

GS1-128기반 생물자원 바코드체계 설계

주민석*, 김대성*, 한복기*
*질병관리본부 국립보건연구원
e-mail:msc202@ngri.go.kr

A Design of the Biological Resources Code based on GS1-128

Minseok Chu*, Daesung Kim*, B.G. Han*
*National Institute of Health, KCDC

요 약

현대 사회의 자동인식의 기술은 매우 발전해 있음에도 보건의료와 관련한 자동인식 기술의 접목은 아직까지 다른 산업기술 전반에 미치지 못하고 있는 실정이다. 특히 보건의료분야 연구에 기반요소로 사용되는 생물자원은 정확성에 대한 중요도가 매우 높음에도 적합한 코드체계와 자동인식기술 연구가 미비한 실정이다. 생물자원의 신뢰성과 정확성은 국제적인 유통물류 바코드 표준인 GS1-128을 적용·확장하여 바코드체계를 설계하고 자동인식기술을 연계하여 발전된 정보환경을 만들 수 있다. 본 연구에서는 생물자원 자동인식 기반을 제공하기 위하여 필요한 바코드 요구사항을 정의하고 GS1-128을 기반으로 한 생물자원 바코드체계를 제시하였으며, 동일한 바코드문자의 표현영역에 보다 많은 정보를 기록하면서도 작은 면적에 표현 가능한 바코드체계를 설계한 것이다.

1. 서론

보건의료분야에서는 제조 수입되는 의약품이나 의료기기에 대해서 바코드를 반드시 표시하도록 대다수의 국가에서 법으로 규정하고 있으며, 최근 들어 부가정보나 고부가 서비스를 제공할 수 있도록 다양한 정보를 바코드에 표시하기 위한 노력이 여러 기관에서 진행되고 있다. 유럽아시아(EAN)와 북미(UCC)로 나뉘어 있는 유통물류 정보표준 관련 기구가 통합되어진 GS1은 HUG(Healthcare User Group)을 구성해 보건의료 정보의 자동식별과 환자의 안전증진 위한 바코드체계를 연구하고 있으며, 미국의 HIBCC(Health Industry Business Communications Council), ABC(American Blood Commission), 일본의 JFMDA(Japan Federation of Medical Devices Associations) 등 각계 각국의 보건의료관련 중요기관들은 의약품, 의료기기, 환자, 장기 및 조직 등 다양한 형태의 보건의료 객체에 대한 바코드체계를 연구하고 표준화하고 있다.

그러나, 현재까지 연구 및 표준화되어 있는 바코

드 형태는 인간, 동·식물, 병원체 등으로부터 추출되어 연구목적으로 사용되는 생물자원의 정보를 표현하고 활용되기에는 부족한 점이 많아 새로운 형태의 바코드 체계 수립이 요구되고 있다.

본 논문의 내용은 GS1-128[1] 바코드를 확장하여 생물자원을 활용하는데 필요한 필수정보를 제공할 수 있도록 요구사항을 분석하고, 바코드체계를 설계한 내용에 관한 것이다. 제 2장에서는 생물자원을 포괄하는 분야인 보건의료산업에서 활용되고 있는 바코드를 분석하였고, 제 3장에서는 생물자원 관련 요구사항을 수렴한 GS1-128기반 생물자원 코드를 설계하였으며, 제 4장에서는 연구 결과와 추후 연구사항에 대하여 다루었다.

2. 보건의료 바코드 분석

보편적으로 보건의료분야에서 사용되는 개체식별 체계는 EAN-13[2]과 UPC[2]가 대표적이다. 그러나 EAN-13이나 UPC의 경우는 유통되는 상품에 대해서 각각의 포장물이 무엇인지 식별하기 위한 식별정

보를 의미할 뿐 개별 객체에 대한 유일한 식별이 불가능하고 객체에 대한 추가적인 정보를 포함하지 않는 단점이 있다. 그 대안으로 GS1은 업태, 업종에 관계없이 각 업체간 또는 업체 내부에서 물류관리를 위해서 요구되는 다양한 정보를 표시하기 위해서 GS1-128 표준 코드를 제공하였다. GS1-128 코드는 상품의 이동, 추적, 보관, 생산관리 등에서 요구되는 다양한 정보를 AI(Application Identifier)를 사용하여 요구되는 다양한 정보를 담는다. 만일 생물개체를 위해 요구되는 식별정보를 위한 AI와 함께 생물개체 코드를 기술할 수 있다면 보편화하여 사용할 수 있는 국제표준바코드로서 사용할 수 있다. 그러나 GS1-128은 산업진반에 걸쳐 사용될 수 있는 보편적인 식별자를 위하여 AI를 제공하고 있으나 전체 산업이 만족하는 무수히 많은 AI를 제공하는 것은 불가능하며, GS1-128코드가 일반화된 표준으로 존재하는 대신 세부 산업분야에서의 정의는 그 특징에 따라 구체화하여 관리되어야 할 것이다. 만일 이러한 GS1-128 디자인의 기본 철학을 지키지 않고 유비쿼터스 컴퓨팅 환경과 같이 매우 다이나믹하고 분산된 환경에서 사용되어지면 코드체계의 일관성과 시스템간 상호운영성과 같은 또 다른 큰 문제를 야기할 수 있다.

