

국가 R&D관련 데이터베이스에서 공통적으로 사용되는 데이터 코드의 표준 선행조사

A Standard Investigation of Commonly Used Data codes at National R&D-related Databases

이병희, 김윤정, 정택영
한국과학기술정보연구원

Lee Byeong-Hee, Kim Yun-Jeong, Chung Taik-Yeong
KISTI

요약

본 논문에서는 국가 R&D관련 기관들의 데이터베이스에서 사용되고 있는 데이터 코드를 표준화하기 위해 R&D관련 4개 주요 기관들의 데이터베이스에서 사용되고 있는 코드를 비교분석하고, 코드 표준을 제정하기 위한 선행조사에 대하여 논한다. 조사결과 4대 주요 R&D관련 기관들의 DB에서 현재 코드로 관리하고 있는 개수는 340여개를 넘고 있었으며 이들 340여개 중에서 공통적이고 일반적인 것들이 150개로 서로 관련성이 있었다. 일반적이고 공통적인 150개 코드들에 대해 본 논문에서는 이들 150개 코드를 코드 의미에 따라 18개로 분류하고 코드 표준화를 위한 선행 조사를 행하였다.

Abstract

For the standardization of data codes at national R&D related organizations, this paper surveys commonly used data codes in national R&D-related databases and presents their standard investigations. The survey says the total number of the data codes in the four major R&D related organizations is 340. Among the 340, 150 data codes are general and interconnected. This paper classifies the 150 data codes into 18 ones by code meanings and checks the standard investigation of the 18 ones.

I. 서론

현대 정보화 사회에서는 정보화 요구가 급증하면서 대량의 데이터를 저장, 관리, 교환하기 위하여 데이터를 코드화하고 있다. 특히, 회사나 정부 부처 등에서는 조직의 정보 자원 관리를 위하여 표준 코드를 만들어 사용하고 있다.

데이터를 코드로 만들어 관리하면 데이터를 입력 할 때 입력이 용이하고 입력자가 실수로 입력할 가능성을 줄일 수 있으며, 데이터를 코드로 만들어 저장하면 데이터베이스의 기억 용량을 절약할 수 있고, 코드화된 데이터는 자동화된 데이터 교환, 처리 속도 등에도 효율적인 장점이 있다.

예를 들어, 미국의 학교 중에 하나인 "University of California at Los Angeles"를 코드화하지 않고 입력자에게 일일이 입력을 하게 하다보면 실수로 글자를 잘못 입력할 가능성이 많아지며, 입력자들이 "U.C.L.A.", "UCLA", "Uni.of California LA" 등으로 입력할 수도 있어 이런 데이터를 저장, 관리하는 컴퓨터의 입장에서는 입력된 데이터를 각각 다른 것으로 잘못 인식할 수 있다.

이런 문제를 해결하기 위하여 국제표준화 기구인 ISO와 IEC에서는 코드화 방법의 원칙과 방법에 대하여 ISO/IEC

TR 9789를 1994년에 제정하여 발표하였다[3].

본 논문에서는 국가 연구개발(R&D) 사업의 투명성을 확보하고 투자 효율성을 극대화하기 위해, R&D 관련 기관에 산재된 정보들을 통합된 데이터베이스로 구축하여 국가 R&D 환경 변화에 대응 가능한 공동 활용 체제를 구축하기 위해 2006년부터 과학기술부와 한국과학기술정보연구원(KISTI)에서 추진 중인 국가과학기술종합정보시스템(NTIS)에서의 코드 표준화를 검토하고자 한다.

이를 위해 R&D관련 기관들의 DB들에서 데이터 코드를 표준화하기 위해 R&D관련 4개 주요 기관들 즉 한국과학재단(KOSEF), 한국학술진흥재단(KRF), 정보통신연구진흥원(IITA), 한국산업기술평가원(ITEP)의 DB에서 사용되고 있는 코드를 비교분석하고, 코드 표준을 제정하기 위한 선행조사에 대하여 기술한다.

II. 기존 연구

국제표준화 기구인 ISO와 IEC에서는 1994년에 코드화 방법의 원칙과 방법에 관하여 ISO/IEC TR 9789를 제정하였다.

