creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

의학석사 학위논문

임상문서의 의미론적 상호운용성 확보를 위한 국제표준 기반 공통 데이터 요소 모델링

Common Data Element modeling

to achieve semantic interoperability

of clinical documents

2023년 2월

서울대학교 대학원

의학과 의료정보학 협동과정

김 나 영

임상문서의 의미론적 상호운용성 확보를 위한 국제표준 기반 공통 데이터 요소 모델링

지도 교수 김 주 한

이 논문을 의학석사 학위논문으로 제출함
2022년 10월

서울대학교 대학원
의학과 의료정보학 협동과정
김 나 영

김나영의 의학석사 학위논문을 인준함
2023년 1월

위 원 장 _____ 박현애 _____ (인)

부위원장 _____ 김주한 _____ (인)

위 원 _____ 박상민 _____ (인)

국문초록

연구 배경 : 환자의 의료 데이터를 효율적으로 관리하고, 기관 간 공유하거나 통합하기 위해서는 데이터 표준화가 선행되어야 한다. 우리나라의 높은 병원정보시스템 보급률에도 불구하고, 병원 별 각기 다른 시스템 적용으로 인해 의료 영역에서의 데이터 표준화 수준은 미비한 실정이다. 본 연구가 수행되기에 앞서, 의료데이터 상호운용성 확보를 위한 ISO/IEC 11179 MDR(Metadata Registry) 국제표준을 이용한 선행연구로 국내의 5개 대학병원에서 입원기록지, 외래초진기록지, 응급기록지, 수술기록지, 퇴원기록지의 5개 임상문서를 수집하여 메타데이터를 추출하고 여러 병원 간 데이터 통합 작업이 수행되었다.

연구 방법 : 본 연구에서는 선행연구에서 추출된 메타데이터들이 신규병원의 임상문서 표현에 얼마나 적용될 수 있는지 확인하고, 이를 효율적으로 관리하기 위해 의미 관계에 따른 그룹화 작업을 하고자 한다. 선행연구의 결과 메타데이터를 검토하며 데이터의 본질적 특성에 따라 더욱 정교하게 정제하고, 정제된 메타데이터를 이용하여 추가로 5개 병원의 임상문서에 적용시킴으로써 높아진 메타데이터 재사용률을 확인하였다.

연구 결과 : 총 10개 병원의 5개 임상문서에 대한 전체 항목은 3,073개이고, 중복 항목을 제거하였을 때 848개의 고유한 메타데이터로 정리되었다. 5개의 임상 문서에 대한 표준화 정도를 평가하기 위해, 본 연구의 10개 병원을 하나씩 추가하면서 증가되는 고유한 메타데이터의 경향성을 확인하였다. 1개 병원이 추가되었을 때 23.36%의 증가율을 보였고, 10개 병원이 모두 추가되었을 때 5.54%의 증가율을 보여 고유하게 늘어나는 메타데이터 개수는 급격히 낮아졌음을 확인하였다. 848개의 고유한 메타데이터의 효율적 검색 및 활용을 위한 방안도 제시되었다. 항목 수준의 관점에서는 ISO/IEC 11179 metamodel의 Object Class를 이용한 그룹화를 진행하고, 임상문서 설계 관점에서는 연관된 메타데이터의 관계를 설명하는 Composite 모델링을 진행하였다.

결론 및 논의 : 본 연구를 통해 병원마다 같은 의미를 가지지만 다르게 표현되었던 항목들 간의 정보 교환을 위한 표준화된 메타데이터가 정리되었다. 이 메타데이터들은 5개의 임상문서 영역에 포함된 데이터에 대해서 의미론적 상호운용성을 확보하여, 향후 표준화된 임상문서 설계에 도움을 줄 수 있다. 또한, 잘 정의된 메타데이터의 재사용성을 개선하기 위한 모델링 방법을 제안하였으므로, 임상문서 설계 시 기관의 필요에 따라 항목을 선정하여 사용할 수 있게 된다.

주요어 : 임상문서, 공통 데이터 요소, MDR, 의미론적 상호운용성
학 번 : 2021-24710

목 차

제 1 장 서 론 7
제 1 절 연구의 배경 및 필요성 7
제 2 절 연구의 목적 9

제 2 장 연구 방법 10
제 1 절 선행연구의 데이터 정제 10
제 2 절 신규 임상문서에 대한 적용 19
제 3 절 메타데이터 재사용성 개선 21

제 3 장 연구 결과 22
제 1 절 선행연구의 데이터 정제 22
제 2 절 신규 임상문서에 대한 적용 25
제 3 절 메타데이터 재사용성 개선 28

제 4 장 논의 35

참고문헌 37

표 목차

[Table 1] 23 basic attributes that describes ISO/IEC 11179 MDR(Metadata Registry) Standard.....13

[Table 2] Number of Common Data Elements extracted from five clinical documents used at 5 teaching hospitals in Korea 23

[Table 3] Number of Common Data Elements extracted from five clinical documents used at 10 teaching hospitals in Korea 24

[Table 4] Composite Common Data Element modeling examples 34

그림 목차

[Figure 1] ISO/IEC 11179 MDR(Metadata Registry) Standard metamodel 12

[Figure 2] Examples of Data Element extraction following the ISO/IEC 11179 MDR(Metadata Registry) Standard 14

[Figure 3] Distinction between Data Element and Common Data Element.....15

[Figure 4] Examples of Heterogeneous Data Type managing.....17

[Figure 5] Flow diagram of Common Data Element reuse and registering method.....20

[Figure 6] Increasing trend of Unique Common Data Elements 26

[Figure 7] Increasing trend of Unique Common Data Elements (by clinical documents) 27

[Figure 8] Data Element Extraction process and approaches to enrich semantics.....29

[Figure 9] Data Element Examples with Object Class 31

[Figure 10] Data Element Query with same Object Classes 32

제 1 장 서 론

제 1 절 연구의 배경 및 필요성

임상 연구 및 의료 분야에서 데이터 공유 및 통합의 가치가 점점 더 중요하게 인식되고 있다. 데이터의 공유 및 통합은 진료정보의 2차 사용을 촉진하여 의학연구의 발전을 도모할 수 있다. 그러나 병원 별 제각기 다르게 구축된 시스템으로 인해 의료정보의 데이터 구조가 달라 병원 간 데이터 공유, 통합은 현실적으로 어렵다. 기관 간 데이터를 공유하고자 할 때에는 데이터의 구문(Syntax) 뿐만 아니라 의미(Semantics)에 대한 손실이 없게 관리되어야 한다.

ISO/IEC 11179는 국제표준화기구(ISO: International Organization for Standardization) 및 국제전기기술위원회(IEC: International Electrotechnical Commission)에서 제정한 메타데이터 레지스트리 표준으로, 데이터를 이해하고 공유할 수 있도록 메타데이터의 표준화 및 등록을 문서화한다. ISO/IEC 11179 메타데이터 모델을 통해 생성되는 MDR(Metadata Registry) 기반 Data Element는 데이터의 상호운용성을 달성하기 위해 적용되는 중요한 방법이다. 이 기술을 통해 서로 다른 병원이 표준적인 방법으로 데이터를 주고받을 수 있고, 의미(Semantics)와 구문(Syntax)이 소통하게 되어 데이터 통합이 가능해진다. 이와 같은 의료데이터의 의미론적 상호운용성 보장을 위해 ISO/IEC 11179 MDR 국제표준을 따르는 메타데이터 저장소인 CHMR(Clinico-Histopathological Metadata Registry)을 운영하고 있다.

병원 별로 다르게 사용되는 임상문서의 데이터를 다기관 통합하기 위해서는 임상문서 항목의 표준화가 필요하다. 이는 의무기록에서 기록할 항목이 무엇인지 정의하고, 각 항목에 어떤 값을 넣을 것인지 약속하는 것을 의미한다. 값은 숫자(int, float…)나 문자(string)와 같은 Free-Text 입력형, 예(Yes) 또는 아니오(No) 중 하나를 선택하는 선택형으로 나뉘어질 수 있다. ISO/IEC 11179 MDR(Metadata

Registry)에서는 Free-Text 입력형 값을 Non-Enumerated Value Domain Type으로, 선택형 값을 Enumerated Value Domain Type으로 구분하고 있다.

이러한 의무기록 항목의 표준화 작업은 해당 항목의 개념에 대한 명확한 정의를 제공한다는 점에서 데이터의 모호성 및 중의성 문제를 해결하는 데에 도움을 준다. 예를 들어, 같은 용어를 진료과마다 다른 의미로 사용하는 것을 예방할 수 있게 되고, 동일한 질병이 기관마다 다르게 표현되고 있더라도 기관 간 데이터 공유 및 통합이 가능해진다. 또한, 값(value)에 대한 제약조건을 충분하게 정의하여 데이터의 입력 오류를 방지할 수 있다.

본 연구에 앞서 ISO/IEC 11179 MDR 국제표준 방법론을 이용한 선행된 연구 내용을 검토하였다. 선행연구에서는 한국의 5개 대학병원(아주대학교병원, 전남대학교병원, 가천대학교길병원, 부산대학교병원, 서울대학교병원)의 실제 임상에서 사용되는 5개 임상문서(입원기록지, 외래초진기록지, 퇴원요약지, 응급기록지, 수술기록지)를 수집하여 임상문서 항목들을 ISO/IEC 11179 Standard 기반의 메타데이터를 이용해 통합시킴으로써 의미론적 상호운용성을 달성하기 위한 시도를 하였다[1, 2]. 이 임상문서들은 Patient, Past History, Admission information, Operation, Family History, Social History, Lab Result, Medication, Vital Sign, Treatments, Physical Examination영역에 대한 데이터를 포함하며, 환자의 입원부터 퇴원까지의 과정에서 필수적으로 사용되는 임상문서여서 선택되었다.

