



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](#)

의학석사 학위논문

국제표준 메타데이터 기반의
서식 생성기 구현을 통한
임상문서의 의미론적
상호운영성 확보.

EasyFormBuilder : Form building
tool based on standardized metadata
to facilitate semantic interoperability

2016 년 8 월

서울대학교 대학원

의학과 의료정보학 협동과정

김 형 준

국문초록

연구 배경: 최근에, 다 기관간의 의료 및 임상 데이터의 통합과 교환에 관련된 연구가 세계적으로 활발히 이루어지고 있다. 각 기관이 저장하고 있는 데이터가 실제로는 같은 의미를 갖고 있다 하더라도 기관과 기관 간에 의미적으로 일치하는 정보를 공유하기는 쉽지 않다. 본 연구를 진행하기 전에 우리는 의미적인 정보를 포함하는 의료 데이터를 통합하기 위해서 국제표준(ISO/IEC 11179)을 따르는 MDR(Metadata registry)를 제공하고 있었다. 그리고 이 저장소를 통해 수많은 메타데이터를 공유하고 임상문서를 설계하는 프로그램도 제공하고 있었다. 메타데이터로 구성된 임상문서를 여러 임상연구에 적용하여 임상 데이터를 통합하고 기관 간 의미적인 데이터 교환을 수행하였다. 우리와 같이 국제표준을 따르는 MDR을 제공하는 기관은 해외에서도 많이 있다.

하지만 메타데이터로 구성된 임상문서를 설계하는 작업은 연구자로 하여금 여러 어려움을 주게 된다. ISO 11179 표준을 따르는 메타데이터 구조는 복잡하기 때문에, 메타데이터를 이용한 임상문서를 생성하는 작업 또한 복잡해진다. 그리고 메타데이터에 대한 구조와 의미를 연구자가 이해하고 있어야 이를 통한 서식을 설계하는 것이 가능하다.

연구 방법: 본 연구는 우리가 기존에 제공하고 있는 MDR에 추가적으로 서식 생성기 프로그램을 개발하여 메타데이터를 기반하는 임상문서 설계를 좀 더 쉽게 하고자 한다. 기존에 웹페이지에서 문서를 생성하고 문항을 하나씩 추가하는 방식을 벗어나 사용자가 정의한 CSV(comma separated value format)파일을 업로드하여 서식을 생성하도록 한다. 생성된 서식에 포함된 각 항목에 대한 메타데이터를 사용자가 쉽게 선택할 수 있도록 list를 제공한다.

연구 결과: 연구자는 생성하고자 하는 문서와 문항에 대한 정보를 CSV 파일에 입력하고 서식 생성기에 업로드한다. 서식 생성기는 업로드된

CSV 파일에서 서식과 문항에 관련된 정보를 추출하고 조합하여 1차적으로 임상문서를 생성한다.

서식 생성기는 문서를 생성한 후, 문서에 포함된 각 문항들에 대한 의미적인 정보를 부여할 수 있는 데이터요소의 목록을 제공한다. 사용자가 생성하기 위해 CSV 파일에 입력한 정보들을 추출하여, 의미적으로 일치하는 메타데이터를 MDR에서 검색한다. 그리고 검색 결과로 나오는 메타데이터의 목록을 사용자에게 보여주고 사용자는 목록에서 의미적으로 일치한다고 생각하는 메타데이터를 선택하여 문항에 쉽게 연결할 수 있다.

결론 및 논의: 이 과정을 통해 생성된 임상문서는 기존 시스템보다 쉽게 만들어지기 때문에 연구자는 문서 설계 단계에서의 부담을 덜어낼 수 있다. 그리고 우리가 MDR을 통해 주기적으로 관리하고 제공하는 정교한 메타데이터를 쉽게 임상문서에 연결할 수 있다. 잘 정의된 메타데이터를 쉽게 연결하는 것은 메타데이터의 재사용성을 증가시킬 수 있다. 아무리 정교하고 잘 정제된 메타데이터를 갖고 있더라도, MDR에서 사용자가 쉽게 찾고 연결할 수 있는 과정이 있지 않으면 사용자는 임상문서를 설계할 때마다 문항에 대한 메타데이터를 일시적, 임시적으로 정의하게 된다. 본 연구는 이런 부분에 있어서 단순한 프로그램을 고안하여 임상문서를 설계하는데 있어서의 사용성만을 증가시킨 것이 아니라, 서식 생성기를 통한 데이터의 의미적인 통합과 교환을 수월하게 이룰 수 있다는 것을 보여주고자 한다.

주요어 : 임상문서, 메타데이터, MDR, 의미적 정보 교환

학 번 : 2014-25084

목 차

제 1 장 서론	1
제 1 절 연구 배경 및 필요성	1
제 2 절 연구 목적	5
제 2 장 연구방법	5
제 1 절 연구 설계	5
제 2 절 임상문서의 구조와 메타데이터의 연관성	7
제 3 절 서식 생성기의 데이터 입력 템플릿	10
제 4 절 List of recommended CDEs	11
제 3 장 연구 결과	14
제 1 절 ‘EasyFormBuilder’ 웹페이지 구성	14
제 2 절 1차적 임상문서 생성	16
제 3 절 문서와 데이터요소의 연결	17
제 4 장 논의	20
참고문헌	22

표 목 차

Table 1. Description about the 11 columns in the CSV template.	9
---	---

그 립 목 차

Figure 1. Data element structure in ISO 11179 and example.	4
Figure 2. 하나의 문항과 데이터요소의 연결 구조	8
Figure 3. The CSV template with sample data in EasyFormBuilder	10
Figure 4. Process of making a list of recommended CDE.	12
Figure 5. Overall workflow of creating a medical form in Easy Form Builder	13
Figure 6. Screen shot of the web page for insering basic information about a clinical form.	14
Figure 7. How to upload a CSV file defined by users and result of generating a form through the data from the CSV file.	15
Figure 8. The list of recommended CDE and semi-automatic anotation	19

제 1 장 서론

제 1 절 연구 배경 및 필요성

오늘 날 의과학 분야에서 의료정보시스템은 임상문서를 사용하여 데이터를 저장한다. 하지만 대다수의 정보시스템에서 사용되는 임상문서 서식은 각 연구나 도메인마다 같은 의미나 내용을 담고 있더라도 개별적으로 다른 형태의 문서로 만들어진다. 이러한 현상은 같은 도메인이나 같은 연구 내에서도 빈번하게 일어난다[1]. 이로 인해 같은 의료정보시스템 내에서 사용되는 여러 용어가 실제로는 같은 의미를 갖더라도 용어 간의 의미가 상호 교환되지 않으며, 의료데이터가 통합하기 어려운 문제점이 발생한다.

이런 문제를 해결하기 위한 방법의 예로, 미국 국립보건원(NIH) 산하 국립암센터(NCI)에서는 ISO/IEC에서 제공하는 국제 메타데이터 표준(ISO 11179)[2]을 따르는 caDSR(Cancer Data Standards Registry and Repository)이라 불리는 메타데이터 저장소를 제공한다[3,4,5]. 이 저장소에는 CRF(Case report form)를 설계할 때 필요한 항목에 대한 데이터요소(data element)를 제공한다[5]. NCI는 이 메타데이터 저장소를 통해 의과학 연구자들에게 다수의 데이터요소를 공개하고 공유하여, 데이터요소를 통한 데이터의 의미적 상호교환을 이루고자 했다. 그리고 NCI에서는 caDSR Form Builder을 개발하여, caDSR내에 등록된 데이터요소를 이용하여 eCRF를 생성할 수 있게 하였다[6,7]. 하지만 caDSR에서 제공하는 데이터요소를 이용하여 전자 임상문서를 생성하는 과정이 쉽지 않다. 그 이유는 ISO 11179 표준에서 정의하는 메타데이터의 구조가 복잡하여, 연구자가 메타데이터 구조에 대해 어느 정도는 숙지해야 임상문서를 생

성할 수 있기 때문이다.