표준 코드 제정을 위해 많은 노력과 투자를 하고 있는 코드 표준으로는 행정자치부의 행정정보체계 표준[4], 교육인적자원부 교육정보 코드관리프로그램[2], 국방부의 국방데이터 부호/병과/제대구분 체계, 노동부 산업인력관리공단의 국가기술자격 코드, 조달청의 정부물품분류코드 등이 있다.

NTIS에서는 각 R&D관련 기관에서 관리되고 있는 데이터 코드를 통합 관리할 필요성을 느끼고 있다. 하지만 현실적으로 국가적 차원에서 R&D관련 데이터 코드의 총괄관리 규정이 미흡하여 각 정보시스템별로 독자적인 코드체계를 사용하고 있다. 이러한 데이터 코드들은 별개의 기관들에서 각각 다른 사람들에 의해서 개발된 것이 대부분이어서 코드들 간에 표현방식에 차이가 나는 문제점을 갖고 있다. 이런 문제점으로 인하여 각 시스템에서 사용되는 국가 R&D정보 데이터가 체계적으로 코드화되지 못함으로 인해 국가 R&D정보화의 효과가 나타나지 못할 뿐 아니라 시스템의 개발시 코드의 중복 개발의 부담을 초래하고 있다. 이런 문제점을 해결하고 국가 R&D정보 자원의 공유 및 효율적 관리를 위해서는 국가차원에서의 표준 데이터 코드를 제정하여 지속적으로 유지 보수할 필요가 있다.

본 논문에서는 NTIS에서 필요한 데이터 코드 표준에 대하여 ISO 표준과 같은 국제표준은 준수하면서, 나머지에 대해서는 현재 각 부처에서 사용하고 있는 표준들을 조사하고자 한다. 그리하여 NTIS의 국가 R&D정보화를 함에 있어 최대한 기존의 데이터 코드 표준 사용을 지향하고 문제가 되는 부분에 대해서만 새로운 데이터 코드 개발을 하는 방식을 취하기로 하고 주요 R&D관련 기관의 코드 사용 실태를 파악하기로 한다. 특히, 본 논문에서는 주요 R&D관련 기관들에서 사용하고 있는 코드들의 현황과, 이들 코드 중에서 공통적이고 일반적인 코드들에 중점을 두고 NTIS 코드 표준을 위한 선행조사를 하도록 한다.

Ⅲ. 4개 R&D관련기관 코드 분석

NTIS 코드 표준을 위한 선행 작업으로 본 논문에서는 주요 4개 R&D관련 기관인 KOSEF, KRF, IITA, ITEP에서 사용하고 있는 코드 현황을 조사·분석하였다. 4개 R&D관련 기관들의 DB에서 사용하고 있는 데이터 코드들은 총 340로 조사되었다. 이들 340여개 중에는 유사 유형으로 그룹화 할 수 있는 것들이 많은데 예를 들어, 기관코드의 경우 <표 1>과 같이 총 30개의 이름으로, 분야분류코드의 경우 <표 2>와 같이 총 26개의 이름으로, 국가코드의 경우 <표 3>과 같이 총 21개의 이름으로 사용되고 있었다. 분야분류코드의 경우는 각 기관들마다 고유의 분류체계를 사용하고 있어 표준화가 시급한 곳이

다. 국가과학기술표준분류체계가 2005년 9월부터 시행되고 있으나 아직은 여러 분류체계가 공존하고 있는 상황이다.

[표 1] 기관코드로 사용되고 있는 이름

국제공동연구기관코드, 근무기관코드, 기관구분코드, 기관코드, 기관코드명, 기관코드직장, 대상기관코드, 방문기관구분코드, 변경전소속기관코드, 변경후소속기관코드, 사업단기관코드, 상대방재단기관코드, 상위기관코드, 선정시점소속기관코드, 소속기관코드, 소속연구소코드, 수출입기관코드, 신청기관코드, 신청연구소코드, 연구소코드, 연구책임자소속기관코드, 연구책임자인력소속기관코드, 연구기관코드, 주관연구기관코드, 지원기관코드, 커뮤니티구성원소속기관코드, 파견기관코드, 파견연구원소속기관코드, 평가자소속기관코드, 의뢰기관코드

