CDISC Transformer: a metadata-based transformation tool for clinical trial and research data into CDISC standards

Yu Rang Park[1], Hye Hyeon Kim[1], Hwa Jeong Seo[2] and Ju Han Kim[1]
[1]Seoul National University Biomedical Informatics (SNUBI), Interdisciplinary Program of Medical Informatics and Systems Biomedical Informatics Research Center, Div. of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea
[e-mail: {kirarang, hyehyeon2, juhan}@snu.ac.kr]
[2]Medical Informatics, Graduate School of Public Health, Gachon University of Medicine and Science, Incheon 405760, Korea
[e-mail: hjseo@gachon.ac.kr]
*Corresponding author: Ju Han Kim

Received March 31, 2011; revised July 18, 2011; accepted July 19, 2011; published October 31, 2011

Abstract

CDISC (Clinical Data Interchanging Standards Consortium) standards are to support the acquisition, exchange, submission and archival of clinical trial and research data. SDTM (Study Data Tabulation Model) for Case Report Forms (CRFs) was recommended for U.S. Food and Drug Administration (FDA) regulatory submissions since 2004. Although the SDTM Implementation Guide gives a standardized and predefined collection of submission metadata 'domains' containing extensive variable collections, transforming CRFs to SDTM files for FDA submission is still a very hard and time-consuming task. For addressing this issue, we developed metadata based SDTM mapping rules. Using these mapping rules, we also developed a semi-automatic tool, named CDISC Transformer, for transforming clinical trial data to CDISC standard compliant data. The performance of CDISC Transformer with or without MDR support was evaluated using CDISC blank CRF as the 'gold standard'. Both MDR and user inquiry-supported transformation substantially improved the accuracy of our transformation rules. CDISC Transformer will greatly reduce the workloads and enhance standardized data entry and integration for clinical trial and research in various healthcare domains.

Keywords: CDISC, SDTM, metadata registry, clinical trial, ISO/IEC 11179, rule base

The part of this paper was presented in the ICONI (International Conference on Internet) 2010, December 16-20, 2010, Philippines. This research was supported by a grant (09182KFDA889-3202) from Korea Food and Drug Administration in 2011. Y.R.P. was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education Science and Technology (2010-0028631).

DOI: 10.3837/tiis.2011.10.009

1. Introduction

Clinical trial is a planned experiment designed to assess the safety and efficacy of health interventions (e.g., drug, diagnostics, devices and therapy protocols) in man [1]. To obtain Food and Drug Administration (FDA) approval on health intervention, entire clinical trial data should be submitted to FDA. The size, quantity and complexity of clinical trial results, however, are enormous; it is a one of the major obstacles in reviews of clinical trial result. Thus, FDA has recommended utilizing Clinical Data Interchange Standards Consortium (CDISC) standard models as regulatory submission since 2004. Following these trends, major Clinical Trial Management Systems (CTMSs) support the CDISC standards on clinical trial data exporting and importing functions [2]. Most of these systems, however, are proprietary and there is no common rule for converting clinical trial data to CDISC standard-compliant data. Without automation, they require labor intensive or time-consuming processes such as re-entering data and manually mapping the data to the standard models [3][4].

According to the purpose of how clinical trial data is used, CDISC provides various standard models. Among the CDISC standards models, the Study Data Tabulation Model (SDTM) is an integral component for submitting clinical trial data to FDA [5][6]; also used in developing the Biomedical Research Integrated Domain Group (BRIDG) model that provides a semantic foundation of clinical research [7]. The Clinical Data Acquisition Standards Harmonization (CDASH), a set of contents standard in clinical trial, is based on the SDTM model as well [8]. Thus, for the purpose of illustration, the present paper focuses on SDTM-related transformation applications.

Case Report Form (CRF) is the primary instrument for data collection in clinical trials [9]. CRF has a free document format, consisting of one or more sections, and each section consists of a group of questions that are logically related. Questions are the atomic elements of CRFs.

The basic difficulty in transforming CRF observations to CDISC SDTM tables comes from the different data structures. One has to first identify the correct domain for each question of a CRF before variable mapping and value transforming processes. SDTM ver. 1.2 has 32 domains. The classification of questions into domains may be ambiguous and have to be determined by contextual information of each CRF.

