

# 국내 연구현황정보 데이터 포맷 개발에 관한 연구

## A Study of Developing Korea Research Information Format(KRIF)

조재인, 한국교육학술정보원  
오삼균, 성균관대학교

Jane Cho, Korea Education & Research Information Service  
Sam-kyun Oh, Sung Kyun Kwan University

본 논문에서는 우리나라 연구지원기관에서 구축한 연구과제 데이터베이스 기술요소를 분석하여 CERIF(Common European Research Information Format)를 준용한 국내 연구과제정보 교환을 위한 데이터 포맷(KRIF)을 정의하였다. KRIF에서는 국내에 구축된 연구과제정보 기술항목의 특성을 고려하여 CERIF를 확장 적용하였으며, 각 엔티티에 로컬정의 필드를 두어 교환이 불필요한 데이터요소를 로컬이 재량껏 정의할 수 있도록 설계하였다. 또한 본 논문에서는 이를 바탕으로 국내에 구축된 이형의 연구정보를 단일 창구로 검색할 수 있는 국내통합연구현황정보 데이터베이스 구축 모형을 제안하였다.

### 1. 서론

연구현황정보는 현재 어떠한 분야의 연구가 어떠한 연구자에 의해 수행되고 있는지에 관한 정보로서, 연구자들이 가장 필요로 하지만 쉽게 획득할 수 없는 정보 중에 하나이다. 연구현황정보서비스를 통해, 연구자들은 자신의 연구방향을 설정하고 최선의 연구결과를 활용할 수 있으며, 연구지원기관에서는 동일 분야 연구에 대한 중복 지원을 사전에 방지함으로써, 국가적 차원에서 중복 연구가 야기할 수 있는 시간과 노력을 절감할 수 있다. 대부분 우리나라 연구지원기관에서는 해당 기관이 지원하고 있는 과제에 대한 전산관리를 주목적으로 연구과제정보 데이터베이스를 구축·운영하고 있으며, 이를 통한 정보서비스는 부차적 기능으로 수행하고 있을 뿐이다. 또한 연구과제정보 데이터베이스가 기관별로 각기 독자적인 형태로 구축되어있고, 일부 기관의 데이터베이스만이

Web상에서 공개되고 있으므로 이러한 상황에서 연구자들이 분산 구축된 이형의 데이터베이스에 효과적으로 접근하여 국내의 연구동향을 파악하는 것은 쉬운 일이 아니다. 각 기관별로 구축되어 있는 연구과제정보를 통합한 국가통합연구현황정보서비스시스템이 개발되어 서비스된다면, 연구자들이 단일 창구로 국내에서 수행중인 연구동향을 파악하고, 최선의 연구정보를 획득하여 활용할 수 있을 것이다. 통합연구현황정보서비스시스템을 구축하기 위해서 가장 먼저 해결해야 할 과제는 우선 구축된 이형의 데이터를 표준화된 교환형태로 변환시키거나 동종의 데이터베이스를 표준화된 방식으로 작성할 수 있는 입력 포맷이 제안되어야 한다는 점이다.

본 연구에서는 유럽국가연합의 표준연구현황정보 입력 포맷인 CERIF (Common European Research Information Format)를 준용하고 국내에서 구축된 연구과제정보 데이터베이스

내용 분석을 바탕으로한 국내 연구현황정보 데이터 포맷(Korea Research Information Format, 이하 KRIF라고 한다)을 제안하고 또한 표준 포맷을 바탕으로 한 통합형연구현황정보서비스시스템 구축 모형을 제안한다.

## 2. 연구의 배경

1987년에 European Working Group on Research Databases가 발족하여 연구 프로젝트 데이터 입력을 위한 표준 포맷인 CERIF(Common European Research Information Format)를 제정하였다. 1991년에 'The Official Journal of the European Communities'에 발표됨으로써 공식적으로 회원국들에게 연구과제정보 데이터 입력을 위한 표준 포맷으로 제안되었다. 개별 회원 국가들에게 기존에 구축된 연구정보 데이터베이스를 통일 포맷으로 변환하고 산업체 및 연구소에서 수행된 연구과제정보의 신규 목록을 CERIF를 적용하여 작성하게 하였다. 1992년 European Working Party on Research Databases가 주최한 회의에서 각 회원국의 CERIF 적용 사례가 교환되었고 회원국들 사이에서 취합된 개정 및 보완 요구기 CREST(Scientific and Technical Research Committee)에 전달되어 전반적인 입력 형태 및 메타데이터 기반의 입력 구조를 위한 보완 작업이 진행중이다. 또한 연구과제정보 기술 표준에만 국한되어 있던 CERIF 포맷을 연구결과물, 연구자, 연구기관 정보 등 다양한 기타 연구관련 정보 입력 표준으로 확장하고자 하는 개정 작업이 추진되고 있다.