이러한 점은 생물개체를 위한 바코드체계를 위해서 EAN-128을 보완하는 새로운 코드체계가 미래의 유비쿼터스 컴퓨팅 응용프로그램을 위해서 필요하다라는 뜻을 내포하고 있다. 생물개체의 GS1-128에 대한 활용은 독립적 산업 표준 기관에서 생물개체 바코드 체계에 대한 활용 지침을 제시하고 표준코드 체계에 대한 세부 정의가 이루어져야 한다. 또한 다양한 기관이 요구하는 생물개체 표현에 대한 세부정보가 다양하기 때문에 효율화를 갖춘 코드체계 알고리즘 디자인이 되어야 한다.

ICCBBA(International Council for Commonality in Blood Banking Automation)의 ISBT128은 Code128을 보완한 코드체계로 6가지의 식별코드를 조합해 혈액 및 혈액 구성요소에 대해 식별하는데 사용되고 있다[3]. 식별자의 종류에 따라 코드 값을 여러 가지 의미로 사용할 수 있는 코드 재사용성이 우수하며 유일키 값을 사용해 식별정보에 대한 유일성을 보장한다. 그러나 이 코드체계는 혈액 유통을 위해 설계된 코드이기 때문에 연구목적으로 사용되는 대다수의 생물개체가 보관되는 매우 작은 크기의 용기에는 표현이 불가능한 코드체계 및 심볼로지

구성되어 생물개체를 위해 일반적으로 사용하기 어려운 코드체계이다.

미 보건의료분야에서는 HIBCC의 주도하에 HIBC SLS(Supplier Labeling Standard)[4]와 HIBC PAS(Provider Application Standard)[5] 두 가지 표준 바코드체계를 제안하였다. 이 바코드는 ISBT128이 혈액유통을 위해 고안된 것과는 달리 대다수의 보건의료 분야에서 사용될 수 있게 고안되었다. 특히 HIBC PAS 표준체계는 'Where'의 개념으로 어디에서 사용되는지를 정의하고 'What'의 개념으로 어떤 객체인지를 정의함으로써 의약품이나 의료기구에 대한 식별 뿐 아니라 환자식별, 생물개체식별 등을 가능하게 하는 범용적인 코드체계이다. 그러나 이 코드체계는 범국가적으로 사용할 수 없게 설계되었다는 단점을 안고 있어 한국과 같이 GS1의 EAN-13코드 GS1-128 코드 등을 국가표준으로 인정하고 있는 국가에서는 정보 간 상호 연계성을 갖기 힘들다.

그밖에 보건의료에서 사용되는 장기 및 조직, 혈관계 조직 등의 경우 GS1-128 코드체계가 사용되고 있기는 하지만 소형 용기에는 바코드를 표현할 수 없어 비교적 큰 형태의 포장물 단위에만 활용하고 있는 것으로 보고되고 있다[6].

3. 생물자원코드 설계

3.1 생물자원코드 요구사항

생물자원을 위한 바코드의 설계는 자원의 생성에서 소멸에 이르기까지 변하지 않는 정적정보와 기관간의 자원 교류에 있어서 필수적인 사항이 바코드 심볼로지에 인쇄되고 판독하기 위한 코드 체계로 정의되어야 하며, 생물자원에 대한 필수적이고 일반적인 속성을 도출하기 위해 아래와 같은 필수적인 요구사항이 정의된다.

- 유일성(uniqueness): 식별코드는 오직 한 객체에 할당되어야 하고 재사용되어서는 안된다.
- 국제성(international): 식별시스템은 국제표준에 의해 지정되어야 한다.
- 중립성(neutral): 특정 응용프로그램에 종속되지 않고 다양한 응용프로그램이나 단체들의 공통된 식별코드로서 역할을 해야 한다.
- 영속성(persistent): 식별코드가 객체를 식별하지 않더라도 수명이 제한되어서는 안된다.
- 유연성(flexible): 식별코드는 객체를 사용하는 기관마다의 사용목적이나 식별기준이 서로 다를

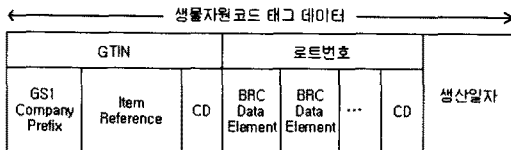
수 있으므로 식별 정보로서의 요소를 유연하게 사용할 수 있어야 한다.