이 외에도, 메타데이터 저장소를 통한 임상문서 서식 생성기를 제공하는 기관이 있다. 독일 뮌스터 대학(Universität Münster)에서는 Medical-Data-Models(MDM) Portal을 개발하였고, MDM-Portal을 통해 open-access가 가능한 메타데이터 저장소를 제공한다. MDM-Portal 또한, ISO 11179 국제표준을 따르는 MDR이다. 저장소에 등록된 450,000개의 데이터요소를 이용하여 CDISC ODM(Operational Data Model)형식의 임상문서를 생성할 수 있는 문서 생성기를 제공한다[8]. 이 문서 생성기를 통해 생성된 임상문서는 MDR에 저장된 데이터요소들로 구성되어 있기 때문에, 추후에 해당 문서로 저장된 데이터를 통합하거나 통합적인 분석을 할 수 있다[9].

Figure 1에서 알 수 있듯이, ISO 11179표준은 데이터의 의미, 구문, 표현을 표준화하기 위한 프레임워크를 제시하며, 데이터요소에 대한 구조를 정의하고 있다. ISO 11179에서 데이터 요소의 구조는 데이터요소 개념(Data element concept)과 값 영역(Value domain)으로 구성된다. 데이터요소 개념은 객체(object)와 객체에 대한 속성(property)이 조합되어 있다. 값 영역은 객체의 값에 대해 정의를 하며, 값으로 허용될 수 있는 범위에 대해 정의하고 있다[2].

이와 같이, 전자 임상문서를 설계하는 과정에서 문항을 추가하기 위해서는 데이터요소 뿐만이 아닌, 데이터요소와 연결된 데이터요소 개념, 값 영역 등의 메타데이터를 함께 고려해야한다(ISO 표준에서 제시하는 데이터요소의 구조는 이 외에도 고려해야할 복잡한 항목들이 더 많이 남아 있지만, 실질적으로 저장소에 저장된 메타데이터를 이용해서 임상문서를 설계할 때는 위에서 언급한 메타데이터 항목들에 한해서만 고려하는 경우가 많다). 메타데이터에 대한 구조를 잘 모르는 연구자는 메타데이터

를 이용한 임상문서를 생성하는 과정에 진입장벽을 느낄 수 있다. 메타데이터의 구조를 잘 알고 있더라도, 하나의 문항을 생성할 때마다 문항에 관련된 데이터요소를 하나씩 검색하고 서식에 등록해야 하는 번거로운 작업이 메타데이터를 이용한 임상문서 설계의 어려운 점을 더한다.

우리는 caDSR이나 MDM-Portal과 같이, ISO 11179 국제 표준을 따르는 메타데이터 저장소 CHMR(Clinico-Histopathological Metadata Registry)을 제공한다[9]. CHMR을 통해 공유되는 메타데이터는 신장투석, 소아암 분야에서 쓰이는 용어 및 개념에 대해 정의되어 있다. 이 메타데이터는 CHMR내에서 점차적으로 관리되고 정제되어, 버전에 따라 분류되어 공유되어 왔다. 그리고 CHMR내에는 자체적인 서식 생성기가 있어 CHMR에 등록된 데이터요소들을 이용해 임상문서를 생성할 수 있었다. 하지만 caDSR Form Builder와 마찬가지로, CHMR의 서식 생성기는 임상문서를 설계할 때의 한계점을 갖고 있었다. 우리는 CHMR에 등록된 데이터요소를 이용하되, 좀 더 쉽게 임상문서를 생성할 수 있는 방안에 대해 모색하고자 한다.

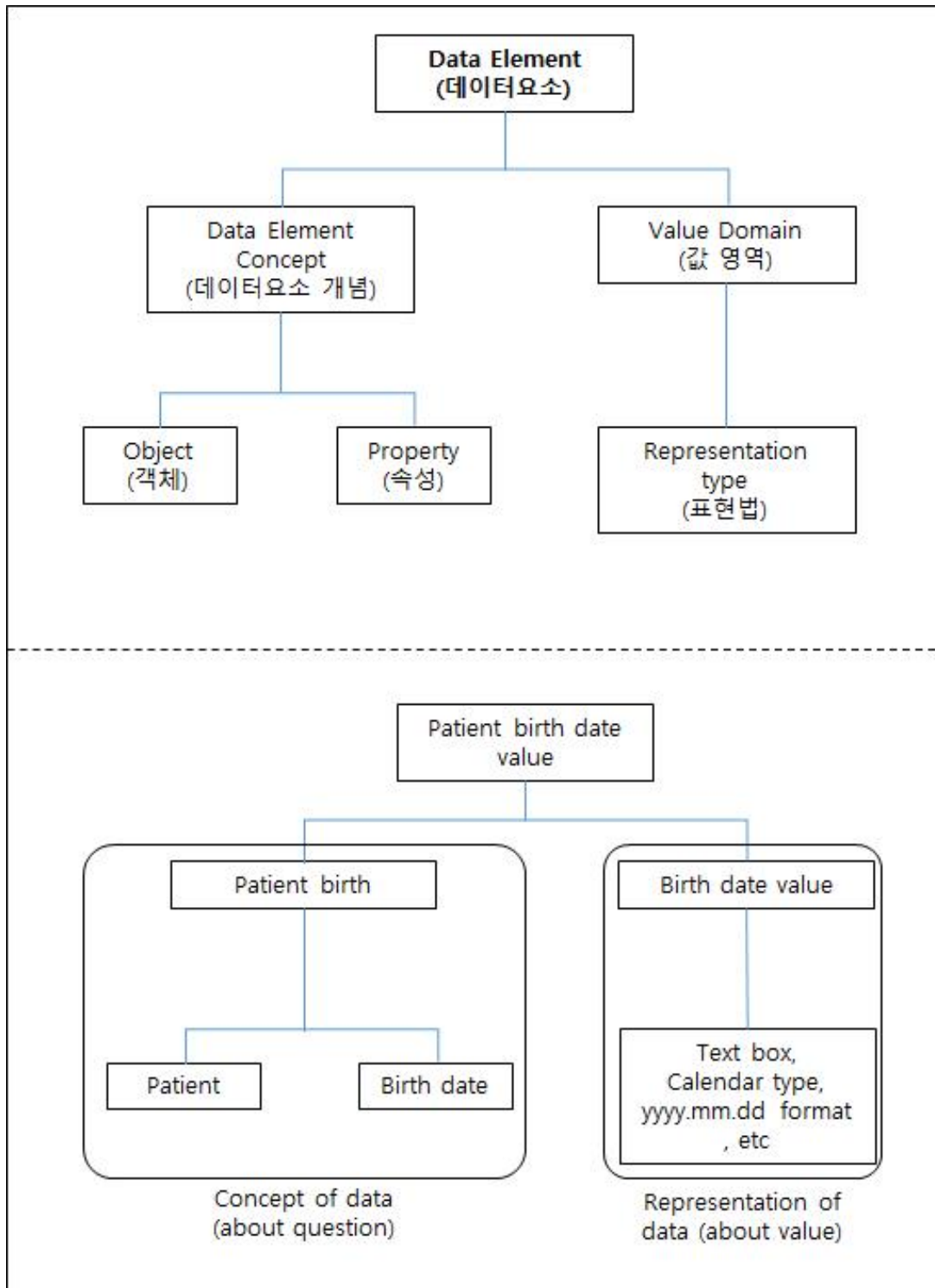


Figure 1. Data element structure in ISO 11179 and its example.