동시에 CERIF 표준 포맷으로 구축된 연구정보 데이터베이스들을 대상으로 통합 검색을 위한 European Research Gateways On-line(ERGO) 프로젝트가 수행 중에 있다. INNOVATION Programme에 의해 창단된 ERGO Working Group이 수행하고 있는 이 사업은 1997년에 시작되었으며 1998년 7월에 30여 유럽국가들의 연구과제 목록데이터베이스를 대상으로 공개 시험 운영에 들어갔다.

(<http://www.cordis.lu/ergo/src/MainInfomation.html>).

## 3. 국내 연구현황정보 서비스 시스템 모형 개발

현재 우리나라의 연구과제 데이터베이스는 각 부처별 연구지원 담당기관에서 구축하여 일부는 Web을 통해 일반에게 공개되어 있지만 대부분이 자체의 연구지원업무를 위한 목적으로 구축되어 있다. 본장에서는 중앙식 모형 개발을 서두르고 있는 유럽국가연합의 연구현황정보서비스시스템을 모델로 국내 기구축 데이터의 유럽 입력표준(CERIF)의 적용 가능성을 분석하고 국내 데이터 교환에 적합한 입력 포맷을 정의한다. 나아가 이용자 서비스를 위한 연구현황정보의 통합데이터베이스 구축시 로컬 기관에서 반출되어야할 핵심 데이터 항목을 추출하며, 교환 포맷으로 변환된 데이터를 바탕으로 한 통합연구현황정보서비스시스템 모형을 설계한다.

### 3.1 데이터 수집 및 분석

본 절에서는 각 부처별로 지원되고 있는 연구사업 중 일부를 대상으로 해당 과제의 관리 정보 요소를 분석하였다. 데이터베이스 스키마 정보는 외부 유출을 금하고 있는 기관이 다수여서, 본 연구에서는 각 부처의 연구 사업별로 데이터베이스 구축을 위한 전산입력 Work Sheet를 입수하여 기술 항목을 분석하였다. 표 1은 조사 대상 기관 및 연구 사업명이다.

부처명	기관명	연구사업명
과기부	과학기술정책연구원	특정연구개발사업
	과학재단	핵심전문연구과제
교육부	학술진흥재단	자유공모과제
정보통신부	정보통신기술연구지원국	학술단체연구활동 지원사업
	정보통신연구관리단	국제공동연구지원
산업자원부	산업기술정책연구소	공업기반기술개발사업
보건복지부	보건의료과학기술연구기획평가단	선도기술개발사업

표 1. 조사 대상 기관 및 연구 사업명

대부분 우리나라 연구과제 데이터베이스의 기술요소들은 CERIF<sup>1)</sup>가 정의한 데이터군에 해당하는 정보를 고루 적용하고 있었으나 데이터베이스 구축 목적이 자관의 연구지원업무를 전산화하기 위함이었으므로 각 기관의 지원 업무 특성을 반영한 고유한 데이터 요소들이 다수 사용되고 있었다. 대학에 소속된 연구자를 중심으로 기초연구 활성화를 위한 지원 업무를 수행하는 한국학술진흥재단과 과학재단은 연구책임자의 논문 지도 실적을 평가하기 위한 데이터필드를 사용하고 있었고, 반면에 과학기술정책연구원, 정보통신연구관리단, 산업기술정책연구원, 보건과학기술연구기획평가단처럼 응용기술 중심의 산·학·연 지원 업무를 위해 관리하는 데이터 요소에는 연구수행기관, 공동 또는 참여기관 상세정보 기술을 위한 필드가 포함되어 있음을 알 수 있었다. 기구축된 연구과제 데이터베이스의 또하나의 특징은 CERIF에서 효율적인 검색을 위한 필드로 정확한 적용을 권장하고 있는 키워드, 시소러스, 분류번호와 같은 데이터요소가 국내 기구축 데이터베이스에서는 사용되고 있지 않다는 점으로, 데이터베이스 구축 목적 자체가 앞에서 언급한 바대로 업무의 전산화일 뿐, 학술연구자를 위한 검색 서비스라고 보기는 어렵다는 것을 알 수 있다.

