Moldawer LL, Wang DZ, Hobson CE, Rashidi P, Li X, Momcilovic P. MySurgeryRisk: development and validation of a machine-learning risk algorithm for major complications and death after surgery. Ann Surg. 2019 Apr;269(4):652-62, doi: 10.1097/SLA.0000000000002706.
[9] Available from: https://digital.ahrq.gov/ahrq-funded-projects/evaluation-scaled-scaling-acceptable-cds-approach-implementation-interoperable-cds-venous.