제 2 절 연구 목적

우리는 기존에 서비스하고 있었던 CHMR과 함께 Easy Form builder라는 웹 기반의 서식 생성기를 개발하려고 한다. Easy Form builder에서는 임상 연구자나 문서 설계자가 데이터요소와 같은 메타데이터의 구조를 잘 알지 못하더라도, 메타데이터를 기반으로한 문서를 쉽게 생성하는 서비스를 제공한다.

임상문서에 메타데이터를 쉽게 연결시키기 위한 연구는 해외의 많은 기관에서도 연구가 되고 있다. 여러 방법 중에서도 반자동으로 용어를 의미론적인 연결하는 방식이 최근에 화두에 오르고 있다[9].

Easy Form Builder에서는 임상 연구자가 생성하고자 하는 임상문서에 대한 메타데이터를 일일이 입력하지 않아도 반자동으로 연결할 수 있는 기능(semi-automatic annotating)을 통해 임상문서를 쉽게 만들 수 있도록 하려고 한다. 이를 통해, 현 시점에서의 메타데이터를 이용한 임상문서 생성에 대한 한계점들을 해결해보고자 한다.

제 2 장 연구 방법

제 1 절 연구 설계

Easy Form Builder가 개발되기 전에 우리는 CHMR내에 서식 생성기(Form Builder)를 자체적으로 서비스하고 있었다. CHMR Form Builder

또한 caDSR의 경우와 같이, 서론에서 언급한 임상 서식을 생성할 때의 한계점을 갖고 있다. 첫 째는, 사용자는 메타데이터 구조에 대해 알고 있어야 하고 메타데이터를 검색하는 법을 숙지해야 임상문서를 생성할 수 있었다. 또 다른 한계점은, 사용자가 메타데이터 구조를 이해하더라도 모든 임상문서에 포함될 모든 문항에 의미적으로 일치하는 데이터요소를 찾는 검색작업에서의 여러 프로세스는 사용자를 번거롭게 했다.

우리는 이전 버전의 문서 생성기를 통해 10개 이상의 research study에서 임상 문서 생성 작업을 하였다. 그 과정에서, 현재 사용자들이 복잡하고 번거롭다고 느끼는 과정에 대해 좀 더 쉽게 할 수 있는 어플리케이션을 설계했다.

기존처럼 웹페이지 상에서 서식을 생성하는 방식 대신에, 사용자가 input file을 업로드하여 서식을 생성하도록 하였다. 대부분의 임상 연구자는 CSV 형식의 파일을 다루는 것에 익숙하다. 그래서 우리는 CSV 템플릿을 제공하고, 사용자는 템플릿에 각 항목에 맞는 정보(임상문서에 등록될 여러 문항에 관련된 정보)를 입력하고 업로드하면 임상 문서를 1차적으로 생성하도록 한다.

CSV 파일이 업로드되면 파일에 입력된 정보를 추출한다. Easy Form Builder 시스템 내부에서는 추출된 정보를 현재 CHMR에 있는 메타데이터와 매칭되는 결과를 리스트로 작성한다. 사용자가 원하는 CDE를 list 내에서 선택할 수 있도록 semi-automatic annotating CDE기능을 고안해 냈다.

Easy Form Builder를 통한 서식의 생성과정은 다음과 같다.

- 1) Easy Form Builder는 문서 내에 문항에 대한 정보를 입력하기 위

한 CSV 템플릿을 제공한다. CSV 템플릿 내에 입력할 정보는 문항의 label과 값에 대한 정보로 11개의 컬럼으로 구성된다. (Section Name, Question label, Representation Type, Datatype, Value item, Dataformat, Unit of Measure, value minimum length, value maximum length, Low value, High value)

- 2) 사용자는 웹페이지에서 서식에 대한 기본정보(서식명, 서식에 대한 설명 등) 입력을 하고, 문항의 정보가 입력된 CSV파일을 업로드한다. 이 후에, EasyFormbuilder에서는 웹에서 입력된 정보와 CSV파일안의 정보를 추출하여 데이터베이스에 저장한다. 데이터베이스에 있는 데이터를 이용해서 1차적으로 서식을 생성한다.
- 3) 새로 생성된 문서에 포함된 여러 문항에 대한 정보(label and Value Domain)를 추출한다. 추출된 정보를 조합하여 MDR에서 매칭될 수 있는 데이터요소가 있는지 검색한다. 매칭되는 결과가 존재하면 해당 데이터요소를 리스트로 작성하여 웹페이지에 보여준다. (List of recommended CDEs)
- 4) 사용자는 원하는 데이터요소가 'List of recommended CDE'에 있는지 여부를 리스트에서 데이터요소와 값 영역을 본 후 선택할 수 있다. 리스트에서 하나의 데이터요소를 선택하면 기존의 문항의 값 영역과 label을 데이터요소에서 이식할 수 있다.

해당 시스템은 JSP, JQuery를 통해 웹을 구성하고, Sphinx search engine[10]을 이용하여 데이터요소를 제안하는 기능을 제공한다.

제 2 절 임상문서의 구조와 메타데이터의 연관성

서론에서 설명한대로 CHMR에 있는 데이터요소 및 메타데이터는 ISO 11179 구조를 따르며, 이 데이터요소를 이용한 문항으로 임상문서를 구

성된다[8]. 예를 들어, Figure 2.에서 ‘생년월일’이란 문항은 ‘생년월일’이란 문항의 label과 텍스트박스에 입력할 수 있는 값으로 이루어진다. 그리고 이 문항은 ‘Patient birth date value’라는 데이터요소와 연결이 되어있다. 이 데이터요소는 환자의 생년월일에 대한 개념적 정보를 데이터요소 개념(‘Patient birth date’)에 저장하고 한다. 그리고 환자의 생년월일 값에 대한 표현방식(Text box)이나 데이터 타입(string) 또는 최대/최소 값과 같은 값에 대한 제약사항을 값 영역(‘date value’)에 저장하고 있다.