### 3.2 CERIF를 준용한 국내연구현황정보 데이터 포맷(KRIF) 제안

통합된 국내 기구축 연구과제 데이터베이스 기술 요소를 근거로 하여, CERIF를 준용한 국내연구현황정보 입력 포맷(KRIF)을 다음과 같

1) Common European Research Information Format(CERIF)은 유럽국가 연합 회원국의 이종의 연구과제정보 레코드 교환을 위한 표준 포맷으로 Essential Data Element 및 Optional Data Element로 구성되어 있다. Essential Data Element에는 제목, 초록, 주소, 연구책임자, 공동연구자, 연구시작시점 및 종료시점, 키워드, 분류번호, 연구지원비 등이 정의되어 있고, Optional Data Element에는 연구협력기관, 잠정적인 결론, 기자재정보 등에 대한 필드를 정의해 놓아 작성 기관에서 재량껏 적용할 수 있도록 구성되어 있다.

이 정의한다. CERIF에서 기정의된 데이터 요소는 CF라고 표기하였으며, 국내연구현황정보에 적합하도록 확장 정의한 요소는 KF라고 표기하였다. KRIF에서는 유사한 데이터요소를 하나의 엔티티로 그룹화하였는데 이를 위해 CERIF에 기정의된 데이터 필드의 번호를 바꾸어 적용한 경우가 있다. 그러한 데이터 요소는 CFX라고 표기하였다.

- Field 00 : 레코드작성기관(CF)
- Field 01 : 최근갱신일(CF)
- Field 02 : 레코드 입력 수준(CF)
- Field 11 : 레코드제어번호(CF)
- Field 12 : 과제번호(CF)
- Field 13 : 연구진행상황(KF)
- Field 21-2 : 과제명
  - Subfield 21.1 : 과제명 기술 언어(CF)  
(데이터의 국제 교환시 자동 셋팅)
  - Subfield 21.2 : 대표 과제명(국문)(CF)
  - Subfield 21.3 : 기타 과제명(국문)(KF)
  - Subfield 22.1 : 과제명(기타어) 기술 언어(데이터의 국제 교환시 자동 셋팅)(CF)
  - Subfield 22.2 : 대표 과제명(기타어)(CF)
  - Subfield 22.3 : 기타 과제명(기타어)(KF)
- Field 23-4 : 초록
  - Subfield 23.1 : 초록문 기술 언어(CF)  
(데이터의 국제 교환시 자동 셋팅)
  - Subfield 23.2 : 초록(국문)(CF)
  - Subfield 24.1 : 초록문(기타어) 기술 언어(CF)  
(데이터의 국제 교환시 자동 셋팅)
  - Subfield 24.2 : 초록(기타어)(CF)
- Field 25 : 연구사업명(CFX 91)
- Field 26-29 : 로컬정의(KF)
- Field 31 : 연구책임자
  - Subfield 31.1 : 연구책임자명(국문)(CF)
  - Subfield 31.2 : 연구책임자명(기타어)(KF)
  - Subfield 31.3 : 주소(KF)
  - Subfield 31.4 : 전화번호(KF)
  - Subfield 31.5 : 팩스(KF)
  - Subfield 31.6 : 전자메일주소(KF)
  - Subfield 31.7 : 홈페이지 URL(KF)
  - Subfield 31.8 : 소속기관명(KF)
  - Subfield 31.9 : 로컬정의(KF)
- Field 32 : 주요공동연구자
  - Subfield 32.1 : 주요공동연구자명(국문)(CF)
  - Subfield 32.2 : 주요공동연구자명(기타어)(KF)