To represent the semantics of CRF questions, we used the International Standard Organization/International Electro-technical Commission (ISO/IEC) 11179 Metadata Registry (MDR) standard model for metadata description and registration [10]. The ISO/IEC 11179 is a standard for representing metadata (semantic and syntactic) of data. National Cancer Institute (NCI) cancer Biomedical Informatics Grid (caBIG) employ for representing CRF and conducting clinical research and trial [11]. It provides two types of standard model; object model and basic attributes by its usage. The object model is used commonly for developing biomedical MDRs. The Agency for Healthcare Research and Quality (AHRQ), United States Health Information Knowledgebase (USHIK) [12] and the NCI cancer Data Standards Registry and Repository (caDSR) [13] implement the object model as public MDR. The other model, the basic attributes are used as descriptors of variables. The Tissue Microarray (TMA) data exchange model, named TMA Data Exchange Specification (DES) is one of the examples based on the basic attributes [14]. In the present study, we selected the ISO/IEC 11179 basic attributes for describing metadata of questions in CRFs for clinical trials and research.

We propose a novel method to convert clinical trial data to CDISC standard-compliant data semi-automatically by using ISO/IEC 11179 metadata standard. To achieve this, we developed SDTM mapping rules, which are consisted of SDTM attributes to ISO/IEC 11179 basic attributes. We also developed CDISC transformer using SDTM mapping rules to construct CDISC standard-compliant data without labor-intensive works.

2. Method

The ISO/IEC 11179 basic attributes consist of 45 attributes for describing metadata. We customized the basic attributes for the purpose of SDTM transformation. We eliminated among the 45 attributes non-essential variables such as optional and provenance-related ones (i.e., '*origins*', '*ownership'*, and '*submitted organization*' of metadata) and added two attributes to describe CRF structure (i.e., '*CRF name*' and '*section name*'), four attributes to describe concepts ('*Concept ID*', '*Concept name*' , '*Concept category*' and '*Concept subcategory*'). Finally, we constructed 17 attributes: 1) *Question label,* 2) *CRF name,* 3) *Section name,* 4) *Question ID,* 5) *Object,* 6) *Property,* 7) *Value domain type (Enumerated/Non-Enumerated),* 8) *Maximum length,* 9) *Minimum length,* 10) *Data type,* 11) *Unit of measure,* 12) *Permissible value,* 13) *Permissible value ID,* 14) *Concept ID,* 15) *Concept name,* 16) *Concept category,* 17) *Concept subcategory*. Among the 17 attributes, *Question label*, *CRF name,* and *Section name* are trivial ones that can easily be obtained without semantic or syntactic analysis, whereas the remaining 14 attributes cannot be.

The SDTM is built around the concept of observations collected from subjects who are participated in a clinical trial. Each observation can be described by a series of variables of the corresponding domain table as a row filled with values. A domain is defined as a collection of observation logically related with a common topic. In version 1.2, the SDTM consists of 32 domains and related variables. According to this SDTM structure, CDISC Transformer has two sub-processes: domain finding and variable mapping.

Each domain in SDTM covers different context of clinical trial data. For appropriate domain finding for a question, the correct semantics concepts of domains are required as well as question labels. We used the National Library of Medicine (NLM) MetaMap, which is a program to map biomedical text to the Unified Medical Language System (UMLS) Metathesaurus or, equivalently, to discover Metathesaurus concepts referred to in text [15].

CDISC Transformer's domain finding process is as follows; first, we map the domain concepts using MMTx (MetaMap Transfer) based on the domain names and their accompanying definitions defined by the SDTM specification. Then, five metadata attributes (i.e., *Question label, CRF name, Section name, Object, and Property*) capable of semantically representing CRF questions are selected among the 17 attributes. For each question, MMTx returns concepts as outputs by means of using the five metadata attributes as inputs. Subsumption test between question concept and domain concept determines candidate domain if a question concept has subsumption relationship with a specific domain concept. After the subsumption test, we obtain the rank orders of the candidate domains for each question. The pseudocode of the domain finding process is shown in **Fig. 1**. Each SDTM domain consists of specific variables, which are categorized into five types; *identifier*, *topic*, *timing*, *qualifier*, and *rule*. Variables of *identifier*, *timing* and *rule* types can be obtained during the study and the value of them are not to be modified by a user, whereas variables of *topic* and *qualifier* types are to be described by a user. These two types of variables are subject to the mapping process. For the *topic* and *qualifier* variables, for each SDTM domain, we manually constructed mapping rules between the previously selected 17 MDR metadata attributes and SDTM

variables as shown in **Fig. 2**.