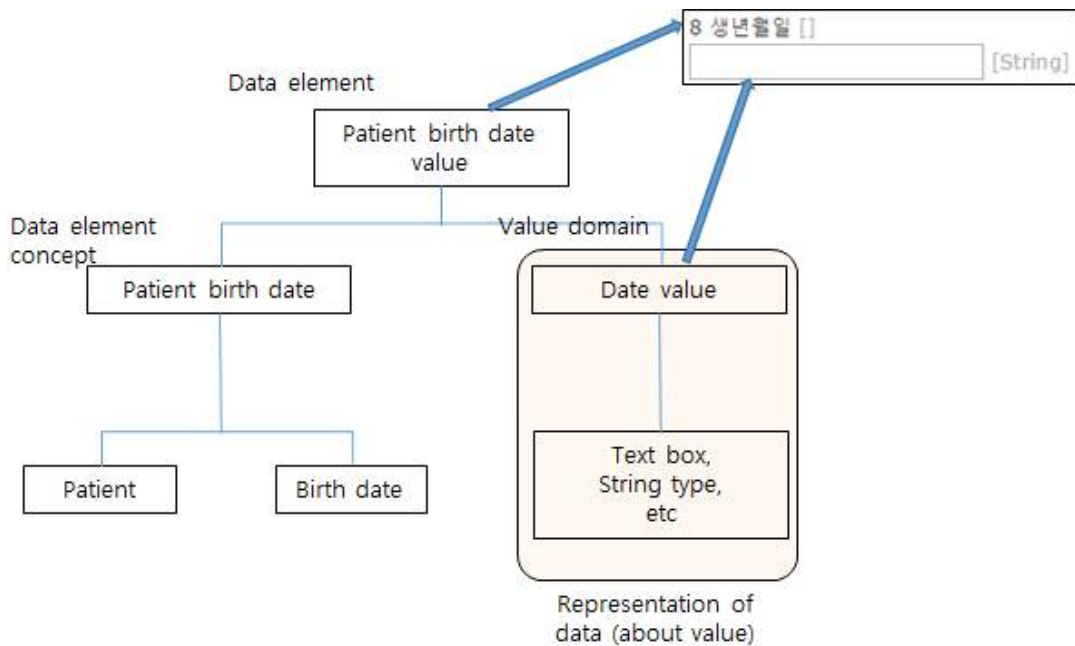


Figure 2. 하나의 문항과 데이터요소의 연결 구조

연구자는 이 데이터요소를 이용하여 ‘생년월일’ 문항의 의미론적 정보를 연결되어 있는 데이터요소 개념에서 확인할 수 있고, 생년월일에 대해 입력된 값에 대한 제약사항을 값 영역으로부터 그대로 적용할 수 있다. 즉, ‘Patient birth date value’ 데이터요소로부터 생성된 문항의 값은 Figure 2 에서 값 영역의 Text box 표현형과 string 타입의 값으로 저장된다는 것을 알 수 있다. 개념적 정보는 ‘Patient’라는 객체와 ‘Birth’라는 속성을 갖는 데이터요소 개념(‘Patient Birth date’)으로부터 알 수 있다.

Column Name (CSV)	Definition	Acceptable value	Example 1	Example2
Section Name	Naming the form's section.	Free text	Demographic	Demographic
Question Label	Naming the label of a question.	Free text	Patient age	Gender
Representation Type	Determining how to represent the question. It could be enumerated or non-enumerated.	Text Box, Text Area, Radio Box, Check Box, Select Box	Text Box	Radio Box
Datatype	Determining the type of a input value.	Number, String, Calendar, Text, Chracter, Int, Float	Number	
Value Item	Listing the permissible values of a question. If the question is enumerated, it required to be typed.	Free texts with the delimiter '^'.		Male^Female
Dataformat	Determining the format of data.	Free text	999	
Unit of Measure	Determining the unit of measure.	Free text	Years old	
Value minimum length	Defining the minimum length of the input value.	Free text	1	
V a l u e maximum length	Defining the maximum length of the input value.	Free text	3	
Low Value	Defining the the lowest input value.	Free text	0	
High Value	Defining the the lowest input value.	Free text	120	

Table 1. Description about the 11 columns in the CSV template.

제 3 절 서식 생성기의 데이터 입력 템플릿

Figure 3.에서도 보여주듯이, CSV 템플릿은 문서의 항목들을 정의하기 위한 정보를 11개의 컬럼으로 나뉘어 입력되도록 되어있다. 11개의 항목들은 생성하고자 하는 문항의 값 영역에 관련된 9개의 항목과 문항 및 섹션의 명명법에 이용될 2개의 항목들로 구성된다. 각 컬럼은 Section Name, Question label, Representation Type, Datatype, Value item, Dataformat, Unit of Measure, value minimum length, value maximum length, Low value, High value로 나뉘어진다. 그 중에서 빨간색으로 표시된 항목들은(Section name, Question label, Representation Type, value item)은 문항을 정의하기 위한 필수항목들이다. 필수항목 이외의 항목들은 선택적이지만 정의할수록 상세하게 문항의 값 영역을 정의할 수 있다. 값 영역이 명확히 정의하면 할수록 해당 문항에 대한 임상 데이터를 잘 검증할 수 있다.

Section_N	Question	Represent	Datatype	value_item	Dataformat	Unit_of_M	value_min	value_max	Low_Value	High_Valu
section1	Dialyzer	Text Box	Number							
	Dialysate	Text Box	Number							
	bruise	Radio Box		No^Yes						
	thrill?	Radio Box		No^Yes						
	cannulatic	Text Box	Number							
	Blood Flow	Text Box	Number			(ml/min)				

Figure 3. The CSV template with sample data in EasyFormBuilder

제 4 절 List of recommended CDEs

초기의 CHMR Form Builder에서는, 하나의 문항을 생성할 때 크게 두 가지의 작업이 요구된다:

1. 추가하고자 하는 문항의 의미에 상응하는 데이터요소를 CHMR에서 검색하는 작업
2. 검색된 데이터요소에 label을 지정하는 작업

이 작업은 생성하려는 임상문서의 문항 수 만큼 반복적으로 이루어져야 하는 작업이다. 그리고 하나의 데이터요소를 검색할 때 CHMR에서 한번에 원하는 검색결과를 찾기는 힘들기 때문에 여러 번의 검색작업이 필요하다. 데이터요소의 검색에 대한 문제는 임상문서 설계자에게 반복작업으로 인한 부담을 주며, 이런 부담은 사용자가 MDR에 있는 데이터요소를 재사용하기보다 새로운 데이터요소를 추가하도록 만든다.

하지만 Easy Form builder에서는 사용자가 업로드한 CSV파일을 기반으로 임상문서와 포함된 여러 문항을 동시에 생성한다. 문서와 문항이 생성되고 나면, 문서에 있는 모든 문항에 대해 의미적으로 비슷한 데이터요소의 목록(List of recommending CDEs)을 사용자에게 제공한다.

이 기능을 통해 사용자는 데이터요소 및 메타데이터에 대해 고려하지 않아도 문항을 생성할 수 있도록 하며, semi-automatic하게 사용자가 만든 문항과 의미적으로 일치하는 CDE를 연결할 수 있다.

시스템 내부적으로는, 하나의 문항에 대한 정보들을 CSV로부터 추출하고 조합하여, Sphinx search engine을 통해 조합된 정보들을 이용해서 CHMR에 저장된 데이터요소를 검색하고 결과를 목록으로 보여준다. 우리는 Easy Form Builder에서 적용될 데이터요소의 범위를 CHMR에 저

장된 모든 데이터요소가 아니다. 값 영역(value domain)이 잘 정의되어 있는 DE를 선별하여 지정하도록 한다.