- 입도(granularity) : 식별 시스템은 시료를 담은 패키지가 단위 객체 단위로 구성된 것처럼 단일항목 또는 복합항목의 의미를 표현할 수 있어야 하고 구성요소들은 단일 객체로서 정의되거나 묘사되어야 한다.
- 용량(capacity) : 식별시스템은 생물자원의 보관용기가 작은 특수성을 지녔어도 시스템의 범위 내에 정의된 객체의 용량을 위해 충분한 데이터 표현 양을 지녀야 한다.
- 부착면적(area) : 식별시스템이 식별하고자하는 생물자원이 매우 작은 보관용기에 담기는 것을 감안하여 최소공간면적에 표현이 가능한 심볼로지 기술로 표현되어야 한다.
- 메타데이터 기록 : 식별코드와 관련된 디렉토리의 발전이 요구된다. 식별하고자 하는 아이템의 식별코드를 관련된 데이터와 연결한 후 적용 가능한 식별 시스템과 연결을 통해 아이템을 식별할 수 있어야 한다.
- 다양한 이해관계 집단의 요구를 만족 : 생물자원의 공급 또는 유통에 포함된 모든 집단들이 제품에 생명 주기가 존재하는 동안 동일한 식별코드와 식별 시스템을 공유하여야 한다. (국제규모의 표준제정협회에서 인정하는 식별코드 필요)

이러한 요구사항을 만족하는 생물자원을 위한 바코드 체계는 GS1-128 코드를 확장하고, 보건의료분야에 특화시키면서도 일반화된 표준에 위배되지 않게 설계함으로써 해결이 가능하다.

3.2 데이터 구조

생물자원은 다양한 부가정보를 가지고 있지만 정보 표준화 과정에서 모든 정보의 속성을 수용하는 것은 적합하지 않다. 이런 이유에 있어서 생물자원 코드의 데이터 구조는 (그림 1)과 같이 GTIN, 로트번호(Lot Number), 생산일자를 데이터 요소로 구성하여 생물자원 정보를 나타낸다.



(그림 1) 생물자원코드 데이터 구조

GTIN은 공급망에서 상품의 정보 검색과 주문이나 계약에 필요한 상품 및 서비스 단위의 식별을 위해 사용되는 번호로서 거래 항목을 정의하기 위한 세계적인 표준이다. 이 데이터 요소는 생물자원에 대한 국제성, 중립성, 유연성, 입도의 요구사항을 만족시킬 수 있으며 범용적인 표준 데이터 식별자이므로 다양한 이해관계 집단의 요구를 만족할 수 있는 데이터 요소라고 볼 수 있다.

GTIN을 이용하여 기본적인 유통 환경을 위한 제품식별은 가능하지만 보건의료 및 생명공학분야에서 사용되는 생물자원은 단순한 제품식별정보만으로는 종종 불충분한 경우가 있다. 연구 목적으로 활용이 되는 생물자원은 자원의 생성에서 소멸에 이르기까지 자원이 담고 있는 세부내용이 더욱 중요하게 여겨지는 경우가 보편적이다. 따라서 생물개체를 위한 개체식별자, 객체의 종류 그리고 객체번호를 <표 1>과 같이 필드화하여 로트번호를 구성함으로써 기관간의 정보교류시 상호 운영성을 높일 수 있다.

<표 1> 필드데이터 구조

필드명	플래그 문자	ASCII	필드설명
개체식별자	#	35	생물개체를 유일하게 식별하기 위한 데이터 필드
객체종류	\$	36	생물체로부터 파생되어 시험, 검사, 분석에 쓰이는 시료의 종류 (예, 세포, DNA, 혈청, 혈장 등)
객체번호	!	33	생물자원을 유일하게 식별하기 위한 데이터 필드

로트번호(Lot)는 하나 또는 그 이상의 생물자원 데이터요소(B)로 구성이 된다. 생물자원 데이터요소는 <표 1>의 필드데이터를 나타내는 요소들로 식별하고자하는 생물자원의 세부항목을 표현한다. 생물자원 데이터요소는 필드데이터(F)와 데이터요소(D)의 조합으로 구성이 된다. 따라서 로트번호는 수식 (1)과 같이 정의할 수 있다.

$$\begin{aligned}
 \text{Number of BRC Data Element} & := n \\
 \text{Lot}(n) & = \text{null} \quad \text{if } n=0 \\
 & = B_1 + B_2 + \dots + B_n + CD \quad \text{if } n > 0 \quad \dots(1) \\
 B & = F + D \quad \dots(2)
 \end{aligned}$$

로트번호에서 체크문자는 마지막 자릿수를 뺀 나머지 문자들이 전산 상에서 올바르게 부여되었는지

확인하는 기능을 한다. 체크문자(CD)의 생성은 제어 문자를 포함하여 사용된 바코드문자의 합을 구한 후에 94 modulo 방법에 의한 나머지 값으로 적용하며 수식(3)에 의해 계산된다.