Input 1: a list of SDTM Domain
Input 2: metadata file which is consisted of 17 metadata attributes for each question from a CRF
Output: a sorted list of candidate SDTM domain for each question
Procedure:

```
For each SDTM domain
   Map domain names and definitions to concepts using MMTx
   Add mapped concepts to the SDTM domain
End
For each question
   Select five metadata attributes
   For each metadata attribute
      Map metadata to concepts using MMTx
      For each mapped metadata concept
         Perform subsumption test between metadata concepts and domain concepts
         If subsumption test returns true
            Set candidate domain flag
         End
      End
    End
   End
End
Return rank sorted candidate domain and question
```

Fig. 1. Pseudocode for finding SDTM domain

Some variables that are in Trial Design Domains (i.e., Trial Arms; TA, Trial Elements; TE, Trial Visits; TV, and Trial Summary; TS) are to be defined by the research investigator in the trial designing phase such that they do not appear in a CRF. CDISC Transformer disregards these 41 variables are 'about' CRFs because they are not available in CRFs. Finally, 307 SDTM variables are subject to CDISC Transformer mapping process.

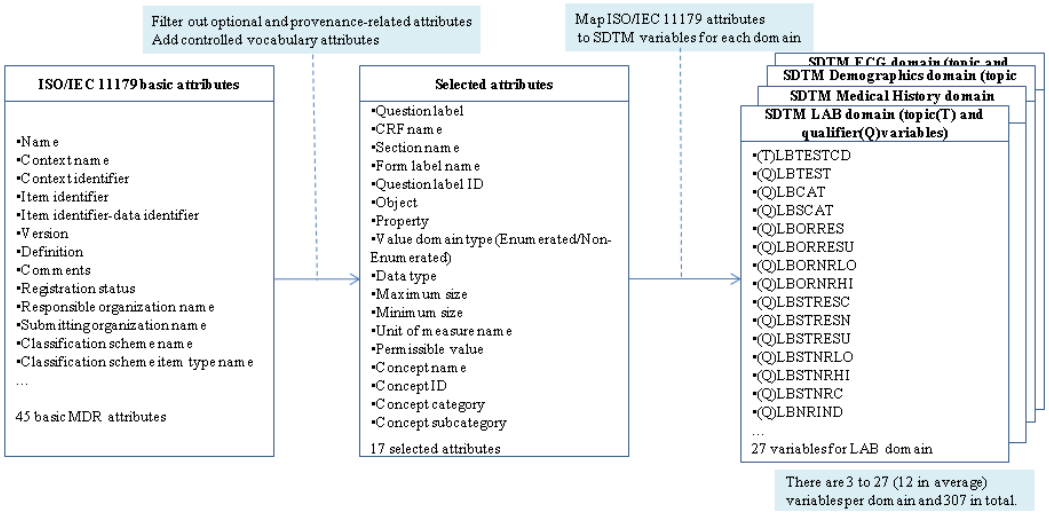

Fig. 2. CDISC Transformer mapping rule creation process. For each SDTM domain, mapping rules

are created by matching the 17 selected and modified attributes from the ISO/IEC 11179 MDR standard, returning 307 variable matching rules. Rules are enhanced by additional rules for better matching.

3. Result

3.1 SDTM mapping rule

We found that 211 (69%) among the 307 target variables perfectly matched with the 17 attributes selected from ISO/IEC 11179 MDR basic attributes. One or more SDTM variables are mapped to one of the 17 attributes, resulting 211 mapping rules. Top five domains showing best mapping are Exposure; EX, Inclusion/Exclusion Criterion Not Met; IE, Substance Use; SU, Subject Characteristics; SC, and Concomitant Medications; CM. Among the 17 metadata attributes, *Question label*, *Result record*, and *Unit of measure* are the most frequently utilized ones for the mapping. Among the 307 target variables, dependent (--REASND), indicator (--PRESP), descriptive and parametric variables are hard to map with 17 metadata attributes.

