APAMI2020 Poster Presentation Sessions | APAMI 2020 | Poster Presentation Sessions
Standardization
Sun. Nov 22, 2020 2:20 PM - 2:50 PM Room E-2 (Congress center 5F - Conference Room 53)

[AP2-E2-3-03] Incorporation of the Korean EDI Vocabulary into the OMOP Vocabulary

*Yeonchan Seong[1,2], Seng Chan You[1], Yeunsook Rho[3], Jimyung Park[4], Jaehyeong Cho[4], Jiwoo Kim[3], Yun-jung Heo[3], Rae Woong Park[1,4] (1. Department of Biomedical Informatics, Ajou University School of Medicine, Korea, 2. Department of Sociology, Yonsei University, Korea, 3. Health Insurance Review &Assessment Service, Korea, 4. Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Korea)
Keywords: Vocabulary, Controlled, National Health Programs

The Electronic Data Interchange (EDI) codes, required for submission of claim to the Korean national insurance system, represent most of approved drug, medical services, and medical devices in South Korea. Despite of its widespread use, the EDI vocabulary has never been regarded or examined as standard medical vocabulary. Here, we audit the EDI vocabulary as a controlled vocabulary. Furthermore, we incorporate the EDI vocabulary into Observational Medical Outcomes and Partnership (OMOP) vocabulary to enhance consistency and maintenance. The proposed process consisted of three steps: Enhancing Maintenance of Vocabulary, Classification of domains and Building hierarchy and Translation. A total of 313,457 concepts were uploaded from the EDI to the OMOP vocabulary. We made an R package public, which automates overall process. The EDI as OMOP vocabulary has improved in quality compared to the current EDI vocabulary.

Incorporation of the Korean EDI Vocabulary into the OMOP Vocabulary

Yeonchan Seong[a, b*], Seng Chan You[a*], Yeunsook Rho[c], Jimyung Park[d], Jaehyeong Cho[d], Jiwoo Kim[c], Yun-jung Heo[c] and Rae Woong Park[a, d]

[a]*Department of Biomedical Informatics, Ajou University School of Medicine, Korea*
[b]*Department of Sociology, Yonsei University, Korea*
[c]*Health Insurance Review & Assessment Service, Korea*
[d]*Deparment of Biomedical Sciences, Ajou University Graduate School of Medicine, Korea*
**First two authors contributed equally.*

Abstract

Despite of its widespread use, the Electronic Data Interchange (EDI) vocabulary has never been regarded or examined as standard medical vocabulary. Here, we audit the EDI vocabulary as a controlled vocabulary. Furthermore, we incorporate the EDI vocabulary into Observational Medical Outcomes and Partnership (OMOP) vocabulary to enhance consistency and maintenance. A total of 313,457 concepts were uploaded from the EDI to the OMOP vocabulary. We made an R package public, which automates overall process. The EDI as OMOP vocabulary has improved in quality compared to the current EDI vocabulary.

Keywords:

Vocabulary, Controlled, National Health Programs

Introduction

The need for standard and controlled vocabularies in the national healthcare system has been widely recognized in South Korea. However, pervious attempts, such as KOSTOM, KCD-7, have not been widely accepted. Despite of its widespread use, the Electronic Data Interchange (EDI) vocabulary has never been regarded or examined as standard medical vocabulary. Here, we aimed to incorporate the Korean EDI vocabulary into the Observational Medical Outcomes Partnership (OMOP) vocabulary which is established by Observational Health Data Sciences and Informatics (OHDSI).

Methods

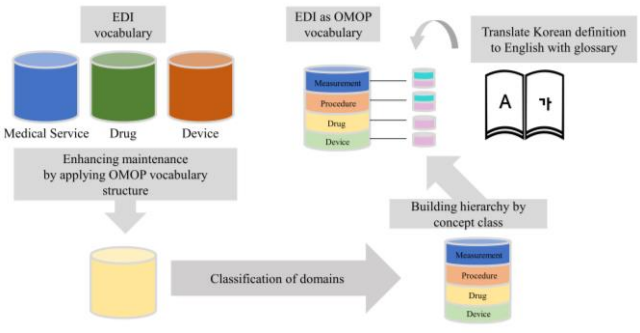

Figure 1- Implementation of transformation process for EDI vocabulary

We built the semi-automated process to incorporate EDI vocabulary into OMOP vocabulary including code cleaning, classification, building hierarchy, and insertion of the vocabulary to the OMOP CDM version 5.3.1 database (Figure 1). We deployed the open-source click-to-run R software, EdiToOmop, on the OHDSI's official github repository [2].

Enhancing Maintenance of Vocabulary

The EDI vocabulary is transformed to have the attributes of the OMOP vocabulary. Once registered as OMOP vocabulary, a permanent, unique, and non-semantic numeric identifier of OMOP vocabulary is assigned to each of EDI vocabularies. In addition, version control is provided as the EDI vocabulary is updated. Lastly, we added further description to the concept definitions for explaining the modifier code of the original EDI code, such as 'emergency use'.