- Subfield 32.3-32.9 : 31.3-31.9와 동일하게 정의(KF)
- Field 33 : 참여연구원수 (CFX 82)
- Field 34 : 보조연구원수(KF)
- Field 35-39 : 로컬정의(KF)
- Field 40 : 연구시작시점(총연구기간 중)(CF)
- Field 41 : 연구종료시점(총연구기간 중)(CF)
- Field 42-49 : 로컬정의(KF)
- Field 50 : 수행연구기관정보
  - Subfield 50.1 : 수행연구기관명(CF)
  - Subfield 50.2 : 수행연구기관명(영문약어)(CF)
  - Subfield 50.3 : 주소(CFX 51.1)
  - Subfield 50.4 : 전화번호(CFX 60.1)
  - Subfield 50.5 : 팩스(CFX 60.2)
  - Subfield 50.6 : 전자메일주소(CFX 60.4)
  - Subfield 50.7 : 홈페이지 URL(KF)
- Field 51-59 : 로컬정의(KF)
- Field 60 : 참여기관정보
  - Subfield 60.1 : 수행연구기관명(CFX 92)
  - Subfield 60.2 : 수행연구기관명(영문약어)(KF)
  - Subfield 60.3 : 주소(KF)
  - Subfield 60.4 : 전화번호(KF)
  - Subfield 60.5 : 팩스(KF)
  - Subfield 60.6 : 전자메일주소(KF)
  - Subfield 60.7 : 홈페이지 URL(KF)
- Field 61-69 : 로컬정의(KF)
- Field 70 : 키워드
  - Subfield 70.1 : 자유기술키워드(CF)
  - Subfield 70.2 : 자유기술키워드 기술언어(CF)  
(데이터의 국제 교환시 자음 셋팅)
- Field 74 : 로컬정의(CF)
- Field 80 : 연구지원기관
  - Subfield 80.1 : 연구지원기관명(CF)
  - Subfield 80.2 : 연구지원기관코드(CF)
- Field 81 : 총연구비
  - Subfield 81.1 : 전체총연구비(CF)
  - Subfield 81.2 : 당해연도총연구비(KF)
- Field 82-89 로컬정의(KF)

위에서 제안한 데이터 포맷은 CERIF와의 호환 가능성을 고려하여 CERIF에 정의된 데이터 요소는 최대한 동일한 필드를 사용할 수 있도록 정의하였다. 그러나 국내에 기구축된 연구과제정보 기술 항목의 특성을 고려하여 CERIF에 확장 적용이 필요한 부분은 서브 필드를 두어 활용할 수 있도록 하였으며 각각의 엔티별로 로컬 태그를 두어 재량껏 정의할 수 있도록

설계하였다. 국내 기구축 연구과제 데이터베이스가 상이한 스키마를 가진 관계형 데이터베이스 형태로 구축되어 있으므로 교환을 위해 RDB에서 데이터를 반출할 때, 로컬이 정의한 태그 및 서브필드를 제외한 데이터 요소만을 쉽게 표준 포맷으로 변환할 수 있도록 설계하였다.

### 3.3 국내 통합연구현황정보데이터베이스 구축 및 서비스 시스템 모형

본고에서는 이형의 데이터를 표준화시키고 표준화된 데이터를 물리적으로 통합하는 중앙 집중식 통합연구현황정보시스템 모형 설정에 초점을 맞추고 있다. 통합연구현황정보데이터베이스를 구축하기 위해서 국내 기구축 DB가 앞절에서 정의한 표준 형태로 구축 또는 변환되었다고 가정하고 국내 표준 입력 형태로 구축 또는 변환된 데이터중 반출(Export)할 데이터 항목을 선정한다. 학술정보서비스를 위한 목적으로 구축되는 통합연구현황정보 데이터베이스는 일반에게 공개 가능하고 학술 목적으로 활용될 수 있는 데이터 필드만을 선정할 필요가 있으므로 표. 2와 같은 항목을 통합형태 데이터베이스 구축을 위한 데이터 요소로 한정하였다.