[표 2] 분야분류코드로 사용되고 있는 이름

14분야코드, 6기술분류코드, 과제유형코드, 과학재단코드, 기술분야코드, 기술코드, 마스터연구분야코드, 매핑분류구분코드, 분야구분코드, 분야코드, 상세기술분류코드, 세부기술코드, 세분야코드, 연구분야코드, 연구분야코드명, 연구중심분야코드, 적용중분야코드, 적용중소분야코드, 적용중심분야코드, 전공코드, 전문분야코드, 중분야코드, 특허구분코드, 학문코드, KOSEF기관구분코드, NTAM코드

[표 3] 국가코드로 사용되고 있는 이름

개최국가코드, 공동과학자소속기관국가코드, 교류국가코드, 국가코드, 기관협력국가코드, 논문게재국가코드, 발표국가코드, 방문연구자소속국가코드, 소재지국가코드, 연구국가코드, 초청과학자국가코드, 최종학교국가코드, 최종학위취득국가코드, 출원등록국가코드, 출판국가코드, 취득국가코드, 파견국가코드, 학술지발행국가코드, 학술회의개최국가코드, 학술회의국가코드, 학위취득국가코드

조사결과 4개 R&D관련 기관들의 DB에서 사용하고 있는 데이터 코드들 340여개 중에서 공통적이고 일반적인 것들이 150개이고 일반적이지 않은 것들이 190개이었다. <표 4>는 공통적이고 일반적인 것들의 코드명과 종류수를 보여준다.

[표 4] 공통적 코드명과 이름수

코드명	이름수	코드명	이름수
기관코드	31	학과코드	4
분야분류코드	26	직위코드	4
국가코드	21	언어코드	3
학교코드	16	지역코드	2
예산비목코드	11	부서코드	2
사업코드	7	병역코드	2
은행코드	6	업체코드	1
학위코드	5	성별코드	1
단과대학코드	5	부처코드	1

IV. 코드 표준 선행조사

본 장에서는 <표 4>에서 조사된 공통적 코드명을 중심으로 코드 표준 선행조사에 대하여 기술한다.

4.1 기관코드

<표 1>에서 보듯 기관코드에는 행정자치부의 행정정보체계 표준(<http://code.gcc.go.kr>)을 활용할 수 있으나 행정정보체계 표준 코드에는 NTIS에 관련된 110여개의 연구관리전문기관들 중 전문기관, 출연연/사업단, 기타 등 90여개 기관들이 빠져 있으므로 NTIS의 목적에 맞게 기관코드를 설계할 필요가 있다.

4.2 분야분류코드

NTIS에서는 표준분야분류체계로 2005년 9월에 과학기술부에서 발표한 국가과학기술표준분류체계를 활용할 계획이다.

국가과학기술표준분류체계는 『과학기술기본법』 <제27조>(국가과학기술표준분류체계의 확립)과 동법 시행령 <제41조>, 『국가연구개발사업의 관리 등에 관한 규정』(이하 공동관리규정) 시행규칙에 따라 적극 활용하여야 한다. 최근 2006년 8월 4일에 개정(안)으로 입법예고된 공동관리규정 7조 4항에서는 “국가과학기술표준분류 등을 알 수 있도록 과제별 고유번호를 부여하여야 한다.”라고 하여 국가과학기술표준분류의 대/중/소분류를 반영하는 과제별 고유번호를 부여하도록 명시하고 있다.

이에 NTIS 2007년도 9개 주요과제에서도 국가과학기술표준분류체계를 적용할 계획이며 국가과학기술인력정보시스템에서는 2005년부터 이미 적용하고 있다. 현재 NTIS사업단에서는 어떤 과제에 대해 자동적으로 국가과학기술표준분류체계의 분야분류코드로 분류해 주는 표준분류지원시스템을 구축 중에 있다.

4.3 국가코드

<표 3>에서 보듯 여러 이름으로 사용되고 있으나 ISO 3166-1에 따라 영어 알파벳 2자리 또는 3자리로 국가를 표시하는 국가코드를 활용하면 NTIS에서도 큰 무리는 없을 것이다. 현재 한국의 국가코드는 'kr'이다.

행정자치부의 행정표준코드관리시스템의 외국명코드를 활용할 수 있다.

