

과학기술전문인력 관리를 위한 인력정보 메타데이터 표준화

Human Resource Metadata Standardization for Managing Science & Technology Personnel

김경옥, 송인석, 표순희*, 이미화**, 이재진***
KISTI, 이화여대 대학원*, 한성대학교**, DPC***

Kim Kyung-Ok, Song In-Seok, Pyo Sun-Hee*,
Lee Mi-Wha**, Lee Jae-Jin***

요약

연구인력정보는 국내 과학기술 관련 기관별로 자관의 수요에 의해 구축되고 있으나, 상호 연계되지 않고 있어 시스템 및 예산의 낭비를 초래하고 있다. 본 연구에서는 국내 과학기술 전문인력을 위한 표준 메타데이터를 제시함으로써 국가 차원의 인력 데이터베이스 구축하고 인력 정보의 생산, 유통 및 관리에 이르는 전 생명주기에서 상호운용성을 지원하고자 한다. 이를 위해 국내 13개 국가연구기관의 인력정보 데이터베이스를 조사·분석하고, 이를 바탕으로 일본 문부과학성과 CERIF 인력 모델을 참조하여 데이터 모델링을 수행하여 11개 대항목과 151개 세부 항목으로 구성된 데이터 요소를 도출하였다. 인력정보와 연계 가능한 타분야 표준 메타데이터 및 데이터베이스를 최대한 고려하고 메타데이터 레지스트리(MDR) 국제 표준인 ISO/IEC 11179를 수용하여 서술하고자 하는 객체, 속성, 표현을 데이터 요소명에 반영하고 이에 대한 정의를 명확히 하여 의미의 모호함이나 중복을 피하고자 하였다. 향후 국가 연구개발정보의 공동 활용을 위하여 추진되고 있는 국가과학기술종합정보시스템(NTIS)에 연구 결과를 활용함으로써 효과적이고 표준화된 국가 과학기술인력의 이용을 기대할 수 있다.

Abstract

R&D Personnel information is constructed locally based on the needs of each institute and therefore is distributed over different databases. It does not support inter operability which makes it difficult to access and update that leads to the problem of ineffective usage. In this study, we have categorized the lower level information that forms the domestic S&T Personnel and defined the relationship between each type of information to suggest a standard for the data elements that guarantee the access to specific information in order to support inter operability. We have analyzed the human resource information database of domestic and foreign research institutes for the data modeling. We have also made reference to the standard metadata and database of other types that can be linked with the human resource information in designing the data elements. ISO/IEC 11179, the international standard for the metadata registry(MDR), was adopted to apply the object, attribute and expression to be described to the name of the data element.

I. 서론

과학기술인력은 과학 지식의 창출과 기술개발의 핵심 자원으로서 과학기술인력의 양성과 관리를 통한

우수한 인력의 확보가 그 나라의 기술수준과 경쟁력의 지표가 된다. 국내 인력정보 시스템은 자관의 수요에 따라 여러 곳에서 구축되고 있으나 통합된 검색

시스템을 제공하지 못하며, 기관별로 데이터 구축항목이 달라 이용자에게 혼동을 주고 있다. 동일한 정보를 여러 기관이 중복 구축하는 것은 자원의 낭비이며, 저장된 정보가 자동 갱신되지 않는 것도 문제이다. 그리고 개별 기관에서 사용하고 있는 요소명이 동일한 개념을 지칭함에도 서로 다른 용어가 사용되고 있으며, 기관에서 따라 입력 형태나, 입력하는 내용이 매우 달라 인력정보의 상호운영성에 장애가 되고 있다. 따라서 본 연구에서는 국가차원의 인력 DB의 구축과 운영을 위한 표준 메타데이터를 제안하며 이를 통하여 국가적 인력 DB의 신뢰성과 유용성을 극대화하고자 한다.

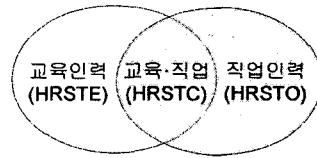
II. 이론적 배경

1. 과학기술과 인력 정보

과학기술인력의 범위는 정의하는 기관이나 국가별 로 약간의 차이를 보이고 있다. OECD는 과학기술인적자원(HRST; Human Resources for Science & Technology)이라는 개념을 도입하여 과학기술 인력을 정의한다. 과학과 기술의 범위 설정에 있어 광의의 개념을 사용하여, 각각 '자연과학'과 '어떠한 목표를 성취하기 위한 계획을 실행하는 도구나 기법'으로 정의하고 있다[2].