Table 1 shows a part of the mapping rule base for Inclusion/Exclusion Criterion Not Met (IE) and Laboratory Test Result (LB) domains of SDTM. Rows in Table 1 have four columns including source attributes, target domain and variable, and additional rule. 'Source attributes' column has 18 variables including the 17 metadata attributes (see method) and 'Result Record' from the result file from a clinical trial. Each row represent basic mapping rule with the optional 'Additional Rule', improving mapping accuracy. For instance, *Question label* for LB domain is mapped both to LBTEST and LBLOINC. But the additional rule for LBLOINC requires to apply the specified controlled vocabulary, Logical Observation Identifiers Names and Codes (LONIC) to map LBLOINC. The blank spaces in the first column of **Table 1** means that no attribute is available in source file to establish a mapping rule for LABTOX and LABTOXGR.

Table 1. Part of SDTM mapping rule base for IE and LB domains

Source Attributes	**Target SDTM Variables**		**Additional Rule**
MDR vs. Trial Result Data	**Domain**	**Variable**	
Concept name	IE	IETESTCD	
Question label	IE	IETEST	
Concept category	IE	IECAT	
Concept subcategory	IE	IESCAT	
Result Record* (Permissible, Question ID)	IE	IEORRES	Refer to SDTM terminology (NY)
Result Record* (Permissible, Question ID)	IE	IESTRESC	Refer to SDTM terminology (NY)
Concept name	LB	LBTESTCD	
Question label	LB	LBTEST	
Concept category	LB	LBCAT	
Concept subcategory	LB	LBSCAT	
Result Record*	LB	LBORRES	
Unit of Measure	LB	LBORRESU	Refer to SDTM terminology (UNIT)
Minimum length	LB	LBORNRLO	
Maximum length	LB	LBORNRHI	
	LB	LBNRIND	
Question label, Result Record* (Question ID)	LB	LBSTAT	If --TESTCD != NULL and --ORRES == NULL, answer NOT

			DONE otherwise NULL, Refer to SDTM terminology (ND)
	LB	LBREASND	
Result Record* (Question ID)	LB	LBNAM	
Question label	LB	LBLONIC	Need controlled vocabulary, LOINC with original Question label
-	LB	LBTOX	
-	LB	LBTOXGR	

*Attributes are from clinical trial result file. Other source attributes are all from the metadata attributes.