결과적으로 전체 25개 임상문서에서 추출된 1,142개의 메타데이터가 통합 과정을 거쳐 586개로 줄어들었음을 확인하였다.

제 2 절 연구의 목적

본 연구는 선행연구에서 개발한 메타데이터가 새로 추가된 병원의 임상문서 항목에 얼마나 적용될 수 있는지 확인하고, 적용 결과를 토대로 5개 임상문서 영역에서의 표준화 정도를 확인하고자 한다. 이후 상호 연관된 메타데이터를 의미 관계에 따라 그룹화하여 MDR(Metadata Registry) 내 검색 효율을 높이고 재사용성을 향상시키고자 하였다[3, 4].

연구의 세부적인 목적은 다음과 같다.

1. 586개의 메타데이터가 데이터의 특성에 따라 정의되어 있는지 확인하고 정제한다.

2. 정제된 586개의 메타데이터를 새로운 5개 병원의 임상문서에 적용시켜 본다(선행연구의 메타데이터로 표현할 수 없는 항목의 경우 새로운 메타데이터를 추출하여 관리한다). 이후 병원이 추가됨에 따라 고유하게 증가하는 메타데이터 개수를 확인하여 5개 임상문서의 표준화 정도를 평가한다.

3. 의미적으로 상호 연관된 메타데이터를 그룹화 시킴으로써 메타데이터의 재사용성 개선 방안을 제시한다.

제 2 장 연구 방법

제 1 절 선행연구의 데이터 정제

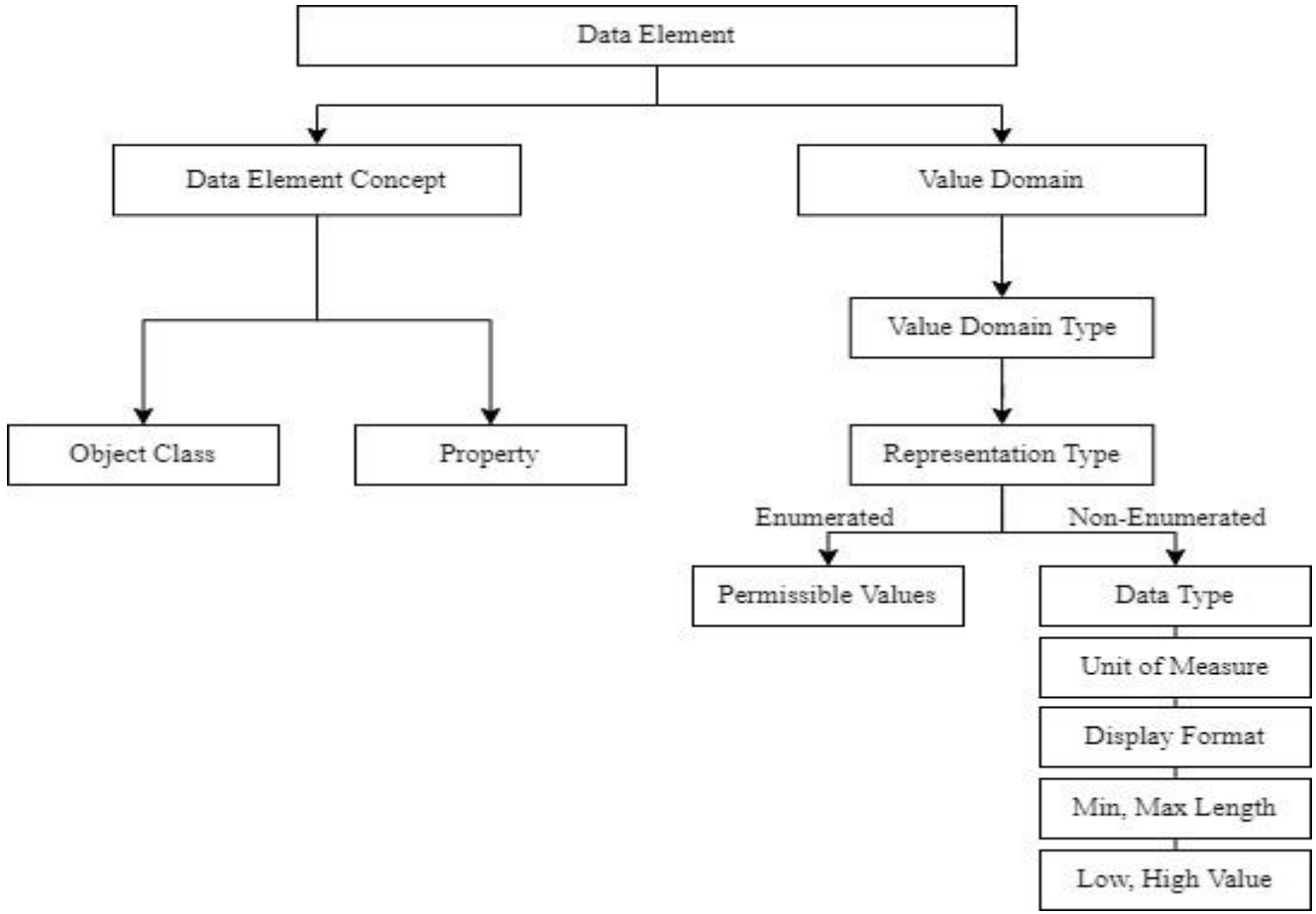
데이터 표준화를 위한 여러 방법 중 ISO/IEC 11179 MDR(Metadata Registry) 국제표준은 MDR을 통해 데이터를 설명하는 표준화된 메타데이터를 등록하고 지속적으로 유지, 관리할 수 있게 해준다. ISO/IEC 11179 MDR(Metadata Registry)에 의해 관리되는 메타데이터의 기본 단위는 "Data Element(DE)"이다. Data Element(DE)의 구조는 Figure 1에서 확인할 수 있다.

Data Element(DE)는 객체(Object)와 속성(Property)을 이용하여 개념영역을 서술하는 Data Element Concept(DEC)과 값에 대한 영역을 서술하는 Value Domain(VD)의 묶음으로 설명된다. Value Domain(VD)은 Enumerated(열거형), Non-Enumerated(비 열거형)의 Value Domain Type을 가지는데 이 Type에 따라 값 영역을 설명하는 항목이 구성된다. ISO/IEC 11179 MDR Standard에서 정의하는 DE의 세부 구성요소는 Table 1에 명시되어 있다[5]. Figure 2에서는 이를 이용한 의료데이터의 표준화 과정을 설명하고 있다.

"Data Element(DE)"와 "Common Data Element(CDE)"는 종종 같은 의미로 사용되고 있으나[6], 본 연구에서는 DE와 CDE에 대한 구분을 명확히 하고자 한다. 이를 위해 Figure 3에서 ISO/IEC 11179 표준에서 정의하는 Data Element의 세부 메타데이터들을 최소한으로 보여주며 병원 별 데이터 수집 방식의 차이를 확인할 수 있도록 하였다.

ISO/IEC 11179 표준의 공식 명칭은 Data Element(DE)이며, 병원마다 각기 다르게 수집되는 데이터에 대하여 DE를 추출하고 관리하게 된다. 이 때, 각 병원의 항목을 메타데이터로 추출하고 관리하는 것을 DE라 하고, 같은 의미에 대한 여러 개의 DE들이 consensus를 이루어 여러 병원에서 공통적으로 사용가능 하도록

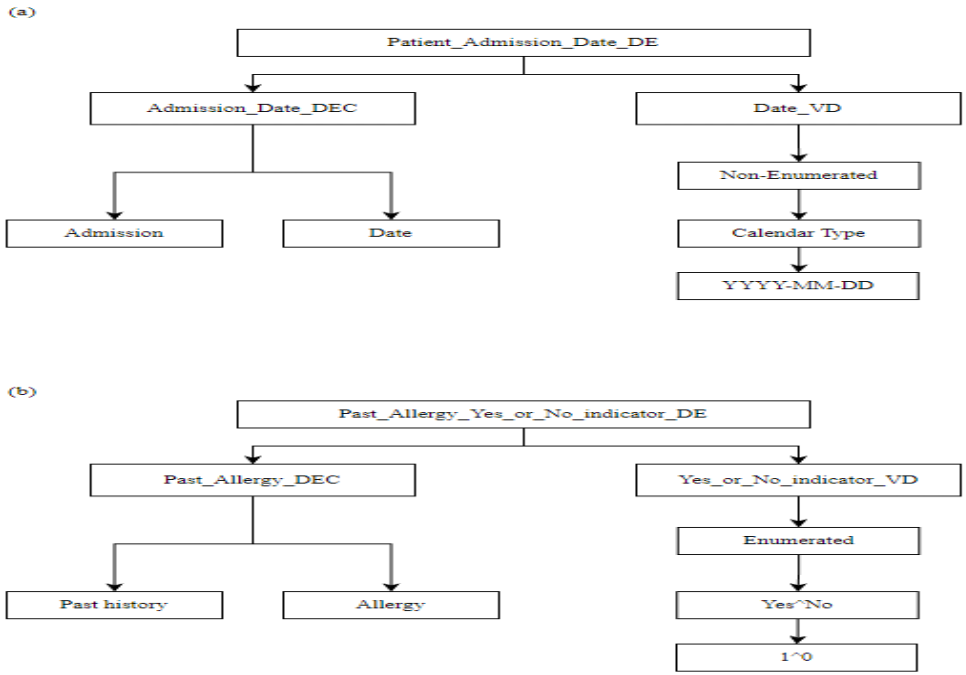
정제된 DE를 CDE로 구분하였다.