[그림 1]은 과학기술인적자원의 구성을 보여준다. 과학기술인적자원을 구성하는 첫 번째 유형은 과학기술인력의 자격요건을 충족시키는 사람들(HRSTE)이고, 두 번째 유형은 자격요건에 관계없이 과학기술과 관련된 직종에 종사하는 사람들(HRSTO)이며, 세 번째 유형은 첫 번째 유형과 두 번째 유형의 교집합으로 이루어진 과학기술인적자원의 핵심을 이루는 사람들(HRSTC)이다[6].

과학기술인력을 정의하기 위해서는 “과학기술 분야의 고등교육”과 “과학기술 관련 직종”이 국제적으로 통용될 수 있는 분류기준에 의해 정의되어야 한다.



▶▶ 그림 1. 과학기술인적자원(HRST)의 구성

본 연구에서는 과학기술의 학문적 범위는 OECD 프레스카티 지침서 상의 학분 분야 분류에 따라 자연과학, 공학 및 기술, 의학, 농림수산학과 사회과학, 인문학을 포함하였고[2], 과학기술인력은 OECD에서 정의하는 핵심 과학기술인력(HRSTC)으로 한정하여 전문대학 이상의 고등 교육과정을 마치고 과학기술 분야에 종사하는 과학자, 공학자, 기술자, 대학교수, 연구원, 국외거주 연구자 및 교수, 석·박사학위 과정생으로 제한하였다.

과학기술인적자원 외에 고학력자(Highly Qualified Man power/Personnel), 과학기술인력(S&T Manpower/Personnel), 연구개발인력(R&D Manpower/Personnel), 연구자 및 공학자(Research Scientists and Engineers) 등의 용어들이 과학기술인력 정의에 사용된다. 특히 UNESCO에서 정의하는 과학기술인력(S&T Personnel)과 OECD의 연구개발인력(R&D Personnel)이 국제적으로 광범위하게 통용되고 있다.

2. 국내외 인력정보 구축 현황

인력정보는 연구자정보, 인력정보, 인물정보 등 다양한 명칭으로 기관별 독자적으로 관리 운영되고 있다. 인력정보는 크게 연구지원기관 및 대학에서 구축한 연구자 중심의 연구과제 인력DB와 언론기관의 정치·사회 유명인 중심의 인력 DB로 구분될 수 있다. 전자의 경우 학력, 경력, 연구실적, 학회활동 등 개인의 연구역량을 나타내는 정보로 이루어지고, 후자는 가족관계, 취미 등의 신변적 정보를 제공하고 있다[7]. 본 연구에서는 전자에 해당하는 인력정보의 구축 현황을 파악하였다.

외국의 OECD, UNESCO, Eurostat, 일본 문부과

학성의 ReaD(Directory Database of Research & Development Activities)는 과학기술인력 분야와 관련된 정황적 자료를 포함한 DB를 구축하여 서비스하고 있다. OECD는 연구개발인력의 국제비교가 가능한 자료를 DB로 운영하고 있어 부문별 연구원·종사자의 숫자까지 파악가능하며, ReaD는 매년 문부과학성에서 실시하는 신규연구자조사표를 통하여 연구자의 현황 파악 및 관리가 신속하다. 또한 CRIS(Current Research Information System)의 데이터 공통 포맷을 위해 개발된 CERIF(Common European Research Information Format)은 데이터교환을 위한 기본 프레임워크로써 과제, 인력, 기관의 1수준에서 엔티티 간 다대다 관계를 나타내는 5수준까지 확장된 E-R모델을 정의한다.

국내 인력정보는 정부 부·청 산하 과제 관리 기관 등에서 구축·관리하는 연구 인력 정보를 중심으로 조사하였다. 이들 기관은 국가 연구 과제 수행 또는 과제 관리 기관으로써 과제 및 관련 정보 시스템을 분산 구축하고 있다. 구축된 인력정보는 연구개발 전 과정에서 과제 수행·관리기관, 연구과제, 연구결과, 연구비, 기자재 등의 관련 정보와 연계되고 있다. 분석된 국내 연구기관과 각 기관별 인력정보 활용 목적은 [표 1]과 같다.

