

# 보건의료분야에서의 유전자원 코드체계 연구

김대성\*, 주민석\*, 김기상\*, 김홍태\*, 한복기\*, 정재두\*\*  
\*질병관리본부 국립보건연구원, \*\*육군본부  
e-mail:kds@ngri.go.kr

## A Study of Genetic Resource Codes for Healthcare

Daesung Kim\*, Minseok Chu\*, Ki Sang Kim\*,  
Hung Tae Kim\*, B.G. Han\*, CHUNG JAE DU\*\*  
\*National Institute of Health, KCDC, \*\*ROK

### 요 약

최근 인간 유전체 연구에 대한 관심이 증대되어 유전자원 정보에 대한 효율적 활용을 위한 정보 수집 및 관리 활동이 지속적으로 이루어지고 있다. 그러나 현재까지 연구된 보건의료분야 표준 코드체계들은 유전자원 정보 식별을 위한 적합한 표현방법을 고려하지 않아 유전자원 정보관리에 적용하기 어려운 문제점을 갖고 있으며, 현재 사용되는 식별체계 또한 기관에 따라 자체적으로 운영되고 있는 것이 현실이다. 따라서 이 논문에서는 보건의료 정보 관리에 적용되는 국내·외 표준 코드체계를 분석하고 유전자원 표준 코드체계 설계를 위한 정보 요구사항 및 기본방향을 제시하였다.

### 1. 서론

인간 게놈 프로젝트의 완성으로 인간 유전체에 대한 모든 정보가 밝혀지면서 유전체 정보를 이용한 연구는 세계 여러 나라에서 막대한 투자와 연구 역량을 집중하고 있는 분야가 되고 있다. 또한 생명과학 분야의 새로운 패러다임인 데이터 기반의 연구가 유전체 연구의 성패를 좌우함에 따라 양질의 유전자원과 자원정보의 확보가 이슈화 되고 있다. 그러나 자원 정보화를 위한 정보 수집 및 처리 기반의 취약과 표준 코드체계의 부재로 정보의 신뢰성 및 상호운용에 상당한 제약이 따르고 있는 실정이다. 또한, 연구기관간의 일관되지 않은 정보 연계로 연구 분야의 다양화에 비해 정보공유 및 관리의 효율성이 심각하게 저하된 상황이다.

현재, 국내 보건의료기관은 유전자원에 대해 기관 자체 식별체계, 또는 단순 일련번호 체계의 사용을 통해 제한적 정보 공유가 진행되고 있으나 연구영역의 확대에 따라 자원 정보의 교환 및 공유가 점차 주된 문제로 대두되고 있다. 따라서 적절한 유전자원 표준 코드체계를 설계하여 자원 정보의 명확한

식별 및 정보교환에 대한 상호 운용성을 보장하는 것이 시급한 과제라 판단된다.[1]

유전자원은 생물 개체가 갖는 정보 객체로서 세포, 혈액, 조직, 체액, 및 뇨 등과 이로부터 추출한 DNA등을 지칭할 수 있다. 이러한 유전자원은 다량, 다형의 정보를 갖고 있으며 신뢰성 있는 정보 획득 및 처리를 위한 방법론이 요구된다. 물리적 자원의 자동식별 및 데이터 수집을 위한 대표적 기술은 바코드, RFID, 스마트 태그, 자기 카드 등을 들 수 있으며 최근 들어서는 RFID기반의 유비쿼터스 지향형 사물식별 연구가 자동인식분야 연구의 주류를 이루고 있다. 현재 유통, 물류에서 일반적으로 사용되는 바코드는 자원에 대한 개별 식별이 어려운 코드체계와 정보 표현에 대한 제약을 갖고 있다. 또한 RFID는 자원 객체가 관리되는 초저온 환경(-192도)의 특수성으로 인해 현존하는 전자 태그들은 초저온 환경을 극복하지 못하는 문제를 안고 있는 실정이다. 따라서 이러한 제약사항을 고려한 코드체계의 설계가 필요하다.

이 논문은 2장 관련연구에서 보건의료분야에 적

용되는 국내·외 바코드 표준 코드체계의 문제점을 분석하고 3장에서는 자원정보 요구사항을 기술하고 바코드 코드체계 설계를 위한 기본방향을 제시한다. 4장에서는 유전자원 표준 코드체계를 고려한 향후연구에 대해 기술하고, 마지막으로 결론을 맺는다.

