

# NTIS 데이터 품질관리 체계와 포털 사이트 구축

이병희\*, 정옥남\*\*, 최희석\*, 임철수\*, 김재수\*

\*한국과학기술정보연구원 NTIS사업단

\*\*교육과학기술부 과학기술정보과

e-mail: {bhlee, choihs, cslim, jaesoo}@kisti.re.kr, \*\*jon77@mest.go.kr

## A Construction of Management System and Its Portal Site for NTIS Data Quality Management

Byeong-Hee Lee\*, Ock-Nam Jung\*\*, Heeseok Choi\*, ChulSu Lim Jaesoo Kim\*

\*NTIS Division, Korea Institute of Science and Technology Information

\*\*Ministry of Education, Science and Technology

### 요 약

데이터가 기업 및 기관 활동의 중요한 자산으로 인식이 높아지고 있지만 저 품질 데이터로 인한 막대한 손실과 비용의 증가는 큰 문제가 되고 있다. 본 논문에서는 교육과학기술부와 KISTI에서 2007년부터 범부처 차원에서 수행중인 국가R&D 정보자원의 데이터 품질관리를 위해 각 부처와 협의하고 수행해 온 국가R&D 데이터 품질관리 체계 수립과 포털 사이트 구축에 관하여 알아본다. 범부처 국가R&D정보 자원의 데이터 품질관리체계 수립과 범부처 적용 지침 및 가이드라인 제시를 위해 NTIS사업단 및 15개 부처(16개 대표전문기관)의 실무팀장 및 DB 관리자 중심으로 총 33명의 설문과 실시하여 품질관리체계 현황을 조사·분석하였다. 또한 부처(기관)별로 국가R&D표준정보 데이터품질 지표 마련과 주기별 데이터 품질 및 개선도 자체점검을 지원하기 위해, 데이터 점검기준과 절차를 마련하고 이를 기반으로 부처(기관)와 협력하여 데이터품질 점검을 기반을 마련하였다. 이렇게 마련된 품질관리체계와 프로세스를 지원하기 위한 자동화 솔루션을 운영하고자 본 논문에서는 NTIS 데이터 품질관리 체계 및 프로세스, 기능이 통합된 웹 포털 구축에 대해서도 알아본다.

### 1. 서론

근래 들어 데이터가 기업 및 기관 활동의 중요한 자산으로 인식이 높아지고 있지만 저품질의 데이터로 인한 막대한 손실과 비용의 낭비는 줄어들지 않고 있다. 저품질 데이터 증가로 인해 관리 관점별 분석 요구의 증가로 정보 연계성이 복잡해지며 데이터 품질저하가 우려되고 데이터 분산 및 중복에 따른 저품질의 원천 데이터가 전체 시스템의 품질을 저하시킬 우려까지 존재하며 또한 시간이 지남에 따라 데이터양이 증가하면서 저품질 데이터에 대한 관리비용도 증가하고 있다[1,2,3].

이러한 문제점을 해결하고 고품질의 데이터를 확보하기 위해서는 데이터 품질관리 활동을 반복적으로 지속 수행하여야 하며 데이터 품질 문제는 복잡한 원인에서 비롯되므로 체계적이고 지속적인 관리 방안을 마련할 필요가 있다. 2009년 초 현재 우리나라의 국가 R&D 정보자원은 15개 부처·청 산하의 연구관리전문기관(예를 들어, 한국산업기술평가원, 한국학술진흥재단, 한국과학재단 등)을 통해 관리·운영되고 있으며, 각 사업을 통해 생산되는 국가 R&D정보는 각 기관의 필요에 따라 데이터의 구조와 형식이 다른 실정이다.

본 논문에서는 교육과학기술부와 한국과학기술정보연구원(KISTI)에서 2007년부터 범부처 차원에서 수행중인 국가R&D 정보자원의 데이터 품질관리를 위해 15개 부처

(16개 대표전문기관)의 실무팀장과 DB관리자를 중심으로 설문과 협의를 거쳐 수행해 온 국가R&D 데이터의 품질관리 체계 수립하여 부처(기관)별로 국가R&D표준정보 데이터 점검기준과 절차를 마련하고 이를 기반으로 데이터품질점검을 실시를 추진한 과정에 대하여 기술하고 NTIS 데이터 품질관리 체계 확보의 자동화 솔루션 운영을 위해 웹 포털 사이트를 구축한 것에 대하여 알아보하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 국가R&D정보 자원 데이터 관리 및 표준화 현황을 살펴보고, 3장에서는 NTIS 데이터 품질관리 체계 구축을 알아보고, 4장에서는 NTIS 데이터 품질관리 포털 구축에 대하여 알아보고, 마지막으로 5장에서는 결론과 향후 방향에 대하여 논하도록 한다.