[Figure 1] ISO/IEC 11179 MDR(Metadata Registry) Standard metamodel

[Table 1] 23 basic attributes that describes ISO/IEC 11179 MDR(Metadata Registry) Standard

No.		**Attribute** (Required: *)	**Definition**
1		Data Element Name (Eng)*	Name of the Data Element (Eng)
2		Data Element definition (Eng)*	Clear definition of the Data Element (Eng)
3		Data Element Name (Kor)*	Name of the Data Element (Kor)
4		Data Element definition (Kor)*	Clear definition of the Data Element (Kor)
5		Data Element Context Name*	Setting in which Data Element is used
6		Conceptual Domain*	Concept that expresses its description or valid instance meanings
7		Data Element Concept Name*	Concept that is an association of a property with an object class
8		Object	Set of ideas, abstractions or things in the real world that are identified with explicit boundaries and meaning and whose properties and behaviour follow the same rules
9		Property	Quality common to all members of an object class
10		Value Domain Name*	Name of the Value Domain in which provides representation and set of permissible values
11		Value Domain Definition*	Semantic content of a possible value
12		Value Domain Type*	Enumerated and Non-Enumerated Value Type
13		Representation Type*	- Enumerated: Check Box, Radio Button, Select Box … - Non-Enumerated: Calendar, Text Box, Text Area …
14	Enumerated	Permissible Values (identifiers) *	Allowable list of values (codes)
15	Non-Enumerated	Value Domain Data Type*	Type of a value: Number, String …
16		Unit of Measure	Actual units in which the associated values are measured
17		Display Format*	Display Format of a value : YYYY-MM-DD, 999 …
18		Value Minimum Length	Minimum length of a value
19		Value Maximum Length	Maximum length of a value
20		Low Value (for Number data type)	Lowest value
21		High Value (for Number data type)	Highest value
22		Classification scheme	Descriptive information for an arrangement or division of objects into groups based on criteria such as characteristics, which the objects have in common
23		Classification scheme items	Metadata item of a type for which classification is supported in a given metadata registry



[Figure 2] Examples of Data Element extraction following the ISO/IEC 11179 MDR(Metadata Registry) Standard

(a) shows an example of medical data with "Non-Enumerated" Value Domain Type. "Patient_Admission_Date_DE" describes the patient's start date of admission.

(b) shows an example of medical data with "Enumerated" Value Domain Type. "Past_Allergy_Yes_or_No_indicator_DE" describes the patient's past allergic history by pre-defined permissible values("Yes", "No")

	Question Label	Representation Type	Data Type	Permissible Values
A병원 DE	성별	Radio Button		남자^여자
B병원 DE	Patient Sex	Text Box	String	
C병원 DE	Sex	Select Box		Male^Female^Unknown
D병원 DE	환자의 성별	Radio Button		Male^Female

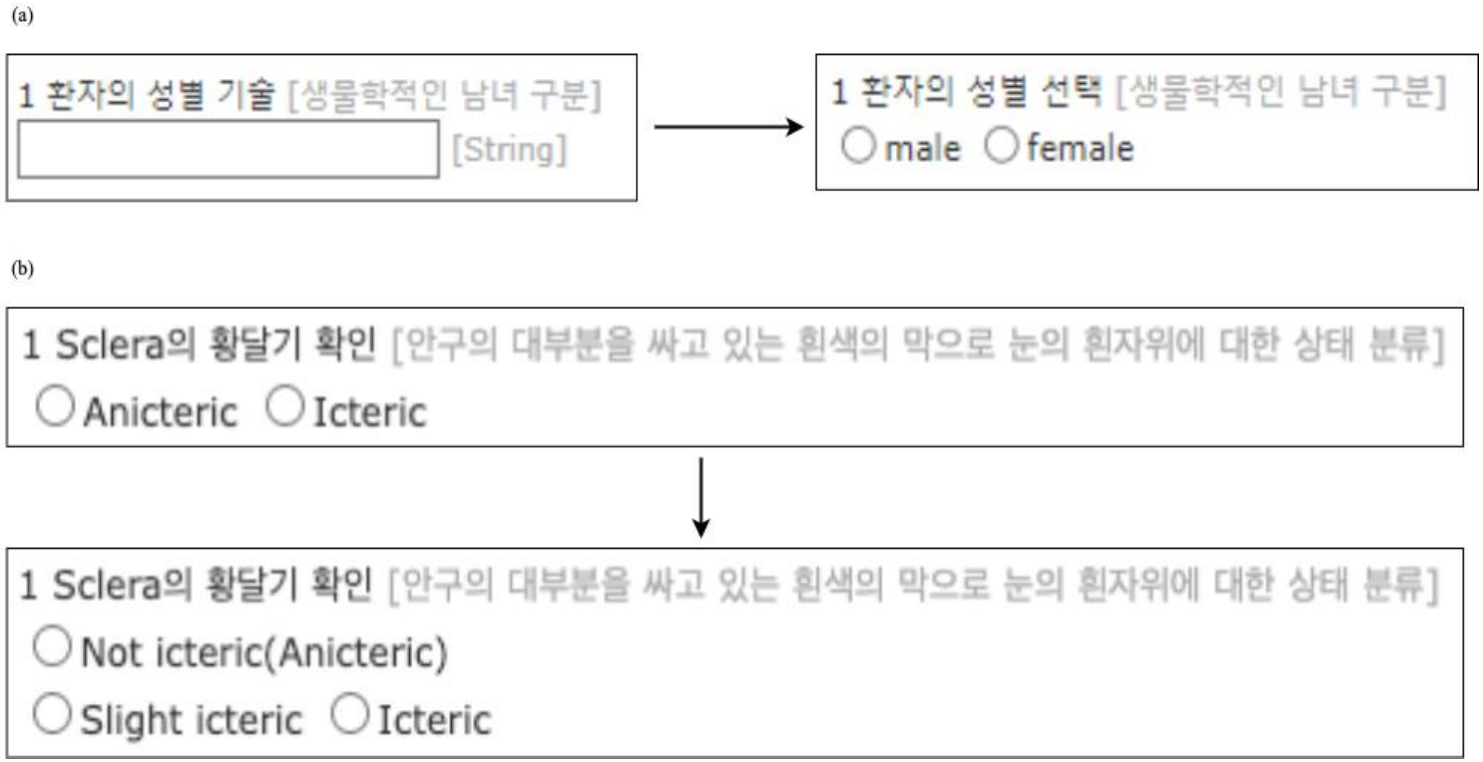
CDE	**Patient Sex**	**Radio Button**		**Male^Female^Unknown**

[Figure 3] Distinction between Data Element and Common Data Element

본 연구는 선행연구의 결과인 586개의 CDE를 정제한 후 이를 사용하여 진행되었다.

586개의 CDE 목록을 조화하는 데 CHMR(Clinico-Histopathological Metadata Registry) 데이터베이스를 이용하였다. 각 병원의 임상문서는 CHMR의 서식 생성 기능인 Form Builder를 통해 등록되고 유지, 관리되므로, Form ID를 이용하여 DE에 접근하는 방법을 취하였다. 전체 25개의 Form Name을 통해 Form ID를 확인한 후, 해당 Form ID를 이용하여 1,142개의 전체 DE를 쿼리하였다. 이후 “distinct” 키워드를 이용하여 중복되는 값을 제거한 결과 586개의 고유한 CDE를 얻을 수 있었다.

이 586개의 CDE를 새로 추가된 병원의 임상문서 항목에 적용시키기 이전에 메타데이터를 구조화하는 작업을 수행하였다. 또한, 기존의 CDE와 같은 의미를 가지더라도 이질적인 값 유형으로 exact mapping이 어려운 경우 기존 CDE를 정제하였다. 이로써 선행 연구에서 개발한 CDE의 질을 높이고, 동일한 의미의 여러 DE가 생성되는 문제—같은 의미를 가지지만 다른 값 유형을 사용하는 경우—를 해결하고자 하였다.



[Figure 4] Examples of Heterogeneous Data Type managing

(a) managing method of the different hospital items which has same meaning but different Value Domain type
(b) managing method of the different hospital items which has same meaning, same Value Domain type, but having difference in permissible values

예를 들어, Figure 4-a "환자의 성별" 항목의 경우, 기존에 정의된 "Patient_Sex_Type_DE (DE-id: 20984)"가 있고, 해당 CDE의 definition은 "생물학적인 남녀 구분(The biological sexual distinction)"이다. 그러나 이 CDE는 Value Domain Type이 Non-Enumerated로 Free-Text로 입력하도록 정의된 값이었다. "성별"이라는 concept의 본질(Ontological Meaning)을 생각해보았을 때 코드화를 통한 구조화가 가능한 항목이므로, 기존 DE의 Value Domain Type를 Enumerated로 업데이트 하고, "Representation Type"은 Radio box, Permissible Values로 "Male"과 "Female"을 부여하였다.