[표 2] 구축기관별 인력정보 활용 목적

기관명	인력정보 활용
한국과학기술정보연구원(KISTI)	교류 및 육성
한국학술진흥재단(KRF)	과제관리
한국과학재단(KOSEF)	과제관리
한국전산원(NCA)	과제관리
한국기술사회(KPEA)	교류 및 육성
중소기업청(SMBA)	정보교환
보건의료기술연구기획평가단(HPEB)	과제관리
국기정생생산지원센터(KNCPC)	정보제공, 평가/기획 위원
농림기술관리센터(ARPC)	평가위원 활용
에너지관리공단(KEMCO)	평가, 자문
정보통신연구진흥원(ITA)	평가위원
한국과학기술평가원(KISTEP)	과제관리
한국환경기술진흥원(KIEST)	평가위원

II. 연구의 방법

1. 인력 정보 데이터 모델링

본 연구에서 구축하는 메타데이터 표준화는 과학기술인력을 대상으로 하며, 기술 대상은 인력정보의 표현을 위한 기술적 메타데이터만을 포함한다. 즉, 인력정보관리와 관련된 관리형 메타데이터는 본 연구에서 제외하며 단지 인력정보 자체를 표현하는 데이터 요소만을 추출하여 기술하였다. 표준 요소 추출을 위해서 다음과 같은 연구 방법이 추진되었다.

첫째, 문헌조사를 통해 국내외 인력정보 관련 문헌 및 표준을 조사하였다.

둘째, 사례조사로 국내외 인력정보의 현황 및 구축된 데이터베이스를 조사하였다. 한국과학기술정보연구원 등 국내 13개 국가연구기관에서 구축·운영하는 과제 데이터베이스의 인력 정보를 기초로, 일본 문부과학성 ReaD와 CERIF 모델의 인력 정보를 참조하였다.

셋째, 문헌 및 사례로 조사된 국내외 과학기술 전문 인력 데이터베이스 데이터요소를 비교·분석하여 인력정보에 필수적인 항목을 추출하였다. 기관별 DB를 비교 분석하기 위한 기준으로서 인력DB를 구성하는 개념항목을 재범주화하였다. 각 개념항목은 전체 DB를 구성하는 세부항목을 조사·분석하여 도출한 것으로, 기관별 인력DB 스키마와의 비교를 통해 개념적으로 동일하나 물리적이거나 논리적으로 혼재·분리되어 기술된 항목을 정리하고 인력 모델을 재정의 하였다.

분석된 인력 데이터베이스는 단순히 인력정보로 존재하는 것이 아니라 다양한 정보요소와 연계되어 활용된다. 즉 과제정보, 연구비, 시설, 특허 등 인력정보와 다양한 개체간의 E-R 모델이 적용되고 있다. 따라서 인력DB를 구성하는 데이터요소는 과학기술 데이터를 구성하는 통합 DB와의 관계 및 역할 파악을 통한 속성의 이해가 전제되어야 할 것이며 연구정보의 일부로서 인력정보가 다양한 개체에서 상호 접근이 가능하도록 고려하였다.

도출된 인력 정보 데이터 모델은 [표 2]와 같다.

[표 3] 인력정보 데이터요소 세부설명

대항목	필수여부	반복	타 DB 연계
기본정보	필수	불가	-
학력사항	필수	반복	-
소속기관	선택	반복	기관 DB
연구분야	필수	반복	-
논문실적	선택	반복	연계 필요
연구과제	선택	반복	연구과제 DB
저역서	선택	반복	연계 필요
산업재산권	선택	반복	특허 DB
학협회활동	선택	반복	
자격사항	선택	반복	
수상사항	선택	반복	

2. 인력 정보 메타데이터 데이터 요소

분석된 데이터 모델을 토대로 11개 대항목과 151개 세부 항목으로 구성된 데이터 요소를 [그림 2]와 같이 도출하였다. 제안된 메타데이터의 특징은 다음과 같다.

첫째, 분석대상인 국내 13개 과제관리기관과 CERIF 모델과 일본 문부과학성 ReaD, CERIF 포맷을 수용하는 CRIS와의 연계가 가능하도록 입력요소를 참조하였으며, E-R 모델의 전제하에 요소를 선정하였다.