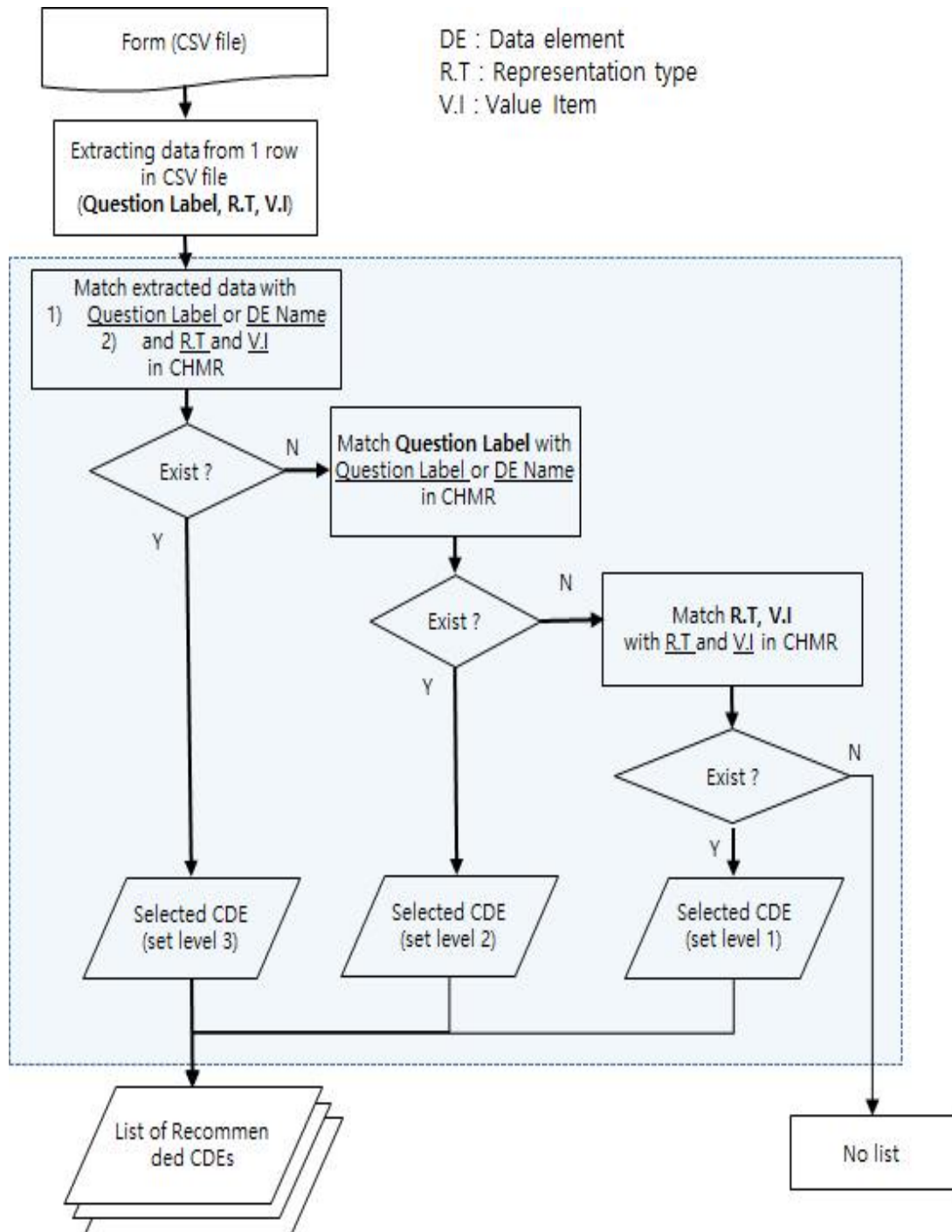


Figure 4. Process of making a list of recommended CDE.

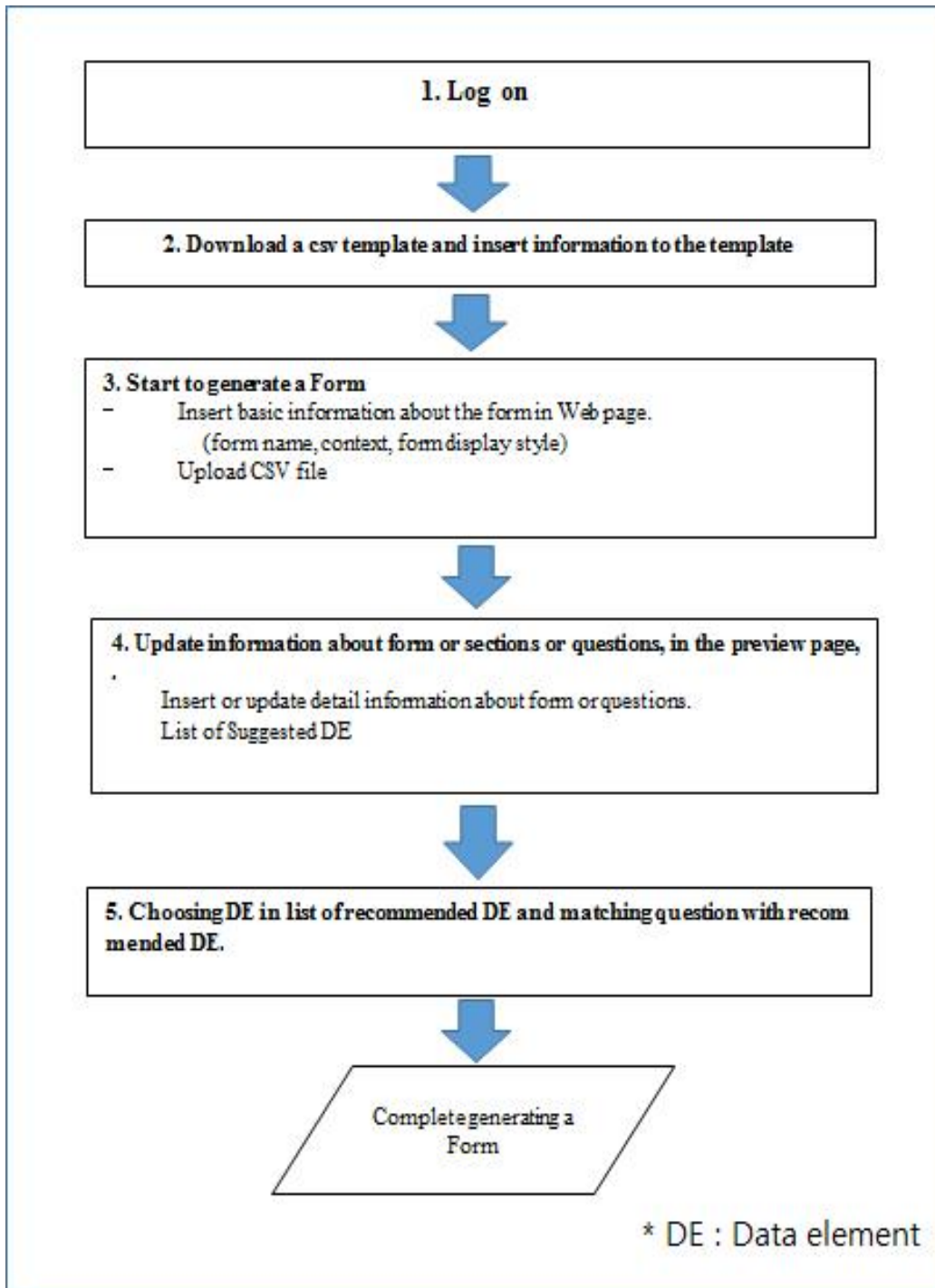


Figure 5. Overall workflow of creating a medical form in Easy Form Builder.

제 3 장 연구 결과

제 1 절 'EasyFormBuilder' 웹페이지 구성

Figure 4 에서 볼 수 있듯이, 사용자가 Easy Form builder에서 임상문서를 생성하는 전체 과정은 5과정으로 나누어지며 기존의 서식 생성기와 비교해보았을 때 복잡하지 않다. 서식을 생성하는 각 5개의 과정은 위 연구방법에서 설계한 내용을 토대로 구현을 하였고, 자세한 설명은 다음과 같다.