2. 관련연구

2.1 GS1

GS1은 유통, 물류 등 다양한 분야에 표준코드 체계를 연구하고 있으며 각 시스템의 특성에 적합한 코드 표현방법을 제공하고 있다. 특히, 보건의료분야 표준 코드체계를 확립하기 위해 HUG(Healthcare User Group)를 구성해 정보의 자동식별과 환자의 안전증진을 위한 표준 코드체계를 연구하고 있다.[2] 현재 보건의료분야에 사용되는 GS1 Version 7의 VSHI(Very Small Healthcare Items) 코드체계는 GTIN, 로트번호, 유효기간으로 구성되며 세부 코드 구조는 <표 1>과 같이 표현한다.

<표 1> VSHI 코드 구조

용용식별자	코드구성요소		
01	물류식별자	GS1 Company Prefix + Item Reference	체크문자
17	유효기간		
10	로트번호		

- GTIN : 물류단위, 제조국가, 제품, 체크문자로 구성되며 VSHI를 식별할 때 부여하는 필수정보이다.
- 유효기간 : 연도, 월, 일의 단위로 구성되며 사용자 요구에 따라 선택적으로 부여한다.
- 로트번호 : 생산 및 유통과정에서 제품을 식별하며 유효기간과 마찬가지로 선택적으로 부여한다.

VSHI 코드체계는 RSS-14 Stacked, RSS Limited, Data Matrix와 같은 2차원 심볼을 통한 정보 표현 방법을 제공해 보건의료 제품의 공간적 제약을 해결하고 있다. 하지만 유전자원 식별시 필수 요구사항인 자원 개별식별 및 모자관계 식별을 위한 정보표현 방법을 제공하고 있지 않아 자원식별의 문제점을 갖고 있으며 자원의 유통, 물류에 대한 정보이외에 부가 정보를 표현할 수 없다.

2.2 ICCBBA

ICCBBA(International Council for Commonality in Blood Banking Automation)는 혈액 및 혈액 구

성요소에 대해 신뢰성 있는 정보수집과 수집 및 관리 분야의 자동식별을 위해 ISBT(International Society of Blood Transfusion) Working Group을 구성해 혈액관리 분야의 코드체계를 연구하고 있다. 현재 혈액관리에 사용되는 ISBT 128 Version 2 코드체계는 6가지의 식별코드를 조합해 혈액 및 혈액 구성요소에 대해 식별하며 세부 코드 구조는 <표 2>와 같이 표현한다.[3]

<표 2> ISBT 128 코드 구조

식별코드명	코드구성요소				
기탁식별번호	데이터 식별자	ICCBBA 승인코드	제품 수집일	일련 번호	플래그 문자
혈액그룹	데이터 식별자	혈액형	표현형	예약문자	
제품코드	데이터 식별자	제품 기술코드	기탁유형	대분류 코드	소분류 코드
유효기간	데이터 식별자	세기	연도	월일	
용기제조업자코드	데이터 식별자	용기식별	상세식별	카탈로그번호	
Special Testing	데이터 식별자	테스트 정보 (표현형, 항체 값 등)			

- 기탁식별번호 : ICCBBA에서 지정한 수집국가 및 기관, 제품수집일, 일련번호, 플래그문자로 구성된다.
- 혈액그룹(ABO,Rh) : 혈액형, 표현형, 예약문자로 구성되며 예약문자는 향후 사용을 위해 지정되어 있다.
- 제품코드 : 혈액정보, 기탁정보, 혈액세부정보로 구성되며 기탁정보와 혈액세부정보는 혈액정보에 의해 종속적으로 코드가 부여된다.
- 유효기간 : 세기, 연도, 월일로 구성되며 사용자의 목적에 따라 다양한 유효기간의 표현방식을 제공한다.
- 용기 제조업자코드 : 물류단위, 세부식별정보, 카탈로그번호로 구성된다.
- Special Testing : 혈액에 대한 실험정보 및 결과값으로 구성되며 현재 표준에 7개 영역으로 분류되어 있다.

ISBT 128은 식별코드에 데이터 식별자를 지정해 식별자의 종류에 따라 코드 값을 표현하여 코드의 재사용성을 보장하며 유일키 값을 사용해 각 식별코드에 대한 유일성을 보장한다. 하지만 자원 식별을 위해 다수의 식별코드가 조합되어 표현함으로써 유전자원 식별시 공간적 제약이 따르며, 혈액 식별을 위한 코드 설계로 유전자원 정보를 표현하는데 어려움이 있다.

2.3 HIBCC

HIBCC(Health Industry Business Communications Council)는 미 보건의료협회의 주도하에 보건의료분야에 특화된 표준 코드체계를 연구하고 있다. HIBCC의 표준 코드체계는 두 가지 코드체제로 구분되는데 HIBC SLS(Supplier Labeling Standard)와 HIBC PAS(Provider Application Standard)로 구성된다. 두 가지 표준 코드체계는 정보 식별에 있어 일차 데이터 구조와 이차 데이터 구조로 구분해 필수정보와 선택정보를 식별하며 HIBC PAS 코드 구조는 <표 3>과 같이 표현한다.

