

# 국가 R&D 정보 공동활용체제 구축 방안에 대한 연구

A Study on a Method to Construct a System for Sharing National R&D Information

김명일, 전성진, 류범중  
한국과학기술정보연구원

Kim Myung-Il, Jhun Sung-Jin, You Beom-Jong  
Korea Institute of Science and Technology Information

## 요약

국가연구개발사업은 각 부처 산하의 연구관리전문기관을 통해 관리·운영되고 있으며, 각 사업을 통해 생산되는 국가 R&D 관련 정보는 각 기관의 특성에 따라 구조적·의미적으로 서로 다르게 설계 및 구축되고 있다. 이러한 이유로 인해 정보 간 상호운용을 통한 국가 차원의 종합적인 수집·분석·활용이 불가능하며, 인력·성과 등의 R&D 기반정보가 중복 저장·운용되는 문제점이 발생한다. 이러한 문제점을 해결하고, 국가연구개발사업의 투명성 및 효율성 제고를 위해 과학기술혁신본부에서는 2006년 1월부터 국가 R&D 정보의 공동활용을 기반으로 하는 국가과학기술종합정보시스템(NTIS, National Science and Technology Information System)을 구축하고 있다. 본 논문에서는 국가 R&D 정보를 관리하고 있는 각 기관의 자치성 침해를 최소화하면서, 공동활용의 효과를 극대화 할 수 있는 국가 R&D 정보의 공동활용체제 구축 방안을 제시한다.

## Abstract

National R&D programs are managed and operated by institutes under different ministries. The R&D information created through these programs are designed and constructed heterogeneously both semantically and structurally. Therefore, integrated collection, analysis and application of the information through the interoperability of information at the national level becomes nearly impossible, and redundant production and operation of R&D information becomes a general problem. The Office of Science and Technology Innovation has started the construction of the National Science and Technology Information System(NTIS), which is based on the interoperability and sharing of R&D information, to resolve these issues. In this paper, we propose a method to construct a system for sharing national R&D information held by different ministries while minimizing the effect on the autonomy of the institutes involved.

## I. 서론

### 1. 국가 R&D 정보 공동활용의 필요성

정부는 과학기술기본법 제 11조의 규정에 의한 국가연구개발사업의 기획·평가 및 관리에 관하여 필요한 사항을 규정하기 위해 「국가연구개발사업의 관리등에 관한 규정(2005년 3월, 대통령령 제 18731호에 의거, 이하 공동관리규정)」을 개정하였으며, NTIS의 정보공동활용체제는 이 규정을 근거로 개발되고 있다. 또한, 수요자 중심의 실효성 있는 시스템을 구축하기 위해 NTIS 사용자 요구사항 조사를 실시하였으며, 다음과 같은 문제점들이 도출되었다.

- 유사하거나 동일한 과제를 서로 다른 연구관리전문기관을 통해 수행하는 중복과제가 발생함
- 고가의 장비·기자재를 중복 구매하는 문제가 발생함
- 인력정보와 같은 기본정보를 서로 다른 연구관리전문기관에 중복해서 입력해야 하는 불편이 있음

- 동일한 성과를 서로 다른 연구관리전문기관에 중복 제출함

이러한 문제는 연구관리전문기관들의 정보를 공동으로 활용하지 못해 발생하는 것으로, 국가 R&D 정보의 단일 뷰(unified view)를 개발하고, 이를 기반으로 정보를 연계통합하는 국가 R&D 정보의 공동활용체제 구축을 통해 해결할 수 있다.

### 2. 국가 R&D 정보 공동활용에 대한 장벽

국가 R&D 정보의 공동활용체제 구축을 어렵게 하는 장벽으로 정보자원의 이질성, 정보 간 연계성 부족, 정보수집의 즉시성 미비 등과 같은 것이 있다.

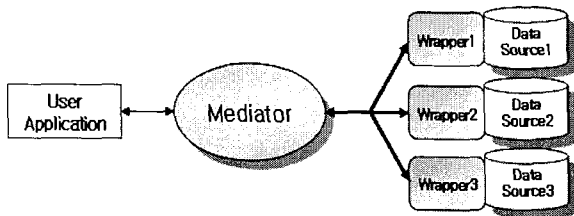
#### 2.1 정보자원의 이질성

서로 다른 목적과 방법에 의해 설계된 정보자원은 다음과 같은 3가지 관점에서의 이질성을 내포하고 있다[1].