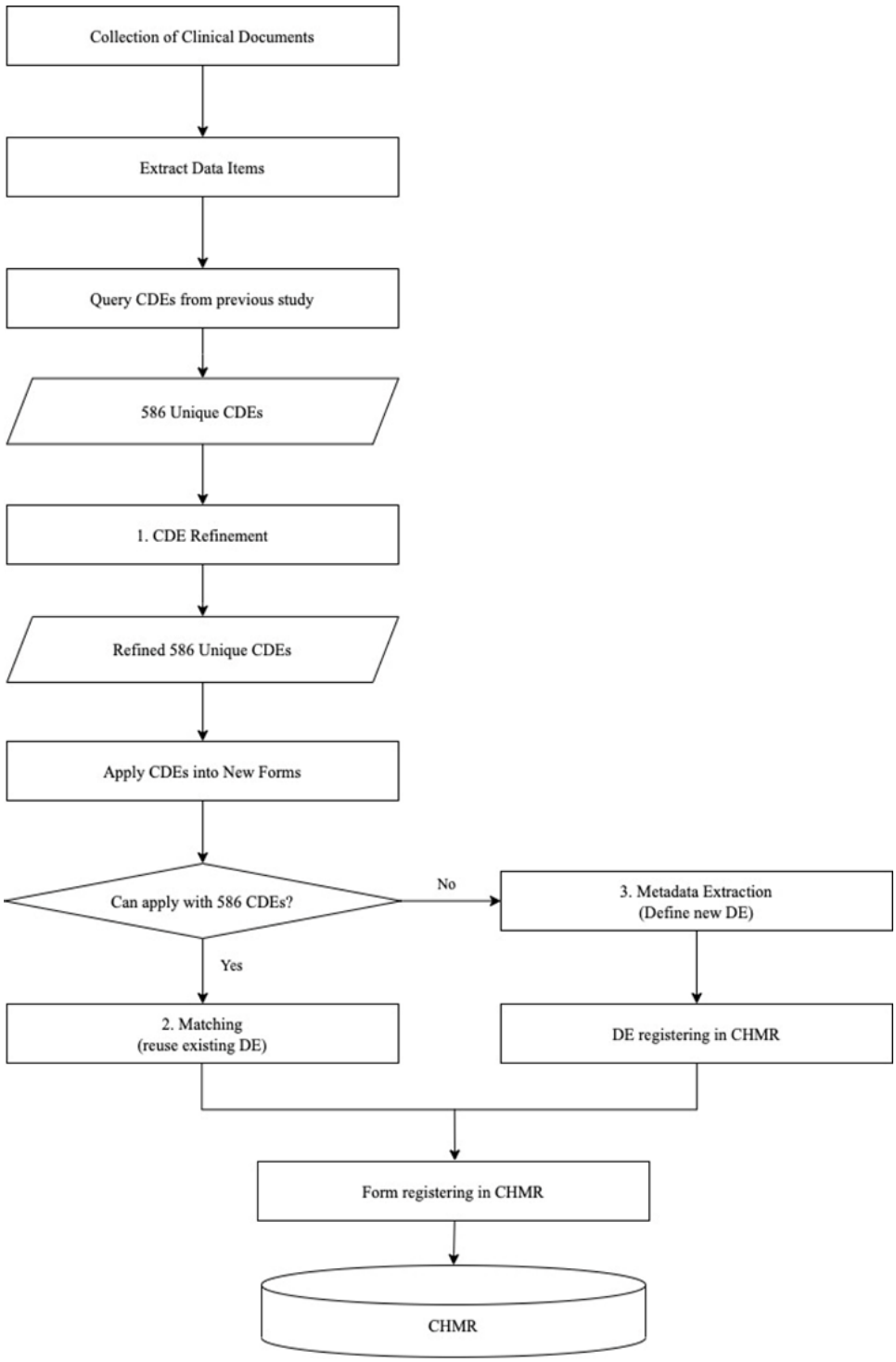
Figure 4-b에서는 동일한 의미를 가지는 항목에서 답항의 허용값이 다를 경우의 해결방안을 제시하고 있다. Sclera(공막)의 황달기를 확인하는 항목에 대하여 선행연구의 CDE로는 "Sclera_Status_Category_DE (DE-id: 21260)"가 있고, 이 CDE는 답항으로 Anicteric과 Icteric 중 하나를 선택하도록 설계되어 있었다. 그런데, 추가된 병원 서식에서는 공막의 상태를 평가함에 있어 3가지 답항 세트(Not icteric, Slight icteric, Icteric)를 가지도록 정의하고 있었다. 따라서 기존 "Sclera_Status_Category_DE (DE-id: 21260)"에 대한 답항의 허용값을 3가지 답항 세트로 수정함으로써, 공막 상태를 평가하는 항목에 대해 기관 간 상호운용성이 확보되도록 하였다.

이번 데이터 정제 작업에서 CDE가 내포하는 개념에 대한 본질적 특성을 파악하고 Figure 4-a와 같이 기존 CDE를 더욱 구조화(Unstructured Data -> Structured Data)하는 작업을 수행하였다. 또한, 동일한 임상 개념을 가지지만 Value Set에서 차이를 보이는 경우, Figure 4-b와 같이 기존 CDE의 Permissible Values를 Comprehensive Value Set으로 수정하여 이 문제를 해결하였다.

제 2 절 신규 임상문서에 대한 적용

ISO/IEC 11179 국제표준 기반의 정제된 586개의 CDE들이 새로 추가된 병원의 새로운 임상문서를 얼마나 잘 표현할 수 있는지 확인하였다. 선행연구에서 수집된 입원기록지, 외래초진기록지, 퇴원요약지, 응급기록지, 수술기록지에 대한 새로운 5개 병원(연세대학교세브란스병원, 인재대학교병원, 한림대학교병원, 을지대학교병원, 고려대학교병원)의 자료가 본 연구를 위해 수집되었다.

본 연구의 핵심 프로세스를 Figure 5를 통해 확인할 수 있다. 총 10개의 병원으로부터 임상문서를 수집한 후 각 임상문서별 항목의 문항과 답항을 확인하였다. 이후 CHMR를 이용하여 선행연구의 CDE를 쿼리하였다. 신규 병원의 항목들을 표현하기에 기존의 CDE를 사용가능한 경우 기존 CDE를 재사용하고, 기존 CDE로 표현하기 어려운 항목의 경우 새로운 DE를 추출하는 방법으로 CHMR를 확장 수정하였다.



[Figure 5] Flow diagram of Common Data Element reuse and registering method

This figure shows three main step of CDE reuse and registering method:
1. CDE refinement 2. CDE Matching (reuse existing CDE) 3. Define new DE

제 3 절 메타데이터 재사용성 개선

총 10개 병원에 대한 기존 CDE 적용 및 신규 DE 추출을 통한 결과로 도출된 Unique CDE는 특정 서식에 종속되지 않는 메타데이터이다. 이러한 잘 정의된 메타데이터를 재사용하는 것은 여러 병원의 임상문서 항목들에 연결되는 CDE를 공유하게 할 수 있으며, 이는 의료데이터 간 상호운용성을 확보하는 데 중요한 작업이다. 본 연구에서는 다음 방법을 사용하여 정교하게 curation된 CDE의 재사용성을 높이는 방법을 제안한다.

1. ISO/IEC 11179 metamodel의 Object Class를 이용한 CDE 그룹화

항목 수준의 관점에서 ISO/IEC 11179 metamodel의 개념 영역을 표현하는 Data Element Concept(DEC)은 잘 정제된 CDE를 효율적으로 검색하고 활발한 재사용을 촉진시킨다. Data Element Concept(DEC)는 Object Class와 Property로 구성된다. 동일한 Object Class를 가진 CDE를 collection으로 관리하게 된다면, 항목 수준에서 정의된 CDE가 해당 개념의 Reference Set로서의 역할을 하게 되고, 기관별 필요에 따라 선택하여 재사용할 수 있다.

2. Composite CDE 모델링

임상문서 구성의 관점에서는 Composite CDE 모델링 방법을 적용한다. 병원의 임상문서 설계 시, 더욱 의미 있는 정보를 생산할 수 있는 연관된 CDE들의 조합을 구성하고 Composite CDE로 모델링하고자 하였다. 예를 들어, 어떤 환자의 흡연 상태에 대한 정보를 얻고 싶은 경우, ‘과거 흡연 여부’, ‘현재 흡연 여부', ‘하루 흡연량', ‘총 흡연 기간' 등의 데이터를 함께 보았을 때 더욱 풍부한 정보가 제공된다. 이러한 Composite CDE 모델링은 다른 Object Class를 가진 CDE들 간의 복합 관계를 설명할 수 있다.

제 3 장 연구 결과

제 1 절 선행연구의 데이터 정제

본 연구는 선행연구의 결과인 586개의 Unique CDE를 재사용하여 신규 임상문서를 얼마나 잘 표현할 수 있을지 확인하고자 하였다. Table 2에서 5개 대학병원(아주대학교병원, 전남대학교병원, 가천대학교길병원, 부산대학교병원, 서울대학교병원)의 5개 임상문서(입원기록지, 외래초진기록지, 퇴원요약지, 응급기록지, 수술기록지)에 대한 각 병원 별, 서식 별 CDE 개수와 Unique CDE 개수를 확인할 수 있다.

환자의 입원에 대한 정보를 기록하는 가장 기본적인 임상문서인 “입원기록지”에 대하여 5개 대학병원의 항목 분석을 한 결과, 부산대학교병원은 204개의 서식 항목으로 구성, 서울대학교병원은 12개의 서식 항목으로 구성되어 기관 간 항목 구성과 범위에 매우 큰 차이가 있어 병원 별 구조화 수준이 상이한 결과를 보였다.