4.4 학교코드

학교코드는 관련해서는 고등학교구분코드, 대학기관코드, 대학코드, 박사취득기관코드, 본교대학교코드, 본교대학코드, 분교대학교코드, 석사취득기관코드, 소속대학기관코드, 전공

취득학교코드, 최종학위취득기관코드, 학교코드, 학사취득기관코드, 학위취득기관코드, 학위취득대학코드, 학위취득학교코드 등의 여러 이름으로 사용되고 있으며, 학교코드는 교육인적자원부에서 주로 관리하고 있다.

하지만 최근 국내 연구자들이 외국에서 대학을 졸업하는 경우와 외국 연구자들이 외국 대학을 졸업하고 국내에 들어오는 경우가 많고 학교명이 변경되는 경우가 많아 학교코드에 대해 더욱 체계적인 관리가 필요하다.

행정자치부의 행정표준코드관리시스템에서도 기관코드에 대학도 대상으로 하고 있으므로 활용가능하다. 또한 NTIS에서 학교코드에 대해 주로 대학을 다루고 있어 대학의 특성을 감안하여 대학코드를 설계할 필요도 있다.

4.5 예산비목코드

예산비목 관련해서는 공동연구비재원구분코드, 급여재원구분코드, 방문연구비재원구분코드, 신청연구비분류코드, 신청연구비항목코드, 연구비구분코드, 연구비항목코드, 재원구분코드, 지급재원구분코드, 학술회의개최재원구분코드, 환입재원구분코드 등의 여러 이름으로 사용되고 있다. NTIS에서는 공동관리규정의 예산비목 유형을 사용하면 될 것이다.

4.6 사업코드

사업코드 관련해서는 기초포털사업코드, 대사업코드, 마스터사업코드, 세부사업코드, 접수사업코드, 중사업 코드 등의 여러 이름으로 사용되고 있다. 사업 주관기관별로 부여한 사업코드를 활용하던지, 사업을 종합적으로 관리하고자 할 경우는 대표기관에서 별도의 사업코드를 만들 필요가 있다.

4.7 은행코드

국내의 은행코드에 있어서는 금융결제원(http://www.kftc.or.kr/kftc/info/bankcode/code_search.jsp)에서 제시하는 은행코드를 활용할 수 있다. 외국계 은행의 경우는 현재 금융결제원의 은행코드에 포함되어 있지 않으므로 별도로 개발할 수도 있다.

4.8 학위코드

교육인적자원부의 『학위의 종류 및 표기방법에 관한 규칙』에서의 국내에서 수여되는 학술학위의 종류구분에 따라 코드화할 수 있다. 행정자치부의 행정표준코드에서는 9개 즉, 국내학사, 국외학사, 국내석사, 국외석사, 국내박사, 국외박사, 명예박사, 전문학사, 기타학위로 구분하고 있다.

연구자의 출신학교에서의 재학, 중퇴, 수료, 졸업 사항을 교육제도의 단계별(학사, 석사, 박사 등)로 연계한 학력코드 KS

X 1501을 활용할 수 있다.

4.9 단과대학코드

단과대학코드 관련해서는 근무단과대코드, 단과대학코드, 소속단과대학코드, 학위취득단과대코드, 학위취득단과대학코드 등의 여러 이름으로 사용되고 있다. 국내 단과대학의 경우는 교육인적자원부의 단과대학 정보를 활용할 수 있다.

4.10 학과코드

행정자치부의 행정표준코드관리시스템의 학과코드나 교육인적자원부의 학과 정보를 활용할 수 있다. 학과명은 자주 변동되므로 수시로 업데이트가 필요하며 최근에는 융합학문의 영향으로 애매모호한 학과들이 자주 등장하고 있다. 또한 외국 학생들이나 외국 학위자들이 국내에 자주 들어오므로 인해서 수시로 업데이트해야 한다.

4.11 직위코드

직위코드 관련해서는 연구책임자인력직급코드, 연구책임자 직위코드, 연수기관지도교수직위코드, 직위코드 등의 여러 이름으로 사용되고 있다. 행정자치부의 행정표준코드관리시스템의 직급/직위코드를 사용하거나 직급/직위는 소속한 기관별로 상이하므로 별도로 개발하여 사용할 수 있다.