1. Login 시스템을 통해 Easy Form Builder 사용자는 개인별 form 생성과정을 중간 저장할 수 있으며, 개인이 만든 form에 대해서는 재사용할 수 있다.

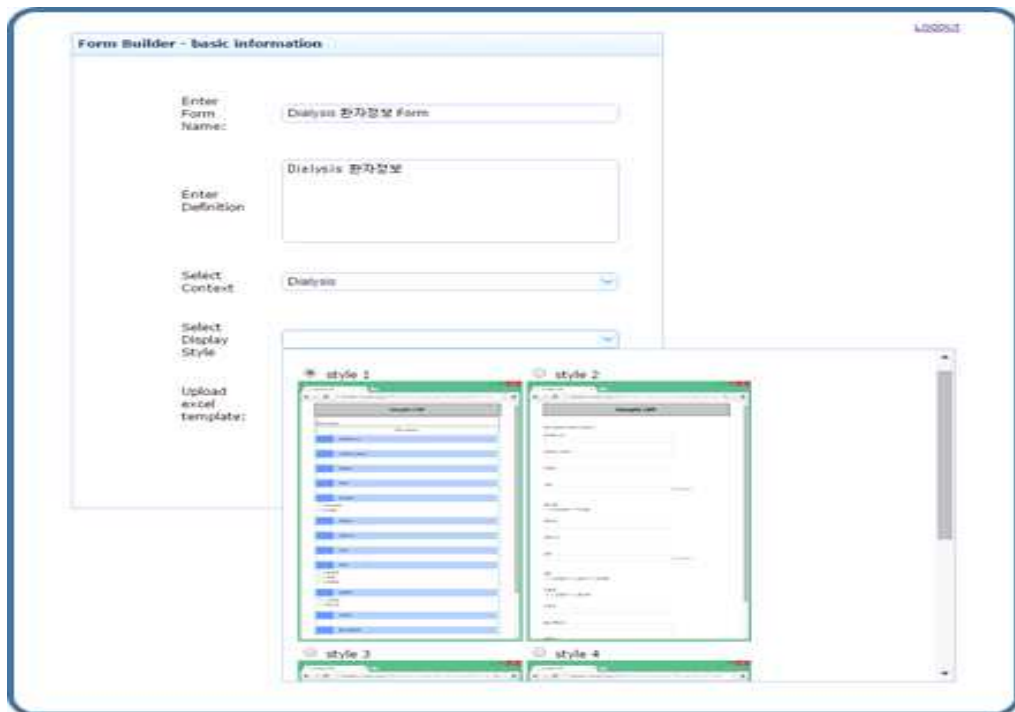


Figure 6. Screen shot of the web page for insering basic information about a clinical form.

2. 작성된 CSV파일을 Easy Form builder에 업로드하면서 생성할 문서에 대한 기본정보를 입력한다. (Form name, Form description, Context, Form display style)
3. 입력 페이지의 데이터와 CSV파일을 통해 1차적으로 만들어진 서식을 미리보기로 보여준다. 미리보기 페이지에서 사용자는 CSV 데이터에서 빠진 내용에 대해서 데이터를 채워나가면서, 최종적으로 임상문서를 완성하기 전에 세부사항을 수정할 수 있다(항목명 변경, 항목이나 섹션의 순서변경 등). 또한, 'List of recommended CDEs'를 확인해 볼 수 있다.

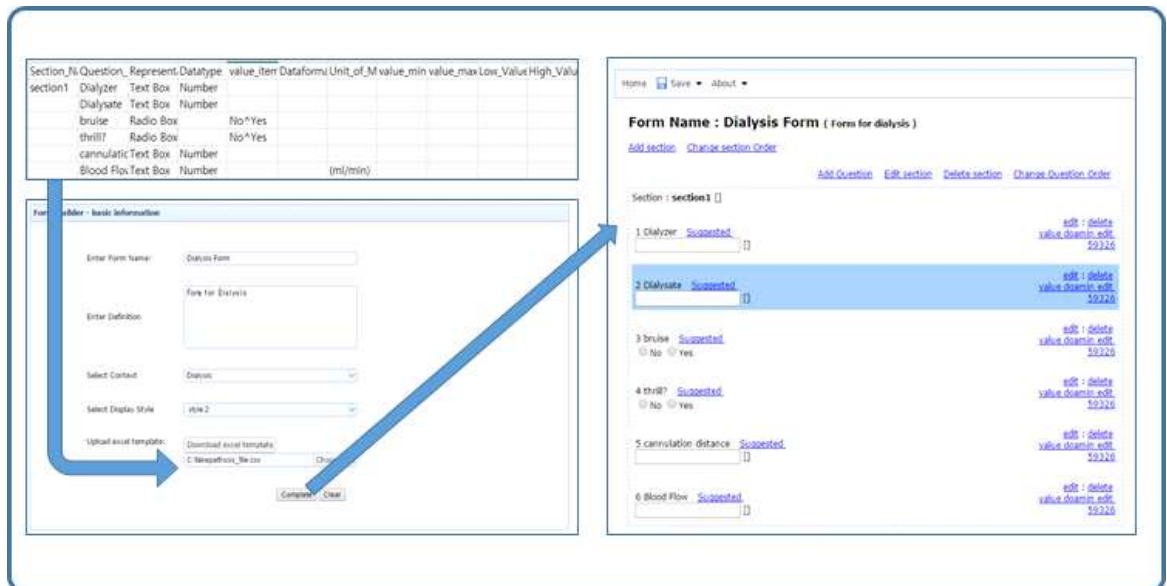


Figure 7. How to upload a CSV file defined by users and result of generating a form through the data from the CSV file.

4. [Suggested Label] 버튼을 통해, 'List of recommended CDEs'를 새 창에서 확인할 수 있다. 사용자는 생성된 문항과 의미적으로 일치하는 데이터요소들을 목록에서 확인할 수 있다. 이 목록에서 사용자는 자신이 추가한 문항의 값 영역과 추천된 데이터요소의

값 영역을 비교 확인한 후, 문항에 데이터요소를 적용시킨다.

5. 모든 수정이 수행되면 [complete]버튼을 통해 최종적으로 임상문서 생성을 완료한다.

제 2 절 1차적 임상문서 생성

CSV 파일 업로드 방식으로 문서를 생성하는 것은 웹페이지에서 문서를 생성하는 방식보다 사용자에게 여러 편의성을 준다. 그 중 하나는 임상문서에 문항을 하나씩 정의하던 방식에서 벗어나 다수의 문항을 일괄적으로 생성할 수 있다. 사용자가 CSV 파일에 여러 문항에 대한 정보를 작성하면, 해당 CSV로 생성된 서식과 같은 형식의 문서들을 쉽게 정의할 수 있다. 이 기능을 통해, 임상문서를 재정의할 때 생기는 반복작업에 대한 부담을 덜어줄 수 있다.

또 다른 이점으로는, CSV template에 있는 각 컬럼들은 사용자가 쉽게 이해할 수 있도록 구성되어 있어 사용자가 추가할 question 정보를 쉽게 입력할 수 있다. Figure 3.에서 볼 수 있듯이, template에는 문항을 생성할 때 필요한 최소한의 정보만을 입력하도록 되어있다. 이는 메타데이터 구조에 대해 잘 모르는 사용자도 문서를 쉽게 정의할 수 있도록 한다.