본 연구를 수행하기 위하여 연세대학교세브란스병원, 인제대학교병원, 한림대학교병원, 을지대학교병원, 고려대학교병원으로부터 선행연구와 같은 5개의 임상문서를 수집하였다. 수집된 각 임상문서 항목의 문항과 답항을 확인하면서 기존 CDE 재사용 및 DE 신규 추출 과정을 진행하였고, 그 결과는 Table 3과 같다.

10개 병원의 5개 임상문서는 3073개의 항목으로 구성되어 있고, DE 통합 과정을 거쳐 848개의 Unique CDE로 줄어들었다. 이는 10개 병원의 5개 임상문서에서 848개의 Unique CDE로 데이터 공유 및 통합이 가능하다는 점을 시사한다.

[Table 2] Number of Common Data Elements extracted from five clinical documents used at 5 teaching hospitals in Korea

Hospital		**Admission Note**	**Initial Medical Examination Note**	**Discharge Summary**	**Emergency Note**	**Operation Note**	**Total No. of DEs**	[a] **No. of Unique CDEs**	[b] **DE Reuse Rate**
Ajou	DE	84	48	70	83	37	322	227	29.5
Chonnam	DE	30	35	20	27	26	138	87	37.0
Gachon	DE	70	28	44	54	11	207	161	22.2
Pusan	DE	204	123	46	43	12	428	266	37.9
Seoul	DE	12	6	9	10	10	47	31	34.0
Total	DE	400	240	189	217	96	1142		48.7
	Unique CDE	297	162	142	178	57		586	

[a] No. of unique CDEs across the five clinical documents

[b] Reuse ratio of DEs across the five clinical documents

[Table 3] Number of Common Data Elements extracted from five clinical documents used at 10 teaching hospitals in Korea

Hospital	Year of collection		**Admission Note**	**Initial Medical Examination Note**	**Discharge Summary**	**Emergency Note**	**Operation Note**	**Total No. of DEs**	[a] **No. of Unique CDEs**	[b] **DE Reuse Rate**
Ajou	2015	DE	84	48	70	83	37	322	227	29.5
Chonnam	2015	DE	30	35	20	27	26	138	87	37.0
Gachon	2015	DE	70	28	44	54	11	207	161	22.2
Pusan	2015	DE	204	123	46	43	12	428	266	37.9
Seoul	2015	DE	12	6	9	10	10	47	31	34.0
[c] Yonsei	2022	DE	162	29	37	53	23	304	245	19.4
[c] Inje	2022	DE	231	147	82	30	16	506	323	36.2
[c] Hallym	2022	DE	67	9	30	80	14	200	175	12.5
[c] Eulji	2022	DE	168	78	71	70	22	409	227	44.5
[c] Korea	2022	DE	155	124	57	148	28	512	231	54.9
Total		DE	1183	627	466	598	199	3073		72.4
		Unique CDE	564	322	212	342	74		848	

[a] No. of unique CDEs across the five clinical documents

[b] Reuse ratio of DEs across the five clinical documents

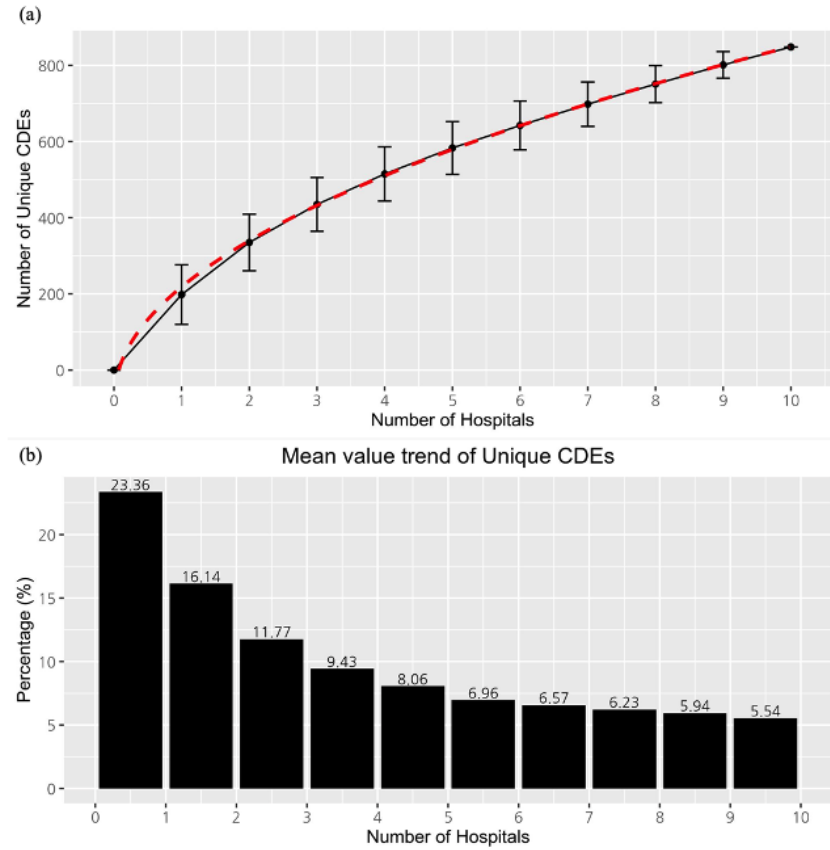
[c] Collected for this study

제 2 절 신규 임상문서에 대한 적용

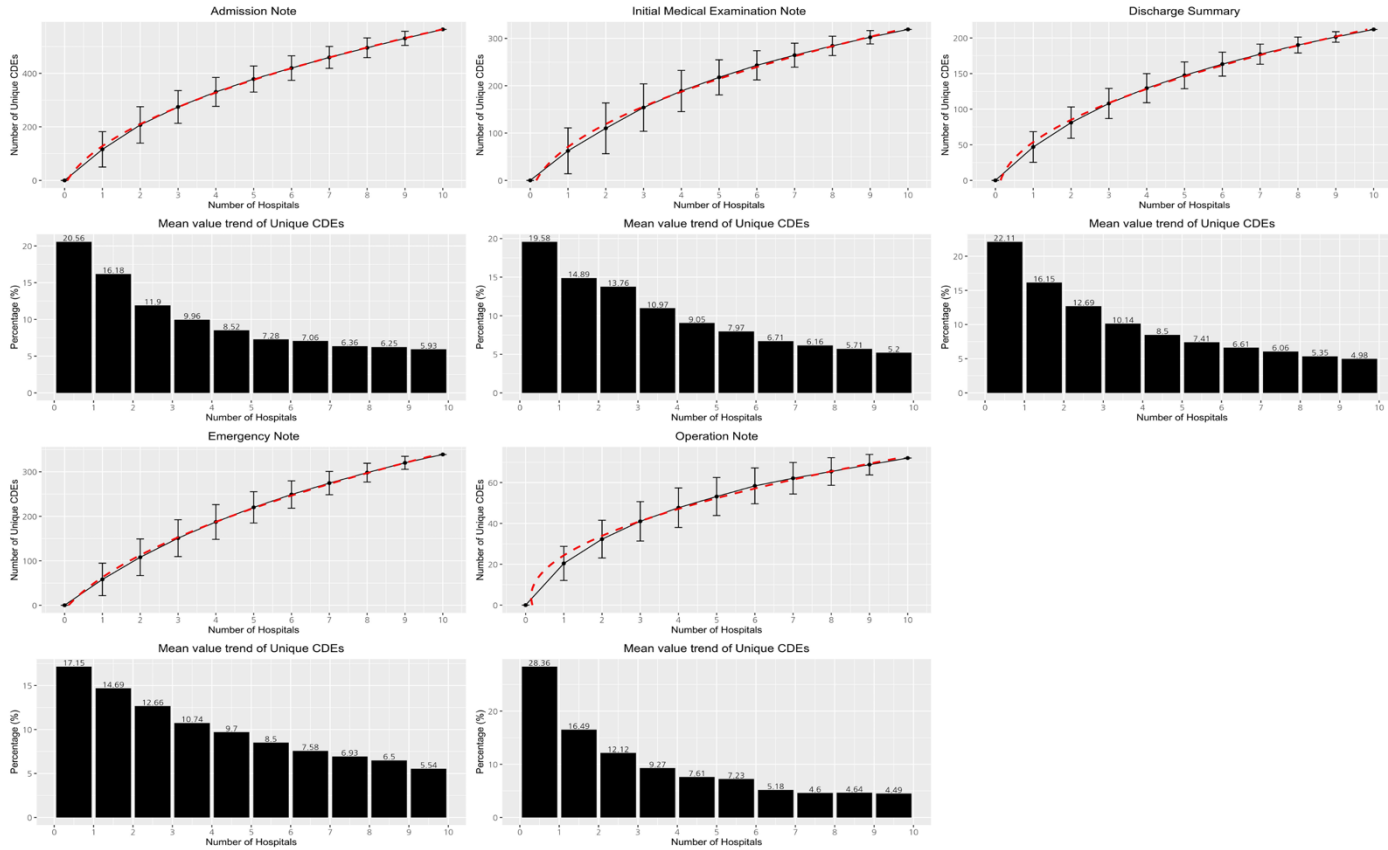
새로운 병원이 추가될 때의 Unique CDE 증가 추세를 확인하기 위해, 병원 추가에 따른 Unique CDE 개수를 꺾은선 그래프로 표현하였다 (Figure 6-a). 이때 사용된 Unique CDE는 10개 병원의 5개 서식으로부터 추출되었다. 추가되는 병원의 순서는 무작위로 1000번 시행하였고, 해당 순서 당 결과 Unique CDE 개수의 평균(Mean)과 표준편차(Standard Deviation, SD)를 구하여 오차 막대로 표시하였다. 이 데이터를 이용하여 2차 함수 곡선을 가지는 추세선으로 표현할 수 있었다. Figure 6-a로부터 병원이 추가될 때마다 구간 간 평균값의 차이가 낮아지는 것을 확인하였다.