그림. 1은 ERGO Pilot Project<sup>2)</sup>에서 구현하고 있는 Catalog model을 부분적으로 적용하여 제안한 연구현황정보서비스시스템 모형이다. ERGO Pilot Project의 Catalog 모델은 검색에 필요한 핵심 정보를 분산 구축된 로컬의 CERIF에서 추출하여 통합한다. 로컬기관이 지속적으로 데이터를 송부함으로써 데이터베이스

2) The European Commission Innovation Programme에서 1996년에 발족한 ERGO Working Group이 단일 창구로 다수 국가의 과학기술정보원에 대한 검색을 가능하게 하기 위한 Pilot Project다. 현재 9개 국가의 12개 R&D 데이터베이스가 통합되었고 20,000여 연구과제 프로젝트 레코드가 로드된 상태다. 데이터 포맷으로 CERIF를 적용하고 있으며 중앙의 데이터베이스는 검색을 위한 핵심 필드로 단순화시키고 상세 정보가 요청될 경우 데이터베이스 제작기관으로 Forward 시키는 Catalogue 모델 적용이 고려되고 있다.

항목	세부항목	태그	반복	필수
관리정보	레코드 작성 기관	00	불가	Y
	최근 레코드 갱신일	01	불가	Y
	입력 수준	02	불가	Y
제어번호	레코드제어번호	11	불가	Y
	과제번호	12	불가	Y
	연구진행상황	13	불가	Y
연구과제 일반정보	대표과제명	21.2	불가	Y
	대표과제명(기타어)	22.2	불가	Y
	기타과제명	21.3	가능	N
	기타과제명(기타어)	22.3	가능	N
	초록	23.2	불가	N
	초록(기타어)	24.2	가능	N
	연구사업명	25	가능	Y
연구자정보	연구책임자명	31.1	불가	Y
	연구책임자명(기타어)	31.2	불가	Y
	연구책임자 주소	31.3	불가	N
	연구책임자 전화	31.4	불가	N
	연구책임자 팩스	31.5	불가	N
	연구책임자 이메일	31.6	불가	N
	연구책임자 홈페이지 URL	31.7	불가	N
	연구책임자 소속 기관명	31.8	불가	N
	주요공동연구자명	32.1	가능	N
	주요공동연구자명(기타어)	32.2	가능	N
	참여연구원수	33	불가	N
연구보조원수	34	불가	N	
연구기간	연구시작시점	40	불가	Y
	연구종료시점	41	불가	Y
수행연구 기관	수행연구기관명	50.1	가능	N
	수행연구기관명(이형)	50.2	가능	N
	수행연구기관주소	50.3	불가	N
	수행연구기관전화	50.4	불가	N
	수행연구기관팩스	50.5	불가	N
	수행연구기관이메일주소	50.6	불가	N
	수행연구기관홈페이지 URL	50.7	불가	N
공동· 참여기관	공동· 참여기관명	60.1	가능	N

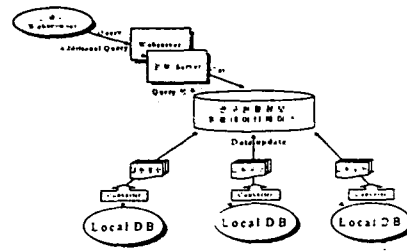
· 언어표시가 없는 항목은 국문으로 표기한다.

표 2. 통합연구현황정보데이터베이스 구축 요소

가 갱신된다. 로컬기관에서는 자관의 데이터베이스에서 표준 교환품으로 데이터를 반출시킬 수 있는 별도의 어플리케이션을 개발하여야 하며 주기적으로 교환품으로 반출받아 중앙식 통합데이터베이스에 송부하여야 한다.

그림 1. 연구현황정보서비스시스템 운영 모델

이용자는 통합데이터베이스를 검색하여 정보를 이용하고 좀더 상세한 정보를 필요로 할 경우 해당 로컬기관에 별도로 요청을 한다. ERGO



서비스는 범 유럽 국가 대상 서비스이므로 상세정보의 요청은 Nation/Regional node를 통하여 접수되고 처리된다.