- 구조적(structural) 이질성 : 동일한 정보 대상에 대해 서로 다른 구조 레벨로 표현하는 것으로, 예를 들어 R&D 참여인력의 학력정보를 엔티티(entity), 속성(attribute), 릴레이션십(relationship)과 같은 레벨로 서로 다르게 설계하는 것을 의미함
- 구문적(syntactical) 이질성 : 동일한 의미를 가지는 정보 항목에 대해 서로 다른 명칭을 부여하는 것으로, 예를 들어 과제정보에 부여하는 고유번호에 대해 과제번호, 과제ID, 과제코드 등과 같은 서로 다른 명칭을 부여하는 것을 의미함
- 의미적(semantic) 이질성 : 동일한 명칭의 정보 항목에 대해 서로 다른 의미를 부여하는 것으로, 예를 들어 '업적'이라는 명칭의 정보항목에 대해 논문 또는 특허, 과제수행경력, 위원회 및 학회활동 등으로 서로 다른 의미를 부여하는 것을 의미함

정보의 이질성은 정보 간 상호운용을 어렵게 하는 요소이며, 일반적으로 다음과 같은 해결 방안이 있다[2].

- 표준화 : DB 스키마에 대한 단일 표준을 정립하고, 이를 기반으로 모든 정보를 통합하는 것으로, 이상적인 방법이지만 비용과 시간이 많이 소요되는 단점이 있음
- 미디어이터(mediator) : 사용자(어플리케이션) 계층과 데이터 소스 계층 사이에 설치되는 조정기로서, 이질적인 데이터 소스들 간의 데이터 항목 매핑(mapping) 및 변환 기능을 수행함
- 래퍼(wrapper) : 그림 1과 같이 데이터 소스와 미디어이터 사이에 설치되는 일종의 질의 실행기로서, 데이터 소스의 수가 많을 경우 미디어이터가 처리해야 하는 정보가 많아져 효율성이 떨어지는 문제를 해결할 수 있음

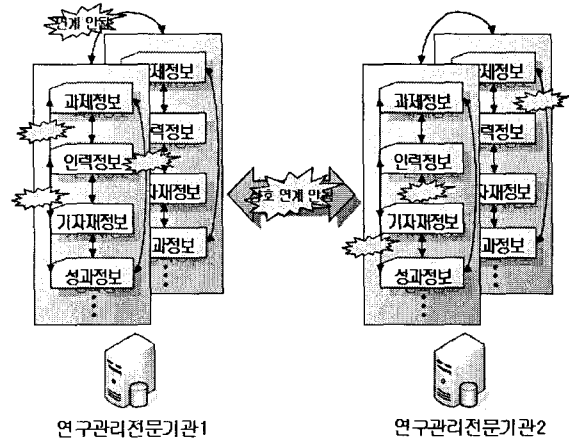


▶▶ 그림 1. 정보간 이질성 해결을 위한 미디어이터-래퍼 구조

## 2.2 정보자원 간 연계성 부족

연구관리전문기관에서 관리하고 있는 국가 R&D 기반정보는 과제, 인력, 기자재, 성과 등이 있으며, 이러한 정보는 독립적으로 존재하는 것이 아니라 서로 밀접한 연관성을 가지고 있

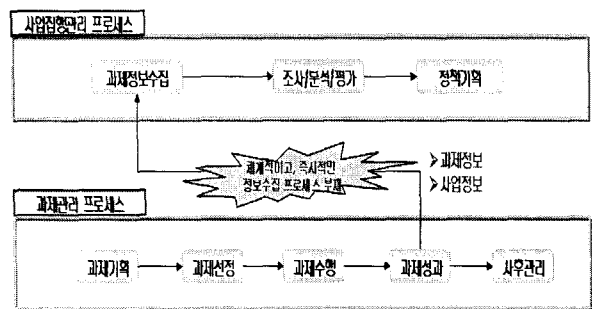
다. 그러나 그림 2와 같이 하나의 연구관리전문기관에서 관리·운영하는 기반정보 사이에도 연계성 설정이 미약한 실정이며, 연구관리전문기관 간의 기반정보 연계성은 전무하다고 할 수 있다. 이러한 정보간 연계성 문제를 해결하기 위해서는 각 정보에 유일성을 보장할 수 있는 글로벌ID를 부여하고, ID와 ID 사이의 관계를 설정해야 한다.