$$\text{Percentage}(\%) = \left(\frac{\textit{다음 구간의 1,000 번 무작위 추출한 Unique CDE 개수의 평균값} - \textit{이전 구간의 1,000 번 무작위 추출한 Unique CDE 개수의 평균값}}{\textit{전체 Unique CDE 개수}}\right) \times 100$$

Figure 6-b는 병원이 추가 됨에 따라 Unique CDE의 증가율을 퍼센트로 표현하였다. 본 연구에서는 10개의 병원에 대한 구간 간 평균값의 차이를 전체 Unique CDE로 나눈 값을 이용하여 CDE 증가량을 평가하였고, 9번째 병원이 추가될 때부터 신규 CDE의 증가율이 약 5 % 정도가 됨을 확인하였다. 해당 그래프를 통해 병원이 하나 추가되었을 때의 Unique CDE의 증가율이 23.36 %에서 병원이 10개 추가되는 경우 5.54 % 까지 감소하였음을 정량적으로 확인하였다. Figure 7에서는 Unique CDE 증가 경향을 각 서식별로 나타내고 있다.



[Figure 6] Increasing trend of Unique Common Data Elements



[Figure 7] Increasing trend of Unique Common Data Elements (by clinical documents)

제 3 절 메타데이터 재사용성 개선

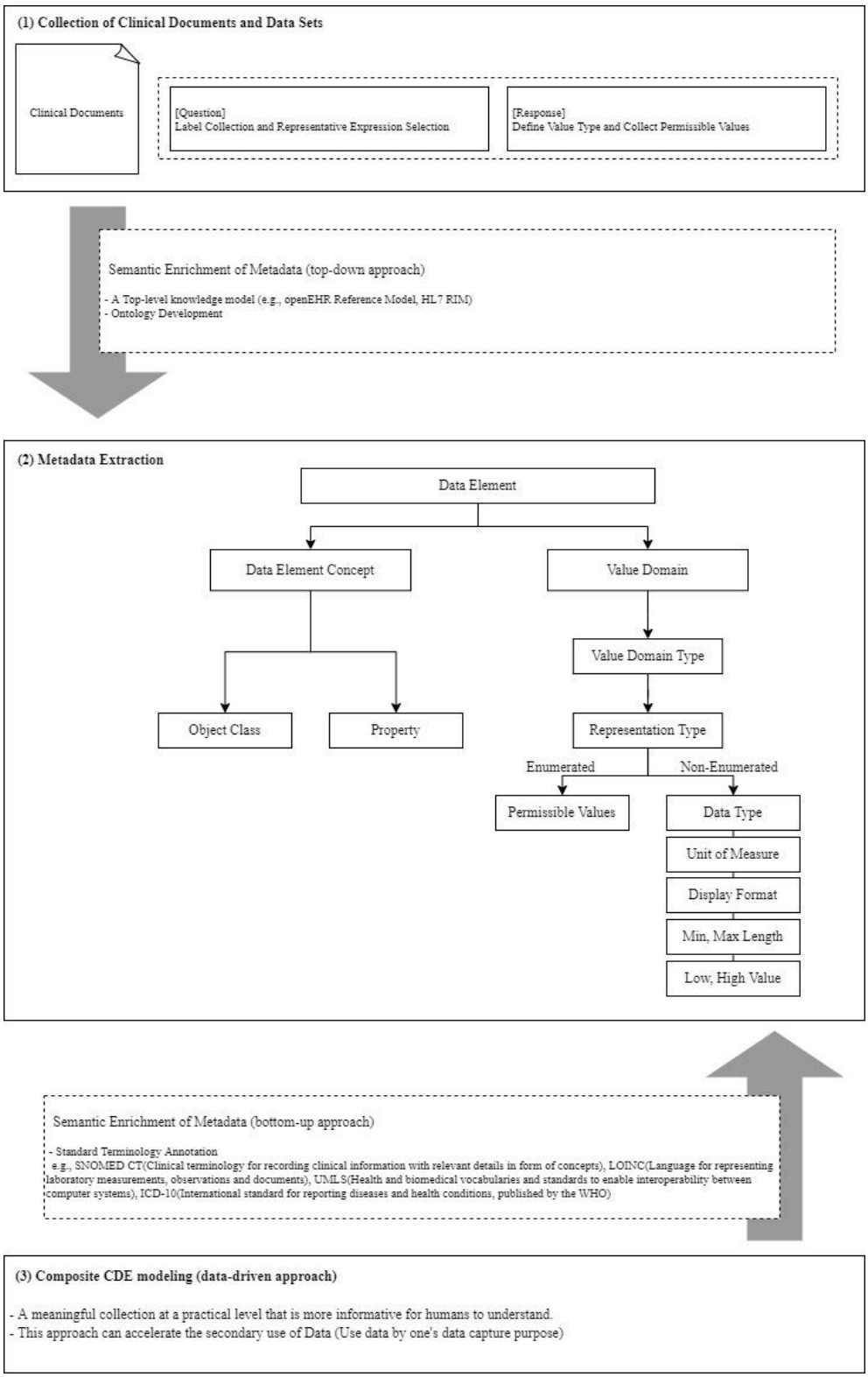
의료 데이터는 표준화된 형식을 적용하여 일관된 방식으로 수집되어야 한다. 그동안 의료 데이터의 표준화를 위한 많은 노력이 있어 왔다[7, 8, 9, 10].

예를 들어, 성공적인 데이터 교환을 보장하기 위해 OpenEHR Archetype이나 HL7 RIM과 같은 최상위 지식모델을 의료데이터에 적용시키는 하향식(Top-down) 접근방식이 있었다. 그러나 이러한 하향식 온톨로지 구축 접근 방식의 가장 큰 문제점은 의료의 모든 가능한 상황들에 이 모델을 적용시키고 사용하는 것에 굉장히 오랜 시간이 걸린다는 것이었다.

또 다른 접근 방식으로 Standard Terminology를 사용하여 의료 개념을 인코딩하도록 하는 상향식(bottom-up) 접근방식이 있다. 그러나 현실 세계의 의료 개념은 매우 정교하고 복잡하여 여러 방법으로 인코딩될 수 있고, 더 나아가 표준용어체계별 사후 조합(post-coordination) 방법론이 다양하여 모호성을 발생시킬 수 있었다.

따라서 본 연구에서는 실제 임상문서에서 추출된 ISO/IEC 11179 국제표준 기반의 CDE를 통한 상호운용성 확보를 제안하였다. 더 나아가, Object Class를 이용한 CDE의 효율적 검색 및 재사용을 가능하게 하고, 실용적인 수준에서 사용가능 하도록 Composite CDE 모델링 접근방식을 적용하였다.

Figure 8에서 (1)임상문서 수집 (2)DE 추출 (3) Composite CDE 모델링의 3단계로 이루어진 본 연구의 전체 진행과정을 확인할 수 있다.



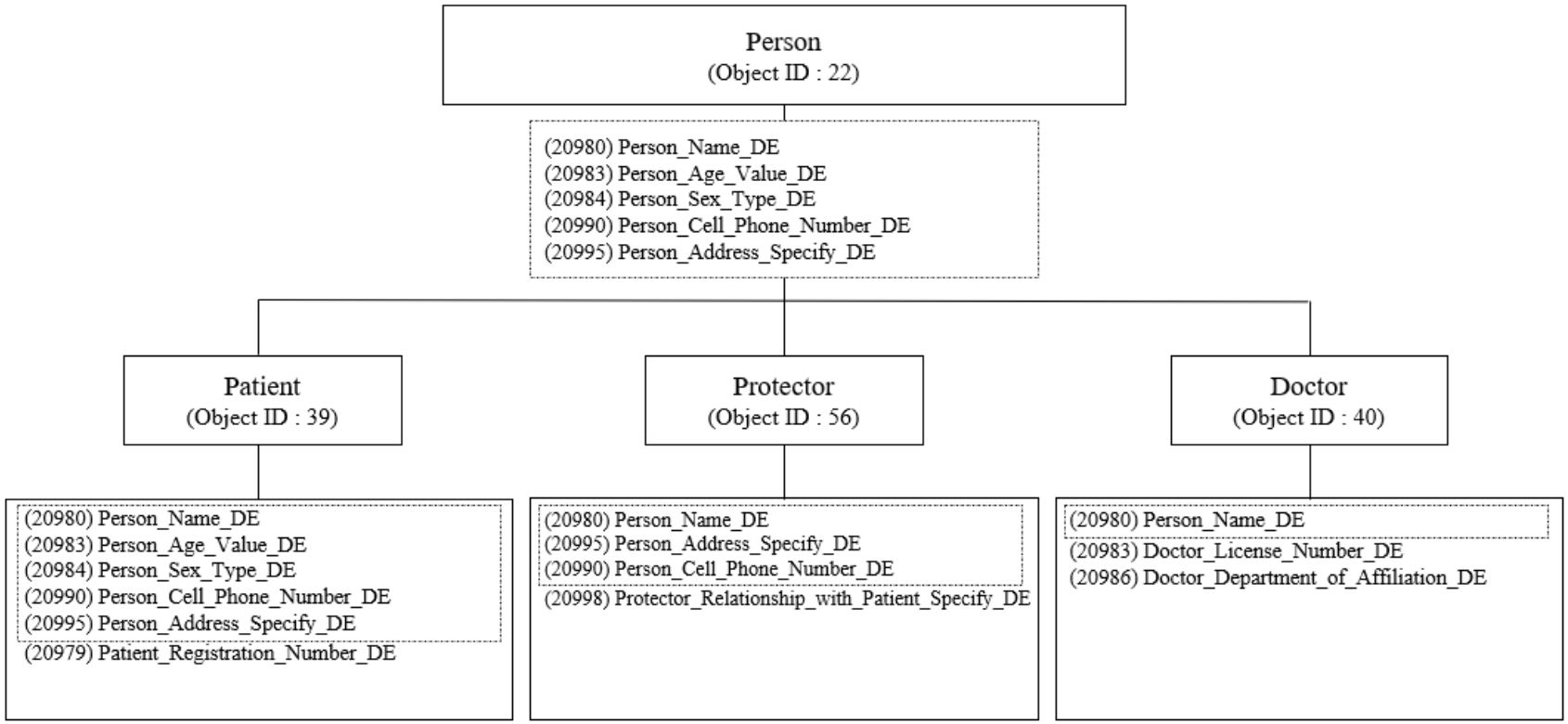
[Figure 8] Data Element Extraction process and approaches to enrich semantics