<표 3> HIBC PAS 코드 구조

코드구성요소	코드설명
Where	자원지정 플래그문자
What	자원속성 플래그문자
Data	사용자 정의 데이터
체크문자	코드 내 오류검출

- Where : HIBCC에 의해 지정된 자원을 식별하며 현재 A부터 K까지 지정되어 있으며 L부터 X는 향후 사용을 위해 예약어로 지정되어 있다. Y는 데이터 필드의 길이를 3자리로 확장해 사용한다.(예, A-환자, B-환자진료카드)
- What : 식별정보의 속성을 표현하며 현재 A부터 T까지 지정되어 있으며 U부터 X는 향후 사용을 위해 예약어로 지정되어 있다. Y는 데이터 필드의 길이를 3자리로 확장해 사용한다.(A-분류식별코드, C-환자 식별코드)
- Data : PAS에 의해 정의된 데이터가 요구된다.
- 체크문자 : 전체 데이터 필드에 대한 정확성을 검증한다.[4]

HIBC SLS는 GS1 표준 코드체계와 HIBCC에서 지정한 HIBC 라벨러 식별코드체계를 병행해 대다수의 보건의료 객체를 식별한다. HIBC LIC 표준 코드체계는 라벨링 플래그 문자, 라벨러 식별코드, 제품 및 카탈로그 번호를 사용해 보건의료 제품 및 서비스를 식별하며 GS1 표준 코드체계는 물류식별자, 제조업체식별코드, 제품식별코드를 사용해 보건의료 제품을 식별한다. HIBC SLS 코드 구조는 <표 4>와 같이 표현한다.

<표 4> HIBC SLS 코드 구조

HIBC LIC	
코드구성요소	코드설명
+	HIBC 라벨링 플래그문자
라벨러 식별코드	HIBC에서 지정한 라벨러 식별코드
제품, 카탈로그 코드	제품 식별코드
물류단위코드	물류단위 식별코드
체크문자	코드 내 오류검출
GS1 Code	
코드구성요소	코드설명
물류식별자	물류단위 식별코드
제조업체, 제품코드	제조업체 및 제품 식별코드
체크문자	코드 내 오류검출

- + : HIBCC에서 지정한 제조업자 데이터 식별자
- 라벨러 식별코드 : HIBCC에서 지정한 라벨러 식별코드
- 제품 및 카탈로그 코드 : 제품 및 카탈로그 식별코드
- 물류단위 식별코드 : 제품의 물류단위를 식별코드(예, 0-아이템, 1-박스)
- 물류식별자 : 보건의료 제품의 물류단위 식별코드
- 제조업체코드 및 제품코드 : 제품을 생산, 유통하는 제조업체 및 제품 식별코드
- 체크문자 : 전체 데이터 필드에 대한 정확성을 검증한다.[5]

HIBCC 표준체계는 PAS와 SLS의 두 가지 표준체계 구조를 갖고 있어 상황에 따라 적절히 코드체계를 구별해 사용해야 하는 문제가 있으며 SLS의 경우 유통 및 물류 정보이외에 다른 정보를 표현할 수 없어 정보표현에 제약이 있어 유전자원 정보를 표현하는데 문제가 있으며 북미지역을 중심으로 사용되고 있어 GS1의 EAN 코드체계를 사용하는 국가들에서는 범용적으로 사용하기 어렵다.

2.4 국내 의료형 바코드

국내에서 생산, 유통되는 모든 의약품에 대해 표준 코드체계를 지원하기 위해 2000년 2월 의약품바코드 표시 및 관리요령 법품을 지정해 국내 보건의료분야 표준화를 진행하고 있다. 국내에서 사용하는 의약품 코드는 의약품 인허가코드, 건강보험약가코드, 의약품생산실적 코드, 생산자 코드 등으로 다원화되어 있을 뿐만 아니라 유통을 위한 의약품 코드 부재에 따른 과도한 물류비용이 발생하고 있다.[6]

이에 따라 식약청에서는 GS1-13코드를 사용해 의약품에 대한 정보를 식별함으로써 의약품에 대한 유통 및 물류 요구조건 및 정보관리를 편의성을 제공하고 있다. GS1-13코드에는 국가, 의약품 제조업체, 의약품에 대한 정보를 표현하고 있다.