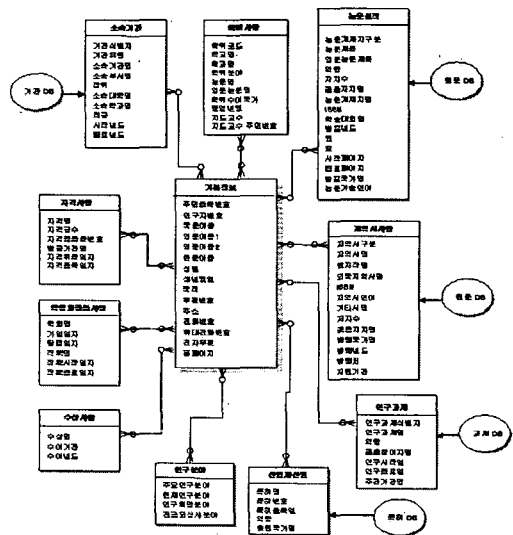
둘째, 인력정보 선발을 위해 필수적인 핵심 요소를 중심으로 선정하였다. 따라서 인력정보에 불필요한 요소는 제외시키고, 다른 정보원의 참조를 통해 이용할 수 있는 요소는 참조를 제시하였다. 기본정보, 학력사항, 연구분야의 3개 대항목은 필수 정보로서 인력 정보의 서술을 위해 반드시 정의되어야 하며, 나머지 8개 대항목은 부가 정보로 구분하여 해당 정보를 최소한 서술하고 타 정보시스템을 참조 가능한 경우 상호 운용이 가능하도록 고려하였다. 특히 과제 DB, 특허DB, 학술지 등의 원문DB와 같이 구축된 정보원과의 연계를 제안하여 데이터의 중복을 줄이

고 데이터베이스의 효율성을 증대시키려 하였다.

셋째, 메타데이터 레지스트리(MDR) 국제 표준인 ISO/IEC 11179를 수용하여 객체 클래스(object class), 속성(property), 표현(representation)을 데이터 요소 명명에 반영하여 서술하고자 하는 정보를 명확히 하며, 데이터 요소 이름의 일관성 확보, 이름의 표현 방식과 의미의 일관성 확보가 가능하다. 표준화된 명명 규칙을 따름으로써 메타데이터 요소명을 통한 조기성 파악이 가능하며, 열거형이나 코드체계로 표현 가능한 데이터 요소 값을 체계적으로 관리할 수 있다.

IV. 결론 및 제언

국가 연구 개발 인력의 선정을 위한 정보원으로서 인력데이터베이스의 역할은 중요하다. 인력정보는 연구 과제와 관련된 다양한 개체와 상호 관계를 가지기 때문에 데이터 상호교환을 위해 표준화가 필요한 대상이나 인력정보의 데이터요소의 표준화가 부재하여 구축된 데이터간의 상호운용성이 어려운 상태이



▶▶ 그림 2. 인력 메타데이터

다. 이에 본 연구에서는 인력정보의 생산, 유통, 관리 측면에서 상호운용성을 위한 공통 데이터요소를 정의하고, 요소의 정의 및 범위, 데이터 값에 대한 구체적인 지침을 제시하였다.

제안된 데이터요소는 다음과 같은 활용이 가능하다.

첫째, 인력 정보 메타데이터 표준화를 통해 효율적이고 다각적인 R&D 정보의 활용 및 공유 체제 구축할 수 있고, 중복구축을 방지한다. 제시한 요소명과 정의를 사용하여 인력 DB를 구축하는 기관간의 데이터 상호교환이 가능하다. 자관의 특수요소를 반영해야 할 경우 해당기관의 요소를 추가하여 데이터를 구축하며, 데이터 교환 시에는 제안된 공통요소만을 대상으로 데이터를 교환할 수 있다.

둘째, 국가 차원의 인력 풀(pool) 운영을 통하여 최적의 인력 선정(recruiting)과 전문가 현황파악에 신뢰성 있는 데이터를 제공할 수 있다.

■ 참고 문헌 ■

- [1] Common European Research Information Format (CERIF) 2004 Full data model <<http://www.cordis.lu/cerif>>
- [2] OECD, Frascati manual: proposed standard practice for surveys on research and experimental development, 2002.
- [3] ISO/IEC JTC1 SC32. ISO/IEC 11179 Part 1-6, 2003.
- [4] UNESCO, Manual for statistics on scientific and technological activities, 1984.
- [5] 고상원, 송성수, 김기국, 과학기술인력 통계지표의 보완 및 개선방안, 한국과학재단, 서울, 2001.
- [6] 조황희, 이은경, 이춘근, 김선우, 한국의 과학기술인력 정책. 과학기술정책연구원, 서울, 2002.
- [7] 이정규. "국가학술연구인력 데이터베이스 구축에 관한 연구". 한국컴퓨터정보학회지, 제10권, 제1호, pp.16-23, 2003.
- [8] 한국과학기술정보연구원, 과학기술 정보의 공동 활용 체제를 위한 유통 정보 표준 프레임워크 개발 중간 보고서, 과학기술부, 서울, 2005.