10개병원의 5개 임상문서의 항목으로부터 얻어진 848개의 Unique CDE의 효율적 검색 및 활용을 위해 Object Class를 이용한 그룹화 및 Composite CDE 모델링을 진행하였다.

Object Class를 이용한 그룹화는 ISO/IEC 11179 표준의 하위 메타데이터인 Data Element Concept(DEC)를 이용하는 방법으로, Figure 9와 Figure 10을 통해 설명될 수 있다.

Figure 9에서는 Person이라는 Object Class에 연결된 DE를 보여준다. Person(Object ID: 22)는 Patient(Object ID: 39), Protector(Object ID: 56), Doctor(Object ID: 40) 등 하위 Object Class를 가질 수 있다. 또한, 하나의 DE는 여러 개의 Object Class에 연결될 수 있다. 예를 들어, 'Person_Name_DE'는 Patient(Object ID: 39), Protector(Object ID: 56), Doctor(Object ID: 40)에서 사용된다.

각각의 DE는 Object Class에 연결되어 있기 때문에, 같은 Object Class를 가지는 DE들을 그룹화하여 관리할 수 있다. Figure 10에서는 Object Name이 "Protector (Object ID = 56)"인 DE를 찾기 위한 질의문을 제공한다. 해당 질의문을 이용하면 Protector에 해당하는 모든 DE를 얻을 수 있다. 이러한 방법은 같은 Object Class를 가지는 DE를 질의한 후, 기관별 data capture 목적에 따라 DE를 선택할 수 있어 메타데이터의 효율적 재사용을 촉진시킬 수 있다.



[Figure 9] Data Element Examples with Object Class

Query

```
SELECT
d.name as 'Object Name',
o.O_ID as 'Object ID',
group_concat(de.D_E_ID) as 'DE_id'
FROM
    data_element de,
    data_element_concept dc,
    dec_o deo,
    object o,
    designation d,
    language_section ls,
    terminological_entry te,
    ai_c aic,
    administered_item ai
WHERE de.Data_Element_Concept_D_E_C_ID = dc.D_E_C_ID
and dc.D_E_C_ID = deo.Data_Element_Concept_D_E_C_ID
and deo.Object_O_ID = o.O_ID
and o.Administration_Record_A_R_ID =  ai.Administration_Record_A_R_ID
and aic.Administered_Item_A_I_ID = ai.A_I_ID
and te.AI_C_ID = aic.AI_C_ID
and ls.Terminological_Entry_T_E_ID = te.T_E_ID
and d.Language_Section_L_S_ID = ls.L_S_ID
and o.O_ID = 56
group by o.O_ID
```

Result

Object Name	Object ID	DE id
Protector	56	20980, 20990, 20995, 20998

[Figure 10] Data Element Query with same Object Classes

임상문서 설계의 관점에서는 서로 다른 Object Class에 연결되어 있는 DE들 간의 관계 모델링이 필요할 수 있다. ‘환자의 흡연 상태’에 대한 정보를 얻고자 할 때, Past Smoking(Object ID: 175)에 연결된 DE와 Present Smoking(Object ID: 203)에 연결된 DE들의 조합을 구성하면 ‘환자의 흡연 상태’에 대한 더욱 풍부한 정보를 얻게 된다. 따라서, 다른 Object Class를 가지는 CDE들 간의 복합 관계를 설명하기 위해 Composite CDE 모델링 방법을 적용하였다.

Table 4에서는 Composite CDE 모델링의 예시를 확인할 수 있다. 예를 들어, 환자의 기본 정보를 수집하는 Patient_Demography_Composite_DE는 Patient(Object ID: 39)에 연결된 DE와 Protector(Object ID: 56)에 연결된 DE의 조합을 통해 생성되었다. 환자와 환자를 보호하는 보호자의 정보를 함께 보여주어 해당 환자에 대한 풍부한 정보를 제공할 수 있게 된다.

[Table 4] Composite Common Data Element modeling examples

Composite CDE Name (No. of aCDE)	Atomic CDE Name	Data Element Concept	Object	Property
Patient_Demography_Composite_DE (8)	Patient_Registration_Number_DE	Patient Registration Number	39: Patient	Registration Number
	Person_Name_DE	Person Name	39: Patient	Name
	Person_Age_Value_DE	Person Age	39: Patient	Age
	Person_Sex_Type_DE	Person Sex	39: Patient	Sex
	Person_Cell_Phone_Number_DE	Person Contact Phone Number	39: Patient	Phone Number
	Person_Address_Specify_DE	Person Address	39: Patient	Address
	Person_Name_Specify_DE	Person Name	56: Protector	Name
	Protector_Relationship_with_Patient_Specify_DE	Protector Relationship	56: Protector	Relationship
Smoking_Status_Composite_DE (6)	Present_Smoking_Yes_or_No_indicator_DE	Present Smoking Status	203: Present Smoking	Status
	Present_Smoking_Specify_DE	Present Smoking Status	203: Present Smoking	Description
	Past_Smoking_Yes_or_No_indicator_DE	Past Smoking Status	175: Past Smoking	Status
	Past_Smoking_Specify_DE	Past Smoking Status	175: Past Smoking	Description
	Smoking_History_Per_Day_Tobacco_Consumption_Count_DE	Past Smoking Tobacco Consumption	175: Past Smoking	Tobacco Consumption
	Smoking_History_Smoking_Period_DE	Past Smoking Period	175: Past Smoking	Period
CBC(Complete_Blood_Count)_Test_Composite_DE (12)	WBC(White_Blood_Cell)_Neutrophil_Count_DE	WBC Neutrophil Count	1237: WBC(White Blood Cell)	Count
	WBC(White_Blood_Cell)_Eosinophill_Count_DE	WBC Eosinophill Count	1237: WBC(White Blood Cell)	Count
	WBC(White_Blood_Cell)_Basophill_Count_DE	WBC Basophill Count	1237: WBC(White Blood Cell)	Count
	WBC(White_Blood_Cell)_Lymphocyte_Count_DE	WBC Lymphocyte Count	1237: WBC(White Blood Cell)	Count
	WBC(White_Blood_Cell)_Monocyte_Count_DE	WBC Monocyte Count	1237: WBC(White Blood Cell)	Count
	RBC(Red_Blood_Cell)_Count_DE	RBC Count	1535: RBC(Red Blood Cell)	Count
	Platelet_Count_DE	Platelet Count	1304: Platelet	Count
	Hemoglobin_Value_DE	Hemoglobin Value	1240: Hemoglobin	Value
	Hematocrit_Value_DE	Hematocrit Value	1302: Hematocirt	Value
	MCV(Mean_Corpuscular_Volume)_Value_DE	MCV Value	1535: RBC(Red Blood Cell)	Value
	MCH(Mean_Corpuscular_Hemoglobin)_Value_DE	MCH Value	1535: RBC(Red Blood Cell)	Value
	MCHC(Mean_Corpuscular_Hemoglobin_Concentration)_Value_DE	MCHC Value	1535: RBC(Red Blood Cell)	Value

제 4 장 논의

의료데이터의 표준화는 건강 정보를 효율적으로 관리하고 활용하기 위해 필수적이다. 국내 병원은 높은 병원정보시스템(Hospital Information System; HIS) 도입률을 보이고 있으나, 개별 병원이 각각 다르게 시스템을 구축함에 따라 데이터 표준화 수준이 미비한 실정이다. 이는 데이터의 활용적 측면을 고려하지 않은 시스템으로 데이터 품질 관리가 어렵다. 이러한 데이터 표준화 문제는 병원 별 진료정보 교류 및 통합을 어렵게 한다.

대부분의 병원들이 사용하는 임상문서의 경우, 자유입력(FREE-TEXT) 형태로 데이터를 저장하고 운영하기 때문에, 데이터의 의미를 관리하는 데 어려움이 있었다. 따라서 CHMR에서 관리하던 임상문서 도메인 항목의 기존 CDE들의 특성을 파악하여 항목을 구조화(Codify)시키는 정제 작업이 수행되었다. 이러한 방법으로 CHMR의 CDE는 점점 더 정교한 메타데이터로 개선될 수 있고, 업데이트된 CDE는 CHMR의 Version Control 기능을 통해 관리된다.