현재 국내 의료형 바코드는 GS1-13 코드체계를 사용하고 있어 소형 객체에 부착하지 못하는 문제를 안고 있어 대부분의 유전자원 객체 용기들이 소형인 점을 감안할 때 자원식별에는 어려움이 있다.

**3. 유전자원 바코드 코드체계**

기초에 연구되었던 보건의료 코드체계들은 개별 객체에 대한 식별이 어려우며 정보표현의 다양성을 지원하지 못하는 단점을 지니고 있어 유전자원 정보 표현에 부적합한 문제를 갖고 있다. 자원 식별을 위해 요구되는 정보는 자원을 제공한 개체를 식별하기 위한 자원제공개체 정보, 유전자원을 나타내기 위한 유전자원 정보, 자원제공개체에 대한 임상, 역학적 자료나 연구대상에 대한 데이터를 식별하는 자료정보, 그리고 관리를 목적으로 만들어지는 데이터를 구분하기 위한 관리정보로 나눌 수 있으며 이 정보를 정리하면 <표 5>와 같이 표현한다.

<표 5> 유전자원 관련정보

정보	식별코드	설 명
자원제공개체	개체식별	자원 개체를 유일하게 식별
유전자원 정보	자원코드	자원 개체를 유일하게 식별
	종분류	생물 종분류 (Ex. 동물, 식물, 인간, RAT, Mouse 등)
	시료종류	생물체로부터 파생되어 시험, 검사, 분석에 쓰이는 시료의 종류 (Ex. Cell, DNA, Serum, Plasma 등)
	보관용기 종류	시험관, 배양접시, 시료채취용기 등의 생물자원 r 객체를 담기 위한 용기의 종류와 박스, 랙 등의 단품 포장물의 종류
	양적정보	자원의 부피, 질량 등의 양적정보
자료 정보	생산일자	자원의 생산날짜를 식별하는 정보
	임상자료	자원제공개체의 임상정보를 유일하게 식별
관리 정보	역학자료	자원제공개체의 병력, 과거력 등에 대한 정보를 유일하게 식별
	저장위치	보관된 유전자원의 저장위치를 식별하는 정보
	실험결과	유전자원 관리상 처리된 실험정보(Ex, QC결과, EBV 결과)
	관련연구 정보	유전자원이 특정 연구에 사용되어 도출된 결과 값
	사업정보	유전자원 수집과 관련된 사업정보(Ex, 사업식별코드, 주관기관, 채취기관등)
담당자	유전자원을 관리하는 담당자를 유일하게 식별	

이와 같은 정보를 연구하거나 관리하는 기관들은 자원을 유일하게 식별하면서도 ‘개체 식별코드’, ‘자

원 생산일자’와 같이 연구와 관련하여 매우 중요하고 빈번하게 사용되는 정적정보 또는 필수정보를 표현하기를 원하고 있다. 따라서 유전자원 코드체계는 유일하게 객체를 식별하는 고유의 속성 외에도 자원을 사용하는데 있어서 중요한 필수정보의 역할을 해야 하며, 특히 이 필수정보는 기관의 사용목적이나 판단기준이 상이하므로 필수 정보로서의 요소를 유연하게 사용될 수 있게 설계되어야 한다.

**4. 결론**

보건의료분야에 있어 미국과 유럽을 중심으로 표준체계 및 정보시스템에 대한 연구가 지속적으로 진행되고 있지만 유전자원에 대한 특화된 표준체계는 현재까지 발표된 사례가 없는 것으로 조사되었다. 이 논문에서는 국제적으로 사용되는 보건의료 표준 코드체계를 분석하여, 유전자원을 위한 코드체계의 문제점 및 적용방안을 파악하였으며, 표준 코드체계 설계를 위한 유전자원 정보 속성 및 코드체계 요구사항을 도출하였다.

향후 연구로는 유전자원 코드체계에 대한 설계를 수행하며 표준 코드체계를 기반으로 한 정보 수집/관리를 위한 자동화 시스템 개발 및 통합 정보시스템에 대한 연구가 필요할 것으로 기대된다.

**참고문헌**

[1] 산업자원부 기술표준원, 기술표준백서, 2005  
 [2] GS1, General Specifications Version 7.0 “Numbering and Symbol Marking for Very Small Healthcare Items”, 2006  
 [3] ICCBBA Inc, “ISBT128 Standard Technical Specification Version 2.1.0”, 2004  
 [4] ANSI Inc, The Health Industry Barcode Provider Applications Standard, 1996  
 [5] ANSI Inc, The Health Industry Barcode Supplier Labeling Standard, 1997  
 [6] 한국보건산업진흥원, <http://barcode.khidi.or.kr>