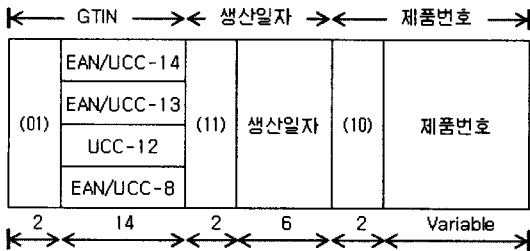
$$\text{Number of Lotnumber Character} := n$$

$$CD(n) = \left(\sum_{i=1}^n C_i \right) \text{ mod } 94 \quad \dots\dots\dots(3)$$

생물자원의 생산일자는 특정 시점을 대표하는 정보로 사용된다. 또한, 생물자원 보존·관리에 관한 안전성에 대한 연구에서는 보존방법, 온도 등과 함께 보존 기간의 정보가 중요한 정보로 사용된다. 보존 기간의 경우는 생물자원 생산일자에서부터 연구 시점까지의 정보이며, 생산일자를 통해 정보를 획득할 수 있다. 이와 같이 제품에 대한 생산일자를 나타내는데 사용되는 생산일자 코드는 년, 월, 일의 데이터를 표기한다.

3.3 데이터 표현

3.2절에서 기술한 데이터 구조는 GS1-128 코드 체계를 이용해서 (그림 2)와 같이 데이터 표현이 가능하다. GS1-128 기반의 생물자원코드는 정보형식과 내용을 이끄는 3개의 AI와 하나의 바코드 심볼로 구성된다. 이와 같이 표현된 생물자원코드는 2차원 바코드 심볼로 자원 단품에 부착하여 면적에 대한 제약조건을 해결한다.



(그림 2) GS1-128기반 생물자원코드 구조

3.4 결과분석

<표 2>에 나타낸 결과는 보건 의료분야에서 활용되고 있는 ISBT128, HIBCC와 본 논문에서 제시하고 있는 생물자원코드에 대해 비교 분석한 내용이다. 생물자원코드는 다른 보편적 코드와 비교하였을 때 같은 자릿수에서 적용문자수가 약 2~3배에 이르러 같은 면적에 표현할 수 있는 정보량이 우수하고 심볼로지를 2차원바코드를 채택하여 바코드를 부착

하기 때문에 난해한 생물자원 고유특성을 해결할 수 있다. 또한 바코드 내에 입력된 정보의 오류검출을 위한 플래그 문자는 정보의 신뢰성을 높여준다. 따라서 생물자원코드를 적용하면 보다 많은 정보의 유형을 정의하여 사용할 수 있을 뿐 아니라 식별자의 오류검출 및 교정방법도 개선되어 복잡 난해한 생물자원관리환경이 개선될 수 있을 것으로 기대된다.

<표 2> 보건의료분야 바코드 비교

구분	ISBT128	HIBCC	생물자원코드
적용범위	혈액	보건의료	생물개체
기반표준	CODE128	CODE128	GS1-128
플래그문자	modulo 37	modulo 43	modulo 94
적용문자수	37	43	94
심볼사이즈	매우큼	일반	아주작음
심볼로지	1차원바코드	1차원바코드	2차원바코드

4. 결론

본 논문에서는 생물자원을 대상객체로 하였을 때의 특수성을 고려한 코드체계를 제시하였으며, GS-128 바코드 심볼을 사용하여 제시된 코드체계를 표현하였다. 이에 따라 보건의료분야에 특화시키면서도 일반화된 표준에 위배되지 않는 생물자원을 위한 자동인식 기반환경을 제공할 수 있게 되었다.

향후 과제로는 제시된 생물자원코드를 활용한 자동인식 시스템을 설계 구현하여 실제 운영환경에 바코드 정보체계를 적용하기 위한 노력이 요구되고 있다.

참고문헌

- [1] "GS1 General Specifications" V7.0 GS1
- [2] "Information technology - Automatic identification and data capture techniques - Bar code symbology specification - EAN/UPC" ISO/IEC
- [3] Edwin Steane "United States Industry Consensus Standard for the Uniform Labeling of Blood and Blood Components Using ISBT 128" V1.2.0 ICCBBA
- [4] "The Health Industry Bar Code Supplier Labeling Standard" ANSI
- [5] "The Health Industry Bar Code Provider Application Standard" ANSI
- [6] "Healthcare Current Situation" GS1 HUG