Classification of domains and Building hierarchy

The concepts in the medical service of the EDI vocabulary are further classified into the 'procedure' and 'measurement' domain in the OMOP vocabulary. Similar to ICD 9/10 code system, the hierarchy of EDI codes is created in such a way that the first 5 digits represent the ancestor term of longer, descent EDI codes, and the rest digits represent the same service for reimbursement.

Translation

For 273,449 EDI concepts that need to be translated, we performed 3-step translation work. First, we used 'Google.Cloud.Translation.V3', a.NET client library in the Google Cloud Translation API, for initial translation. Then we retranslated 272,559 terms that are not translated into proper English words using a glossary as constraint for frequent Korean words. Finally, doublechecking and revising by a human reviewer was performed, and if necessary manual revision was followed by a final third-party reviewer.

Auditing vocabulary

The evidence that our restructuring process of the EDI vocabulary improves the quality of data can be proved by the qualitative criteria of the health terminology system. Based on previous research on terminology evaluation, we defined the evaluation criteria for comparing EDI vocabulary and EDI as OMOP [3-6].

Result

A total of 313,431 concepts were incorporated and uploaded to the OMOP, which are published in OHDSI's public and official

vocabulary website, Athena [7]. OHDSI vocabulary team assigned the unique OMOP vocabulary identifier to all of these updated EDI concepts in February 2020.

On all criteria except compositionality, the EDI in OMOP vocabulary showed better quality index than the original ED vocabulary as shown in Table1.

Table 1– The comparison of EDI vocabulary and EDI in OMOP Vocabulary based on the terminology *evaluation*

Criteria	Explanation	EDI vocabulary	EDI in OMOP Vocabulary
Concept orientation	A concept must be linked with only one term.	△	O
Concept permanence	Even if the used term is updated, the previously used term should not be deleted.	O	O
Coverage	The domain covered by the terminology system must be consistent and obvious.	O	O
Relation	The relation of each concept should be defined.	X	O
Multiple hierarchy	A concept can have multiple hierarchies	X	△
Compositionality	Terms can be separated into atomic units and has compositional extensibility	X	X
Non-semantic concept identifiers	There must be unique code representing a concept.	X	O
Version control	When terminology is updated, version information including changes must be specified	X	O
Formal definitions	Having a structure and definition that can be indexed and processed by computer	X	O
Synonyms uniquely identified and mapped to relevant concept	Synonyms, including abbreviations, are managed by unique identifiers, and related concepts are mapped	X	O
Multi-language	The terminology system supports multiple languages	△	O

Discussion

By August 2020, a total of 447 EDI IDs' synonyms were changed, 10,320 were newly added and 7,873 were deleted. However, only the EDI vocabulary until October 2019 was updated until now. To update the newly added EDI concepts, Korean definitions should be translated into English with an aid of human translator again. Furthermore, the official mapping between the Korean EDI vocabulary and global standard terminology system, such as SNOMED-CT, remains a future work.

References

[1] E. J. Hwang, H. A. Park, S. K. Sohn, H. B. Lee, H. K. Choi, S. Ha, H. J. Kim, T. W. Kim, W. Youm, Mapping Korean EDI Medical Procedure Code to SNOMED CT. *Studies in health technology and informatics*. 2019; 264:178-82.

[2] EdiToOmop Repository; [accessed 2020 June 29]. Available from: *https://github.com/OHDSI/EdiToOmop*

[3] J.J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century. *Methods of information in medicine*. 1998; 37:394-403. doi: 10.1055/s-0038-1634558.

[4] C. G. Chute, S. P. Cohn, J. R. Campbell. A framework for comprehensive health terminology systems in the United States: development guidelines, criteria for selection, and public policy implications. *Journal of the American Medical Informatics Association. 1998;* 5(6):503-10. DOI: 10.1136/jamia.1998.0050503.

[5] S. T. Rosenbloom, R. A. Miller, K. B. Johnson, P. L. Elkin, S. H. Brown. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. *Journal of the American Medical Informatics Association*. 2006; 13(3):277-88. DOI: 10.1197/jamia.m1957

[6] K. C. Lee. A study on Enhancing Terminology & its Major Issues (2012). Doctoral dissertation, Dept. of Hospital Administration, Yonsei University. ATHENA-OHDSI VOCABULARIES Repository; [accessed 2020 June 29]. *http://athena.ohdsi.org*

Acknowledgements

We greatly thank the OHDSI vocabulary subgroup including Christian Reich (IQVIA), Anna Ostropolets (Columbia University), and Dmitry Dymshyts (Odysseus) for supporting this project and release of EDI vocabulary in the ATHENA.

Address for co-correspondence

Yun Jung Heo, PhD
Health Insurance Review & Assessment Service, Korea
E-mail: yunjungheo@naver.com

Rae Woong Park, MD, PhD
Department of Biomedical Informatics
Ajou University School of Medicine, Korea
E-mail: veritas@ajou.ac.kr