### 2. 국가R&D정보자원 데이터 관리 및 표준화 현황

NTIS가 출범되기 전인 2006년까지 우리나라의 국가 R&D 정보자원은 15개 부처·청 산하의 대표전문기관(예를 들어, 한국산업기술평가원, 한국학술진흥재단, 한국과학재단 등)을 통해 관리·운영되고 있었으며, 각 사업을 통해 생산되는 국가 R&D정보는 각 기관의 필요에 따라 데이터의 항목, 구조, 형식이 달랐다.

2007년 7월말 범부처 차원에서 국가R&D사업 현황 파악, 종합조정 및 공동활용을 하기 위해 NTIS 세부서비스를 통해 서비스될 항목을 중심으로 대표전문기관의 정보 DB

현황 및 정책적 수요 등을 고려하여 필요 최소한의 정보 항목 341개를 선정하였다[4]. 범부처 국가R&D정보 표준항목은 크게 기존 국가R&D사업 조사·분석을 통해 수집되는 정보와 그 외 공동활용 정보로 구성되어 있다. 이를 수집형태에 따라 분류하면 각 부처(대표전문기관)로부터 수집되는 항목 248개, 성과물전담기관 등 연계항목 61개, NTIS 자체관리 항목이 32개이다. 업무분야별로 분류하면 과제 81개, 성과 129개, 인력 57개, 장비·기자재 74개 항목이다.

341개 범부처 국가R&D정보 표준항목 중에서 코드 관련 항목은 코드표준화 작업을 통해 입력 표준매뉴얼을 작성하고, NTIS홈페이지를 통해 배포·공유할 수 있는 체계를 마련하였다.

한편, 341개 범부처 국가R&D정보 표준항목 대비 각 부처(대표전문기관)별 관리항목을 조사한 결과 <표 1>과 같이 부처(대표전문기관)별 정보화수준이 상이하여 평균 100여개 항목을 보유하고 있는 것으로 파악되었고, 무엇보다 데이터 품질수준이 상이하여 범부처 차원에서 데이터 품질 제고가 요구되었다.

<표 1> 대표기관별 관리항목 현황('07.12월 기준)

농진청	보건	건기평	환경	해양수산	학진	과학재단		농림센터	산기평	정보통신	문화콘텐츠	평균
						특정연구	기초					
74	183	121	112	118	80	138	139	89	74	112	48	100

※ 대부분의 기관은 인력상세정보, 성과, 장비 등 정보 미 관리 중

이러한 요구에 따라 범부처 차원에서 데이터 품질을 제고하기 위해 주기별로 데이터품질을 점검하고 이를 지속적으로 개선·유지관리를 위한 체계수립이 필요하여 각 부처 실·국장급으로 구성된 NTIS사업추진위원회 안건으로 이러한 문제를 상정(08.5월)하고 부처 차원에서 데이터 품질제고를 위한 협력체계를 마련하였다.

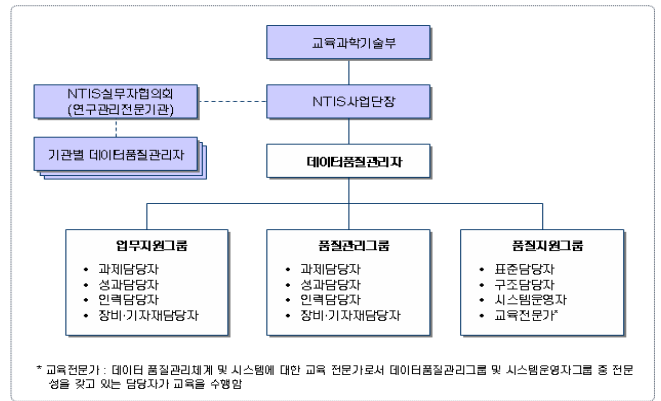
### 3. NTIS 데이터 품질관리 체계 구축

NTIS 데이터품질 제고를 위한 본격적인 작업은 대표전문기관별 실무팀장으로 구성된 NTIS실무자 협의회를 통해 이루어졌다[5,6,7]. 우선 범부처 차원 데이터 품질관리체계수립을 위해서는 부처(대표전문기관)별 데이터품질관리체계 현황조사와 품질 자체점검결과가 필요하였다. 부처(대표전문기관)별 데이터품질관리체계 현황을 조사하기 위해 데이터품질관리체계에 대한 인식수준 및 관리현황 등에 대한 설문을 실시되었다.