▶▶ 그림 2. 국가 R&D 기반정보의 연계성 부족

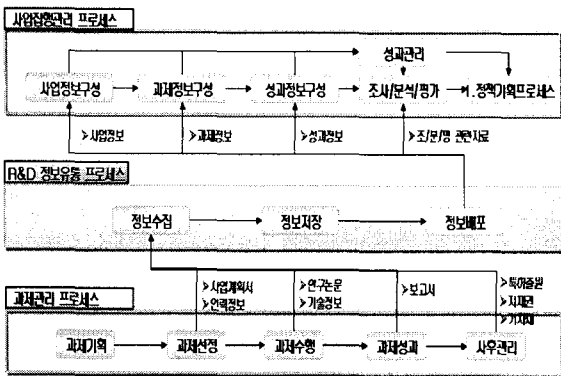
## 2.3 정보 수집의 즉시성 미비

국가 R&D 정보의 효율적인 유통을 위해서는 체계적이고 즉시성을 갖춘 정보유통 프로세스의 정립이 필요하다. 즉, 연구관리전문기관의 과제관리 프로세스와 종합적인 정보유통을 위한 프로세스가 서로 연계되어야 한다. 현재는 그림 3과 같이 국가연구개발사업의 조사·분석·평가 기능을 통해 연간 1회의 정보수집이 이루어지고 있는 실정이다.



▶▶ 그림 3. 체계적이고 즉시성을 갖춘 정보유통 프로세스의 부재

따라서 그림 4와 같이 과제관리 프로세스의 세부 단계별로 수집 가능하며 필요한 정보가 생성되는 시점에 즉시 수집·연계하여 활용할 수 있는 정보유통 프로세스가 요구된다.

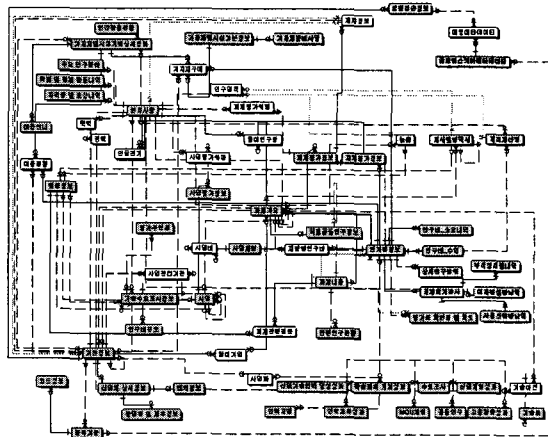


▶▶ 그림 4. 국가 R&D 정보유통 프로세스

II. 본론

1. 공동활용정보의 정의

공동활용정보는 정보의 소유에 관계없이, 응용서비스를 통해 사용자 간에 공동으로 활용될 수 있는 정보를 의미하며, 사업·과제·인력·기자재 등의 정보가 이에 해당한다. 공동활용정보의 범위는 고정되어 있지 않으며, 새로운 응용서비스의 데이터 수요에 따라 확장 가능하다.



▶▶ 그림 5. NTIS 글로벌스키마의 구조

공동활용정보는 글로벌스키마(global schema)와 데이터사전(data dictionary)로 서술될 수 있다. 글로벌스키마는 엔티티, 속성, 릴레이션쉽으로 표현되는 국가 R&D 정보의 가상 단일 뷰이며, 데이터사전은 엔티티, 속성, 릴레이션쉽에 대한 상세한 설명서이다. 현재 NTIS의 글로벌스키마는 다음과 같은 4가지 요소를 근거로 개발되었으며, 계속해서 수정·보완 작업이 진행 중이다. NTIS 글로벌스키마의 구성은 그림 5와 같다.

- 응용서비스의 데이터소요
- 국가연구개발사업의 관리 등에 관한 규정

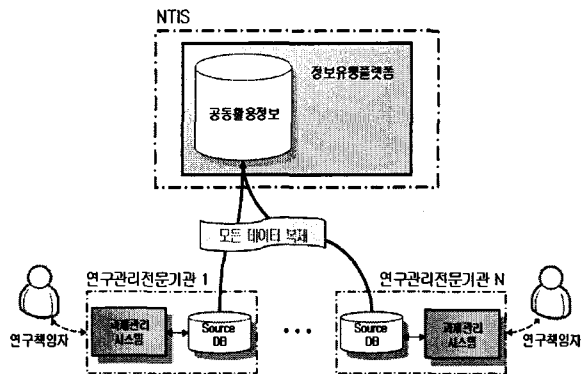
- 국가연구개발사업 자체평가 지침
- 연구관리전문기관에서 운용중인 DB 스키마

2. 정보공유기법

국가 R&D 정보의 효율적인 공유를 위해 정보의 복제 및 연계 기반으로 하는 다음의 3가지 방안을 제시한다.