3.2 CDISC transformer

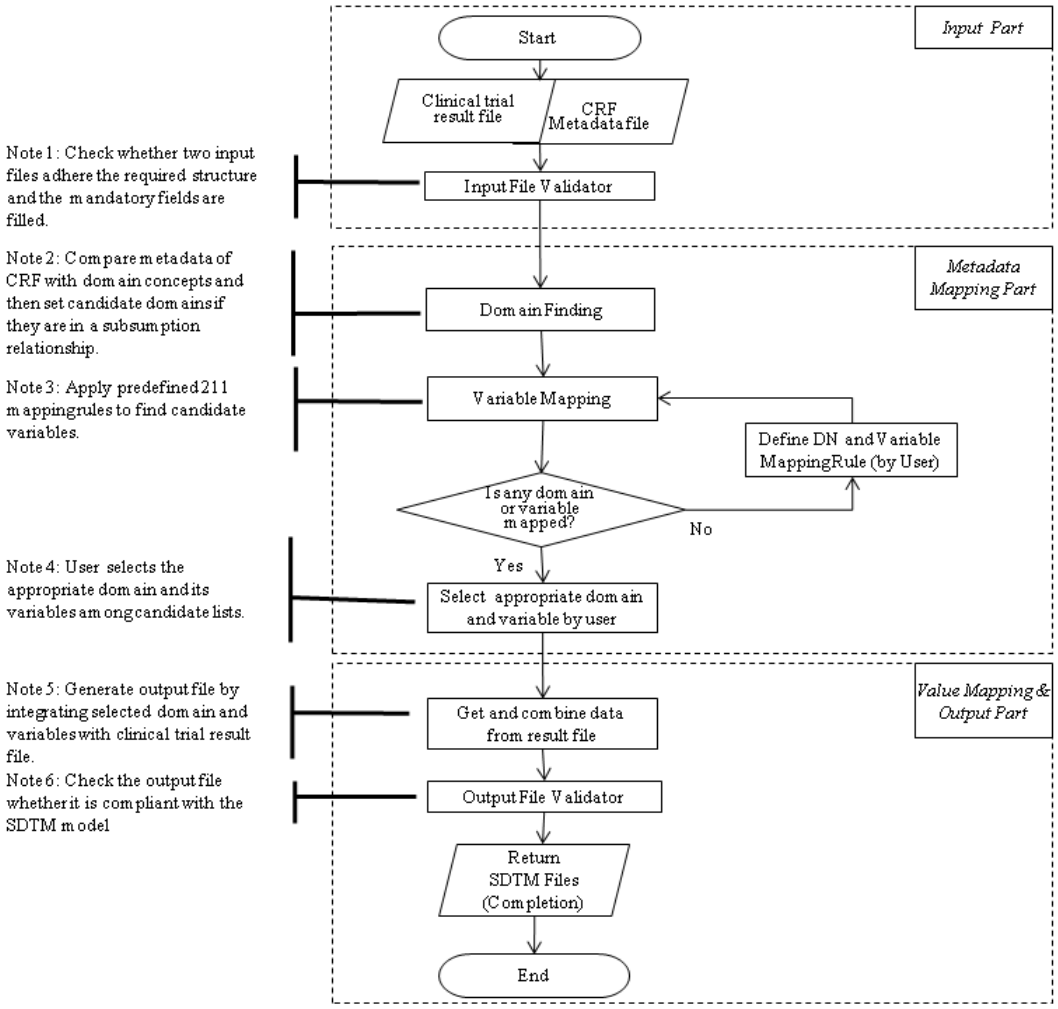

Fig. 3. CDISC Transformer's workflow diagram.

CDISC Transformer consists of a set of processes including the domain finding process, CRF question to SDTM mapping rule, interactive user inquiry interface, and integrated algorithms for the semi-automated creation of CDISC SDTM tables from user input CRFs accompanied by the observed results of a clinical trial and research. CDISC Transformer consists of input file validator, metadata mapping, value mapping, and output file validator **(Fig. 3)**. Input files to CDISC transformer are clinical trial result file and the metadata for the corresponding CRFs

and questions. These two input files have tabulation formats like SDTM. One row of a result file consists of identifier variables (i.e., subject, site, etc.), timing variables (i.e., examine time, starting time, etc.), question ID and result of specific question. Metadata file consist of previously defined 17 metadata attributes.

Input file validator checks the attributes and values of input files, returning warnings if a user omit mandatory attributes. Metadata mapping consists of domain finding and variable mapping processes. CDISC Transformer suggests a rank ordered candidate domains and variables for each CRF question as shown in **Fig. 4** demonstrating three components, Domain & Attribute Mapping, Target Domains, and Target Attributes. Then a user can manually select the correct domain and/or variables.