그림 2.에서는 이용자가 통합형 데이터베이스 검색 결과의 상세 정보를 원하는 경우, 중앙의 검색 시스템에서 로컬기관의 데이터베이스로 해당 검색 결과의 상세 정보를 가져올 수 있는 질의를 보낼 수 있도록 설계하였다. 다시 말해 중앙의 검색 시스템이 각각의 데이터베이스에서 상세정보를 가져올 수 있는 질의문을 모두 내장하고 있어 특정 데이터베이스가 선택되면 해당 데이터베이스에 적합한 질의문이 선택되어 보내질 수 있도록 하였다. 연구현황정보검색에 관한 표준 프로토콜이 개발되거나, 또는 서지데이터 검색 표준 프로토콜인 Z39.50을 확장하여 적용한다면 굳이 데이터를 물리적으로 통합할 필요도 없고 복수의 질의문을 내장하고 있을 필요도 없겠으나 여하튼 검색을 분산시스템에 모두 의존하는 것은 대상 데이터베이스가 복수개로 증가할 경우 늦은 검색 속도 및 응답 시간으로 검색시스템의 안정성이 떨어진다는 문제가 있을 수 있다.

#### 4. 결론 및 제언

이형으로 분산 구축되어 있는 국내의 연구정보데이터가 표준 형태로 구축 또는 변환 가능하다면, 단일창구로 모든 연구정보를 효율적으로 검색 할 수 있고, 또 나아가 국제적인 연구정보의 교환을 용이하게 할 수 있다. 국가 단위의 통합데이터베이스를 구축하기 위해서는 먼저 분산 구축되어있는 이형의 데이터베이스를 변환할 수 있는 어플리케이션이 개발되어야 하며 각 연구지원기관의 실무자들이 통일된 데

이타 기술 방식에 대한 지속적인 논의와 작성 그리고 국내에서는 아직 적용된 바 없는 표준 분류법 및 시소러스 개발 노력도 지체되어서는 안될 것이다.

많은 연구 정보가 인쇄물 형태에서 네트워크를 통한 인터넷 자원으로 그 형태가 이전됨에 따라 자원을 효율적으로 관리하고 검색하기 위해서는 새로운 기술요소의 범위와 조직 기법, 또한 접근방안을 모색해야 한다. CERIF도 메타데이터 기반의 자원 기술 구조로 개정 방안을 모색 중에 있다. CERIF를 준용하여 작성된 KRIF는 새로운 데이터요소를 추가하거나 기존의 형식을 변경해야 하는 경우 그 구조의 경직성으로 새로운 유형의 자료나 내용을 표현하는데 한계가 있으며 데이터 요소간의 연관관계나 구조적 특성이 명확하게 표현되지 않는다. 또한 연구결과로 산출되는 연구 보고서의 서지사항, 또는 관련 연구정보를 수용하지 못한다는 단점이 있다. RDF가 메타데이터를 위한 공식적인 구문 모형이므로 KRIF가 RDF 데이터 모형을 바탕으로 재설계 된다면 상이한 다른 메타데이터 간에도 상호 운용성을 지원하고 유연성있게 통합 가능한 구조를 제공할 수 있을 것이다.

#### 참 고 문 헌

- 안중호. 1997. 학술연구지원의 효율화를 위한 제도 비교 연구. 한국학술진흥재단
- 한려연, 조재인, 이지원, 오정선, 김성희. 1998. 12. 하이브리드 방식 국가종합목록 구축에 관한 연구. 정보관리학회지
- European Commission DG XIII/D2-Community Research & Development. European Research Gateways On-Line. <http://www.cordis.lu/ergo/home.html>
- European Union. 1991. "Concerning the harmonization within Community of research and technical development database" in Official Journal of the European Communities", OJ L 198, The 14th.
- Finch, Peter. 1998. "European Research Gateways On-line: ERGO pilot" in proceedings of the CRIS'98 Conference.
- Irmgard, Lankenau. 1998. "Quality of Research Information on the net : from the point of users" in proceedings of the CRIS'98 Conference.
- Woensel, L. V. 1998. "Common European Research Information Format (CERIF) Revision : State of the ongoing CERIF revision activities." in proceedings of the CRIS'98 Conference. <http://www.cordis.lu/cris98>

- Hubbard, Bethan. "Code of Good Practice for Current Research Information System", <http://www.cordis.lu/cris98>
- Lavieter, L. D., "DRIS, Directory of Research Information Systems : Inventory and analysis of available research information systems " <http://www.cordis.lu/cris98>.