2.1 물리적통합

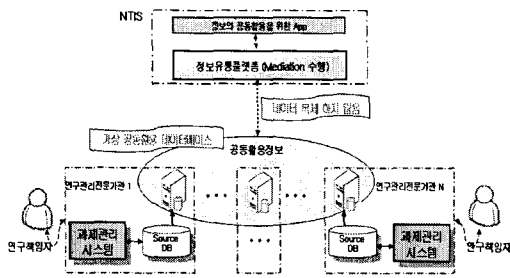
연구관리전문기관에 분산·관리되고 있는 모든 정보를 NTIS 시스템으로 복제하여 정보를 통합하는 것으로, 시스템 구성은 그림 6과 같다. 이 방식은 일종의 웨어하우스(warehousing) 방식으로 비교적 간단하게 시스템을 개발할 수 있는 장점이 있으나, 데이터소스를 소유하고 있는 기관의 자치성을 침해할 수 있는 문제점이 있다. 또한 복제된 데이터와 소스데이터 사이의 동기화를 위한 별도의 기능을 개발해야 하는 단점이 있다.



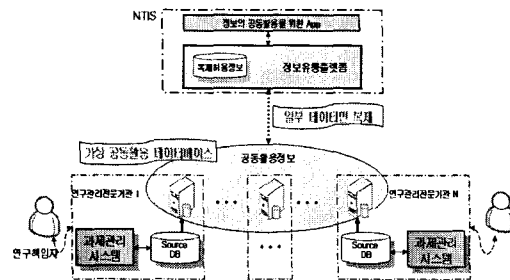
▶▶ 그림 6. 물리적통합에 의한 정보공유

2.2 논리적통합

연구관리전문기관의 소스데이터를 NTIS 시스템 내로 복제하지 않고, 그림 7과 같이 별도의 논리적 정보통합 기술을 사용하여 정보를 연계하는 방식이다. 즉, 가상의 공동활용 데이터베이스를 구성하는 것으로, 일반적인 논리적 정보통합 기술로 미디어이터-래퍼 방식이 있다[3][4]. 미디어이터는 응용서비스의 데이터 요청질을 적절히 분석·재작성·전송하는 기능을 수행하며, 적절한 결과를 얻기 위해 소스데이터 대상의 질의실행 기능을 수행한다. 이 방식은 정보 소유기관에 대한 자치성 침해가 없는 장점이 있으나, 원격 데이터에 대한 질의 실행을 전제하므로 응답시간이 길어질 수 있는 단점이 있다. 또한, 글로벌스키마와 로컬스키마 사이의 호환을 위한 매핑테이블과 별도의 관리도구 등이 필요하다.



▶▶ 그림 7. 논리적통합에 의한 정보공유



▶▶ 그림 8. 혼합방식에 의한 정보공유

2.3 혼합(hybrid) 방식의 통합

물리적통합과 논리적통합 방식의 혼합 형태로, 그림 8과 같이 가상의 공동활용 데이터베이스를 구축하고, 이 중에서 일부 필수 데이터를 NTIS 시스템 내로 복제하는 방식이다. 복제 데이터는 시스템 성능, 통합관리의 필요성, 정보의 특성 등에 따라 선택될 수 있다. 이 방식은 정보 소유 기관의 자치성 침해를 최소화 할 수 있고, 논리적 통합 방식의 성능저하 문제를 어느 정도 해결할 수 있는 장점이 있으나, 매핑테이블과 별도의 관리도구는 필요하다. NTIS는 정보공동활용을 위해 이 방식을 채택한다.

III. 결론

국가 R&D 정보의 공동활용체제 구축을 통한 NTIS의 개발은 연구개발사업의 효율성과 투명성을 제고시켜 국가 경쟁력 향상에 기여하는 방향으로 구축되고 있다. 정보공동활용체제의 효과적인 구축을 위해서는 정보소유기관의 이해관계를 떠나 국가 차원의 필요성을 반영해야 한다. 또한, 누락정보의 보완·데이터 오류수정 등을 통한 데이터 정제와 품질관리 프로세스의 개발·성과지표 정의 등을 통한 정보의 고품질화 방안을 개발해야 한다.

■ 참고 문헌 ■

[1] 정의석 "상호운용성 확보를 위한 지식정보 표준화 동향", IITA 주간기술동향 제 1230호, 2006.

[2] Amit P. Sheth., "Changing Focus on Interoperability in Information Systems : From System, Syntax, Sturcture to Semantics," KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE Interoperating geographic Information Systems, 1999.  
 [3] Garcia-Molina, et. al, "The TSIMMIS Approach to Mediation : Data Models and Languages," Journal of Intelligent Information Systems, Vol.8, No.2, pp.117-132, 1997.  
 [4] J. A. R. Castillo, et. al, "Information Extraction and Integration from Heterogeneous, Distributed, Autonomous Information Sources ," IEEE International Conference Reuse and Integration, pp.183-191, 2003.