대표기관별 데이터 품질관리체계 수립을 위한 설문에서는 대표기관별 데이터 품질관리체계에 대한 인식수준 및 관리현황 파악, 데이터 품질관리체계에 대한 향후 방향성 정립 의견 수렴, 데이터 품질점검 기준에 대한 의견 수렴이 조사되었다. NTIS사업단 및 15개 부처(16개 대표전문기

관)의 관련자 33명을 대상으로 설문을 실시하여 대표전문기관별로 품질관리체계 현황을 조사·분석한 결과는 다음과 같았다. 부처(기관)별 데이터품질관리 성숙도(DMM: Data Management Maturity Model) 단계에서 NTIS사업단은 전사적 표준에 의해 데이터를 관리하는 DMM Level 3단계 진입 가능한 시점이고 16개 대표전문기관은 DMM Level 0단계(데이터모델에 대한 문서화가 안 된 단계)부터 Level 4단계(전사 데이터품질관리 조직체계가 있는 단계)까지 다양하게 분포하였다. 이중 8개 기관(50%)이 데이터요소 정의에 대한 문서화가 되어 있는 Level 2단계로 조사되었다. 설문을 통해 문제점 및 개선방향도 도출되었다. 공통적으로는 부처(기관)별 데이터품질관리 조직(담당자 지정) 마련요구와 데이터품질점검 및 평가체계 수립·적용이 도출되었다.

(그림 1)은 NTIS 데이터 품질관리 체반 활동을 수행할 조직체계이다.



(그림 1) NTIS 데이터 품질관리 조직체계

또한 부처(기관)별로 국가R&D표준정보 데이터품질 지표 마련과 주기별 데이터품질 및 개선도 자체점검을 지원하기 위해, 데이터 점검기준과 절차를 마련하고 이를 기반으로 부처(기관)와 협력하여 데이터품질점검을 실시하였다. 먼저 데이터 품질점검을 위한 업무규칙(BR: Business Rule)을 도출하고, NTIS사업단에서는 도출된 업무규칙을 가지고 각 대표기관에서 수집한 데이터를 가지고 데이터 품질점검을 실시하였다. 품질점검은 2008년 8월초 부처(기관)와 합의된 완전성, 유효성, 일관성, 유일성, 정확성 등 5개 품질지표를 기준으로 품질이 측정되었다. 기관별로 수집된 정보수준(항목, 건수)이 상이하여 수집된 항목의 데이터 질만을 고려하여 점검한 결과 총 4%정도의 오류율이 나와 공공기관 DB 평균오류율 5%보다 좋게 나타났다. 정보별, 품질지표별 오류율은 다음 <표 2>와 같다.

<표 2> 대표전문기관 국가R&D 정보별, 품질지표별 오류율 예시

구분	인력	성과	과제	장비	
오류율	0.96%	9.27%	10.96%	11.49%	
지표별	· 완전성	0.18%	0.00%	2.88%	12.32%
	· 유효성	1.37%	13.27%	5.78%	5.44%
	· 일관성	0.00%	0.00%	35.08%	23.93%

#### 4. NTIS 데이터 품질관리 포털 구축

범부처 차원에서 추진되고 있는 NTIS 데이터품질에 대한 효율적인 관리와 관련기관(담당)간 원활한 커뮤니케이션 채널을 확보하기 위해서 NTIS데이터품질관리 포털을 구축하였다[8,9]. NTIS 데이터 품질관리 포털 개념도는 (그림 2)와 같다.



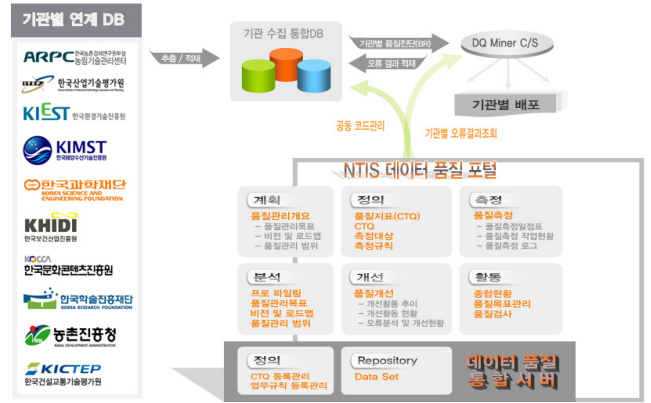
(그림 2) NTIS 데이터 품질관리 포털 개념도

NTIS 데이터 품질관리 포털의 구축은 NTIS에서 수집, 가공되는 데이터 품질진단을 수행하는 16개 기관의 분석 결과 및 분석·개선을 통하여 국민에게 서비스되는 데이터베이스 품질진단 내역을 포함하며 표준, 구조, 값의 정보에 대하여 의사소통을 하고 품질과 관련된 담당자들이 해당 내역을 공유하는 공간으로 (그림 3)은 NTIS 데이터 품질관리 프레임워크를 보여준다.