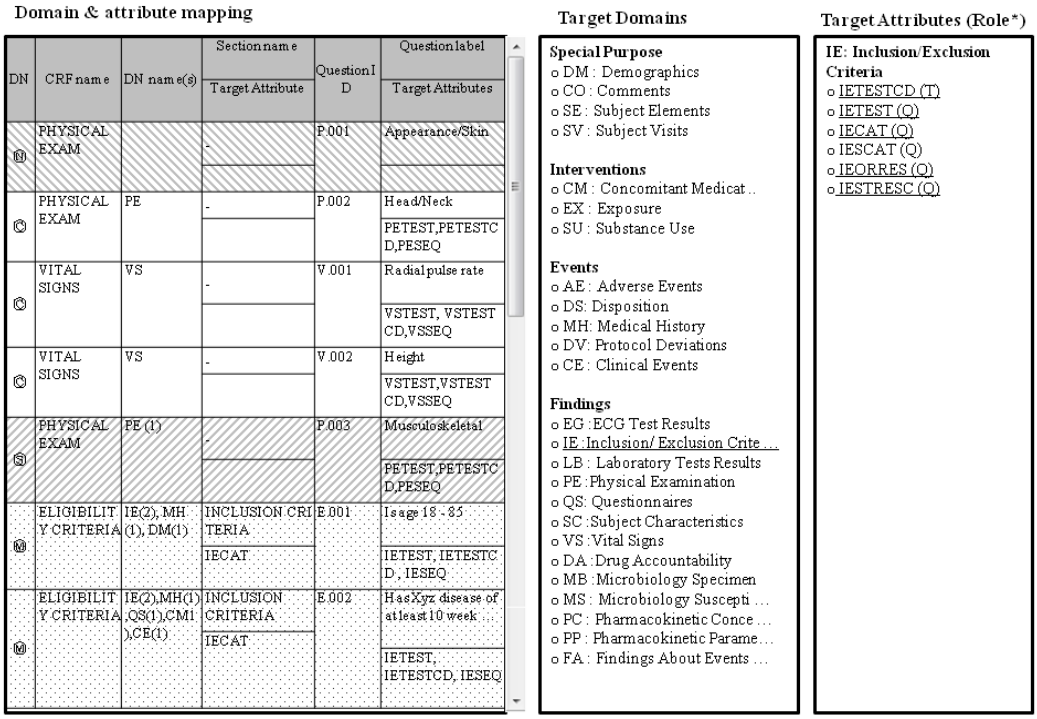

Fig. 4. Metadata mapping result and interface for approving or selecting the correct SDTM domains and variables. N (Ⓝ: No Domain, filled with down diagonal), S (Ⓢ: Single Domain, rilled with up diagonal), and M (Ⓜ: Multiple Domains, filled with dots).

The first domain and attribute mapping part rates the mapping results by an automatic mapping algorithm into three categories, N (Ⓝ: No Domain), S (Ⓢ: Single Domain), and M (Ⓜ: Multiple Domains), representing no, single, and multiple domain mapping, respectively. When one approves or completes the selection process among the candidates, the row become C category (Ⓒ: Domain Mapping Complete). In the Multiple Domain category, the numbers in parentheses represent the number of mapped metadata concepts to the question in the corresponding domain. By default, the domain with the highest number of mapped metadata concepts is suggested as the best candidate. The second Target Domain part underlines the suggested domains when a user selects a row in the first part. The third Target Attribute part lists the variables associated with the target domain selected by a user. Here the underlined variables represent ‘required’ values and (T) represent topic and (Q) qualifier roles. Finally, CDISC Transformer generates output files compliant with SDTM formats by integrating

selected domain and variables with clinical trial result file. Among the four SDTM variable types, values of the *timing* and *identifier* variables can trivially be mapped from directly the clinical trial result file to the output files and those of *topic* and *qualifier* variables are extracted by applying the appropriate mapping rule that determines the SDTM target domain and variable (see Table 1) and the corresponding data from the MDR or clinical trial result file. Output file validator checks whether it follows SDTM structure, whether the variable value follows SDTM terminology, and whether value follows value length, and whether the required fields each domain are filled up.

4. Evaluation

We applied CDISC blank CRF as a gold standard for an evaluation. The Blank CRF is an example document that includes which item on the CRF map to the corresponding domains and variables in the CDISC SDTM [16]. It consists of nine visits, 16 CRFs, and 125 questions. All of these questions are mapped to 13 domains and 304 redundant variables of SDTM model.

As mentioned in method section, CDISC Transformer consists of two sub-processes: domain finding and variable mapping. Domain finding is the core process of finding candidate domain based on three structural (i.e., *Question label*, *Section name*, *CRF name*) and two metadata (*Object, and Property*) attributes. Variable mapping process applying one of the predefined mapping rules starts after the target domain is determined. To test the effectiveness of metadata support, we evaluated the performance of domain finding processes with (i.e. with five variables) and without (i.e. with three variables) the two metadata attributes. The three structural attributes (*Question label*, *Section name, and CRF name*) can straightforwardly be obtained from the CRF itself and the two metadata attributes (*Object* and *Property*) can also be easily obtained from non-expert users. For a fair comparison, the following variable mapping process was equally applied to both comparison sets.