CSV template에서 ‘Section_Name’, ‘Question_Label’ 두 항목을 제외한 나머지 항목들은 문서가 생성되고 나서 문항에 대한 값의 영역(Value Domain)에 대한 정의를 하는 부분이다. 이 값 영역에 관련된 컬럼들은

ISO 11179에서 정의하는 값 영역에 포함되는 내용들로, 컬럼의 표현법만 사용자가 보기 쉽도록 변경하였다. 예를 들어, 값 영역에 대한 컬럼 중에서 ‘Value Item’은 ISO 11179에서의 ‘Permissible Value’에 대한 내용과 일치하는 것으로, 열거형 문항에서 허용될 수 있는 값이다.

값 영역이 잘 정의된 문항들로 구성된 임상문서는 저장된 데이터에 대한 validation을 잘 할 수 있다. Easy Form Builder의 사용자가 CSV 파일에 생성할 문항의 값 영역에 대한 정보를 잘 정의를 할수록 차후에 데이터에 대한 검증이 잘 이뤄질 수 있고, 분석도 잘 이뤄질 수 있다.

제 3 절 문서와 데이터요소의 반자동 연결

기존의 문서 생성방식에서는 문서에 문항을 추가하는 작업과 문항의 의미와 일치하는 데이터요소를 사용자가 직접 찾는 작업이 항상 같이 이뤄져야 했다. 그래서 메타데이터 및 데이터요소에 대해 잘 모르는 사람은 임상문서를 설계하는 과정에서 진입장벽을 느낄 수밖에 없었다. 또한, 메타데이터의 구조를 잘 알고 있는 사용자라도 하나의 문서를 만들기 위한 여러 과정이 복잡하다고 느낄 수 있다.

위에서 언급한 CSV 파일의 업로드를 통한 문서를 생성하는 기능은 사용자가 정의한 문항을 일괄적으로 쉽게 만들어 주는 과정으로, 아직 이 단계에서는 문항에 대한 의미적인 부분은 고려되지 않았다. 문항의 의미론적인 정보를 담기 위해서는 데이터요소가 연결(semantic annotation)되어야 한다. 문항의 의미적으로 일치하는 데이터요소를 사용자가 쉽게 선택하고 연결할 수 있도록 우리는 ‘제안된 데이터요소들의 목록’을 제공한다(‘List of recommend CDEs’).

‘List of recommended CDEs’을 제공하는 가장 주된 목적은 생성된 하나의 문항에 대한 DE를 찾는 작업을 덜어주면서 쉽게 DE를 반자동으로 연결하는 것이다. 이 기능을 통해, 문서를 생성하고 문항을 추가하는 작업과 생성된 여러 문항들에 대한 DE를 찾는 작업을 분리할 수 있도록 해준다. 이를 통해, 1차적으로 문서를 생성하는 과정에서의 진입장벽을 낮출 수 있다.

‘List of suggested CDEs’을 제공하는 또 다른 중요한 목적은 MDR에 저장되어 있는 잘 정의된 데이터요소와 메타데이터를 재사용하는 것을 가속화하는 것이다. 생성한 문서의 항목들이 잘 정의된 메타데이터와 연결되면 해당 문서가 다른 문서의 생성과정에서 재사용될 수 있도록 할 수 있다.

ISO 11179 표준에서 제공하는 메타데이터의 구조에는 필수적으로 정의되어야 하는 항목과 선택적으로 정의될 수 있는 항목이 있다. 잘 정의된 메타데이터(well curated metadata)란 필수 항목뿐만 아닌 선택적 항목들도 정의가 되어있고 주기적으로 정제되고 수정되어지는 메타데이터이다. 우리가 가지고 있는 MDR에는 Dialysis context, childhood cancer context내에 있는 DE가 Data element concept과 value domain이 잘 정의되어 있다. 이 두 context의 DE를 공유하고 재사용할 수 있도록 한다

CSV 파일로부터 생성된 모든 항목이 MDR에 저장된 데이터요소와 100%의 확률로 연결되기는 현실적으로 어렵다. 사용자가 CSV template에서 주어지는 각 컬럼에 대한 데이터를 잘 정의한다는 가정에 메타데이터와 문항 간 연결이 수월하게 이뤄지기 때문이다. 데이터요소와 연결되지 않은 문항은, 해당 문항에 관련된 정보가 있는 CSV 파일에서 한 줄의 데이터들을 추출하고 조합하여 임시 데이터요소를 생성한다. 우리는

사용자에게 정의하고자 하는 문항에 대한 label, 값 영역을 CSV에 명확하게 입력하도록 권고한다.

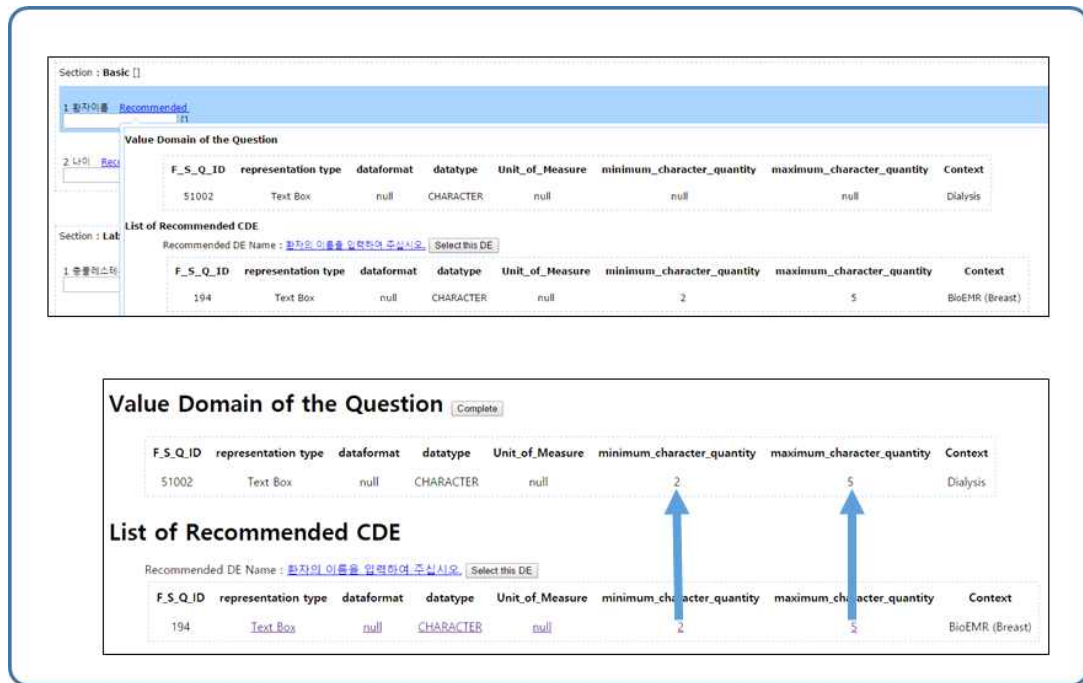


Figure 8. The list of recommended CDE and semi-automatic anotation.