(그림 3) NTIS 데이터 품질관리 프레임워크

NTIS 데이터 품질관리 포털의 주요 기능은 NTIS 데이터 품질(값) 진단 결과의 조회 및 출력, NTIS 데이터 품질(값) 기준 신청, 승인 등의 관련 프로세스, NTIS 데이터 품질(표준) 관리 현황 조회 및 출력, NTIS 데이터 품질(표준) 용어 신청, 승인 등의 관련 프로세스이다. 이런 기능을 수행할 수 있는 NTIS 데이터 품질관리 포털 시스템 구성도는 (그림 4)와 같다.



(그림 4) NTIS 데이터 품질관리 포털 시스템 구성도

포털로 구현된 시스템은 (그림 5)와 같다. 개발 구성은 언어 JAVA 1.4, Java 개발 툴 eclipse europa, 3th part DHX, fusion Chart, JXL으로 개발되었으며, DQ 통합 Repository Oracle 10g로 데이터베이스가 구성되었고, WAS는 Apache Tomcat 4.1.30으로 개발되었다.



(그림 5) 구현된 NTIS 데이터 품질관리 포털

데이터 품질관리 포털(또는 DQMS: Data Quality Management System) 구축 성과지표는 (그림 6)와 같다.



(그림 6) NTIS-DQMS 성과지표

NTIS에 대한 데이터 품질 활동의 성과지표는 첫째, 범부처/국가 R&D정보 중심의 데이터 품질개선, 둘째, 범부처/국가R&D정보 중심의 데이터 표준화, 셋째, 데이터

품질관리 시스템의 내재화가 도출되었다. 범부처/국가 R&D정보 중심의 데이터 품질개선은 각 기관에서 입력되는 데이터의 중복입력 절감과 오류 데이터의 개선이 포함되며, 범부처/국가R&D정보 중심의 데이터 표준화는 정보 수집 및 검색의 의사소통 표준화가 발생하는 것으로 도출되었다. 데이터 품질관리 시스템 내재화는 데이터의 단일화된 단일 뷰(Single View) 마련, 데이터관리 업무 효율 향상 등이 발생하는 것으로 도출되었다.

## 5. 결론

지금까지 본 논문에서는 교육과학기술부와 KISTI에서 2007년부터 범부처 차원에서 수행중인 국가R&D 정보자원의 데이터 품질관리를 위해 각 부처와 협의하고 수행해온 국가 R&D 데이터 품질관리 체계 수립과 포털 사이트 구축에 관하여 알아보았다.

NTIS는 여러 부처(기관)와 연계하여 데이터를 수집·관리하고 있어 범 부처 차원에서 지속적인 데이터 품질관리 활동이 수행되어야 한다. NTIS 데이터 품질관리 포털은 이러한 범부처 차원에 추진하고 있는 품질관리업무를 효율적으로 지원할 뿐만 아니라 NTIS사업단 및 대표전문기관 데이터 품질담당자간 커뮤니케이션 창구로서의 역할을 담당할 것으로 기대 된다.

향후에는 NTIS사업단에서 데이터 품질관리 인증(DQMC)획득을 추진할 예정이며, 더불어 NTIS에 대한 데이터 품질 활동의 성과지표로 도출된 NTIS 데이터 품질관리 포털을 통한 범부처/국가 R&D정보 중심의 데이터 품질개선, 데이터 표준화, 품질관리 시스템의 내재화의 꾸준한 노력이 필요하다.

## 참고문헌

- [1] 한국데이터베이스진흥센터, 데이터 품질관리 성숙 모형(ver 1.0), 2006.
- [2] 이병희 외 8인, 국가R&D 데이터 품질관리체계 구축 방안, 2007년 기술동향보고서, 한국과학기술정보연구원, 2007.
- [3] 박미영, 조용한, 승현우, “데이터베이스 품질 성숙도 모델 및 핵심 수준 영역 개발”, 정보기술 논문지, 제3권, pp69-76, 2005.
- [4] 과학기술관계장관회의, “범부처 국가R&D정보 유통기반 구축방안”, 2007.8.2.
- [5] 교육과학기술부 과학기술정보과, “3. 국가R&D표준정보 데이터품질점검 및 관리체계 수립방안“, ‘08년 제1회 NTIS실무자 협의회 회의 안건, pp.10-11, 2008.
- [6] 교육과학기술부 과학기술정보과, “1. 부처(기관)별 NTIS 데이터품질관리체계 현황분석 결과“, ‘08년 제2회 NTIS실무자 협의회 회의 안건, pp.2-4, 2008.
- [7] 교육과학기술부 과학기술정보과, “1. 부처(기관)별 데이터품질점검 결과 및 기관별 평가방안“, ‘08년 제3회

- [8] NTIS실무자 협의회 회의 안건, pp.2-4, 2008. 한국과학기술정보연구원, NTIS 데이터 품질관리 체계 수립 최종보고서, 2008
- [9] 한국과학기술정보연구원, NTIS 데이터 품질 고도화 및 데이터 품질 포털 구축 최종보고서, 2008.