Table 2. Evaluation of CDISC Transformer

	Three trivial attributes only		**Five metadata attributes**	
	Domain	**Variable**	**Domain**	**Variable**
Correct	58	114	98	208
Incorrect	67	190	27	57
Accuracy	46.4%	37.5%	78.4%	68.4%

Table 2 demonstrates the advantage of using metadata for better mapping. It seems that metadata improved the mapping accuracies about twice. Domain mapping showed better results than variable mapping probably because of the higher coverage of metadata for domains and the more general concepts than variables. The accuracy can greatly be improved by the following user selections from the candidate lists. **Table 2** shows results from a purely automatic process without the user selection process.

Thirteen SDTM domains were mapped by the evaluation study. Interestingly, most variables in Findings General Observation class (Laboratory Test Results; LB, Vital Signs; VS, and Questionnaires; QS) were correctly mapped in metadata supported evaluation. On the contrary, most variables in Special-Purpose domains (Demographics; DM and Subject Elements; SE) were incorrectly mapped. This result implies that similar structures of the

Findings General Observation class (one question or test per row) and MDR (one data element per row) resulted high accuracy in contrast to the different structure of the Special-Purpose domains (one question or test per column).

5. Conclusion

CDISC Standards are crucial to streamlining clinical research and trial processes, and to improving data quality [17]. Many CTMSs have been used CDISC standards in clinical trial data managing and submitting. This study tried to establish common transforming rules for CDISC standard compliant clinical trial data using ISO/IEC 11179 metadata description standard. To the best of our knowledge, none of the previous works has considered the common transforming rules in CDISC standard support. Based on the transforming rules, we developed an algorithm, named CDISC Transformer, for transforming clinical trial data to CDISC standard compliant data. We also evaluated the performance of CDISC Transformer. MDR supported mapping process is demonstrated to be effective.

Current version of CDISC Transformer applies very basic text mining techniques and tools including subsumption test and NLM MetaMap, leaving room for improving performance. The other limitation of our research is that we focused only on CDISC SDTM model. Other CDISC standard models such as Operational Data Model (ODM), Analysis Data Model (ADaM), and Case Report Tabulation Data Definition Specification (CRT-DDS) will be included in the future study.

Recently, CDISC Standard model is not only used in submission of clinical trial data to FDA but also used in many biomedical research areas such as electronic data archiving during clinical trial [18] and designing element of clinical research system [19]. Moreover, CDISC standards are supporting not only the pharmaceutical industry but also other initiatives and services in healthcare. Following these trends, CDISC Transformer and CDISC SDTM transforming rule can be applied in applications for future research.

References

[1] C.L. Meinert, “Clinical Trials: Design, Conduct, and Analysis,” *Oxford University Press*, New York, pp. 3-4, 1986.
[2] CDISC, “Registered Solutions Providers Chart,” http://www.cdisc.org/rsp-chart.
[3] XML4Pharma, “The user-friendly ODM to SDTM Mapping software,” SDTM-ETL Version 1.4. http://www.xml4pharma.com/SDTM-ETL/index.html.
[4] Business & Decision, “CDISC Legacy Data Conversion,” http://cro.businessdecision.com/1651-cdisc-legacy-data-conversion.htm.
[5] CDISC, “Study Data Tabulation Model,” Version 1.2. http://www.cdisc.org/sdtm.
[6] U.S. FDA, “Study Data Standards Resources,” http://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm.
[7] D.B. Fridsma, J. Evans, S. Hastak, C.N. Mead, “The BRIDG project: a technical report,” *J Am Med Inform Assoc*, vol. 15, no. 2, pp.130-137, Mar.-Apr. 2007.
[8] C. Ohmann, W. Kuchinke, “Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and integration,” *Methods Inf Med*, vol. 48, no. 1, pp. 45-54, 2009.
[9] N. Meradith, S. John, B. Ann, R. Reza, C. Andrew, M. Jonathan, P. Ricardo, “Design and implementation of an institutional case report form library,” Clinical Trials, vol. 8, pp. 94-102, 2011. Article (CrossRef Link)
[10] ISO/IEC JTC 1/SC 32, “ISO/IEC 11179, Information Technology -- Metadata registries (MDR),” http://metadata-stds.org/11179/.