Figure 8.에서는 ‘환자이름’이란 명칭을 가지는 문항에 대한 ‘List of recommended CDE’를 보여주고 있다. 먼저, 상단에는 사용자가 정의한 문항의 값 영역에 대해 보여주기 때문에 사용자는 리스트에 있는 데이터 요소들의 값 영역과 비교를 할 수 있다. [Select thie DE] 버튼을 클릭하면 제안된 데이터요소의 모든 항목을(question label, representation type, dataformat, datatype, etc.) 사용자가 정의한 문항에 적용할 수 있다. 만약 사용자가 문항의 값 영역 일부에만 적용을 하고 싶을 땐, 제안된 데이터요소 값 영역의 각 항목을 클릭하여 원하는 영역만 적용을 할 수 있다. 예를 들어, 사용자가 추가한 문항의 ‘minimum_character_quantity’와 ‘maximum_character_quantity’가 없을

경우(null)에 제안된 데이터요소의 ‘minimum_character_quantity’, ‘maximum_character_quantity’를 클릭하면 다른 값 영역은 변경하지 않고 해당 두 항목만 적용할 수 있다.

제 4 장 논의

Easy Form Builder를 통해 임상문서를 생성하면 모든 문항은 이미 MDR에 저장된 DE(데이터요소)와 연결되거나, 해당 문항에 대한 임시 데이터요소가 생성된다. 두 가지 경우 모두 정의된 데이터요소에 따라 저장된 값에 대한 검증(validation)이 가능하다. 사용자가 정의한 문항의 임시 DE도 ‘min/max value’나 ‘min/max length’, ‘permissible value’, ‘unit of measure’과 같은 값에 대한 제약사항(value domain)이 반영되어 MDR에 저장이 된다. 즉, 사용자가 CSV template에 정보를 각 항목에 빠짐없이 잘 정의할수록 임시 DE를 통해 저장된 값의 검증을 정확하게 수행할 수 있다.

Easy Form Builder에서는 CHMR에 있는 모든 DE를 공유하는 것이 아니라, 메타데이터가 잘 정의된 DE를 공유하여 반자동으로 임상문서와 연결한다. 그리고 CHMR 시스템에서는 임상문서를 다른 임상문서에 복사하는 기능이 있다. Easy Form Builder를 통해 생성되어 정교한 DE와 연결된 임상문서가 CHMR상에서 지속적으로 재사용이 이루어질 수 있다. 이런 과정이 여러 번 수행되면 정교하게 정의된 메타데이터 위주의 재사용이 활발하게 이루어지는 시스템을 갖출 수 있다. 이 부분은 Easy Form Builder를 통한 연구에서의 중요한 목표이며, 하나의 어플리케이션을 개발한 것에 국한되지 않는다는 것을 보여준다.

이 연구를 시작하는 단계에서, CSV template에 어떤 항목들로 구성할지 결정하는 것이 어려운 관건이었다. 최대한 항목을 줄여서 11개로 결정했지만, 실제로 ISO/IEC 11179 MDR 표준을 완전히 충족하는 위해서는 더 많은 항목이 필요했다. 고심한 끝에, Template의 11개 항목 'Section name', 'Question label'을 제외하고는 문항의 value domain에 대한 내용으로 구성하기로 하고, 개념적인 정보나 의미론적인 정보는 'List of recommended CDEs'에서 제공하는 DE를 연결하여 충족하기로 하였다.

현재는 사용자가 문항에 대한 DE만 반자동적인 연결을 할 수 있다. 하지만, 추후에는 문서에 포함된 콘텐츠를 분석하고 CHMR상에 저장되어 있는 문서들과 비교해서 문서자체를 제안 및 추천하여 재사용할 수 있도록 개발하고자 한다.

우리가 MDR에 저장된 것 중에서 신장내과와 소아암에 관련된 범위에서 DE를 공개하는 것으로 시작하였다. CHMR에는 이 두 도메인 이외의 데이터요소를 20000개 이상 저장하고 관리하고 있다. DE를 정제하고 수정하는 작업이 완료되면, 다른 context내에서의 DE도 쉽게 연결할 수 있을 것이다.

추후에는, 연구자들이 실제 임상연구에 진행할 서식을 Easy Form Builder를 통해 생성하고 이전 버전의 서식 생성기에 비교해서 피드백을 받을 수 있도록 한다. 그 과정에서 현장에서의 사용자들을 통해 Easy Form Builder의 장단점을 파악해 장점을 높이고 단점을 보완할 수 있도록 할 수 있다.

참 고 문 헌

- [1] Establishing semantic interoperability in the course of clinical document exchange using international standard for metadata registry (J Korean Med Assoc 2012)
- [2] http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=61932
- [3] Stausberg, Jürgen, et al. "Foundations of a metadata repository for databases of registers and trials." MIE. 2009.
- [4] Warzel, Denise B., et al. "Common data element (CDE) management and deployment in clinical trials." AMIA annual symposium proceedings. Vol. 2003. American Medical Informatics Association, 2003.
- [5] Richesson, Rachel L., and Prakash Nadkarni. "Data standards for clinical research data collection forms: current status and challenges." Journal of the American Medical Informatics Association 18.3 (2011): 341-346.
- [6] Breil, B., et al. "Multilingual Medical Data Models in ODM Format. A novel form-based approach to semantic interoperability between routine healthcare and clinical research." Appl Clin Inf 3 (2012): 276-289.
- [7] <https://wiki.nci.nih.gov/display/caDSR/Form+Builder+FAQs>
- [8] Dugas, Martin, et al. "Portal of medical data models: information infrastructure for medical research and healthcare." Database 2016 (2016): bav121.
- [9] Park, Yu Rang, and Ju Han Kim. "Metadata registry and management system based on ISO 11179 for Cancer Clinical Trials Information System." AMIA. 2006.
- [10] <http://sphinxsearch.com/docs/current.html>

Abstract

EasyFormBuilder: Form building tool based on standardized metadata to facilitate semantic interoperability.

Hyeong Joon Kim

Interdisciplinary Program of Medical Informatics

The Graduate School

Seoul National University

Introduction: When researchers or clinicians add a question using the ISO/IEC 11179 standardized metadata, called CDE in the medical form, there are limitations. This is due to the fact that metadata structure of ISO/IEC 11179 is difficult for beginners to understand, and is difficult for even form designers to retrieve CDEs from the metadata registry.

Methods: We created a number of eCRFs for the several clinical research studies. In this process, we designed an application that makes users easy to generate eCRFs. This form building application has the following functions: 1) Users can input information about a form and questions to given CSV(comma separated value format) template. A medical form is generated by uploading the CSV file. 2) The application extract information about a form and questions from user defined CSV file, and make a 'list of recommend DE(data element)s' of each question defined in CSV. Researchers can select a data element about each question from the list of recommended data elements

Results: We developed a tool, named Easy Form Builder, to overcome the limitations mentioned above to generate a form without metadata search process, and match semi-automatically the CDEs with the questions in the form. Many physicians and researchers are used to manage information. So, defining questions through the way of filling out and uploading CSV file may make the form building task easy. Easy Form Builder provides the CSV template file required to get information about the questions in a form. After generating form, the tool provides a list of suggested CDEs, which make users easy to retrieve and map CDEs on the user-defined questions. These processes facilitate not only usability of form building task but also consistent semantics interoperability in clinical data.

Discussion: The scope of sharing CDE is beginning within dialysis and childhood care domains. Once curating the CDEs of more than 20000 stored in the CHMR is completed, the CDEs in other domains

can be on the list of suggesting CDEs and those are able to be shared in the Easy Form Builder.

keywords : Medical meta data, Common data element, Metadata registry and repository, electronic Case Report Form(eCRF), Form Building

Student Number : 2014-25084