본 연구는 다양한 유형의 이질적 데이터 통합을 위한 방안을 제시하고, 실용적인 관점에서 정보 교환을 위한 표준화 기반을 마련했다는 점에서 의의가 있다. 총 10개 병원의 5개 임상문서 영역에 대해 병원이 추가될 때마다 고유하게 늘어나는 CDE 증가율을 확인한 결과(Figure 6), 초기 23.36%의 증가율에서 5.54%의 증가율로 감소하는 경향을 확인하였다. 이 결과는 5개 임상문서 도메인에서의 CDE 표준화 수준을 평가함에 있어 본 연구에서 수행된 10개 병원이 어느 정도 충분함을 의미하며, 앞으로 새로운 병원의 임상문서를 기존 CDE 데이터셋을 사용하여 표현할 때 추가적으로 생성해야 하는 신규 DE의 개수가 많지 않다는 것을 시사한다.

더 나아가, 정리된 CDE 데이터를 기반으로 실용적인 수준에서 그룹화하는 접근 방법을 취함으로써 CDE의 재사용성을 개선하고자 하였다. 이 방법을 통해 개별 항목 수준에서 정의된 CDE를 더욱

효율적으로 검색할 수 있게 되었고, 결과적으로 재사용을 용이하게 한다는 점을 확인하였다.

본 연구를 통해 10개 병원의 5개 임상문서 도메인에서 사용되는 CDE는 의료데이터의 의미론적 상호운용성이 확보된 핵심 메타데이터 셋이다. 이 결과는 앞으로 데이터의 의미론적 관리를 위한 임상문서를 설계할 때 도움을 줄 수 있으며, 병원 간 데이터 교환 시에도 효율적으로 사용될 수 있다.

향후에는 표준화된 메타데이터셋을 공개하여 각 기관의 연구자들이 공개된 메타데이터와 병원 서식 항목을 매핑하는 작업을 수행하면서 각 병원의 임상문서 항목에 대한 구조화 정도를 평가하고 지속적으로 모니터링하는 지표로 활용할 수 있다.

또한, 하나의 DE에 연결된 Question Label들은 SNOMED CT Korean Extension 구축 시 FSN(Fully Specified Name)에 대한 동의어(Synonym) 참조 세트로도 활용될 수 있다. Synonym은 사용자가 검색 및 디스플레이 목적으로 인터페이스에서 사용하는 용어로 한 개의 고유한 개념이 여러 개의 동의어를 가질 수 있다.

본 연구의 “Name_of_Recording_Doctor_DE(DE-id: 21013)”는 다른 서식에서 여러번 재사용되었는데, 각각 ‘작성의’, ‘기록자’, ‘작성자’, ‘평가자’, ‘교수’, ‘기록자명’의 Question Label 값을 가지고 있다. Question Label은 CHMR을 통해 서식을 생성할 때마다 수집되기 때문에, 하나의 의료 데이터(DE)에 대한 여러 개의 동의어가 자동으로 연결되어 있다. 따라서 하나의 개념에 대한 동의어 목록을 쉽게 얻을 수 있다.

5개의 임상문서 영역에서의 다기관 데이터 통합 방법론을 제시한 본 연구를 통해 의료데이터 관리가 필수적인 다양한 영역에서 ISO/IEC 11179 MDR 표준 기반의 CDE를 이용한 의미 있는 정보 교류가 가능해질 것이다.

참고 문헌

1. Kim HH, Park YR, Lee S, Kim JH. Composite CDE: modeling composite relationships between common data elements for representing complex clinical data. BMC Med Inform Decis Mak. 2020 Jul 3;20(1):147. doi: 10.1186/s12911-020-01168-0. PMID: 32620117; PMCID: PMC7333279.

2. Jeong S, Kim HH, Park YR, Kim JH. Clinical Data Element Ontology for Unified Indexing and Retrieval of Data Elements across Multiple Metadata Registries. Healthc Inform Res. 2014 Oct;20(4):295-303. doi: 10.4258/hir.2014.20.4.295. Epub 2014 Oct 31. PMID: 25405066; PMCID: PMC4231180.

3. de Mello BH, Rigo SJ, da Costa CA, da Rosa Righi R, Donida B, Bez MR, Schunke LC. Semantic interoperability in health records standards: a systematic literature review. Health Technol (Berl). 2022;12(2):255-272. doi: 10.1007/s12553-022-00639-w. Epub 2022 Jan 26. PMID: 35103230; PMCID: PMC8791650.

4. Ulrich H, Kock-Schoppenhauer AK, Deppenwiese N, Gött R, Kern J, Lablans M, Majeed RW, Stöhr MR, Stausberg J, Varghese J, Dugas M, Ingenerf J. Understanding the Nature of Metadata: Systematic Review. J Med Internet Res. 2022 Jan 11;24(1):e25440. doi: 10.2196/25440. PMID: 35014967; PMCID: PMC8790684.

5. https://www.iso.org/obp/ui/#iso:std:iso-iec:11179:-3:ed-3:v1:en

6. Kim HH, Park YR, Lee KH, Song YS, Kim JH. Clinical MetaData ontology: a simple classification scheme for data elements of clinical data based on semantics. BMC Med Inform Decis Mak. 2019 Aug 20;19(1):166. doi: 10.1186/s12911-019-0877-x. PMID: 31429750; PMCID: PMC6701018.

7. Leroux H, Metke-Jimenez A, Lawley MJ. Towards achieving semantic interoperability of clinical study data with FHIR. J Biomed Semantics. 2017 Sep 19;8(1):41. doi: 10.1186/s13326-017-0148-7. PMID: 28927443; PMCID: PMC5606031.

8. Kock-Schoppenhauer AK, Kroll B, Lambarki M, Ulrich H, Stahl-Toyota S, Habermann JK, Duhm-Harbeck P, Ingenerf J, Lablans M. One Step Away from Technology but One Step Towards Domain Experts-MDRBridge: A Template-Based ISO 11179-Compliant Metadata Processing Pipeline. Methods Inf Med. 2019 Dec;58(S 02):e72-e79. doi: 10.1055/s-0039-3399579. Epub 2019 Dec 18. PMID: 31853911.

9. Hegselmann S, Storck M, Gessner S, Neuhaus P, Varghese J, Bruland P, Meidt A, Mertens C, Riepenhausen S, Baier S, Stöcker B, Henke J, Schmidt CO, Dugas M. Pragmatic MDR: a metadata repository with bottom-up standardization of medical metadata through reuse. BMC Med Inform Decis Mak. 2021 May 17;21(1):160. doi: 10.1186/s12911-021-01524-8. PMID: 34001121; PMCID: PMC8130274.

10. Yang Z, Jiang K, Lou M, Gong Y, Zhang L, Liu J, Bao X, Liu D, Yang P. Defining health data elements under the HL7 development framework for metadata management. J Biomed Semantics. 2022 Mar 18;13(1):10. doi: 10.1186/s13326-022-00265-5. PMID: 35303946; PMCID: PMC8932333.

Abstract

Common Data Element modeling to achieve semantic interoperability of clinical documents

Na Young Kim
Interdisciplinary Program of Medical Informatics
The Graduate School
Seoul National University

Introduction : Data standardization is a prerequisite for the effective management of patient medical data in order to share or integrate data between institutions. Despite the high adoption rate of hospital information systems in Korea, the level of data standardization is insufficient due to the application of different systems by institutions. Prior to this study, there was a preceding study using the ISO/IEC 11179 MDR (Metadata Registry) international standard to ensure semantic interoperability of medical data. Five clinical documents–Admission Note, Initial Medical Examination Note, Discharge Summary, Emergency Note, Operation Note–were collected from five teaching hospitals in Korea, metadata was extracted, and data integration work was performed.

Methods : The goal of this study is to evaluate how well the metadata derived from pervious study can be applied to the clinical documents from new hospitals and to organize them into groups based on semantic linkages for effective management. The metadata reuse rate increased by reviewing the results of the previous study, refining the metadata more precisely in accordance with the key

concepts of the data, and applying the refined metadata to the clinical documents of five new hospitals.

Results : In total, 3,073 metadata were presented for 5 clinical records from 10 hospitals, and 848 unique metadata were sorted after duplicated items were eliminated. The propensity for unique metadata to increase as hospitals were added one at a stage was validated for the 10 hospitals in this study, which enabled us to examine the level of standardization for the five clinical documents. An increase rate of 23.36% was shown when one hospital was added, and an increase rate of 5.54% was shown when all 10 hospitals were included, indicating that the number of unique metadata was decreasing as more hospitals were added. Additionally, a strategy for effectively utilizing the 848 distinct pieces of metadata was described. From the perspective of grouping items, the ISO/IEC 11179 Object Class metamodel was used, and from the perspective of clinical document design, Composite CDE modeling was used to link relevant CDEs.

Discussion : This study organized standardized metadata that was varied in expression but had the same meaning for each hospital, allowing for the sharing and integrating of data between institutions. Further designs of standardized clinical documents may benefit from these metadata since the organized metadata have the semantic compatibility for the five clinical document domains. Additionally, as a modeling approach was suggested to enhance the reuse of well-defined metadata, it is now possible to select and choose the essential components from the relevant concept area in accordance with the requirements of the institution when generating clinical documents.

Keywords : Clinical Document, Common Data Element, Metadata Registry, Semantic interoperability
Student Number : 2021-24710