[11] L.B. Boyd, S.P. Hunicke-Smith, G.A. Stafford, E.T. Freund, M. Ehlman, U. Chandran, R. Dennis, A.T. Fernandez, S. Goldstein, D. Steffen, B. Tycko, J.D. Klemm, "The caBIG® Life Science Business Architecture Model", Bioinformatics, vol. 27, no. 10, pp. 1429-1435, May 2011. Article (CrossRef Link)
[12] U.S. Agency for Healthcare Research and Quality, "United States Health Information Knowledgebase," http://ushik.ahrq.gov/whats_new.html?Referer=What.
[13] G.A. Komatsoulis, D.B. Warzel, F.W. Hartel, et al., "caCORE version 3: Implementation of a model driven, service-oriented architecture for semantic interoperability," J Biomed Inform, vol. 48, no. 1, pp. 106-123, Feb. 2008. Article (CrossRef Link)
[14] D.G. Nohle, L.W. Ayers, "The tissue microarray data exchange specification: a document type definition to validate and enhance XML data," BMC Med Inform Decis Mak, vol. 5, no. 12, May 2005. Article (CrossRef Link)
[15] U.S. National Library of Medicine (NLM), "MetaMap and MetaMap Transfer (MMTx)," http://www.nlm.nih.gov/research/umls/implementation_resources/metamap.html.
[16] CDISC, "Metadata submission guidelines draft version 1.0 for SDTM IG V3.1.2 call for public review," http://www.cdisc.org/msg-draft.
[17] T. Souza, R. Kush, J.P. Evans, "Global clinical data interchange standards are here!," Drug Discov Today, vol. 12, no. 3-4, pp. 174-181, Jan. 2007. Article (CrossRef Link)
[18] W. Kuchinke, J. Aerts, S.C. Semler, C. Ohmann,. "CDISC standard-based electronic archiving of clinical trials," *Methods Inf Med*, vol.48, no.5, pp.408-413, 2009.
[19] K.Y. Megan, C. Dahlke, Q. Xiang, Y. Qian, D. Karp, R.H. Scheuermann, "Toward an ontology-based framework for clinical research databases," *J Biomed Inform*, vol. 44, no. 1, pp. 48-58, May 2011.

Yu Rang Park obtained B.S. degree in Multimedia Communication Engineering from Seoul Women's University, Seoul, Korea, in 2003, and M.S. degree in Molecular and Genomic Medicine from Seoul National University College of Medicine, Seoul, Korea, in 2006. She is currently a Ph.D. candidate in Molecular and Genomic Medicine at Seoul National University College of Medicine, Seoul, Korea. Her research interests include metadata based semantic interoperability in biomedical domain with ontology creation and correction.

Hye Hyeon Kim obtained B.S. and M.S degrees in Computer Science from Seoul Women's University, Seoul, Korea, in 2008 and 2010, respectively. She is currently a Ph.D. candidate in Biomedical Informatics at Seoul National University College of medicine, Seoul, Korea. Her research interests translational bioinformatics.

Hwa Jeong Seo obtained B.S. and M.S degrees in Computer Science from Seoul Women's University, Seoul, Korea, in 1997 and 1999, respectively. She obtained Ph.D. degrees in Medical Science from Suzuka University, Suzuka, Japan, in 2002. She is currently an assistant professor in Medical informatics, Graduate School of Public Health, Gachon University of Medicine and Science, Incheon, Korea.

Ju Han Kim obtained his M.D. and Ph.D. (Brain Imaging) degrees in Seoul National University and M.S. degree in Biomedical Informatics at MIT. After his fellowship at Beth Israel Deaconess Medical Center, he joined as a faculty member the Children's Hospital, Harvard Medical School, Boston, MA. He is currently a professor and chair of the Div. of Biomedical Informatics and director of Systems Biomedical Informatics National Core Research Center, Seoul National University College of Medicine, Seoul, Korea.