4.12 언어코드

ISO에서는 세계의 언어들을 영문자 소문자 2자리(ISO 639-1) 코드와 수화나 역사적(고대/중세/현대 등) 언어까지 표현할 수 있는 3자리(ISO 639-2) 코드를 제정하여 발표했으며, 최근에는 어족과 관련된 코드(ISO 639-5)와 언어 변이형까지 나타낼 수 있는 ISO 639-6이 나오고 있다. 현대 한국어의 2자리 코드는 'ko'이며, 3자리 코드는 'kor'이다. 하지만 3자리 코드에 영어나 불어와 같은 언어는 고대어, 중세어 등이 포함되어 있으나 한국어는 아직 고대어, 중세어, 근대어 등의 코드가 ISO 639-2에 포함되어 있지 않다.

대체로 언어코드는 ISO 639를 활용하면 언어코드에 관해서는 거의 모두를 관리할 수 있다. ISO 639보다는 작은 규모로 전 세계 언어가 망라되어 있지는 않지만 행정자치부의 행정표준코드관리시스템에서 외국어 코드를 활용할 수도 있다.

4.13 지역코드

KS X 1505에 따른 시군구 코드 또는 행정자치부 행정표준코드관리시스템의 우편번호코드와 법정동코드를 활용할 수 있다.

4.14 부서코드

부서코드 관련해서는 소속부서코드, 평가자소속부서코드 등의 여러 이름으로 사용되고 있다. NTIS에서 부서코드는 소속한 기관별로 상이하므로 별도로 개발하여 사용할 수 있다.

4.15 병역코드

국방부에서는 역종/군별/병과/제대구분/미필사유 코드를 만들어 발표하고 있으며 이들 코드들은 행정자치부의 행정표준코드에서도 서비스 중이므로 이들을 활용하면 된다.

4.16 업체코드

국세청의 사업자등록번호에 따른 업체코드나 한국전화번호부의 업종/상호 편을 활용하여 업체코드를 생성할 수 있다.

4.17 성별코드

행정자치부의 행정표준코드관리시스템에서 성별코드를 활용할 수 있다. ISO/IEC 5218을 활용할 수도 있다.

4.18 부처코드

행정기관 관련해서는 행정자치부의 행정표준코드관리시스템에서 기관코드를 활용할 수 있다. NTIS에서 부처코드는 소속한 기관별로 상이할 수 있으므로 별도로 개발하여 사용할 수도 있다.

V. 결론

지금까지 본 논문에서는 R&D관련 기관들의 DB에서 데이터 코드를 표준화하기 위해 R&D관련 4개 주요 기관들 KOSEF, KRF, IITA, ITEP의 DB에서 사용되고 있는 코드를 비교분석하고, 코드 표준을 제정하기 위한 선행조사에 대하여 논하였다. 특히 4대 주요 R&D관련 기관들의 DB에서 현재 사용하고 있는 코드로 관리하고 있는 개수가 340여개를 넘고 있었으며 이들 340여개 중에서 공통적이고 일반적인 것들이 150개이고 일반적이지 않은 것들이 190개이었다. 본 논문에서는 일반적이고 공통적인 150개를 다시 18개로 분류하여 코드명을 중심으로 코드 선행 조사를 행하였다.

향후 본 논문에서 선행 조사된 코드항목들에 기반하여 R&D관련 기관 간에 실질적인 표준화가 이루어져 NTIS에서 정보 공유 및 교환에 유용하게 사용되길 기대한다.

■ 참고 문헌 ■

- [1] 과학기술정보 표준자료집(II), “과학기술정보 유통을 위한 코드 체계”, STI-S.2005-1.05, 2005.
- [2] 김정겸, “교육정보의 효율적 관리를 위한 표준데이터 코드관리 프로그램 설계”, 한국컴퓨터교육학회 논문지. 제3권 제1호, pp.13-22, 2000.
- [3] 기술표준원, KS X 1520: 1997 (ISO/IEC TR 9789 : 1994) 「데이터 교환용 데이터 요소의 구조와 표현 지침 - 코드화 방법 및 원칙」.
- [4] 행정자치부 정부전산정보관리소, 행정정보체계 표준, <http://code.gcc.go.kr>.